# Polysaccharide degradation by marine flavobacteria

Inauguraldissertation

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Universität Greifswald

Vorgelegt von

Irena Beidler

Greifswald, 04.09.2023

Dekan: Prof. Dr. Gerald Kerth

1. Gutachter: Prof. Dr. Thomas Schweder
2. Gutachter: Prof. Dr. Dirk Schüler
3. Gutachter: Prof. Dr. Matthias Labrenz

Tag der Promotion: 15.02.2024

"*We each exist for but a short time, and in that time explore but a small part of the whole universe*"

— Stephen Hawking

# Zusammenfassung

In den Weltmeeren findet rund die Hälfte der jährlichen globalen Kohlenstofffixierung statt, davon ein großer Anteil in küstennahen Regionen. Hier kommt es zu wiederkehrenden saisonalen Algenblüten, die durch eine zeitlich begrenzte explosionsartige Vermehrung von Mikroalgen (hauptsächlich Diatomeen und Coccolithophoren) charakterisiert sind. Vor allem Frühjahrsblüten (März-Mai) haben aufgrund ihrer zeitlichen und räumlichen Vorhersagbarkeit einen hohen Stellenwert als Modellsysteme, anhand deren sich der Kohlenstoffkreislauf der Meere untersuchen lässt.

Mikroalgen produzieren eine große Vielfalt an Makromolekülen, die für die mit ihnen vergesellschafteten Bakterien als Nahrungsgrundlage dienen. Besonders im Fokus stehen hier die für den Kohlenstoffkreislauf relevanten Polysaccharide. Im Gegensatz zu anderen natürlichen Makromolekülen wie DNA oder Proteinen können Polysaccharide aus vielen verschiedenen Monomeren mit unterschiedlichsten Bindungen bestehen. Zusätzlich finden sich an diesen Zuckermonomeren viele Modifikationen wie Acetylierungen, Methylierungen oder Sulfatierungen, die die Komplexität weiter erhöhen. Diese Variabilität bedingt eine hohe strukturelle und funktionale Diversität. So können Polysaccharide Speicherstoffe, Zellwandbestandteile oder Teile der extrazellulären Matrix darstellen.

Komplementär hierzu besitzen Polysaccharid-verwertende Bakterien entsprechend komplexe, enzymatische Abbaumechanismen. Besonders hervorzuheben sind hier die Bakterien des Phylums *Bacteroidota*, die sich in verschiedensten Nischen auf den Abbau von Polysacchariden spezialisiert haben. Sie finden sich in Bodenproben, als Teil der menschlichen Darmflora, oder eben auch als bedeutende Begleiter von Algenblüten.

*Bacteroidota* (und in marinen Systemen hauptsächlich die zu ihnen gehörenden Flavobakterien) besitzen zum Abbau diverser Polysaccharide sogenannte *Polysaccharide utilization loci* (PULs), genomische Inseln, die alle notwendigen Proteine zur Aufnahme und Abbau eines bestimmten Polysaccharids codieren. Hierzu gehören hochspezifische Enzyme (*Carbohydrate-active enzymes,* CAZymes), transkriptionelle Regulatoren sowie Transportersysteme, die initial gespaltene Oligosaccharide über die Membran in das Bakterium transportieren, wo sie von weiteren Enzymen vollständig abgebaut werden. Diese Co-Lokalisation der benötigten Gene und deren gemeinsame Regulation stellt einen enormen Selektionsvorteil der *Bacteroidota* dar und ist der Grund, warum sie, ähnlich wie Algen, einer jährlich wiederkehrenden Sukzession folgen, die sich gut untersuchen lässt.

Die Forschungsartikel, die Teil dieser Doktorarbeit sind, untersuchen das Zusammenspiel von Polysaccharid-produzierenden Algen mit den Bakterien, die sie abbauen, aber auch darauf basierende Beziehungen der Bakterien untereinander. Die erste Publikation beschäftigt sich mit dem weit verbreiteten Speicherpolysaccharid α-Glucan, für das der Großteil der blütenbegleitenden Bakterien einen spezifischen aktiven PUL besitzt. Eine Untersuchung der in der Blüte vorhandenen Algenarten bestätigte, dass die Blüte von β-Glucan-produzierenden Algen dominiert wird. Da Bakterien aber selbst α-Glucane als Speicherpolysaccharide verwenden, konnte gezeigt werden, dass nicht die Algen selbst, sondern die Bakterien Hauptproduzent dieser Polysaccharide während einer Phytoplanktonblüte sind. Bakterielle Proteine, die dem Abbau von Algen-β-Glucan und dem daraus folgenden Aufbau von bakteriellem α-Glucan dienen, waren in Umweltproben und in Laborkulturen unter ähnlichen Bedingungen abundant. Die Untersuchung von extrahiertem bakteriellem Polysaccharid bewies, dass dieses nicht nur α-Glucan enthält, sondern dass dieses Polysaccharid auch in der Lage war, α-Glucan PULs mariner Bakterien zu induzieren. Hier zeigte sich ein innerhalb des marinen Kohlenstoffkreislaufs bisher wenig berücksichtigter Kreislauf, indem Bakterien Polysaccharide anderer Bakterien nutzen, die z.B. durch Viren lysiert wurden.

Die anderen zwei Artikel dieser Arbeit befassen sich mit dem Abbau von Zellwandpolysacchariden durch blütenassoziierte Modellbakterien. In einer der Studien wird detailliert der Abbau eines β-Mannans (ein Polysaccharid das hauptsächlich aus dem Monosaccharid Mannose besteht) durch ein Bakterium des Genus *Muricauda* beschrieben. Die PUL-Struktur dieses Bakteriums kam in mehreren anderen Phytoplanktonblüten-assoziierten Bakterien vor. Diese Beobachtung wies darauf hin, dass es sich hier um ein Mannan mit zusätzlichen Galactose- und Glucose-Substitutionen handelte. Proteom-Untersuchungen bestätigten, dass das Bakterium derartige Substrate unter Induktion des β-Mannan-PULs nutzen können. β-Mannan konnte durch Antikörpermarkierung in Blütenproben sowie spezifischen Mikroalgenarten (*Chaetoceros*, *Coscinodiscus*) nachgewiesen werden. Die in dieser Publikation charakterisieren β-Mannan-PUL-codierten Enzyme waren in der Lage, dieses Signal zu löschen, was bewies, dass *Muricauda* sp. Mannan-basierte Zellwandpolysaccharide bestimmter Arten von Mikroalgen abbauen kann.

Die dritte Studie geht näher auf den Abbau von Xylanen (bestehend aus Xylose) durch ein blütenassoziiertes Bakterium des Genus *Flavimarina* ein. In diesem Bakterium wurden anhand der enthaltenen Xylanasen zwei putative Xylan-PULs annotiert. Wachstumsexperimente und Proteom-Untersuchungen zeigten, dass einer dieser PULs hauptsächlich bei Wachstum auf Glucoronoxylan induziert wird, während der andere PUL auf

Arabinoxylane stärker reagierte. Untersuchung der PUL-CAZymes bestätigte diese Ergebnisse durch Charakterisierung mehrerer Xylanasen sowie Glucoronidasen und Arabinofuranosidasen. Zusätzlich codierten beide PULs für Esterasen, die eine Modifikation der natürlichen Substrate durch Acetylierungen oder Methylierungen nahelegen. Da all diese Merkmale von terrestrischen Xylanen geteilt werden und in Blütenproben aus Küstennahen Regionen Xylane nachgewiesen wurden, ist es möglich, dass Bakterien aus solchen Regionen sowohl Xylane terrestrischen Ursprungs (z.B. durch Flusseinspeisung) sowie marinen Ursprungs abbauen können.

# Contents

## List of commonly used abbreviations

| | |
|---|---|
| AA | auxiliary activity |
| ANTS | 8-aminonaphthtalene-1,3,6-trisulfonic acid |
| BX | beechwood xylan |
| CAZy | carbohydrate-active-enzymes database |
| CAZyme | carbohydrate-active enzyme |
| CBM | carbohydrate binding module |
| CE | carbohydrate esterase |
| CPX | *Caulerpa prolifera* xylan |
| DMS | dimethyl sulphide |
| DMSP | dimethylsulfoniopropionate |
| DNS | di-nitro salicylic acid |
| DOM | sissolved organic matter |
| FACE | flurophore-assisted carbohydrate electrophoresis |
| FISH | fluorescence *in-situ* hybridization |
| GH | glycoside hydrolase |
| Glc | glucose |
| GT | glycosyl transferase |
| HAB | harmful algal bloom |
| HMWDOM | high-molecular-weight dissolved organic matter |
| HPAEC-PAD | High-performance anionic exchange chromatography with pulsed amperometric detection |
| HPLC | high-performance liquid chromatography |
| iBAQ | identity-based absolute quantification |
| LPMO | lytic polysaccharide monooxygenase |
| MAG | metagenome-assembled genome |
| ORF | open reading frame |
| PL | polysaccharide lyase |
| POM | particulate organic matter |
| PPX | *Palmaria palmata* xylan |
| PUL | polysaccharide utilization locus |
| RAX | rhy arabinoxylan |
| riBAQ | relative identity-based absolute quantification |
| RMSD | root mean square deviation |

| | |
|---|---|
| SGBP | surface glycan binding protein |
| Sus | Starch utilization system |
| TBDT | TonB-dependent transporter |
| TCC | total cell count |
| TPM | transcripts per million |
| WAX | wheat arabinoxylan |

# Part I

# Background & Summary of Research Articles

# 1. Background and Summary

Around 70 % of the earth's surface is covered with water. Within, marine phytoplankton accounts for roughly half of the world's primary production through carbon fixation [1, 2], producing enormous amounts of biomass. Much of this biomass is formed in coastal near-surface regions during biotic and abiotic factor-driven rapid increases in phytoplankton population size over a relatively short time - algal blooms. These blooms grow so large that they are visible from space (**Figure 1.1**), and their extent and intensity are monitored by earth observation satellites such as the European Space Agency's Sentinel 3 mission within the Copernicus program [3]. Disproportionate to representing just 7% of the oceans' area [4], it is estimated that around 19% of global carbon fixation occurs in coastal regions [1], underlining their importance to this process. Algae in coastal regions also benefit from nutrient input via rivers, especially rich in nitrogen and phosphorous [5–7]. It becomes clear that while algal blooms have been studied for over 100 years and are naturally occurring phenomena, there is a significant anthropogenic impact over the past decades causing them to become larger and longer-lasting. Agricultural waste entering coastal areas through river runoff significantly alters the available nutrient levels that would otherwise become limiting [8] support higher abundances of phytoplankton. These global processes are further boosted by climate change driven warming of the ocean surface water [9].

Such algae blooms either directly or indirectly cause harm to the ecosystem (Reviewed in [10]). Classed as harmful algal blooms (HAB), these events are characterized by species that e.g. produce large amounts of toxins that can harm fish or accumulate in shellfish [11], making them unfit for human consumption. Furthermore, the industrial farming of shellfish and fish in coastal waters has taken more and more prevalence around the world in order to meet higher demands while combating overfishing. Such farms naturally operate with higher densities of fish or shellfish than would naturally occur which in turn produce more waste that promote algal blooms. The algae can kill fish by producing mucus that clogs their gills [12] or pierces them with parts of their exoskeleton [13]. Generally harmless species can also become problematic when they bloom to such extents that their inevitable decay and degradation by associated heterotrophic bacteria causes oxygen depletion which can indiscriminately kill fish and invertebrates.

These factors make it obvious that algal blooms and their impacts on ecosystems and industries have global relevance and represent a challenge. Simultaneously, they offer many chances if they can be used rather than just having to be mitigated. Algae produce a wide

variety of different biomolecules that may be of biotechnological interest, some of the most diverse being polysaccharides which can make up over 50 % of the algae's dry biomass. These polysaccharides have been shown to have a wide variety of functions as storage [14] or structural [15, 16] components.

**Figure 1.1 Satellite data helps visualize algal blooms**. Integrated Sentinel 3 data (2017) of the North Sea. Blooming microalgae appear light blue or green. (Copyright: European Space Agency)

Unlike other bio-macromolecules like nucleic acids, which are in total comprised of a limited number of different monomers and always exhibit the same linkage types, polysaccharides vary not only in the array of monosaccharides that they may be comprised of. Additionally, they generate structural and functional diversity through various linkage types and the addition of modifications. It becomes clear that, while the applicable potential of the wide variety of marine polysaccharides may be high, they are difficult to study simply by their enormous variety.

This is where heterotrophic marine bacteria fit into the equation. Tightly associated with algal blooms, members of the phylum *Bacteroidota* have specialized themselves as primary polysaccharide degraders. They allow us to identify relevant polysaccharides by studying their degradation mechanisms, achieving depolymerization of what can be incredibly

complex structures. The research articles that are a part of this thesis attempt to bridge the gap between identification of key polysaccharides and the understanding of where their relevance comes from. We address complex polysaccharides such as mannans and xylans (comprised mainly of the monosaccharides mannose and xylose, respectively) that require highly specialized protein machineries to degrade. Only bacteria equipped with a dedicated array of enzymes are able to degrade them, giving such organisms a decided advantage as bloom responders. On the other hand, we also discuss a comparatively simple storage polysaccharide in α-glucan that can be degraded by most of the bloom-associated bacteria but for which we were previously lacking a definitive source. Together, these results further our understanding of the intricate interplay between blooming microalgae and the bacteria that degrade them.

## 1.1. The marine carbon cycle

Most of the carbon (~94%) in the oceans is dissolved inorganic carbon [17]. This reservoir largely consists of carbonic acid, bicarbonate and carbonate alongside dissolved carbon dioxide [18] and is the basis for all carbon fixation by marine phototrophs (mainly photosynthetic eukaryotes and cyanobacteria). Via photosynthesis, these organisms convert inorganic to organic carbon as dissolved or particulate organic matter [19]. This flux of carbon is what forms the basis of marine food webs. Organic matter can be transferred to higher trophic levels via grazing by zooplankton to larger eukaryotes like fish. It can also form longer-lived organic matter that does not remain in the surface ocean but rather sinks. Particles like this are generally referred to as “marine snow” due to their small size and sinking characteristics and consist of inorganic matter and high molecular weight organic matter like polysaccharides and proteins (reviewed in [20]). On their way down, they are partially degraded [21] but some percentage sinks out to the ocean floor where it stays, removed from the carbon cycle.

Back in the surface water, about 10-50% of the carbon fixed by phytoplankton is released as dissolved organic matter [22]. This can be due to active processes but also factors such as viral lysis [23] or senescence [24] and forms the main food source for marine heterotrophic bacteria. Due to their small size in comparison to their surface area, the uptake of dissolved organic matter is as beneficial to them as it is to no other group of organisms and so they have specialized themselves on it. From this nutrient source, they in turn produce their own organic matter, which, due to its larger size is now again available for grazing by higher

trophic levels (The marine carbon cycle is summarized in **Fig. 1.2**). This makes marine heterotrophic bacteria a focal point in the marine carbon cycle without which large amounts of dissolved organic matter would go unutilized. Their part in the oceans' carbon cycle is therefore termed the microbial loop [25].

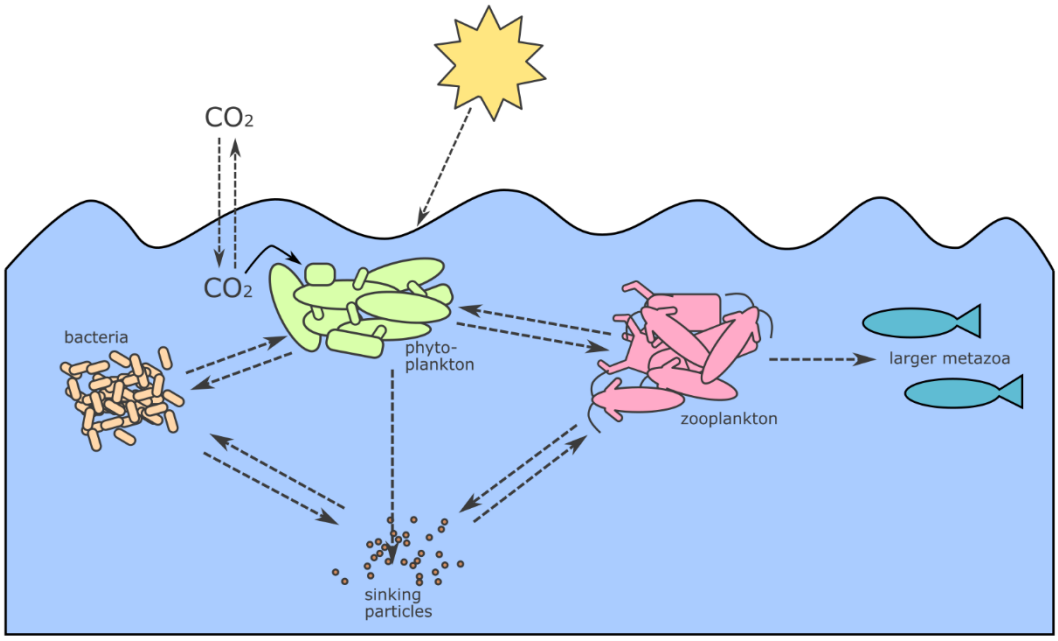



**Figure 1.2 The marine carbon cycle.** Inorganic carbon is fixed by phototrophic eukaryotes and cyanobacteria (green) and converted to organic carbon. Dissolved organic matter released via viral lysis or senescence serves as food source for heterotrophic marine bacteria (tan). Both, bacteria and marine autotrophs are grazed on by zooplankton (red) which in turn are fed on by larger metazoa (blue). Some organic matter of high molecular weight forms aggregates that sink trough the water column (brown). Depending on size, they can be a nutrient source for bacteria as well as zooplankton. The fraction that is not remineralized during descent sinks to the bottom of the ocean and is removed from the carbon cycle. Arrows signify relationships between different groups. Simplified from [26].

## 1.2. Algal blooms

In order to describe bloom dynamics, one must first understand where they come from. As a general rule, algal blooms are driven by an imbalance between the division and loss rates of phytoplankton. Abiotic factors such as light intensity and water temperature play an important role in influencing this balance and higher values for both will generally favor a bloom. Extensive mixing of the water column by storms such as often seen in winter will lead

to higher nutrient influx from deeper waters but will also potentially disperse phytoplankton below the photic zone where it can no longer grow [27, 28]. When water stratifies again in early spring, phytoplankton thus benefits from these nutrients as well as higher temperatures and better light conditions, forming widely observed spring blooms all over the world until nutrients become limiting [29].

When looking at abiotic factors alone, one might assume a bloom could occur whenever light and temperature are favorable and would last until nutrients ran out, at which point they would gradually decline. However, that does not reflect what is observed in nature. Re-occurring annual blooms normally form rather quickly, but also quickly recede once they reach a maximum. This can only be explained when looking at abiotic and biotic factors combined. Biotic factors such as zooplankton grazing or viral lysis exert significant pressure on phytoplankton communities by increasing their mortality rate [30]. Their impact cannot be overlooked, as in winter, through the aforementioned mixing of the water column, phytoplankton and their zooplankton predators both become dispersed and therefore less likely to meet – they are de-coupled [31]. This allows phytoplankton to assume net positive growth rates even in winter where abiotic factors may be less favorable, simply by a significantly lower loss rate through grazing. Thus, spring blooms are initiated much earlier than they can be readily observed [32]. When abiotic conditions become more favorable, this head-start leads to massive blooms of phytoplankton before the zooplankton catches up, effectively ending the bloom by significantly increasing the mortality rate.

The aforementioned factors are the reason that seasonal algal blooms are relatively predictable in both the area and time frame they occur in. This makes them excellent for scientific study aimed at understanding the dynamics not only of the algae themselves but their interplay with their surrounding biotic and abiotic factors.

### 1.2.1. Algal organic matter

Algae produce a wide variety of different particulate and dissolved organic matter of which 10-20 % are estimated to be released into the water column in coastal systems [33, 34]. This can be due to passive leakage [35] but also part of active processes. For example, algae may exude the osmolyte dimethylsulfoniopropionate (DMSP), which can constitute up to 10% of the cells total carbon, to keep their redox balance [36, 37]. Additionally, this compound can act as a defense mechanism, deterring protozoan herbivores by releasing dimethyl sulphide (DMS) from DMSP upon lysis [38, 39]. Algae also release glycolate as a waste product of

photosynthesis, which can be detected at low micromolar levels in sea water samples to serve as an estimate for active photorespiration [40, 41].

Lysis through grazing, viral infection or senescence inevitably releases algal intracellular organic matter into the water column. The bulk of this algal organic matter is made up of protein and polysaccharide [42, 43]. While proteins fulfill a wide variety of cellular functions including catalytic activity, transport and structural properties, they are not as structurally diverse as polysaccharides. Made up of a plethora of different monosaccharides, linkages and modifications, polysaccharides can have many functions ranging from structural cell-wall components to energy storage and parts of the extracellular matrix [44, 14, 45]. This diversity is due to their structural heterogeneity. Where proteins are exclusively made up of amino acids linked by peptide bonds and rarely display extended repeats of the same monomers, polysaccharides most often do exactly that. Most polysaccharides are made up of one or few main monosaccharides. These monomers, usually five- or six-carbon ring structures, may have saccharide bonds at various or multiple positions. Together with their, and in turn their bonds' stereochemical variability, polysaccharides display immense variability [46]. Additionally, many polysaccharides carry functional groups such as acetyl- [47] or sulfate-groups [48, 49], further diversifying their potential chemical properties and the mechanisms needed to degrade them, as will be detailed later. Which monosaccharides coupled by which main linkage type are chosen greatly impact the resulting polysaccharide's chemical and structural properties. For example, storage glucans such as glycogen [50], starch [51] or laminarin [52] consist solely of glucose with few specific linkage types, making them easy to synthesize and degrade. However, there are also incredibly complex polysaccharides such as fucose-containing cell wall polysaccharides (reviewed in [53]) that are very hard to degrade and may require over 100 enzymes to fully remineralize [54].

With algal organic matter being as complex as it is, a variety of bacteria have specialized themselves on its degradation. DMSP can activate chemotaxis in bacteria, leading them towards algae [55] and an estimated 59% of marine bacteria code for the necessary enzyme to degrade it, showing that a majority of marine bacteria utilize this compound [56]. Likewise, proteins can be degraded and utilized by employing peptidases cleaving the peptide bond, again making this compound accessible to bacteria. The previously discussed structural variability of polysaccharides makes their degradation far less simple. More importantly, the variety of monomers, linkages and modifications makes it mandatory to have dedicated cleavage and uptake machinery for every single different polysaccharide. For simpler substrates like storage glucans, such machineries exist in different marine bacterial clades

(e.g. *Alphaproteobacteria* & *Gammaproteobacteria*) but are by far most sophisticated and abundant in members of the *Bacteroidota*. Enzymes for the degradation of more complex polysaccharides such as mannans are almost exclusively found in this group and their specialized degradation mechanism allows members of this phylum to be main algal bloom responders.

## 1.3. The *Bacteroidota*

Members of the phylum *Bacteroidota* fill the niche of specialized degraders of high molecular weight organic matter in every environment they appear in and are abundant in a multitude of habitats. They make up around 10% of the microbial community in soils [57] including fields [58] and greenhouses [59] and have been shown to possess the degradation pathways for polysaccharide substrates such as cellulose, glucans and chitin [60, 61]. In the gut, they have been extensively studied for their role in food glycan utilization [62, 63]. Together with bacteria of the phylum *Firmicutes*, they make up >98% of all detectable 16S rRNA sequences in the gut of mammals [64]. Their presence here has major health implications as they produce butyrate that has been shown to reduce number and size of aberrant crypt foci in the colon that have been linked to the development of colon cancer [65, 66]. *Bacteroidota* have also been detected in polar snow and freshwater samples where they make up significant parts of the population [67] and multiple isolates from freshwater exist [68, 69].
In the oceans, *Bacteroidota* constitute one of the main phyla associated with marine snow and can even be the most abundant group detected attached to these particles [70]. Their prevalence here underlines their role as degraders of complex organic matter, especially polysaccharides, that make up large fractions of the marine snow particles [71]. Similarly, they appear wherever large amounts of organic matter are released over short periods of time, such as during seasonal algal blooms. Here, members of the *Flavobacteriia* within the *Bacteroidota* have been shown to continuously follow peaks in algal abundances [72] and are directly associated with the appearance of phytoplankton [73].

### 1.3.1. Polysaccharide utilization – the PUL advantage

As mentioned above, many polysaccharides require tailor-made degradation machineries to address the different linkages, monosaccharides and modifications they consist of. The major advantage that allows the *Bacteroidota* to utilize increasingly complex polysaccharide

substrates lies within their genome composition. The genes coding for the necessary carbohydrate-active enzymes (CAZymes) are often grouped together with their respective uptake and regulation machinery. Such genomic islands solely dedicated to the utilization of one polysaccharide are called polysaccharide utilization loci (PULs, first introduced by Xu *et al.* [74]). This means as soon as a specific polysaccharide is sensed, the entire degradation machinery can be activated as one. It is especially important that only the enzymes for a nearby substrate are produced as it is a major investment of energy to produce large numbers of enzymes and secrete or imbed them in the outer membrane [75, 76]. This tight regulation is what allows some members of the *Flavobacteriia* to encode large numbers of PULs dedicated to many different substrates such as the macroalgae-associated strains *Formosa agariphila* KMM3901, which harbors 13 different PULs [77] or *Zobellia galactanivorans* Dsij$^{T}$, that encodes as much as 50 PULs [78].

The first PUL to be functionally characterized was the *Sus*-operon of the gut bacterium *Bacteroides thetaiotaomicron* (summarized in **Fig. 1.3**). The Sus herein stands for "starch utilization system" and some of the proteins characterized from this system have gone on to lend their name to proteins of similar functions encoded in other PULs [79, 80]. Of the eight genes comprising the Sus (*susRABCDEFG*), the ones most often used to describe functional and structural homology in other PULs are *susC* and *susD*. These genes code for the transporter system shuttling specific oligosaccharides through the outer membrane into the periplasm. SusC is the actual transporter integrated in the outer membrane belonging to the TonB-dependent transporters (TBDTs), a group of active transporters also facilitating the translocation of siderophores, vitamin $B_{12}$ and nickel chelates [81]. Structurally, it is a 22-stranded β-barrel with an additional globular plug-domain that, as the name suggests, blocks the formed pore until the right substrate is recognized (reviewed in [82]). Upon binding, the plug undergoes a conformational change, exposing its conserved TonB-box to the periplasm where it can interact with TonB, initiating transport.

SusD is a lipid-anchored surface glycan-binding protein (SGBP) that forms a complex with SusC, sitting above it like a lid. The lipid anchor hereby serves as a pivot point [84] upon recognition and binding of the correct substrate. It therefore aids uptake by shuttling oligosaccharides towards SusC. Both, SusC- and SusD-like proteins have structurally well-conserved domains that allow easy annotation, making their genes hallmarks for the identification of PULs within a given genome. Their substrate specificity is derived from modifications at the binding sites. The original SusD from *B. thetaiotaomicron* binds malto-oligosaccharides derived from starch via a single binding site recognizing the amylose helices

[85] but other SusD-like with a structurally distinct binding site in similar position have been identified. SusD-like from gut *Bacteroidota* bind substrates like xyloglucan [86] and yeast mannan [62]. A SusD-like binding the algal storage polysaccharide laminarin by a planktonic *Flavobacterium* was also described, once again underlining the prevalence of these systems in multiple habitats [87].

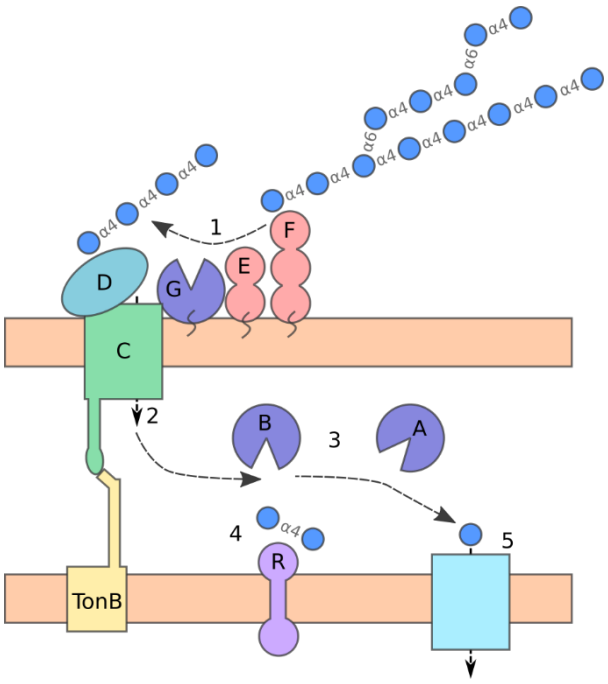



**Figure 1.3. Degradation of starch by the starch utilization system of *B. thetaiotaomicron*.** Starch is recognized by SusE/F enzymes and shuttled towards SusC/D with initial degradation by SusG (**1**). Produced oligosaccharides are transported into the periplasm by the SusC/D complex (**2**). They are further degraded to maltose and glucose by SusA/B (**3**). Maltose is recognized by the regulator SusR, initiating PUL-upregulation (**4**). Glucose units are transported into the cytosol and used for fermentation (**5**). Adapted from Foley *et al*. [83].

Recognition of the right oligosaccharides is often not only facilitated by SusC/D. In the *B. thetaiotaomicron* Sus, two additional SGBPs were described: SusE and SusF. Both are lipoproteins like SusD but are structurally much less conserved in other PULs. They carry multiple carbohydrate binding modules (CBMs) that are structures usually attached to CAZymes in order to enhance the accessibility of the substrate [88], but neither SusE nor SusF have catalytic activity. As both proteins bind larger malto-oligosaccharides as well as starch, it is assumed that they increase the availability of the substrate at the cell surface by binding

and bringing it into contact with extracellular CAZymes as well as SusC/D [89]. It has been observed that in contrast to the other surface-associated proteins within the Sus, SusE and SusF are not mobile at the surface level and seem to group the other proteins around them. This indicates a role for both proteins in forming the translocation complex, aiding oligosaccharide uptake [90]. In contrast to SusC and SusD, that are structurally well-conserved within the *Bacteroidota*, SusE and SusF only appear in PULs dedicated to starch/α-glucan uptake. Other PULs with different target substrates have been shown to also encode SGBPs with similar functions such as a *B. thetaiotaomicron* lipoprotein binding fructan [91] or *Bacteroides ovatus* SGBPs binding xylan or β-mannan [89]. It is therefore likely that, although no strong sequence conservation exists between SGBPs with similar functions as SusE and SusF, many PULs do encode for additional SGBPs to aid in polysaccharide recognition and capture.

The oligosaccharides transported into the periplasm are further degraded and eventually shuttled into the cytoplasm either as di- or monosaccharides by dedicated transporters that can be but are not necessarily encoded within the PUL [63]. However, the oligosaccharides within the intermembrane space fulfill another important function. At this stage, oligosaccharides derived from initial degradation of a specific polysaccharide are recognized by sensor proteins that act as transcriptional regulators. In the classic case of the Sus, this is facilitated by SusR. Specific for linkage and monosaccharide, this membrane-spanning regulator recognizes maltose and larger malto-oligosaccharides. Upon binding, the entire *Sus* regulon excluding *susR* is upregulated, resulting in the production of high levels of proteins dedicated to the degradation of starch [92, 93]. PULs generally operate on base levels of expression, allowing the signal oligosaccharide to be cleaved from the recognized polysaccharide and imported into the periplasm while at the same time conserving energy by only mass-producing enzymes for recognized substrates [94]. A mechanism like this is a wide-spread feature of PULs, even if the regulator is not necessarily a SusR-homolog. Other regulator types include classical or hybrid two-component systems or, as is the case for the alginate PUL of *Z. galactanivorans* Dsij$^{T}$, a GntR-type regulator [95].

For the actual degradation of the polysaccharide, the Sus encodes three CAZymes. Two of them, SusA and SusB, are located within the periplasm and degrade the imported oligosaccharides into smaller di- and monosaccharides [93]. A third CAZyme, SusG, is attached to the outer membrane and associated with SusC/D/E/F [96]. The attachment of CAZymes to the outer membrane to facilitate initial degradation is another PUL-hallmark. Most polysaccharides are too large to be imported into the cell without being made bite-

sized by specific enzymes first. As many *Bacteroidota* are located within open systems such as bodies of water, simply secreting enzymes for this purpose would most likely result in them being carried away. By attaching them to the membrane, the bacteria ensure that they a) reap the full benefit of the enzyme and b) gain an additional protein that helps in glycan capture [90]. To this purpose, *Bacteroidota* have developed a unique translocation system for outer-membrane-attached proteins: the type IX secretion system. First detected in the oral pathogen *Porphyromonas gingivalis* [97, 98], they are often virulence-factor associated but proteins associated with gliding motility and many CAZymes also display a type IX secretion system target sequence [99]. The existence of this translocation system underlines the sophistication regarding the acquisition of nutrients (by activating motility towards a substrate or its degradation by CAZymes) employed by the *Bacteroidota*.

### 1.3.2. Carbohydrate-active enzymes: tailor-made

Equivalent to the enormous variety of polysaccharides, enzymes targeting these structures are just as diverse. Five different groups of CAZymes are defined: glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and enzymes with auxiliary activity (AA). They are classed in the carbohydrate active enzymes database (CAZy) according to their sequence and, by extension, function [100]. Additionally, the CAZy database also lists CBMs that bind polysaccharides without catalytic activity of their own. A new family is added whenever a new enzymatic reaction is biochemically described, and homologs of this enzyme are then added [101]. Families may also have subfamilies, where evolutionary diversification has led to enzymes with different functions grouping within the same family and small differences lead to a diversified substrate spectrum.

Glycoside hydrolases form the largest group of CAZymes. They hydrolyze the glycosidic bond between two carbohydrates and have a wide variety of substrates. The CAZy database currently (2023) lists 184 distinct GH families. Aside from their substrate specificity, they can be classed by different means. Firstly, a GH can either be specific to cleaving off residues at the (often but not always, non-reducing) end of existing polysaccharide chains (*exo*-enzymes) or cleave a glycosidic bond within the chain (*endo*-enzymes) [102]. Both types are required for the complete depolymerization of a polysaccharide, as *endo*-enzymes typically do not accept very small substrates, even if they consist of the correct linkage type and monosaccharide (e.g. [103]). Additionally, GHs can by classed by the mechanism they employ for cleavage. As carbohydrates exists as half-acetal rings, they form anomers defined by the

carbon sitting next to the oxygen-atom within the ring (the anomeric carbon). Depending on the position of the hydroxyl-group at this carbon, a monosaccharide is either in α- or β-conformation. Based on this distinction, a glycosidic bond at this position is also either an α- or a β-bond, leading to differential secondary structures within the oligo- or polysaccharide. GHs cleaving this bond can either be retaining, meaning they keep the O-bearing group at the anomeric carbon in the position it started out in, or inverting, meaning they convert it from α- to β-conformation or *vice versa* [104]. The active center of GHs is generally quite simple and usually requires only two active residues, a general acid and a general base, often a glutamate and an aspartate [105]. Substrate specificity is mediated by other residues within the active site, making it so that only the preferred poly- or oligosaccharide can enter. A relatively small modification at the wrong site or a different monosaccharide can already lead to rejection of the substrate [106]. GHs using alternative cleavage mechanisms have also been described. Enzymes belonging to the GH33 and GH34 families, sialidases and trans-sialidases, respectively, as well as GH143 (2-keto-3-deoxy-D-lyxo-heptulosaric acid hydrolases) are known to use a tyrosine as nucleophile [107]. A different mechanism that involves NAD as cofactor is also known from enzymes of families GH4, GH109, GH177 and GH179 [108, 109].

Enzymes catalyzing the formation of the glycosidic bond rather than its cleaving are called glycosyl transferases. These can be used by polysaccharide-degrading bacteria in order to build their own storage or structural polysaccharides from the ones they have taken up and degraded, for example the formation of storage glycogen from glucose [110]. GTs usually require sugar phosphates as glycosyl donors and transfer the glycosyl group to an alcohol, forming an O-, N-, S-, or C- glycosidic bond [111]. Much like GHs, GTs can be either retaining or inverting, providing linkage specificity to the produced sugar.

Polysaccharide lyases use a different mechanism for cleaving glycosidic bonds than GHs. Instead of hydrolysis, these enzymes use lytic β-elimination, a process that can only occur when there is an acidic group bound to the carbon next to the carbon forming the glycosidic bond [112]. PLs are therefore specific to polysaccharides containing uronic acids. These include alginate, which mainly consists of guluronic and mannuronic acid [113], pectin made from galacturonic acid [114] or ulvan, that includes both, glucuronic and iduronic acid [115]. Due to the limited number of substrates, PLs are currently (2023) only differentiated into 42 families. During cleavage, the abstracted proton and the bridging oxygen can either be on the same or opposite sides of the uronic acid ring, therefore named *syn*- or *anti*-mechanisms, respectively. These mechanisms are the only ones currently classed as PLs [116].

The enzymes discussed above focus on the cleavage or formation of glycosidic bonds that form the backbone of all polysaccharides. In contrast, carbohydrate esterases and enzymes classed as such with auxiliary activity both focus on removing modifications from the main polysaccharide chain, thus enabling degradation by GHs and PLs. Carbohydrate esterases remove *O*- or *N*-acetylations from specific poly- or oligosaccharides. This is especially utilized for cell wall polymers that often carry such modifications in order to reduce solubility and cross-link the backbones of multiple chains, increasing stability [117, 118]. As an ester is cleaved into an acid and an alcohol, two classes of esterases are defined. In polysaccharides including modifications attached to an uronic acid-containing backbone, the sugar is the acid. For neutral polysaccharides, such as xylan, the sugar acts as the alcohol. The cleavage of acetyl- or methyl-groups robs the polysaccharide of its crosslinking properties, making it more susceptible to GHs. The nature of CEs as the first line of attack on an insoluble polysaccharide makes it clear why many of them, such as all members of the CE2 family, also code for CBMs in order to increase binding efficiency to a very inaccessible substrate [119].

In marine environments, sulfatases also play a very important role in polysaccharide degradation (reviewed in [120]). Sulfatases are also esterases but, due to their relevance and number, they have been classed as their own group with four different families [121]. Uncommon in land plants, excessive sulfation of the polysaccharide backbone is thought to be a specific adaptation to marine habitats. In macroalgae, sulfation can account for up to 40% of the entire polysaccharide's dry weight [122, 123], making it clear that sulfatases need to be employed in order to degrade them efficiently. Sulfatases are active on a multitude of different polymers such as carragenans, fucoidans and ulvans but their individual substrate spectrum is highly specific, usually even requiring different enzymes to cleave sulfations at different positions of the same oligosaccharide (e.g. [124–127]). In addition to a large number of GHs or PLs, PULs of marine bacteria therefore often also encode specific sulfatases in order to address the high sulfation level of marine polysaccharides [128].

Lastly, enzymes with auxiliary activity are a group of redox enzymes with ligninolytic or monooxygenase-activity. Ligninolytic enzymes are technically not active on the polysaccharides themselves but rather lignin, the main non-polysaccharide component of plant cell walls that is secondarily embedded into the polysaccharide scaffold, providing more stability (reviewed in [129]). It is made up of different phenolic compounds which further hydrophobicity and therefore insolubility. Additionally, lignin can be linked to the cellulose-backbone and is then called ligno-cellulose. An efficient degradation of plant cell walls thus requires the presence of ligninolytic enzymes. In contrast to highly specific CAZymes,

ligninolytic enzymes are usually unspecific and degrade lignin structures by generating free radicals that indiscriminately break the lignin bonds [130].

Polysaccharide degradation is also aided by lytic polysaccharide monooxygenases (LPMOs) such as those originally classed as GH61. Rather than hydrolysis, these enzymes use low molecular weight compounds such as ascorbate or gallate as reducing agents for the cleavage of cellulose [131]. This is beneficial as it doesn't require the polysaccharide to be in solution or even in isolated strands. It can be degraded directly in its natural crystalline structure. The initial degradation by LPMOs enables other enzymes to then further process the freed oligosaccharides [132].

Both, ligninolytic enzymes and LPMOs are mostly found in degradation mechanisms targeting cell walls of terrestrial plants. However, some significant marine examples of AA enzymes exist. Marine *Flavobacteria* capable of degrading the polysaccharide porphyran encode a zinc-dependent alcohol dehydrogenase that seemingly aids in the utilization of the porphyran component 6-O-methly-D-galactose as deletion of the gene led to growth defects on this substrate [133]. Similarly, a P450 monooxygenase was found to catalyze the demethylation of this compound [134]. These findings expand the already enormous repertoire of enzymes specifically tailored towards the degradation of highly complex polysaccharides.

## 1.4. Interpreting diversity – how meta-omics inform what to study

Cell counting performed on environmental samples via e.g. fluorescence *in-situ* hybridization (FISH) [135] can tell us which bacteria are not only part of the response to an algae bloom, but time-sampling enables us to know which bacteria are most abundant during which phases of the bloom. This information in crucial to our understanding of the microbial community during bloom events and likewise can give us a foundation of what needs to be studied. For example, as discussed before, the *Flavobacteriia* are most abundant during times in which the algae reach their peak abundances. Genomes obtained from isolates of these bacteria and experiments regarding their substrate spectrum show us that they are highly-skilled degraders of high molecular weight organic matter [136]. However, the ability to study specific isolates derives from our ability to successfully isolate them. Most bacteria are not easily cultivatable in the lab, thus making it very likely that working with isolates alone introduces a bias that doesn't accurately depict the microbial community investigated [137].

FISH can help guide us in the right direction by pre-selecting specific taxa that are highly abundant but can't tell us what actually makes these bacteria so successful.

This is where metagenomics step in. Sequencing techniques were first developed for pure cultures of single organisms like viruses [138] and bacteria [139], allowing a new understanding of how complex eukaryotic cells were initially formed and even how organelles of both eu- and prokaryotes were seemingly derived from endosymbiosis [140, 141]. This led to the discovery of the significance of the 16S rDNA as a phylogenetic tool [142], as is used in FISH to flag bacteria belonging to different taxa. The development of newer sequencing techniques with higher accuracy and throughput (reviewed in [143]) allowed the expansion of genomics towards not only the study of a single organism in pure culture but also that of entire communities. Where a genome is the entire genetic information of a single organism, a metagenome encompasses that of an entire environmental sample. Importantly, these samples can be processed more or less directly without the need for lab cultivation. This removes the cultivation bias and allows a clear view of which bacteria inhabit distinct ecosystems and, with time sampling, how this community changes over time or under specific conditions. It also provides a higher resolution image than FISH, in which by definition we can only identify what we look for by designing specific probes.

Via metagenomics, we can also see what the genetic potential of the detected organisms is. To this end, the reads generated by sequencing an entire environmental sample at once need to be assigned to single taxa or species. This is done by bioinformatic tools that generate so-called metagenome-assembled genomes (MAGs) [144–146]. A MAG is not necessarily the genome of a single bacterium within the sample, but rather contains sequences of a group of bacteria that are very closely related (e.g. belong to the same species) and likely inhabit similar niches. These are then taxonomically and functionally annotated in order to gain an understanding of what these bacteria are capable of [147]. Coming back to algae blooms, metagenomics allows an in-depth view at which bacteria accompany which parts of the bloom and what they are likely doing there. For example, *Bacteroidota*-MAGs associated with peak bloom phases usually contain a much higher number of CAZymes and protases, underlining their role as primary degraders of high molecular weight organic matter [72, 148].

Using just metagenomics, we are not able to definitively say which genes a specific MAG is actively using at any given time, just what its potential is. It does however enable the use of other -omics technologies to find out exactly that. Just as a genome is the entirety of an organism's genetic information, a transcriptome is the entirety of its RNA, especially its

mRNA serving as a template for protein translation. Mapping the reads obtained by sequencing the mRNA back to the genome tells us exactly which genes are being actively transcribed by the organism [149]. This method also works for environmental samples (meta-transcriptomics). As transcripts can only be mapped to an existing library of genomes, a high-quality metagenome is needed to generate the best results. But one can go one step further. A proteome is the entirety of an organism's proteins and likewise, a metaproteome that of a complex sample. These data are recorded by first separating the proteins of the sample by molecular weight and then digesting the proteins into analyzable peptides with a protease. Peptides are then analyzed using a mass spectrometer in order to identify them and subsequently mapped to the protein they belong to using computational tools as well as the aforementioned high-quality metagenome. As transcripts as well as proteins have varying lifespans, a look at both allows for the most comprehensive functional picture of an environment. The question of 'who is there' is already addressed by FISH and metagenomics but now, the question 'and what are they doing' can also be answered.

## 1.5. Aims of the thesis

For algal blooms, meta-omics approaches tell us which polysaccharides are actively being degraded at which time, by looking at which PULs the bacteria associated are actively expressing. In the case of annual spring blooms around the island of Helgoland in the North Sea (54°11'N 7°54'E), metagenomics showed PULs for different polysaccharides to be present, mainly such for the degradation of alginates, α- and β-glucans, fucans, xylans and mannans. By far the most abundant were genes for the TBDTs encoded in PULs dedicated to algal storage polysaccharide laminarin (a β-glucan) but when looking at genomics and proteomic combined, that lead visibly shrunk. They were still the transporters with the overall highest abundance among polysaccharide degradation-associated proteins, but at some time points, α-glucan targeting TBDTs were even more abundant, even though their abundance in the metagenomes was much lower [150]. Transcriptomics performed on a Helgoland bloom from a different year corroborated this prevalence of activity associated with the degradation of storage glucans [151]. Results like these are what enable us to go back to isolated strains matching those found to be abundant and active during a bloom and ask very specific questions. How are certain polysaccharides degraded by marine bacteria? What is their prevalence and which role do they play in marine carbon cycling? Which algae

produce them? All of which are questions that the research articles comprising this thesis aim to shed light on.

So, how do we get from meta-omics and strain libraries to elucidating the role of specific polysaccharides and their impact on the bacterial community? This thesis combines -omics technologies (metagenomics, -proteomics and -transcriptomics) with an extensive collection of bacterial isolates taken from algal blooms. Strains either belonging to taxa directly relevant to the bloom (e.g. *Polaribacter* spp., [152]) or harboring putative degradation mechanisms for relevant polysaccharides (mannans, xylans) were chosen based on previous data [148, 150, 72]. While genome annotation can lead us in the right direction as to what the role of certain CAZymes encoded within their PULs play, it is only via biochemical characterization of these enzymes that we get the full and accurate picture. This is why characterization of enzymes specific to certain polysaccharides was a main focal point of all three research articles comprising this thesis. A second was the use of single-stain proteomics to elucidate which substrates were able to elicit specific responses from our chosen model bacteria, thus shining light on which natural substrates might be degraded. All of these methods were combined with computational analysis to gain more understanding about how prevalent a certain degradation mechanism or polysaccharide is in nature, tailored to fit the questions meta-omics posed for each polysaccharide investigated.

## 1.6. Article I: Alpha-glucans from bacterial necromass indicate an intra-population loop within the marine carbon cycle

The first article is on the role of α-glucans for marine bacteria during marine phytoplankton bloom situations. Bloom-responding bacteria belonging to the *Flavobacteriia* target polysaccharides that are produced by blooming microalgae, mainly α- and β-glucans, alginates and mannose- or xylose-containing polysaccharides of which genes targeting glucans are most abundant. We found, however, that even during blooms of exclusively β-glucan-producing microalgae (diatoms) around the island of Helgoland in the German Bight in 2020, abundant flavobacteria still highly expressed their α-glucan targeting genes. Via 18S rRNA gene sequencing of the mainly eukaryotic >10 µm and 10-3 µm fractions, we could show that α-glucan producing microalgae such as *Rhodophytes* and *Chlorophytes* were indeed not abundant enough to explain this flavobacterial response, hinting that the source might not be of algal origin.

Bacteria themselves use α-glucans as storage polysaccharides and expression of α-glucan-targeting genes was highest during times where the bacteria directly responded to blooming algae. Additionally, metaproteomics of three spring bloom events (2016, 2018, 2020) showed that the relative abundance of α-glucan synthesizing proteins increased together with the response to blooming algae, pointing towards an intra-population recycling loop.

Analysis of the α-glucan polysaccharide utilization loci (PULs) of marine bacteria revealed a variety of different modularities centered around one or more GH13 α-amylases/glucosidases grouped with a *susCD* transporter pair. SusD-sequence alignment showed differences between two main groups and, together with the variety of α-glucan-targeting functions of different expressed marine GH13s, pointed towards variability in marine α-glucans.

The marine isolate *Polaribacter* sp. Hel_I_88 accumulated α-glucan when grown on laminarin and analysis of the extracted intracellular polysaccharide showed large amounts of glucose. This extract could be degraded by the previously characterized marine enzymes, proving it to contain α-glucan utilizable by marine bacteria. With this extract as sole carbon source, marine isolates were able to grow and proteomics showed high expression of the α-glucan PUL. This proved that the intracellular polysaccharide formed by marine bacteria can also be taken up and degraded by them when released into the water column, shining light in how these bacteria combat increased mortality by viral lysis or predation that greatly impacts bloom-responding bacteria reaching high cell densities.

## 1.7. Article II: Marine bacteroidetes use a conserved enzymatic cascade to digest diatom β-mannan

The second article focuses on the degradation of the polysaccharide β-mannan by specialized marine bacteria. Common in terrestrial plants but previously unknown to occur in microalgae, β-mannan structures were detected in diatom bloom samples, but the exact source remained obscure. The marine isolate *Muricauda* sp. MAR_2010_75 possesses three GH26s predicted to be β-mannanases clustered in a PUL. Comparison to other isolates with similar clusters showed rearrangements, but showed all of them to at least encode for two GH26s and a GH130 β-1,4-mannosylglucose phosphorylase. Searches in marine databases as well as metagenomes obtained at the island of Helgoland in the German Bight showed that this cluster type was prevalent and seemed to be of significance in marine habitats. As in *Muricauda* sp., many of these similar clusters encoded a GH5 and a GH27, proposed to be a

glucanase and a galactosidase, respectively. The unknown marine substrate therefore was likely to be a mannose-rich polymer also containing glucose and galactose.

Growing the strain on different β-mannan containing substrates, proteomics revealed a significant upregulation of the entire PUL compared to a culture grown on pectin. The upregulation was more pronounced on galacto- and glucomannan compared to homomannan, further indicating that these were preferred substrates.

A recombinantly produced GH26 from the PUL showed β-mannanase activity on different β-mannans, also accepting galactose modifications. The proposed glucanase GH5 was active on different β-1,4 glucans and also released glucose from glucomannan while the galactosidase GH27 released galactose from β-mannan backbones substituted with galactose, including galactomannan. This proved that both enzymes indeed assist in the degradation of complex β-mannans.

The active GH26 as well as another from the PUL were analyzed structurally. For the active mannanase, the typical TIM-barrel hydrolase fold underlined its function. For the other GH26, a loop blocking one end of the active site indicated *exo*-activity. As we could detect no activity on terrestrial substrates, we chose to test this GH26 as well as the other enzymes on bloom samples, where we assumed the natural substrate originated from. Indeed, both GH26s were able to delete or decrease the signal for β-1,4 mannan obtained in epitope deletion assays of bloom samples. Using this assay, we also tested extracts from different diatom species for the existence of β-1,4 mannan targetable by the *Muricauda* sp. enzymes. Such polysaccharides were detected in both, *Cheatoceros affinis* and *Coscinodiscus wailesii* extracts. For *C. wailesii*, we additionally isolated a strain from Helgoland waters which also proved to contain β-mannan. This allowed us to link the polysaccharide β-mannan directly from the producing algae to the degrading bacteria.

## 1.8. Article III: Marine *Bacteroidetes* enzymatically digest xylans from terrestrial plants

The third article explores the degradation of diverse xylans by marine bacteria using the model strain *Flavimarina* sp. Hel_I_48. This strain was chosen as it possesses two distinct putative xylan-targeting PULs encoding a multitude of putative GH10 xylanases and SusD surface glycan binding proteins. Growth experiments using xylans of terrestrial and algal origin showed that the strain is able to accept a number of different modifications including glucorono- and arabinoxylans. Proteomics revealed divergent expression of the two xylan PULs depending on the substrate type. While PULI mainly responded to glucuronoxylan,

proteins of PULII were more abundant on different arabinoxylans. Additionally, a third PUL encoding GH43 arabinofuranosidases likely assisting in the degradation of arabinoxylans could be identified as it was upregulated during growth on these substrates.

As SusD proteins are used for the specific recognition of extracellular polysaccharides, we assumed that the six different SusDs encoded in the *Flavimarina* sp. PULs indicated a high variability in substrate. Indeed, we could via affinity gel electrophoresis show that one SusD belonging to PULI specifically binds glucoronoxylans, corroborating the proteomics results. Sequence alignments of all six PUL-encoded SusDs gave relatively low identities with large numbers of gaps, underlining that the individual SusDs likely bind different substrates and thus contribute to the ability of *Flavimarina* sp. to degrade diverse xylans.

Biochemical characterization of PUL-encoded GH10s revealed xylanase activity on different β-1,4-linked xylans for three enzymes. Only a purely β-1,3-linked xylan extracted from the red algae *Caulerpa prolifera* was not degraded, proving specificity towards β-1,4 xylans. Combination of the catalytic activity of these xylanases with those of other PUL-encoded enzymes proved further digestion and the removal of side chains such as glucuronic acid and arabinose. We could also detect galactosidase-activity for an additional enzyme encoded in PULII, pointing towards the substrate spectrum of *Flavimarina* sp. also containing xylans modified with galactose.

Both PULs encoded multiple carbohydrate esterases, which are known to increase the accessibility of hemicelluloses such as xylan to degradation by cleaving off side groups that facilitate crosslinking. We could show esterase activity for three encoded esterases including the removal of methyl groups, *O*-acetylation and ferulic acid from model substrates.

All of the side chains and modifications shown to be targeted by the *Flavimarina* sp. enzymes occur in terrestrial plants. Comparative PUL analysis showed clusters containing similar functions to occur not only in marine habitats but also in terrestrial environments and the human gut. This allows for two possible sources of xylan substrates for marine bacteria such as *Flavimarina* sp. The substrates could be of algal origin and are structurally close to those known from terrestrial plants, or could be terrestrial biomass introduced into coastal waters via rivers. This shows the adaptability of bacteria such as *Flavimarina* sp., who have evolved to target a multitude of different xylans containing diverse modifications.

