Genomic Analyses of Polysaccharide Utilization in Marine *Flavobacteriia*

Dissertation

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

- **Dr. rer. nat**. -

dem Fachbereich 2 Biologie/Chemie der

Universität Bremen

vorgelegt von

Lennart Kappelmann

Bremen, April 2018

Die vorliegende Doktorarbeit wurde im Rahmen des Programms **International Max Planck Research School of Marine Microbiology**“ (MarMic) in der Zeit von August 2013 bis April 2018 am **Max-Planck-Institut für Marine Mikrobiologie** angerfertigt.

This thesis was prepared under the framework of the **International Max Planck Research School of Marine Microbiology** (MarMic) at the **Max Planck Institute for Marine Microbiology** from August 2013 to April 2018.

Gutachter: Prof. Dr. Rudolf Amann

Gutachter: Prof. Dr. Jens Harder

Prüfer: Dr. Jan-Hendrik Hehemann

Prüfer: Prof. Dr. Rita Groß-Hardt

Tag des Promotionskolloquiums: 16.05.2018

Inhaltsverzeichnis

Summary

In marine surface waters, the majority of carbon fixation, which accounts for half of the global carbon fixation, is carried out by microalgae at high turnover rates. In temperate coastal shelf seas like the North Sea, large amounts of carbon dioxide are fixed during spring phytoplankton blooms, which are often dominated by diatoms. When these blooms collapse, heterotrophic bacteria utilize the algal-derived organic matter that consists in large parts of polysaccharides. Members of the *Flavobacteriia* are polysaccharide degradation specialists and dominate the bacterioplankton during these blooms. They possess polysaccharide utilization loci (PULs), gene arrangements encoding proteins for binding, uptake and breakdown of polysaccharides. The carbohydrate-active enzymes (CAZymes) encoded within these PULs cleave specific polysaccharides. PUL analysis hence allows addressing the carbohydrate degradation capacities of the bacterium and the composition of the targeted marine polysaccharide pool. Although many of these polysaccharides likely are important components of the marine carbon cycle, their structures are still poorly understood.

The main objectives of this doctoral thesis were thus to gain new insights into the polysaccharide degradation capacities of *Flavobacteriia* and, indirectly, the composition of marine polysaccharides. In chapter II, I investigated in a culture independent approach how two contrasting North Atlantic surface water sites characterized by different polysaccharide sources select for specifically adapted *Bacteroidetes,* mostly *Flavobacteriia*. In chapter III, the taxonomic and functional recurrence of bacterioplankton in four consecutive spring blooms was investigated. We examined which bacterial clades annually reappeared and which associated CAZymes were found in high gene frequencies to successfully degrade the algae-derived organic matter. Finally, in chapter IV, I investigated the PULs of 53 new flavobacterial isolates from the North Sea, many of them

isolated from surface waters during the phytoplankton blooms. I found a wide range of PUL repertoires, suggesting substrate partitioning among *Flavobacteriia* likely providing ecological niche space. Many of the predicted PULs were new and some of them likely target diatom-derived abundant substrates. In order to assess the relevance of these predicted substrates, I analyzed the expression levels of PUL-associated, substrate-specific *susCD*-like genes for substrate binding and transport and thereby provide first hints that these could be used as environmental proxies to determine the utilization of the targeted polysaccharides. Detailed functional characterizations of PULs selected on the basis of this thesis have the potential to advance our knowledge on the marine carbon cycle.

Zusammenfassung

In den Oberflächengewässern der Meere wird der Großteil der marinen Kohlenstoff-Fixierung, welcher rund die Hälfte der globalen Kohlenstoff-Fixierung ausmacht, von Mikroalgen mit hohen Umsatzraten durchgeführt. In Küstengewässern der gemäßigten Zonen wie der Nordsee werden zudem große Mengen Kohlendioxid während sogenannter Frühjahrsphytoplanktonblüten fixiert, welche häufig von Kieselalgen dominiert werden. Wenn diese Blüten einbrechen, konsumieren heterotrophe Bakterien die organische Substanz der Algen, welche zu großen Teilen aus Polysacchariden besteht. Mitglieder der Bakterienklasse der *Flavobacteriia* sind besonders gut an den Abbau der Polysaccharide der Algen angepasst und dominieren daher den Bakterioplankton während dieser Blüten. Sie besitzen sogenannte „Polysaccharide Utilization Loci" (PULs), genetische Bereiche, welche Proteine zum Binden, Zerkleinern und Aufnehmen dieser Polysaccharide kodieren. Die in solchen PULs kodierten „Carbohydrate-active Enzymes" (CAZymes) spalten spezifische Polysaccharide. Die Analyse von PULs ermöglicht daher Schlussfolgerungen auf das Potential von Bakterien zum Kohlenhydratabbau und indirekt auch auf die Zusammensetzung mariner Polysaccharide. Obwohl viele dieser Polysaccharide für den marinen Kohlenstoffkreislauf wichtig sind, ist über ihre Struktur noch immer verhältnismäßig wenig bekannt.

Das Hauptziel dieser Doktorarbeit war es daher, neue Erkenntnisse über die Kapazitäten der *Flavobacteriia* zum Kohlenhydratabbau und den damit zusammenhängenden Aufbau mariner Polysaccharide zu gewinnen. Für das Kapitel II untersuchte ich dafür mit Kollegen mit einem kultivierungsunabhängigen Verfahren in zwei unterschiedlichen ozeanischen Provinzen des Nordatlantiks, die durch unterschiedliche Polysaccharidquellen gekennzeichnet sind, Genomabschnitte von *Flavobacteriia*. In Kapitel III wurde in Frühjahrsalgenblüten in vier aufeinander folgenden Jahren das Wiederauftreten

bestimmter Taxa und Funktionselemente im Bakterioplankton untersucht. In einer groß angelegten Studie analysierten wir, welche Bakterientypen jährlich wieder auftraten und welche Gene (CAZymes) zum erfolgreichen Abbau der Algenpolysaccharide in hohen Frequenzen gefunden werden konnten. In Kapitel IV habe ich schließlich die PULs von 53 neu isolierten *Flavobacteriia* aus der Nordsee untersucht. Viele dieser Isolate waren dafür aus den Oberflächengewässern während der Algenblüten gewonnen worden. Es zeigte sich eine große Vielfalt an PULs, welche die Aufteilung bestimmter Polysaccharid-Substrate unter den verschiedenen Stämmen nahelegt, was vermutlich zur funktionellen Differenzierung und damit zur Nischenbildung beiträgt. Viele der mittels Funktionsvorhersagen charakterisierten PULs waren zuvor unbekannt und einige von ihnen sind wahrscheinlich für den Abbau von in Kieselalgen häufigen Polysacchariden verantwortlich. Um auf die Relevanz dieser vorhergesagten Substrate zu schließen, habe ich die Expression von PUL-assoziierten, substratspezifischen *susCD*-Genen analysiert, die für das Binden von Polysacchariden und ihren Transport in die Zelle verantwortlich sind. Diese Arbeit liefert erste Hinweise darauf, dass aus den Expressionsdaten dieser Gene auf die Häufigkeit spezifischer Polysaccharide in der Umwelt geschlossen werden kann. Die detaillierte funktionelle Beschreibung der auf Grundlage meiner Daten ausgewählten PULs wird es in der Zukunft erlauben, unser Wissen zum marinen Kohlenstoffkreislauf zu erweitern.

Abbreviations

AA	amino acid
BCP	biological carbon pump
CAZyme	carbohydrate-active enzyme
CBM	carbohydrate-binding module
CE	carbohydrate esterase
COGITO	Coastal Microbe Genomic and Taxonomic Observatory
DIC	dissolved inorganic carbon
DOC	dissolved organic carbon
DOM	dissolved organic matter
EPS	extracellular polymeric substances
FCSP	Fucose-containing sulfated polysaccharide
FISH	fluorescence in situ hybridization
GH	glycoside hydrolase (family)
HMW	high molecular weight
LMW	low molecular weight
MIMAS	Microbial Interactions in Marine Systems
NPP	net primary production
OM	organic matter
PCP	physical carbon pump
POC	particulate organic carbon
PL	polysaccharide lyase
PUL	polysaccharide utilization locus
SGBP	surface glycan binding protein
TBDR	TonB-dependent receptor
TEP	transparent exopolymers

Chapter I: General introduction

1.1 The marine carbon cycle

Earth‘s surface is covered to about 70% by oceans, which compose large ecosystems. The oldest and most abundant organisms within these ecosystems are microorganisms, which contribute substantially to the global biogeochemical cycles, including the largest of them all, the carbon cycle. Life is based on carbon molecules, also called organic matter (OM). The carbon cycle describes the movement of carbon between the terrestrial, atmospheric and oceanic reservoirs. Major carbon pools are known as dissolved inorganic carbon (DIC, the sum of carbon dioxide, carbonic acid, carbonate and bicarbonate), dissolved organic carbon (DOC) and the particulate organic carbon (POC). Carbon turnover directly influences life on Earth and its climate (Falkowski *et al*., 2000).

Behind the huge, but very inactive lithosphere (7.5×10^6 gigatons of carbon), oceans are the world‘s second largest carbon reservoir at 3.8×10^4 gigatons of carbon (Ciais *et al*., 2013). Oceans have shorter carbon turnover times and are in constant exchange with the atmosphere through the physical carbon pump (PCP) and the biological carbon pump (BCP) (Raven and Falkowski, 1999; Sarmiento and Le Quéré 1996; Volk and Hoffert 1985). In the PCP, CO_2 is exchanged according to the difference in partial pressure between the atmosphere and the water (Raven and Falkowski, 1999). CO_2 is dissolved in oceanic water as DIC, making up 97% of oceanic carbon (Ciais et al., 2013). This process prominently occurs in cold, polar water which due to its high density sinks to the ocean basins as part of the thermohaline circulation, also known as the global (ocean) conveyor belt (Broecker, 1997).

In the BCP, carbon sequestration mainly occurs in surface oceanic waters where photosynthetic algal primary producers annually fix an estimated 45-55 gigatons of CO_2 as soluble (DOC) and or insoluble (POC) organic matter (Azam and Malfatti, 2007; Behrenfeld and Falkowski 1997, Field *et al*., 1998; Finkel. 2014; Westberry *et al*., 2008). This OM makes up roughly half of the annual global net primary production (NPP) of an estimated 105 gigatons of carbon (Falkowski *et al.*, 1998; Field *et al.*, 1998; Sarmento and Gasol, 2012). Remineralization in the food web by marine microorganisms turns most of this OM back into CO_2, however, an estimated 0.6-2.4% sinks to the deep ocean as POC (Legendre *et al*., 2015; Polimene *et al.*, 2016; Turner, 2015). A fraction of an estimated 0.5-0.6% is turned into recalcitrant DOC in a process termed the microbial carbon pump (Jiao *et al.*, 2010; Legendre *et al*., 2015). A process that is sometimes overlooked in carbon budgeting is the export of particulate inorganic carbon (PIC) through carbonate-building microorganisms such as coccolithophores, foraminifera, but also by corals and mollusks. Even though this process is only estimated to amount to 1 gigaton per year (Field and Raupach, 2004), it has built up thick carbonate sediments over geological timescales, sometimes making up entire mountains like the cliffs of Dover. Through all these processes, the BCP contributes to keeping surface DIC concentrations low and thus, together with the PCP, creates a vertical gradient causing atmospheric CO_2 to enter the oceans, impacting the global carbon cycle and climate (Ciais *et al*., 2013).

It is important to note that these estimates are difficult to determine, and some authors report higher numbers, for example Field and Raupach (2004) report an annual total of up to 12 gigatons of DIC sinking to the deep sea. One of the reasons is that the amount of carbon fixation by low-light adapted *Prochlorococcus* spp. living in subsurface waters down to about -200 m is hard to quantify, as it cannot be assessed via satellite chlorophyll absorption measurements (Zinser *et al*., 2007).

1.2 Marine microbial life and the marine food web

Microorganisms are the most important players in the marine ecosystem both in overall mass and total abundance at an average of 10^6 cells per milliliter in surface waters (Whitman *et al.*, 1998). They are usually categorized into autotrophs (producers) and heterotrophs (consumers). This distinction is not always clear, as it has become evident that abundant marine "autotrophs" such as *Prochlorococcus* spp. and "heterotrophs" such as *Pelagibacter* spp. can, to a certain degree, feature mixotrophic lifestyles. *Prochlorococcus* cyanobacteria are capable of supplemental uptake of amino acids (Zubkov *et al*., 2003) and glucose (del Carmen Muñoz-Marín *et al*., 2013) and *Pelagibacter* uses proteorhodopsin for supplemental phototrophy (Giovannoni *et al*., 2005).

Photolithoautotrophs use the sun's energy to fix CO_2 into glucose using reducing equivalents from inorganic compounds (e.g. water or hydrogen sulfide). The phytoplankton forms the basis of the food web, including zooplankton and fish (Pomeroy, 1974). By sloppy feeding and defecation, organic matter becomes available to heterotrophic bacteria and archaea in the microbial loop (purple arrow in Figure 1.1, Azam et al., 2007), Phytoplankton also actively exports up to half of the originally fixed organic matter as DOM, which is then almost exclusively available to marine heterotrophic microorganisms who funnel part of the DOM back to the food web (Teira et al., 2001; Thornton, 2014). They turn over the majority (75-95%) of all phytoplankton-derived OM within days to weeks of its production, respiring it back to CO_2 (Cho and Azam 1988; Moran *et al.*, 2016; Piontek *et al.*, 2014). This remineralization also releases other elements from the OM (e.g. nitrogen, phosphorus and iron), influencing their elemental cycles and making them available to higher trophic levels of the marine food web (Moran *et al.*, 2016).

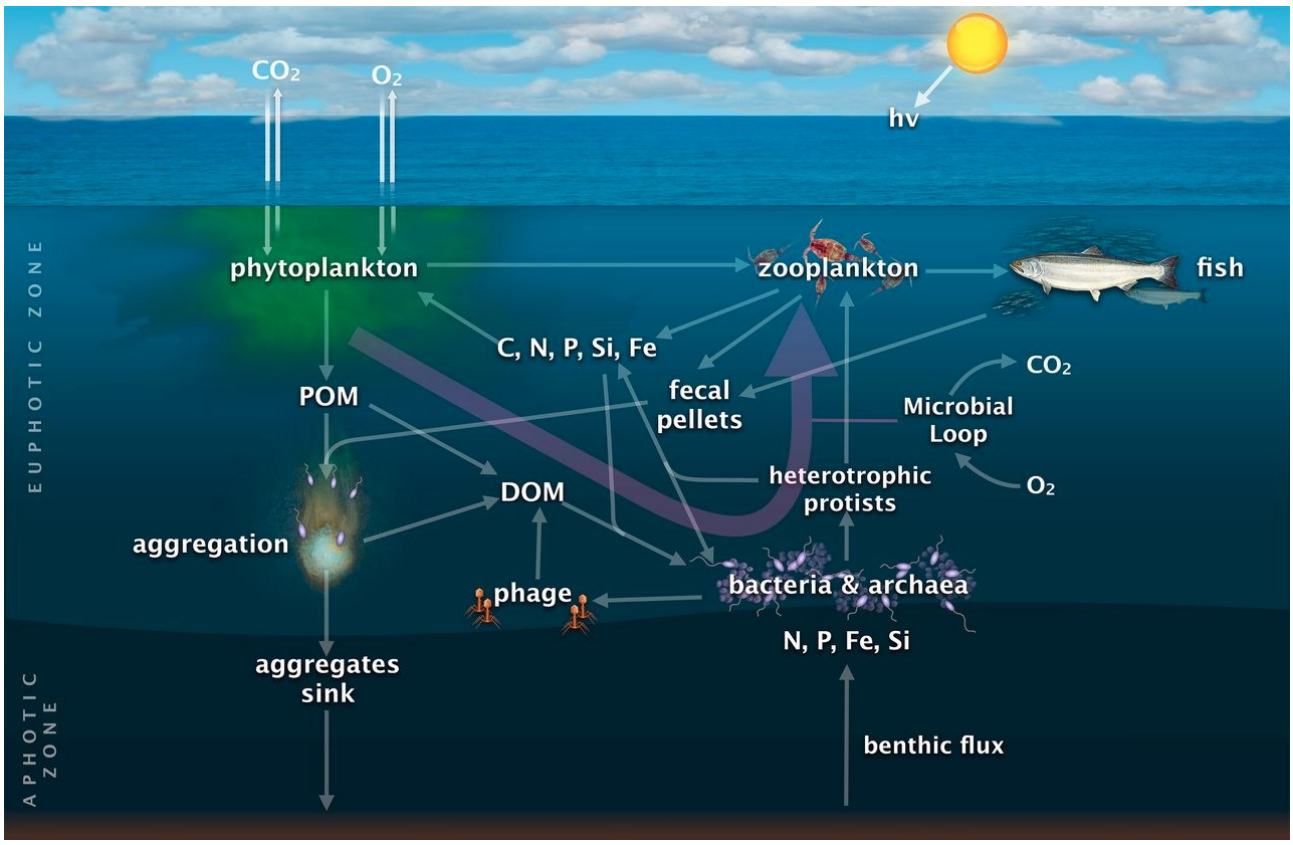


Figure 1.1: Schematic depiction of the marine microbial loop (purple arrow), describing the turnover of dissolved and particulate organic matter (DOM and POM) by marine microbes. They degrade and remineralize this OM, making essential nutrients available again for primary production in the form of CO_2 and DIC and also funnel back DOM to the food web. (Adapted with permission from Worden *et al.*, 2015, who modified from Azam *et al.*, 2007)

1.3 The North Sea as a coastal shelf sea

Photosynthetic primary production is dependent on the availability of sunlight, sufficient temperatures and essential nutrients, prominently nitrogen, iron and phosphorus (for review, see Howarth, 1988, also Geider *et al*., 1997; Marañón *et al*., 2000). While surface open oceans tend to have low essential nutrient concentrations (oligotrophy; Smith, 1984, Raimbault *et al*., 2008) coastal regions, particularly in the higher latitudes of the Northern hemisphere, are rich in essential nutrients due to land-derived input and upwelling of

nutrient-rich bottom waters (Cloern, 1996; Ishizaka *et al*., 1983; Van Dongen-Vogels *et al*., 2012).

The North Sea is such a eutrophic coastal shelf sea. It is localized between the British, Norwegian, Swedish, Danish, German, Dutch and Belgian coastlines. Atlantic water exchange and nutrient-rich discharge of rivers such as Rhine, Elbe, Glomma, Ijssel, Meuse and Weser support strong phytoplankton blooms in spring when solar irradiance and temperatures rise. Although coastal shelf seas only constitute 7.5% of oceanic surfaces (Wollast, 2002), they are thus hotspots for marine carbon fixation into carbohydrates, or (poly-) saccharides, by macro- and microalgae.

1.4 Algal primary production

1.4.1 Macroalgae

Marine macroalgae, colloquially also referred to as seaweed, are a polyphyletic group of multicellular organisms which grow in the intertidal und subtidal zones. They require regular or constant presence of seawater, sufficient light irradiation for photosynthesis and for most species a firm attachment point. Macroalgae are divided into three major subtypes: green, red and brown algae. Green and red algae (green and red boxes (1) and (2) in Figure 1.2), together with land plants, belong to the A*rchaeplastids*, or kingdom Planta *sensu lato* and store energy as starch. Green algae possess the pigments chlorophyll a and b for photosynthesis and store starch inside their chloroplasts. Red algae possess chlorophyll a and phycobiliproteins (which are responsible for the red color), and accumulate starch outside of their chloroplasts. Brown algae (*Phaeophytes*) on the other hand are the sole multicellular organisms within the superphylum of *Stramenopiles* (or heterokonts), which also contains the unicellular diatoms (brown box (3) in Figure 2). They

contain chlorophyll *a* and *c* and in most cases fucoxanthin, giving them their brownish color.

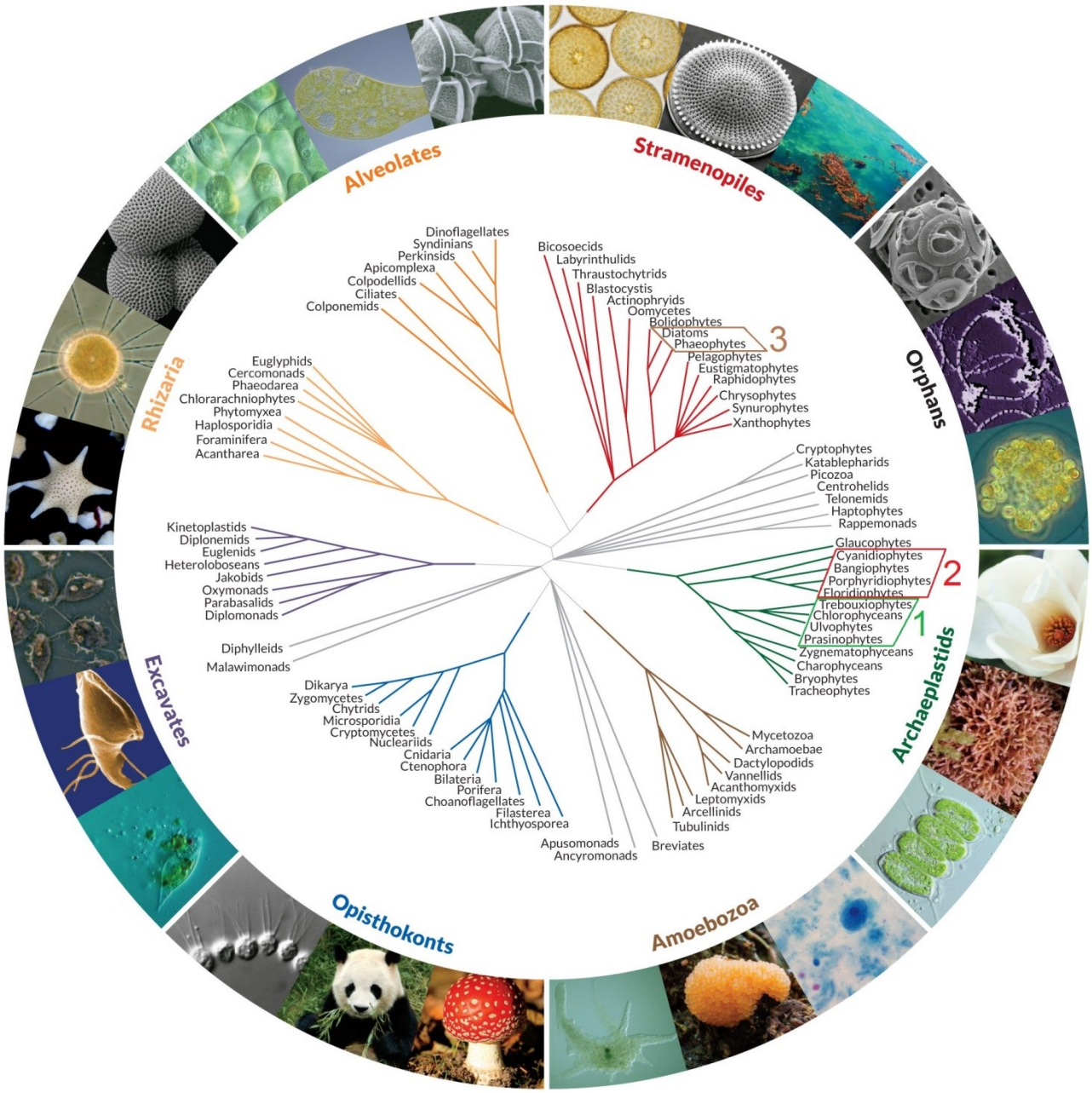


Figure 1.2: Phylogenetic tree of Eukaryotes, indicating the positions of (1) green algae, (2) red algae and the closely related (3) brown algae and diatoms. (Adapted with permission from Worden *et al*., 2015)

Like in land plants, the cells of algae are surrounded by a cell wall providing them mechanical strength and shape. Cell wall polysaccharides exist in the skeletal phase and

the matrix phase. Unlike in land plants, it is the matrix phase which dominates in algae (Kloareg and Quatrano, 1988), which is due to their respective environments: Land plants are much more subjected to gravity and must develop a solid network of skeletal phase polysaccharides. The aquatic environment of algae allows them to have a predominating matrix phase, giving them the flexibility needed to grow in currents. Additionally, the matrix phase retains water well, allowing algae of the intertidal zone to avoid desiccation during low tide (Kloareg and Mabeau, 1987). The dominance of the matrix phase is significant because its polysaccharides can have very different compositions in green, red and brown algae (Table 1). They can be broadly categorized into neutral, acidic and sulfated matrix phase polysaccharides.

1.4.1 Microalgae

Microalgae contribute the dominating fraction of fixed carbon in coastal sea shelves. In spring, when temperatures and solar irradiance rise, carbon fixation peaks in massive phytoplankton blooms (Lignell *et al*. 1993; Shinada *et al*., 1999; Mann, 1999; Teeling *et al*., 2012, Teeling *et al*., 2016). Spring phytoplankton blooms are recurrent, large-scale phenomena during which immense amounts of carbon dioxide are fixed into organic carbon. Phytoplankton fixes a significant fraction of CO_2 despite its comparatively low total biomass of 3 gigatons of carbon (of a total of 700 gigatons of carbon in the oceans, Ciais *et al*., 2013) due to very high turnover rates (Arnosti *et al*., 2011; Kirchman *et al*., 2001; Moran *et al*., 2016; Piontek *et al*., 2011). In the North Sea, spring phytoplankton blooms typically form during March when, due to solar irradiance, the temperature of nutrient-rich water surpasses 3-5 °C, and reach their highest intensity in April or early May.

Table 1.1: Major cell-wall polysaccharides of green, red and brown algae and selected known polysaccharides found in diatoms. (Modified after Popper *et al*., 2011)

		Chlorophyta green algae	Rhodophyta red algae	Phaeophyceae brown algae
skeletal phase		Cellulose	Cellulose β-1,4-mannan β-1,3/4-xylans	Cellulose
matrix phase	neutral	Hemicellulose	Hemicellulose	Hemicellulose
	acidic	Ulvans	-	alginate
	sulfated	Ulvans	Agar Carrageenans Porphyran	Fucans

	Bacillariophyceae diatoms
cell-wall	sulfated α-glucuronomannan and more
storage compound	chrysolaminarin
exopoly-saccharides	highly diverse, often sulfated

Spring phytoplankton blooms are frequently dominated by diatoms, which are the predominating photosynthetic organisms in highly productive oceanic regions (Mann *et al*., 1999), while small *Prochlorococcus* cyanobacteria, likely the most abundant photosynthetic organisms on Earth, dominate in temperate, oligotrophic regions. Diatoms are unicellular eukaryotes that prominently feature a silica cell wall called frustule. They constitute a very diverse phylogenetic group of more than 250 known genera and an estimated 10^5 to 10^7 species (Guiry, 2012; Mann, 1999, Norton *et al*., 1996). In his 1999 review, diatom expert Prof. David G. Mann estimates that while “diatoms account for no more than 25-30% of primary production in oligotrophic regions”, they predominate in the highly productive oceanic regions, estimating that diatoms account for “between 40 and 45% of oceanic production (marine NPP), producing perhaps 20 Pg (or gigatons) of carbon per year, making them more productive than all the world's tropical rainforests”. Studies like Nelson *et al* (1995) and Armbrust (1995) even estimate up to 40% of global NPP for diatoms, which makes this microalgal group a meaningful target to study. Much of the carbon produced by diatoms exists in the form of chrysolaminarin, whose annual production is estimated to amount to 5-15 Gt (Alderkamp *et al*., 2007). Other dominant

microalgal groups in phytoplankton blooms include haptophytes and autotrophic dinoflagellates.

Phytoplankton blooms typically last only few weeks and then collapse due to factors such as nutrient starvation and bacterial and viral lysis. Based on which phytoplankton species dominates the bloom, specifically composed organic matter becomes available to the microbial community, resulting in distinct changes in its composition on a taxonomic and functional level (Rooney-Varga *et al*., 2005; Teeling *et al*., 2012; Rinta-Kanto *et al*., 2012; Sarmento and Gasol, 2012; Wemheuer *et al*., 2015; Bunse and Pinhassi, 2016; Sison-Mangus *et al*., 2016; Teeling *et al*., 2016). Prominent bacterial groups which have frequently been reported to reach high abundances include the classes of *Flavobacteriia* and G*ammaproteobacteria* and the alphaproteobacterial genus *Roseobacter*. Conversely, certain bacterial clades decrease in relative abundance, like members of the oligotrophy-specialists SAR11 (McCarren *et al*., 2010; Teeling *et al*., 2016).

1.5 Carbohydrates

In primary production, carbon is fixed into energy-rich, sometimes complex carbohydrates, which constitute the largest metabolically accessible repository of carbon in the biosphere (Field *et al*., 1998; Malhi, 2002). Carbohydrates are essential building blocks of life and exist as structural cell wall and matrix compounds, and as energy storage compounds, but can also be secreted as extracellular transparent exopolymeric particles (TEP). Carbohydrates consist mostly of carbon, hydrogen and oxygen but can also contain other atoms like nitrogen or phosphorus.

Monosaccharides are the basic units of carbohydrates, with a chemical composition of roughly $C_nH_{2n}O_n$ and can be linked via glycosidic bonds. The glycosidic bond is formed between the anomeric carbon of one monosaccharide and a hydroxy group of another. Glycosidic bonds exist in two stereoisomeric forms: alpha and beta. The stereochemistry of the bond is based on the position of the stereocenter furthest away from the C1. The alpha conformation of the glycosidic bond leads to an axial conformation and beta conformation creates an equatorial conformation. Maltose and cellobiose, for example, are 1,4-linked disaccharides made up of two glucose monomers (figure 1.3). Cellobiose has a β-1,4 glycosidic linkage between the C1 and the C4 of the two monomers which results in equatorial orientation of the linkage and the ring plane, resulting in straight cellulose chains that can associate side by side to form fibers of great strength. In the α-1,4-linked maltose, the glycosidic bond is orientated axially to the ring plane, which results in helical shapes that stabilize polymers in shape, for example in starch.

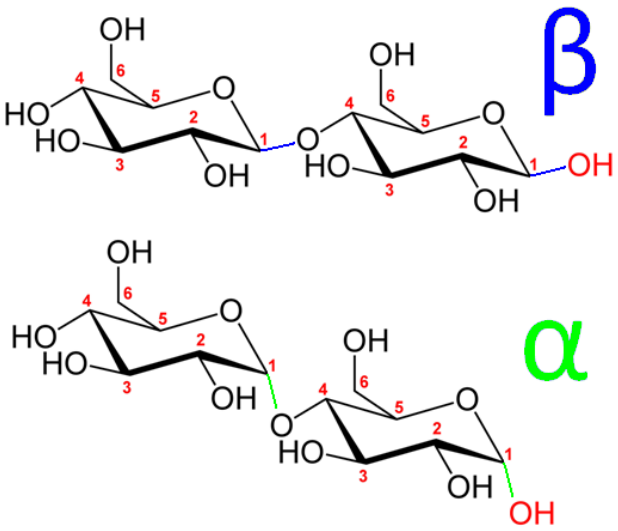


Figure 1.3 Glycosidic linkage types in cellobiose (β) and maltose (α).

As seen in figure 1.3, glycosidic bonds link different monosaccharides at different carbon positions (1,2-, 1,3-, 1,4- and 1,6-) to oligo- and polysaccharides. Polysaccharides can contain additional branches and can be substituted by different moieties (e.g. sulfate,

methyl or acetyl groups), making them the chemically most diverse macromolecules on Earth (Laine, 1994). For example, a hexasaccharide could theoretically have up to 10^{12} possible linear or branched isomers (Laine *et al*., 1994).

Many members of the *Bacteroidetes* phylum, including marine representatives of the class *Flavobacteriia*, are specialized on the degradation these complex polysaccharides. A particularly active field of research in this respect is how plant polysaccharides are degraded by human gut *Bacteroidetes*, using model species such as *Bacteroides thetaiotaomicron* and *Bacteroides ovatus* (Martens *et al.*, 2011). Together, these two *Bacteroidetes* can break down a major fraction of human dietary plant polysaccharides, including starch (see Figure 3), α-mannan, various hemicelluloses (e.g. arabinogalacturonans, xyloglucans), and pectic polysaccharides (e.g. arabinans, rhamnogalacturonans, Martens *et al.*, 2009; Larsbrink *et al.*, 2014; Cuskin *et al.*, 2015). Likewise, marine *Flavobacteriia* are abundant in polysaccharide-rich habitats, including coastal and open ocean photic zones and the upper sediment layers. While research in terrestrial polysaccharides, especially their breakdown in the gut as part of human diet and health research, has become a major research field in the recent past, their marine counterparts have only just begun to be investigated in greater detail. Known examples of macroalgal cell wall polysaccharides include ulvans in green algae (Lahaye *et al*., 2007), agars, carrageenans and porphyrans in red algae (Michel *et al*., 2006; Barbeyron *et al*., 1998; Hehemann *et al*., 2010; Ficko-Blean *et al*., 2017) and laminarin, alginate and fucans in brown algae (Read *et al*., 1996; Michel *et al*., 2010; Berteau and Mulloy, 2003), all of which are presumably absent in land plants. These cell wall polysaccharides have become the recent focus of research beyond their ecological importance, for example their potential economic value in food and fiber industries, the production of biofuel and pharmaceuticals (Niklas *et al*., 2004). In diatoms, the main storage compound is

chrysolaminarin (Beattie and Percival, 1961). It constitutes large fractions of the cellular carbon in exponentially growing diatoms (Handa, 1969; Vårum *et al*., 1986; Kroth *et al*., 2008), reaching up to 70% in *Chaetoceros pseudocurvisetus* (Myklestad and Granum, 2009). (Chryso-) laminarin has a β-1,3-glucan backbone which is occasionally branched by β-1,6 and, less frequently, β-1,2 linked glucose side chains (Gügi *et al*., 2015)

Although studies resolving the structures of algal polysaccharides have recently begun to be conducted more regularly, a full inventory of even the dominating structures of algal polysaccharides has not yet been achieved. Particularly lacking are the structural elucidations of microalgal polysaccharides, where pioneering studies on monosaccharide composition were performed in the 60s and 70s (Ford and Percival, 1965; Rees and Welsh, 1977). Their precise structures however remain, with few exceptions, unresolved (for review, see Hoagland *et al*., 1993). Complete structural analysis of algal polysaccharides requires sophisticated methods and instruments (Le Costaouëc *et al*., 2017, Ficko-Blean *et al*., 2017), which remain rarely used in marine sciences. To date, only a limited number of marine polysaccharides have been characterized beyond their monosaccharide composition. However, it is possibly exactly this detailed structural composition that selects for the marine heterotrophic bacteria (MHB) possessing the enzymatic machinery for quick and efficient degradation of these substrates, similar to how it was recently shown for complex polysaccharides in the human gut (Cuskin *et al*., 2015).

1.6 Carbohydrate-active enzymes

Carbohydrate-active enzymes, also termed CAZymes, are the enzymes involved in the carbohydrate turnover in all domains of life (Coutinho *et al*., 2003) through the degradation, modification or creation of oligo- and polysaccharides through glycosidic bonds (Lombard *et al*., 2014). Degradative CAZymes are of particular interest in the study

of algal polysaccharides decomposition. These are glycoside hydrolases (GH), polysaccharide lyases (PL) or carbohydrate esterases (CE). GHs catalyze the hydrolysis of a specific glycosidic bond, while the reaction for PLs is non-hydrolytic. CEs catalyze the degradation of carbohydrate esters. CAZymes possess catalytic modules and frequently harbor a variable number of other discrete modules, like carbohydrate-binding modules (CBMs) to recognize and bind their substrate (Boraston *et al*., 2004). In order to carry out its function, the integrity of the catalytic active site of the CAZyme is of utmost importance. For example, the carboxylic residues of two glutamates are essential for hydrolyzing the glycosylic bond between two β-1,3-linked glucose monomers in a GH16 laminarinase as described by Labourel et al. (2014).

CAZymes are categorized into families in the CAZy database project (Lombard *et al*., 2014) based on sequential and resulting structural similarities (Davies and Henrissat, 1995). This classification scheme allows easy first assessment of the phylogeny of enzymes as well as the glycobiological profile of an organism based on its “CAZome”. However, polyspecific CAZyme families exist, which feature similar catalytic mechanisms but may target structurally related but ultimately differently composed polysaccharides, demonstrating that the acquisition of novel specificity has been commonplace during evolution (Lombard *et al*., 2014). For that reason, subfamily classifications have already been established for the families GH5, GH13, GH30 and GH43, and more are inevitably going to follow. Since CAZymes can be multi-modular, they are sometimes assigned to several families if their constitutive modules belong to separate families.

Table 1.1: Number of families and modules in the CAZy database as of 13.03.2018

CAZyme class	Number of families	Number of modules
GH	152	477,170
GT	105	401,876
PL	28	12,110
CE	16	53,848
AA	15	13,274
CBM	83	119,285

Depending on the complexity of the substrate, many degradative CAZymes are usually required to synergistically degrade a complex polysaccharide into monosaccharides that can be channeled into metabolic pathways, e.g. glucose into glycolysis. For example the storage glucan laminarin, which occurs in brown algae and in diatoms as chrysolaminarin, is a β-1,3-linked glucan featuring β-1,6-linked glucose side chains. It requires the synergistic actions of endo- and exo-active β-1,3-glucanases (GH17, GH16) and β-1,6-glucosidases (GH30) to be fully degraded (Figure 1.3).

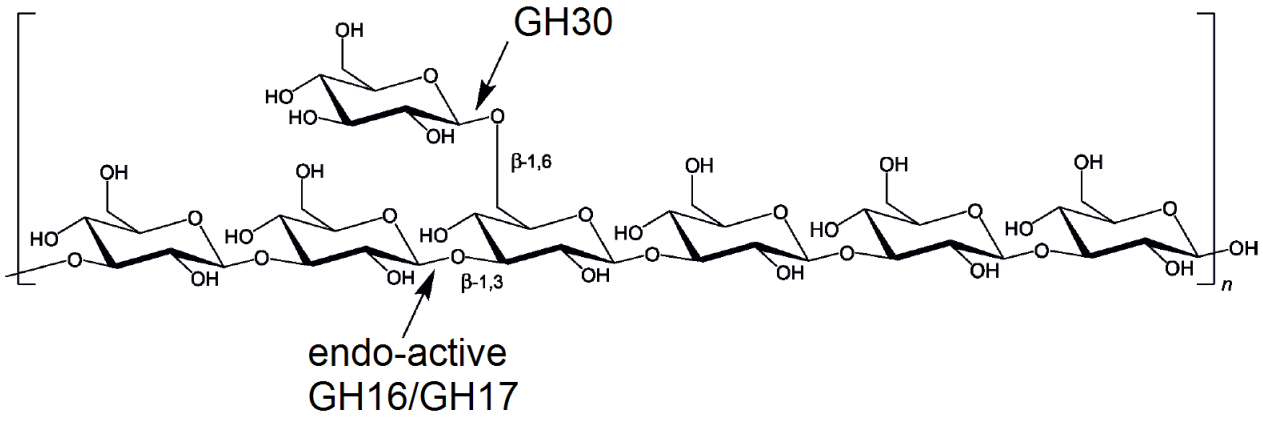


Figure 1.3: Molecule of a (chryso-) laminarin polysaccharide, the major storage compound of diatoms and brown algae. The GH30 β-1,6-glucosidase removes glucose side chains from laminarin, while endo-active β-1,3-glucanases cleaves the laminarin chain into smaller oligosaccharides. Finally, an exo-acting β-1,3-glucosidase (e.g. GH17) cleaves of glucose units from the oligosaccharide (not shown, modified from Becker *et al.*, 2017).

1.7 Polysaccharide utilization loci

In *Flavobacteriia*, genes encoding the enzymatic machinery to sense, bind, import and incrementally degrade a polysaccharide are often co-located in distinct, regulated gene clusters termed polysaccharide utilization loci (PUL, Bjursell *et al.*, 2006). PULs most prominently comprise substrate-specific degradative CAZymes and a *susCD*-like gene pair for substrate binding and transport into the periplasm. *SusCD*-like genes are named after their structurally conserved counterparts in the archetypal starch utilization system (Sus), the first PUL featuring eight genes (*susRABCDEFG*) described in detail by the group of Abigal Salyers in *Bacteroides thetaiotaomicron* (Figure 1.3). The SusC-like protein is an integral membrane protein of the family of TonB-dependent transporters (TBDT). TBDTs are typically found in Gram-negative bacteria. Their beta-barrel transport channel is capped by the N-terminally lipidated SusD-like cell surface glycan binding protein (SGBP),

forming a “pedal bin”-like protein complex (Glenwright *et al*., 2017). In the absence of a ligand, the substrate-binding site is exposed to the extracellular space. Upon substrate binding, the SusD lid closes and conformational changes, driven by the energy of the inner membrane, lead to substrate transport and release into the periplasm (Hickman *et al*., 2017). For larger substrates, endo-active cell surface anchored CAZymes cleave the polysaccharides into smaller oligosaccharide suitable for transport into the periplasm. Quick oligosaccharide uptake has been shown and has been speculated to represent a “selfish” strategy to avoid diffusive substrate loss in often highly competitive communities in the environment (Cuskin *et al*., 2015; Reintjes *et al*., 2017). In the periplasm, further saccharification of oligosaccharide substrates to component monosaccharides takes place which are finally taken up into the cytoplasm through dedicated transporters. Oligosaccharides in the periplasm may also bind carbohydrate sensor system proteins leading to transcriptional activation of PUL genes. Thus, sensing of early break-down products of targeted polysaccharides informs the bacterium that a specific polysaccharide is present in its surroundings and allows for a rapid, specific response towards this substrate, as well as co-regulated responses towards other substrates that naturally co-occur in the environment, as previously described for different glucans in *Gramella forsetii* KT0803 (Kabisch *et al.*, 2014).

Additionally, PULs frequently encode proteases for glycoprotein substrates, phosphatases or sulfatases, which are prominent in PULs targeting marine polysaccharides. Most likely, many PULs also feature additional SGBPs lacking sequence conservation to be considered as SusE/F homologs (Grondin *et al*., 2017). Overall, the complexity of PULs scales with their corresponding substrate. It has also become apparent that reduced versions of PULs exist, which do not feature SusD-like SGBPs, have TBDTs with sequence similarities too low to be characterized as SusC-like (Blanvillain *et al*., 2007) or

are not co-localized in one cluster but possess dedicated, upregulated genes elsewhere in the genome (Ficko-Blean *et al*., 2017; Unfried *et al*., in review, see appendix). Considering these findings, the original definition of the term “PUL”, referring to clusters of “genes encoding GH and PL activities [that frequently] co-localize adjacent to SusC and SusD paralogues” (Bjursell *et al*., 2006) is not timely and accurate any more. It does not cover the common understanding that a “bacterial polysaccharide utilization locus (PUL) is a set of physically linked genes that orchestrate the breakdown of a specific glycan” (Terrapon *et al*., 2015). Such loci have also been referred to as “carbohydrate utilization containing TBDT” (CUT) loci (Blanvillain *et al*., 2007), and can be considered PUL sub-types (Grondin *et al*., 2017).

1.8 PUL analyses provide insights into the structure of polysaccharides and their decomposition by adapted microbes

The PUL repertoires of different bacterial species can be highly individual and allow for resource partitioning and inhabiting specific niches. Determining the polysaccharide degradation capacities through PUL analyses hence reveals key features about an organism’s lifestyle and ecological niche. MHB can for example live attached to a surface (e.g. macroalgae) or pelagically (free-floating in the water column). Characteristically, macroalgae-attached bacteria feature many dedicated PULs targeting the specific structural and storage polysaccharides synthesized by their host (Barbeyron *et al*., 2016). On the other hand, pelagic bacteria typically possess a more limited set of PULs often targeting more generic, abundant polysaccharides of microalgae such as diatom-derived laminarin also known as chrysolaminarin (e.g. *Formosa* sp. Hel1_33_131, Appendix). Some MHB can only target marine polysaccharides whereas others can additionally

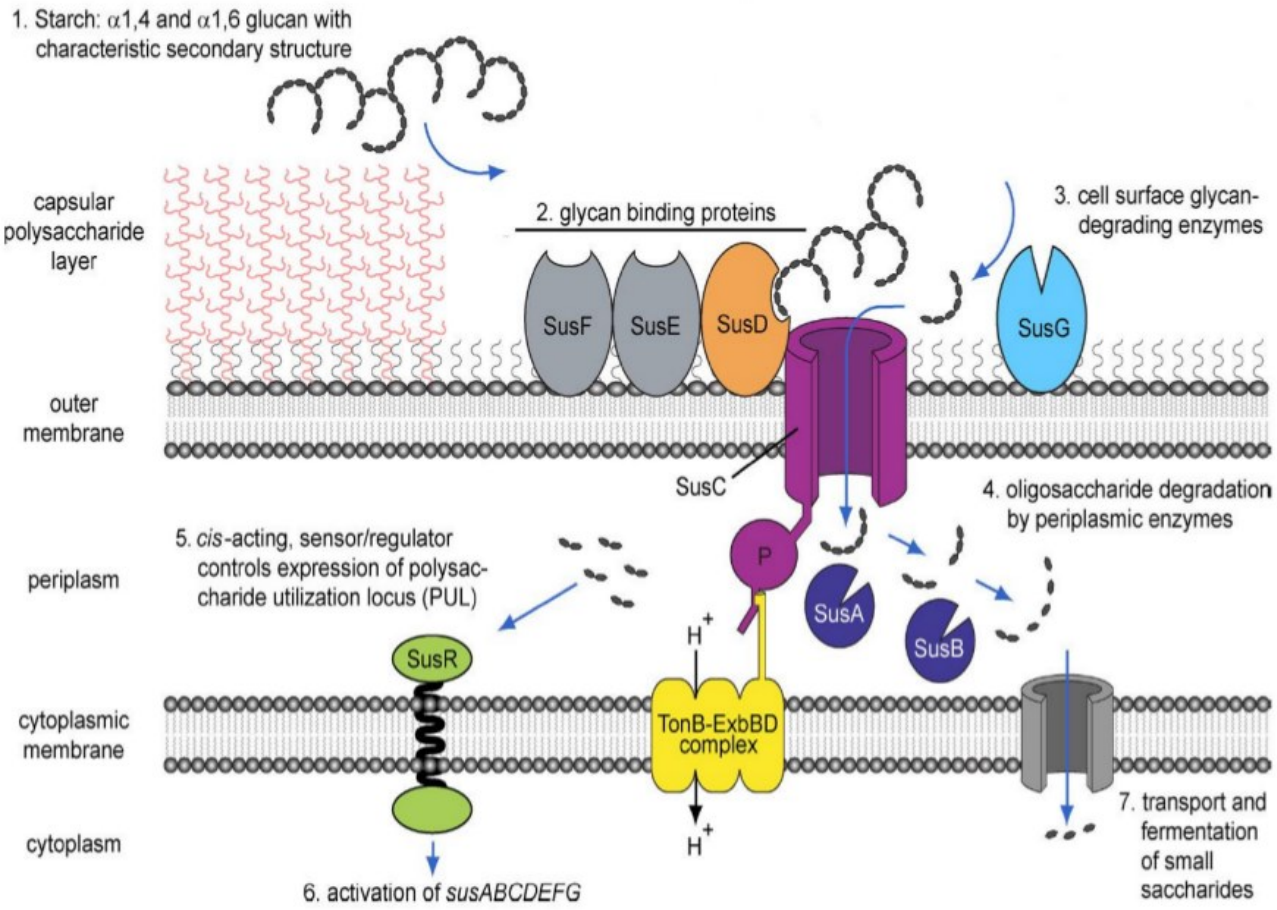


Figure 1.4: Functional model of the starch utilization system (*susRABCDEFG*) in *Bacteroides thetaiotaomicron,* depicting the incremental breakdown of starch. (1) Large starch molecules are bound by the cell surface proteins SusD, SusE and SusF (2) and hydrolyzed into smaller oligosaccharides by an endo-amylase (SusG, 3). The TonB-dependent transporter complex encoded by SusC imports these oligosaccharides, where they are hydrolyzed into glucose monomers by SusA neopullulanases and SusB glucosidases (4). Oligosaccharides also bind (5) to the N-terminal segment of a hybrid two-component system (HTCS; SusR) leading to transcriptional activation of the *susABCDEFG* genes (6). Finally, mono- or disaccharides are taken up into the cytoplasm (7) where they serve as a carbon or energy source. (Adapted with permission from Martens *et al*., 2009)

degrade terrestrial carbohydrates like (hemi-)celluloses or pectins, which makes them well adapted to coastal environments where terrestrial and riverine input is often prevalent (Xing *et al*., 2015).

The number of PULs and degradative CAZymes (GH, PL, CE) in a microbial genome often provides a clear indication whether the organism is more of a carbohydrate generalist or specialist. In *B. thetaiotaomicron*, a commensal of the extremely polysaccharide-rich human gut, 88 putative PULs have been identified, which make up almost 20% of the genome (Martens, *et al*. 2011). Likewise, the macroalgae-associated flavobacterial isolate *Z. galactanivorans* DsiJ^{T} has 50 PULs and 174 degradative CAZymes encoded in its large 5.5 Mbp genome (Barbeyron *et al*., 2016). Contrarily, the most abundant marine bacterium in the ocean, the oligotrophic *alphaproteobacterium Candidatus Pelagibacter ubique* features only 7 degradative CAZymes (strain HTCC1062) in its streamlined 1.3 Mbp genome.

