

# Proteomic insight into arabinogalactan utilization by particle-associated *Maribacter* sp. MAR\_2009\_72

Saskia Kalenborn<sup>1</sup>, Daniela Zühlke<sup>2</sup>, Katharina Riedel<sup>2</sup>, Rudolf I. Amann<sup>1</sup>, Jens Harder<sup>1\*</sup>

<sup>1</sup>Department of Molecular Ecology, Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany

<sup>2</sup>Department for Microbial Physiology and Molecular Biology, University of Greifswald, Felix-Hausdorff-Str. 8, D-17489 Greifswald, Germany

\*Corresponding author. Department of Molecular Ecology, Max Planck Institute for Marine Microbiology, Celsiusstr. 1, D-28359 Bremen, Germany. E-mail:

[jharder@mpi-bremen.de](mailto:jharder@mpi-bremen.de)

Editor: [Cindy Nakatsu]

## Abstract

Arabinose and galactose are major, rapidly metabolized components of marine particulate and dissolved organic matter. In this study, we observed for the first time large microbiomes for the degradation of arabinogalactan and report a detailed investigation of arabinogalactan utilization by the flavobacterium *Maribacter* sp. MAR\_2009\_72. Cellular extracts hydrolysed arabinogalactan *in vitro*. Comparative proteomic analyses of cells grown on arabinogalactan, arabinose, galactose, and glucose revealed the expression of specific proteins in the presence of arabinogalactan, mainly glycoside hydrolases (GH). Extracellular glycan hydrolysis involved five  $\alpha$ -L-arabinofuranosidases affiliating with glycoside hydrolase families 43 and 51, four unsaturated rhamnogalacturonylhydrolases (GH105) and a protein with a glycoside hydrolase family-like domain. We detected expression of three induced TonB-dependent SusC/D transporter systems, one SusC, and nine glycoside hydrolases with a predicted periplasmic location. These are affiliated with the families GH3, GH10, GH29, GH31, GH67, GH78, and GH115. The genes are located outside of and within canonical polysaccharide utilization loci classified as specific for arabinogalactan, for galactose-containing glycans, and for arabinose-containing glycans. The breadth of enzymatic functions expressed in *Maribacter* sp. MAR\_2009\_72 as response to arabinogalactan from the terrestrial plant larch suggests that *Flavobacteriia* are main catalysts of the rapid turnover of arabinogalactans in the marine environment.

**Keywords:** alpha-L-arabinofuranosidase; *Flavobacteriia*; GH43; marine bacteria; proteomics; unsaturated rhamnogalacturonylhydrolase

## Introduction

Marine environments contain many different polysaccharides as dissolved organic matter (DOM) or in particulate organic matter (POM). These are a vital carbon source for microorganisms, released from algae as exudates or during lysis by zooplankton predation or viral infection. Monosaccharide analysis of planktonic biomass from the North Sea revealed already in 1982 a dominance of glucose followed by arabinose, galactose, and mannose (Ittekkot et al. 1982, Urbani et al. 2005, Alderkamp et al. 2007, Scholz and Liebezeit 2013, Huang et al. 2021). These monomers are the building blocks of algal polysaccharides: the abundant beta-homoglycans laminarin, cellulose, and xylan are often complemented with species-specific glycans such as agar, alginate, carrageenan, fucoidan, mannan, pectin, porphyran, and ulvan. The degradation of these glycans has been studied intensively in marine systems, however, details for arabinogalactan are missing (Bäumgen et al. 2021). Recently arabinogalactan was detected in the high molecular weight dissolved organic matter (HMWDOM) and POM fraction using monoclonal antibodies during the algal spring bloom in the North Sea (Vidal-Melgosa et al. 2021). This coincides with the high arabinose and galactose content of *Phaeocystis* spp., a haptophyte blooming in the North Sea (Alderkamp et al. 2007, Sato et al. 2018). The antibody-based quantification also showed a decrease in arabinogalactan content towards the end of the spring bloom, suggesting a fast turnover of the compound—

contrasting with the accumulation of fucose-containing sulfated polysaccharides (Vidal-Melgosa et al. 2021). The major source of arabinose and galactose in algae are likely arabinogalactan proteins, which anchor polysaccharide cell walls in the outer membrane of plants and algae (Silva et al. 2020, Leszczuk et al. 2023). The model compound for arabinogalactan type II is arabinogalactan from larch wood. It contains D-galactose and L-arabinose in a 6:1 molar ratio as well as traces of rhamnose, fucose, mannose, xylose, and D-glucuronic acid (Fujita et al. 2019, Villa-Rivera et al. 2021, Leszczuk et al. 2023). Type II arabinogalactans have a complex backbone structure consisting of  $\beta$ -1,3-linked galactan backbone with  $\beta$ -1,6-linked galactan side chains (Kelly 1999, Wang and LaPointe 2020). Type I has a  $\beta$ -1,4-linked galactan backbone, whereby C3 can be linked with L-arabinofuranose (Hinz et al. 2005).

Plant arabinogalactan is degraded by aerobic bacteria and fungi as well as by anaerobic fermenting bacteria in gut systems, including *Bifidobacterium* and *Bacteroidetes* (Shulami et al. 2011, Ndeh et al. 2017, Cartmell et al. 2018, Luis et al. 2018, Wang and LaPointe 2020, Sasaki et al. 2021). The latter phylum encompasses also aerobic *Flavobacteriia* that have been identified as specialists for polysaccharide degradation in marine systems (Sidhu et al. 2023). For this first study on the degradation of arabinogalactan by marine microorganisms, we selected a flavobacterial strain with a published genome and a particle-associated lifestyle, *Maribac-*

Received 5 February 2024; revised 13 March 2024; accepted 2 April 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of FEMS. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

ter sp. MAR\_2009\_72 (Kappelmann et al. 2018, Heins et al. 2021a). Strains of the genus *Maribacter* are rarely isolated from sea water, but they are more abundant in particle fractions (Nedashkovskaya et al. 2004, Heins and Harder 2023, Lu et al. 2023, Sidhu et al. 2023). Abundances of up to 4% were detected in the oxic surface layer of sandy sediments (Probandt et al. 2018, Miksch et al. 2021). Even higher abundances were observed in micro- and macroalgae phycosphere populations (Heins et al. 2021b, Lu et al. 2023). This makes *Maribacter* strains ideal candidates for studying the degradation of algal cell wall polysaccharides.

The uptake and degradation of polysaccharides in *Bacteroidetes* is often encoded in polysaccharide utilization loci (PULs). The first PUL was described for *Bacteroides thetaiotaomicron* for starch utilization (Shipman et al. 2000). Polysaccharide utilization starts with the extracellular hydrolysis of polysaccharides into oligosaccharides on the surface of the cell. The oligosaccharides are transported into the periplasm via the SusC/D transport system, which is energized by a proton gradient via an ExbB/D-TonB system in the cytoplasmic membrane and by a domain in the periplasm to open the  $\beta$ -barrel channel of SusC for the transport (Noinaj et al. 2010). The hydrolysis of polysaccharides is achieved by glycoside hydrolases (GH), glycoside transferases, polysaccharide lyases, and carbohydrate esterases with a high specificity, sometimes assisted by carbohydrate binding modules. These five groups of proteins are classified as carbohydrate active enzymes (CAZymes) (Bäumgen et al. 2021, Drula et al. 2022). For the degradation of arabinogalactan from larch wood, PULs were so far characterized for gut bacteria including *Bifidobacterium longum* ssp. *longum* NCC2705, *Bacteroides caccae* ATCC 43185, and *Bacteroides thetaiotaomicron* (Ndeh et al. 2017, Cartmell et al. 2018, Luis et al. 2018, Wang and LaPointe 2020). Here, we analyzed *Maribacter* sp. MAR\_2009\_72 proteomes using cells grown on arabinogalactan, arabinose, galactose, and glucose. Those proteomes were compared to identify the proteins induced by arabinogalactan. This study expands a recent *in silico* study that did not report on arabinogalactan-specific PULs (Kappelmann et al. 2018) and provides experimental observations for a better interpretation of marine metagenomes.

## Material and methods

### Growth experiments

*Maribacter* sp. MAR\_2009\_72 (DSM 29384), originally isolated from a phytoplankton catch in the Wadden Sea near the island Sylt, Germany, was revived from glycerol stocks that had been preserved in the laboratory since the initial isolation (Hahnke and Harder 2013). The strain was grown in the liquid medium HaHa\_100 V with 0.3 g/l of casamino acids as the sole carbon source (Hahnke et al. 2015). This limited growth to an optical density (OD) at 600 nm below 0.2. Growth beyond an OD of 0.3 was achieved by adding 2 g/l of a carbohydrate source, here arabinose, galactose, glucose (Sigma Aldrich/Merck KGaA, Darmstadt, Germany), and larch arabinogalactan (The Dairy School, Auchincruive, Scotland). The supplier of arabinogalactan had specified the monosaccharide composition as 81% galactose, 14% arabinose, and 5% other, whereby the other fraction was not defined. For proteomics, three cultures of 50 ml were inoculated with 0.4% v/v of a pregrown culture in the same medium and incubated at room temperature at 110 r/m. A fourth culture per substrate was maintained to monitor bacterial growth by measuring OD at 600 nm beyond the harvest point. Cells were harvested at an OD of 0.25. Cells were pelleted by centrifugation in 50 ml tubes with 3080

$\times g$  for 30 min at 4°C. Pellets were resuspended in 1 ml medium and centrifuged in 1.5 ml tubes at 15870  $\times g$  for 15 min at 4°C. The wet biomass was weighed and stored at  $-20^{\circ}\text{C}$ .