## 1.9. References


1. Field CB, Behrenfeld MJ, Randerson JT et al. (1998) Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science 281:237–240
2. Longhurst A, Sathyendranath S, Platt T et al. (1995) An estimate of global primary production in the ocean from satellite radiometer data. J Plankton Res 17:1245–1271
3. Donlon C, Berruti B, Buongiorno A et al. (2012) The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. Remote Sensing of Environment 120:37–57
4. Gattuso J-P, Frankignoulle M, Wollast R (1998) Carbon and Carbonate Metabolism in Coastal Aquatic Ecosystems. Annual Review of Ecology and Systematics 29:405–434
5. Hänninen J, Vuorinen I (2015) Riverine tot-P loading and seawater concentrations in the Baltic Sea during the 1970s to 2000—transfer function modelling based on the total runoff. Environmental Monitoring and Assessment 187:343
6. Barker JD, Sharp MJ, Fitzsimons SJ et al. (2006) Abundance and Dynamics of Dissolved Organic Carbon in Glacier Systems. Arctic, Antarctic, and Alpine Research 38:163–172
7. Nakajima T, Sugimoto R, Kusunoki T et al. (2021) Nutrient fluxes from rivers, groundwater, and the ocean into the coastal embayment along the Sanriku ria coast, Japan. Limnol Oceanogr 66:2728–2744
8. Ryther JH, Dunstan WM (1971) Nitrogen, phosphorus, and eutrophication in the coastal marine environment. Science 171:1008–1013
9. Martinez E, Raitsos DE, Antoine D (2016) Warmer, deeper, and greener mixed layers in the North Atlantic subpolar gyre over the last 50 years. Glob Change Biol 22:604–612
10. Hallegraeff GM (1993) A review of harmful algal blooms and their apparent global increase. Phycologia 32:79–99
11. Stabell OB, Pedersen K, Aune T (1993) Detection and separation of toxins accumulated by mussels during the 1988 bloom of Chrysochromulina polylepis in Norwegian coastal waters. Marine Environmental Research 36:185–196
12. MacKenzie LA, Smith KF, Rhodes LL et al. (2011) Mortalities of sea-cage salmon (*Oncorhynchus tshawytscha*) due to a bloom of *Pseudochattonella verruculosa* (*Dictyochophyceae*) in Queen Charlotte Sound, New Zealand. Harmful Algae 11:45–53
13. Rensel JE (1991) Severe blood hypoxia of Atlantic salmon (*Salmo salar*) exposed to the marine diatom Chae toceros concavicornis. Fifth International Conference on Toxic Marine Phytoplankton
14. Becker S, Tebben J, Coffinet S et al. (2020) Laminarin is a major molecule in the marine carbon cycle. Proc Natl Acad Sci U S A 117:6599–6607

15. Pauly M, Gille S, Liu L et al. (2013) Hemicellulose biosynthesis. Planta 238:627–642
16. Domozych D (2016) Algal Cell Walls. In: Lauc G, Wuhrer M (eds) High-throughput glycomics and glycoproteomics: Methods and protocols. Humana Press, New York, pp 1–11
17. P. Ciais, C. Sabine, G. Bala et al. (2014) Carbon and Other Biogeochemical Cycles. In: O. Edenhofer, R. Pichs-Madruga, Y. Sokona et al. (eds) Climate Change 2013. Cambridge University Press, United Kingdom, pp 465–570
18. Millero FJ (2007) The Marine Inorganic Carbon Cycle. Chemical Reviews 107:308–341
19. Kirchman DL, Suzuki Y, Garside C et al. (1991) High turnover rates of dissolved organic carbon during a spring phytoplankton bloom. Nature 352:612–614
20. Alldredge AL, Silver MW (1988) Characteristics, dynamics and significance of marine snow. Progress in Oceanography 20:41–82
21. Simon M, Alldredge AL, Azam F (1990) Bacterial carbon dynamics on marine snow. Marine Ecology Progress Series 65:205–211
22. Cole JJ, Findlay S, Pace ML (1988) Bacterial production in fresh and saltwater ecosystems: a cross-system overview. Marine Ecology Progress Series 43:1–10
23. Middelboe M, Brussaard CPD (2017) Marine Viruses: Key Players in Marine Ecosystems. Viruses 9
24. Beevers L (1976) 23 - Senescence. In: Bonner J, Varner JE (eds) Plant Biochemistry (Third Edition). Academic Press, San Diego, pp 771–794
25. Azam F, Fenchel T, Field JG et al. (2022) Foundations of Ecology II: The Ecological Role of Water-Column Microbes in the Sea. In: Miller TE, Travis J (eds) Classic Papers with Commentaries. University of Chicago Press, pp 384–390
26. Steinberg DK, Landry MR (2017) Zooplankton and the Ocean Carbon Cycle. Annual Review of Marine Science 9:413–444
27. Gran HH On the conditions for the production of plankton in the sea. Rapports et Proces-Verbaux des Reunions, Conseil Internationalpour l'Exoploration de la Mer 1931:37–46
28. Harvey HW (1937) The Supply of Iron to Diatoms. Journal of the Marine Biological Association of the United Kingdom 22:205–219
29. Sverdrup HU (1953) On Conditions for the Vernal Blooming of Phytoplankton. ICES J Mar Sci 18:287–295
30. Riley GA (1946) Factors controlling phytoplankton populations on Georges Bank. Journal of Marine Research 6:54–72

31. Behrenfeld MJ (2010) Abandoning Sverdrup's Critical Depth Hypothesis on phytoplankton blooms. Ecology 91:977–989
32. Boss E, Behrenfeld M (2010) In situ evaluation of the initiation of the North Atlantic phytoplankton bloom. Geophys Res Lett 37
33. Thornton DC (2014) Dissolved organic matter (DOM) release by phytoplankton in the contemporary and future ocean. European Journal of Phycology 49:20–46
34. Fogg GE (1983) The Ecological Significance of Extracellular Products of Phytoplankton Photosynthesis 26:3–14
35. Flynn KJ, Clark DR, Xue Y (2008) Modeling the Realease of Dissolved Organic Matter by Phytoplantkon (1). J Phycol 44:1171–1187
36. Stefels J, Steinke M, Turner S et al. (2007) Environmental constraints on the production and removal of the climatically active gas dimethylsulphide (DMS) and implications for ecosystem modelling. Biogeochemistry 83:245–275
37. Seymour JR, Simó R, Ahmed T et al. (2010) Chemoattraction to dimethylsulfoniopropionate throughout the marine microbial food web. Science 329:342–345
38. Strom S, Wolfe G, Slajer A et al. (2003) Chemical defense in the microplankton II: Inhibition of protist feeding by β-dimethylsulfoniopropionate (DMSP). Limnol Oceanogr 48:230–237
39. Wolfe GV, Steinke M, Kirst GO (1997) Grazing-activated chemical defence in a unicellular marine alga. Nature 387:894–897
40. Tolbert NE, Zill LP (1956) Excretion of glycolic acid by algae during photosynthesis. J Biol Chem 222:895–906
41. Leboulanger C, Descolas-Gros C, Jupin H (1994) HPLC determination of glycolic acid in seawater. An estimation of phytoplankton photorespiration in the Gulf of Lions, western Mediterranean Sea. J Plankton Res 16:897–903
42. Haug A, Myklestad S (1976) Polysaccharides of marine diatoms with special reference to *Chaetoceros* species. Marine Biology 34:217–222
43. Villacorte LO, Ekowati Y, Neu TR et al. (2015) Characterisation of algal organic matter produced by bloom-forming marine and freshwater algae. Water Res 73:216–230
44. Sverre M. Myklestad (1995) Release of extracellular products by phytoplankton with special emphasis on polysaccharides. Science of The Total Environment 165:155–164
45. Pauly M, Gille S, Liu L et al. (2013) Hemicellulose biosynthesis. Planta 238:627–642

46. Laine RA (1994) A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10(12) structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. Glycobiology 4:759–767
47. Pauly M, Ramírez V (2018) New Insights Into Wall Polysaccharide *O*-Acetylation. Frontiers in Plant Science 9
48. Qi H, Zhang Q, Zhao T et al. (2005) Antioxidant activity of different sulfate content derivatives of polysaccharide extracted from *Ulva pertusa* (*Chlorophyta*) in vitro. Int J Biol Macromol 37:195–199
49. Mourão PA, Pereira MS, Pavão MS et al. (1996) Structure and Anticoagulant Activity of a Fucosylated Chondroitin Sulfate from Echinoderm: Sulfated Fucose Branches on the Polysaccharide account for its high Anticoagulant Action*. J Biol Chem 271:23973–23984
50. Preiss J (1984) Bacterial glycogen synthesis and its regulation. Annu Rev Microbiol 38:419–458
51. Buléon A, Colonna P, Planchot V et al. (1998) Starch granules: structure and biosynthesis. Int J Biol Macromol 23:85–112
52. Biersmith A, Benner R (1998) Carbohydrates in phytoplankton and freshly produced dissolved organic matter. Marine Chemistry 63:131–144
53. Deniaud-Bouët E, Hardouin K, Potin P et al. (2017) A review about brown algal cell walls and fucose-containing sulfated polysaccharides: Cell wall context, biomedical properties and key research challenges. Carbohydr Polym 175:395–408
54. Sichert A, Corzett CH, Schechter MS et al. (2020) Verrucomicrobia use hundreds of enzymes to digest the algal polysaccharide fucoidan. Nat Microbiol 5:1026–1039
55. Seymour JR, Amin SA, Raina J-B et al. (2017) Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. Nat Microbiol 2:17065
56. Howard EC, Sun S, Biers EJ et al. (2008) Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. Environmental Microbiology 10:2397–2410
57. Fierer N (2017) Embracing the unknown: disentangling the complexities of the soil microbiome. Nature Reviews Microbiology 15:579–590
58. Borneman J, Skroch PW, O'Sullivan KM et al. (1996) Molecular microbial diversity of an agricultural soil in Wisconsin. Appl Environ Microbiol 62:1935–1943

59. Kim B-Y, Weon H-Y, Cousin S et al. (2006) *Flavobacterium daejeonense* sp. nov. and *Flavobacterium suncheonense* sp. nov., isolated from greenhouse soils in Korea. Int J Syst Evol Microbiol 56:1645–1649
60. Taillefer Marcel, Arntzen Magnus Ø., Henrissat Bernard et al. (2018) Proteomic Dissection of the Cellulolytic Machineries Used by Soil-Dwelling *Bacteroidetes*. mSystems 3:10.1128/msystems.00240-18
61. Alteio L. V., Schulz F., Seshadri R. et al. (2020) Complementary Metagenomic Approaches Improve Reconstruction of Microbial Diversity in a Forest Soil. mSystems 5:10.1128/msystems.00768-19
62. Cuskin F, Lowe EC, Temple MJ et al. (2015) Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. Nature 517:165–169
63. Larsbrink J, Rogers TE, Hemsworth GR et al. (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. Nature 506:498–502
64. Ley RE, Peterson DA, Gordon JI (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 124:837–848
65. Kim YS, Milner JA (2007) Dietary modulation of colon cancer risk. J Nutr 137:2576S-2579S
66. Alrawi SJ, Schiff M, Carroll RE et al. (2006) Aberrant crypt foci. Anticancer Res 26:107–119
67. Møller AK, Søborg DA, Abu Al-Soud W et al. (2013) Bacterial community structure in High-Arctic snow and freshwater as revealed by pyrosequencing of 16S rRNA genes and cultivation. Polar Research 32:17390
68. Lim JH, Baek S-H, Lee S-T (2009) *Ferruginibacter alkalilentus* gen. nov., sp. nov. and *Ferruginibacter lapsinanis* sp. nov., novel members of the family '*Chitinophagaceae'* in the phylum Bacteroidetes, isolated from freshwater sediment. Int J Syst Evol Microbiol 59:2394–2399
69. O'Sullivan LA, Rinna J, Humphreys G et al. (2005) *Fluviicola taffensis* gen. nov., sp. nov., a novel freshwater bacterium of the family *Cryomorphaceae* in the phylum '*Bacteroidetes'*. Int J Syst Evol Microbiol 55:2189–2194
70. DeLong EF, Franks DG, Alldredge AL (1993) Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. Limnol Oceanogr 38:924–934
71. Azam F, Malfatti F (2007) Microbial structuring of marine ecosystems. Nature Reviews Microbiology 5:782–791

72. Teeling H, Fuchs BM, Becher D et al. (2012) Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. Science 336:608–611
73. Gómez-Pereira PR, Fuchs BM, Alonso C et al. (2010) Distinct flavobacterial communities in contrasting water masses of the north Atlantic Ocean. ISME J 4:472–487
74. Xu J, Bjursell MK, Himrod J et al. (2003) A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis. Science 299:2074–2076
75. Wallenstein MD, Burns RG (2011) Ecology of Extracellular Enzyme Activities and Organic Matter Degradation in Soil: A Complex Community-Driven Process. In: Methods of Soil Enzymology, pp 35–55
76. Traving SJ, Thygesen UH, Riemann L et al. (2015) A model of extracellular enzymes in free-living microbes: which strategy pays off? Appl Environ Microbiol 81:7385–7393
77. Mann AJ, Hahnke RL, Huang S et al. (2013) The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides. Appl Environ Microbiol 79:6813–6822
78. Barbeyron T, Thomas F, Barbe V et al. (2016) Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: example of the model algae-associated bacterium *Zobellia galactanivorans* DsijT. Environmental Microbiology 18:4610–4627
79. Tancula E, Feldhaus MJ, Bedzyk LA et al. (1992) Location and characterization of genes involved in binding of starch to the surface of Bacteroides thetaiotaomicron. J Bacteriol 174:5609–5616
80. Martens EC, Roth R, Heuser JE et al. (2009) Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. J Biol Chem 284:18445–18457
81. Schauer K, Rodionov DA, Reuse H de (2008) New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? Trends Biochem Sci 33:330–338
82. Noinaj N, Guillier M, Barnard,, Travis J. et al. (2010) TonB-Dependent Transporters: Regulation, Structure, and Function. Annu Rev Microbiol 64:43–60
83. Foley MH, Cockburn DW, Koropatkin NM (2016) The Sus operon: a model system for starch uptake by the human gut *Bacteroidetes*. Cell Mol Life Sci 73:2603–2617
84. Gray DA, White JBR, Oluwole AO et al. (2021) Insights into SusCD-mediated glycan import by a prominent gut symbiont. Nature Communications 12:44

85. Koropatkin NM, Martens EC, Gordon JI et al. (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure 16:1105–1115
86. Tauzin AS, Kwiatkowski KJ, Orlovsky NI et al. (2016) Molecular Dissection of Xyloglucan Recognition in a Prominent Human Gut Symbiont. mBio 7:e02134-15
87. Mystkowska AA, Robb C, Vidal-Melgosa S et al. (2018) Molecular recognition of the beta-glucans laminarin and pustulan by a SusD-like glycan-binding protein of a marine *Bacteroidetes*. FEBS J 285:4465–4481
88. Boraston AB, Bolam DN, Gilbert HJ et al. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 382:769–781
89. Cameron EA, Maynard MA, Smith CJ et al. (2012) Multidomain Carbohydrate-binding Proteins Involved in *Bacteroides thetaiotaomicron* Starch Metabolism*. J Biol Chem 287:34614–34625
90. Tuson HH, Foley MH, Koropatkin NM et al. (2018) The Starch Utilization System Assembles around Stationary Starch-Binding Proteins. Biophys J 115:242–250
91. Sonnenburg ED, Zheng H, Joglekar P et al. (2010) Specificity of Polysaccharide Use in Intestinal Bacteroides Species Determines Diet-Induced Microbiota Alterations. Cell 141:1241–1252
92. Cho KH, Cho D, Wang GR et al. (2001) New regulatory gene that contributes to control of *Bacteroides thetaiotaomicron* starch utilization genes. J Bacteriol 183:7198–7205
93. D'Elia JN, Salyers AA (1996) Effect of regulatory protein levels on utilization of starch by *Bacteroides thetaiotaomicron*. J Bacteriol 178:7180–7186
94. Pereira GV, Abdel-Hamid AM, Dutta S et al. (2021) Degradation of complex arabinoxylans by human colonic *Bacteroidetes*. Nature Communications 12:459
95. Dudek M, Dieudonné A, Jouanneau D et al. (2020) Regulation of alginate catabolism involves a GntR family repressor in the marine flavobacterium *Zobellia galactanivorans* DsijT. Nucleic Acids Res 48:7786–7800
96. Koropatkin NM, Smith TJ (2010) SusG: a unique cell-membrane-associated alpha-amylase from a prominent human gut symbiont targets complex starch molecules. Structure 18:200–215
97. Shah HN, Seddon SV, Gharbia SE (1989) Studies on the virulence properties and metabolism of pleiotropic mutants of *Porphyromonas gingivalis* (*Bacteroides gingivalis*) W50. Oral Microbiology and Immunology 4:19–23

98. Mckee AS, Mcdermid AS, Wait R et al. (1988) Isolation of colonial variants of *Bacteroides gingivalis* W50 with a reduced virulence. Journal of Medical Microbiology 27:59–64
99. Kharade Sampada S., McBride Mark J. (2014) Flavobacterium johnsoniae Chitinase ChiA Is Required for Chitin Utilization and Is Secreted by the Type IX Secretion System. J Bacteriol 196:961–970
100. Lombard V, Golaconda Ramulu H, Drula E et al. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490-5
101. Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280 (Pt 2):309–316
102. Davies G, Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. Structure 3:853–859
103. Salyers A A, Palmer J K, Wilkins T D (1977) Laminarinase (beta-glucanase) activity in Bacteroides from the human colon. Appl Environ Microbiol 33:1118–1124
104. KOSHLAND Jr. DE (1953) STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. Biological Reviews 28:416–436
105. McCarter JD, Stephen Withers G (1994) Mechanisms of enzymatic glycoside hydrolysis. Current Opinion in Structural Biology 4:885–892
106. Kumar R, Henrissat B, Coutinho PM (2019) Intrinsic dynamic behavior of enzyme:substrate complexes govern the catalytic action of β-galactosidases across clan GH-A. Scientific Reports 9:10346
107. Ndeh D, Rogowski A, Cartmell A et al. (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. Nature 544:65–70
108. Rajan SS, Yang X, Collart F et al. (2004) Novel Catalytic Mechanism of Glycoside Hydrolysis Based on the Structure of an NAD+/Mn2+-Dependent Phospho-α-Glucosidase from *Bacillus subtilis*. Structure 12:1619–1629
109. Liu QP, Sulzenbacher G, Yuan H et al. (2007) Bacterial glycosidases for the production of universal red blood cells. Nat Biotechnol 25:454–464
110. Cifuente JO, Comino N, Madariaga-Marcos J et al. (2016) Structural Basis of Glycogen Biosynthesis Regulation in Bacteria. Structure 24:1613–1622
111. Lairson LL, Henrissat B, Davies GJ et al. (2008) Glycosyltransferases: Structures, Functions, and Mechanisms. Annual Review of Biochemistry 77:521–555
112. Garron M-L, Cygler M (2010) Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. Glycobiology 20:1547–1573

113. Rønne ME, Madsen M, Tandrup T et al. (2023) Gut bacterial alginate degrading enzymes. Essays Biochem 67:387–398
114. Lei S P, Lin H C, Wang S S et al. (1987) Characterization of the *Erwinia carotovora pelB* gene and its product pectate lyase. J Bacteriol 169:4379–4383
115. Reisky L, Stanetty C, Mihovilovic MD et al. (2018) Biochemical characterization of an ulvan lyase from the marine flavobacterium *Formosa agariphila* KMM 3901T. Applied Microbiology and Biotechnology 102:6987–6996
116. Lombard V, Bernard T, Rancurel C et al. (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. Biochem J 432:437–444
117. Hettiarachchi SA, Kwon Y-K, Lee Y et al. (2019) Characterization of an acetyl xylan esterase from the marine bacterium *Ochrovirga pacifica* and its synergism with xylanase on beechwood xylan. Microbial Cell Factories 18:122
118. Williamson G, Kroon PA, Faulds CB (1998) Hairy plant polysaccharides: a close shave with microbial esterases. Microbiology (Reading) 144 (Pt 8):2011–2023
119. Montanier C, Money VA, Pires VMR et al. (2009) The Active Site of a Carbohydrate Esterase Displays Divergent Catalytic and Noncatalytic Binding Functions. PLOS Biology 7:e1000071
120. Hettle AG, Vickers CJ, Boraston AB (2022) Sulfatases: Critical Enzymes for Algal Polysaccharide Processing. Frontiers in Plant Science 13
121. Barbeyron T, Brillet-Guéguen L, Carré W et al. (2016) Matching the Diversity of Sulfated Biomolecules: Creation of a Classification Database for Sulfatases Reflecting Their Substrate Specificity. PLOS ONE 11:e0164846
122. Anderson NS, Dolan TC, Rees DA (1965) Evidence for a common structural pattern in the polysaccharide sulphates of the Rhodophyceae. Nature 205:1060–1062
123. Aquino RS, Grativol C, Mourão PAS (2011) Rising from the sea: correlations between sulfated polysaccharides and salinity in plants. PLOS ONE 6:e18862
124. Genicot SM, Groisillier A, Rogniaux H et al. (2014) Discovery of a novel iota carrageenan sulfatase isolated from the marine bacterium *Pseudoalteromonas carrageenovora*. Frontiers in Chemistry 2
125. Ficko-Blean E, Préchoux A, Thomas F et al. (2017) Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. Nature Communications 8:1685
126. Reisky L, Préchoux A, Zühlke M-K et al. (2019) A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan. Nat Chem Biol 15:803–812

127. Silchenko AS, Rasin AB, Zueva AO et al. (2018) Fucoidan Sulfatases from Marine Bacterium *Wenyingzhuangia fucanilytica* CZ1127(T). Biomolecules 8
128. Bauer M, Kube M, Teeling H et al. (2006) Whole genome analysis of the marine *Bacteroidetes 'Gramella forsetii'* reveals adaptations to degradation of polymeric organic matter. Environmental Microbiology 8:2201–2213
129. Levasseur A, Drula E, Lombard V et al. (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnology for Biofuels 6:41
130. Wong DWS (2009) Structure and Action Mechanism of Ligninolytic Enzymes. Applied Biochemistry and Biotechnology 157:174–209
131. Langston JA, Brown K, Xu F et al. (2012) Cloning, expression, and characterization of a cellobiose dehydrogenase from Thielavia terrestris induced under cellulose growth conditions. Biochim Biophys Acta 1824:802–812
132. Vaaje-Kolstad G, Westereng B, Horn SJ et al. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. Science 330:219–222
133. Brott S, Nam KH, Thomas F et al. (2023) Unique alcohol dehydrogenases involved in algal sugar utilization by marine bacteria. Applied Microbiology and Biotechnology 107:2363–2384
134. Reisky L, Büchsenschütz HC, Engel J et al. (2018) Oxidative demethylation of algal carbohydrates by cytochrome P450 monooxygenases. Nat Chem Biol 14:342–344
135. Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169
136. Kappelmann L, Krüger K, Hehemann J-H et al. (2019) Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. ISME J 13:76–91
137. Staley JT, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev Microbiol 39:321–346
138. Baer R, Bankier AT, Biggin MD et al. (1984) DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature 310:207–211
139. Fleischmann RD, Adams MD, White O et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512
140. Cavalier-Smith T (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of *Protozoa*. Int J Syst Evol Microbiol 52:297–354
141. Zimorski V, Ku C, Martin WF et al. (2014) Endosymbiotic theory for organelle origins. Current Opinion in Microbiology 22:38–48

142. Schmidt TM, DeLong EF, Pace NR (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J Bacteriol 173:4371–4378
143. Ambardar S, Gupta R, Trakroo D et al. (2016) High Throughput Sequencing: An Overview of Sequencing Chemistry. Indian Journal of Microbiology 56:394–404
144. Krakau S, Straub D, Gourlé H et al. (2022) nf-core/mag: a best-practice pipeline for metagenome hybrid assembly and binning. NAR Genom Bioinform 4:lqac007
145. Uritskiy GV, DiRuggiero J, Taylor J (2018) MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6:158
146. Tamames J, Puente-Sánchez F (2018) SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline. Front Microbiol 9:3349
147. Dong X, Strous M (2019) An Integrated Pipeline for Annotation and Visualization of Metagenomic Contigs. Frontiers in Genetics 10
148. Krüger K, Chafee M, Ben Francis T et al. (2019) In marine Bacteroidetes the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. ISME J 13:2800–2816
149. Mutz K-O, Heilkenbrinker A, Lönne M et al. (2013) Transcriptome analysis using next-generation sequencing. Current Opinion in Biotechnology 24:22–30
150. Francis TB, Bartosik D, Sura T et al. (2021) Changing expression patterns of TonB-dependent transporters suggest shifts in polysaccharide consumption over the course of a spring phytoplankton bloom. ISME J 15:2336–2350
151. Sidhu C, Kirstein IV, Meunier CL et al. (2023) Dissolved storage glycans shaped the community composition of abundant bacterioplankton clades during a North Sea spring phytoplankton bloom. Microbiome 11:77
152. Avcı B, Krüger K, Fuchs BM et al. (2020) Polysaccharide niche partitioning of distinct *Polaribacter* clades during North Sea spring algal blooms. ISME J 14:1369–1383

# 2. Article information

This thesis comprises three research articles. The first article, “Alpha-glucans from bacterial necromass indicate an intra-population loop within the marine carbon cycle”, has been submitted and is under review at Nature Communications. A preprint is available at doi.org/10.21203/rs.3.rs-3205445/v1. The second, “Marine bacteroidetes use a conserved enzymatic cascade to digest diatom β-mannan” was published in the ISME Journal in 2023 and the third, “Marine *Bacteroidetes* enzymatically digest xylans from terrestrial plants” in Environmental Microbiology in 2023.

## 2.1. Author contributions

*Article I*

Authors: **I Beidler**, N Steinke, T Schulze, C Sidhu, D Bartosik, J Krull, T Dutschei, B Ferrero-Bodera, J Rielicke, V Kale, T Sura, A Trautwein-Schult, IV Kirstein, KH Wiltshire, H Teeling, D Becher, MM Bengtsson, J-H Hehemann, UT Bornscheuer, RI Amann, T Schweder

**I Beidler** drafted the manuscript, characterized marine enzymes, performed single strain proteomics, prepared and analyzed polysaccharide extracts and performed 18S rRNA sequence analysis. T Schweder designed the study and supported the writing of the manuscript together with H Teeling. N Steinke and J Krull performed additional glycan extract analysis. T Schulze assisted with growth experiments and enzyme characterization for which T Dutschei provided resources. Sample measurements for single strain proteomics were performed by B Ferrero-Bordera and J Rielicke. Metaproteome samples were measured by T Sura and V Kale. Mass spectrometry analyses were performed by A Trautwein-Schult. Metaproteome data were analysed by D Bartosik. C Sidhu performed metatranscriptome analyses and provided data. MM Bengtsson procured 18S rRNA sequence data and performed analyses. KH Wiltshire and IV Kirstein provided HPLC chl *a* data and ensured the sampling and laboratory logistics on Helgoland. Figures were prepared by **I Beidler** and D Bartosik. T Schweder, J-H Hehemann, D Becher, UT Bornscheuer, and RI Amann coordinated the project. All authors read and contributed to the manuscript.

*Article II*

Authors: **I Beidler**, CS Robb, S Vidal-Melgosa, M-K Zühlke, D Bartosik, V Solanki, S Markert, D Becher, T Schweder, J-H Hehemann

J-H Hehemann and T Schweder initiated the study and directed the project. CS Robb and **I Beidler** conducted the cloning experiments, the enzyme and structural biological analyses. V Solanki supported the protein structural biological analyses. **I Beidler**, S Markert and M-K Zühlke performed the proteome analysis for which D Bartosik provided resources. D Bartosik analyzed the genomic data. S Vidal-Melgosa conducted the epitope-deletion assays. CS Robb, **I Beidler**, S Markert, J-H Hehemann and T Schweder wrote the manuscript. All co-authors critically reviewed and approved the submitted manuscript.

*Article III*

Authors: T Dutschei, **I Beidler**, D Bartosik, J-M Seeßelberg, M Teune, M Bäumgen, S Querido Ferreira, J Heldmann, F Nagel, J Krull, L Berndt, K Methling, M Hein, D Becher, P Langer, M Delcea, M Lalk, M Lammers, M Höhne, J-H Hehemann, T Schweder, UT Bornscheuer

T Schweder: Conceptualization (equal); funding acquisition (equal); project administration (equal); resources (equal); supervision (equal); validation (equal); writing – review and editing (equal). T Dutschei: Formal analysis (equal); investigation (equal); methodology (equal); writing – original draft (equal). **I Beidler**: Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); visualization (equal); writing – review and editing (equal). D Bartosik: Data curation (equal); formal analysis (equal); investigation (equal); software (equal); visualization (equal). J-M Seeßelberg: Formal analysis (equal); investigation (equal); methodology (equal). M Teune: Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); writing – review and editing (equal). M Bäumgen: Formal analysis (equal); investigation (equal). S Querido: Formal analysis (equal); investigation (equal). J Heldmann: Formal analysis (equal); investigation (equal). F Nagel: Formal analysis (equal); investigation (equal). J Krull: Formal analysis (equal); investigation (equal); methodology (equal). Leona Berndt: Formal analysis (equal); investigation (equal). K Methling: Formal analysis (equal); investigation (equal); methodology (equal). M Hein: Formal analysis (equal); investigation (equal). D Becher: Funding acquisition (equal); methodology (equal); resources (equal). P Langer: Formal analysis (equal); investigation (equal); methodology (equal). M Delcea: Methodology (equal); resources (equal). M Lalk:

Author contributions

Data curation (equal); investigation (equal); methodology (equal). M Lammers: Investigation (equal); methodology (equal). M Höhne: Funding acquisition (equal); investigation (equal); methodology (equal); resources (equal). J-H Hehemann: Funding acquisition (equal); investigation (equal); methodology (equal); resources (equal); writing – review and editing (equal). UT Bornscheuer: Conceptualization (equal); funding acquisition (equal); project administration (equal); resources (equal); writing – review and editing (equal)

_____

Irena Beidler

_____

Prof. Dr. Thomas Schweder

# Part II

# Research Articles

# 3. Alpha-glucans from bacterial necromass indicate an intra-population loop within the marine carbon cycle

**Authors**: Irena Beidler[1], Nicola Steinke[2,3], Tim Schulze[1], Chandni Sidhu[2], Daniel Bartosik[1,4], Joris Krull[2,4], Theresa Dutschei[5], Borja Ferrero-Bordera[6], Julia Rielicke[6], Vaikhari Kale[6], Thomas Sura[6], Anke Trautwein-Schult[6], Inga V. Kirstein[7], Karen H. Wiltshire[7], Hanno Teeling[2], Dörte Becher[6], Mia Maria Bengtsson[8], Jan-Hendrik Hehemann[2,3], Uwe. T. Bornscheuer[5], Rudolf I. Amann[2], Thomas Schweder[1,4]

[1] Pharmaceutical Biotechnology, Institute of Pharmacy, University of Greifswald, 17489 Greifswald, Germany

[2] Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany

[3] University of Bremen, Center for Marine Environmental Sciences, MARUM, 28359 Bremen, Germany

[4] Institute of Marine Biotechnology, 17489 Greifswald, Germany

[5] Biotechnology and Enzyme Catalysis, Institute of Biochemistry, University of Greifswald, 17489 Greifswald, Germany

[6] Microbial Proteomics, Institute of Microbiology, University of Greifswald, 17489 Greifswald, Germany

[7] Alfred Wegener Institute for Polar and Marine Research, Biologische Anstalt Helgoland, 20 27483, Helgoland, Germany

[8] Microbial Physiology and Molecular Biology, Institute of Microbiology, University of Greifswald, 17489 Greifswald, Germany

Submitted and under review at *Nature Communications*, August 2023

## 3.1. Abstract

Phytoplankton blooms initiate bacterioplankton blooms, from which bacterial biomass is released via grazing zooplankton and viral lysis. Bacterial consumption of algal biomass during blooms is well studied, but little is known about the simultaneous reuse of bacterial necromass. Alpha- and beta-glucans are abundant dissolved organic macromolecules during blooms. We demonstrate algal laminarin-fueled alpha-glucan synthesis in marine

Bacteroidota strains, as well as bacterial reuse of these alpha-glucans as major carbon source in vitro and during a diatom-dominated bloom. We highlight two types of genomic loci and the encoded protein machineries with structurally distinct SusD substrate-binding proteins that may target alpha-glucans of different complexities. It is demonstrated that these encoded machineries can be specifically induced by extracted alpha-glucan-rich bacterial polysaccharides. This bacterial alpha-glucan synthesis and recycling from bacterial necromass constitutes a large-scale intra-population energy conservation mechanism redirecting substantial amounts of carbon in an essential part of the microbial loop.

## 3.2. Introduction

Marine microalgae (phytoplankton) account for an estimated 40-50% of the global photosynthetic primary production [1]. Phytoplankton blooms in particular entail fixation of large amounts of carbon, of which considerable amounts are converted to various polysaccharides [2, 3]. Secretion, leakage or lysis of algal cells release these polysaccharides as dissolved or particulate organic matter (DOM, POM), providing a diverse carbon and energy source for heterotrophic bacteria specialized in their uptake and degradation [4, 5]. This remineralization of dissolved algal glycans represents an essential part of the marine microbial loop [6] and thus the global carbon cycle. Particularly prominent in this process are marine members of the phylum *Bacteroidota* [7–9]. These bacteria target algal polysaccharides by specific sets of enzymes and transporters encoded in genomic islands, termed polysaccharide utilization loci (PULs) [10]. PULs enable specialized *Bacteroidota* to thrive in close succession to phytoplankton primary producers during blooms [11, 12], providing a link to higher trophic levels as they are grazed on by, e.g., bacterivorous flagellates [13] or ciliates [14]. The high cell densities that are reached by some bloom-associated bacteria also render them susceptible to viral infections. It has been shown that phage numbers correspond well to bacterial cell counts during algal blooms, and that phages infect key polysaccharide degraders such as *Polaribacter* spp [15, 16]. Viral lysis of such bacteria therefore influences the carbon flux and fuels the DOM pool by releasing soluble bacterial organic matter, including their bacterial storage glucans [17]. However, it is so far unclear, whether or not these released bacterial storage glucans are simultaneously remineralized with dissolved algal polysaccharides under bloom conditions.

During previous studies of seasonal spring phytoplankton blooms off the North Sea island Helgoland in the southern German Bight, we found that most active and abundant planktonic

bacteria possess dedicated PULs for β-glucans (laminarins) and α-glucans [18]. In a study of 53 sequenced North Sea *Flavobacteriia* strains, we furthermore found laminarin and α-glucan specific PULs in 62% and 75% of the strains, respectively, whereas only 37% coded for alginate PULs [19]. Likewise, spring bloom metaproteome and -transcriptome data obtained at Helgoland Roads revealed that glucan metabolism outweighs that of any other polysaccharide in DOM [17, 20]. However, while the high frequency and expression of laminarin-targeting genes in bloom-associated bacteria are readily explained by the abundance of algal laminarin, the role of α-glucans remains more elusive [5].

Marine phototrophs use glucans as storage polysaccharides. Red algae (*Rhodophyta*), green algae (*Chlorophyta*) and dinoflagellates (*Dinophyceae*) form α-1,4 glucans with varying degrees of α-1,6 branching, while *Stamenopiles*, e.g., diatoms as well as haptophytes (*Haptophyta*), form β-1,3/ β-1,6 laminarin or chrysolaminarin [21]. It is reasonable to assume that marine heterotrophic bacteria utilize both storage glucan types. However, in a recent study we observed that bacteria associated with an overwhelmingly diatom-dominated spring algal bloom expressed their α-glucan PULs during peak bloom phases [17]. This led to the assumption that these bacteria may be specifically adapted towards recycling bacterial α-glucans of lysed bacteria. In this study, we show that marine *Flavobacteriia* associated with phytoplankton blooms degrade the microalgal β-glucan laminarin and uses excess glucose to synthesize α-glucan storage polysaccharides. We reveal that these bacteria at the same time employ a highly specialized protein machinery to take up and thereby recycle α-glucan storage polysaccharides from bacterial necromass, i.e., from lysed bacterial community members. This process acts as an intra-population energy conservation mechanism under bloom conditions retaining a large amount of glucans in a permanent loop and represents an as yet neglected part of the marine carbon cycle.

## 3.3. Materials and Methods

### *3.3.1. Sampling site*

Subsurface seawater (1 m depth) was collected at 52 time points between 2nd of March and 26th of May 2020 at the station Helgoland Roads near Helgoland in the southern North Sea. Since 1962 bucket water samples have been taken as part of a long-term monitoring program Helgoland Roads (54°11'N 7°54'E; DEIMS.iD: https://deims.org/1e96ef9b-0915-4661-849f-b3a72f5aa9b1) [22]. Water samples for the current study were taken here.

### *3.3.2. Sequence analysis*

Sequence analysis was carried out on the basis of the existing annotation [19] with additional reannotation as describes previously [20]. Sequences were aligned using ClustalO [23] in Unipro UGENE [24]. Trees were visualized using iTOL [25]. Structural models of SusD proteins were generated with AlphaFold [26].

### *3.3.3. Chlorophyll a measurements and cells counts of total bacteria and dominant bacterial clades*

Sample filtration was carried out in a laboratory under dim light to avoid the loss of pigments during the filtration procedure. For 2016 and 2018 samples, pigment extraction and analysis was carried out using a combined protocol from Zapata *et al* [27]. and Garrido *et al* [28]. For 2020 samples, we followed the extraction and analysis method as described previously [29]. Subsequently, pigments were separated via high-performance liquid chromatography (HPLC) (Waters 2695 Separation Module), and detected with a Waters 996 Photodiode Array Detector.

Total bacterial cell counts (TCC) and cell numbers of the dominant clades *Aurantivirga* (CARD-FISH probe AUR452) and *Polaribacter* (POL740) were described and published previously [12].

### *3.3.4. 18S rRNA gene sequencing*

Sampled water was filtered sequentially onto polycarbonate membrane filters with different pore sizes (10 μm, 3 μm and 0.2 μm). For 18S rRNA gene amplicon sequencing, the two larger size fractions, >10 μm and 3-10 μm were analyzed. DNA was extracted from the filters using the DNeasy PowerSoil kit for DNA (Qiagen). Mechanical lysis was achieved by bead beating in a FastPrep 24 5G (MP Biomedicals). The V7 region of the 18S rRNA gene was amplified using the primers [F-1183mod: 5'-AATTTGACTCAACRCGGG-3', R-1443mod: 5'-GRGCATCACAGACCTG-3'] [30] coupled to custom adaptor-barcode constructs. PCR amplification and Illumina MiSeq library preparation and sequencing (V3 chemistry) was carried out by LGC Genomics in Berlin. Sequences have been submitted to the European Nucleotide Archive under the accession number PRJEB51816. Amplicon Sequence Variants (ASVs) were obtained using DADA2 [31] and taxonomically classified as described previously [32]. ASVs classified as *Metazoa* were removed before downstream analyses to reduce the effect of crustacean zooplankton on community composition. For analysis, 10 μm and 3-10

µm counts were combined, set to 1 and relative abundances calculated. Classification of *Dinophyceae* into majorly heterotroph and majorly autotroph taxa was done for dominant groups according to literature [13, 33–36].

### *3.3.5. Metatranscriptomics*

Metagenome and metatranscriptome sequencing were performed as described in [17]. Briefly, thirty metagenomes were sequenced using PacBio Sequel II (1 SMRT cell/sample) (Menlo Park, CA, USA) while corresponding metatranscriptomes were obtained using Illumina HiSeq 3,000 (~100 million reads/sample) (San Diego, CA, USA). The metagenomes were then processed to reconstruct metagenome-assembled genomes (MAGs), and mRNA reads were mapped to these genomes to identify highly expressed MAGs. The mapping and annotation of the MAGs were carried out using the SqueezeMeta v1.3.1 pipeline [37]. To predict open reading frames (ORFs), FragGeneScan v1.31 with the parameters "w1" and "sanger_5" as described by Rho *et al*. [38] was used. The predicted ORFs were searched against various databases including GenBank r239 [39], eggNOG v5.0 [40], KEGG r58.0 [41], and CAZy (as of 30/07/2020) [42] using Diamond v0.9.24.125 [43]. HMM homology searches against the Pfam 33.0 database [44] were conducted using HMMER3 [45]. The combined annotations were utilized for the manual prediction of polysaccharide utilization loci (PULs) and carbohydrate-active enzyme (CAZyme) clusters. Bowtie2 [46] was employed to map mRNA reads to the MAGs, and transcripts per million (TPM) values were calculated for all MAGs in a given sample using the formula: (sum of reads successfully mapping to a MAG in the sample x 10^6) / (sum of contig lengths of the MAG x sum of reads in the sample). Metagenome, metatranscriptome and MAG sequence data are available from the European Nucleotide Archive (accession PRJEB52999).

### *3.3.6. Metaproteomics*

#### *3.3.6.1. Sample preparation*

All information about the metaproteomics analysis of the 0.2 µm fraction from the spring phytoplankton bloom in 2016 has been described in detail in [20] and the free-living bacteria of the blooms from 2018 and 2020 were prepared as previously described [47]. Briefly, proteins were extracted from one-eighth of a filter (Millipore Express PLUS Membrane, polyethersulfone, hydrophilic, 0.2 µm pore size, diameter 142 mm) by cutting the filter into small pieces before transfer to 15 mL low binding tubes containing 1 mL resuspension buffer

1 (50 mM Tris-HCl (pH 7.5), 0.1 mg mL1 chloramphenicol, 1 mM phenylmethylsulfonyl fluoride (PMSF)) and 1,5 mL resuspension buffer 2 (20 mM Tris-HCl pH 7.5, 2% SDS (w/v)). After heating (10 min at 60 °C at 1,000 rpm in a thermo-mixer), 5 mL DNAse buffer (20 mM Tris-HCl pH 7.5, 0.1 mg mL-1 MgCl2, 1 mM PMSF, 1 µg mL-1 DNAse I) was added, and cells lysis was carried out by ultra-sonication (amplitude 51-60%; cycle 0.5; 3x 2 min) on ice before incubation for 10 min at 37 °C at 1,000 rpm. After centrifugation (10 min at 4 °C at 10,000 × *g*), the supernatant was collected and the pelleted filter pieces were stirred and centrifuged again for 1 min at 4 °C at 5,000 × *g*. Pre-cooled trichloroacetic acid (20% TCA (v/v)) was added for protein precipitation to the supernatant and after inverting the tube approximately 10x, the precipitate was pelleted via centrifugation (45 min, 4 °C, 12,000 × *g*) and the protein pellet was washed 3x in pre-cooled (-20 °C) acetone (10 min, 4 °C, 12,000 × *g*) before drying at room temperature. The proteins were resuspended in 2× SDS sample loading buffer (4% SDS (w/v), 20% glycerol (w/v), 100 mM Tris-HCl pH 6.8, bromphenol blue (tip of a spatula, to add color), 3.6% 2 mercaptoethanol (v/v) (freshly added before use)), incubated for 5 min at 95 °C before vortexing and separated via SDS-PAGE (Criterion TG 4-20% Precast Midi Gel, BIO-RAD Laboratories, Inc., USA). The proteins were fixated, stained with Coomassie, and each gel lane was cut into 20 pieces [48]. Gel pieces were destained 3x for 10 min with 1 mL of gel washing buffer (200 mM ammonium bicarbonate in 30% acetonitrile (v/v)) at 37 °C under vigorous shaking, dehydrated in 1 mL 100% acetonitrile (v/v) for 20 min and the supernatant was removed before drying the gel pieces in a vacuum centrifuge at 30 °C. Proteins were in-gel reduced with 100 µL 10 mM dithiothreitol in 25 mM ammonium bicarbonate buffer (1 h at 56 °C) and alkylated with 100 µL 55 mM iodoacetamide in 25 mM ammonium bicarbonate buffer (without light for 45 min at room temperature) before the supernatant was removed. The gel pieces were washed with 1 mL 25 mM ammonium bicarbonate buffer (10 min, 1,000 rpm at room temperature), dehydrated with 500 µL (2018 bloom) or 800 µL (2020 bloom) 100% acetonitrile for 10 min. The supernatant was removed before gel pieces were dried in a vacuum centrifuge (20 min) and finally covered with 120 µL trypsin solution (2 µg/mL Trypsin (Promega). After incubation for 20 min at room temperature, excess trypsin solution was removed and incubated in a thermo-mixer 15 h at 37 °C without shaking. Peptides were eluted with 120 µL solvent A (water MS grade in 0,1% acetic acid (v/v)) by sonication for 15 min before protein containing supernatant was transferred into a new tube. Peptide elution was repeated with 120 µL 30% acetonitrile (v/v) by sonication for 15 min. The eluates were pooled, and eluate volume was reduced in a vacuum centrifuge to a maximum of 15 to 20 µL. The peptides were desalted via ZipTips µC18

(Merck Millipore, P10 tip size) according to the manufacturer's protocol. The eluted samples were dried in a vacuum centrifuge and resuspended in 10 µL 0.5x Biognosys iRT standard kit in solvent A.

### *3.3.6.2. LC-MS/MS measurement and data analysis*

All information about the LC MS/MS measurement and data analysis of the 0.2 µm fraction from the spring phytoplankton bloom in 2016 has been described in detail in [20] and are briefly described here for the free-living bacteria of the blooms from 2018 and 2020. An Easy-nLC1000 (Thermo Fisher Scientific, Waltham, MA, USA) was coupled to a Q Exactive mass spectrometer (Thermo Fisher Scientific) and peptides were loaded onto in-house packed capillary columns (20 cm length,75 µm inner diameter) filled with Dr. Maisch ReproSil Pur 120 C18-AQ 1.9 µm (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) and separated using a 131 min nonlinear binary gradient from 1% to 99% solvent B (99.9% acetonitrile(v/v), 0.1% acetic acid (v/v)) in solvent A (0.1% acetic acid (v/v)) at a constant flow rate of 300 nL $min^{-1}$. The MS1 scan was recorded with a mass window of 300–1,650 m/z and a resolution of 140,000 at 200 m/z. The 15 most intense precursor ions were selected for HCD fragmentation (ions with an unassigned charge or a charge of 1,7,8, >8 were excluded) with a normalized collision energy of NCE 27. The resulting MS/MS spectra were recorded with a resolution of 17,500 at 200 m/z. Dynamic exclusion and lock mass correction were enabled.

All MS/MS spectra were analyzed using Mascot (version 2.7.0.1; Matrix Science, London, UK) and a bloom-specific metagenome-derived database containing all protein sequences from the 18 metagenomes obtained during the spring bloom in 2018 or 15 metagenomes obtained during the spring bloom 2020 assuming the digestion enzyme trypsin. Redundant proteins were removed using cd-hit [49] with a clustering threshold of 97% identity. The non-redundant database was added by a set of common laboratory contaminants and reverse entries, amounting to 81,874,922 (bloom 2018) or 4,221,978 (bloom 2020) sequences in the final database.

For database search with Mascot [50], the following parameters were used: fragment ion mass tolerance and parent ion tolerance of 10 ppm, none missed cleavages, variable modification on methionine (oxidation), and fixed modification on cysteine (carbamidomethylation). Scaffold (version 4.11.1 (bloom 2018) or version 5.0.1 (bloom 2020); Proteome Software Inc., Portland, OR) was used to merge the search results and to validate MS/MS-based peptide and protein identifications [51]. During data analysis in Scaffold, an additional X! Tandem search was performed for validation (version 2017.2.1.4;

The GPM, thegpm.org; version X!Tandem Alanine) [52] with default settings (fragment ion mass tolerance and parent ion tolerance of 10 ppm, carbamidomethyl on cysteine as fixed modification, Glu->pyro-Glu of the N-terminus, ammonia-loss of the N-terminus, Gln->pyro-Glu of the N-terminus and oxidation on methionine for 2018 and 2020 bloom, and additional carbamidomethyl of cysteine as variable modifications for 2018 bloom). Peptide identifications were accepted if they could be established at greater than 95% probability. Peptide probabilities from Mascot were assigned by the Peptide Prophet algorithm (bloom 2018) [53] or the Scaffold Local FDR algorithm (bloom 2020). Peptide Probabilities from X! Tandem were assigned by the Peptide Prophet algorithm [53] with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 99% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm [54]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

For (semi-)quantitative analysis of 2016 [20], 2018 and 2020 metaproteomic datasets, percent normalized weighted spectra (%NWS) were calculated by dividing Scaffold's 'Quantitative Value' for normalized, weighted (i.e. protein size-adjusted) spectra for each protein group, by the sum of all quantitative values for the sample. Average values were calculated from three biological replicates, including '0' for proteins that were not identified within a replicate. To make Bacteria-specific %NWS readily comparable across all samples, all bacterial spectra were normalized to 100% (%BacNWS) using taxonomic assignment for protein groups provided by GhostKOALA v2.0 [55] (genus_prokaryotes + family_eukaryotes + viruses database).

The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [56] with the dataset identifier PXD019294 (bloom 2016), PXD042676 (bloom 2018) (**Reviewer access** Username: reviewer_pxd042676@ebi.ac.uk, Password: 7pX6or0p), PXD042805 (bloom 2020) (**Reviewer access** Username: reviewer_pxd042805@ebi.ac.uk, Password: fEknpDcw).

### *3.3.7. Comparative genomics*

To avoid redundancy across the sampling years, MAGs from 2010-2012, 2016 [20] (European Nucleotide Archive project accession PRJEB28156), 2018 (PRJEB38290) and 2020 [17] (PRJEB52999) were dereplicated using dRep [57] v3.2.0 with minimum completeness of 70%

and contamination lower than 5% at 0.95 ANI (average nucleotide identity). Protein sequences for representative MAGs were predicted with Prokka [58] v1.14.6 . PULs, CAZymes, SusC-like and SusD-like proteins were predicted as described previously [20], using hmmscan v3.3.2 against dbCAN-HMMdb-V11 and diamond [43] v2.1.1.155 against CAZyDB.08062022 provided by dbCAN [59]. GH13-encoding PULs were classified as „α-glucan-targeting“, whereas PULs carrying combinations of GH3, GH16, GH17, GH30 and/or GH5 enzymes were annotated as „β-glucan-targeting“. These substrate predictions were curated manually according to further PUL encoded CAZymes.

For identification of enzymes involved in α-glucan synthesis, K numbers were assigned to each sequence by GhostKOALA v2.0 [55] (genus_prokaryotes + family_eukaryotes + viruses) and KofamScan [60] (ver. 2023-04-01, KEGG release 106) with an E-value ≤ 0.01. Proteins with hits for K00963, K00975, K00693, K00750, K16150, K16153, K13679, K20812, K00703, K16148, K16147, K00700 and K16149 were kept as part of the α-glucan synthesis pathway.

### *3.3.8. Strain and cultivation conditions*

We used the North Sea flavobacterial strains *Polaribacter* sp. Hel_I_88 (isolated from seawater off Helgoland island) and *Muricauda* sp. MAR_2010_75 (isolated from seawater at Sylt island), as model organisms [61]. For pre-cultures and polysaccharide extractions *Polaribacter* sp. Hel_I_88 and *Muricauda* sp. MAR_2010_75 (20 °C) were grown over night (200 rpm) in Marine Broth (MB 2216, Difco). *Polaribacter* sp. and *Muricauda* sp. were tested for growth on specific carbon sources in MPM medium [62] containing 0.1% (w/v) of a single polysaccharide or 0.2% (w/v) bacterial polysaccharide extract. Growth was assessed via measurement of optical density at 600 nm. Cultures for proteome analysis of were carried out in 25 mL MPM medium with 0.1% (w/v) carbon source using biological triplicates.

### *3.3.9. Proteomics*

Cultures of *Polaribacter* sp. Hel_I_88 for time sampling and *Polaribacter* sp. and *Muricauda* sp. MAR_2010_75 for extract characterization were grown in 100 mL and 25 mL cultures, respectively. For time sampling, 25 mL samples were sequentially filtered through 3 and 0.2 μm polycarbonate filters (Ø 47 mm, Merck) using a vacuum pump (PC 3002 VARIO, VACCUBRAND) after 16, 24 and 48 h. Filters were stored at -80 °C until further use. Protein was extracted from ¼ of each 0.2 μm filter and prepared for mass spectrometry as described for metaproteomics but using 10% 1D-SDS polyacrylamide gels.

For extract characterization, cultures were grown on either polysaccharide extract, glycogen and alginate (*Polaribacter* sp.) or xylan from beechwood (*Muricauda* sp.). After 72 h, 25 mL cultures were harvested via centrifugation at 4,000 x *g* and stored at -80 °C until further use. Protein was extracted by resuspending the pellet in 2 mL 50 mM TEAB Buffer containing 4% (w/v) SDS. Samples were incubated at 95 °C and 600 rpm for 5 min, cooled on ice and sonicated in an ultrasonic bath for 5 min. Debris was removed by centrifugation (14,000 x *g*, 10 min) and protein concentration was determined using the Pierce BCA Protein Assay Kit (ThermoFischer Scientific, Waltham, MA, USA). Per sample, 25 µg protein were used. Proteins were separated on a 10% 1D-SDS polyacrylamide gel at 120 V for 90 min.
Samples were measured using an easy nLCII HPLC system applying a 100 min gradient coupled to an LTQ Orbitrap Velos mass spectrometer (Resolution 30,000, Scan range 300-1 700) (Thermo Fisher Scientific Inc., Waltham, MA, USA) [63]. Using MaxQuant [64], spectra were matched using a target-decoy protein sequence database with sequences and reverse sequences of *Polaribacter* sp. Hel_I_88 (NCBI ASM68793v1) or *Muricauda* sp. MAR_2010_75 (NCBI ASM74518v1) and common laboratory contaminants. A protein and peptide level FDR of 0.01 (1%) with at least two identified peptides per protein was applied. Only proteins that were detected in at least two replicates were classed as identified. Relative iBAQ (intensity based absolute quantification) values were manually calculated from automatically calculated iBAQs. Data and Results are available through the ProteomeXchange Consortium via the PRIDE partner repository (http://proteomecentral.proteomexchange.org) [56] with the identifiers PXD043390 (Username: reviewer_pxd043390@ebi.ac.uk; Password: x3FhiYi1).

### *3.3.10. Cloning, protein expression and purification*

Genes coding for the proteins GH13 (1) (P161_RS0117435), GH13 (2) (P161_RS0117440) and GH13 (3) (P161_RS0117455) of *Polaribacter* sp. Hel_I_88 were codon optimized and synthesized by *de novo* gene synthesis (BioCat GmbH, Heidelberg, Germany)**.** They were cloned into pET22b+ in *Escherichia coli* BL21 (DE3) for protein production. Proteins were produced in 200 mL LB cultures (30 µg/mL ampicillin) by induction with IPTG and incubation over night at 20 °C. Cells were harvested by centrifugation (5,000 x *g*, 20 min), lysed using BugBuster Protein Extraction Reagent (Merck) and centrifuged (9,500 x *g*, 20 min) to remove debris. Proteins were purified by loading the lysate onto a prepacked 5 mL IMAC column (HisTrap HP 5 mL, Cytiva) equilibrated with IMAC Buffer A (100 mM NaCl, 20 mM Imidazole, 20 mM TRIS-HCl, pH 8) using an ÄKTA Pure 25 L (Cytiva). Proteins were eluted with a step

gradient of IMAC Buffer B (100 mM NaCl, 500 mM Imidazole, 20 mM TRIS-HCL, pH 8). The resulting protein was desalted using a prepacked Sepharose-based desalting column (HiPrep Desalting 26/10, Cytiva) with PBS Buffer (pH 7.4) and concentrated via spin column (Pierce Protein Concentrator PES, 30K MWCO, 2-6 mL, Thermo Fischer).