Corresponding to the sparse understanding of the structural diversity of algal polysaccharides, we do not have a detailed understanding of the PUL diversity in MHB. Among the few well characterized PULs are an porphyran-specific PUL which was transferred to human gut *Bacteroidetes* from MHB (Hehemann *et al*., 2010 and 2012), alginate-specific PULs in *Zobellia galactanivorans* DsiJ^{T} and *Gramella forsetii* KT0803 (Thomas *et al*., 2012; Kabisch *et al*., 2014) and laminarin-specific PULs in *Gramella forsetii* KT0803 and *Formosa* spp. (Kabisch *et al*., 2014; Becker *et al*., 2017; Unfried *et al*. in review, see appendix). Comparative genomic studies (for example Xing *et al*., 2015; Barbeyron *et al*., 2016) so far focused on overall CAZyme family repertoires and not dedicated PULs, which is necessary to elucidate the structures of their targeted marine polysaccharides and to understand their decomposition by microbes.

In this thesis, I will show how in-depth genomic PUL analyses can provide insights into the structural composition of marine polysaccharides and help to understand their decomposition by microbes. Understanding these polysaccharide degradation capacities, especially those of abundant bacteria associated with phytoplankton-blooms, will significantly advance the knowledge on processes influencing and regulating major carbon fluxes in the marine (and global) carbon cycle.

1.9 Aims of this thesis

The overarching scope of this doctoral thesis was to investigate how marine heterotrophic bacteria degrade which kinds of marine polysaccharides. I particularly focused on *Flavobacteriia*, which are known as polysaccharide specialists and hence abundant in spring phytoplankton blooms. Since the complex structures of these polysaccharides are not directly determinable *in situ*, I examined in great detail the degradative genetic repertoires encoded in genomes, metagenomes and fosmids of different habitats, but with a focus on the degradation of the spring phytoplankton bloom-derived organic matter, as it occurs in the Southern North Sea. Since significant amounts of carbon are fixed and remineralized in these blooms, they contribute significantly to the fluxes in the marine carbon cycle.

In chapter II, "**Polysaccharide Utilization Loci of *Bacteroidetes* from two contrasting open ocean sites in the North Atlantic**", I investigated the role of so far uncultivated marine *Bacteroidetes* in the degradation of high molecular weight organic matter in the surface waters of a mesotrophic polar and an oligotrophic gyre open ocean site. I explored whether the two locations of DNA isolation, which are likely characterized by different

polysaccharide sources (microalgae, animals, bacteria) and composition select for specifically adapted *Bacteroidetes.*

In chapter III, "**Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms**", we investigated to what degree recurrent patterns can be determined in response to North Sea spring phytoplankton blooms over four years on the taxonomic and functional level. We examined which bacterial clades reappeared every year despite substantial inter-annual variation between spring phytoplankton blooms and what functional genes they were equipped with to successfully degrade the algae-derived organic matter.

In chapter IV, "**Polysaccharide Utilization Loci of 53 North Sea *Flavobacteriia* as Basis for Using SusC/D-Protein Expression for Predicting Major Phytoplankton Glycans**", my main piece of work, I investigated the polysaccharide degradation capacities encoded in the genomes of 53 *Flavobacteriia* isolates from the North Sea; many of them surface seawater isolates sampled during spring phytoplankton blooms. My goal was to find polysaccharide utilization loci (PULs) targeting relevant and as yet undescribed marine substrates and contribute to establishing a PUL inventory in the process. Additionally, I examined whether strongly expressed *susC* and *susD*-like genes for polysaccharide binding and transport are substrate-specific and can be used as proxies to determine microbial utilization of corresponding polysaccharides in metaproteomes.

Chapter II: Polysaccharide Utilization Loci of *Bacteroidetes* from two contrasting open ocean sites in the North Atlantic

Published in *Environmental Microbiology*

Christin M. Bennke[1‡], Karen Krüger[1‡], **Lennart Kappelmann[1‡]**, Sixing Huang[2], Angélique Gobet[3], Margarete Schüler[4], Valérie Barbe[5], Bernhard M. Fuchs[1], Gurvan Michel[3]*, Hanno Teeling[1]*, Rudolf I. Amann[1]*

[1] Max-Planck-Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany

[2] Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Inhoffstraße 7B, 38124 Braunschweig, Germany

[3] Sorbonne Université, UPMC Univ Paris 06, CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff cedex, Bretagne, France

[4] University of Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany

[5] Laboratoire de Biologie Moléculaire pour l'Étude des Génomes, C.E.A., Institut de Génomique - Genoscope, 2 rue Gaston Crémieux, 91057 Évry cedex, France

‡ These authors made equal contributions

Contribution to the manuscript:

Experimental concept and design	10%
Acquisition of experimental data	10%
Data analysis and interpretation	30%
Preparation of figures and tables	30%
Drafting of manuscript	30%

***Corresponding authors:**

Prof. Dr. Rudolf Amann
Max-Planck-Institute for Marine Microbiology
Department of Molecular Ecology
Celsiusstraße 1
28359 Bremen, Germany
Phone: +49 421 2028 930
Fax: +49 421 2028 790
Email: ramann@mpi-bremen.de

Dr. Gurvan Michel
UMR 8227 (CNRS-UPMC)
Station Biologique de Roscoff
Place Georges Teissier
29682 Roscoff, France
Phone: +33 298 2923 30
Email: gurvan.michel@sb-roscoff.fr

Dr. Hanno Teeling
Max-Planck-Institute for Marine Microbiology
Department of Molecular Ecology
Celsiusstraße 1
28359 Bremen, Germany
Phone: +49 421 2028 976
Email: hteeling@mpi-bremen.de

Email addresses and telephone numbers of all authors:

C. M. Bennke	christin.bennke@io-warnemuende.de	+49 381 5197 3437
K. Krüger	kkrueger@mpi-bremen.de	+49 421 2028 545
L. Kappelmann	lkappelm@mpi-bremen.de	+49 421 2028 545
S. Huang	sih13@dsmz.de	+49 531 2616 381
A. Gobet	agobet@sb-roscoff.fr	+33 298 2923 62
M. Schüler	margarete.schüler@uni-bayreuth.de	+49 921 55 2226
V. Barbe	vbarbe@genoscope.cns.fr	+33 160 8725 00
B. M. Fuchs	bfuchs@mpi-bremen.de	+49 421 2028 935
G. Michel	gurvan.michel@sb-roscoff.fr	+33 298 2923 30
H. Teeling	hteeling@mpi-bremen.de	+49 421 2028 976
R. Amann	ramann@mpi-bremen.de	+49 421 2028 930

2.1 Summary

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

2.2 Introduction

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

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

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

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

most functional studies on PULs have been conducted for land plant polysaccharide-specific PULs in human gut bacteria, for example recently for xyloglucan decomposition by human gut *Bacteroidetes* (Larsbrink *et al.*, 2014). Only few PULs have been characterized for polysaccharides of marine origin, such as agar/porphyran-specific PULs (Hehemann *et al.*, 2012a) that have been laterally transferred from marine *Bacteroidetes* to *Bacteroidetes* of the human gut (Hehemann *et al.*, 2010; Hehemann *et al.*, 2012b) and alginate-specific PULs (Thomas *et al.*, 2012). Notably, alginate induction experiments with *Zobellia galactanivorans* DsiJ^T demonstrated that its alginate-specific PUL is a genuine operon (Thomas *et al.*, 2012). A recent proteomic study on the coastal marine bacteroidetes *Gramella forseti*i KT0803 confirmed expression of proteins encoded by a homologous alginate-specific PUL, and identified an additional laminarin-induced PUL (Kabisch *et al.*, 2014). The latter was also shown to be present and inducible by laminarin in a proteomic study of the coastal marine bacteroidetes *Polaribacter* sp. Hel1_33_49 (Xing *et al.*, 2015). These findings notwithstanding, we still have little knowledge on the PUL repertoire and associated degradation potential of marine *Bacteroidetes*, in particular for those thriving in the mostly oligotrophic open oceans.

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

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

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

2.3 Experimental Procedures

2.3.1 Study sites and fosmid library preparation

Samples were taken in the North Atlantic Ocean during the VISION cruise MSM03/01 on board of R/V Maria S. Merian in September 2006 (Fig. S1, Table S1). Fosmid metagenome libraries were constructed from surface water of two contrasting oceanic provinces. Samples from station 3 (S3) were collected in the Boreal Polar province

(65°52.64’ N, 29°56.54’ W) and samples from station 18 (S18) in the North Atlantic Subtropical province (34°04.43’ N, 30°00.09’ W). Libraries of 35,000 (S3) and 50,000 (S18) fosmids were constructed for both sites. Subsequently, 16,938 (S3) and 16,266 (S18) high-quality end sequences were generated by sequencing inserts from both sites using the Sanger technique. Details have been described elsewhere (Gómez-Pereira *et al.*, 2010; Gómez-Pereira *et al.*, 2012).

2.3.2 Selection and sequencing of fosmids

End sequences were mapped on the 76 previously sequenced fosmids from both libraries in order to detect connecting fosmids. Using a sequence identity threshold of 94.5% or higher, 43 (S3) and 27 (S18) connecting fosmid candidates were identified. Twenty of these had the potential to prolong partial PULs on the previously sequenced fosmids. These were sequenced at LGC Genomics (LGC Genomics GmbH, Berlin, Germany) using the 454 FLX Ti platform and assembled using Newbler. Further 104 putative *Bacteroidetes* fosmids were selected based on BLASTx hits of end-sequences to the NCBI non-redundant protein sequence databases with a rank-based evaluation similar as proposed by Podell and Gasterland (Podell and Gaasterland, 2007), phylogenetic reconstructions based on HMMer 3 searches of all-frame translated end sequences against the Pfam v. 25 database (Krause *et al.*, 2008) and an evaluation of end sequence tetranucleotide usage patterns (Teeling *et al.*, 2004). These fosmids and the remaining 50 connecting fosmid candidates were sequenced at Genoscope (Évry Cedex, France) using the 454 FLX Ti platform and assembled using Newbler as described previously (Gómez-Pereira *et al.*, 2010; Gómez-Pereira *et al.*, 2012). The final dataset comprised 250 fosmids.

2.3.3 Fosmid re-assembly

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

2.3.4 Taxonomic classification

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

2.3.5 Automated gene prediction and annotation

Gene prediction and annotation of all 226 contigs was done via the RAST server (Aziz *et al.*, 2008). The RAST gene calls and annotations of the included 76 published fosmids differed only marginally from the published ones. Results for all contigs were downloaded and subsequently imported into a local installation of the GenDB (v. 2.2) annotation system (Meyer *et al.*, 2003) for curation. CAZymes were annotated based on HMMER searches against the dbCAN database (Yin *et al.*, 2012), BLASTp (Altschul *et al.*, 1990) searches against the CAZy database (Cantarel *et al.*, 2009; Lombard *et al.*, 2014) and HMMER searches against the Pfam v. 28 database (Finn *et al.*, 2010) using E-values derived from manual annotations of test data (Table S5). CAZymes were only annotated when at least two of the three database searches yielded positive results. Peptidases were automatically annotated based on batch BLASTp searches against the MEROPS 9.13 databases (Rawlings *et al.*, 2012) using the default E-value cutoff criterion of 10^{-4}. ABC transporter, TonB-dependent receptor, *susD* genes and sulfatase genes were automatically predicted based on HMMer 3 hits to the Pfam v. 28 database at $E \leq 10^{-5}$

using the following profiles: ABC_tran, ABC_membrane, ABC_membrane_2, ABC_membrane_3, ABC_tran_2, ABC2_membrane, ABC2_membrane_2, ABC2_membrane_3, ABC2_membrane_4, ABC2_membrane_5, ABC2_membrane_6, TonB_dep_Rec, SusD, SusD-like, SusD-like_2, SusD-like_3 and sulfatase. Annotated sequences were deposited at NCBI's Genbank (BioSample accessions SAMN04870880 and SAMN04870884).

2.3.6 Manual CAZyme annotation

CAZymes were identified based on homology with a selected subset of characterized enzymes from each CAZyme family. Initial annotations were validated by BLASTp searches against the UniProtKB/SwissProt database as of February 2014 (The UniProt Consortium, 2014) and HMMer searches against the Pfam v. 27 database (Punta *et al.*, 2012). Each CAZyme was assigned to a CAZY family and, when possible, to an EC number. Abundant glycoside hydrolases from the multi-functional families GH3, GH5, GH13, and GH16 were subjected to an in-depth phylogenetic analysis to determine their substrate-specificities. Experimentally characterized proteins (Table S6) were selected from the CAZy database for each activity within a given GH family and aligned to their contig homologs using MAFFT (FFT-NS-i iterative refinement method; BLOSUM62 amino acid substitution matrix) (Katoh and Standley, 2013). These alignments were used to calculate model tests and maximum likelihood trees with MEGA v. 6.0.6 (Kumar *et al.*, 2004) with bootstrapping (100 resamplings). Annotation of ambiguous proteins was refined based on the proximity to characterized proteins in the phylogenetic trees.

Subcellular locations were predicted using CELLO v. 2.5 (Yu *et al.*, 2006), PSORTb v. 3.0.2 (Yu *et al.*, 2010) and HMMer searches against the TIGRfam profile (Selengut *et al.*,

2007) TIGR04183 (Por secretion system C-terminal sorting domain). Only unambiguous consensus predictions were considered as reliable.

2.4 Results and Discussion

2.4.1 Characterization of the dataset

The initial 76 fosmids (S3: 40; S18: 36) were pooled with 154 newly sequenced putative *Bacteroidetes* fosmids (S3: 84; S18: 70) and re-assembled. This way, 107 contigs were obtained from S3 and 95 contigs from S18. Some of the new fosmids were selected for sequencing due to high similarity of their end-sequences to previously sequenced fosmids. Thereby it was possible to extend some of the initial fosmids and to obtain assemblies of up to 85.4 kbp (S3) and 72.2 kbp (S18). In addition, 14 putative PULs were retrieved that were not obtained before. Additional fosmids (S3: 11; S18: 9) were selected in order to possibly extend these PUL-carrying fosmids, which after assembly yielded additional 24 contigs (S3: 14; S18: 10). In summary, the total dataset comprised 250 fosmids (S3: 135; S18: 115) that were re-assembled to 226 contigs (S3: 121; S18: 105) of 8.8 Mbp (S3: 4.7 Mbp; S18: 4.1 Mbp).

Based on gene content 96% (S3) and 92% (S18) of the contigs affiliated with *Bacteroidetes.* Thus the error of selecting *Bacteroidetes* fosmids based on information from combined ~1.4 kbp Sanger sequenced end sequences was comparable to the ~5% error of PCR-based screening (Gómez-Pereira *et al.*, 2012). Sequenced non-*Bacteroidetes* contigs affiliated with the *Planctomycetes-Verrucomicrobia-Chlamydia* (PVC)-cluster and with *Proteobacteria*. On class level 95% of the *Bacteroidetes* contigs affiliated with *Flavobacteriia* at S3 and 94% at S18, of which 96% affiliated with the family *Flavobacteriaceae* at both stations. On genus level (Fig. 1) S3 contigs affiliated most fre

quently with *Polaribacter* (39)%, *Flavobacterium* (13%), *Dokdonia* (6%) and *Gramella* (6%), whereas S18 contigs most frequently affiliated with *Dokdonia* (25%), *Leeuwenhoekiella* (17%), *Flavobacterium* (11%) *Robiginitalea* (8%), *Polaribacter* (7%), and *Croceibacter* (7%). The numbers obtained for *Polaribacter* are in good agreement with *in situ Polaribacter* abundances that were previously determined by catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) of the same samples (Gómez-Pereira *et al.*, 2010). However, *Gramella* was not detected using CARD-FISH at both stations, and *Leeuwenhoekiella* and *Dokdonia* were only detected at station 18 with abundances below 1% (Gómez-Pereira *et al.*, 2010). Such discrepancies are expected, since our taxonomic affiliations are based on gene BLASTp and HMMer database similarity searches and thus biased towards publically available sequenced *Flavobacteriia*.

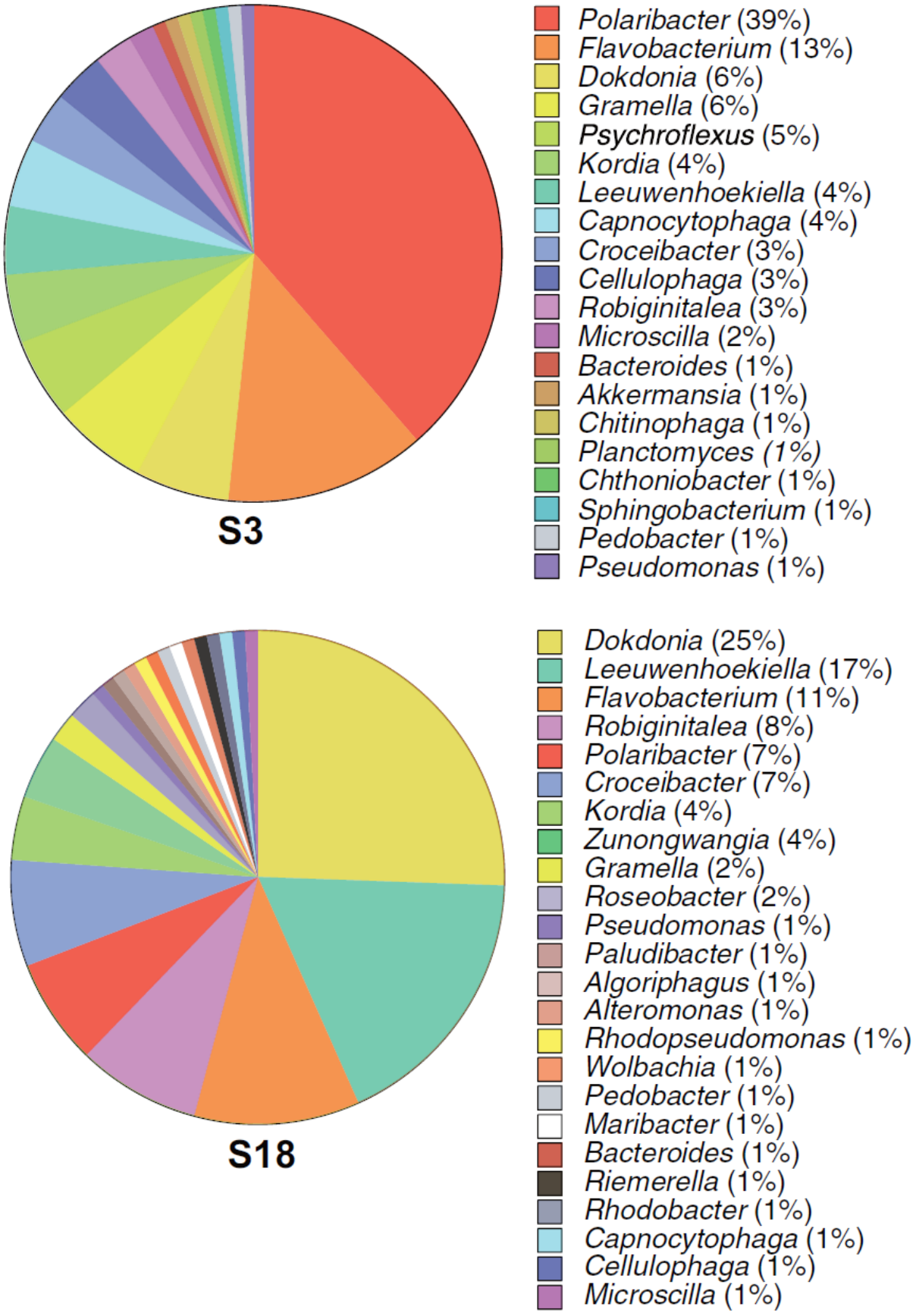


Figure 2.1: Genus-level taxonomic affiliation of contigs obtained from re-assembled fosmid sequences of metagenomic libraries from station S3 (121 contigs) and S18 (105 contigs).

2.4.2 Bacteroidetes' peptidases and CAZymes

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

Genes for polysaccharide binding (*susD*) and GH genes exhibited about equal frequencies in both datasets, while TonB-dependent receptor and sulfatase genes were 1.6 to 1.7-fold

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

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

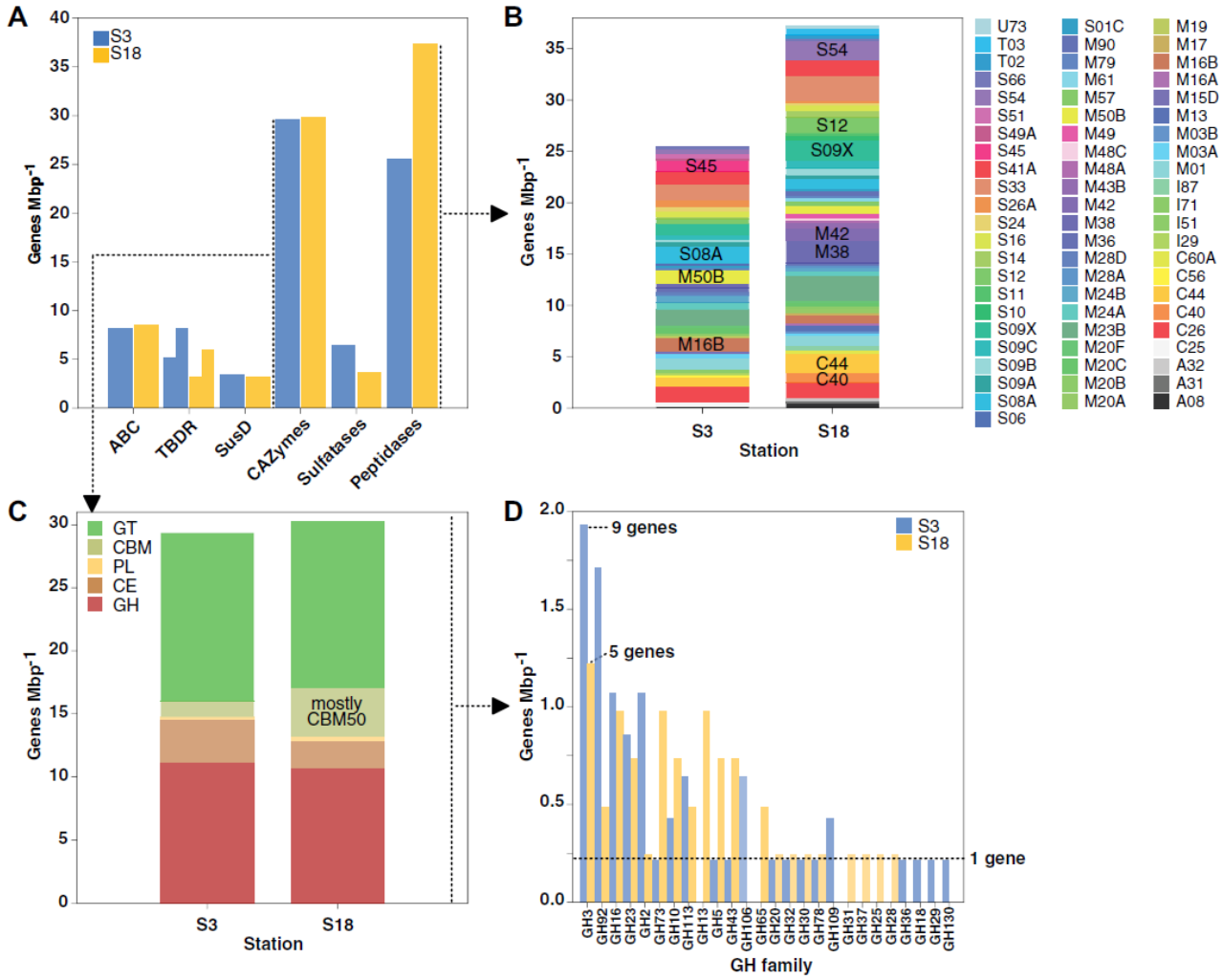


Figure 2.2: Comparison of selected genes on the 121 BPLR station 3 contigs (4.7 Mbp) and the 105 NAST station 18 contigs (4.1 Mbp): [A] Genes for ABC-transporters, TonB-dependent receptors (TBDRs), SusD-like proteins, CAZymes, sulfatases and peptidases. [B] Peptidase families. [C] CAZyme classes: glycoside hydrolases (GH), carbohydrate esterases (CE), polysac-charide lyases (PL), carbohydrate-binding modules (CBM) and glycosyltransferases (GT) [D] GH families ordered by decreasing averages.

Table 2.1: Comparison of frequencies of genes involved in organic matter degradation and uptake between the S3 and S18 contig datasets.

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

[a] HMMer 3 searches against the Pfam v. 28 database with E ≤ 10^{-5}. Values for ABC transporter and susD genes were determined by combining all genes with hits to any of the following profiles: ABC_tran, ABC_membrane, ABC_membrane_2, ABC_membrane_3, ABC_tran_2, ABC2_membrane, ABC2_membrane_2-6, as well as SusD, SusD-like, SusDlike_2 and SusD-like_3. Values for TonB-dependent receptor genes were determined using the TonB_dep_rec profiles from October 2014 and 2012 (values in brackets). Both TonB_dep_rec profiles predicted lower numbers as was suggested by BLAST-based database similarity searches.

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

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

2.4.3 Commonalities between PULs from both stations

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

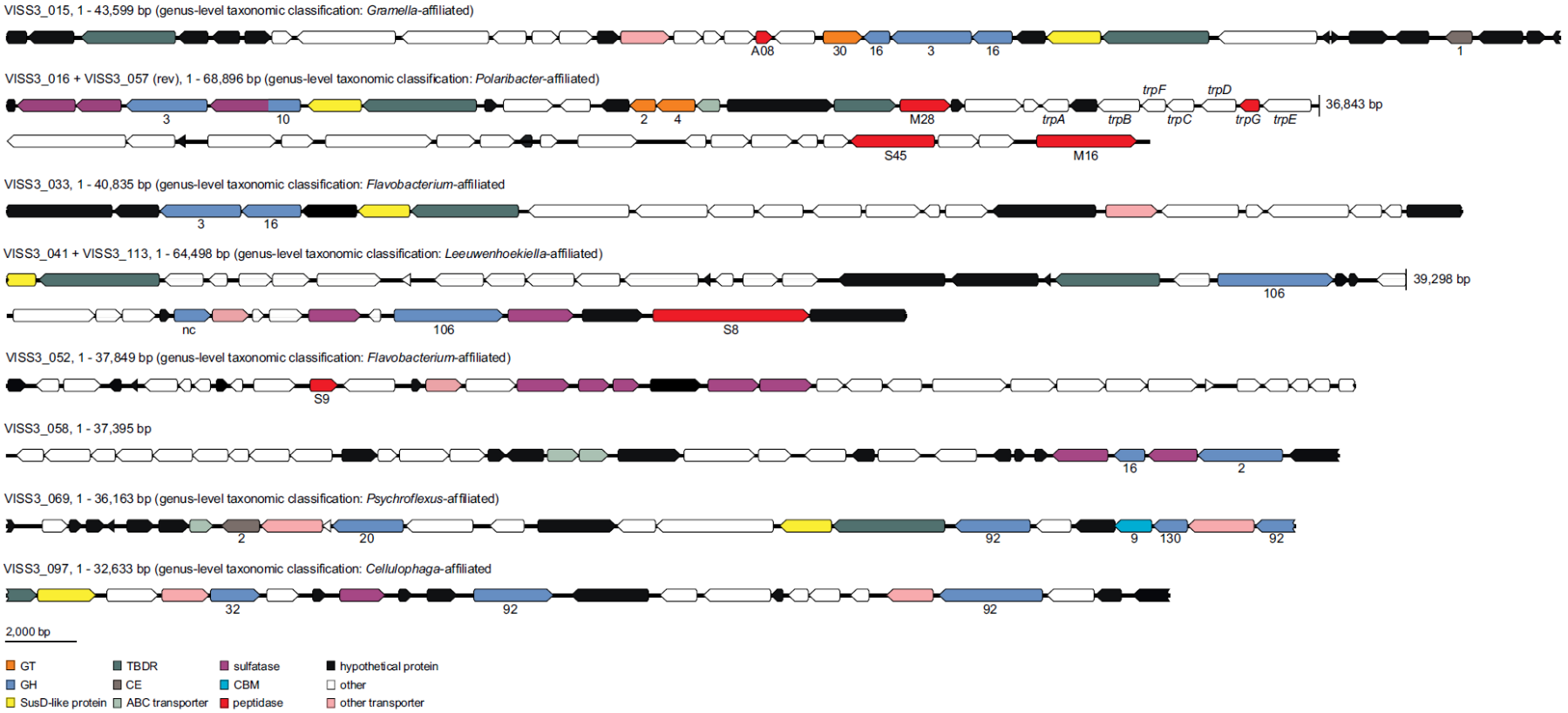


Figure 2.3: PUL-containing contigs from BPLR station 3. Names, presumed taxonomic affiliations, lengths and the gene contents are provided for each contig.

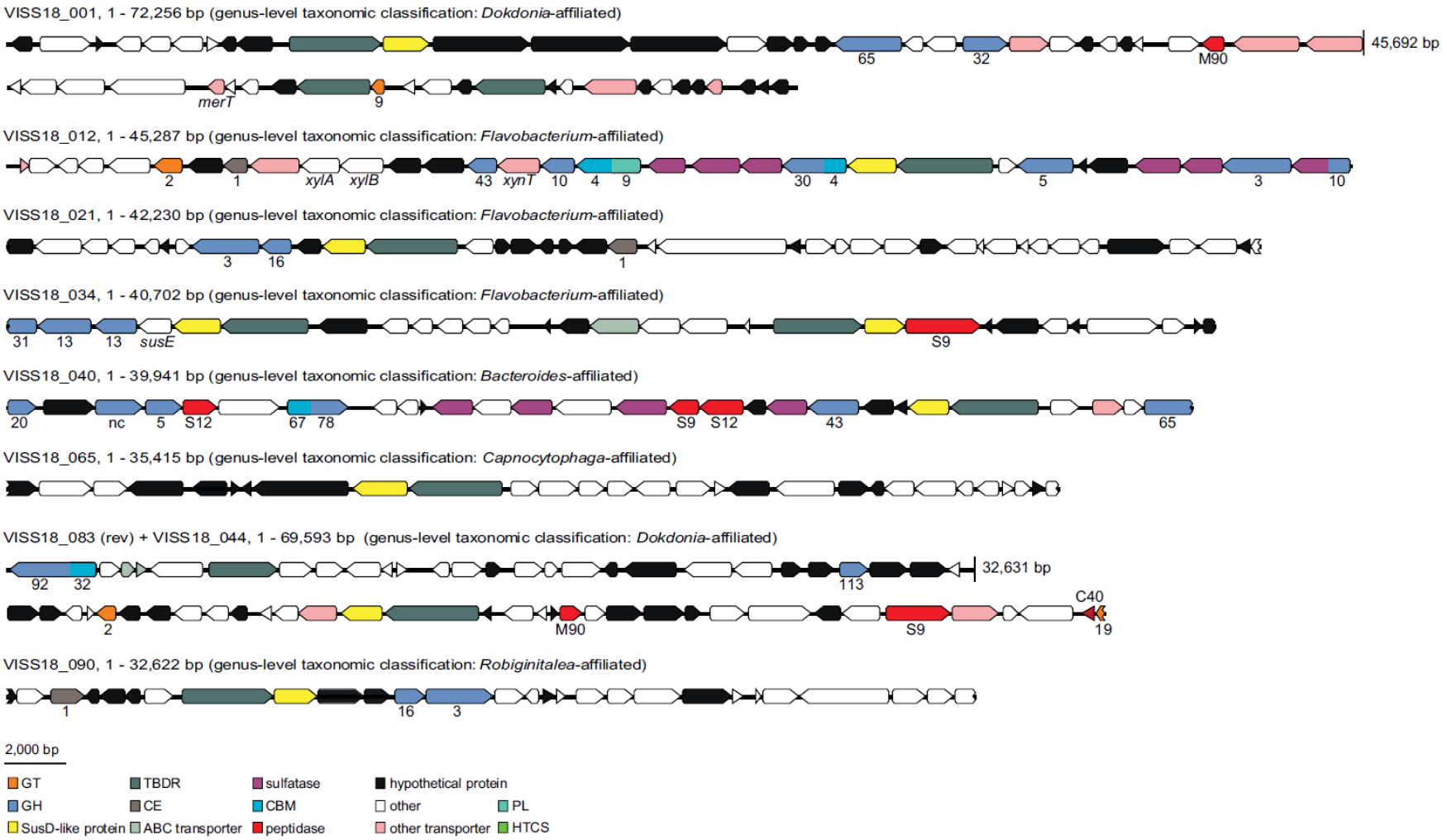


Figure 2.4: PUL-containing contigs from NAST station 18. Names, presumed taxonomic affiliations, lengths and the gene contents are provided for each contig.

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

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

via xylose and xylulose. The respective genes *xylA* (xylose-isomerase) and *xylB* (xylulose-kinase) were identified downstream of the PUL. Neutral xylan occurs in red and green algae (Popper *et al.*, 2011), and sulfated xylan has been found in the red macroalga *Palmaria palmata* (Deniaud *et al.*, 2003). Likewise, exopolysaccharides (EPS) of some diatoms and bacteria are sulfated and rich in xylose.

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

2.4.4 Additional PULs from BPLR station 3

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

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

Contig VISS3_097 has been classified as *Cellulophaga*-related. *Cellulophaga* species are known to be associated with diatoms and macrophytes from cold marine waters, e.g.

C. algicola (Bowman, 2000). This PUL on this contig contains a sulfatase (ORF VISS3_097_09), two putative GH92 family alpha-1,2-mannosidases (ORFs VISS3_097_13, VISS3_097_22) and a putative GH32 family levanase (ORF VISS3_097_05). The latter might hydrolyze 2,6-beta-D-fructofuranosidic linkages in 2,6-beta-D-fructans. Fructans (or levans for bacteria) can be synthesized by green algae or bacteria. Bacterial fructans are produced extracellularly and generally composed of beta-2,6-linked fructosyl residues linked to a terminal glucose, such as for example in *Lactobacillus* and *Streptococcus* species (Corrigan and Robyt, 1979; Hendry, 1993; Van Geel-Schutten *et al.*, 1999). Therefore, this PUL might be dedicated to the degradation of sulfated EPS containing mannose and/or fructose residues.

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

2.4.5 Additional PULs from NAST station 18

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

that participates in dimer formation (Lee *et al.*, 2002; Stam *et al.*, 2006). Cyclomaltodextrinases effectively hydrolyze cyclomaltodextrin, a circular sugar derived from starch degradation, whereas the degradation of starch and pullulan is less effective (Lee *et al.*, 2002). The PUL also contains a putative GH31 family alpha-glucosidase (ORF VISS18_034_01), which hydrolyzes the oligosaccharides released by alpha-amylases. Therefore, this PUL likely targets an alpha-1,4-glucan, which is for example produced by some bacteria as storage compound during stationary growth (Preiss, 1984; Field *et al.*, 1998). Usage of bacterial polysaccharides by *Bacteroidetes* may be of higher relative importance at NAST station S18 since algae were less abundant than at the BPLR station S3 (Table S1).

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

The PUL on contig VISS18_040 contains enzymes involved in starch and sucrose metabolism, such as a predicted GH65 family maltose phosphorylase (ORF VISS18_040_30) and GH43 family xylosidase/arabinofuranosidase (ORF VISS18_040_21). This PUL also contains a putative GH20 beta-N-acetylhexosaminidase (ORF VISS18_040_01), a GH5_13 glycoside hydrolase that may target beta-mannan (ORF VISS18_040_04; Fig. S4; Aspeborg *et al.*, 2012), a putative GH109 alpha-N-acetylgalactosaminidase (ORF VISS18_040_13), and a GH of unknown specificity (ORF VISS18_040_03). Furthermore, the PUL contains a predicted GH78 family alpha-L-

rhamnosidase (ORF VISS18_040_07) and four sulfatases (ORFs VISS18_040_12, _14, _16, _20; family S1 and subfamilies 8, 9, 16, 20; Table S2), similar to the PUL of S3 contigs VISS3_113 + VISS3_041. These enzymes may be involved in the hydrolysis of sulfated polysaccharides (e.g. algal ulvans).

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

2.4.6 Comparative analysis of PULs

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

BPLR contigs VISS3_016 + VISS3_057 and the NAST contig VISS18_012 have a highly similar region harboring homologous GH3 family genes, neighbored by three non-homologous sulfatase genes on both contigs (Fig. S6C). Similarly four of the putative laminarinase-containing contigs from both stations (BPLR: VISS3_015, VISS3_033; NAST: VISS18_021, VISS18_090) exhibited a high level of sequence conservation (Fig. S6D).

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

2.5 Conclusions

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

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

PUL comparisons indicated that sulfated xylan-rich polysaccharides and algal laminarin are possibly among the more frequent polysaccharide substrates for open ocean *Bacteroidetes*. One PUL at each station contained putative xylan-specific genes and sulfatases, suggesting xylan-rich polysaccharide of marine origin as substrates (e.g. from algal EPS). Four out of the 14 PULs in our dataset were likely laminarin-specific. Similar PULs have been identified in other members of the *Bacteroidetes* (Kabisch *et al.*, 2014), suggesting that such PULs are widespread and of global importance. This underpins the importance of laminarin as substrate in the marine realm, also in the open ocean.

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

At oligotrophic open-ocean sites algal and bacterial polysaccharides are produced in much lower amounts than at eutrophic sites. Therefore these energy-rich compounds are

particularly valuable at open ocean sites, and consequently heterotrophic bacteria exist at such sites that can consume these polysaccharides when they become available. As in other habitats, the *Bacteroidetes* seem to play a key role in such turnover of complex organic matter also in open oceans. It will be up to future systematic studies to inventory the PUL repertoire of marine *Bacteroidetes* in a comprehensive manner and to explore, which individual PULs are ubiquitously distributed and thus most important in marine habitats.

2.6 Acknowledgements

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

2.7 References

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

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

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

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

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

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

Bauer, M., Kube, M., Teeling, H., Richter, M., Lombardot, T., Allers, E. et al. (2006) Whole genome analysis of the marine *Bacteroidetes* '*Gramella forsetii*' reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8**: 2201-2213.

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

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

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

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

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

Bowman, J.P. (2000) Description of *Cellulophaga algicola* sp. nov., isolated from the surfaces of Antarctic algae, and reclassification of *Cytophaga uliginosa* (ZoBell and

Upham 1944) Reichenbach 1989 as *Cellulophaga uliginosa* comb. nov. *Int J Syst Evol Microbiol* **50 Pt 5**: 1861-1868.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gómez-Pereira, P.R., Schüler, M., Fuchs, B.M., Bennke, C., Teeling, H., Waldmann, J. et al. (2012) Genomic content of uncultured *Bacteroidetes* from contrasting oceanic provinces in the North Atlantic Ocean. *Environ Microbiol* **14**: 52-66.

González, J.M., Fernández-Gómez, B., Fernàndez-Guerra, A., Gómez-Consarnau, L., Sánchez, O., Coll-Lladó, M. et al. (2008) Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (*Flavobacteria*). *Proc Natl Acad Sci U S A* **105**: 8724-8729.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Meyer, F., Goesmann, A., McHardy, A.C., Bartels, D., Bekel, T., Clausen, J. et al. (2003) GenDB--an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31**: 2187-2195.

Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., and Dideberg, O. (2001) The kappa-carrageenase of P. carrageenovora features a tunnel-shaped active site: a novel insight in the evolution of Clan-B glycoside hydrolases. *Structure* **9**: 513-525.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Van Geel-Schutten, G.H., Faber, Smit, Bonting, Smith, Ten Brink, B. et al. (1999) Biochemical and structural characterization of the glucan and fructan

exopolysaccharides synthesized by the Lactobacillus reuteri wild-type strain and by mutant strains. *Appl Environ Microbiol* **65**: 3008-3014.

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

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

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

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

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

Chapter III: Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms

Published in *eLife*

Hanno Teeling[1]*, Bernhard M. Fuchs[1]*, Christin M. Bennke[1,2], Karen Krüger[1], Meghan Chafee[1], Lennart Kappelmann[1], Greta Reintjes[1], Jost Waldmann[1], Christian Quast[1], Frank Oliver Glöckner[1], Judith Lucas[3], Antje Wichels[3], Gunnar Gerdts[3], Karen H. Wiltshire[4], Rudolf I. Amann[1†]

[1] Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany
[2] current address: Leibniz Institute for Baltic Sea Research, Section Biology, Seestraße 15, 18119 Rostock-Warnemünde, Germany
[3] Alfred Wegener Institute for Polar and Marine Research, Biologische Anstalt Helgoland, 27483 Helgoland, Germany
[4] Alfred Wegener Institute for Polar and Marine Research, Hafenstraße 43, 25992 List/Sylt, Germany

* These authors contributed equally to this work.

Contribution to the manuscript:

Experimental concept and design	5%
Acquisition of experimental data	5%
Data analysis and interpretation	10%
Preparation of figures and tables	10%
Drafting of manuscript	10%

This manuscript represents the main result of the 3 year COGITO project. Fifteen authors collaborated on the processing and analysis of an immense dataset. The percental contributions therefore correspond to extensive work investment.

† Corresponding author:

Rudolf I. Amann, Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, e-mail: ramann@mpi-bremen.de, phone: +49 421 2028 930

E-mail addresses and telephone numbers of all authors:

Hanno Teeling	hteeling@mpi-bremen.de	+49 421 2028 976
Bernhard M. Fuchs	bfuchs@mpi-bremen.de	+49 421 2028 935
Christin M. Bennke	christin.bennke@io-warnemuende.de	+49 381 5197 3437
Karen Krüger	kkrueger@mpi-bremen.de	+49 421 2028 545
Meghan Chafee	mchafee@mpi-bremen.de	+49 421 2028 545
Lennart Kappelmann	lkappelm@mpi-bremen.de	+49 421 2028 545
Greta Reintjes	greintje@mpi-bremen.de	+49 421 2028 943
Jost Waldmann	jwaldman@mpi-bremen.de	+49 421 2208 253
Christian Quast	cquast@mpi-bremen.de	+49 421 2028 980
Frank Oliver Glöckner	fgloeckn@mpi-bremen.de	+49 421 2020 970
Judith Lucas	Judith.Lucas@awi.de	+49 4725 819 3200
Antje Wichels	awichels@awi-bremerhaven.de	+49 4725 819 3257
Gunnar Gerdts	Gunnar.Gerdts@awi.de	+49 4725 819 3245
Karen H. Wiltshire	Karen.Wiltshire@awi.de	+49 4651 956 4112
Rudolf I. Amann	ramann@mpi-bremen.de	+49 421 2028 930

3.1 Abstract

A process of global importance in carbon cycling is the remineralization of algae biomass by heterotrophic bacteria, most notably during massive marine algae blooms. Such blooms can trigger secondary blooms of planktonic bacteria that consist of swift successions of distinct bacterial clades, most prominently members of the *Flavobacteriia*, *Gammaproteobacteria* and the alphaproteobacterial *Roseobacter* clade. We investigated such successions during spring phytoplankton blooms in the southern North Sea (German Bight) for four consecutive years. Dense sampling and high-resolution taxonomic analyses allowed the detection of recurring patterns down to the genus level. Metagenome analyses also revealed recurrent patterns at the functional level, in particular with respect to algal polysaccharide degradation genes. We therefore hypothesize that even though there is substantial inter-annual variation between spring phytoplankton blooms, the accompanying succession of bacterial clades is largely governed by deterministic principles such as substrate-induced forcing.

3.2 Introduction

Pelagic zones of the world's oceans seemingly constitute rather homogenous habitats, however, they feature enough spatial and temporal variation to support a large number of species with distinct niches. This phenomenon has been termed 'paradox of the plankton' by G. Evelyn Hutchinson (Hutchinson, 1961). Interactions within planktonic microbial communities are manifold and complex (see Amin *et al.*, (2012) and Worden *et al.* (2015) for reviews). Still, planktonic microbial communities are simple in comparison to benthic or terrestrial soil communities and thus particularly suitable for the study of microbial community composition dynamics. In recent years, continuous biodiversity studies at long-term sampling stations have started to reveal discernible deterministic patterns within

marine microbial plankton communities (see Fuhrman *et al.* (2015) for a recent review). This is particularly true for less dynamic oligotrophic oceanic regions that are dominated by members of the alphaproteobacterial *Pelagibacteriaceae* (SAR11 clade) and the cyanobacterial *Prochlorococcaceae* (*Prochlorococcus marinus*). By contrast, more dynamic eutrophic coastal regions are often subject to frequent system perturbations and thus seldom in a state of equilibrium. This can lead to apparently stochastic changes in bacterioplankton community composition. To capture recurrence of biodiversity patterns in such coastal areas, sampling must occur at the order of weekly to sub-weekly time scales over multiple years. Owing to the lack of such intensively sampled long-term time series data, our current understanding of the extent and predictability of recurring microbial biodiversity patterns for such marine habitats is still limited.

A particularly important connection in the marine carbon cycle exists between marine microalgae as primary producers and heterotrophic bacteria that feed on algal biomass. Global photosynthetic carbon fixation is estimated to exceed 100 Gigatons yearly, of which marine algae contribute about half (Falkowski *et al.*, 1998; Field *et al.*, 1998; Sarmento and Gasol, 2012). Planktonic uni- to pluricellular algae such as diatoms, haptophytes, and autotrophic dinoflagellates are the most important marine primary producers. Diatoms alone are estimated to contribute 20-40% to global carbon fixation (Nelson *et al.*, 1995; Mann, 1999; Armbrust, 2009).

Primary production by planktonic microalgae differs from primary production by sessile macroalgae or land plants as it is much less constant, but culminates in blooms that are often massive, as occurs worldwide during spring blooms from temperate to polar regions. These blooms are highly dynamic phenomena that are time-limitd by nutrients, predator grazing and viral infections. Bloom termination results in a short-lived massive release of

algal organic matter that is consumed by dedicated clades of heterotrophic bacterioplankton. This trophic connection leads to synchronized blooms of planktonic bacteria during phytoplankton blooms, as has been described in various studies (Bell and Kuparinen, 1984; Niu *et al.*, 2011; Tada *et al.*, 2011; Teeling *et al.*, 2012; Yang *et al.*, 2015; Tan *et al.*, 2015).

The activities of these heterotrophic bacteria impact the proportion of algal biomass that is directly mineralized and released back into the atmosphere mostly as carbon dioxide, and the algae-derived biomass that sinks out to the bottom of the sea as carbonaceous particles. These are further remineralized by particle-associated bacteria while sinking and by benthic bacteria when reaching the sediment, even in the deep sea (e.g. Ruff *et al.*, 2014). The remainder is buried for a long time as kerogen and forms the basis for future oil and gas reservoirs. The ratio between bacterial mineralization and burial of algae-derived organic matter thus has a profound influence on the atmospheric carbon dioxide concentration (Falkowski *et al.*, 1998). However, the bulk of bacteria during phytoplankton blooms are free-living and not attached to particles or algae. These bacteria play a pivotal role in the mineralization of algae-derived non-particulate dissolved organic matter (DOM).
The bacterial clades that respond most to phytoplankton blooms belong to the classes *Flavobacteriia* (phylum *Bacteroidetes*) and *Gammaproteobacteria,* and the *Roseobacter* clade within class *Alphaproteobacteria* (Buchan *et al.*, 2014). This response is typically not uniform, but consists of a series of distinct clades that bloom one after another. In the year 2009 we investigated the response of bacterioplankton to a diatom-dominated spring phytoplankton bloom in the German Bight (Teeling *et al.*, 2012). Within the free-living bacteria (0.2 to 3 µm) we observed a swift succession of bacterial clades that was dominated by *Flavobacteriia* and *Gammaproteobacteria*, with consecutively blooming *Ulvibacter* (*Flavobacteriia*), *Formosa* (*Flavobacteriia*), *Reinekea* (*Gammaproteobacteria*),

Polaribacter (*Flavobacteriia*) genera and SAR92 (*Gammaproteobacteria*) as prominent clades.

Using time-series metagenome and metaproteome analyses we demonstrated that the substrate-spectra of some of these clades were notably distinct. The succession of bacterioplankton clades hence constituted a succession of distinct gene function repertoires, which suggests that changes in substrate availability over the course of the bloom were among the forces that shaped the bacterioplankton community. Dominance of bottom-up over top-down control is assumed to be characteristic for the initial phases of spring phytoplankton blooms. After winter, inorganic nutrients are aplenty and the overall abundance of microbes is low. When suitable temperature and sunlight conditions are met in spring, algae and subsequently bacteria can enter an almost unrestricted proliferation. In contrast, predators such as flagellates, protists and zooplankton can only start proliferating when their food sources are available in larger numbers. Hence, top-down control by predation sets in only during later bloom phases. This situation is distinct from summer and fall phytoplankton blooms.

Pronounced differences between blooming clades were found in the gene frequencies and protein expression profiles of transporters and carbohydrate-active enzymes (CAZymes; (Cantarel *et al.*, 2009; Lombard *et al.*, 2014)), such as glycoside hydrolase (GH), polysaccharide lyase (PL), carbohydrate esterase (CE), or carbohydrate-binding module (CBM) containing genes. The latter indicates a pronounced niche partitioning with respect to algal polysaccharide degradation. Marine algae produce large quantities of distinct polysaccharides, e.g. storage, cell matrix and cell wall constituents, or as part of extracellular transparent exopolymer particles (TEP). It has been recently shown that in particular *Flavobacteriales* and *Rhodobacterales* respond to TEP availability (Taylor *et al.*,

2014). The diversity of algal polysaccharides is too high for a single bacterial species to harbor all the genes required for the complete degradation of all naturally occurring variants. Thus, polysaccharide-degrading bacteria specialize on dedicated subsets of polysaccharides, which is why the decomposition of algal polysaccharides during and after algal blooms is a concerted effort among distinct bacterial clades with distinct glycan niches (e.g. Xing *et al.* (2015)).

