Niches of abundant heterotrophic bacteria during North Sea spring algal blooms

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

-Dr. rer. nat.-

dem Fachbereich Biologie/Chemie

der Universität Bremen

vorgelegt von

Burak Avcı

Bremen, September 2018

Die vorliegende Doktorarbeit wurde in der Zeit von September 2014 bis September 2018 im Rahmen des Programms „International Max Planck Research School of Marine Microbiology, MarMic“ in der Abteilung Molekulare Ökologie am Max-Planck-Institut für marine Mikrobiologe in Bremen angefertigt.

Gutachter: Prof. Dr. Rudolf Amann

Gutachter: Prof. Dr. Thorsten Brinkhoff

Tag des Promotionskolloquiums: 24 September 2018

“Simplicity is the ultimate sophistication”

Leonardo da Vinci

Summary

The remineralization of algal biomass by heterotrophic bacteria during spring phytoplankton blooms is a globally important process in carbon cycling. Yet, the ecophysiology of environmentally relevant bacteria occurring during these events is largely unknown. In this thesis, I investigated the niches of the gammaproteobacterial genus *Reinekea* and the flavobacterial genera *Polaribacter* and *Formosa*, which are recurrently abundant during the 2009-2012 North Sea spring algal blooms off Helgoland. Firstly, I studied '*Reinekea forsetii*', the representative of North Sea *Reinekea* clade, using genomic and physiological analyses. The results suggested a versatile opportunistic lifestyle comprising (i) utilization of relevant algal polysaccharides, (ii) potential toxin production, and (iii) strategies to deal with phosphorus limitation. Secondly, temporal dynamics and large niche space of diverse North Sea *Polaribacter* spp. was investigated using fluorescence *in situ* hybridization with novel oligonucleotide probes, binning of time-series metagenomes and re-analysis of *in situ* expression data. These analyses showed the succession of four major *Polaribacter* clades with varying degradation capacity for high molecular weight compounds and suggested a polysaccharide-driven niche partitioning between these closely-related bacteria. Finally, the polysaccharide niche of North Sea *Formosa* clade was examined using meta'omic approaches and complementary physiological and biochemical experiments. An efficient laminarin uptake and degradation mechanism, which is coupled to peptide utilization, was revealed. This thesis demonstrated how heterotrophic bacteria employ different ecological strategies to utilize diverse high molecular weight compounds released during spring algal blooms and increased our knowledge on microbially mediated carbon turnover in the surface ocean.

Zusammenfassung

Die Remineralisierung von Algenbiomasse durch heterotrophe Bakterien während der Algenblüten im Frühjahr ist von zentraler Bedeutung für den globalen Kohlenstoffkreislauf. Die Ökophysiologie der für diese Prozesse relevanten Bakterien ist bisher jedoch weitgehend unbekannt. Zu dieser Gruppe von Bakterien gehören Gammaproteobakterien der Gattung *Reinekea* und Flavobakterien der Gattungen *Polaribacter* und *Formosa*, die wiederkehrend während der Frühlingsalgenblüten in der Nordsee vor Helgoland von 2009 bis 2012 auftraten. Ziel dieser Arbeit war die Charakterisierung der Nische dieser Gattungen. Im ersten Teil meiner Arbeit untersuchte ich mittels genomischer und physiologischer Analysen '*Reinekea forsetii*', einen Repräsentanten der Nordsee vorkommenden Bakteriengruppe *Reinekea*. Meine Ergebnisse zeigten einen vielseitig opportunistischen Lebensstil, welcher (i) die Verwendung von Algen-Polysacchariden, (ii) potentielle Toxinproduktion, und (iii) Strategien zum Umgehen von Phosphorlimitierungen umfasst. Im zweiten Teil meiner Arbeit untersuchte ich die Dynamik und die Nischenbreite verschiedener Arten der Gattung *Polaribacter*. Dafür führte ich Fluoreszenz *in situ* Hybridisierungen mit neu entwickelten Oligonukleotidsonden durch, erstellte „Bins" aus Metagenomzeitserien und re-analysierte bereits vorhandene Expressionsdaten. Die Ergebnisse zeigten eine Aufeinanderfolge von vier *Polaribacter* Hauptgruppen. Diese wiesen ein divergierendes Potential zum Abbau von Polymeren mit einem hohen Molekulargewicht auf und wiesen auf eine Nischenpartitionierung bezüglich des Polysaccharidabbaus hin. Zuletzt war ich an einer Studie zur Polysaccharid Nische von zwei Nordsee *Formosa* Stämmen beteiligt. Die Ergebnisse zeigten eine effiziente Aufnahme und Verwertung von Laminarin, wobei die Induktion den Peptidmetabolismus umfasste. Diese Arbeit zeigte wie heterotrophe Bakterien verschiedene ökologische Strategien anwenden, um Polymere mit hohem Molekulargewicht für sich zu nutzen, welche während der Algenblüte im Frühjahr freigesetzt werden. Dadurch ergänzt diese Arbeit unser bisheriges Wissen über den mikrobiellen Kohlenstoffumsatz innerhalb der Oberflächenschichten des Ozeans.

Table of contents

Abbreviations

ANI: average nucleotide identity

BSA: bovine serum albumin

CARD-FISH: catalyzed reporter deposition fluorescence *in situ* hybridization

CAZymes: carbohydrate active enzyme

CBM: carbohydrate binding module

CE: carbohydrate esterase

DAPI: 4',6-diamidino-2-phenylindole

DGGE: denaturant gel gradient electrophoresis

DMSP: dimethylsulfoniopropionate

DNA: deoxyribonucleic acid

dNTP: deoxyribose nucleotide triphosphate

DOM: dissolved organic matter

DOC: dissolved organic carbon

FACS: fluorescence-activated cell sorting

FISH: fluorescence *in situ* hybridization

GH: glycoside hydrolase

GT: glycoside transferase

HCR-FISH: hybridization-chain reaction fluorescence *in situ* hybridization

PCR: polymerase chain reaction

PL: polysaccharide lyase

PUL: polysaccharide utilization locus

POM: particulate organic matter

rRNA: ribosomal ribonucleic acid

Sus: starch utilization system

MAG: metagenome assembled genome

MDA: multiple displacement amplification

MED: minimum entropy decomposition

TBR: TonB-dependent receptor

Chapter 1

Introduction

"If I have seen further, it is by standing upon the shoulders of giants"

Isaac Newton

1) Introduction

1.1) Renaissance of microbiology

Sighting microscopic, single-celled organisms for the first time, Robert Hooke and Antonie van Leeuwenhoek pioneered the field of microbiology in the 17th century (1). Two centuries elapsed between these observations and the isolation of microorganisms in pure culture (Fig. 1.1) (2). Since then, traditional culture-based techniques have been used for the discovery and characterization of microbial strains and made significant contributions towards understanding the physiology of important microorganisms such as *Escherichia coli* (3) and *Vibrio cholerae* (4). However, many microorganisms are not able to grow in artificial media as used in contemporary cultivation studies. This phenomenon is demonstrated by the “great plate count anomaly”, which refers to the discrepancy between the total counts of microorganisms detected under the microscope and the fraction that can be grown on cultivation plates under laboratory conditions (5). Thus, only a small proportion of microorganisms are accessible via cultivation-dependent methods, and the “uncultivated majority” remains to be characterized. The development of molecular approaches in the 20th century eliminated the need for cultivation to explore the diversity and functioning of individual microbes and paved the way towards the study of entire microbial communities.

Microbial ecology has led the application of cultivation-independent approaches in the field of microbiology (Fig.1.1). The comparative sequence analysis of the 16S rRNA as a molecular clock, first introduced by Carl Woese (6), was developed into a cultivation-independent method for the investigation of the diversity and composition of microbial communities (7) (8). Fluorescence *in situ* hybridization of 16S rRNA has been implemented as a phylogenetic stain for the identification of single microbial cells in environmental samples through microscopy (9) (10). Subsequently, shot-gun metagenomics (11) in conjunction with high throughput sequencing has enabled access to the genetic material of most or potentially even all organisms present in a given complex community (12). This approach was a groundbreaking achievement for the comprehensive investigation of taxonomic composition and functional potential of microbial communities (13).

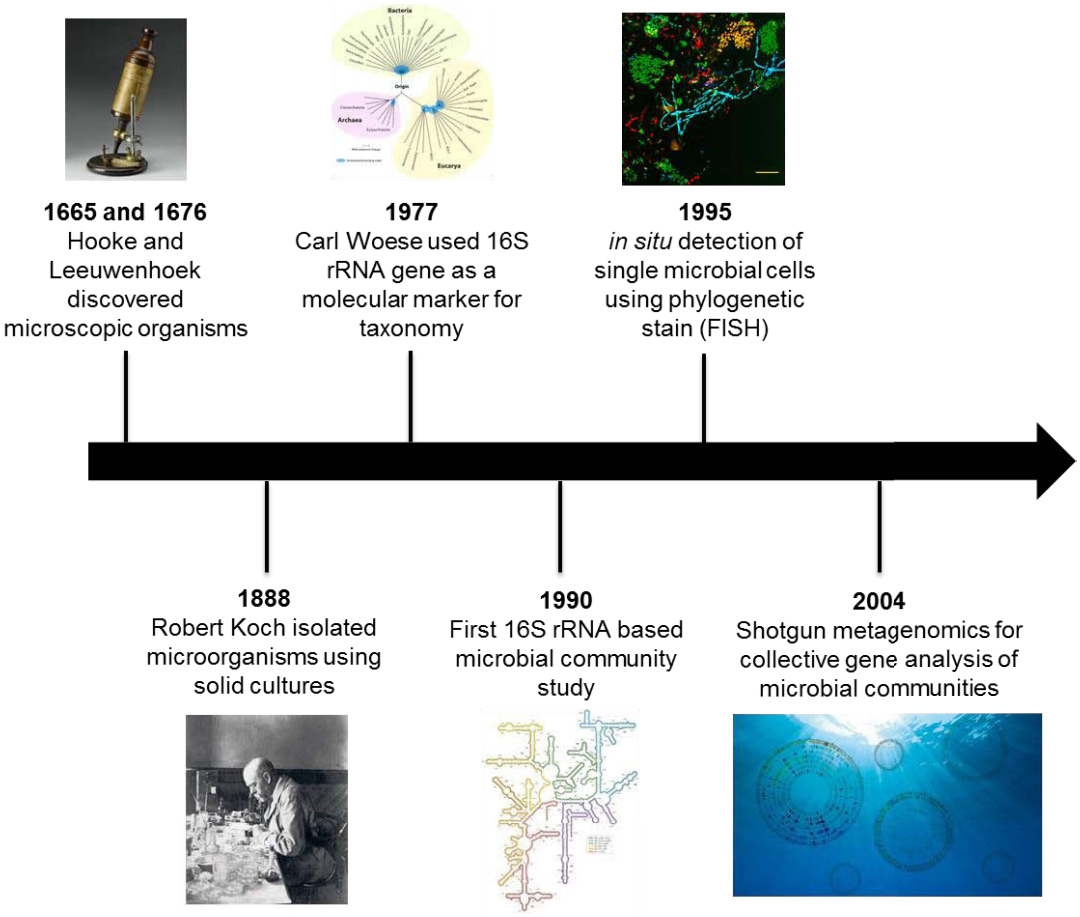


Fig. 1.1. Timeline of microbiological techniques enabling some major advances in the study of microbial diversity.

A multidisciplinary approach with a theoretical background is essential to answer the central questions of microbial ecology (Fig. 1.2). High-throughput sequencing methods have yielded tremendous amounts of data (14). Holistic data integration approaches are necessary to transform this data deluge into biological knowledge. For example, shotgun metagenome data are coupled to metatranscriptomics and metaproteomics to investigate *in situ* gene expression, as well as to metabolomics to elucidate the chemical fingerprints of specific cellular processes (15). Such multi-omics approaches have the potential to generate novel hypotheses, which can then be tested in physiological experiments (Fig. 1.2). Therefore, traditional pure culture studies are particularly important in the “omics age” (16). Furthermore, a strong theoretical background builds a significant link between the environment and microbial species composition (17). Besides being a method-driven discipline, microbial ecology also tackles the long-standing ecological questions with respect to community structure and function at the microscale. However,

interpretation of experimental data in the light of ecological theories is claimed to be not yet well established (18). This connection would allow a holistic understanding of the ecological and evolutionary processes structuring microbial communities with an ultimate aim of creating a common language to bridge ecology of microscopic and macroscopic organisms (Fig. 1.2).

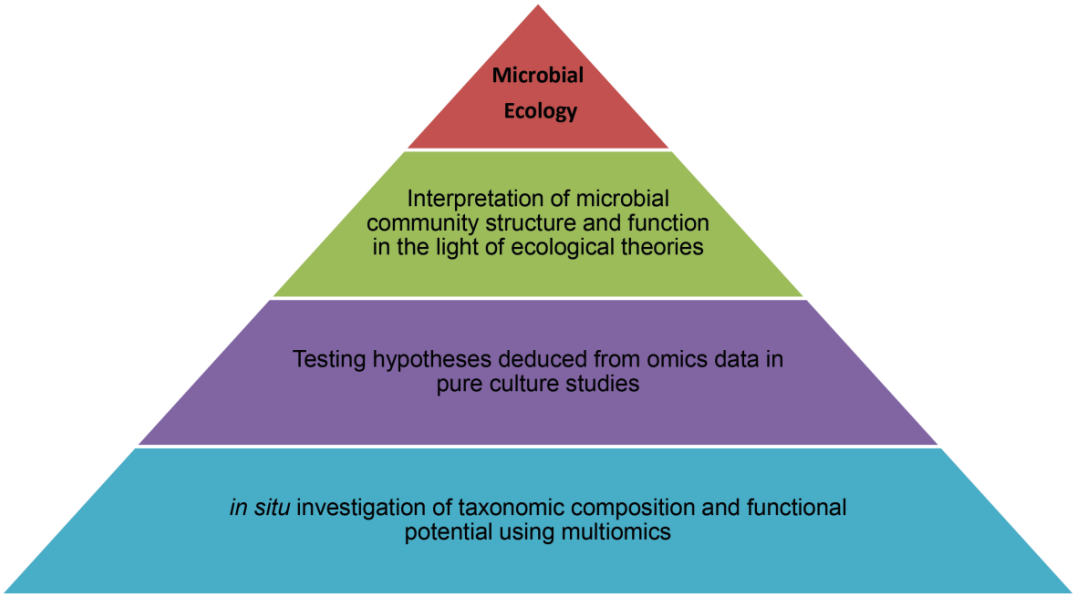


Fig. 1.2. Hierarchy of multiomics, pure culture studies, and ecological theories for a holistic understanding of microbial community structure and function.

1.2) The niche: history, concept, and applications

"What are the principles underlying the assembly and structure of complex microbial communities" is a fundamental question in microbial ecology. To address this question, neutral and niche theories have emerged. The population geneticist Motoo Kimura originally proposed the neutral theory in the context of molecular evolution. He hypothesized that most observed genetic variation enters the next generation independent of its allelic type. Therefore such variation is adaptively neutral, and the majority of evolutionary changes at the molecular level are caused by random fixation (19). Stephen P. Hubbell applied this theory to forest ecology without invoking Darwinian adaptive divergence and assumed that each tree is equally likely to reproduce (20). In his model, the tree's species does not affect its prospects of death or reproduction (21). The neutral theory, therefore, claims that each species is ecologically equivalent and stochastic processes regulate the species

composition. However, this theory cannot explain adaptive divergence and ecosystem function, which ecologists must understand (21)

In contrast, the niche theory states that environmental conditions shape the community structure through habitat filtering and resource partitioning (Fig. 1.3). The kernel of the niche concept is found in the writings of early naturalists such as Charles Robert Darwin (22). Demonstrating the effect of environment on species morphology for the first time, Darwin observed phenotypic adaptation of finch species with different morphological traits to utilize the variety of resources available in Galapagos Island (23). Sixty-five years later, ecologist Joseph Grinnell used the term “niche” first time to explain the ecological position of an animal (24), while Charles Sutherland Elton further explained this term as an animal’s place in the abiotic environment and its relations to food and enemies (25). Subsequently, George Evelyn Hutchinson developed a conceptual framework for the niche theory with the terms of “fundamental” and “realized” niches (26). According to Hutchinson, abiotic conditions dictate a fundamental niche for an organism and select a group of organisms sharing similar traits and competing for the same resources in a given environment (Fig. 1.3). However, if two species have overlapping niches, one could competitively exclude the other (27). This tension in biotic conditions is eliminated by resource (or niche) partitioning in a habitat and creates the realized niche of an organism (Fig. 1.3). Therefore, similar organisms occupy different ecological niches and co-occur in the same environment.

In light of the niche theory, considerable efforts have been made to explain microbial community assembly and structure. Lourens G.M. Baas Becking and G. Evelyn Hutchinson did noteworthy observations to highlight the importance of habitat filtering. In 1934, Baas Becking published the article with the well-known quotation “Everything is everywhere, but, the environment selects” (28). Implementing the niche-based theories into microbial ecology for the first time, this study emphasized the biogeographical pattern of microbial diversity (29). Afterwards, Hutchinson described a problem termed “The Paradox of the Plankton”, asking how it is possible for different phytoplankton species to coexist in a relatively homogeneous environment while competing for the

same sorts of resources (30). This question has challenged microbial ecologists for decades and shifted their focus towards niche partitioning, which could address this problem.

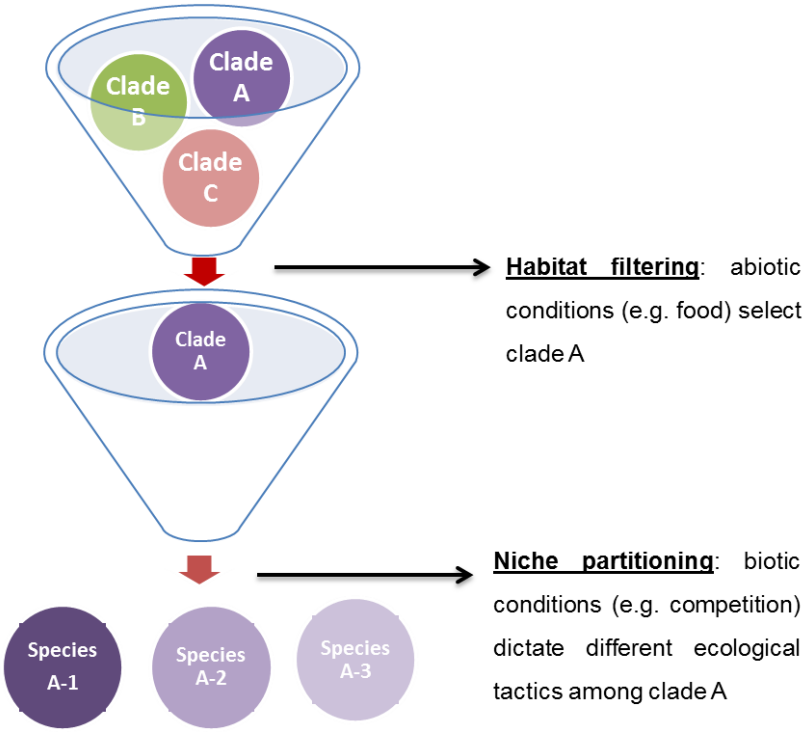


Fig. 1.3. Simplified scheme of habitat filtering and niche partitioning processes shaping community composition.

Cultivation-independent techniques have paved the way for studies investigating the effect of niche partitioning on microbial community structure. For example, Schramm and colleagues demonstrated the activity and spatial distribution of *Nitrosospira* and *Nitrospira* spp. in bacterial aggregates from a nitrifying fluidized- bed reactor using micro-sensor measurements and rRNA-based techniques (31). This study explored the spatial heterogeneity of these bacteria in a varying chemical gradient throughout the aggregates and suggested that different ecological niches could form in a microbial community from an artificial environment with low diversity. Furthermore, within a complex microbial community in a natural environment, Hunt and colleagues showed sympatric differentiation among closely related members of bacterioplankton (32). This study revealed niche partitioning between *Vibrio* spp. by combining traditional cultivation techniques and phylogenetic analyses. These investigations overall could not define the ecological niches of environmental microorganisms in great detail since methodological limitations did not allow

in-depth functional analyses. Advancements of high-throughput sequencing technologies have enabled collective analysis of the genetic repertoire, and allow us to explore which metabolic strategies drive ecological processes in complex microbial communities in the wild (33).

1.3) An ocean of microbes

Microorganisms are ubiquitous in the marine environment. Accounting for ~ 20% of the global biomass in the planet Earth (34), microbes are abundant and diverse inhabitants of the ocean ecosystem. The total number of bacteria in the open ocean exceeds 10^{29} cells, with an average of 10^{6} cells per milliliter of surface water (35). A global microbiome survey suggested > 35,000 prokaryotic species in the epipelagic (down to 200 m depth) and mesopelagic (200 m - 1,000 m depth) ocean waters (36). These observations indicate the complex community structure in the world oceans and generate a plethora of questions to investigate the ecological processes shaping the enormous diversity of marine microbes.

The surface waters of coastal marine zones represent a dynamic environment with high microbial diversity. Complex hydrodynamic and meteorological factors drive the water movement and lead to oscillations in nutrient concentrations, which support the growth of diverse microbes (37). Oceanographic time series in the coastal ocean provide valuable information on the microbial community dynamics in temporal and spatial scales. An example of such variation is the seasonal change in community composition. Surface microbial communities in the western English Channel were sampled monthly for six years and investigated using high-resolution 16S rRNA gene tag pyrosequencing. This study revealed a diverse microbial community, which varied according to seasons (38) (39). Moreover, examination of Group II *Euryarchaeota* and Group I *Thaumarchaeota* abundances at the Blanes Bay Microbial Observatory in the northwestern Mediterranean Sea also showed a seasonal pattern of variation (40). Overall, these seasonal community dynamics suggested distinct ecological niches of coastal bacterioplankton. To have a holistic understanding of microbial community structure, these niches should be investigated in conjunction with metabolic repertoires and linked to biogeochemical cycles in the ocean.

1.4) Organic carbon hot-spots in the coastal ocean: algal blooms

Marine microorganisms drive the cycling of biogenic elements (e.g. carbon, nitrogen, sulfur, and phosphorus) in the ocean. These biogeochemical cycles are tightly associated with primary productivity, which largely refers to the production of organic matter by photosynthesis. Primary production in the euphotic zone of the surface ocean is dominated by phytoplankton (henceforth termed algae), most of which are single-celled photoautotrophic organisms. Algae harvest light to convert inorganic carbon into organic carbon, which subsequently enters a complex food web (Fig. 1.4). This trophic network includes diverse heterotrophic organisms such as bacteria, zooplankton and nekton (e.g. fish) that obtain their energy solely from the respiration of organic matter (Fig. 1.4). In the traditional food web, the algal biomass is consumed by zooplankton, which is grazed by multicellular eukaryotic nekton (41). Algal-derived organic matter is therefore transferred to higher trophic levels (Fig. 1.4). Furthermore, organic matter is continuously released back to the environment by viral lysis, excretion of waste products, and exudation of exopolymers (42). This released organic matter is composed of both dissolved organic matter (DOM, particle size < 0.45 μm) and particulate organic matter (POM, particle size > 0.45 μm). A small portion of POM (e.g. fecal pellets or debris) aggregates and sinks to the aphotic layer providing a nutrient source to the inhabitants of the deep ocean (Fig. 1.4) (42). A large fraction of POM (about 90-95%) is remineralized by heterotrophic bacteria or solubilized enzymatically and thus is recycled as DOM to the food web (Fig. 1.4) (42). DOM is the largest carbon reservoir in the oceans but is only available to bacterioplankton (41). Diverse heterotrophic bacteria utilize algal-derived DOM within their catabolic metabolisms and release CO_2. A fraction of these bacteria is also consumed by zooplankton, and the organic matter is thereby reintroduced to higher levels of the food web (Fig. 1.4). These interactions are called the “microbial loop” and play an important role in balancing carbon remineralization and sequestration (43). Therefore, the trophic connection between algae and heterotrophic bacteria forms the foundation of carbon turnover in the world oceans.

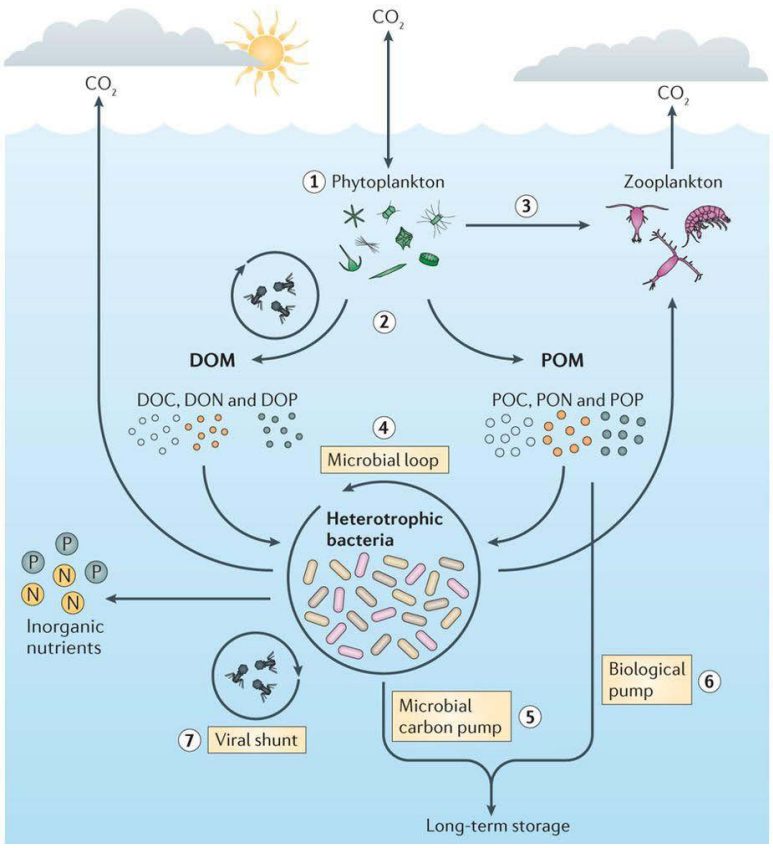


Fig. 1.4. Bacterial transformation of algal-derived organic matter. The conversion of CO_2 to organic carbon by photosynthetic phytoplankton (step 1); the release of both dissolved organic matter (DOM) and particulate organic matter (POM) from phytoplankton (step 2); the consumption of phytoplankton biomass by zooplankton grazers (step 3) and the mineralization and recycling of organic matter by diverse heterotrophic bacteria, which is known as the microbial loop (step 4). A fraction of the heterotrophic bacteria is consumed by zooplankton, and the carbon is further transferred up the food web. Transformation of organic carbon into recalcitrant DOC (step 5). The export of phytoplankton-derived POM from the surface oceans to deeper depths via sinking (step 6). The contributions of viral-mediated cell lysis to the release of dissolved and particulate matter from both the phytoplankton and bacterial pools (step 7). The figure was taken from (43).

Algal blooms are important hotspots for carbon recycling in the surface waters of the coastal marine zones. In the higher latitudes, primary production in the coastal ocean peaks during annually recurring spring algal blooms. As temporal and highly dynamic events, these blooms develop in the spring months in response to the longer duration of light exposure, higher seawater temperature, reduced grazing pressure, and elevated nutrient concentrations (43). These conditions promote the growth of different phytoplankton species depending on the physiochemical parameters (44). This diverse community declines in response to grazing pressure, viral lysis (“top-down” control), or nutrient limitation (“bottom-up” control) (43). Termination of such blooms

results in the massive release of high molecular weight compounds such as polysaccharides, proteins, and lipids.

Polysaccharides constitute the major component of organic matter released during algal blooms. These polymeric carbohydrate molecules are composed of long chains of monosaccharide units, which are linked together by glycosidic bonds, and can form highly branched structures. Algae use these structurally diverse macromolecules as storage compounds as well as cell wall and matrix components. Depending on the growth stage, the polysaccharide content of algae can reach up to ~90% of the cell's dry weight in extreme cases (45), but more often accounts for 30-65% (46). Thus, large amounts of algal polysaccharides are released during algal blooms. However, it is challenging to measure relevant algal polysaccharides *in situ* with structural elucidation. The identification of the monomer units and a compositional analysis of the polysaccharides is possible using acid hydrolysis in combination with chromatographic assays (47). These analyses revealed glucose, mannose, fucose, arabinose, xylose, rhamnose and galactose as the major monosaccharide units of algal glycans (48). Yet these methods cannot yield insights into the structural information since polysaccharides occur in marine samples as complex mixtures with various types of linkages (49). To overcome this problem, Becker and colleagues proposed a new approach employing microbial enzymes as a tool to quantify polysaccharides in POM (50). They used three enzymes purified from marine bacteria to selectively digest laminarin, the major storage polysaccharide of diatoms (51). Demonstrating polysaccharide release during an algal bloom, the laminarin concentration was 0.48±0.09 mg/L and 0.13±0.02 mg/L in 10 µm and 3 µm POM fractions, respectively. Therefore, microbial enzymes specifically targeting various linkage types in algal polysaccharides are particularly important as a biochemical proxy for the presence and utilization of a given polysaccharide.

1.5) Polysaccharide utilization: genes, enzymes, and mechanisms

The biochemical and genomic characterization of polysaccharide utilization during algal blooms sheds light on the trophic connection between algal-derived organic matter and heterotrophic bacteria. Large amounts of algal

polysaccharides released upon bloom termination are consumed by heterotrophic bacteria using carbohydrate-active enzymes (CAZymes). These enzymes include catalysts for polysaccharide binding, degradation, and modification, and are classified as carbohydrate-binding modules (CBMs), glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and glycosyltransferases (GTs) (52). These classified enzymes are further subdivided into families based on sequence and structural similarity (53). CAZymes are multimodular and enzymes targeting a particular polysaccharide could be assigned to separate families. For example, GH families 16, 17 and 30 are required to decompose laminarin, which consists of a β-1,3-D-glucose polysaccharide with β-1,6 linked monomer side chains (Fig. 1.5) (50). The GH30 β-1,6-glucosidase removes the β-1,6-linked side chains (54), while endo-active GH16 cleaves the β-1,3-D-linked main glucan chain of laminarin into smaller oligosaccharides (55). Finally, exo-acting GH17 β-1,3-glucosidase hydrolyzes these oligosaccharides to glucose units (Fig. 1.5) (56). Thus, different degradative CAZyme families are required to cleave diverse linkages found in a single complex polysaccharide.

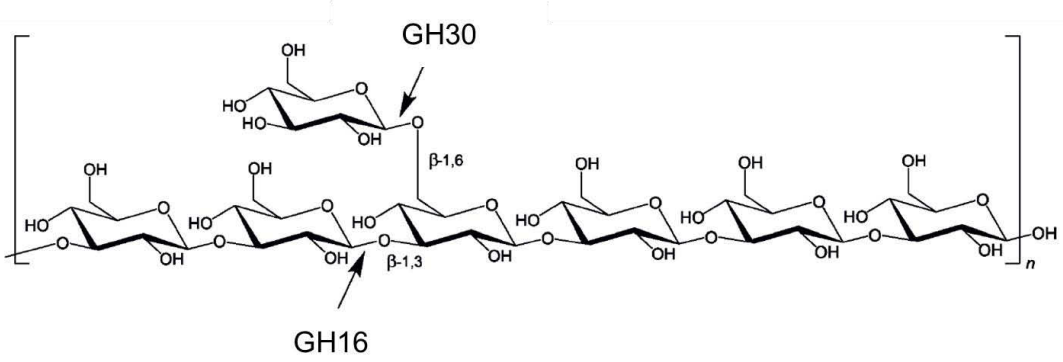


Fig. 1.5. Laminarin structure and enzyme activities. Laminarin consists of β-1,3-D-glucose polysaccharide with β-1,6 linked monomer side chains. GH30 cleaves the β-1,6 linkages, while GH16 hydrolyzes β-1,3-D-linked main chain. GH17 is an exo-acting enzyme (not shown in the figure). Figure adapted from (50).

Diverse CAZyme genes encoding the binding and degradation of a particular polysaccharide are often co-localized in polysaccharide utilization loci (PULs). PULs are distinct genomic islands, which enable sensing, binding, importing and degrading a polysaccharide (Fig. 1.6) (57) (58). The archetypal PUL encoding the starch utilization system (Sus) was first demonstrated in

Bacteroides thetaiotaomicron (59). Here, the polysaccharide is recruited to the cell surface by three Sus proteins (SusD, E, and F), some of which form immobile complexes (60) and facilitate the selection of particular glycans from the environment (61). The recruited polysaccharide is cleaved into oligosaccharides by the outer membrane localized glycoside hydrolases (Fig. 1.6). These oligosaccharides are subsequently transported across the outer membrane into the periplasm via SusC, an integral membrane protein of TonB-dependent transporters (TBDT) (Fig. 1.6). The beta-barrel transport channel is capped by the N-terminally lapidated SusD-like cell surface glycan-binding protein (SGBP), forming a “pedal-bin” like protein complex (62). Upon substrate binding, the SusD lid closes and leads the substrate transport into the periplasm (63). Oligosaccharides transported into the periplasm are further degraded into monosaccharides by linkage-specific exo-glucosidases before they are imported into the cytoplasm for primary metabolism (Fig. 1.6). Therefore, from binding to the cell surface to the degradation into monomers, PULs orchestrate the catabolism of a particular polysaccharide.

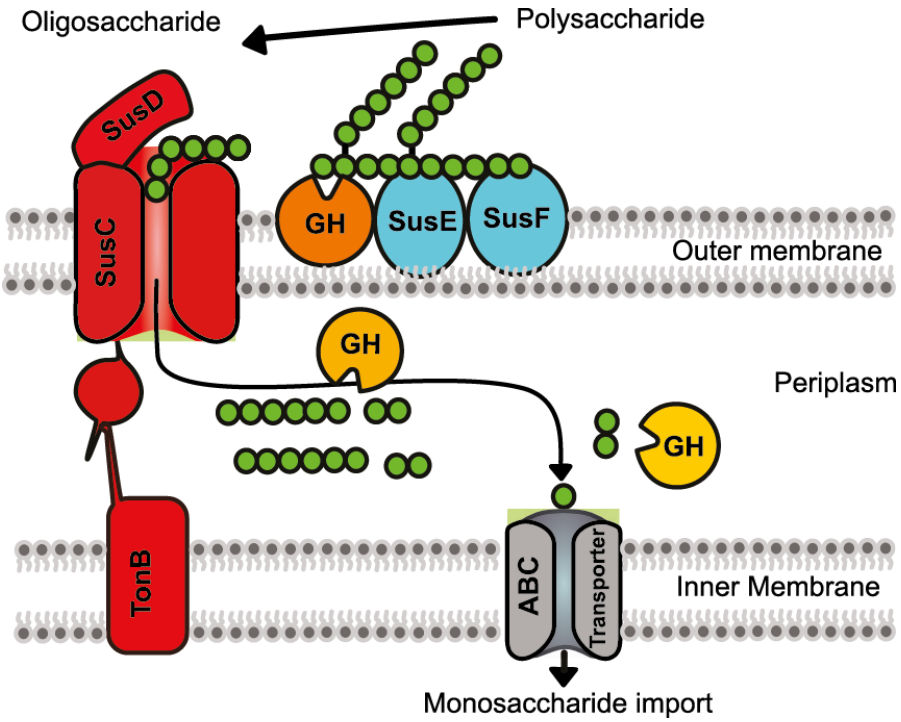


Fig. 1.6. Sensing, binding, importing and degrading a polysaccharide via the starch utilization (Sus) system. The recognition, binding, and initial hydrolysis of a given polysaccharide occur at the outer membrane via SusD, SusE, and SusF. The oligosaccharides are subsequently transported into the periplasm via SusC/D; further hydrolysis occurs in the periplasm via glycosyl hydrolases (GH). Figure modified after (64). SusR is used for substrate sensing (not shown in the figure).