### *3.3.11. Enzyme characterization*

Activity profiles for all enzymes were generated by 3,5-dinitrosalicylic acid (DNS) reducing-end assay [65] as well as fluorophore-assisted carbohydrate electrophoresis (FACE) [8]. 25 μg purified protein were incubated with 0.5% (w/v) poly-/oligosaccharide for 24 h. Samples were heat-inactivated at 80 °C for 10 min and centrifuged (13,000 x *g*, 10 min) to remove precipitated protein.

For the reducing-end assay, samples were incubated with DNS-reagent solution (30% (w/v) Potassium sodium tartrate tetrahydrate, 10 mg/mL DNS, 0.4 M NaCl) for 15 min at 95 °C and cooled to RT before measurement at 540 nm (Infinite 200 PRO M PLEX, Tecan, Männedorf, Switzerland). Values were compared against those of solutions containing only polysaccharide or only enzyme.

FACE was performed with 8-aminonaphthtalene-1,3,6-trisulfonic acid (ANTS) as fluorophore. 100 μL of the reaction samples were dried in a SpeedVac (Concentrator Plus, Eppendorf) and dissolved in 4 μL 0.05 M ANTS (in DMSO, 15% (V/V) acetic acid) and 4 μL 1 M $NaCNBH_3$ (in DMSO). They were incubated over night at 37 °C before being loaded onto a FACE-Gel [66] and separated at 400 V for 1 h.

### *3.3.12. Polysaccharide extraction*

Polysaccharide was extracted from intracellular fractions of *Polaribacter* sp. Hel_I_88. 200 mL culture were harvested via centrifugation (4,000 x *g*, 20 min, 4 °C) and washed once with 10 mL MOPS buffer before being resuspended in $ddH_2O$. Polysaccharide extraction was carried out according to a protocol modified from literature [67]. In short, attached particles were removed by centrifugation (500 x *g*, 10 min) and cells were lysed by sonication on ice (3 x 2 min, 50% cycle) and two more centrifugation steps (1,100 x *g*, 30 min & 27,000 x *g*, 15 min) to remove unbroken cells and membrane fragments. Three volumes of glycine-buffer (0.2 M, pH 10.5) and two volumes of chloroform (both 4 °C) were added to the supernatant and shaken vigorously for 30 seconds. Phase separation was achieved by centrifugation (100 x *g*, 2 min) and the aqueous phase removed. The remaining organic phase was re-extracted twice with 2 volumes of glycine buffer and all aqueous phases pooled. They were centrifuged

at 47,000 x *g* for 3 h until a gelatinous pellet remained. After resuspending the pellet in 5 mL $ddH_2O$, 6 volumes of ethanol (4 °C) were added to precipitate the polysaccharides over night. Precipitate was centrifuged (14,000 x *g*, 1 h), resolubilized in $ddH_2O$, dialyzed against $ddH_2O$ over night to remove residual salts and dried in a SpeedVac. Extracts were weighed and stored at -20 °C until further use.

### *3.3.13. Bacterial glycan extract characterization*

To determine specific components of the bacterial polysaccharide extracts, 5 mg extract resuspended in PBS were incubated with 25 mg of the characterized enzymes GH13 (1), GH13 (2), GH13 (3) and GH16, respectively. Samples were analyzed by reducing-end assay and FACE as described above. Mono- & oligosaccharide release was compared to samples containing either untreated extract or extract after acid hydrolysis. For acid hydrolysis, 5 mg glycan extract were boiled with 1 M HCl for 2 h, and neutralized using NaOH. Monosaccharide composition of all samples was determined via HPAEC-PAD using a Dionex CarboPac PA10 column (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and monosaccharide mixtures as standards [68].

### *3.3.14. Determination of glucan concentrations on filters*

Polysaccharide extraction was performed from 3 and 0.2 μm bloom filters of environmental spring bloom samples and from bacterial single-cultures. Analysis of the extracts was carried out as described [69]. In short, filters were cut into small pieces and extracted using hot $ddH_2O$ with sonication treatment and debris was removed via centrifugation (4 500 x *g*, 15 min). α-glucan content was determined via incubation with amylase and amyloglucosidase in sodium acetate buffer (0.1 M, pH 4.5) followed by PAHBAH-Assay [70].

## 3.4. Results

### *3.4.1. Sources of α-glucans during algal blooms*

Recently, we conducted a study on the response of free-living planktonic bacteria (0.2-3 μm) to a diatom-dominated spring bloom at Helgoland Roads (54°11'N 7°54'E German Bight). Chlorophyll *a* and microscopic bacterial count data across a three-month time period from the beginning of March to the end of May 2020 revealed a biphasic bloom with a first phase around the end of March until mid-April followed by a main phase from the end of April until

the end of May, during which the dominant flavobacterial clades *Polaribacter* and *Aurantivirga* responded (**Fig. 3.1A&B**).

Diatoms globally produce substantial amounts of β-glucans in the form of laminarin [5], but are not known to produce notable amounts of α-glucans [21]. To identify α-glucan sources, we analyzed 18S rRNA gene amplicon data, obtained from > 10 µm and 3-10 µm biomass size fractions (53 time points). Corresponding to microscopic biovolume data obtained in the framework of the Helgoland Roads time series [17, 71], 18S rRNA gene sequences confirmed the centric diatoms *Dytilum bightwellii* and *Ceratulina pelagica* as dominant algae. Importantly, algae with α-glucans such as *Rhodophyta*, *Chlorophyta* and *Cryptophyceae* [21] were rare (**Fig. 3.1C**). However, *Dinophyceae* (dinoflagellates), also known to contain α-glucans [21], were detected throughout the sampling period and continuously made up 7 to 47% of the eukaryotic population (>10 µm and 3-10 µm fractions). Dinoflagellate autotrophs (e.g. *Karenia* spp.) and heterotrophs (e.g. *Gyrodinium* spp.) were both abundant during the first bloom phase (up to 27% and 20% of the eukaryotic population, respectively), whereas heterotrophic dinoflagellates dominated the main bloom phase (up to 30%), likely in response to higher bacterial cell numbers. Likewise, choanoflagellates, known to feed on bacteria, became more prominent during the main bloom phase (**Fig. 3.1C**, **Tab. S3.1**). Grazing by heterotrophic flagellates therefore likely represents a factor that promotes the release of bacterial storage α-glucan into the DOM pool. This would also corroborate an observed tight correlation between bacteria responding to blooming diatoms and the expression of their α-glucan PULs, in particular in top-expressed metagenome assembled genomes (MAGs) of the dominant responder clades *Aurantivirga* and *Polaribacter* [14] (**Fig. 3.1D**).

### *3.4.2. Marine* Flavobacteriia *degrade different types of α-glucans*

In a previous study, we sequenced 53 coastal North Sea *Flavobacteriia* strains, 75% of which featured α-glucan PULs, more than for any other polysaccharide [19]. Analysis of PUL genes coding for carbohydrate-active enzymes (CAZymes) revealed a variety of gene modules centered around one or more family 13 glycoside hydrolases (GH13) that involved also GH65, GH31 and GH97 genes alongside the characteristic *susCD* gene pair (**Fig. S3.1A**).

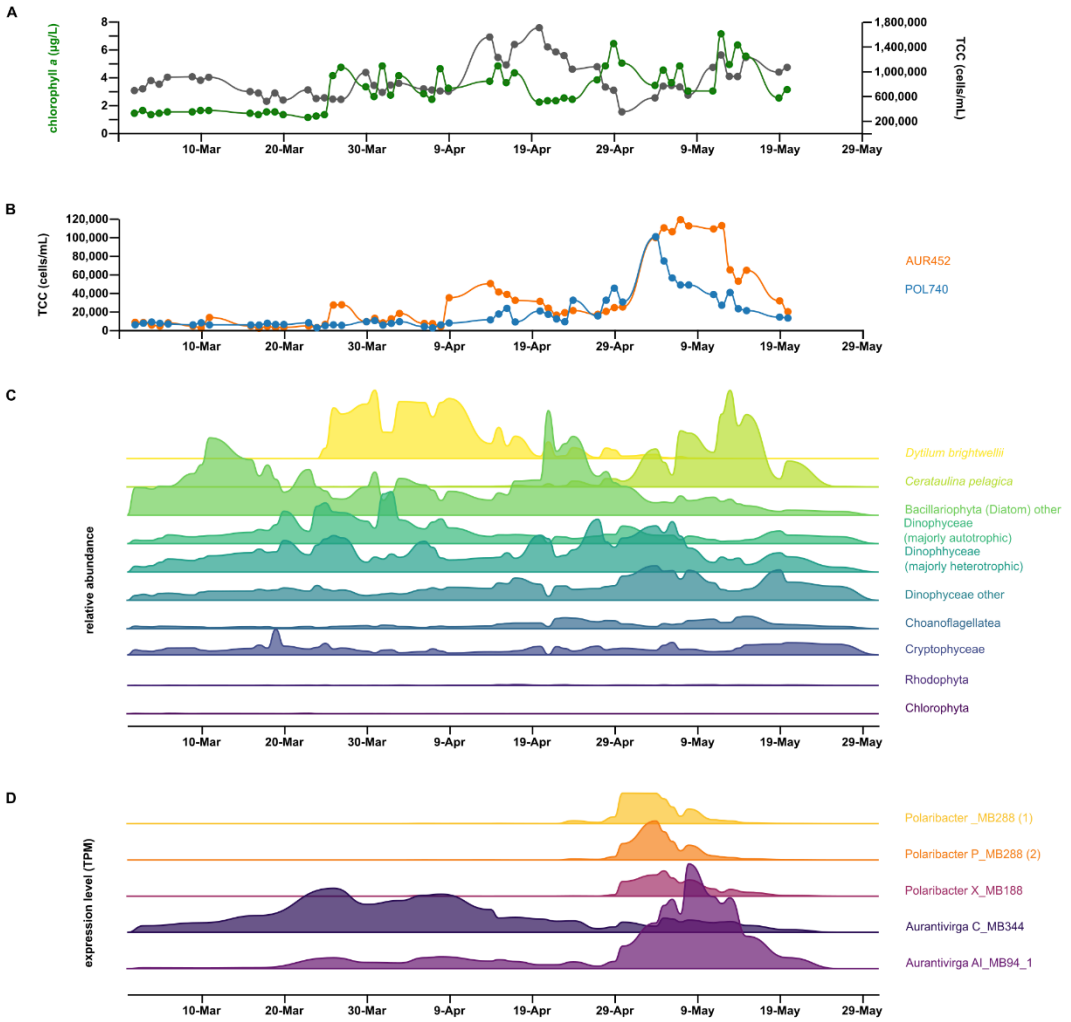



**Figure 3.1. Flavobacterial abundance and activity peaks align with increases in diatom abundance across the 2020 Helgoland spring phytoplankton bloom.** (**A**) Cell counts of total bacteria (TCC) and corresponding chlorophyll *a* measurements. (**B**) Cell counts of the dominant flavobacterial clades *Polaribacter* (POL740) and *Aurantivirga* (AUR452) as detected by FISH. (**C**) Relative abundance of the main eukaryotic taxa as detected by 18S rRNA gene sequencing. Dominant diatom (*Bacillariophyceae*) taxa as well as largely heterotrophic and autotrophic *Dinophyceae* are plotted individually for clarity. Larger zooplankton was removed for this analysis. See also Table S1. (**D**) $\alpha$-glucan PUL expression levels of dominant flavobacterial MAGs belonging to the clades *Polaribacter* and *Aurantivirga*. Expression is given as transcripts per million (TPM) of all PUL-associated CAZymes as well as SusC/D-like proteins combined. Figure was visualized using RawGraphs.

Sequence alignment of these SusD-like substrate binding proteins together with α-glucan-binding SusD proteins from bacterial MAGs of the 2020 spring bloom at Helgoland Roads [17] and the well-characterized α-glucan-binding SusD from *Bacteroides thetaiotaomicron* [72] uncovered two distinct SusD types. Corresponding AlphaFold2 structure predictions revealed a separation into two functional groups, one of which was characterized by an about 17 amino-acid-containing loop close to the binding site, similar to one described in *B. thetaiotaomicron* [73]. The second group lacked this loop along with two residues shown to be substrate-binding in *B. thetaiotaomicron*. The result is a more open binding site that may serve as an adaptation to a structurally distinct α-glucans (**Fig. S3.1B**). MAG-derived SusD sequences (0.2-3 μm fraction) including such with expression during the 2020 bloom clustered with both groups, indicating ecological relevance for both variants (**Fig.** 3.**2**). Gene composition analysis showed that PULs with an open type SusD almost exclusively coded for basic enzymes such as GH13, whereas PULs with a looped SusD comprised a wider variety of CAZymes, most notably more GH13s as well as at least one SusE-like protein. The widest variety of GH13s was found in *Polaribacter* strains, representing one of the most recurrent bloom-associated bacterial clades at Helgoland Roads [12].

We conducted growth experiments with North Sea strains *Polaribacter* sp. Hel_I_88 (PUL with looped SusD) and *Muricauda* sp. MAR_2010_75 (PUL with open SusD). These experiments revealed a correlation between α-glucan substrate complexity and growth efficiency when α-glucan was offered as sole carbon source. While *Polaribacter* sp. grew well on glycogen and pullulan, which has a α-1,4- α-1,4- α-1,6 repeating unit, *Muricauda* sp. showed a strong preference for glycogen and grew poorly on pullulan (**Fig. S3.1C**). This demonstrates the presence of distinct α-glucans niches among marine bacteria that specialize on different types of α-glucans.

### *3.4.3. Marine bacteria contain multiple enzymes targeting α-glucans*

Protein sequence alignment of all PUL-encoded GH13s within the two studied isolates showed the *Polaribacter* sp. enzymes to represent the majority, as roughly 70% of all isolate- and MAG-associated GH13s grouped with them. Additionally, they represented three of the GH13s encoded in highly expressed bloom-associated MAGs (**Fig** 3.**3A, Tab. S3.2**). A notable exception was a GH13_31 with proposed α-1,6 activity. While underrepresented in the isolates, the respective gene was highly expressed during the 2020 bloom, indicating presence of this linkage type in marine bacterial α-glucans.

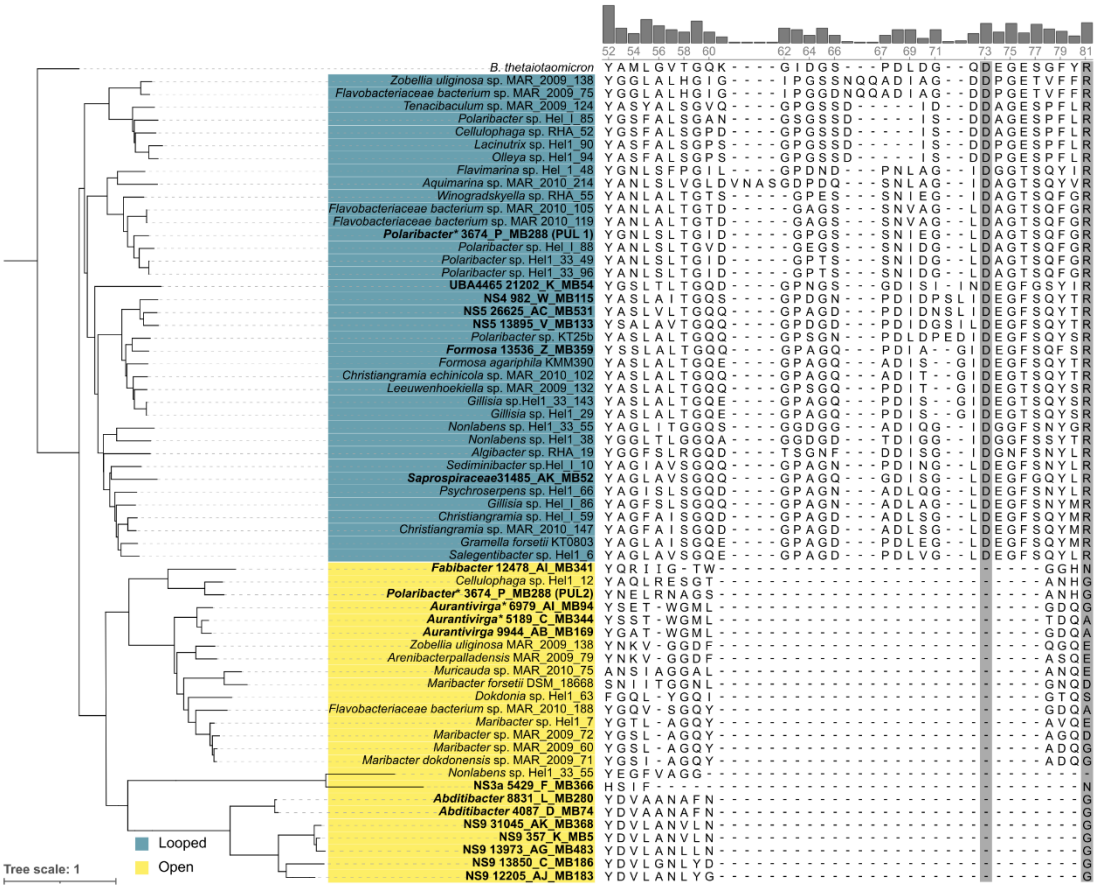



**Figure 3.2. α-glucan PUL clusters according to their SusD-like protein sequences.** Alignment of sequences from North Sea *Flavobacteriia* MAGs from the 2020 spring bloom at Helgoland Roads. α-glucan PUL SusD-like sequences revealed two groups that differed in the presence or absence of a ~17 amino acid loop close to the binding site, leading to one group missing two residues characterized as binding in *B. thetaiotaomicron* (outgroup, grey). MAG sequences are highlighted in bold. An asterisk for such sequences that were found highly expressed in 2020 bloom metatranscriptomes. Conservation of amino acids at their position is represented as bars on top. Numbering corresponds to amino acids of *B. thetaiotaomicron* SusD.

We heterologously expressed three GH13s encoded in the *Polaribacter* sp. Hel_I_88 PUL in *E. coli* and purified the enzymes for biochemical characterization. Via 3,5-dinitrosalicylic acid (DNS) reducing end assay and fluorophore-assisted carbohydrate electrophoresis (FACE), the enzymes GH13A (P161_RS0117435) and GH13C (P161_RS0117455) were shown to act on the α-1,4 linked glucans glycogen and pullulan. GH13A preferred the mostly α-1,4-linked glycogen, producing a dimer and smaller amounts of glucose, whereas GH13C was more

active on pullulan, releasing primarily products with a degree of polymerization (dp) of three. The enzymes were inactive on α-1,6-linked dextran and β-1,3-linked laminarin. Selectivity for alpha-1,4 linkages was confirmed with malto-oligosaccharides. These oligosaccharides were degraded by both enzymes until dp2. α-1,6-linked isomalto-oligosaccharides were not hydrolyzed. Experiments with mixed-link α-1,6/α-1,4 oligosaccharides revealed minor activity on isopanose (α-1,4-α-1,6) and panose (α-1,6-α-1,4), indicating that an α-1,6-bond next to the 1,4 connected glucose monomers in -1 and +1 subsites disables hydrolysis. Additionally, both enzymes acted on β-cyclodextrin, releasing dp1 and dp2 (GH13A) and dp1, 2 and dp3 (GH13C), respectively. Interestingly, similar activity could not be detected on α-cyclodextrin, indicating that the enzymes recognize a specific substrate diameter (**Fig. 3.3B**, **S3.2A&C**).

GH13B (P161_RS0117440) showed only minor activity on glycogen and pullulan, releasing a dimer. Of the tested oligosaccharides with only one linkage type, only such containing α-1,4-linkages were acted upon, but all activities remained minimal. However, from isopanose, GH13B released notable amounts of dp1 and dp2, clearly indicating a preference for α-1,4-bonds situated next to α-1,6 bonds. No activity could be detected on panose, and as the enzyme released only dimers from polysaccharides it can be assumed that the enzyme specifically releases isomaltose from the reducing end (**Fig.** 3.**3B**, **S3.2B**). These results support an adaptation of marine bacteria towards α-1,4/α-1,6 substrates.

### *3.4.4. Marine bacteria synthesize alpha glucans*

α-glucans are known as major storage compounds of marine bacteria. They should therefore be formed during peak bloom phases, when excess organic carbon from algae outweighs the availability of other essential nutrients such as nitrogen, as has been shown under nitrogen limitation *in vitro* [75]. Bacteria synthesize glucose-based storage polysaccharides via proteins encoded in the *glg*-operon, by addition of glucose-1-phosphate or maltose-1-phosphate to ADP-glucose (via GlgA or GlgE, respectively). This results in linear glycogen, which is branched with α-1,6 linkages by the branching enzyme GlgB [76]. Metaproteome analysis of bacteria-dominated 0.2-3 μm filters from spring blooms at Helgoland Roads during 2016, 2018 and 2020 revealed spikes in bacterial *glg*-operon protein abundances. These correlated well with bloom progression as determined by chlorophyll *a* concentration measurements as well as laminarin- and α-glucan targeting protein abundances (**Fig. S3.3**). Except for 2018, where laminarin-degradation proteins remained comparatively low during the bloom (**Fig. S3.3A&D**), a correlation of higher abundances of laminarin-targeting Glg-

proteins and α-glucan uptake proteins could be observed (**Fig. S3.3B&C**). This suggests that α-glucan synthesis is a general mechanism of bloom-associated *Flavobacteriia* in response to growth on laminarin. Consequently, when growing *Polaribacter* sp. Hel_I_88 on laminarin as sole carbon source, we detected a significant increase in α-glucan over time via specific enzymatic hydrolysis (**Fig. S3.4A**). Thus, laminarin degradation coupled with simultaneous α-glucan-synthesis could be confirmed with an isolated strain *in vitro*.

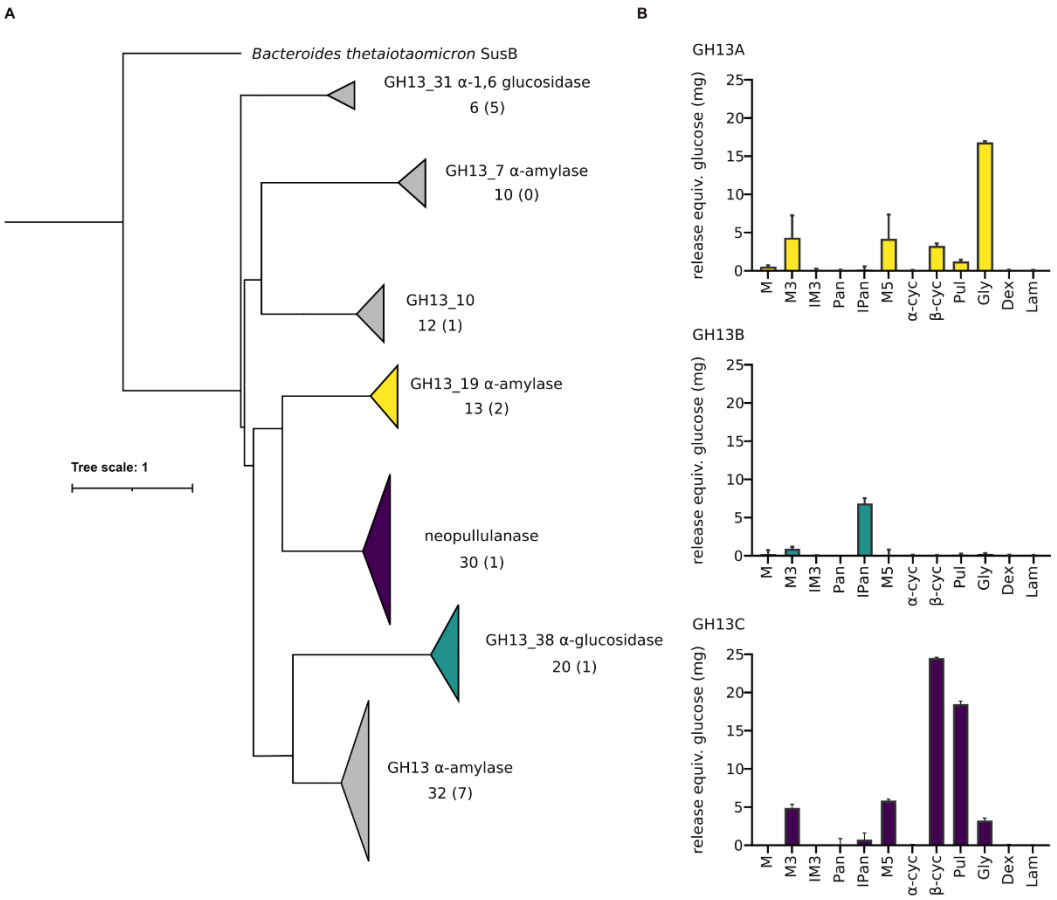



**Figure 3. Phytoplankton bloom associated bacteria encode a multitude of α-glucan-degrading enzymes.** (**A**) Maximum likelihood tree of the main GH13s encoded in the α-glucan PULs of 53 bloom-associated flavobacterial isolates as well as PUL-associated GH13s from the top 50 expressed MAGs of the 2020 Helgoland spring bloom [17]. The tree is rooted to the characterized *B. thetaiotaomicron* glucosidase SusB [74]. Groups are clustered for clarity with arrow size indicating group size. Numbers under the enzyme descriptions represent the number of sequences in the group with the included number of MAG-associated sequences in parenthesis. Colored arrows correspond to (**B**) the characterization of representative *Polaribacter* sp. Hel_I_88 GH13 enzymes. Shown is the mean activity of recombinantly produced enzyme on different oligo- and polysaccharides

measured via DNS-assay (all values corrected against oligo-/polysaccharide without enzyme, n=3). M: Maltose, M3: Maltotriose, IM3: Isomaltotriose, Pan: Panose, IPan: Isopanose, α-cyc: α-cyclodextrin, β-cyc: β-cyclodextrin, Pul: Pullulan, Gly: Glycogen, Dex: Dextrin, Lam: Laminarin. See also Fig. S3.2 for corresponding degradation patterns investigated via FACE.

Proteomics revealed that proteins encoded by the *glg*-operon (P161_RS0109480, RS0109490, RS0109495 & RS0109500) were expressed continuously during growth on laminarin (**Fig. S3.5A**), promoting bacterial α-glucan formation. As expected, overall protein abundance was dominated by the laminarin-PUL (P161_RS0117335-P161_RS0117415) (**Fig. S3.5B**). Yet, both the SusC/D-like protein pair (P161_RS0117480/85) and a GH13 (P161_RS0117500) of the α-glucan PUL became more abundant in later growth phases for which we showed increased amounts of α-glucan present in the culture (**Fig. S5C**). A similar induction could not be shown for proteins of other PULs, such as the alginate PUL (P161_RS0107490- P161_RS0107540) (**Fig. S3.5D**, **Tab. S3.3**). These findings support the view that, as bacteria of the culture begin to lyse, the released organic matter was sensed, taken up and utilized. Analysis of 3 and 0.2 μm filters sampled during the 2020 Helgoland spring bloom showed that α-glucans were more abundant on 0.2 μm filters, which largely represent the free-living planktonic bacterial population. Concentrations rose to around 50 μg/L at the end of March, coinciding with the first bloom event the bacteria responded to. A spike to over 150 μg/L was observed at the beginning of May, coinciding with the main bloom phase (**Fig. S3.4B**). Taken together, these results corroborate that marine bacteria produce significant amounts of α-glucan storage polysaccharide during microalgal blooms.

### *3.4.5. Bacterial polysaccharide contains α-1,4-glucans*

Monosaccharide analysis of polysaccharide extracts from *Polaribacter* sp. Hel_I_88 cultures via high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) revealed high proportions of glucose (29 mol%) (**Fig. 3.4A**). Incubation of this polysaccharide with recombinant *Polaribacter* sp. GH13A, GH13B and GH13C, showed visible degradation in reducing end assays (**Fig. 3.4B**), corroborating that the extract contained α-glucans. FACE-analysis of incubations with GH13A and GH13C yielded oligosaccharides of different dp, but predominantly dp2, which was also the main product formed by incubation of either enzyme with glycogen. This was supported by GH13B releasing only dp2 from the

extract, indicating that the extracted polysaccharide indeed contained bacterial α-glucans (**Fig. 3.4C**).

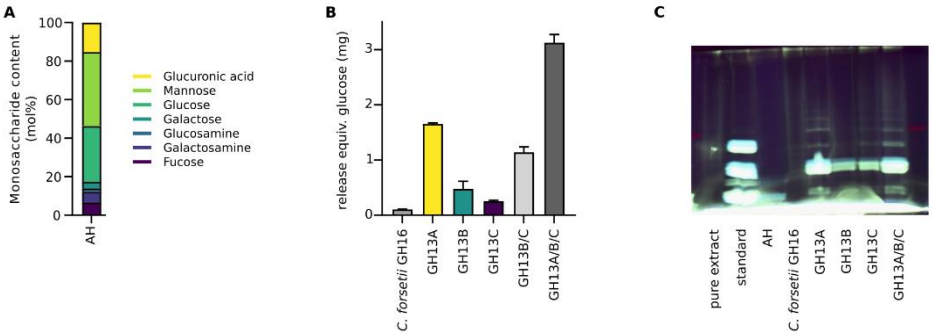



**Figure 3.4. Polysaccharide extracted from *Polaribacter* sp. Hel_I_88 contains high amounts of glucose in the form of α-glucan.** Extract was incubated with different recombinantly expressed *Polaribacter* sp. GH13s or *Christiangramia forsetii* GH16 glycoside hydrolases and analyzed via (**A**) HPAEC-PAD (**B**) reducing end assay or (**C**) FACE. Pure polysaccharide extract and extract after acid hydrolysis (AH) were used as controls. Values of the reducing end assay are corrected against untreated extract (n=3).

### *3.4.6. Bacterial polysaccharide induces α-glucan PUL expression*

*Polaribacter* sp. Hel_I_88 grew on polysaccharide extract from lysed cells as sole carbon source (**Fig. S3.1B**). Comparisons of culture lysate activity to glycogen or alginate-grown cultures showed increased activity on α-1,4-glucan-containing substrates for cultures that were grown on polysaccharide extract (**Fig 3.5A, Fig. S3.6**). Corresponding proteomics revealed an induction of the entire α-glucan PUL (P161_RS0117430-P161_RS0117500) compared to alginate controls, with SusC and SusD proteins making up 0.46% and 1% of the entire proteome, respectively. Interestingly, this induction was higher than for a culture grown on glycogen as positive control, showing that the extracted polysaccharide elicited a more pronounced response (**Fig. 3.5A, Tab. S3.4**).

While growth on extracted polysaccharide could also be observed for *Muricauda* sp. MAR_2010_75 (**Fig. S3.1B**), no significant activity was visible in reducing end assays with culture lysate (**Fig 3.5B**, **Fig. S3.6**). FACE analysis showed the degradation of α-glucan from bacterial lysates under all tested conditions, indicating a basal activity of the α-glucan PUL expression in *Muricauda* sp. MAR_2010_75 but no differential induction of the α-glucan degradation machinery under these conditions. Proteomics revealed the α-glucan SusC- and SusD-like proteins (FG28_RS04375, FG28_RS04380) as most abundant during growth on

polysaccharide extract, but these were also highly expressed during growth on either glycogen or xylan. This high abundance was not mirrored by the PUL's associated CAZymes, a GH13 (FG28_RS04360) and a GH65 (FG28_RS04365), and suggests a difference in α-glucan utilization by bacteria with open-type SusD-containing α-glucan PULs (**Fig. 3.5B, Tab. S3.4**).

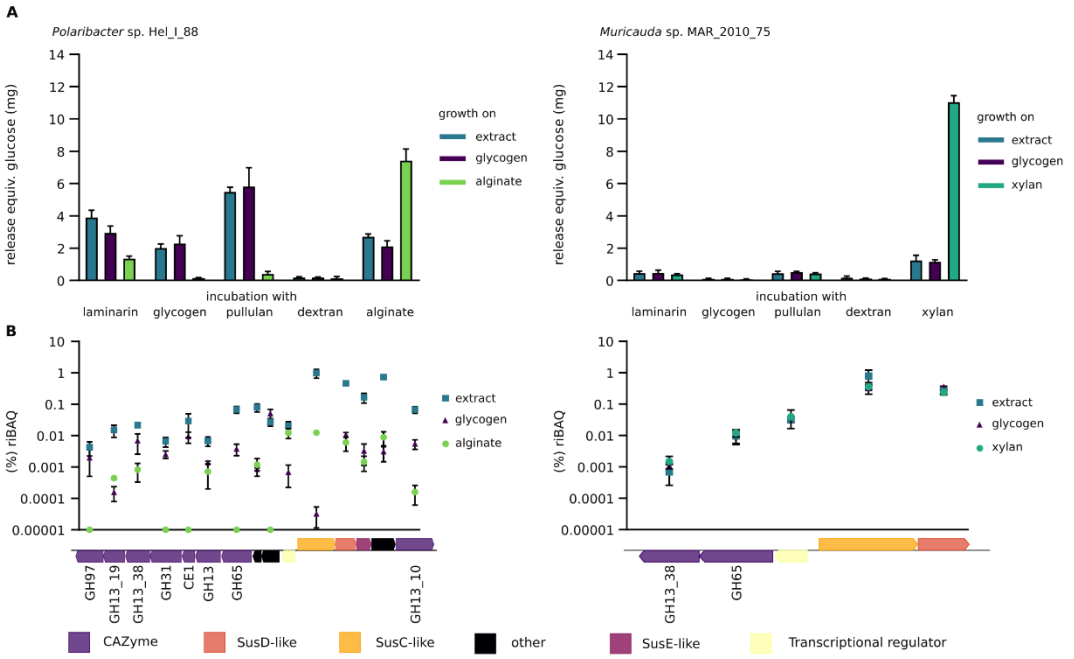



**Figure 5. The α-glucan PUL expression is specifically induced by bacterial α-glucan extracts.** (**A**) Lysate activity and α-glucan (**B**) PUL-encoded protein abundance of *Polaribacter* sp. Hel_I_88 and *Muricauda* sp. MAR_2010_75 grown on extracted bacterial polysaccharide as sole carbon source. Samples were taken from growing cultures (n=3) after 72 h and activity determined by incubating pure culture lysate with different polysaccharides. Alginate and xylan were used as controls, respectively. The α-glucan PULs of both bacteria are depicted under the protein abundances and specific CAZyme annotation are provided underneath.

## 3.5. Discussion

Bacteria release carbon dioxide from algal glycans and other organic molecules in the marine carbon cycle. The high cell densities of bacterioplankton during phytoplankton blooms entail increased mortalities as a consequence of elevated viral lysis and zooplankton predation [16, 77]. The consequence is that bacteria are rather short-lived during bloom events. In fact, it has been estimated that during phytoplankton blooms about half of the bacterial biomass is

recycled on a daily basis [78]. Hence, bacterial glycans are continuously released to the water column, including bacterial α-glucan storage polysaccharides.

As we show, both model bacteria studied can quickly take up and recycle α-glucans. Prevalence and expression of corresponding α-glucan PULs in sampled bloom-associated *Bacteroidota* suggest that this is also a common and highly relevant process *in situ*. Thus, it seems that during diatom-dominated phytoplankton blooms, bacteria employ an α-glucan cycle that is constantly refueled by algal β- and bacterial α-glucans. While viral lysis and zooplankton predation diminish bacterial cell numbers, and predation furthermore shifts bacterial biomass towards higher trophic levels, rapid α-glucan recycling allows bacteria to achieve high growth rates during blooms and thereby to partially offset the loss due to increased mortality rates.

PUL analysis of North Sea *Bacteroidota* revealed two types of α-glucan PULs, a simpler PUL encoding less enzymes and a structurally open SusD (as in *Muricauda* sp. MAR_2010_75) seemingly targeting only simple α-glucans, and a more complex PUL encoding a looped SusD (as in *Polaribacter* sp. Hel_I_88) capable of targeting structurally more complex glucans. The latter is more frequent in environmental metagenomes from bloom-associated North Sea bacteria and also regularly includes a SusE-like protein. SusE was shown to be starch-binding and essential in establishing the SusCD protein complex in *B. thetaiotaomicron* and may play a role in fine-tuning glycan uptake [79]. The two PUL types coded for diverse enzymes around one or more GH13 genes, supporting the view that the glucan substrate exhibits structural variability. We showed activity on different predominantly α-1,4-linked glucans for representatives of three types of GH13s, all of which were present in prevalent MAGs obtained during phytoplankton bloom events. Combined with GH13_31 presence during blooms, which has a proposed α-1,6-hydrolytic activity, this indicates a specific adaptation towards α-1,4/α-1,6-glucan storage polysaccharides of marine bacteria.

Our results suggest that marine α-glucan cycling is unexpectedly complex and provides space for multiple distinct ecophysiological niches. Such niches may also include the decomposition of α-glucans that are produced by *Dinophyceae* [21, 80]. We detected higher abundances of phototrophic *Dinophyceae*, such as *Karenia* spp., during the analyzed 2020 North Sea spring bloom. We suppose that marine bacteria are capable of also utilizing these microalgal α-glucans, as they are structurally similar to the α-glucans that are targeted by the enzymes we investigated [81]. However, close association of α-glucan PUL expression with peaks in diatom abundance lead us to conclude that this was of minor importance during the analyzed diatom-dominated bloom of 2020.

The here described bacterial α-glucan loop is fueled by laminarin, indicated by the detected simultaneous activity of α- and β-glucan PULs in pure culture experiments and during diatom-driven phytoplankton blooms *in situ*. Such intra-population bacterial α-glucan cycling is likely not a feature unique to diatom-dominated blooms, since diatoms are not the only microalgae that produce laminarin. Wide-spread laminarin-synthesizing *Prymnesiophyceae* such as *Phaeocystis* and coccolithophorids such as *Emiliania huxleyi* also form massive blooms [82]. As a consequence, laminarin is one of the most abundant macromolecules on Earth with an annual estimated production 12 ± 8 gigatons and a prevalence of 26 ± 17% in the particulate organic carbon pool [5]. Our data suggest that a substantial fraction of this laminarin is not immediately remineralized, but rather converted to bacterial α-glucans. This suggests that α-glucans likewise represent a significant portion of the marine carbon pool, also ranging in gigatons. Intra-population bacterial α-glucan cycling, in particular during phytoplankton blooms, therefore may constitute a substantial process within the global carbon cycle in terms of carbon turnover and fluxes that has so far not been well-recognized (**Fig. 3.6**).

The role of bacterial necromass turnover has so far been mainly demonstrated for terrestrial microbiomes, such as grassland soil ecosystems [83] or groundwater mesocosms [84]. Recently, the importance of such processes for the marine biogeochemical cycles could also be detected in marine sediments [85, 86]. Here we show that such necromass turnover processes are also relevant during phytoplankton blooms and represent a relevant facet of the marine carbon cycle. Similar to the microalgal β-glucans, abundant bacterial storage α-glucans do not sink, but remain soluble in the water column as accessible energy and carbon sources for planktonic bacteria with specific α-glucan-degradation machineries. Our findings suggest that uptake and recycling of bacterial α-glucans is a wide-spread intra-population energy conservation mechanism of abundant polysaccharide-degrading bacteria during phytoplankton bloom situations in the world's oceans.

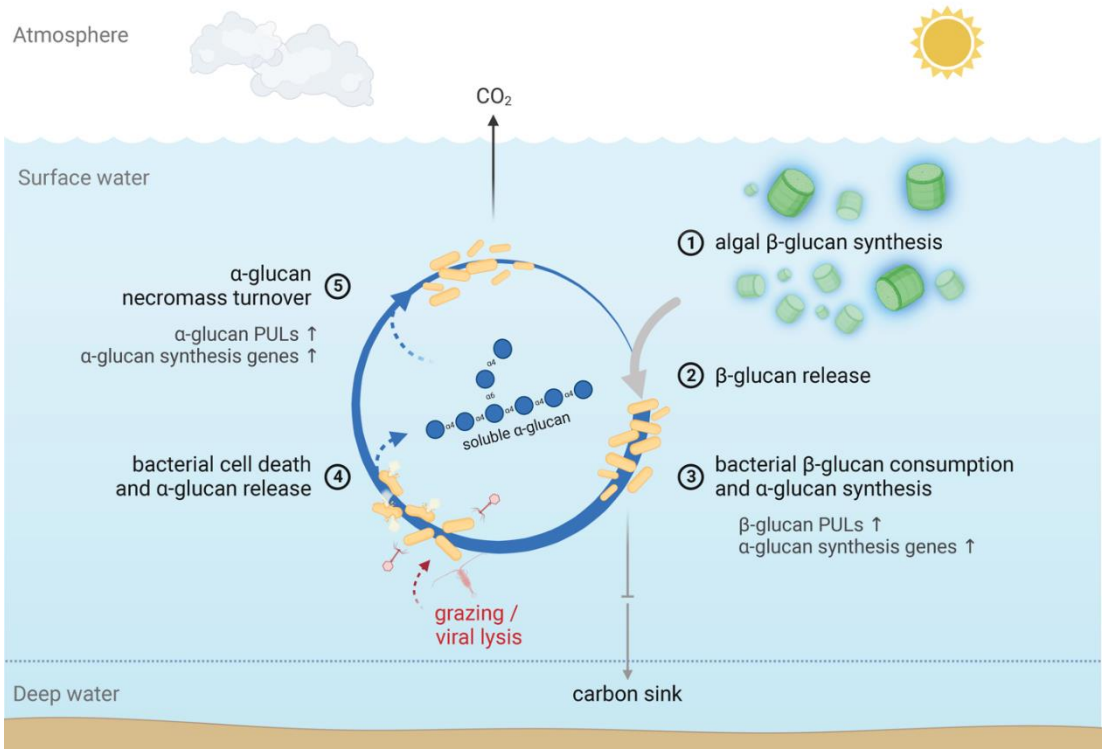



**Figure 3.6. Proposed succession model of the intra-population bacterial glucan flow during phytoplankton blooms.** Microalgae (such as diatoms) produce and release gigatons of laminarin (**1** & **2**). *Flavobacteriia* degrade this carbon source via their PUL-encoded enzymes, leading to the synthesis of bacterial storage α-glucan within the population (**3**). Viral infection and predator feeding cause lyses of a significant part of the microbial population and thus releasing bacterial storage polysaccharides as soluble DOM (**4**). This additional carbon source is recycled by the bacterial population using uptake and degradation pathways specifically adapted towards their own α-glucan storage polysaccharides (**5**). Figure was created with Biorender.

## 3.6. Acknowledgements

We thank the German Research Foundation (DFG) for funding via the Research Unit FOR 2406 'Proteogenomics of Marine Polysaccharide Utilization' (POMPU) by grants to Uwe T. Bornscheuer (BO 1862/17-3), Jan-Hendrik Hehemann (HE 7217/2-3), Rudolf I. Amann (AM 73/9-3) Hanno Teeling (TE 813/2–3), Mia M. Bengtsson (RI 969/9-2), Dörte Becher (BE 3869/4-3) and Thomas Schweder (SCHW 595/10-3, SCHW 595/11-3). This work was supported by a DFG Heisenberg grant and through the Cluster of Excellence "The Ocean Floor — Earth's Uncharted Interface" project 390741603 to Jan-Hendrik Hehemann. We thank the

staff of the Biological Station Helgoland, Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (AWI_BAH_o1) for help with sampling, analyses, logistics, and providing lab space. We are grateful to Jana Matulla and Tina Trautmann for technical assistance and especially thank Lilly Franzmeyer for environmental sampling and sample processing.

## 3.7. References

1. Falkowski PG, Barber RT, Smetacek V. Biogeochemical Controls and Feedbacks on Ocean Primary Production. Science. 1998; 281:200–206.
2. Myklestad S, Haug A. Production of carbohydrates by the marine diatom *Chaetoceros affinis* var. willei (Gran) Hustedt. I. Effect of the concentration of nutrients in the culture medium. Journal of Experimental Marine Biology and Ecology. 1972; 9:125–136.
3. Bligh M, Nguyen N, Buck-Wiese H, Vidal-Melgosa S, Hehemann J-H. Structures and functions of algal glycans shape their capacity to sequester carbon in the ocean. Curr Opin Chem Biol. 2022; 71:102204.
4. Becker S, Tebben J, Coffinet S, Wiltshire K, Iversen MH, Harder T et al. Laminarin is a major molecule in the marine carbon cycle. Proc. Natl. Acad. Sci. USA. 2020; 117:6599–6607.
5. Moran MA, Kujawinski EB, Schroer WF, Amin SA, Bates NR, Bertrand EM et al. Microbial metabolites in the marine carbon cycle. Nature Microbiology. 2022; 7:508–523.
6. Azam F. Microbial Control of Oceanic Carbon Flux: The Plot Thickens. Science. 1998; 280:694–696.
7. Beidler I, Robb CS, Vidal-Melgosa S, Zühlke M-K, Bartosik D, Solanki V et al. Marine bacteroidetes use a conserved enzymatic cascade to digest diatom β-mannan. ISME J. 2023; 17:276–285.
8. Reisky L, Préchoux A, Zühlke M-K, Bäumgen M, Robb CS, Gerlach N et al. A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan. Nat Chem Biol. 2019; 15:803–812.
9. Ferrer-González FX, Widner B, Holderman NR, Glushka J, Edison AS, Kujawinski EB et al. Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. ISME J. 2021; 15:762–773.

10. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J Biol Chem. 2009; 284:24673–24677.
11. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012; 336:608–611.
12. Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. Elife. 2016; 5:e11888.
13. Jeong HJ, Seong KA, Du Yoo Y, Kim TH, Kang NS, Kim S et al. Feeding and grazing impact by small marine heterotrophic dinoflagellates on heterotrophic bacteria. J Eukaryot Microbiol. 2008; 55:271–288.
14. Epstein SS, Shiaris MP. Size-selective grazing of coastal bacterioplankton by natural assemblages of pigmented flagellates, colorless flagellates, and ciliates. Microb Ecol. 1992; 23:211–225.
15. Bartlau N, Wichels A, Krohne G, Adriaenssens EM, Heins A, Fuchs BM et al. Highly diverse flavobacterial phages isolated from North Sea spring blooms. ISME J. 2022; 16:555–568.
16. Bratbak G, Heldal M, Norland S, Thingstad TF. Viruses as partners in spring bloom microbial trophodynamics. Appl Environ Microbiol. 1990; 56:1400–1405.
17. Sidhu C, Kirstein IV, Meunier CL, Rick J, Fofonova V, Wiltshire KH et al. Dissolved storage glycans shaped the community composition of abundant bacterioplankton clades during a North Sea spring phytoplankton bloom. Microbiome. 2023; 11
18. Krüger K, Chafee M, Ben Francis T, Del Glavina Rio T, Becher D, Schweder T et al. In marine Bacteroidetes the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. ISME J. 2019; 13:2800–2816.
19. Kappelmann L, Krüger K, Hehemann J-H, Harder J, Markert S, Unfried F 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.
20. Francis TB, Bartosik D, Sura T, Sichert A, Hehemann J-H, Markert S et al. Changing expression patterns of TonB-dependent transporters suggest shifts in polysaccharide consumption over the course of a spring phytoplankton bloom. ISME J. 2021; 15:2336–2350.
21. Suzuki E, Suzuki R. Variation of Storage Polysaccharides in Phototrophic Microorganisms. J. Appl. Glycosci. 2013; 60:21–27.

22. Wiltshire KH, Kraberg A, Bartsch I, Boersma M, Franke H-D, Freund J et al. Helgoland Roads, North Sea: 45 Years of Change. Estuaries and Coasts. 2010; 33:295–310.
23. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7:539.
24. Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics. 2012; 28:1166–1167.
25. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019; 47:W256-W259.
26. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596:583–589.
27. Zapata M, Rodríguez F, Garrido JL. Separation of chlorophylls and carotenoids from marine phytoplankton: a new HPLC method using a reversed phase C8 column and pyridine-containing mobile phases. Marine Ecology Progress Series. 2000; 195:29–45.
28. Garrido JL, Rodríguez F, Campaña E, Zapata M. Rapid separation of chlorophylls a and b and their demetallated and dephytylated derivatives using a monolithic silica C18 column and a pyridine-containing mobile phase. J Chromatogr A. 2003; 994:85–92.
29. Wiltshire K, Harsdorf S, Smidt B, Blöcker G, Reuter R, Schroeder F. The determination of algal biomass (as chlorophyll) in suspended matter from the Elbe estuary and the German Bight: A comparison of high-performance liquid chromatography, delayed fluorescence and prompt fluorescence methods. Journal of Experimental Marine Biology and Ecology. 1998; 222:113–131.
30. Ray JL, Althammer J, Skaar KS, Simonelli P, Larsen A, Stoecker D et al. Metabarcoding and metabolome analyses of copepod grazing reveal feeding preference and linkage to metabolite classes in dynamic microbial plankton communities. Mol Ecol. 2016; 25:5585–5602.
31. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods. 2016; 13:581–583.
32. R. Siebers, D. Schultz, M. S. Farza, A. Brauer, D. Zühlke, K. Hoff et al. Sulfate reducing bacteria are active and co-occur with diatoms during a spring phytoplankton bloom. bioRxiv. 2023

33. Gómez F, Moreira D, López-García P. Life cycle and molecular phylogeny of the dinoflagellates *Chytriodinium* and *Dissodinium*, ectoparasites of copepod eggs. European Journal of Protistology. 2009; 45:260–270.
34. Kyeong Ah Seong, Hae Jin Jeong, Shin Kim, Gwang Hoon Kim, Jung Hoon Kang. Bacterivory by co-occurring red-tide algae, heterotrophic nanoflagellates, and ciliates. Marine Ecology Progress Series. 2006; 322:85–97.
35. Takano Y, Horiguchi T. Surface ultrastructure and molecular phylogenetics of four unarmored heterotrophic dinoflagellates, including the type species of the genus Gyrodinium (Dinophyceae). Phycological Research. 2004; 52:107–116.
36. McKay L, Kamykowski D, Milligan E, Schaeffer B, Sinclair G. Comparison of swimming speed and photophysiological responses to different external conditions among three Karenia brevis strains. Harmful Algae. 2006; 5:623–636.
37. Tamames J, Puente-Sánchez F. SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline. Front Microbiol. 2018; 9:3349.
38. Rho M, Tang H, Ye Y. FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res. 2010; 38:e191.
39. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2016; 44:D67-D72.
40. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019; 47:D309-D314.
41. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28:27–30.
42. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009; 37:D233-8.
43. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods. 2021; 18:366–368.
44. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016; 44:D279-85.
45. Eddy SR. A new generation of homology search tools based on probabilistic inference. Genome Inform. 2009; 23:205–211.

46. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012; 9:357–359.
47. Deusch S, Seifert J. Catching the tip of the iceberg - evaluation of sample preparation protocols for metaproteomic studies of the rumen microbiota. Proteomics. 2015; 15:3590–3595.
48. Bonn F, Bartel J, Büttner K, Hecker M, Otto A, Becher D. Picking vanished proteins from the void: how to collect and ship/share extremely dilute proteins in a reproducible and highly efficient manner. Anal. Chem. 2014; 86:7421–7427.
49. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006; 22:1658–1659.
50. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999; 20:3551–3567.
51. Searle BC. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. Proteomics. 2010; 10:1265–1269.
52. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics. 2004; 20:1466–1467.
53. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 2002; 74:5383–5392.
54. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. Anal. Chem. 2003; 75:4646–4658.
55. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016; 428:726–731.
56. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022; 50:D543-D552.
57. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017; 11:2864–2868.
58. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014; 30:2068–2069.

59. Zheng J, Ge Q, Yan Y, Zhang X, Le Huang, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023; 51:W115-W121.
60. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020; 36:2251–2252.
61. Hahnke RL, Harder J. Phylogenetic diversity of Flavobacteria isolated from the North Sea on solid media. Syst Appl Microbiol. 2013; 36:497–504.
62. Schut F, Vries EJ de, Gottschal JC, Robertson BR, Harder W, Prins RA et al. Isolation of Typical Marine Bacteria by Dilution Culture: Growth, Maintenance, and Characteristics of Isolates under Laboratory Conditions. Appl Environ Microbiol. 1993; 59:2150–2160.
63. Otto A, Bernhardt J, Meyer H, Schaffer M, Herbst F-A, Siebourg J et al. Systems-wide temporal proteomic profiling in glucose-starved *Bacillus subtilis.* Nat Commun. 2010; 1:137.
64. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology. 2008; 26:1367–1372.
65. Miller GL. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. Anal. Chem. 1959; 31:426–428.
66. Hehemann J-H, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature. 2010; 464:908–912.
67. Sambou T, Dinadayala P, Stadthagen G, Barilone N, Bordat Y, Constant P et al. Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis:* biosynthesis and impact on the persistence in mice. Mol Microbiol. 2008; 70:762–774.
68. Engel A, Händel N. A novel protocol for determining the concentration and composition of sugars in particulate and in high molecular weight dissolved organic matter (HMW-DOM) in seawater. Marine Chemistry. 2011; 127:180–191.
69. Steinke N, Vidal-Melgosa S, Schultz-Johansen M, Hehemann J-H. Biocatalytic quantification of α-glucan in marine particulate organic matter. MicrobiologyOpen. 2022; 11:e1289.
70. Lever M. A new reaction for colorimetric determination of carbohydrates. Anal Biochem. 1972; 47:273–279.