In this study we provide evidence that the succession of bacterioplankton clades that we reported for the 2009 North Sea spring phytoplankton bloom re-occurred during the spring blooms from 2010 to 2012. We tested whether the bacterioplankton clades and their associated CAZyme repertoires differ from year to year or exhibit recurrent patterns. We analyzed spring bacterioplankton community composition via 16S rRNA catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) and 16S rRNA gene tag sequencing, as well as gene function repertoires by deep metagenome sequencing. Our efforts have culminated into the as of yet highest resolved dataset capturing the response of planktonic bacteria to marine spring phytoplankton blooms and have allowed identification of recurring patterns that might ultimately lead to an explanatory model for bacterioplankton succession dynamics during spring algae blooms.

3.3 Materials and methods

3.3.1 Phytoplankton and physicochemical data

Physicochemical parameters (supplementary file 1) and phytoplankton data (supplementary file 2) were assessed in subsurface water on a weekday basis as part of the Helgoland Roads LTER time series. Details on the acquisition of these data have been described previously (Teeling *et al.*, 2012). The Helgoland Roads time series is accessible

via the public database Pangaea (http://www.pangaea.de).

3.3.2 Bacterioplankton

Sampling of bacterioplankton was carried out as described previously (Teeling *et al.*, 2012). In brief, surface seawater samples were taken at the long-term ecological research station 'Kabeltonne' (54° 11.3' N, 7° 54.0' E) at the North Sea island Helgoland using small research vessels (http://www.awi.de/en/infrastructure/ships/aade_and_diker) and processed in the laboratory of the Biological Station Helgoland within less than two hours after sampling.

Biomass of free-living bacteria for DNA extraction was harvested on 0.2 µm pore sized filters after pre-filtration with 10 µm and 3 µm pore sized filters to remove large debris and particle-associated bacteria. By contrast, cells for microscopic visualization methods were first fixed by addition of formaldehyde to sampled seawater, which was then filtered directly onto 0.2 µm pore sized filters. All filters were stored at -80 °C until further use.

3.3.3 Microscopy: total cell counts, CARD-FISH

Assessment of absolute cell numbers and bacterioplankton community composition was carried out as described previously (Thiele *et al.*, 2011). To obtain total cell numbers, DNA of formaldehyde fixed cells filtered on 0.2 µm pore sized filters was stained with 4',6-diamidino-2-phenylindole (DAPI). Fluorescently labeled cells were subsequently counted on filter sections using an epifluorescence microscope. Likewise, bacterioplankton community composition was assessed by catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) of formaldehyde fixed cells on 0.2 µm pore sized filters.

DAPI and CARD-FISH cell counts are summarized in supplementary file 3 and the corresponding probes in supplementary file 4.

3.3.4 16S rRNA V4 gene tag sequencing

Surface seawater samples were collected on bi-monthly to bi-weekly time scales from January 2010 to December 2012 at Helgoland roads. 500 ml of each sample were subjected to fractionating filtration as described above using 10, 3 and 0.2 µm pore size polycarbonate membrane filters (Millipore, Schwalbach, Germany). DNA of the 0.2-3 µm fraction was extracted from filters as described previously (Sapp *et al.*, 2007) and quantified using the Invitrogen (Carlsbad, CA, USA) Quant-iT™ PicoGreen® dsDNA reagent as per manufacturer's instructions. Concentrations ranged from <1 to 20 µg DNA/ml.

50 µl aliquots of each sample were pipetted into 96-well plates and sent to the Department of Energy (DOE) Joint Genome Institute (JGI, Walnut Creek, CA, USA) for amplification and sequencing as follows: Sample prep was done on a PerkinElmer (Waltham, MA, USA) Sciclone NGS G3 Liquid Handling Workstation capable of processing 96 plate-based samples in parallel, utilizing the 5 PRIME (Gaithersburg, MD 20878, USA) HotMasterMix amplification kit and custom amplification primers targeting the V4 region of the 16S rRNA gene using 515F (5' GTGCCAGCMGCCGCGGTAA 3') and 806R (5' GGACTACHVGGGTWTCTAAT 3') (Caporaso *et al.*, 2011). Primers also contained Illumina adapter sequences and a barcode index. PCR reactions were set up in 75 µl with 1x HotMasterMix (5 PRIME) with final concentrations of 0.4 µg/µl BSA and 0.2 µM of each primer. This volume was split into triplicate 25 µl reactions for independent amplification and then pooled to reduce PCR bias. Prepared amplicon libraries were normalized, multiplexed into a single pool per plate and quantified using the KAPA Biosystems

(Wilmington, MA, USA) next-generation sequencing library qPCR kit on a Roche (San Francisco, CA, USA) LightCycler 480. Libraries were sequenced on an Illumina (San Diego, CA, USA) MiSeq sequencer using the Reagent Kit v3 and 2x250 bp chemistry. The resulting sequences are available from the DOE-JGI GOLD database (Reddy *et al.*, 2015) as part of the COGITO project (Gp0056779) and from the NCBI short read archive (SRA) (SRA278189).

3.3.5 16S rRNA gene tag analysis

Roche 454 16S rRNA gene tags from 2009 MIMAS (Microbial Interactions in Marine Systems) project (Teeling *et al.*, 2012) were reanalyzed for comparison with the Illumina-based COGITO extension project from subsequent years 2010-2012. The 2009 datasets was generated using the primers Bakt_314F (5' CCTACGGGNGGCWGCAG 3') and Bakt_805R (5’ GACTACHVGGGTATCTAATCC 3') (Herlemann *et al.*, 2011). The forward primers of both datasets target distinct, but the reverse primers target the same region. Hence only those 454 reads sequenced from the 805 direction were reanalyzed for comparison. For 2010-2012, raw MiSeq paired-end reads (2x250 bp) were merged and filtered using illumina-utils (https://github.com/meren/illumina-utils) to retain only read pairs without mismatches in the overlapping regions. These high-quality Illumina tags and the 454 tags were then processed separately but with the same methods via the SILVAngs pipeline (Quast *et al.*, 2013), which includes additional quality filtering steps via alignment as well as length, ambiguity and homopolymer filters. Sequences were dereplicated at 100% identity and then globally clustered at 98%. Representative OTUs were classified to genus level against the SILVA (Quast *et al.*, 2013) v119 database using BLAST with a similarity threshold = (sequence identity + alignment coverage) / 2 >=93%. The SAR92 clade was reclassified according to SILVA v123. Reads were mapped against representative OTUs to obtain final abundance counts. For the purpose of this study,

OTUs were collapsed based on shared taxonomy no higher than the genus level.

For MIMAS samples we retained a total of 110,995 454 reads across 7 samples with an average of 16,000 per sample. After SILVAngs quality filtering, 110,866 remained for clustering. 6,102 representative OTUs were identified and 107,708 total sequences were assigned to a relative in the database during classification within the 93% similarity threshold. The final abundance matrix collapsed on shared taxonomic classification contained 500 unique taxa.

In total, 20,869,432 paired raw MiSeq reads were obtained across 142 samples from 2010-2012 COGITO samples. 15,016,350 merged reads with no mismatches in the overlapping region were retained with an average of 106,000 per sample. Reads were randomly sub-sampled to 40,000 tags per sample to reduce computational demands. In total 6,120,000 tags were submitted to the SILVAngs pipeline. After additional quality filtering, 6,116,021 sequences were clustered at 98% and the resulting 935,006 representative OTUs were classified. A total of 5,676,259 sequences were assigned to a relative in the database within the 93% similarity threshold. The final abundance matrix collapsed on shared taxonomy no higher than the genus level contained 1,995 unique taxa (Supplementary file 5).

3.3.6 Metagenome sequencing

Total community DNA of 2009 samples (02/11/09; 03/31/09; 0407/09; 04/14/09; 06/16/09) was sequenced on the 454/Roche FLX Ti platform as described previously (Teeling *et al.*, 2012). Metagenome sequencing of 2010-12 samples (03/03/2010; 04/08/10; 05/04/10; 05/18/10; 03/24/11; 04/28/11; 05/26/11; 03/08/12; 04/16/12; 05/10/12) was performed at the DOE Joint Genome Institute on the Illumina HiSeq2000 platform. Libraries were

created from 100 ng environmental DNA per sample that was sheared to 270 bp using a Covaris E210 (Covaris, Woburn, MA, USA) and size selected using SPRI beads (Beckman Coulter, Indianapolis, IN, USA). The fragments were treated with end-repair, A-tailing, and ligation of Illumina compatible adapters (IDT, Coralville, IA, USA) using the KAPA-Illumina library creation kit. The libraries were quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq PE Cluster Kit v3, and Illumina's cBot instrument to generate a clustered flowcell for sequencing. Sequencing was performed on the Illumina HiSeq2000 sequencer using TruSeq SBS sequencing Kits, v3, following a 2x150 bp indexed run recipe.

Raw reads were screened against Illumina artifacts with kmer size of 28, step size of 1. Reads were subsequently trimmed from both ends using a minimum quality cutoff of 3; reads with three or more N's or with average quality score <Q20 were removed. In addition, reads <50 bp were removed. The remaining quality-filtered Illumina reads were assembled using SOAPdenovo v1.05 (Luo *et al.*, 2012) at a range of kmers (81, 85, 89, 93, 97, 101) with default settings (options: -K 81 -p 32 -R -d 1). Contigs generated by each assembly (6 total contig sets), were de-replicated using JGI in house Perl scripts. Contigs were then sorted into two pools based on length. Contigs <1,800 bp were re-assembled using Newbler (Life Technologies, Carlsbad, CA, USA) in attempt to generate larger contigs (options: -tr, -rip, -mi 98, -ml 80). Contigs >1,800 bp as well as the contigs from the Newbler assembly were combined using minimus 2 (options: -D MINID=98 -D OVERLAP=80) from the AMOS package (http://sourceforge.net/projects/amos). Read depths were estimated based on read mapping with bbmap (http://bio-bwa.sourceforge.net/). The metagenome study information is available from the DOE-JGI

GOLD database (study: Gs0000079). The unassembled reads are available from the NCBI SRA (see Supplementary file 8), and the assembled and annotated metagenome datasets from the IMG/M system (Markowitz *et al.*, 2014).

3.3.7 Metagenome analysis

The DOE-JGI MAP v.4 annotation pipeline (Huntemann *et al.*, 2015) was used for initial metagenome gene prediction and annotation. The annotated metagenomes were loaded in the IMG/M system as of mid 2014, and subsequently imported into a GenDB v2.2 annotation system (Meyer *et al.*, 2003) for taxonomic classification and data mining.

All genes were searched against the NCBI non-redundant protein database (as of June 17th, 2014) using USEARCH v6.1.544 (Edgar, 2010), against the Pfam v25 database (Finn *et al.*, 2014) using HMMER v3 (Punta *et al.*, 2012), for signal peptides using SignalP v3.0 (Nielsen *et al.*, 1999) and for transmembrane helices using TMHMM v2.0c (Krogh *et al.*, 2001). CAZymes were automatically annotated based on HMMER searches against the Pfam v25 and dbCAN (Yin *et al.*, 2012) databases and BLAST (Altschul *et al.*, 1990) searches against the CAZy database (Cantarel *et al.*, 2009; Lombard *et al.*, 2014) using E-value cut-offs that were specifically adjusted for each CAZyme family (Supplementary file 11). Genes were only annotated as CAZymes when at least two of the search results were congruent, and CAZymes were only analyzed for contigs ≥500 bp.

Taxonomic classification of the metagenome sequences into taxonomically coherent bins ('taxobins') was carried out with a modified version of the Taxometer approach described in (Teeling *et al.*, 2012). Taxometer consolidates predictions of a set of individual sequence classification tools into a consensus using a weighted assessment on seven selected ranks (superkingdom, phylum, class, order, family, genus, species) of the NCBI taxonomy (http://www.ncbi.nlm.nih.gov/Taxonomy/). We combined taxonomic information inferred from (i) Pfam hits using the CARMA3 approach (Gerlach and Stoye, 2011), (ii)

BLASTp hits using the KIRSTEN approach (Teeling *et al.*, 2012; supp. data), and (iii) mapping of quality-filtered (illumina-utils; https://github.com/meren/illumina-utils/) Illumina reads to selected reference sequences. In contrast to the original Taxometer approach we omitted signature-based classification with Self-Organizing Maps and mapping of reads containing partial 16S rRNA gene sequences. The prediction tools that were used are outlined below.

We used the HMMER-based module of CARMA3 (not the BLAST-based module) that infers taxonomy of sequences by post-processing genes with HMMER3 hits to the Pfam database. The basic principle is to apply a reciprocal search technique to reduce the number of identified matches and thereby to improve taxonomic classification quality.
KIRSTEN (Kinship Relationship Reestablishment) infers taxonomy of sequences by post-processing BLASTp hits to the NCBI nr database by means of rank-based evaluations on all levels of the NCBI taxonomy with an increasing stringency from the superkingdom down to the species level. On each taxonomic level, all occurring taxa are weighted by the sum of their BLASTp bit scores. When the taxon with the highest weight exceeds an adjustable threshold, the process continues towards the next taxonomic level. The threshold increases with each taxonomic level, i.e. the algorithm becomes more critical while it progresses. For this study, we substituted BLASTp by the UBLAST module of USEARCH 6.1 (Edgar, 2010) with an E-value cutoff of E-10 and maximum hit count of 500.

SMALT (http://www.sanger.ac.uk/resources/software/smalt/) was used to map metagenome reads on a manually compiled set of 49 reference genomes and a streamlined version of the NCBI nr database. Both, the reference genomes and the sequences selected from the NCBI nr database were selected based on habitat-specific

information. This was done manually for the reference genomes and automatically for the NCBI nr database as follows: Each of the hits from the UBLAST search during KIRSTEN analysis can be associated with multiple taxa, since in nr redundant sequences from different taxa are merged. We used this information to extract all taxa associated with a given hit, then combined the taxa of all hits and finally extracted all sequences of these taxa from Genbank. This way a sample-specific streamlined subset of Genbank was generated that greatly sped up the mapping process. Only metagenome reads with at least 95% identity to any sequence in the sample-specific sequence database were used. Reads were subsequently back-mapped to contigs. Since contigs consist of many reads, this way, contigs were associated with multiple taxonomic paths. Taxonomic paths resulting from classification with reference genomes and the habitat-specific streamlined nr were concatenated. Paths representing less than 1% of the reads of a given contig were discarded.

Finally, Taxometer was used to combine all gene-based predictions from CARMA3 and KIRSTEN and the mapping results for each contig, and to infer a consensus taxonomy. Taxonomic predictions were possible for 94.6% of the contigs above 1 kbp. The results are summarized in supplementary file 9.

3.3.8 Statistical analyses

Spearman rank correlation test (Supplementary file 6) were used to test for correlations between *Bacteroidetes* clade abundances and environmental variables (chlorophyll a, temperature, salinity, silicate, phosphate, nitrate, nitrite, ammonia) and phytoplankton abundances (classes: diatoms, dinoflagellates, coccolithophorids, silicoflagellates, flagellates, ciliates, green algae; species: *Mediophyxis helysia*, *Thalassiosira nordenskioeldii*, *Chaetoceros debilis* and *C. minimus*, *Rhizosolenia styliformis*, *Chattonella*

and *Phaeocystis*). *Bacteroidetes* and phytoplankton numbers were transformed to log-scale for better comparison. Linear regressions (Supplementary file 7) were done using stepwise forward regression model by using the log transformed *Bacteroidetes* and phytoplankton abundances. All statistical analyses were performed using the software Sigma-Plot 12 (SYSTAT, Santa Clara, CA, USA).

3.4 Results

3.4.1 Sampling site characteristics

Samples were taken at Helgoland Island about 40 km offshore in the southeastern North Sea in the German Bight at the station 'Kabeltonne' (54° 11.3' N, 7° 54.0' E; Figure 1) between the main island and the minor island, Düne (German for 'dune'). Water depths at this site fluctuate from 6 to 10 m over the tidal cycle. During most of the year a westerly current transports water from the English Channel alongside the Dutch and Frisian coast to Helgoland, but waters around the island are also influenced by nutrient inputs from the rivers Weser and Elbe and from the northern North Sea (Wiltshire *et al.*, 2010). During the 2009 to 2012 study period the lowest water temperatures were measured in mid to late February (min. 2010: 1.1 °C; max. 2009: 3.4 °C), followed by a continuous increase until a peak in August (min. 2011: 18.0 °C; max. 2009: 18.7 °C) (Supplementary file 1).

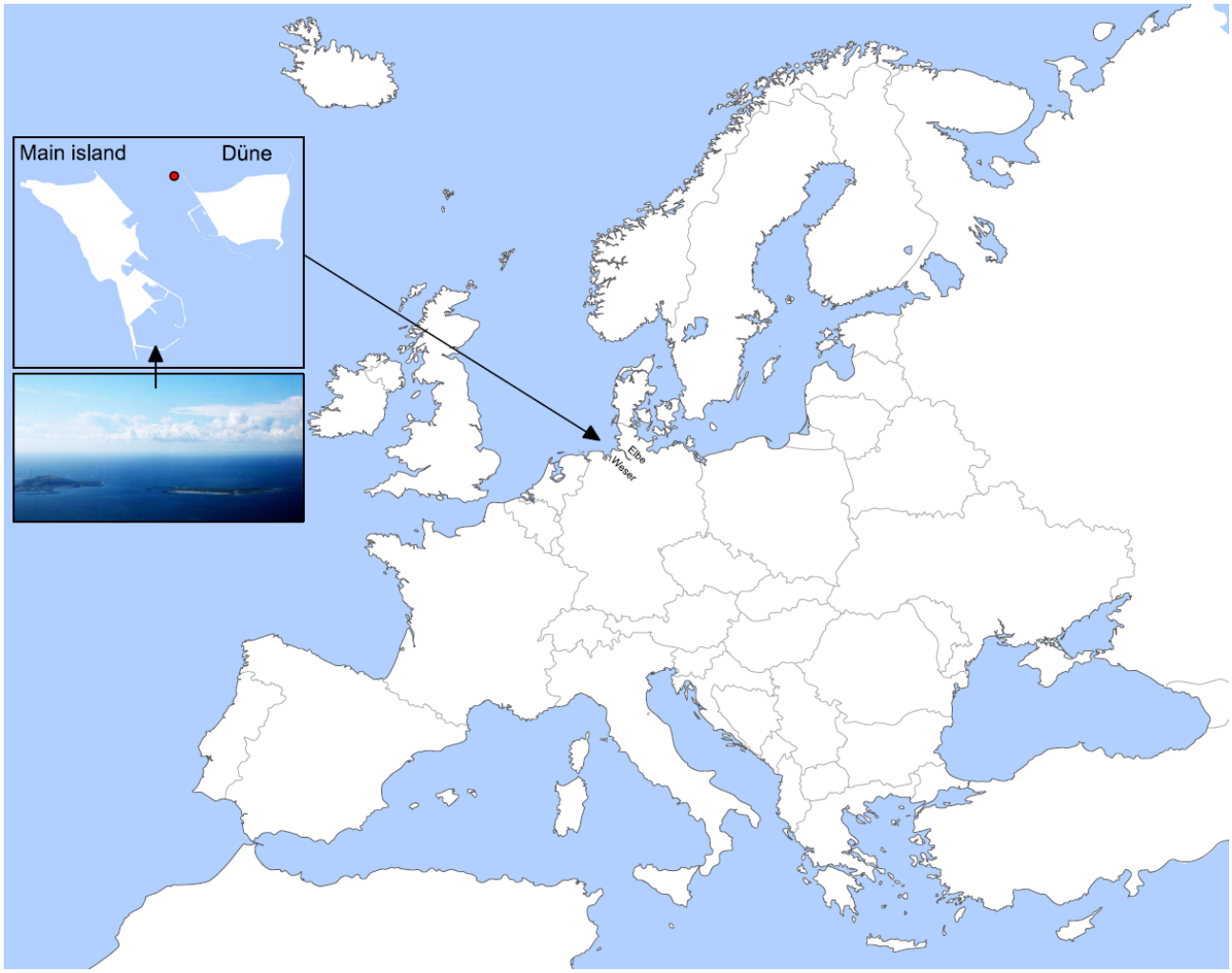


Figure 3.1: Location of Helgoland Island (ca. 40 km offshore the northern German coastline) and the long-term ecological research site 'Kabeltonne' (red circle: 54° 11.3' N, 7° 54.0' E) in the German Bight of the North Sea.

3.4.2 Phytoplankton - diversity and bloom characteristics

Spring phytoplankton blooms in the North Sea typically develop during March and reach highest intensities during April and May. The highest chlorophyll *a* concentrations are usually observed at the coastlines including the area around Helgoland Island (Figure 2). North Sea spring blooms are thus large-scale phenomena that are, however, influenced by local conditions, such as riverine inputs. At Helgoland island, spring phytoplankton blooms started around mid-March when water temperatures surpassed 3 to 5 °C (Figure 3A-H; Supplementary file 1). The diatoms *Chaetoceros debilis and Chaetoceros minimus*,

Mediopyxis helysia, *Rhizosolenia* styliformis and *Thalassiosira nordenskioeldii,* the silicoflagellate *Chattonella,* the haptophyte *Phaeocystis* and dinoflagellates dominated these blooms in terms of cell numbers (Figure 3I-L; Supplementary file 2). Relative abundances of these algae varied in no apparent order during the observed blooms, and we have yet to understand the factors that determine these variations.

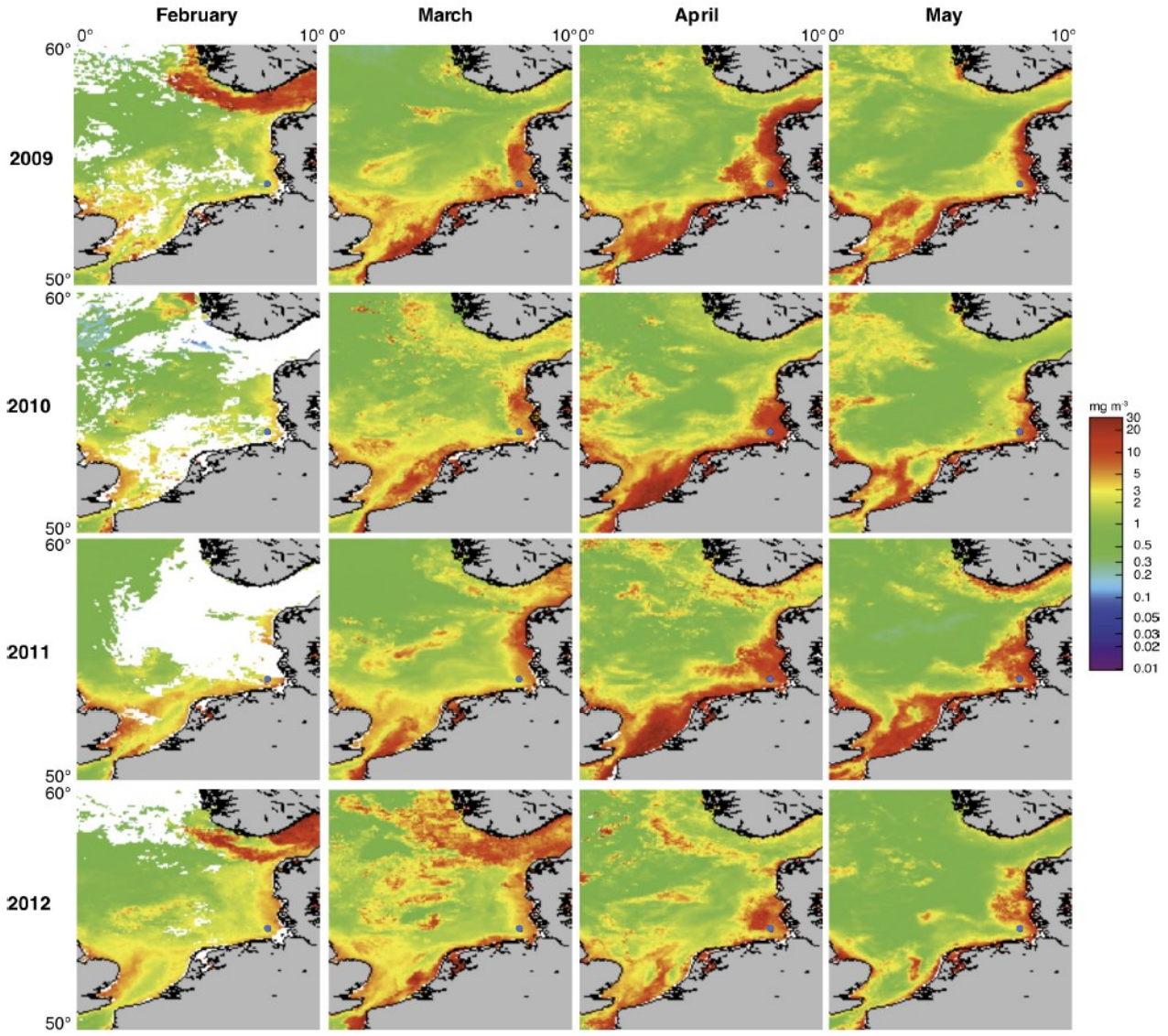


Figure 3.2: Satellite chlorophyll *a* measurements. Data are shown for the southern North Sea for the months February to May (monthly averages) of the years 2009 to 2012. Images were retrieved from the GlobColour website using the 'extended Europe area' at full resolution (1 km) as merged products of weighted averages from the following sensors: MERIS, MODIS AQUA, SeaWIFS and VIIRS. See GlobColour website for details (http://hermes.acri.fr). The position of Helgoland Island is indicated by a blue dot.

The sizes of the dominant algae taxa are different, with *Chaetoceros minimus* and in particular *Phaeocystis* spp. featuring the smallest cells and *Mediopyxis helysia* and *Rhizosolenia styliformis* featuring the largest cells. Spherical *Phaeocystis* spp. cells for example have estimated biovolumes of ~50 to 250 μm^3, whereas elipsoid cylindrical *Mediopyxis helysia* cells have a biovolume of ~ 82,000 μm^3 and for *Rhizosolenia styliformis* even a biovolume of ~282,000 μm^3 has been reported (Olenina, 2006; Loebl *et al.*, 2013). Considering biomass, the blooms were largely dominated by the diatoms *T. nordenskioeldii* and *M. helysia* and the silicoflagellate *Chattonella*. Blooms of these three algae were bimodal in all years with dominance of first *T. nordenskioeldii* followed by *Chattonella* in 2009 (Figure 3I), *T. nordenskioeldii* followed by *M. helysia* in 2010 (Figure 3J), *M. helysia* followed by *Chattonella* in 2011 (Figure 3K) and a pronounced bimodal bloom of *Chattonella* species in 2012 (Figure 3L). All blooms were accompanied by a notable decrease of silicate (Figure 3A-D; Supplementary file 1), which diatoms use for frustule formation (see Yool and Tyrrell (2003) for the controlling effect of silicate on diatom abundance).

Bloom maximum intensities decreased from 2009 to 2012 with chlorophyll *a* maxima (measured by fluorescence) reaching 28 mg m^{-3} (day 82), 18 mg m^{-3} (day 125), 15 mg m^{-3} (day 116), and 11 mg m^{-3} (day 114) in each respective year (Figure 3A-D; Supplementary file 1). Maximum bacterioplankton cell counts were observed either close to or after the chlorophyll maxima (Supplementary file 3). In 2009, the bacterioplankton peaked at 3.5 $x10^6$ cells ml^{-1} (day 118), 36 days after the Chl *a* maximum. In 2010 the peak in Chl *a* was broader and only four days ahead of the bacterioplankton peak abundance of 2.4 $x10^6$ cells ml^{-1} (day 129). In 2011 the Chl *a* peak was only two days ahead of the bacterioplankton peak abundance of 2.3 $x10^6$ cells ml^{-1} (day 118), whereas it was nine days ahead in 2012, where the bacterioplankton peaked at 3.2 $x10^6$ cells ml^{-1} (day 123).

3.4.3 Bacterioplankton - diversity and bloom characteristics

DAPI and CARD-FISH cell staining (Supplementary files 3 and 4) showed that SAR11 dominated the bacterioplankton community in winter, but with the onset of each spring bloom relative abundances of *Bacteroidetes* followed by *Gammaproteobacteria* increased and finally surpassed those of the SAR11 (Figure 3M-P). *Bacteroidetes* reached higher maximum relative abundances than *Gammaproteobacteria,* 40% (2012) to 60% (2010) as compared to 10% (2012) to 30% (2010), respectively.

Bacteroidetes genera *Polaribacter*, *Formosa*, and VIS6, a genus-level clade within the family *Cryomorphaceae* (Gómez-Pereira *et al.*, 2010), peaked each year with relative abundances well above 5%, reaching relative abundances of up to 25% sometimes within less than a week (Figure 3Q-T). *Ulvibacter* reached similar peak abundances with the exception of 2012, where this genus never surpassed 2% and ranged below 1% most of the time. Within *Gammaproteobacteria* the genus-level SAR92 clade responded notably in all years increasing from background levels below 1% to relative abundance of 8% to 16%. Members of the *Alteromonadales* families *Alteromonadaceae* and *Colwelliaceae* bloomed in three (2010: 8%; 2011: 6%; 2012: 7%), and genus *Reinekea* in two (2009: 16%; 2010: 5%) of the years (Figure 3U-X). Some less abundant, but nevertheless recurrent taxa included the genus *Balneatrix* within *Gammaproteobacteria* with relative abundances up to 2% (Figure 3U-X). Minor recurring groups of *Bacteroidetes* (Figure 3Y-β) included the NS3a marine group (9% in 2009 and 6% in 2011), the genus-level VIS1 clade within the NS5 marine group detected before the Chl *a* peaks of 2011 (7%) and 2012 (5%), and the *Cytophagia* clade *Marinoscillum* that reached 1-3% abundance most years after initial blooms.

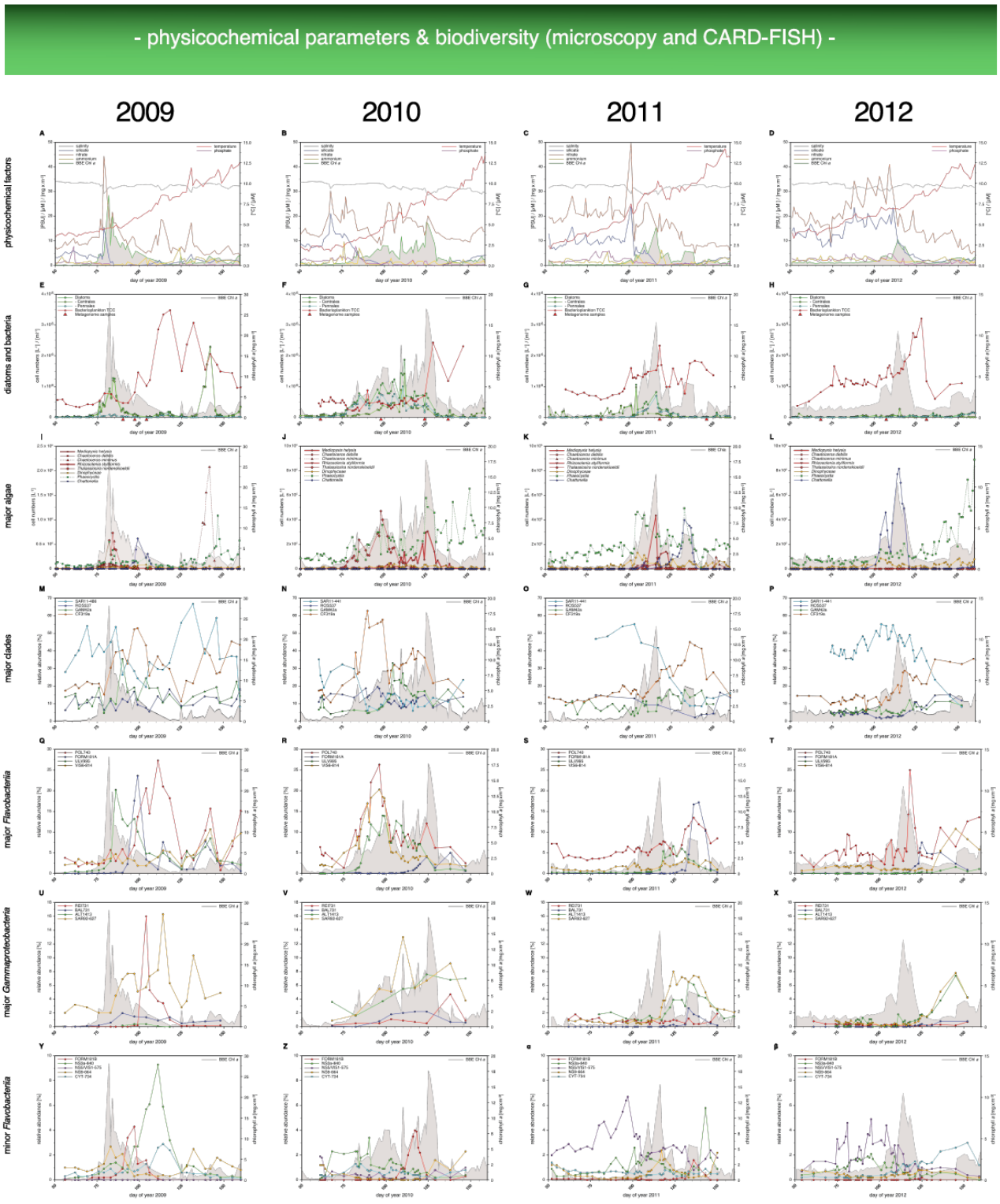


Figure 3.3: Physicochemical parameters, phytoplankton composition and bacterioplankton composition as assessed by CARD-FISH.

Sampling: **Surface seawater samples were taken at the North Sea island Helgoland between the main island and the minor island 'Düne' (station 'Kabeltonne', 54°11'03''N, 7°54'00''E) using small research vessels (http://www.awi.de/en/expedition/ships/more-ships.html) and**

processed in the laboratory of the Biological Station Helgoland within less than two hours after sampling. Cells for microscopic visualization methods were first fixed by the addition of formaldehyde to sampled seawater, which was then filtered directly onto 0.2 mm pore sized filters.

***Physicochemical and phytoplankton data*: Physicochemical parameters and phytoplankton data were assessed in subsurface water on a weekday basis as part of the Helgoland Roads LTER time series as described in Teeling et al. (2012). The Helgoland Roads time series is accessible via the public database Pangaea (http://www.pangaea.de) and can be used to assess long-term changes of the North Sea pelagic ecosystem. Left-hand side legends correspond to ordinates on the left, and right-hand side legends to ordinates on the right. A-D: Physicochemical measurements including measurements of BBE Chl *a* (chlorophyll *a* fluorescence by algal group analyzer sensor). Left ordinate: salinity [PSU], silicate [mM], nitrate [mM], ammonium [mM] and chlorophyll *a* [mg/m3]; right ordinate: temperature [°C] and phosphate [mM]. E-H: Counts of the diatom groups. I-L: Microscopic cell counts of the most abundant algae genera (red: diatoms; orange: dinoflagellates: green: haptophytes; blue: silicoflagellates). Algae with large cells and thus large biovolumes are depicted by bold solid lines and algae with small cells are represented by dotted lines. *Rhizosolenia styliformis* and *Mediopyxis helysia* feature large cells, whereas *Chaetoceros minimus* and in particular *Phaeocystis* species feature small cells. The latter typically reaches lengths of below 10 mm and *Phaeocystis* spp. biovolumes therefore typically are hundreds to thousands fold smaller than those of *R. styliformis* and *M. helyisa* cells (Olenina, 2006; Loebl et al., 2013). Physicochemical data are summarized in Supplementary file 1, and data on the major phytoplankton clades in Supplementary file 2.**

***Total cell counts and CARD-FISH of bacterioplankton*: E-H: TCC (total cell counts); red triangles depict sampling of metagenomes. M-b: Recurrent bacterioplankton clades as assessed by CARD-FISH (Catalyzed Reporter Deposition-Fluorescence in situ Hybridization) with the following probes: M-P (major bacterial groups): SAR11-486 and**

SAR11-441: alphaproteobacterial SAR11-clade; ROS537: alphaproteobacterial *Roseobacter* clade; GAM42a: *Gammaproteobacteria*; CF319a: *Bacteroidetes*. Q-T (major *Flavobacteriia* clades): POL740: genus *Polaribacter*; FORM181A: genus *Formosa*; ULV995: genus *Ulvibacter*; VIS6-814: genus-level clade VIS6 within the family *Cryomorphaceae-Owenweeksia*; U-X (major *Gammaproteobacteria* clades): REI731: genus *Reinekea*; BAL731: genus *Balneatrix*; ALT1413: families *Alteromonadaceae* and *Colwelliaceae*; SAR92-627: genus-level clade SAR92. Y-b (minor *Bacteroidetes* clades): FORM181B: species-specific for *Formosa* sp. Hel1_33_131; NS3a-840: NS3 marine group; NS5/VIS1-575: VIS1 genus-level clade within the NS5 marine group; NS9-664: NS9 marine group; CYT-734: *Cytophagia* clade *Marinoscillum*. Total and CARD-FISH cell counts are summarized in Supplementary file 3 and the corresponding probes in Supplementary file 4.

For high resolution image, go to https://elifesciences.org/articles/11888#fig3

We used complementary 16S rRNA gene tag sequencing for the detection of bacterioplankton clades that were not recovered by CARD-FISH probes (Supplementary file 5). Relative proportions of 16S tags from distinct clades correlated for the most part those inferred from CARD-FISH cell counts (Figure 4A-P), but members of SAR11 were substantially underreported - a known limitation of the 806R primer used in v4 amplification for 2010 to 2012 samples (Apprill *et al.*, 2015). Additional abundant clades detected in the 16S amplicon data comprised the *Flavobacteriia* genus *Tenacibaculum* (Figure 4U-X) that bloomed in 2010 (read frequencies of max. ~12%) and 2011 (max. ~5%). Within *Gammaproteobacteria*, clades with read frequencies ≥5% in at least one year comprised the genera *Aeromonas*, *Glaciecola*, *Pseudoalteromonas*, *Pseudomonas*, *Psychrobacter* and the SAR86 and ZD0405 clades (Figure 4Q-T). Within the alphaproteobacterial *Rhodobacteriaceae*, high abundances of '*Candidatus* Planktomarina temperata' (DC5-80-3 lineage) and the NAC11-7 clade were detected reaching 6-21% and 7-19% of the tag data, respectively (Figure 4U-X). Also within *Alphaproteobacteria* the genus *Sulfitobacter* peaked with a read frequency of ~7% in 2010 (Figure 4V), and within *Betaproteobacteria* the order *Methylophilales* (dominated by OM43 clade members) was detected with high relative abundances of up to ~10% before blooms, which decreased with bloom progression. *Verrucomicrobia* (Figure 4U-X) were detected with decreasing peak read frequencies of 7.7% (2010), 5.0% (2011) and 2.9% (2012). This decrease corresponds to decreasing bloom intensities, which supports a proposed role of *Verrucomicrobia* in polysaccharide decomposition (e.g. Martinez-Garcia *et al.*, 2012).

Within *Bacteroidetes*, *Gammaproteobacteria* and *Rhodobacterales* a total of eleven clades peaked during at least two of the four spring blooms with relative cell abundances or, for those clades that were not assessed by CARD-FISH, relative read frequencies ≥5%.

These were six *Flavobacteriia* clades (*Formosa*, *Polaribacter*, NS3a marine group, *Tenacibaculum*, *Ulvibacter*, VIS6 clade *Cryomorphaceae*), three *Gammaproteobacteria* clades (*Alteromonadaceae/Colwelliaceae*, *Reinekea* and SAR92), and two *Roseobacter* clades (DC5-80-3 and NAC11-7).

Each year a succession was observed within the *Flavobacteriia* and *Gammaproteobacteria* clades. The succession in the *Flavobacteriia* was more pronounced than in the *Gammaproteobacteria*, but the sequence of clades varied. Spearman rank correlation analyses revealed that the abundances of the most prominent *Flavobacteriia* clades were for the most part correlated with multiple algae groups and physiochemical factors (Supplementary file 6). According to linear regression analyses the strongest abiotic predictors were temperature, salinity, silicate and nitrate, and the strongest biotic predictors were *Phaeocystis* spp. haptophytes, *Rhizosolenia* spp., *Chaetoceros debilis*, and *Chaetoceros minimus* diatoms and the silicoflagellate *Chattonella* (Supplementary file 7). It should be noted though that linear regressions were computed based on log-transformed abundance data and not algae volumes (which were not measured). Thus the influence of the rather small cell-sized algae such as *Chaetoceros minimus* is likely overestimated. Such limitations notwithstanding it is noteworthy that in no case a simple one-to-one relationship between specific algae and specific bacterioplankton groups was detected. The strongest significant ($p < 0.05$) correlations were obtained for the *Ulvibacter* clade that was positively correlated with diatoms and haptophytes and negatively correlated with silicoflagellates. Further results comprised an opposite trend for the VIS1 clade of the NS5 marine group, and a correlation of *Polaribacter* and *Chattonella* abundances (see Supplementary file 6 for details).

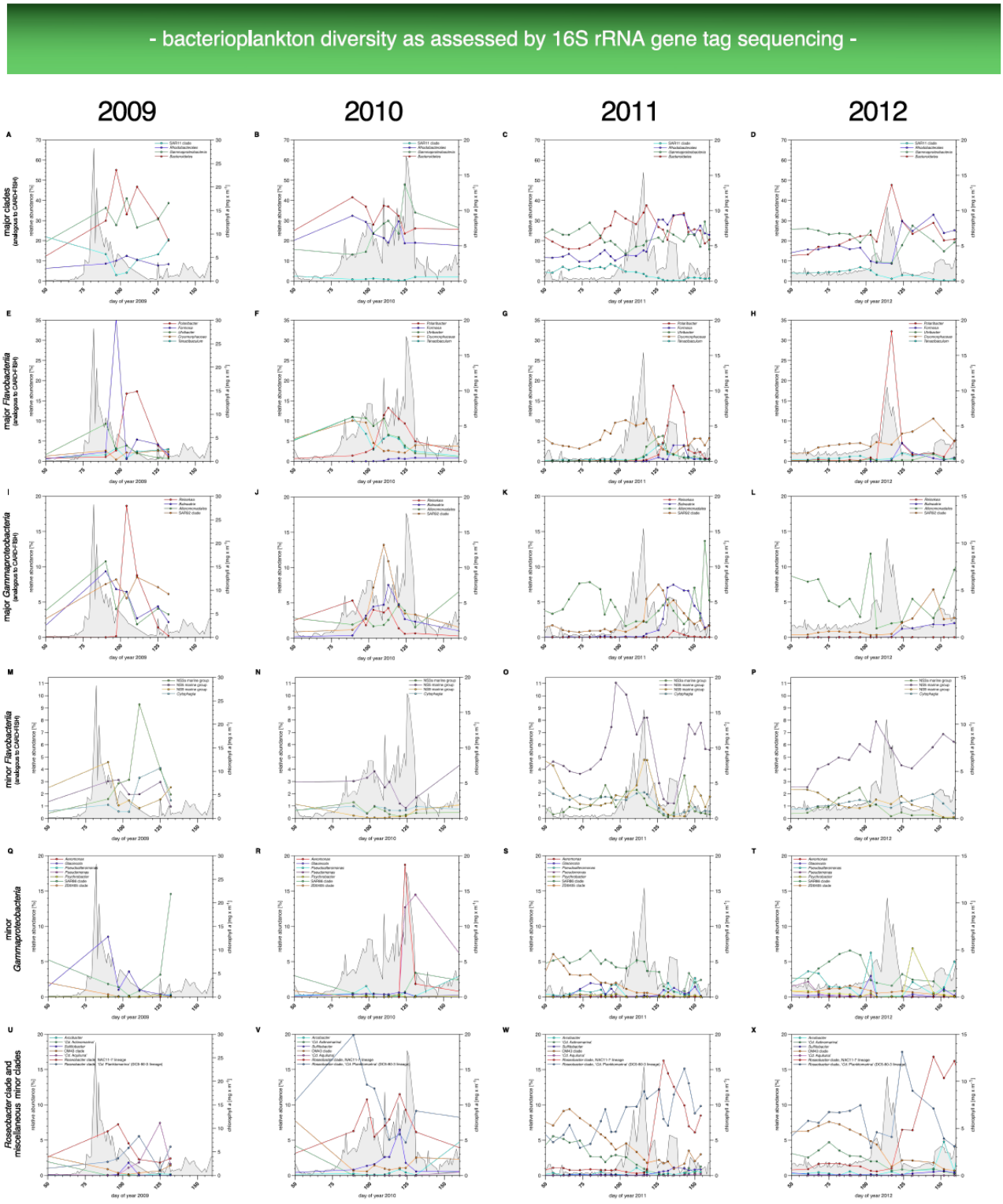


Figure 3.4: Bacterioplankton diversity as assessed by 16S rRNA gene tag sequencing.

***Sampling*: Surface seawater samples were taken at the North Sea island Helgoland between the main island and the minor island 'Düne' (station 'Kabeltonne', 54°11'03''N, 7°54'00''E) using small research vessels (http://www.awi.de/en/expedition/ships/more-ships.html) and**

processed in the laboratory of the Biological Station Helgoland within less than two hours after sampling. Biomass of free-living bacteria for DNA extraction was harvested on 0.2 mm pore sized filters after pre-filtration with 10 mm and 3 mm pore sized filters to remove large debris and particle-associated bacteria. Biomass of the 0.2–3 mm bacterioplankton fraction was used for DNA extraction and subsequent 16S rRNA gene tag sequencing.

16S rRNA gene tag sequencing: **A total of 142 samples were collected for the years 2010 to 2012. After DNA extraction, the V4 region of the 16S rRNA gene was amplified using the primers 515F (5' GTGCCAGCMGCCGCGGTAA 3') and 806R (5' GGACTACHVGGGTWTC TAAT 3') (Caporaso *et al*., 2011). Sequencing was carried out on an Illumina (San Diego, CA, USA) MiSeq sequencer with and 2x250 bp chemistry. This dataset was complemented by 16S rRNA gene tags from 7 samples from our initial study on the 2009 spring bloom (Teeling *et al*., 2012). DNA of these samples was amplified using the primers 314F (5' CCTACGGGNGGCWGCAG 3') and 805R (5' GACTACHVGGGTATCTAATCC 3') (Herlemann et al., 2011) and sequenced on the 454 FLX Ti platform.**

Data analysis: **All tag data were analyzed using the SILVAngs pipeline with the SILVA (Quast et al., 2013) v119 database. The SAR92 clade was subsequently reclassified to comply with the recently released SILVA v123, where the SAR92 no longer belong to the order *Alteromonadales*. The corresponding abundance data is summarized in Supplementary file 5. Time points from days 50 to 160 were plotted for all years. Panel A-P depict data that are analogous to the CARD-FISH data presented in Figure 3, with addition of the *Flavobacteriia* genus *Tenacibaculum* (E-H). Panels Q-X show minor *Gamma-proteobacteria* clades (Q-T) and *Roseobacter* clades together with miscellaneous other minor clades (U-X) that were not tested by CARD-FISH probes.**

3.4.4 Bacterioplankton - genetic repertoires

In total 16 metagenomes of free-living bacterioplankton (0.2 - 3 µm) were generated from time points before, during and after spring phytoplankton blooms, six during 2009 using the 454 FLX Ti platform that were published previously (Teeling *et al.*, 2012) and ten during 2010-2012 using the Illumina HiSeq2000 platform. Most of the 454 (0.5 - 4 pico titer plates / metagenome) and all of the Illumina metagenomes (1 lane / metagenome; 2x150 bp) were deeply sequenced (Supplementary file 8) with final assembled contigs of up to 96 kbp and 458 kb, respectively.

Taxonomic classification of the metagenome contigs resulted in identification of major bloom-associated clade sequence bins (Supplementary file 9), including *Formosa*, *Polaribacter*, the NS3a and NS5 marine groups, and *Cryomorphaceae* of the *Flavobacteriia* and *Alteromonadales, Reinekea, Glaciecola* and the SAR92 clade of the *Gammaproteobacteria*. Classification was poor, however, for *Ulvibacter* (*Flavobacteriia*) and *Balneatrix* (*Gammaproteobacteria*), most likely since the only available reference genomes (unidentified eubacterium SCB49; *Balneatrix alpica*) were too distant from North Sea representatives. Clone libraries from 2009 (Teeling *et al.*, 2012) indicated 16S rRNA similarities of only 94% and 91%, respectively. Other abundant clades comprised the betaproteobacterial *Burkholderiales* and *Methylophilales* (including the OM43 clade), the alphaproteobacterial SAR116 and *Roseobacter* NAC11-7 clade, and the gammaproteobacterial SAR86 and ZD0405 clades. Lower abundant clades comprised, amongst others, the OM60 (NOR5) group, the AEGEAN-169 group, and *Sulfitobacter*.