Diverse canonical PULs targeting specific polysaccharides have been identified in marine heterotrophic bacteria using the SusC/SusD tandem as a signature. For example, polysaccharide utilization in the marine *Bacteroidetes Gramella forsetii* KT0803^T was characterized using a combined application of genomics, proteomics and isotope labeling (Fig. 1.7) (65). This study identified two PULs specifically induced, one by alginate and the other one by laminarin. In addition to encoding GH3 and two GH16 (β-1,3-glucanases) genes, the laminarin PUL also contained a SusCD pair together with a Polycystic Kidney Disease (PKD) domain containing protein. Upregulation of these genes in the presence of laminarin was also demonstrated by proteomic analysis (Fig. 1.7). PUL analysis thus represents an important genomic proxy to detect the carbohydrate degradation potential of a bacterium and to define the polysaccharide niches of heterotrophic bacteria. This approach would shed light on ecological processes shaping the microbial communities during algal blooms, in which the large amounts of diverse polysaccharides are released.

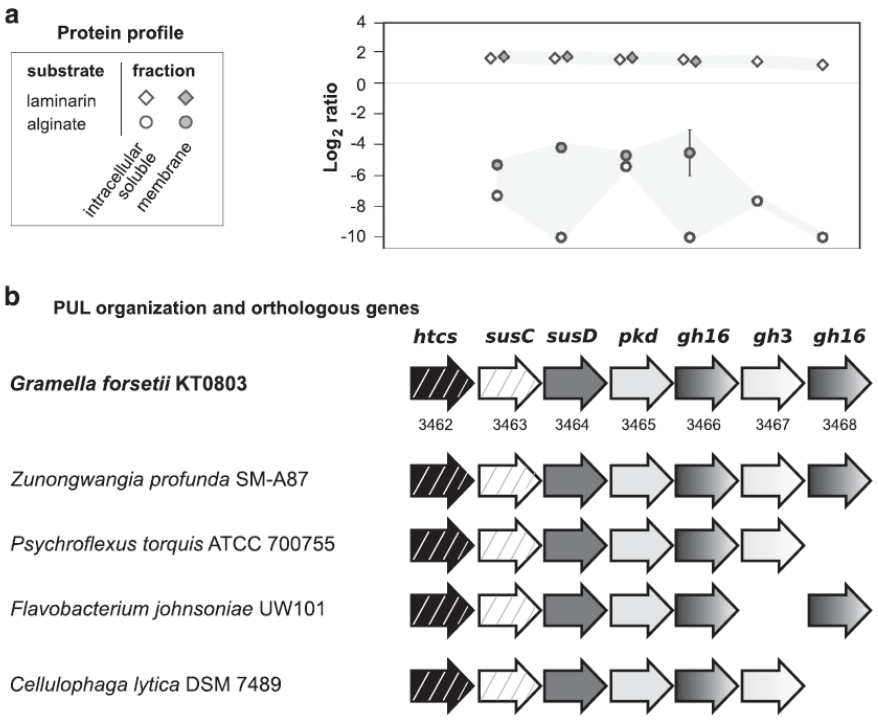


Fig. 1.7. Expression profile **(a)** and gene organization **(b)** of laminarin utilization operon in *Gramella forsetii* KT0803^T. **(a)** Protein ratios of the intracellular soluble and membrane protein fractions of cultures grown on laminarin compared with alginate are shown. Error bars refer to the biological replicates (n=3). **(b)** The gene organization of laminarin-specific PUL in *G. forsetii* together with others in some marine *Flavobacteriaceae*. htcs, hybrid two-component system; susC, SusC-like TBDR; susD, SusD-family protein; pkd, PKD-domain-containing protein; gh, glycoside hydrolase. The figure was taken from (65).

1.6) Key heterotrophic bacteria abundant during algal blooms

The organic matter released during algal blooms triggers a significant shift in the microbial community composition. Diverse heterotrophic bacteria become abundant and degrade algal-derived organic compounds. This bacterial response was initially investigated in microcosm or mesocosm experiments. For example, Riemann and colleagues studied the bacterial community composition, enzymatic activities, and carbon dynamics during diatom blooms in seawater mesocosms (66). This study demonstrated extensive colonization of heterotrophic bacteria of the phyla *Alphaproteobacteria* and *Bacteroidetes* with high aminopeptidase, β-glucosidase, and lipase activities. Another study by Pinhassi and colleagues revealed links between phytoplankton and bacterioplankton in microcosm experiments using denaturing gradient gel electrophoresis (DGGE) of 16S rRNA gene fragments (67). Members of the alphaproteobacterial *Roseobacter* clade, *Gammaproteobacteria* and the class *Flavobacteriia* (phylum *Bacteroidetes*) were detected in response to the algal-derived organic matter.

Recent advances in high-throughput sequencing have allowed *in situ* investigations of microbial species composition during algal blooms. Pyrosequencing-based analysis of 16S rRNA gene amplicons from environmental RNA revealed a *Roseobacter* clade-affiliated cluster, and the gammaproteobacterial SAR92 clade as the most abundant and active bacterial clades during an algal bloom in the southern North Sea (68). Furthermore, Delmont and his colleagues investigated bacterial communities following a phytoplankton bloom in the Amundsen Sea, Antarctica (69). This study demonstrated the power of state-of-the-art omics technologies to dissect the identity and function of diverse microorganisms with genome reconstructions and indicated the ecological roles of the SAR92 and flavobacterial *Cryomorphaceae* clades with a coarse niche description. Overall these studies repeatedly highlighted that the members of the *Roseobacter*, *Flavobacteriia* and *Gammaproteobacteria* clades are key responders to algal blooms.

The alphaproteobacterial *Roseobacter* clade is widespread in world oceans, comprising up to 20% of coastal bacterioplankton communities (70). For

example, *Planktomarina temperata* RCA23, the first described species of *Roseobacter* clade affiliated cluster, is abundant in world oceans and has a streamlined genome with genes for aerobic anoxygenic photosynthesis and carbon monoxide oxidation (71). *Roseobacter* clade is also associated with marine phytoplankton (72). Hahnke and colleagues explored the physiological diversity of the *Roseobacter* clade co-occurring during an algal bloom in the North Sea (73). A great metabolic diversity with respect to the decomposition of an algal osmolyte dimethylsulfoniopropionate (DMSP) was demonstrated using seven strains. These observations reflected different ecological niches within this clade during algal blooms. Furthermore, *Bacteroidetes*, in general, and marine *Flavobacteriia,* in particular, are associated with algal-derived organic matter. For example, the genomic content of uncultured *Bacteroidetes* from contrasting oceanic provinces in the North Atlantic Ocean was investigated using a targeted metagenomic approach with fosmid libraries (74). The results showed a large genomic potential of flavobacterial fosmids for polysaccharide degradation and cell surface attachment in the phytoplankton-rich polar biome. A follow-up study at the same sampling site with the further analysis of 174 fosmids demonstrated polysaccharide utilization loci (PULs) from diverse *Flavobacteriia* putatively targeting polysaccharides such as laminarin, xylan, and mannan (75). *Gammaproteobacteria* also degrade algal-derived polysaccharides. Utilization of alginate and other algal polysaccharides was shown in the gammaproteobacterial *Alteromonas macleodii* strain 83-1 (76). An alginate-induced PUL harboring five alginate lyases and TonB-dependent receptors was shown together with putative PULs hydrolyzing laminarin, pullulan, and xylan. These studies overall demonstrated the broad algal-derived organic matter degradation potential of *Roseobacter*, *Flavobacteriia*, and *Gammaproteobacteria*. The metabolic potential of these key players should be investigated *in situ* during algal blooms with a high taxonomic and functional resolution.

1.7) Bacterial response to spring algal blooms in the North Sea

The bacterial response to spring algal blooms off Helgoland Island in the Southern North Sea has been studied in great detail. The Helgoland Roads

represents one of the richest time series dataset for a coastal marine environment (Fig. 1.8) (77). Phytoplankton, salinity, Secchi disc depths, and macronutrients have been documented since 1962 and are further augmented by zooplankton, intertidal macroalgae, macrozoobenthos, and bacterioplankton data. Together with these metadata, we investigated the microbial community structure and function during the 2009-2012 spring algal blooms at Helgoland using a comprehensive suite of state-of-the-art techniques (78) (79). Microbial community composition was determined using 16S rRNA gene tag sequencing and catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) with dense sampling in multiple years. Metagenome sequencing enabled us to access the genetic repertoire of microorganisms, while metaproteomics was used to monitor *in situ* gene expression. Furthermore, follow-up studies aimed to isolate heterotrophic bacteria that were abundant during bloom events (80) and to analyze the micro-diversity of recurrent bacteria at a fine scale (81). The synergy between these studies has yielded an immense dataset to holistically understand bacterial dynamics and niches during spring algal blooms.

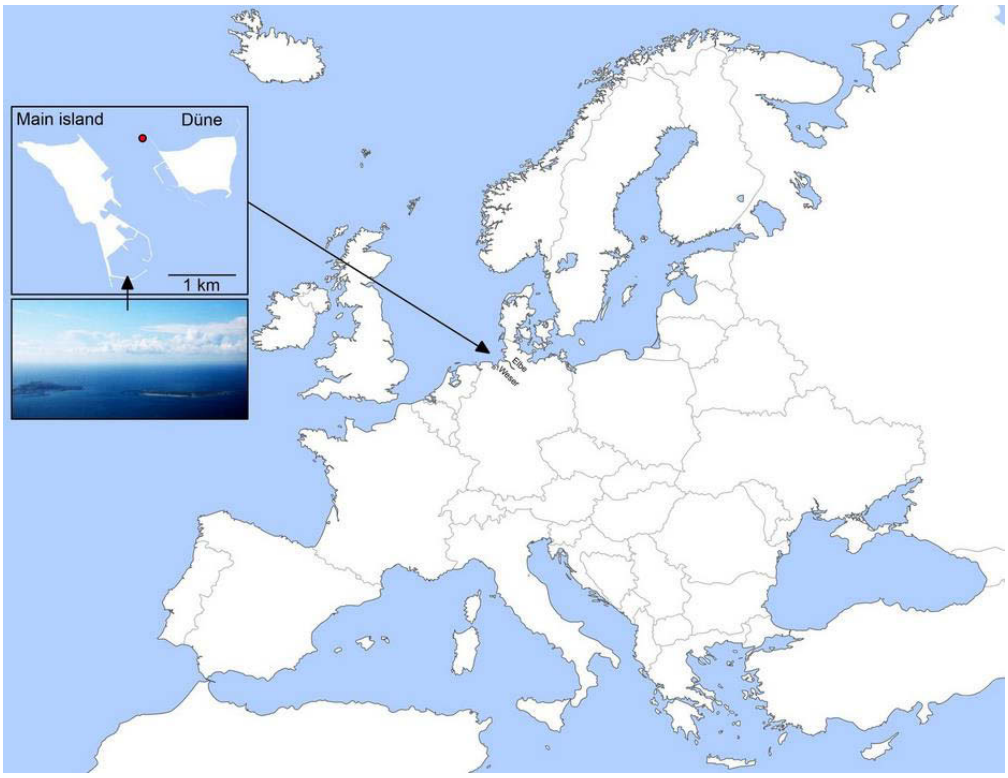


Fig. 1.8. Location of Helgoland Island and the long-term ecological research site 'Kabeltonne' (red circle: 54° 11.3' N, 7° 54.0' E) in the German Bight of the North Sea. The figure was taken from (79).

Succession and recurrence of *Flavobacteriia* and *Gammaproteobacteria* were detected during the 2009-2012 spring algal blooms off Helgoland (Fig. 1.9) (78) (79). In the 2009 spring bloom, the flavobacterial genera *Ulvibacter* (recently reclassified as *Cd.* 'Prosiliicoccus'), *Formosa,* and *Polaribacter* successively reached high relative abundances in the bacterioplankton (0.2–3 µm fraction) together with the gammaproteobacterial *Reinekea* and SAR92 clades (Fig. 1.9) (78). Metagenomic analyses demonstrated the differentiated transporter and CAZyme contents of these bacterial clades (78). With a pronounced expression of TonB-dependent transporter (TBDT) components, *Flavobacteriia* possessed various CAZyme genes to utilize diverse algal polysaccharides such as laminarin, fucose, and α-mannose containing polysaccharides. For example, *Polaribacter* spp. encoded glycoside hydrolase (GH) family 92, which targets α-mannose-rich polysaccharides, yet genus-specific CAZyme assignment for *Formosa* spp. was not possible. In contrast, *Reinekea* spp. exhibited high expression of ABC and TRAP transporters, indicating a nutritional strategy based on the uptake of low molecular weight compounds. In a follow-up study, we revealed the recurrence of these bacterial clades during 2010-2012 spring blooms at the same sampling site (79). Deeply sequenced metagenomes also allowed us to detect the genomic contents of relevant bacterioplankton down to genus level, and to further investigate the CAZyme repertoires of the *Reinekea*, *Polaribacter*, and *Formosa* clades.

Reinekea spp. occupied a limited glycan niche with a clonal proliferation during the North Sea spring algal blooms. During 2009-2012 North Sea spring algal blooms, *Reinekea* spp. were recurrently present albeit with decreasing abundances (~16%, ~5%, ~2%, and ~1%) based on CARD-FISH counts (Fig. 1.9) (79). Minimum entropy decomposition (MED) analysis of 16S rRNA tag sequences, which can detect microdiversity, demonstrated only one dominant oligotype and suggested a quasi-clonal *Reinekea* population during the bloom events (81). Metagenomic analyses also showed the carbohydrate degradation potential of the North Sea *Reinekea* clade. The genus-specific CAZyme repertoire of this clade contained high proportions of GH23 (e.g. peptidoglycan lyases) and GH13 (e.g. α-amylase) with various carbohydrate

binding modules (CBMs) to bind α-1,4-glucans and peptidoglycans (79). This was a relatively limited and homogeneous glycan niche compared to other *Flavobacteriia* abundant during spring blooms. Furthermore, Hahnke and colleagues isolated a *Reinekea* strain (Hel1_31_D35) during the 2010 spring bloom at the same sampling site (80) and suggested that this strain could be the representative of *Reinekea* clade, which is abundant during the bloom events. This paved the way for exploring the niche of North Sea *Reinekea* clade in detail and testing the hypotheses deduced from (meta)genomic analyses in pure culture studies.

The genus *Polaribacter* was the most abundant and diverse recurring bacterial clade with a broad glycan niche during North Sea spring algal blooms from 2009 to 2012. CARD-FISH analysis with a *Polaribacter*-specific oligonucleotide probe (POL740) detected ~27%, ~26%, ~14%, and ~25% maximum relative abundances in the 0.2-3 µm planktonic fraction during 2009 to 2012 spring algal blooms in Helgoland (Fig. 1.9) (79). MED analysis also demonstrated four different *Polaribacter* oligotypes, which exceeded 1% relative abundance (81). These results indicated a highly diverse and abundant clade, which was recurrently associated with North Sea spring algal blooms. Furthermore, the metagenomic analysis showed the enrichment of GH families 16, 17, and 30 (all targeting laminarin), GH13 (α-amylases), GH92 (α-mannosidases) and suggested a broad glycan niche for the bloom-associated *Polaribacter* clade (79). Isolation and characterization of North Sea *Polaribacter* strains (Hel1_33_49 and Hel1_85), which were obtained from the same sampling site (80), showcased two distinct niches in this large niche space (82). Strain Hel1_33_49 was a planktonic isolate feeding on proteins and a small subset of algal polysaccharides, whereas strain Hel1_85 could decompose a larger spectrum of high molecular weight compounds, and was likely associated with algae. Yet, these isolates were not members of the most abundant *Polaribacter* clades from North Sea spring algal blooms. Therefore, the broad niche space of environmentally relevant *Polaribacter* spp. should be investigated in detail by separating their taxonomic and functional diversity into ecologically meaningful niches.

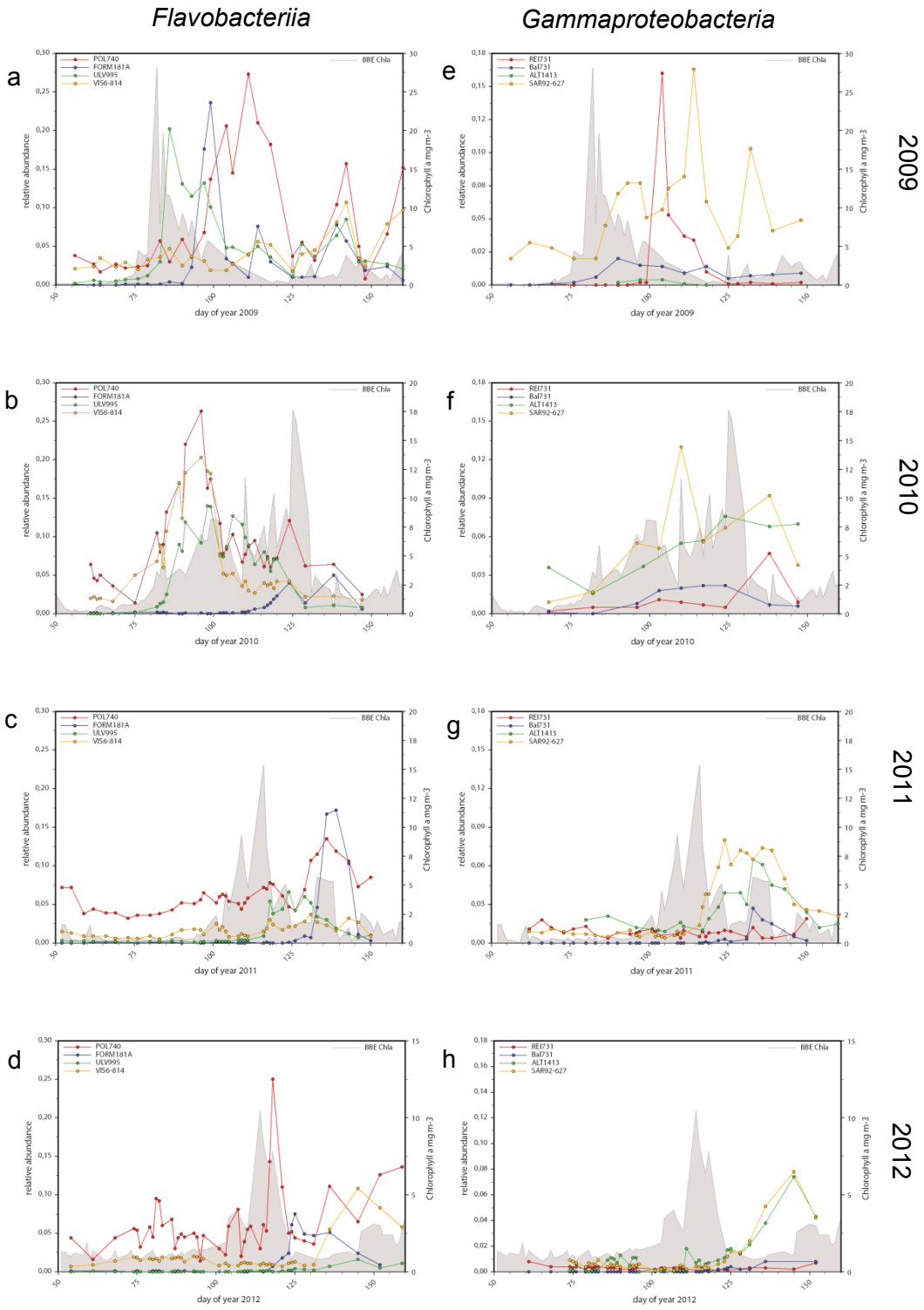


Fig. 1.9. The recurrent abundance of major *Flavobacteriia* **(a-d)** and *Gammaproteobacteria* **(e-h)** during 2009-2012 North Sea spring algal blooms off Helgoland as assessed by CARD-FISH analyses. POL740: genus *Polaribacter*; FORM181A: genus *Formosa*; ULV995: genus *Ulvibacter*; VIS6-814: genus-level clade VIS6 within the family *Cryomorphaceae-Owenweeksia*. REI731: genus *Reinekea*; BAL731: genus *Balneatrix*; ALT1413: families *Alteromonadaceae* and *Colwelliaceae*; SAR92-627: genus-level clade SAR92. Figure modified after (79).

Formosa was another abundant and recurrent flavobacterial genus during North Sea spring algal blooms. Mainly belonging to three dominant 16S rRNA oligotypes (81), *Formosa* spp. reached relative abundances of up to ~23%, ~5%, ~17%, ~7% during 2009-2012 North Sea spring algal blooms off Helgoland (Fig. 1.9) (79). Similar to *Polaribacter* spp., the CAZyme spectrum of bloom-associated *Formosa* clade contained the GH13 and GH92 families, which decompose laminarin and mannan (79). Furthermore, five *Formosa* strains were isolated from the surface waters of Helgoland Roads during the 2010 bloom event (80). Metagenome read recruitment on the initial draft genomes showed that 4.1% of the metagenome reads were aligned to the strain Hel1_31_131 (clade B), while the strain Hel3_A1_48 (clade A) recruited 0.15% of the all reads. This result indicated that the clades A and B could represent *Formosa* spp., which were abundant during the bloom events and their genomic analyses in conjunction with pure culture studies and biochemical characterization could shed light on the adaptive metabolisms allowing heterotrophic bacteria to thrive during algal blooms.

Overall these results suggest that the release of diverse polysaccharides during North Sea spring algal blooms creates successive niche spaces and recurrently selects a particular group of heterotrophic bacteria. The niches of these environmentally relevant bacteria should be investigated in great detail combining (meta)genomic, (meta)proteomic, pure culture and biochemical studies to yield a holistic understanding of ecological and evolutionary processes shaping carbon cycling in the ocean.

1.8) Aims and scope of the thesis

In this thesis, I aim to explore the ecological niches of abundant heterotrophic bacteria in the North Sea spring algal blooms.

The second chapter describes the ecological niche of the gammaproteobacterium '*Reinekea forsetii*'. Using phylogenetic and metagenome analyses, I first showed that '*R. forsetii*' is the representative of the North Sea *Reinekea* clade. Then physiological experiments were conducted to test ecological hypotheses generated from the genome of '*R. forsetii*' and accompanying *in situ* environmental data.

The third chapter explores the large niche space of the diverse North Sea *Polaribacter* clade. I identified distinct *Polaribacter* sub-clades using phylogenetic analyses and quantified their abundances with novel oligonucleotide probes. Metabolic potentials of these environmentally relevant clades were demonstrated by retrieval of almost complete representative metagenome assembled genomes and assessing their *in situ* gene expression with metaproteome data.

The fourth chapter represents the polysaccharide niche of the North Sea *Formosa* clade. Cultivation and biochemical studies were used to complement metagenomic and metaproteomics data and demonstrate the molecular details and ecological implications of the laminarin degradation and protein utilization metabolisms of *Formosa*.

1.9) References

1. Gest H. The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society. Notes Rec Roy Soc. 2004;58(2):187-201.
2. Blevins SM, Bronze MS. Robert Koch and the 'golden age' of bacteriology. Int J Infect Dis. 2010;14(9):E744-51.
3. Vila J, Saez-Lopez E, Johnson JR, Romling U, Dobrindt U, Canton R, et al. *Escherichia coli*: an old friend with new tidings. FEMS Microbiol Rev. 2016;40(4):437-63.
4. Reidl J, Klose KE. *Vibrio cholerae* and cholera: out of the water and into the host. FEMS Microbiol Rev. 2002;26(2):125-39.
5. Staley JT, Konopka A. Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev of Microbiol. 1985;39(1):321-46.
6. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA. 1977;74(11):5088-90.
7. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid-determination of 16S ribosomal-RNA sequences for phylogenetic analyses. Proc Natl Acad Sci USA. 1985;82(20):6955-9.

8. Giovannoni SJ, DeLong EF, Schmidt TM, Pace NR. Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. Appl Environ Microb. 1990;56(8):2572-5.

9. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in-situ detection of individual microbial cells without cultivation. Microbiol Rev. 1995;59(1):143-69.

10. DeLong E, Wickham G, Pace N. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. Science. 1989;243(4896):1360-3.

11. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol. 1998;5(10):245-9.

12. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 2004;304(5667):66-74.

13. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017;35(9):833-44.

14. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458(7239):719-24.

15. Vilanova C, Porcar M. Are multi-omics enough? Nat Microbiol. 2016;1(8):16101.

16. Giovannoni S, Stingl U. The importance of culturing bacterioplankton in the 'omics' age. Nat Rev Microbiol. 2007;5(10):820-6.

17. Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics. ISME J. 2016;10(11):2557-68.

18. Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH. Relative roles of niche and neutral processes in structuring a soil microbial community. ISME J. 2010;4(8):337-45.

19. Kimura M. Evolutionary rate at the molecular level. Nature. 1968;217:624-6.

20. Hubbell SP. The unified neutral theory of biodiversity and biogeography. Princeton: Princeton Univ. Press; 2001.

21. Leigh EG. Neutral theory: a historical perspective. Journal of Evolutionary Biology. 2007;20(6):2075-91.

22. Vandermeer JH. Niche Theory. Annu Rev Ecol Evol and Syst. 1972;3(1):107-32.

23. Darwin C. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life: London: John Murray; 1859.

24. Grinnel J. Geography and Evolution. Ecology. 1924;5(3):225-9.

25. Elton C. Animal Ecology. London: Sidgewick & Jackson; 1927.

26. Hutchinson GE. Population Studies - Animal Ecology and Demography - Concluding Remarks. Cold Spring Harb Sym. 1957;22:415-27.

27. Gauze GF. The struggle for existence. Baltimore: The Williams & Wilkins Company; 1934.

28. Baas Becking LGM. Gaia of leven en aarde. The Hague: Nijhoff. Ambtsrede R.U. Leiden; 1931.

29. Rutger DW, Thierry B. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environ Microbiol. 2006;8(4):755-8.

30. Hutchinson GE. The Paradox of the Plankton. Amer Nat. 1961;95(882):137-45.

31. Schramm A, de Beer D, Wagner M, Amann R. Identification and activities *in situ* of *Nitrosospira* and *Nitrospira* spp. as dominant populations in a nitrifying fluidized bed reactor. Appl Environ Microb. 1998;64(9):3480-5.

32. Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science. 2008;320(5879):1081-5.

33. Coutinho FH, Meirelles PM, Moreira APB, Paranhos RP, Dutilh BE, Thompson FL. Niche distribution and influence of environmental parameters in marine microbial communities: a systematic review. PeerJ. 2015;3:e1008.

34. Bar-On YM, Phillips R, Milo R. The biomass distribution on Earth. Proc Natl Acad Sci USA. 2018; doi:10.1073/pnas.1711842115

35. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: The unseen majority. Proc Natl Acad Sci USA. 1998;95(12):6578-83.

36. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. Science. 2015;348(6237):1261359.

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

38. Gilbert JA, Steele JA, Caporaso JG, Steinbrück L, Reeder J, Temperton B, et al. Defining seasonal marine microbial community dynamics. ISME J. 2011;6(2):298-308.

39. Gilbert JA, Field D, Swift P, Newbold L, Oliver A, Smyth T, et al. The seasonal structure of microbial communities in the Western English Channel. Environ Microbiol. 2009;11(12):3132-9.

40. Galand PE, Gutierrez-Provecho C, Massana R, Gasol JM, Casamajor EO. Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). Limnol Oceanogr. 2010;55(5):2117-25.

41. Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial realm. Nat Microbiol. 2018;3(7):754-66.

42. Azam F, Malfatti F. Microbial structuring of marine ecosystems. Nat Rev Microbiol. 2007;5:782-91.

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

44. Chang FH, Zeldis J, Gall M, Hall J. Seasonal and spatial variation of phytoplankton assemblages, biomass and cell size from spring to summer across the north-eastern New Zealand continental shelf. J Plankton Res. 2003;25(7):737-58.

45. Myklestad S. Production of carbohydrates by marine planktonic diatoms. I. Comparison of nine different species in culture. J Exp Mar Biol Ecol. 1974;15(3):261-74.

46. Markou G, Angelidaki I, Georgakakis D. Microalgal carbohydrates: an overview of the factors influencing carbohydrates production, and of main bioconversion technologies for production of biofuels. Appl Microbiol Biotechnol. 2012;96(3):631-45.

47. Engel A, Händel N. A novel protocol for determining the concentration and composition of sugars in particulate and in high molecular weight dissolved organic matter (HMW-DOM) in seawater. Mar Chem. 2011;127(1):180-91.

48. Mühlenbruch M, Grossart H-P, Eigemann F, Voss M. Phytoplankton-derived polysaccharides in the marine environment and their interactions with heterotrophic bacteria. Environ Microbiol. 2018;doi:10.1111/1462-2920.14302.

49. Hedges JI, Baldock JA, Gélinas Y, Lee C, Peterson M, Wakeham SG. Evidence for non-selective preservation of organic matter in sinking marine particles. Nature. 2001;409:801-4.

50. Becker S, Scheffel A, Polz MF, Hehemann J-H. Accurate quantification of laminarin in marine organic matter with enzymes from marine microbes. Appl Environ Microb. 2017;83(9): e03389-16.

51. Alderkamp AC, van Rijssel M, Bolhuis H. Characterization of marine bacteria and the activity of their enzyme systems involved in degradation of the algal storage glucan laminarin. FEMS Microbiol Ecol. 2007;59(1):108-17.

52. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(D1):D490-5.

53. Davies G, Henrissat B. Structures and mechanisms of glycosyl hydrolases. Structure. 1995;3(9):853-9.

54. Sano K, Amemura A, Harada T. Purification and properties of a β-1,6-glucosidase from *Flavobacterium*. Biochim Biophys Acta. 1975;377(2):410-20.

55. Chesters CGC, Bull AT. The enzymic degradation of laminarin. 2. The multicomponent nature of fungal laminarinases.Biochem J. 1963;86(1):31-8.

56. Barras DR, Stone BA. β-1,3-glucan hydrolases from *Euglena gracilis*: II. Purification and properties of the β-1,3-glucan exo-hydrolase. Biochim Biophys Acta. 1969;191(2):342-53.

57. Hemsworth GR, Déjean G, Davies Gideon J, Brumer H. Learning from microbial strategies for polysaccharide degradation. Biochem Soc Trans. 2016;44(1):94-108.

58. Martens EC, Koropatkin NM, Smith TJ, Gordon JI. Complex glycan catabolism by the human gut microbiota: the *Bacteroidetes* Sus-like paradigm. J Biol Chem. 2009;284(37):24673-7.

59. Shipman JA, Berleman JE, Salyers AA. Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. J Bacteriol. 2000;182(19):5365-72.

60. Tuson HH, Foley MH, Koropatkin NM, Biteen JS. The starch utilization system assembles around stationary starch-binding proteins. Biophys J. 2018;115(2):242-50.

61. Foley MH, Martens EC, Koropatkin NM. SusE facilitates starch uptake independent of starch binding in *B. thetaiotaomicron*. Mol Microbiol. 2018;108(5): 551-66.

62. Glenwright AJ, Pothula KR, Bhamidimarri SP, Chorev DS, Baslé A, Firbank SJ, et al. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. Nature. 2017;541(7637):407-11.

63. Hickman SJ, Cooper REM, Bellucci L, Paci E, Brockwell DJ. Gating of TonB-dependent transporters by substrate-specific forced remodelling. Nat Commun. 2017;8:14804.

64. Reintjes G, Arnosti C, Fuchs BM, Amann R. An alternative polysaccharide uptake mechanism of marine bacteria. ISME J. 2017;11(7):1640-50.

65. Kabisch A, Otto A, König S, Becher D, Albrecht D, Schüler M, et al. Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes '*Gramella forsetii*' KT0803. ISME J. 2014;8(7):1492-502.

66. Riemann L, Steward GF, Azam F. Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. Appl Environ Microb. 2000;66(2):578-87.

67. Pinhassi J, Sala MM, Havskum H, Peters F, Guadayol O, Malits A, et al. Changes in bacterioplankton composition under different phytoplankton regimens. Appl Environ Microb. 2004;70(11):6753-66.

68. Wemheuer B, Güllert S, Billerbeck S, Giebel H-A, Voget S, Simon M, et al. Impact of a phytoplankton bloom on the diversity of the active bacterial community in the southern North Sea as revealed by metatranscriptomic approaches. FEMS Microbiol Ecol. 2014;87(2):378-89.

69. Delmont TO, Eren AM, Vineis JH, Post AF. Genome reconstructions indicate the partitioning of ecological functions inside a phytoplankton bloom in

the Amundsen Sea, Antarctica. Front Microbiol. 2015: doi:10.3389/fmicb .2015.01090.

70. Buchan A, González JM, Moran MA. Overview of the Marine *Roseobacter* Lineage. Appl Environ Microb. 2005;71(10):5665-77.

71. Voget S, Wemheuer B, Brinkhoff T, Vollmers J, Dietrich S, Giebel HA, et al. Adaptation of an abundant *Roseobacter* RCA organism to pelagic systems revealed by genomic and transcriptomic analyses. ISME J. 2014;9(2):371-84.

72. Alavi M, Miller T, Erlandson K, Schnider R, Belas R. Bacterial community associated with *Pfiesteria*-like dinoflagellate cultures. Environ Microbiol. 2001;3(6):380-96.

73. Hahnke S, Brock NL, Zell C, Simon M, Dickschat JS, Brinkhoff T. Physiological diversity of *Roseobacter* clade bacteria co-occurring during a phytoplankton bloom in the North Sea. Syst Appl Microbiol. 2013;36(1):39-48.

74. Gómez-Pereira PR, Schüler M, Fuchs BM, Bennke C, Teeling H, Waldmann J, et al. Genomic content of uncultured *Bacteroidetes* from contrasting oceanic provinces in the North Atlantic Ocean. Environ Microbiol. 2012;14(1):52-66.

75. Bennke CM, Krüger K, Kappelmann L, Huang S, Gobet A, Schüler M, et al. Polysaccharide utilisation loci of *Bacteroidetes* from two contrasting open ocean sites in the North Atlantic. Environ Microbiol. 2016;18(12):4456-70.

76. Neumann AM, Balmonte JP, Berger M, Giebel HA, Arnosti C, Voget S, et al. Different utilization of alginate and other algal polysaccharides by marine *Alteromonas macleodii* ecotypes. Environ Microbiol. 2015;17(10):3857-68.

77. Wiltshire KH, Kraberg A, Bartsch I, Boersma M, Franke HD, Freund J, et al. Helgoland Roads, North Sea: 45 years of change. Estuaries Coast. 2010;33(2):295-310.

78. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012;336(6081):608-11.

79. Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L, et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. eLIFE. 2016;5:e11888.

80. Hahnke RL, Bennke CM, Fuchs BM, Mann AJ, Rhiel E, Teeling H, et al. Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. Environ Microbiol. 2015;17(10):3515-26.

81. Chafee M, Fernandez-Guerra A, Buttigieg PL, Gerdts G, Eren AM, Teeling H, et al. Recurrent patterns of microdiversity in a temperate coastal marine environment. ISME J. 2018;12(1):237-52.

82. Xing P, Hahnke RL, Unfried F, Markert S, Huang S, Barbeyron T, et al. Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. ISME J. 2015;9(6):1410-22.