71. Armonies W, Asmus H, Buschbaum C, Lackschewitz D, Reise K, Rick J. Microscopic species make the diversity: a checklist of marine flora and fauna around the Island of Sylt in the North Sea. Helgoland Marine Research. 2018; 72:11.
72. Foley MH, Cockburn DW, Koropatkin NM. The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. Cell Mol Life Sci. 2016; 73:2603–2617.
73. Koropatkin NM, Martens EC, Gordon JI, Smith TJ. Starch Catabolism by a Prominent Human Gut Symbiont Is Directed by the Recognition of Amylose Helices. Structure. 2008; 16:1105–1115.
74. Gloster TM, Turkenburg JP, Potts JR, Henrissat B, Davies GJ. Divergence of Catalytic Mechanism within a Glycosidase Family Provides Insight into Evolution of Carbohydrate Metabolism by Human Gut Flora. Chemistry & Biology. 2008; 15:1058–1067.
75. Barbeyron T, Thomas F, Barbe V, Teeling H, Schenowitz C, Dossat C 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. 2016; 18:4610–4627.
76. Preiss J. Bacterial Glycogen Synthesis and its Regulation. Annual Review of Microbiology. 1984; 38:419–458.
77. Pernthaler J. Predation on prokaryotes in the water column and its ecological implications. Nature Reviews Microbiology. 2005; 3:537–546.
78. Brüwer Jan D., Orellana Luis H., Sidhu Chandni, Klip Helena C. L., Meunier Cédric L., Boersma Maarten et al. In situ cell division and mortality rates of SAR11, SAR86, Bacteroidetes, and Aurantivirga during phytoplankton blooms reveal differences in population controls. mSystems. 2023; 0:e01287-22.
79. Foley MH, Martens EC, Koropatkin NM. SusE facilitates starch uptake independent of starch binding in B. thetaiotaomicron. Mol Microbiol. 2018; 108:551–566.
80. Osbeck CMG, Lundin D, Karlsson C, Teikari JE, Moran MA, Pinhassi J. Divergent gene expression responses in two Baltic Sea heterotrophic model bacteria to dinoflagellate dissolved organic matter. PLOS ONE. 2022; 17:e0243406.
81. Plancke C, Colleoni C, Deschamps P, Dauvillée D, Nakamura Y, Haebel S et al. Pathway of cytosolic starch synthesis in the model glaucophyte Cyanophora paradoxa. Eukaryot Cell. 2008; 7:247–257.
82. Hopkins J, Henson SA, Painter SC, Tyrrell T, Poulton AJ. Phenological characteristics of global coccolithophore blooms. Global Biogeochem. Cycles. 2015; 29:239–253.

83. Greenlon Alex, Sieradzki Ella, Zablocki Olivier, Koch Benjamin J., Foley Megan M., Kimbrel Jeffrey A. et al. Quantitative Stable-Isotope Probing (qSIP) with Metagenomics Links Microbial Physiology and Activity to Soil Moisture in Mediterranean-Climate Grassland Ecosystems. mSystems. 2022; 7:e00417-22.
84. Geesink P, Taubert M, Jehmlich N, Bergen M von, Küsel K. Bacterial Necromass Is Rapidly Metabolized by Heterotrophic Bacteria and Supports Multiple Trophic Levels of the Groundwater Microbiome. Microbiol Spectr. 2022; 10:e0043722.
85. Pelikan C, Wasmund K, Glombitza C, Hausmann B, Herbold CW, Flieder M et al. Anaerobic bacterial degradation of protein and lipid macromolecules in subarctic marine sediment. ISME J. 2021; 15:833–847.
86. Orsi WD, Richards TA, Francis WR. Predicted microbial secretomes and their target substrates in marine sediment. Nature Microbiology. 2018; 3:32–37.

## 3.8. Competing interests

The authors declare no conflict of interest.

## 3.9. Supplementary information

### *3.9.1. Supplementary tables*

Large supplementary tables can be viewed online at https://doi.org/10.21203/rs.3.rs-3205445/v1.

**Table S1. 18S rDNA read counts of the Helgoland spring phytoplankton bloom from March to May 2020.** Read counts from 3 and 10 μm fractions (Sheets 1 & 2) were summed. Relative abundances of groups of significant eukaryotes were plotted in **Figure 1**.

**Table S2. List of all isolate and MAG-encoded PUL-associated GH13**

**Table S3. Proteomics of *Polaribacter* sp. Hel_I_88 grown on laminarin** with samples taken at 18, 24 and 48 h.

**Table S4. Proteomics of *Polaribacter* sp. Hel_I_88 and *Muricauda* sp. MAR_2010_75 grown on extracted polysaccharides from *Polaribacter* sp**. as well as glycogen, alginate and xylan as controls.

### *3.9.2. Supplementary Figures*

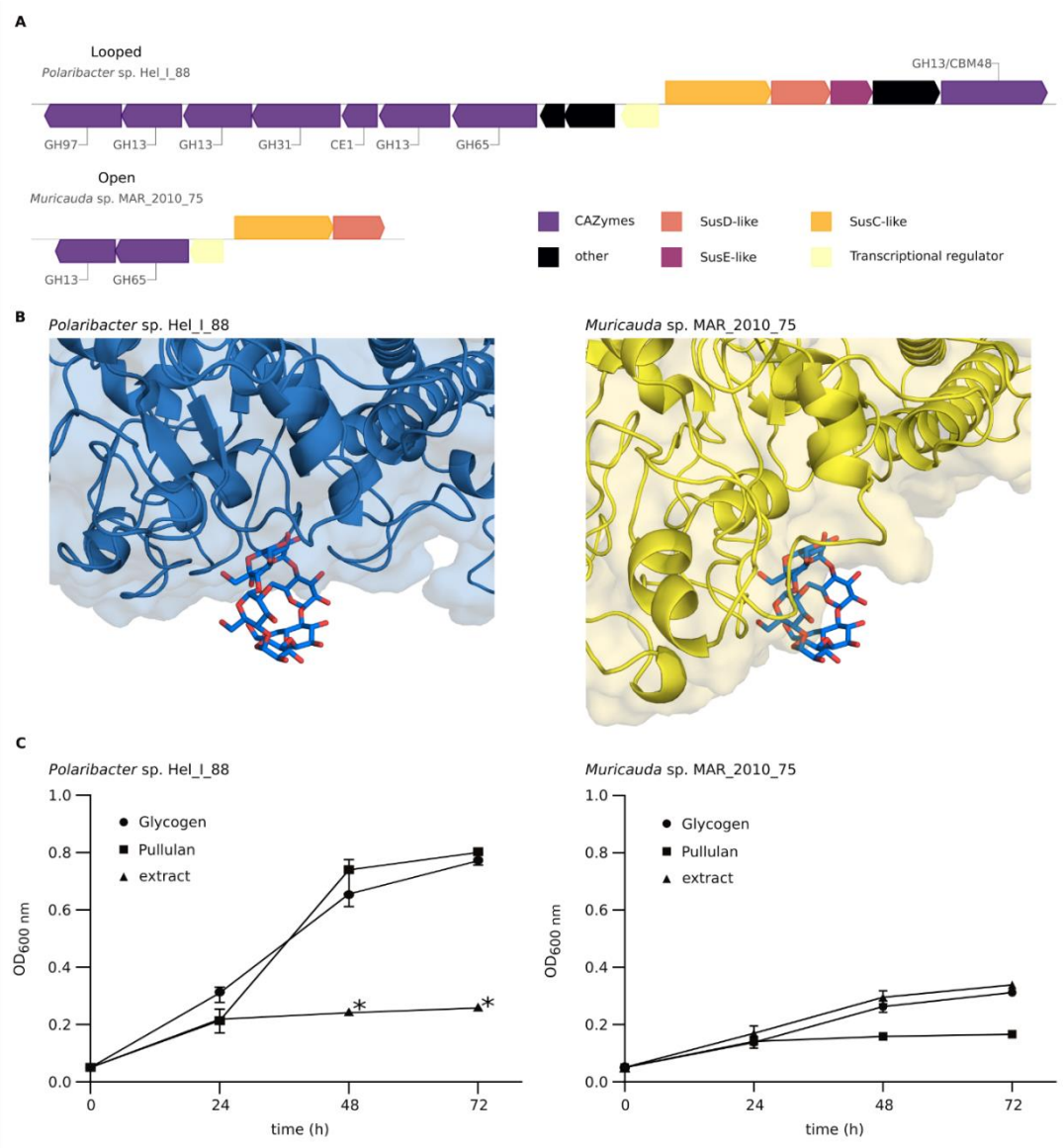



**Figure S3.1. Differences in PUL complexity, predicted SusD-structure and growth efficiency on α-glucan-containing substrates between *Polaribacter* sp. Hel_I_88 and *Muricauda* sp. MAR_2010_75.** (**A**) PUL composition of *Polaribacter* sp. (looped SusD) and *Muricauda* sp. (open SusD), representing differences between PULs belonging to the groups outlined in Figure 2. (**B**) AlphaFold structure prediction of the encoded SusDs. Shown is a surface model alongside the structure cartoon overlayed with a cyclodextrin

substrate as observed in *B. thetaiotaomicron* (PBD: 3CK8). (**C**) Growth of both strains on glycogen, pullulan and extracted intracellular polysaccharide of *Polaribacter* sp. Growth was determined (n=3) in marine minimal medium (MPM) containing the tested polysaccharide or extract as sole carbon source. *indicates the formation of aggregates during growth, hindering an accurate measurement.

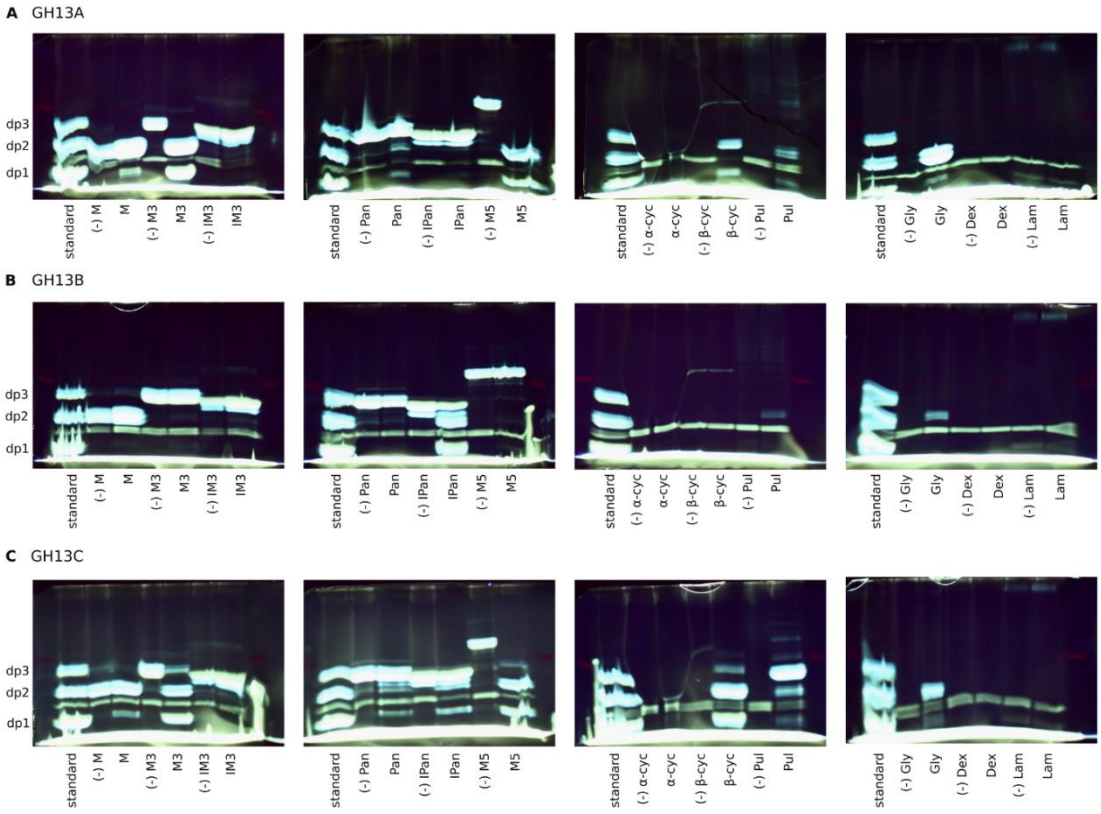



**Figure S3.2. FACE-analysis of GH13s encoded in the *Polaribacter* sp. PUL.** Degradation profiles of (**A**) GH13A, (**B**) GH13B and (**C**) GH13C on different poly- and oligosaccharides. M: Maltose, M3: Maltotriose, IM3: Isomaltotriose, Pan: Panose, IPan: Isopanose, α-cyc: α-cyclodextrin, β-cyc: β-cyclodextrin, Pul: Pullulan, Gly: Glycogen, Dex: Dextrin, Lam: Laminarin.

Top next page:

**Figure S3.3. Glg-protein abundance relates to those of α-glucan and β-glucan-targeting proteins.** (**A**) Progression of three sperate Helgoland bloom events (2016, 2018, 2020) as observed via chlorophyll *a* measurements. For each bloom, metaproteomes of the 0.2 μm fraction were searched for proteins of the Glg-pathway (**B**), α-glucan PUL (**C**) and β-glucan

PUL-encoded (**D**) proteins. Relative abundances are given as normalized weighted spectra of the bacterial fraction (BacNWS%).

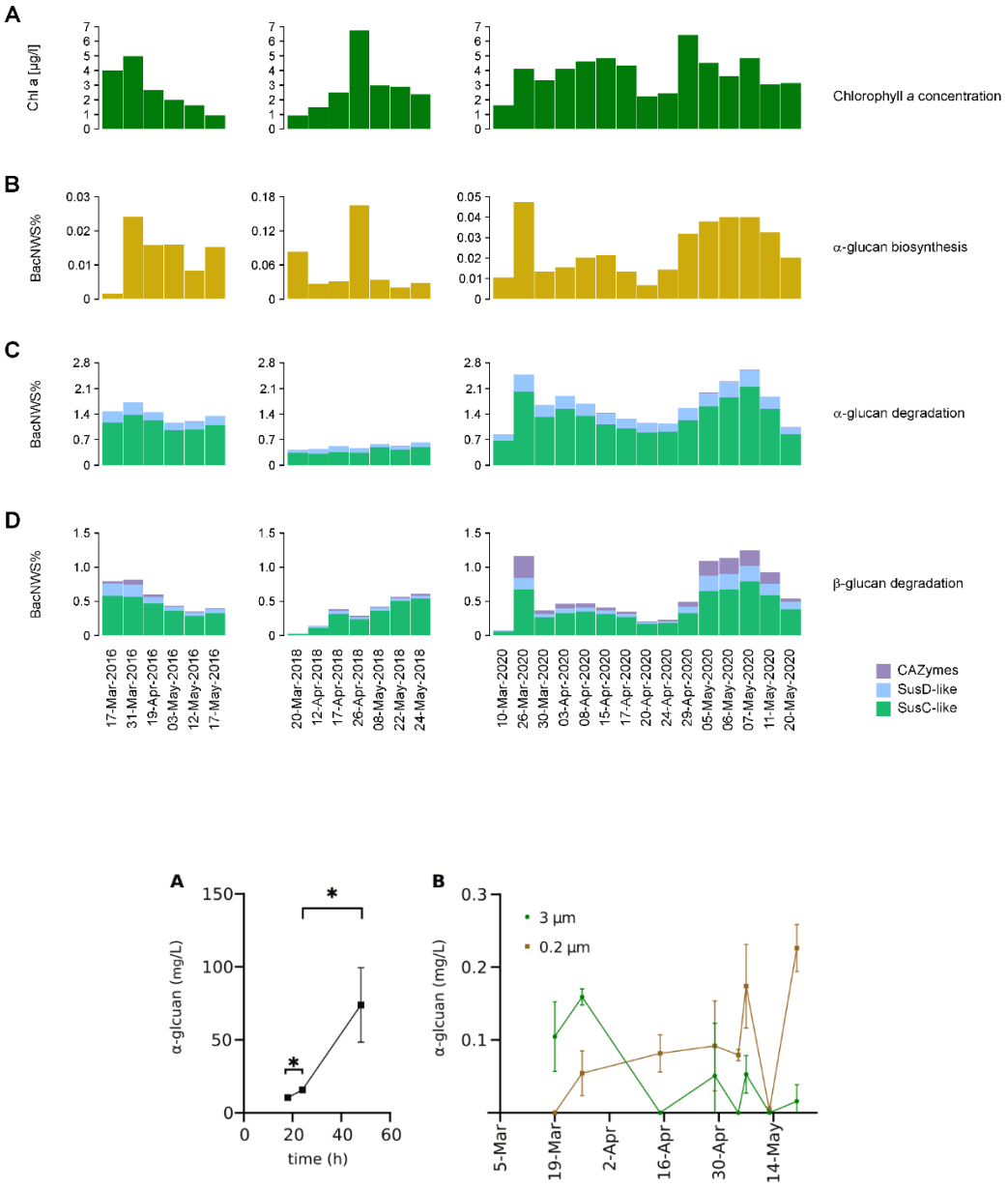



**Figure S3.4. Bacteria produce significant amounts of α-glucan during the 2020 phytoplankton bloom.** Accumulated α-glucan amounts from (**A**) three biological replicates of *Polaribacter* sp. Hel_I_88 grown on laminarin as sole carbon source (* = unpaired t-test P value < 0.05) and (**B**) 0.2 µm and 3 µm fraction 2020 Helgoland bloom filters (three technical replicates) were determined via specific enzymatic hydrolysis followed by reducing end assay. All values are corrected for volume filtered.

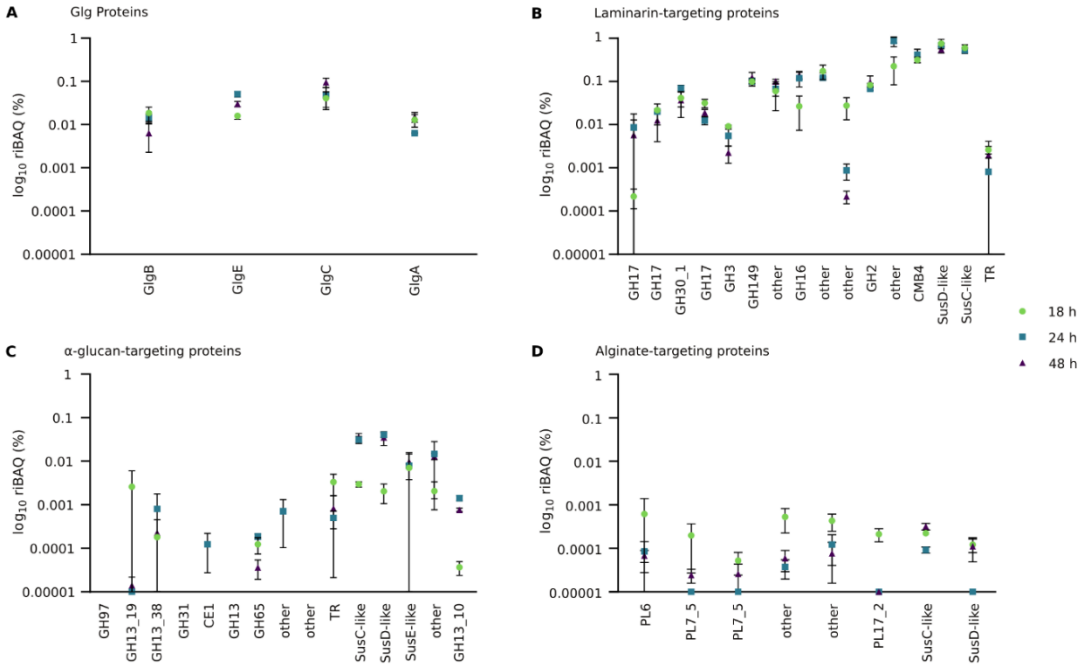



**Figure S3.5. During growth on laminarin, *Polaribacter* sp. shows expression of proteins responsible for glycogen synthesis and degradation.** *Polaribacter* sp. was grown on laminarin as sole carbon source and samples were taken at 18, 24 and 48 h, respectively. Protein abundances of Glg-proteins (**A**), laminarin PUL proteins (**B**), α-glucan PUL proteins (**C**) and alginate PUL proteins (**D**) are shown over time. Abundances are given as manually calculated riBAQ (%) values.

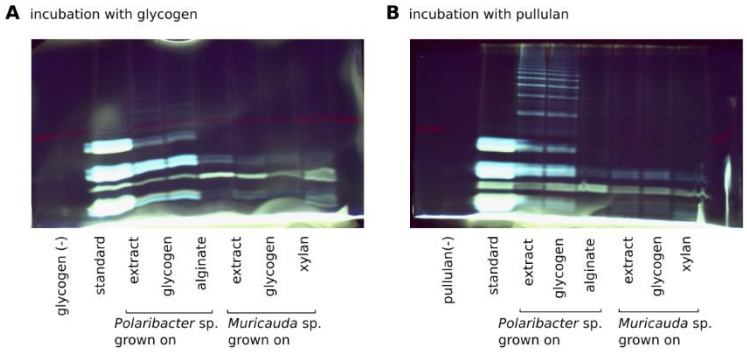



**Figure S3.6. FACE-based enzyme activity assays of culture lysates of the two model bacteria *Polaribacter* sp. Hel_I_88 and *Muricauda* sp. MAR_2010_75 grown with different polysaccharides.** Lysate of cultures grown on extracted bacterial α-glucans, glycogen and alginate or xylan were incubated (**A**) with glycogen and (**B**) with pullulan to test for substrate-induced differential PUL expression activities.

# 4. Marine bacteroidetes use a conserved enzymatic cascade to digest diatom β-mannan

**Authors**: Irena Beidler[1,] Craig S. Robb[2,3], Silvia Vidal-Melgosa[2,3], Marie-Katherin Zühlke[1,4], Daniel Bartosik[1,4], Vipul Solanki[2,3], Stephanie Markert[1,4], Dörte Becher[5], Thomas Schweder[1,4], Jan-Hendrik Hehemann[2,3]

[1] Pharmaceutical Biotechnology, Institute of Pharmacy, University Greifswald, 17489 Greifswald, Germany

[2] Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany

[3] University of Bremen, Center for Marine Environmental Sciences, MARUM, 28359 Bremen, Germany

[4] Institute of Marine Biotechnology, 17489 Greifswald, Germany

[5] Institute of Microbiology, University Greifswald, 17489 Greifswald, Germany

Published in *The ISME Journal*, Volume 2, 2023, Pages 276-285

## 4.1. Abstract

The polysaccharide β-mannan, which is common in terrestrial plants but unknown in microalgae, was recently detected during diatom blooms. We identified a β-mannan polysaccharide utilization locus (PUL) in the genome of the marine flavobacterium Muricauda sp. MAR_2010_75. Proteomics showed β-mannan induced translation of 22 proteins encoded within the PUL. Biochemical and structural analyses deduced the enzymatic cascade for β-mannan utilization. A conserved GH26 β-mannanase with endo-activity depolymerized the β-mannan. Consistent with the biochemistry, X-ray crystallography showed the typical TIM-barrel fold of related enzymes found in terrestrial β-mannan degraders. Structural and biochemical analyses of a second GH26 allowed the prediction of an exo-activity on shorter manno-gluco oligosaccharides. Further analysis demonstrated exo-α-1,6-galactosidase- and endo-β-1,4-glucanase activity of the PUL-encoded GH27 and GH5_26, respectively, indicating the target substrate is a galactoglucomannan. Epitope-deletion assays with mannanases as analytic tools indicate the presence of β-mannan in the diatoms Coscinodiscus *wailesii* and *Chaetoceros affinis*. Mannanases from the PUL were active on diatom β-mannan and

polysaccharide extracts sampled during a microalgal bloom at the North Sea. Together these results demonstrate that marine microorganisms use a conserved enzymatic cascade to degrade β-mannans of marine and terrestrial origin and that this metabolic pathway plays a role in marine carbon cycling.

## 4.2. Introduction

Polysaccharides are abundant biomolecules in the marine carbon cycle. By combining various monosaccharides, glycosidic linkage configurations, connections and functional groups, polysaccharides create molecular diversity [1]. They act as energy storage, cell wall components, extracellular defense and communication agents [2–6]. To yield energy and carbon from the diversity of existing polysaccharides, microorganisms use a proportionally complex arsenal of carbohydrate-active enzymes (CAZymes) targeting the glycosidic linkages and other chemical groups. CAZymes include the enzyme classes glycoside hydrolases (GHs) and carbohydrate esterases (CEs), which are further classified in families according to the CAZy database by their sequence, fold and substrate specificity [7]. CAZymes catalyzing degradation of a type of polysaccharide are often organized in genomic islands, named polysaccharide utilization loci (PULs). Alongside CAZymes, these clusters contain genes for specific uptake and recognition (SusC- and SusD-like) proteins as well as transcriptional regulators [8]. Together, they form highly specific uptake machineries for a multitude of substrates.

Members of the phylum Bacteroidetes have been studied as polysaccharide degraders of terrestrial ecosystems especially in the human gut [8, 9]. They are also prominent in aquatic habitats, especially during marine algae blooms. Microalgae bloom in near-surface, nutrient-rich regions and contribute to the global carbon cycle by fixing carbon on a scale roughly equal with terrestrial phototrophs [10, 11]. Most of this biomass is re-mineralized by heterotrophic bacteria, such as members of the Flavobacteriia within the Bacteroidetes. Metagenomics and metaproteomics of such bacterioplankton revealed a genetic potential for the decomposition of structurally diverse polysaccharides [12–14]. Although this genetic potential appears remarkably similar to human gut bacteroidetes and their well-studied role in food polysaccharide breakdown, many marine polysaccharides relevant in microalgal blooms remain to be characterized.

Recently, we discovered over 20 different polysaccharide structures in particulate organic matter (POM) and high molecular weight dissolved organic matter (HMWDOM) during spring

blooms around the North Sea island of Helgoland [15], including a β-1,4-mannan. β-Mannans are a complex group of glycans found in plants as homo- and heteropolymers. While homomannan consists of a linear backbone of β-(1,4)-linked mannose [16], β-mannans can also exists as a backbone of β-(1,4)-linked mannose and glucose, known as glucomannan [17]. Both linear mannan and glucomannan can have α-(1,6)-galactosyl branch points and are then referred to as galactomannan and galactoglucomannan, respectively [18]. Additionally, some mannans are acetylated at O2 or O3 of the mannose monomer [19, 20].

The presence of mannan during microalgae blooms suggested that this substrate might be the potential target for a PUL type with multiple GH26 β-mannanases previously identified in marine isolates (Salegentibacter sp. Hel_1_6, Leeuwenhoekiella sp. MAR_2009_132 and Sediminibacter sp. Hel_1_10). These bacteria were shown to also consume terrestrial mannan, leading to the expression of PULs related to those in the human gut symbiont B. ovatus specific for galactomannan [21, 22]. The marine PULs, however, have additional enzymes such as a predicted GH27 α-galactosidase and a predicted glucanase (GH5), indicating adaptation to a more complex polysaccharide. Here we show that the marine *Muricauda* sp. MAR_2010_75 strain possesses a conserved PUL, which is specifically induced by β-mannan. The core enzymes of the PUL catalyze β-mannan utilization and can degrade β-mannan from HMWDOM sampled during a diatom bloom at the North Sea as well as β-mannans extracted from diatoms of the genera Coscinodiscus and Chaetoceros, both relevant to marine microalgae blooms.

## 4.3. Materials and methods

### *4.3.1. Comparative genomics*

RefSeq assemblies of genomes deposited in MarRef (v1.7) and MarDB (v1.6) [23] were downloaded from NCBI web server [24] and screened for co-occurrence of *Muricauda* sp. MAR_2010_75 β-mannan PUL coding sequences using the “hmm” search function of cblaster (v1.3.14) [25] with the HMM profiles “TIGR04056” (SusC-like), “PF12741.10”, “PF12771.10”, “PF14322.9” and “PF07980.14” (all SusD-like) as well as “GH5”, “GH26”, “GH27”, “GH130” and “CE2” from the dbCAN-HMMdb-V10 database [26]. Assemblies containing putative clusters with at least one GH26 and a SusCD-pair were kept for further analysis. CAZymes were annotated using the hmmscan function of HMMer (v3.3.2) against the dbCAN-HMMdb-V10 database. Results were parsed using the hmmscan-parser.sh script provided by dbCAN and additionally confirmed using Protein-Protein BLAST (v2.11.0+) [27] against the CAZyDB

(release 09242021) with an e-value threshold of E-20, minimum query coverage of 40% and at least 30% sequence identity [28]. PULs were extracted by screening the genetic context of SusCD-pairs for CAZyme annotations using a seven-gene frame, excluding glycosyl transferases. SusCD genes and CAZyme repertoire of previously published MAGs from 2010 to 2012 and 2016 (available in ENA project PRJEB28156, see [29] for details) were predicted as described above, regardless of the genetic context. Results were visualized with UpSetR [30, 31] and Circos [32]. Synteny between selected reference PULs and the *Muricauda* sp. MAR_2010_75 PUL were analyzed using Protein-Protein BLAST with an e-value threshold of E-5. For phylogenetic analysis of taxa encoding β-mannan targeting PULs, rpoB-sequences were aligned using T-Coffee web service of EMBL-EBI [33] (v13.41.0.28bdc39) with default settings. Maximum-likelihood phylogeny was estimated by PhyML 3.0 [34] using automatic model selection by SMS [35] with Bayesian Information Criterion. The resulting tree was visualized using iTOL [36].

### *4.3.2. Diatom isolation*

An algae net sample with a cut-off of 80 µm was collected from the Helgoland Bight (54˚ 11.3' N, 7˚ 54.0' E) in 2017. From this sample, Coscinodiscus wailesii cells were transferred to 24 well plates (Thermo Fisher Scientific, Waltham, MA, USA) in F/2 media at room temperature (RT) with a 12:12 h light:dark cycle and transferred until no other diatoms were growing in co-culture. Once in monoculture, they were routinely cultured in 25 mL tissue culture flasks under the same conditions.

### *4.3.3. Strain and cultivation conditions*

*Muricauda* sp. MAR_2010_75 was isolated as previously reported [14, 37]. Growth experiments were performed in synthetic seawater medium (MPM) [38] with 0.2% (w/v) defined carbon sources at 21 °C and 200 rpm. Growth was determined by optical density (600 nm). As homomannan is insoluble in water, growth on this substrate was determined via protein concentration (Pierce BCA Protein Assay Kit, Thermo Fisher Scientific, Waltham, MA, USA).

### *4.3.4. Proteomics*

Cultures of *Muricauda* sp. MAR_2010_75 grown to late exponential phase with glucomannan, galactomannan, homomannan, mannose and citrus pectin (control) as sole carbon source were used for proteome analyses. All substrates were purchased from

Megazyme. Details of protein extraction and subproteome enrichment can be found in the Supplementary Information.

Peptides were separated using reversed phase C18 column chromatography on a nano ACQUITY-UPLC (Waters Corporation, Milford, MA, USA) online-coupled to an LTQ-Orbitrap Classic mass spectrometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) [39]. Spectra were searched against a target-decoy protein sequence database including sequences and reverse sequences of *Muricauda* sp. MAR_2010_75 and of common laboratory contaminants using MaxQuant [40], applying a protein FDR as well as peptide level FDR of 0.01 (1%). Only proteins that could be detected in at least two of three replicates were considered identified. Automatically calculated iBAQ values (intensity-based absolute quantification) were used to manually calculate % riBAQ values (relative iBAQ; giving the relative protein abundance of all proteins in the same sample) for semiquantitative comparisons between samples from different conditions. Tests for differential expression were performed using Perseus v. 1.6.2.3 [41] with Welch's two-sided t-test (permutation-based FDR 0.05). Subcellular protein location was deduced using subproteome data combined with PSORTb 3.0 [42] and CELLO [43] analysis. Data and Results are available through the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [44] with the identifier PXD033586.

### *4.3.5. Diatom and HMWDOM polysaccharide extraction and microarray analysis*

Diatom species (Table S1) were grown in lab cultures as previously described [15], with two 1 L non-axenic monospecific cultures per strain. Diatom cultures were harvested 10 days after inoculation by centrifugation at 6 800 x g for 20 min at 15 ºC. Cell pellets were freeze dried, homogenized using a pestle, and alcohol extraction was performed to obtain the alcohol-insoluble residue (AIR) enriched in polysaccharides. Polysaccharides from the diatom dry biomass (AIR) were sequentially extracted with MilliQ water, 50 mM EDTA and 4 M NaOH with 0.1% w/v $NaBH_4$ as previously described [15].

Each extract was printed in duplicates onto nitrocellulose membrane (pore size of 0.45 µm, Whatman, UK) using a microarrayer (Sprint, Arrayjet, UK) obtaining several identical microarrays populated with all the diatom extracts. Two polysaccharide standards, fucoidan from Laminaria sp. (Glycomix) and glucomannan from Amorphophallus konjac (Megazyme), were dissolved in MilliQ (0.5 mg/mL) and included in the print as controls.

For the HMWDOM samples, sampling was performed in 2016 during a spring microalgae bloom period (over 2.5 months) at the North Sea (54°11.3'N, 7°54.0'E) near the island of

Helgoland, Germany. 100 L of seawater (1 m depth) were filtered through 0.2 µm pore size polycarbonate filters and the filtrate concentrated by tangential flow filtration using 3 filter cassettes with a cut-off of 1 kDa. Specifics on HMWDOM field sampling, processing of the samples, polysaccharide extraction and microarray printing of the HMWDOM extracts is described in (Vidal-Melgosa, 2021). Details of microarray enzymatic treatment for diatom and HMWDOM arrays as well as microarray analysis can be found in Supplementary Information.

### *4.3.6. Cloning, protein expression and purification by chromatography*

The genes for GH26A (WP_036379595.1), GH26B (WP_197062540.1), GH26C (WP_036379585.1), GH5_26 (WP_197062539.1) and GH27 (WP_036379578.1) were amplified from genomic DNA without their signal peptides by PCR using gene-specific primers (Table S2) (Biomers, Ulm, Germany) and cloned using Gibson assembly in Escherichia coli DH5α (New England Biolabs, Ipswich, MA, USA). Clones were in house sequenced by Sanger sequencing using Bigdye (Thermofisher, Waltham, MA, USA). For protein production, plasmid DNA was transformed into E. coli BL21 (DE3) (New England Biolabs, Ipswich, MA, USA) and the proteins were produced in 1 L batches of autoinduction media (ZYP5052) incubated at 20 °C for 4 days [45]. Cells were harvested by centrifugation and stored at -20 °C until processing. Cell lysis was conducted by chemical lysis. Frozen cell pellets were resuspended in 20 mL sucrose solution (25% sucrose, 50 mM TRIS (pH 8.0)). Lysozyme was added at a concentration of 1 mg/mL and the sample was incubated 10 min at room temperature with spinning. 40 mL deoxycholate solution (1% deoxycholate, 1% Triton X-10, 100 mM NaCl) was added followed by MgCl2 to a final concentration of 1 mM and DNase to a concentration of 1 mg/mL. The resulting lysate was centrifuged at 16,000 x g for 45 min at 4 °C. For purification, clarified lysate was applied to a 5 mL prepacked IMAC column (GE Healthcare Life Sciences, Marlborough, MA, USA) previously equilibrated in Buffer A (20 mM TRIS (pH 8) and 500 mM NaCl) using an ÄKTA start FPLC (fast protein liquid chromatography) system (Cytiva, Marlborough, MA, USA). The column was washed extensively with Buffer A and His-tagged protein was eluted using a gradient of imidazole to 500 mM in Buffer B. Purified protein was concentrated using a stirred cell ultrafiltration device with a 10 kDa membrane. For crystallization, the concentrated protein was polished using size exclusion chromatography [HiPrep Sephacryl S200 HR column (Cytiva, Marlborough, MA, USA)] in 20 mM TRIS (pH 8) with 200 mM NaCl. Protein was concentrated to 20 mg/mL prior to further experiments.

#### *4.3.7. Crystallization, X-ray diffraction data collection, structure solution and refinement*

Crystals of GH26C were obtained by hanging drop vapour diffusion with protein mixed 1:1 with a solution of 0.2 M Li2SO4, 0.1 M Bis-Tris:HCl (pH 5.5), 24% w/v PEG3350 and incubating at 16 °C. Crystals of GH26A were obtained in an identical manner but the well solution was composed of 0.15 M MgCl2, 0.1 M Tris:HCl (pH 8.5), 21% w/v PEG4000, 20% v/v glycerol. Crystals were cryo-protected in mother liquor supplemented up to 30% glycerol prior to freezing by being submersed in liquid nitrogen. Diffraction data were collected at DESY P11 (Hamburg, Germany), processed using XDS, and Aimless in CCP4 [46, 47]. Molecular replacement was carried out using Phaser using the coordinates of Cellvibrio japonicus CjGH26C (pdb id: 2VX6) [48] for GH26C and Podospora anserina GH26-CBM35 (pdb: 3ZM8) for GH26A [49]. Model building was carried out automatically in Buccaneer, manually in Coot and refined using Phenix.refine and REFMAC [50–53]. Data were validated and deposited at the Protein Data Bank (PDB). Figures were made using PyMOL v.2.3.2 (Schrödinger, New York, NY, USA) [54].

#### *4.3.8. Enzyme characterization*

Activity profiles of GH26A, GH26C and GH5_26 were generated using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Oligosaccharide and polysaccharide standards for enzymology were acquired from Megazyme. 50 µM purified enzyme was incubated at RT in 50 mM phosphate buffer containing 50 mM NaCl and 1% w/v substrate for 2 h. Samples were inactivated at 90 °C for 10 min and centrifuged at 13,000 rpm for 10 min to remove debris. Supernatant was diluted 1:1,000 in HPLC-grade H2O and products were detected using HPAEC-PAD (Dionex ICS-5000+, ThermoFischer Scientific Inc., Waltham, MA, USA). Controls for each substrate containing no enzyme were treated in similar fashion.

### 4.4. Results

#### *4.4.1. β-mannan PULs are a specialized adaptation in marine habitats*

To examine marine β-mannan degradation we used *Muricauda* sp. MAR_2010_75 as a model organism. This bacterium was isolated from seawater samples taken at the island of Sylt in the German Bight. The 4.4 Mbp genome (GCF_000745185.1) contains a β-mannan PUL related to those in two distant marine relatives that were recently examined using proteomics without further biochemical characterization [22]. While significant PUL

rearrangements compared to *Muricauda* sp. are visible in these and other recently isolated bloom-associated strains (*Salegentibacter* sp. Hel_1_6, *Leeuwenhoekiella* sp. MAR_2009_132, *Sediminibacter* sp. Hel_1_10, *Flavimarina* sp. Hel_I_48) [14], they all contain at least two GH26 β-mannanases and a GH130 β-1,4-mannosylglucose phosphorylase (**Fig. S1**). This presence, together with a putative GH5 as well as a conserved GH27 α-galactosidase and an epimerase suggests the substrate to be a mannose-rich polysaccharide also containing glucose and galactose.

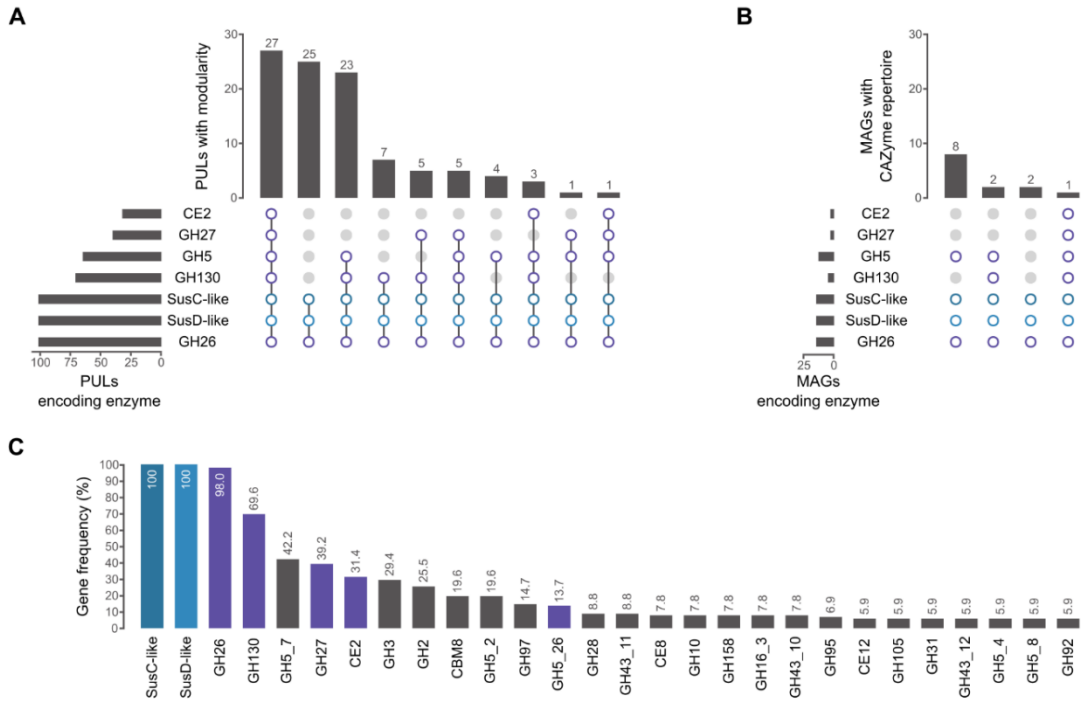



**Figure 4.1. Genes for β-mannan degradation are present in marine habitats.** Clusters/MAGs had to encode for at least a SusC/D pair and a GH26. CAZyme families shared with the *Muricauda* sp. PUL are colored purple, SusC/D pairs blue. **A**) Modularity of different β-mannan PULs found in assemblies from marine databases. **B**) MAGs from Helgoland spring bloom metagenomes containing similar CAZyme sets as the *Muricauda* sp. PUL. **C**) CAZyme families contained in at least 5% of all screened β-mannan degrading clusters, highlighting their variability.

In order to analyze the relevance of β-mannan-containing substrates in marine habitats, we screened the genomes contained in publicly available marine databases (MarRef, MarDb) for clusters containing at least a SusC/D pair and a GH26. This yielded 101 clusters from 82 assemblies, 27 of which feature the same modularity as the *Muricauda* sp. PUL (**Fig.4.1A, Fig. S2**). It also revealed a series of additional CAZymes often associated with GH26-containing

clusters, including GH10, GH95 and GH97, which indicate xylose, galactose and fucose-containing substrates, respectively (**Fig. 4.1C**). The high level of co-occurrence with CAZymes not encoded within the *Muricauda* sp. PUL shows both the specificity of our PUL as well as the high variability β-mannan-containing substrates that marine habitats are likely to display. To account for relevance during bloom situations, we additionally screened Helgoland spring bloom-associated metagenome-assembled genomes (MAGs) from 2010, 2011, 2012 & 2016 [13, 28, 29], yielding 13 MAGs that encoded for at least a SusC/D pair and a GH26. Only 1 MAG encoded for the entire repertoire of the *Muricauda* sp. PUL (**Fig. 4.1B**), where it was also ordered in a cluster. While not among the most common substrates, our results suggest β-mannan-containing polysaccharides to be relevant during microalgal blooms where they are targeted by highly specialized bacteria such as *Muricauda* sp.

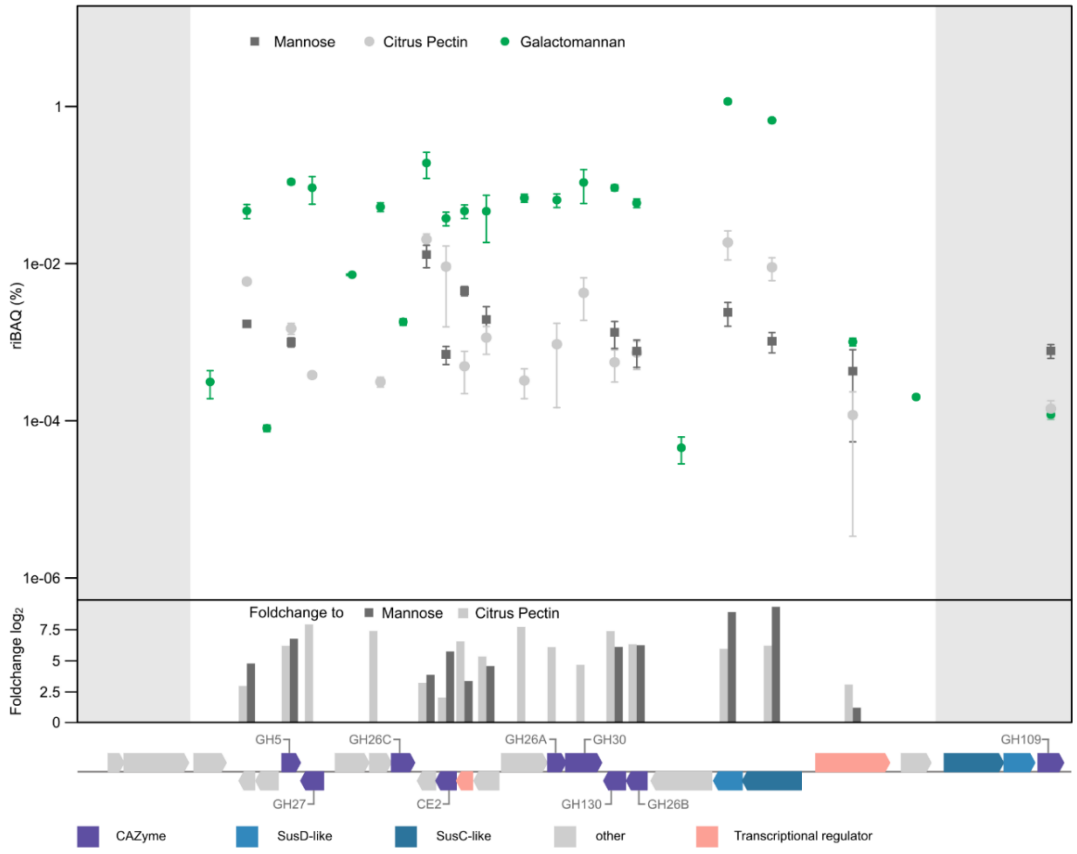



**Figure 4.2. *Muricauda* sp. MAR_2010_75 PUL is specifically induced by galactomannan as sole carbon source.** Abundance of PUL-encoded proteins (mean %riBAQ values and standard deviations, n=3) using galactomannan, citrus pectin and mannose as substrates. Shown are the proteins FG28_RS02265-RS02395 of the β-mannan PUL. Positive foldchanges from galactomannan to other conditions are given as bars. Genes upstream and downstream of the PUL are shown for comparison (highlighted in gray).

### *4.4.2. PUL is specifically upregulated by β-mannans*

*Muricauda* sp. MAR_2010_75 uses mannose-rich substrates as sole carbon source. This ability includes the metabolism of homo-, gluco- or galactomannan, which contain glucose in the backbone and galactose sidechains, respectively (**Fig. S3**). Whole cell proteomics offers experimental evidence for the genomic boundaries of the PUL and thus further indicates the conserved cluster's function for mannan degradation. Abundance of mannan-targeting PUL-encoded proteins was increased with the three tested mannans compared to citrus pectin and the monosaccharide mannose. Except for the HTCS regulator (FG28_RS02375), the 22 proteins of our predicted β-mannan PUL (FG28_RS02275-FG28_RS02380) were significantly upregulated during growth on galactomannan with some being among the most abundant expressed proteins in the whole proteome (**Fig. 4.2, Fig. S4 & Table S3**). They represent proteins directed at mannan digestion, including necessary CAZymes as well as a SusC/D-like pair and two transcriptional regulators. Some proteins, like GH26A (FG28_RS02340), were only detected during growth on mannan. Proteins encoded by genes upstream or downstream of the PUL were either not detected in the proteome or in lower abundance compared to pectin and mannose controls (**Fig. 4.2**). This includes a SusC/D-like pair as well as a predicted α-N-acetylgalactosaminidase of family GH109 located directly downstream of the PUL. Thus, these proteins encoded outside of the PUL likely do not participate in mannan degradation. The proteome data demonstrate that the PUL with its 22 genes is a regulatory unit which specifically responds to and is activated by the presence of β-mannan. As the strain grew faster and to a higher optical density on monosaccharides such as mannose (**Fig. S3**), it can be assumed that a bottleneck exists within the degradation cascade. Our results, which show that β-mannan degradation is performed by highly specialized bacteria (**Fig. 4.1**), make it likely that the ability to target such a complex substrate is of itself an advantage.

### *4.4.3. PUL encodes active mannanases as well as a glucanase and galactosidase*

Consecutive stepwise β-mannan degradation begins with extracellular degradation by endo-active mannanases from GH families 26 and 5 [21]. Similarly, subproteomics combined with Cello and PSORTb analysis (Table S4) predicts that *Muricauda* sp. MAR_2010_75 displays GH26_C and GH5_26 on the extracellular surface. We recombinantly produced the extracellular endo-mannanase (GH26C) and found that it is active on β-1,4-homomannan, galactomannan and glucomannan as well as on different defined manno-oligosaccharides. Galactomannan degradation with GH26C yields a wide range of products due to the galactose branches that interfere with complete digestion (**Fig. 4.3A**). Using HPAEC-PAD we obtained

peaks up to a degree of polymerization (dp) of 11. From defined manno-oligosaccharides mannose and oligosaccharides up to dp4 were released. Of the available galactosylated oligosaccharides — G2M5 and GM3 — only the former is hydrolyzed, releasing mannose and a single peak for the putative product G2M4 while the latter is not affected by GH26C treatment (**Fig 4.3A, Fig. S5A**).

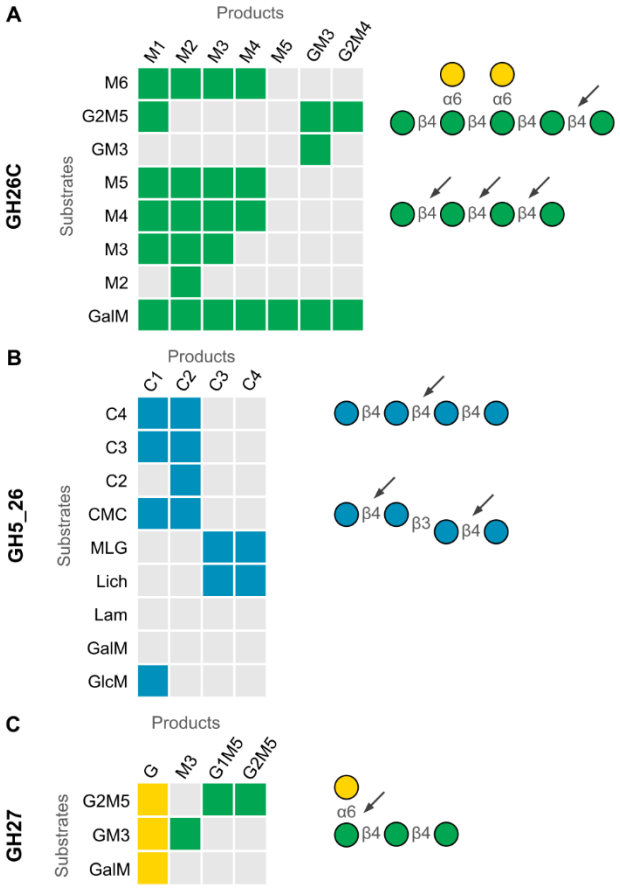



**Figure 4.3. Activities of the β-mannan PUL encoded CAZymes hydrolyze different poly- and oligosaccharides**. Products formed by incubation of specific substrates with recombinantly produced **A**) GH26C, **B**) GH5_26 and **C**) GH27 were detected using HPAEC-PAD. They are shown in color and shape according to the Symbol Nomenclature for Glycans (SNFG) [65]. Examples for each enzyme are depicted on the right. Arrows point to cleavage sites specific for each enzyme.

As evidenced by its activity on carboxymethyl cellulose (CMC) and barley β-glucan (β-1,3/β-1,4) but not laminarin (β-1,3) or galactomannan, the predicted cellulase GH5_26 is a β-1,4-glucanase. This is in accordance with activities described for members of the GH5_26 family

[55]. The enzyme hydrolyzed glucomannan, releasing glucose, consistent with activity on β-1,4-mannogluco-saccharides. On CMC and short cello-oligosaccharides dp3 and dp4, the enzyme produced a mixture of glucose and cellobiose but showed no activity on cellobiose alone. Barley glucan digestion with GH5 resulted in 4 peaks representing oligosaccharides with dp3 and dp4 containing both β-1,3- and β-1,4 linkages (**Fig 4.3B, Fig. S5B**).

Finally, GH27 showed α-1,6-galactosidase activity as it was active on GM3 and G2M5 as well as on galactomannan, releasing galactose branches from the mannan backbone (**Fig. 4.3C, Fig. S5C**). Overall, this biochemical data shows that, although marine and terrestrial β-mannan PULs share enzymatic activities, their differences indicate the marine mannan to be more complex.

### *4.4.4. GH26C and GH26A structures are conserved between distant protein family members*

Characterized glycoside hydrolases from family 26 include both mannanases and glucanases. GH26C shares only 30% sequence identity with its nearest structurally characterized homolog CjGH26C [48], so in order to determine the effects of the sequence dissimilarities, we solved the structure of GH26C to 1.50 Å from residues 47 to 420 for two chains in the asymmetric unit (Table S5). The overall structure is composed of a monomeric catalytic domain in the form of the familiar (α/β)8 (TIM-barrel) fold common to Clan D glycoside hydrolases (**Fig. 4.4A**) [56]. The scaffold is modified with elongated loops at the C-termini of the β-strands at the core of the structure to create a pair of lobes that form the walls of the active site groove. Modelling of the pentasaccharide G1M4 ligand from CjGH26C yielded a picture of an open active site. If anything, the groove is even larger than in CjGH26C as the BA 2 loop is orientated away from the active site groove (**Fig. 4.4A**). The key catalytic residues are conserved providing structural support for the observed biochemical activity (**Fig. 4.4B**). Overall, the marine mannanase shares a high degree of structural identity with its distant homologs from terrestrial ecosystems.

GH26A from *Muricauda* sp. MAR_2010_75 is most likely located in the periplasm (Table S4). This enzyme is 34% identical to the nearest structurally characterized GH26 (*Podospora anserina* GH26) [49] and shares the conserved catalytic residues. Nevertheless, this enzyme shows no similar activity on polysaccharides or oligosaccharide substrates. The X-ray crystal structure revealed a loop blocking the +2 subsite of the active site (**Fig. 4.4C**). These data suggest that GH26A may be a reducing end-specific exo-glycoside hydrolase. However, this does not explain why the recombinant protein is not active on the tested substrates. The

structure may imply a more limited substrate specificity for which we have not yet found the correct oligosaccharide substrate such as shorter manno-gluco oligosaccharides.

A structural comparison of the two new GH26 to the enzymes available in the protein data bank reveals that they are more similar to characterized members than to each other. The closest structural homolog of GH26C is CjGH26C (2VX6) [48] with root mean square deviation (RMSD) of 1.81 Angstroms (Å) over 342 residues aligned. The closest structural homolog of GH26A is a mannanase from *Podospora anserina* (3ZM8) [49] with a C-alpha root mean square deviation of 2.41 Å over 290 residues aligned. On the other hand, M_GH26C and M_GH26A share an RMSD of only 3.12 Å over 301 residues and a sequence identity of 27%.