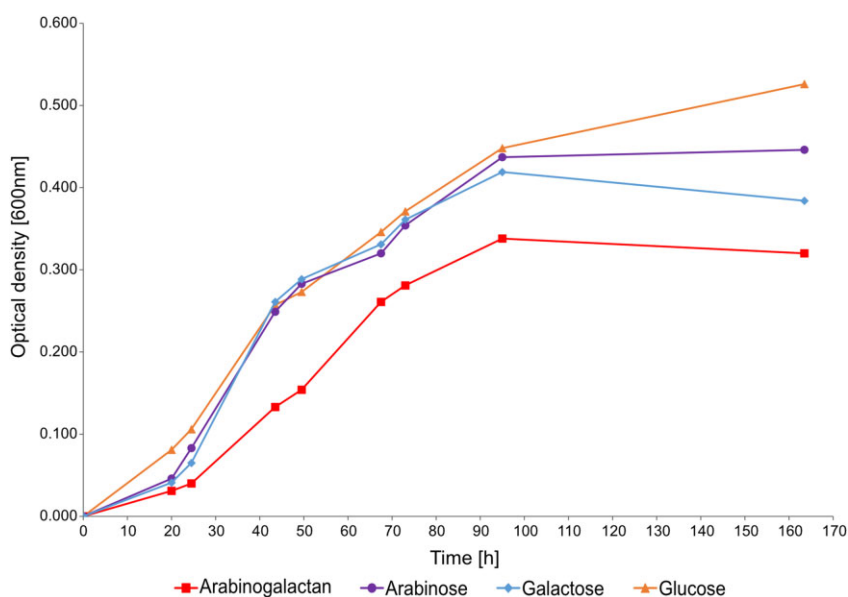
For microbiome size determinations, colony-forming units (CFU) were determined with 4 g/l larch wood arabinogalactan as organic carbon source on marine plates (Hahnke and Harder, 2013), using 4 g/l glucose or ZoBell's 2216 marine agar plates as reference. Inoculation of serial diluted sea or sediment pore water was performed with a 96 pin-holder. Inoculations were at room temperature. Partial 16S rRNA gene sequences of strains were obtained by colony PCR and Sanger sequencing (Hahnke and Harder 2013). Partial 16S rRNA gene sequences have been deposited at GenBank under the accession numbers PP600029 to PP600099.

### Protein preparation and mass spectrometry

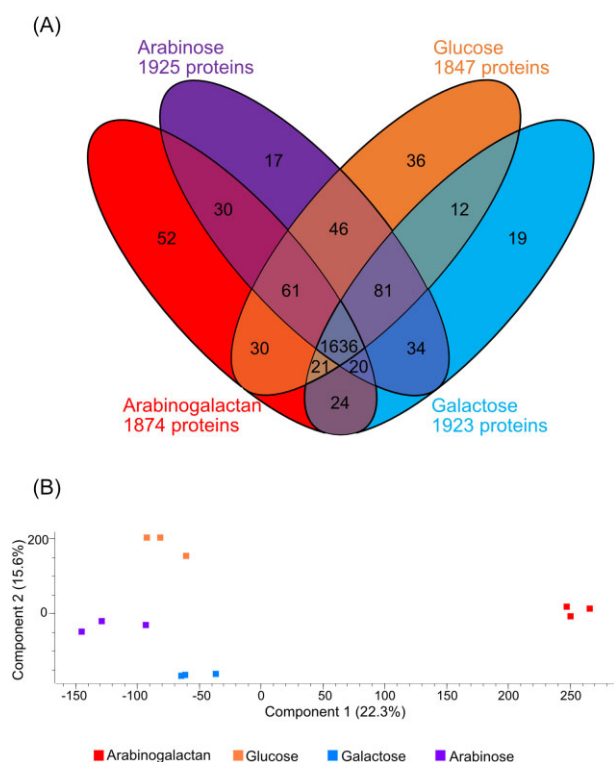
Proteins were extracted from cells using a bead-beating method following the protocol by Schultz et al. (2020). A pellet of wet weight ranging from 20 to 200 mg was disrupted using 0.25 ml glass beads in 500  $\mu\text{l}$  of lysis buffer. The protein content was quantified using the Roti Nanoquant assay (Carl Roth, Karlsruhe, Germany). For protein purification on denaturing polyacrylamide gels (SDS-PAGE), 50  $\mu\text{g}$  of protein was combined with 10  $\mu\text{l}$  of 4x SDS buffer [composed of 20% glycerol, 100 mM Tris/HCl, 10% (w/v) SDS, 5%  $\beta$ -mercaptoethanol, 0.8% bromophenol blue, pH 6.8] and loaded onto Tris-glycine-extended precast 4%–20% gels (Bio-Rad, Neuried, Germany). Electrophoresis was conducted at 150 V for 8 min. Subsequently, the gel was fixed in a solution of 10% v/v acetic acid and 40% v/v ethanol for 30 min, stained with Brilliant Blue G250 Coomassie, and the desired protein band was excised. The proteins were extracted from the gel in one piece and then washed with a solution of 50 mM ammonium bicarbonate in 30% v/v acetonitrile. The gel pieces were dried using a SpeedVac (Eppendorf, Hamburg, Germany), and then rehydrated with 2 ng/ $\mu\text{l}$  trypsin (sequencing grade trypsin, Promega, USA). After a 15-min incubation at room temperature, excess liquid was removed, and the samples were digested overnight at 37°C. Following digestion, the gel pieces were covered with water suitable for mass spectrometry (MS), and peptides were eluted using ultrasonication. The peptides were subsequently desalted using Pierce™ C18 Spin Tips (Thermo Fisher, Schwerte, Germany) in accordance with the manufacturer's guidelines. The eluted peptides were dried using a SpeedVac and stored at  $-20^{\circ}\text{C}$ . For MS analysis, the samples were thawed and reconstituted in 10  $\mu\text{l}$  of Buffer A (99.9% acetonitrile + 0.1% acetic acid).

Tryptic peptides of *Maribacter* sp. MAR\_2009\_72 were analyzed using an EASYnLC 1200 system coupled to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, located in Waltham, USA). Peptides were loaded onto a custom-packed analytical column containing 3  $\mu\text{m}$  C18 particles (Dr. Maisch GmbH, Ammerbuch, Germany). The loading was performed using buffer A (0.1% acetic acid) at a flow rate of 2  $\mu\text{l}/\text{min}$ . Peptide separation was achieved through an 85-min binary gradient, transitioning from 4% to 50% buffer B, composed of 0.1% acetic acid in acetonitrile, at a flow rate of 300 nl/min. Samples were measured in parallel mode; survey scans in the Orbitrap were recorded with a resolution of 60 000 with a m/z range of 333 to 1650. The 15 most intense peaks per scan were selected for fragmentation. Precursor ions were dynamically excluded from fragmentation for 30 s. Single-charged ions as well as ions with unknown charge state were rejected. Internal lock mass calibration was applied (lock mass 445.12003 Da).

The MS files were analyzed in MaxQuant version 2.2.0.0 in the standard settings against the strain specific protein



**Figure 1.** Growth curve of *Maribacter* sp. MAR\_2009\_72 in presence of four different carbon sources; arabinogalactan, arabinose, galactose, and glucose. MAR\_2009\_72 was grown in 50 ml of modified HaHA100V with 2 g/l of the respective carbon source at room temperature at 110 r/m. The OD was measured at 600 nm.



**Figure 2.** Comparison of the number of detected proteins in arabinogalactan, arabinose, galactose, and glucose. (A) Venn diagram showing the overlap of detected proteins in at least one of three biological replicates. (B) Principal component analysis shows the differences between the expression intensities of the four proteomes of MAR\_2009\_72.

database downloaded from NCBI: *Maribacter* sp. MAR\_2009\_72 NZ\_VISB01000001.1 (27 September 2022) and common laboratory contaminants (Tyanova and Cox 2018, Sayers et al. 2022). Statistical analysis was performed in Perseus version 2.0.7.0 (Tyanova and Cox 2018). Proteins were recognized as being expressed when

they had label free quantification intensities (LFQ) intensities in one out of three biological replicates.

## Bioinformatic analyses

Protein annotation was refined using several databases. CAZymes were considered to be identified, if two out of three search algorithms in dbCAN3 were positive (Zheng et al. 2023). The conserved domain database (CDD) (Lu et al. 2019), the SulfAtlas web interface (Stam et al. 2022), InterPro (Paysan-Lafosse et al. 2022), PULDB (Terrapon et al. 2018), deepTMHMM (Hallgren et al. 2022), SignalP (Teufel et al. 2022), Blastkoala (Kanehisa et al. 2016), and UniProt (The UniProt 2023) provided additional information.

The search pattern used in PULDB were the following for PUL 1: GH10+GH43\_1+GH67; for PUL 7: GH43\_18+GH78+GH105+GH105+GH115; and for PUL 8: GH43\_19+GH43\_34+GH51+GH105+GH105.