Chapter 2

Genomic and physiological analyses of '*Reinekea forsetii*' reveal a versatile opportunistic lifestyle during spring algae blooms

Published in *Environmental Microbiology*

volume 19, issue 3, pages 1209-1221

doi.org/10.1111/1462-2920.13646

Contributions to the manuscript:

Experimental concept and design: 50%

Experimental work and/or acquisition of data: 60%

Data analysis and interpretation: 60%

Preparation of figures and tables: 80%

Drafting of the manuscript: 80%

environmental microbiology

Environmental Microbiology (2017) **19**(3), 1209–1221 doi:10.1111/1462-2920.13646

Genomic and physiological analyses of '*Reinekea forsetii*' reveal a versatile opportunistic lifestyle during spring algae blooms

Burak Avcı,[1] Richard L. Hahnke,[1,2] Meghan Chafee,[1] Tanja Fischer,[1] Harald Gruber-Vodicka,[1] Halina E. Tegetmeyer,[1,3] Jens Harder,[1] Bernhard M. Fuchs,[1] Rudolf I. Amann[1]* and Hanno Teeling[1]*

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

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

[3]*Institute for Genome Research and Systems Biology, Center for Biotechnology, University of Bielefeld, Universitätsstraße 27, Bielefeld 33615, Germany.*

Summary

Gammaproteobacterial *Reinekea* spp. were detected during North Sea spring algae blooms in the years 2009-2012, with relative abundances of up to 16% in the bacterioplankton. Here, we explore the ecophysiology of *'R. forsetii'* strain Hel1_31_D35 that was isolated during the 2010 spring bloom using (i) its manually annotated, high-quality closed genome, (ii) re-analysis of *in situ* data from the 2009–2012 blooms and (iii) physiological tests. High resolution analysis of 16S rRNA gene sequences suggested that *'R. forsetii'* dominated *Reinekea* populations during these blooms. This was corroborated by retrieval of almost complete Hel1_31_D35 genomes from 2009 and 2010 bacterioplankton metagenomes. Strain Hel1_31_D35 can use numerous low-molecular weight substrates including diverse sugar monomers, and few but relevant algal polysaccharides such as mannan, α-glucans, and likely bacterial peptidoglycan. It oxidizes thiosulfate to sulfate, and ferments under anoxic conditions. The strain can attach to algae and thrives at low phosphate concentrations as they occur during blooms. Its genome encodes RTX toxin and secretion proteins, and in cultivation experiments Hel1_31_D35 crude cell extracts inhibited growth of a North Sea *Polaribacter* strain. Our data suggest that the combination of these traits make strain Hel1_31_D35 a versatile opportunist that is particularly competitive during spring phytoplankton blooms.

Received 29 March, 2016; revised 9 December, 2016; accepted 10 December, 2016. *For correspondence: E-mail ramann@mpi-bremen.de; Tel. +49 421 2028 930; Fax +49 421 2028 790 and E-mail hteeling@mpi-bremen.de; Tel. +49 421 2028 976; Fax +49 421 2028 790.

Introduction

The carbon cycle constitutes the largest element cycle on Earth. Global photosynthetic carbon fixation exceeds 100 Gt/a, of which marine microalgae and cyanobacteria significantly contribute (Falkowski *et al.*, 1998; Field *et al.*, 1998; Sarmento and Gasol, 2012). Primary production in higher latitudes culminates during massive microalgae blooms. Such blooms are temporary, highly dynamic and terminated as nutrients become limiting and predator grazing and viral infections take effect. Upon bloom termination there is a massive release of algal organic matter that is consumed by distinct clades of heterotrophic bacterioplankton. This trophic connection is reflected in synchronized blooms of bacteria after phytoplankton blooms and has been observed in various studies (e.g. Bell and Kuparinen, 1984; Niu *et al.*, 2011; Tada *et al.*, 2011; Teeling *et al.*, 2012; Tan *et al.*, 2015; Yang *et al.*, 2015; Teeling *et al.*, 2016).

The most prominent bacterial clades that respond to phytoplankton blooms include members of *Bacteroidetes* (class *Flavobacteriia*), *Alphaproteobacteria* (*Roseobacter* clade) and *Gammaproteobacteria* (Buchan *et al.*, 2014). Bloom-associated *Flavobacteriia* are typically specialized in the degradation of high molecular weight (HMW) macromolecules such as polysaccharides and proteins (e.g. Thomas *et al.*, 2011), whereas members of the *Roseobacter* clade have been shown to feature 'opportunitrophic' (generalist) lifestyles where carbon is mostly derived from low molecular weight (LMW) substrates such as the algal osmolyte dimethylsulfoniopropionate, glyoxylate, aromatic compounds and amino acids (e.g. Moran *et al.*, 2007; Newton *et al.*, 2010). In comparison, little is known about

the ecophysiology of bloom-associated *Gammaproteobacteria* (Buchan *et al.*, 2014).

In a previous study we reported on high abundances of the gammaproteobacterial genus *Reinekea* during a diatom-dominated spring phytoplankton bloom in the year 2009 at the North Sea island Helgoland (Teeling *et al.*, 2012). *Reinekea* abundances within the planktonic fraction (0.2–3 μm) increased within five days from below 3×10^3 (<1%) to 1.6×10^5 cells/ml (16%), and subsequently returned to near initial levels within three weeks. In a follow-up study we also found *Reinekea* during subsequent spring blooms from 2010 to 2012 at the same location, but at decreasing abundances of ~5%, ~2% and < 1%, or $\sim 5.5 \times 10^4$, $\sim 2.6 \times 10^4$, and $\sim 4.1 \times 10^3$ cells/ml, respectively (Teeling *et al.*, 2016). V4 16S rRNA gene tag data (Teeling *et al.*, 2016) indicated that *Reinekea* (i) was abundant only during spring phytoplankton blooms, (ii) decreased in abundance from 2009 to 2012 coinciding with declining chlorophyll *a* maxima, and (iii) conspicuously co-occurred with *Thalassiosira nordenskioeldii* (Supporting Information Fig. S1), a cold water adapted diatom species that is frequently found within North Sea phytoplankton in spring. However, the ecophysiology of the North Sea *Reinekea* clade and potential explanations for its recurrence during spring algae blooms remained elusive.

The genus *Reinekea* belongs to *Oceanospirillales*. So far four species have been described: *R. marinisedimentorum* KMM 3655^T from coastal sediments of the Russian Reineke Island in the Sea of Japan (Romanenko *et al.*, 2004), *R. blandensis* MED297^T from seawater of the northwestern Mediterranean (Pinhassi *et al.*, 2007), *R. aestuarii* IMCC4489^T from a Yellow Sea tidal flat (Choi and Cho, 2010) and more recently *R. marina* HME8277^T from surface water of the Yellow Sea (Kang *et al.*, 2016). These species feature curved rod-shaped cells that are motile or non-motile and exhibit chemoheterotrophic, aerobic and facultative anaerobic metabolisms. Physiological tests indicated the ability to utilize a broad range of substrates, including polymers such as starch, chitin and casein. *Reinekea* representatives are also involved in the decomposition of algal and plant biomass. *Reinekea* sp. KIT-YO10, an isolate from the Kanazawa port in Japan, is known to degrade seaweeds and exhibits β-mannanase activity (Hakamada *et al.*, 2014). This strain thus might be able to degrade a mannan-type algal polysaccharide. Likewise, *Reinekea* was detected in biofilms on the leaves of the seagrass *Enhalus acoroides* in Papua New Guinea with 16S rRNA gene tag frequencies of up to 7.2% (Hassenrück *et al.*, 2015). *Reinekea* species have also been found in a wood decomposing microbial community at a marine oak wood fall, where *Reinekea* constituted one of the most abundant clades reaching up to ~10% of the total community (Kalenitchenko *et al.*, 2015).

During the 2010 spring bloom a *Reinekea* strain was isolated at Helgoland (Hel1_31_D35; DSM 29653). Coarse sequencing of a first draft genome and subsequent recruitment of metagenomic reads from 2009 spring bloom bacterioplankton suggested that strain Hel1_31_D35 is representative of the *Reinekea* clade that was abundant during the 2009 bloom (Hahnke *et al.*, 2015). Here we report physiological experiments with *Reinekea* sp. Hel1_31_D35 in which we test hypotheses retrieved from its high-quality, manually annotated complete genome and accompanying *in situ* environmental data. We also further investigated the representativeness of strain Hel1_31_D35 through analyses of (i) full-length 16S rRNA gene sequences from the 2009 spring bloom, (ii) Minimum Entropy Decomposition (MED) of 16S rRNA gene tags from 2010 to 2012 spring blooms, and (iii) targeted sub-assemblies of the *Reinekea* genome from metagenomes obtained during the 2009–2012 spring blooms. We show that *Reinekea* strain Hel1_31_D35 represents a distinct, ecologically relevant bloom-associated species that we propose to name '*Reinekea forsetii*' (for.setti.i. N.L. gen. masc. n. *forsetii*, of Forseti, a god in Frisian mythology that lived on Helgoland, the German island from where the bacterium was isolated).

Results

Genome properties and phylogeny

Strain Hel1_31_D35 has a single circular chromosome (3 660 060 bp, 53% G + C). 3350 genes were predicted, including 3277 CDS, 51 tRNAs for all 20 standard amino acids and five rRNA operons (Fig. 1). Three and two of the 16S rRNA sequences were identical, respectively, with a single nucleotide difference between variants. They share ≤ 96.2% similarity to those from *Reinekea* type strains of validly published species (Fig. 2). Average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) calculations between '*R. forsetii*' Hel1_31_D35 and *R. blandensis* MED297^T supported that this isolate is representative of a new species (ANIm: 83.6%, dDDH: 21%).

Representativeness of strain Hel1_31_D35

For the year 2009 we obtained 75 full-length *Reinekea* 16S rRNA sequences by random colony picking and sequencing of a respective clone library constructed from 0.2 to 3 μm 2009 spring bloom bacterioplankton (Teeling *et al.*, 2012). 74 out of 75 16S rRNA sequences were more than 99% identical (Supporting Information Table S1), suggesting that the *Reinekea* population was dominated by a single species. Consistently, MED analysis of 2010 to 2011 V4 16S rRNA tag data revealed that *Reinekea* tags were dominated by a single oligotype (Supporting Information Fig. S1D). Moreover, 74% of the cluster of sequences represented by this oligotype was identical, suggesting

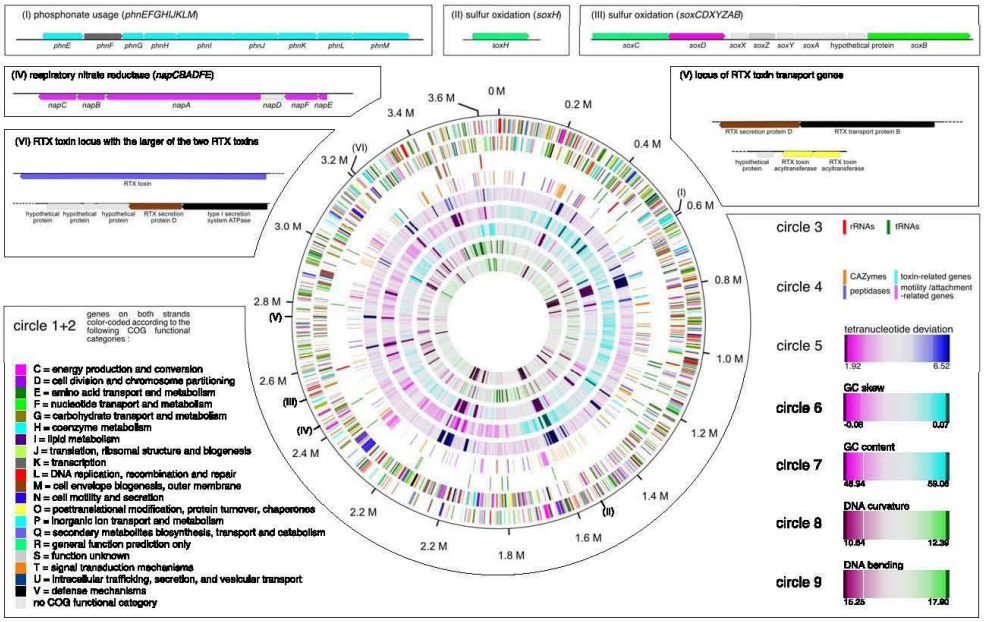


Fig. 1. Atlas of the '*Reinekea forsetii*' Hel1_31_D35 genome. From the outmost circle inwards: (1 + 2): genes on both strands; (3) RNAs; (4) CAZymes, peptidases, toxin-related genes and motility/attachment-related genes; (5) sum of deviations for all 256 tetranucleotide frequency z-scores from respective mean values of the genome (window: 5 kbp); (6) GC skew (window 10 kbp); (7) GC content (window: 10 kbp); (8) DNA curvature (window: 10 kbp); (9) DNA bending (window: 10 kbp). Feature (5) was computed according to Teeling *et al.* (2004). Features (6) and (7) were computed with self-written scripts, and features (8) and (9) with the program btwisted of the EMBOSS package (Rice *et al.*, 2000). (8) is an important aspect for DNA-protein interactions. The dark areas in (5)-(9) represent transposase-rich regions.

that the *Reinekea* population largely consisted of a single species in 2010 and 2011 as well. Seventy-four out of 75 cloned full-length 16S rRNA gene sequences from the 2009 spring bloom exhibited >99% identity to the 16S rRNA genes of '*R. forsetii*' isolate and the dominant 253 bp V4 16S rRNA oligotype (oligotype 3330) from the 2010 and 2011 blooms was even 100% identical to the corresponding '*R. forsetii*' sequences, suggesting that strain Hel1_31_D35 is representative for the highly abundant *Reinekea* from the 2009 to 2011 spring blooms (Supporting Information Table S2). Representativeness of strain Hel1_31_D35 was further investigated by analysis of 18 published metagenomes from Helgoland bacterioplankton (Teeling *et al.*, 2016) (Supporting Information Table S3). At a ≥ 95% identity threshold, the high-quality, complete Hel1_31_D35 genome recruited up to 4.9% of metagenome reads from the peak of the *Reinekea* bloom in 2009, 0.82% for 2010, 0.03% for 2011 and 0.0017% for 2012 (Supporting Information Table S3), corroborating the decreasing *Reinekea* peak abundances observed by CARD-FISH and 16S rRNA gene tag analyses. Metagenome reads covered up to 97.6% and 99.8% of the Hel1_31_D35 genome in 2009 and 2010, respectively, and the sparse reads that were mapped to the 2011 dataset still covered up to 52.7% of the genome (Supporting Information Fig. S1A).

In order to test the representativeness of '*R. forsetii*', we combined the reads which mapped onto the Hel1_31_D35 genome from the four metagenomes with the highest *Reinekea* spp. representation (Supporting Information Table S3) and reads of metagenome contigs predicted to belong to the *Reinekea* genus by taxonomic classification (Supporting Information Table S4). We then performed *de novo* assemblies and assessed completeness of genome recovery from the metagenomes. An assembly of all combined reads allowed the almost complete retrieval of the '*R. forsetii*' Hel1_31_D35 genome (Fig. 3A). This was also possible for three individual of the four metagenomes (Fig. 3B–D), while the fourth, the one with the highest recruited read number, had overlapping contigs, indicating strain heterogeneity (Fig. 3E), but still yielded assemblies that were highly collinear to the Hel1_31_D35 genome. All assemblies exhibited more than 99% ANI with the Hel1_31_D35 genome.

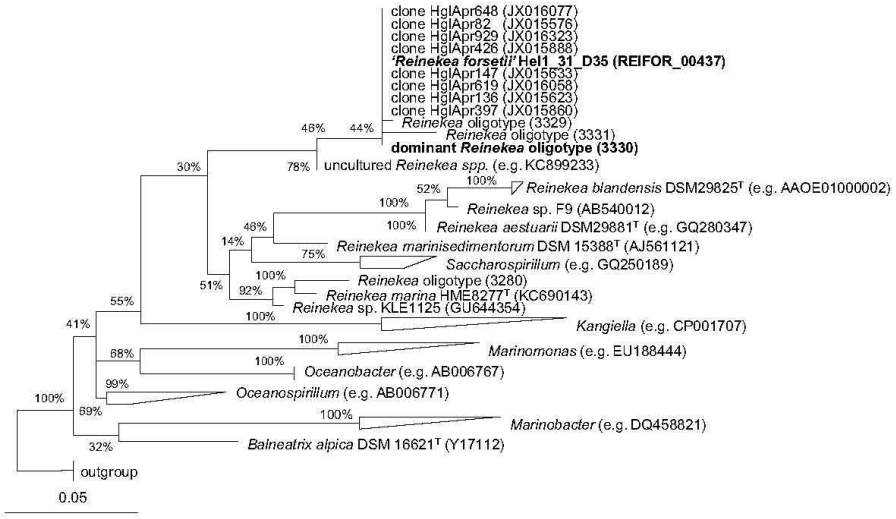


Fig. 2. Phylogenetic tree of cloned *Reinekea* full-length 16S rRNA gene sequences obtained during the 2009 spring bloom at Helgoland, '*R. forsetii* Hel1_31_D35 16S rRNA gene sequence and *Reinekea* oligotype sequences. The tree was computed with the RAxML (Stamatakis and Ott, 2008) maximum likelihood method (GTR gamma substitution model, 1 000 repetitions) using a 50% base conservation filter. *Alphaproteobacteria* were used as outgroup. Bar: 0.05 substitutions per nucleotide position. *Saccharospirillum* interspersed with *Reinekea* sequences as clade with low bootstrap support, which was also the case when using the maximum parsimony method and other positional conservation filter thresholds (40% and 45%). This does not affect the phylogenetic affiliation of strain Hel1_31_D35, but it indicates that the commonly accepted phylogeny in this area might need revision. An alternative neighbor-joining tree did resolve the currently accepted phylogeny and is provided in Supporting Information Fig. S7.

Physiology

Energy metabolism. Strain Hel1_31_D35 is a facultative anaerobe and grows between 4 and 19°C with an optimum at 12°C. Its genome harbors the complete Embden-Meyerhof-Parnas (EMP) and pentose-5-phosphate (PP) pathways. For pyruvate oxidation both components of a pyruvate dehydrogenase complex and a pyruvate formate lyase were found. Strain Hel1_31_D35 has a complete TCA cycle with glyoxylate shunt. Annotations revealed a redox chain for oxygen respiration including a NAD(H):ubiquinone oxidoreductase (complex I), succinate dehydrogenase (complex II), cytochrome cbb_3 and aa_3 type (complex IV) terminal oxidases and a F1F0-type ATPase, but no genes for complex III (cytochrome bc1) and d-type cytochromes. The genome also encodes a *napCBADFE* respiratory nitrate reductase (Fig. 1), though nitrate reduction was not confirmed with the API 20 NE quick test.

Regarding sulfur metabolism, the Hel1_31_D35 genome features the sulfur oxidation genes *soxH-CDXYZA-B* (Fig. 1) and a sulfide-quinone reductase (*sqr*) gene. Consistent with this finding, strain Hel1_31_D35 could oxidize thiosulfate to sulfate in a mixotrophic growth experiment under oxic conditions (Supporting Information Fig. S2).

Under anoxic conditions, strain Hel1_31_D35 fermented D-fructose and yeast extract to succinate, acetate, propionate, lactate and formate, but did not ferment D-galactose and N-acetyl-D-glucosamine. The Hel1_31_D35 genome also harbors a gene for the light-driven proton pump proteorhodopsin (PR) that might be used as a supplemental energy source. In a phylogenetic tree reconstructed from PR amino acid sequences of *Gammaproteobacteria*, the Hel1_31_D35 PR was most closely related to the partial sequence of the SAR92 strain HTCC2207 (81% amino acid similarity; Supporting Information Fig. S3). Energy gain from PR has been assumed for this strain, but promotion of growth by light has not yet been proven (Stingl *et al.*, 2007).

Carbohydrate utilization. Strain Hel1_31_D35 can use a wide spectrum of mono-, di- and trisaccharides (Table 1). Annotations also suggest usage of D-ribose, but this was not tested. In contrast, growth on polysaccharides was only observed with glycogen, but not with agar, κ- or ι-carrageenan, cellulose, laminarin or xylan. The strain could hydrolyze galactomannan (endo-1,4-β-mannanase activity), xanthan and amylose, but not arabinoxylan, cellulose, galactan, pectin or xylan (Table 1).

Algae use polysaccharides as storage compounds as well as components in their cell matrices, cell walls and exudates. The polysaccharide content of algae can vary

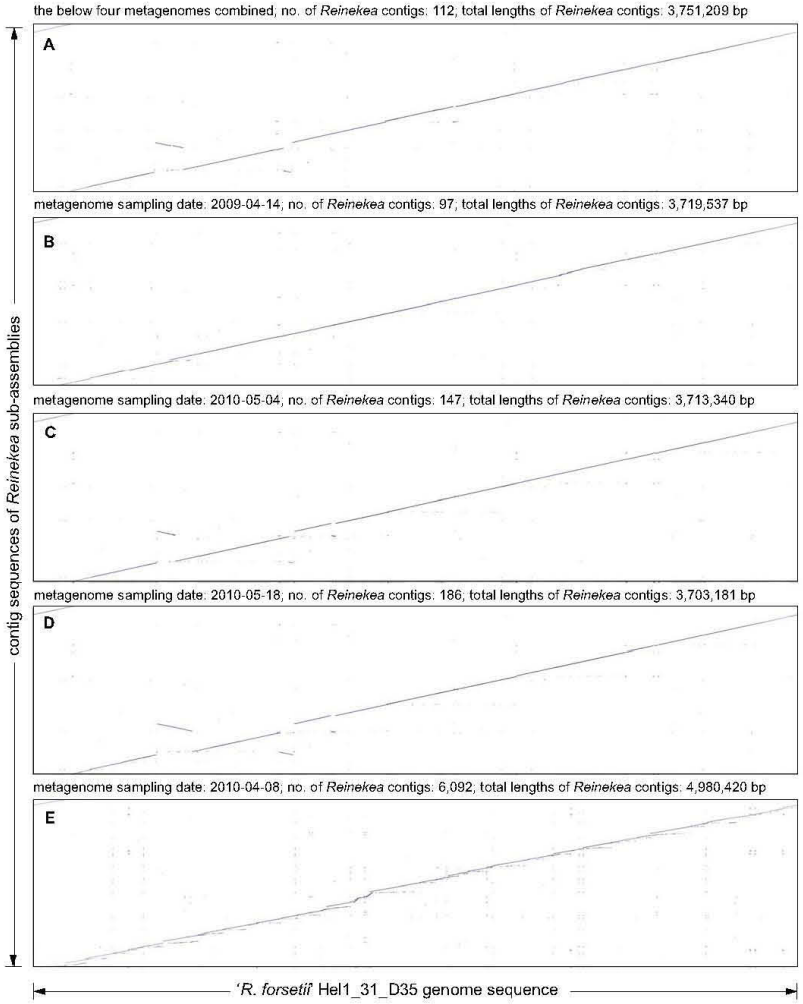


Fig. 3. Colinearity plots of *Reinekea* spp. sub-assemblies from metagenome data (ordinate) and the *'R. forsetii'* Hel1_31_D35 genome (abscissa). Sub-assemblies were computed using both the metagenome reads which mapped onto the Hel1_31_D35 genome and reads of metagenome contigs taxonomically classified as the genus *Reinekea*. Sampling dates of corresponding metagenomes are shown above the panels together with total contig numbers and sizes of each sub-assembly. For the sub-assembly in panel (A), all of the above mentioned reads from four metagenomes were combined. Retrieval of almost the entire *'R. forsetii'* Hel1_31_D35 genome was possible from three of the metagenomes. The metagenome of 2010-04-08 with the highest coverage retrieved more heterogeneity and yielded a larger number of overlapping contigs. Still also these contigs aligned almost perfectly to the *'R. forsetii'* Hel1_31_D35 genome.

widely and depends on their growth state. In extreme cases polysaccharides can reach up to ~90% of the cell dry weight (Myklestad, 1974). Consequently, phytoplankton blooms lead to the release of large amounts of algal polysaccharides. Decomposition of these polysaccharides requires carbohydrate-binding modules (CBMs) (see e.g. Carvalho *et al.*, 2015) that mediate polysaccharide recognition and binding, in concert with glycoside hydrolases (GHs) that subsequently hydrolyze glycosidic bonds. The Hel1_31_D35 genome harbors 26 CBMs from 14 families with highest frequencies of CBM48 and CBM50, and 44 GHs from 15 families with families GH13, GH23, GH26,

Table 1. Substrate spectrum of '*R. forsetii*' Hel1_31_D35 as assessed by (a) growth experiments, AZO-CL substrate assays in (b) HaHa medium and (c) marine broth 2216, and (d) the API 20 NE test system. (–) negative; (w) weak; (+) positive; (++) strongly positive.

Utilization of	a	b	c	d
Amino acids				
Casamino acids	+			
Casein		++	++	
Gelatin	+	w	–	w
Peptone tryptone	+			
Yeast extract	+			
Monosaccharides and sugar alcohols				
D-arabinose	+			
L-arabinose	+			+
D-fructose	+			
D-galactose	+			
D-gluconate				+
D-glucose	+			+
D-mannitol	+			+
D-mannose	+			+
L-rhamnose	+			
DL-xylose	+			
Disaccharides				
D-cellobiose	+			
D-maltose	+			+
D-sucrose	+			
D-trehalose	+			
Trisaccharides				
L-raffinose	+			
Aminosugars				
N-acetyl-D-glucosamine	++			+
N-acetylneuraminic acid	++			
Polysaccharides				
Agar	–			
Amylose		w	w	
Arabinoxylan		–	–	
ι-carrageenan	–			
κ-carrageenan	–			
Cellulose (Avicel)		–	–	
Cellulose (carboxymethyl cellulose)	–			
Cellulose (filter paper)	–			
Cellulose (hydroxyethyl cellulose)		–	–	
Galactan		–	–	
Galactomannan		+	+	
Glycogen	+			
Laminarin	–			
Pectin		–	–	
Xanthan	w			
Xylan	–	–	–	
Mono-, di- and tricarboxylic acids				
Adipic acid				+
Citrate	+			+
Malic acid	+			+
Phenylacetic acid				+
Fatty acids				
Capric acid				+

GH73 and GH109 showing highest abundance (Supporting Information Table S5). GH26 enzymes act predominantly on mannans and a few other substrates, e.g. lichenan (Taylor *et al.*, 2005). Members of the GH13 family can utilize α-glucans such as glycogen, and CBM48 domains are known to have a glycogen-binding function. Families GH23 and GH73 comprise predominantly peptidoglycan lyases, while CBM50 domains predominantly bind chitin and peptidoglycan. Presence of these carbohydrate-active enzymes (CAZymes) is consistent with observed growth on galactomannan, glycogen and peptidoglycan monomers.

R. aestuarii can hydrolyze chitin (Choi and Cho, 2010), whereas '*R. forsetii*' Hel1_31_D35 does not possess a chitinase gene. A sole GH33 family gene encodes a neuraminidase (endo-α-sialidase) that likely participates in the removal of N-acetyl-D-neuraminic acid from extracellular glycoproteins, gangliosides and sialic acids that typically occur in eukaryotes. Interestingly, the genome harbors not only genes for the degradation of N-acetyl-D-neuraminic acid, but also some genes required for its biosynthesis (Supporting Information Fig. S4), which might be involved in capsule formation (Roberts, 1996). Furthermore, sialic acids have been recently detected in polysaccharide-rich capsules surrounding members of *Polaribacter*, *Ulvibacter* and *Formosa* of the 2011 North Sea spring bloom (Bennke *et al.*, 2013).

We compared the GH and CBM repertoires of 21 related marine *Gammaproteobacteria* genomes. GH26 family genes used in mannan utilization were present only in strain Hel1_31_D35 (2 genes, Supporting Information Table S5), and GH13 genes involved in α-glucan utilization showed higher abundance in '*R. forsetii*' Hel1_31_D35 and *R. blandensis* MED297^T (both 14 genes) compared to other genomes (0 to 7 genes). No differences between compared genomes were observed with respect to CBM50, GH23 and GH73 frequencies (Supporting Information Table S6).

Protein utilization. Strain Hel1_31_D35 hydrolyzed casein and grew on casamino acids, gelatin, peptone tryptone and yeast extract (Table 1). Its genome harbors 94 peptidases, the majority of which are metalloproteinases with M16, M23, M24 and M48 being the most frequent families. M23 family members have been shown to partake in extracellular degradation of bacterial peptidoglycan (Baba and Schneewind, 1996; Bamford *et al.*, 2010).

Nitrogen, sulfur and phosphorus assimilation. The Hel1_31_D35 genome has assimilatory nitrate reductase (*nas*) and nitrite reductase (*nir*) genes. It also codes for assimilatory sulfate reduction proteins such as sulfate adenylyltransferase, adenylylsulfate kinase, phosphoadenylylsulfate reductase, sulfite reductase and cysteine synthetase (CysK).

Genes for phosphate uptake (*phoU-R* and *pstSCAB*), polyphosphate synthesis and hydrolysis (*ppK and ppX*) and the associated transcriptional regulator (*phoB*) were annotated. Polyphosphate inclusions were also detected in

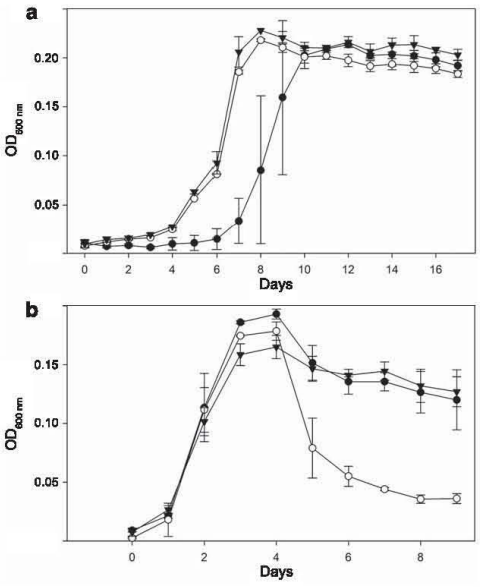


Fig. 4. Growth of *Formosa* sp. Hel1_33_131 (a) and *Polaribacter* sp. Hel1_88 (b) in HaHa_100V medium without (•) and with addition of lysed (○) or lysed and heat-treated (▼) '*R. forsetii*' Hel1_31_D35 crude cell extracts. Error bars refer to biological duplicates.

Hel1_31_D35 cells by methylene blue and DAPI staining (Supporting Information Fig. S5). Additionally, the genome also encodes a large *phn* gene cluster for the usage of phosphonates (Fig. 1).

Transporters. Strain Hel1_31_D35 features a broad range of ABC transporters for high-affinity uptake of LMW substrates. 180 genes were assigned to ABC transporter components, with the highest abundances of genes coding for phosphonates, nucleosides, glycerol-3-phosphate, ribose and N-acetyl-D-glucosamine transport. Other substrates included inorganic iron, nitrate, zinc and molybdenum, the organic compounds fructose, maltose/maltodextrin, glycine betaine, oligopeptides and arginine/ornithine and vitamins B_1 and B_{12}. The latter is consistent with the observation that strain Hel1_31_D35 is vitamin auxotrophic (Hahnke *et al.*, 2015). The genome also features 31 genes belonging to ATP-independent periplasmic (TRAP) transporters, five genes for phosphotransferase system components (at least one fructose PTS) and one Tol-type biopolymer transporter system.

Toxins. The Hel1_31_D35 genome encodes a smaller (REIFOR_00579; 1285 aa) and a larger (REIFOR_02945; 2130 aa; Fig. 1) RTX family toxin, and respective RTX modification and excretion genes (e.g. *tolC*). Likewise, genes for a type VI secretion system were found. Some bacteria employ such systems to prey on other bacteria and scavenge their DNA (Borgeaud *et al.*, 2015). We tested for toxicity of Hel1_31_D35 crude cell extracts on cultures of North Sea representatives of the genera *Formosa* and *Polaribacter* (Hahnke *et al.*, 2015). Both genera were highly abundant during the *Reinekea* bloom in 2009 (Teeling *et al.*, 2012). No effect on the optical density was found for *Formosa* sp. Hel1_33_131 cultures (Fig. 4), but for *Polaribacter* sp. Hel1_88, a sharp decrease in optical density was detected after the cells entered stationary phase and thus energy limitation (Fig. 4). It is expected that the toxic effect of a membrane-decoupling agent such as an RTX family porin takes effect once the targeted cells become energy limited. By contrast, heat-treated *Reinekea* cell extracts showed no such effect.

Motility and attachment. Strain Hel1_31_D35 is motile by a sole flagellum (Hahnke *et al.*, 2015). The respective genes are clustered and comprise flagellar biosynthesis and movement (*flh*, *flg*, *fli*, *fle*, *fla*, *mot*), pilus and fimbrial biogenesis (*pil*) and chemotaxis and twitching motility. Furthermore, a cluster of mannose-sensitive haemagglutinin (MSHA)-like pilus genes was found (*mshACDOPQ-I_2JK_3LMNEG*). Such pili are known to promote bacterial attachment to surfaces of the green alga *Ulva australis* (Dalisay *et al.*, 2006) and various phyto- and zooplankton (Chiavelli *et al.*, 2001). The presence of motility and attachment genes suggests that strain Hel1_31_D35 features both free-living and algae-associated lifestyles. *Reinekea* cells could indeed be detected on particles collected by fractionated filtration in the size range of 3–10 μm, although they were much more abundant in the 0.2–3 μm fraction (Supporting Information Fig. S6).

Discussion

Just as the 16S rRNA approach to bacterial taxonomy has opened a new window for the discovery of yet uncultured *Bacteria* and *Archaea* in the 1990s, metagenomics is today generating a plethora of hypotheses on the physiological potential and ecological role of microbial populations. Not surprisingly, the testing of these hypotheses in the laboratory is lagging behind due to limitations intrinsic to the enrichment and isolation of representative pure cultures. This paper is intended to provide a small step towards ameliorating this problem by a (meta)genome inspired ecophysiological characterization of a recently isolated strain of the marine gammaproteobacterial clade *Reinekea*.

It was postulated that *Reinekea* acted as an effective scavenger of LMW compounds within a swift substrate-based succession of bacterioplankton clades during the 2009 North Sea spring phytoplankton bloom (Teeling *et al.*, 2012). In a subsequent publication on the recurrence of bacterial clades during North Sea spring algae blooms

(Teeling *et al.*, 2016) comparative metagenomics indicated *Reinekea* had a limited glycan niche involving the decomposition of external α-1,4-glucans and possibly peptidoglycan. We here provide multiple lines of evidence supporting that *'R. forsetii'* strain Hel1_31_D35 constitutes a valid representative of the *Reinekea* clade peaking during North Sea spring algae blooms in three subsequent years. This included almost complete genome retrieval from metagenomes of 2009 and 2010 (Fig. 3), almost identical full-length 16S rRNA gene sequences (Supporting Information Table S2) and a single dominating 16S rRNA oligotype detected by minimum entropy decomposition (Supporting Information Fig. S1). We therefore performed a detailed ecophysiological study of *'R. forsetii'* strain Hel1_31_D35 by testing hypotheses retrieved both from a manually annotated, high-quality closed genome and *in situ* environmental data.

In temperate regions with strong seasonality, temperature adaptation is likely an important niche-determining trait which can only be determined on representative pure cultures. Reflecting its restricted occurrence in spring phytoplankton blooms, Hel1_31_D35 was shown to be a psychrophile with a temperature optimum of 12°C and range of growth temperatures from 4 to 19°C. The co-occurrence with the cold-water adapted diatom *T. nordenskioldii* might therefore indicate a similar temperature preference, but not constitute a direct biotic interaction.

Growth experiments showed that strain Hel1_31_D35 is capable of utilizing a wide array of substrates including mono- and disaccharides, the polysaccharides mannan and α-glucans, as well as proteins and amino acids (Table 1). Growth on both N-acetyl-D-glucosamine and N-acetylneuraminic acid suggested that peptidoglycan is likely degraded, since it is composed of these two aminosugars. These results are consistent with a nutritional strategy relying strongly on LMW compounds, as suggested in previous studies.

Strain Hel1_31_D35 did not grow on complex polysaccharides such as laminarin, xylan, various types of cellulose, carrageenan and agar. It was also evident from genome annotations that genetic equipment for the degradation of these typical algal or plant polysaccharides is missing. A notable exception was the detection of GH26 suggesting decomposition of mannan, a polysaccharide consisting of mannose units (Supporting Information Table S5). Interestingly, GH26 genes were not found in the 20 other marine *Gammaproteobacteria* genomes that are closely related to *Reinekea* (Supporting Information Table S6). Strain Hel1_31_D35 indeed grew on galactomannan composed of a mannose backbone with galactose side chains. Both monosaccharides supported growth as well (Table 1). In diatoms mannans are found in close association with the silicified cell wall and may play a role in biomineralization (Chiovitti *et al.*, 2005). Bloom intensities and diatom *T. nordenskioeldii* abundances successively declined during the 2009 to 2012 study period and coincided with dwindling *Reinekea* abundances (Supporting Information Fig. S1). Altogether, these results indicated that in addition to degradation of LMW compounds, mannan utilization could create another window of opportunity for '*R. forsetii*' during spring algae blooms.