### *4.4.5. Muricauda enzymes degrade β-mannan and β-glucan from microalgae*

Since the bacterium containing the investigated β-mannan PUL was isolated in the German Bight during a phytoplankton bloom, we hypothesized its targeted substrate originates from microalgae. In order to investigate this, we performed epitope deletion assays. Single carbohydrate microarrays (each containing microalgae related extracts) were treated with either recombinant *Muricauda* enzymes, buffer (control) or commercial enzymes. This assay is based on the fact that an antibody binds to one specific polysaccharide epitope, thus if an enzyme hydrolyzes the polysaccharide then the epitope will be lost and the antibody binding will be reduced or completely abolished. We treated the abovementioned arrays populated with diatom extracts. We observed mannan epitope deletion when the array-immobilized diatom extracts were treated with GH26C from the PUL compared with the buffer-treated control, demonstrating this enzyme targeted the microalgal β-mannan (**Fig. 4.5A**). The monoclonal antibody (mAb) LM21 is an anti-β-1,4-mannan probe and binds to β-1,4-homomannan, glucomannan and galactomannan [57]. The abolishment of LM21 signal when treating the array with GH5_26 and with a β-1,4-glucanase, suggests that *C. affinis* and *C. wailesii* produce glucomannan. Note that the commercial β-glucanase has reported activity against glucomannan and β-glucan, as described by the manufacturer (Megazyme). The LM21 signal decrease in C. affinis extract treated with β-galactanase may be due to certain enzyme side activity, as none of the known mannan types contains β-linked galactose. A standard glucomannan from Konjac was included as control and the two *Muricauda* enzymes catalyzed its degradation. Moreover, GH5_26 digested β-1,3/1,4-glucan from *C. wailesii*, as shown by the loss of signal for mAB BS-400-3 [58] compared to the buffer control (**Fig. 4.5A**). The two polysaccharides were only detected in the NaOH extracts, which is the solvent

commonly used to release hemicelluloses in plants [59, 60], suggesting they might be part of the diatom cell wall.

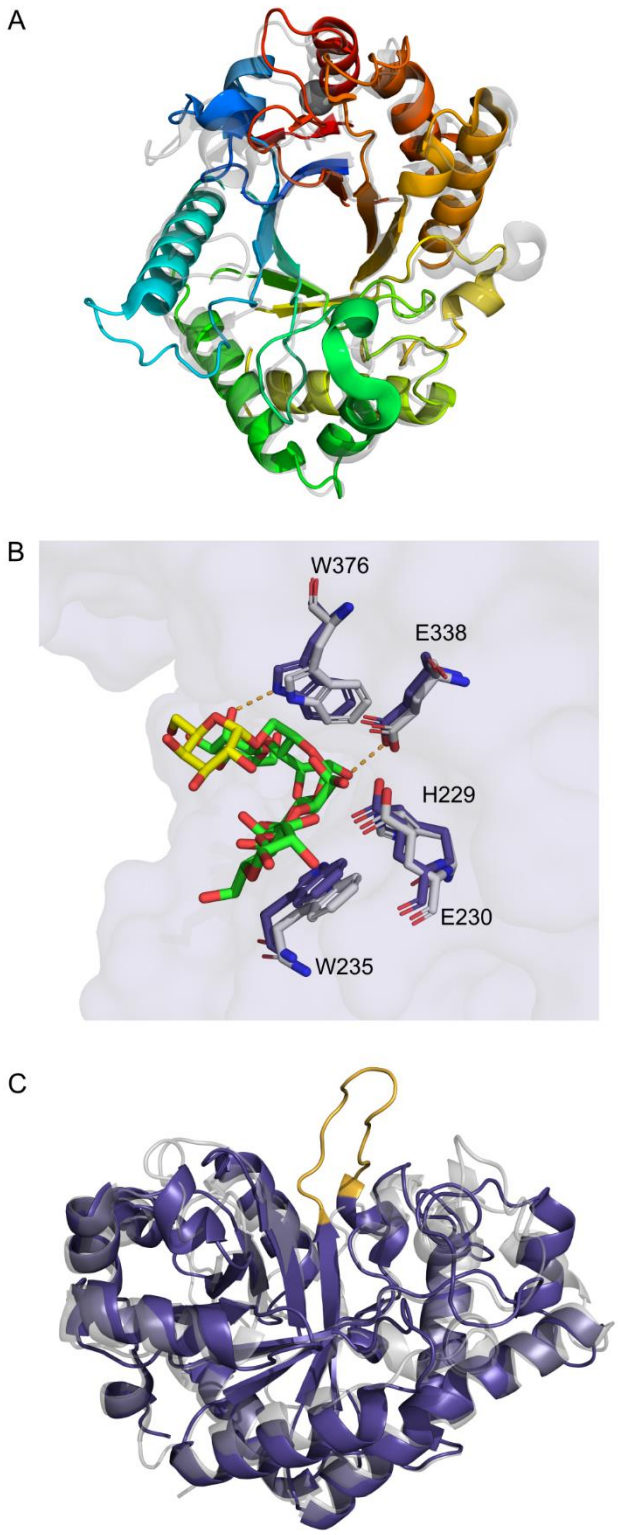



**Figure 4.4. Structural analyses of *Muricauda* sp. MAR_2010_75 GH26A and GH26C allow insights into substrate binding. A**) Structure of GH26C ramped from N- to C-terminus

from blue to red and overlaid with CjGH26C (2VX6, gray). **B**) Active site residues of GH26C (purple) looking into the active site cleft from the side. Enzyme surface is shown in light purple. Residues are superimposed with those observed in CjGH26C (2VX6, gray) and their pentasaccharide substrate. Residues are numbered as observed in the GH26A structure. Putative hydrogen bonds stabilizing the enzyme-substrate complex are shown in orange. C-atoms belonging to mannose moieties of the bound oligosaccharide are colored green, those of galactose yellow. **C**) Overview of GH26A (purple) superimposed with *P. anserina* GH26 (3ZM8, gray) looking into the active site cleft. The additional loop found in GH26A is colored orange.

We treated arrays populated with extracts of HMWDOM sampled during a spring diatom-dominated microalgae bloom at the North Sea. The HMWDOM extracts contained β-1,4-mannan, as shown by binding of mAb LM21 (**Fig. 4.5B**, third column). Presence of β-mannan in the dissolved fraction could be a result of diatom decay. Alternatively, it may be that diatoms contain β-1,4-mannan as part of their cell wall but also actively release some of it into the surrounding seawater. Treatment of the extracts with GH26B showed minor signal decrease compared to the control, thus no clear activity was observed for this enzyme. In contrast, when treating the extracts with GH26A the LM21 signals were abolished. While GH26A showed no activity on commercial β-mannan polysaccharide or oligosaccharide substrates, it showed activity on a β-1,4-mannan present in HMWDOM from a microalgae bloom dominated by diatoms, supporting the hypothesis that the commercially available manno-oligosaccharides were not the correct substrate (**Fig. 4.5B**).

Overall, our data support diatom β-mannan and β-glucan were targeted by widely conserved enzymes belonging to the *Muricauda* β-mannan PUL. Additionally, the epitope deletion shown by our enzymes as well as by commercial GHs further underline the presence of these two polysaccharides in marine diatoms.

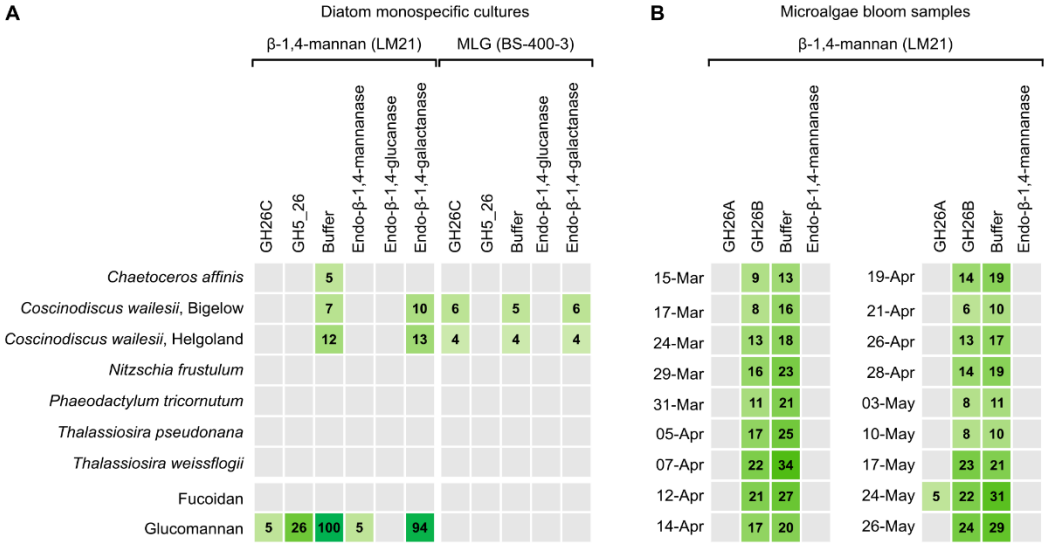



**Figure 4.5. β-1,4-mannan and β-1,3/1,4-glucan substrates targeted by *Muricauda* sp. enzymes can be found in different diatom isolates.** Two types of microarrays were used for epitope deletion experiments. A) Polysaccharides from different diatom species were extracted and those extracts were printed onto microarrays. B) High molecular weight dissolved organic matter (HMWDOM) was sampled during a diatom-dominated microalgae bloom in 2016 at the North Sea, dates are specified at the left of the sub-heatmaps. Polysaccharides from the HMWDOM were extracted and printed onto microarrays. A,B) Identical arrays were treated with either recombinant *Muricauda* sp. MAR_2010_75 GH26C, GH5_26, GH26A, GH26B, buffer (control) or commercial enzymes (Megazyme). After treatment, arrays were incubated with the antibody LM21 specific to β-1,4-mannan or with the antibody BS-400-3 specific to β-1,3/1,4-glucan (mixed-linkage glucan, MLG). For both array types, signals obtained in the NaOH-extracts are shown. Values in the heatmaps correspond to the mean signal intensities of a minimum of two replicates (each extract represented by at least two spots in each array). The highest mean signal in each of the datasets was set to 100 and all other values were normalized accordingly.

## 4.5. Discussion

Diatom cell wall polysaccharides represent an important yet underexplored carbon pool. Besides the specific structures of these microalgal sugar compounds, the biochemical mechanisms bacteria use to break down these polymers remain poorly understood. Based on our results, the process of bacteria-driven marine β-mannan degradation is evolutionarily

and functionally similar to terrestrial gut bacteria with regards to both, enzymatic reactions and protein structures. Genomic content of experimentally determined PULs and predicted biochemical activities of our PUL-encoded enzymes align with those of gut microbiome pathways [21]. On the other hand, there also are specific enzyme adaptations to more complex marine β-mannan sources.

Based on the proteomic and biochemical data of this study, we propose an enzymatic cascade for the degradation of β-mannan-containing substrates by *Muricauda* sp. MAR_2010_75 (**Fig. 4.6**). Our subproteome analysis revealed the PUL-encoded enzymes GH26C and GH5_26 to be most abundant in the extracellular protein fraction, indicating a position attached to but not embedded in the outer membrane. GH26C, the main conserved mannanase of the PUL, was experimentally shown to be an endo-acting β-1,4 mannanase, producing oligosaccharides of varying lengths when incubated with galactomannan. The solved structure shows the classic hydrolase TIM-barrel conformation and shares active residues with structural and functional homologs of terrestrial ß-mannan-degrading bacteria [21, 48]. Homologs of this enzyme are found in similarly structured PULs of bacteroidetes found in marine databases as well as bloom isolates, underlining the importance of β-mannan-containing substrates (**Fig. 4.1, Fig. S1**). GH5_26 showed β-1,4 glucanase activity on substrates containing different β-linkages as well as activity on glucomannan, indicating that the polysaccharide targeted by this enzyme contains at least some amounts of glucose as described for terrestrial glucomannans [17]. A limitation of our study is the use of terrestrial mannans as substrates because beta-mannans from marine micro- or macroalgae are not yet available. It should be considered that the terrestrial mannans and the oligos used here for bacterial growth and enzymology are not the optimal substrates. Slow growth and incomplete digestion support this argument suggesting that the PUL with its enzymes could be targeting algal mannans with similar but also different structures compared to terrestrial mannans.

In the periplasm, further depolymerization is facilitated by an exo-α-1,6 galactosidase of family GH27 able to release galactose from β-1,4-mannan oligosaccharide backbones of different length. GH26A, which was also predicted to be present in the periplasm, showed no activity on tested manno-oligosaccharides but revealed activity on β-1,4-mannan produced during a microalgae bloom. Structural analysis revealed an additional loop blocking the +2 site of the active site that cannot be observed in homologs with high structural similarities [49, 61]. This suggests a narrower substrate specificity for which we could not determine a suitable oligosaccharide. From the other deduced enzyme activities encoded by

the PUL it can, however, be speculated that the substrate is an oligosaccharide containing both mannose and glucose. Activities like this have previously been demonstrated for family 26 glycoside hydrolases [62] but additional data is needed to test this hypothesis.

Whereas proteome analyses indicated the SusC (FG28_RS02370) to channel oligosaccharides into the cell, a PUL-enocoded symporter (FG28_RS02300) transports manno- and manno-gluco-disaccharides into the cytosol. Here, they are targeted by a predicted GH130 phosphorylase, allowing the released monosaccharides to enter glycolysis [63].

The presence of an expressed transcription factor of the AraC-type (FG28_RS02325) together with the regulatory up- and downstream boundaries of the PUL show this cluster to be a genomic island for the degradation of β-mannan. The upregulation of the PUL, observed for three tested mannans (galactomannan, glucomannan and homomannan), was highest on galactomannan and glucomannan (Table S3). The abundance of a PUL-encoded predicted esterase in the membrane fraction suggests the substrate to be modified by acetylation, which is common for β-mannans [19, 20].

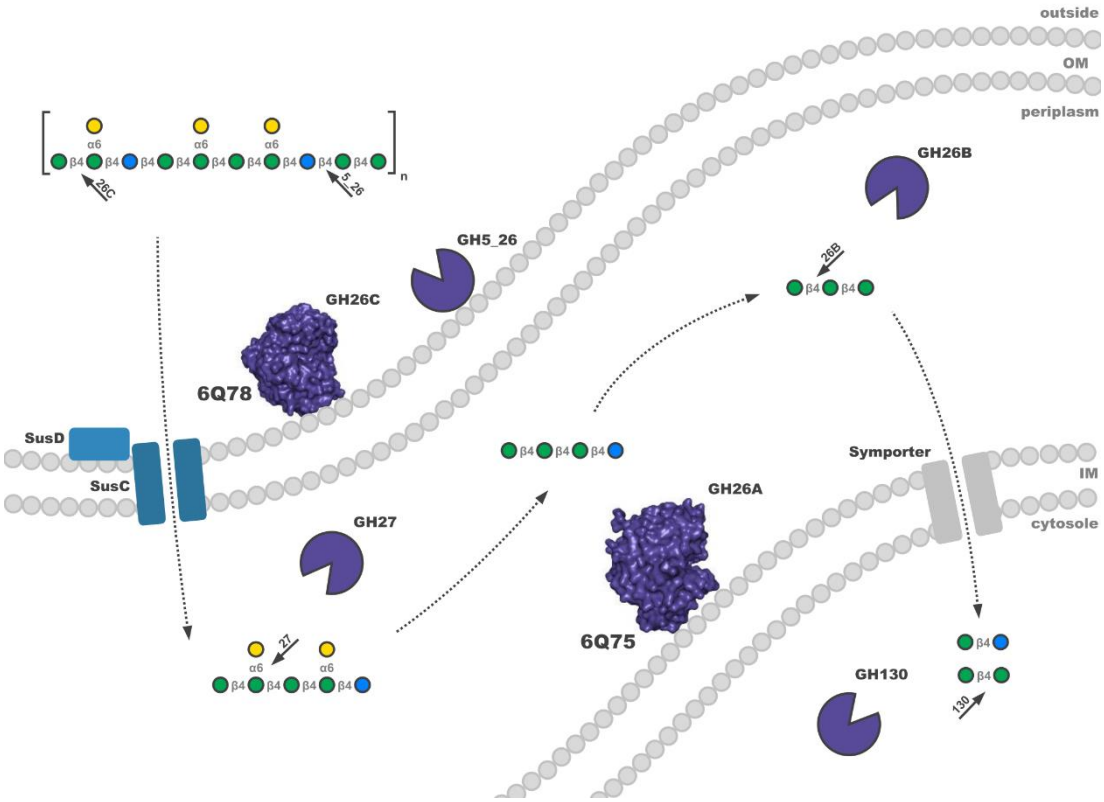



**Figure 4.6. Schematic of the proposed enzymatic cascade for the degradation of galactoglucomannan substrates by *Muricauda* sp. MAR_2010_75.** Key proteins and enzymes of the pathway coded for in the β-mannan PUL (FG28_RS02275-FG28_RS02380) are colored according to Fig. 4.3. The polysaccharide substrate sketch is shown according to Fig. 4.4. Regulatory elements and enzymes directed at monosaccharide processing and

deacetylation were omitted for clarity. Localization of proteins was predicted via subproteome analyses by the pSORTb 3.0 and CELLO tools [42, 43]. GH5_26, GH26A, GH26C and GH27 were assigned functions based on biochemical and/or structural evidence. Functions of other proteins and CAZymes were assigned via a combined BLAST and HMMER search [66, 67]. IM: inner membrane, OM: outer membrane

β-mannans, especially galactoglucomannans, have been widely characterized as a main group of hemicelluloses in plants [19, 64]. Terrestrial mannans such as galactomannan are capable of activating not only the β-mannan PUL of *Muricauda* sp. MAR_2010_75 but, as we previously showed, also similar PULs of marine bloom isolates such as those of *Salegentibacter* sp. Hel_1_6 and *Leeuwenhoekiella* sp. MAR_2009_132 [22]. With the comparative genomics of this study, we show that our proposed degradation pathway is conserved in many marine bacteria as well as bloom-associated isolates. Observed CAZyme modularities that differ from the *Muricauda* sp. PUL also suggest a β-mannan-containing substrates to show a degree of variability (Fig 1, Fig S1).

Only recently, we detected a β-mannan of microalgal origin in HMWDOM and POM [15]. In the present study, we now for the first time link these β-mannan-targeting PULs of marine bacteria to a specific microalgae source. By analyzing carbohydrates extracted from different diatom species relevant to microalgae blooms, we showed the existence of β-1,4-mannan in *C. affinis* and in two different isolates of *C. wailesii*. This diatom-based β-mannan was degraded by recombinant versions of the GH26C and GH5_26 of *Muricauda* sp. MAR_2010_75. Furthermore, the HMWDOM β-mannan from a natural diatom bloom was degraded by a recombinant version of the GH26A. These enzymatic activities demonstrate that the enzymes conserved in PULs of different marine isolates target microalgal β-mannan.

In conclusion, we show that the β-mannan PUL of *Muricauda* sp. MAR_2010_75 can be viewed as a model for how marine bacteria target β-mannan-containing substrates. Parts of this general degradation mechanism are similar to those of gut bacteria but also bear significant differences, such as the addition of the GH5_26 (**Fig. 4.6**). The detected β-mannan links *C. affinis* and *C. wailesii* to marine bacteria containing β-mannan PULs such as *Muricauda* sp. MAR_2010_75. The specificity of this glycan-enzyme link at species resolution uncovers hypothetical connections between primary producers and secondary consumers of organic matter in the carbon cycle.

## 4.6. Acknowledgements

We are grateful to Sebastian Grund for mass spectrometry analyses and Tina Trautmann and Alek Bolte for HPAEC-PAD analyses. We thank the Biological Institute Helgoland of the Alfred-Wegener Institute for logistic support during our sampling at the Kabeltonne Helgoland. We also thank Rudolf Amann and the Max-Planck-Institute for Marine Microbiology for financially supporting research visits for this project. This work was financially supported by the German Research Foundation (DFG) for the research unit FOR2406 “Proteogenomics of Marine Polysaccharide Utilization” by grants of D. Becher (BE 3869/4-2), J.-H. Hehemann (HE 7217/2-2), and T. Schweder (SCHW 595/10-2).

## 4.7. References

1. Laine RA. A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 x 10(12) structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. Glycobiology. 1994; 4:759–767.
2. Becker S, Tebben J, Coffinet S, Wiltshire K, Iversen MH, Harder T et al. Laminarin is a major molecule in the marine carbon cycle. Proceedings of the National Academy of Sciences. 2020; 117:6599.
3. Pauly M, Gille S, Liu L, Mansoori N, Souza A de, Schultink A et al. Hemicellulose biosynthesis. Planta. 2013; 238:627–642.
4. Domozych D. Algal Cell Walls. In: Lauc G, Wuhrer M. High-throughput glycomics and glycoproteomics. Humana Press, New York, 2016. pp 1–11.
5. Donlan RM. Biofilms: microbial life on surfaces. Emerg Infect Dis. 2002; 8:881–890.
6. Popper ZA, Michel G, Hervé C, Domozych DS, Willats WGT, Tuohy MG et al. Evolution and diversity of plant cell walls: from algae to flowering plants. Annual Review of Plant Biology. 2011; 62:567–590.
7. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. 1991; 280 (Pt 2):309–316.
8. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J Biol Chem. 2009; 284:24673–24677.

9. Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron E, Pudlo NA et al. Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. Nature. 2015; 517:165–169.
10. Falkowski PG, Barber RT, Smetacek VV. Biogeochemical Controls and Feedbacks on Ocean Primary Production. Science. 1998; 281:200–207.
11. Smetacek V. Seeing is Believing: Diatoms and the Ocean Carbon Cycle Revisited. Protist. 2018; 169:791–802.
12. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012; 336:608–611.
13. Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. Elife. 2016; 5:e11888.
14. Kappelmann L, Krüger K, Hehemann J-H, Harder J, Markert S, Unfried F 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.
15. Vidal-Melgosa S, Sichert A, Francis TB, Bartosik D, Niggemann J, Wichels A et al. Diatom fucan polysaccharide precipitates carbon during algal blooms. Nat Commun. 2021; 12:1150.
16. Chanzy H, Dube M, Marchessault RH, Revol JF. Single crystals and oriented crystallization of ivory nut mannan. Biopolymers. 1979; 18:887–898.
17. Katsuraya K, Okuyama K, Hatanaka K, Oshima R, Sato T, Matsuzaki K. Constitution of konjac glucomannan: chemical analysis and 13C NMR spectroscopy. Carbohydr Polym. 2003; 53:183–189.
18. L. Melton, Bronwen G. Smith, R. Ibrahim, R. Schröder, P. Harris, U. Schmitt. Mannans in primary and secondary plant cell walls. New Zealand journal of forestry science. 2009; 39:153–160.
19. Hannuksela T, Du Hervé Penhoat C. NMR structural determination of dissolved O-acetylated galactoglucomannan isolated from spruce thermomechanical pulp. Carbohydr Res. 2004; 339:301–312.
20. Gilbert HJ, Stålbrand H, Brumer H. How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. Curr Opin Plant Biol. 2008; 11:338–348.

21. Bågenholm V, Reddy SK, Bouraoui H, Morrill J, Kulcinskaja E, Bahr CM et al. Galactomannan Catabolism Conferred by a Polysaccharide Utilization Locus of Bacteroides ovatus: ENZYME SYNERGY AND CRYSTAL STRUCTURE OF A β-MANNANASE. J Biol Chem. 2017; 292:229–243.
22. Chen J, Robb CS, Unfried F, Kappelmann L, Markert S, Song T et al. Alpha- and beta-mannan utilization by marine Bacteroidetes. Environ Microbiol. 2018; 20:4127–4140.
23. Klemetsen T, Raknes IA, Fu J, Agafonov A, Balasundaram SV, Tartari G et al. The MAR databases: development and implementation of databases specific for marine metagenomics. Nucleic Acids Res. 2018; 46:D692-D699.
24. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022; 50:D20-D26.
25. Gilchrist CL, Booth TJ, van Wersch B, van Grieken L, Medema MH, Chooi Y-H. cblaster: a remote search tool for rapid identification and visualisation of homologous gene clusters. bioRxiv. 2021
26. Zhang H, Yohe T, Le Huang, Entwistle S, Wu P, Yang Z et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018; 46:W95-W101.
27. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402.
28. Krüger K, Chafee M, Ben Francis T, Del Glavina Rio T, Becher D, Schweder T et al. In marine Bacteroidetes the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. ISME J. 2019; 13:2800–2816.
29. Francis TB, Bartosik D, Sura T, Sichert A, Hehemann J-H, Markert S et al. Changing expression patterns of TonB-dependent transporters suggest shifts in polysaccharide consumption over the course of a spring phytoplankton bloom. ISME J. 2021; 15:2336–2350.
30. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph. 2014; 20:1983–1992.
31. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017; 33:2938–2940.
32. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19:1639–1645.

33. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022:gkac240.
34. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol. 2010; 59:307–321.
35. Lefort V, Longueville J-E, Gascuel O. SMS: Smart Model Selection in PhyML. Mol Biol Evol. 2017; 34:2422–2424.
36. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019; 47:W256-W259.
37. Hahnke RL, Harder J. Phylogenetic diversity of Flavobacteria isolated from the North Sea on solid media. Syst Appl Microbiol. 2013; 36:497–504.
38. Schut F, Vries EJ de, Gottschal JC, Robertson BR, Harder W, Prins RA et al. Isolation of Typical Marine Bacteria by Dilution Culture: Growth, Maintenance, and Characteristics of Isolates under Laboratory Conditions. Appl Environ Microbiol. 1993; 59:2150–2160.
39. Otto A, Bernhardt J, Meyer H, Schaffer M, Herbst F-A, Siebourg J et al. Systems-wide temporal proteomic profiling in glucose-starved Bacillus subtilis. Nat Commun. 2010; 1:137.
40. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology. 2008; 26:1367–1372.
41. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nature Methods. 2016; 13:731–740.
42. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010; 26:1608–1615.
43. Yu C-S, Lin C-J, Hwang J-K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci. 2004; 13:1402–1406.
44. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022; 50:D543-D552.

45. Studier F. Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. 2005; 207

46. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr. 1994; 50:760–763.

47. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–132.

48. Cartmell A, Topakas E, Ducros VM-A, Suits MDL, Davies GJ, Gilbert HJ. The Cellvibrio japonicus mannanase CjMan26C displays a unique exo-mode of action that is conferred by subtle changes to the distal region of the active site. J Biol Chem. 2008; 283:34403–34413.

49. Couturier M, Roussel A, Rosengren A, Leone P, Stålbrand H, Berrin J-G. Structural and Biochemical Analyses of Glycoside Hydrolase Families 5 and 26 β-(1,4)-Mannanases from Podospora anserina Reveal Differences upon Manno-oligosaccharide Catalysis*. Journal of Biological Chemistry. 2013; 288:14624–14635.

50. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr. 2012; 68:352–367.

51. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011; 67:355–367.

52. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of \it Coot. Acta Crystallographica Section D. 2010; 66:486–501.

53. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr. 2006; 62:1002–1011.

54. DeLano WL. The PyMOL Molecular Graphics System Version 2.3.4. Schrödinger, LLC, New York; 2010

55. Fontes CM, Clarke JH, Hazlewood GP, Fernandes TH, Gilbert HJ, Ferreira LM. Possible roles for a non-modular, thermostable and proteinase-resistant cellulase from the mesophilic aerobic soil bacterium Cellvibrio mixtus. Appl Microbiol Biotechnol. 1997; 48:473–479.

56. Brändén C-I. The TIM barrel—the most frequently occurring folding motif in proteins: Current Opinion in Structural Biology 1991, 1:978–983. Current Opinion in Structural Biology. 1991; 1:978–983.

57. Marcus SE, Blake AW, Benians TAS, Lee KJD, Poyser C, Donaldson L et al. Restricted access of proteins to mannan polysaccharides in intact plant cell walls. Plant J. 2010; 64:191–203.
58. Meikle PJ, Hoogenraad NJ, Bonig I, Clarke AE, Stone BA. A (1--3,1--4)-beta-glucan-specific monoclonal antibody and its use in the quantitation and immunocytochemical location of (1--3,1--4)-beta-glucans. Plant J. 1994; 5:1–9.
59. Lauc G, Wuhrer M. High-throughput glycomics and glycoproteomics. In: ; 1503
60. Moller I, Sørensen I, Bernal AJ, Blaukopf C, Lee K, Øbro J et al. High-throughput mapping of cell-wall polymers within and between plants using novel microarrays. Plant J. 2007; 50:1118–1128.
61. Yan X-X, An X-M, Gui L-L, Liang D-C. From structure to function: insights into the catalytic substrate specificity and thermostability displayed by Bacillus subtilis mannanase BCman. J Mol Biol. 2008; 379:535–544.
62. Tailford LE, Ducros VM-A, Flint JE, Roberts SM, Morland C, Zechel DL et al. Understanding how diverse beta-mannanases recognize heterogeneous substrates. Biochemistry. 2009; 48:7009–7018.
63. Nakae S, Ito S, Higa M, Senoura T, Wasaki J, Hijikata A et al. Structure of Novel Enzyme in Mannan Biodegradation Process 4-O-β-d-Mannosyl-d-Glucose Phosphorylase MGP. J Mol Biol. 2013; 425:4468–4478.
64. Scheller HV, Ulvskov P. Hemicelluloses. Annual Review of Plant Biology. 2010; 61:263–289.
65. Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD et al. Symbol Nomenclature for Graphical Representations of Glycans. Glycobiology. 2015; 25:1323–1324.
66. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215:403–410.
67. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011; 39:W29-W37.

## 4.8. Competing interests

The authors declare no competing interests.

## 4.9. Supplementary information

### *4.9.1. Supplementary Methods*

#### *4.9.1.1. Protein extraction and subproteome enrichment*

Triplicates of 100 mL cultures were harvested via centrifugation (4 000 x g, 20 min). Cells were washed twice with TE buffer (10 mM Tris-HCl (pH 7.5), 10 mM EDTA (pH 8.0)) and resuspended in 50 mM Tris-HCl (pH 7.5) containing Roche 'cOmplete Protease Inhibitor (Lysis Buffer) (Roche, Basel, Switzerland). Cells were disrupted via sonication (3 x 30 s) (HD/UV 2070, Bandelin, Berlin, Germany) and cell debris was removed by centrifugation. Integral membrane and cytosolic proteins for subproteomes from glucomannan cultures were separated by ultracentrifugation at 100 000 x g and 4 °C for 65 min. The pelleted membrane was resuspended in TE buffer and thoroughly ground using a glass homogenizer.

For the extraction of the extracellular fraction, the culture supernatant was filtered using a 0.2 µm filter to remove remaining cells and then incubated with StrataClean beads at 4 °C and 150 rpm over night to bind cell-detached protein [1]. The mixture was centrifuged at 10 000 x g and 4 °C and washed once with TE buffer. Remaining buffer was evaporated using a SpeedVac (Concentrator Plus, Eppendorf, Hamburg, Germany).

Protein concentration of all samples except the extracellular fraction was determined using the Pierce™ BCA Protein Assay Kit (ThermoFischer Scientific, Waltham, MA, USA). 25 µg of protein were loaded on a 10% 1D-SDS polyacrylamide gel and separated for 90 min at 120 V. After fixing with 40% ethanol/10% acetic acid followed by Coomassie G-250 staining [2], the proteins were in-gel digested overnight (16 h) using trypsin [3].

#### *4.9.1.2. Microarray analysis*

The printed arrays were first blocked for 1 h with phosphate-buffered saline (PBS) containing 5% (w/v) low fat milk powder (MPBS). The MPBS was washed three times with PBS. Then single arrays were individually incubated with either: recombinant GH26C or GH5_26 at 5 µg/mL, or 100 mM sodium phosphate buffer pH 7.0, or β-mannanase (E-BMACJ), cellulase (E-CELTR) or β-galactanase (E-GALCJ) from Megazyme at 1 U/mL, at 37 °C and 100 rpm overnight. After the treatment, arrays were extensively washed with PBS followed by 2 h incubation with the primary monoclonal antibodies (mAbs) LM21 (PlantProbes, UK) and BS-400-3 (BioSupplies, Australia) diluted 1/10 and 1/1000 in MPBS, correspondingly. After a washing step with PBS, arrays were incubated for 2 h with the anti-rat or anti-mouse

secondary antibodies conjugated to alkaline phosphatase (A8438 and A3562, Sigma-Aldrich) both diluted 1/5000 in MPBS. After washing with PBS and MilliQ, arrays were developed with alkaline phosphatase substrate. The developed arrays were analyzed as described previously [4]. The highest mean signal intensity value obtained in the whole data set was set to 100 and all other values were normalized accordingly. A cut-off of 4 arbitrary units was applied. Controls for the extraction solvents and for the secondary antibodies showed no unspecific binding of the probes.

### *4.9.2. Supplementary references*

1. Bonn F, Otto A. Protein Enrichment from Highly Dilute Samples with StrataClean. Methods Mol Biol. 2018; 1841:11–18.
2. Candiano G, Bruschi M, Musante L, Santucci L, Ghiggeri GM, Carnemolla B et al. Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. Electrophoresis. 2004; 25:1327–1333.
3. Grube M, Cernava T, Soh J, Fuchs S, Aschenbrenner I, Lassek C et al. Exploring functional contexts of symbiotic sustain within lichen-associated bacteria by comparative omics. ISME J. 2015; 9:412–424.
4. Vidal-Melgosa S, Sichert A, Francis TB, Bartosik D, Niggemann J, Wichels A et al. Diatom fucan polysaccharide precipitates carbon during algal blooms. Nat Commun. 2021; 12:1150.
5. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nature Methods. 2016; 13:731–740.
6. Yu C-S, Lin C-J, Hwang J-K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci. 2004; 13:1402–1406.
7. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010; 26:1608–1615.

### *4.9.3. Supplementary tables*

Large supplementary tables can be viewed online at https://doi.org/10.1038/s41396-022-01342-4.

**Table S4.1.** Information on diatom species used for the search of marine β-mannan sources

| **Diatom species** | **Strain information** |
|---|---|
| *Chaetoceros affinis* | CCMP158 (NCMA, Bigelow Laboratory) |
| *Coscinodiscus wailesii* (strain 1) | CCMP2513 (NCMA, Bigelow Laboratory) |
| *Coscinodiscus wailesii* (strain 2) | Isolated at Helgoland (54˚11.3'N,7˚54.0'E), North Sea |
| *Nitzschia frustulum* | CCMP558 (NCMA, Bigelow Laboratory) |
| *Phaeodactylum tricornutum* | CCMP2561 (NCMA, Bigelow Laboratory) |
| *Thalassiosira pseudonana* | provided by André Scheffel (Potsdam, Germany) |
| *Thalassiosira weissflogii* | provided by André Scheffel (Potsdam, Germany) |

**Table S4.2** Primer sequences

| **Name** | **Protein ID** | **Sequence** |
|---|---|---|
| GH5_fw | WP_197062539.1 | GTTTAACTTTAAGAAGGAGATATACCATGGATGGGGCGTATATCAGTCGATG |
| GH5_rv | WP_197062539.1 | ATCTCAGTGGTGGTGGTGGTGGTGCTCGAGCTTATTGCCCTTGTCATAGCTCTG |
| GH27_fw | WP_036379578.1 | ATCTCAGTGGTGGTGGTGGTGGTGCTCGAGAAATCTATTCTTTTTCTCAAGCAACACCATT |
| GH27_rv | WP_036379578.1 | GTTTAACTTTAAGAAGGAGATATACCATGGTGTGGTCAAAAATTTAATGGGCTTGCC |
| GH26A_fw | WP_036379595.1 | GTTTAACTTTAAGAAGGAGATATACCATGGTCAATGGGAATGCCACGGAAG |
| GH26B-fw | WP_197062540.1 | GTTTAACTTTAAGAAGGAGATATACCATGGGCAATACCTCCCTCACCGACA |
| GH26B_rev | WP_197062540.1 | ATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTTTAATTTTTTTGATATCCTTTAAAAACA |
| GH26A_rv | WP_036379595.1 | ATCTCAGTGGTGGTGGTGGTGGTGCTCGAGGCGCTCGTTGGTAAACTGTATCTCATC |
| GH26C_fw | WP_036379585.1 | GTTTAACTTTAAGAAGGAGATATACCATGGGAAGTTTCATGGTAAACCCGGATG |
| GH26C_rv | WP_036379585.1 | ATCTCAGTGGTGGTGGTGGTGGTGAGAATTGGGCGGTAGCCGGTAC |

**Table S4.3.** Whole cell proteomics results. *Muricauda* sp. was grown on homomannan, galactomannan, glucomannan, mannose and citrus pectin as sole carbon source. Only

proteins that could be detected in at least two of three replicates were considered identified. Automatically calculated iBAQ values were used to manually calculate % riBAQ for semiquantitative comparisons between samples from different conditions. Tests for differential expression were performed using Perseus v. 1.6.2.3 [5] with Welch's two-sided t-test (permutation-based FDR 0.05). (see separate Excel file)

**Table S4.4.** Subproteomics results with location prediction. Tests for differential expression were performed using Perseus v. 1.6.2.3 [5] with Welch's two-sided t-test (permutation-based FDR 0.05). Localization of the proteins was additionally analyzed via the pSORTb 3.0 and CELLO tools [6, 7]. (see separate Excel file)

**Table S4.5**. X-ray data collection, processing and model refinement statistics for M_GH26C and M_GH26A

| Dataset | GH26A | GH26C |
|---|---|---|
| Resolution (Å) | 45.12-1.75 (1.84-1.75) | 49.40-1.50 (1.52-1.50) |
| Space group | P1 | C121 |
| Unit Cell (Å) (˚) | 47.524, 49.342, 71.711 (77.999, 89.831, 69.641) | 94.119, 60.358, 148.99 (90, 95.925, 90) |
| No. of Reflections | 248361 (36556) | 884001 (39440) |
| No. Unique | 58106 (8362) | 132761 (6137) |
| $R_{sym}$ (%) | 0.089 (0.482) | 0.117 (0.945) |
| Completeness (%) | 9.2 (2.7) | 9.5 (2.2) |
| Redundancy | 96.7 (95.2) | 99.5 (93.6) |
| <*I*/□(*I*)> | 4.3 (4.4) | 6.7 (6.4) |
| Mosaicity | 0.13 | 0.06 |
| *Refinement* | | |
| $R_{work}/R_{free}$ (%) | 15/19 | 16/19 |
| Number of Atoms | 5418 | 6747 |
| Protein | 4970 | 6189 |
| Ligand (TRS; GOL) | 20 | 18.3 |
| Water | 428 | 842 |
| *B* factors | | |
| Overall | 26.09 | 18.3 |
| Protein | 25.12 | 16.49 |
| Ligand (TRS; GOL) | 32.04 | 18.68 |
| Water | 37.01 | 31.23 |
| RMSZ | | |

| | | |
|---|---|---|
| Bond Lengths | 0.009 | 0.011 |
| Bond Angles | 0.95 | 1.1 |
| Ramachandran Statistics (%) | | |
| Favored | 96.4 | 97.33 |
| Allowed | 3.56 | 2.4 |
| Outliers | 0 | 0.27 |
| | | |
| PDB accession code | 6Q75 | 6Q78 |

*4.9.4. Supplementary figures*

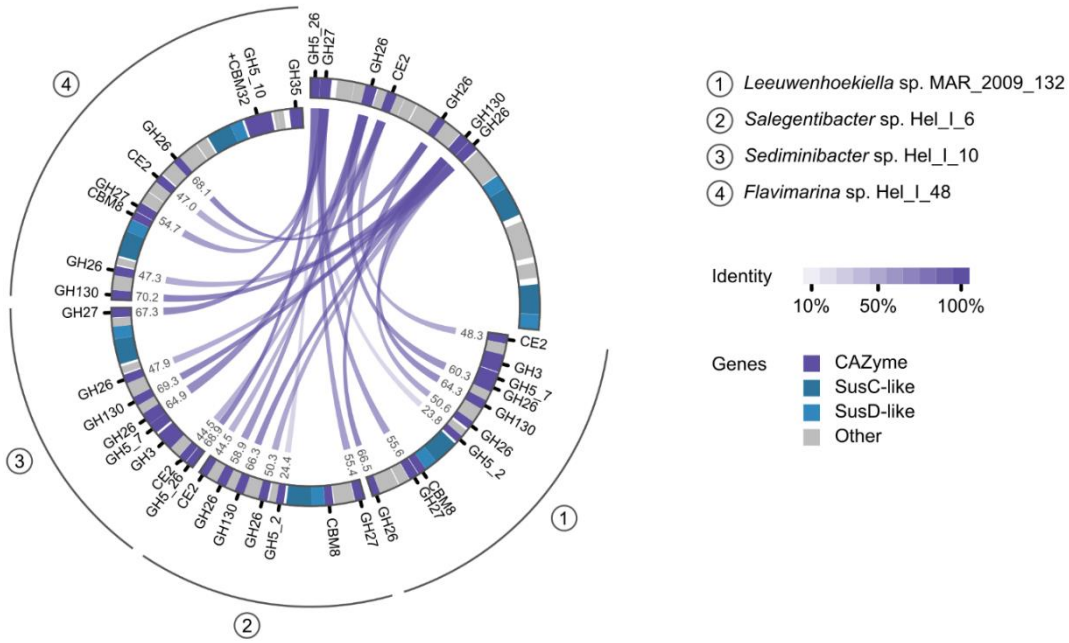
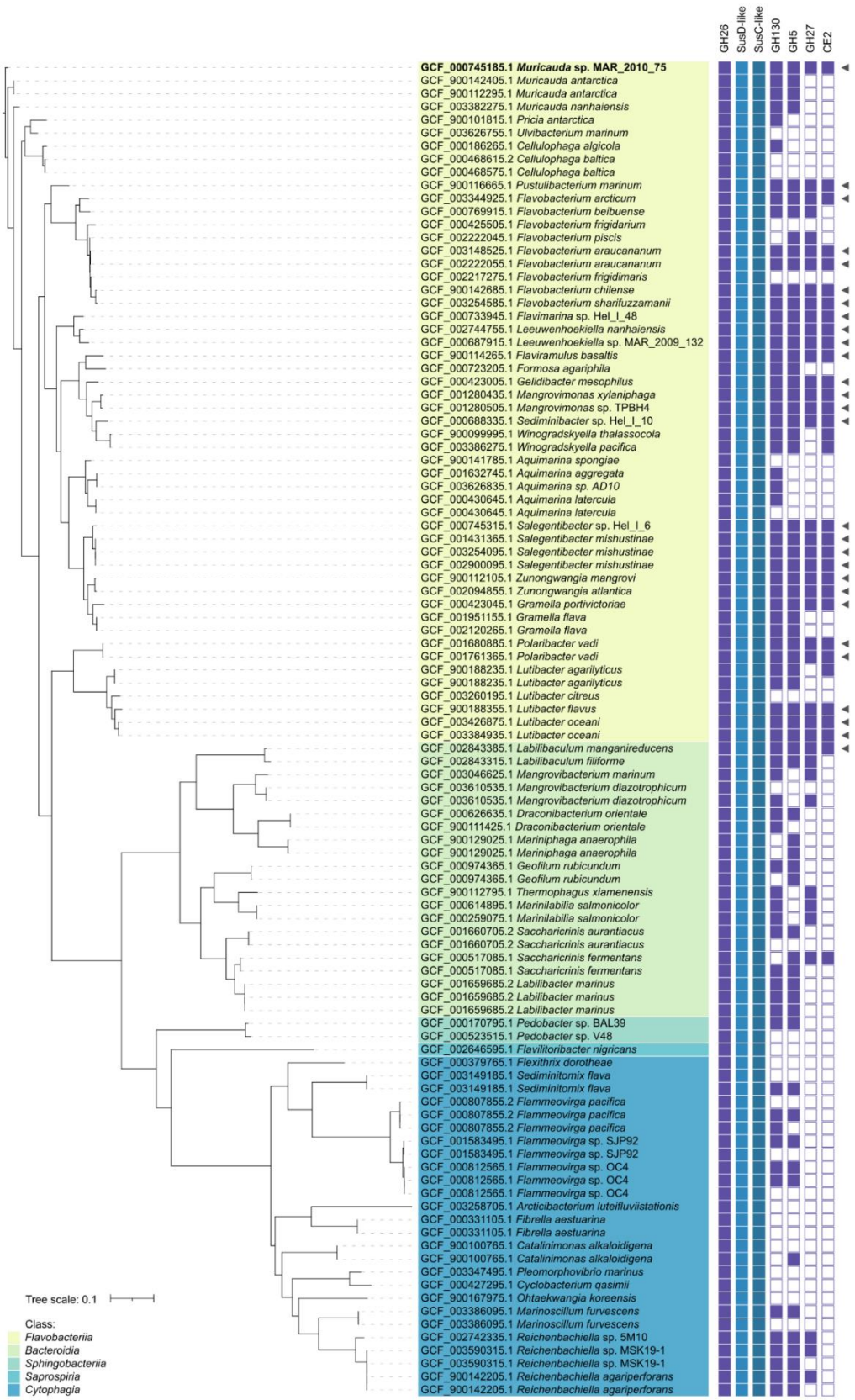



**Figure S4.1.** CAZyme synteny of the *Muricauda* sp. β-mannan PUL with those of other bloom-associated strains. Only the highest scoring hit for each CAZyme in a PUL is shown. Ribbon color intensity indicates degree of similarity; cutoff 10-5. Identities (%) are given for each link.

*Next page*

**Figure S4.2.** Maximum likelihood phylogeny of all clusters similar to the *Muricauda* sp. PUL found in marine databases. Class of the bacteria encoding for the found clusters is given in color. The modularity of each cluster as it compares to the *Muricauda* sp. PUL is shown on the right.


GH26
SusD-like
SusC-like
GH130
GH5
GH27
CE2
GCF_000745185.1 Muricauda sp. MAR_2010_75
GCF_900142405.1 Muricauda antarctica
GCF_900112295.1 Muricauda antarctica
GCF_003382275.1 Muricauda nanhaiensis
GCF_900101815.1 Pricia antarctica
GCF_003626755.1 Ulvibacterium marinum
GCF_000186265.1 Cellulophaga algicola
GCF_000468615.2 Cellulophaga baltica
GCF_000468575.1 Cellulophaga baltica
GCF_900116665.1 Pustulibacterium marinum
GCF_003344925.1 Flavobacterium arcticum
GCF_000769915.1 Flavobacterium beibuense
GCF_000425505.1 Flavobacterium frigidarium
GCF_002222045.1 Flavobacterium piscis
GCF_003148525.1 Flavobacterium araucananum
GCF_002222055.1 Flavobacterium araucananum
GCF_002217275.1 Flavobacterium frigidimaris
GCF_900142685.1 Flavobacterium chilense
GCF_003254585.1 Flavobacterium sharifuzzamanii
GCF_000733945.1 Flavimarina sp. Hel_I_48
GCF_002744755.1 Leeuwenhoekiella nanhaiensis
GCF_000687915.1 Leeuwenhoekiella sp. MAR_2009_132
GCF_900114265.1 Flaviramulus basaltis
GCF_000723205.1 Formosa agariphila
GCF_000423005.1 Gelidibacter mesophilus
GCF_001280435.1 Mangrovimonas xylaniphaga
GCF_001280505.1 Mangrovimonas sp. TPBH4
GCF_000688335.1 Sediminibacter sp. Hel_I_10
GCF_900099995.1 Winogradskyella thalassocola
GCF_003386275.1 Winogradskyella pacifica
GCF_900141785.1 Aquimarina spongiae
GCF_001632745.1 Aquimarina aggregata
GCF_003626835.1 Aquimarina sp. AD10
GCF_000430645.1 Aquimarina latercula
GCF_000430645.1 Aquimarina latercula
GCF_000745315.1 Salegentibacter sp. Hel_I_6
GCF_001431365.1 Salegentibacter mishustinae
GCF_003254095.1 Salegentibacter mishustinae
GCF_002900095.1 Salegentibacter mishustinae
GCF_900112105.1 Zunongwangia mangrovi
GCF_002094855.1 Zunongwangia atlantica
GCF_000423045.1 Gramella portivictoriae
GCF_001951155.1 Gramella flava
GCF_002120265.1 Gramella flava
GCF_001680885.1 Polaribacter vadi
GCF_001761365.1 Polaribacter vadi
GCF_900188235.1 Lutibacter agarilyticus
GCF_900188235.1 Lutibacter agarilyticus
GCF_003260195.1 Lutibacter citreus
GCF_900188355.1 Lutibacter flavus
GCF_003426875.1 Lutibacter oceani
GCF_003384935.1 Lutibacter oceani
GCF_002843385.1 Labilibaculum manganireducens
GCF_002843315.1 Labilibaculum filiforme
GCF_003046625.1 Mangrovibacterium marinum
GCF_003610535.1 Mangrovibacterium diazotrophicum
GCF_003610535.1 Mangrovibacterium diazotrophicum
GCF_000626635.1 Draconibacterium orientale
GCF_900111425.1 Draconibacterium orientale
GCF_900129025.1 Mariniphaga anaerophila
GCF_900129025.1 Mariniphaga anaerophila
GCF_000974365.1 Geofilum rubicundum
GCF_000974365.1 Geofilum rubicundum
GCF_900112795.1 Thermophagus xiamenensis
GCF_000614895.1 Marinilabilia salmonicolor
GCF_000259075.1 Marinilabilia salmonicolor
GCF_001660705.2 Saccharicrinis aurantiacus
GCF_001660705.2 Saccharicrinis aurantiacus
GCF_000517085.1 Saccharicrinis fermentans
GCF_000517085.1 Saccharicrinis fermentans
GCF_001659685.2 Labilibacter marinus
GCF_001659685.2 Labilibacter marinus
GCF_001659685.2 Labilibacter marinus
GCF_000170795.1 Pedobacter sp. BAL39
GCF_000523515.1 Pedobacter sp. V48
GCF_002646595.1 Flavilitoribacter nigricans
GCF_000379765.1 Flexithrix dorotheae
GCF_003149185.1 Sediminitomix flava
GCF_003149185.1 Sediminitomix flava
GCF_000807855.2 Flammeovirga pacifica
GCF_000807855.2 Flammeovirga pacifica
GCF_000807855.2 Flammeovirga pacifica
GCF_001583495.1 Flammeovirga sp. SJP92
GCF_001583495.1 Flammeovirga sp. SJP92
GCF_000812565.1 Flammeovirga sp. OC4
GCF_000812565.1 Flammeovirga sp. OC4
GCF_000812565.1 Flammeovirga sp. OC4
GCF_003258705.1 Arcticibacterium luteifluviistationis
GCF_000331105.1 Fibrella aestuarina
GCF_000331105.1 Fibrella aestuarina
GCF_900100765.1 Catalinimonas alkaloidigena
GCF_900100765.1 Catalinimonas alkaloidigena
GCF_003347495.1 Pleomorphovibrio marinus
GCF_000427295.1 Cyclobacterium qasimii
GCF_900167975.1 Ohtaekwangia koreensis
GCF_003386095.1 Marinoscillum furvescens
GCF_003386095.1 Marinoscillum furvescens
GCF_002742335.1 Reichenbachiella sp. 5M10
GCF_003590315.1 Reichenbachiella sp. MSK19-1
GCF_003590315.1 Reichenbachiella sp. MSK19-1
GCF_900142205.1 Reichenbachiella agariperforans
GCF_900142205.1 Reichenbachiella agariperforans
Tree scale: 0.1
Class:
Flavobacteriia
Bacteroidia
Sphingobacteriia
Saprospiria
Cytophagia

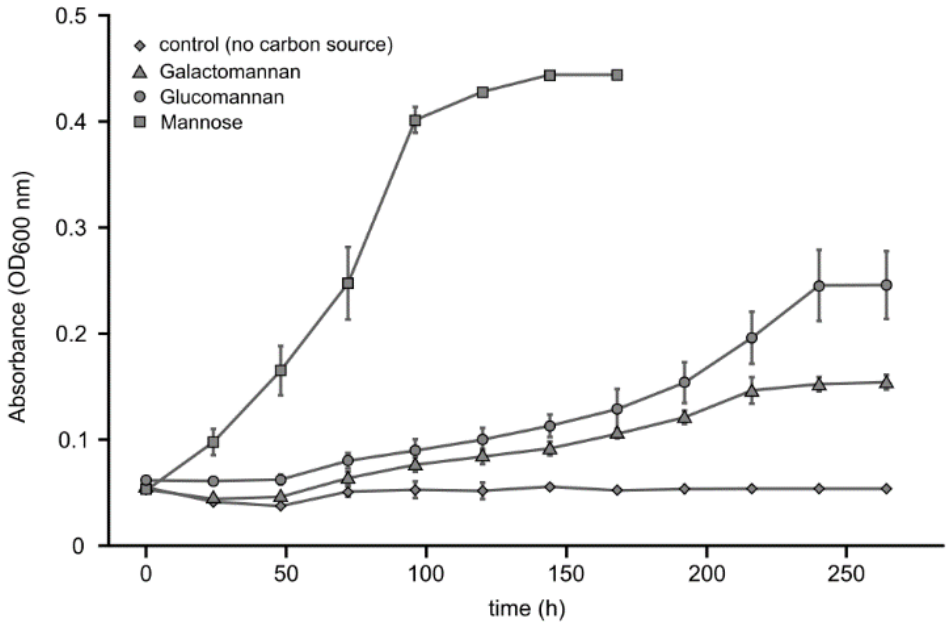



**Figure S4.3.** *Muricauda* sp. MAR_2010_75 is able to use multiple mannose-containing substrates as sole carbon source. The strain was grown in MPM-medium with 0.2% (w/v) specific poly- or monosaccharides until no significant changes in optical density could be detected over the course of 24 h after exponential growth (time points omitted). Values shown the mean of three replicates with standard deviation.

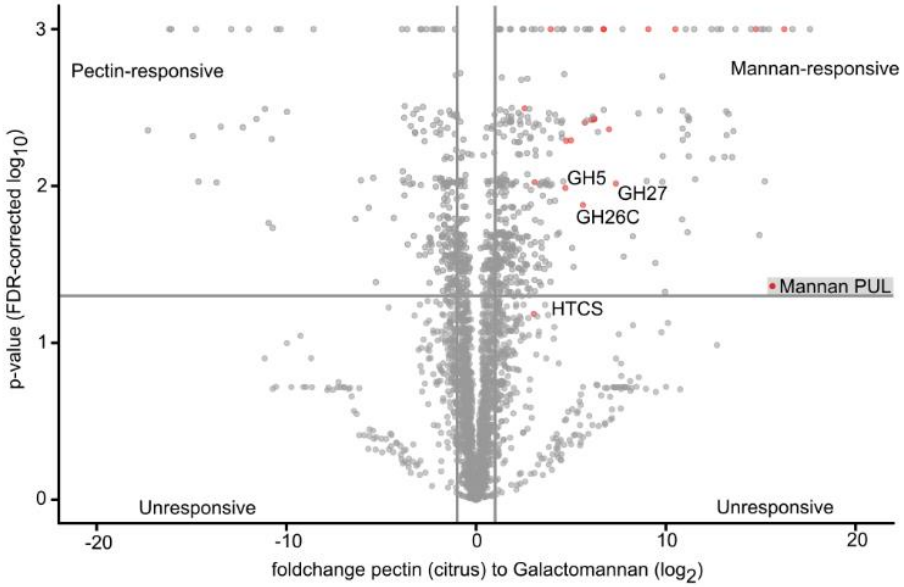



**Figure S4.4.** Growth on β-mannan-containing substrates specifically induces proteins of the *Muricauda* sp. MAR_2010_75 mannan-PUL in a statistically significant manner. Fold changes and p-values were calculated from MS-measurements of three biological replicates per substrate using the Perseus statistical software [5]. Proteins not detected in one condition were given an artificial value 1.25 times lower than the lowest detected

value in the sample to show their significance. Proteins belonging to the β-mannan PUL are shown in red.