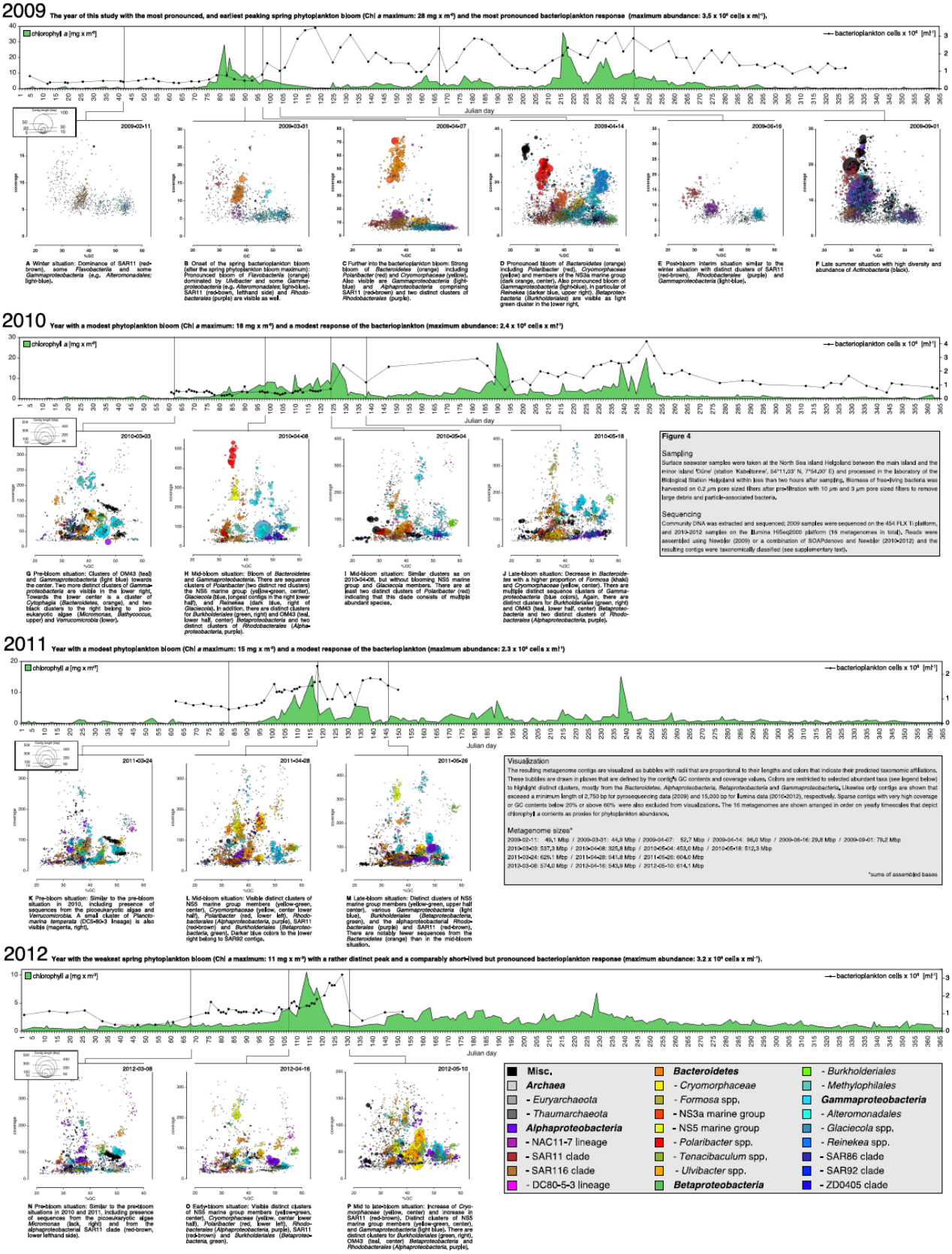


Figure 3.5: Taxonomic classification of bacterioplankton metagenomes

Sampling: **Surface seawater samples were taken at the North Sea island Helgoland between the main island and the minor island 'Düne' (station 'Kabeltonne', 54°11.03' N, 7°54.00' E) and processed in the laboratory of the Biological Station Helgoland within less than two hours after sampling. Biomass of free-living bacteria was harvested on 0.2 mm pore sized filters after pre-filtration with 10 mm and 3 mm pore sized filters to remove large debris and particle-associated bacteria.**

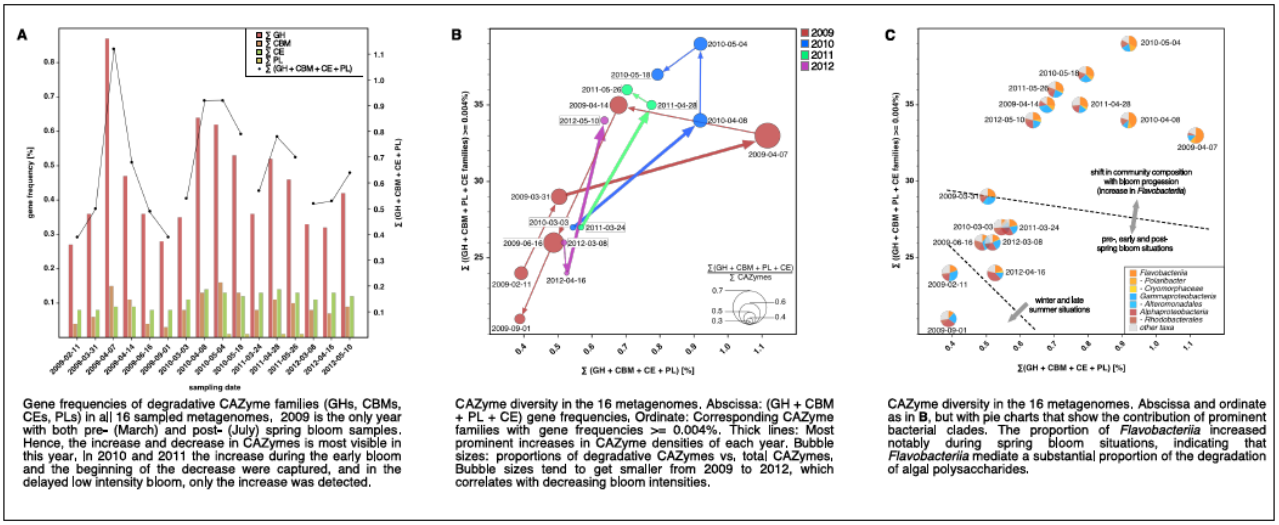
Sequencing: **Community DNA was extracted and sequenced; 2009 samples were sequenced on the 454 FLX Ti platform, and 2010-2012 samples on the Illumina HiSeq2000 platform (16 metagenomes in total). Reads were assembled using Newbler (2009) or a combination of SOAPdenovo and Newbler (2010-2012) and the resulting contigs were taxonomically classified (Supplementary file 9).**

Visualization: **The resulting metagenome contigs are visualized as bubbles with radii that are proportional to their lengths and colors that indicate their predicted taxomomic affiliations. These bubbles are drawn in planes that are defined by the contig's GC contents and coverage values. Colors are restricted to selected abundant taxa (see legend below) to highlight distinct clusters, mostly from the *Bacteroidetes, Alphaproteobacteria, Betaproteobacteria* and *Gammaproteobacteria*. Likewise only contigs are shown that exceed a minimum length of 2750 bp for pyrosequencing data (2009) and 15,000 bp for Ilumina data (2010-2012), respectively. Sparse contigs with very high coverage or GC contents below 20% or above 60% were also excluded from visualizations. The 16 metagenomes are shown arranged in order on yearly timescales that depict chlorophyll *a* contents as proxies for phytoplankton abundance.**

Metagenome sizes*: **2009-02-11: 49.1 Mbp / 2009-03-31: 44.9 Mbp / 2009-04-07: 52.7 Mbp / 2009-04-14: 96.0 Mbp / 2009-06-16: 29.8 Mbp / 2009-09-01: 79.2 Mbp 2010-03-03: 537.3 Mbp / 2010-04-08: 325.8 Mbp / 2010-05-04: 453.0 Mbp / 2010-05-18: 512.3 Mbp 2011-03-24: 629.1 Mbp / 2011-04-28: 541.8 Mbp / 2011-05-26: 604.0 Mbp 2012-03-08: 574.0 Mbp / 2012-04-16: 543.9 Mbp / 2012-05-10: 614.1 Mbp *sums of assembled bases**

We plotted contig GC contents versus coverage to evaluate our taxonomic classification, which in some cases allowed assessing the coherence of some of the clades (Figure 5). For example, *Reinekea* (Figure 5D,H,I) and the NS5 marine group (Figure 5H,L,M,O) were mostly represented by distinct clusters, whereas *Polaribacter* (Figure 5D,H,I) was almost always represented by at least two clusters indicating presence of sub-populations. In general, the number of clusters increased from pre-bloom to mid-bloom situations and decreased slightly towards late bloom situations and notably towards post-bloom situations. This tendency was most evident in 2009, the year with the highest bloom intensity and the largest number of metagenome samples spanning a broader timespan (Figure 5A-F). It is noteworthy that high *in situ* abundance did not always correlate with good metagenome assemblies. SAR11 for example, while highly abundant in all metagenome datasets, yielded few large contigs, possibly due to population heterogeneity and presence of hyper-variable regions described in sequenced SAR11 genomes (Wilhelm *et al.*, 2007).

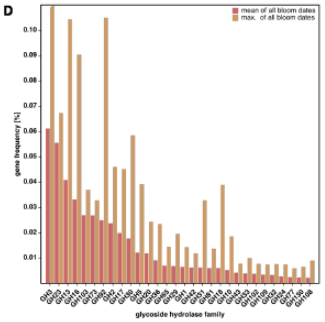
- metagenome functional analyses: CAZyme & transporter genes -



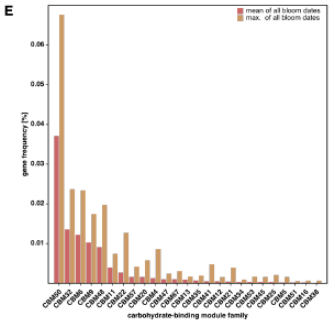
Gene frequencies of degradative CAZyme families (GHs, CBMs, CEs, PLs) in all 16 sampled metagenomes, 2009 is the only year with both pre- (March) and post- (July) spring bloom samples. Hence, the increase and decrease in CAZymes is most visible in this year. In 2010 and 2011 the increase during the early bloom and the beginning of the decrease were captured, and in the delayed low intensity bloom, only the increase was detected.

CAZyme diversity in the 16 metagenomes. Abscissa: (GH + CBM + PL + CE) gene frequencies, Ordinate: Corresponding CAZyme families with gene frequencies >= 0.004%. Thick lines: Most prominent increases in CAZyme densities of each year. Bubble sizes: proportions of degradative CAZymes vs. total CAZymes. Bubble sizes tend to get smaller from 2009 to 2012, which correlates with decreasing bloom intensities.

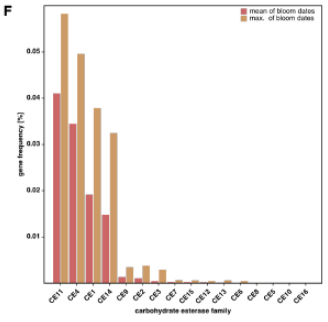
CAZyme diversity in the 16 metagenomes. Abscissa and ordinate as in **B**, but with pie charts that show the contribution of prominent bacterial clades. The proportion of *Flavobacteriia* increased notably during spring bloom situations, indicating that *Flavobacteriia* mediate a substantial proportion of the degradation of algal polysaccharides.



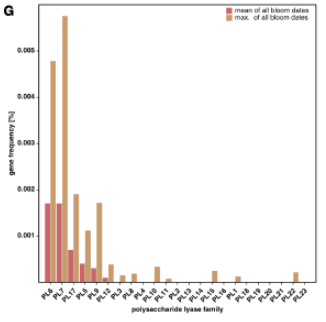
The 30 most abundant GH families in all spring bloom metagenomes (i.e. all metagenomes combined excluding the winter and late summer samples of 2009; see **C**). These families are indicative for the hydrolysis of various glucans, xylans, mannans (GH92), cellulose (GH5), fucans (GH29) and peptido-glycan (GH23, 103, 73, 18). The abundant families GH3, 16, 17 and 5 comprise β-1,3-glucanases that are used to cleave chrysolaminarin, the storage polysaccharide in diatoms.



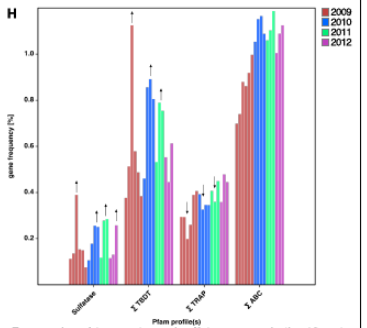
The 25 most abundant CBM families in all spring bloom metagenomes (as in **D**). These CBMs indicate binding of peptidoglycan (CBM50, 32), galactans (CBM32), glycogen (CBM48), β-1,3- and β-1,4-glucans (CBM6, 11, 4), xylan (CBM6, 9, 22, 4), and cellulose (CBM6, 4), i.e. storage and cell wall polysaccharides of algae and bacteria. CBMs for the binding of fucose- and rhamnose-rich algal TEP were only found in lesser abundances, e.g. CBM47 (fucose) and CBM67 (rhamnose).



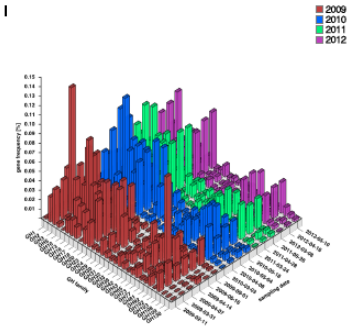
Abundances of all 16 CE families in all spring bloom metagenomes (as in **D**). The only known function for family CE11 is that of UDP-3-O-acyl-N-acetylglucosamine deacetylase, an enzyme involved in lipid A biosynthesis. The two second most abundant CE families 4 and 1 comprise a number of enyzmatic functions, for example deacetylation of peptidoglycan, xylan and chitin. Family CE14 contains known deacetylases in inositol metabolism in *Eukaryota* and *Actinobacteria*.



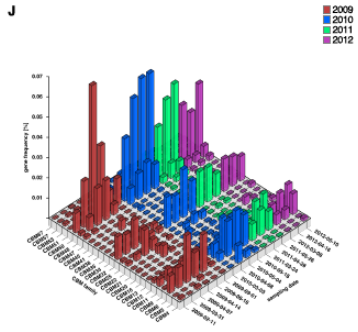
Abdunances of all 23 PL families in all spring bloom metagenomes (as in **D**). The most abundant families PL7, 6, 17 and 5 are involved in the degradation of alginate - a cell wall polysaccharide in brown macroalgae. Frequencies of these PLs were an order of magnitude lower than those of GHs, CBMs, and CEs, indicating that alginate degradation does not play a vital role in North Sea spring phytoplankton blooms that are dominated by microalgae such as diatoms and raphidophytes.



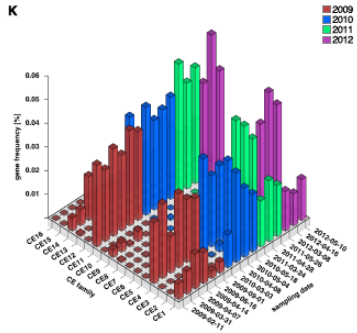
Frequencies of transporter and sulfatase genes in the 16 metagenomes according to Pfam profiles (see **M** for details). TBDT = TonB-dependent transporters; TRAP = Tripartite ATP-independent periplasmic transporters; ABC = ABC type transporters. As indicated by arrows, relative frequencies of sulfatase and TBDT genes tended to peak during blooms, whereas TRAP transporter genes tended to decrease (see **R** for details).



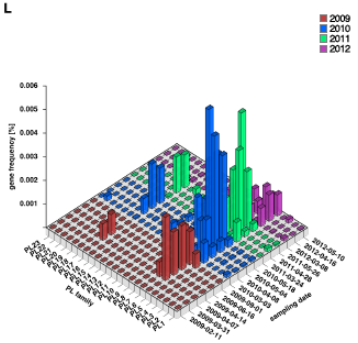
Gene frequencies of the 30 most abundant GH families in all 16 metagenomes. Patterns of higher vs. lower GH abundances were similar across years. Some GH families exhibited prounced peaks during all blooms, e.g. GH3, 13, 16, 23, and 30.



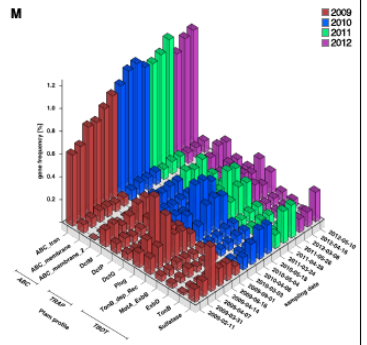
Gene frequencies of the 25 most abundant CBM families in all 16 metagenomes. Similar patterns of higher vs. lower abundances were observed in all years.



Frequencies of genes of all 16 CE families in all 16 metagenomes. Similar patterns of highly vs. lowly abundant CE families were present in all years. The families CE1 and 14 exhibited clear bloom-associated peaks.



Frequencies of genes of all 23 PL families in all 16 metagenomes. PL genes were notably less frequent than GHs, CBMs and CEs. Similar patterns were observed in all years. Known members of the most abundant families PL7, 6, 17, and 5 are involved in alginate degradation.



Pfam-based analysis of the frequencies of transporter genes and sulfatases. ABC transporter profiles: ABC_trans, ABC_membrane, ABC_membrane_2; TRAP transporter profiles: DctM, DctP, DctQ; TBDT profiles: Plug, TonB-dep_Rec, MotA_ExbB, ExbD, TonB.

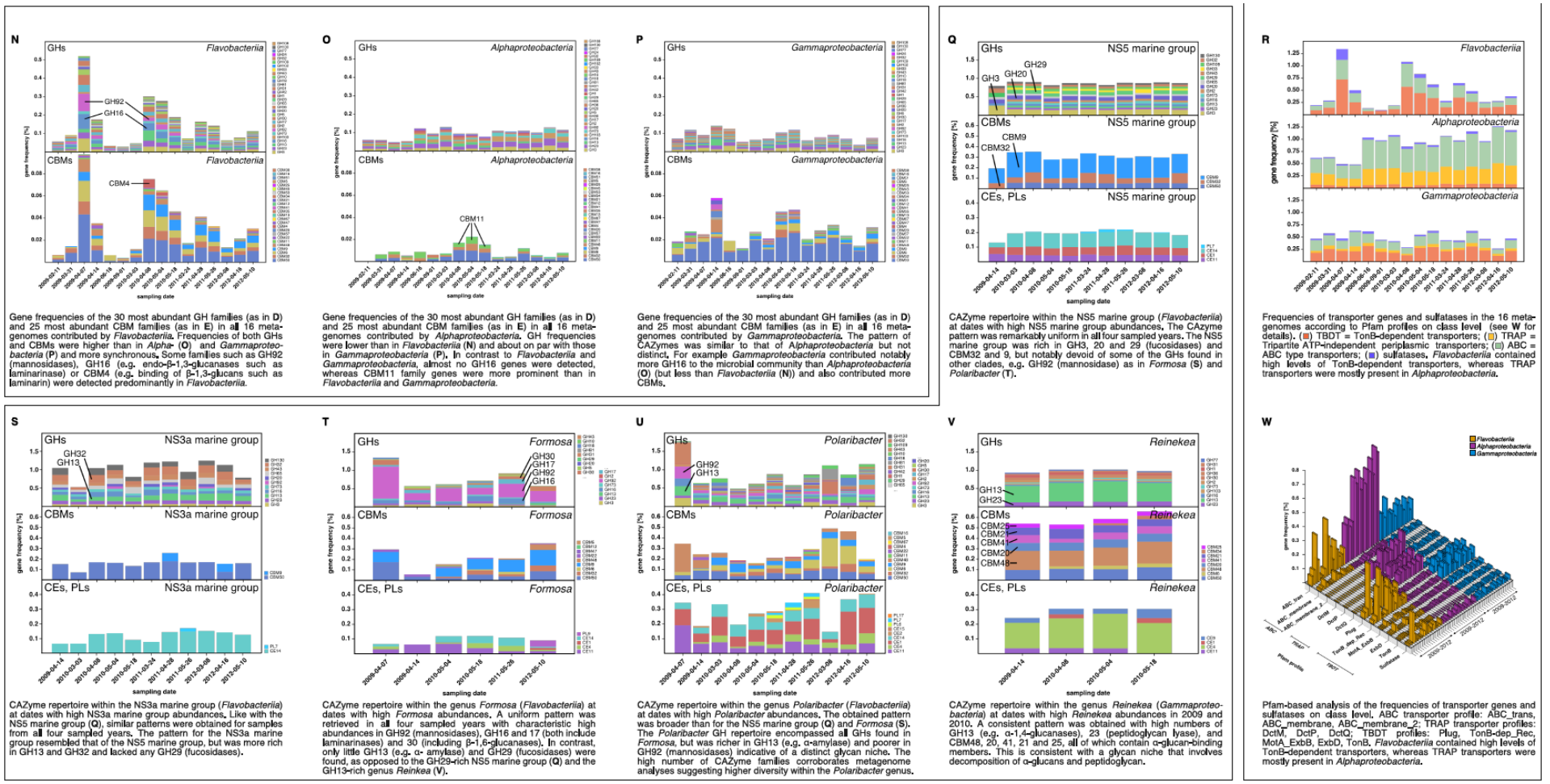


Figure 3.6: Metagenome functional analyses: CAZyme, sulfatase and transporter gene frequencies.

Sampling: **Surface seawater samples were taken at the North Sea island Helgoland between the main island and the minor island Düne' (station 'Kabeltonne', 54°11'03''N, 7°54'00''E) and processed in the laboratory of the Biological Station Helgoland within less than two hours after sampling. Biomass of free-living bacteria was harvested on 0.2 mm pore sized filters after pre-filtration with 10 mm and 3 mm pore sized filters to remove large debris and particle-associated bacteria.**

Sequencing: **Community DNA was extracted and sequenced. 2009 samples were sequenced on the 454 FLX Ti platform, and 2010–2012 samples on the Illumina HiSeq2000 platform (16 metagenomes in total). Reads were assembled using Newbler (2009) or a combination of SOAPdenovo and Newbler (2010–2012) and the resulting contigs were taxonomically classified (Supplementary file 9).**

Metagenome sizes*: **2009-02-11: 49.1 Mbp / 2009-03-31: 44.9 Mbp / 2009-04-07: 52.7 Mbp / 2009-04-14: 96.0 Mbp / 2009-06-16: 29.8 Mbp / 2009-09-01: 79.2 Mbp 2010-03-03: 537.3 Mbp / 2010-04-08: 325.8 Mbp / 2010-05-04: 453.0 Mbp / 2010-05-18: 512.3 Mbp 2011-03-24: 629.1 Mbp / 2011-04-28: 541.8 Mbp / 2011-05-26: 604.0 Mbp 2012-03-08: 574.0 Mbp / 2012-04-16: 543.9 Mbp / 2012-05-10: 614.1 Mbp *sums of assembled bases**

Data Analysis: **CAZymes were predicted as consensus of searches against the CAZy, dbCAN and Pfam databases with custom E-value cutoffs (Supplementary file 11). Sulfatase and transporter genes were predicted based on HMMER searches against the Pfam databases with an E-value cutoff of E-5. Gene frequencies were computed as [(sum of average coverage of target genes) *100 / (sum of average coverage of all genes)]. All dates in the graphs are in the format [yyyy-mm-dd].**

Analyses of functional genes identified in the bacterioplankton metagenomes revealed that increases in *Flavobacteriia* relative abundance during blooms was always accompanied by an increase in community-wide CAZyme gene frequency as well as an increase in the diversity of CAZyme families (Figure 6A-C). As blooms subsided, CAZyme frequencies also declined. This was most pronounced for more densely sampled years in 2009 and 2010. In 2012 the decline in CAZymes was not captured as the last metagenome sample was taken before the bloom decline (Figure 6A, x-axis).

The 20 glycoside hydrolase families with the highest mean abundances during bloom dates were, in descending order, GH3, 23, 13, 16, 103, 73, 92, 2, 17, 30, 5, 20, 36, 65, 29, 1, 42, 31, 81, and 18 (Figure 6D). While most of these families comprise a diverse range of functions, they are indicative for the hydrolysis of certain glucans, xylans, mannans (GH92), cellulose (GH5), fucans (GH29), and peptidoglycan (GH23, 103, 73, and 18). The abundant families GH3, 16, 17 and 5 comprise β-1,3-glucanases and the family GH30 β-1,6-glucanases. Both enzymes are involved in the cleavage of chrysolaminarin. Chrysolaminarin, a mostly linear β-1,3-glucan with occasional β-1,6 branches, is the storage polysaccharide in diatoms and thus one of the most abundant polysaccharides on Earth.

The ten families of carbohydrate-binding modules with the highest mean abundances during blooms were, in descending order, CBM50, 32, 6, 9, 48, 11, 22, 57, 20, and 4 (Figure 6E). These CBM domains are indicative for binding of peptidoglycan (CBM50), galactans (CBM32), glycogen (CBM48), β-1,3- and β-1,4-glucans (CBM6, 11, 4), xylan (CBM6, 9, 22, 4), and cellulose (CBM6, 4). This suggests a pronounced specialization of the bacterial community in the acquisition of storage polysaccharides (β-1,3-glucans such as chrysolaminarin; α-1,4-glucans such as starch and glycogen) and cell wall

polysaccharides (xylose, cellulose and peptidoglycan) of both algae and bacteria. By contrast, CBMs for the binding of algal TEP, which consists predominantly of fucose- and rhamnose-rich anionic sulfated heteropolysaccharides (Passow, 2002), were only found in lower abundances, e.g. CBM47 (fucose) and CBM67 (rhamnose).

Among carbohydrate esterase families, CE11, 4, 1, and 14 exhibited the highest mean abundances during blooms (Figure 6F). The only known function for family CE11 is that of the UDP-3-O-acyl-N-acetylglucosamine deacetylase, an enzyme involved in lipid A biosynthesis. The second and third most abundant CE families 4 and 1 comprise a number of enzymatic functions including deacetylation of peptidoglycan, xylan, and chitin.

Finally, the polysaccharide lyase families PL6, 7, 17 and 5 constituted the most abundant PL families during bloom dates (Figure 6G). Known members of these PL families are all involved in the usage of alginate. However, gene frequencies of these PL families were an order of magnitude below those of abundant GHs, CBMs, and CEs, indicating that alginate degradation does not play a vital role in North Sea spring phytoplankton blooms. Alginate is a cell wall constituent in brown macroalgae, however, the microalgae that dominate North Sea spring blooms are devoid of alginate.

For all of these GH, CBM, PL, and CE families, we observed remarkably similar gene frequency patterns during all four spring blooms, often with peaks in the same families (Figure 6I-L). Alongside CAZymes, sulfatase and TonB-dependent transporter (TBDT) gene frequencies also increased during blooms while tripartite ATP-independent periplasmic (TRAP) transporter genes showed an almost opposite trend (Figure 6H,M). A class-level analysis revealed that sulfatases and TBDT genes were predominantly present in *Bacteroidetes*, whereas TRAP transporters were mostly present in *Alphaproteobacteria*

(Figure 6R,W). The observed shifts in sulfatase, TBDT and TRAP transporter frequencies hence reflect the shifts in relative abundance between these two classes. This is in agreement with our study on the 2009 spring bloom (Teeling *et al.*, 2012) and furthermore demonstrates recurrence of this phenomenon during four consecutive years.

Class-level analyses of the most abundant GH and CBM families (Figure 6N-P) showed that *Flavobacteriia* not only contributed more total CAZymes to the microbial community than *Alpha-* and *Gammaproteobacteria*, but also exhibited a tighter coupling between GH and CBM genes with highly similar abundances profiles (Figure 6N). The distribution of families was also more uneven in *Flavobacteriia*, indicative of a more pronounced substrate specialization compared to *Alpha-* and *Gammaproteobacteria*. GH92 (mannosidase) and GH16 (including β-1,3-glucanase) genes, for example, were predominantly found in *Flavobacteriia*, possibly indicating more readiness to decompose mannans and chrysolaminarin.

Metagenome taxonomic classification provided sufficient data for analysis of CAZyme repertoires of the flavobacterial NS5 marine group, NS3a marine group, *Formosa*, *Polaribacter,* and *Cryomorphaceae*, and the gammaproteobacterial *Alteromonadales* and *Reinekea* clades (Supplementary file 10). For most of these clades, the analyses revealed fingerprint-like patterns, which corroborate the hypothesis that these clades have distinct glycan niches that are relatively stable across years. For example, the NS5 marine group (Figure 6Q) was rich in GH3, 20 and 29 (fucosidases), but notably devoid of GH92 (mannosidases). By contrast, *Formosa* and *Polaribacter* clades (Figure 6T,U) contained higher abundances of GH92 genes. The *Formosa* CAZyme profile was also characterized by high proportions of GH16, 17, and 30 families, which all contain enzymes that can decompose chrysolaminarin. *Polaribacter* contained a broader set of CAZymes that

included all families found in *Formosa*, however, *Polaribacter* was richer in GH13 (e.g. α-amylase) and poorer in GH92 (mannosidases) than *Formosa*. Likewise, the GH repertoires of the NS3a and NS5 marine groups were similar (Figure 6Q,S), but the NS3a marine group was richer in GH13 and 32 and devoid of GH29 family fucosidases. The high number of CAZyme families in *Polaribacter* corroborated metagenome bin analyses that suggested a higher diversity within this clade. CAZyme gene frequencies were much lower in the *Cryomorphaceae* than in the other investigated *Flavobacteriia* clades with GH frequencies barely exceeding 0.5% (Figure 6–figure supplement 1). This suggests a different ecophysiological niche and a distinct role of the *Cryomorphaceae* during phytoplankton blooms.

For *Gammaproteobacteria*, recurring patterns were detected for the prominent *Alteromonadales* and *Reinekea* clades. *Alteromonadales* contained some of the GH families that play important roles during phytoplankton blooms, such as GH13 and 16, but were notably poor in or even devoid of others, such as GH29 and GH92, respectively (Figure 6–figure supplement 2). In contrast to other prominent clades, we did not obtain sufficient metagenome sequences for *Reinekea* for all four years, but only for 2009 and 2010 (Figure 6V). However, the *Reinekea* CAZyme patterns of 2009 and 2010 were well conserved with high proportions of GH23 and 13, and CBM48, 20, 41, 21, and 25. The GH23 family comprises peptidoglycan lyases and the GH13 family contains α-1,4-glucanases (e.g. α-amylase). CBM48, 20, 41, 21, and 25 all bind α-1,4-glucans such as starch and glycogen, and the ubiquitous CBM50 contains peptidoglycan-binding members. These results are consistent with a glycan niche that involves decomposition of external α-1,4-glucans and possibly peptidoglycan.

3.5 Discussion

Nutrient-poor marine surface waters are dominated by clades such as SAR11 and *Prochlorococcus*. Both feature small, reduced genomes and can use sunlight and small organic molecules. The otherwise heterotrophic SAR11 use proteorhodopsin for supplemental phototrophy (Giovannoni *et al.*, 2005) and phototrophic *Prochlorococcus* cyanobacteria are capable of supplemental uptake of amino acids (Zubkov *et al.*, 2003) and glucose (Muñoz-Marín Mdel *et al.*, 2013). Microbial communities in nutrient-rich 'green' surface oceans by contrast feature higher proportions of heterotrophic species that feed on more complex organic substrates. In particular during phytoplankton blooms, the release of algae-derived organic matter selects for fast growing species with genomic adaptations towards algal biomass remineralization. These are typically members of the *Flavobacteriia* and *Gammaproteobacteria* classes and the alphaproteobacterial *Roseobacter* clade. Similarly adapted species from these clades compete for substrates during phytoplankton blooms with variation in which species prevail. Despite this stochastic effect, the most well-adapted species will be successful more often and thus exhibit patterns of annual recurrence.

During spring phytoplankton blooms at Helgoland in the North Sea, we observed recurrent bloom-associated abundance peaks of in particular flavobacterial clades, namely *Formosa*, *Polaribacter*, the NS3a marine group, *Tenacibaculum*, *Ulvibacter*, and the *Cryomorphaceae* VIS6 clade. Within *Gammaproteobacteria Alteromonadaceae/ Colwelliaceae, Reinekea*, and the SAR92 clade were clearly bloom-associated and recurrent as was *Methylophilales* within *Betaproteobacteria*, and 'Cd. *Planktomarina temperata*' from the DC5-80-3 lineage (a.k.a *Roseobacter* clade affiliated = RCA group) and the NAC11-7 lineage within the *Roseobacter* clade. It has already been shown that the abundant North Sea isolate 'Cd. *Planktomarina temperata*' RCA23^{T} is associated with

decaying phytoplankton (Giebel *et al.*, 2011) and high abundances and in particular high activity have been reported during a spring phytoplankton bloom event in the North Sea of 2010 (Voget *et al.*, 2014; Wemheuer *et al.*, 2015). High activities of members of the RCA and the SAR92 clades during North Sea spring phytoplankton blooms have also been reported in 2009 (Klindworth *et al.*, 2014) and 2010 (Wemheuer *et al.*, 2015), just as an increase of *Bacteroidetes* of the genera *Marinoscillum* and *Polaribacter* during 2010 (Wemheuer *et al.*, 2015).

Many of the other clades we report here (including some low abundance groups) have been found during blooms of dinoflagellates, including AEGEAN-169, *Alteromonadales*, NS3a marine group, NS5 marine group, OM43, OM60 (NOR5), SAR116, SAR86, and ZD0405 (Yang *et al.*, 2015) or *Cryomorphaceae, Glaciecola* and *Sulfitobacter* (Tan *et al.*, 2015). The OM43 clade (order *Methylophilales*) comprises methylotrophs known to feed on algae C1 compounds (Halsey *et al.*, 2012) and it has been reported that *Sulfitobacter* species SA11 shares a mutually beneficial exchange of compounds with the diatom species *Pseudo-nitzschia multiseries* (Amin *et al.*, 2015).

There is of course unaddressed diversity in all these clades. Some of the genera might be dominated by a single species while others might be more diverse with considerable variation in competitive success between bloom events. Nevertheless, the high level of recurrence in particular of flavobacterial clades indicates a strong selection of few clades highly adapted for the manipulation and uptake of specific and complex polysaccharides (and likely other biopolymers) and disagrees with substantial levels of sloppy-feeding by these bacteria that would allow other less adapted clades to arbitrarily reach high abundances via cross-feeding. This might be attributed to the capability of *Bacteroidetes* for very efficient macromolecule uptake as it has been recently shown for uptake of α-

mannan by the human gut bacterium *Bacteroides thetaiotaomicron* (Cuskin *et al.*, 2015). This bacterium binds α-mannan macromolecules to its surface, followed by rapid cleavage into larger oligomers that are immediately imported via a TonB-dependent transporter (TBDT) into the periplasm without detectable loss. The bulk of the degradation into smaller molecules takes place in the periplasm where the substrate is secure from outside competitors before transport into the cytoplasm.

TonB-dependent transporters are not specific to *Bacteroidetes*, but it seems that only *Bacteroidetes* have evolved a functional coupling of SusC-like TBDT porins with SusD-like TonB-dependent receptors (TBDRs) that bind and guide the substrate to the porin. At least so far, only *Bacteroidetes* genomes feature characteristic *susCD* gene tandems. Within *Bacteroidetes* genomes such tandems are frequently found in so-called polysaccharide utilization loci (PULs; (Sonnenburg *et al.*, 2010)). PULs are operons or regulons where one or more *susCD* gene tandems are co-located with CAZymes. Further frequent accessory genes in PULs encompass transcriptional regulators, proteases, transporter components and sulfatases. The latter are required for the desulfation of sulfated polysaccharides, which marine algae produce in large quantities. The diversity of PULs in marine *Bacteroidetes* genomes is high and largely unexplored and so far only few PULs have been linked to dedicated algal polysaccharides (e.g. Hehemann *et al.* (2012), Kabisch *et al.* (2014), Xing *et al.* (2015)). The large, complex and efficient PUL uptake systems might explain why *Flavobacteria* consistently outcompeted *Gammaproteobacteria* during the onset of all blooms.

It is noteworthy that we observed a shift in bacterioplankton biodiversity alongside a shift in functional gene repertoires for the major clades in all four years of this study. The abundance of CAZymes and sulfatases increased from pre- to mid-bloom situations and

leveled off post-bloom. Likewise, similar abundance patterns were observed for the most abundant CAZyme families in all studied years.

We did observe an increase in the abundance of TBDRs during bloom situations, and we have shown previously that TBDRs are among the most abundantly expressed proteins during the bacterial mineralization of algae biomass (Teeling *et al.*, 2012). The relevance of TBDRs in nutrient-rich oceanic regions has been also supported by *in situ* metaproteome studies of samples from the South Atlantic Ocean (in particular at coastal upwelling zones; (Morris *et al.*, 2010)) and from the Antarctic Southern Ocean (Williams *et al.*, 2013).

The recurrent patterns in bacterioplankton diversity and functional repertoires during the four studied spring blooms are remarkable in view of the variation among algae taxa. For example, even though algal biomass of the 2012 spring bloom was dominated by silicoflatellate *Chattonella* spp. and not by diatoms as was true for the three preceding years, the respective bacterioplankton communities were strikingly similar. Likewise only few bacterioplankton taxa seemed to be weakly correlated with dedicated phytoplankton taxa, and no clear correlation was found between distinct bacterioplankton taxa and individual distinct diatom clades in statistical analyses (Supplementary files 6 and 7). It seems that phytoplankton community composition did not exert a strong effect on the composition of the free-living non-phycosphere bacterioplankton community. Instead, the dominating algae (in terms of biomass) seemed to produce similar or perhaps identical types of substrates for specifically adapted clades of heterotrophic bacterioplankton. It is therefore conceivable that recurrence is more pronounced on the functional level than on the taxonomic level, since species from different taxa with similar ecophysiological niches might functionally substitute each other in different years. This hypothesis is supported by

the bacterioplankton communities' similar CAZyme gene repertoires in the 2009 to 2012 spring blooms and in particular the consistency on class level that was almost unaffected by distinct blooming clades, yet needs to be further tested by deep metatranscriptome sequencing and metaproteomics during multiple spring phytoplankton blooms in future studies. Considering the extent of recurrence, our combined metagenome data (>5 million predicted proteins) should provide sufficient search space for such an analysis.

The existence of recurrent key players during North Sea spring phytoplankton blooms suggests that the bacterioplankton community composition during and after such blooms is governed by deterministic effects. We have shown before that temperature exerts a strong effect on North Sea bacterioplankton as it selects for temperature-dependent guilds, for example when comparing spring and summer blooms (Lucas *et al.*, 2015). Within short-lived spring blooms, however, the supply of algae-derived organic matter is among the main factors that shape the bacterioplankton composition. In particular the different types of structurally distinct polysaccharides that algae produce in large quantities seem to exert such substrate-induced forcing. Since *Flavobacteriia* are more specialized on polysaccharides than *Gammaproteobacteria*, this would also explain, why *Flavobacteriia* dominated the recurrent clades (Figure 3Q-T, Figure 4E-J) and *Gammaproteobacteria* clades exhibited more stochastic peaks (Figure 4Q-T).

At the beginning of a bloom, most available polysaccharides will be exopolysaccharides, but as the bloom commences and algae become senescent, more and more cellular algal substrates are released, culminating in the bloom's final die off phase. Bacteria will naturally consume the more degradable substrates such as storage polysaccharides (e.g. chrysolaminarin) first, and more recalcitrant substrates (e.g. branched and sulfated polysaccharides) later. TEP for example seems to undergo such selective feeding, as it

has been suggested that in particular fucose-rich TEP is less readily degraded than mannose and galactose rich TEP (for review see Passow (2002)). Such selective feeding creates an additional change in substrate availability and leads to a succession of substrate niches for specifically adapted bacterioplankton clades to grow.

3.6 Concluding remarks

Bacterioplankton communities during spring phytoplankton blooms in the coastal North Sea undergo swift and dynamic composition changes and thus are difficult to investigate. Nonetheless, we found clades that recurrently reached high abundances among *Flavobacteriia* (*Formosa*, *Polaribacter*, NS3a marine group, *Ulvibacter*, VIS6 clade *Cryomorphaceae, Tenacibaculum*), *Gammaproteobacteria* (*Alteromonadaceae*/*Colwelliaceae*, SAR92, *Reinekea*) and *Roseobacter* clade *Alphaproteobacteria* (DC5-80-3, NAC11-7). Recurrence was not only detectable on the taxonomic but also on the functional level with a highly predictable increase in TonB-dependent polysaccharide uptake systems and distinct CAZyme patterns. The niches of abundant bacterioplankton clades are more complex and manifold than the glycan niches that we explore in this study. CAZymes, however, have the advantage that they allow linking of gene repertoires and possible environmental functions in a way currently not feasible for other macromolecules such as proteins and lipids. Our results suggest that besides stochastic also deterministic effects influence phytoplankton-bacterioplankton coupling during blooms. They indicate that during spring phytoplankton blooms similar principles of resource partitioning and specialization are at play as within human gut microbiota that decompose fiber-rich plant material, albeit at a much larger scale. Rather the availability of substrates commonly occurring in microalgae than one-to-one interactions of particular phytoplankton and bacterioplankton species caused the succession of free-living bacterioplankton clades.

3.7 Acknowledgements

This study was funded by the Max Planck Society. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231.

3.8 References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215:** 403-10.

Apprill A, McNally S, Parsons R, Weber L. 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol* **75:** 129-37.

Amin SA, Parker MS, Armbrust EV. 2012. Interactions between diatoms and bacteria. *Microbiol Mol Biol Rev* **76**: 667-84.

Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, Morales RL, Berthiaume CT, Parker MS, Djunaedi B, Ingalls AE, Parsek MR, Moran MA, Armbrust EV. 2015. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**: 98-101.

Armbrust EV. 2009. The life of diatoms in the world's oceans. Nature 459: 185-92.

Bell RT, Kuparinen J. 1984. Assessing phytoplankton and bacterioplankton production during early spring in Lake Erken, Sweden. *Appl Environ Microbiol* **48:** 1221-30.

Buchan A, LeCleir GR, Gulvik CA, Gonzalez JM. 2014. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* **12:** 686-98.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37:** D233-8.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* **108** Suppl 1: 4516-22.

Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron EA, Pudlo NA, Porter NT, Urs K, Thompson AJ, Cartmell A, Rogowski A, Hamilton BS, Chen R, Tolbert TJ, Piens K, Bracke D, Suits MD, Boraston AB, Atherly T, Ziemer CJ, Williams SJ, Davies GJ, Abbott DW, Martens EC, Gilbert HJ. 2015. Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature* **517:** 165-9.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460-1.

Eilers H, Pernthaler J, Glöckner FO, Amann R. 2000. Culturability and *in situ* abundance of pelagic bacteria from the North Sea. *Appl Environ Microbiol* **66**: 3044-51.

Eilers H, Pernthaler J, Peplies J, Glöckner FO, Gerdts G, Amann R. 2001. Isolation of novel pelagic bacteria from the German bight and their seasonal contributions to surface picoplankton. *Appl Environ Microbiol* **67**: 5134-42.

Falkowski PG, Barber RT, Smetacek VV. 1998. Biogeochemical controls and feedbacks on ocean primary production. *Science* **281:** 200-7.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281:** 237-40.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* **42:** D222-30.

Fuhrman JA, Cram JA, Needham DM. 2015. Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* **13:** 133-46.

Gerlach W, Stoye J. 2011. Taxonomic classification of metagenomic shotgun sequences with CARMA3. *Nucleic Acids Res* **39:** e91.

Giebel HA, Kalhoefer D, Lemke A, Thole S, Gahl-Janssen R, Simon M, Brinkhoff T. 2011. Distribution of *Roseobacter* RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. *ISME J* **5:** 8-19.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappe MS, Short JM, Carrington JC, Mathur EJ. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242-5.

Gómez-Pereira PR, Fuchs BM, Alonso C, Oliver MJ, van Beusekom JE, Amann R. 2010. Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. *ISME J* **4:** 472-87.

Gómez-Pereira PR, Schüler M, Fuchs BM, Bennke C, Teeling H, Waldmann J, Richter M, Barbe V, Bataille E, Glöckner FO, Amann R. 2012. Genomic content of uncultured *Bacteroidetes* from contrasting oceanic provinces in the North Atlantic Ocean. *Environ Microbiol* **14**: 52-66.

Gómez-Pereira PR, Hartmann M, Grob C, Tarran GA, Martin AP, Fuchs BM, Scanlan DJ, Zubkov MV. 2013. Comparable light stimulation of organic nutrient uptake by SAR11 and Prochlorococcus in the North Atlantic subtropical gyre. *ISME J* **7**: 603-14.

Halsey KH, Carter AE, Giovannoni SJ. 2012. Synergistic metabolism of a broad range of C1 compounds in the marine methylotrophic bacterium HTCC2181. *Environ Microbiol* **14**: 630-40.

Hehemann JH, Kelly AG, Pudlo NA, Martens EC, Boraston AB. 2012. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc Natl Acad Sci U S A* **109:** 19786-91.

Herlemann DP, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, Andersson AF. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J* **5**: 1571-9.

Huntemann M, Ivanova NN, Mavromatis M, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen IM-A, Pati A, Markowitz VM, Kyrpides NC. 2015. The standard operating procedure of the DOE-JGI metagenome annotation pipeline (MAP v.4). *Stand Genomic Sci*, **10:** 86.

Hutchinson GE. 1961. The paradox of the plankton. *Amer Nat* **95:** 137-45.

Kabisch A, Otto A, König S, Becher D, Albrecht D, Schüler M, Teeling H, Amann RI, Schweder T. 2014. Functional characterization of polysaccharide utilization loci in the marine *Bacteroidetes* '*Gramella forsetii*' KT0803. *ISME J* **8:** 1492-1502.

Klindworth A, Mann AJ, Huang S, Wichels A, Quast C, Waldmann J, Teeling H, Glöckner FO. 2014. Diversity and activity of marine bacterioplankton during a diatom bloom in the North Sea assessed by total RNA and pyrotag sequencing. *Mar Genomics* **S1874-7787**: 00101-9.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol S* **305:** 567-80.

Loebl M, van Beusekom JEE, Philippart CJM. 2013. No microzooplankton grazing during a Mediopyxis helysia dominated diatom bloom. *Journal of Sea Research* **82**: 80-5.

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

Lucas J, Wichels A, Teeling H, Chafee M, Scharfe M, Gerdts G. 2015. Annual dynamics of North Sea bacterioplankton: seasonal variability superimposes short-term variation. *FEMS Microbiol Ecol* **91**: doi: 10.1093/femsec/fiv099.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1:** 18.

Malmstrom RR, Straza TRA, Cottrell MT, Kirchman DL. 2007. Diversity, abundance, and biomass production of bacterial groups in the western Arctic Ocean. *Aquat Microb Ecol* **47**: 45-55.

Mann DG. 1999. The species concept in diatoms. *Phycologia* **38**: 437-95.

Manz W, Amann R, Ludwig W, Vancanneyt M, Schleifer KH. 1996. Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum cytophaga-flavobacter-bacteroides in the natural environment. *Microbiology* **142**: 1097-106.

Manz W, Amann R, Ludwig W, Wagner M, Schleifer K-H. 1992. Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria: problems and solutions. *Syst Appl Microbiol* **15**: 593-600.

Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Pillay M, Ratner A, Huang J, Pagani I, Tringe S, Huntemann M, Billis K, Varghese N, Tennessen K, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC. 2014. IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res* **42:** D568-73.

Martinez-Garcia M, Brazel DM, Swan BK, Arnosti C, Chain PS, Reitenga KG, Xie G, Poulton NJ, Gomez ML, Masland DE, Thompson B, Bellows WK, Ziervogel K, Lo CC, Ahmed S, Gleasner CD, Detter CJ, Stepanauskas R. 2012. Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of *Verrucomicrobia*. *PLoS One* **7**: e35314.

Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, Puhler A. 2003. GenDB--an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31:** 2187-195.

Morris RM, Nunn BL, Frazar C, Goodlett DR, Ting YS, Rocap G. 2010. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J* **4:** 673-85.

Muñoz-Marín M del C, Luque I, Zubkov MV, Hill PG, Diez J, García-Fernández JM. 2013. *Prochlorococcus* can use the Pro1404 transporter to take up glucose at nanomolar concentrations in the Atlantic Ocean. *Proc Natl Acad Sci U S A* **110:** 8597-602.

Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B. 1995. Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles* **9:** 359-72.

Nielsen H, Brunak S, von Heijne G. 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng* **12:** 3-9.

Niu Y, Shen H, Chen J, Xie P, Yang X, Tao M, Ma Z, Qi M. 2011. Phytoplankton community succession shaping bacterioplankton community composition in Lake Taihu, China. *Water Res* **45:** 4169-82.

Olenina I. 2006. Biovolumes and size-classes of phytoplankton in the Baltic Sea. *HELCOM BaltSea Environ Proc* **106**: 144pp.

Passow U. 2002. Transparent exopolymer particles (TEP) in aquatic environments. Progress in Oceanography **55:** 287-333.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Res* **40:** D290-301.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* D590-96.

Rappé MS, Connon SA, Vergin KL, Giovannoni SJ. 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**: 630-3.

Reddy TB, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos EA, Kyrpides NC. 2015. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res* **43**: D1099-106.

Ruff SE, Probandt D, Zinkann AC, Iversen MH, Klaas C, Würzberg L, Krombholz N, Wolf-Gladrow D, Amann R, Knittel K. 2014. Indications for algae-degrading benthic microbial communities in deep-sea sediments along the Antarctic Polar Front. *Deep Sea Research Part II: Topical Studies in Oceanography* doi:10.1016/j.dsr2.2014.05.011.

Sapp M, Wichels A, Wiltshire KH, Gerdts G. 2007. Bacterial community dynamics during the winter-spring transition in the North Sea. *FEMS Microbiol Ecol* **59:** 622-37.

Sarmento H, Gasol JM. 2012. Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environ Microbiol* **14:** 2348-60.

Schattenhofer M, Fuchs BM, Amann R, Zubkov MV, Tarran GA, Pernthaler J. 2009. Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ Microbiol* **11**: 2078-93.

Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, Sonnenburg JL. 2010. Specificity of polysaccharide use in intestinal *Bacteroides* species determines diet-induced microbiota alterations. *Cell* **141:** 1241-52.