While strain Hel1_31_D35 did not use algal β-glucans such as callose and chrysolaminarin, it grew on α-glucans such as glycogen, and on amylose, albeit weakly (Table 1). The '*R. forsetii*' genome was indeed rich in genes for GH13 family proteins, which catalyze the degradation of α-glucans (Supporting Information Table S5). Such α-glucans can for example be found in the exudate slimes of some *Flavobacteriia* (R. L. Hahnke, personal communication). Thus it could be that bloom-associated *Flavobacteriia* that feed on algal β-glucans such as laminarin dispose excess glucose as α-glucan-rich exudates that in turn are consumed by opportunistic bacteria such as *Reinekea*. This, however, remains to be tested. It is also possible that the acquisition of substrates from both algae and bacteria by strain Hel1_31_D35 is an active process, as suggested by presence of RTX toxin genes and the inhibitory effect of Hel1_31_D35 crude cell extracts on stationary phase *Polaribacter* sp. Hel1_88 cultures (Fig. 4). Furthermore, during the 2009 bloom, *Polaribacter* abundances started to decline when the *Reinekea* abundances started to rise (Teeling *et al.*, 2012), which suggests that such an antagonistic interaction might also be present during the blooms *in situ*. In this respect, it should be also noted that '*R. forsetii*' was shown to grow on peptidoglycan monomers, which are released during bacterioplankton successions when blooming bacterial clades are supplanted by new clades and bacteria disintegrate.

Mannose-sensitive haemagglutinin (MSHA)-like pilus genes and lectin staining (Supporting Information Fig. S6) suggested that '*R. forsetii*' has a potential to thrive on algal-derived particles, which improves access to substrates. Yet a downside of particle-attachment could be oxygen-depletion either within marine aggregates or by sedimentation to the sea floor. We annotated both nitrate respiration and fermentation genes; however we could only verify fermentation in the pure culture studies. The niche space of '*R. forsetii*' is further enlarged by the oxidation of reduced sulfur compounds, which was predicted in the genome and also demonstrated experimentally (Supporting Information Fig. S2). Strain Hel_31_D35 is thus well equipped to deal with anoxic conditions occurring on algal-derived particles, potentially gaining another competitive advantage over other scavengers of LMW compounds.

Phosphate may also play a critical role in niche differentiation, as it is quickly consumed during phytoplankton blooms and was below detection limit by the time *Reinekea* peaked in abundance in 2009 (Teeling *et al.*, 2012). Strain

Hel1_31_D35 possesses genes for polyphosphate synthesis and hydrolysis. We also microscopically detected polyphosphate inclusions in '*R. forsetii*' cells (Supporting Information Fig. S5). Furthermore, genes encoding phosphonate utilization were annotated (Fig. 1). Phosphonate is an alternative phosphorus source for abundant marine taxa such as SAR11 (Sowell *et al.*, 2009) and *Roseobacter* (Moran *et al.*, 2007) and encompasses a non-negligible fraction of dissolved organic phosphorus in the oceans (Clark Kolowith *et al.*, 2001). Thus, any potential to access this secondary pool could bolster the selective advantage of '*R. forsetii*', which seems to be well adapted to phosphate limitations typical of algal blooms.

Conclusions

We are just beginning to understand the ecological strategies of distinct bacterioplankton species during phytoplankton blooms. Our analyses of '*R. forsetii*' Hel1_31_D35 did not uncover a single specific trait that could explain its success during North Sea spring phytoplankton blooms, but rather a combination of multiple traits such as temperature adaptation, ability to use a large substrate spectrum encompassing many LMW and few HMW substrates, toxin production, and strategies to deal with phosphate limitation. This demonstrates that our capability to understand the ecology of bacterial clades can be greatly enhanced by laboratory studies of cultivable representatives. In this sense, the central microbiological tool of cultivation is not supplanted by modern 'omics, but necessary as ever as a prerequisite for the testing of hypotheses deduced from large-scale 'omics data. Further isolation of strains of abundant bloom-associated clades should provide a more complete picture regarding their distinct niches, and might begin to unravel the mutual interdependencies during blooms that determine the limited windows of opportunity in which these strains thrive.

Experimental procedures

Physiology

Hel1_31_D35 cells were cultivated in artificial seawater medium with vitamins, and micro-molar substrate and nutrient concentrations, as described previously (Hahnke *et al.*, 2015). Oxic and anoxic growth on different carbon sources was tested in HaHa medium with vitamins at 12°C. Fermentation products were identified by HPLC as described in Zhilina *et al.* (2004). Gelatin hydrolysis was tested using charcoal-gelatin discs as in chapter 15.3.32.3, method 3, of Tindall *et al.* (2007), and hydrolysis of additional substrates was tested with AZO-CL-substrates (Megazyme, Wicklow, Ireland) embedded in an autoclaved matrix of artificial seawater with 4% alginate in marine broth 2216 or HaHa medium with vitamins at 15°C for 14 days. The API 20 NE system (API bioMérieu, Marcy-l'Etoile, France) was used for further substrate tests with suspensions of cells in the provided medium supplemented with vitamins and 2.5% (v/v) sea salts (Biomaris seawater, Biomaris, Bremen, Germany) at 15°C for 2 to 14 days. All results are summarized in Table 1.

Toxicity tests

A Hel1_31_D35 culture (OD_{600} ~0.3) was lysed using a french press (SIMO-Aminco Industry, Rochester, NY) with 70 bar twice, and subsequently filtered through a 0.22 μm sterile polyethersulfone filter (Merk Millipore, Cork, Ireland). One ml aliquots of this cell extract were incubated with 0.5 ml of either *Polaribacter* sp. Hel1_88 or *Formosa* sp. Hel1_33_131 cultures (both in stationary phase) in 8.5 ml HaHa_100V medium. As controls, these experiments were also conducted with heat-inactivated (80°C, 10 min) and filtered Hel1_31_D35 cell extracts. Bacterial growth was monitored (OD_{600}) using an UVmini 1240 UV-Vis spectrophotometer (Shimadzu Scientific Instruments, Tokyo, Japan).

Thiosulfate oxidation

A sulfate-free HaHa_100V medium was prepared by replacing $MgSO_4$ x 7 H_2O from the original recipe (Hahnke *et al.*, 2015) with $MgCl_2$ × 6 H_2O. Hel1_31_D35 cultures were grown in this medium with and without 10 mM $Na_2S_2O_3$ × 5 H_2O. Growth was monitored as described above. Sulfate production was measured using a Metrohm 761 Compact IC equipped with a conductivity detector (Metrohm AG, Herisau, Switzerland) with a carbonate eluent (3.2 mM Na_2CO_3/1 mM $NaHCO_3$ in deionized water) at a flow rate of 0.7 ml min^{-1}.

Genome sequencing and assembly

DNA was extracted from a Hel1_31_D35 cell culture according to the protocol of Zhou *et al.* (1996). One SMRTcell was sequenced on the PacBio RS II platform, yielding 27 850 reads (N50: 9976 bp) that were assembled using the PacBio HGAP pipeline into a single contig of 3 670 008 bp. In addition, a paired-end library (insert size 432 ±72 bp) was sequenced on the Illumina MiSeq platform. The raw reads were filtered to a minimum Phred quality of two and all reads containing ambiguous bases were removed. In total 2 456 271 read pairs were mapped onto the PacBio assembly using bowtie2 (Langmead and Salzberg, 2012). A consensus was generated using Pilon (Walker *et al.*, 2014). Manual editing resulted in a high-quality circular genome of 3 660 060 bp.

Genome annotation and metabolic reconstruction

Gene predictions and annotations were performed as described previously (Mann *et al.*, 2013). 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.25 database (Finn *et al.*, 2010) using E-values derived from manual annotations of test data (Supporting Information Table S7). CAZymes were only annotated when at least two of the three database searches yielded positive results. The annotated genome sequence was deposited at NCBI's Genbank (bioproject PRJNA283582).

Metagenome read recruitment, taxonomic classification and sub-assemblies

The Hel1_31_D35 genome was used to recruit reads (minimum identity: 95%) from 18 metagenomes from North Sea bacterioplankton as described previously (Xing *et al.*, 2015). These metagenomes were obtained from samples taken from 2008 to 2012 at the buoy 'Kabeltonne' at the island Helgoland (Helgoland roads, North Sea, Germany, 54° 11.3' N, 7° 54.0 'E) as part of the Microbial Interactions in Marine Systems (MIMAS) project (Teeling *et al.*, 2012) and the Coastal Microbe Taxonomic & Genomic Observatory (COGITO) community sequencing project of the U.S. Department of Energy (DOE) Joint Genome Institute (JGI, Walnut Creek, CA, USA). Metagenomes were taxonomically classified as described previously (Teeling *et al.*, 2016). Sub-assemblies were computed from metagenomes using SPAdes v.3.1.1 (Bankevich *et al.*, 2012) with kmers 21, 33, 55 and 77 with recruited reads plus reads of contigs classified as belonging to the *Reinekea* genus (Supporting Information Table S3).

16S rRNA clone libraries and phylogenetic analysis

Reinekea spp. 16S rRNA sequences were obtained from a clone library constructed from North Sea spring bloom bacterioplankton (0.2–3 μm) sampled at a depth of 1m on April 14th, 2009 (Supporting Information Table S2). In total 822 clones were sequenced. Sequences >1 500 bp and those from strain Hel1_31_D35 were aligned to the 16S rRNA sequences in the SILVA SSURef v.119 dataset (Quast *et al.*, 2013) with the Silva Incremental Aligner (Pruesse *et al.*, 2012). ARB v. 5.5 (Ludwig *et al.*, 2004) was used for manual curation of the alignment. Phylogenetic tree reconstruction (Fig. 2) was done with the RAxML (Stamatakis and Ott, 2008) maximum likelihood method (GTR gamma substitutions model, 1 000 repetitions) and the Neighbor-Joining method (Jukes-Cantor substitution model, 1 000 repetitions) with a *Gammaproteobacteria* 50% positional conservation filter. Partial *Reinekea* oligotype sequences were subsequently added by parsimony using the same filter. Average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) were calculated using JSpecies (Richter and Rossello-Mora, 2009) and the Genome-to-Genome Distance Calculator (Auch *et al.*, 2010) respectively.

16S rRNA gene tag sequencing and oligotyping analysis

Sample collection, DNA extraction and sequencing of V4 16S rRNA gene tags were performed as described previously (Lucas *et al.*, 2015; Teeling *et al.*, 2016). In total 142 surface seawater samples were collected on bi-monthly to bi-weekly time scales from 2010 to 2012 at Helgoland roads for tag sequencing. In addition, two samples from the 2009 spring bloom were selected to complement the dataset. 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).

Raw paired-end reads were merged and filtered using scripts from illumina-utils (https://github.com/meren/illumina-utils) to retain only sequences without mismatches in the overlapping region. These reads were processed using the MED part of the oligotyping package (https://github.com/meren/oligotyping). MED uses Shannon entropy to identify information-rich positions along an alignment without implementation of pairwise sequence alignments and arbitrary global similarity thresholds (i.e. 97%) used in traditional OTU clustering methods (Eren *et al.*, 2015). A minimum substantive abundance (-M) of 100 was used to filter low-abundant oligotypes and was decomposed one nucleotide position at a time (-d 1). Representative node sequences were classified using the SILVAngs pipeline (Quast *et al.*, 2013) against the SILVA v123 database (BLAST threshold = (sequence identity + alignment coverage)/2 ≥ 93%). For this study, only oligotypes classified as belonging to the *Reinekea* genus of greater than 0.1% abundance were used.

CARD-FISH, polyphosphate and lectin staining

CARD-FISH analyses were performed as described previously (Thiele *et al.*, 2011) using probe REI731 (Teeling *et al.*, 2012). For staining of polyphosphate inclusions in cultured *Reinekea*, cells were fixed by short heat treatment on a glass slide and incubated for 10 min in Loeffler methylene blue solution (0.5 g methylene blue chloride, 100 ml distilled water, 30 ml ethanol (96% v/v), 1 ml potassium hydroxide (1% w/v)). DAPI staining of polyphosphate inclusions was carried out as described in Romano *et al.* (2015). Lectin staining was performed as described by Bennke *et al.* (2013). *Aleuria aurantia* type lectin (Vector Laboratories, Burlingame, Canada) which has a specific binding to α-fucose was used. Samples were examined by a LSM 780 with super resolution system ELYRA PS.1 (Zeiss, Oberkochen, Germany).

Acknowledgements

We thank Lennart Kappelmann and Karen Krüger for support with bioinformatics, Greta Reintjes for polyphosphate staining, Mariette Kasabgy, Yilung Chen and Dirk Brekelmann for CARD-FISH, and Irene Panschin for physiological tests. Richard L. Hahnke was supported by the German Federal Ministry of Food and Agriculture (no. 22016812), and the work by the U.S. DOE JGI was supported under contract no. DE-AC02-05CH11231. This study was funded by the Max Planck Society.

References

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.

Auch, A.F., von Jan, M., Klenk, H.P., and Goker, M. (2010) Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* **2:** 117–134.

Baba, T., and Schneewind, O. (1996) Target cell specificity of a bacteriocin molecule: a C-terminal signal directs lysostaphin to the cell wall of *Staphylococcus aureus*. *Embo J* **15:** 4789–4797.

Bamford, C.V., Francescutti, T., Cameron, C.E., Jenkinson, H.F., and Dymock, D. (2010) Characterization of a novel family of

fibronectin-binding proteins with M23 peptidase domains from *Treponema denticola*. *Mol Oral Microbiol* **25:** 369–383.
Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19:** 455–477.
Bell, R.T., and Kuparinen, J. (1984) Assessing phytoplankton and bacterioplankton production during early spring in Lake Erken, Sweden. *Appl Environ Microbiol* **48:** 1221–1230.
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.
Borgeaud, S., Metzger, L.C., Scrignari, T., and Blokesch, M. (2015) The type VI secretion system of *Vibrio cholerae* fosters horizontal gene transfer. *Science* **347:** 63–67.
Buchan, A., LeCleir, G.R., Gulvik, C.A., and Gonzalez, J.M. (2014) Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol* **12:** 686–698.
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.
Carvalho, C.C., Phan, N.N., Chen, Y., and Reilly, P.J. (2015) Carbohydrate-binding module tribes. *Biopolymers* **103:** 203–214.
Chiavelli, D.A., Marsh, J.W., and Taylor, R.K. (2001) The mannose-sensitive hemagglutinin of *Vibrio cholerae* promotes adherence to zooplankton. *Appl Environ Microbiol* **67:** 3220–3225.
Chiovitti, A., Harper, R.E., Willis, A., Bacic, A., Mulvaney, P., and Wetherbee, R. (2005) Variations in the substituted 3-linked mannans closely associated with the silicified walls of diatoms. *J. Phycol* **41:** 1154–1161.
Choi, A., and Cho, J.C. (2010) *Reinekea aestuarii* sp. nov., isolated from tidal flat sediment. *Int J Syst Evol Microbiol* **60:** 2813–2817.
Clark Kolowith, L., Ingall, E.D., and Benner, R. (2001) Composition and cycling of marine organic phosphorus. *Limnol Oceanogr* **46:** 309–320.
Dalisay, D.S., Webb, J.S., Scheffel, A., Svenson, C., James, S., Holmström, C., *et al.* (2006) A mannose-sensitive haemagglutinin (MSHA)-like pilus promotes attachment of *Pseudoalteromonas tunicata* cells to the surface of the green alga Ulva australis. *Microbiology* **152:** 2875–2883.
Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud J., Vineis, J.H., and Sogin, M.L. (2015) Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J* **9:** 968–979.
Falkowski, P.G., Barber, R.T., and Smetacek, V.V. (1998) Biogeochemical controls and feedbacks on ocean primary production. *Science* **281:** 200–207.
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–D222.
Hahnke, R.L., Bennke, C.M., Fuchs, B.M., Mann, A.J., Rhiel, E., Teeling, H., *et al.* (2015) Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. *Environ Microbiol* **17:** 3515–3526.
Hakamada, Y., Ohkubo, Y., and Ohashi, S. (2014) Purification and characterization of β-Mannanase from *Reinekea* sp. KIT-YO10 with transglycosylation activity. *Biosci, Biotechnol, Biochem* **78:** 722–728.
Hassenrück, C., Hofmann, L.C., Bischof, K., and Ramette, A. (2015) Seagrass biofilm communities at a naturally CO_2-rich vent. *Environ Microbiol Rep* **7:** 516–525.
Kalenitchenko, D., Fagervold, S.K., Pruski, A.M., Vétion, G., Yücel, M., Le Bris, N., and Galand, P.E. (2015) Temporal and spatial constraints on community assembly during microbial colonization of wood in seawater. *ISME J* **9:** 2657–2670.
Kang, H., Kim, H., Joung, Y., and Joh, K. (2016) *Reinekea marina* sp. nov., isolated from seawater, and emended description of the genus *Reinekea*. *Int J Syst Evol Microbiol* **66:** 360–364.
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.
Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.
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* **42:** D490–D495.
Lucas, J., Wichels, A., Teeling, H., Chafee, M., Scharfe, M., and Gerdts, G. (2015) Annual dynamics of North Sea bacterioplankton: seasonal variability superimposes short-term variation. *FEMS Microbiol Ecol* **91:** fiv099. doi: 10.1093/femsec/fiv099.
Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar., *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32:** 1363–1371.
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.
Moran, M.A., Belas, R., Schell, M.A., Gonzalez, J.M., Sun, F., Sun, S., *et al.* (2007) Ecological genomics of marine Roseobacters. *Appl Environ Microbiol* **73:** 4559–4569.
Myklestad, S. (1974) Production of carbohydrates by marine planktonic diatoms. I. Comparison of nine different species in culture. *J Exp Mar Biol Ecol* **15:** 261–274.
Newton, R.J., Griffin, L.E., Bowles, K.M., Meile, C., Gifford, S., Givens, C.E., *et al.* (2010) Genome characteristics of a generalist marine bacterial lineage. *ISME J* **4:** 784–798.
Niu, Y., Shen, H., Chen, J., Xie, P., Yang, X., Tao, M., *et al.* (2011) Phytoplankton community succession shaping bacterioplankton community composition in Lake Taihu, China. *Water Res* **45:** 4169–4182.
Pinhassi, J., Pujalte, M.J., Macian, M.C., Lekunberri, I., Gonzalez, J.M., Pedros-Alio, C., and Arahal, D.R. (2007) *Reinekea blandensis* sp. nov., a marine, genome-sequenced gammaproteobacterium. *Int J Syst Evol Microbiol* **57:** 2370–2375.
Pruesse, E., Peplies, J., and Glöckner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28:** 1823–1829.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* D590–D596.

Reddy, T.B., Thomas, A.D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., *et al.* (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–D1106.

Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16:** 276–277.

Richter, M., and Rossello-Mora, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* **106:** 19126–19131.

Roberts, I.S. (1996) The biochemistry and genetics of capsular polysaccharide production in bacteria. *Ann Rev Microbiol* **50:** 285–315.

Romanenko, L.A., Schumann, P., Rohde, M., Mikhailov, V.V., and Stackebrandt, E. (2004) *Reinekea marinisedimentorum* gen. nov., sp. nov., a novel gammaproteobacterium from marine coastal sediments. *Int J Syst Evol Microbiol* **54:** 669.

Romano, S., Schulz-Vogt, H.N., González, J.M., and Bondarev, V. (2015) Phosphate limitation induces drastic physiological changes, virulence-related gene expression, and secondary metabolite production in *Pseudovibrio* sp. strain FO-BEG1. *Appl Environ Microbiol* **81:** 3518–3528.

Sarmento, H., and Gasol, J.M. (2012) Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environ Microbiol* **14:** 2348–2360.

Stamatakis, A., and Ott, M. (2008) Efficient computation of the phylogenetic likelihood function on multi-gene alignments and multi-core architectures. *Philos Trans R Soc Lond B Biol Sci* **363:** 3977–3984.

Stingl, U., Desiderio, R.A., Cho, J.C., Vergin, K.L., and Giovannoni, S.J. (2007) The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microbiol* **73:** 2290–2296.

Sowell, S.M., Wilhelm, L.J., Norbeck, A.D., Lipton, M.S., Nicora, C.D., Borofsky, D.F., *et al.* (2009) Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J* **3:** 93–105.

Tada, Y., Taniguchi, A., Nagao, I., Miki, T., Uematsu, M., Tsuda, A., and 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:** 4055–4065.

Tan, S.L., Xiaoshan Zhu, J.S., Yu, S., Zhan, W., Wang, B., and Cai, Z. (2015) An association network analysis among microeukaryotes and bacterioplankton reveals algal bloom dynamics. *J Phycol* **5:** 120–132.

Taylor, E.J., Goyal, A., Guerreiro, C.I., Prates, J.A., Money, V.A., Ferry, N., *et al.* (2005) How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides: structure and activity of a *Clostridium thermocellum* lichenase, CtLic26A. *J Biol Chem* **280:** 32761–32767.

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.

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* **5:** e11888.

Thiele, S., Fuchs, B.M., and Amann, R.I. (2011) Identification of microorganisms using the ribosomal RNA approach and fluorescence in situ hybridization. In *Treatise on Water Science*. Wilderer, P. (ed). Oxford: Elsevier Science, pp. 171–189.

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

Tindall, B.J., Sikorski, J., Smibert, R.M., and Krieg, N.R. (2007) Phenotypic characterization and the principles of comparative systematics. In *Methods for General and Molecular Microbiology*. Reddy, C.A., Beveridge, T.J., Breznak, J.A., Marzluf, G.A., Schmidt, T.M., and Snyder, L.R. (eds). Washington DC: ASM Press, pp. 330–393.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., *et al.* (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9:** e112963.

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.

Yang, C., Li, Y., Zhou, B., Zhou, Y., Zheng, W., Tian, Y., *et al.* (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., and Xu, Y. (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40:** W445–W451.

Zhilina, T.N., Appel, R., Probian, C., Brossa, E.L., Harder, J., Widdel, F., and Zavarzin, G.A. (2004) *Alkaliflexus imshenetskii* gen. nov. sp. nov., a new alkaliphilic gliding carbohydrate-fermenting bacterium with propionate formation from a soda lake. *Arch Microbiol* **182:** 244–253.

Zhou, J., Bruns, M.A., and Tiedje, J.M. (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62:** 316–322.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Recurrence of *Reinekea* spp. during spring phytoplankton blooms in the North Sea in the years 2009 to 2012. (a) Chlorophyll *a* concentration, (b) co-occurring abundances of the diatom *Thalassiosira nordenskioeldii*, (c) *Reinekea* spp. abundances as assessed by CARD-FISH (probe REI731), (d) *Reinekea* spp. abundances as assessed by 16S rRNA gene tag sequencing and MED analysis. Periods of two months centered around spring blooms are highlighted in light gray. Dotted lines indicate decreasing peak abundances. *Reinekea* oligotype 3280 is hardly visible in the figure since it has very low abundances only in 2012. Sampling dates are indicated by small dark

gray boxes below x-axes. Panel (a), (b), (c) are adapted from Teeling *et al.* (2016). Metagenome sampling dates are indicated by red boxes at the top of (a), including coverage and average depth values of read recruitments with the *'Reinekea forsetii'* Hel1_31_D35 genome.

Fig. S2. Sulfate concentration (bars) and optical density (dots) of '*R. forsetii*' Hel1_31_D35 cultures over time in sulfate-free HaHa_100V medium with (black) and without (white) thiosulfate addition. Initial sulfate concentrations are due to the inocula. Error bars refer to biological triplicates.

Fig. S3. Phylogenetic tree of proteohodopsin (PR) amino acid sequences of '*Reinekea forsetii*' Hel1_31_D35 and other *Gammaproteobacteria*. Sequences were retrieved from GenBank and aligned using MAFFT v. 7 (Katoh and Standley, 2013). Full-length PR sequences were used to construct a phylogenetic tree in ARB v. 5.5 (Ludwig *et al.*, 2004) using a 40% base conservation filter with RAxML (1,000 repetitions). The partial PR sequence of the SAR92 strain HTCC2207 was subsequently added by parsimony. Bar: 0.1 substitutions per nucleotide position. *Alphaproteobacteria* PR sequences were used as outgroup.

Fig. S4. N-acetyl-D-neuraminic acid degradation and biosynthesis genes in '*R. forsetii*' Hel1_31_D35.

Fig. S5. Polyphosphate inclusions in '*R. forsetii*' Hel1_31_D35 cells as visualized by (**a**) methylene blue and (**b**) DAPI staining.

Fig. S6. Free-living and particle-associated lifestyles of *Reinekea* spp. visualized by CARD-FISH and lectin staining. Water samples were taken from the isolation site of '*R. forsetii*' Hel1_31_D35 on April 14th, 2009. The samples were filtered and stained by DAPI, the CARD-FISH probe REI731 and lectin. (a) Free-living cells collected in the 0.2–3 μm fraction, (b) cells in the 3–10 μm fraction. Arrows point toward selected *Reinekea* cells. Scale bar: 5 μm.

Fig. S7. Phylogenetic tree of cloned *Reinekea* 16S rRNA gene sequences obtained during the 2009 spring bloom at Helgoland and '*R. forsetii*' Hel1_31_D35 16S rRNA gene sequences. The tree was computed with the Neighbor-Joining method (Jukes-Cantor substitution model; 1,000 repetitions) using a 50% base conservation filter. *Alphaproteobacteria* were used as outgroup. Bar: 0.01 substitutions per nucleotide position.

Table S1. 16S rRNA sequence similarity between 75 *Reinekea* clones obtained during the 2009 spring phytoplankton bloom.

Table S2. 16S rRNA sequence similarity between '*R. forsetii*' Hel1_31_D35 and 75 clones obtained during the 2009 spring phytoplankton bloom.

Table S3. Read recruitment of North Sea metagenomes by the '*R. forsetii*' Hel1_31_D35 genome.

Table S4. List of metagenome contigs classified as affiliating with the genus *Reinekea*.

Table S5. CAZyme gene repertoire of the '*R. forsetii*' Hel1_31_D35 genome.

Table S6. Comparison of GH and CBM repertoires of 21 gammaproteobacterial genomes of marine origin.

Table S7. E-value thresholds used for automated CAZyme family detection. Searches were performed against the CAZy database, the dbCAN database and the Pfam database using the indicated E-value thresholds. CAZymes were only annotated when at least two of the three database searches yielded positive results.

Chapter 3

Polysaccharide-driven niche differentiation between distinct *Polaribacter* clades during North Sea spring algal blooms

The manuscript is in preparation for *The ISME Journal*

Contributions to the manuscript:

Experimental concept and design: 60%

Experimental work and/or acquisition of data: 60%

Data analysis and interpretation: 80%

Preparation of figures and tables: 100%

Drafting of the manuscript: 100%

The ISME Journal – Original Article

Polysaccharide-driven niche differentiation between distinct *Polaribacter* clades during North Sea spring algal blooms

Burak Avcı[1], Karen Krüger[1], Bernhard M. Fuchs[1], Hanno Teeling[1], Rudolf I. Amann[1*]

[1]Department of Molecular Ecology, Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359, Bremen, Germany

[*]Corresponding author: Rudolf I. Amann, Department of Molecular Ecology, Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany.

E-mail: ramann@mpi-bremen.de, Phone: +49 421 2028 930, Fax: +49 421 2028 790

Running title: *Polaribacter* niche partitioning during algal blooms

E-mail addresses and telephone numbers of all authors:

Burak Avcı	bavci@mpi-bremen.de	+49 421 2028 545
Karen Krüger	kkrueger@mpi-bremen.de	+49 421 2028 942
Bernhard M. Fuchs	bfuchs@mpi-bremen.de	+49 421 2028 935
Hanno Teeling	hteeling@mpi-bremen.de	+49 421 2028 976
Rudolf I. Amann	ramann@mpi-bremen.de	+49 421 2028 930

Conflict of Interest

The authors declare no conflict of interest.

Abstract

Algal blooms create dynamic environments with the release of structurally complex molecules and trigger the growth of diverse heterotrophic bacteria. Yet ecological processes driving the assembly of these bacteria are largely unknown. In this study, we explore the niches of the genus *Polaribacter*, an abundant bacterial clade during spring algal blooms in the North Sea. In the 2009 to 2012 blooms we (i) identified six distinct *Polaribacter* clades using phylogenetic analyses, (ii) quantified their abundances using fluorescence *in situ* hybridization, (iii) retrieved largely complete genomes via metagenome binning, and (iv) assessed *in situ* expression using metaproteomics. The genomes of the four major clades revealed distinct niches. *Polaribacter* 2-a represented a typical first responder with a small genome and limited polysaccharide utilization capacity. *Polaribacter* 3-a was abundant only in 2010 and possessed a distinct degradation potential for mannose-rich sulfated polysaccharide. *Polaribacter* 3-b was a late responder with a pronounced utilization capability for sulfated xylan. *Polaribacter* 1-a had the largest genome encoding diverse glycan degradation genes and was particularly abundant following *Chattonella* algae blooms. Our study suggests a polysaccharide-driven niche partitioning between different *Polaribacter* clades during algal blooms and highlights how diverse algal substrates could create distinct niches in which specialized bacterioplankton thrive.

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

Keywords: North Sea; phytoplankton; bacterioplankton; Bacteroidetes; Polaribacter; niche differentiation

3.1) Introduction

One of the fundamental questions in microbial ecology is which factors shape the composition of a given community. To address this question, neutral and niche theories have emerged (1). According to the neutral theory, all species are ecologically equivalent and the microbial community structure is shaped only by stochastic processes (2). On the other hand, the niche theory postulates that environmental conditions regulate the species composition by a process known as habitat filtering (3). The niche is defined as a set of biotic and abiotic factors that an organism needs to thrive (4). Thus a given habitat selects a group of organisms sharing similar niches and competing for the same resources (5). Stable conditions provided, the best-adapted species ultimately prevails and supplants less adapted species (6). However, many habitats are subjected to constant changes, which preclude the establishment of stable communities. An important mechanism structuring the community assembly in such dynamic habitats is resource partitioning, which enables the coexistence of similar organisms with distinct ecological niches (7).

Marine algal blooms are an essential part of the global carbon cycle and are characterized by a dynamically changing environment with a large supply of diverse and structurally complex substrates (8). These molecules require a concerted breakdown by more than a single species and therefore provide conditions allowing co-occurrence of various heterotrophic bacteria with adjacent ecological niches. In particular, the terminal phases of these blooms result in the massive release of organic matter triggering significant and often successive shifts in microbial community composition (9) (10) (11). Different clades of heterotrophic bacteria become abundant and degrade algal-derived organic substrates (12) (13). Previous studies on spring algal blooms in the southern North Sea detected the succession and annual recurrence of *Flavobacteriia*, *Gammaproteobacteria*, and alphaproteobacterial *Roseobacter* clade (14) (15). Metagenome analyses suggested that these clades possessed distinct polysaccharide niches and were likely selected by the successive availability of particular algal glycans or glycan classes (14) (15).

Members of the genus *Polaribacter* (class *Flavobacteriia*) constituted the most abundant and diverse recurring bacterial clade during spring algal blooms in the southern North Sea from 2009 to 2012 (15). Catalyzed reporter deposition

fluorescence *in situ* hybridization (CARD-FISH) analysis using a *Polaribacter*-specific oligonucleotide probe (POL740) showed that *Polaribacter* relative abundances within the planktonic fraction (0.2 – 3 µm) reached up to ~27%, ~26%, ~14%, and ~25% from 2009 to 2012, respectively (15). Four different *Polaribacter* oligotypes with relative read abundances exceeding 1% were detected during the blooms using minimum entropy decomposition (MED) of 16S rRNA tag sequences (16). These data demonstrated the high intra-genus diversity of the recurrent *Polaribacter*. The detected oligotypes were also confined to spring and summer algal blooms with rapid shifts in their abundance, suggesting that algal-derived substrates triggered the growth of distinct *Polaribacter* clades (16). This was corroborated by analyses of the carbohydrate-active enzyme (CAZyme) gene repertoires of distinct bacterioplankton clades (15). These analyses revealed that *Polaribacter* spp. were enriched in glycoside hydrolase (GH) families 16, 17, and 30, all of which all contain enzymes that can decompose laminarin, as well as GH13 (e.g. α-amylases) and GH92 (α-mannosidases) (15). However, a metagenomic analysis with higher taxonomic resolution than genus-level is required to separate the taxonomic and functional diversity of distinct *Polaribacter* clades into ecologically meaningful niches.

Sympatric taxa within the same genus tend to have distinct niches and respond differently to the same environmental conditions (17) (18). Members of the genus *Polaribacter* are widely distributed in marine habitats. To date, 23 *Polaribacter* species have been described from various environments such as sea ice (19), seawater (20), marine sediments (21), marine animals (22) and macro-algae (23). *Polaribacter* spp. were also detected during the algal blooms in temperate and polar regions (24) (25) (26). Likewise, we previously isolated two *Polaribacter* strains (Hel1_33_49 and Hel1_85) during a phytoplankton bloom off the coast of Helgoland island in the southern North Sea (27). Strain Hel1_33_49 is a planktonic isolate feeding on proteins and a small subset of algal polysaccharides, whereas strain Hel1_85 is likely associated with algae and can decompose a larger spectrum of polysaccharides (27). This study showcased only two niches in the large niche space of diverse spring bloom-associated *Polaribacter* species. However, these two isolates were not representative of the dominant *Polaribacter* clades during North Sea algal blooms and a possible niche differentiation between environmentally relevant *Polaribacter* remained elusive.

In this study, we explore niche partitioning between *Polaribacter* clades that are recurrently abundant during North Sea spring algal blooms. We (i) identified six distinct *Polaribacter* clades using phylogenetic analyses, (ii) quantified their abundances using fluorescence *in situ* hybridization with novel oligonucleotide probes and metagenome read recruitment (iii) retrieved almost complete *Polaribacter* genomes for all six clades via binning of a time series of metagenomes, and (iv) assessed the *in situ* gene expression by re-analyzing published metaproteome data. Our results revealed the differential abundance of distinct *Polaribacter* clades with varying glycan utilization capabilities, suggesting that release of diverse polysaccharides during algal blooms contribute to the niche differentiation among these closely related bacteria.

3.2) Materials and Methods

16S rRNA clone libraries and phylogenetic analysis

Full-length *Polaribacter*-related 16S rRNA gene sequences were retrieved from a clone library constructed from Helgoland spring bloom bacterioplankton (0.2 - 3 µm) on April 8th, 2010 as described in (14). Using the Silva Incremental Aligner (SINA) (28), 73 *Polaribacter* 16S rRNA sequences (>1,450 bp) were aligned to all 16S rRNA sequences in the Silva SSURef v.132 dataset (29) and manually curated in Arb v.6.1 (30). Together with all *Polaribacter*-related sequences in the Silva SSURef NR 99 v.132 dataset, phylogenetic tree reconstruction was done with the RAxML v.7 (31) maximum likelihood method (GTR-GAMMA rate distribution model, rapid bootstrap algorithm, 100 repetitions) and the Neighbor-joining method (Jukes-Cantor substitution model, 1,000 repetitions). Both phylogenetic treeing methods were run on the sequences filtered with and without 10%, 30%, and 50% positional conversation for all *Flavobacteriia*. A consensus tree was generated following the recommendations described in (32).

Oligonucleotide probe design and CARD-FISH analysis

Four oligonucleotide probes (POL405, POL1270, POL183a, and POL180) were designed using the probe design tool implemented in Arb with Silva SSURef NR 99 v.128 (Table S1). These probes were tested on environmental samples using varying formamide concentrations from 0% to 50% at 46°C hybridization temperature. The highest possible formamide concentration providing sufficient brightness for signal

detection was used for further hybridizations (Table S1). The sum of cell numbers detected by these four probes was also cross-checked with the genus-specific probe POL740 to evaluate the probe's specificities. For CARD-FISH analysis, samples were taken as described in (14) and were analyzed according to (33) using the newly designed four *Polaribacter* oligonucleotide probes. Cell counting was performed with an automatic microbial cell enumeration system (34) and ACMETOOL2.0 image analysis software (http://www.technobiology.ch/index.php?id=acmetool). Cell counting results and additional abundance data are summarized in Table S2.