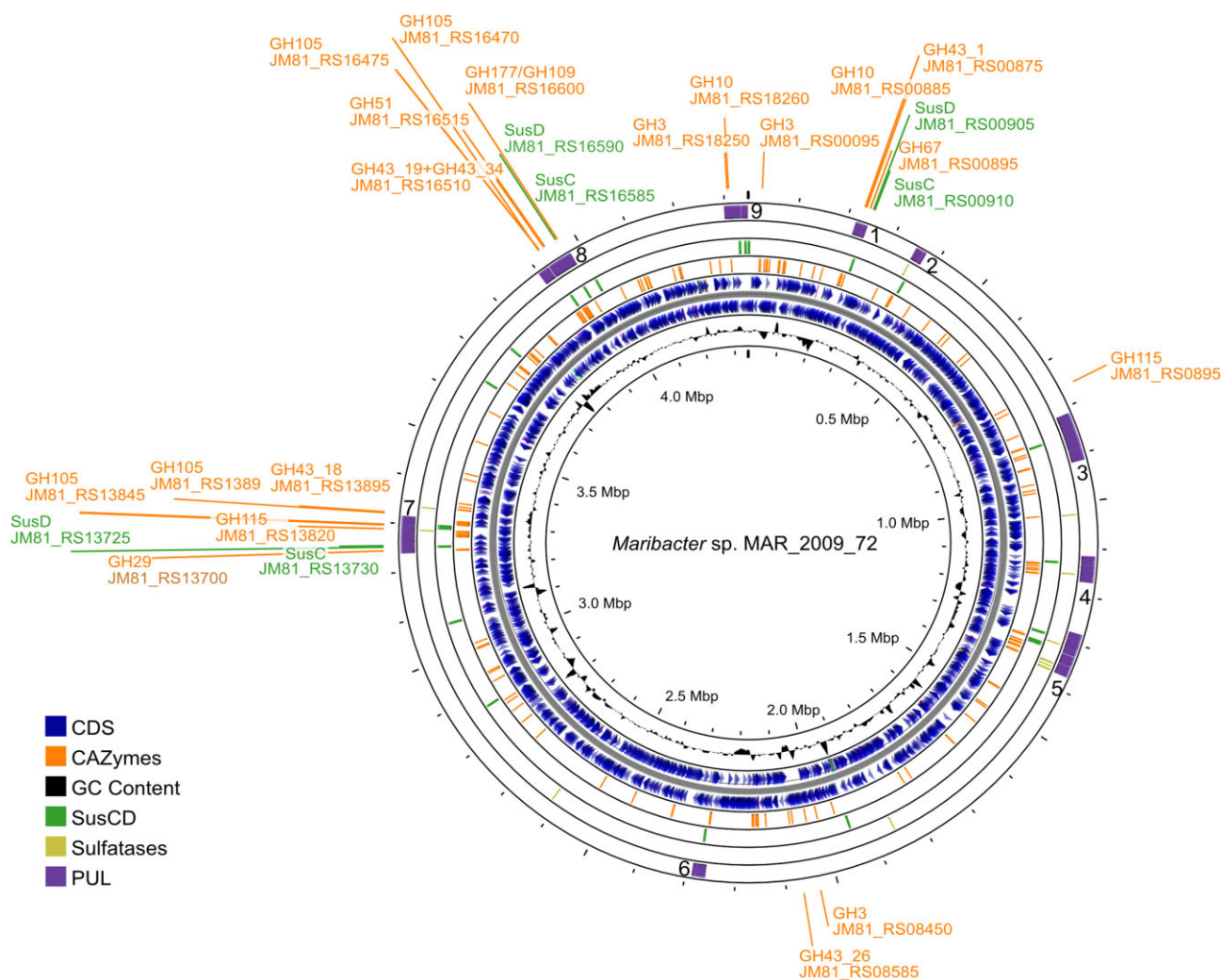
For the visualization of the data the following programs and packages were used: R version 4.3.2 (R Core Team 2023), ggplot2 (Wickham 2016), gggenes (Wilkins 2023), and Proksee (Grant et al. 2023).

The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al. 2022) partner repository with the dataset identifier PXD049074 and 10.6019/PXD049074.

## Results

### Growth on arabinogalactan

*Maribacter* sp. MAR\_2009\_72 grew in presence of larch wood arabinogalactan to a maximum OD of 0.338 and at a maximum growth rate  $\mu = 0.06 \text{ h}^{-1}$  (Fig. 1). When 2 g/l of galactose or arabinose were provided in the medium a maximum OD of 0.419 and 0.446 was measured with respective growth rates of  $0.07 \text{ h}^{-1}$  and  $0.06 \text{ h}^{-1}$ . Glucose supported the largest biomass formation, with an OD of 0.526 and  $\mu = 0.05 \text{ h}^{-1}$ . The arabinogalactan cultures required more time to enter the exponential growth phase than the cultures with monosaccharides as substrates. The physiological



**Figure 3.** Full genome overview of *Maribacter* sp. MAR\_2009\_72 showcasing the GC content (ring one (most inner ring)), all annotated coding genes (CDS, ring two and three) in forward and reverse direction, CAZymes identified by dbCAN3 (ring four), SusC/D (ring five), sulfatases (ring six), and PULs (ring seven). Furthermore, we highlighted CAZymes and SusC/Ds that might be important for arabinogalactan utilization.

reaction was confirmed *in vitro*. Soluble cell extracts hydrolysed larch wood arabinogalactan to an abundant molecule with the size of a hexose (galactose) and a minor molecule with the size of a pentose (arabinose) in fluorophore-assisted carbohydrate electrophoresis gels (Fig. S1, Supporting Information).

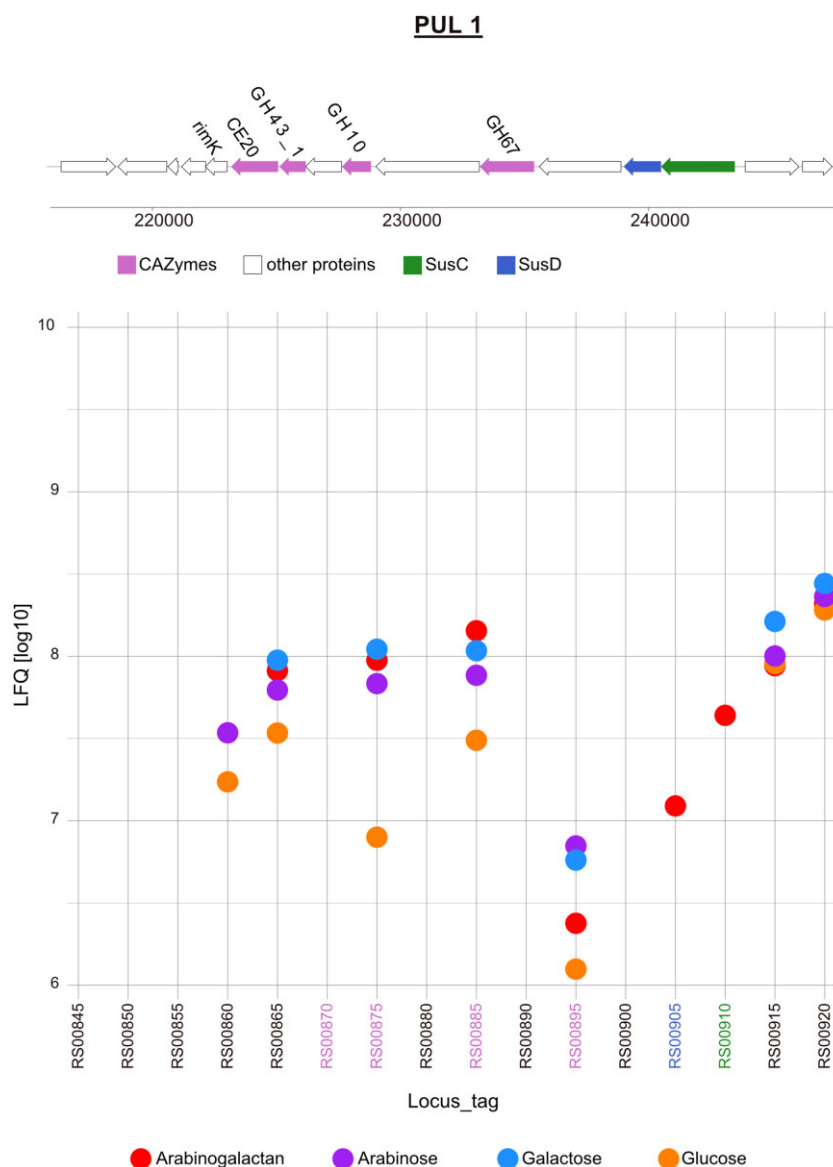
### Protein expression in *Maribacter* sp. MAR\_2009\_72

The comparative proteomic analysis was based on glucose as reference against arabinose, galactose, and arabinogalactan. We identified 1874 proteins in the arabinogalactan proteome (Fig. 2A). Overall, these four conditions shared 1636 proteins. Only a small number of proteins were found to be unique to a particular growth condition. The glucose proteome had 36 unique proteins, the arabinose proteome 17 proteins, and the galactose proteome 19 proteins. Arabinogalactan had 52 unique proteins. We used the expression data, here label free quantification intensities (LFQ), to visualize the difference between the four conditions in a principal component analysis (PCA) (Fig. 2B). The PCA plot indicated that the arabinogalactan proteome had the most contrasting expression pattern. The PCA analysis documented that the differences

between monosaccharide proteomes were less pronounced than to the arabinogalactan proteome.