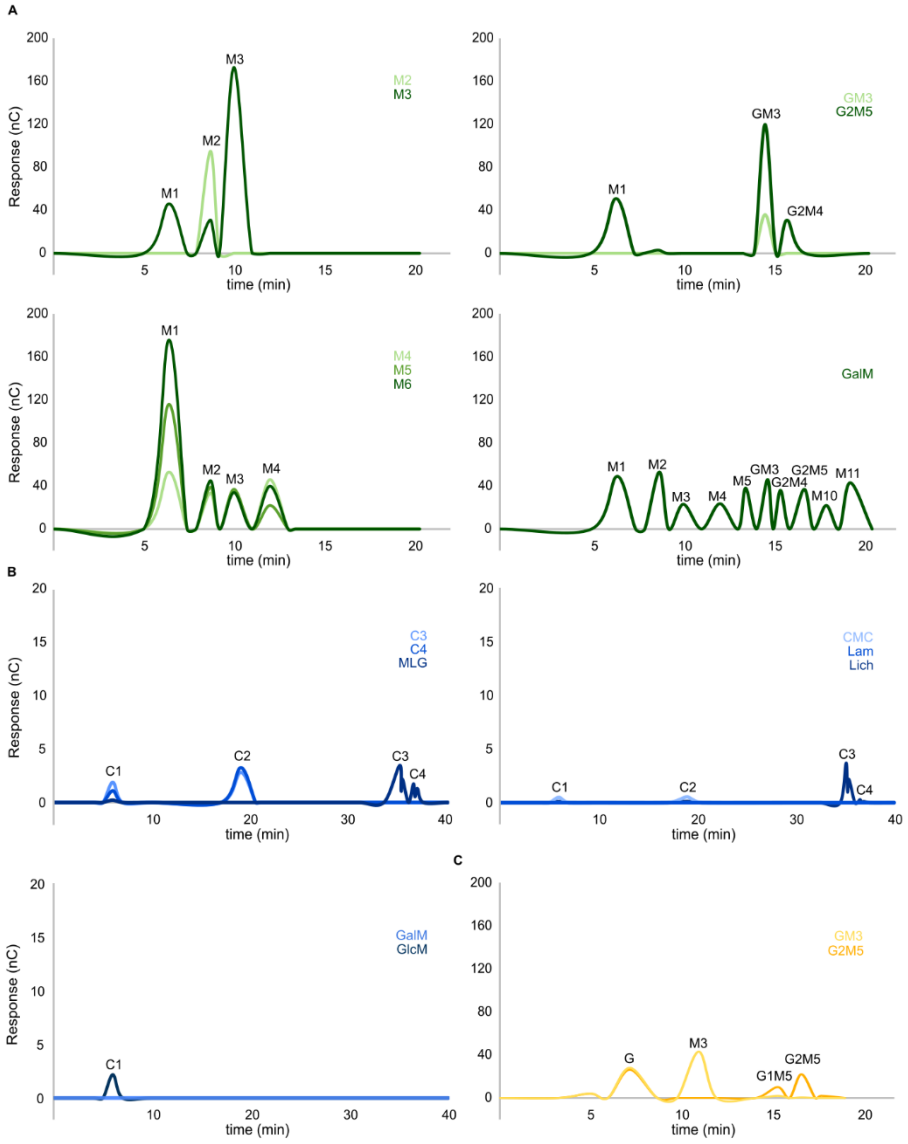



**Figure S4.5.** HPAEC-PAD data of recombinantly produced enzymes with different poly- and oligosaccharides. Enzymes GH26C (**A**), GH5_26 (**B**) and GH27 (**C**) were incubated with the respective substrate (in color) for 2 h. Samples were heat inactivated and then measured diluted 1:1000. Observed peaks were compared to mono- and oligosaccharide standards (black). M manno-, C cellu-, G galactomono or -oligosaccharides with dp given as the number behind them.

# 5. Marine *Bacteroidetes* enzymatically digest xylans from terrestrial plants

**Authors**: Theresa Dutschei[1], Irena Beidler[2], Daniel Bartosik[2,3] Julia-Maria Seeßelberg[4], Michelle Teune[1], Marcus Bäumgen[1], Soraia Querido Ferreira[1], Julia Heldmann[1], Felix Nagel[5], Joris Krull[3,6], Leona Berndt[7], Karen Methling[8], Martin Hein[9], Dörte Becher[10], Peter Langer[9], Mihaela Delcea[5], Michael Lalk[8], Michael Lammers[7], Matthias Höhne[4], Jan-Hendrik Hehemann[3,6], Thomas Schweder[2,3], Uwe T. Bornscheuer[1]

[1]Department of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, University Greifswald, Greifswald, Germany

[2]Department of Pharmaceutical Biotechnology, Institute of Pharmacy, University of Greifswald, Greifswald, Germany

[3]Institute of Marine Biotechnology e.V., Greifswald, Germany

[4]Department of Protein Biochemistry, Institute of Biochemistry, University of Greifswald, Greifswald, Germany

[5]Department of Biophysical Chemistry, Institute of Biochemistry, University of Greifswald, Greifswald, Germany

[6]Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany

[7]Department of Synthetic and Structural Biochemistry, Institute of Biochemistry, University of Greifswald, Greifswald, Germany

[8]Department of Cellular Biochemistry and Metabolomics, Institute of Biochemistry, University of Greifswald, Greifswald, Germany

[9]Department of Organic Chemistry, Institute of Chemistry, University of Rostock, Rostock, Germany

[10]Department of Microbial Proteomics, Institute of Microbiology, University of Greifswald, Greifswald, Germany

Published in *Environmental Microbiology*, Online ahead of print, May 2023

## 5.1. Abstract

Marine Bacteroidetes that degrade polysaccharides contribute to carbon cycling in the ocean. Organic matter, including glycans from terrestrial plants, might enter the oceans

through rivers. Whether marine bacteria degrade structurally related glycans from diverse sources including terrestrial plants and marine algae was previously unknown. We show that the marine bacterium *Flavimarina* sp. Hel_I_48 encodes two polysaccharide utilization loci (PULs) which degrade xylans from terrestrial plants and marine algae. Biochemical experiments revealed activity and specificity of the encoded xylanases and associated enzymes of these PULs. Proteomics indicated that these genomic regions respond to glucuronoxylans and arabinoxylans. Substrate specificities of key enzymes suggest dedicated metabolic pathways for xylan utilization. Some of the xylanases were active on different xylans with the conserved β-1,4-linked xylose main chain. Enzyme activity was consistent with growth curves showing *Flavimarina* sp. Hel_I_48 uses structurally different xylans. The observed abundance of related xylan-degrading enzyme repertoires in genomes of other marine Bacteroidetes indicates similar activities are common in the ocean. The here presented data show that certain marine bacteria are genetically and biochemically variable enough to access parts of structurally diverse xylans from terrestrial plants as well as from marine algal sources.

## 5.2. Introduction

Complex carbohydrates photosynthetically produced by marine algae or of terrestrial origin [87] can sequester carbon in the ocean [88, 89]. Such polysaccharides are major components of land plants and marine algal biomass [4, 90]. They serve as intracellular energy storage, structural cell wall components [91] and extracellular matrix [88, 92]. Many marine bacteria are known to degrade these complex polymeric structures, especially those produced by algae [93–95]. Members of the phylum Bacteroidetes are selected in such glycan-rich environments [96]. Recent studies showed the utilization of a diverse range of algal glycans by members of this phylum [97–103]. Their carbohydrate-active enzymes, or CAZymes, are often co-located with transporters and regulatory proteins in polysaccharide utilization loci (PULs) within bacterial genomes [104]. CAZymes are specific towards their target polysaccharide and its decorations [88, 101, 105], raising the question whether marine bacteria can also degrade glycans from terrestrial plants that reach the ocean.

Rivers transport terrestrial plant-derived organic matter into marine coastal regions [106–109]. The climate change boosters this carbon input into the ocean by mobilizing soil-bound organic matter [110]. In the Arctic, the warming of permafrost soils leads to an increased mobilization of organic material via river runoff [111–114]. This export of carbon from land

includes material from vascular plants, which contribute unknown quantities of organic carbon to the ocean [115]. Glycan structures common in plants, including xylans, have been detected in sediments close to the coast of the Baltic Sea [116]. Xylans and other glycans were also found to sequester carbon in over 1000-year old sediments [117] of the Red Sea. Presence of xylans in these two different oceanic regions indicates their degradation remains incomplete. Considering that marine bacteria apparently use them as a carbon and energy source, as suggested by import of fluorescent xylans by marine Bacteroidetes [118], and enzymatic degradation of fluorescent xylans added to seawater [119], this presence of xylans in marine sediments remains puzzling. To resolve the role glycans play for cycling versus sequestration of carbon, different perspectives are required, including biochemical characterization of xylan degrading pathways of marine bacteria.

Terrestrial xylans occur in cell walls of grains and wood in form of arabinoxylan, galactoarabinoxylan and glucuronoxylan sharing a β-1,4-linked D-xylopyranose main chain [120]. Besides the combination with other monosaccharides like arabinose, uronic acids and galactose, xylans available in the ocean may also be sulphated, acetylated or phosphorylated. Structural variability remains inaccessible unless bacteria have appropriate enzymes to react [121–126]. It has been shown that xylans are also present in green algae (*Chlorophyta/Charophyta*) and red algae (*Rhodophyta*) [127]. In red macroalgae, like *Palmaria* spp., the backbone is composed of β-1,4- or β-1,3-linked D-xylopyranose, depending on species and source [124, 128]. β-1,3-xylan structures were found in cell walls of green algae, like *Caulerpa* spp. [129, 130]. Xylose-containing polysaccharides have also been detected in microalgae [125, 131, 132]. There are potentially many different, unknown sources of xylans in the coastal ocean.

Their depolymerization requires dedicated enzymatic cascades. Degradation of terrestrial xylans by bacteria has been progressively researched due to its abundance and ecological relevance [133–140]. This extensive previous work led to the discovery that the marine Bacteroidetes strain *Flavimarina* sp. Hel_I_48 contains two PULs with homologues to xylanases from terrestrial bacteria (PUL I: P162_RS02310-P162_RS02395; PUL II: P162_RS04015-P162_RS04080). These two PULs might provide the ability to consume xylans of unknown origin [141]. While both PULs share related xylanases and xylosidases of the same CAZyme family, they differ in some of the associated CAZymes. The two PULs also contain genes for an unusually high number of six different pairs of transporter and receptor proteins for polysaccharide uptake (SusC/D-like pairs).

In this study, we analysed xylan degradation by these two PULs, which are common in other marine Bacteroidetes [142]. CAZymes and SusD-like binding proteins of these PULs were investigated with four terrestrial and two marine xylans as substrates, covering different xylan structures and common motifs of L-arabinose and D-glucuronic acid decorations. The work shows that marine bacteria can not only consume xylans from algae but also plant xylans that bring terrestrial carbon energy into the ocean.

## 5.3. Materials and Methods

### *5.3.1. Bioinformatics and comparative genomics*

Databases were created using NCBI RefSeq assemblies [143] of prokaryotic genomes stored in the RefSoilv1 database [144], Mar-Refv1.7 and MarDBv1.6 [145] as well as the NIH Human Microbiome Project [146, 147] catalogue with isolation body site 'gastrointestinal_tract'. Genomes were screened for *Flavimarina* sp. Hel_I_48 PUL I- and PUL II-like gene clusters with the 'hmm' search function of cblaster [148] v1.3.14 (default settings) using the HMM profiles 'GH67.hmm', 'GH115.hmm', 'CE15.hmm', 'GH43_1. hmm' (all PUL I), 'GH43_10.hmm', 'GH97.hmm', 'GH43_12.hmm', 'CE6.hmm', 'GH8.hmm', 'GH95. hmm' (all PUL II) and 'GH10.hmm' as marker profile from the dbCAN-HMMdb-V10.hmm database [149]. CAZyme context of clusters encoding a GH10 with at least two other glycoside hydrolases from PUL I and/or PUL II was predicted using the hmmscan function of HMMer v3.3.2 [150] against the dbCAN-HMMdb-V10.hmm database. Hits were filtered using the hmmscan-parser.sh script from dbCAN and validated using Protein–Protein BLAST (v2.11.0+) [151] against CAZyDB (release 09242021) with an e-value threshold of E-20, a minimum sequence identity of 30% and a query coverage of at least 40% [142]. The resulting gene clusters were visualized with UpSetR [152, 153] and RIdeogram [154]. For the phylogenetic tree, rpoB genes were aligned using the ClustalW [155] web service in 'slow/accurate' mode and default settings. Maximum-likelihood phylogenies were estimated by the PhyML 3.0 web server [156] with default settings and visualized with iTOL [157].

The *Flavimarina* sp. Hel_I_48 PUL repertoire was annotated as described above using additionally the TIGRFAM profile 'TIGR04056.hmm' for prediction of SusC-like proteins and the PFAM models 'PF07980.11.hmm', 'PF12741.7.hmm', 'PF14322.6.hmm' or 'PF12771.7.hmm' for prediction of SusD-like proteins. Final PULs were predicted as described previously [131], excluding sulphatase-encoding genes and visualized using Circos [158].

### *5.3.2. Proteome analysis*

*Flavimarina* sp. Hel_I_48 was grown to the late exponential phase in modified MPM medium [159] containing 0.1% beechwood xylan (BX), Palmaria palmata xylan (PPX), Caulerpa prolifera xylan (CPX), rye arabinoxylan (RAX), and wheat arabinoxylan of medium viscosity (WAX-M), insoluble wheat arabinoxylan (WAX-I), or apple pectin (Pec) as sole carbon sources. Triplicates of 50 mL cultures were harvested at 4000 xg, 20 min and 4°C. Cells were resuspended in 50 mM TEAB buffer containing 4% SDS and incubated at 95°C and 600 rpm for 5 min (Thermomixer C, Eppendorf, Hamburg, Germany). Samples were cooled to room temperature before being placed in an ultrasonic bath for 5 min. Cell debris was removed via centrifugation at 14,000 xg and 4xC for 10 min.

Protein concentration was measured using the BCA Pierce Protein assay kit (Thermo Fisher Scientific Inc., Waltham, MA, USA). Twenty-five microgrammes protein per sample was loaded on a 10% 1D-SDS polyacrylamide gel and separated at 120 V for 90 min. Gels were fixed with a 40% ethanol/10% acetic acid solution and stained overnight using Coomassie G-250 (Candiano et al., 2004). Each sample was divided into 10 subsamples, de-stained using 30% acetonitrile in 200 mM (NH4)2CO3 and digested for 16 h using trypsin.

Peptides were separated as described previously [160] by reverse phase C18 column chromatography on a nano ACQUITY-UPLC (Waters Corporation, Milford, MA, USA) online-coupled to an LTQOrbitrap Classic mass spectrometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). Spectra were searched against a target-decoy protein sequence database including sequences and reverse sequences of *Flavimarina* sp. Hel_I_48 and common laboratory contaminants using MaxQuant [161]. Only proteins that could be detected in at least two out of three replicates were considered identified. Relative protein abundance values in % of all proteins in the same sample were manually calculated from iBAQ values as %riBAQ (relative intensity based absolute quantification). Data and results are available via the PRIDE partner repository [162] with identifier PXD033600.

### *5.3.3. Gene cloning and enzyme production*

Expression constructs of FI1_GH67, FI2_GH10, FI4_GH10, FI5_hyp, FI7_GH43_1, FI8_CE6 and FI9_hyp (**Tab. S5.2**) were prepared using the FastCloning strategy (primer sequences are listed in **Tab. S5.3**) [163] with genomic DNA from *Flavimarina* sp. Hel_I_48 as template for the amplification of the inserts. The genomic DNA was extracted as described previously [164]. The pET28 constructs were prepared as described recently [101].

*Escherichia coli* BL21(DE3) was transformed with pET28-based plasmids (expression constructs and gene cloning described in the supplementary information) harbouring the required genes. For expression, 50 mL LB or TB medium with 100 µg/mL kanamycin were inoculated from an overnight culture in LB containing 50 µg/mL1 kanamycin. The culture was grown at 37 °C and 180 rpm until optical density at 600 nm reached 0.8. Expression was then induced by adding 0.5 or 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the culture was cooled to 20 °C and incubated for 24 h. Enzyme purification is described in the supplementary information.

Plasmids containing the genes encoding for SusD-like proteins were introduced into chemo-competent E. coli BL21(DE3) cells. 4 mL of overnight culture in LB media was used to inoculate 600 mL TB media both containing 50 µg/ mL kanamycin. The cells were grown to an optical density at 600 nm (OD600) between 1–1.5 at 37 °C and 180 rpm. Temperature was lowered to 20 °C and IPTG was added to a final concentration of 1 mM as the OD600 reached 2–3. Cells were harvested after 16 h by centrifugation at 4000 g and 4 °C for 20 min including washing with 20 mM sodium phosphate, 500 mM sodium chloride at pH 8. Washed pellets were flash frozen in liquid nitrogen and stored at 20 °C until purification. The purification of the SusD-like proteins is described in the supplementary information.

### *5.3.4. Purification of Xylan*

*Palmaria palmata* dulse was purchased at Algenladen (Gießen, Germany). After milling the dry algae (25 g) biomass, it was extracted two times with dH2O (1 L) for 2 h at 70 °C [122]. Afterwards, the solid particles were removed and the water content reduced to a viscous consistency. The polysaccharide was precipitated by adding four volumes of cold ethanol. The precipitate was then separated from the ethanol fraction. Afterwards the alcohol insoluble fraction was dissolved in deionized water and freeze-dried. *Caulerpa prolifera* was extracted with the same protocol. This algal material was provided by the Ozeaneum (Stralsund, Germany). Monosaccharide composition of the self-extracted polysaccharides, charge and size of the used polymers were analysed (Tables S3 and S4). The xylan degradation products from enzymatic reactions and the conversion of purified oligomers were analysed by FACE. Untreated xylan was generally used at a concentration of 1 g/L, while purified sugar oligomers were used at 0.25 g/L. Incubation with the enzymes was performed overnight at room temperature. Additional enzymatic assays are described in the supplementary information.

### *5.3.5. Determination of reducing ends (DNS-assay)*

The dinitrosalicylic acid-assay (DNS-assay) was used to determine the reducing ends of the carbohydrates [165]. A 20 µL reaction sample of the biocatalysis and 20 µL of the colour reagent were combined and incubated at 100 °C for 15 min. After the samples were cooled down to room temperature, 180 µL of water was added, and the 200 µL were transferred to a microtiter plate to measure the absorption at 540 nm in a plate reader (Infinite M200 Pro, Tecan Group, Swiss).

### *5.3.6. Flurophore-assisted carbohydrate electrophoresis*

Fluorophore-assisted carbohydrate electrophoresis (FACE) was performed with 8-aminonaphthtalene-1,-3,6-trisulphonic acid (ANTS) as fluorophore adapted from [166]. Ten microlitres aliquots of the biocatalysis reaction were lyophilized and dissolved in 4 µL ANTS (0.2 M in water) solution and 4 µL NaCNBH3 (1 M in DMSO) solution. After incubation at 37 °C overnight in the dark, the samples were mixed with 20 µL loading buffer and 4 µL were loaded on a FACE-gel [167].

### *5.3.7. HPLC determination of oligosaccharide degradation products*

HPLC analysis for the determination of xylan oligosaccharide standards was performed using a Chrommaster HPLC system from Hitachi (equipped with a Hitachi Chrommaster 5310 column oven) and a detector (Hitachi Chrommaster 5450 RI detector). Analyses were performed with a mobile phase consisting of H2O with 0.005 M H2SO4 on a styrene/polyvinyl benzene copolymer column (SugarSep-H 10 µm 3008 mm) at 70 °C for 20 min. The flow rate was set to 0.5 mL/min.

## 5.4. Results

### *5.4.1.* Flavimarina *sp. Hel_I_48 grows on multiple xylans*

To test whether *Flavimarina* sp. Hel_I_48 (Figure 1) uses xylans of terrestrial and marine origin, we employed growth experiments and proteomics with different xylans as sole carbon source (**Fig. S5.2**). We could show that the bacterium is able to grow on different xylans (**Fig. S5.1**). Proteome analyses showed divergent expression patterns for the two PULs depending on the xylan type. PUL I (P162_RS02310-RS02390) was expressed with beechwood xylan (BX), while PUL II (P162_RS04015-RS04080) remained silent with this substrate (**Fig. 5.1**). Notably,

the expression of SusD_I_2 (P162_RS02355) and SusD_I_3 (P162_RS02370) of PUL I was significantly higher on beechwood xylan compared to the other xylans. SusD_I_3 (P162_RS02370) accounted for 1% of the proteome. This high abundance indicates a specific adaptation of PUL I towards glucuronoxylan, which was recently identified in the coastal ocean with input from the Elbe river [132]. In contrast, higher protein levels of SusD_I_1 (P162_RS02310) from PUL I were observed on all investigated xylans, most notably on the marine β-1,4/1,3-mix linked xylan of *P. palmata* (PPX). The broad expression response indicates that this SusD-like protein enables the utilization of the general xylan main-chain structures shared among xylans of different origin. Expression of SusD_II_1 (P162_RS04065) and SusD_II_2 (P162_RS04075) of PUL II was not induced on PPX, BX or *C. prolifera* xylan (CPX), but was high with arabinoxylans from rye (RAX) and wheat (WAX-M) (**Fig. 5.1**). These xylans contain a β-1,4-xylan main chain with side chains of L-arabinose at the C3 or C2 position [168], indicating preference of this PUL region towards such xylans. A separate PUL a with putative arabinofuranosidases (P162_RS00625-RS00655) was significantly upregulated during growth on arabinoxylans compared to the other substrates (**Fig. 5.1** & **Tab. S5.1**), which revealed that *Flavimarina* sp. can also consume arabinoxylan. The arabinoxylan was also detected in diatoms [125] by using a wheat arabinoxylan-based antibody, which identified xylan in samples of microalgal blooms. Thus, these growth and proteomic data demonstrated that *Flavimarina* sp. is able to access xylans from different sources including algae and terrestrial plants.

### *5.4.2. SusD-like proteins select diverse xylans*

PULs I and II of *Flavimarina* sp. encode six SusD-like proteins that were divergently expressed during growth on different xylan substrates (**Fig. 5.1**). We investigated the interaction of recombinantly expressed and purified PUL-encoded SusD-like proteins with diverse xylans (BX, PPX, WAX-M, RAX, laminarin and no substrate as controls). SusD-like proteins assist in the transport of glycan oligosaccharides through SusClike pores into the periplasm [169, 170]. Alternative SusD-like proteins might be required to select for xylan oligosaccharides with different molecular architectures. Affinity gel electrophoresis (**Fig. S5.3**) showed that SusD_I_2 (P162_RS02355) interacts with glucuronoxylan BX, as it was slower in the electrophoresis with BX. This slower migration correlates with the upregulation observed in the proteome analysis (**Fig. 5.1**). SusD_I_2 migrated as three bands or as one in the presence of PPX and RAX (**Fig. S5.3**). These multiple bands, which might represent oligomeric states of

the SusD-like proteins, potentially resolved by concentration into one band by the stronger binding of xylan, or xylan types altering oligomerization states [171, 172].

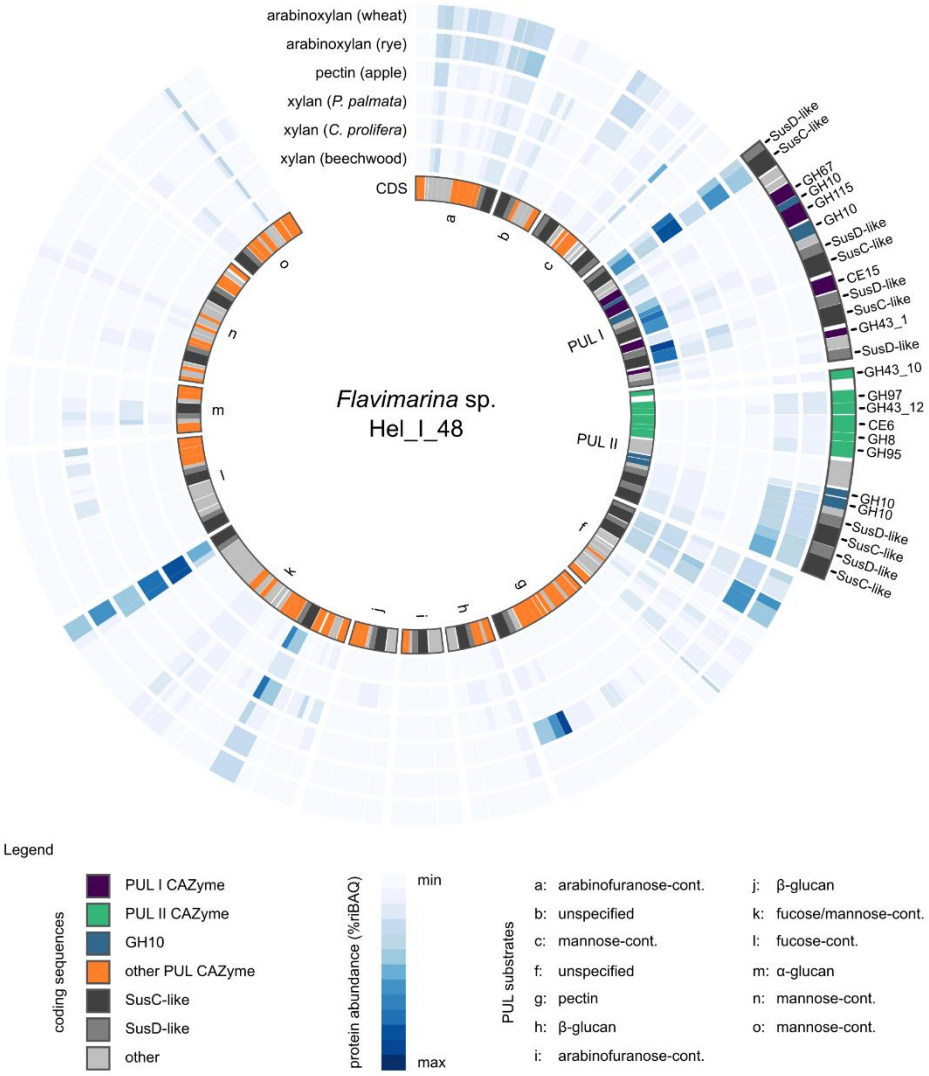



**Figure 5.1 PULs and proteomic profiles of *Flavimarina* sp. Hel_I_48.** The bacterium was cultivated on different xylose-containing substrates as sole carbon source and with pectin (apple) as control (see **Fig. S5.1**). Shown are the abundances (%riBAQ) of all PUL-encoded proteins on each substrate (see **Tab. S5.1**) with darker blue colour indicating a higher abundance. Proteins encoded in either of the two xylan-PULs are annotated on the outside of the ring.

The absence of detectable interaction with the laminarin control indicates that SusD_I_2 selects for the structure of β-1,4-linked xylan and against other glycans (**Fig. S5.3**). With

increased hydrodynamic diameter of the polymer (BX < WAXM < RAX < PPX; see **Tab. S5.5**) three bands merged into one as also seen for PPX (**Fig. S5.3**). In general, ligand size contributes to recognition by SusD-like proteins [171, 172]. AlphaFold based structural comparison of SusD_I_2 with its closest characterized homologue, a SusD from *Bacteroides ovatus* (Bacova_02651, PDB 5E76) [173], revealed a similar overall structure with a sequence identity of 39% (**Fig. S5.4A**). Previously, Bacova_02651 was shown to bind xyloglucan, structurally related to the glucuronoxylan bound here by SusD_I_2. Comparing binding site hydrophobicity revealed similarities, consistent with SusD_I_2 to select linear conformation xylan with smaller side chains such as glucuronoxylan (**Fig. S5.4B**). Sequence alignments of the six PUL-encoded SusD-like proteins gave pairwise identities between 19.5% (SusD_II_1 and Sus-D_II_2) and 27.6% (SusD_I_1 and SusD_I_4) with a high number of gaps (22.6%–39.7%) (**Fig. S5.5** & **S5.6**). The detected sequence diversity might indicate recognition of different xylan structures by these six SusD-like proteins.

### *5.4.3.* Flavimarina *sp. xylanases hydrolyse different xylans*

Extracellular enzymes often catalyse initial degradation of polysaccharides into oligosaccharides. Transporters import oligosaccharides for downstream processing through other specialized CAZymes [93]. We studied the hydrolytic activity of the PUL-encoded enzymes on xylans from different sources (RAX, WAX-M, WAX-I, BX, PPX and β-1,3-linked CPX) using DNS-reducing end assay and fluorophore-assisted carbohydrate electrophoresis (FACE) (**Fig. S5.7–S5.10**). Three xylanases contained a lipoprotein signal peptide cleaved by signal peptidase II (Sec/SPII) for extracellular localization on the cell surface [174]. FI4_GH10 (P162_RS02345) and FII8_GH10 (P162_RS04060) degraded β-1,4-xylan polymers RAX, WAX-M, WAX-I and BX as well as the β-1,4/1,3 linked PPX [122, 129], into ladder type oligosaccharides (**Fig. S5.7** & **S5.8**). This is commonly observed for enzymes with an endo-mode of action [175]. No activity was observed with the β-1,3-linked CPX. FII7_GH10 (P162_RS04050) was inactive on all xylans. Xylanase activity was detected for FI2_GH10 (P162_RS032335) on RAX, WAX-M, WAX-I, BX and PPX (**Fig. 5.2, S5.7** & **S5.8**). FI2_GH10 was inactive on CPX.

FI2_GH10 removed D-xylose and xylobiose (XBI) from xylan-oligosaccharides. Kinetic assisted subside mapping on successively shortened, defined oligosaccharides in conjunction with substrate complex crystal structures and additional experiments would be required to unambiguously ascertain endo- or exomode of action (**Fig. S5.11b**) [134, 176, 177]. While FI2_GH10 was able to cleave xylose from arabinoxylooligosaccharides with an arabinose

substitution in the middle of the oligosaccharide (**Fig. S5.11c**, $XA^3XX$), it was inactive on oligosaccharides with arabinose side chain at the non-reducing end (**Fig. S5.11c**, $A^2XX$, $A^3X$). FI4_GH10 and FII8_GH10 hydrolysed different xylans (**Fig. S5.7**) potentially explained by the presence of segments of an undecorated β-1,4-xylose main chain in the xylan or the acceptance of decorations in some of the unknown subsites. Acceptance of decorations enables productive binding of different xylans including mixed β-1,4-/1,3-xylans and branches with D-glucuronic acid or L-arabinose, present in BX or arabinoxylans (RAX, WAX-M, WAX-I). Extracellular localization and attachment to the outer membrane, in combination with broad activity shows that *Flavimarina* sp. can access a variety of xylans to extract carbon and energy.

### *5.4.4. Intracellular enzymatic xylan degradation*

The promiscuous extracellular xylanases from both PULs were active on different xylans and deliver the oligosaccharides for the remaining intracellular CAZymes encoded by the two PULs. PUL I encodes two putative α-glucuronidases, FI1_GH67 (P162_RS02330) and FI3_GH115 (P162_RS02340), potentially active on glucuronoxylan. Enzymatic assays with *p*NP-α-D-glucuronide showed glucuronidase activity for FI1_GH67 (**Fig. 5.3** & **S5.12**).

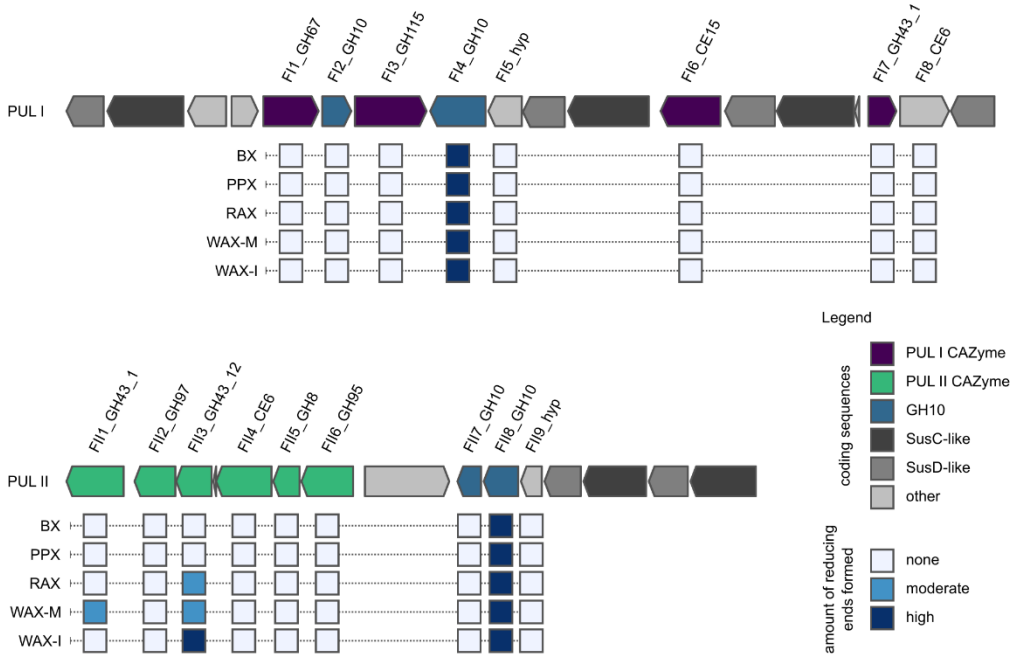



**Figure 5.2. PUL organization and initial xylan degradation of *Flavimarina* sp. Hel_I_48.** The xylan PULs of *Flavimarina* sp. were defined by [141]. Genome loci refer to the RefSeq assembly (GCF_000733945.1) of PUL I (P162_RS02310-RS02395) and PUL II

(P162_RS04015-RS04080). Enzyme activity was determined on the following polysaccharides: beechwood xylan (BX), *Palmaria palmata* xylan (PPX), rye arabinoxylan (RAX), wheat arabinoxylan of medium viscosity (WAX-M) and insoluble wheat arabinoxylan (WAX-I). After incubation, the formed reducing ends were measured using the DNS-reducing end assay (**Fig. S5.8**). Enzyme activity was designated as moderate when the average absorbance at 540 nm (after subtracting the negative control without enzyme) was between 0.1 and 0.2 and as high for values >0.2.

Oligosaccharides made by FI4_GH10, were not active substrates for GHs from PUL I (**Fig. S5.9**). FII5_GH8 (P162_RS04035) was active on oligosaccharides generated by FII8_GH10 (**Fig. S5.10**). Combinations of enzymes shifted the bands of BX and PPX derived oligosaccharides to lower molecular size (**Fig. S5.13**). For RAX, subtle band shifts without intensification of lower molecular weight band were observed. No further activity could be shown on WAX-M and WAX-I. A FACE experiment showed the presence of a band on the same position of the xylose and xylobiose standards. HPLC and TLC revealed that FII5_GH8 removed D-xylose from arabinoxylooligosaccharide $A^2XX$ (**Fig. S5.11** & **S5.14b**).

FII3_GH43 (P162_RS04025), which was upregulated on arabinoxylans, removed L-arabinose from arabinoxylo-oligosaccharides (**Fig. S5.14c**) and *p*NP- α-L-arabinofuranoside (**Fig. S5.12b**), in line with a predicted α-L-arabinofuranosidase function. Synthetic substrates revealed further activities of the PUL II-encoded enzyme FII2_GH97, annotated as α-galactosidase (EC.3.2.1.131). FII2_GH97 was active on *p*NP-α-Dgalactose (**Fig. S5.12c**). Terrestrial xylans contain a β- 1,4-linked D-xylose main chain with sugar modifications including galactose amounts varying by source [168]. D-galactose was almost absent in the here investigated xylans (**Fig. S5.2**), which therefore provided no substrate for FII2_GH97. A compilation of these findings is presented in the xylan degradation scheme in **Fig. 5.3**.

### *5.4.5. Carbohydrate esterases increase xylan degradation*

Both xylan PULs encode carbohydrate esterases (CE) known to increase the solubility of hemicellulose polymers by cleavage of acetyl groups or phenolic acids. As side groups are removed that cross-link different cell wall polymers, the polysaccharide backbone becomes more accessible to GHs [123, 178].

PUL I encodes FI6_CE15 (P162_RS02365), which was annotated as a CE15, a family containing 4-Omethyl-glucuronyl methylesterase activity [179, 180]. CE15 enzymes are involved in wood degradation, where they separate hemicelluloses and lignin-like moieties [181, 182].

FI6_CE15 was active on the model substrates *p*NP-acetate (**Fig. S5.12a**), benzyl-D-glucuronic acid and allyl-D-glucuronic acid (**Fig. S5.15c**). Also, in the ocean this enzyme might cleave methyl esters opening access for CAZymes to the xylan structure. The additional CBM9 domain of FI6_CE15 (**Tab. S5.2**), which is described as a xylan-targeting domain for birch and beechwood xylan, could aid in binding of the polysaccharide. CBM9 family modules can bind insoluble xylan polysaccharides and amorphous or crystalline cellulose [183, 184]. PUL I contains one additional carbohydrate esterase, FI8_CE6 (P162_RS02385), which is a CE6 acetyl xylan esterase [179, 180]. Deacetylase activity was shown by removal of *O*-acetylation from 6-*O*-acetyl-D-glucose and 6-*O*-acetyl-D-galactose detected with NMR-spectroscopy and an acetate assay (**Fig. S5.15b**). These two esterase activities from PUL I might be relevant in the degradation of glucuronoxylans. Hardwood xylans contain D-glucuronic acid side groups with methyl and acetylation of the β-1,4-linked D-xylose backbone [185].

PUL II contains two multimodular esterases, FII1_GH43_10 (P162_RS04015) and FII4_CE6 (P162_RS04030). FII1_GH43_10 consists of a putative acetyl xylan esterase CE3 domain, and an additional GH43_10 module, which is annotated as a xylan β-1,-4-xylosidase/α-L-arabinofuranosidase [179, 180]. This domain combination is similar to an enzyme described in Bacteroides eggerthii, containing an esterase and a GH43 domain involved in arabinoxylan degradation [186]. The putative CE3 and GH43_10 domains were separately analysed here as FII1A_CEnc and FII1B_GH43_10 constructs. Only FII1B_GH43_10 showed xylosidase activity in ANTS-FACE analysis (**Fig. S5.9**) and a DNS-assay (**Fig. 5.2**), but no esterase activity was found for FII1A_CEnc.

FII4_CE6 is an enzyme with two CE domains, a CE6/acetyl xylan esterase and an uncharacterized carbohydrate esterase module CEnc, which is likely to be a CE1/feruloyl esterase (**Tab. S5.2**) [179, 180]. Ferulic acid xylan esterases target phenolic groups bound to the Larabinose moieties in arabinoxylans, which enables α-Larabinofuranosidases to further degrade these polysaccharides [186]. This activity was very likely supported by the substantial amounts of ferulic acid released from WAX-I (**Fig. S5.15d**). FII4_CE6 also showed activity towards *p*NP-acetate and released acetate from partially acetylated birchwood xylan (**Fig. S5.15a**). A related multimodular CE6/CE1 protein was found in Bacteroides intestinalis and was described with a similar activity profile [186]. Both CE activities indicate removal of phenolic esters and acetylations, enabling PUL II to potentially use arabinoxylans with such modifications. The PUL-encoded esterases indicate that xylans targeted by *Flavimarina* sp. Hel_I_48 might include acetylations detected in macroalgal polysaccharides [187].

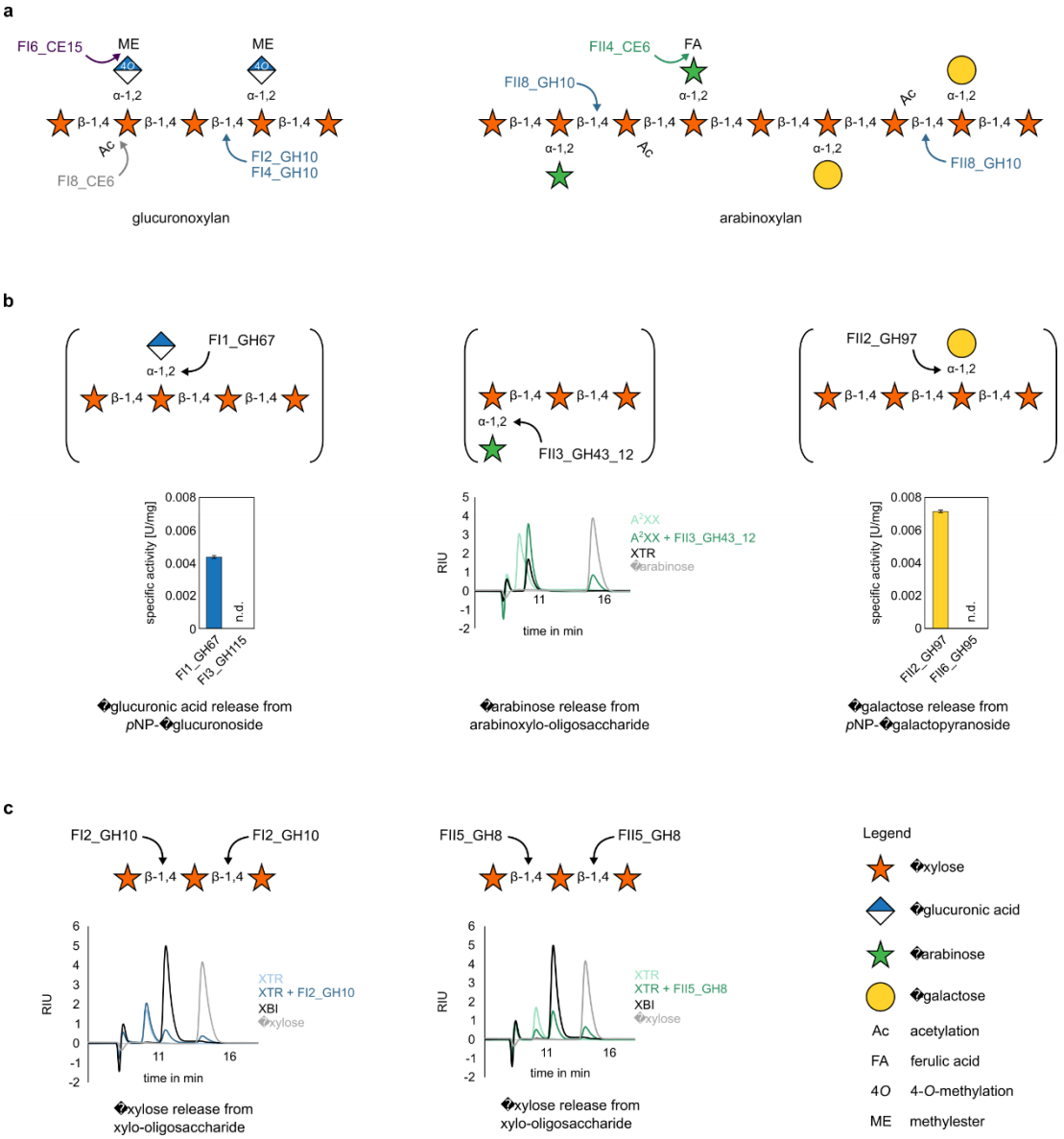



**Figure 5.3 Putative enzymatic cascade for the degradation of divergent xylan substrates by PUL I- and PUL II-encoded CAZymes.** The targeted structures are assumed to be similar to glucuronoxylan (PUL I) and arabinoxylan (PUL II). Summarized initial degradation (**a**) of polysaccharides by GH10 xylanases (FI2_GH10, FI4_GH10 and FII8_GH10 (**Fig. S5.7**), esterase FI8_CE6 cleaving acetate moieties (**Fig. S5.15b**), FII4_CE6 cleaving acetate and ferulic acid moieties (**Fig. S5.15a&d**) and FI6_CE15 cleaving 4-*O*-methyl-glucuronyl methyl esters (violet arrow; **Fig. S5.15c**). The resulting xylo-oligosaccharides can be further digested by other PUL-encoded enzymes, which was investigated using not only the degradation products but also artificial substrates (**b**). Representatively shown is the release of glucuronic acid (FI1_GH67; **Fig. S5.14d**) and galactose (FII2_GH97; **Fig. S5.14c**)

from *p*NP-derivates, while arabinose release from artificial oligos was detected upon digestion with Fll3_GH43_12 (**Fig. S5.11c** & **S5.14b**). Lastly the undecorated xylo-oligosaccharides can be further degraded (**c**) by xylosidaseactivity detected for e.g. Fl2_GH10 and Fll5_GH8 (**Fig. S5.13b**). $A^2XX$, $2^3$-α-L-arabinofuranosyl-xylotriose; XBI, xylobiose; XTR, xylotriose.

### *5.4.6. Ecological relevance of the xylan degradation pathways and PUL-architecture*

*Flavimarina* sp. degrades a multitude of xylans. This ability is largely facilitated by the broad activity of multiple PUL-associated GH10s on different xylan main chains containing various modifications. A GH10 with additional processing enzymes is likely required for degradation of complex xylans. We identified 226 PULs in marine, terrestrial and human digestive system-associated databases, sharing at least one GH10 and two additional CAZymes with either of the *Flavimarina* sp. PULs, indicating that complex xylans are important for diverse marine habitats (**Fig. S5.16**). A modularity search revealed 81 different xylan-PUL architectures, most commonly containing only a few shared enzymes with *Flavimarina* sp. (**Fig. 5.4a**). The most widely distributed PUL group contains only the enzymes known from *Flavimarina* sp. PUL I, which are GH43 and GH67 alongside a GH10 hydrolase. This enzyme combination was invariable and therefore the core of CAZymes required for xylan degradation. In some cases, bacteria containing such a PUL also contain a second, more complex PUL, such as the terrestrial bacterium *Flavobacterium johnsoniae*, which likely facilitates niche-specific adaptations (**Fig. 5.4b**).

PUL sequence diversity points towards the ability to degrade different xylan structures that requires specific CAZyme repertoires. Most identified PULs showed no distinct separation between CAZymes homologous to *Flavimarina* sp. PUL I and PUL II. In fact, *Fibrisoma limi* and *Fibrella aestuarina*, the only other species shown to have two enlarged *Flavimarina*-like PULs, are also associated with marine coastal ecosystems (**Fig. 5.4b**). Some marine bacteria as well as human gut symbionts possess the entire CAZyme spectrum of the *Flavimarina* sp. PULs consolidated into a single PUL (**Fig. 5.4b**), underlining that similar xylans are carbon sources for microbes from different ecosystems. Notably, many of the related PULs also encode additional CAZymes associated with arabinose-containing polysaccharide degradation, including GH39, GH51 or GH146. *Flavimarina* sp. does not encode CAZymes of these families within its two main xylan-targeting PULs. Some of these GHs are encoded in the genome and associated with the arabinofuranosidase-containing PUL a that was specifically upregulated during growth on RAX and WAX (**Fig. 5.1**). These observations further indicate that

*Flavimarina* sp. has the genetic potential to degrade structurally diverse xylans from different sources available in the dynamic marine land interface of the coastal ocean.

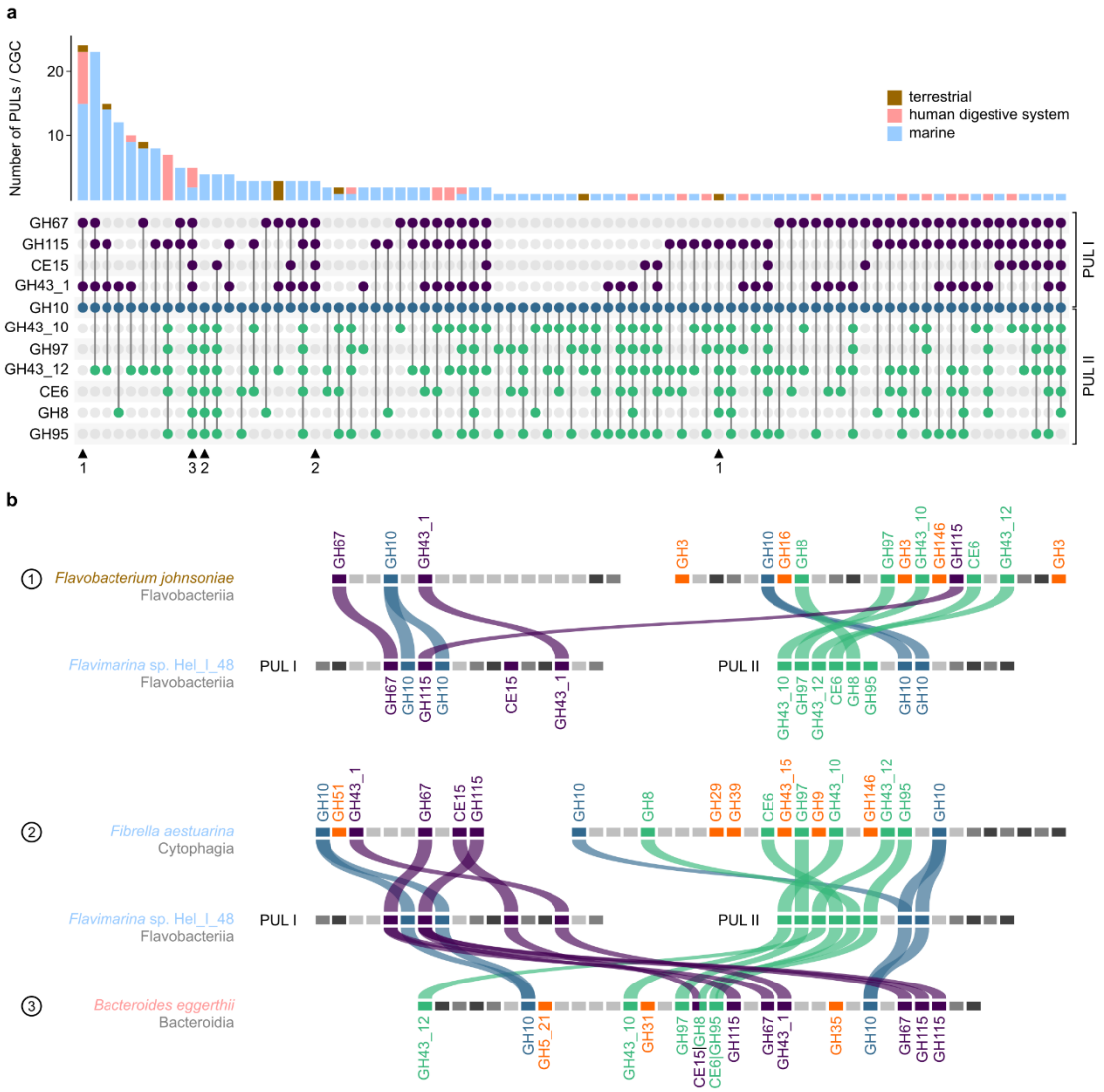



**Figure 5.4 Xylan PUL modularity of bacteria from different environments.** (**a**) Upset plot showing different xylan PUL compositions as well as their prevalence in different habitats. (**b**) Comparison of *Flavimarina* sp. PULs to other significant cluster types, such as those (1) containing a second, more complex PUL, (2) all CAZymes also encoded by *Flavimarina* sp. or (3) all CAZymes but encoded in a single larger PUL. CAZymes of *Flavimarina* sp. PUL I are depicted in purple, those of PUL II in green. CAZymes of families or subfamilies not encoded by either of the *Flavimarina* sp. PULs are marked in orange. See also the complementary phylogenetic tree in **Fig. S5.16** of the supplementary information.

## 5.5. Discussion

This study explores xylan-specific metabolic pathways of a marine *Bacteroidetes* strain. *Flavimarina* sp. Hel_I_48 contains two separate xylan PULs, active on glucuronoxylans (PUL I) and arabinoxylans (PUL II) with similarities to PULs of human gut bacteria and soil bacteria [177].

The general occurrence and abundance of xylan PULs in Bacteroidetes was previously described for gut microbiota [139], and xylan degradation was shown for enzymes of microbes from many different environments [135]. For example, the gut bacterium *B. ovatus* contains two xylan-targeting PULs [139], which encode CAZymes showing partially related activities to enzymes of *Flavimarina* sp. Hel_I_48. Our systematic biochemical analysis of PUL-encoded enzymes revealed promiscuous extracellular xylanases, which are active on diverse xylans and generate dedicated oligomers [177]. An overview of the putative enzymatic cascade for xylan degradation in *Flavimarina* sp. Hel_I_48 is depicted in **Fig. 5.3**. It is illustrated that these distinct oligosaccharides are further degraded by more specialized intracellular carbohydrate esterases, which remove acetylations and phenolic esters (**Fig. 5.3**). These enzymatic activities are supported by arabinofuranosidases and glucuronidases, which hydrolyse L-arabinose- and D-glucuronic acid-containing xylans.

Xylans used for the characterization of *Flavimarina* sp. Hel_I_48 PUL-encoded enzymes were from terrestrial plants and marine macroalgae covering structures from both ecosystems and providing a spectrum of xylan structures potentially available in marine habitats. The fact that not all of the PUL-encoded enzymes were found to be active in this study is to be expected, because we are missing xylan substrates from unknown sources. The extent of the potential xylan structure diversity space is likely proportional to the number of species that synthesize xylans and remains to be uncovered. The strain *Flavimarina* sp. Hel_I_48 was isolated during a phytoplankton bloom in the North Sea [141] near the island Helgoland, which is close to the river mouth of the Elbe, Germany's largest river that brings terrestrial organic matter into the ocean. Utilization of terrestrial organic matter, including carbohydrates that enter the ocean through rivers [106, 109] appears to be a selected strategy for the here tested bacterium. This assumption is supported by the detection of two other putative xylan-utilizing bacteria from marine coastal habitats, *F. limi* [188] and *F. aestuarina* [189], which share the same two enlarged *Flavimarina* sp. PUL architectures (see **Fig. 5.4**). The lack of sulphatase-encoding genes in these xylan-specific PULs of marine *Bacteroidetes* is another indication that they preferentially target terrestrial-like xylan structures. The simultaneous utilization of macroalgal and terrestrial xylans by these marine

PULs might be explained by conserved xylan structures with similar linkages and modifications present in marine algae and terrestrial plants [124, 190].