Metagenome sequencing and binning

38 surface seawater metagenome samples were sequenced at the DOE Joint Genome Institute (JGI) (Table S3) (15). Quality filtering and trimming of raw reads, metagenome assembly and binning were performed as described previously (35). Briefly, BBDuk v35.14 (http://bbtools.jgi.doe.gov) was used to remove adapters and low quality reads. Metagenomic datasets were assembled individually using SPAdes v3.10.0 (36) Separate binning of contigs from each assembly was performed using CONCOCT (37) integrated into the anvi'o v.3 metagenomic workflow (38). BBMap v35.14 (http://bbtools.jgi.doe.gov) was used for read mapping. Phylogenomic placement and quality estimates of metagenome assembled genomes (MAGs) were investigated in CheckM v1.0.7 (39) and a subset of automatically binned CONCOCT MAGs was selected for manual bin refinement in anvi'o. Refined bins were analyzed a second time with further *Bacteroidetes* reference genomes, chosen based on 16S rRNA sequences. Mash v.1.1.1 (40) was used to reduce redundancy and to cluster MAGs into approximate species clusters and the ones placed within the genus *Polaribacter* by CheckM were kept. The MAGs having (i) relative abundance >0.4% based on metagenome read recruitment, (ii) contamination <5% and (iii) completeness >69% were selected for further analyses of environmentally relevant *Polaribacter* (Table S4).

Metagenome-assembled genome (MAG) analyses

Average Nucleotide Identity (ANI) values between *Polaribacter* MAGs and genomes were calculated using the ani Ruby script from the enveomics toolbox (41). For phylogenomic analysis, the protein sequences of 43 conserved marker genes were extracted and aligned in CheckM v.1.0.8 (39). Using the concatenation of these gene

sequences, a protein tree was computed with the RAxML v.7 (31) maximum likelihood algorithm (GAMMA-WAG substitution model, rapid bootstrap algorithm, 100 repetitions) in Arb v.6.1 (30). Metagenome reads from 2009 to 2012 North Sea spring phytoplankton blooms were recruited by *Polaribacter* MAGs and genomes with BBMap v.35.14 using “fast” mode. For 38 metagenomes sequenced on the Illumina HiSeq 2500 platform (Illumina Inc., San Diego, CA, USA) (Table S3), reads recruitment was done with minimum mapping identity of one (minid= 1) and identity filter for reporting mappings of one (idfilter= 1) to prevent read cross-mapping between closely-related MAGs or genomes. For the clades with low micro-diversity (*Polaribacter* 2-a and *Polaribacter* 2-b), the most complete MAGs were used and both parameters were set to 0.99. For four metagenomes sequenced on the now defunct 454 FLX Ti platform (454 Life Sciences, Branford, CT, USA) (Table S3), all read recruitments were carried out with minid=0.97 and idfilter= 0.99 using the most complete MAG or genome from each clade. The read recruitment results are shown in Table S5. For each *Polaribacter* clade, the correlation between the relative abundances obtained from (i) metagenome read recruitment (ii) CARD-FISH analysis and (iii) minimum entropy decomposition (MED) was calculated using the Spearman rank correlation test implemented in R v.3.2.3 (42) (Table S6).

Gene prediction and annotation

Within the MAGs having ANI >99%, the most complete and the least contaminated one was selected as representatives. To demonstrate the validity of this approach, MAGs with ANI >99% in the *Polaribacter* clades 2-a and 2-b were also analyzed. Gene prediction was performed using Prodigal (43) as implemented in Prokka v.1.12 (44) including prediction of partial genes (removal of –c and –m options). Genes were annotated in RAST v.2 (45) (Table S7). CAZymes were annotated as described in (46) using the dbCAN v.6 (47), Pfam v.31 (48) and CAZy (as of March 15th, 2017) (49) databases. Peptidases were annotated based on best-hits of BLASTp v.2.6.0+ (50) searches against the MEROPS database (merops_scan database v.12) (51) with default settings. For annotations of *SusC*-like proteins, *SusD*-like proteins, and sulfatases, HMMer v.3.1 (52) searches with TIGRFAM (profile TIGR04056), Pfam (profiles PF07980 and PF00884) were used (e-value: E-10). PULDB (53) was used to predict the putative substrate-specificity based on CAZyme patterns of some polysaccharide utilization loci (PULs) found in *Polaribacter* MAGs and genomes. To

cluster these PULs, translated nucleotide sequences of genes of *SusC*-like proteins were aligned in MAFFT v.7 (54) and subjected to a phylogenetic analysis with the Neighbor-joining method (JTT substitution model).

Core and pan genome analysis

Core and pan genomes of *Polaribacter* clades were determined in EDGAR v.2.2 (55), which implements both reciprocal best BLAST hits and BLAST Score Ratio Values (SRVs) (56) to calculate gene orthology. Core (shared by all MAGs) and dispensable genomes (shared by at least two MAGs) together with singletons (non-shared, unique to a MAG) were calculated via the "fractional pan genome" analysis tool in EDGAR using the most complete MAG or genome in four major *Polaribacter* clades (Table S8). Genes other than CAZymes, peptidases, and hypothetical proteins were clustered according to the COG categories (57) using the eggNOG-mapper v.4.5.1 (58).

Metaproteomics

Metaproteome data from 2009 (14) and 2010-2012 (59) Helgoland spring blooms were re-analyzed to investigate *in situ* gene expression of *Polaribacter* clades. Using standalone BLASTp v.2.6.0+ (50), metaproteome sequences were mapped on the protein gene sequences from *Polaribacter* MAGs and genomes with specific identity and e-value thresholds (percent identity > 99% and e-value = 0) (Table S9).

3.3) Results

Phylogeny

Phylogenetic analysis of all 16S rRNA sequences taxonomically classified as *Polaribacter* in the Silva database (v.132) revealed four distinct clusters (*Polaribacter* 1-4) (Fig. 1). The type species for the genus, *Polaribacter filamentus* (19), was grouped into *Polaribacter* 1, while macro-algae related species, *P. reichenbachii* (23), was in *Polaribacter* 4. For clusters *Polaribacter* 2 and 3, no isolates have been validly described yet. *P. huanghezhanensis* branched separately with maximum sequence similarity of 94.2% to the other described *Polaribacter* species. This was below the taxonomic threshold (94.5%) suggested for the rank genus (60) and indicates the need for a further taxonomic revision.

Polaribacter-related 16S rRNA clone sequences obtained during North Sea spring algal blooms in 2009 (14) and 2010 (this study) were grouped into three clusters. 58 clone sequences retrieved on April 14, 2009 formed three distinct branches in the clusters *Polaribacter* 1 and 3 (Fig. 1), indicating a large diversity of *Polaribacter* species during the 2009 bloom. Furthermore, 15 sequences obtained on April 8, 2010 were clustered into a single monophyletic clade in *Polaribacter* 2 (Fig. 1), which demonstrated the dominance of a different *Polaribacter* clade in the early phase of the 2010 spring bloom. Mean similarities between sequences within North Sea *Polaribacter* clades were in the range of 96.5% - 98.4% (Table S10). This supported the presence of distinct species within the genus *Polaribacter* (60), which was individually targeted by novel oligonucleotide probes (Table S1).

Classification of Polaribacter *MAGs*

We obtained 35 *Polaribacter*-related MAGs (Table S4) via binning of 38 metagenomes sampled from 2010-2012 bloom events (Table S3). Phylogenetic survey of 43 conserved marker genes and ANI comparisons of the *Polaribacter* MAGs with seven genomes of North Sea *Polaribacter* strains demonstrated presence of six distinct clades (henceforth termed *Polaribacter* 1-a, 1-b, 2-a, 2-b, 3-a, and 3-b) during spring algal blooms. The clade *Polaribacter* 3-a also included three of the North See strains (Hel1_33_49, Hel1_33_78, and Hel1_33_96). The overall tree topology was similar to the one obtained through 16S rRNA gene analysis (Fig. 1) (Fig. 2). ANI values in between these six clades were < 95% (Fig. S1), suggesting that they represent distinct species (61). ANI values within these clades varied between ~94.5 - 99% and indicated some intra-clade heterogeneity (Fig. S1). For example, four sub-clades were detected in *Polaribacter* 1-a, whereas two sub-clades were found both in *Polaribacter* 3-a and 3-b. In contrast, MAGs in *Polaribacter* 2-a and 2-b were seemingly homogeneous clades with ANI values ~99%. Altogether, these results suggested six distinct spring-bloom associated *Polaribacter* clades with various degrees of intraspecific diversity in the bloom events.

Polaribacter *in situ abundances*

Various methods were used to monitor the *in situ* abundance of the six distinct *Polaribacter* clades during the 2009 to 2012 North Sea spring algal blooms. These methods included (i) CARD-FISH analysis using newly designed oligonucleotide

probes (Table S1) (Fig. 3b), (ii) reassessment of previously published minimum entropy decomposition (MED) analysis of 16S rRNA amplicon data (Fig. 3c) (16) (iii) metagenome read recruitment on *Polaribacter* MAGs and genomes (Table S5) (Fig. 3d). *Polaribacter* 1-a dominated the bacterial community in 2009, 2011, and 2012 spring blooms (19%, 7%, and 13%) based on the cell counts with the clade-specific probe POL405 (Fig. 3b). Spearman tests suggested a high level of correlation between individual abundance data (Table S6) and enabled the connection of the taxonomic units revealed by the different methods (Fig. 4). For example, MED node 3321 (Fig. 3c) and metagenome read recruitment on *Polaribacter* 1-a MAGs (Fig. 3d) yielded a similar abundance pattern to POL405 counts (Fig. 3b) with a high correlation coefficient (Spearman's rho = 0.96 and 0.88, respectively) (Table S6). Furthermore, MAGs within the *Polaribacter* 1-a clade exhibited a differential abundance pattern. The MAG POL1A_74 was abundant in 2009 and 2012, while the MAGs POL1A_42, POL1A_60, and POL1A_84 were detected in 2011. POL1A_42 was also found in lower abundances in the late phase of the 2010 bloom (Table S4). Interestingly, the peak abundances of *Polaribacter* 1-a in the three blooms events occurred after the dominance *Chattonella* (Fig. 3a). These algae are rather distinct from diatoms, which are usually present in the Helgoland spring blooms. In the 2010 bloom, the phytoplankton community was dominated by *Phaeocystis* spp., *Thalassiosira nordenskioeldii*, and *Mediopyxis helysia*. In this year, *Polaribacter* 2-a, *Polaribacter* 3-a, and *Polaribacter* 3-b successively reached high abundances (6%, 8%, and 10%) (Fig. 3d). *Polaribacter* 2-a and *Polaribacter* 3-b were also present albeit with lower abundances, respectively, in the early and late phase of bloom events in 2009, 2011, and 2012 (1%, 1%, 0.4% and 4%, 0.5%, 3%), whereas *Polaribacter* 3-a was only abundant in 2010 (Fig. 3d). In contrast, *Polaribacter* 1-b and *Polaribacter* 2-b were the "rare" clades, reaching lower relative abundances than other clusters (Fig. 3d). *Polaribacter* 1-b was detected in relatively low numbers during the 2009 and 2011 spring blooms (2% and 3%) together with *Polaribacter* 1-a, while *Polaribacter* 2-b reached abundances up to 3% only in 2010. Overall, our results coherently showed the presence of four major (1-a, 2-a, 3-a, and 3-b) and two rare (1-b and 2-b) *Polaribacter* clades in spring algal blooms off Helgoland.

Core and pan genomes

We investigated the common and variable genes within and between the *Polaribacter* clades. These genes were classified in the core genome (shared by all), dispensable genome (shared at least by two MAGs), and singletons (non-shared, unique for a MAG). The proportion of the genes detected in the core genome of individual clades (1-b, 2-a, 2-b, 3-a, and 3-b) (Fig. S2) is significantly higher (83-99%) than the core genome of four major *Polaribacter* clades (50-67%) (Fig. 5). This suggested a higher level of inter-clade metabolic diversification. 1327 genes were found in the core genome of the most complete MAGs or genomes in four major *Polaribacter* clades (POL1A_74, POL2A_63, PHEL_96, and POL3B_70). These genes dominantly carry out housekeeping cellular functions, and are involved in metabolisms such as gliding motility, proteorhodopsin, and polyphosphate hydrolysis (*ppX*) (Fig. 5) (Table S8). Gliding motility enables the movement on surfaces and some proteorhodopsin types are used for light-dependent supplementary energy production. Moreover, polyphosphate metabolism is a key to thrive in phosphate-limiting conditions, which is typical of algal blooms. These traits are frequent in marine *Bacteroidetes* (62) (63) and could provide certain ecological advantages to survive in algal bloom conditions.

Genes responsible for carbohydrate degradation and transport were enriched in dispensable genome and singleton genes. Large amounts of polysaccharides released during algal blooms are decomposed by carbohydrate-active enzymes (CAZymes), which include catalysts for polysaccharide binding, degradation, and modification. These enzymes are classified in carbohydrate-binding modules (CBMs), glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and glycosyltransferases (GTs) (49). Furthermore, *Bacteroidetes* possess a specialized carbohydrate transport system including a tandem of a SusC-like TonB-dependent transporter and a SusD-like protein for glycan-binding (64). These genes, which are responsible for carbohydrate degradation and transport, represented the most abundant metabolic category in the dispensable genome and singletons with higher relative numbers than the core genome (Fig. 5b). The MAG POL1A_74 had 54 genes responsible for carbohydrate metabolism in its singletons including families GH20 (e.g. N-acetylglucosaminidase) and GH2 (e.g. β-galactosidase) (Table S8). Analysis of representative MAGs of the four most abundant sub-clades within *Polaribacter* 1-a indicated a core genome smaller than other clades (59-72%), and

the enrichment of carbohydrate utilization and transport genes mostly in singletons. This suggested an additional intra-clade metabolic diversification with respect to polysaccharide utilization (Fig. S3). Furthermore, in their singletons, the POL2A_63 encoded GH13 (e.g. α-amylase), GH16 (e.g. laminarinase), while PHEL_96 (*Polaribacter* 3-a) had three GH92 (e.g. alpha-mannosidase). POL3B_70 also possessed GH10 (e.g. β-xylanase) and GH3 (e.g. β-xylosidase) together with other 31 genes in the singletons (Fig. 5a) (Table S8). These analyses overall suggested that *Polaribacter* clades differed in terms of their carbohydrate degradation potentials.

Peptidase and CAZyme distribution

Comparison of peptidase and degradative CAZyme abundances between *Polaribacter* clades showed different protein and carbohydrate utilization potentials. Together with polysaccharides, large amounts of proteins are also present in marine phytoplankton (65) and heterotrophic bacteria are able to utilize proteins using peptidases (66). All *Polaribacter* clades encoded high numbers of serine (S) and metallo (M) peptidases (~80% of all peptidases), which mediate the degradation and uptake of extracellular proteins (Table S11) (27). Furthermore, glycoside hydrolases (GHs) that hydrolyze glycosidic bonds between carbohydrate molecules were the most abundant CAZyme family genes (Table S11). Remarkably, the range of CAZyme proportions (8-39 per Mbp) is two times higher than the one with peptidases (37-51 Mbp). Together with *Polaribacter* isolates from temperate seawater and polar regions, *Polaribacter* 2-a and *Polaribacter* 2-b had the highest peptidase and the lowest degradative CAZyme abundances per megabase with the smallest estimated MAG sizes (Fig. 6). In contrast, *Polaribacter* 1-a possessed the highest CAZyme and lowest peptidase proportions with the largest MAG sizes and grouped into species, which are associated with macro-algae and marine animals (Fig. 6). *Polaribacter* 3-a and *Polaribacter* 3-b contained moderate CAZyme and peptidase repertoires together with North Sea spring bloom isolates, while *Polaribacter* 1-b harbored the lowest numbers (Fig. 6). Overall, the niche space of North Sea *Polaribacter* clades was clearly separated into three axes (peptidase, CAZyme, and genome size), which revealed pronounced differentiation towards carbohydrate degradation.

Polysaccharide utilization locus (PUL) repertoires

Differences in the polysaccharide utilization spectrum of *Polaribacter* spp. suggested common and variable algal substrates targeted by each clade (Fig. 7). Polysaccharide degradation in *Bacteroidetes* is usually encoded in specialized genomic islands referred to as polysaccharide utilization loci (PULs). To orchestrate the binding, uptake, and degradation of a particular polysaccharide, PULs contain a gene tandem coding for a SusD-like protein for glycan-binding and a SusC-like TonB-dependent transporter as well as dedicated CAZyme and sulfatase genes (64). Therefore, the genetic composition of a PUL provides hints at the potential to utilize a specific polysaccharide and enables prediction of the carbohydrate degradation potential of a bacterium (67). In *Polaribacter* MAGs and genomes, we detected 78 putative PULs targeting a large array of substrates (Fig. S5). Phylogenetic analysis of the translated *SusC* gene sequences in these PULs suggested a substrate-specific clustering (59) and enabled us to further classify the different variants of a PUL type putatively targeting the same substrate (Fig. S4). All *Polaribacter* clades possessed PULs to putatively degrade β-glucans (e.g. laminarin) and each clade harbored its own β-glucan PUL variant having GH16, GH17, and GH30 (all targeting β-glucans) in common (Fig. 7a) (Fig S3) (Fig. S5). The PUL detected in *Polaribacter* 3-a (strain Hel1_33_49) was also shown to be upregulated with laminarin in a previous study (27). Among the seven variants of β-glucan PULs, *Polaribacter* 2-a and *Polaribacter* 2-b had the most complex one, which included peptidases M01 (e.g. aminopeptidase activity) and S51 (dipeptidase activity) and CBM4 (e.g. binding to β-1,3-glucan) (Fig. 7b). According to metaproteomic analyses, GH16, CBM4, and *SusCD* genes in this PUL were expressed during the 2009 and 2010 spring blooms (Fig. S6).

Four *Polaribacter* clades (1-a, 2-a, 2-b, and 3-a) harbored PULs putatively targeting α-glucans (e.g. starch and glycogen) (Fig. 7a). Having two variants, these PULs contained GH13 (α-amylase), GH65 (maltose phosphorylase), GH31 (e.g. α-glucosidase) as core GH family genes (Fig. S4) (Fig. S5). A glucose-induced PUL with high synteny was shown in the North Sea isolate *Gramella forsetii* KT0803^{T} (68). α-glucan PULs also encoded two *SusE* genes, which direct the uptake of maltooligosaccharides of specific lengths and likely facilitate the selection of particular glycans from the environment (69). However, expression of the genes related to α-glucan PULs was not detected during the bloom events (Fig. S6).

Moreover, all *Polaribacter* clades, except *Polaribacter* 2-a and 2-b, had putative PULs to utilize α-mannose-rich polysaccharides, which is one of the major constituents of diatom frustules (70) (Fig. 7a). These PULs had two variants and possessed CAZymes families such as GH92 (e.g α-mannosidase), GH130 (e.g mannooligosaccharide phosphorylase) (Fig. S4). α-mannan specificity of a PUL containing GH92, GH130, and GH76 was shown in a bacterium from the human gut (71). The MAG POL1A_42 also expressed the *SusCD* genes of putative α-mannose-rich polysaccharide PUL in the late phase of the 2009 bloom (Fig. S6).

Polaribacter 1-a, 3-a, and 3-b harbored PULs presumably targeting mannose-rich sulfated polysaccharides (Fig. 7a). Among these PULs, *Polaribacter* 3-a encoded a distinctive one, which was the most complex PUL structure found in all clades. This PUL contained twelve CAZymes including five GH92, two GH3 (α-mannosidases) and a GH99 (putative endo-α-mannanase) together with nine sulfatases (Fig. 7b). However, expression of the genes related to this PUL was not detected, likely since the metaproteome sampling dates in 2010 were not coinciding with the peak abundance of *Polaribacter* 3-a (Fig. S6). Furthermore, *Polaribacter* clades 1-a, 1-b, and 3-b possessed putative PULs targeting sulfated xylan with a high level of expression (Fig. 7a). *Polaribacter* 3-b harbored all variants and the most complex form of sulfated xylan PULs (Fig. 7). This PUL encoded GH3 (e.g. xylosidase), GH10 (e.g. β-xylanase) and two sulfatases, together with an adjacent putative PUL containing GH128 (β-glucanase) and four sulfatases (Fig. 7b). A PUL with GH3 and GH10 was verified to be upregulated with xylan in human gut bacterium *Bacteroides xylanisolvens* (72). In addition to these substrates, *Polaribacter* clades further differentiated their glycan niches utilizing N-acetyl-D-glucosamine, alginate, arabinogalactan, and trehalose rich polysaccharides (Fig. 7a) (Supplementary information). These findings overall demonstrated that *Polaribacter* clades possessed varying gene repertoires encoding the binding, transport, and degradation of different algal polysaccharides.

3.4) Discussion

In this study, we investigated niche partitioning of different *Polaribacter* clades thrived during the North Sea algal blooms with a focus on the utilization of high molecular weight compounds. We could dissect niche space and temporal dynamics in high

taxonomic resolution, which allowed detecting the abundance and genomic content of six distinct *Polaribacter* clades.

With differing genome sizes, peptidase numbers and PUL spectra, the six *Polaribacter* clades occupied different niches during the North Sea spring algal blooms. The range of relative CAZyme abundances was remarkably higher than peptidase proportions. This was corroborated by early studies (27) and suggested a more pronounced specialization for carbohydrate degradation. All *Polaribacter* clades targeted β-glucans as a common substrate with different combinations of GH16, GH17, and GH30 family CAZymes. In total, seven variants of putative β-glucan PULs were detected (Fig. 7a) (Fig S3), indicating that each clade likely had its own strategy to utilize these compounds. β-glucans are diverse and widespread molecules in world oceans (73). For example, laminarin, a storage polysaccharide in diatoms and brown algae, has a β-1,3-glucan backbone that sometimes includes β-1,2 or β-1,6-glucose side chains (74). Future experiments should investigate whether β-glucan PULs with different composition are specific for particular β-glucans.

Polaribacter 2-a was abundant in the early phase of bloom events and could be characterized as a typical flavobacterial first responder (Fig. 3). Harboring the highest peptidase proportion, small genome size and limited polysaccharide utilization capacity (Fig. 6), *Polaribacter* 2-a also encoded a unique laminarin PUL variant with the combination of peptidases and CAZymes (Fig. 7b), which was expressed in 2009 and 2010 blooms (Fig. S6). This suggested a coupled carbohydrate and protein mechanisms, which could enable rapid growth during the onset of the bloom termination. Such a coupled mechanism was also found in spring-bloom associated *Formosa* spp. (75).

Polaribacter 3-a and *Polaribacter* 3-b responded later and targeted more complex sulfated polysaccharides to differentiate their glycan niche (Fig. 7a). *Polaribacter* 3-a was highly abundant only in the mid-phase of the 2010 bloom (Fig. 3) and had a distinctive PUL to utilize mannose-rich sulfated polysaccharides. Such substrates are, for example, found in diatom cell walls (76). Furthermore, *Polaribacter* 3-b was detected in the late phase of bloom events (Fig. 3). This clade had a distinctive sulfated xylan PUL together with a putative GH128-containing PUL (Fig. 7b), which was expressed in the 2009 spring bloom (Fig. S6). Besides being a component of marine phytoplankton (77), high xylan (β-1,4-xylose) hydrolysis rates were also

reported in many ocean provinces (78). Sulfated polysaccharides are likely more complex to degrade and maximum sulfatase expression was also found in the final stage of algal blooms (14). Therefore, distinctive sulfated polysaccharide degradation capacity could provide an ecological advantage for *Polaribacter* 3-a and 3-b to dominate the late phase of bloom events.

The *Polaribacter* 1-a clade represented “polysaccharide specialist” with the largest genome size, highest micro-diversity and broadest carbohydrate utilization capacity. Low peptidase proportions (Fig. 6) and high CAZyme gene abundances (Fig. S2) further demonstrated the distinct glycan niche of *Polaribacter* 1-a. Four sub-clades within *Polaribacter* 1-a possessed in total 38 putative PULs targeting 10 different substrates such as β-glucan, α-mannose-rich polysaccharides, sulfated xylan, N-acetyl-D-glucosamine, and alginate (Fig. 7a) (Fig. S5). The latter has so far only been described in brown macro-algae, indicating that *Polaribacter* 1-a does not only associated with planktonic microalgae during the bloom events. This observation was also corroborated by peptidase and CAZyme abundance pattern of *Polaribacter* 1-a, which was similar to macro-algae and marine animal associated species in the same genus. Remarkably, the peak abundance of *Polaribacter* 1-a was detected after blooms of *Chattonella* microalgae in 2009, 2011, and 2012 (Fig. 3). This could suggest a specific interaction, which remains to be tested in future studies using representative *Polaribacter* and *Chattonella* strains.

Polaribacter 1-b and 2-b represented “rare” clades with lower abundances and distinctive polysaccharide utilization potentials. *Polaribacter* 1-b was detected together with clade 1-a during 2009 and 2011 blooms and (Fig. 3) expressed the genes related to sulfated xylan PUL (Fig. S6). Furthermore, *Polaribacter* 2-b was detected in the late phase of 2010 bloom with a prominent expression of its putative GH37-containing PUL (Fig. S6) (Fig. 7a). This PUL could utilize trehalose rich polysaccharides, an abundant metabolite both in micro- (79) and macro-algae (80). Although many *Polaribacter* clades possessed a putative PUL for α-glucan usage (Fig. 7), expression of the genes related to this PUL was not detected in our metaproteome data (Fig. S6). Other heterotrophic bacteria abundant during the bloom events might be more competitive to degrade α-glucans, which are common storage compounds in marine algae, bacteria, and animals (81). For example,

'*Reinekea forsetii*', a gammaproteobacterial species isolated from the same sampling site, harbored fourteen GH13 genes possibly targeting α-glucans (46).

Highlighting the ecological importance of intra-genus diversity and metabolic diversification, our study shows how niche partitioning could enable different *Polaribacter* clades to avoid direct competition in a nutrient-rich environment with rapidly changing conditions. Our analyses also provide omics-based hypotheses, which will become testable in physiological experiments as soon as representative isolates are available. Such studies will allow for a more complete picture of the distinct niches of closely related bacteria co-occurring during algal blooms and lead to a better understanding of the ecological and evolutionary processes that drive microbial community composition in such dynamic environments.

3.5) Acknowledgements

We thank Jörg Wulf for technical support, T. Ben Francis for help in bioinformatics, and Carol Arnosti for critical reading. Jochen Blom's support in core and pan genome analysis is acknowledged. This study was funded by the Max Planck Society. Burak Avcı and Karen Krüger are members of the International Max Planck Research School of Marine Microbiology (MarMic).

3.6) Sequence Information

The North Sea *Polaribacter* MAG and genome sequences are available in this link: https://owncloud.mpi-bremen.de/index.php/s/aOXHIoXLB2E10to

3.7) References

1. Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH. Relative roles of niche and neutral processes in structuring a soil microbial community. ISME J. 2010;4(8):337-45.

2. Hubbell SP. The unified neutral theory of biodiversity and biogeography. Princeton: Princeton Univ. Press; 2001.

3. Keddy PA. Assembly and response rules: two goals for predictive community ecology. Journal of Veg Sci. 1992;3(2):157-64.

4. Hutchinson GE. Population Studies - Animal Ecology and Demography - Concluding Remarks. Cold Spring Harb Sym. 1957;22:415-27.

5. Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW. The merging of community ecology and phylogenetic biology. Ecol Lett. 2009;12(7):693-715.

6. Gauze GF. The struggle for existence. Baltimore: The Williams & Wilkins Company; 1934.

7. Meier DV, Pjevac P, Bach W, Hourdez S, Girguis PR, Vidoudez C, et al. Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. ISME J. 2017;11(7):1545-58.

8. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: Integrating terrestrial and oceanic components. Science. 1998;281(5374):237-40.

9. Camarena-Gomez MT, Lipsewers T, Piiparinen J, Eronen-Rasimus E, Perez-Quemalinos D, Hoikkala L, et al. Shifts in phytoplankton community structure modify bacterial production, abundance and community composition. Aquat Microb Ecol. 2018;81(2):149-70.

10. Niu Y, Shen H, Chen J, Xie P, Yang X, Tao M, et al. Phytoplankton community succession shaping bacterioplankton community composition in Lake Taihu, China. Water Res. 2011;45(14):4169-82.

11. Tada Y, Taniguchi A, Nagao I, Miki T, Uematsu M, Tsuda A, et al. Differing growth responses of major phylogenetic groups of marine bacteria to natural phytoplankton blooms in the Western North Pacific Ocean. Appl Environ Microb. 2011;77(12):4055-65.

12. Williams TJ, Wilkins D, Long E, Evans F, DeMaere MZ, Raftery MJ, et al. The role of planktonic *Flavobacteria* in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. Environ Microbiol. 2013;15(5):1302-17.

13. Georges AA, El-Swais H, Craig SE, Li WKW, Walsh DA. Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. ISME J. 2014;8(6):1301-13.

14. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton Bloom. Science. 2012;336(6081):608-11.

15. Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L, et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. eLIFE. 2016;5: e11888.

16. Chafee M, Fernandez-Guerra A, Buttigieg PL, Gerdts G, Eren AM, Teeling H, et al. Recurrent patterns of microdiversity in a temperate coastal marine environment. ISME J. 2018;12(1):237-52.

17. Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science. 2008;320(5879):1081-5.

18. Simek K, Kasalicky V, Hornak K, Hahn MW, Weinbauer MG. Assessing niche separation among coexisting *Limnohabitans* strains through interactions with a competitor, viruses, and a bacterivore. Appl Environ Microb. 2010;76(5):1406-16

19. Gosink JJ, Woese CR, Staley JT. *Polaribacter* gen. nov., with three new species, *P. irgensii* sp. nov., *P. franzmannii* sp. nov., and *P. filamentus* sp. nov., gas vacuolate polar marine bacteria of the Cytophaga-Flavobacterium-Bacteroides group and reclassification of *'Flectobacillus glomeratus*' as *Polaribacter glomeratus* comb. nov. Int J Syst Bacteriol. 1998;48:223-35.

20. Yoon JH, Kang SJ, Oh TK. *Polaribacter dokdonensis* sp nov., isolated from seawater. Int J Syst Evol Micr. 2006;56:1251-5.

21. Park S, Yoon SY, Ha MJ, Yoon JH. *Polaribacter litorisediminis* sp nov., isolated from a tidal flat. Int J Syst Evol Micr. 2017;67(6):2036-42.

22. Kim E, Shin SK, Choi S, Yi H. *Polaribacter vadi* sp nov., isolated from a marine gastropod. Int J Syst Evol Micr. 2017;67(1):144-7.

23. Nedashkovskaya OI, Kukhlevskiy AD, Zhukova NV. *Polaribacter reichenbachii* sp. nov.: a new marine bacterium associated with the green alga *Ulva fenestrata*. Curr Microbiol. 2013;66(1):16-21.

24. Bunse C, Bertos-Fortis M, Sassenhagen I, Sildever S, Sjoqvist C, Godhe A, et al. Spatio-temporal interdependence of bacteria and phytoplankton during a Baltic Sea Spring Bloom. Front Microbiol. 2016;7(517):doi:10.3389/fmicb.2016.00517.

25. Delmont TO, Eren AM, Vineis JH, Post AF. Genome reconstructions indicate the partitioning of ecological functions inside a phytoplankton bloom in the Amundsen Sea, Antarctica. Front Microbiol. 2015;6(1090):doi:10.3389/fmicb.2015.01090.

26. Rapp JZ, Fernández-Méndez M, Bienhold C, Boetius A. Effects of ice-algal aggregate export on the connectivity of bacterial communities in the central Arctic Ocean. Front Microbiol. 2018;9(1035):doi:10.3389/fmicb.2018.01035.

27. Xing P, Hahnke RL, Unfried F, Markert S, Huang SX, Barbeyron T, et al. Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. ISME J. 2015;9(6):1410-22.

28. Pruesse E, Peplies J, Glöckner FO. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28(14):1823-9.

29. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(D1):D590-6.

30. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al. ARB: a software environment for sequence data. Nucleic Acids Res. 2004;32(4):1363-71.

31. Stamatakis A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688-90.

32. Peplies J, Kottmann R, Ludwig W, Glöckner FO. A standard operating procedure for phylogenetic inference (SOPPI) using (rRNA) marker genes. Syst Appl Microbiol. 2008;31(4):251-7.

33. Thiele S, Fuchs BM, and Amann RI. Identification of microorganisms using the ribosomal RNA approach and fluorescence *in situ* hybridization. In: Wilderer P, editor. Treatise on Water Science. Oxford: Elsevier Science; 2011. p. 171–89.

34. Bennke CM, Reintjes G, Schattenhofer M, Ellrott A, Wulf J, Zeder M, et al. Modification of a high-throughput automatic microbial cell enumeration system for shipboard analyses. Appl Environ Microb. 2016;82(11):3289-96.

35. Francis TB, Krüger K, Fuchs BM, Teeling H, Amann RI. *Candidatus* Prosiliicoccus vernus, a spring bloom associated member of the *Flavobacteriaceae*. Syst Appl Microbiol. 2018;doi:10.1016/j.syapm.2018.08.007.

36. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77.

37. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11(11):1144-6.

38. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3:e1319.

39. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043-55.

40. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17(1):132.

41. Rodriguez-R LM, Konstantinidis K. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints. 2016;4:e1900v1.

42. Team RC. R: A language and environment for statistical computing. 2013;2.

43. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.

44. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.

45. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST server: Rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.

46. Avcı B, Hahnke RL, Chafee M, Fischer T, Gruber-Vodicka H, Tegetmeyer HE, et al. Genomic and physiological analyses of '*Reinekea forsetii*' reveal a versatile opportunistic lifestyle during spring algae blooms. Environ Microbiol. 2017;19(3):1209-21.

47. Huang L, Zhang H, Wu PZ, Entwistle S, Li XQ, Yohe T, et al. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Res. 2018;46(D1):516-21.

48. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44(D1):279-85.

49. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(D1):D490-5.

50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol. 1990;215(3):403-10.

51. Rawlings ND, Alan J, Thomas PD, Huang XD, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 2018;46(D1):624-32.

52. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 2013;41(12)e121.

53. Terrapon N, Lombard V, Drula E, Lapebie P, Al-Masaudi S, Gilbert HJ, et al. PULDB: the expanded database of Polysaccharide Utilization Loci. Nucleic Acids Res. 2018;46(D1):677-83.

54. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2017;DOI 10.1093/bib/bx108.

55. Blom J, Kreis J, Spanig S, Juhre T, Bertelli C, Ernst C, et al. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. Nucleic Acids Res. 2016;44(W1):22-8.

56. Lerat E, Daubin V, Moran NA. From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-proteobacteria. PLOS Biol. 2003;1(1):101-9.

57. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28(1):33-6.

58. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 2017;34(8):2115-22.

59. Kappelmann L, Krüger K, Hehemann J-H, Harder J, Markert S, Unfried F, et al. Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. ISME J. 2018;doi: 10.1038/s41396-018-0242-6.

60. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol. 2014;12(9):635-45.

61. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Micr. 2007;57(1):81-91.

62. Fernández-Gómez B, Richter M., Schüler M, Pinhassi J, Acinas SG, González JM & Pedrós-Alió, C. Ecology of marine *Bacteroidetes*: a comparative genomics approach. ISME J. 2013;7(5):1026–37.

63. Temperton B, Gilbert JA, Quinn JP, McGrath JW. Novel analysis of oceanic surface water metagenomes suggests importance of polyphosphate metabolism in oligotrophic environments. PLOS One. 2011;6(1):e16499.