*Maribacter* sp. MAR\_2009\_72 has a genome of 4.35 Mb encoding 3635 proteins (Fig. 3). Nine PULs contain one or several SusC/D transporter and neighboring CAZymes. We labelled the PULs based on the arrangement in the genome, with PUL 1 being closest to the origin of replication (Table S1, Supporting Information). The expression values revealed a proteomic response to arabinogalactan in PUL 1, 7, and 8 and outside of PULs.

PUL 1 encodes 13 proteins of which three out of four CAZymes and one SusC/D pair were expressed in arabinogalactan grown cells (Fig. 4). The SusC/D pair (JM81\_RS00910 and JM81\_RS00905) was only expressed in the arabinogalactan proteome. Four other proteins were clearly induced by arabinogalactan, arabinose, and galactose. The  $\alpha$ -L-arabinofuranosidase GH43\_1 (JM81\_RS00875) was 10-fold induced relative to the glucose proteome. A GH10, an endo- $\beta$ -1,4-xylanase, showed a similar expression pattern with a 5-fold difference to glucose. The third enzyme was a GH67, an  $\alpha$ -glucuronidase, which had the strongest induction in arabinose and galactose proteomes. The fourth induced protein of the operon with an expression in the arabinogalactan proteome affiliated to the superfamily of protein or cofactor modifying



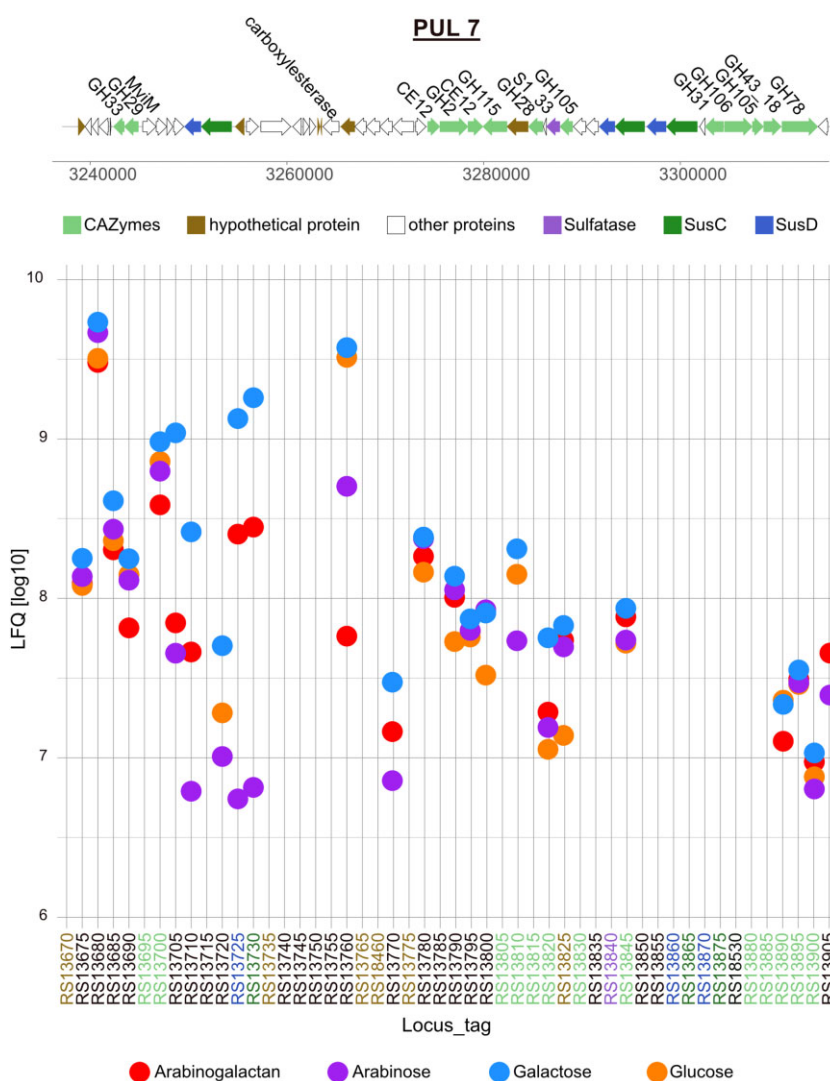
**Figure 4.** Gene organization and expression of polysaccharide utilization locus 1 of *Maribacter* sp. MAR\_2009\_72 grown in the presence of arabinogalactan, arabinose, galactose, and glucose. Expression intensities in the plot are the mean values of three biological replicates of each condition shown in LFQ values [log10].

RimK-type glutamate ligases with an ATP-grasp binding domain (JM81\_RS00865).

PUL 7 contains a single SusC/D pair and a tandem of SusC/D pair in one genetic region. It encodes 42 enzymes, 13 being classified as CAZymes, three SusC/D pairs, and one sulfatase (Fig. 5). One SusC/D pair and 6 CAZymes were expressed in arabinogalactan grown cells. SusC (JM81\_RS13730) and SusD (JM81\_RS13725) were expressed in the galactose and arabinogalactan proteomes 100-fold and 10-fold stronger than in the arabinose proteome, respectively, and not in the glucose proteome. The tandem SusC/D pairs were not detected in any of the proteomes. An  $\alpha$ -L-fucosidase of the GH29 family (JM81\_RS13700) had the highest expression among the CAZymes in this PUL. The GH29 was expressed in similar intensities in all four conditions, suggesting a constitutive expression of this periplasmic enzyme. Less intense, but also expressed in all proteomes was a xylan- $\alpha$ -1,2-glucuronidase belonging to the GH115 family (JM81\_RS13820), with the strongest expression on galactose. Two GH105 unsaturated rhamnogalactur-

onyl hydrolases (EC 3.2.1.172) (JM81\_RS13845 and JM81\_RS13890) were expressed in all four growth conditions, with the exception of JM81\_RS13890, which was not detected in the arabinose proteome. A GH43\_18 (JM81\_RS13895) was expressed in all four proteomes with similar expression intensities. An  $\alpha$ -L-rhamnosidase GH78 (JM81\_RS13900) was expressed under all growth conditions. During our analysis a hypothetical protein (JM81\_RS13825) with a six-hairpin GH like family domain sparked our interest. It was expressed in all four proteomes, with higher intensities in arabinogalactan, arabinose, and galactose proteomes.

PUL 8 encodes a total of 58 proteins, including 11 CAZymes and two SusC/D pairs (Fig. 6). A total of five CAZymes, two SusCs but only one SusD were expressed in arabinogalactan grown cells. SusC (JM81\_RS16585) and SusD (JM81\_RS16590) were expressed in the arabinose and arabinogalactan proteome. Another SusC (JM81\_RS16455) showed expression, slightly lower than the other SusC, in the arabinose proteome and slightly less for arabinogalactan. Two GH105 proteins (JM81\_RS16470 and JM81\_RS16475)

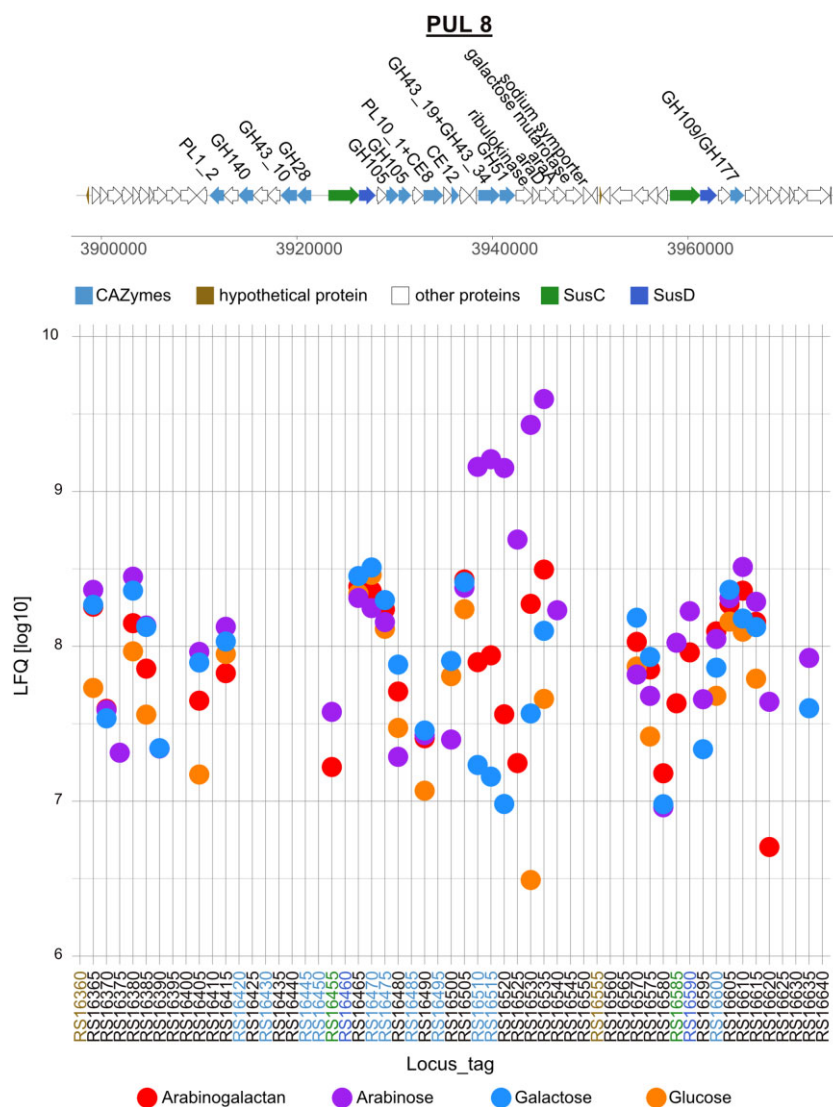


**Figure 5.** Gene organization and expression of polysaccharide utilization locus 7 of *Maribacter* sp. MAR\_2009\_72 grown in the presence of arabinogalactan, arabinose, galactose, and glucose. Expression intensities in the plot are the mean values of three biological replicates of each condition shown in LFC values [log10].