The comparative analysis of available *Bacteroidetes* genomes demonstrates that xylan utilization is ecologically relevant in marine habitats and ensured by conserved sets of related xylan-degrading enzymes (**Fig. 5.4** & **S5.16**). The documented functional activity of xylan utilization from the North Sea isolate *Flavimarina* sp. indicates that coastal habitats might provide heterogenous xylans with glucurono- and arabino-side groups and acetylation. Marine bacteria that can degrade such xylans from marine algae and terrestrial plants are adapted to access temporally dynamic sources of organic matter in the costal ocean. Their ability to degrade xylans contributes to the unknown magnitude of cycling or sequestration of this type of carbon.

## 5.6. Acknowledgements

We thank the German Research Foundation (DFG) for funding through the Research Unit FOR 2406 'Proteogenomics of Marine Polysaccharide Utilization' (POMPU) by grants to Uwe T. Bornscheuer (BO 1862/17-3), Jan-Hendrik Hehemann (HE 7217/2-3), Matthias Höhne (HO 4754/5-2), and Thomas Schweder (SCHW 595/10-3). We are grateful to the Ozeaneum Stralsund, which kindly provided the *C. prolifera* algal biomasses. Open Access funding enabled and organized by Projekt DEAL.

## 5.7. References

1. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. Science. 1998; 281:237–240.
2. Bligh M, Nguyen N, Buck-Wiese H, Vidal-Melgosa S, Hehemann J-H. Structures and functions of algal glycans shape their capacity to sequester carbon in the ocean. Current Opinion in Chemical Biology. 2022; 71:102204.
3. Repeta DJ, Aluwihare LI. Radiocarbon analysis of neutral sugars in high-molecular-weight dissolved organic carbon: Implications for organic carbon cycling. Limnol Oceanogr. 2006; 51:1045–1053.
4. Becker S, Tebben J, Coffinet S, Wiltshire K, Iversen MH, Harder T et al. Laminarin is a major molecule in the marine carbon cycle. Proc Natl Acad Sci U S A. 2020; 117:6599–6607.

5. Stefan Kraan. Algal Polysaccharides, Novel Applications and Outlook. In: Chuan-Fa Chang. Carbohydrates. IntechOpen, Rijeka, 2012. Ch. 22.
6. Kloareg B, Quatrano RS. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. Oceanography and Marine Biology. 1988; 26:259–315.
7. Hoagland KD, Rosowski JR, Gretz MR, Roemer SC. DIATOM EXTRACELLULAR POLYMERIC SUBSTANCES: FUNCTION, FINE STRUCTURE, CHEMISTRY, AND PHYSIOLOGY. Journal of Phycology. 1993; 29:537–566.
8. Arnosti C. Microbial Extracellular Enzymes and the Marine Carbon Cycle. Annual Review of Marine Science. 2010; 3:401–425.
9. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012; 336:608–611.
10. Yeh Y-C, Fuhrman JA. Effects of phytoplankton, viral communities, and warming on free-living and particle-associated marine prokaryotic community structure. Nature Communications. 2022; 13:7905.
11. McKee LS, La Rosa SL, Westereng B, Eijsink VG, Pope PB, Larsbrink J. Polysaccharide degradation by the Bacteroidetes: mechanisms and nomenclature. Environ Microbiol Rep. 2021; 13:559–581.
12. Barbeyron T, Thomas F, Barbe V, Teeling H, Schenowitz C, Dossat C et al. Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: example of the model algae-associated bacterium Zobellia galactanivorans Dsij(T). Environ Microbiol. 2016; 18:4610–4627.
13. Beidler I, Robb CS, Vidal-Melgosa S, Zühlke M-K, Bartosik D, Solanki V et al. Marine bacteroidetes use a conserved enzymatic cascade to digest diatom β-mannan. The ISME Journal. 2023; 17:276–285.
14. Ficko-Blean E, Préchoux A, Thomas F, Rochat T, Larocque R, Zhu Y et al. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. Nature Communications. 2017; 8:1685.
15. Hehemann J-H, Kelly AG, Pudlo NA, Martens EC, Boraston AB. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. Proc Natl Acad Sci U S A. 2012; 109:19786–19791.

16. Reisky L, Préchoux A, Zühlke M-K, Bäumgen M, Robb CS, Gerlach N et al. A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan. Nat Chem Biol. 2019; 15:803–812.
17. Robb CS, Hobbs JK, Pluvinage B, Reintjes G, Klassen L, Monteith S et al. Metabolism of a hybrid algal galactan by members of the human gut microbiome. Nat Chem Biol. 2022; 18:501–510.
18. Unfried F, Becker S, Robb CS, Hehemann J-H, Markert S, Heiden SE et al. Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms. The ISME Journal. 2018; 12:2894–2906.
19. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J Biol Chem. 2009; 284:24673–24677.
20. Hehemann J-H, Boraston AB, Czjzek M. A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. Curr Opin Struct Biol. 2014; 28:77–86.
21. Brockmann UH. Organic matter in the Elbe estuary. Netherland Journal of Aquatic Ecology. 1994; 28:371–381.
22. Herrmann N, Boom A, Carr AS, Chase BM, Granger R, Hahn A et al. Sources, transport and deposition of terrestrial organic material: A case study from southwestern Africa. Quaternary Science Reviews. 2016; 149:215–229.
23. Opsahl S, Benner R. Distribution and cycling of terrigenous dissolved organic matter in the ocean. Nature. 1997; 386:480–482.
24. Schefuß E, Schouten S, Schneider RR. Climatic controls on central African hydrology during the past 20,000 years. Nature. 2005; 437:1003–1006.
25. Hemingway JD, Schefuß E, Spencer RG, Dinga BJ, Eglinton TI, McIntyre C et al. Hydrologic controls on seasonal and inter-annual variability of Congo River particulate organic matter source and reservoir age. Chemical Geology. 2017; 466:454–465.
26. Peterson BJ, Holmes RM, McClelland JW, Vörösmarty CJ, Lammers RB, Shiklomanov AI et al. Increasing River Discharge to the Arctic Ocean. Science. 2002; 298:2171–2173.
27. Raymond PA, McClelland JW, Holmes RM, Zhulidov AV, Mull K, Peterson BJ et al. Flux and age of dissolved organic carbon exported to the Arctic Ocean: A carbon isotopic study of the five largest arctic rivers. Global Biogeochem. Cycles. 2007; 21

28. Shogren AJ, Zarnetske JP, Abbott BW, Iannucci F, Medvedeff A, Cairns S et al. Arctic concentration–discharge relationships for dissolved organic carbon and nitrate vary with landscape and season. Limnol Oceanogr. 2021; 66:S197-S215.
29. Tranvik LJ, Jansson M. Terrestrial export of organic carbon. Nature. 2002; 415:861–862.
30. Cragg SM, Friess DA, Gillis LG, Trevathan-Tackett SM, Terrett OM, Watts JEM et al. Vascular Plants Are Globally Significant Contributors to Marine Carbon Fluxes and Sinks. Annual Review of Marine Science. 2020; 12:469–497.
31. Salmeán AA, Willats WGT, Ribeiro S, Andersen TJ, Ellegaard M. Over 100-Year Preservation and Temporal Fluctuations of Cell Wall Polysaccharides in Marine Sediments. Front Plant Sci. 2022; 13:785902.
32. Silvia Vidal-Melgosa, Matija Lagator, Andreas Sichert, Taylor Priest, Jürgen Pätzold, Jan-Hendrik Hehemann. Not digested: algal glycans move carbon dioxide into the deep-sea. bioRxiv. 2022:2022.03.04.483023.
33. D'Ambrosio L, Ziervogel K, MacGregor B, Teske A, Arnosti C. Composition and enzymatic function of particle-associated and free-living bacteria: a coastal/offshore comparison. The ISME Journal. 2014; 8:2167–2179.
34. Reintjes G, Arnosti C, Fuchs BM, Amann R. An alternative polysaccharide uptake mechanism of marine bacteria. The ISME Journal. 2017; 11:1640–1650.
35. Ebringerová A. Structural Diversity and Application Potential of Hemicelluloses. Macromol. Symp. 2005; 232:1–12.
36. Bäumgen M, Dutschei T, Bornscheuer UT. Marine Polysaccharides: Occurrence, Enzymatic Degradation and Utilization. ChemBioChem. 2021; 22:2247–2256.
37. Deniaud E, Quemener B, Fleurence J, Lahaye M. Structural studies of the mix-linked beta-(1--3)/beta-(1--4)-D-xylans from the cell wall of Palmaria palmata (Rhodophyta). Int J Biol Macromol. 2003; 33:9–18.
38. Hettiarachchi SA, Kwon Y-K, Lee Y, Jo E, Eom T-Y, Kang Y-H et al. Characterization of an acetyl xylan esterase from the marine bacterium Ochrovirga pacifica and its synergism with xylanase on beechwood xylan. Microbial Cell Factories. 2019; 18:122.
39. Hsieh YSY, Harris PJ. Xylans of Red and Green Algae: What Is Known about Their Structures and How They Are Synthesised? Polymers (Basel). 2019; 11
40. Huang G, Vidal-Melgosa S, Sichert A, Becker S, Fang Y, Niggemann J et al. Secretion of sulfated fucans by diatoms may contribute to marine aggregate formation. Limnol Oceanogr. 2021; 66:3768–3782.

41. Viana AG, Noseda MD, Gonçalves AG, Duarte MER, Yokoya N, Matulewicz MC et al. β-D-(1→4), β-D-(1→3) 'mixed linkage' xylans from red seaweeds of the order Nemaliales and Palmariales. Carbohydr Res. 2011; 346:1023–1028.
42. Popper ZA, Michel G, Hervé C, Domozych DS, Willats WGT, Tuohy MG et al. Evolution and diversity of plant cell walls: from algae to flowering plants. Annu Rev Plant Biol. 2011; 62:567–590.
43. Jensen JK, Busse-Wicher M, Poulsen CP, Fangel JU, Smith PJ, Yang J-Y et al. Identification of an algal xylan synthase indicates that there is functional orthology between algal and plant cell wall biosynthesis. New Phytol. 2018; 218:1049–1060.
44. Lahaye M, Rondeau-Mouro C, Deniaud E, Buléon A. Solid-state 13C NMR spectroscopy studies of xylans in the cell wall of Palmaria palmata (L. Kuntze, Rhodophyta). Carbohydr Res. 2003; 338:1559–1569.
45. Mackie IM, Percival E. 227. The constitution of xylan from the green seaweed caulerpa filiformis. Journal of the Chemical Society (Resumed). 1959:1151–1156.
46. Francis TB, Bartosik D, Sura T, Sichert A, Hehemann J-H, Markert S et al. Changing expression patterns of TonB-dependent transporters suggest shifts in polysaccharide consumption over the course of a spring phytoplankton bloom. The ISME Journal. 2021; 15:2336–2350.
47. Vidal-Melgosa S, Sichert A, Francis TB, Bartosik D, Niggemann J, Wichels A et al. Diatom fucan polysaccharide precipitates carbon during algal blooms. Nature Communications. 2021; 12:1150.
48. Ferreira LM, Durrant AJ, Hall J, Hazlewood GP, Gilbert HJ. Spatial separation of protein domains is not necessary for catalytic activity or substrate binding in a xylanase. Biochem J. 1990; 269:261–264.
49. Gilbert HJ, Hazlewood GP, Laurie JI, Orpin CG, Xue GP. Homologous catalytic domains in a rumen fungal xylanase: evidence for gene duplication and prokaryotic origin. Mol Microbiol. 1992; 6:2065–2072.
50. Gilbert HJ, Hazlewood GP. Bacterial cellulases and xylanases. Microbiology. 1993; 139:187–194.
51. Gloster TM, Williams SJ, Roberts S, Tarling CA, Wicki J, Withers SG et al. Atomic resolution analyses of the binding of xylobiose-derived deoxynojirimycin and isofagomine to xylanase Xyn10A. Chemical Communications. 2004:1794–1795.

52. Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O et al. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. Nature. 2014; 506:498–502.
53. Larsbrink J, Thompson AJ, Lundqvist M, Gardner JG, Davies GJ, Brumer H. A complex gene locus enables xyloglucan utilization in the model saprophyte Cellvibrio japonicus. Mol Microbiol. 2014; 94:418–433.
54. Rogowski A, Briggs JA, Mortimer JC, Tryfona T, Terrapon N, Lowe EC et al. Glycan complexity dictates microbial resource allocation in the large intestine. Nature Communications. 2015; 6:7481.
55. Tull D, Withers SG. Mechanisms of cellulases and xylanases: a detailed kinetic study of the exo-beta-1,4-glycanase from Cellulomonas fimi. Biochemistry. 1994; 33:6363–6370.
56. Kappelmann L, Krüger K, Hehemann J-H, Harder J, Markert S, Unfried F et al. Polysaccharide utilization loci of North Sea Flavobacteriia as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. The ISME Journal. 2019; 13:76–91.
57. Krüger K, Chafee M, Ben Francis T, Del Glavina Rio T, Becher D, Schweder T et al. In marine Bacteroidetes the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. The ISME Journal. 2019; 13:2800–2816.
58. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022; 50:D20-D26.
59. Choi J. RefSoil Database. 2017
60. Klemetsen T, Raknes IA, Fu J, Agafonov A, Balasundaram SV, Tartari G et al. The MAR databases: development and implementation of databases specific for marine metagenomics. Nucleic Acids Res. 2018; 46:D692-D699.
61. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT et al. Structure, function and diversity of the healthy human microbiome. Nature. 2012; 486:207–214.
62. Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C et al. A framework for human microbiome research. Nature. 2012; 486:215–221.
63. Gilchrist CLM, Booth TJ, van Wersch B, van Grieken L, Medema MH, Chooi Y-H. cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters. Bioinform Adv. 2021; 1:vbab016.

64. Zhang H, Yohe T, Le Huang, Entwistle S, Wu P, Yang Z et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018; 46:W95-W101.
65. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011; 39:W29-37.
66. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402.
67. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph. 2014; 20:1983–1992.
68. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017; 33:2938–2940.
69. Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G et al. RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput Sci. 2020; 6:e251.
70. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22:4673–4680.
71. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol. 2010; 59:307–321.
72. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021; 49:W293-W296.
73. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19:1639–1645.
74. Schut F, Vries EJ de, Gottschal JC, Robertson BR, Harder W, Prins RA et al. Isolation of Typical Marine Bacteria by Dilution Culture: Growth, Maintenance, and Characteristics of Isolates under Laboratory Conditions. Appl Environ Microbiol. 1993; 59:2150–2160.
75. Otto A, Bernhardt J, Meyer H, Schaffer M, Herbst F-A, Siebourg J et al. Systems-wide temporal proteomic profiling in glucose-starved Bacillus subtilis. Nature Communications. 2010; 1:137.
76. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology. 2008; 26:1367–1372.

77. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022; 50:D543-D552.
78. Li C, Wen A, Shen B, Lu J, Huang Y, Chang Y. FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. BMC Biotechnology. 2011; 11:92.
79. Chen WP, Kuo TT. A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. Nucleic Acids Res. 1993; 21:2260.
80. Bernfeld P. [17] Amylases, α and β. In: Methods in Enzymology. Academic Press, 1955; 1. pp 149–158.
81. Starr CM, Masada RI, Hague C, Skop E, Klock JC. Fluorophore-assisted carbohydrate electrophoresis in the separation, analysis, and sequencing of carbohydrates. J Chromatogr A. 1996; 720:295–321.
82. Hehemann J-H, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature. 2010; 464:908–912.
83. Bastawde KB. Xylan structure, microbial xylanases, and their mode of action. World J Microbiol Biotechnol. 1992; 8:353–368.
84. Pollet RM, Martin LM, Koropatkin NM. TonB-dependent transporters in the Bacteroidetes: Unique domain structures and potential functions. Mol Microbiol. 2021; 115:490–501.
85. Koropatkin NM, Martens EC, Gordon JI, Smith TJ. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure. 2008; 16:1105–1115.
86. Gray DA, White JBR, Oluwole AO, Rath P, Glenwright AJ, Mazur A et al. Insights into SusCD-mediated glycan import by a prominent gut symbiont. Nature Communications. 2021; 12:44.
87. Glenwright AJ, Pothula KR, Bhamidimarri SP, Chorev DS, Baslé A, Firbank SJ et al. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. Nature. 2017; 541:407–411.
88. Tauzin AS, Kwiatkowski KJ, Orlovsky NI, Smith CJ, Creagh AL, Haynes CA et al. Molecular Dissection of Xyloglucan Recognition in a Prominent Human Gut Symbiont. mBio. 2016; 7:e02134-15.

89. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nature Biotechnology. 2022; 40:1023–1025.
90. Davies G, Henrissat B. Structures and mechanisms of glycosyl hydrolases. Structure. 1995; 3:853–859.
91. Pell G, Taylor EJ, Gloster TM, Turkenburg JP, Fontes, Carlos M G A, Ferreira LMA et al. The mechanisms by which family 10 glycoside hydrolases bind decorated substrates. J Biol Chem. 2004; 279:9597–9605.
92. Gilbert HJ. The biochemistry and structural biology of plant cell wall deconstruction. Plant Physiol. 2010; 153:444–455.
93. Williamson G, Kroon PA, Faulds CB. Hairy plant polysaccharides: a close shave with microbial esterases. Microbiology (Reading). 1998; 144 (Pt 8):2011–2023.
94. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014; 42:D490-5.
95. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009; 37:D233-8.
96. Santi C de, Gani OA, Helland R, Williamson A. Structural insight into a CE15 esterase from the marine bacterial metagenome. Scientific Reports. 2017; 7:17278.
97. Charavgi MD, Dimarogona M, Topakas E, Christakopoulos P, Chrysina ED. The structure of a novel glucuronoyl esterase from Myceliophthora thermophila gives new insights into its role as a potential biocatalyst. Acta Crystallogr D Biol Crystallogr. 2013; 69:63–73.
98. Notenboom V, Boraston AB, Kilburn DG, Rose DR. Crystal structures of the family 9 carbohydrate-binding module from Thermotoga maritima xylanase 10A in native and ligand-bound forms. Biochemistry. 2001; 40:6248–6256.
99. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J. 2004; 382:769–781.
100. Kmezik C, Krska D, Mazurkewich S, Larsbrink J. Characterization of a novel multidomain CE15-GH8 enzyme encoded by a polysaccharide utilization locus in the human gut bacterium Bacteroides eggerthii. Scientific Reports. 2021; 11:17662.
101. Pereira GV, Abdel-Hamid AM, Dutta S, D'Alessandro-Gabazza CN, Wefers D, Farris JA et al. Degradation of complex arabinoxylans by human colonic Bacteroidetes. Nature Communications. 2021; 12:459.

102. Wong T, Brault L, Gasparotto E, Vallée R, Morvan P-Y, Ferrières V et al. Formation of Amphiphilic Molecules from the Most Common Marine Polysaccharides, toward a Sustainable Alternative? Molecules. 2021; 26
103. Filippini M, Kaech A, Ziegler U, Bagheri HC. Fibrisoma limi gen. nov., sp. nov., a filamentous bacterium isolated from tidal flats. Int J Syst Evol Microbiol. 2011; 61:1418–1424.
104. Filippini M, Svercel M, Laczko E, Kaech A, Ziegler U, Bagheri HC. Fibrella aestuarina gen. nov., sp. nov., a filamentous bacterium of the family Cytophagaceae isolated from a tidal flat, and emended description of the genus Rudanella Weon et al. 2008. Int J Syst Evol Microbiol. 2011; 61:184–189.
105. Martone PT, Estevez JM, Lu F, Ruel K, Denny MW, Somerville C et al. Discovery of Lignin in Seaweed Reveals Convergent Evolution of Cell-Wall Architecture. Current Biology. 2009; 19:169–175.

## 5.8. Competing interests

The authors declare no conflicts of interest.

## 5.9. Supplementary information

### *5.9.1. Supplementary methods*

#### *5.9.1.1. Gene constructs for SusD-like proteins*

Codon-optimized genes encoding for SusD_I_1, SusD_I_2, SusD_I_3, SusD_I_4, SusD_II_1, SusD_II_2 and GM_SusD were ordered as “GenePart” fragments (GenScript Biotech (Netherlands) B.V., Leiden, Netherlands) containing a N-terminal hexahistidine-tag, a tobacco etch virus cleavage site, and N- and C-terminal overhangs for a modified pET28-based vector (**Tab. S5.2**). The amino acids leading up to the SPII cleavage site were excluded from the constructs as they just serve as signal peptide and are likely to be removed upon integration of SusD in the outer membrane [191]. Cloning was performed using the SLiCE method [192] with 125 ng of insert and a vector to insert ratio of 1:7. For preparation of the modified pET28 vector, the isolated plasmid was linearized with the restriction endonucleases FastDigest™ NcoI and XbaI (Thermo Scientifi, Thermo Fisher Scientific Inc., Waltham, USA) followed by heat inactivation of the restriction enzymes at 65 °C for 15 min. Due to observation of no expression of SusD_I_1, the tobacco etch virus cleavage site was

substituted with the small ubiquitin-like modifier (SUMO) fusion protein SMT3 from *S. cerevisiae* [193]. Therefore, a hexahistidine-tag containing the SMT3-fusion protein coding gene fragment was amplified from a pBAD vector with primers listed in **Tab. S5.2**. The PCR was conducted using Pfu Plus! DNA polymerase and its related buffer (EURx Ltd., Gdansk, Poland), 0.2 mM dNTPs, 0.5 µM primers (see **Tab. S5.3**) and 10 ng SMT3 containing plasmid DNA under the following conditions: 5 min of initial denaturation at 95 °C, followed by 30 cycles of 30 s of denaturation at 95 °C, 30 s of annealing at 56.5 °C and 24 s of extension at 72 °C. Final extension was conducted at 72 °C for 7 min. Subsequently, residual plasmid was digested with DpnI in CutSmart™ Buffer (NEB, New England Biolabs Inc., Ipswich, USA) at 37 °C for 1 h. 2 µg of the amplified gene fragment and SusD_I_1 containing pET28 plasmid, respectively, were incubated with 10 U SalI and 20 U XbaI (NEB, New England Biolabs Inc., Ipswich, USA) at 37 °C for 2 h and enzymes were heat inactivated at 80 °C for 2 h. The restriction digested plasmid was mixed with Gel Loading Dye, Purple (6X) (NEB®, New England Biolabs Inc., Ipswich, USA) and separated from potential circular plasmid DNA by gel electrophoresis in a 0.8% (w/v) agarose gel containing 0.005% (v/v) ROTISafe (Carl Roth, Karlsruhe, Germany). The agarose gel electrophoresis was conducted at 90 V for one hour and the linearized plasmid was processed with the Monarch DNA gel extraction kit (NEB, New England Biolabs Inc., Ipswich, USA). The insert was purified with a NucleoSpin gel and PCR Clean-up kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany). Hexahistidine-tagged SMT-3 was introduced into the SusD_I_1 containing pET28a plasmid using 50 ng of vector and 200 U T4 DNA Ligase (NEB, New England Biolabs Inc., Ipswich, USA) at 24 °C for two hours and a 1:5 ratio of vector (0.02 pmol) to insert (0.1 pmol).

#### *5.9.1.2. Protein purification*

CAZymes

Cell pellets for the purification of CAZymes from a 50 mL culture were thawed on ice and resuspended in 10 mL of ice-cold resuspension buffer (Tris-HCl 50 mM, pH 8 + 300 mM NaCl + 10 mM imidazole). The cells were lysed by ultrasonication on ice (2 x 3 min, 50% power, 50% cycle time) and the cell debris was removed by centrifugation (15 min at 10,000 x g). Rotigarose-His/Ni beads (Carl Roth, Karlsruhe, Germany) incubated with the clarified lysate were used in gravity flow columns. After washing, the protein was eluted with elution buffer (Tris-HCl-50 mM, pH 8 + 100 mM NaCl + 300 mM imidazole). Fractions containing the protein of interest were pooled and desalted using PD-10 columns (GE Healthcare, Freiburg, Germany) equilibrated with 50 mM Tris-HCl (pH 8 + 10 mM NaCl). The desalted enzymes

were aliquoted in tubes, flash frozen in liquid nitrogen and stored at -20 °C. The protein concentration was determined with the Roti®-Nanoquant kit with an albumin standard (0100 µg/mL).

FI8A_CEnc and FII1A_CEnc were additionally purified via size exclusion chromatography for the NMR-analysis. The gravity flow purified enzyme was loaded to a Superdex 200 16/600 (Cytiva, Marlborough, MS, USA) (50 mM TRIS-HCl pH 7.5, 100 mM NaCl, 2 mM β-Mercaptoethanol) and the resulting protein fraction was then dialysed overnight in 20 mM K2HPO4/KH2PO4 buffer pH 7.7 with 100 mM NaCl.

SusD-like proteins

The SusD-expression cell pellets of 600 mL culture were thawed on ice and resuspended in 20 mL of cooled lysis buffer (20 mM sodium phosphate, 500 mM sodium chloride, 20 mM imidazole, 5% (v/v) glycerol, pH 8). Cell lysis was performed at 1,000 atm with a Maximator HPL6 device (MAXIMATOR GmbH, Nordhausen, Germany) and the cell suspension was lysed twice. Cell debris was removed by centrifugation at 10,000 x g and 4 °C for 1 h. The supernatant was filtered using a 0.45 µm filter. Protein purification was performed at room temperature using a Cytiva HisTrapHP 1 mL column (Cytiva Europe GmbH, Freiburg, Germany) at an Äkta Pure device (Cytiva Europe GmbH, Freiburg, Germany) with a flow rate of 1 mL min-1 and an imidazole gradient ranging from 20 mM to 250 mM imidazole within 7 min. Purification buffers contained 20 mM sodium phosphate, 500 mM sodium chloride and no or 500 mM imidazole respectively at pH 8. SusD-like proteins eluted between 65-86 mM imidazole. SDS-PAGE using 10% (v/v) polyacrylamide gels was performed to identify fractions containing the protein of interest. Proteins were stained with ROTIBlue (Carl Roth, Karlsruhe, Germany). Those fractions containing the desired protein were united and the volume was adjusted to 1 mL with Vivaspin 10,000 MWCO ultrafiltration units (Sartorius Stedim Lab Ltd., Stonehouse, UK) at 4,000 x g and 4 °C for the required time. Subsequently, the samples were centrifuged at 17,000 x g at 4 °C for 10 min in a table top centrifuge to obtain aggregate free samples for injection. Size exclusion chromatography was performed at room temperature with a flow rate of 0.75 mL min-1 using a Cytiva Superdex 200 Increase 10/300 GL column (Cytiva Europe GmbH, Freiburg, Germany) in 20 mM sodium phosphate, 250 mM sodium chloride at pH 8. The proteins were concentrated again with Vivaspin 10,000 MWCO ultrafiltration units (Sartorius Stedim Lab Ltd., Stonehouse, UK). Protein concentrations were measured under consideration of their particular molecular weight and extinction calculated with the Expasy ProtParam tool [194] using a NanoDrop 1000 Spectrophotometer (Thermo

Fisher Scientific Inc., Waltham, USA). Aliquots of 1 mg mL-1 were flash frozen in liquid nitrogen and stored at -20 °C.

### *5.9.1.3. Supplementary carbohydrate analyses*

Monosaccharide composition analysis

The self-extracted polysaccharides were chemically hydrolyzed (1 M HCl for 24 h at 100 °C). Afterwards, the samples were filtered using a 0.2 µm Spin-X filter prior to HPAEC-PAD analyses (**Tab. S5.4**) using a Dionex CarboPac PA10 column (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and monosaccharide mixtures as standards for column calibration [195].

Dynamic light scattering

Hydrodynamic diameters of higher order polysaccharide structures and zeta potentials were determined using dynamic light scattering (DLS). Backscattering was recorded at 173 nm and samples were equilibrated at 25 °C. Zeta potential measurements were carried out using a maximum voltage of 10 V. Data represent mean and standard deviation of at least three independent experiments (**Tab. S5.5**).

### *5.9.1.4. Supplementary enzymatic assays*

pNP-assay chromogenic substrates

screening 1 mM solutions of 4-nitrophenyl-/D-galacturonide/L-arabinose/acetate were prepared in 50 mM HEPES buffer, pH 7.4 with 100 mM NaCl. 10 µL of purified enzyme solution were added to 200 µL of the substrate solutions and the absorbance at 410 nm was measured over 30 min to detect the formation of p-nitrophenolate. The latter was taken from Bowers *et al*. [196] with a concentration of 18.3 mM. Autohydrolysis was determined by no addition of enzyme.

Affinity gel electrophoresis

Affinity gel electrophoresis (AGE) was performed using Tris-acetate based gels with 10% (v/v) acrylamide [197] casted as native gels without SDS but containing 0.5% (w/v) of polysaccharide (PPX, BX, RAX, WAX-M and Laminarin from *Eisenia bicyclis* from TCI Europe N.V., Zwijndrecht, Belgium) or no polysaccharide as reference, respectively. 5 µg protein in native charge buffer (63 mM Tris-HCl pH 6.8, 10% (v/v) glycerol, 0.01% (w/v) bromophenole blue) were applied and AGE was conducted on ice at 80 V for 6 hours in cooled buffer

containing 25 mM Tris, 192 mM glycine at pH 8.3. Proteins were stained with ROTIBlue (Carl Roth, Karlsruhe, Germany).

Acetate release

The release of acetate from polymeric xylan and from the partially acetylated beechwood xylan from Megazymes (Bray, Ireland) was measured after the biocatalytic reactions with the esterases performed at room temperature for 16 h. Aacetic acid detection was performed with the acetic acid kit from R-Biopharm (Darmstadt, Germany). The release of acetate from 6-Oacetylated D-glucose and 6-*O*-acetylated D-galactose were measured with the acetate detection kit from Megazymes (Bray, Ireland). 2 µM enzyme were incubated with 10 mM of the substrate and incubated at 25 °C overnight. The enzyme was removed via Ni-NTA-beads (Cube Biotech, Monheim, Germany) before the sample was analyzed.

Ferulic acid release

The release of ferulic acid from WAX-I was measured by incubating the FII4_CE6 with 10 mg mL-1 of the substrate, in a 50 mM sodium phosphate 100 mM buffer pH 6.5 taking time samples (2 h, 4 h, 8 h, 1 d, 2 d) incubating at 37 °C and 1,000 rpm. After incubation the enzymes were heat inactivated at 90 °C for 10 min. An equal volume of methanol was added to the cooled down samples, vortexed for 30 s and centrifuged at 13,000 x g for 5 min. The resulting supernatant was analyzed via ultra-high-performance liquid chromatography with pulse amperometric detector (U-HPLC, Agilent) injected with 5 µL, flow rate 0.8 mL min-1 with the liquid phase water / acetonitrile / formic acid. UV detection was carried out at 325 nm. A standard curve using ferulic acid was used for determining its concentration in the hydrolysate.

D-glucuronic acid release

The measurement of CE-activity towards D-glucuronic acid-derivatives (methyl-, benzyl-, allyl- D-glucuronic acid) was performed in a discontinuous attempt using the K-URONIC Assay Kit (Megazyme) as described previously [198]. The enzyme reactions were performed for 30 min at 25 °C in 25 mM sodium phosphate buffer, measuring 3-4 different dilutions of each enzyme. To quantify the release of GlcA, the reaction solutions were transferred to a 96-well plate (200 µL each) and detection-solution (50 µL, 40 % NAD+, 6% UDH) was added to measure the release of NADH at 340 nm for 30-60 min. Preparation of 6-*O*-acetylated derivatives of D-glucose and D-galactose (6-*O*-acetyl-α,β-D-glucopyranose and 6-*O*-acetyl-

α,β-D-galactopyranose) The preparation of the 6-*O*-acetylated sugars was carried out according to a modified procedure described by Duff *et al*. [199]. 10 g of the sugar (D-glucose or D-galactose) were mixed with 67% aqueous acetic acid (20 ml) and stirred at 100 °C for 16-18 h. Then silica gel (30 g) was added and the solvent was removed at 40 °C in the RotaVap. The dried silica gel support was directly used for column chromatographic purification (eluent: toluene/methanol v/v =2.5:1) yielding the corresponding 6-*O*-acetylated sugar as a mixture which mainly consisted of the anomeric pyranoses (yield: 27% for 6-*O*-acetyl-D-glucose and 24% for 6-*O*-acetyl-Dgalactose). The products were identified by comparison of their NMR spectroscopic data with those from the literature. The glucose derivative can be further purified by crystallization from methanol-ethyl acetate. All used chemicals are commercially available and were used without further purification. Analytical TLC on Merck silica gel 60 F254 plates was visualized by using anisaldehyde-sulfuric acid colouring reagent in methanol. Column chromatography was performed on MerckGeduran Si 60 (0.063-0.200 mm). 1H NMR and 13C NMR spectra were recorded on a Bruker AVANCE DRX-500 or AVANCE 300 III. Chemical shifts in ppm were calibrated by residual solvent signals methanol-d4 (1H, 3.31 ppm, 13C, 49.00 ppm), D2O (1H, 4.79 ppm) or DMSO-d6 (1H, 2.50 ppm, 13C, 39.52 ppm).

1H-NMR spectroscopic based acetate quantification

1H-NMR analysis was performed as previously described with some modifications [200]. In brief, 400 µL of samples and 200 µL of 0.2 mol L-1 sodium hydrogen phosphate buffer solution, containing 30 % D2O (Euriso-Top, St-Aubin Cedex, France) and 1.5 mmol L-1 3-trimethylsilyl[2,2,3,3-D4]-1-propionic acid (Sigma-Aldrich, St. Louis, USA) were mixed in 5 mm glass tubes (103.5 mm length, Bruker Biospin GmbH, Rheinstetten, Germany). The Bruker AVANCE-NEO 600 NMR spectrometer equipped with a SampleJet autosampler and a 5 mm QCI cryo probe was operated by TOPSPIN 4.0.9 software (Bruker Biospin GmbH, Rheinstetten, Germany). Metabolite quantification was done using AMIX Viewer 3.9.15 software (Bruker Biospin GmbH, Rheinstetten, Germany). Integrals of the acetate peak were compared to the integral of the ERETIC signal for absolute quantification. The ERETIC signal was generated by external calibration with the ERETIC quantification tool based on PULCON [201].

Thin layer chromatography

In thin layer chromatography (TLC), (arabino-) xylo-oligosaccharides (Megazyme®, 2 mg mL1) were incubated with purified enzyme in 50 mM Tris pH 8.0, 100 mM NaCl at room temperature overnight. The negative control contained no enzyme. Xylo-oligosaccharides (1 mg mL-1) were mixed in reaction buffer and used as a standard (X1-6) for TLC analysis. Under analogous conditions the formation of oligosaccharides from natural substrates (5 mg mL-1 BX, PPX, RAX, WAX-M, WAX-I) was observed after incubation with Fll8_GH10 and/or Fll5_GH8. The hydrolysates were analysed as described before [202]) using silica gel plates (60 F245) and a mixture of 1-butanol, acetic acid and water (2:1:1) as a solvent. After spraying with staining solution (4 g α-diphenylamine, 4 mL aniline, 200 mL acetone, 30 mL phosphoric acid 80% (v/v)) the spots were visualized by heating over 100 °C.

*5.9.1.5. Supplementary computational analyses*

SusD-like proteins multiple sequence alignments

PUL I- and PUL II-encoded SusD-like protein sequences were aligned by ClustalO, Kalign, MAFFT, MUSCLE (using EMBL-EBI sequence analysis tools services, all default settings) [203] and COBALT [204]. Resulting multiple sequence alignments were combined using the T-Coffee [205] web server and visualized with Jalview 2 [206] and AlignmentViewer (release 1.0).

Sequence similarity analysis

For sequence similarity networks, all-by-all global sequence alignments of PUL I- and PUL IIencoded SusD-like protein sequences were created using EMBOSS Needle (default settings) [203]. Pairwise identity, similarity and the percentage of gaps were plotted using Cytoscape v3.9.0 [207].

Structural alignment

Structure models of *Flavimarina* sp. PUL I and PUL II encoded SusDs were predicted using AlphaFold [208]. The model of SusD_I_2 was aligned to its closest characterized homolog Bacova_02651 (5E76) [173] and visualized using PyMOL [209].

### *5.9.2. Supplementary tables*

Large supplementary tables can be viewed online at https://doi.org/10.1111/1462-2920.16390.

**Table S5.1: Proteomics results.** *Flavimarina* sp. Hel_I_48 was grown on different xylans and pectin from apple (control) as sole carbon source. Automatically calculated iBAQ values were used to determine relative % riBAQ values for semiquantitative comparison. Only proteins identified in at least two of three replicates were classed as identified.

**Table S5.2: Summary of the investigated carbohydrate active enzymes.** The annotation was performed using DBCan, Pfam, Interpro, and Hmmer. Only annotations provided by at least two tools were used. The genes were either ordered as synthetic genes at Biocat (Heidelberg, Germany) (1) or GenScript (NJ, USA) (2) or cloned via FastCloning technique (3)*: FI8A and FI8B differentiate between two optional lengths, FII1A and FII1B are separate modules of the gene P162_RS04015 and were ordered as separate codon optimized genes.

| Name | Gene Locus Taq | Annotation and modularity | Functional annotation | Gene origin |
|---|---|---|---|---|
| FI1_GH67 | P162_RS02330 | GH67 | n.d. | (3) |
| FI2_GH10 | P162_RS02335 | GH10 | Exo-1,4-xylanase | (3) |
| FI3_GH115 | P162_RS02340 | GH115\|GH115 | Alpha-(4-O-methyl)-glucuronidase (EC3.2.1.) | (2) |
| FI4_GH10 | P162_RS02345 | CBM4\|GH10 | Endo-xylanase | (3) |
| FI5_hyp | P162_RS02350 | hyp | n.d. | (3) |
| FI6_CE15 | P162_RS02365 | CE15\|CBM9 | 4-O-Methyl-glucuronyl methylesterase | (2) |
| FI7_GH43_1 | P162_RS02380 | GH43_1 | n.d. | (3) |
| FI8A_CEnc* | P162_RS02385 | Putative CE6 | Acetyl-xylan esterase | (3) |
| FI8B_CEnc* | P162_RS02385 | Putative CE6 | Acetyl-xylan esterase | (3) |
| FII1A_CEnc | P162_RS04015 | CE3\|GH43_10 | Acetyl-xylan esterase/ (xylosidase/arabinose) | (1) |
| FII1B_GH43_10 | P162_RS04015 | CE3\|GH43_10 | n. d. | (1) |
| FII2_GH97 | P162_RS04020 | GH97 | α-D-galactosidase | (1) |

| | | | | |
|---|---|---|---|---|
| FII3_GH43_12 | P162_RS04025 | GH43 | α-L-arabinofuranosidase | (1) |
| FII4_CE6 | P162_RS04030 | CE6\|CEnc\|CEnc (CEnc putative CE1) | Feruloyl xylan esterase /acetyl xylan esterase | (1) |
| FII5_GH8 | P162_RS04035 | GH8 | Exo-xylanase | (1) |
| FII6_GH95 | P162_RS04040 | GH95 | n.d. | (1) |
| FII7_GH10 | P162_RS04050 | GH10 | n.d. | (1) |
| FII8_GH10 | P162_RS04055 | CBM4\|GH10 | Endo-1,4-xylanase | (1) |
| FII9_hyp | P162_RS04060 | Hyp (DUF1735) | n.d. | (1) |
| GM SusD | BLT93_RS06685 | RagB/SusD family nutrient uptake outer membrane protein | Target substrate laminarin | (2) |
| SusD_I_1 | P162_RS02310 | RagB/SusD family nutrient uptake outer membrane protein | Target substrate xylan | (2) |
| SusD_I_2 | P162_RS02355 | RagB/SusD family nutrient uptake outer membrane protein | Target substrate xylan | (2) |
| SusD_I_3 | P162_RS02370 | RagB/SusD family nutrient uptake outer membrane protein | Target substrate xylan | (2) |
| SusD_II_1 | P162_RS04065 | RagB/SusD family nutrient uptake outer membrane protein | Target substrate xylan | (2) |
| SusD_II_2 | P162_RS04075 | RagB/SusD family nutrient uptake outer membrane protein | Target substrate xylan | (2) |

**Table S5.3: Primers used in this study.** The primers were purchased from Invitrogen Life Technologies (CA, USA).

| Primer name | Sequence 5' → 3' | Purpose |
|---|---|---|
| T7 pET mode | CCCGCGAAATTAATACGACTCAC | Sequencing |
| T7_term | CTAGTTATTGCTCAGCGGT | Sequencing |
| FI1 seq fw | GCGCTGGCTGATGTGTTTCGACC | Sequencing |
| FI1 seq fw2 | CCAAAAAGGAAAAGGATCCACC | Sequencing |
| FI4 seq fw | GTAAAGGCCGAATTTCTTTTG | Sequencing |
| FI8A seq fw | GATACTGATCAGGGAATCATTAATAACT | Sequencing |
| FI9 seq fw | GATCATGGTGGTGGTGGTGTATAT | Sequencing |
| FI9 seq fw2 | CGAGATGATTGATGGTACATTAAA | Sequencing |
| FI9 seq fw3 | CTAGCGCTAACGGTAATTTTGCTG | Sequencing |
| FI1A fw | CTGCTTCACATAACAGCTTCTGATGGC | Fast Cloning |
| FI1A rv | TTAATCCCATTGTGGCCGTATTCCC | Fast Cloning |
| FI1B fw | TTGAAACCGCATCCTACATTCCTTTTTATACTC | Fast Cloning |
| FI1B rv | ATGCTTGGGCAAGAAATAGAACTGCTACC | Fast Cloning |
| FI2 fw | TGCAAAAACGAGACAAAAACCAC | Fast Cloning |
| FI2 rv | TTATTCGTTGATGTCGGTTACTTTATAG | Fast Cloning |
| FI3 fw | CAGAAATCTGGTGATTATGTATCAAAAACAC | Fast Cloning |
| FI3 rv | CTATTTCACAACTTTACTTTCAGGAGGTCC | Fast Cloning |
| FI4 fw | TGTGAAGACGATATTATGGAGTGGCAGG | Fast Cloning |
| FI4 rv | CTAATCTTCAAGCTCTCCAGTGAAATCCTC | Fast Cloning |
| FI5 fw | TGTTCCAACGATGATGATGCTG | Fast Cloning |
| FI5 rv | TTATTCAGGAAAATCGGTAACGGTAGG | Fast Cloning |
| FI6 fw | CAACTTCCGTTGGTCTATAACTCTGAAAATACGG | Fast Cloning |
| FI6 rv | TTAAAGTGTTTCTCCTGCCAGCACCAC | Fast Cloning |
| FI7 fw | TGCAAAAATAACACAGATAAAGATTCCG | Fast Cloning |
| FI7 rv | TTATGGATTTTCTACCTTGGCATCAATAG | Fast Cloning |
| FI8A rv | CTACCAGTTATCTGTCCTAAAAGGTGAGGCAG | Fast Cloning |
| FI8B fw | CAGATCAAACTGCCAAAATTAGTTTCTGACG | Fast Cloning |
| FI9 fw | CAGGTAGTGACCAGCGGGGCAG | Fast Cloning |
| FI9 rv | CTAATTCAGTTGAACCGTTCCTCCTCTTGATG | Fast Cloning |

| | | |
|---|---|---|
| SMT3_amplification_fwd | GCTAGCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACGATGGGTCATCATCATCATCATCACGGCAGCG | Amplification of SMT3 with overhangs |
| SMT3_amplification_rev | GCACTACCATGGAACCACCAATCTGTTCTCTGTGAGC | Amplification of SMT3 with overhangs |

**Table S5.4: Elementary analysis of the self-extracted polysaccharides.**

| | N [%] | C [%] | S [%] | C/N ratio |
|---|---|---|---|---|
| *C. prolifera* (CPX) | 0.83 ± 0.10 | 35.05 ± 1.19 | 1.18 ± 0.15 | 42.31 ± 5.30 |
| *P. palmata* (PPX) | 1.09 ± 0.06 | 35.43 ± 0.50 | 2.04 ± 0.13 | 32.43 ± 1.77 |
| Beechwood (BX) | 0.10 ± 0.02 | 40.76 ± 0.39 | 0.16 ± 0.07 | 410.34 ± 96.70 |

**Table S5.5: Hydrodynamic diameters and ζ-potentials of polysaccharides determined by dynamic light scattering.**

| Sample | Hydrodynamic diameter (nm) | Zeta potential (mV) |
|---|---|---|
| Beechwood xylan (BX) | 189 ± 29 | -14.5 ± 4.3 |
| Palmaria palmata xylan (PPX) | 510 ± 37 | -33 ± 5.75 |
| Caulerpa prolifera xylan (CPX) | 172 ± 4 | -20.9 ± 1.9 |
| Wheat arabinoxylan medium viscosity (WAX-M) | 300 ± 21 | -4.7 ± 0.3 |
| Wheat arabinoxylan insoluble fraction (WAX-I) | 699 ± 509 | -4.3 ± 0.5 |
| Rye arabinoxylan (RAX) | 459 ± 8.7 | -5.25 ± 0.5 |

*5.9.3. Additional figures*

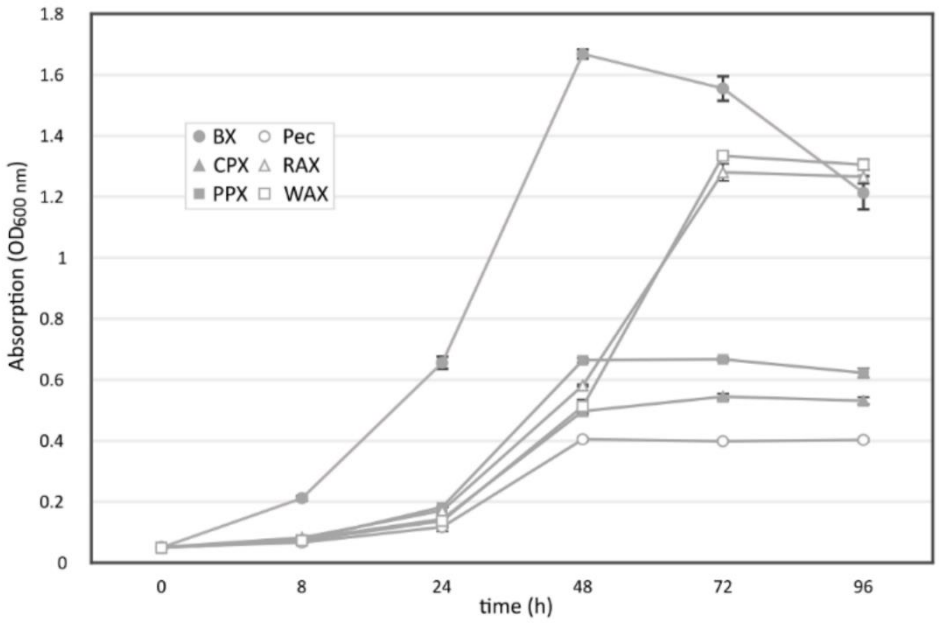



**Fig. S5.1. Growth pattern of *Flavimarina* sp. Hel_I_48 on different substrates.** Cultures were grown to stationary phase in MPM Medium with 0.1% of a specific carbon source to determine optimal sampling points for proteomics. The following substrates were tested: Beechwood xylan (BX), *Caulerpa prolifera* xylan (CPX), *Palmaria palmata* xylan (PPX), apple pectin (Pec), rye arabinoxylan (RAX), wheat arabinoxylan medium viscosity (WAX-M).

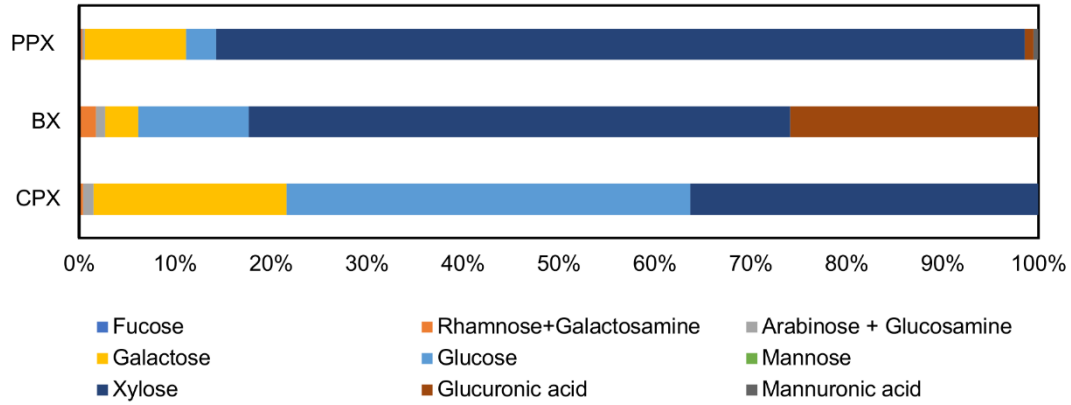



**Fig. S5.2. Monosaccharide composition analysis of the self-extracted polysaccharides** from PPX, BX and CPX.

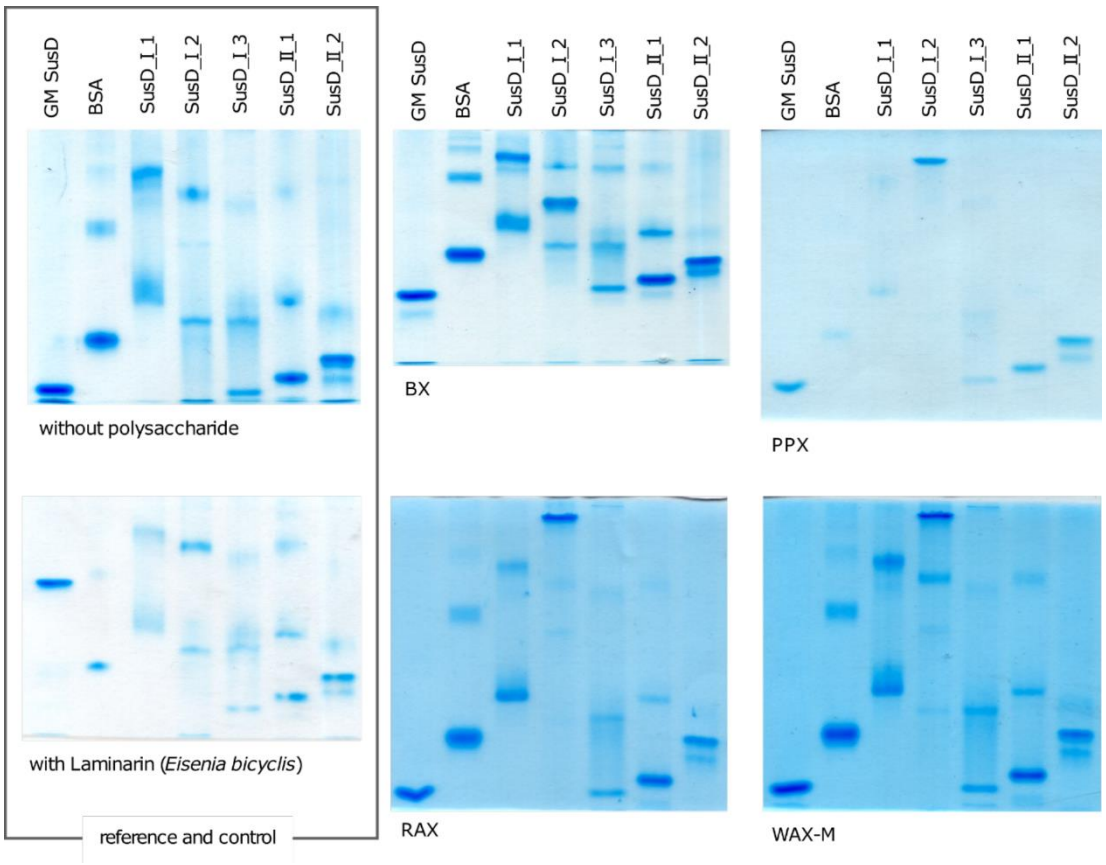



**Fig. S5.3. Tris-acetate based affinity gel electrophoresis of SusD-like proteins from PUL I and II on different xylan polysaccharides and laminarin from *Eisenia bicyclis*.** 0.5% of polysaccharides (BX, PPX, RAX, WAX-M, laminarin from *E. bicyclis*) were added to native PAGE gels before polymerization. A gel without polysaccharide served as reference. The gel with 0.5% laminarin from *E. bicyclis* was made to confirm functionality of this method. BSA acts as a reference marker and GM SusD [191] acts as a positive control showing a shift on the laminarin containing gel. 5 µg of protein were loaded and showed a shift upon binding to the polysaccharide and a separation of their multimers.