64. Hemsworth GR, Déjean G, Davies Gideon J, Brumer H. Learning from microbial strategies for polysaccharide degradation. Biochem Soc T. 2016;44(1):94-108.

65. Roy S. Distributions of phytoplankton carbohydrate, protein and lipid in the world oceans from satellite ocean colour. ISME J. 2018;12(6):1457-72.

66. Orsi WD, Smith JM, Liu S, Liu Z, Sakamoto CM, Wilken S, et al. Diverse, uncultivated bacteria and archaea underlying the cycling of dissolved protein in the ocean. ISME J. 2016;10(9):2158-73.

67. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. Polysaccharide utilization loci: fueling microbial communities. J Bacteriol. 2017;199(15):e00860-16.

68. Kabisch A, Otto A, König S, Becher D, Albrecht D, Schüler M, et al. Functional characterization of polysaccharide utilization loci in the marine *Bacteroidetes 'Gramella forsetii'* KT0803. ISME J. 2014;8(7):1492-502.

69. Foley MH, Martens EC, Koropatkin NM. SusE facilitates starch uptake independent of starch binding in *B. thetaiotaomicron*. Mol Microbiol.108(5):551-66.

70. Chiovitti A, Harper RE, Willis A, Bacic A, Mulvaney P, Wetherbee R. Variations in the substituted 3-linked mannans closely associated with the silicified walls of diatoms. J Phycol. 2005;41(6):1154-61.

71. Cuskin F, Lowe EC, Temple MJ, Zhu YP, Cameron EA, Pudlo NA, et al. Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. Nature. 2015;517(7533):165-U86.

72. Despres J, Forano E, Lepercq P, Comtet-Marre S, Jubelin G, Chambon C, et al. Xylan degradation by the human gut *Bacteroides xylanisolvens* XB1A^T involves two distinct gene clusters that are linked at the transcriptional level. BMC Genomics. 2016;17:326.

73. Alderkamp AC, van Rijssel M, Bolhuis H. Characterization of marine bacteria and the activity of their enzyme systems involved in degradation of the algal storage glucan laminarin. Fems Microbiol Ecol. 2007;59(1):108-17.

74. Gügi B, Le Costaouëc T, Burel C, Lerouge P, Helbert W, Bardor M. Diatom-specific oligosaccharide and polysaccharide structures help to unravel biosynthetic capabilities in diatoms. Mar Drugs. 2015;13(9):5993-6018.

75. Unfried F, Becker S, Robb C, Hehemann J-H, Markert S, Heiden S, et al. Adaptive mechanisms that provide competitive advantages to marine *Bacteroidetes* during microalgal blooms. ISME J. 2018;doi:10.1038/s41396-018-0243-5.

76. Le Costaouëc T, Unamunzaga C, Mantecon L, Helbert W. New structural insights into the cell-wall polysaccharide of the diatom *Phaeodactylum tricornutum*. Algal Res. 2017;26:172-9.

77. Parsons TR, Stephens K, Strickland JDH. On the chemical composition of eleven species of marine phytoplankters. J Fish Res Board of Can. 1961;18(6):1001-16.

78. Arnosti C. Microbial Extracellular Enzymes and the Marine Carbon Cycle. Annu Rev of Mar Sci. 2011;3(1):401-25.

79. Hirth M, Liverani S, Mahlow S, Bouget FY, Pohnert G, Sasso S. Metabolic profiling identifies trehalose as an abundant and diurnally fluctuating metabolite in the microalga *Ostreococcus tauri*. Metabolomics. 2017;13(6)68.

80. Michel G, Tonon T, Scornet D, Cock JM, Kloareg B. Central and storage carbon metabolism of the brown alga *Ectocarpus siliculosus*: insights into the origin and evolution of storage carbohydrates in Eukaryotes. New Phytol. 2010;188(1):67-81.

81. Viola R, Nyvall P, Pedersen M. The unique features of starch metabolism in red algae. P Roy Soc B-Biol Sci. 2001;268(1474):1417-22.

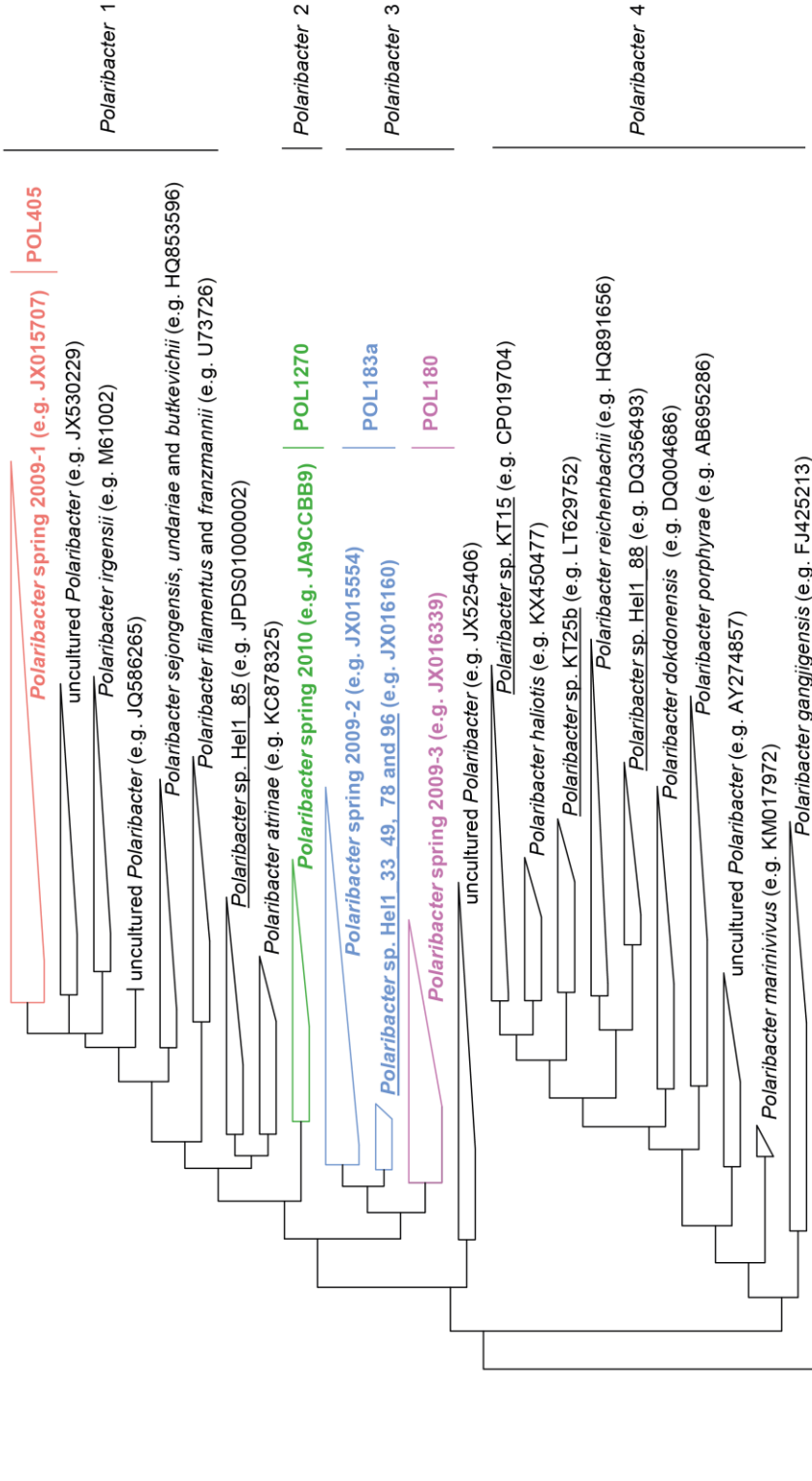


Fig. 1. Consensus phylogenetic tree of *Polaribacter*-related 16S rRNA sequences including clones and isolates obtained during spring blooms at Helgoland. Brackets indicate the four *Polaribacter* clusters and the specificity of newly designed oligonucleotide probes. Relative abundances of representative FISH probes are shown in Fig. 3. Representative MAGs in Fig. 2 are depicted by the same color. North Sea *Polaribacter* isolates are underlined. Bar: 0.1 substitutions per nucleotide position. *Tenacibaculum*-related sequences were used as the outgroup.

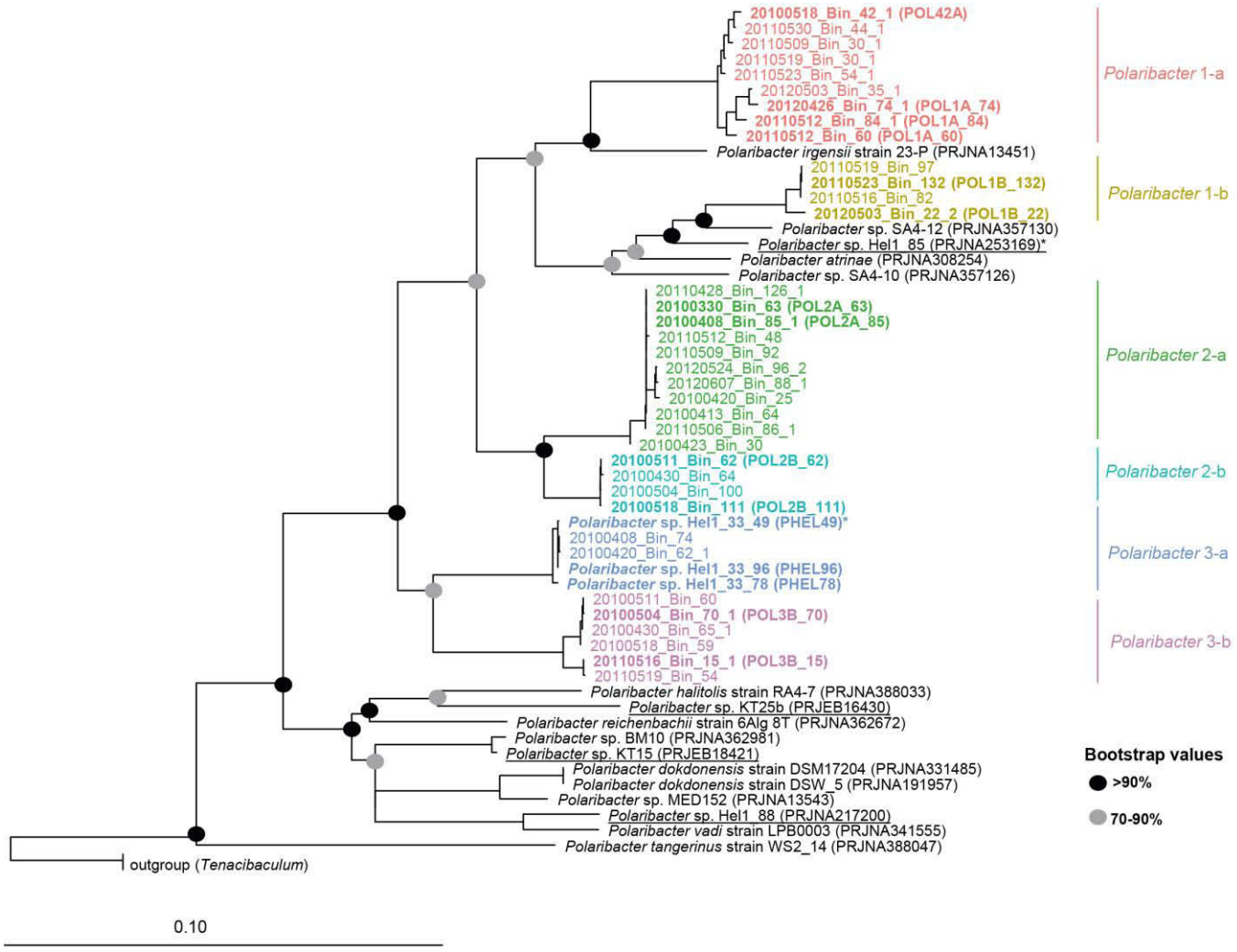


Fig. 2. Maximum likelihood phylogenetic tree of 43 conserved marker gene protein sequences from spring bloom associated *Polaribacter* genomes and MAGs. The MAGs which are used in further analyses are shown in bold with their abbreviations. For the genomes retrieved from NCBI, BioProject numbers are shown in brackets. The genomes of North Sea *Polaribacter* isolates are underlined. The range of bootstrap values is indicated with black and grey circles. The results of metagenome read recruitment on the corresponding MAGs are depicted in Fig. 3. North Sea *Polaribacter* isolates Hel1_85 and Hel1_33_49 were previously compared in (27). Bar: 0.1 substitutions per aminoacid position. *Tenacibaculum* were used as the outgroup.

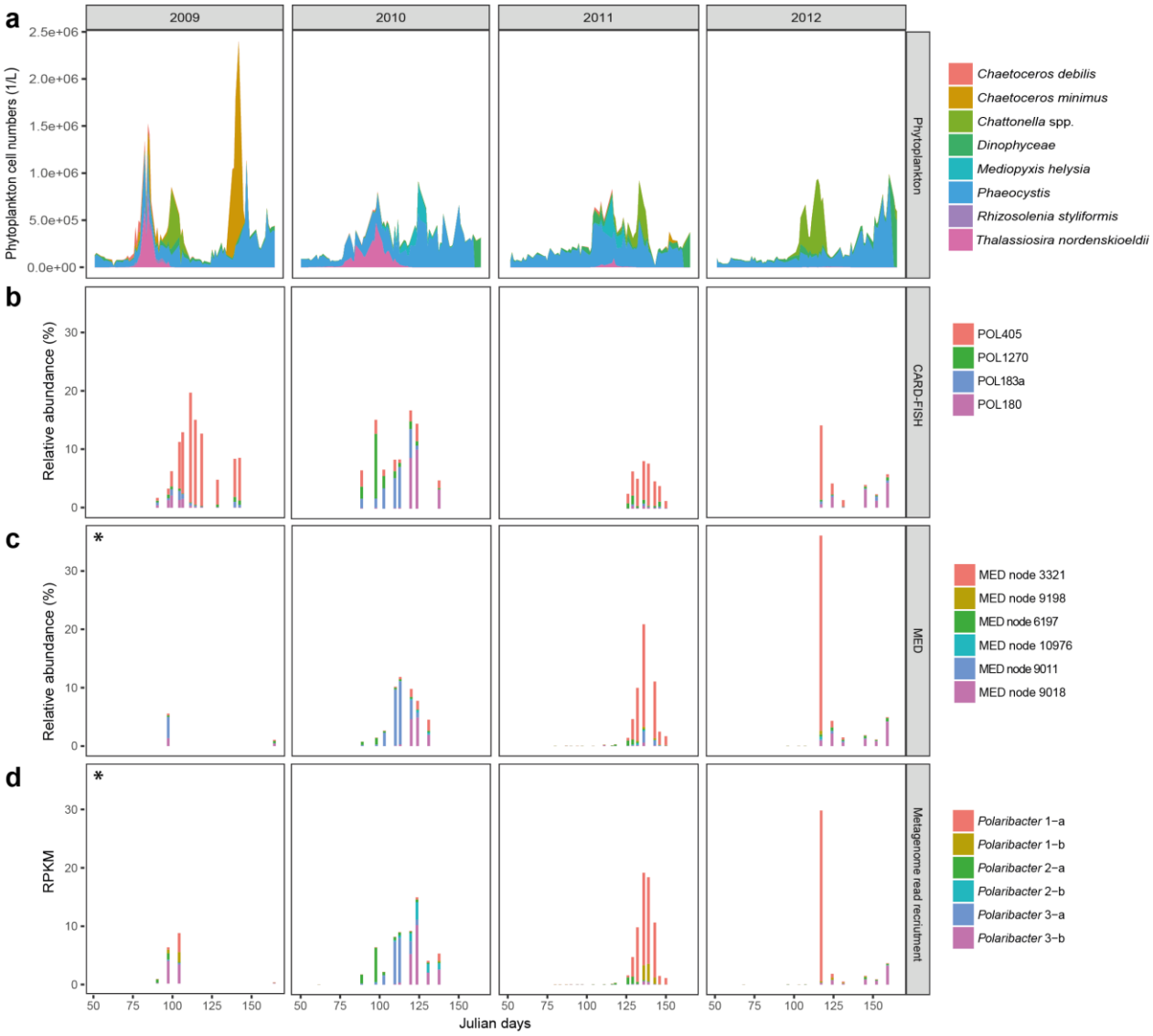


Fig. 3. Temporal succession of six distinct *Polaribacter* clades during North Sea spring algal blooms. **(a)** Cell counts of eight dominant phytoplankton as reported in (15). **(b)** CARD-FISH analysis using newly designed oligonucleotide probes targeting four major *Polaribacter* clades. **(c)** Relative abundances of the six most abundant MED nodes as retrieved from (16). **(d)** Metagenome read recruitment on the *Polaribacter* MAGs and genomes. Number of recruited reads are normalized to the bin/genome size and reported as reads per kilobase million (RPKM). Taxonomic units in each method with a high level of correlation based on the Spearman test (Table S6) are shown in the same color. (*) Data from 2009 spring bloom are sparse, consisting of only two and four sampling dates for MED and metagenome analyses, respectively.

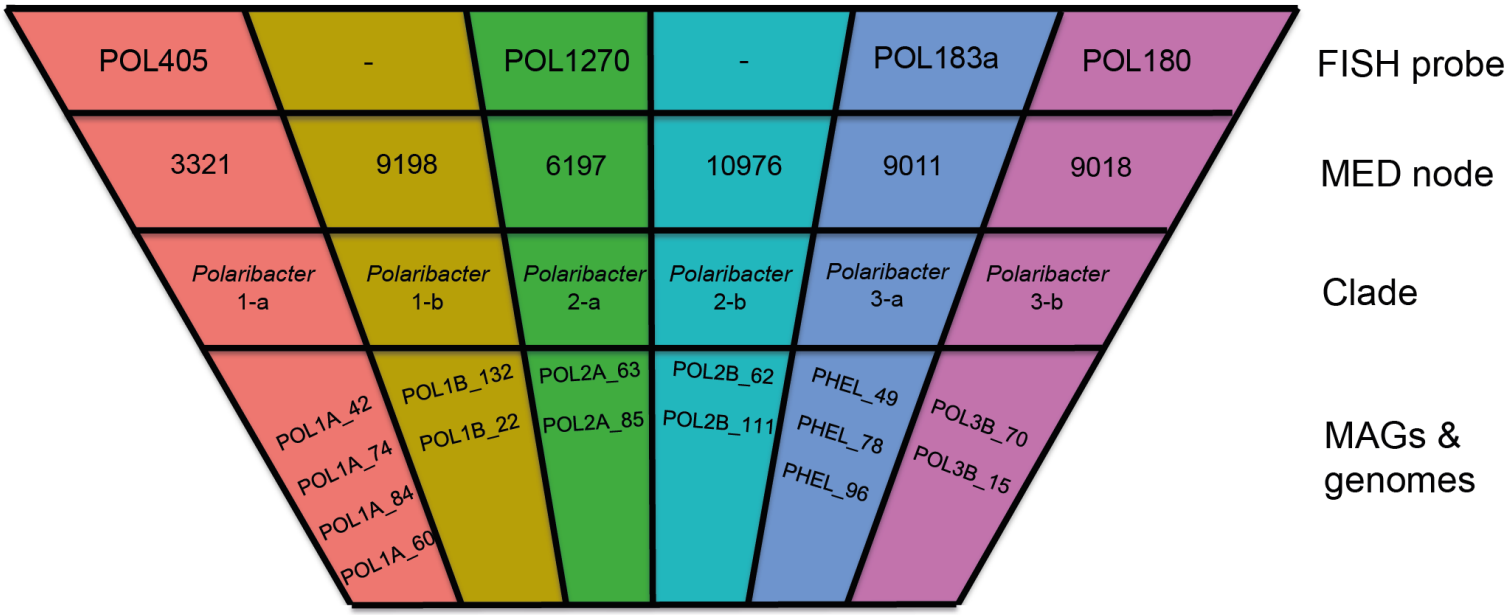


Fig. 4. The key for *Polaribacter* nomenclature in this study. Distinct taxonomic units detected by various methods are linked to each other according to Spearman correlation (Table S6).

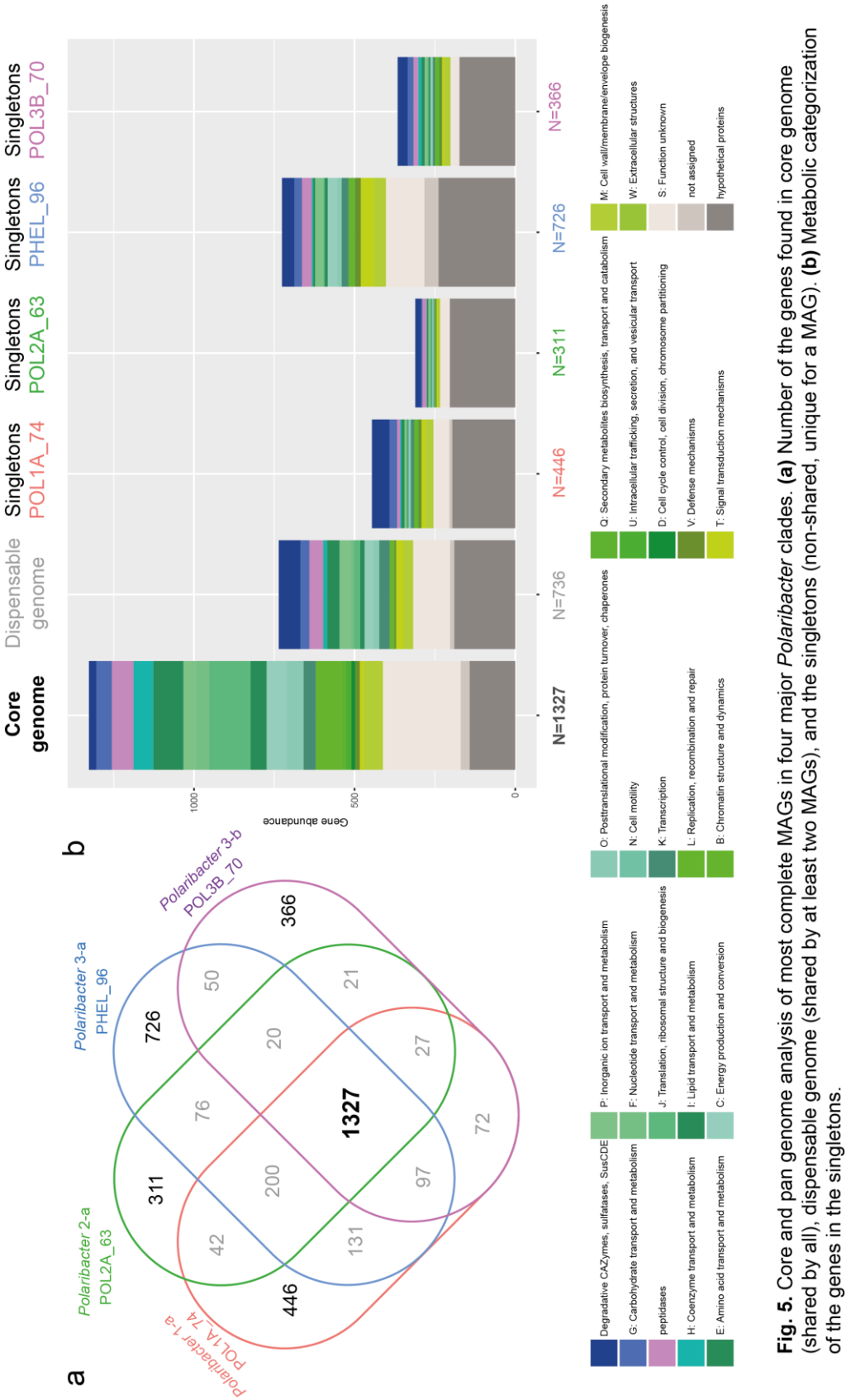


Fig. 5. Core and pan genome analysis of most complete MAGs in four major *Polaribacter* clades. **(a)** Number of the genes found in core genome (shared by all), dispensable genome (shared by at least two MAGs), and the singletons (non-shared, unique for a MAG). **(b)** Metabolic categorization of the genes in the singletons.

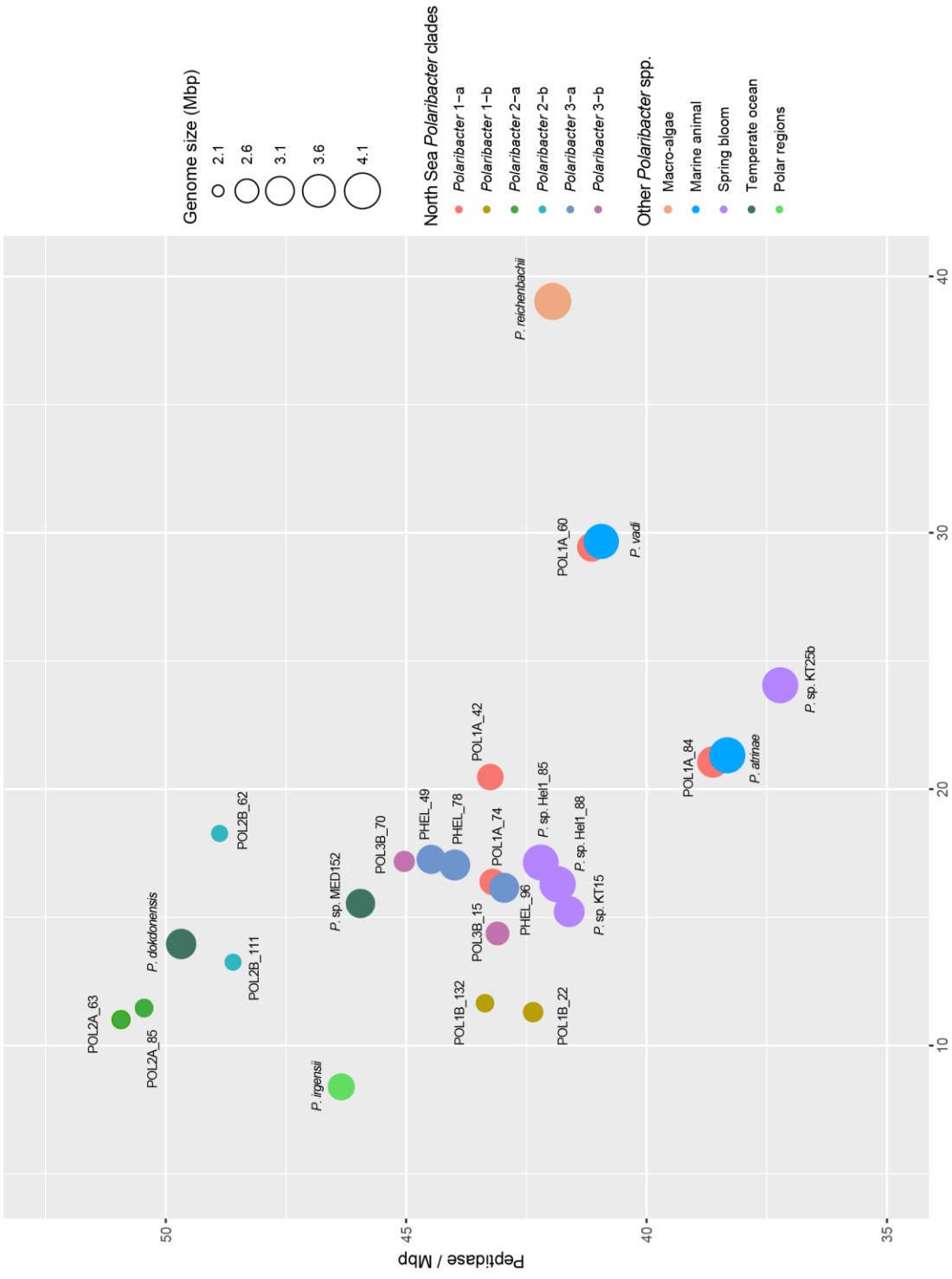


Fig. 6. Comparison of peptidase and degradative CAZyme (GH+PL+CE) repertoires of North Sea *Polaribacter* clades and the described species or isolates with sequenced genomes in the same genus. Isolation site of other *Polaribacter* spp. are also depicted by different colors. Gene abundances are normalized with the MAG or genome sizes. Full MAG sizes are estimated based on the completeness values calculated via CheckM (39).

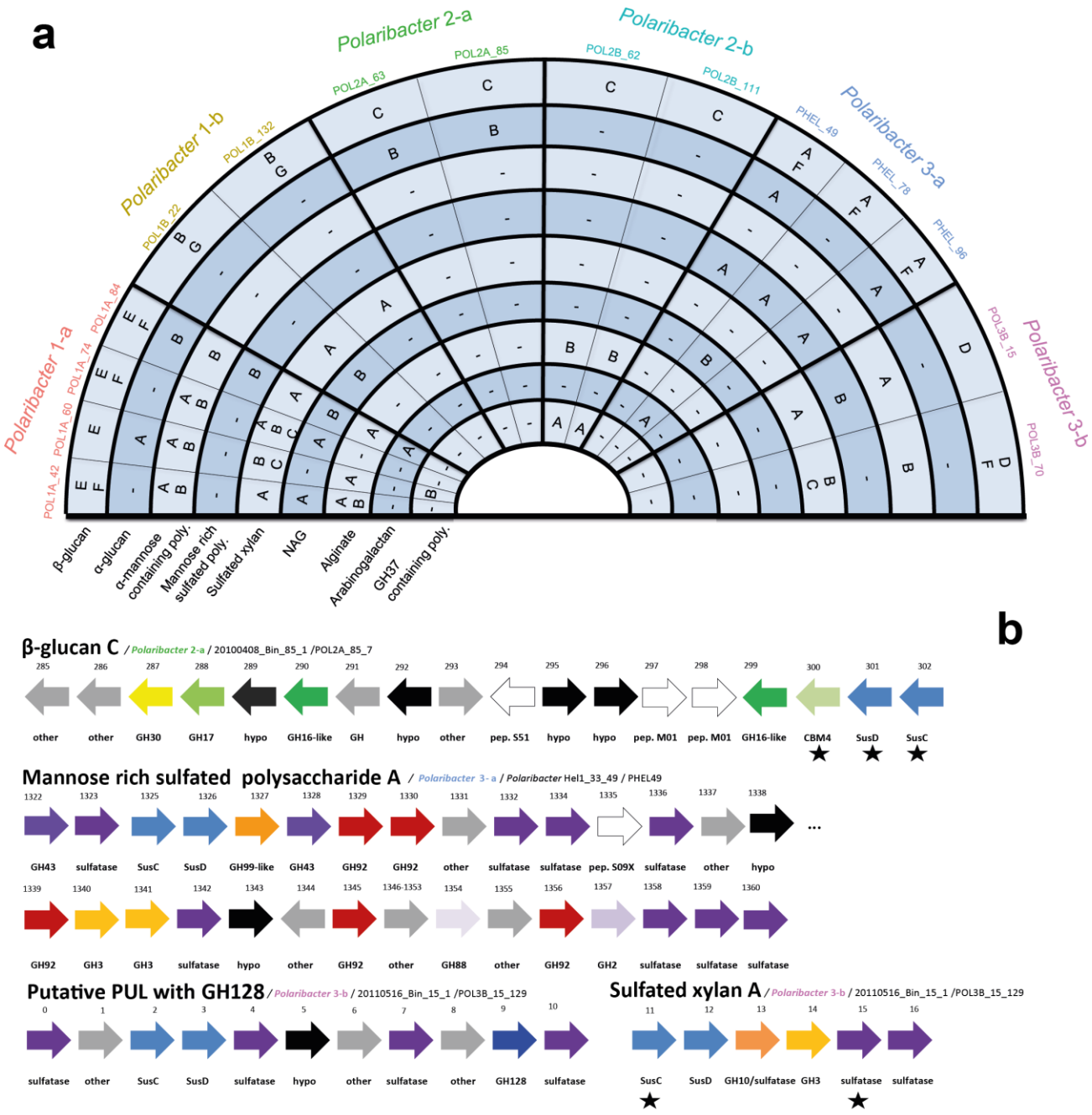


Fig. 7. Putative PUL spectra of *Polaribacter* clades. **(a)** Distribution of PULs with predicted substrates across six distinct *Polaribacter* clades. Letters in the circle slices represent the PUL variants for individual substrates as it is shown with phylogenetic analysis of *SusC* gene protein sequences in Fig. S4. **(b)** Gene content of some distinctive PULs in *Polaribacter* clades with the PUL variants shown in panel a. Black stars show the expressed genes based on metaproteome analysis (Fig. S6). Contig and gene IDs are shown for each PUL. All PUL repertoires are presented in Fig S5. CAZymes, peptidases, and Sus transport system genes are depicted with different colors. Hypothetical proteins and genes involved in other metabolic functions are abbreviated with "hypo" and "other" respectively.

3.8) Supporting Information

Polysaccharide-driven niche differentiation between distinct *Polaribacter* clades during North Sea spring algal blooms

Burak Avcı[1], Karen Krüger[1], Bernhard M. Fuchs[1], Hanno Teeling[1], Rudolf I. Amann[1]*

[1] Department of Molecular Ecology, Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359, Bremen, Germany

* Corresponding author: Rudolf I. Amann, Department of Molecular Ecology, Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany, E-mail: ramann@mpi-bremen.de, Phone: +49 421 2028 930, Fax: +49 421 2028 790

Other polysaccharides targeted by North Sea Polaribacter *clades*

N-acetyl-D-glucosamine containing polysaccharides, alginate, and arabinogalactan were likely degraded by particular *Polaribacter* taxa as well (Fig. 7a). MAG annotations suggested that *Polaribacter* clades 1-a, 1-b, and 3-a were able to utilize peptidoglycan monomer N-acetyl-D-glucosamine (NAG), a cell-wall component which is released when bacteria disintegrate during algal blooms (1). A putative PUL consisting of GH20 (e.g. N-acetylglucosaminidase) and NAG transporter components (Fig. S4) was also encoded and expressed in *Polaribacter* 1-a during the late phase of 2009 bloom (Fig. S5). Furthermore, alginate is a cell wall polysaccharide (2) and an intracellular component of brown algae (3), and *Polaribacter* 1-a and 2-b encoded putative alginate PULs containing PL6, PL7, and PL17 (all include alginate lyases) in common (Fig. 7a) (Fig. S4). The MAGs POL1A_42 and POL1A_60 also translated some *SusCD* genes in these PULs during the late phase of 2009, 2010 and 2011 blooms (Fig. S5). A similar PUL with high synteny was shown to be upregulated with alginate in *Gramella forsetii* KT0803^T (4). Harboring endo-beta-1,4-galactanase (GH family 53) genes, *Polaribacter* clades 1-a and 3-a also had putative PULs to target arabinogalactan (Fig. 7a) (Fig. S4). It is noteworthy that microcolonies of North Sea *Polaribacter* spp. were also detected with galactose-specific lectins, which is a monomer of arabinogalactan (5).

References

1. Avcı B, Hahnke RL, Chafee M, Fischer T, Gruber-Vodicka H, Tegetmeyer HE, et al. Genomic and physiological analyses of '*Reinekea forsetii*' reveal a versatile opportunistic lifestyle during spring algae blooms. Environ Microbiol. 2017;19(3):1209-21.

2. Davis TA, Volesky B, Mucci A. A review of the biochemistry of heavy metal biosorption by brown algae. Water Res. 2003;37(18):4311-30.

3. Haug A, Larsen B, Smidsrod O. A study of the constitution of alginic acid by partial acid hydrolysis. Acta Chem Scand. 1966;20:183–90.

4. Kabisch A, Otto A, König S, Becher D, Albrecht D, Schüler M, et al. Functional characterization of polysaccharide utilization loci in the marine *Bacteroidetes* '*Gramella forsetii*' KT0803. ISME J. 2014;8(7):1492-502.

5. Bennke CM, Neu TR, Fuchs BM, Amann R. Mapping glycoconjugate-mediated interactions of marine *Bacteroidetes* with diatoms. Syst Appl Microbiol. 2013;36(6):417-25.