annotated as unsaturated rhamnogalacturonyl hydrolases were expressed similar in all proteomes. JM81\_RS16510 includes two domains, GH43\_19 and GH43\_34. It was expressed in the arabinose, arabinogalactan, and galactose proteome, whereby the highest intensities were measured for arabinose. Another  $\alpha$ -L-arabinofuranosidase, a GH51 (JM81\_RS16515), was expressed in a similar pattern to the GH43\_19+GH43\_34 protein. These two genes are followed by genes of the arabinose metabolism to the pentose phosphate pathway—ribulokinase, L-ribulose-5-phosphate 4-epimerase, and L-arabinose isomerase—and a gene for a galactose mutarotase. All proteins in this operon were expressed in the arabinose, arabinogalactan, and galactose proteome, with highest intensities in arabinose proteomes. Unknown is the function of a GH109, a member of the Gfo/Iah/MocA superfamily of NAD(P) dependent oxidoreductases, that had the highest expression in the arabinogalactan proteome. The expression of a mannonate dehydratase (JM81\_RS16615) hinted at a sugar acid metabolism. Interestingly, PUL 8 is preceded by an operon with sugar acid metabolizing enzymes. The following enzymes were induced in the arabinogalactan proteome in comparison to glucose: 5-dehydro-4-deoxy-D-glucuronate isomerase,

gluconate-5-dehydrogenase, a sugar kinase, 2-dehydro-3-deoxy-phosphogluconate aldolase, and tagaturonate reductase.

An analysis with dbCAN3 identified 153 CAZymes in the genome, of which 106 were detected in the proteomes. Outside of the PULs 1, 7, and 8, several CAZymes were expressed in arabinogalactan degradation. Many expressed CAZymes had a signal peptide for export out of the cytosol (Table S1 and Fig. S2, Supporting Information). Three of the CAZymes were annotated as GH family 3 enzymes. JM81\_RS00095 was expressed in all four conditions, the highest intensities were measured in the arabinogalactan proteome (Fig. S2A, Supporting Information). The second GH3 (JM81\_RS08450) was expressed in all four conditions, but with a three to four times larger expression in arabinogalactan, arabinose, and galactose (Fig. S2B, Supporting Information). A third GH3 (JM81\_RS18250) was as well expressed in all four conditions, but the highest intensities were measured for arabinose and galactose. It was part of an operon also including an endo-1,4- $\beta$ -xylanase (GH10) expressed only in arabinose and galactose grown cells (Fig. S2C, Supporting Information). All three GH3 were annotated as galactosidases. A GH43\_26 (JM81\_RS08585) was expressed in all four datasets, whereby the highest intensities



**Figure 6.** Gene organization and expression of polysaccharide utilization locus 8 of *Maribacter* sp. MAR\_2009\_72 grown in the presence of arabinogalactan, arabinose, galactose, and glucose. Expression intensities in the plot are the mean values of three biological replicates of each condition shown in LFQ values [log<sub>10</sub>].

were recorded for arabinose and nearly identical LFQs for glucose and arabinogalactan (Fig. S2D, Supporting Information). A GH115 xylan- $\alpha$ -1,2-glucuronidase (JM81\_RS03245) was only expressed in arabinose and arabinogalactan grown cells (Fig. S2E, Supporting Information).

The transport of the monosaccharides across the inner membrane may be facilitated by an ABC transport system consisting of ABC substrate-binding (JM81\_RS03610), ABC permease (JM81\_RS16840), and ABC ATP binding proteins (JM81\_RS01625).

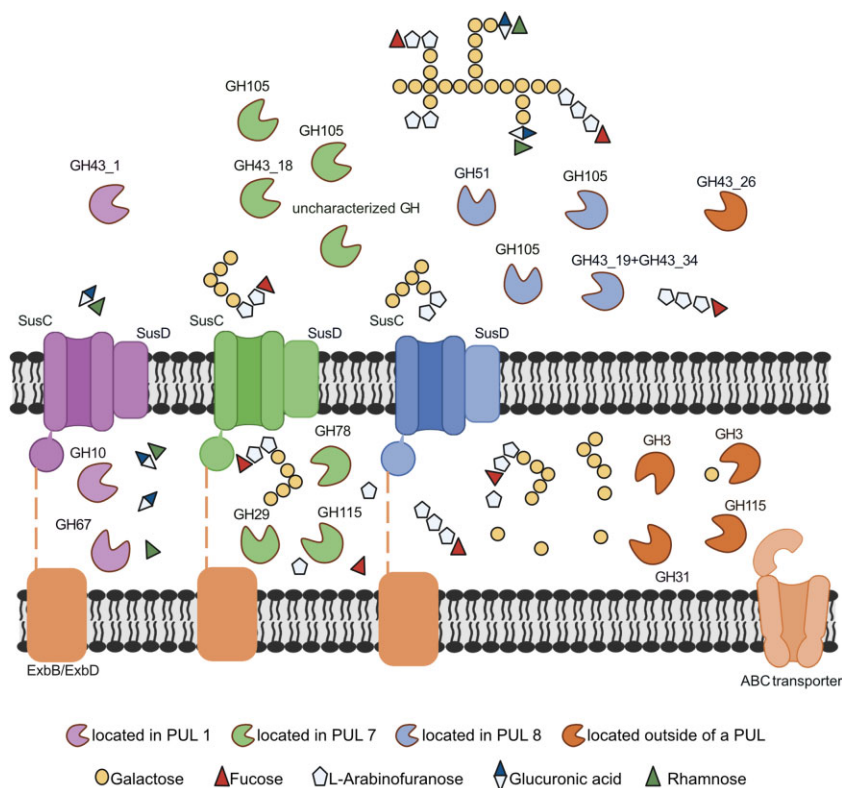
Marine glycans are often decorated with sulfate. We identified 13 sulfatases in the genome of MAR\_2009\_72, of which three were expressed in arabinogalactan grown cells. JM81\_RS05685, JM81\_RS05692, and JM81\_RS076760 were equally expressed in all four proteomes. All three were previously affiliated with the utilization of mucin, which contains to some extent galactose (Tailford et al. 2015, Glover et al. 2022).

We also detected expressed proteins of the degradation pathway that were not classified as CAZymes. For the degradation of galactose, a galactose mutarotase (JM81\_RS1635), galactokinase (JM81\_RS01495), UDP-glucose-4-epimerase (JM81\_RS00155), and nucleotidyltransferases (JM81\_RS00330/14030/10945) were

induced in the arabinogalactan proteome. Rhamnose and glucuronic acid are minor sugars in arabinogalactan from larch. Besides the aforementioned CAZymes and enzymes, three proteins related to the metabolism of sugar acids were induced in the arabinogalactan proteome. Gluconate dehydrogenases JM81\_RS04310 and JM81\_RS04965 yield 2-dehydro-D-gluconate. A 3-dehydro-L-gulonate 2-dehydrogenase (JM81\_RS04965) was also induced.