Stingl U, Desiderio RA, Cho JC, Vergin KL, Giovannoni SJ. 2007. The SAR92 clade: an

abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microbiol* **73**: 2290-6.

Tada Y, Taniguchi A, Nagao I, Miki T, Uematsu M, Tsuda A, Hamasaki K. 2011. Differing growth responses of major phylogenetic groups of marine bacteria to natural phytoplankton blooms in the western North Pacific Ocean. *Appl Environ Microbiol* **77:** 4055-65.

Tan SL, Xiaoshan Zhu JS, Yu S, Zhan W, Wang B, Cai Z. 2015. An association network analysis among microeukaryotes and bacterioplantkon reveals algal bloom dynamics. *J Phycol* **5:** 120-32.

Taylor JD, Cottingham SD, Billinge J, Cunliffe M. 2014. Seasonal microbial community dynamics correlate with phytoplankton-derived polysaccharides in surface coastal waters. *ISME J* **8:** 245-48.

Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy M, Huang S, Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J, Bernhardt J, Reinsch C, Hecker M, Peplies J, Bockelmann FD, Callies U, Gerdts G, Wichels A, Wiltshire KH, Glöckner FO, Schweder T, Amann R. 2012. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336:** 608-11.

Thiele S, Fuchs BM, Amann RI. 2011. Identification of microorganisms using the ribosomal RNA approach and fluorescence in situ hybridization. In: Wilderer P, editor. *Treatise on Water Science*, **vol. 3**. Elsevier Science. pp. 171-89.

Voget S, Wemheuer B, Brinkhoff T, Vollmers J, Dietrich S, Giebel H-A, Beardsley C, Sardemann C, Bakenhus I, Billerbeck S. 2015. Adaptation of an abundant *Roseobacter* RCA organism to pelagic systems revealed by genomic and transcriptomic analyses. *The ISME journal* **9**: 371-84.

Wemheuer B, Wemheuer F, Hollensteiner J, Meyer F-D, Voget S, Daniel R. 2015. The green impact: bacterioplankton response towards a phytoplankton spring bloom in the southern North Sea assessed by comparative metagenomic and metatranscriptomic approaches. *Front Microbiol* **6**: 805.

Wilhelm LJ, Tripp HJ, Givan SA, Smith DP, Giovannoni SJ. 2007. Natural variation in

SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol Direct* **2:** 27.

Williams TJ, Wilkins D, Long E, Evans F, DeMaere MZ, Raftery MJ, Cavicchioli R. 2013. The role of planktonic *Flavobacteria* in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. *Environ Microbiol* **15:** 1302-17.

Wiltshire KH, Kraberg A, Bartsch I, Boersma M, Franke H-D, Freund J, Gebühr C, Gerdts G, Stockmann K, Wichels A. 2010. Helgoland Roads, North Sea: 45 years of change. *Estuar Coast* **33**: 295-310.

Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. 2015. Environmental science. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347:** 1257594.

Xing P, Hahnke RL, Unfried F, Markert S, Huang S, Barbeyron T, Harder J, Becher D, Schweder T, Glockner FO, Amann RI, Teeling H. 2015. Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. *ISME J* **9:** 1410-1422.

Yang C, Li Y, Zhou B, Zhou Y, Zheng W, Tian Y, Van Nostrand JD, Wu L, He Z, Zhou J, Zheng T. 2015. Illumina sequencing-based analysis of free-living bacterial community dynamics during an Akashiwo sanguine bloom in Xiamen sea, China. *Sci Rep* **5:** 8476.

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

Yool A, Tyrrell T. 2003. Role of diatoms in regulating the ocean's silicon cycle. *Global Biogeochem Cy* **17**: 1003-24.

Zubkov MV, Fuchs BM, Tarran GA, Burkill PH, Amann R. 2003. High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Applied and Environmental Microbiology* **69**: 1299-304.

Chapter IV: Polysaccharide Utilization Loci of 53 North Sea Flavobacteriia as Basis for Using SusC/D-Protein Expression for Predicting Major Phytoplankton Glycans

In review for *ISME Journal*

Lennart Kappelmann[1], Karen Krüger[1], Jan-Hendrik Hehemann[1,2], Jens Harder[1], Stephanie Markert[3,4], Frank Unfried[3,4], Dörte Becher[5], Nicole Shapiro[6], Thomas Schweder[3,4*], Rudolf I. Amann[1*], Hanno Teeling[1*]

[1] Max Planck Institute for Marine Microbiology, Bremen, Germany
[2] Zentrum für Marine Umweltwissenschaften, Bremen, Germany
[3] Pharmaceutical Biotechnology, Ernst-Moritz-Arndt-University, Greifswald, Germany
[4] Institute of Marine Biotechnology, Greifswald, Germany
[5] Institute for Microbiology, Ernst-Moritz-Arndt-University, Greifswald, Germany
[6] DOE Joint Genome Institute, Walnut Creek, CA, USA

Contribution to the manuscript:

Experimental concept and design	45%
Acquisition of experimental data	10%
Data analysis and interpretation	90%
Preparation of figures and tables	100%
Drafting of manuscript	80%

* Corresponding authors:
Hanno Teeling, Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, email: hteeling@mpi-bremen.de, phone: +49 421 2028 976

Thomas Schweder, Ernst-Moritz-Arndt-University Pharmaceutical Biotechnology, Friedrich-Ludwig-Jahn-Str. 17, 17487 Greifswald, email: schweder@uni-greifswald.de, phone: +49 3834 420 4212

Rudolf I. Amann, Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, email: ramann@mpi-bremen.de, phone: +49 421 2028 930

E-mail addresses and telephone numbers of all authors:

Lennart Kappelmann	lkappelm@mpi-bremen.de	+49 421 2028545
Karen Krüger	kkrueger@mpi-bremen.de	+49 421 2028942
Jan-Hendrik Hehemann	jhhehemann@marum.de	+49 421 21865775
Jens Harder	jharder@mpi-bremen.de	+49 421 2028750
Stephanie Markert	stephanie.markert@uni-greifswald.de	+49 3834 4204892
Frank Unfried	fu072203@uni-greifswald.de	+49 3834 4204892
Dörte Becher	dbecher@uni-greifswald.de	+49 3834 4205903
Nicole Shapiro	nrshapiro@lbl.gov	+1 925 296 5767
Thomas Schweder	schweder@unigreifswald.de	+49 3834 4204212
Rudolf I. Amann	ramann@mpi-bremen.de	+49 421 2028930
Hanno Teeling	hteeling@mpi-bremen.de	+49 421 2028976

Conflict of interest

The authors declare no conflict of interest.

Subject category: Integrated genomics and post-genomics approaches in microbial ecology

Keywords: *Flavobacteriia;* marine Polysaccharides; North Sea; Phytoplankton; Proteomics; PUL; SusC/D

4.1 Abstract

Marine algae convert a substantial fraction of fixed carbon dioxide into various polysaccharides. *Flavobacteriia* that are specialized on algal polysaccharide degradation feature genomic clusters termed polysaccharide utilization loci (PULs). Since knowledge on extant PUL diversity is sparse, we sequenced the genomes of 53 North Sea *Flavobacteriia*. We obtained 400 PULs, suggesting usage of a large array of polysaccharides, including laminarin, α- and β-mannans, fucose-, xylose-, galactose-, rhamnose- and arabinose-containing substrates, pectins, and chitins. Many of the PULs exhibit new genetic architectures and suggest substrates rarely described for marine environments. The isolates' PUL repertoires often differed considerably within genera, corroborating ecological niche-associated glycan partitioning. Polysaccharide uptake in *Flavobacteriia* is mediated by SusCD-like transporter complexes. Respective protein trees revealed clustering according to polysaccharide specificities predicted by PUL annotations rather than phylogenetic affiliation. Using the trees, we analyzed expression of SusC/D homologs in multiyear phytoplankton bloom-associated metaproteomes and found indications for profound changes in microbial utilization of laminarin, α-glucans, ß-mannan and sulfated xylan. We hence suggest the suitability of SusC/D-like transporter protein expression within heterotrophic bacteria as a proxy for the temporal utilization of discrete polysaccharides.

4.2 Introduction

Half of global net primary production is oceanic and carried out mostly by small, unicellular phytoplankton such as diatoms (Field *et al.,* 1998). Polysaccharides account for up to 50% of algal biomass (Kraan *et al*., 2012) and can be found as intracellular energy storage compounds, as structural components of their cell walls (Kloareg and Quatrano, 1988), or as secreted extracellular transparent exopolymeric substances (TEP, Hoagland *et al.,* 1993). They can be composed of different cyclic sugar monomers linked by either α- or β-glycosidic bonds at different positions and can be substituted by different moieties (e.g. sulfate, methyl or acetyl groups), making them the most structurally diverse macromolecules on Earth (Laine, 1994).

Many members of the bacterial phylum *Bacteroidetes*, including marine representatives of the class *Flavobacteriia*, are specialized on polysaccharide degradation. They feature distinct polysaccharide utilization loci (PULs, Bjursell *et al.,* 2006), i.e. operons or regulons that encode the protein machinery for binding, degradation and uptake of a type or class of polysaccharides. Polysaccharides are initially bound by outer membrane proteins and cleaved by endo-active enzymes into oligosaccharides suitable for transport through the outer membrane. Oligosaccharides are bound at the interface of SusCD complexes. SusD-like proteins are extracellular lipoproteins and SusC-like proteins constitute integral membrane beta-barrels termed TonB-dependent transporters (TBDTs). Glenwright and colleagues (2017) showed that these two proteins form a “pedal bin” complex in *Bacteroides thetaiotaomicron*, with SusD acting as a lid on top of the SusC-like TBDT. Upon binding of a ligand, the SusD lid closes and conformational changes lead to substrate release into the periplasm. Here, further saccharification to sugar monomers takes place that are taken up into the cytoplasm via dedicated transporters.

Besides the characteristic *susCD*-like gene pair, *Bacteroidetes* PULs contain various substrate-specific carbohydrate-active enzymes (CAZymes), such as glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), carbohydrate binding modules (CBMs), and proteins with auxiliary functions. PULs of human gut *Bacteroidetes* and their capacity to degrade various land plant polysaccharides have been thoroughly investigated (e.g. Martens *et al.,* 2011), but knowledge on marine polysaccharide degradation is sparse. Many polysaccharides in marine algae differ from those in land plants. Green algae contain ulvans, red algae contain agars, carrageenans and porphyrans, brown algae contain alginates, fucans and laminarin, and diatoms contain chrysolaminarin and sulfated mannans, all of which are presumably absent in land plants (Popper *et al.,* 2011). Likewise, many algae feature anionic, sulfated polysaccharides that require sulfatases for degradation.

A systematic inventory of the structural diversity of algal polysaccharides has not yet been achieved. We do not have a good understanding of the associated diversity of PULs in marine *Bacteroidetes*. Also only few PULs have so far been linked to their polysaccharide substrate. Examples include an agar/porphyran-specific PUL (Hehemann *et al.,* 2012) that human gut *Bacteroidetes* acquired from marine counterparts (Hehemann *et al.,* 2010), an alginate-specific PUL in *Zobellia galactanivorans* DsiJ^{T} (Thomas *et al.,* 2012; Hehemann *et al.,* 2012), alginate- and laminarin-specific PULs in *Gramella forsetii* KT0803 (Kabisch *et al.,* 2014), and a similar laminarin-specific PUL in *Polaribacter* sp. Hel1_33_49 (Xing *et al.,* 2014). Few overarching comparative genomic studies exist (Xing *et al.,* 2014; Barbeyron *et al.,* 2016), focusing largely on overall CAZyme repertoires.

Pioneering studies on structural elucidation of polysaccharides from microalgae were performed (Ford and Percival, 1965; Rees and Welsh, 1977), but precise microalgal

polysaccharide structures remain mostly unresolved (for review see Hoagland *et al.*, 1993), because they require sophisticated methods (Le Costaouëc *et al.*, 2017). PUL analysis of heterotrophic bacteria co-occurring with phytoplankton could serve as an alternative starting point to advance insight into the structures of marine polysaccharides and to understand their microbial decomposition.

Here we present a comparative analysis of PULs from 53 newly sequenced *Flavobacteriia* isolated from the German Bight, comprising a total of 400 manually determined PULs. Based on these data we investigated whether SusC- and SusD-like sequences can be linked to distinct predicted polysaccharides. Using environmental metaproteome data we show how SusC/D homolog expression can be used to assess the presence of marine polysaccharides during North Sea spring blooms.

4.3 Materials and methods

4.3.1 Isolation and sequencing of North Sea Flavobacteriia

Flavobacteriia were sampled at the North Sea Islands Helgoland and Sylt as described previously (Hahnke and Harder, 2013; Hahnke *et al.,* 2015, Supplementary Table S1). Also included were the previously sequenced *Gramella forsetii* KT0803 (Bauer *et al.,* 2006), *Polaribacter* spp. Hel1_33_49 and Hel1_85 (Xing *et al.,* 2014) and the *Formosa* spp. Hel1_33_131 and Hel3_A1_48. The remaining 48 genomes were sequenced at the Department of Energy Joint Genome Institute (DOE-JGI, Walnut Creek, CA, USA) in the framework of the Community Sequencing Project No. 998 COGITO (Coastal Microbe Genomic and Taxonomic Observatory). 40 genomes were sequenced using the PacBio RSII platform exclusively, while eight isolates were sequenced using a combination of Illumina HiSeq 2000/2500 and PacBio RSII. All these genomes are GOLD certified at level

3 (improved high-quality draft) and are publicly available at the DOE-JGI Genomes OnLine Database (GOLD, Reddy *et al.,* 2015) under the GOLD Study ID Gs0000079.

4.3.2 Gene and PUL annotation

Initial annotations of the genomes of *Polaribacter* spp. Hel1_44_49 and Hel1_85 and *Formosa* spp. Hel1_33_131 and Hel3_A1_48 were performed using the RAST annotation system (Aziz *et al.,* 2008). All other genomes were annotated using the DOE-JGI Microbial Annotation Pipeline (MGAP, Huntemann *et al.,* 2015). These annotations were subsequently imported into a GenDB v2.2 annotation system (Meyer *et al.,* 2003) for refinement and additional annotations based on similarity searches against multiple databases as described previously (Mann *et al.,* 2013).

SusC and SusD-like proteins were annotated by the DOE-JGI MGAP, which uses the TIGRfam model TIGR04056 to detect SusC-like proteins and the Pfam models 12741, 12771, and 14322 to detect SusD-like proteins. CAZymes were annotated based on HMMer searches against the Pfam v25 (Finn *et al.,* 2013) and dbCAN 3.0 (Yin *et al.,* 2012) databases and BLASTp searches (Altschul *et al.,* 1990) against the CAZy database (Lombard *et al*., 2014). CAZymes were annotated only as such when at least two of the database searches were positive based on family-specific cutoff criteria that were described previously (Teeling *et al.,* 2016). Peptidases were annotated using BLASTp searches against the MEROPS 9.13 database (Rawlings *et al.,* 2012) using the default settings of $E \leq 10^{-4}$.

PULs were manually detected based on the presence of CAZyme clusters, which in most cases also featured co-occurring *susCD-like* gene pairs previously suggested (Bjursell *et al.,* 2006). In some cases, the sequence similarity of a TBDT was too low to be considered

SusC-like, no SusD homolog was present or the entire *susCD*-like gene tandem was missing. These operons were still counted as PULs and are regarded as incomplete sub-types (Hemsworth *et al.,* 2016).

4.3.3 Gene expression analyses of Flavobacteriia-rich North Sea bacterioplankton using metaproteomics

During spring phytoplankton blooms of 2009 to 2012, 16 surface seawater biomass samples were collected at the long-term ecological research station 'Kabeltonne' (54° 11.3' N, 7° 54.0' E) off the German North Sea island Helgoland as previously described in detail (Teeling *et al.*, 2012 and 2016). Biomass was collected on 0.2 µm pore sized filters after pre-filtration with 10 and 3 µm pore sized filters. Metagenome sequencing was done using the 454 FLX Ti platform for 2009 and the Illumina HiSeq 2000 platform for 2010 to 2012 samples (Teeling *et al.,* 2016).

Corresponding metaproteome analyses were performed from biomass obtained from the same water samples. Protein extraction from 0.2 µm filtered bacterioplankton biomass and separation was carried out as described previously (Teeling *et al.,* 2012) with the modification that gel lanes were cut into 10 equal pieces prior to tryptic digestion (1 µg/ml, Promega, Madison WI, USA) and subsequent mass spectrometric detection in an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher, Bremen, Germany). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Vizcaíno *et al.,* 2016; dataset identifiers: PXD008238, 10.6019/PXD008238).

Mass spectrometric data was analyzed using Sequest v27r11 (Thermo Fisher Scientific, San Jose, CA, USA). Searches were carried out against a forward-decoy database of all

proteins from all metagenome samples combined. This non-redundant database was constructed from all predicted protein-coding genes of all metagenomes (6,194,278 sequences) using the uclust option of USEARCH v6.1.544 (Edgar, 2010; options: cluster_fast; nucleotide identity 0.99; maxhits 5; maxrejects 30) and contained 3,212,324 sequences. Common laboratory contaminants were included in all databases. Technical duplicates of each sample were searched together (including all 20 subsamples) to obtain averaged spectral counts. Validation of protein- and peptide identifications was performed with Scaffold v4 (Proteome Software Inc, Portland, OR, USA) using the parameters previously described (Teeling *et al.,* 2012), and normalized spectral abundance factors (%NSAF) were calculated (Zhang *et al.,* 2010) to allow for semi-quantitative analyses (Supplementary Table S2).

4.3.4 SusC/D homolog tree reconstruction

We constructed trees from SusC- and SusD-like protein sequences of the isolates' PULs (SusC: 369; SusD: 361). Sequences were aligned using MAFFT v7.017 (Katoh *et al.,* 2013) using the G-INS-i algorithm and BLOSUM62 matrix with default gap open penalty (1.53) and offset (0.123) values. Maximum-likelihood trees were constructed using FastTree 2.1.5 (Price *et al.,* 2010) with default settings.

4.4 Results

4.4.1 High genomic and phylogenetic diversity in isolated marine Flavobacteriia

The 53 flavobacterial isolates cover a broad range of the *Flavobacteriia* class within the phylogenetic tree based on full length 16S rRNA genes (Figure 1). The strains fall into several clusters that can be linked to characteristic genomic features (Supplementary

Table S1). Genome sizes ranged from 2.02 Mbp (*Formosa* sp. Hel3_A1_48) to 5.98 Mbp (*Aquimarina* sp. MAR_2010_214), with an average of 3.83 Mbp. One of the clusters was dominated by macroalgae-associated isolates (8 out of 12; Figure 1) and contained mostly larger genomes (average 4.5 Mbp). 47 of the 53 strains have two to four 16S rRNA operons, with the notable exception of the three *Tenacibaculum* strains possessing six (strains MAR_2009_124 and MAR_2010_205) and seven (strain MAR_2010_89), respectively.

The capacity of the isolates to degrade polysaccharides varied widely as indicated by the number of degradative CAZymes per Mbp and predicted PULs per genome. On average, we identified 7.5 PULs per genome and 55 degradative CAZymes (Supplementary Table S1). Strains of the macroalgae-associated cluster differed with on average 83.3 degradative CAZymes, almost twice as many PULs per genome (14.2) and many sulfatase genes, indicating an extended capacity for the degradation of sulfated polysaccharides (average of 28.2 sulfatases, with a maximum of 95 sulfatases in *Zobellia amurskyensis* MAR_2009_138). The other strains had an average of 46.8 degradative CAZymes and 5.5 PULs. Eleven isolates possessed less than three PULs, contained few (≤3) or no sulfatases and were exclusively isolated from surface seawater or pore water. They likely target rather simple, non-sulfated polysaccharides and peptides. This strategy is emphasized by their high peptidase:CAZyme ratio of 1.81, compared to an average ratio of 0.95 for isolates with more than 10 PULs. Still it is noteworthy that numbers of PULs and degradative CAZymes varied considerably, even within isolates of the same genus.

	Organism	Isolation source	Cultivation medium	Genome size [Mbp]	16S rRNA operons	No. of PULs	Sulfa-tases	CAZymes	Degrad. CAZYmes (GH, PL, CE)	Pepti-dases	Peptidase: CAZyme ratio
	Polaribacter sp. Hel1_88	seawater	HaHa agar	4.0	3	7	26	127	70	167	1.31
	Polaribacter sp. KT25b	seawater	ASM	4.0	3	11	40	146	90	142	0.97
	Polaribacter sp. Hel1_85	seawater	HaHa agar	3.9	4	8	31	139	73	162	1.17
	Polaribacter sp. Hel1_33_49	seawater	liquid Haha	3.0	3	5	12	105	50	136	1.30
	Polaribacter sp. Hel1_33_78	seawater	liquid Haha	3.2	3	6	7	107	51	149	1.39
	Polaribacter sp. Hel1_33_96	seawater	liquid Haha	3.1	4	4	7	104	50	135	1.30
	Tenacibaculum sp. MAR_2009_124	particle net	NAG	5.5	6	7	8	117	55	226	1.93
	Tenacibaculum sp. MAR_2010_89	porewater	HaHa agar	4.2	7	2	1	84	26	200	2.38
	Tenacibaculum sp. MAR_2010_205	seawater	HaHa agar	3.2	6	1	1	47	15	128	2.72
	Lutibacter sp. Hel1_33_5	seawater	liquid Haha	3.1	2	4	4	97	38	131	1.35
	Flavobacteriaceae bacterium sp. Hel1_10	seawater	HaHa agar	4.1	2	10	1	182	88	167	0.92
	Flavobacteriaceae bacterium sp. MAR_2010_10	porewater	HaHa agar	3.3	2	1	0	88	19	177	2.01
	Flavobacteriaceae bacterium sp. MAR_2010_118	porewater	HaHa agar	3.3	2	1	0	97	23	202	2.08
	Lacinutrix sp. Hel1_90	seawater	HaHa agar	3.8	3	4	1	109	35	169	1.55
	Olleya sp. Hel1_94	seawater	HaHa agar	3.6	6	5	0	112	37	145	1.29
	Winogradskyella sp. RHA_55	mussel surface	2216E	3.7	3	6	4	131	40	162	1.24
	Psychroserpens sp. Hel1_66	seawater	HaHa agar	3.8	2	6	2	146	48	173	1.18
	Flavobacteriaceae bacterium sp. MAR_2010_105	porewater	HaHa agar	3.3	2	11	4	133	64	158	1.19
	Flavobacteriaceae bacterium sp. MAR_2010_119	porewater	HaHa agar	3.2	2	9	2	119	50	162	1.36
	Flavobacteriaceae bacterium sp. MAR_2010_188	porewater	HaHa agar	3.8	2	6	1	148	49	196	1.32
	Formosa sp. Hel1_33_131	seawater	liquid Haha	2.7	2	5	7	63	24	114	1.81
	Formosa sp. Hel3_A1_48	seawater	liquid Haha	2.0	2	7	16	66	33	91	1.38
	Formosa sp. Hel1_31_208	seawater	liquid Haha	3.1	2	1	0	92	25	181	1.97
	Dokdonia sp. Hel1_5	seawater	HaHa agar	3.6	3	0	0	98	19	182	1.86
	Dokdonia sp. Hel1_53	seawater	HaHa agar	2.9	4	0	0	89	22	140	1.57
	Dokdonia sp. Hel1_63	seawater	HaHa agar	3.5	6	2	0	103	28	156	1.51
	Aquimarina sp. MAR_2010_214	seawater	HaHa agar	6.0	3	9	0	133	55	269	2.02
Macroalgae associated cluster	*Muricauda* sp. MAR_2010_75	seawater	HaHa agar	4.4	2	10	5	140	72	194	1.39
	Flavobacteriaceae bacterium sp. MAR_2009_75	particle net	SYL	4.8	2	22	44	189	112	186	0.98
	Zobellia amurskyensis MAR_2009_138	particle net	cellobiose	5.4	3	40	95	236	154	181	0.77
	Maribacter forsetii DSM_18668	seawater	2216E	4.5	3	9	12	118	59	195	1.65
	Maribacter dokdonensis MAR_2009_71	particle net	NAG	4.6	3	9	28	145	77	207	1.43
	Maribacter dokdonensis MAR_2009_60	particle net	NAG	4.5	3	8	20	146	74	189	1.29
	Maribacter sp. MAR_2009_72	particle net	glucose	4.3	3	9	14	159	82	197	1.24
	Maribacter sp. Hel1_7	seawater	HaHa agar	4.8	4	12	8	144	72	204	1.42
	Arenibacter palladensis MAR_2009_79	particle net	glucose	5.4	3	25	60	189	117	186	0.98
	Cellulophaga sp. Hel1_12	seawater	HaHa agar	4.0	3	5	11	96	46	202	2.10
	Cellulophaga sp. RHA_19	*P. lanosa*	CAA	3.9	4	12	27	133	73	143	1.08
	Cellulophaga sp. RHA_52	particle net	2216E	3.7	4	9	14	119	62	145	1.22
	Ulvibacter sp. MAR_2010_11	porewater	HaHa agar	3.0	2	0	3	83	16	156	1.88
	Gramella forsetii KT0803	seawater	ASM	3.8	3	10	2	151	63	160	1.06
	Gramella sp. MAR_2010_102	porewater	HaHa agar	3.5	3	5	2	140	46	169	1.21
	Gramella sp. MAR_2010_147	porewater	HaHa agar	3.2	3	7	5	128	54	150	1.17
	Gramella sp. Hel1_59	seawater	HaHa agar	3.4	3	3	0	128	42	163	1.27
	Salegentibacter sp. Hel1_6	seawater	HaHa agar	4.2	3	9	1	178	81	167	0.94
	Gillisia sp. Hel1_29	seawater	HaHa agar	4.0	4	4	0	125	30	167	1.34
	Gillisia sp. Hel1_33_143	seawater	liquid Haha	3.5	5	3	1	119	31	157	1.32
	Gillisia sp. Hel1_86	seawater	HaHa agar	4.2	3	8	0	135	52	161	1.19
	Leeuwenhoekiella sp. MAR_2009_132	seawater	HaHa agar	4.3	4	17	6	205	117	160	0.78
	Flavimarina sp. Hel1_48	particle net	cellobiose	4.1	2	15	0	214	104	174	0.81
	Nonlabens sp. Hel1_56	seawater	HaHa agar	4.0	3	2	1	116	40	146	1.26
	Nonlabens sp. Hel1_38	seawater	HaHa agar	3.0	2	1	0	80	20	136	1.70
	Nonlabens sp. Hel1_33_55	seawater	liquid Haha	3.3	2	4	8	117	44	144	1.23
	min			2.02	2	0	0	47	15	91	0.77
	max			5.98	7	40	95	236	154	269	2.72
	average			**3.83**	**3.2**	**7.5**	**10.2**	**126.7**	**55.0**	**167.2**	**1.41**

Supplementary Table S4.1: Selected genomic information on the 53 North Sea *Flavobacteriia* of this study. Isolate order corresponds to the 16S rRNA phylogenetic tree (Figure 4.1A).

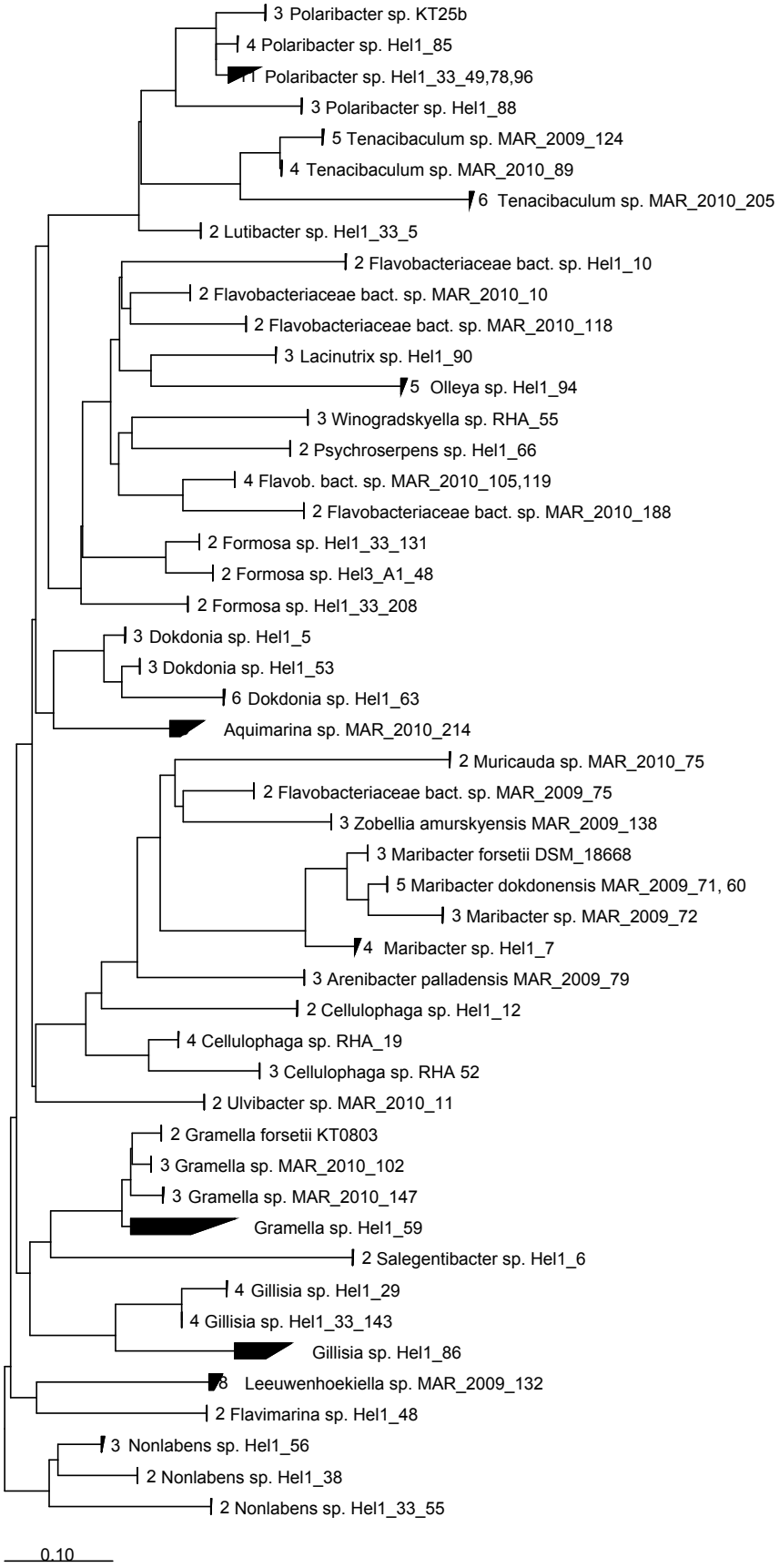


Figure 4.1A: Maximum-likelihood tree of 53 North Sea *Flavobacteriia* isolates based on full-length 16S rRNA gene sequences. Scale bar: 10 nucleotide substitutions per 100 nucleotides.

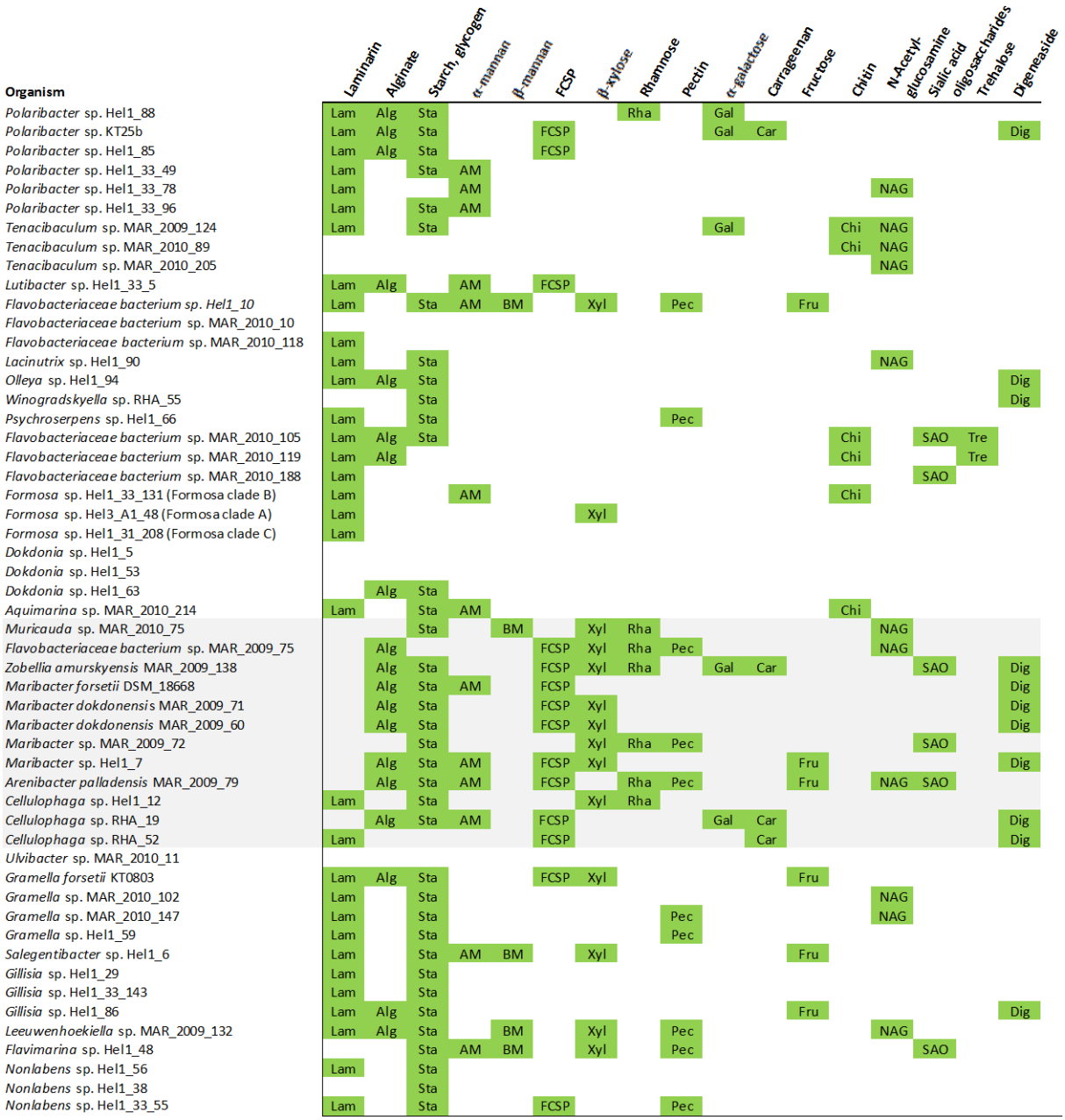


Figure 4.1B: Predicted degradation capacities of polysaccharide classes based on PUL-associated CAZyme annotations. Isolate order corresponds to the 16S rRNA phylogenetic tree (Figure 4.1A).

4.4.2 Substrate specificities

The 53 genomes revealed a wide range of as yet undescribed PULs. In total, 400 PULs were annotated, 259 of which could be linked to either dedicated polysaccharides or polysaccharide classes by in-depth annotations (Supplementary Table S3).

4.4.3 Laminarin

Laminarins are β-1,3-linked glucans that are abundant since they act as storage compounds in diatoms and brown algae. 46 PULs (~12%) could be linked to laminarin degradation, featuring four variants (Figure 2): Variant A is a highly conserved, short PUL containing one predicted GH3 β-1,3-glucosidase framed by two GH16 endo-active β-1,3(4)-glucanases (Figure 2A). This arrangement was first described in *Gramella forsetii* KT0803 and shown to be up-regulated by laminarin (Kabisch *et al.,* 2014). It was speculated that a cell surface-associated GH16 glucanase cleaves branched laminarin polysaccharides into oligosaccharides (Labourel *et al*., 2015), which can be transported through the SusC-like TBDT into the periplasm. Here, the GH3 β-1,3-glucosidase may further cleave off glucose units (Tamura *et al.,* 2017), which are imported into the cytoplasm.

Variant B is a larger, more variable PUL (Figure 2B) shown to be induced by laminarin in *Polaribacter* sp. Hel1_33_49 (Xing *et al.,* 2014). This PUL additionally features a GH30 exo-β-1,6-glucanase and at least two GH17 β-1,3-glucan hydrolases with predicted endo- and exo-activities, respectively. The GH30 exo-β-1,6-glucanase removes β-1,6-glucose side chains from laminarin (Becker *et al.,* 2017). While GH16 enzymes can hydrolyze both β-1,3- and β-1,4-linked glucans, GH17 glucan hydrolases are highly specific to undecorated β-1,3 glucans and can have endo- (Reese and Mantels, 1959) and exo-activity (Barras and Stone, 1969). The β-1,3-glucan endohydrolase thus likely cleaves laminarin into oligosaccharides, which may be further degraded into glucose by the β-1,3-glucan exohydrolase.

Variants C and D PULs are likewise predicted to be capable of laminarin degradation but have not been described before (Figure 2C and D). They feature an additional putative

GH5 glucan hydrolase with a carbohydrate-binding domain that binds β-1,3- and β-1,4-glucans (CBM6c; Michel *et al.,* 2009). They furthermore contain GH16 and GH30 family enzymes as described in variant B, but no GH17 enzymes.

In total, almost two thirds (33/53) of all isolates and 78% (25/32) of surface water isolates contained at least one laminarin PUL. Variant A occurred 21 times, B 17 times, C five times and D two times (Supplementary Table S3). Eight isolates possessed two laminarin PULs (*Flavobacteriaceae bacterium* spp. MAR_2010_105 and MAR_2010_119, *Gramella* sp. MAR_2010_102, *Polaribacter* spp. Hel1_33_49, 78, 96 and Hel1_88 and *Psychroserpens* sp. Hel1_66) and three isolates contained three (*Formosa* spp. Hel3_A1_48 and Hel1_33_131, *Flavobacteriaceae bacterium* sp. Hel1_10). In contrast, laminarin PULs were far less prevalent in isolates of the macroalgae-associated cluster (2/12). Laminarins are composed of a β-1,3-glucan backbone ramified by β-1,6 and, less frequently, β-1,2-linked glucose side chains (Gügi *et al.,* 2015). The backbone length and ramification degree varies in different species. Laminarin of brown algae is capped at the reducing end by a 1-linked D-mannitol (Read *et al.,* 1996). Only three isolates with laminarin PULs, namely the *Polaribacter* spp. Hel1_85 and KT25b and *Gramella* sp. MAR_2010_102, also possessed an annotated mannitol-2-dehydrogenase, which likely enables them to utilize brown algal laminarin. All other isolates probably only target diatom-type non-mannitol-capped chrysolaminarins, indicating that these are the major available laminarins in the southern North Sea.

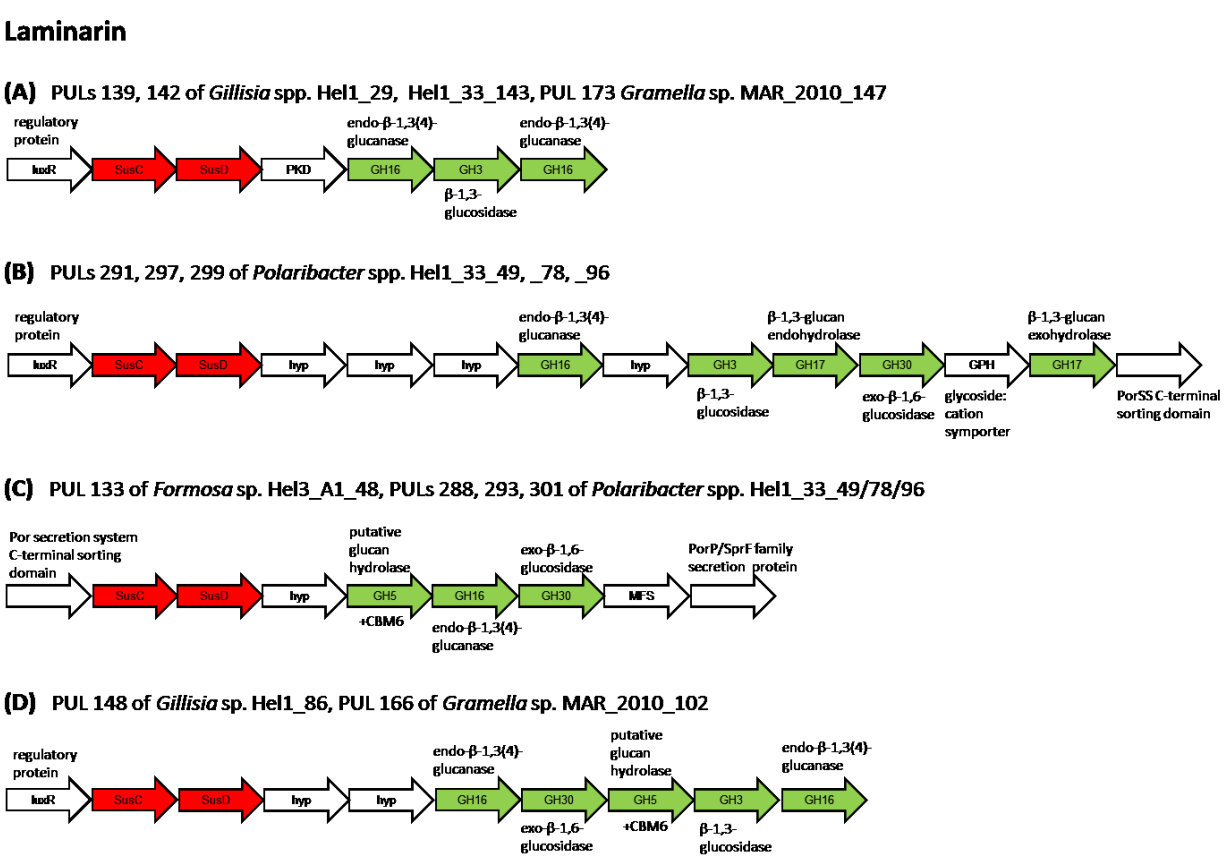


Figure 4.2: Four variants of conserved PULs predicted to target laminarin.

4.4.4 α-1,4-glucan (starch, glycogen)

PULs targeting α-1,4-glucans, such as starch, glycogen and amylose were also highly abundant (43/400 PULs, 37/53 isolates; Supplementary Table S3). Respective PULs often featured a *susCDE*-like gene triplet, at least one predicted GH13 α-glycosidase and frequently GH65 phosphorylases and GH31 hydrolases acting on α-glucosidic linkages (Figure 3A). In some cases, these PULs also included a GH97 family gene, known to hydrolyze diverse α-1,2-, α-1,3-, α-1,4- and α-1,6-linked glycosidic bonds (Smith *et al.,* 1991; Kitamura *et al.* 2008). A similar PUL was first described for *Gramella forsetii* KT0803 and found to be up-regulated under glucose-polymer substrates (Kabisch *et al.,* 2014). The PUL depicted in Figure 3A likely facilitates utilization of α-1,4-glucans featuring α-1,6-branches, such as the starch molecule amylopectin or potentially bacterial glycogen. Contrastingly, some isolates featured reduced versions of this PUL without *susE* gene and

only one GH13 and GH65 gene, respectively (e.g. all *Maribacter* isolates). These isolates may only target non-branched α-1,4-glucans such as maltodextrin or amylose.

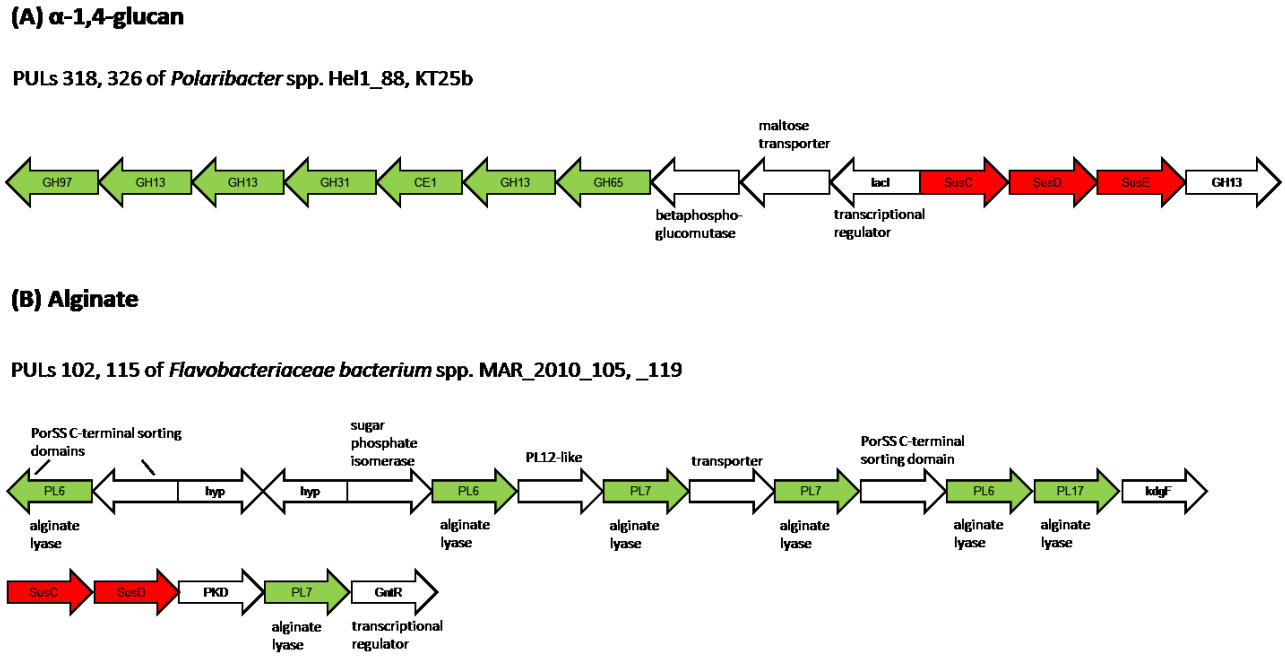


Figure 4.3: PULs predicted to target (A) α-1,4-glucans and (B) alginate.

4.4.5 Alginate

Twenty isolates featured in total 27 alginate-specific PULs (Figure 1). Alginate consists of blocks of β-D-mannuronic acid and α-L-guluronic acid, forming a linear β-1,4-linked chain (Fischer and Dörfel, 1955). Corresponding PULs encode family PL6, 7 and 17 alginate lyases (Figure 3B). Six of the alginate PULs also contained genes with sequence similarities to the sparsely investigated PL12, whose heparin-sulfate lyase activity has not been described in alginate PULs yet. Heparin and alginate are both linear, β-1,4-linked C5-uronans (Arlov and Skjåk-Bræk, 2017). Sulfated alginates have not yet been reported in nature, but are artificially synthesized for potential use as therapeutics and biomaterials (Arlov *et al.,* 2014; Mhanna *et al.,* 2017). No PL15 and only one potential PL14 family alginate lyase (*Lutibacter* sp. Hel1_33_5, not PUL-associated) were annotated. The

macroalgae-associated cluster had a higher prevalence (8/12) of alginate PULs as compared to the other isolates (12/41) (Figure 1).

4.4.6 Mannan

Seventeen isolates showcased PULs dominated by mannose-targeting CAZymes (Figure 1, Supplementary Table S3). Mannan PULs are less conserved in terms of gene arrangements and can be separated into α-mannan- (13/400) and β-mannan-targeting PULs (5/400). Alpha-mannan-PULs can be further subdivided into two types: (A) PULs containing multiple GH92 exo-acting α-mannosidases (α-1,2/3/4/6) that are often rich in sulfatase genes (Figure 4A), hinting at a sulfated α-mannan (e.g. Supplementary Table S3: PULs 289, 296, 300 of *Polaribacter* spp. Hel1_33_49/78/96; PUL126 of *Formosa* sp. Hel1_33_131) and (B) PULs with α-mannan-targeting CAZymes of diverse additional families, such as GH76 endo-α-1,6-mannanases, GH125 exo-α-1,6-mannosidases or GH38 α-mannosidases (α-1,2/3/6). These type (B) α-mannan PULs are notably devoid of sulfatase-coding genes, indicating an unsulfated α-mannan substrate (Figure 4B). A PUL with a similar CAZyme repertoire in *B. thetaiotaomicron* facilitates the utilization of yeast cell wall α-mannan (Cuskin *et al.,* 2015). Type (A) sulfatase- and GH92-rich PULs have been observed previously in *Polaribacter*-affiliated North Atlantic fosmids (Gómez-Pereira *et al.,* 2012) and *Polaribacter* sp. Hel1_33_49 (Xing *et al.,* 2014). However, we found a PUL that contains additional GH2, 3 and 88 family enzymes (Figure 4A). While GH families 2 and 3 are functionally diverse, GH88 enzymes are unsaturated β-glucuronyl hydrolases. This functional combination of CAZymes strongly suggests degradation of α-glucomannans such as glucuronomannan, a polysaccharide that has been reported for diatoms (Percival *et al.,* 1965; Gügi *et al.,* 2015, Le Costaouëc *et al.,* 2017) and brown algae (Wu *et al.,* 2015). Finally, co-located peptidases and a gene distantly related to GH99, a family which is reported to contain glycoprotein endo-α-mannosidases, indicate

that this hypothesized glucuronomannan substrate might be a glycoprotein. Le Costaouëc and colleagues (2017) recently revealed the main cell wall polysaccharide of the diatom *Phaeodactylum tricornutum* and possibly many other diatoms (Chiovitti *et al.,* 2005) to be a 'linear poly-α-1-3-mannan decorated with sulfate ester groups and β-D-glucuronic residues'.