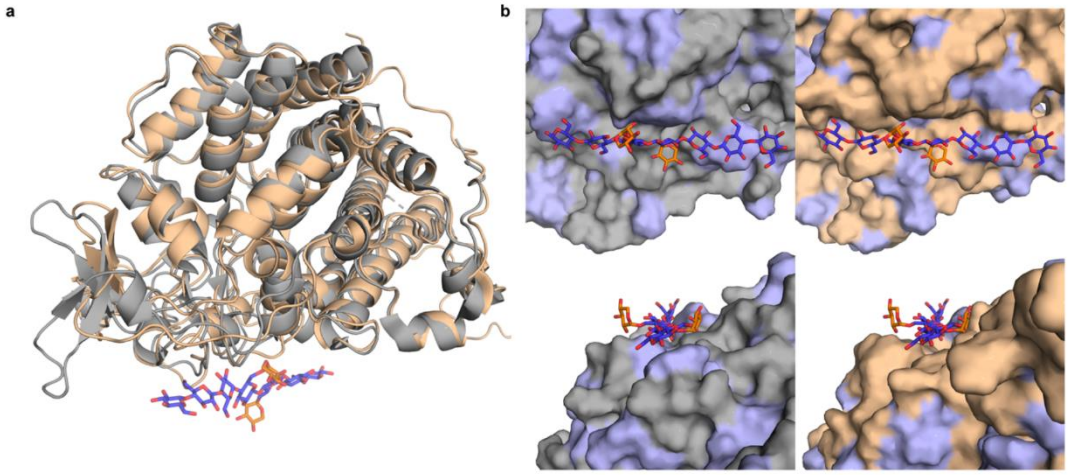
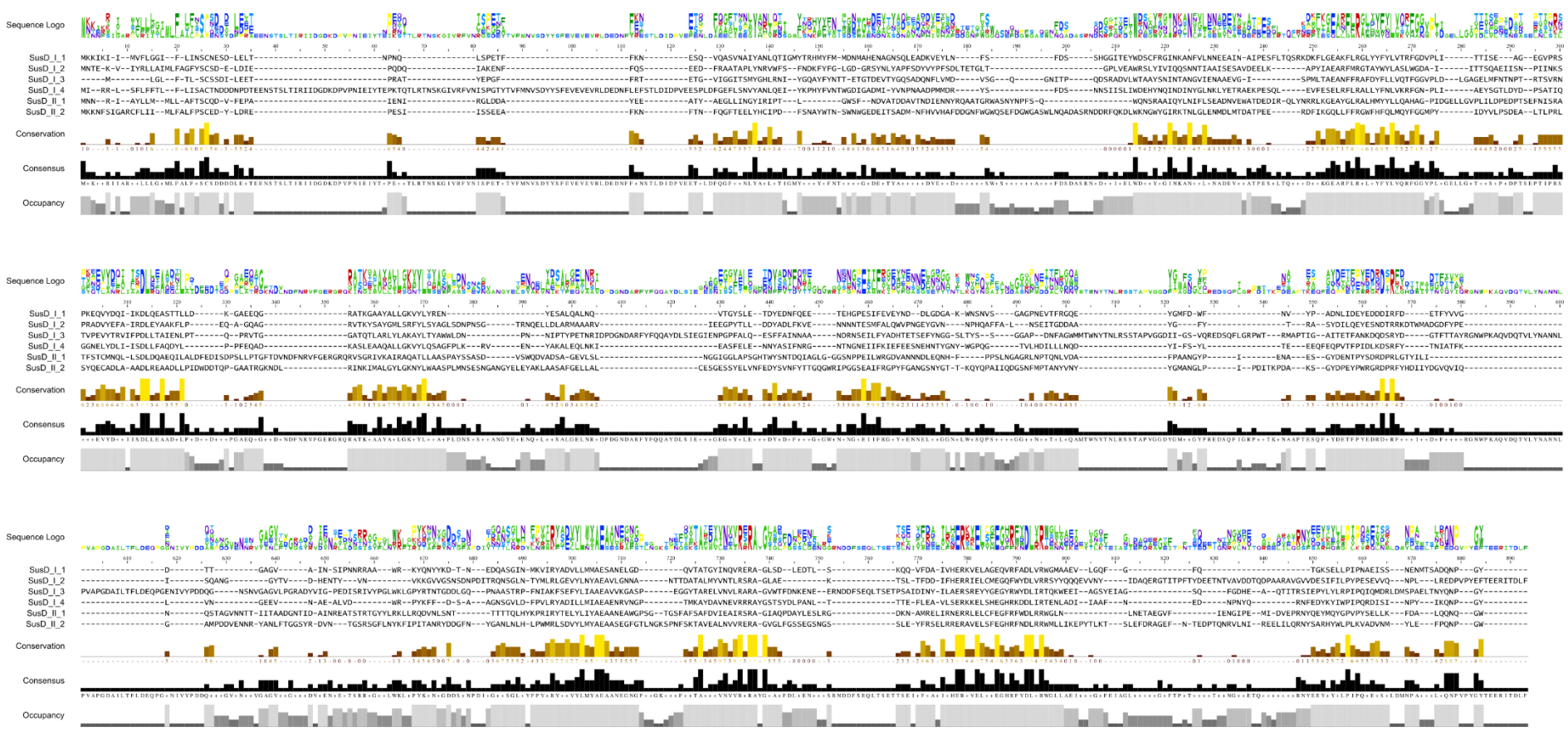



**Fig. S5.4. SusD_I_2 binding site is similar to characterized homologs.** (**a**) Overlay of an AlphaFold2 model of *Flavimarina* SusD_I_2 (grey) with its closest characterized homolog Bacova_02651 (5E76, light brown) [19], that was shown to bind xyloglucan. (**b**) Front (top) and side (bottom) views of binding sites of *Flavimarina* SusD_I_2 (grey) and Bacova_02651 (light brown), both shown with the latter's xylogluco-oligosaccharide ligand (glucose blue, xylose orange sticks). Position of the ligand within the SusD_I_2 model was generated by structural alignment. Amino acids with hydrophobic side chains (F, L, M, P, V, W) are marked light blue to underline similarities within the binding cleft.

next page

**Fig. S5.5. Multiple sequence alignment of the six SusDs from *Flavimarina* sp. xylan PULs I and II.** Shown is conservation of all individual amino acids (low/brown to high/yellow), consensus (black) as well as occupancy (grey) of all positions in the alignment. A logo graph details the conservation of one or more amino acids at a specific position (top). Gaps in alignment are marked by dashes.


Sequence Logo
SusD_I_1
SusD_I_2
SusD_I_3
SusD_I_4
SusD_II_1
SusD_II_2
Conservation
Consensus
Occupancy

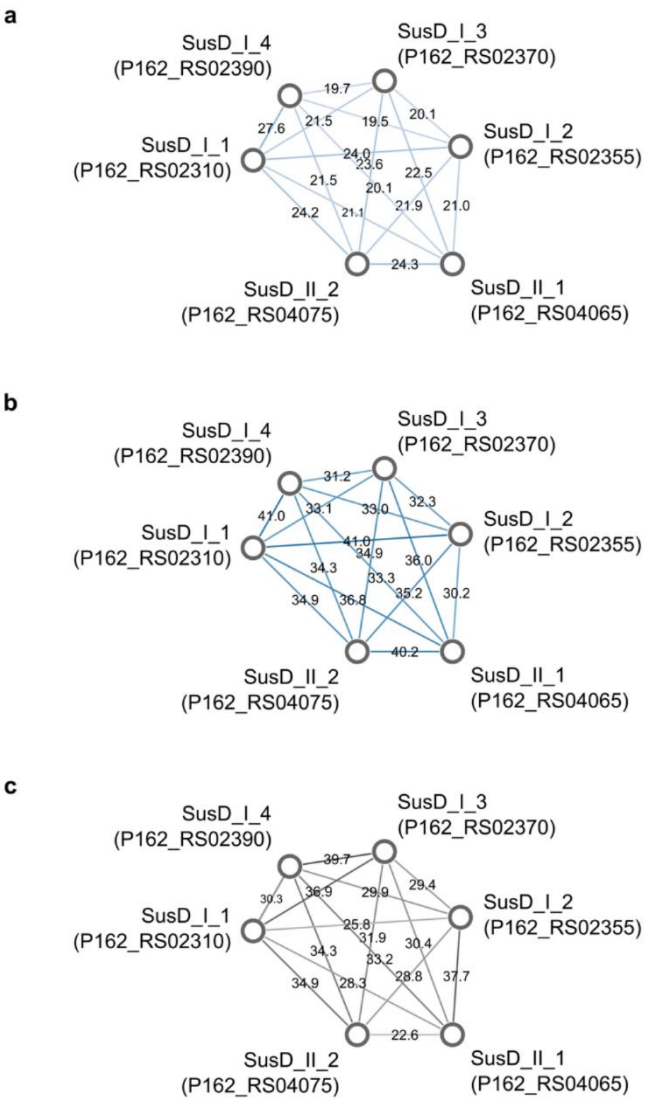



**Fig. S5.6. Sequence similarity analysis by All-by-all global pairwise alignment** using EMBOSS Needle of the six SusD amino acid sequences from the Flavimarina sp. xylan PULs I and II. Pairwise identity (a), sequence similarity (b) and percentage of gaps (c) were plotted.

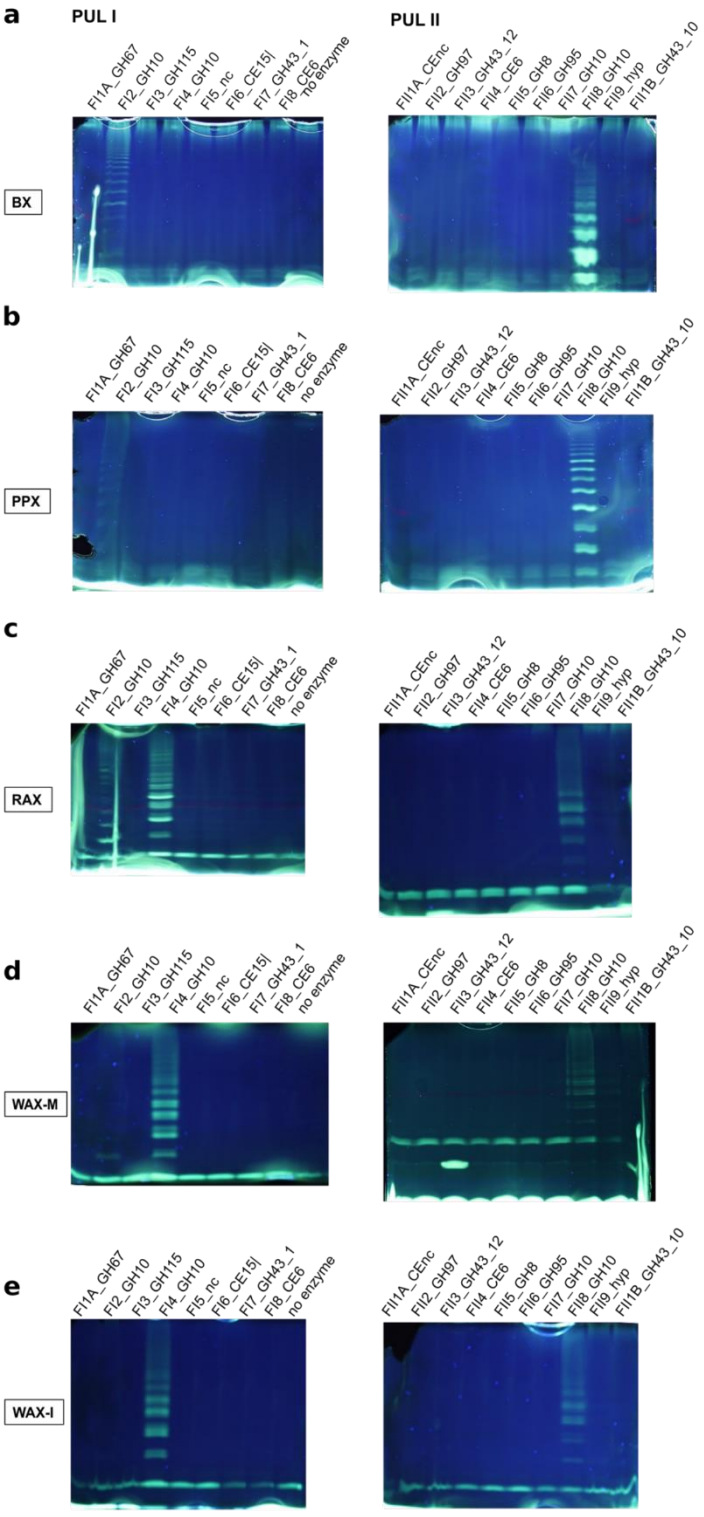



**Fig. S5.7. ANTS-FACE from the initial degradation of different polymeric xylan substrates.** Biocatalysis of the recombinantly expressed flavobacterial CAZymes (20 µg mL-1) with 10 mg mL-1 BX (**a**), PPX (**b**), RAX (**c**), WAX-M (**d**) and WAX-I (**e**). Small degradation products of the polymeric substrate can be fluorescently labelled with the

ANTS dye and separated in an electric field. The ANTS-FACE gel with CPX is not shown as no degradation products

were visible and CPX was excluded from further analysis. The biocatalysis of FI4_GH10 with BX and PPX also results in a ladder pattern which is seen in Figure 3. A positive reducing end assay is shown Figure S8.

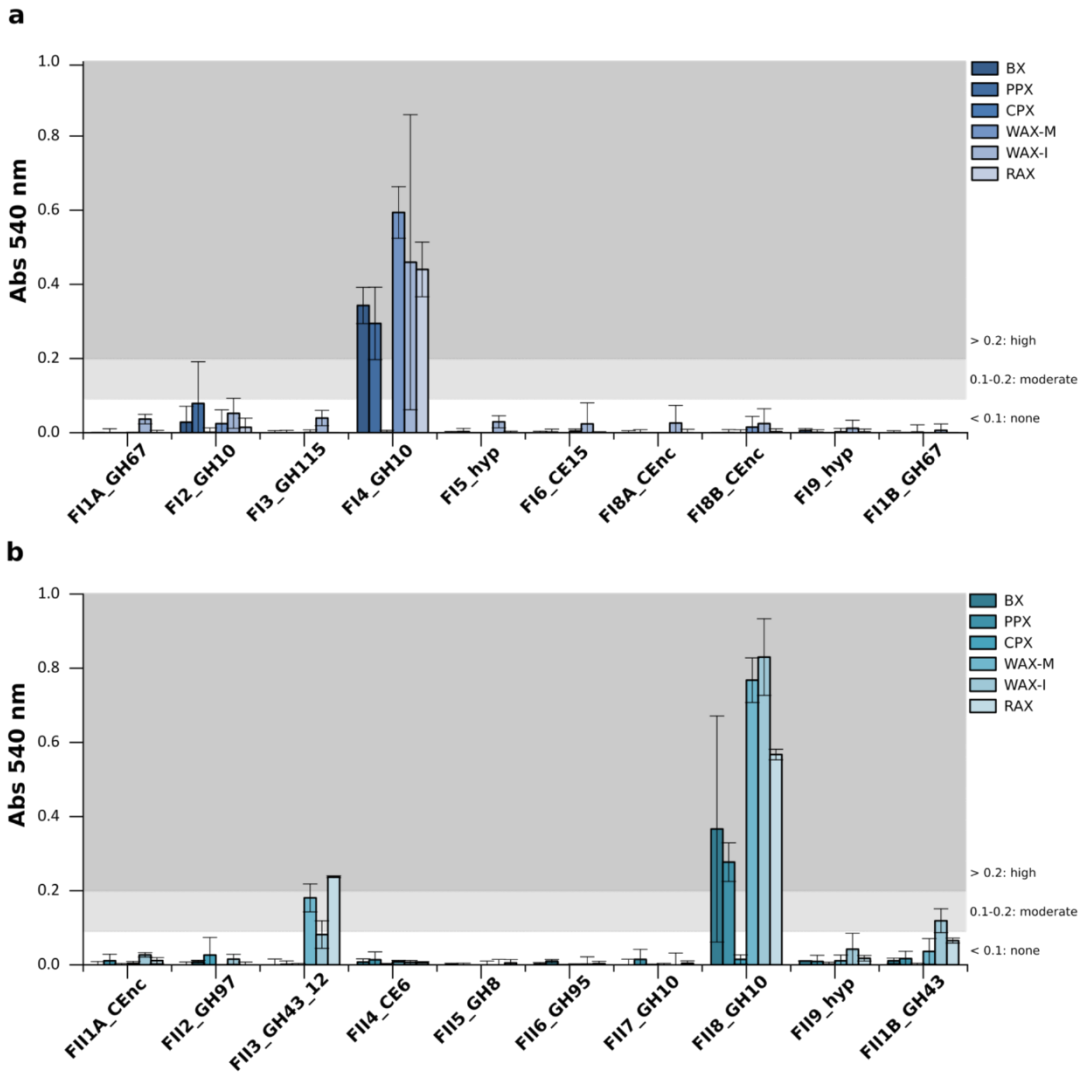



**Fig. S5.8. DNS-reducing end assay screening of the *Flavimarina* sp. PUL proteins.** (**a**) PUL I and (**b**) PUL II enzymes (15 µg mL-1) were incubated overnight with a 1% xylan solution. The reducing ends were measured via DNS-assay [165]. The increase of the absorption at 540 nm indicates the increase of reducing ends and thus degradation of the carbohydrate. The absorbance of the negative control (without enzyme) was subtracted. Enzyme activity was designated as moderate when the average absorbance (after subtracting the negative control without enzyme) was between 0.1 and 0.2 and as high for values > 0.2.

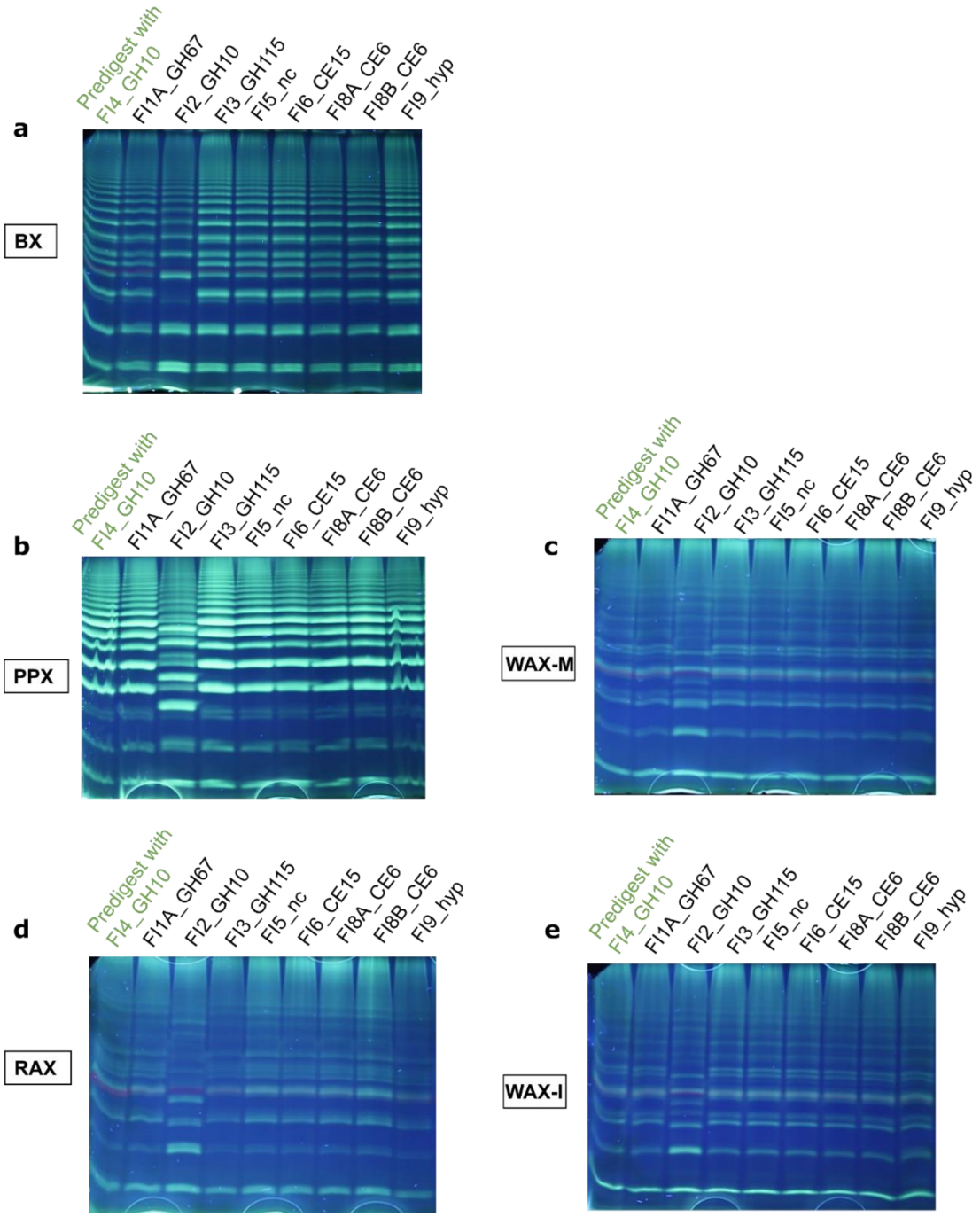



**Fig. S9. ANTS-FACE analysis of samples from the biocatalytic reactions utilizing different xylan substrates pre-digested with the PUL I-encoded FI4_GH10 with further PUL I CAZymes.** BX (**a**), PPX (**b**), RAX (**c**), WAXM (**d**) WAX-I (**e**) were pre-digested with FI4_GH10 and combined with further enzymes of the PUL I after heat inactivation of FI4_GH10. A shift in the gel pattern indicates further activity.

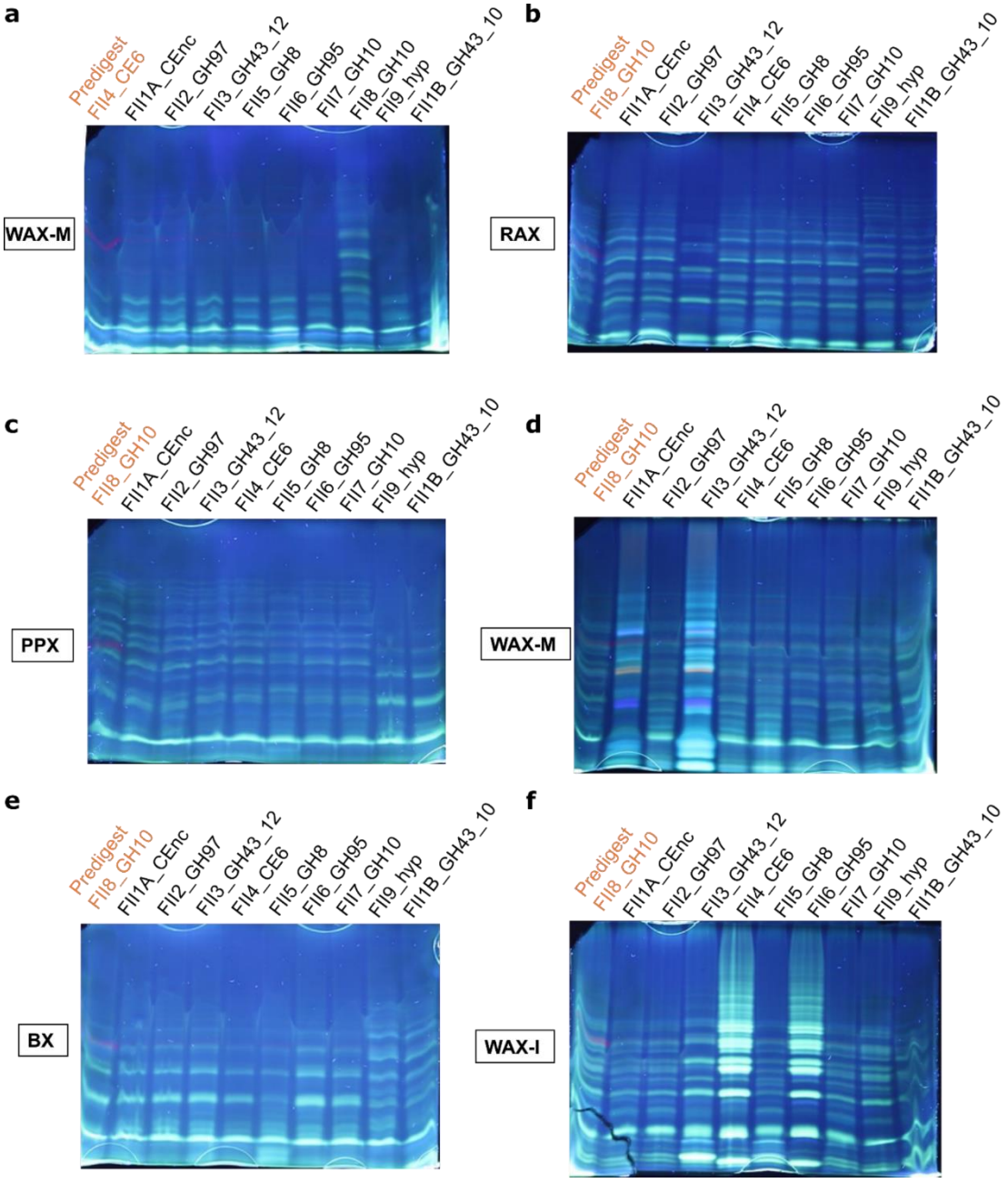



**Fig. S5.10. ANTS-FACE analysis of the xylan pre-digested with the PUL II enzyme FII8_GH10.** Wheat arabinoxylan medium viscosity (WAX-M) was predigested with FII4_CE6 and combined with further PULII enzymes after heat inactivation (**a**). RAX (**b**), PPX (**c**) WAX-M (**d**) BX (**e**) and WAX-I (**f**) were pre-digested with FII8_GH10 and combined with further enzymes of PUL II after heat inactivation of FI4_GH10. A shift in the gel pattern indicates further enzyme activity.

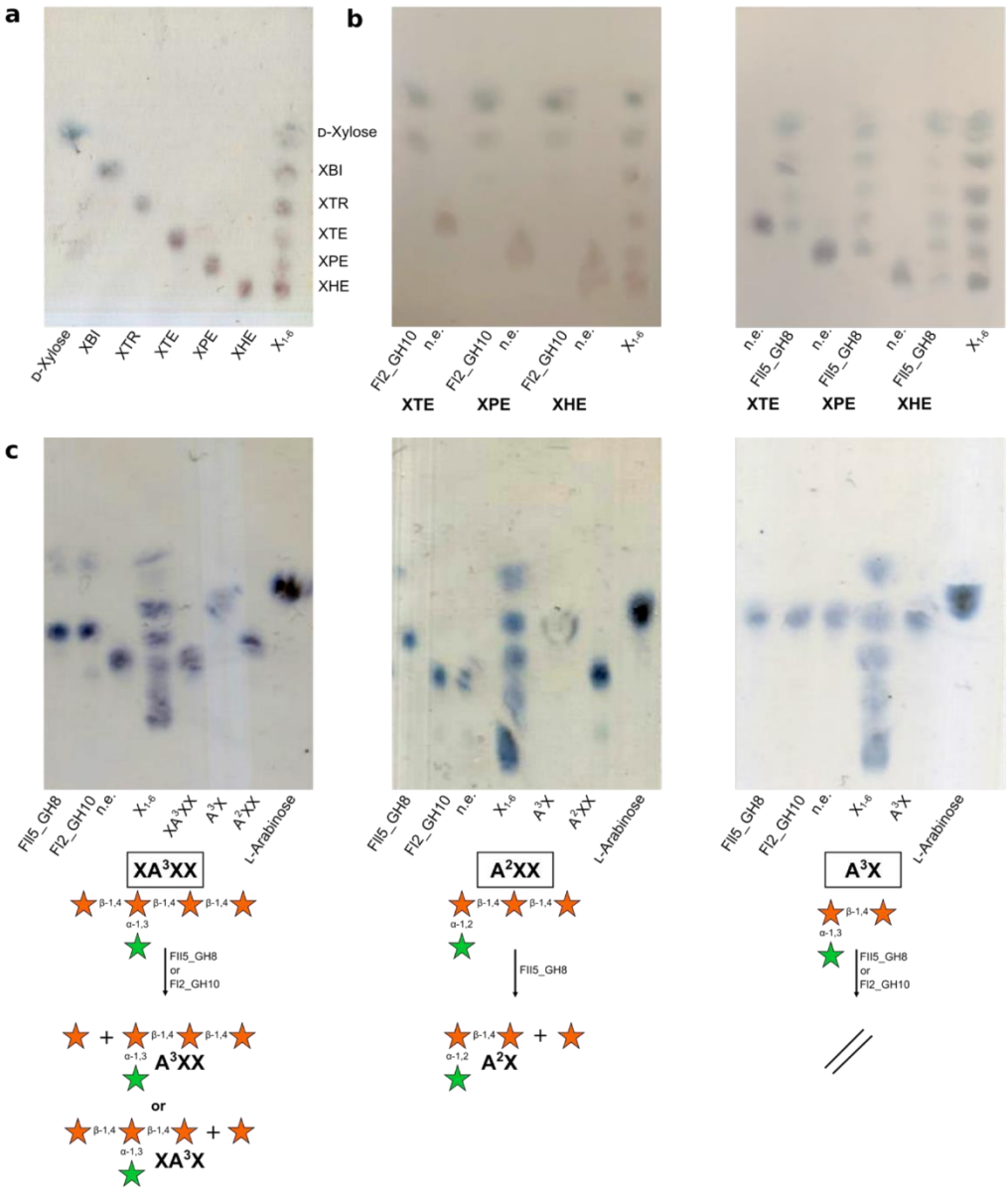



**Fig. S5.11. TLC analysis of (arabino-) xylo-oligosaccharides digested with Fll5_GH8 and Fl2_GH10.** Purchased xylo-oligosaccharides were mixed to generate the standard $X_{1-6}$ containing D-xylose, xylobiose (XBI), xylotriose (XTR, xylotetraose (XTE), xylopentaose (XPE) and xylohexaose (XHE) (**a**). The standard was used to identify generated oligos after incubation of xylotetraose, xylopentaose and xylohexaose with Fl2_GH10 and Fll5_GH8 (**b**) where only Fl2 indicates a degradation to mono- and disaccharide level. Acceptance towards arabinose modification was tested by incubating Fll5_GH8 and Fl2_GH10 with $3^3$-α-L-Arabinofuranosyl-xylotetraose ($XA^3XX$), $2^3$-α-LArabinofuranosyl-xylotriose ($A^2XX$) and $3^2$-α-L-Arabinofuranosyl-xylobiose ($A^3X$), leading to the here visualized assumed reaction schemes (**c**). Due to a missing standard for the degradation products of $XA^3XX$, it can only be assumed that the potentially released $A^3XX$ would show similar Rf than $A^2XX$, while the

released product is not running parallel to the $A^2XX$ standard. This allows no clear assignment of the released products, resulting in the two illustrated potentially possible reaction schemes.

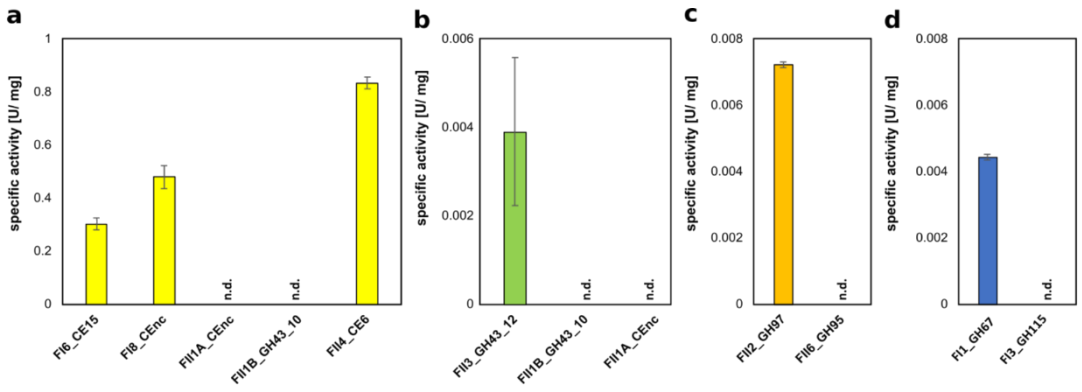



**Fig. S5.12. Screening enzymatic activity with chromogenic substrates (pNP).** Autohydrolysis was deducted from each value. An increase of absorption corresponds to the hydrolytic activity of the enzyme. (**a**) Activity of the carbohydrate esterases towards the artificial substrate pNP-acetate. (**b**) Arabinase activity towards the pNP-α-Larabinose substrate. (**c**) Activity of putative galactosidases towards pNP-α-D-galactoside and (**d**) activity towards pNP-α-D-glucuronide.

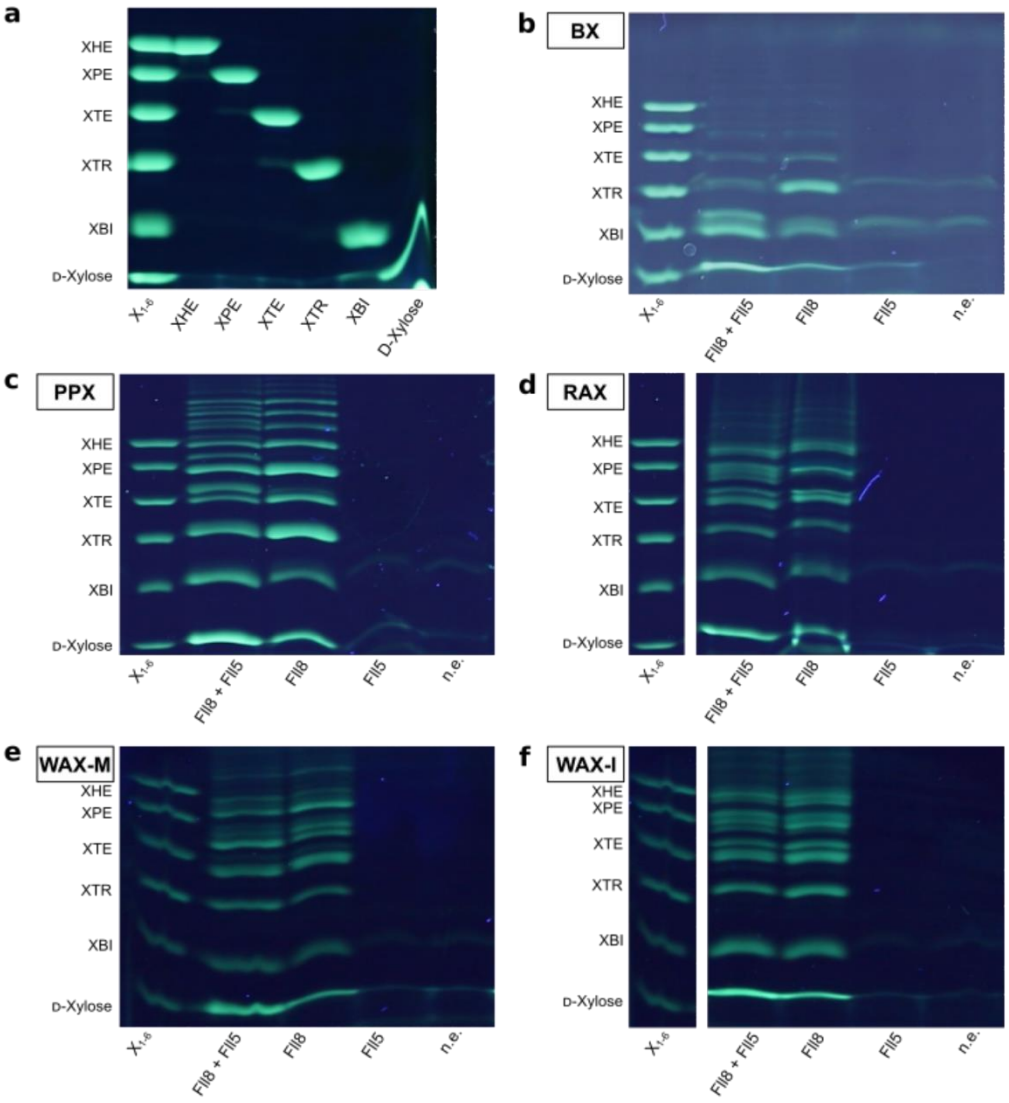



**Fig. S5.13. ANTS-FACE analysis of the Fll5 activity towards Fll8_GH10 hydrolysates of different model xylans.** As a standard xylo-oligosaccharides were combined and used as a xylo-oligosaccharide ladder ($X_{1-6}$, **a**). The xylans were digested with Fll8_GH10 and/ or Fll5_GH8 while the negative control contained no enzyme (n.e.). For BX (**b**) and PPX (**c**) a shift in intensities and slightly altered band pattern could be observed upon digestion by both enzymes compared to only Fl8_GH10. Also, for RAX (**d**) a slight shift occurred while no significant intensity differences were detected. WAX-M (**e**) and WAX-I (**f**) showed no difference upon further digestion with Fll5_GH8. No degradation was observed for incubation with Fll5_GH8 alone.

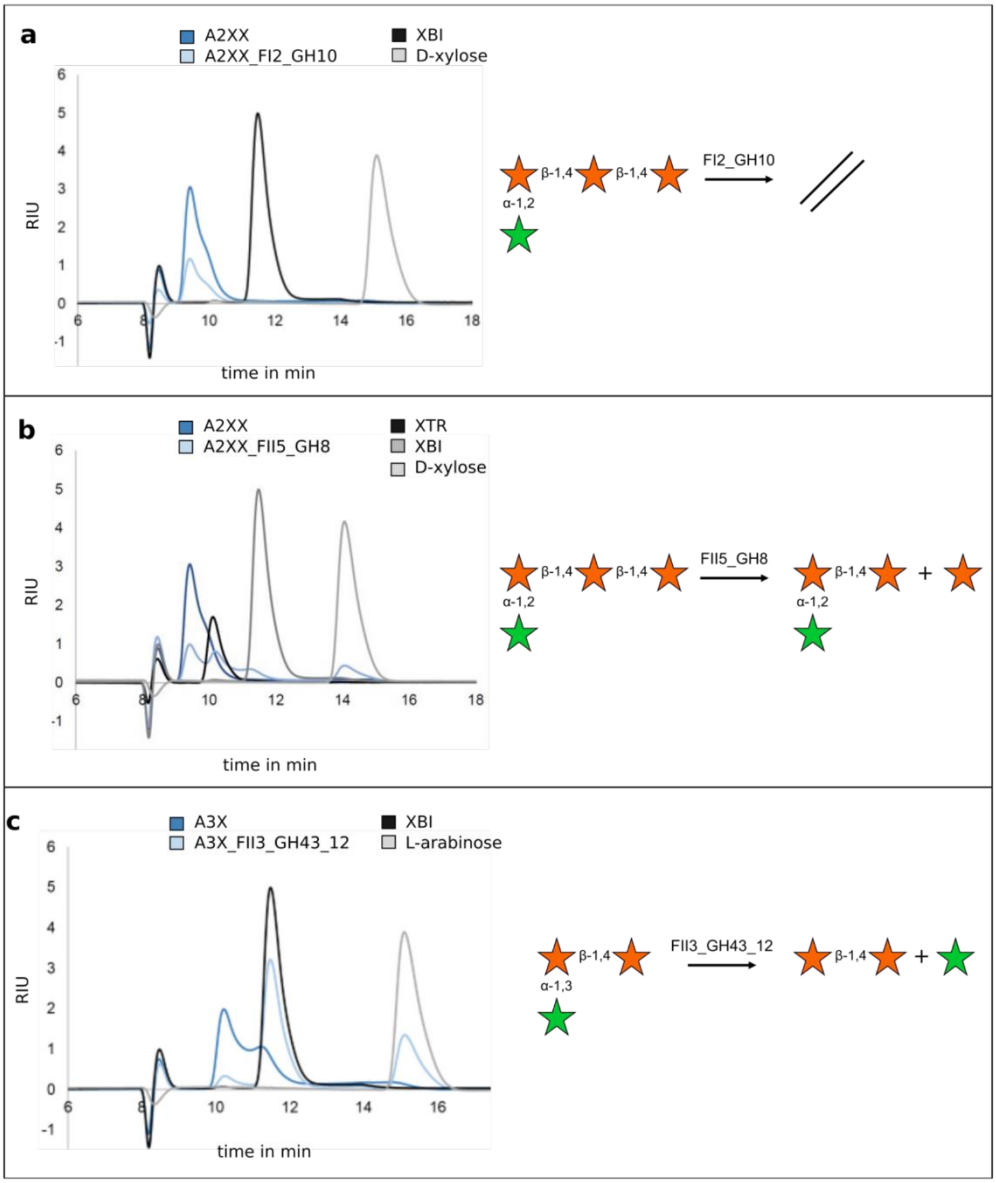



**Fig S5.14. HPLC analysis of arabinoxylo-oligosaccharides digested with FI2_GH10, FII5_GH8 and FII3_GH43_12.** The reactions with 1 mM standard substance purchased from Megazymes (Wicklow, Irland) were analysed after 20 µM enzyme was added and incubation was performed overnight at room temperature. The sugars were measured via HPLC-RI (Hitachi Chrommaster 5310 column oven, Hitachi Chromaster 5450 RI detector; SugarSep-H 10 µm 300 x 8 mm) while the formed products were compared with the single standard solutions. (**a**) The xylanase activity of FI2_GH10 was verified via the biocatalysis with XTR releasing xylobiose (XBI) and D-xylose (see also Fig. 3). Furthermore, the arabinose is blocking the xylanase cleaving of the D-xylose from the non-reducing end, which is the reason for the missing release of D-xylose from $2^3$-α-L-arabinofuranosyl-

xylotriose ($A^2XX$). (**b**) The arabinofuranosidase activity and xylosidase activity of FII5_GH8 was verified by the biocatalysis of xylotriose (XTR) (see also Fig. 3) and $A^2XX$ and the formation of L-arabinose. (**c**) Arabinofuranosidase activity of FII3_GH43_12 was verified via biocatalysis with $A^2XX$ (see also Fig. 3) and ($A^3X$) $3^2$-α-L-arabinofuranosyl-xylobiose releasing L-arabinose.

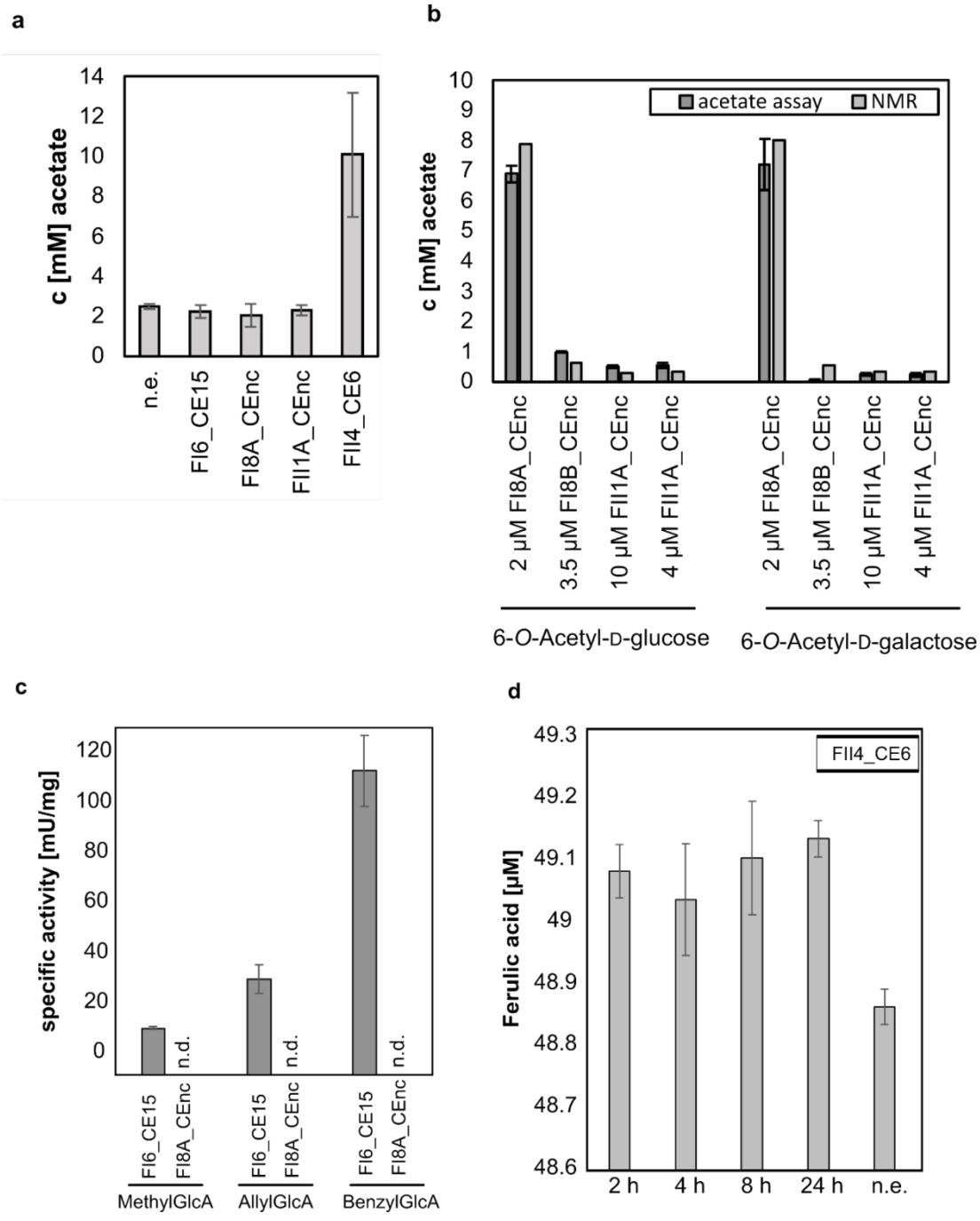



**Fig. S5.15. Analysis of the carbohydrate esterase activities.** (**a**) Release of acetate from polymeric partially acetylated birchwood xylan (Megazymes) measured with the R-

Biopharm acetate assay, no release of acetate was measurable with BX, PPX, WAX-M, WAX-I and RAX, and therefore not shown. (**b**) Release of acetate from monomeric 6-*O*-acetyl-D-glucose and 6-*O*-acetyl-D-galactose measured with the Acetate-Kit (Megazymes) with additional proof via NMR quantification. (**c**) Measurement of D-glucuronic acid hydrolysis with the esterases FI6_CE15 and FI8_CE6 with methyl-D-glucuronic acid (MethylGlcA), allyl-D-glucuronic acid (AllylGlcA) and benzyl- D-glucuronic acid (BenzylGlcA) with the K-URONIC Assay Kit (Megazymes). (**d**) Release of ferulic acid by FII4_CE6 WAX-I measured and quantified via U-HPLC DAD 320 nm.

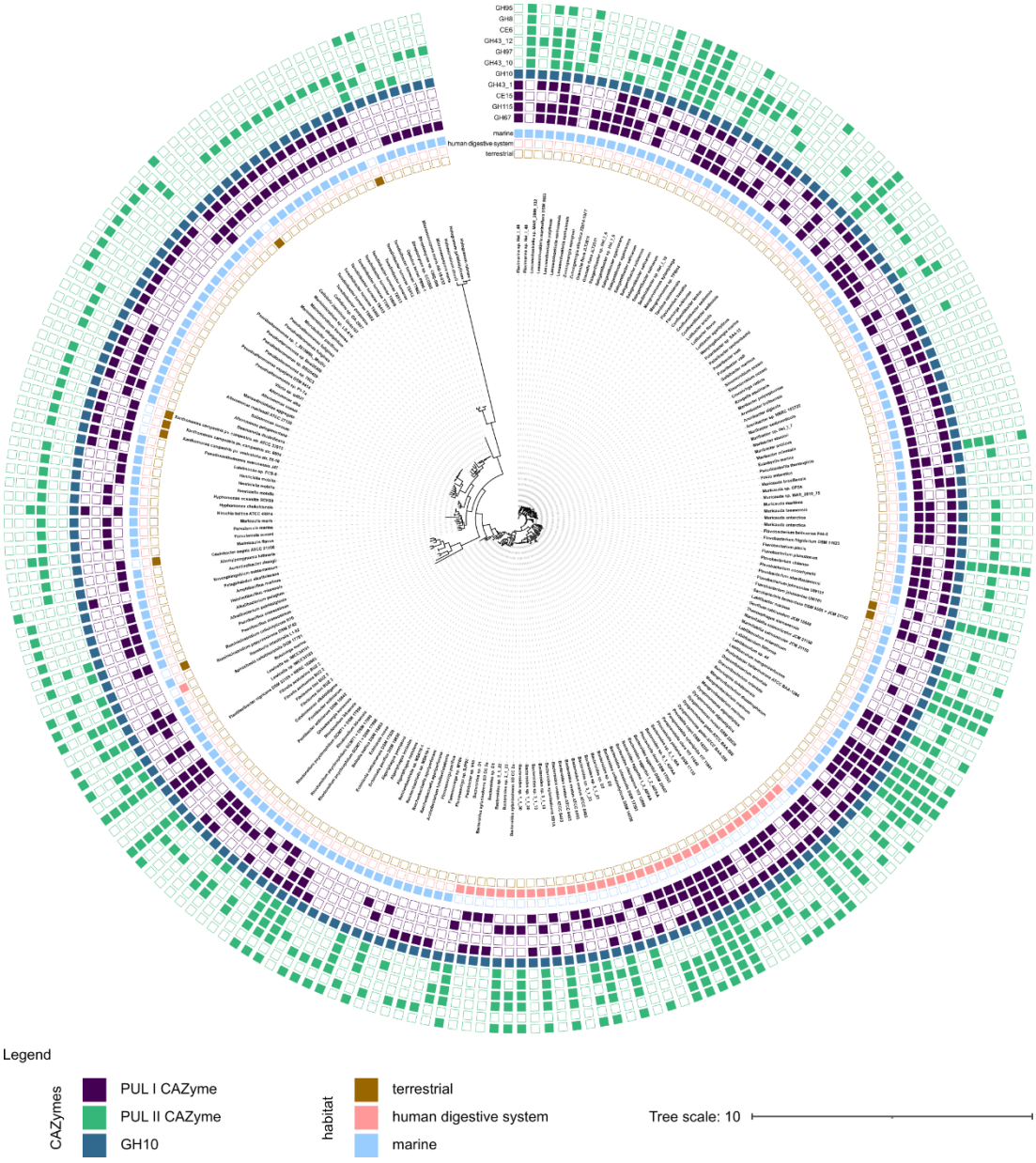


**Fig. S5.16. Taxonomy and xylan PUL modularity of 226 bacteria with PULs similar to *Flavimarina* sp.** The phylogenetic tree is based on *rpoB* sequences. The inner rings represent the habitat from which the individual bacteria were initially isolated. Expanded from Figure 4, modularity of all identified clusters sharing at least a GH10 (blue) and two additional enzymes with the *Flavimarina* sp. PULs (PUL I purple, PUL II green) is depicted on the outer rings.

### *5.9.4. Supplementary References*


1. Mystkowska AA, Robb C, Vidal-Melgosa S, Vanni C, Fernandez-Guerra A, Höhne M et al. Molecular recognition of the beta-glucans laminarin and pustulan by a SusD-like glycan-binding protein of a marine Bacteroidetes. FEBS J. 2018; 285:4465–4481.
2. Zhang Y, Werling U, Edelmann W. SLiCE: a novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res. 2012; 40:e55.
3. Butt TR, Edavettal SC, Hall JP, Mattern MR. SUMO fusion technology for difficult-to-express proteins. Protein Expr Purif. 2005; 43:1–9.
4. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker J M. The Proteomics Protocols Handbook. Humana Press, Totowa, NJ, 2005. pp 571–607.
5. Engel A, Händel N. A novel protocol for determining the concentration and composition of sugars in particulate and in high molecular weight dissolved organic matter (HMW-DOM) in seawater. Marine Chemistry. 2011; 127:180–191.
6. Bowers, G N, Jr, McComb RB, Upretti A. 4-nitrophenyl phosphate--characterization of high-purity materials for measuring alkaline phosphatase activity in human serum. Clin Chem. 1981; 27:135–143.
7. Cubillos-Rojas M, Amair-Pinedo F, Tato I, Bartrons R, Ventura F, Rosa JL. Tris-Acetate Polyacrylamide Gradient Gels for the Simultaneous Electrophoretic Analysis of Proteins of Very High and Low Molecular Mass. Methods Mol Biol. 2019; 1855:269–277.
8. Sunner H, Charavgi M-D, Olsson L, Topakas E, Christakopoulos P. Glucuronoyl Esterase Screening and Characterization Assays Utilizing Commercially Available Benzyl Glucuronic Acid Ester. Molecules. 2015; 20:17807–17817.
9. Duff RB. 954. Esterification of the primary alcoholic groups of carbohydrates with acetic acid: a general reaction. Journal of the Chemical Society (Resumed). 1957:4730–4734.

10. Troitzsch A, van Loi V, Methling K, Zühlke D, Lalk M, Riedel K et al. Carbon Source-Dependent Reprogramming of Anaerobic Metabolism in Staphylococcus aureus. J Bacteriol. 2021; 203
11. Wider G, Dreier L. Measuring Protein Concentrations by NMR Spectroscopy. Journal of the American Chemical Society. 2006; 128:2571–2576.
12. Phakeenuya V, Ratanakhanokchai K, Kosugi A, Tachaapaikoon C. A novel multifunctional GH9 enzyme from Paenibacillus curdlanolyticus B-6 exhibiting endo/exo functions of cellulase, mannanase and xylanase activities. Applied Microbiology and Biotechnology. 2020; 104:2079–2096.
13. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022; 50:W276-W279.
14. Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics. 2007; 23:1073–1079.
15. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 2000; 302:205–217.
16. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009; 25:1189–1191.
17. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–2504.
18. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596:583–589.
19. Tauzin AS, Kwiatkowski KJ, Orlovsky NI, Smith CJ, Creagh AL, Haynes CA et al. Molecular Dissection of Xyloglucan Recognition in a Prominent Human Gut Symbiont. mBio. 2016; 7:e02134-15.
20. Schrödinger LL. The PyMOL Molecular Graphics System, Version 1.8. 2015
21. Bernfeld P. [17] Amylases, α and β. In: Methods in Enzymology. Academic Press, 1955; 1. pp 149–158.

# List of publications

**Beidler I**, Steinke N, Schulze T, Sidhu C, Bartosik C, Krull J, Dutschei T, Ferrero-Bodera B, Rielicke J, Kale V, Sura T, Trautwein-Schult A, Kirstein IV, Wiltshire KH, Teeling H, Becher D, Bengtsson MM, Hehemann J-H, Bornscheuer UT, Amann RI, Schweder T. 3. Heterotrophic marine bacteria employ a laminarin-fueled intra-population α-glucan recycling loop. Under review in *Nature Communications*, July 2023. Preprint: https://doi.org/10.21203/rs.3.rs-3205445/v1.

**Beidler I**, Robb CS, Vidal-Melgosa S, Zühlke M-K, Bartosik D, Solanki V, Markert S, Becher D, Schweder T, Hehemann J-H. Marine bacteroidetes use a conserved enzymatic cascade to digest diatom β-mannan. *ISME J* 17, 276–285 (2023). https://doi.org/10.1038/s41396-022-01342-4

Dutschei T, **Beidler I**, Bartosik D, Seeßelberg J-M, Teune M, Bäumgen M, Querido Ferreira S, Heldmann J, Nagel F, Krull J, Berndt L, Methling K, Hein M, Becher D, Langer P, Delcea M, Lalk M, Lammers M, Höhne M, Hehemann J-H, Schweder T, Borscheuer UT. Marine Bacteroidetes enzymatically digest xylans from terrestrial plants. *Environmental Microbiology*, 1– 15. Published online May 2023. https://doi.org/10.1111/1462-2920.16390

# Acknowledgements

First and foremost, I would like to thank Thomas Schweder for the opportunity to work in the amazing world of marine sugars for the last few years and for helping me develop my skills as a scientist and a critical thinker.

I would also like to thank the German Research Foundation (DFG) for funding POMPU FOR2406.

I thank the other members of my thesis committee, Uwe Bornscheuer and Jan-Hendrik Hehemann for many fruitful discussions and helping to guide my research towards the ever next interesting question. I would also like to thank Rudolf Amann for his support of my research stays at the Max-Planck Institute in Bremen and for helping me put my work into the right perspectives.

I'd like to thank all my colleagues and co-authors, both in the AG Schweder and throughout all of the POMPU consortium, especially Marie for teaching me the basics of proteomics and how to handle marine bacteria, Alex for our FACE-Gels and protein expression talks as well as being a great office-mate and Daniel for every "I need another database search", every astonishing figure and every coffee he made for me.

Throughout POMPU, I especially thank Theresa, Anke, Mia, Chandni and Joris for help with many smaller and bigger data sets and samples. None of the incredible conclusions we came to would have been possible with even one part missing.

Finally, I would like to thank my friends Vincent, Laura, Stephie and Betti. We started all of this together ten years ago and I'm proud to summarize that none of us actually dropped out of our studies (we were told that was the likely outcome for about half of us).