### CFU microbiomes

North seawater sampled on 28 May 2022, at the beach of Sahlenburg, Cuxhaven, Germany, (N53.863607°, E8.593004°) yielded after 12 days of incubation a CFU microbiome of  $2.57 \times 10^4$ /ml on arabinogalactan and of  $2.32 \times 10^4$ /ml on ZoBell's 2216 marine agar. Partial 16S rRNA gene sequences of abundant CFU on arabinogalactan revealed 27 different genera within 88 colonies, with *Algoriphagus*, *Leeuwenhoekiella*, *Maribacter*, *Sulfitobacter*, *Yoonia-Loktanella*, and *Pseudalteromonas* being present with 6–13 CFU. Only two of 63 CFU on ZoBell's 2216 marine agar, but 29 of 88 CFU on arabinogalactan plates belonged to *Flavobacteriaceae*. Addition-



**Figure 7.** Proposed utilization pathway for arabinogalactan by *Maribacter* sp. MAR\_2009\_72. Poly- and oligosaccharide structures are not experimentally verified. PUL 1: GH43\_1 (JM81\_RS00875), SusC (JM81\_RS00910), SusD (JM81\_RS00905), GH10 (JM81\_RS00885), and GH67 (JM81\_RS00895). PUL 7: GH105 (JM81\_RS13845), GH105 (JM81\_RS13890), GH43\_18 (JM81\_RS13895), uncharacterized GH (JM81\_RS13825), SusC (JM81\_RS13730), SusD (JM81\_RS13725), GH29 (JM81\_RS13700), GH78 (JM81\_RS13900), and GH115 (JM81\_RS13820). PUL 8: GH105 (JM81\_RS16475), GH105 (JM81\_RS16470), GH51 (JM81\_RS16515), GH43\_19+GH43\_34 (JM81\_RS16510), SusC (JM81\_RS16585), and SusD (JM81\_RS16470). Outside of the PULs: GH43\_26 (JM81\_RS08590), GH3 (JM81\_RS00095), GH3 (JM81\_RS08450), GH31 (JM81\_RS06015), CBM48+GH13\_9 (JM81\_RS06020), GH115 (JM81\_RS03245), ExbB (JM81\_RS07935), ExbD (JM81\_RS07930), ABC substrate-binding (JM81\_RS03610), ABC permease (JM81\_RS16840), and ABC ATP binding (JM81\_RS01625). The figure was created with BioRender.com.

ally, we investigated autumn samples. Seawater sampled 18–21 October 2022 on Sylt, Germany (N55.031417°, E8.441802°), showed after 18 days of incubation  $5.06 (+/-0.88) \times 10^5$  CFU/ml for glucose and of  $8.32 (+/- 2.21) \times 10^5$  CFU/ml for arabinogalactan ( $n = 2$ ). At the sandy west beach (N55.037505°, E8.383341°) plating of pore water yielded  $1.75 (+/-1.30) \times 10^6$  CFU/ml on for glucose and  $0.90 (+/- 0.64) \times 10^6$  CFU/ml for arabinogalactan ( $n = 3$ ).

## Discussion

Galactose belongs to the four abundant monosaccharides in planktonic organic matter, mainly as part of polysaccharides and more complex molecules, i.e. arabinogalactan proteins. Plating sea and sediment pore water on arabinogalactan medium showed a large microbiome with the capacity to utilize arabinogalactan for growth. Together with the recent finding that particle-associated bacteria dominate the readily culturable fraction of seater microbiomes (Heins and Harder 2023) this observation indicates that arabinogalactan is a common carbon source for particle-associated bacteria.

Arabinogalactan degradation pathways were so far only described for bacteria from gut and plant systems, but not for marine bacteria (Shulami et al. 2011, Ndeh et al. 2017, Cartmell et al. 2018, Luis et al. 2018, Fujita et al. 2019, Wang and LaPointe 2020, Sasaki et al. 2021). These studies provided information regarding enzymes involved in arabinogalactan utilization, which includes GH families GH43, GH51, GH27, and GH28, often orga-

nized in PULs (Shulami et al. 2011, Cartmell et al. 2018, Luis et al. 2018). Hence, we inspected first the upregulated proteins in arabinogalactan grown cells in comparison to glucose grown cells. After a discussion of the SusC/D systems, we analyzed the uniqueness of marine PULs for arabinogalactan degradation in *Maribacter* sp. MAR\_2009\_72.

Enzymes annotated as members of the GH family 43 are most frequently linked with arabinogalactan utilization. This protein family is one of the largest GH families and has currently 39 subfamilies, whereby only a part has been characterized enzymatically. Most common enzymatic functions are  $\beta$ -D-xylosidase (3.2.1.37),  $\alpha$ -L-arabinofuranosidase (3.2.1.55), endo- $\alpha$ -L-arabinonase (3.2.1.99), and 1,3- $\beta$ -galactosidase (3.2.1.145) (Mewis et al. 2016), with no correlation between subfamily classification and enzyme activity (Kelly 1999, Wang and LaPointe 2020), but as common feature the presence of three active sites in the enzymes. Still, many need a biochemical characterization (Pons et al. 2004, Mewis et al. 2016). The genome of *Maribacter* sp. MAR\_2009\_72 encodes five GH43 and four of these were shown to be expressed in this study. Subfamilies GH43\_1 (JM81\_RS00875) and GH43\_26 (JM81\_RS08585) are characterized so far solely as  $\alpha$ -L-arabinofuranosidases. They hydrolyze via retention of configuration (Pitson et al. 1996) the terminal non-reducing residues in  $\alpha$ -L-arabinoside side chains of hemicelluloses, like arabinoxylan, arabinogalactan, and arabinan (Shallom et al. 2002a, b). JM81\_RS16510 encodes a protein that consists out of two GH43 domains, GH43\_19 and GH43\_34. Over 70% of subfam-



ily GH43\_34 proteins contain a second GH43 domain (Mewis et al. 2016). GH43\_34 domains are usually located at the C-terminus and considered to be membrane-spanning (Mewis et al. 2016). The function is not clear, but they may be involved in carbohydrate binding (Mewis et al. 2016). Unlike the GH43\_1 and GH43\_26 proteins, the GH43\_19 domain is annotated as a catalytic  $\alpha$ -L-arabinofuranosidase domain. The fourth GH43 protein is a member of subfamily GH43\_18 (JM81\_RS13895), which mostly includes  $\alpha$ -L-arabinofuranosidases.

The  $\alpha$ -L-arabinofuranosidases function is also annotated to GH51 proteins, although the amino acid sequence differs by large from GH43 enzymes (Sevim et al. 2017). The acid/base catalyst of GH51 is E176 in *Thermobacillus xylanticus* and E175 in *Geobacillus stearothermophilus* T-6 (Debeche et al. 2002, Shallom et al. 2002a). The GH51 from MAR\_2009\_72 (JM81\_RS16515) has an E177, which aligns perfectly with the catalytic glutamic acid of *T. xylanticus* and *G. stearothermophilus* T-6. The amino acid E294 in *G. stearothermophilus* T-6 was identified as catalytic nucleophile (Shallom et al. 2002b), this glutamate is also conserved in GH51 of MAR\_2009\_72. In contrast to the inverting  $\alpha$ -L-arabinofuranosidases of family GH43,  $\alpha$ -L-arabinofuranosidases of GH51 hydrolyze bonds retaining the anomeric configuration (Pitson et al. 1996).

The family GH105 comprises unsaturated rhamnogalacturonyl hydrolases, which have been suggested to remove L-rhamnose- $\alpha$ -1,4-D-galacturonic acid on arabinogalactan (Munoz-Munoz et al. 2021). Four GH105 proteins were expressed. The catalysis involves hydrolysis of bonds through syn-hydration of double bonds between C4 and C5 carbons of enopyranuronosyl residues of substrates (Itoh et al. 2006a, b). The catalytic aspartate side chain D143 of GH105 is conserved in GH105 proteins of MAR\_2009\_72 (Itoh et al. 2006a).

The transport of the oligosaccharides involved several SusC/D pairs. PULs 1, 7, and 8 encode the three SusC/D systems that had the highest expression intensities of all SusC/Ds in the arabinogalactan proteome. On the basis of the dedicated substrate specificity of SusC/D transport systems, we propose two explanations for the induction of several SusC/D pairs: (i) the extracellular hydrolysis of larch wood arabinogalactan generates a mixture of structurally different oligosaccharides which need dedicated transport system and (ii) a signal molecule derived from larch wood arabinogalactan may induce the expression of proteins that may not be necessary for larch wood arabinogalactan, but for the degradation of marine arabinogalactans. The structural diversity of arabinogalactans in terrestrial system is well characterized (Fujita et al. 2019, Villa-Rivera et al. 2021, Leszczuk et al. 2023), but marine arabinogalactans are understudied.

In the periplasm the oligosaccharides are further hydrolyzed by a range of CAZymes. Some PULs (1 and 7) expressed enzymes that can generate monomers. Furthermore, the proteome detected CAZymes that are not encoded in PULs and are predicted to be periplasmic. The GH10 of PUL 1 was annotated as an endo-1,4- $\beta$ -xylanase, which indicates that arabinoxylans may also be a substrate for the enzymes of PUL 1. The expression of an  $\alpha$ -glucuronidase annotated to GH67, coincides with the presence of glucuronic acid in side chains of arabinogalactan. GH67 removes glucuronic acid from side chains by a single displacement mechanism using an inverting mechanism (Shulami et al. 1999, Biely et al. 2000, Nagy et al. 2002). But it only removes glucuronic acid from nonreducing ends of the oligo- and polysaccharides. A broader substrate range is known for GH115 proteins, which remove glucuronic acid from terminal and internal regions of oligosaccharides (Ryabova et al. 2009, Aalbers et al. 2015). The presence of both GH families, GH67 and two GH115, suggests that glucuronic

acid is part of the decoration of arabinogalactans. The expression of the GH29 argues for fucose as a decorating sugar. Enzymes of the family GH29 are exo- $\alpha$ -fucosidases and cleave via an retaining mechanism (Grootaert et al. 2020). Also, rhamnose as specific substrate is supported by expression of a GH78,  $\alpha$ -L-rhamnosidase. This GH family solely includes rhamnosidases, which use an inverting mechanism to hydrolyze bonds in cooperation with their catalytic residues (Cui et al. 2007). The galactan backbone hydrolysis requires a  $\beta$ -D-galactosidase. This enzymatic function is frequent among members of the GH family GH3. The proteome detected three expressed GH3 proteins. Final steps of the arabinogalactan pathway include the translocation through the inner membrane, likely via an ABC transport system, and cytoplasmic transformations to channel galactose, arabinose, glucuronic acid, rhamnose, and fucose into the pentose phosphate pathway and the glycolysis.