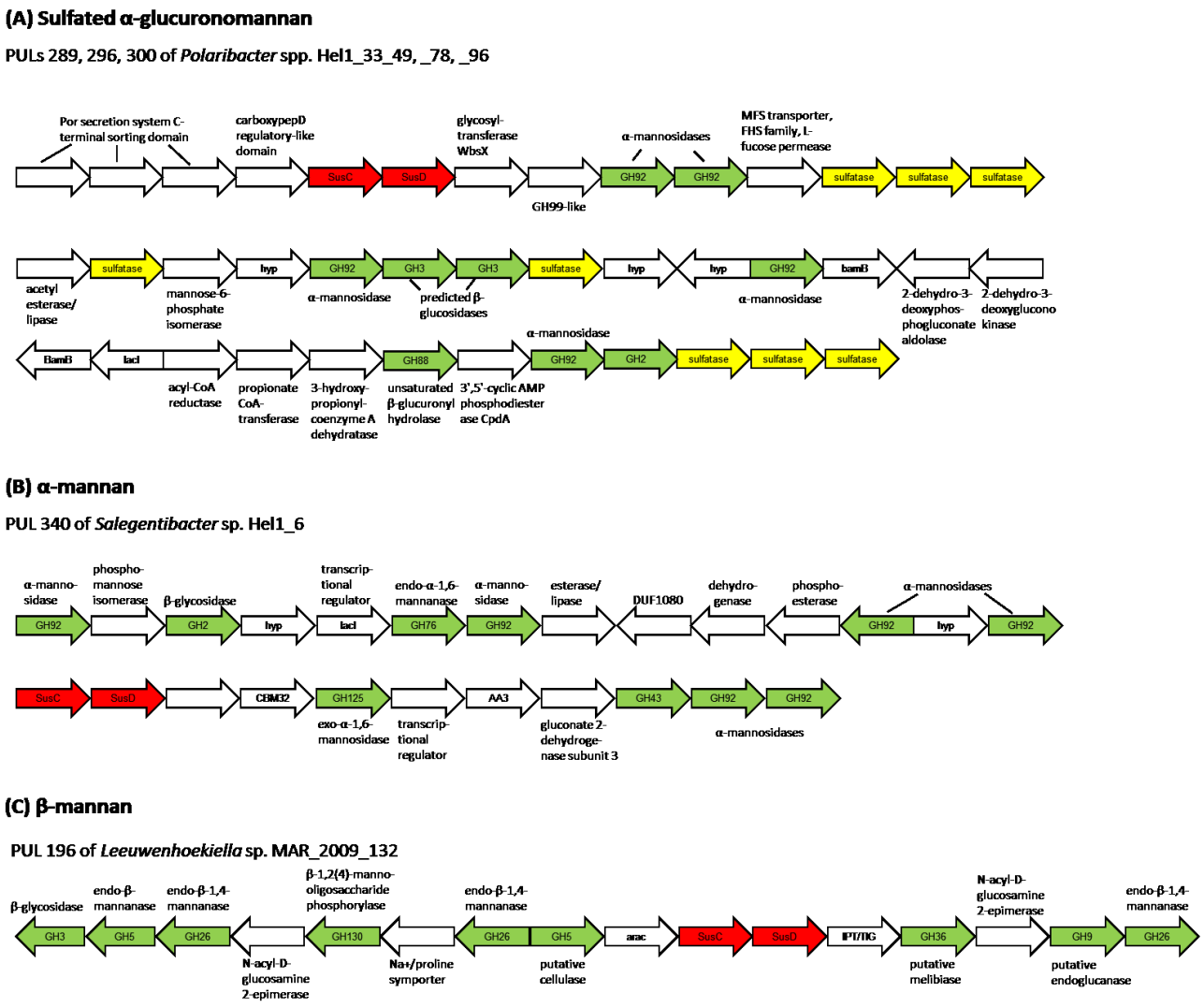


Figure 4.4: Selected PUL variants predicted to target (A, B) α-mannans and (C) β-mannans.

The β-mannan-PULs (Figure 4C) all contained GH130 β-1,2(4)-mannooligosaccharide phosphorylases and GH26 CAZymes which are primarily composed of endo-β-1,4-mannanases (Supplementary Table S3: PUL 6 of *Flavobacteriaceae bacterium* sp.

Hel1_10; PUL 196 of *Leeuwenhoekiella* sp. MAR_2009_132; PUL 211 of *Flavimarina* sp. Hel1_48; PUL 266 of *Muricauda* sp. MAR_2010_75; PUL 342 of *Salegentibacter* sp. Hel1_6). Beta-mannans have been reported in the red alga *Porphyra umbilicalis* and in various species of the green alga *Codium*. Moreira and Filho (2008) proposed that 'in some algae species, linear (beta-) mannan seems to replace cellulose as the main cell wall glycan'.

4.4.7 Fucose-containing sulfated polysaccharides (FCSP)

Twenty PULs in fourteen isolates suggest that FCSPs are also common substrates to marine *Flavobacteriia*. A prominent substrate of this group is fucoidan, a highly diverse polysaccharide prominent in brown algae. It contains L-fucose and sulfate ester groups due to its backbone of α-1,3 or alternating α-1,3/1,4 linked L-fucopyranosyl residues (Ale *et al.,* 2011). This backbone has side chains containing diverse other monosaccharides, uronic acids, acetyl groups and proteins (Li *et al.,* 2008). In accordance with the structural complexity of FCSPs, PULs display equally complex gene repertoires, averaging 38 genes/PUL. A relatively short PUL is shown in Figure 5A. Characteristic CAZymes of FCSP PULs in the isolated *Flavobacteriia* were GH29 and GH95 family α-L-fucosidases or potentially α-L-galactosidases. Other regularly occurring CAZymes included GH117 family enzymes, β-xylosidases mostly of the family GH43, but also GH30, 39 and 120, and diverse α- and β-glucosidases of the families GH2, 3, 31 and 97. Sulfated FCSPs such as xylofucoglucans or -glucuronans have been reported for brown algal hemicelluloses (Kloareg *et al.,* 1988; Popper *et al.,* 2011) and might also occur in diatoms (Gügi *et al.,* 2015).

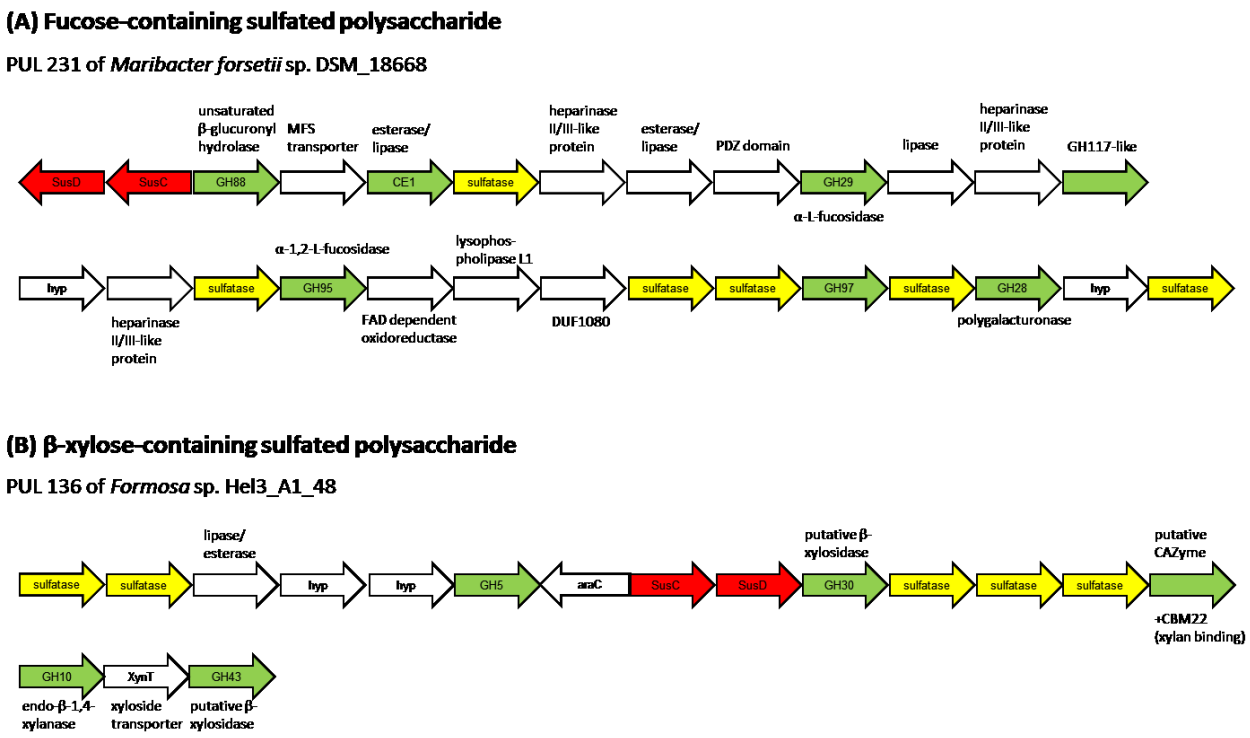


Figure 4.5: PULs predicted to target sulfated substrates rich in (A) fucose (FCSP) and (B) xylose.

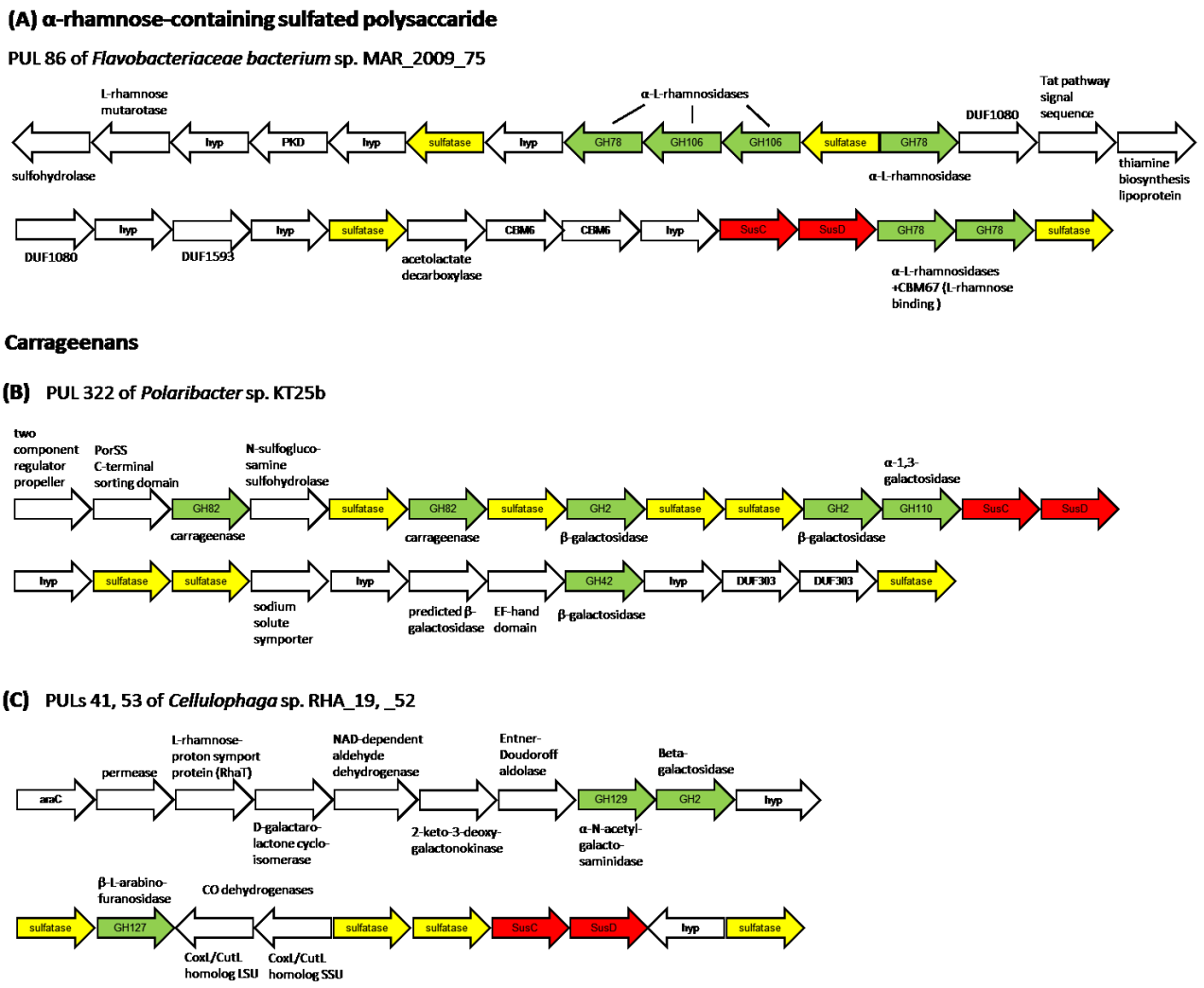
4.4.8 β-xylose-containing substrates

Twenty-two PULs targeting β-xylose-rich substrates were found in 14 isolates (Figure 5B, Supplementary Table S3). Likely substrates are heterogeneous β-xylans such as arabinoxylans, glucuronoxylans and sulfated xyloglucans. These PULs encode GH10 and GH43 enzymes targeting xylans, as well as GH30, 115 and 67 enzymes. GH10 enzymes are endo-β-1,4-xylanases capable of cleaving large β-1,4-β-xylan backbones into oligosaccharides. GH30 and GH43 enzymes have broader degradation capacities, and while both families contain β-xylosidases, they were also reported to target mixed xylose-containing substrates such as arabinoxylan and α-L-arabinofuranosides (GH43) or even completely different substrates such as β-glucosylceramidase or β-1,6-glucanase (GH30). GH67 and 115 can cleave glucuronic acid side chains from native xylans and are present in four PULs that likely target glucuronoxylans (PULs 243, 256 of *Maribacter* spp. Hel1_7,

MAR_2009_72; PUL 269 of *Muricauda* sp. MAR_2010_75; PUL 338 of *Salegentibacter* sp. Hel1_6; Supplementary Table S3). Two PULs were predicted to target arabinoxylans through a GH51 α-L-arabinofuranosidase (PUL 155 of *Gramella forsetii* KT0803; PUL 199 of *Flavimarina* sp. Hel1_48; Supplementary Table S3). Four PUL targeting β-xylose-rich substrates encode sulfatases (PUL136 of *Formosa* sp. Hel1_33_131, Figure 5B; PULs 363, 364, 366 of *Zobellia amurskyensis* MAR_2009_138; Supplementary Table S3). Marine xylans have been reported as hemicellulose components in green (Sørensen *et al.,* 2011), red and brown algae (Painter, 1983; Kloareg *et al.,* 1988), and as cell wall components in some diatoms (Wustman *et al.,* 1998; Murray *et al.,* 2007).

4.4.9 Sulfated α-rhamnose-containing substrates

In seven isolates we detected eleven PULs likely targeting sulfated α-rhamnose-containing substrates. These PULs feature predicted GH78 α-L-rhamnosidases and were often complemented by GH106 family α-L-rhamnosidases (Supplementary Figure S1A). Sulfated rhamnans have been reported for green algae (Wang *et al.,* 2014). Some of these rhamnose-PULs additionally contained predicted GH105 family rhamnogalacturonyl hydrolases, which cleave rhamnose from uronic acids. As GH78 family enzymes have also been shown to act on rhamnogalacturonans, it seems likely that these GH105 complemented rhamnose-PULs target rhamnogalacturonans, e.g. the matrix polysaccharide ulvan in green algae (Lahaye *et al.,* 2007).



Supplementary Figure S1A-C: PULs predicated to target a (A) sulfated, rhamnose-rich substrate and (B, C) sulfated, galactose-rich carrageenans

4.4.10 Sulfated α-galactose-containing substrates

Four PULs possessed characteristic GH110 family α-galactosidases, were rich in sulfatases and likely target sulfated, galactose-rich substrates. Sulfated galactans have been described in red algae and seagrass (Aquino *et al.,* 2005). Two of those PULs additionally featured GH82 family carrageenases and predicted GH2 family β-galactosidases (Supplementary Figure S1B), indicating degradation of the red algal cell wall constituent carrageenan that consists of D-galactose-4-sulfate and 3,6-anhydro-d-galactose-2-sulfate units bound by alternating α-1,3 and β-1,4 linkages (Michel *et al.,*

2001). A different carrageenan is likely targeted in three different sulfatase-rich PULs (PULs 43, 51 of *Cellulophaga* spp. RHA19, _52, PUL 361 of *Zobellia amurskyensis* MAR_2009_138) that possess a combination of CAZymes distantly related to families GH127 and GH129 and a GH2 β-galactosidase (Supplementary Figure S1C). A similar, carrageenan-specific PUL, which lacks the GH2 and *susCD*-like genes in the locus itself, was recently characterized in *Z. galactanivorans* DsiJ^T (Ficko-Blean *et al.,* 2017). The GH127 and GH129 proteins encoded in that PUL showed exo-lytic α-1,3-(3,6-anhydro)-D-galactosidase activity on carrageenan oligosaccharides, hence displaying novel enzymatic activities for the two CAZyme families.

4.4.11 Pectin

GH105 family CAZymes also frequently occur in another type of PUL which is equipped with a rich set of CAZymes suggesting degradation of pectin. 14 PULs (Supplementary Table 3) were identified in ten isolates (Supplementary Table 3), which characteristically feature GH28 α-1,4-polygalacturonases and GH88 unsaturated β-glucuronyl hydrolases, pectate lyases of the families PL1, PL9 and PL10 and carbohydrate esterase of the families CE8 and CE12 (Supplementary Figure S1D). CE8 are pectin methylesterases and CE12 family enzymes have been described as pectin acetylesterases, rhamnogalacturonan acetylesterases and acetyl xylan esterases. All CAZymes observed in this type of PUL were identified in fungal degradation of pectins (Benoit *et al.*, 2015). Barbeyron and colleagues found a PUL with similar pectin-specific CAZymes (PL1 and GH28) in the marine flavobacterium *Z. galactanivorans* DsiJ^T, yet no pectin degradation could be shown in culture. They hypothesized that these enzymes could therefore display new, uncharacterized specificities towards an as yet undescribed, pectin-like marine substrate.

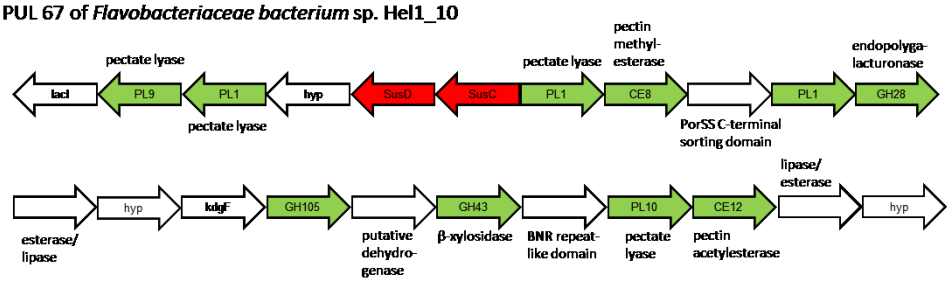
4.4.12 Further potential substrates: Digeneaside, N-acetylglucosamine, chitin, fructose, arabinan and trehalose

The flavobacterial isolates featured plenty of additional PULs (Supplementary Table 3). Eleven isolates can assimilate digeneaside (Supplementary Figure S1E), which occurs in red algae exudates (Kremer, 1980), through a PUL-associated GH63 family mannosylglycerate hydrolase and a co-occurring glycerate kinase, which convert digeneaside into D-mannose 6-phosphate and 3-phospho-D-glycerate (Barbeyron *et al*., 2016). This PUL does not encode a SusD-like protein and the TBDT is not annotated as SusC-like.

Ten isolate genomes contained PULs targeting β-N-acetylglucosamine (GlcNAc) through GH20 family β-hexosaminidases (Supplementary Figure S1F). GlcNAc is the monomeric unit of chitin, which is targeted by another set of CAZymes, GH18 and GH19 family chitinases (Supplementary Figure S1G). Both these CAZymes frequently possess chitin-binding CBM5 and CBM12 domains. Chitin targeting loci were found in six isolates although interestingly, two of those, belonging to the genus *Tenacibaculum*, did not contain an adjacent *susCD*-like gene pair. Both isolates, however, feature a GlcNAc-targeting (GH20) PUL. In fact, for *Tenacibaculum* sp. MAR_2010_89, it is the only PUL with a dedicated *susCD*-like gene pair. This indicates a GlcNAc-containing substrate, potentially chitin, as a polysaccharide of major importance to the *Tenacibaculum* clade. Chitin is found in the exoskeletons of crustaceans, but was also detected in crystalline fibers excreted by diatoms (Blackwell *et al.*, 1967) and in the silica frustule of *Thalassiosira pseudonana* (Tesson *et al.*, 2008).

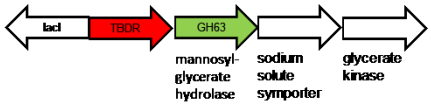
(D) Pectin

PUL 67 of *Flavobacteriaceae bacterium* sp. Hel1_10



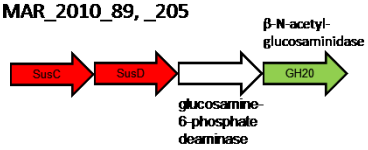
(E) Digeneaside

PULs 44, 55 of Cellulophaga spp. RHA_19, _52,
PULs 239, 222, 254 Maribacter spp. Hel1_7, MAR_2009_71, _60



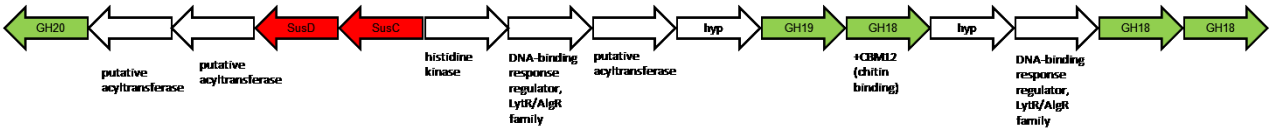
(F) N-acetylglucosamin (NAG)

PULs 349, 351, 354 of *Tenacibaculum* spp. MAR_2009_124,
MAR_2010_89, _205



(G) Chitin

PUL 5 of *Aquimarina* sp. MAR_2010_214

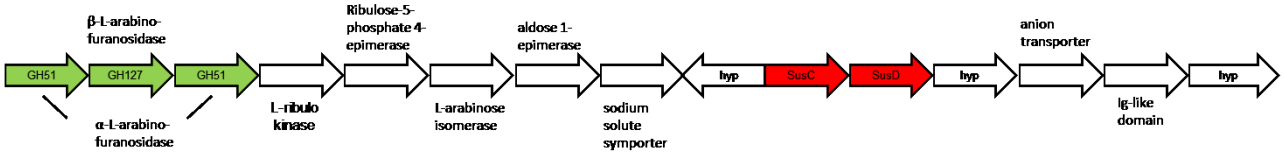


(H) Fructose

PULs 64 of *Flavobacteriaceae bacterium* sp. *Hel1_10,* PUL 145 of *Gillisia* sp. Hel1_86, PUL 237 of *Maribacter* sp. Hel1_7

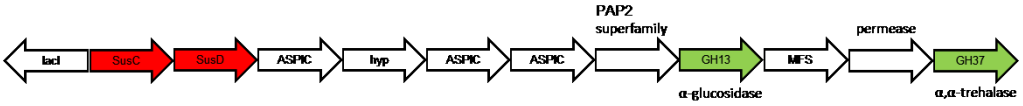
(I) Arabinan

PUL 267 of *Muricauda* sp. MAR_2010_75



(J) Trehalose (α-1,1-glucan)

PULs 103, 116 of *Flavobacteriaceae bacterium* sp. MAR 2010 105, _119



Supplementary Figure S1D-J: PULs targeting pectin, digeneaside, N-acetylglucosamine, chitin, fructose, arabinan and trehalose.

A fructose-containing substrate is targeted by a conserved PUL in six isolates featuring a GH32 β-fructosidase and a sugar kinase, likely a fructokinase (Supplementary Figure S1H). Five isolates showed capacities to tackle diverse sialic acid oligosaccharides through PUL-associated GH33 family sialidases (Supplementary Table 3). Another PUL featuring a predicted GH127 β-L-arabinofuranosidase framed by two GH51 α-L-arabinofuranosidases (Supplementary Figure S1I) was found in *Muricauda* sp. MAR2010_75, likely targeting a mixed α- and β-linked, unsulfated arabinan. Finally, an α-1,1-linked glucan, likely trehalose, is predicted to be targeted by two isolates that possess PULs containing a GH37 family α,α-trehalase and a GH13 family α-glucosidase (Supplementary Figure S1J).

4.4.13 Trees of SusC- and SusD-like proteins reveal substrate-specific clusters

We computed trees for all SusC- and SusD-like protein sequences of the 400 isolate PULs and obtained pronounced clusters for many of the predicted polysaccharide substrates (Figure 6). For clarity, functionally heterogeneous or undefined clusters are depicted as gray clusters (complete trees: Supplementary files 1, 2). Well-defined clusters in both trees included the structurally simple polysaccharides laminarin, α-1,4-glucans and alginate. For example SusD-like proteins of laminarin-targeting PULs of *Cellulophaga* sp. RHA_52, *Flavobacteriaceae bacterium* sp. Hel1_10, *Formosa* sp. Hel1_33_131 and *Psychroserpens* sp. Hel1_66 (PULs 58, 71, 128, 331, Supplementary Table S3), exhibited between 64% and 78% identity (Supplementary Table S4), while identity to SusD-like sequences from other PULs within the same respective genome was only 10-25% (data not shown).

SusC/D-like proteins from conserved PULs for these structurally simple substrates were more closely related than those from more variable PULs targeting structurally more

diverse substrate classes such as FCSPs or xylose-rich substrates (Supplementary Table S4). This is visible in the trees by shorter and longer respective branch lengths (Figure 6). Some substrates formed multiple clusters, for example xylose-rich substrates. This might indicate either rather different xylose-containing substrates or multiple ways of attack and uptake for a given class of xylose-containing substrate.

The topologies of the SusC- and SusD-like protein trees were noteworthy congruent regarding branching patterns of the identified substrate-specific clusters. Only the pectin cluster was located at distinctly different positions. SusC- and SusD-like proteins from the same PULs exhibited a strong tendency to occur in corresponding substrate-specific clusters in both trees. This applied to >70% of the SusC and SusD sequences within identified substrate-specific clusters (Supplementary Figures 1A and B).

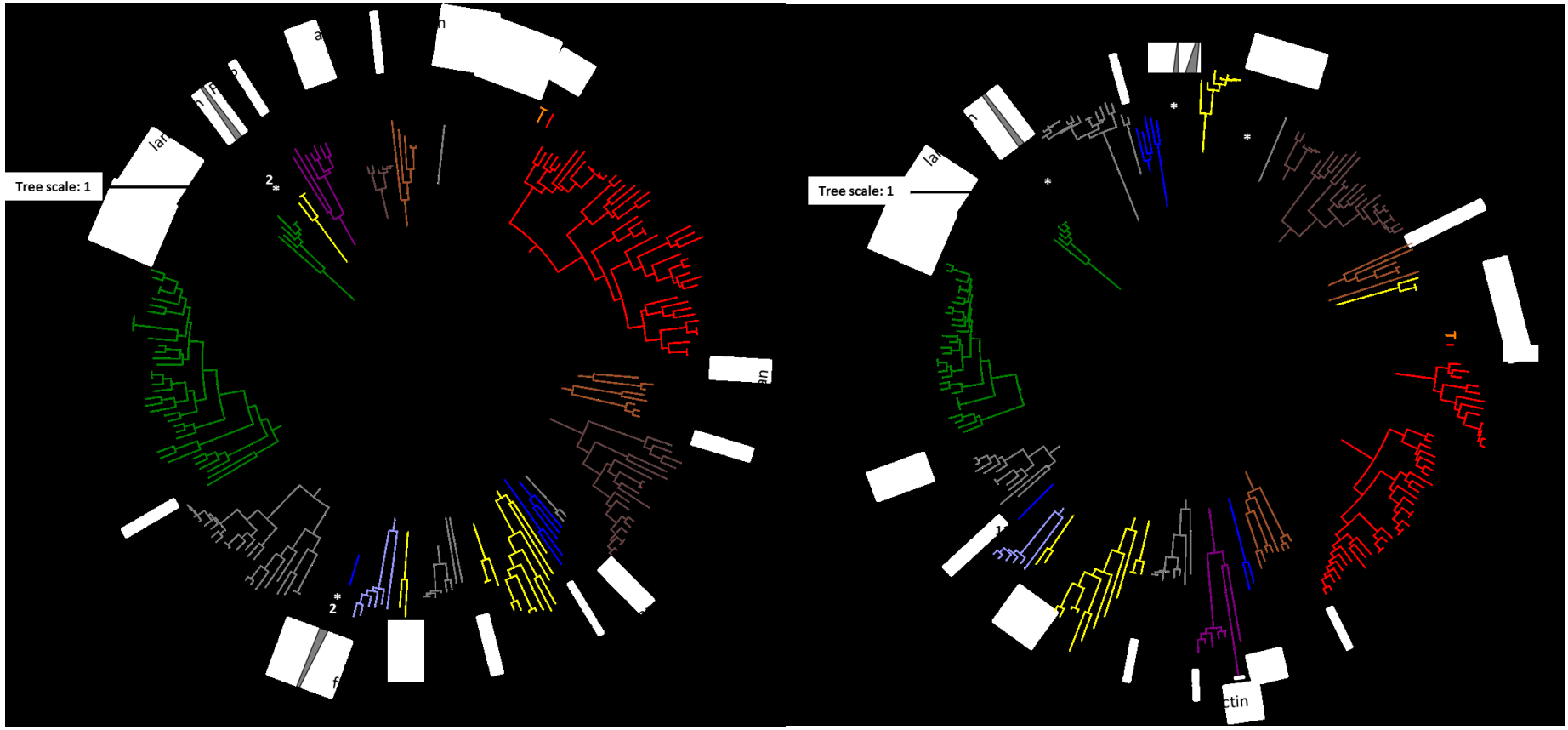


Figure 4.6: Trees of all PUL-associated SusC- (A) and SusD-like (B) proteins of the *Flavobacteriia* isolates showing functional, substrate-specific clustering. Protein sequences were aligned using the MAFFT G-INS-i algorithm and trees were calculated using FastTree 2.1.5 approximate-maximum likelihood (SusC-like: 369; SusD-like: 361). Substrate predictions are depicted in colors. Proteins with expressed homologs in North Sea bacterioplankton blooms of more than 40% sequence identity are marked with asterisks (and number of homologs if x>1). Corresponding figures labeled with protein sequence identifiers, originating species and PUL-associated CAZymes are provided as supplementary material.

4.4.14 SusC/D-like protein expression of bacterioplankton during phytoplankton blooms supports temporal variations of polysaccharide abundances in situ

Highly expressed bacterial proteins usually comprise proteins of the replication and translation machinery, such as ribosomal proteins and chaperonins. CAZymes in comparison mostly feature only medium expression levels. Therefore, state-of-the-art metaproteomics can only detect meaningful CAZymes patterns, when specific bacterioplankton clades are abundant and active. This was the case in a previous study of a 2009 North Sea spring bloom, where samples were taken at weekly intervals (Teeling *et al.,* 2012). We continued an about monthly sampling during the subsequent spring blooms of 2010 to 2012. However, these blooms were less pronounced and so was the bacterial response (Teeling *et al.,* 2016). As a result, only few expressed CAZymes were detected using metaproteomics in a total of 23,917 identified proteins.

However, SusC/D-like proteins range among the highest expressed proteins in bacterioplankton metaproteomes from productive oceans (Morris *et al.,* 2012; Teeling *et al.,* 2012; Williams *et al.,* 2013). Accordingly, we detected high numbers of expressed SusC/D-like proteins in our metaproteomes across all sampled years (Supplementary Table S2). Studies on flavobacterial isolates have identified SusC/D-like proteins as the highest expressed proteins within PULs that are furthermore co-regulated with other PUL-encoded proteins including CAZymes (Xing *et al.,* 2014; Kabisch *et al.,* 2014). SusC/D expression thus represents a suitable proxy for overall PUL expression.

To identify potential substrates, we aligned all expressed SusC/D-like sequences (SusC: 390; SusD: 118) to the SusC/D-tree constructed from isolate PULs. Isolate sequences with highest similarities (>=40%) to expressed sequences are indicated in Figure 6. Further

semi-quantitative analyses were confined to SusC/D-like proteins where at least one related homolog reached expression levels of ≥0.05 %NSAF (Figure 7).

Laminarin: Homologs to laminarin-binding SusC-like proteins were detected amidst the 2009 and 2010 phytoplankton blooms, with one homolog reaching a notable maximum of 0.13 %NSAF on May 4th, 2010 (Figure 7B). Respective SusD-like homologs were detected in the same years and highest expression was observed at the same date in 2010 (0.07 %NSAF, Figure 7D). Amino acid identities of expressed SusC homologs and laminarin PUL SusC-like proteins from isolates ranged from 48-68%, and for SusD homologs from 40-78% (Supplementary Table S4). Our data suggest that laminarin occurred at the bloom peaks in 2009 and 2010 and directly thereafter. This is supported by detection of expressed GH3 β-glucosidases and GH16 β-glucanases in 2009 (Teeling *et al.,* 2012) and, to a lesser degree, in 2010 (Supplementary Table S2). Chrysolaminarin is produced by microalgae such as *Thalassiosira nordenskioeldii* diatoms or representatives of *Phaecystis* haptophytes (Alderkamp *et al.,* 2007; Myklestad, 1989), which both were among the dominating microalgae in 2009 and 2010 (Teeling *et al*., 2016).

Alpha-1,4-glucan: Respective SusD-like proteins were most abundantly detected in 2010, peaking on April 8th (0.11 %NSAF, Figure 7D), but also in 2009 and 2011. Sequence identities to isolate α-1,4-glucan PUL SusD-like proteins ranged from 43-46% (Supplementary Table S4). These data indicate that α-1,4-glucans, potentially starch or glycogen, represented a recurring substrate from 2009 to 2011 during early to late phytoplankton bloom stages.

Alpha-1,1-glucan: An α-1,1-glucan-binding SusC-like protein potentially targeting a trehalose-like substrate was strongly expressed on April 28th 2011 (0.19 %NSAF;

Figure 7B), but also in all other years except 2010. Its sequence identity to the SusC-like proteins of the trehalose PULs of *Flavobacteriaceae bacterium* spp. MAR_2010_105 and MAR_2010_119 was 43% (Supplementary Figure S2J, PUL103 and PUL116, Supplementary Tables S3 and S4). Corresponding SusD-like proteins were detected in all years, but most strongly throughout the blooms of 2009 and 2011. Their protein identities to the isolate PULs ranged from 58-59% (Supplementary Table S4).

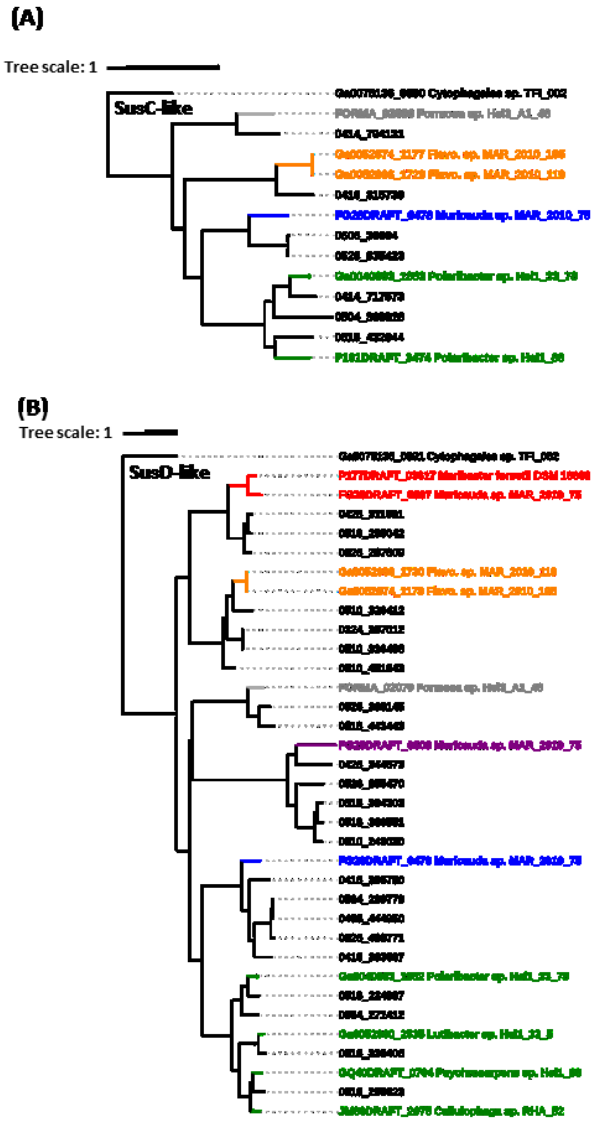
Sulfated β-xylan: One SusC and two SusD-like proteins likely targeting a sulfated β-xylan were expressed in the mid and late stages of the phytoplankton bloom of 2009, peaking at 0.08 %NSAF for SusC-like proteins and 0.07 %NSAF for SusD-like proteins. Their identities to homologs of the sulfated β-xylan PUL of *Formosa* sp. Hel3_A1_48 (Figure 5B, PUL136, Supplementary Table S3) was 53% and 49-53%, respectively (Supplementary Table S4).

Beta-mannan: Homologs with high identities to SusC/D-like proteins occurring in a β-mannan PUL from *Muricauda* sp. MAR_2010_75 were strongly expressed during blooms from 2009 to 2011, peaking on April 28th 2011 for SusC (0.22 %NSAF) and SusD (0.25 %NSAF) homologs. No SusC-like proteins of putative β-mannan PULs were detected in 2012 and SusD-like expression was likewise much weaker. The expressed SusC-like proteins were 36-41% identical to the ones from the β-mannan PUL of *Muricauda* sp. MAR_2010_75 (PUL266, Supplementary Table S3) and the SusD-like proteins showed 45-51% identity (Supplementary Table S4).

Arabinan: SusD-like proteins potentially targeting an arabinan were expressed at late phytoplankton blooms stages during all four years with at least 0.05 %NSAF. However, their identities to the SusD-like protein of a predicted arabinan PUL from *Muricauda* sp.

MAR_2010_75 were only 26-30% (Supplementary Figure S2I, PUL267, Supplementary Tables S3 and S4).

In summary, comparative analyses of SusC/D homolog expression are indicative of a successive utilization of different polysaccharides over the course of phytoplankton blooms. This agrees with successive changes in the microbial community composition during bloom events that we reported earlier on (Teeling *et al.,* 2012 and 2016).



(C)

%NSAF (=SAF/sum all SAFs*100)														
	2009				2010				2011			2012		
PP bloom state	MID	MID	MID	LATE	PRE	EARLY	MID	LATE	PRE	MID	LATE	PRE	EARLY	LATE
date / substrate	03.31.	04.07.	04.14.	04.21.	03.03.	04.08.	05.04.	05.18.	03.24.	04.28.	05.26.	03.08.	04.16.	05.10.
sulfated xylan	0.00	0.00	0.08	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
α-1,1-glucan	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.01	0.00	0.00	0.02
β-mannan	0.04	0.05	0.00	0.02	0.00	0.00	0.09	0.05	0.00	0.22	0.00	0.00	0.00	0.00
	0.03	0.04	0.02	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00
laminarin	0.00	0.04	0.07	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	0.08	0.05	0.01	0.01	0.00	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(D)

α-1,4-glucan	0.01	0.04	0.01	0.02	0.00	0.11	0.06	0.02	0.00	0.01	0.01	0.00	0.00	0.01
	0.02	0.01	0.00	0.01	0.00	0.05	0.01	0.01	0.00	0.07	0.00	0.00	0.00	0.01
	0.01	0.01	0.00	0.00	0.00	0.00	0.06	0.06	0.00	0.00	0.03	0.00	0.00	0.00
α-1,1-glucan	0.00	0.00	0.01	0.03	0.00	0.01	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.07	0.08	0.03	0.04	0.01
	0.03	0.08	0.06	0.12	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.03
sulfated Xylan	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.03	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
arabinan	0.01	0.02	0.02	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.05	0.00	0.00	0.01
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07
β-mannan	0.04	0.02	0.05	0.09	0.00	0.02	0.01	0.02	0.00	0.03	0.01	0.00	0.01	0.02
	0.11	0.09	0.03	0.04	0.00	0.07	0.11	0.02	0.00	0.25	0.01	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.01	0.00
laminarin	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.01	0.01	0.00	0.00	0.00	0.07	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.00	0.01	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Figure 4.7: (A, B) Trees of expressed SusC and SusD-like proteins identified in 3 to 0.2 µm bacterioplankton during North Sea spring phytoplankton blooms in 2009-2012 using proteomics. The most closely related SusC/D-like sequences from North Sea *Flavobacteriia* isolates in this study were integrated in the tree. Protein names correspond to sequence identifier and isolate name. Sequences were aligned using the MAFFT G-INS-i algorithm. The tree was calculated using FastTree 2.1.5 approximate-maximum likelihood. (C, D) Corresponding expression levels as Normalized Spectral Abundance Factors (%NSAF) for the four consecutive blooms. Metaproteomic samples were classified as pre-, early-, mid-, and late-bloom based on chlorophyll *a* concentrations during the spring phytoplankton blooms (PP bloom state). Expression levels are highlighted by green color.

4.5 Discussion

We observed diverse polysaccharide degradation capacities among North Sea *Flavobacteriia* with no distinct correlation to taxonomy. Even isolates from identical genera often featured notably diverging PUL repertoires and genome sizes (e.g. *Polaribacter*, *Maribacter* and *Cellulophaga*), substantiating earlier data (Xing *et al.,* 2014). Our findings suggest that a species' PUL repertoire is more dependent on its distinct ecological niche while its phylogeny is of secondary importance. This corroborates the hypothesis that PULs are exchanged between *Flavobacteriia* through horizontal gene transfer (Hehemann *et al.,* 2012).

The isolates' PUL repertoires showcase that abundant, structurally simple substrates such as laminarin, α-1,4-glucans and alginate are targeted by likewise conserved and frequent PULs. These substrates are likely so common that preserving the respective catabolic machinery is favorable for many marine *Flavobacteriia*. Diatom-derived chrysolaminarin has been estimated to amount to 5-15 petagrams of organic carbon annually (Alderkamp *et al.,* 2007) and accordingly laminarin-specific PULs were frequent in our surface water isolates. The four predicted laminarin-PUL variants we identified indicate that different laminarin types (Gügi *et al*., 2015) may be targeted by different PULs. Variant B contains predicted endo- and exo-acting β-1,3-glucan hydrolases (GH17) highly specific to laminarin degradation (Becker *et al*., 2017). Variants A, C and D only contain GH16 endo-1,3(4)-β-glucanases and may not be restricted to laminarin, but are potentially capable of degrading further mixed-linkage β-1,3/1,4-glucans, as recently shown for a similar conserved PUL in human gut *Bacteroidetes* (Tamura *et al.,* 2017).

Alginate and α-1,4-glucan degradation capacities were prevalent in macroalgae-associated, but were also common in many seawater isolates. Overall, laminarin, α-1,4-

glucan and alginate PULs are fairly conserved and make up over a quarter of all PULs in the isolates (115/400), suggesting that these are abundant polysaccharide substrates in North Sea coastal habitats that many microbes can consume and likely compete for.

Other substrates are utilized by fewer isolates, which implies that algal polysaccharide degradation is usually carried out by multiple resource-partitioning bacterioplankton species. In macroalgae-associated isolates, these substrates include new FCSP variants, and xylose- and rhamnose-rich polysaccharides. Among surface water isolates, these substrates are sulfated α-mannans (likely an α-glucuronomannan glycoprotein), β-mannans, sulfated α-galactans and β-xylans, chitin and (trehalose-like) α-1,1-glucans.

A major result of this study is the substrate-specific clustering of both SusC- and SusD-like proteins. The strong tendencies of SusC and SusD homologs to occur in corresponding substrate-specific clusters in both trees, resulting in similar tree topologies, suggest coevolution of these two proteins. This hypothesis is corroborated by recent X-ray crystallography findings showing complex formation of two SusC- and SusD-like proteins of *B. thetaiotaomicron* (Glenwright *et al.,* 2017). Clustering was more pronounced for structurally conserved, simple polysaccharides than for the heterogeneous and partially new substrates described in this study. This is expected, as heterogeneous substrates are attacked at multiple points resulting in a variety of structurally different oligosaccharides for uptake. Furthermore, broad substrate classes that currently can only be defined as e.g. FCSPs or xylose-containing substrates might actually represent multiple chemically rather different substrates. Hence, improvement of functional clustering is to be expected once more detailed knowledge on algal polysaccharides structures is available.

We here provide first metaproteomic data indicating that high-resolution expression analysis of SusC/D homologs may be used for monitoring changes in microbial polysaccharide degradation activity. This provides a proxy on which polysaccharides are important at a given time and space in marine carbon cycling. Considering our still incomplete knowledge, only expressed SusC/D homologs exhibiting a high level of sequence identity to functionally annotated or characterized SusC/D sequences should be considered. Absence of such expressed homologs, however, does not preclude that a respective substrate may be targeted by an as yet unknown SusC/D system. This current limitation notwithstanding, our approach provides a new method to identify environmentally relevant polysaccharide substrates that due to their structural complexity are still difficult to identify by direct chemical analysis.

Conflict of interest

The authors declare no conflict of interest.

4.6 Acknowledgements

We thank Sabine Kühn for the DNA extractions and Bernhard Fuchs for critical reading. Genome sequencing and assembly was conducted in the framework of the COGITO project (Contract No. DE-AC02-05CH11231) by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, and is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This study was funded by the Max Planck Society and supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the research unit FOR2406 'Proteogenomics of Marine Polysaccharide Utilization (POMPU)' by grants of Rudolf Amann (AM 73/9-1), Hanno Teeling (TE 813/2-1), and Thomas Schweder (SCHW 595/10-1) and the DFG Emmy Noether program (Jan-Hendrik Hehemann grant HE 7217/1-1).

Lennart Kappelmann is a member of the International Max Planck Research School of Marine Microbiology (MarMic).

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej).

4.7 References

Alderkamp, AC, Van Rijssel, M, & Bolhuis, H (2007). Characterization of marine bacteria and the activity of their enzyme systems involved in degradation of the algal storage glucan laminarin. *FEMS Microbiology Ecology*, **59**: 108-117.

Ale, MT, Mikkelsen, JD, Meyer, AS (2011). Important determinants for fucoidan bioactivity: A critical review of structure-function relations and extraction methods for fucose-containing sulfated polysaccharides from brown seaweeds. *Marine Drugs* **9***:* 2106-2130.

Altschul, SF, Gish, W, Miller, W, Myers, EW, Lipman, DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Arlov, Ø, Aachmann, FL, Sundan, A, Espevik, T, Skjåk-Bræk, G (2014). Heparin-like properties of sulfated alginates with defined sequences and sulfation degrees. *Biomacromolecules* **15***:* 2744-2750.

Arlov, Ø, Skjåk-Bræk, G (2017). Sulfated Alginates as heparin analogues: a review of chemical and functional properties. *Molecules* **22***:* 778.

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

Barras, DR, Stone, BA (1969). β-1,3-glucan hydrolases from *Euglena gracilis*: II. Purification and properties of the β-1,3-glucan exo-hydrolase. *Biochimica et Biophysica Acta (BBA)-Enzymology* **191***:* 342-353.

Barbeyron, T, Thomas, F, Barbe, V, Teeling, H, Schenowitz, C, Dossat, C, *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* Dsij^T. *Environmental Microbiology* **18:** 4610-4627.

Bauer, M, Kube, M, Teeling, H, Richter, M, Lombardot, T, Allers, *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.

Becker, S, Scheffel, A, Polz, MF, Hehemann, JH (2017). Accurate quantification of laminarin in marine organic matter with enzymes from marine microbes. *Applied and Environmental Microbiology* **83***:* e03389-16.

Bjursell, MK, Martens, EC, Gordon, JI (2006). Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *Journal of Biological Chemistry* **281**: 36269-79.

Chiovitti, A, Harper, RE, Willis, A, Bacic, A, Mulvaney, P, Wetherbee, R (2005). Variantions in the substituted 3-linked mannans closely associated with the silicified walls of diatoms. *Journal of Phycology* **41**: 1154-1161.

Cuskin, F, Lowe, EC, Temple, MJ, Zhu, Y, Cameron, EA, Pudlo, NA *et al.* (2015). Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature **517***: 165-169.

Edgar, RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26***:* 2460-2461.

Finn, RD, Bateman, A, Clements, J, Coggill, P, Eberhardt, RY, Eddy, SR *et al.* (2013). Pfam: the protein families database. *Nucleic Acids Research* **42***:* D222-D230.

Fischer, FG, Dörfel, H (1955). Die Polyuronsäuren der Braunalgen (Kohlenhydrate der Algen I). *Hoppe-Seylers Zeitschrift für Physiologische Chemie* **302***:* 186-203.

Field, CB, Behrenfeld, MJ, Randerson, JT, Falkowski, P (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**: 237-240.

Ford, CW, Percival, E (1965). 1299. Carbohydrates of *Phaeodactylum tricornutum*. Part II. A sulphated glucuronomannan. *Journal of the Chemical Society*, 7042-7046.

Glenwright, AJ, Pothula, KR, Bhamidimarri, SP, Chorev, DS, Baslé, A, Firbank, SJ *et*

al. (2017). Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**: 407-411.