Table S2, Table S5, Table S8, Table S9, and Table S11 are available in the CD-ROM, which is provided with the thesis.

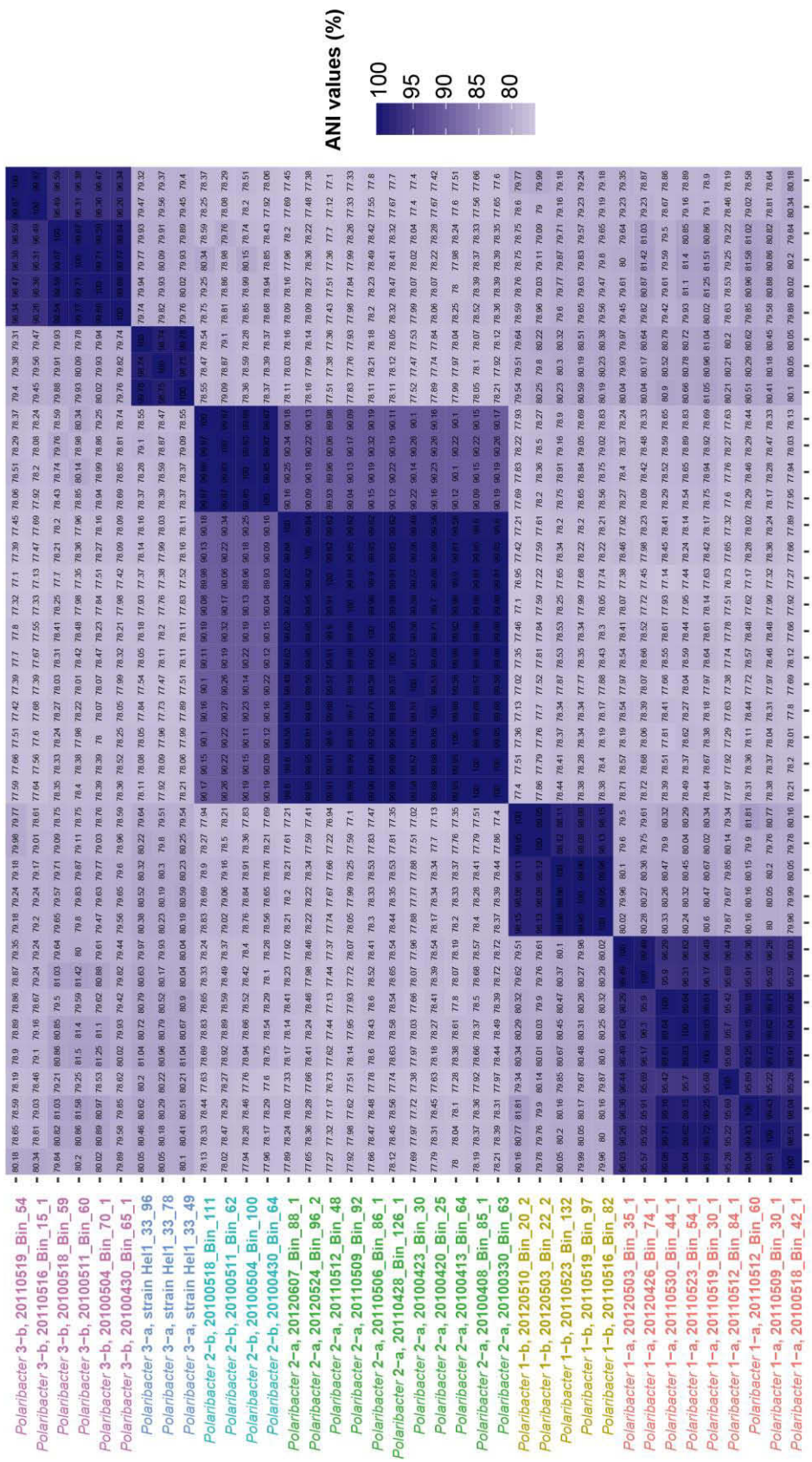


Figure S1. ANI comparison of *Polaribacter* MAGs and genomes.

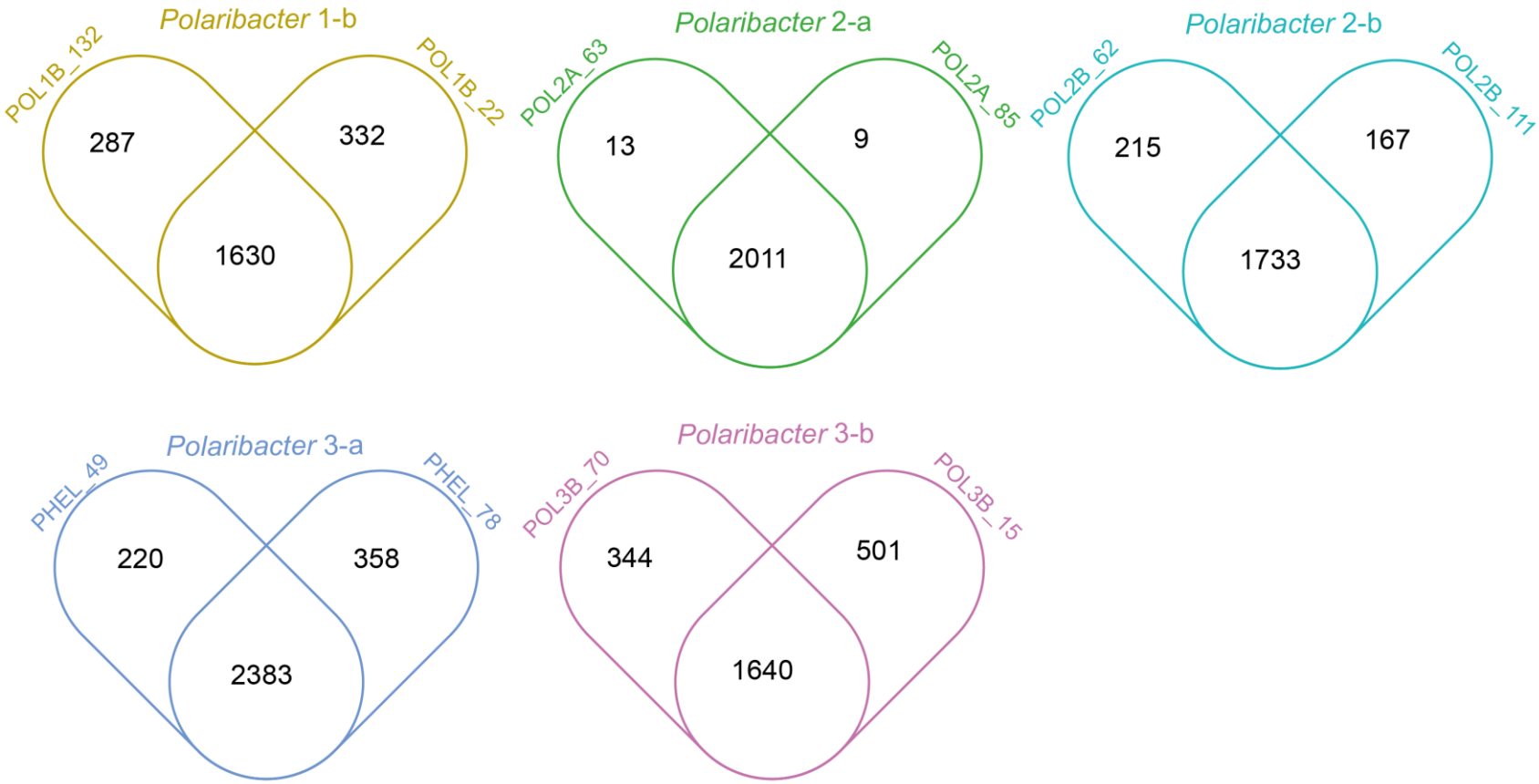


Fig. S2. Core and pan genome analysis of the representative MAGs and genomes in *Polaribacter* clades 1-b, 2-a, 2-b, 3-a, 3-b. Number of the genes detected in the core genome (shared by all) and singletons (unique for a MAG or genome) are shown.

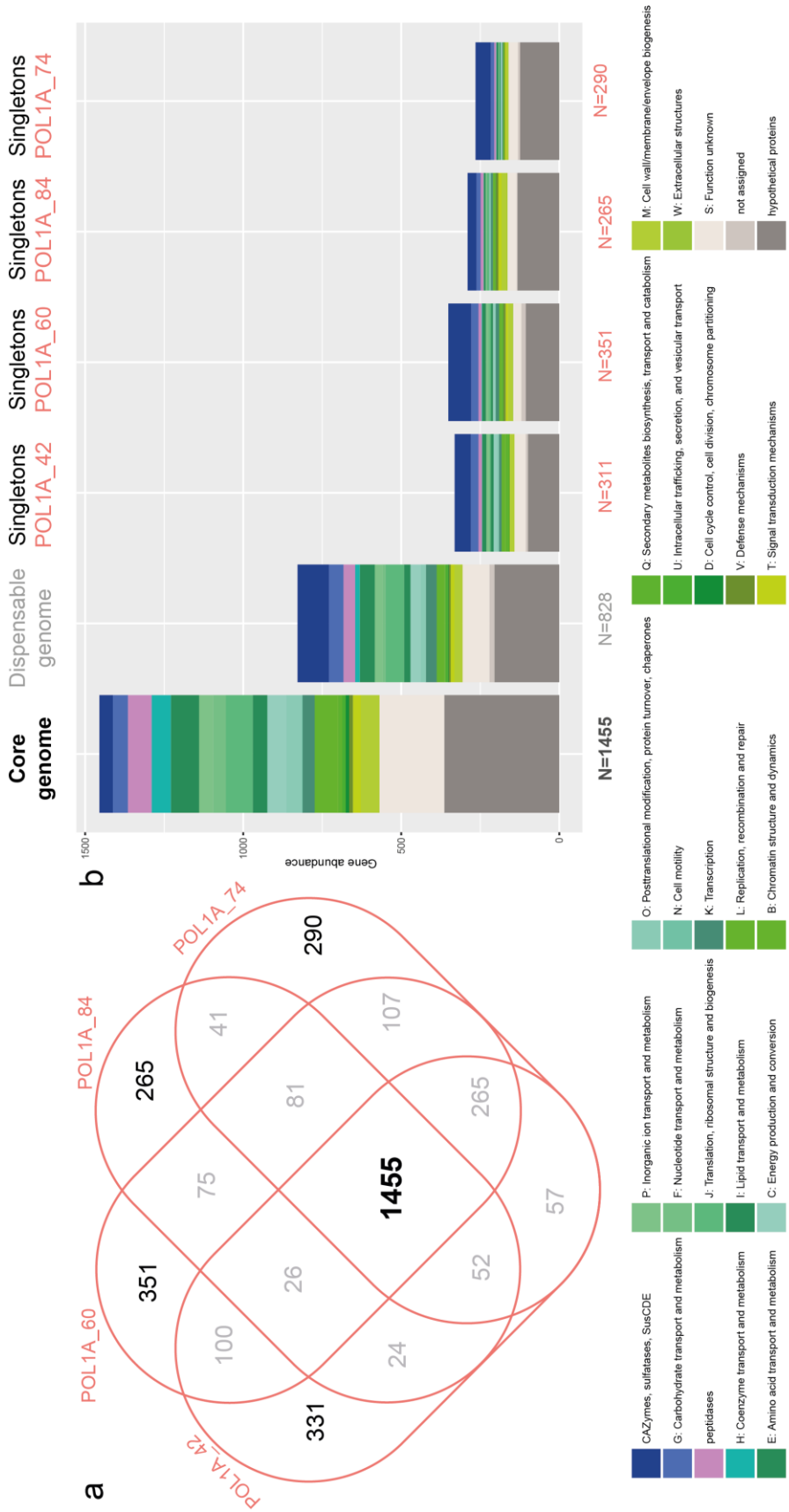


Fig. S3. Core and pan genome analysis of all MAGs in *Polaribacter* 1-a. **(a)** Number of the genes found in core genome (shared by all MAGs), dispensable genome (shared by at least two MAGs), and the singletons (non-shared, unique for a MAG). **(b)** Metabolic categorization of the genes in the singletons.

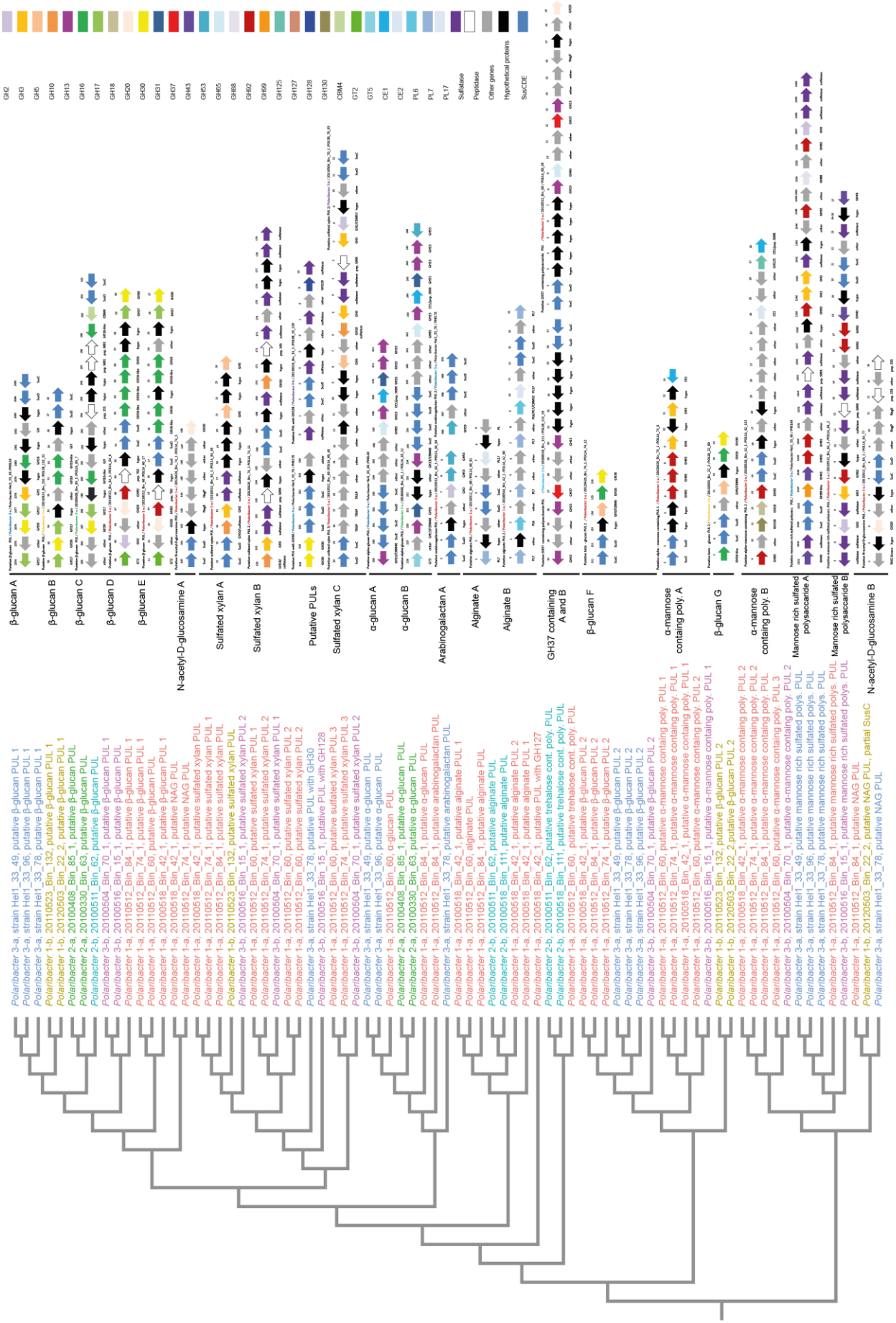
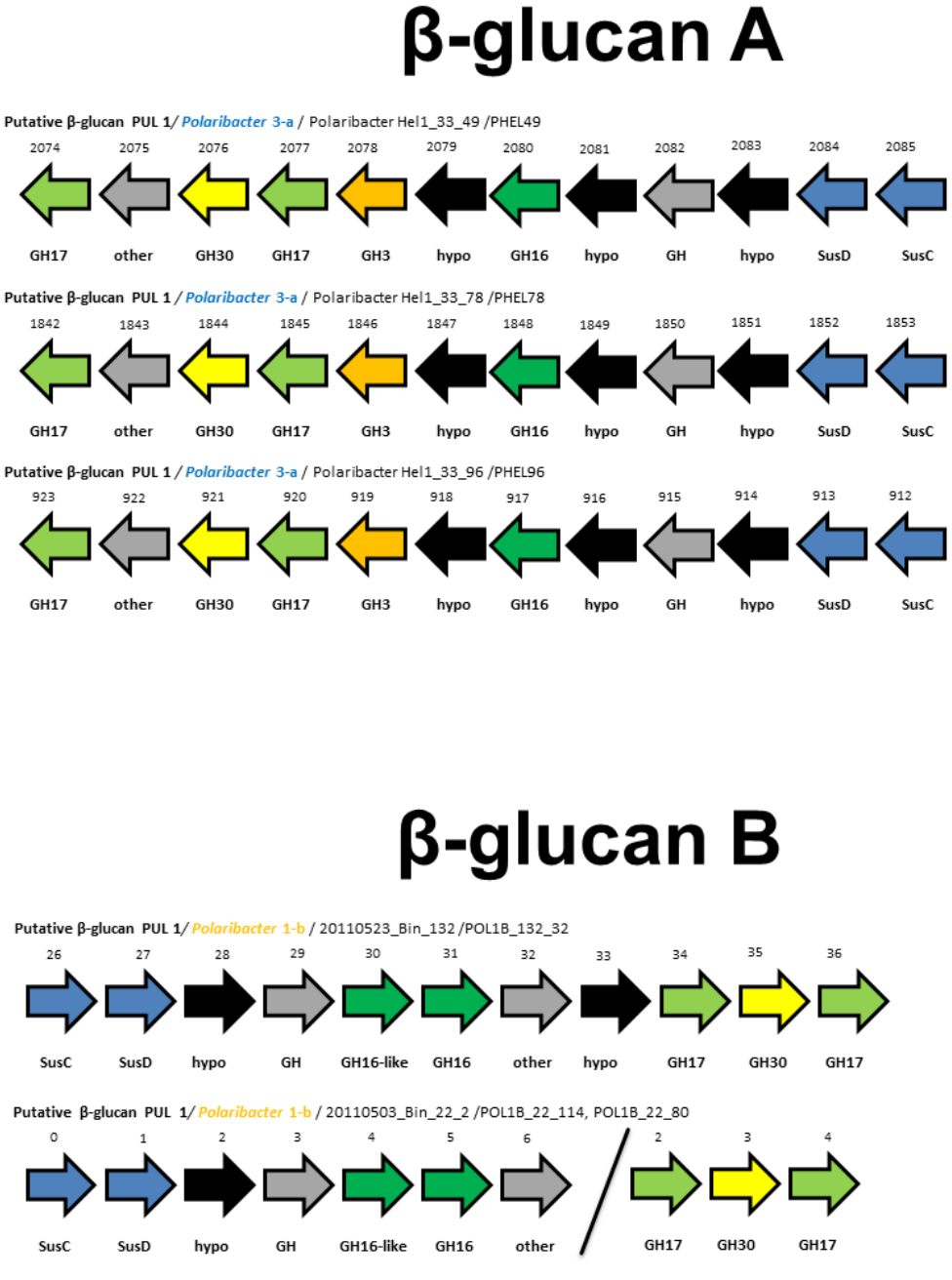


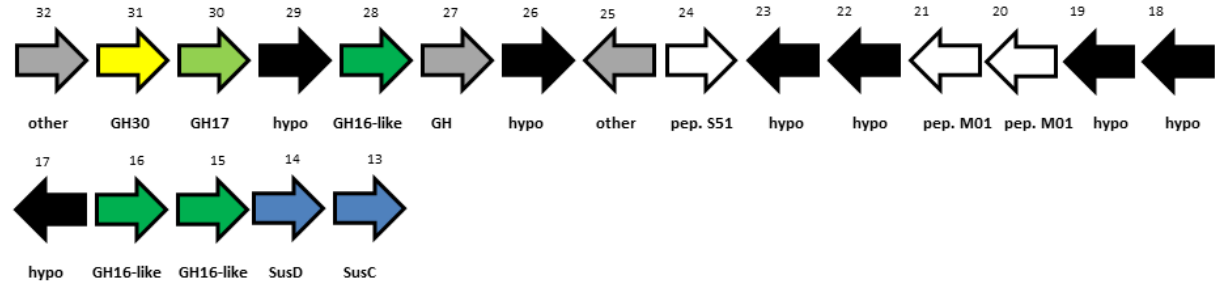
Fig. S4. Neighbor joining phylogenetic tree of *SusC* gene protein sequences found in *Polaribacter* PULs showing the substrate-dependent clustering and different variants of the PULs targeting the same substrate. PULs with partial *SusC* sequences were not considered in the tree computation. The longest PUL in each substrate cluster is shown as a cluster representative. All PUL repertoire of *Polaribacter* clades are shown in Fig. S5.

Fig. S5. Cumulative PUL repertoire of all *Polaribacter* clades. All PULs annotated in *Polaribacter* clades are shown with the predicted substrate together with its variant as provided in Fig. S4. Gene clusters containing CAZymes without any SusCD pair and the PULs with partial *SusC* or *SusD* gene sequences are also provided. Contig and gene IDs are also written for each PUL. CAZymes, peptidases, and Sus transport system genes are depicted with different colors. Partial PULs at the contig boundaries are shown with dashes. Hypothetical proteins and genes involved in other metabolic functions are abbreviated with "hypo" and "other" respectively.

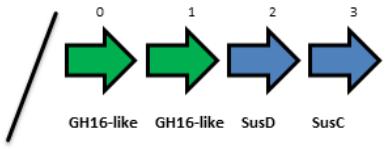


β-glucan C

Putative β-glucan PUL / *Polaribacter* 2-b / 20100511_Bin_62 /POL2B_62_168

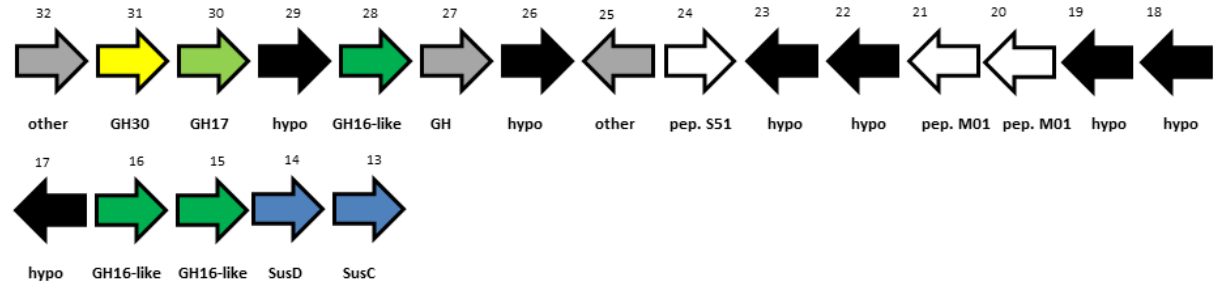


Putative β-glucan PUL with partial SusC / *Polaribacter* 2-b / 20100518_Bin_111 /POL2B_12

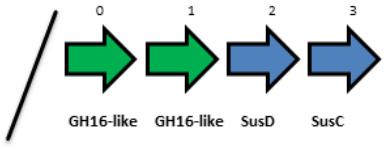


β-glucan C

Putative β-glucan PUL / *Polaribacter* 2-b / 20100511_Bin_62 /POL2B_62_168

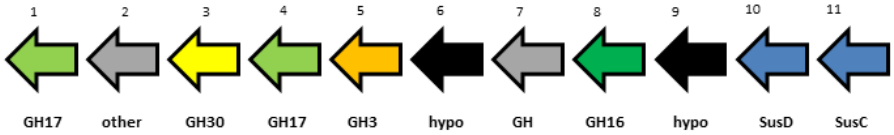


Putative β-glucan PUL with partial SusC / *Polaribacter* 2-b / 20100518_Bin_111 /POL2B_12

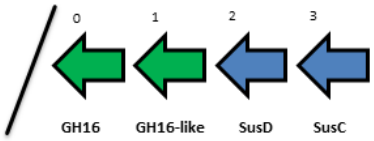


β-glucan D

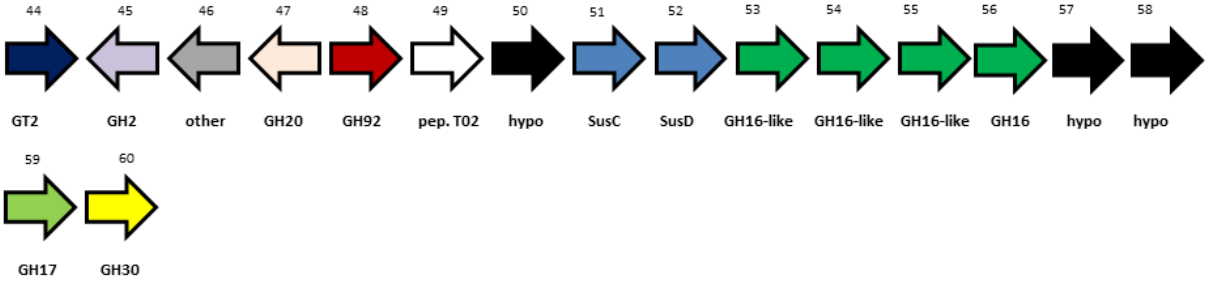
Putative β-glucan PUL 1 / *Polaribacter* 3-b / 20110504_Bin_70_1 /POL3B_70_4



Putative β-glucan PUL / *Polaribacter* 3-b / 20110516_Bin_15_1 /POL3B_15_170

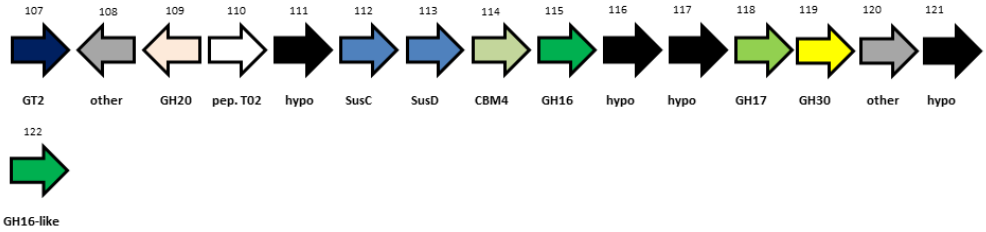


Putative β-glucan PUL 1/ *Polaribacter* 1-a / 20110512_Bin_84_1 /POL1A_84_6

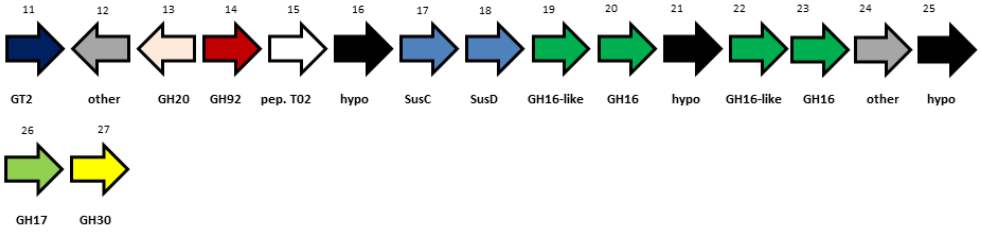


β-glucan E

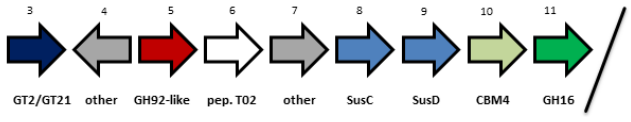
Putative β-glucan PUL 1 / *Polaribacter* 1-a / 20120426_Bin_74_1 /POL1A_74_12



Putative β-glucan PUL / *Polaribacter* 1-a / 20110512_Bin_60 /POL1A_60_17

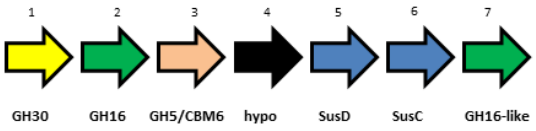


Putative β-glucan PUL 1 / *Polaribacter* 1-a / 20100518_Bin_42_1 /POL1A_42_9

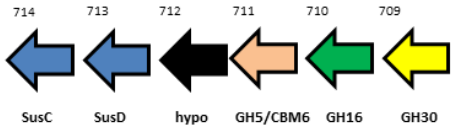


β-glucan F

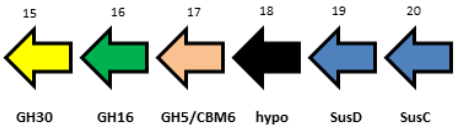
Putative β-glucan PUL 2 / *Polaribacter* 1-a / 20100518_Bin_42_1 /POL1A_42_119



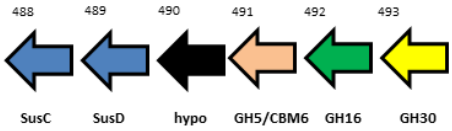
Putative β-glucan PUL 2/ *Polaribacter* 3-a / Polaribacter_Hel1_33_49 /PHEL49



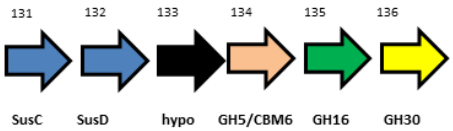
Putative β-glucan PUL 2 / *Polaribacter* 1-a / 20110512_Bin_84_1 /POL1A_84_78



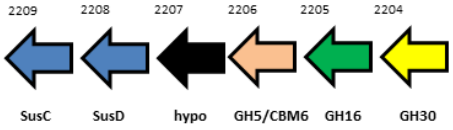
Putative β-glucan PUL 2/ *Polaribacter* 3-a / Polaribacter_Hel1_33_78 /PHEL78



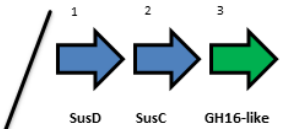
Putative β-glucan PUL 2/ *Polaribacter* 1-a / 20120426_Bin_74_1 /POL1A_74_12



Putative β-glucan PUL 2 / *Polaribacter* 3-a / Polaribacter_Hel1_33_96 /PHEL96

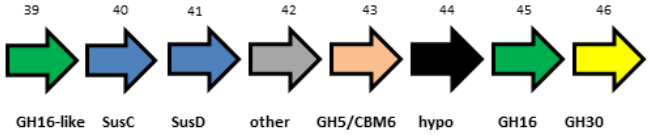


Putative β-glucan PUL 2 / *Polaribacter* 3-b / 20110504_Bin_70_1/ POL3B_70_73

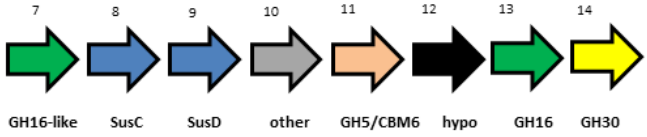


β-glucan G

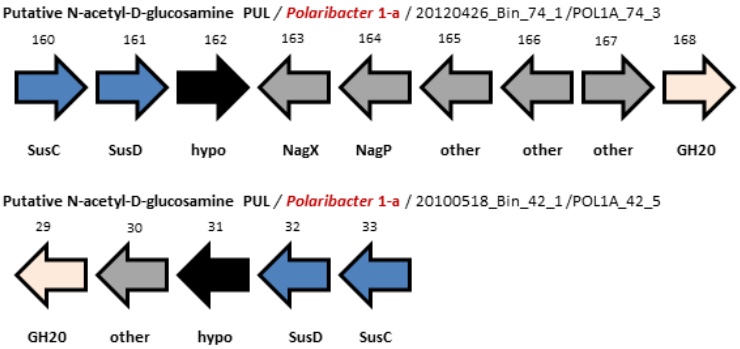
Putative β-glucan PUL 2/ *Polaribacter* 1-b / 20110523_Bin_132 /POL1B_132_32



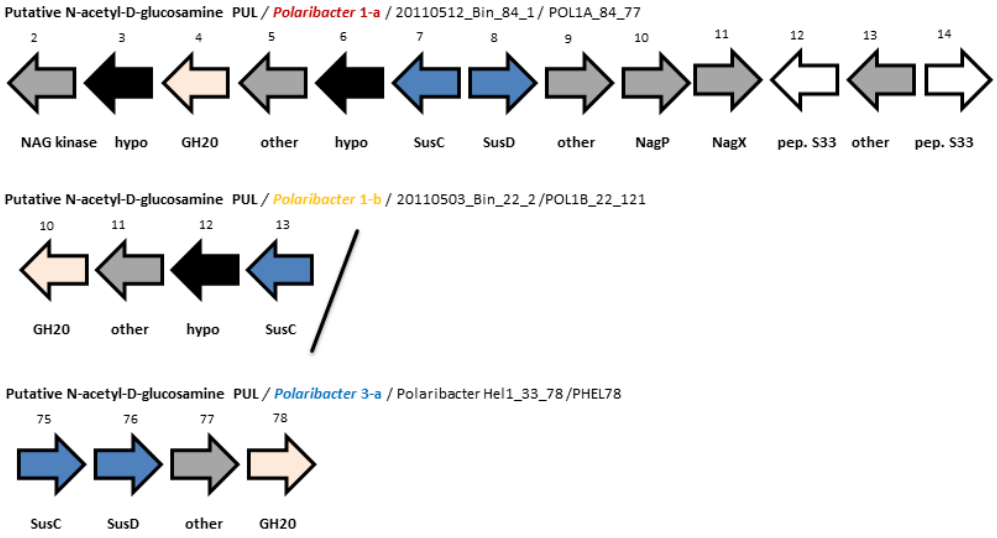
Putative β-glucan PUL 2 / *Polaribacter* 1-b / 20110503_Bin_22_2 /POL1B_22_80



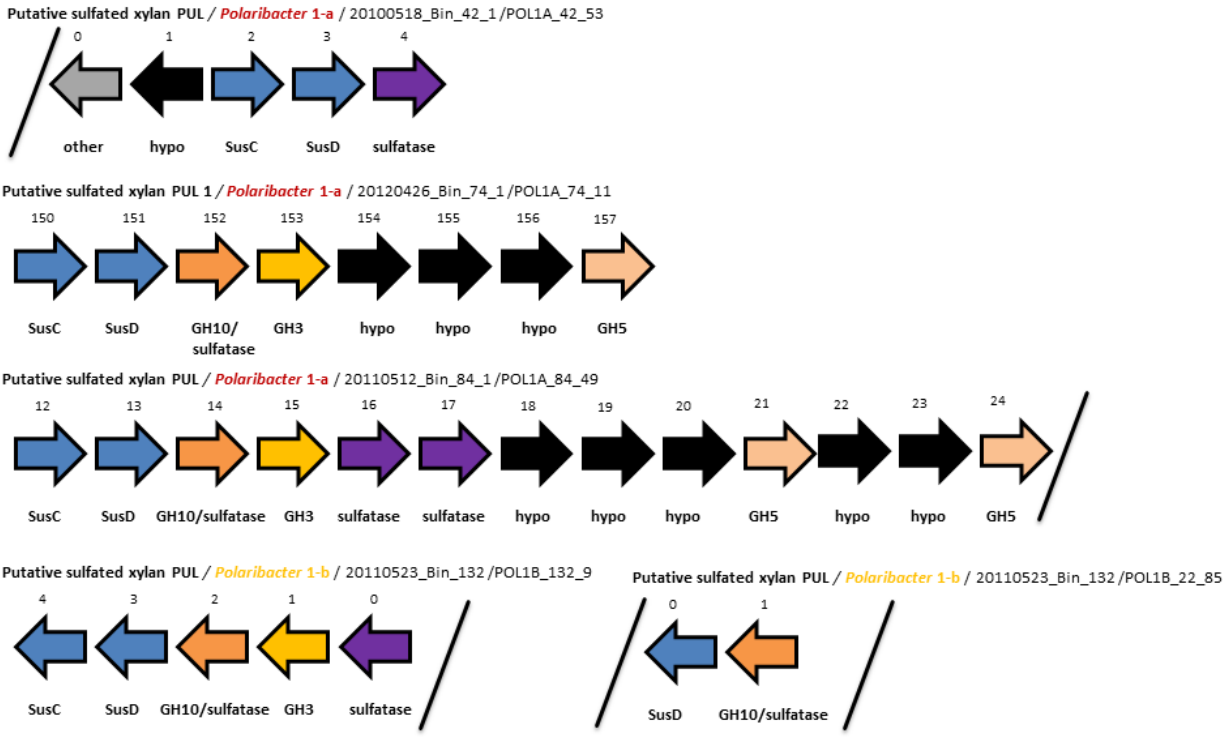
NAG A


Putative N-acetyl-D-glucosamine PUL / Polaribacter 1-a / 20120426_Bin_74_1 /POL1A_74_3
160 161 162 163 164 165 166 167 168
SusC SusD hypo NagX NagP other other other GH20
Putative N-acetyl-D-glucosamine PUL / Polaribacter 1-a / 20100518_Bin_42_1 /POL1A_42_5
29 30 31 32 33
GH20 other hypo SusD SusC