We investigated the distribution of PUL 1, 7, and 8 of *Maribacter* sp. MAR\_2009\_72 in the PULDB database using the expressed CAZymes (Terrapon et al. 2018). Homologs of PUL 1 have been characterized for human gut bacteria and *Bacteroides* spp. for the utilization of a range of xylan polysaccharides including arabinoxylan (Martens et al. 2008, Rogowski et al. 2015, Wang et al. 2016). The PUL was *in silico* detected in genomes of a large variety of *Bacteroidota*. In contrast, PUL 7 has so far not been studied experimentally. An *in silico* search detected a homologous PUL structure in *Maribacter sedimenticola* DSM19840 (Nedashkovskaya et al. 2004). PUL 8 has also a homolog in *M. sedimenticola* DSM19840 and other *Bacteroidota*.

A recent metagenomic study of particle-associated bacteria detected a GH43-rich PUL in a *Maribacter* MAG, which the authors annotated as an arabinogalactan PUL (Wang et al. 2024). This PUL is different to the PULs we identified for arabinogalactan in the genome of *Maribacter* sp. MAR\_2009\_72.

Our observations revealed a substrate specificity of the three PULs. In PUL 1, arabinogalactan is the only inducer for SusC/D, and the expression of a glucuronidase and a xylanase suggests that also glycans with these sugars are substrates for the PUL (Fig. S3, Supporting Information). This hypothesis is supported by previous studies with gut bacteria (Martens et al. 2008, Rogowski et al. 2015, Wang et al. 2016). PULs 7 and 8 have so far not been experimentally observed. PUL 7 is characterized by a very strong induction of SusC/D by galactose and arabinogalactan (Fig. S4, Supporting Information). Galactose is for several proteins the strongest inducer, suggesting galactans as substrate. The presence of fucosidase, glucuronidase, and rhamnosidase suggests a decoration of the marine galactans with the corresponding monosaccharides. PUL 8 is dedicated to arabinose containing glycans. The SusC/D is induced by arabinose and arabinogalactan (Fig. S5, Supporting Information). Besides GHs, the genetic region of PUL 8 includes also monosaccharide-transforming cytoplasmic enzymes for arabinose and sugar acids. This PUL shows that the consideration of cytosolic carbohydrate-transforming enzymes in the bioinformatic analysis of PULs may improve predictions of substrate specificity.

The comparative proteomic analysis of larch wood arabinogalactan degradation by *Maribacter* sp. MAR\_2009\_72 identified expressed proteins encoded in three PULs and outside of PULs (Fig. 3). In summary, members of the GH families 43, 51, and 105 may produce a variety of oligosaccharides. At least three SusC/D systems are involved in the transport into the periplasm, where enzymes belonging to the GH families 3, 10, 29, 67, 78, and 115 produce monosaccharides. The interplay of all these enzymes allows for the utilization of arabinogalactan, which we have summarized

in a graph (Fig. 7). The plant polysaccharide structure is expected to be less complex than the variety of arabinogalactans present in the marine habitat (Pfeifer et al. 2020). This may explain why not all CAZymes of each PUL were detected as expressed proteins. A difference between this study of a marine bacterium and previous studies on gut and plant associated bacteria was the presence of GH105 enzymes and the absence of GH27 and GH28 enzymes. Future studies might characterize marine arabinogalactans and enzymatic studies will resolve the individual functions of the induced proteins to provide further information on the microbial utilization.

## Acknowledgement

We thank Nina Endt, Helene Bardella, Dirk Albrecht, and Sabine Kühn for their technical assistance. We thank the Alfred Wegener Institute (Sylt) for support in the field campaign. S.K. is a member of the International Max Planck Research School of Marine Microbiology (MarMic).

## Author contributions

Saskia Kalenborn (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing), Daniela Zühlke (Investigation, Methodology, Supervision, Writing – review & editing), Katharina Riedel (Funding acquisition, Resources, Writing – review & editing), Rudolf I. Amann (Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing), and Jens Harder (Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing)

## Supplementary data

Supplementary data is available at [FEMSEC Journal](#) online.

*Conflict of interest:* We declare no conflict of interest.

## Funding

We acknowledge the support by the Max Planck Society and the funding by the Deutsche Forschungsgemeinschaft (DFG) in the framework of research unit FOR2406 “Proteogenomics of Marine Polysaccharide Utilization (POMPU)” with the grants RI 969/9-2 (K. R.), AM73/9-3 (R.I.A.), and HA 1673/9-2 (J.H.).

## References

- Aalbers F, Turkenburg JP, Davies GJ et al. Structural and functional characterization of a novel family GH115 4-O-methyl- $\alpha$ -glucuronidase with specificity for decorated arabinogalactans. *J Mol Biol* 2015;**427**:3935–46.
- Alderkamp A-C, Buma AGJ, van Rijssel M. The carbohydrates of *Phaeocystis* and their degradation in the microbial food web. *Biogeochemistry* 2007;**83**:99–118.
- Bäumgen M, Dutschei T, Bornscheuer UT. Marine polysaccharides: occurrence, enzymatic degradation and utilization. *ChemBioChem* 2021;**22**:2247–56.
- Biely P, de Vries RP, Vršanská M et al. Inverting character of  $\alpha$ -glucuronidase a from *Aspergillus tubingensis*. *Biochim Biophys Acta* 2000;**1474**:360–4.
- Cartmell A, Muñoz-Muñoz J, Briggs JA et al. A surface endogalactanase in *Bacteroides thetaiotaomicron* confers keystone status for arabinogalactan degradation. *Nat Microbiol* 2018;**3**:1314–26.
- Cui Z, Maruyama Y, Mikami B et al. Crystal structure of glycoside hydrolase family 78  $\alpha$ -L-rhamnosidase from *Bacillus* sp. GL1. *J Mol Biol* 2007;**374**:384–98.
- Debeche T, Bliard C, Debeire P et al. Probing the catalytically essential residues of the  $\alpha$ -L-arabinofuranosidase from *Thermobacillus xylanilyticus*. *Protein Eng Design Select* 2002;**15**:21–8.
- Drula E, Garron M-L, Dogan S et al. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* 2022;**50**:D571–7.
- Fujita K, Sasaki Y, Kitahara K. Degradation of plant arabinogalactan proteins by intestinal bacteria: characteristics and functions of the enzymes involved. *Appl Microbiol Biotechnol* 2019;**103**:7451–7.
- Glover JS, Ticer TD, Engevik MA. Characterizing the mucin-degrading capacity of the human gut microbiota. *Sci Rep* 2022;**12**:8456.
- Grant JR, Enns E, Marinier E et al. Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res* 2023;**51**:W484–92.
- Grootaert H, Van Landuyt L, Hulpiu P et al. Functional exploration of the GH29 fucosidase family. *Glycobiology* 2020;**30**:735–45.
- Hahnke RL, Harder J. Phylogenetic diversity of Flavobacteria isolated from the North Sea on solid media. *Syst Appl Microbiol* 2013;**36**:497–504.
- Hahnke RL, Bennke CM, Fuchs BM et al. Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. *Environ Microbiol* 2015;**17**:3515–26.
- Hallgren J, Tsigos KD, Pedersen MD et al. Preprint DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *Biorxiv* 2022. <https://doi.org/10.1101/2022.04.08.487609>.
- Heins A, Amann RI, Harder J. Cultivation of particle-associated heterotrophic bacteria during a spring phytoplankton bloom in the North Sea. *Syst Appl Microbiol* 2021a;**44**:126232.
- Heins A, Harder J. Particle-associated bacteria in seawater dominate the colony-forming microbiome on ZoBell marine agar. *FEMS Microbiol Ecol* 2022;**99**:fiac151.
- Heins A, Reintjes G, Amann RI et al. Particle collection in Imhoff sedimentation cones enriches both motile chemotactic and particle-attached bacteria. *Front Microbiol* 2021b;**12**:643730.
- Hinz SWA, Verhoef R, Schols HA et al. Type I arabinogalactan contains  $\beta$ -d-galp-(1→3)- $\beta$ -d-galp structural elements. *Carbohydr Res* 2005;**340**:2135–43.
- Huang G, Vidal-Melgosa S, Sichert A et al. Secretion of sulfated fucans by diatoms may contribute to marine aggregate formation. *Limnol Oceanogr* 2021;**66**:3768–82.
- Itoh T, Ochiai A, Mikami B et al. A novel glycoside hydrolase family 105: the structure of family 105 unsaturated rhamnogalacturonyl hydrolase complexed with a disaccharide in comparison with family 88 enzyme complexed with the disaccharide. *J Mol Biol* 2006a;**360**:573–85.
- Itoh T, Ochiai A, Mikami B et al. Structure of unsaturated rhamnogalacturonyl hydrolase complexed with substrate. *Biochem Biophys Res Commun* 2006b;**347**:1021–9.
- Ittekkot V, Degens ET, Brockmann U. Monosaccharide composition of acid-hydrolyzable carbohydrates in particulate matter during a plankton bloom. *Limnol Oceanogr* 1982;**27**:770–6.
- Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016;**428**:726–31.