Gügi, B, Le Costaouec, T, Burel, C, Lerouge, P, Helbert, W, Bardor, M (2015). Diatom-specific oligosaccharide and polysaccharide structures help to unravel biosynthetic capabilities in diatoms. *Marine Drugs* **13**: 5993-6018.

Hahnke, RL, Harder, J (2013). Phylogenetic diversity of *Flavobacteria* isolated from the North Sea on solid media. *Systematic and Applied Microbiology* **36**: 497-504.

Hahnke, RL, Bennke, CM, Fuchs, BM, Mann, AJ, Teeling, H, Harder, J (2015). Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. *Environmental Microbiology* **17**: 3515-3526.

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

Hehemann, JH, Kelly, AG, Pudlo, NA, Martens, EC, Boraston, AB (2012). Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proceedings of the National Academy of Sciences* **109***:* 19786-19791.

Hemsworth, GR, Déjean, G, Davies, GJ, Brumer, H (2016). Learning from microbial strategies for polysaccharide degradation. *Biochemical Society Transactions* **44**: 94-108.

Huntemann, M, Ivanova, NN, Mavromatis, K, Tripp, HJ, Paez-Espino, D, Palaniappan, K *et al.* (2015). The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v. 4). *Standards in Genomic Sciences **10***: 86.

Hoagland, KD, Rosowski, JR, Gretz, MR, Roemer, SC (1993). Diatom extracellular polymeric substances: function, fine structure, chemistry, and physiology. *Journal of Phycology* **29**: 537-566.

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

Katoh, K, Standley, DM (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772-780.

Kloareg, B, Quatrano, RS (1988). Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanography and Marine Biology: An Annual Review* **26***:* 259-315.

Kraan, S (2012). Algal polysaccharides, novel applications and outlook. In *Carbohydrates-comprehensive studies on glycobiology and glycotechnology*. InTech.

Labourel, A, Jam, M, Legentil, L, Sylla, B, Hehemann, JH, Ferrières *et al*. (2015). Structural and biochemical characterization of the laminarinase ZgLamCGH16 from Zobellia galactanivorans suggests preferred recognition of branched laminarin. *Acta Crystallographica Section D: Biological Crystallography* **71**: 173-184.

Laine, RA (1994). A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: The Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* **4**: 759-767.

Le Costaouëc, T, Unamunzaga, C, Mantecon, L, Helbert, W (2017). New structural insights into the cell-wall polysaccharide of the diatom *Phaeodactylum tricornutum*. *Algal Research* **26**: 172-179.

Li, B, Lu, F, Wei, X, Zhao, R (2008). Fucoidan: structure and bioactivity. *Molecules* **13**: 1671-1695.

Lombard, V, Golaconda Ramulu, H, Drula, E, Coutinho, PM, Henrissat, B (2013). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**: D490-D495.

Martens, EC, Lowe, EC, Chiang, H, Pudlo, NA, Wu, M, McNulty, NP *et al.* (2011). Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biology* **9**: e1001221.

Meyer, F, Goesmann, A, McHardy, AC, Bartels, D, Bekel, T, Clausen, J *et al.* (2003). GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Research* **31**: 2187-2195.

Mhanna, R, Kashyap, A, Palazzolo, G, Vallmajo-Martin, Q, Becher, J, Möller, S *et al.* (2014). Chondrocyte culture in three dimensional alginate sulfate hydrogels promotes proliferation while maintaining expression of chondrogenic markers. *Tissue Engineering Part A* **20**, 1454-1464.

Michel, G., Barbeyron, T., Kloareg, B., & Czjzek, M. (2009). The family 6 carbohydrate-binding modules have coevolved with their appended catalytic modules toward similar substrate specificity. *Glycobiology*, ***19***: 615-623.

Moreira, LRS, Filho, EXF (2008). An overview of mannan structure and mannan-degrading enzyme systems. *Applied Microbiology and Biotechnology* **79**: 165.

Morris, RM, Nunn, BL, Frazar, C, Goodlett, DR, Ting, YS. Rocap, G (2010). Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *The ISME Journal* **4**: 673-685.

Murray, AE, Arnosti, C, Christina, L, Grossart, HP, Passow, U (2007). Microbial dynamics in autotrophic and heterotrophic seawater mesocosms. II. Bacterioplankton community structure and hydrolytic enzyme activities. *Aquatic Microbial Ecology* **49**: 123-141.

Myklestad, SM (1989). Production, chemical structure, metabolism, and biological function of the (1→3)-linked, β-3-D-glucans in diatoms. *Biological Oceanography* **6***:* 313-326.

Painter, TJ (1983). Algal polysaccharides. *The Polysaccharides* **2**, 195-285.

Popper, ZA, Michel, G, Hervé, C, Domozych, DS, Willats, WG, Tuohy, MG *et al.* (2011). Evolution and diversity of plant cell walls: From algae to flowering plants. *Annual Review of Plant Biology* **62**: 567-590.

Price, MN, Dehal, PS, Arkin, AP (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS One* **5**: e9490.

Rawlings, ND, Barrett, AJ, Bateman, A (2011). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, **40***:* D343-D350.

Read, SM, Currie, G, Bacic, A (1996). Analysis of the structural heterogeneity of laminarin by electrospray-ionisation-mass spectrometry. *Carbohydrate Research* **281**:

187-201.

Reddy, TB, Thomas, AD, Stamatis, D, Bertsch, J, Isbandi, M, Jansson, J *et al.* (2014). The Genomes OnLine Database (GOLD) v. 5: a metadata management system based on a four level (meta) genome project classification. *Nucleic Acids Research*, **43***:* D1099-D1106.

Reese, ET, Mandels, M (1959). β-D-1,3-glucanases in fungi. *Canadian Journal of Microbiology* **5**: 173-185.

Smith, KA, Salyers, AA (1991). Characterization of a neopullulanase and an alpha-glucosidase from *Bacteroides thetaiotaomicron* 95-1. *Journal of Bacteriology* **173**: 2962-2968.

Sørensen, I, Pettolino, FA, Bacic, A, Ralph, J, Lu, F, O'Neill, MA *et al.* G (2011). The charophycean green algae provide insights into the early origins of plant cell walls. *The Plant Journal* **68**: 201-211.

Tamura, K, Hemsworth, GR, Déjean, G, Rogers, TE, Pudlo, NA, Urs, K *et al.* (2017). Molecular Mechanism by which Prominent Human Gut *Bacteroidetes* Utilize Mixed-Linkage Beta-Glucans, Major Health-Promoting Cereal Polysaccharides. *Cell Reports* **21**: 417-430.

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

Teeling, H, Fuchs, BM, Bennke, CM, Krüger, K, Chafee, M, Kappelmann, L *et al*. (2016). Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *Elife* **5**, e11888.

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

Vizcaíno, JA, Csordas, A, Del-Toro, N, Dianes, JA, Griss, J, Lavidas, I *et al.* (2015). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Research **44***,

D447-D456.

Williams, TJ, Wilkins, D, Long, E, Evans, F, DeMaere, MZ, Raftery, MJ, Cavicchioli, R (2013). The role of planktonic *Flavobacteria* in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. *Environmental Microbiology* **15**: 1302-1317.

Wu, J, Lv, Y, Liu, X, Zhao, X, Jiao, G, Tai, W *et al.* (2015). Structural study of sulfated fuco-oligosaccharide branched glucuronomannan from *Kjellmaniella crassifolia* by ESI-CID-MS/MS. *Journal of Carbohydrate Chemistry* **34**: 303-317.

Wustman, BA, Lind, J, Wetherbee, R, Gretz, MR (1998). Extracellular matrix assembly in diatoms (*Bacillariophyceae*) III. Organization of fucoglucuronogalactans within the adhesive stalks of *Achnanthes longipes*. *Plant Physiology* **116**: 1431-1441.

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

Yin, Y, Mao, X, Yang, J, Chen, X, Mao, F, Xu, Y (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **40***:* W445-W451.

4.8 Supplementary Table S3. Predicted PULs and their substrates within the isolate genomes.

isolate	# PUL	substrate	locus_tag start	locus_tag stop	genes in PUL
Aquimarina sp. MAR_2010_214	1	unknown	Ga0075137_0612	Ga0075137_0626	14
	2	unknown	Ga0075137_1266	Ga0075137_1274	8
	3	alpha-mannan	Ga0075137_2202	Ga0075137_2216	15
	4	laminarin A	Ga0075137_2229	Ga0075137_2237	9
	5	chitin	Ga0075137_2352	Ga0075137_2367	15
	6	alpha-glucan (starch-like)	Ga0075137_4121	Ga0075137_4130	10
	7	alpha-mannan	Ga0075137_4505	Ga0075137_4515	11
	8	unknown	Ga0075137_4699	Ga0075137_4707	9
	9	unknown	Ga0075137_5035	Ga0075137_5041	7
Arenibacter palladensis sp. MAR_2009_79	10	unknown	GQ41DRAFT_0178	GQ41DRAFT_0181	4
	11	unknown	GQ41DRAFT_0277	GQ41DRAFT_0296	19
	12	alpha-mannan	GQ41DRAFT_1277	GQ41DRAFT_1289	13
	13	unknown	GQ41DRAFT_1718	GQ41DRAFT_1743	23
	14	rhamnose-containing substrate	GQ41DRAFT_2079	GQ41DRAFT_2095	17
	15	rhamnose-containing substrate	GQ41DRAFT_2096	GQ41DRAFT_2109	13
	16	rhamnose-containing substrate	GQ41DRAFT_2111	GQ41DRAFT_2117	7
	17	unknown	GQ41DRAFT_2140	GQ41DRAFT_2168	24
	18	unknown	GQ41DRAFT_2169	GQ41DRAFT_2182	14
	19	fructose	GQ41DRAFT_2158	GQ41DRAFT_2163	6
	20	unknown	GQ41DRAFT_2239	GQ41DRAFT_2253	15
	21	alginate	GQ41DRAFT_2577	GQ41DRAFT_2592	16
	22	unknown	GQ41DRAFT_2946	GQ41DRAFT_2976	31
	23	unknown	GQ41DRAFT_2980	GQ41DRAFT_3005	26
	24	unknown	GQ41DRAFT_3008	GQ41DRAFT_3016	9
	25	unknown	GQ41DRAFT_3023	GQ41DRAFT_3034	12
	26	alpha-mannan	GQ41DRAFT_3065	GQ41DRAFT_3102	38
	27	FCSP	GQ41DRAFT_3258	GQ41DRAFT_3291	33
	28	pectin	GQ41DRAFT_3311	GQ41DRAFT_3333	23
	29	sialic acid polysaccharide	GQ41DRAFT_3695	GQ41DRAFT_3702	8
	30	alpha-glucan (starch-like)	GQ41DRAFT_4087	GQ41DRAFT_4095	9
	31	unknown	GQ41DRAFT_4213	GQ41DRAFT_4225	13
	32	unknown	GQ41DRAFT_4243	GQ41DRAFT_4255	12
	33	pectin	GQ41DRAFT_4256	GQ41DRAFT_4272	17
	34	unknown	GQ41DRAFT_4274	GQ41DRAFT_4281	8
	35	unknown	GQ41DRAFT_4313	GQ41DRAFT_4329	17

Cellulophaga sp. Hel1_12	36	unknown	Q45DRAFT_0818	Q45DRAFT_0842	25
	37	laminarin B	Q45DRAFT_1308	Q45DRAFT_1318	11
	38	alpha-glucan (pot. starch)	Q45DRAFT_1756	Q45DRAFT_1764	9
	39	unknown	Q45DRAFT_2458	Q45DRAFT_2469	12
	40	rhamnose-containing substrate	Q45DRAFT_3222	Q45DRAFT_3244	23
Cellulophaga sp. RHA_19	41	carrageenan	Ga0075144_0334	Ga0075144_0353	20
	42	unknown	Ga0075144_0890	Ga0075144_0899	10
	43	unknown	Ga0075144_0944	Ga0075144_0954	11
	44	digeneaside	Ga0075144_1455	Ga0075144_1459	5
	45	FCSP	Ga0075144_1723	Ga0075144_1773	51
	46	alginate	Ga0075144_1890	Ga0075144_1900	11
	47	galactose-containing polysaccharide	Ga0075144_2571	Ga0075144_2578	8
	48	unknown	Ga0075144_3036	Ga0075144_3054	19
	49	alpha-glucan (potentially starch)	Ga0075144_3073	Ga0075144_3083	11
	50	carrageenan	Ga0075144_3138	Ga0075144_3162	25
	51	galactose-containing sulfated polysaccharide	Ga0075144_3178	Ga0075144_3196	19
	52	alpha-mannan	Ga0075144_3202	Ga0075144_3226	25
Cellulophaga sp. RHA_52	53	carrageenan	JM80DRAFT_0303	JM80DRAFT_0315	13
	54	unknown	JM80DRAFT_0864	JM80DRAFT_0874	11
	55	digeneaside	JM80DRAFT_1401	JM80DRAFT_1405	5
	56	FCSP	JM80DRAFT_1668	JM80DRAFT_1694	27
	57	alginate	JM80DRAFT_1820	JM80DRAFT_1826	7
	58	laminarin A	JM80DRAFT_2168	JM80DRAFT_2173	6
	59	unknown	JM80DRAFT_2973	JM80DRAFT_2976	4
	60	alpha-glucan (potentially starch)	JM80DRAFT_2973	JM80DRAFT_2992	20
	61	unknown	JM80DRAFT_3054	JM80DRAFT_3072	19
Dokdonia sp. Hel1_63	62	alpha-glucan (potentially starch)	JM84DRAFT_0206	JM84DRAFT_0218	13
	63	alginate	JM84DRAFT_1364	JM84DRAFT_1372	9
Flavobacteriaceae bacterium sp. Hel1_10	64	fructose	P176DRAFT_0362	P176DRAFT_0367	6
	65	xylan	P176DRAFT_0736	P176DRAFT_0790	54
	66	laminarin C	P176DRAFT_1080	P176DRAFT_1088	9
	67	pectin	P176DRAFT_2075	P176DRAFT_2093	19
	68	alpha-glucan (potentially starch)	P176DRAFT_2348	P176DRAFT_2363	16
	69	beta-mannan	P176DRAFT_2530	P176DRAFT_2544	15
	70	unknown	P176DRAFT_2742	P176DRAFT_2747	6
	71	laminarin A	P176DRAFT_2792	P176DRAFT_2796	5
	72	unknown	P176DRAFT_2797	P176DRAFT_2810	14
	73	laminarin B	P176DRAFT_2816	P176DRAFT_2824	9

Flavobacteriaceae bacterium sp. MAR_2009_75	74	FCSP	Ga0075135_0377	Ga0075135_0391	15
	75	unknown	Ga0075135_0462	Ga0075135_0475	14
	76	unknown	Ga0075135_1036	Ga0075135_1048	13
	77	unknown	Ga0075135_1134	Ga0075135_1145	11
	78	unknown	Ga0075135_1146	Ga0075135_1153	8
	79	NAG	Ga0075135_1154	Ga0075135_1162	9
	80	unknown	Ga0075135_1163	Ga0075135_1169	7
	81	FCSP	Ga0075135_1363	Ga0075135_1377	15
	82	alginate	Ga0075135_1385	Ga0075135_1402	18
	83	unknown	Ga0075135_1408	Ga0075135_1415	8
	84	unknown	Ga0075135_1420	Ga0075135_1437	18
	85	unknown	Ga0075135_1486	Ga0075135_1502	14
	86	rhamnose-containing substrate	Ga0075135_1503	Ga0075135_1535	29
	87	pectin	Ga0075135_1544	Ga0075135_1553	10
	88	unknown	Ga0075135_1829	Ga0075135_1840	12
	89	rhamnose-containing substrate	Ga0075135_2218	Ga0075135_2230	13
	90	pectin	Ga0075135_2231	Ga0075135_2274	44
	91	xylan	Ga0075135_2280	Ga0075135_2286	7
	92	FCSP	Ga0075135_3394	Ga0075135_3417	23
	93	unknown	Ga0075135_3430	Ga0075135_3439	10
	94	unknown	Ga0075135_3465	Ga0075135_3495	31
	95	unknown	Ga0075135_3578	Ga0075135_3596	19
Flavobacteriaceae bacterium sp. MAR_2010_10	96	unknown	LZ17DRAFT_0401	LZ17DRAFT_0407	7
Flavobacteriaceae bacterium sp. MAR_2010_105	97	sialic acid polysaccharide	Ga0052874_0002	Ga0052874_0020	19
	98	unknown	Ga0052874_0334	Ga0052874_0345	12
	99	alpha-glucan	Ga0052874_0386	Ga0052874_0399	14
	100	laminarin A	Ga0052874_0467	Ga0052874_0472	6
	101	chitin	Ga0052874_1138	Ga0052874_1146	8
	102	alginate	Ga0052874_1153	Ga0052874_1171	19
	103	trehalose	Ga0052874_1176	Ga0052874_1187	12
	104	unknown	Ga0052874_1190	Ga0052874_1210	21
	105	alpha-glucan	Ga0052874_2403	Ga0052874_2406	4
	106	laminarin A	Ga0052874_2834	Ga0052874_2843	10
	107	unknown	Ga0052874_2953	Ga0052874_2964	11
Flavobacteriaceae bacterium sp. MAR_2010_118	108	laminarin B	LZ20DRAFT_0758	LZ20DRAFT_0770	13
Flavobacteriaceae bacterium sp. MAR_2010_119	109	alpha-glucan	Ga0052906_0062	Ga0052906_0065	4
	110	laminarin A	Ga0052906_0483	Ga0052906_0492	10
	111	unknown	Ga0052906_0882	Ga0052906_0894	13
	112	alpha-glucan	Ga0052906_0935	Ga0052906_0947	13
	113	laminarin A	Ga0052906_1015	Ga0052906_1020	6
	114	chitin	Ga0052906_1689	Ga0052906_1698	9
	115	alginate	Ga0052906_1705	Ga0052906_1723	19

Flavobacteriaceae bacterium sp. MAR_2010_119	116	trehalose	Ga0052906_1728	Ga0052906_1739	12
	117	unknown	Ga0052906_1745	Ga0052906_1753	9
Flavobacteriaceae bacterium sp. MAR_2010_188	118	laminarin	Ga0052905_0415	Ga0052905_0433	19
	119	alpha-glucan	Ga0052905_1230	Ga0052905_1241	12
	120	alpha-glucan	Ga0052905_1755	Ga0052905_1768	13
	121	sialic acid polysaccharide	Ga0052905_2352	Ga0052905_2361	10
	122	unknown	Ga0052905_2819	Ga0052905_2832	14
	123	unknown	Ga0052905_3095	Ga0052905_3117	21
Formosa sp. Hel1_31_208	124	laminarin B	Ga0075143_1713	Ga0075143_1730	18
Formosa sp. Hel1_33_131	125	laminarin B	Formosa_661	Formosa_669	9
	126	alpha-mannan	Formosa_1137	Formosa_1155	19
	127	chitin	Formosa_1560	Formosa_1571	12
	128	laminarin A	Formosa_1822	Formosa_1828	7
	129	laminarin	Formosa_2161	Formosa_2164	4
Formosa sp. Hel3_A1_48	130	laminarin B	FormosaA_12	FormosaA_17	6
	131	unknown	FormosaA_508	FormosaA_514	7
	132	laminarin A	FormosaA_534	FormosaA_540	7
	133	laminarin C	FormosaA_643	FormosaA_653	11
	134	unknown	FormosaA_770	FormosaA_782	13
	135	unknown	FormosaA_1202	FormosaA_1225	24
	136	Sulfated, xylose-containing polysaccharide	FormosaA_1248	FormosaA_1264	17
Gillisia sp. Hel1_29	137	alpha-glucan	P163DRAFT_0839	P163DRAFT_0849	11
	138	unknown	P163DRAFT_1034	P163DRAFT_0004	15
	139	laminarin A	P163DRAFT_0607	P163DRAFT_0613	7
	140	unknown	P163DRAFT_3317	P163DRAFT_3324	8
Gillisia sp. Hel1_33_143	141	unknown	Ga0052908_0480	Ga0052908_0487	8
	142	laminarin A	Ga0052908_0707	Ga0052908_0713	7
	143	alpha-glucan	Ga0052908_3173	Ga0052908_3183	11
Gillisia sp. Hel1_86	144	alpha-glucan	JM83DRAFT_0208	JM83DRAFT_0218	11
	145	fructose	JM83DRAFT_0438	JM83DRAFT_0446	9
	146	unknown	JM83DRAFT_0793	JM83DRAFT_0807	12
	147	unknown	JM83DRAFT_0821	JM83DRAFT_0830	10
	148	laminarin D	JM83DRAFT_1080	JM83DRAFT_1090	11
	149	alginate	JM83DRAFT_1670	JM83DRAFT_1681	12
	150	digeneaside	JM83DRAFT_2179	JM83DRAFT_2183	5
	151	unknown	JM83DRAFT_2184	JM83DRAFT_2208	25
Gramella forsetii sp. KT0803	152	fructose	GFO_0008	GFO_0014	7
	153	unknown	GFO_0027	GFO_0036	10
	154	unknown	GFO_0355	GFO_0375	21
	155	arabinoxylan	GFO_0687	GFO_0705	19
	156	unknown	GFO_1027	GFO_1072	46
	157	alginate	GFO_1143	GFO_1158	16
	158	FCSP	GFO_1698	GFO_1738	41
	159	alpha-glucan	GFO_2130	GFO_2141	12
	160	alpha-glucan	GFO_2989	GFO_2998	10
	161	laminarin A	GFO_3462	GFO_3468	7

Gramella sp. Hel1_59	162	laminarin A	JM79DRAFT_0027	JM79DRAFT_0033	7
	163	pectin	JM79DRAFT_0175	JM79DRAFT_0206	32
	164	alpha-glucan	JM79DRAFT_1217	JM79DRAFT_1226	10
Gramella sp. MAR_2010_102	165	unknown	Ga0075139_0791	Ga0075139_0799	9
	166	laminarin D	Ga0075139_1340	Ga0075139_1349	10
	167	laminarin A	Ga0075139_1356	Ga0075139_1362	7
	168	NAG	Ga0075139_1488	Ga0075139_1500	13
	169	alpha-glucan	Ga0075139_2463	Ga0075139_2477	15
Gramella sp. MAR_2010_147	170	alpha-glucan	Ga0075141_0214	Ga0075141_0227	14
	171	unknown	Ga0075141_1368	Ga0075141_1374	7
	172	unknown	Ga0075141_1445	Ga0075141_1455	11
	173	laminarin A	Ga0075141_1946	Ga0075141_1952	7
	174	NAG	Ga0075141_2085	Ga0075141_2098	14
	175	pectin	Ga0075141_2732	Ga0075141_2750	19
	176	pectin	Ga0075141_2754	Ga0075141_2783	30
Lacinutrix sp. Hel1_90	177	laminarin A	GQ46DRAFT_0203	GQ46DRAFT_0214	11
	178	NAG	GQ46DRAFT_1622	GQ46DRAFT_1639	18
	179	alpha-glucan	GQ46DRAFT_2445	GQ46DRAFT_2461	16
	180	unknown	GQ46DRAFT_2490	GQ46DRAFT_2496	7
Leeuwenhoekiella sp. MAR_2009_132	181	NAG	P162DRAFT_2505	P162DRAFT_2523	19
	182	pectin	P162DRAFT_2533	P162DRAFT_2563	31
	183	unknown	P162DRAFT_2903	P162DRAFT_2943	41
	184	unknown	P162DRAFT_3223	P162DRAFT_3229	7
	185	unknown	P162DRAFT_0078	P162DRAFT_0111	34
	186	laminarin A	P162DRAFT_0225	P162DRAFT_0231	7
	187	unknown	P162DRAFT_0293	P162DRAFT_0303	11
	188	unknown	P162DRAFT_0572	P162DRAFT_0589	18
	189	alpha-glucan	P162DRAFT_0963	P162DRAFT_0968	6
	190	unknown	P162DRAFT_1094	P162DRAFT_1102	9
	191	unknown	P162DRAFT_1154	P162DRAFT_1160	7
	192	xylan	P162DRAFT_1214	P162DRAFT_1239	24
	193	unknown	P162DRAFT_1311	P162DRAFT_1324	14
	194	unknown	P162DRAFT_1328	P162DRAFT_1337	9
	195	alginate	P162DRAFT_1609	P162DRAFT_1616	8
	196	beta-mannan	P162DRAFT_1668	P162DRAFT_1685	18
	197	galactose-containing polysaccharide	P162DRAFT_1700	P162DRAFT_1708	9
Flavimarina sp. Hel1_48	198	alpha-mannan	P164DRAFT_3387	P164DRAFT_3402	16
	199	arabinoxylan	P164DRAFT_0131	P164DRAFT_0140	10
	200	sialic acid polysaccharide	P164DRAFT_0312	P164DRAFT_0318	7
	201	unknown	P164DRAFT_0435	P164DRAFT_0446	10
	202	xylan	P164DRAFT_0473	P164DRAFT_0491	19
	203	xylan	P164DRAFT_0830	P164DRAFT_0844	14
	204	unknown	P164DRAFT_0993	P164DRAFT_1006	14
	205	pectin	P164DRAFT_1077	P164DRAFT_1095	18
	206	unknown	P164DRAFT_1644	P164DRAFT_1648	5
	207	unknown	P164DRAFT_1690	P164DRAFT_1697	8

Flavimarina sp. Hel1_48	208	fucomannan	P164DRAFT_2366	P164DRAFT_2380	14
	209	unknown	P164DRAFT_2696	P164DRAFT_2702	7
	210	alpha-glucan	P164DRAFT_2858	P164DRAFT_2864	7
	211	beta-mannan	P164DRAFT_2911	P164DRAFT_2924	14
	212	unknown	P164DRAFT_2920	P164DRAFT_2932	11
Lutibacter sp. Hel1_33_5	213	alpha-mannan	Ga0052900_0647	Ga0052900_0658	12
	214	FCSP	Ga0052900_0894	Ga0052900_0917	24
	215	laminarin B	Ga0052900_2533	Ga0052900_2544	12
	216	alginate	Ga0052900_2632	Ga0052900_2649	18
	217	unknown	Ga0052900_2652	Ga0052900_2660	9
Maribacter dokdonensis MAR_2009_71	218	alginate	Ga0040978_0453	Ga0040978_0465	13
	219	unknown	Ga0040978_0522	Ga0040978_0531	10
	220	alginate	Ga0040978_0536	Ga0040978_0548	13
	221	alpha-glucan	Ga0040978_0714	Ga0040978_0723	10
	222	digeneaside	Ga0040978_1404	Ga0040978_1407	4
	223	unknown	Ga0040978_1626	Ga0040978_1644	19
	224	unknown	Ga0040978_2075	Ga0040978_2093	19
	225	xylan	Ga0040978_2701	Ga0040978_2713	13
	226	FCSP	Ga0040978_3697	Ga0040978_3799	102
Maribacter forsetii sp. DSM_18668	227	digeneaside	P177DRAFT_00429	P177DRAFT_00433	5
	228	unknown	P177DRAFT_01100	P177DRAFT_01104	5
	229	unknown	P177DRAFT_01105	P177DRAFT_01116	12
	230	unknown	P177DRAFT_01117	P177DRAFT_01131	15
	231	FCSP	P177DRAFT_02775	P177DRAFT_02801	27
	232	alginate	P177DRAFT_03270	P177DRAFT_03276	7
	233	unknown	P177DRAFT_03387	P177DRAFT_03412	26
	234	alginate	P177DRAFT_03419	P177DRAFT_03431	13
	235	alpha-glucan	P177DRAFT_03609	P177DRAFT_03617	9
Maribacter sp. Hel1_7	236	unknown	P178DRAFT_0359	P178DRAFT_0375	16
	237	fructose	P178DRAFT_0828	P178DRAFT_0833	6
	238	unknown	P178DRAFT_0877	P178DRAFT_0904	28
	239	digeneaside	P178DRAFT_1079	P178DRAFT_1082	4
	240	alpha-glucan	P178DRAFT_1805	P178DRAFT_1814	10
	241	alginate	P178DRAFT_2004	P178DRAFT_2013	10
	242	unknown	P178DRAFT_2024	P178DRAFT_2033	10
	243	glucuronoxylan	P178DRAFT_2646	P178DRAFT_2655	10
	244	unknown	P178DRAFT_2768	P178DRAFT_2778	11
	245	alpha-mannan	P178DRAFT_2779	P178DRAFT_2795	17
	246	FCSP	P178DRAFT_3731	P178DRAFT_3759	29
	247	alginate	P178DRAFT_4207	P178DRAFT_4213	7
Maribacter sp. MAR_2009_60	248	alginate	Ga0040979_0145	Ga0040979_0157	13
	249	unknown	Ga0040979_0165	Ga0040979_0172	8
	250	alginate	Ga0040979_0231	Ga0040979_0243	13
	251	FCSP	Ga0040979_0687	Ga0040979_0794	108
	252	xylan	Ga0040979_1801	Ga0040979_1812	12
	253	unknown	Ga0040979_2379	Ga0040979_2392	14
	254	digeneaside	Ga0040979_3145	Ga0040979_3149	5
	255	alpha-glucan	Ga0040979_3839	Ga0040979_3848	10

Maribacter sp. MAR_2009_72	256	glucuronoxylan	JM81DRAFT_0176	JM81DRAFT_0188	13
	257	unknown	JM81DRAFT_0271	JM81DRAFT_0280	9
	258	unknown	JM81DRAFT_0753	JM81DRAFT_0765	13
	259	unknown	JM81DRAFT_0994	JM81DRAFT_1012	19
	260	rhamnose-containing substrate	JM81DRAFT_1105	JM81DRAFT_1161	53
	261	alpha-glucan	JM81DRAFT_1936	JM81DRAFT_1945	10
	262	sialic acid polysaccharide	JM81DRAFT_2789	JM81DRAFT_2801	13
	263	pectin	JM81DRAFT_2819	JM81DRAFT_2835	17
	264	pectin	JM81DRAFT_3351	JM81DRAFT_3367	16
Muricauda sp. MAR_2010_75	265	unknown	FG28DRAFT_0350	FG28DRAFT_0370	19
	266	beta-mannan	FG28DRAFT_0451	FG28DRAFT_0484	34
	267	arabinan	FG28DRAFT_0505	FG28DRAFT_0519	15
	268	alpha-glucan	FG28DRAFT_0880	FG28DRAFT_0887	8
	269	glucuronoxylan	FG28DRAFT_1588	FG28DRAFT_1602	15
	270	unknown	FG28DRAFT_1894	FG28DRAFT_1899	6
	271	xylan	FG28DRAFT_2756	FG28DRAFT_2771	16
	272	unknown	FG28DRAFT_2775	FG28DRAFT_2787	13
	273	unknown	FG28DRAFT_2817	FG28DRAFT_2839	22
	274	unknown	FG28DRAFT_3032	FG28DRAFT_3038	7
Nonlabens sp. Hel1_33_55	275	FCSP	Ga0052909_0637	Ga0052909_0660	24
	276	laminarin B	Ga0052909_1665	Ga0052909_1676	12
	277	pectin	Ga0052909_2762	Ga0052909_2776	14
	278	alpha-glucan	Ga0052909_2827	Ga0052909_2837	11
Nonlabens sp. Hel1_38	279	alpha-glucan	LY99DRAFT_0128	LY99DRAFT_0138	11
Nonlabens sp. Hel1_56	280	laminarin B	LZ06DRAFT_2562	LZ06DRAFT_2570	9
	281	galactose-containing polysaccharide	LZ06DRAFT_3549	LZ06DRAFT_3556	7
Olleya sp. Hel1_94	282	laminarin A	JM82DRAFT_0745	JM82DRAFT_0751	7
	283	alginate	JM82DRAFT_1098	JM82DRAFT_1107	10
	284	digeneaside	JM82DRAFT_1900	JM82DRAFT_1904	5
	285	unknown	JM82DRAFT_2583	JM82DRAFT_2593	11
	286	alpha-glucan	JM82DRAFT_3125	JM82DRAFT_3136	12
Polaribacter sp. Hel1_33_49	287	alpha-glucan	PHEL49_359	PHEL49_372	14
	288	laminarin C	PHEL49_707	PHEL49_717	11
	289	alpha-mannan	PHEL49_1317	PHEL49_1360	43
	290	unknown	PHEL49_1439	PHEL49_1446	8
	291	laminarin B	PHEL49_2073	PHEL49_2086	14
Polaribacter sp. Hel1_33_78	292	NAG	Ga0040983_0075	Ga0040983_0081	7
	293	laminarin C	Ga0040983_0487	Ga0040983_0499	13
	294	unknown	Ga0040983_0911	Ga0040983_0922	12
	295	unknown	Ga0040983_0923	Ga0040983_0944	20
	296	alpha-mannan	Ga0040983_1057	Ga0040983_1097	41
	297	laminarin B	Ga0040983_1840	Ga0040983_1854	14
Polaribacter sp. Hel1_33_96	298	alpha-glucan	Ga0040984_0369	Ga0040984_0382	14
	299	laminarin B	Ga0040984_0911	Ga0040984_0925	14
	300	alpha-mannan	Ga0040984_1648	Ga0040984_1687	40
	301	laminarin C	Ga0040984_2204	Ga0040984_2210	7

Polaribacter sp. Hel1_85	302	alginate	PHEL85_1558	PHEL85_1584	27
	303	alginate	PHEL85_1573	PHEL85_1584	12
	304	FCSP	PHEL85_1983	PHEL85_2011	29
	305	FCSP	PHEL85_2077	PHEL85_2114	37
	306	unknown	PHEL85_156	PHEL85_163	7
	307	laminarin B	PHEL85_889	PHEL85_900	12
	308	alpha-glucan	PHEL85_902	PHEL85_915	14
	309	FCSP	PHEL85_3144	PHEL85_3178	35
Polaribacter sp. Hel1_88	310	unknown	P161DRAFT_0037	P161DRAFT_0042	6
	311	laminarin A	P161DRAFT_0706	P161DRAFT_0710	5
	312	sulfated substrate rich in galactose	P161DRAFT_1142	P161DRAFT_1176	33
	313	alginate	P161DRAFT_1497	P161DRAFT_1510	14
	314	unknown	P161DRAFT_2311	P161DRAFT_2325	15
	315	galactose-containing sulfated polysaccharide	P161DRAFT_2458	P161DRAFT_2485	26
	316	rhamnose-containing substrate	P161DRAFT_2488	P161DRAFT_2523	32
	317	laminarin B	P161DRAFT_3459	P161DRAFT_3475	17
	318	alpha-glucan	P161DRAFT_3477	P161DRAFT_3492	16
Polaribacter sp. KT25b	319	FCSP	Ga0114009_0001	Ga0114009_0050	49
	320	digeneaside	Ga0114009_0079	Ga0114009_0083	5
	321	unknown	Ga0114009_0466	Ga0114009_0491	23
	322	carrageenan	Ga0114009_0492	Ga0114009_0529	37
	323	unknown	Ga0114009_0799	Ga0114009_0817	19
	324	laminarin B	Ga0114009_1227	Ga0114009_1239	13
	325	alginate	Ga0114009_1868	Ga0114009_1877	10
	326	alpha-glucan	Ga0114009_1916	Ga0114009_1930	15
	327	FCSP	Ga0114009_2381	Ga0114009_2415	34
	328	alginate	Ga0114009_3492	Ga0114009_3501	10
	329	unknown	Ga0114009_3528	Ga0114009_3545	18
Psychroserpens sp. Hel1_66	330	unknown	GQ40DRAFT_0301	GQ40DRAFT_0311	11
	331	laminarin A	GQ40DRAFT_0702	GQ40DRAFT_0711	10
	332	unknown	GQ40DRAFT_1713	GQ40DRAFT_1720	8
	333	laminarin B	GQ40DRAFT_2819	GQ40DRAFT_2823	5
	334	pectin	GQ40DRAFT_3221	GQ40DRAFT_3238	18
	335	alpha-glucan	GQ40DRAFT_3242	GQ40DRAFT_3256	15
Salegentibacter sp. Hel1_6	336	unknown	FG27DRAFT_0952	FG27DRAFT_0967	16
	337	xylose-containing polysaccharide	FG27DRAFT_1286	FG27DRAFT_1296	10
	338	glucuronoxylan	FG27DRAFT_1304	FG27DRAFT_1309	6
	339	xylan	FG27DRAFT_1324	FG27DRAFT_1343	20
	340	alpha-mannan	FG27DRAFT_1659	FG27DRAFT_1700	39
	341	laminarin A	FG27DRAFT_1942	FG27DRAFT_1947	6
	342	beta-mannan	FG27DRAFT_2043	FG27DRAFT_2059	17
	343	fructose	FG27DRAFT_2307	FG27DRAFT_2317	11
	344	alpha-glucan	FG27DRAFT_3244	FG27DRAFT_3253	10

Tenacibaculum sp. MAR_2009_124	345	unknown	Ga0075138_2336	Ga0075138_2352	17
	346	laminarin B	Ga0075138_3522	Ga0075138_3533	12
	347	galactose-containing sulfated polysaccharide	Ga0075138_4394	Ga0075138_4403	10
	348	chitin	Ga0075138_0072	Ga0075138_0076	5
	349	NAG	Ga0075138_0346	Ga0075138_0350	5
	350	alpha-glucan	Ga0075138_1001	Ga0075138_1013	13
	351	NAG	Ga0075138_1520	Ga0075138_1525	6
Tenacibaculum sp. MAR_2010_205	352	NAG	Ga0052903_0010	Ga0052903_0016	6
Tenacibaculum sp. MAR_2010_89	353	chitin	Ga0052902_0174	Ga0052902_0193	19
	354	NAG	Ga0052902_0648	Ga0052902_0656	9
Winogradskyella sp. RHA_55	355	unknown	Ga0075142_0818	Ga0075142_0825	8
	356	unknown	Ga0075142_1037	Ga0075142_1042	6
	357	unknown	Ga0075142_1046	Ga0075142_1060	15
	358	alpha-glucan	Ga0075142_1141	Ga0075142_1156	16
	359	digeneaside	Ga0075142_1277	Ga0075142_1281	5
	360	unknown	Ga0075142_1726	Ga0075142_1740	15
Zobellia amurskyensis MAR_2009_138	361	carrageenan	FG20DRAFT_0058	FG20DRAFT_0072	15
	362	unknown	FG20DRAFT_0113	FG20DRAFT_0124	11
	363	xylose-containing sulfated polysaccharide	FG20DRAFT_0127	FG20DRAFT_0132	6
	364	xylose-containing sulfated polysaccharide	FG20DRAFT_0134	FG20DRAFT_0151	17
	365	unknown	FG20DRAFT_0153	FG20DRAFT_0160	8
	366	xylose-containing sulfated polysaccharide	FG20DRAFT_0164	FG20DRAFT_0191	14
	367	xylose-containing polysaccharide	FG20DRAFT_0192	FG20DRAFT_0205	14
	368	unknown	FG20DRAFT_0206	FG20DRAFT_0219	12
	369	unknown	FG20DRAFT_0247	FG20DRAFT_0258	12
	370	sialic acid polysaccharide	FG20DRAFT_0262	FG20DRAFT_0276	15
	371	unknown	FG20DRAFT_0291	FG20DRAFT_0297	7
	372	xylan	FG20DRAFT_0368	FG20DRAFT_0374	7
	373	unknown	FG20DRAFT_0650	FG20DRAFT_0667	18
	374	unknown	FG20DRAFT_0674	FG20DRAFT_0689	16
	375	Alpha-mannose-containing substrate	FG20DRAFT_0711	FG20DRAFT_0715	5
	376	alginate	FG20DRAFT_0818	FG20DRAFT_0823	6
	377	unknown	FG20DRAFT_1180	FG20DRAFT_1191	12
	378	unknown	FG20DRAFT_1199	FG20DRAFT_1206	8
	379	unknown	FG20DRAFT_1284	FG20DRAFT_1292	8

Zobellia amurskyensis MAR_2009_138	380	unknown	FG20DRAFT_1305	FG20DRAFT_1316	12
	381	unknown	FG20DRAFT_1383	FG20DRAFT_1393	11
	382	unknown	FG20DRAFT_1475	FG20DRAFT_1488	14
	383	FCSP	FG20DRAFT_1584	FG20DRAFT_1614	29
	384	unknown	FG20DRAFT_1621	FG20DRAFT_1634	14
	385	unknown	FG20DRAFT_1652	FG20DRAFT_1659	8
	386	unknown	FG20DRAFT_1697	FG20DRAFT_1704	7
	387	FCSP	FG20DRAFT_1710	FG20DRAFT_1746	37
	388	rhamnose-containing substrate	FG20DRAFT_1747	FG20DRAFT_1752	6
	389	rhamnose-containing substrate	FG20DRAFT_1753	FG20DRAFT_1758	6
	390	rhamnose-containing substrate	FG20DRAFT_1794	FG20DRAFT_1841	44
	391	unknown	FG20DRAFT_1865	FG20DRAFT_1875	11
	392	unknown	FG20DRAFT_2485	FG20DRAFT_2506	21
	393	unknown	FG20DRAFT_2528	FG20DRAFT_2534	7
	394	alpha-glucan	FG20DRAFT_2714	FG20DRAFT_2728	15
	395	unknown	FG20DRAFT_2828	FG20DRAFT_2834	7
	396	unknown	FG20DRAFT_3429	FG20DRAFT_3437	9
	397	galactose-containing sulfated polysaccharide	FG20DRAFT_3652	FG20DRAFT_3667	16
	398	unknown	FG20DRAFT_3668	FG20DRAFT_3681	14
	399	digeneaside	FG20DRAFT_3685	FG20DRAFT_3689	5
	400	alginate	FG20DRAFT_4039	FG20DRAFT_4055	16

Chapter V: General discussion

This thesis contributed to several large projects in the Max Planck Institute for Marine Microbiology that investigated the utilization of organic matter by heterotrophic surface bacteria. For the year 2009, a spring phytoplankton bloom in the North Sea was investigated in great detail in the MIMAS project. It described how phytoplankton-derived organic matter gave rise to a succession of heterotrophic bacterioplankton clades, equipped with specialized proteins for substrate uptake and degradation (Teeling *et al.*, 2012). It was hypothesized that the dominant driving factor for the rapid cell number increase of MHB was bottom-up organic matter supply by the phytoplankton.

In the following years, investigations continued in the COGITO project. It tested the hypothesis that these phytoplankton-induced blooms of heterotrophic bacteria follow recurring, predictable patterns at the taxonomic and functional level (Teeling *et al.*, 2016, chapter III). A large team of scientists sampled surface waters throughout the spring seasons of 2010 to 2012 and subsequently sequenced genomes, metagenomes, metaproteomes and 16S rRNA tags, complementing the 2009 data and adding up to four consecutive years of extensive data. We found that the spring blooms were dominated by different diatom species of the genera *Chaetoceros*, *Thalassiosira* and *Mediopyxis*, but also the silicoflagellate *Chattonella* in varying abundances (Figure 3.3I-L). The phytoplankton blooms were followed by recurrent, swift and dynamic successions of heterotrophic bacterioplankton dominated by representatives of the families *Flavobacteriia*, *Gammaproteobacteria* and, in later stages, the alphaproteobacterial *Roseobacter* clade (Figure 3.3M-P). Among the highly abundant *Flavobacteriia* genera were *Formosa*, *Polaribacter* and *Ulvibacter* (Figure 3.3Q-T; *Ulvibacter* is currently reclassified as *Prosiliobacter* by Thomas Benjamin Francis and colleagues (unpublished)). In several

cases, relative bacterial abundances of these genera reached 25% and more in less than a week. In *Gammaproteobacteria*, members of the SAR92 clade, *Reinekea* and *Alteromonadales* likewise reached abundances of up to 16% (Figure 3.3U-X). The recurrence of many of these key players was remarkable regarding the preceding microalgal diversity. Even though the abundances of certain prominent *Flavobacteriia* clades were correlated with multiple algae groups and physiochemical factors according to Spearman rank correlation analyses, simple one-to-one relationship between specific algae and specific bacterioplankton groups could not be detected. This suggests that the different blooming algal species produce similar substrate types for specifically adapted clades of heterotrophic bacterioplankton. These substrate types likely exert a deterministic effect on the bacterioplankton community composition. This effect was most pronounced for *Flavobacteriia*, which are also the most specialized polysaccharide degraders with the most distinct CAZyme repertoires, and less pronounced for *Alpha-* or *Gammaproteobacteria*.

With respect to the taxonomic recurrence observed in chapter III, it is important to note that the resolution of the taxonomic classification of metagenome contigs “only” reached down to genus level (or sometimes less), due to a lack of reference sequences from sufficiently close cultured relatives. Hence, there is microdiversity in the observed clades. For example certain *Flavobacteriia* clades, such as *Polaribacter* and *Formosa*, were shown to be represented by different “oligotypes” (closely related but distinct bacterial taxa) sharing >99% 16S rRNA sequence identity in the V4 region (Chafee *et al.*, 2017). Two *Polaribacter* strains (Hel1_33_49 and Hel1_85, Xing *et al.*, 2015) isolated during the 2010 spring diatom bloom from Helgoland surface seawaters even showed 100% V4 region identity, however, their genomes only shared 85% average nucleotide identity with distinct ecological niches (see chapter 5.2).

Not all bacterioplankton clades recurred every year. Assuming our previous hypothesis that similar substrate types are produced by the blooming algae, a possible explanation could be that species from different taxa with similar ecophysiological niches can functionally substitute each other in subsequent years. An example for this could be laminarin degradation, which as I showed in chapter IV many MHB are genetically capable of (Figure 4.1). In this case, additional adaptations likely enable certain clades to dominate the initial bacterial bloom phase in which laminarin becomes available from decaying algae. Such adaptations were described in detail for two strains of the *Formosa* genus analyzed in chapter IV (spp. Hel1_33_131 and Hel3_A1_48), sampled at the same location of the spring blooms (Unfried *et al*., in review, see appendix). The strains featured small, streamlined genomes (Supplementary Table S4.1) and enhanced uptake of diatom-derived peptides to sustain high growth rates and therefore outcompete other bacteria. After the easily degradable storage polysaccharides are consumed, more recalcitrant substrates such as branched and sulfated polysaccharides, requiring more sophisticated degradation enzyme assemblies, may be tackled by specifically adapted microbes.

It is the linking element of all chapters of this doctoral thesis that *Flavobacteriia* are highly specialized for the degradation of specific polysaccharides and peptides, which provide them with distinct ecological niches. These heterotrophic bacteria live pelagically or attached in environments of different trophic levels from oligotrophic to eutrophic. In many cases, these characteristics are interrelated. For example a typical pelagic MHB living in the open ocean will likely face oligotrophic conditions with lower concentrations of available polysaccharides and hence focus more on peptides, while an algae-attached bacterium will be highly adapted towards a eutrophic environment due to the constant supply of algal exudates rich in polysaccharides.

In chapter II, I observed genetic adaptions for the degradation of algal and plant polysaccharides for *Bacteroidetes*-affiliated fosmids from the eutrophic Boreal Polar region (Station 3 of the VISION cruise 2006). These fosmids showcased a high prevalence of algae and plant-polysaccharide targeting CAZyme genes (e.g. GH92) as well as sulfatase genes at a site where marine algae that produce sulfated polysaccharides are more abundant. Conversely, higher gene frequencies of peptidases and CAZymes for the degradation of bacterial- and animal-derived substrates (e.g. GH31 and CBM50) were found on the fosmids from the contrasting oligotrophic North Atlantic Subtropical region (Station 18). Interestingly, degradation capacities for laminarin and xylan-rich polysaccharides were found at both stations, indicating that these possibly belong to the more frequent substrates for *Bacteroidetes* in both mesotrophic and oligotrophic open oceans.

In chapter III, the annual relative abundance increase of *Flavobacteriia* during blooms was always accompanied by increases in community-wide CAZyme gene frequencies as well as increases in the diversity of CAZyme families from pre- to mid-bloom situations which both leveled off after the bloom. The most frequent families were indicative of the hydrolysis of laminarin (GH3, 5, 16, 17 and 30, CBM4 and 6), xylans (CE1, 4), alpha-mannans (GH92), fucans (GH29), alpha-1,4-glucans (GH13, 31, 65) and peptidoglycans (GH23, 103, 73, 18, CBM50; Figure 3.6I-L). Likewise, gene frequencies of SusC-like TBDT and sulfatases increased (Figure 3.6M). The CAZyme repertoires of many *Flavobacteriia* clades showed fingerprint-like patterns indicating individual glycan niches, for example alpha-mannan degradation (GH92) on *Formosa* and *Polaribacter*-affiliated contigs. I found corresponding PULs in many *Flavobacteriia* seawater isolates at the same sites, indicating that these polysaccharides likely represent typical phytoplankton-derived substrates that

adapted, free-living *Flavobacteriia* compete for (laminarin: Figure 4.2; xylan: Figure 4.5B; alpha-mannans: Figures 4.4A and B; fucans: Figure 4.5A; alpha-1,4-glucan: Figure 4.3A; isolates obtained by Richard Hahnke and Jens Harder)

In contrast to these free-living representatives, I could show that many *Flavobacteriia* isolated on solid media featured large, CAZyme-rich genomes with many PULs likely encoding genes for the degradation of polysaccharides typically occurring in macroalgae (chapter IV, Figure 4.1). Among these are sulfated α-galactose-containing substrates such as carrageenan (red algae, Aquino *et al.,* 2005), sulfated α-rhamnose-containing substrates (green algae, Wang *et al.,* 2014), diverse FCSPs (Kloareg *et al.,* 1988; Popper *et al.,* 2011), digeneaside (red algae exudates, Kremer, 1980) and pectin or pectin-like substrates (Barbeyron *et al*., 2016). Many of these PULs were notably larger and showed little conservation. I categorized them based on recurring CAZyme families (as described in chapter IV, paragraphs 4.4.7 to 4.4.11), but many of these PULs feature different additional CAZyme genes, and it is likely that they target polysaccharides of slightly or even significantly different structures. For example, FCSPs are prominently found in brown algal hemicelluloses, but were also reported in the EPS of diatoms (Gügi *et al.,* 2015). Only profound biochemical investigations can ultimately confirm what algal substrates these new PULs actually target.