- Kappelmann L, Krüger K, Hehemann J-H et al. Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J* 2019;**13**:76–91.
- Kelly GS. Larch arabinogalactan: clinical relevance of a novel immune-enhancing polysaccharide. *Alternat Med Rev J Clin Therap* 1999;**4**:96–103.
- Leszczuk A, Kalaitzis P, Kulik J et al. Review: structure and modifications of arabinogalactan proteins (AGPs). *BMC Plant Biol* 2023;**23**:45.
- Lu D-C, Wang F-Q, Amann RI et al. Epiphytic common core bacteria in the microbiomes of co-located green (*Ulva*), brown (*Saccharina*) and red (*Grateloupia*, *Gelidium*) macroalgae. *Microbiome* 2023;**11**:126.
- Lu S, Wang J, Chitsaz F et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* 2020;**48**:D265–8.
- Luis AS, Briggs J, Zhang X et al. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides*. *Nat Microbiol* 2018;**3**:210–9.
- Martens EC, Chiang HC, Gordon JI. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* 2008;**4**:447–57.
- Mewis K, Lenfant N, Lombard V et al. Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. *Appl Environ Microb* 2016;**82**:1686–92.
- Miksch S, Meiners M, Meyerdierks A et al. Bacterial communities in temperate and polar coastal sands are seasonally stable. *ISME Commun* 2021;**1**:29.
- Munoz-Munoz J, Ndeh D, Fernandez-Julia P et al. Sulfation of arabinogalactan proteins confers privileged nutrient status to *Bacteroides plebeius*. *mBio* 2021;**12**:e0136821. <https://doi.org/10.1128/mbio.01368-21>.
- Nagy T, Emami K, Fontes Carlos MGA et al. The membrane-bound  $\alpha$ -glucuronidase from *Pseudomonas cellulosa* hydrolyzes 4-O-methyl-d-glucuronoxyloligosaccharides but not 4-O-methyl-D-glucuronoxylan. *J Bacteriol* 2002;**184**:4925–9.
- Ndeh D, Rogowski A, Cartmell A et al. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* 2017;**544**:65–70.
- Nedashkovskaya OI, Kim SB, Han SK et al. *Maribacter* gen. nov., a new member of the family *Flavobacteriaceae*, isolated from marine habitats, containing the species *Maribacter sedimenticola* sp. nov., *Maribacter aquivivus* sp. nov., *Maribacter orientalis* sp. nov. and *Maribacter ulvicola* sp. nov. *Int J Syst Evol Microbiol* 2004;**54**:1017–23.
- Noinaj N, Guillier M, Barnard TJ et al. TonB-dependent transporters: regulation, structure, and function. *Annu Rev Microbiol* 2010;**64**:43–60.
- Paysan-Lafosse T, Blum M, Chuguransky S et al. InterPro in 2022. *Nucleic Acids Res* 2023;**51**:D418–27.
- Perez-Riverol Y, Bai J, Bandla C et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 2022;**50**:D543–52.
- Pfeifer L, Shafee T, Johnson KL et al. Arabinogalactan-proteins of *Zostera marina* L. contain unique glycan structures and provide insight into adaptation processes to saline environments. *Sci Rep* 2020;**10**:8232.
- Pitson SM, Voragen AGJ, Beldman G. Stereochemical course of hydrolysis catalyzed by arabinofuranosyl hydrolases. *FEBS Lett* 1996;**398**:7–11.
- Pons T, Naumoff DG, Martínez-Fleites C et al. Three acidic residues are at the active site of a  $\beta$ -propeller architecture in glycoside hydrolase families 32, 43, 62, and 68. *Proteins* 2004;**54**:424–32.
- Probandt D, Eickhorst T, Ellrott A et al. Microbial life on a sand grain: from bulk sediment to single grains. *ISME J* 2018;**12**:623–33.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2023.
- Rogowski A, Briggs JA, Mortimer JC et al. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun* 2015;**6**:7481.
- Ryabova O, Vršanská M, Kaneko S et al. A novel family of hemicellulolytic  $\alpha$ -glucuronidase. *FEBS Lett* 2009;**583**:1457–62.
- Sasaki Y, Horigome A, Odamaki T et al. Novel 3-O- $\alpha$ -d-galactosyl- $\alpha$ -l-arabinofuranosidase for the assimilation of gum arabic arabinogalactan protein in *Bifidobacterium longum* subsp. *Longum*. *Appl Environ Microb* 2021;**87**:e02690–20.
- Sato K, Hara K, Yoshimi Y et al. Yariv reactivity of type II arabinogalactan from larch wood. *Carbohydr Res* 2018;**467**:8–13.
- Sayers EW, Bolton EE, Brister JR et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res* 2022;**50**:D20–6.
- Scholz B, Liebezeit G. Biochemical characterisation and fatty acid profiles of 25 benthic marine diatoms isolated from the Solthörn tidal flat (southern North Sea). *J Appl Phycol* 2013;**25**:453–65.
- Schultz D, Zühlke D, Bernhardt J et al. An optimized metaproteomics protocol for a holistic taxonomic and functional characterization of microbial communities from marine particles. *Environ Microbiol Rep* 2020;**12**:367–76.
- Sevim E, Inan Bektas K, Sevim V et al. Purification and characterization of  $\alpha$ -L-arabinofuranosidases from *Geobacillus stearothermophilus* strain 12. *Biologia* 2017;**72**:831–9.
- Shallom D, Belakhov V, Solomon D et al. The identification of the acid-base catalyst of  $\alpha$ -arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *FEBS Lett* 2002a;**514**:163–7.
- Shallom D, Belakhov V, Solomon D et al. Detailed kinetic analysis and identification of the nucleophile in  $\alpha$ -L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *J Biol Chem* 2002b;**277**:43667–73.
- Shipman JA, Berleman JE, Salyers AA. Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J Bacteriol* 2000;**182**:5365–72.
- Shulami S, Gat O, Sonenshein Abraham L et al. The glucuronic acid utilization gene cluster from *Bacillus stearothermophilus* T-6. *J Bacteriol* 1999;**181**:3695–704.
- Shulami S, Raz-Pasteur A, Tabachnikov O et al. The L-arabinan utilization system of *Geobacillus stearothermophilus*. *J Bacteriol* 2011;**193**:2838–50.
- Sidhu C, Kirstein IV, Meunier CL et al. Dissolved storage glycans shaped the community composition of abundant bacterioplankton clades during a North Sea spring phytoplankton bloom. *Microbiome* 2023;**11**:77.
- Silva J, Ferraz R, Dupree P et al. Three decades of advances in arabinogalactan-protein biosynthesis. *Front Plant Sci* 2020;**11**:610377.
- Stam M, Lelièvre P, Hoebeke M et al. SulfAtlas, the sulfatase database: state of the art and new developments. *Nucleic Acids Res* 2023;**51**:D647–53.
- Tailford LE, Crost EH, Kavanaugh D et al. Mucin glycan foraging in the human gut microbiome. *Front Genet* 2015;**6**:81.
- Terrapon N, Lombard V, Drula É et al. PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res* 2018;**46**:D677–83.
- Teufel F, Almagro Armenteros JJ, Johansen AR et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* 2022;**40**:1023–5.

- The UniProt Consortium. UniProt: the Universal Protein knowledge-base in 2023. *Nucleic Acids Res* 2023;**51**:D523–31.
- Tyanova S, Cox J. Perseus: a bioinformatics platform for integrative analysis of proteomics data in cancer research. Stechow von (ed.), In: *Cancer Systems Biology: Methods and Protocols*. New York: Springer, 2018, pp. 133–48.
- Urbani R, Magaletti E, Sist P et al. Extracellular carbohydrates released by the marine diatoms *Cylindrotheca closterium*, *Thalassiosira pseudonana* and *Skeletonema costatum*: effect of P-depletion and growth status. *Sci Total Environ* 2005;**353**:300–6.
- Vidal-Melgosa S, Sichert A, Francis TB et al. Diatom fucan polysaccharide precipitates carbon during algal blooms. *Nat Commun* 2021;**12**:1150.
- Villa-Rivera MG, Cano-Camacho H, López-Romero E et al. The role of arabinogalactan type II degradation in plant-microbe interactions. *Front Microbiol* 2021;**12**:730543.
- Wang F-Q, Bartosik D, Sidhu C et al. Particle-attached bacteria act as gatekeepers in the decomposition of complex phytoplankton polysaccharides. *Microbiome* 2024;**12**:32.
- Wang W, Yan R, Nocek BP et al. Biochemical and structural characterization of a five-domain GH115  $\alpha$ -glucuronidase from the marine bacterium *Saccharophagus degradans* 2-40T. *J Biol Chem* 2016;**291**:14120–33.
- Wang Y, LaPointe G. Arabinogalactan utilization by *Bifidobacterium longum* subsp. *Longum* NCC 2705 and *Bacteroides caccae* ATCC 43185 in monoculture and coculture. *Microorganisms* 2020;**8**:1703.
- Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag, 2016.
- Wilkins D. *gggenes: Draw Gene Arrow maps in 'ggplot2.'* CRAN, 2023.
- Zheng J, Ge Q, Yan Y et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res* 2023;**51**:W115–21.