In accordance with the above-mentioned microdiversity within distinct clades occurring during North Sea spring phytoplankton blooms, we observed diverse polysaccharide degradation capacities even among closely related North Sea *Flavobacteriia* isolates, indicating distinct glycan niches with no distinct correlation to taxonomy. For marine *Flavobacteriia*, this was described for the first time for the two Helgoland surface seawater *Polaribacter* strains Hel1_33_49 and Hel1_85 (Xing *et al.*, 2015). The relatively small 3.0

Mbp genome of strain Hel1_33_49 reflects a typical pelagic lifestyle with a higher peptidase:CAZyme ratio and a limited number of PULs targeting likely diatom-derived substrates (e.g. laminarin, sulfated α-glucuronomannan, Figure 4.1). The 3.9 Mbp genome of strain Hel1_85 in contrast featured PULs for the degradation of alginate and FCSPs (Figures 4.3B and 4.5A), as they for example occur in brown algae (Fischer and Dörfel, 1955; Kloareg *et al.,* 1988; Popper *et al.,* 2011), as well as genes to thrive under low oxygen conditions such as biofilms, indicating an algae-attached lifestyle. As described in chapter IV, I could detect such genus-level diversity also in several isolates of the genera *Cellulophaga*, *Gramella* and *Polaribacter*. They featured highly varying PUL numbers (*Cellulophaga:* 5-10, *Gramella*: 3-10, *Polaribacter*: 4-11, Supplementary Figure S4.1) targeting diverse polysaccharide classes. Such significant PUL repertoire diversity between species of the same genera supports the hypothesis that PULs are frequently exchanged between *Flavobacteriia* through horizontal gene transfer (Hehemann *et al.,* 2012).

Although many *Bacteroidetes* are „polysaccharide specialists“, certain isolates showcased only a limited polysaccharide degradation capacity that was mostly restricted to a few, likely abundant glycan substrates (e.g. laminarin and α-1,4-glucan). Instead, these isolates were more specialized on the degradation of peptides (Supplementary table S4.1). In isolate genomes of the *Flavobacteriia Formosa*, *Ulvibacter* (*Cd.* '*Prosiliobacter*') and *Polaribacter*, which recurrently outcompeted other bacteria in early phytoplankton bloom phases, I detected above-average Peptidase:CAZyme ratios in *Formosa* sp. Hel1_33_131 (1.81), *Ulvibacter* sp. MAR_2010_11 (1.88) and *Polaribacter* sp. Hel1_33_49 (1.39, compared to e.g. 0.97 in *Polaribacter* sp. KT25b). Peptidases were also found to be expressed in *Formosa* sp. Hel1_33_131 (Unfried *et al.*, in review, see appendix).

Polysaccharides are generally nitrogen-depleted, and the amine (NH_2) groups of peptides offer a nitrogen source that allows sustaining fast growth rates.

A major goal of this thesis was to gain new insights into the utilization of marine polysaccharides, which are structurally so diverse that no single bacterium can degrade them all. Instead, different bacterial clades fill distinct glycan niches (Xing *et al.*, 2015; Hehemann *et al.*, 2016). My analyses revealed certain PULs such as those encoding degradation of laminarin and α-1,4-glucan to be very widespread among *Flavobacteriia* isolates. The respective CAZymes were present in high gene frequencies in the phytoplankton bloom metagenomes and open-ocean fosmids (laminarin and α-1,4-glucan). Other PULs possibly targeting environmentally abundant substrates were not widespread in the isolates, which can be attributed to a cultivation bias (see Outlook). These PULs were found in liquid culture isolates with high abundances in the spring phytoplankton blooms, suggesting they might be of environmental importance (α-glucuronomannan, sulfated xylans). In this chapter, I will discuss these selected algal polysaccharides and their PULs in greater detail.

Laminarin

PULs encoding CAZymes required for laminarin degradation were abundant in all genomic data investigated in this thesis. They were present in the metagenomes of all spring phytoplankton bloom years, on the fosmids retrieved from both open-ocean sampling stations and in PULs of more than three quarters of the 53 *Flavobacteriia* genomes which I analyzed. The *susCD*-like genes (encoding substrate-binding and uptake) of laminarin PULs were shown to be expressed during the spring blooms in 2009 and 2010. These findings suggest that my approach to monitor glycan substrates through the expression levels of the respective SusCD homologs is meaningful, since laminarin is indeed the most

abundant marine polysaccharide on Earth. The annual production of diatom-derived chrysolaminarin is estimated to amount to 5-15 Gt (Alderkamp *et al.*, 2007).

In the *Flavobacteriia* isolates, I detected four variants of laminarin PULs, two of which had been experimentally verified (Kabisch *et al.,* 2014; Xing *et al.,* 2014). In their review on diatom-specific polysaccharide structures, Gügi and colleagues summarize that the known diatom-derived chrysolaminarins are “usually composed of a β-1,3-glucan backbone chain ramified with β-1,6-glucose and sometimes with β-1,2-glucose.” The length of the backbone chain and the degree of ramification vary with the diatom species (Gügi *et al*., 2015). It hence seems possible that different laminarin subtypes are targeted by different PUL subtypes. Due to its significant share in the marine NPP, it would be interesting to biochemically elucidate the differences of the laminarin-targeting PUL subtypes.

Alpha-1,4-glucans

Alpha-1,4-glucans occur as an energy storage compound in green plants as starch and in animals, fungi and bacteria as glycogen. I found α-1,4-glucan-targeting CAZymes and PUL to be equally prominent as laminarin in metagenomes, fosmids and isolates, and corresponding substrate-binding proteins were expressed from early to late bloom stages. In contrast to laminarin, they were also very prominent in macroalgae-attached isolates. While it seems plausible that some of the bacterial isolates, especially the macroalgae-attached ones, populate coastal subtidal and intertidal habitats and are therefore exposed to starch of terrestrial origin, genes encoding α-1,4-glucan-specific CAZymes were also found in an open ocean fosmid PUL (VISS18_034, Figure 2.4) from the oligotrophic open ocean North Atlantic Subtropical region. Hence, these α-1,4-glucan PULs likely target bacterial glycogen. It has been estimated that on average, about 20% of marine heterotrophic bacteria are infected by viruses and 10–20% of the bacterial community is

lysed daily by viruses (Suttle, 1994). In the phytoplankton blooms, we observed the rapid collapse of highly abundant bacterial groups, during which large amounts of glycogen must become available to the bacterial community.

Sulfated alpha-glucuronomannan

Alpha-glucuronomannan is one of the more complex substrates that might be targeted after the structurally simpler, easily degradable storage polysaccharides are consumed. This PUL is only encoded in the genomes of two *Flavobacteriia* which belonged, however, to clades abundant during the spring phytoplankton blooms, namely in *Polaribacter* spp. Hel1_33_49, _78, _96 (abundant in 2010) and, as a reduced PUL, in *Formosa* sp. Hel1_33_131 (abundant in 2009). In both strains, the PUL encodes multiple GH92 α-mannosidases and sulfatases. In *Polaribacter* spp. Hel1_33_49, _78, it additionally encodes a predicted GH88 β-glucuronyl hydrolase, a peptidase and a GH99-like glycoprotein endo-α-mannosidase. I found GH92 α-mannosidases to be recurrently present at high frequencies during spring phytoplankton blooms. Costaouëc and colleagues (2017) recently revealed the main cell wall polysaccharide of the diatom *Phaeodactylum tricornutum* to be a “linear poly-α-1,3-mannan decorated with sulfate ester groups and β-D-glucuronic residues” showing ‘strong interactions with cell-wall proteins” (Le Costaouëc *et al*., 2017), indicating that the targeted substrate is possibly a glycoprotein. This PUL might hence not only supply its host with valuable energy (carbon), but also nitrogen-rich amino acids to sustain the high growth rates observed for *Polaribacter* and *Formosa*, both reaching ~25% bacterial abundance within days in response to the phytoplankton blooms.

Sulfated xylose-rich substrate

In addition I predict a sulfated xylose-rich substrate to be important based on two highly similar PULs found in the genome of isolate *Formosa* sp. Hel3_A1_48 and on a *Bacteroidetes* fosmid obtained from the polar station 18 in the North Atlantic (Fosmid-PUL VISS18_012). Homologs of the SusCD-like proteins of this PUL were expressed in the late phase of the 2009 diatom bloom. This PUL encodes a GH10 β-xylanase, putative GH43 and GH30 β-xylosidases and no less than 5 sulfatases. Between the GH43 and GH10 genes, a predicted xyloside transporter gene (*xynT*) likely encodes xyloside transport into the cytoplasm. While a sulfated xylan has been described in the red macroalga *Palmaria palmata* (Deniaud *et al.*, 2003), the organisms harboring this PUL type - a bloom-associated *Formosa* sp. and an open-ocean North Atlantic surface water bacteroidete - make it more plausible that this substrate could be originating from diatom EPS, for which sulfatation of xylose-rich EPS have been reported (Gügi *et al*., 2015). This includes the genera *Chaetoceros* and *Thalassiosira*, which were among the abundant diatoms in the phytoplankton blooms.

Reduced PUL variants

In many cases, the investigated *Flavobacteriia* featured reduced versions of PULs (e.g. laminarin, alginate and α-1,4-glucan PULs). These PULs lacked enzymes to fully degrade the targeted polysaccharide, such as the debranching GH30 β-1,6-glucosidase in laminarin PULs or GH97 α-1,6-glucosidase in α-1,4-glucan PULs. A possible explanation is that these organisms scavenge on substrates that were extracellularly “predigested” by other pioneering bacteria. This hypothesis was described for alginate utilization in *Vibrio* spp. (Hehemann *et al.*, 2016). It was suggested that horizontal gene transfer of PULs may even lead to fine-scale ecophysiological differentiation.

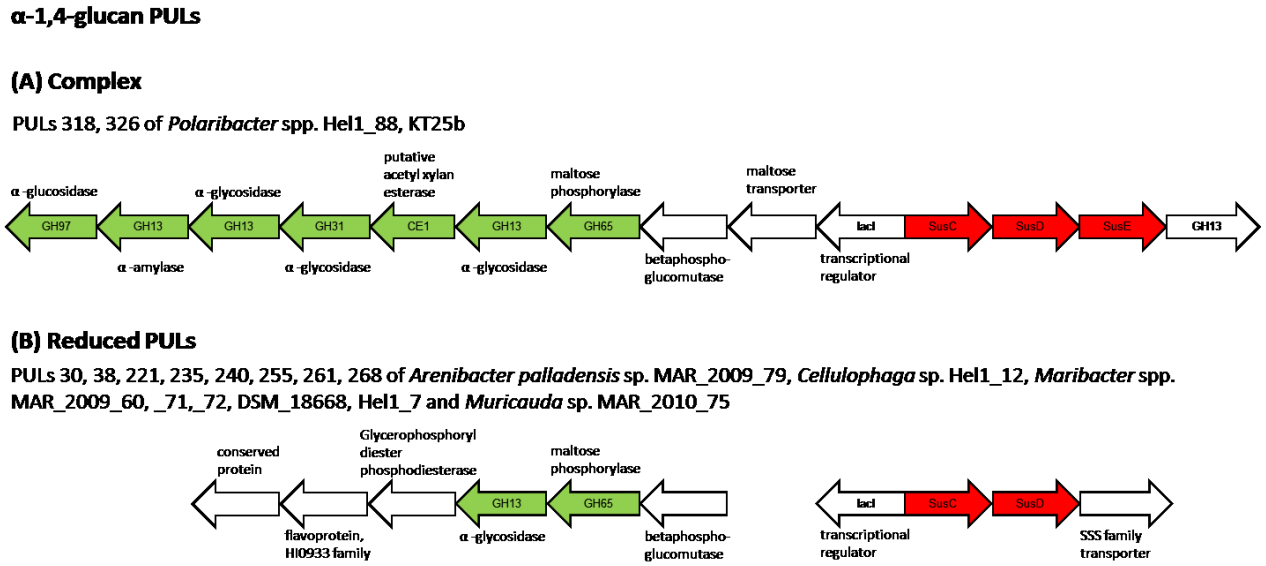


Figure 5.1: (A) Complex and (B) reduced α-1,4-glucan PUL variants found in different *Flavobacteriia* isolates

Conclusions and outlook

The work presented in this thesis demonstrates how sequencing-based PUL analyses can provide predictions for the polysaccharide utilization capacities of heterotrophic bacteria and possibly lay the foundation for understanding their ecological niche. If (meta-) proteomic analyses show high expression levels of corresponding substrate uptake and/or degradation proteins, predictions can also be made for the substrate's significance in the environment.

It is, however, important to note that these substrate predictions can only be as good as the functional annotation of CAZyme families. The predictions should be carefully evaluated, for example considering the specificity of CAZyme families and the integrity of the catalytic site of these enzymes (e.g. key amino acids for hydrolysis). *In silico* CAZyme family affiliation and the corresponding functional prediction for the degradation of a particular substrate are based on sequence similarities, hence only the continuing efforts

of biochemically characterizing these CAZymes will continue to augment the databases and hence raise the accuracy of these predictions.

The validity of the *in silico* predictions done in this thesis were verified through cultivation of the PUL-harboring isolates with the predicted substrate as sole carbon source for laminarin with *Polaribacter* and *Formosa* spp. (Xing *et al*., 2015; Unfried *et al*., in review, appendix), α-1,4-glucan in *Gramella forsetii* KT0803 (Kabisch *et al*., 2014), β-galactomannan in *Flavimarina* sp. Hel1_48, *Leeuwenhoekiella* sp. MAR_2009_132 and *Salegentibacter* sp. Hel1_6 and (unsulfated) yeast-derived α-mannan in *Salegentibacter* sp. Hel1_6 (Chen *et al*., in review, see appendix). For some of the more complex predicted polysaccharides, such as α-glucuronomannan (-glycoprotein) and the sulfated xylose-rich substrate, the targeted substrate may first have to be obtained through cultivation and extraction of the correct microalgal compounds.

Some of these substrates may represent important intermediates in the marine carbon cycle. Considering its role, the carbon cycle has been neglected for a long time, because environmental microbiologists focused on other element cycles that feature easily measurable inorganic educts and products (e.g. methane, lactate, acetate or propionate), while the degradation of complex organic substrates such as polysaccharides is more difficult to study. It is just now that we start to get a better understanding of the bacterial degradation of complex organic matter in the marine environment by the tools of genomics and metagenomics.

Metagenome-binning seems today to be the most straight-forward approach to investigate which functional capacities are encoded in environmental microorganisms. Nonetheless, in order to ultimately test and verify these hypotheses, laboratory-based cultivation of isolates

is essential. Alternatively, CAZymes or entire PULs can be heterologous expressed and biochemically characterized. For future experiments that aim to isolate further functional key-players, I suggest to focus isolation efforts of heterotrophic bacteria on liquid cultivation, possibly using diatom extracts as substrates. It is, however, possible that relevant bacteria continue to evade isolation efforts, for example because they require the activity of other organisms predigesting the microalgae. An alternative option to enrich such missing clades is through cell sorting, which can be technologically challenging. Additionally, proteomic analyses, fluorescently labeled nucleic acid probes targeting mRNA or antibodies that directly target substrate-specific CAZyme or *susC/D*-like transporter genes can quantify the expression levels of these genes and thus infer conclusions on the importance of the targeted substrate in the environment.

The most interesting and likely environmentally relevant PULs were found in the surface seawater isolates, and among them the liquid culture isolates, which provided the *Formosa* and *Polaribacter* spp. isolates that were abundant in the phytoplankton blooms of 2009 and 2010. While the *Flavobacteriia* isolated on solid media did provide many new PULs, they had a clear isolation bias towards bacteria of surface-attached, and in many cases macroalgae-associated lifestyles. Thus, these PULs are likely not involved in the turnover of the main polysaccharides in the marine realm, which are of microalgal origin.

Future studies of the 53 *Flavobacteriia* isolates characterized in this thesis for their PUL repertoire might focus on the strains listed in table 5.1. First and foremost, growth studies to verify the substrate predictions should be performed with the predicted substrates as sole carbon source, or as close structural equivalents as possible. With the confirmed polysaccharide substrates of potential environmental relevance, experiments need to be

extended to proteomics analyses and finally to the in-depth biochemical and structural characterization of the involved CAZymes and transporter proteins.

Table 5.1: PULs and corresponding substrates of suggested for future in-depth investigation.

Priority	Predicted substrate	encoded CAZymes	Organism	PUL # *	Figure	Note
1	α-glucuronomannan (-glycoprotein)	GH92 (5x), GH88, GH99-like, GH2, GH3 (2x), GH5	*Polaribacter* sp. Hel1_33_49/78/96, *Formosa* sp. Hel1_33_131	289, 296, 300, 126	4.3	GH92-rich PUL, GH92 abundant in metagenomes, isolates abundant in 2009 & 2010 MGs
	sulfated β-xylan	GH5, GH10, GH30, GH43	Formosa sp. Hel3_A1_48	136	4.5	SusC/D homologs in metaproteomes**
2	laminarin (uncharacterized variant C)	GH5, GH16, GH30	*Formosa sp. Hel3_A1_48, Polaribacter* sp. Hel1_33_49/78/96,	133, 288, 293, 301	4.2	SusC/D homologs in metaproteomes**
	laminarin (uncharacterized variant D)	GH3, GH5, GH16, GH30	*Gillisia* sp. Hel1_86, *Gramella* sp. MAR_2010_102	148, 166	4.2	-
	arabinan	GH51, GH127	*Muricauda* sp. MAR_2010_75	267	S1	SusC/D homologs in metaproteomes**
	β-mannan	GH26 (3x), GH130, GH3, GH5 (2x), GH36	*Muricauda sp. MAR_2010_75, Leeuwenhoekiella* sp. MAR_2009_132	196	4.4	SusC/D homologs in metaproteomes**
3	FCSP	GH28, GH29, GH95, GH97, GH117	*Maribacter* forsetii sp. DSM_18668	231	4.5	relatively short FCSP PUL
	Chitin	GH18, GH19, GH20	*Aquimarina* sp. MAR_2010_214	5	S1	Chitin reported in diatom fibres and silica frustule of *Thalassiosira pseudonana* (Tesson et al., 2008)
	Trehalose	GH13, GH37	*Flavobacteriaceae* bacterium spp. MAR 2010 105, _119	103, 116	S1	SusC/D homologs in metaproteomes**

* The PUL # identifiers correspond to Supplementary Table S3 (chapter 4.9), which lists all predicted PULs and their substrates within the isolates' genomes. This table also lists the locus_tags (start and stop) of the respective PULs.

** A figure depicting the expressed SusC and SusD-like proteins identified in 3 to 0.2 µm bacterioplankton during the North Sea spring phytoplankton blooms in 2009-2012 using proteomics can be found in chapter 4.4.14 (Figure 4.7).

Figures and detailed descriptions of the gene composition of the PULs can be found in chapters 4.4.3 to 4.4.12.

Bibliography

Alderkamp, A. C., Van Rijssel, M., & Bolhuis, H. (2007). Characterization of marine bacteria and the activity of their enzyme systems involved in degradation of the algal storage glucan laminarin. *FEMS Microbiology Ecology*, *59*(1), 108-117.

Aquino, R. S., Landeira-Fernandez, A. M., Valente, A. P., Andrade, L. R., & Mourao, P. A. (2004). Occurrence of sulfated galactans in marine angiosperms: evolutionary implications. *Glycobiology*, *15*(1), 11-20.

Armbrust, E. V. (2009). The life of diatoms in the world's oceans. *Nature*, *459*(7244), 185.

Arnosti, C., Steen, A. D., Ziervogel, K., Ghobrial, S., & Jeffrey, W. H. (2011). Latitudinal gradients in degradation of marine dissolved organic carbon. *PLoS One*, *6*(12), e28900.

Azam, F., & Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nature Reviews Microbiology*, *5*(10), 782.

Barbeyron, T., Gerard, A., Potin, P., Henrissat, B., & Kloareg, B. (1998). The kappa-carrageenase of the marine bacterium Cytophaga drobachiensis. Structural and phylogenetic relationships within family-16 glycoside hydrolases. *Molecular Biology and Evolution*, *15*(5), 528-537.

Barbeyron, T., Thomas, F., Barbe, V., Teeling, H., Schenowitz, C., Dossat, C., Goesmann, A., Leblanc, C., Glöckner, F.O., Czjzek, M., Amann, R., Michel, G. (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*(12), 4610-4627.

Biddanda, B., & Benner, R. (1997). Carbon, nitrogen, and carbohydrate fluxes during the production of particulate and dissolved organic matter by marine phytoplankton. *Limnology and Oceanography*, *42*(3), 506-518.

Beattie, A., Hirst, E. L., & Percival, E. (1961). Studies on the metabolism of the Chrysophyceae. Comparative structural investigations on leucosin (chrysolaminarin)

separated from diatoms and laminarin from the brown algae. *Biochemical Journal*, *79*(3), 531.

Becker, S., Scheffel, A., Polz, M. F., & Hehemann, J. H. (2017). Accurate quantification of laminarin in marine organic matter with enzymes from marine microbes. *Applied and Environmental Microbiology*, *83*(9), e03389-16.

Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*, *42*(1), 1-20.

Berteau, O., & Mulloy, B. (2003). Sulfated fucans, fresh perspectives: structures, functions, and biological properties of sulfated fucans and an overview of enzymes active toward this class of polysaccharide. *Glycobiology*, *13*(6), 29R-40R.

Blanvillain, S., Meyer, D., Boulanger, A., Lautier, M., Guynet, C., Denancé, N., *et al*. (2007). Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PLoS One*, *2*(2), e224.

Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal*, *382*(3), 769-781.

Broecker, W. S. (1997). Thermohaline circulation, the Achilles heel of our climate system: Will man-made CO2 upset the current balance. *Science*, *278*(5343), 1582-1588.

Bunse, C., & Pinhassi, J. (2017). Marine bacterioplankton seasonal succession dynamics. *Trends in Microbiology*, *25*(6), 494-505.

Chafee, M., Fernàndez-Guerra, A., Buttigieg, P. L., Gerdts, G., Eren, A. M., Teeling, H., & Amann, R. I. (2017). Recurrent patterns of microdiversity in a temperate coastal marine environment. *The ISME Journal*, *12*(1), 237.

Cho, B. C., & Azam, F. (1988). Major role of bacteria in biogeochemical fluxes in the ocean's interior. *Nature*, *332*(6163), 441.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., *et al*. (2013). Carbon and other biogeochemical cycles. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 465-570). Cambridge University Press. Cambridge, United Kingdom and New York, NY, USA. pp 465-570.

Cloern, J. E. (1996). Phytoplankton bloom dynamics in coastal ecosystems: a review with some general lessons from sustained investigation of San Francisco Bay, California. *Reviews of Geophysics*, *34*(2), 127-168.

Coutinho, P. M., Stam, M., Blanc, E., & Henrissat, B. (2003). Why are there so many carbohydrate-active enzyme-related genes in plants. *Trends in Plant Science*, *8*(12), 563-565.

Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y., Cameron, E. A., Pudlo, N. A., *et al*. (2015). Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature*, *517*(7533), 165.

Davies, G., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, *3*(9), 853-859.

del Carmen Muñoz-Marín, M., Luque, I., Zubkov, M. V., Hill, P. G., Diez, J., & García-Fernández, J. M. (2013). Prochlorococcus can use the Pro1404 transporter to take up glucose at nanomolar concentrations in the Atlantic Ocean. *Proceedings of the National Academy of Sciences*, *110*(21), 8597-8602.

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

Falkowski, P., Scholes, R. J., Boyle, E. E. A., Canadell, J., Canfield, D., Elser, J., *et al*. (2000). The global carbon cycle: a test of our knowledge of earth as a system. *Science*, *290*(5490), 291-296.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237-240

Ficko-Blean, E., Préchoux, A., Thomas, F., Rochat, T., Larocque, R., Zhu, Y., *et al*. (2017). Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nature Communications*, *8*(1), 1685.

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

Field, C. B., & Raupach, M. R. (2004). The Global Carbon Cycle: Integrating Humans, Climate, and the Natural World. Field C.B., Raupach, M. R. (eds). Island Press, pp. 243-257.

Finkel, Z. V. (2014). Marine net primary production. Global Environmental Change. Freedman B. (ed). Springer Netherlands, pp. 117-124.

Fischer, F. G., & Dörfel, H. (1955). Die Polyuronsäuren der Braunalgen (Kohlenhydrate der Algen I). *Hoppe-Seyler´ s Zeitschrift für physiologische Chemie*, *302*(1-2), 186-203.

Ford, C. W., & Percival, E. (1965). 1299. Carbohydrates of *Phaeodactylum tricornutum*. Part II. A sulphated glucuronomannan. *Journal of the Chemical Society*, 7042-7046.

Geider, R. J., MacIntyre, H. L., & Kana, T. M. (1997). Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a: carbon ratio to light, nutrient-limitation and temperature. *Marine Ecology Progress Series 148*, 187-200.

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., *et al*. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science*, *309*(5738), 1242-1245.

Glenwright, A. J., Pothula, K. R., Bhamidimarri, S. P., Chorev, D. S., Baslé, A., Firbank, S. J., *et al*. (2017). Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature*, *541*(7637), 407.

Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W., & Brumer, H. (2017). Polysaccharide utilization loci: fueling microbial communities. *Journal of Bacteriology*, *199*(15), e00860-16.

Guiry, M. D. (2012). How many species of algae are there? *Journal of Phycology*, *48*(5), 1057-1063.

Gügi, B, Le Costaouec, T, Burel, C, Lerouge, P, Helbert, W, Bardor, M (2015). Diatom-specific oligosaccharide and polysaccharide structures help to unravel biosynthetic capabilities in diatoms. *Marine Drugs* 13, 5993-6018.

Handa, N. (1969). Carbohydrate metabolism in the marine diatom *Skeletonema costatum*. *Marine Biology*, *4*(3), 208-214.

Hehemann, J. H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., & Michel, G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature*, 464(7290), 908.

Hehemann, J. H., Kelly, A. G., Pudlo, N. A., Martens, E. C., & Boraston, A. B. (2012). Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proceedings of the National Academy of Sciences*, *109*(48), 19786-19791.

Hehemann, J. H., Arevalo, P., Datta, M. S., Yu, X., Corzett, C. H., Henschel, A., *et al*. (2016). Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. *Nature Communications*, 7, 12860.

Hickman, S. J., Cooper, R. E., Bellucci, L., Paci, E., & Brockwell, D. J. (2017). Gating of TonB-dependent transporters by substrate-specific forced remodelling. *Nature Communications*, *8*, 14804.

Hoagland, K. D., Rosowski, J. R., Gretz, M. R., & Roemer, S. C. (1993). Diatom extracellular polymeric substances: function, fine structure, chemistry, and physiology. *Journal of phycology*, *29*(5), 537-566.

Howarth, R. W. (1988). Nutrient limitation of net primary production in marine ecosystems. *Annual Review of Ecology and Systematics*, *19*(1), 89-110.

Ishizaka, J., Takahashi, M., & Ichimura, S. (1983). Evaluation of coastal upwelling effects on phytoplankton growth by simulated culture experiments. *Marine Biology*, *76*(3), 271-278.

Jiao, N., Herndl, G. J., Hansell, D. A., Benner, R., Kattner, G., Wilhelm, S. W., ... & Azam, F. (2010). Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nature Reviews Microbiology*, *8*(8), 593.

Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M., *et al*. (2014). Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes *Gramella forsetii* KT0803. *The ISME Journal*, *8*(7), 1492.

Kirchman, D. L., Meon, B., Ducklow, H. W., Carlson, C. A., Hansell, D. A., & Steward, G. F. (2001). Glucose fluxes and concentrations of dissolved combined neutral sugars (polysaccharides) in the Ross Sea and Polar Front Zone, Antarctica. *Deep Sea Research Part II: Topical Studies in Oceanography*, *48*(19-20), 4179-4197.

Kloareg, B., Mabeau, S. (1987). Isolation and analysis of the cell walls of brown algae: Fucus spiralis, F. ceranoides, F. vesiculosus, F. serratus, Bifurcaria bifurcata and Laminaria digitata. *Journal of Experimental Botany*, *38*(9), 1573-1580.

Kloareg, B., Quatrano, R.S. (1988). Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanography and Marine Biology: An Annual Review* 26, 259-315.

Kremer, B. P. (1980). Taxonomic implications of algal photoassimilate patterns. *British Phycological Journal*, *15*(4), 399-409.

Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Parker, M. S., *et al*. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PloS One*, *3*(1), e1426.

Labourel, A., Jam, M., Jeudy, A., Hehemann, J. H., Czjzek, M., & Michel, G. (2014). The β-glucanase ZgLamA from *Zobellia galactanivorans* evolved a bent active site adapted for efficient degradation of algal laminarin. *Journal of Biological Chemistry*, *289*(4), 2027-2042.

Lahaye, M., & Robic, A. (2007). Structure and functional properties of ulvan, a polysaccharide from green seaweeds. *Biomacromolecules*, *8*(6), 1765-1774.

Laine, R. A. (1994). Invited Commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields $1.05x10^{12}$ structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology*, *4*(6), 759-767.

Larsbrink, J., Rogers, T. E., Hemsworth, G. R., McKee, L. S., Tauzin, A. S., Spadiut, O., *et al*. (2014). A discrete genetic locus confers xyloglucan metabolism in select human gut *Bacteroidetes*. *Nature*, *506*(7489), 498.

Le Costaouëc, T., Unamunzaga, C., Mantecon, L., & Helbert, W. (2017). New structural insights into the cell-wall polysaccharide of the diatom Phaeodactylum tricornutum. *Algal Research*, *26*, 172-179.

Legendre, L., Rivkin, R. B., Weinbauer, M. G., Guidi, L., & Uitz, J. (2015). The microbial carbon pump concept: Potential biogeochemical significance in the globally changing ocean. *Progress in Oceanography*, *134*, 432-450.

Lignell, R., Heiskanen, A. S., Kuosa, H., Gundersen, K., Kuuppo-Leinikki, P., Pajuniemi, R., & Uitto, A. (1993). Fate of a phytoplankton spring bloom: sedimentation and carbon flow in the planktonic food web in the northern Baltic. *Marine Ecology Progress Series*, 239-252.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2013). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*(D1), D490-D495.

Malhi, Y. (2002). Carbon in the atmosphere and terrestrial biosphere in the 21st century. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *360*(1801), 2925-2945.

Mann, D. G. (1999). The species concept in diatoms. *Phycologia*, *38*(6), 437-495.

Marañón, E., Holligan, P. M., Varela, M., Mouriño, B., & Bale, A. J. (2000). Basin-scale variability of phytoplankton biomass, production and growth in the Atlantic Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, *47*(5), 825-857.

Martens, E. C., Koropatkin, N. M., Smith, T. J., & Gordon, J. I. (2009). Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *Journal of Biological Chemistry*, *284*(37), 24673-24677.

Martens, E. C., Lowe, E. C., Chiang, H., Pudlo, N. A., Wu, M., McNulty, N. P., *et al*. (2011). Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biology*, *9*(12), e1001221.

Martens, E. C., Kelly, A. G., Tauzin, A. S., & Brumer, H. (2014). The devil lies in the details: how variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. *Journal of Molecular Biology*, *426*(23), 3851-3865.

McCarren, J., Becker, J. W., Repeta, D. J., Shi, Y., Young, C. R., Malmstrom, R. R., *et al*. (2010). Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proceedings of the National Academy of Sciences*, *107*(38), 16420-16427.

Michel, G., Nyval-Collen, P., Barbeyron, T., Czjzek, M., & Helbert, W. (2006). Bioconversion of red seaweed galactans: a focus on bacterial agarases and carrageenases. *Applied Microbiology and Biotechnology*, *71*(1), 23-33.

Michel, G., Tonon, T., Scornet, D., Cock, J. M., & Kloareg, B. (2010). The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes. *New Phytologist*, *188*(1), 82-97.

Myklestad, S., & Granum, E. (2009). Biology of (1, 3)-β-glucans and related glucans in protozoans and chromistans. *Chemistry, Biochemistry, and Biology of (1-3)-β-Glucans and Related Polysaccharides.* Bacic A., Fincher G. (eds). Academic Press, Boston, pp. 353-385.

Moran, M. A., Kujawinski, E. B., Stubbins, A., Fatland, R., Aluwihare, L. I., Buchan, A., *et al*. (2016). Deciphering ocean carbon in a changing world. *Proceedings of the National Academy of Sciences*, *113*(12), 3143-3151.

Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A., & Quéguiner, B. (1995). Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles*, *9*(3), 359-372.

Niklas, K. J. (2004). The cell walls that bind the tree of life. *AIBS Bulletin*, *54*(9), 831-841.

Norton, T. A., Melkonian, M., & Andersen, R. A. (1996). Algal biodiversity. *Phycologia*, 35(4), 308-326.

Piontek, J., Händel, N., De Bodt, C., Harlay, J., Chou, L., & Engel, A. (2011). The utilization of polysaccharides by heterotrophic bacterioplankton in the Bay of Biscay (North Atlantic Ocean). *Journal of Plankton Research*, *33*(11), 1719-1735.

Piontek, J., Sperling, M., Nöthig, E. M., & Engel, A. (2014). Regulation of bacterioplankton activity in Fram Strait (Arctic Ocean) during early summer: the role of organic matter supply and temperature. *Journal of Marine Systems*, *132*, 83-94.

Polimene, L., Sailley, S., Clark, D., Mitra, A., & Allen, J. I. (2016). Biological or microbial carbon pump? The role of phytoplankton stoichiometry in ocean carbon sequestration. *Journal of Plankton Research*, *39*(2), 180-186.

Pomeroy, L. R. (1974). The ocean's food web, a changing paradigm. *Bioscience*, *24*(9), 499-504.

Popper, Z. A., Michel, G., Hervé, C., Domozych, D. S., Willats, W. G., Tuohy, M. G., ... & Stengel, D. B. (2011). Evolution and diversity of plant cell walls: from algae to flowering plants. *Annual Review of Plant Biology*, *62*, 567-590.

Raimbault, P., Garcia, N., & Cerutti, F. (2008). Distribution of inorganic and organic nutrients in the South Pacific Ocean− evidence for long-term accumulation of organic matter in nitrogen-depleted waters. *Biogeosciences*, *5*(2), 281-298.

Raven, J. A., & Falkowski, P. G. (1999). Oceanic sinks for atmospheric CO2. *Plant, Cell & Environment*, *22*(6), 741-755.

Read, S. M., Currie, G., & Bacic, A. (1996). Analysis of the structural heterogeneity of laminarin by electrospray-ionisation-mass spectrometry. *Carbohydrate Research*, *281*(2), 187-201.

Rees, D. A., & Welsh, E. J. (1977). Secondary and tertiary structure of polysaccharides in solutions and gels. *Angewandte Chemie International Edition*, *16*(4), 214-224.

Reintjes, G., Arnosti, C., Fuchs, B. M., & Amann, R. (2017). An alternative polysaccharide uptake mechanism of marine bacteria. *The ISME Journal*, *11*(7), 1640.

Rinta-Kanto, J. M., Sun, S., Sharma, S., Kiene, R. P., & Moran, M. A. (2012). Bacterial community transcription patterns during a marine phytoplankton bloom. *Environmental Microbiology*, *14*(1), 228-239.

Rooney-Varga, J. N., Giewat, M. W., Savin, M. C., Sood, S., LeGresley, M., & Martin, J. L. (2005). Links between phytoplankton and bacterial community dynamics in a coastal marine environment. *Microbial Ecology*, *49*(1), 163-175.

Sarmiento, J. L., & Le Quere, C. (1996). Oceanic carbon dioxide uptake in a model of century-scale global warming. *Science*, *274*(5291), 1346-1350.

Sarmento, H., & Gasol, J. M. (2012). Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environmental Microbiology*, *14*(9), 2348-2360.

Shinada, A., Shiga, N., & Ban, S. (1999). Origin of Thalassiosira diatoms that cause the spring phytoplankton bloom in Funka Bay, southwestern Hokkaido, Japan. *Plankton Biology and Ecology*, 46(2), 89-93.

Smith, S. V. (1984). Phosphorus versus nitrogen limitation in the marine environment. *Limnology and Oceanography*, *29*(6), 1149-1160.

Sison-Mangus, M. P., Jiang, S., Kudela, R. M., & Mehic, S. (2016). Phytoplankton-associated bacterial community composition and succession during toxic diatom bloom and non-bloom events. *Frontiers in Microbiology*, *7*, 1433.

Suttle, C. A. (1994). The significance of viruses to mortality in aquatic microbial communities. *Microbial Ecology*, *28*(2), 237-243.

Teeling H., Fuchs B. M., Becher D., Klockow C., Gardebrecht A., Bennke C. M., Kassabgy M., Huang S., Mann A. J., Waldmann J., Weber M., Klindworth A., Otto A., Lange J., Bernhardt J., Reinsch C., Hecker M., Peplies J., Bockelmann F. D., Callies U., Gerdts G., Wichels A., Wiltshire K. H., Glöckner F. O., Schweder T., Amann R. (2012). Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science*, *336*(6081), 608-611.

Teeling, H., Fuchs, B. M., Bennke, C. M., Krueger, K., Chafee, M., Kappelmann, L., Reintjes, G., Waldmann, J., Quast, C., Glöckner, F. O., Lucas, J., Wichels, A., Gerdts, G., Wiltshire, K. H. Amann, R. (2016). Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *Elife*, *5*, e11888.

Teira, E., José Pazó, M., Serret, P., & Fernández, E. (2001). Dissolved organic carbon production by microbial populations in the Atlantic Ocean. *Limnology and Oceanography*, *46*(6), 1370-1377.

Terrapon, N., Lombard, V., Gilbert, H. J., & Henrissat, B. (2014). Automatic prediction of polysaccharide utilization loci in *Bacteroidetes* species. *Bioinformatics*, *31*(5), 647-655.

Thomas, F., Barbeyron, T., Tonon, T., Génicot, S., Czjzek, M., & Michel, G. (2012). Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine Flavobacteriia to their independent transfers to marine Proteobacteria and human gut Bacteroides. *Environmental Microbiology*, *14*(9), 2379-2394.

Thornton, D. C. (2014). Dissolved organic matter (DOM) release by phytoplankton in the contemporary and future ocean. *European Journal of Phycology*, *49*(1), 20-46.

Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress in Oceanography*, *130*, 205-248.

Van Dongen-Vogels, V., Seymour, J. R., Middleton, J. F., Mitchell, J. G., & Seuront, L. (2012). Shifts in picophytoplankton community structure influenced by changing upwelling conditions. *Estuarine, Coastal and Shelf Science*, *109*, 81-90.

Vårum, K. M., Østgaard, K., & Grimsrud, K. (1986). Diurnal rhythms in carbohydrate metabolism of the marine diatom Skeletonema costatum (Grev.) Cleve. *Journal of Experimental Marine Biology and Ecology*, *102*(2-3), 249-256.

Volk, T., & Hoffert, M. I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO2 changes. *The Carbon Cycle and Atmospheric CO: Natural Variations Archean to Present*, American Geophysical Union; Geophysical Monograph 32, 99-110.

Wang, L., Wang, X., Wu, H., & Liu, R. (2014). Overview on biological activities and molecular characteristics of sulfated polysaccharides from marine green algae in recent years. *Marine Drugs*, *12*(9), 4984-5020.

Wemheuer, B., Wemheuer, F., Hollensteiner, J., Meyer, F. D., Voget, S., & Daniel, R. (2015). The green impact: bacterioplankton response toward a phytoplankton spring bloom in the southern North Sea assessed by comparative metagenomic and metatranscriptomic approaches. *Frontiers in Microbiology*, *6*, 805.

Westberry, T., Behrenfeld, M. J., Siegel, D. A., & Boss, E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*, *22*(2).

Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences*, *95*(12), 6578-6583.

Wollast, R. (2002). Continental margins—review of geochemical settings. In *Ocean Margin Systems.* Wefer, G., Billett, D., Hebbeln, D., Jørgensen, B.B., Schlüter, M., and van Weering, T.C.E. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 15-31.

Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., & Keeling, P. J. (2015). Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science*, *347*(6223), 1257594.

Xing, P., Hahnke, R. L., Unfried, F., Markert, S., Huang, S., Barbeyron, T. Harder J., Becher D., Schweder T., Glöckner F.O., Amann R., Teeling H. (2015). Niches of two polysaccharide-degrading Polaribacter isolates from the North Sea during a spring diatom bloom. *The ISME Journal*, *9*(6), 1410.

Zinser, E. R., Johnson, Z. I., Coe, A., Karaca, E., Veneziano, D., & Chisholm, S. W. (2007). Influence of light and temperature on Prochlorococcus ecotype distributions in the Atlantic Ocean. *Limnology and Oceanography*, *52*(5), 2205-2220.

Zubkov, M. V., Fuchs, B. M., Tarran, G. A., Burkill, P. H., & Amann, R. (2003). High rate of uptake of organic nitrogen compounds by Prochlorococcus cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Applied and Environmental Microbiology*, *69*(2), 1299-1304.

Appendix I: Adaptive mechanisms that provide competitive advantages to marine Bacteroidetes during microalgal blooms

Submitted to *ISME Journal*

Frank Unfried[1,2,3], Stefan Becker[2,4], Craig S. Robb[2,4], Jan-Hendrik Hehemann[2,4], Stephanie Markert[1,3], Stefan E. Heiden[1], Tjorven Hinzke[1,3], Dörte Becher[3,5], Greta Reintjes[2], Karen Krüger[2], Burak Avcı[2], Lennart Kappelmann[2], Richard L. Hahnke[2,6], Tanja Fischer[2], Jens Harder[2], Hanno Teeling[2], Bernhard Fuchs[2], Tristan Barbeyron[7], Rudolf I. Amann[2], Thomas Schweder[1,3,#]

[1] Pharmaceutical Biotechnology, University Greifswald, Germany
[2] Max Planck Institute for Marine Microbiology, Bremen, Germany
[3] Institute of Marine Biotechnology, Greifswald, Germany
[4] MARUM, Center for Marine Environmental Sciences at the University of Bremen, Bremen, Germany
[5] Institute for Microbiology, University Greifswald, Germany
[6] DSMZ, Braunschweig, Germany
[7] National Center of Scientific Research/Pierre and Marie Curie University Paris 6, UMR 7139 Marine Plants and Biomolecules, Station Biologique de Roscoff, Roscoff, Bretagne, France

Contribution to the manuscript:

I participated in the assembly and annotation of the genomes of the two *Formosa* strains Hel3_A1_48 and Hel1_33_131 (in the manuscript referred to as (*Formosa* A and *Formosa* B), as well as metagenomic read mappings at different nucleotide identities to show how representative these two strains were for the recurrent Formosa clade during North Sea spring microalgae blooms.

Abstract

Polysaccharide degradation by heterotrophic microbes is a key process within Earth's carbon cycle. Here, we use environmental proteomics and metagenomics in combination with cultivation experiments and biochemical characterizations to investigate the molecular details of in situ polysaccharide degradation mechanisms during microalgal blooms. For this, we use laminarin as a model polysaccharide. Laminarin is an ubiquitous marine storage polymer of marine microalgae and is particularly abundant during phytoplankton blooms. In this study, we show that highly specialized bacterial strains of the Bacteroidetes phylum repeatedly reached high abundances during North Sea algal blooms and dominated laminarin turnover. These genomically streamlined bacteria of the genus Formosa have an expanded set of laminarin hydrolases and transporters that belonged to the most abundant proteins in the environmental samples. In vitro experiments with cultured isolates allowed us to determine the functions of in situ expressed key enzymes and to confirm their role in laminarin utilization. It is shown that laminarin consumption of Formosa spp. is paralleled by enhanced uptake of diatom-derived peptides. This study reveals that genome reduction, enzyme fusions, transporters and enzyme expansion, as well as a tight coupling of carbon- and nitrogen-metabolism provide the tools which make Formosa spp. so competitive during microalgal blooms.

Appendix II: Alpha- and beta-mannan utilization by marine Bacteroidetes

In preparation

Jing Chen[1,2,3,4], Craig Robb[1,5], Frank Unfried[6,7], Lennart Kappelmann[1], Stephanie Markert[6,7], Tao Song[5], Jens Harder[1], Dörte Becher[8], Ping Xie[2], Rudolf Amann[1], Jan-Hendrik Hehemann[1,5], Thomas Schweder[6,7*], Hanno Teeling[1*]

[1] Max Planck Institute for Marine Microbiology, Celsiusstraße 1, D-28359 Bremen, GER
[2] Donghu Experimental Station of Lake Ecosystems, State Key Laboratory of Freshwater Ecology and Biotechnology of China, Institute of Hydrobiology, Chinese Academy of Sciences, Donghu South Road No.7, 430072 Wuhan, China
[3] University of Chinese Academy of Sciences, Yuquan Road No.19 (A), 100049 Beijing, China
[4] College of Ocean, Hebei Agricultural University, Qinhuangdao, 066003, PR China
[5] MARUM, Center for Marine Environmental Sciences, University of Bremen, Leobener Straße, D-28359 Bremen, Germany
[6] Pharmaceutical Biotechnology, Institute of Pharmacy, Ernst-Moritz-Arndt-University, Felix-Hausdorff-Straße 3, D-17487 Greifswald, Germany
[7] Institute of Marine Biotechnology, Walther-Rathenau-Straße 49a, D-17489 Greifswald, Germany
[8] Institute of Microbiology, Ernst-Moritz-Arndt-University, Friedrich-Ludwig-Jahn-Straße 15, D-17487 Greifswald, Germany

Contribution to the manuscript:

I annotated the genomes and predicted the mannan PULs of the isolates investigated in this study. Based on these results, I suggested the isolates for mannan growth experiments and proteomic analyses.

Abstract

Marine microscopic algae carry out about half of the global carbon dioxide fixation into organic matter. On lysis they provide organic substrates for marine microbes such as members of the *Bacteroidetes* that degrade algal polysaccharides using carbohydrate-active enzymes (CAZymes). In *Bacteroidetes* genomes CAZymes are mostly grouped in distinct regions termed polysaccharide utilization loci (PULs). While the involvement of PULs in algal polysaccharide degradation is well established, their specific substrates are for the most part still unknown. We investigated marine *Bacteroidetes* isolated from the southern North Sea that harbor putative mannan-specific PULs. These PULs are related to genomic clusters of human gut *Bacteroides* strains, which digest α- and β-mannans from yeasts and plants, respectively. Using proteomics and defined growth experiments with polysaccharides as sole carbon sources we could show that these *Bacteroidetes'* PULs indeed mediate α- and β-mannan degradation. Our data hence suggest that biochemical principles established for gut or terrestrial microbes also apply to marine bacteria, even though their PULs are evolutionarily distant from each other. Moreover, our data indicate that algal mannans play an as yet unknown important role in the marine carbon cycle.

Acknowledgements

I would like to thank:

- My direct supervisor **Dr. Hanno Teeling** for the opportunity to work in his group. You gave me valuable feedback, support, ideas and guidance. Our talks were always interesting inside and outside of science. Not often have I met a person with such an encyclopedic knowledge on so many topics. I wholeheartedly wish you all the best for your future health.

- **Prof. Dr. Rudolf Amann** for giving me the chance of doing research in his Molecular Ecology (MolEcol) department within the Max Planck Society. I always admired your analytical, ecological way of thinking, but also your friendly, relaxed temper and ability to find joy outside of science, for example during our group retreats. Thank you for jumping in as my direct supervisor despite your busy schedule when I needed it.

- **Dr. Jan-Hendrik Hehemann** for offering his biochemical expertise and giving me critical feedback on many occasions. I still regret we didn't manage to go fishing together.

- **Prof. Dr. Jens Harder** for being a great and interesting scientist and person since the MarMic Master program and for providing the *Flavobacteriia* isolates through your microbiology group.

- **Prof. Dr. Rita Groß-Hardt** for being a member of my examination board on short notice.

- All people involved in the **COGITO project** and gathering the datasets I worked on. This includes Jens Harder and Richard Harder for the isolation of the North Sea *Flavobacteriia*, Dr. Bernhard Fuchs, who also frequently gave me feedback and without whom the MolEcol department is unimaginable to me, Prof. Dr. Thomas Schweder, Dr. Antje Wichels, Dr. Gunner Gerdts and Prof. Dr. Karen Wiltshire.

- All my coworkers in the **MolEcol department** for the amazing work environment. Within the **Teeling group**, I would like to thank the old guard (”The IT Crowd”) of Sixing Huang, Johannes Werner and Alexander Mann, as well as the “next generation” of Meghan Chafee, Karen Krüger, Burak Avci and Ben Francis.

- My “**sunshine class**” of the MarMic Master program. I will never forget these intense yet fun 18 months.

- My **family** for the love and support they always gave me and my **private friends** for taking my mind off science during the most intense times.

- **Betty**, por hacer de los últimos 9 meses estresantes de esta tesis doctoral los momentos más asombrosos de mi vida. Gracias a ti, siempre supe que después de un intenso día de escritura, me esperaba una velada agradable. Tu sonrisa puede borrar todas las preocupaciones en un instante.

Erklärung der selbstständigen Erarbeitung

Hiermit erkläre ich, dass ich die Doktorarbeit mit dem Titel

Genomic Analyses of Polysaccharide Utilization in Marine *Flavobacteriia*

selbstständig verfasst und geschrieben habe und außer den angegebenen Quellen keine weiteren Hilfsmittel verwendet habe.

Ebenfalls erkläre ich hiermit, dass es sich bei den von mir abgegebenen Arbeiten um drei identische Exemplare handelt.

_____

Unterschrift, Datum