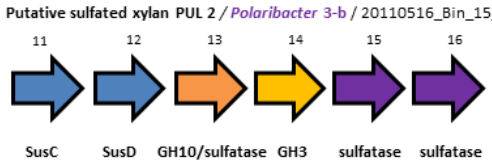
NAG B


Putative N-acetyl-D-glucosamine PUL / Polaribacter 1-a / 20110512_Bin_84_1 / POL1A_84_77
2 3 4 5 6 7 8 9 10 11 12 13 14
NAG kinase hypo GH20 other hypo SusC SusD other NagP NagX pep. S33 other pep. S33
Putative N-acetyl-D-glucosamine PUL / Polaribacter 1-b / 20110503_Bin_22_2 /POL1B_22_121
10 11 12 13
GH20 other hypo SusC
Putative N-acetyl-D-glucosamine PUL / Polaribacter 3-a / Polaribacter Hel1_33_78 /PHEL78
75 76 77 78
SusC SusD other GH20

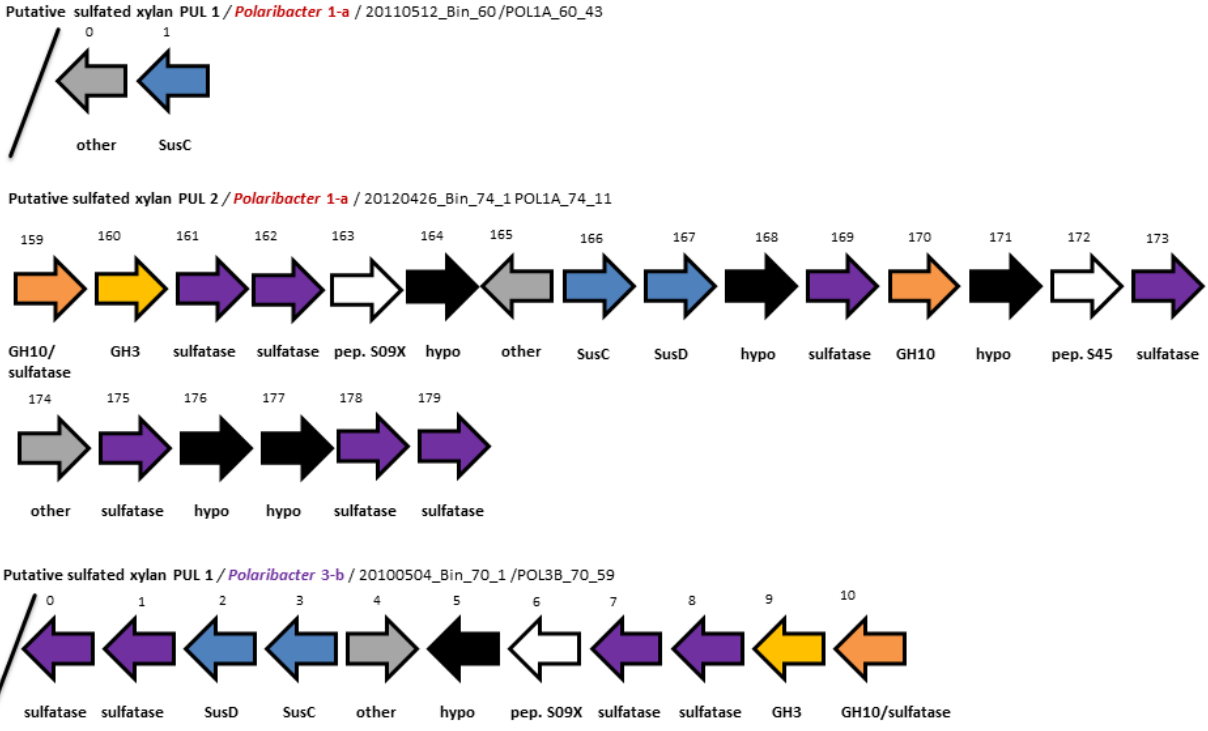
Sulfated xylan A



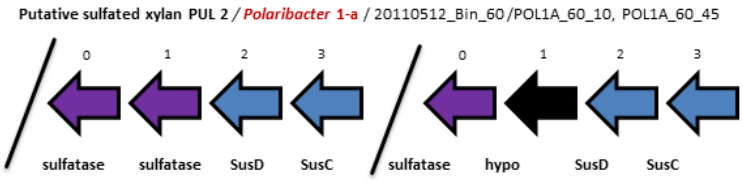
Sulfated xylan A



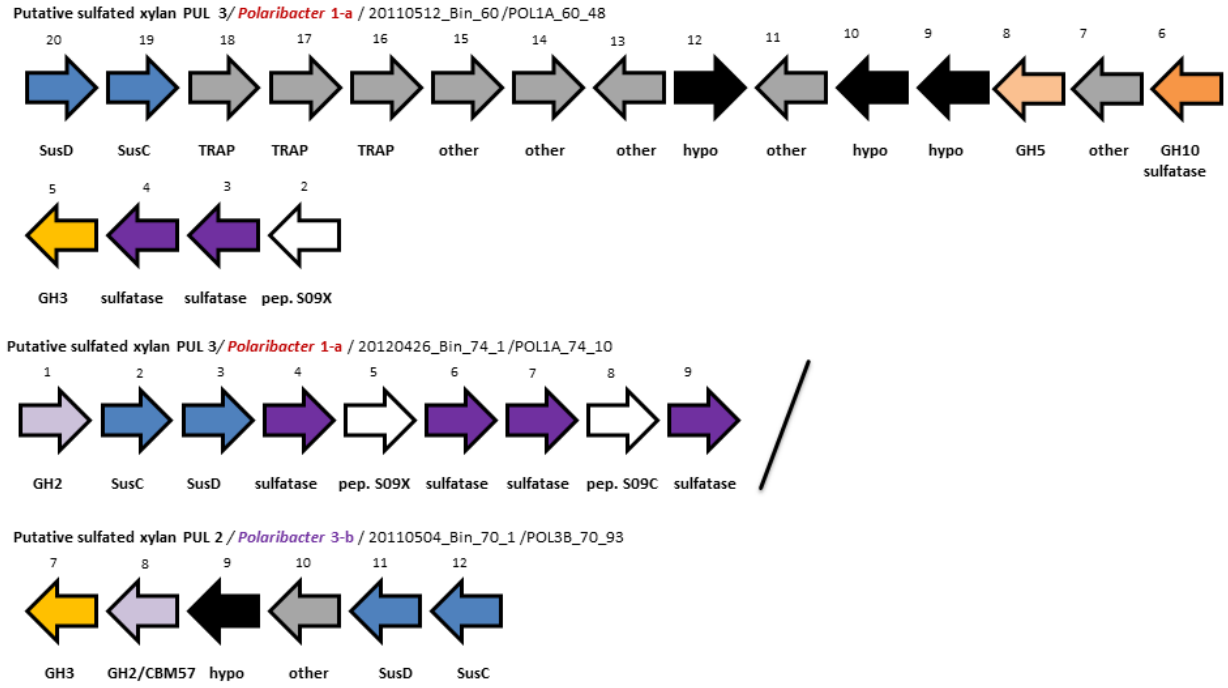
Sulfated xylan B



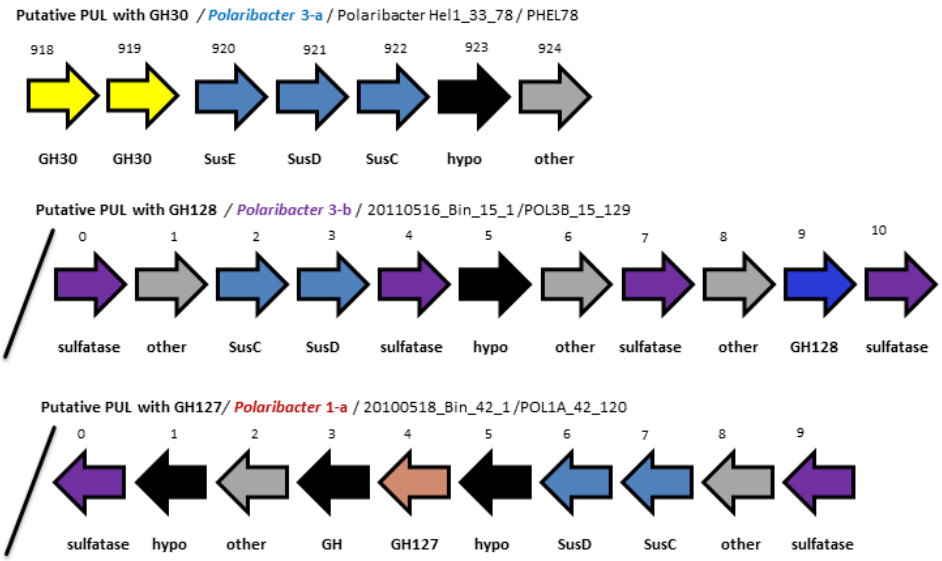
Sulfated xylan B



Sulfated xylan C

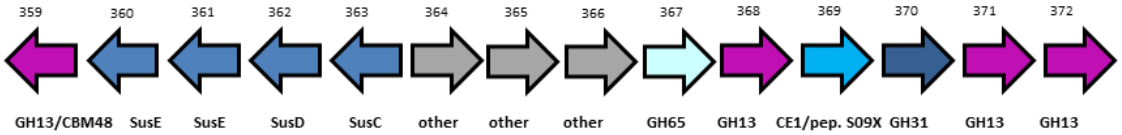


Putative PULs

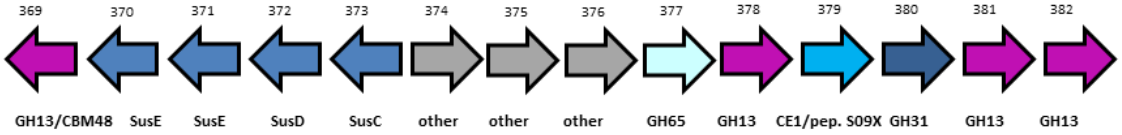


α-glucan A

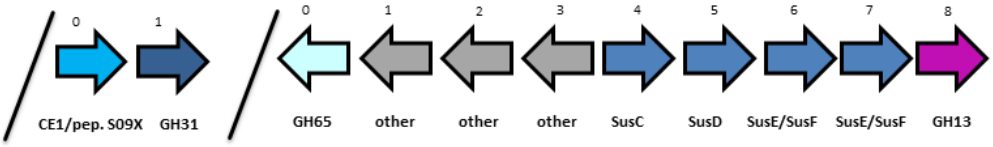
Putative α-glucan PUL / *Polaribacter* 3-a / Polaribacter Hel1_33_49 /PHEL49



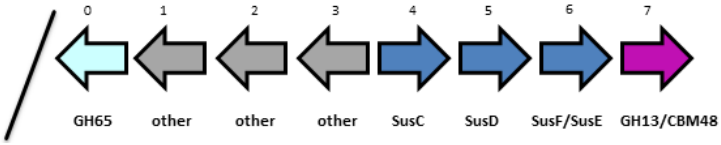
Putative α-glucan PUL / *Polaribacter* 3-a / Polaribacter Hel1_33_96 /PHEL96



Putative α-glucan PUL / *Polaribacter* 1-a / 20110512_Bin_60 /POL1A_60_72 and POL1A_60_142

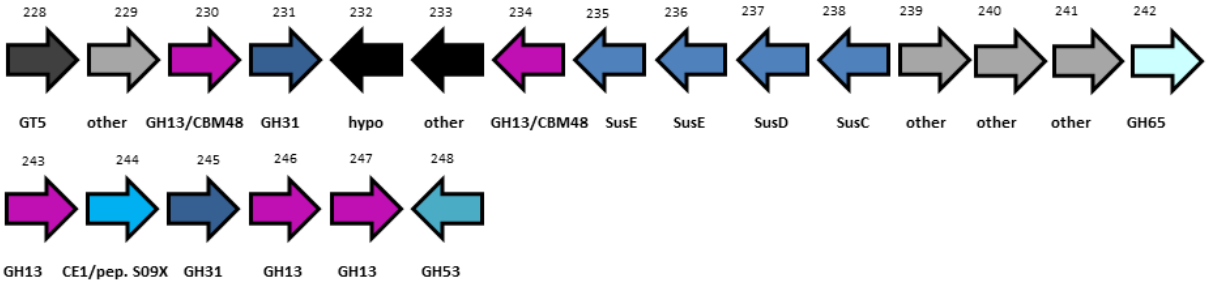


Putative α-glucan PUL / *Polaribacter* 1-a / 20110512_Bin_84_1 /POL1A_84_38

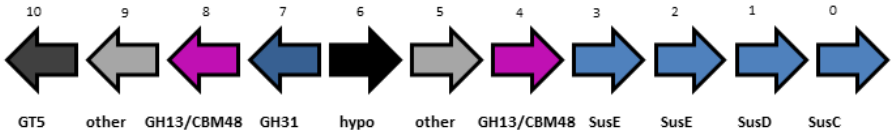


α-glucan B

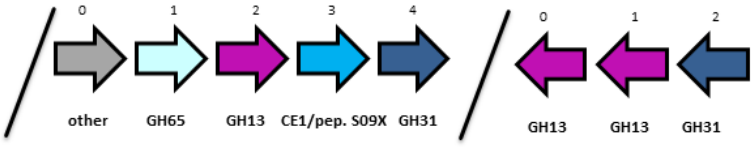
Putative α-glucan PUL / *Polaribacter* 2-a / 20100408_Bin_85_1 / POL2A_85_11



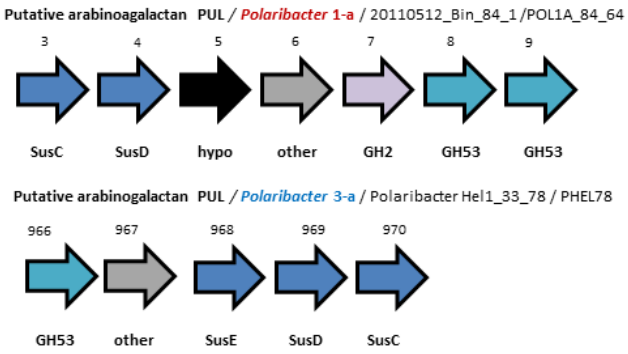
Putative α-glucan PUL / *Polaribacter* 2-a / 20100330_Bin_63 /POL2A_63_1



Putative partial α-glucan PUL / *Polaribacter* 2-b / 20100511_Bin_62 / POL2B_62_155, POL2B_62_120

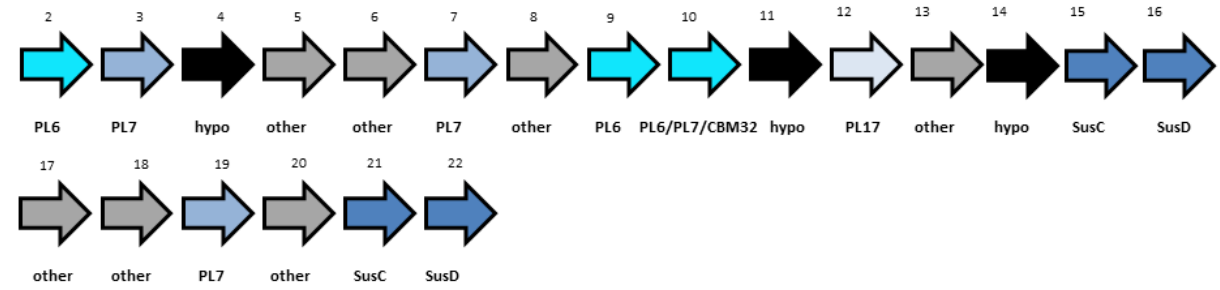


Arabinogalactan

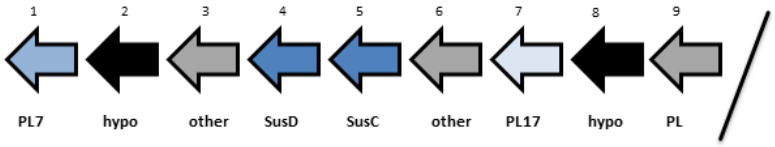


Alginate A

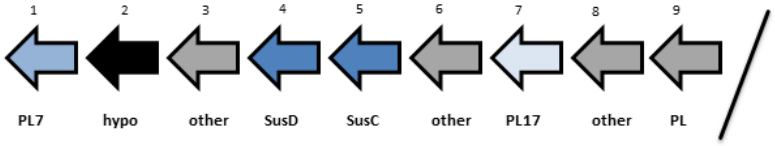
Putative alginate PUL 1 / *Polaribacter* 1-a / 20100518_Bin_42_1 /POL1A_42_113



Putative alginate PUL / *Polaribacter* 1-a / 20110512_Bin_60 /POL1A_60_91

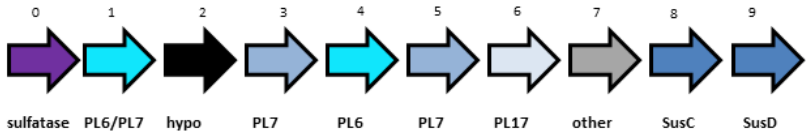


Putative alginate PUL / *Polaribacter* 1-a / 20110512_Bin_84_1 /POL1A_84_61

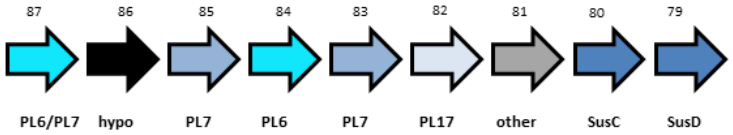


Alginate B

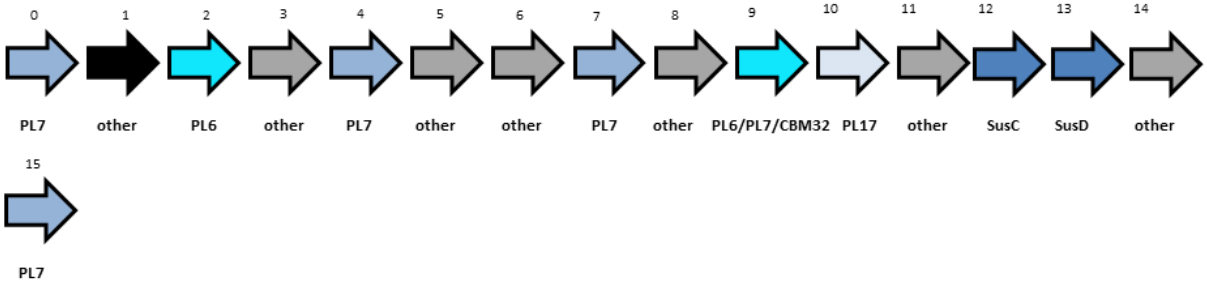
Putative alginate PUL / *Polaribacter* 2-b / 20100511_Bin_62 /POL2B_62_133



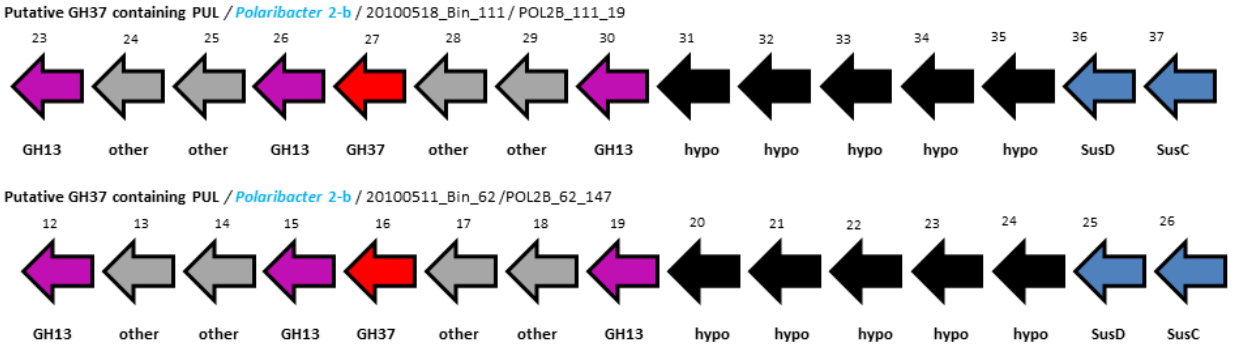
Putative alginate PUL / *Polaribacter* 2-b / 20100518_Bin_111 / POL2B_19



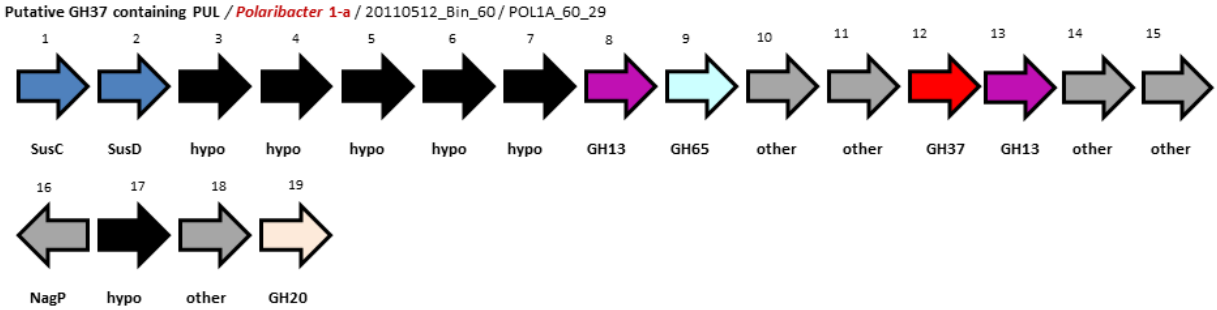
Putative alginate PUL 2 / *Polaribacter* 1-a / 20100518_Bin_42_1 /POL1A_42_90



GH37 containing A

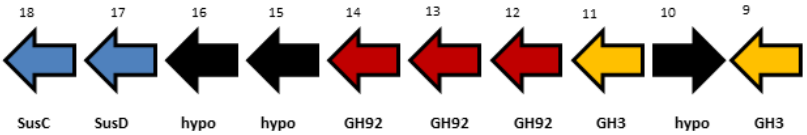


GH37 containing B

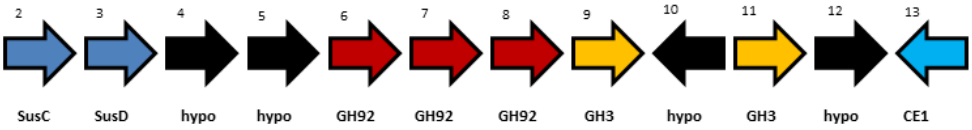


α mannose containing A

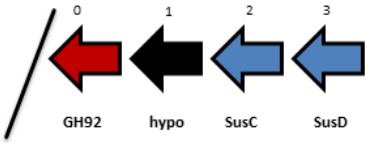
Putative α-mannose containing PUL 1 / *Polaribacter* 1-a / 201105212_Bin_60 / POL1A_60_64



Putative α-mannose containing PUL 1 / *Polaribacter* 1-a / 20120426_Bin_74_1 /POL1A_74_8

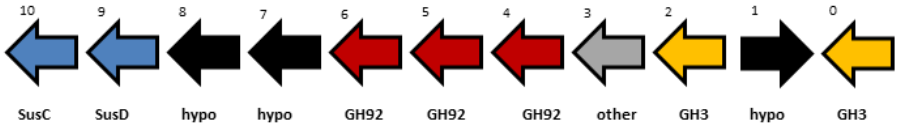


Putative α-mannose containing PUL 1 / *Polaribacter* 1-a / 20100518_Bin_42_1 /POL1A_42_45

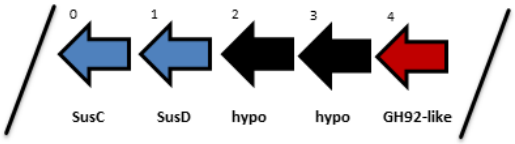


α mannose containing A

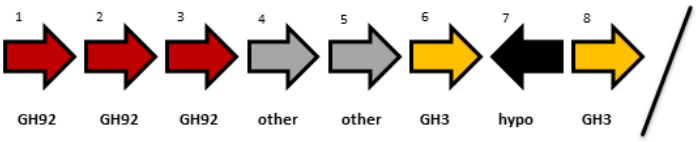
Putative α-mannose containing PUL 2 / *Polaribacter* 1-a / 201105212_Bin_60 /POL1A_60_23



Putative α-mannose containing PUL 1/ *Polaribacter* 3-b / 20110516_Bin_15_1 /POL3B_15_64

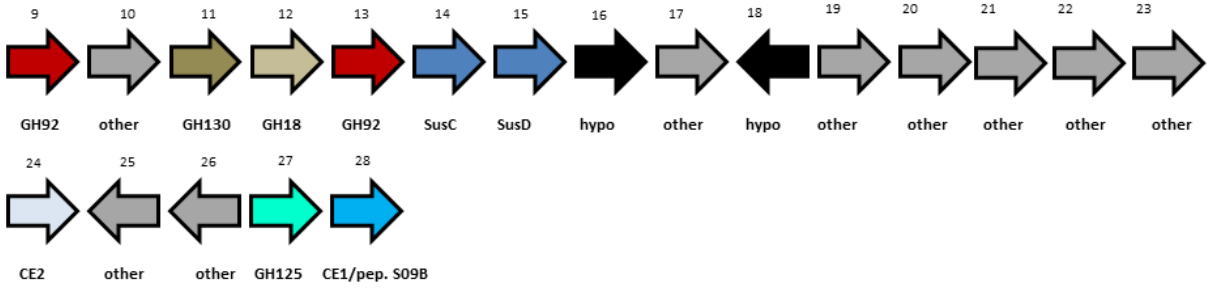


Putative partial α-mannose containing PUL 1/ *Polaribacter* 3-b / 20100504_Bin_70_1 /POL3B_70_19

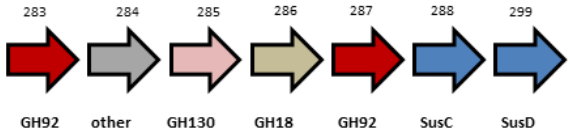


α mannose containing B

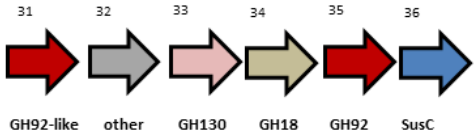
Putative α-mannose containing PUL 2/ *Polaribacter* 1-a / 20100518_Bin_42_1 /POL1A_42_123



Putative α-mannose containing PUL 2 / *Polaribacter* 1-a / 20120426_Bin_74_1 /POL1A_74_5

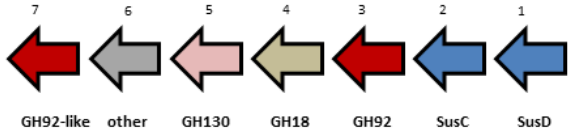


Putative α-mannose containing PUL / *Polaribacter* 1-a / 20110512_Bin_84_1 /POL1A_84_2

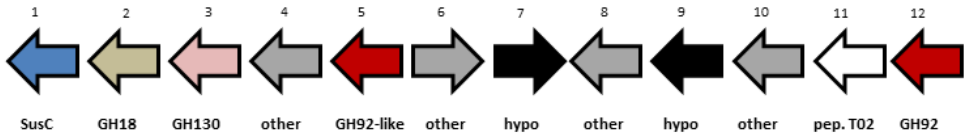


α mannose containing B

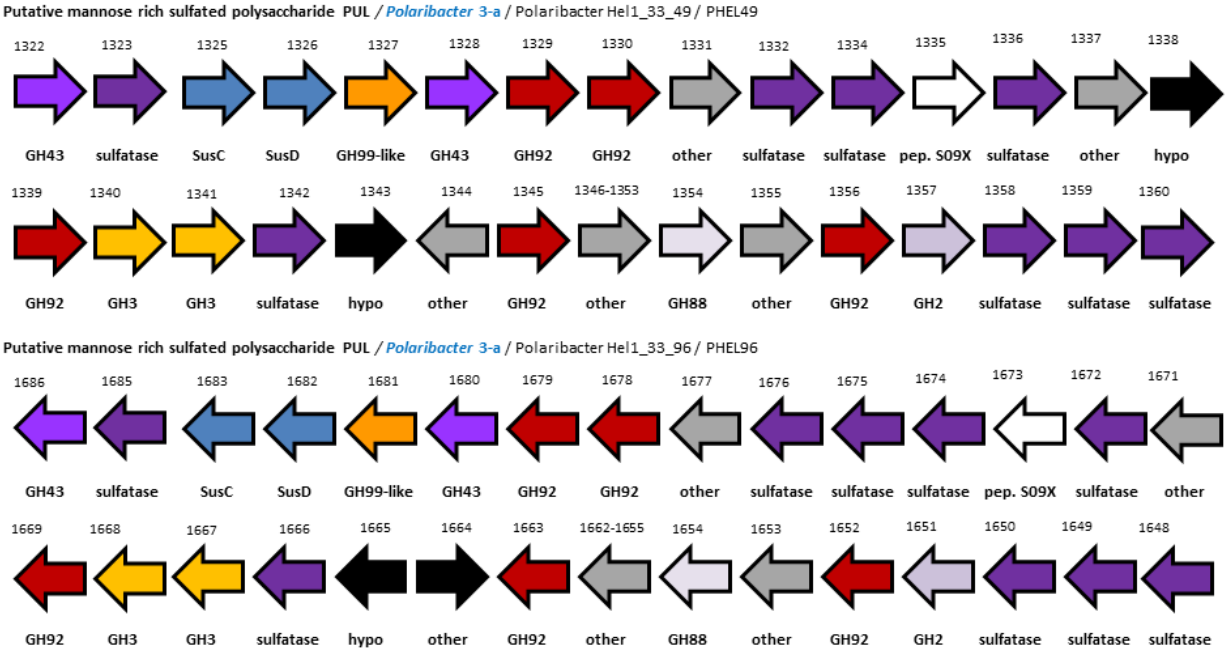
Putative α-mannose containing PUL 3 / *Polaribacter* 1-a / 20110512_Bin_60 /POL1A_60_104



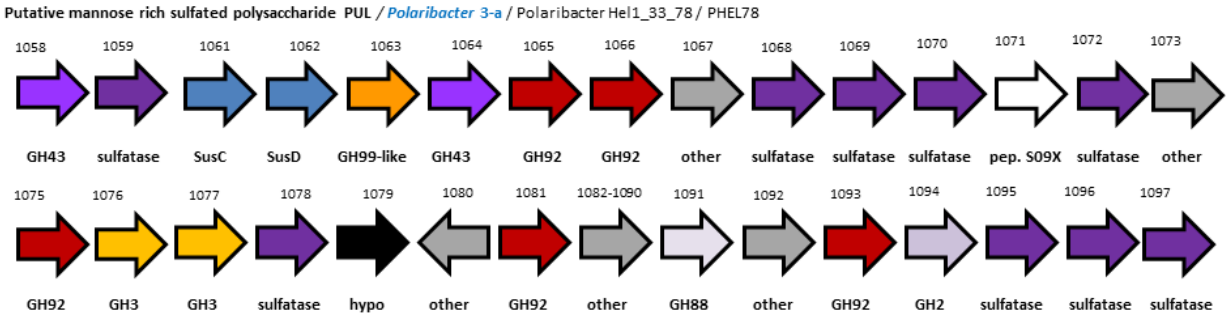
Putative α-mannose containing PUL 2/ *Polaribacter* 3-b / 20100504_Bin_70_1 /POL3B_70_30



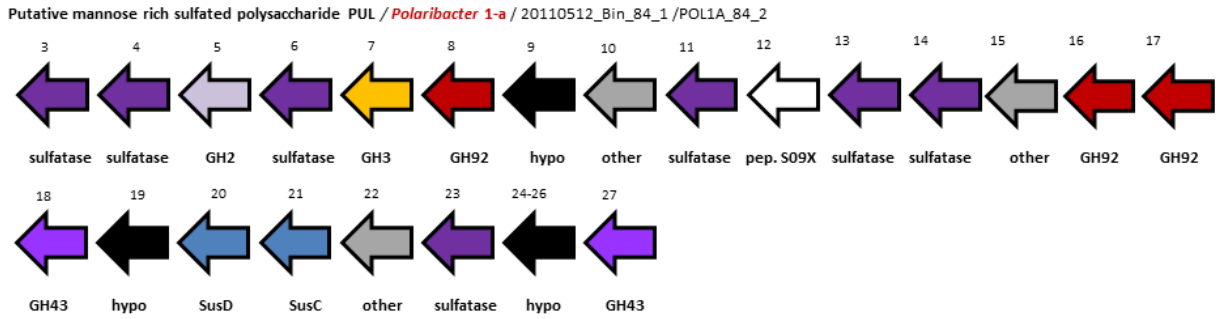
Mannose rich sulfated polysaccharide A



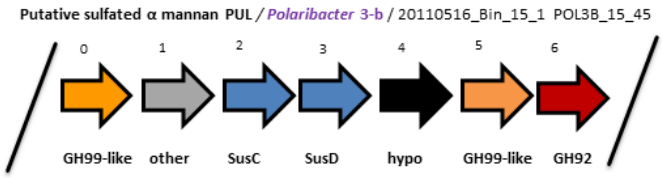
Mannose rich sulfated polysaccharide A



Mannose rich sulfated polysaccharide B

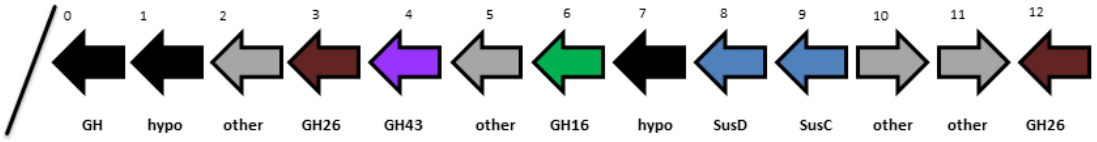


Mannose rich sulfated polysaccaride C

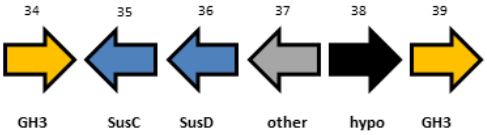


Putative PULs with partial or no SusC sequences

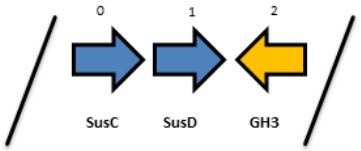
Putative PUL with GH26 and GH43/ *Polaribacter* 1-a / 20100518_Bin_42_1 /POL1A_42_20



Putative PUL 2 with GH3/ *Polaribacter* 1-b / 20110523_Bin_132 /POL1B_132_30



Putative PUL with GH3/ *Polaribacter* 1-b / 20110503_Bin_22_2 /POL1B_22_60



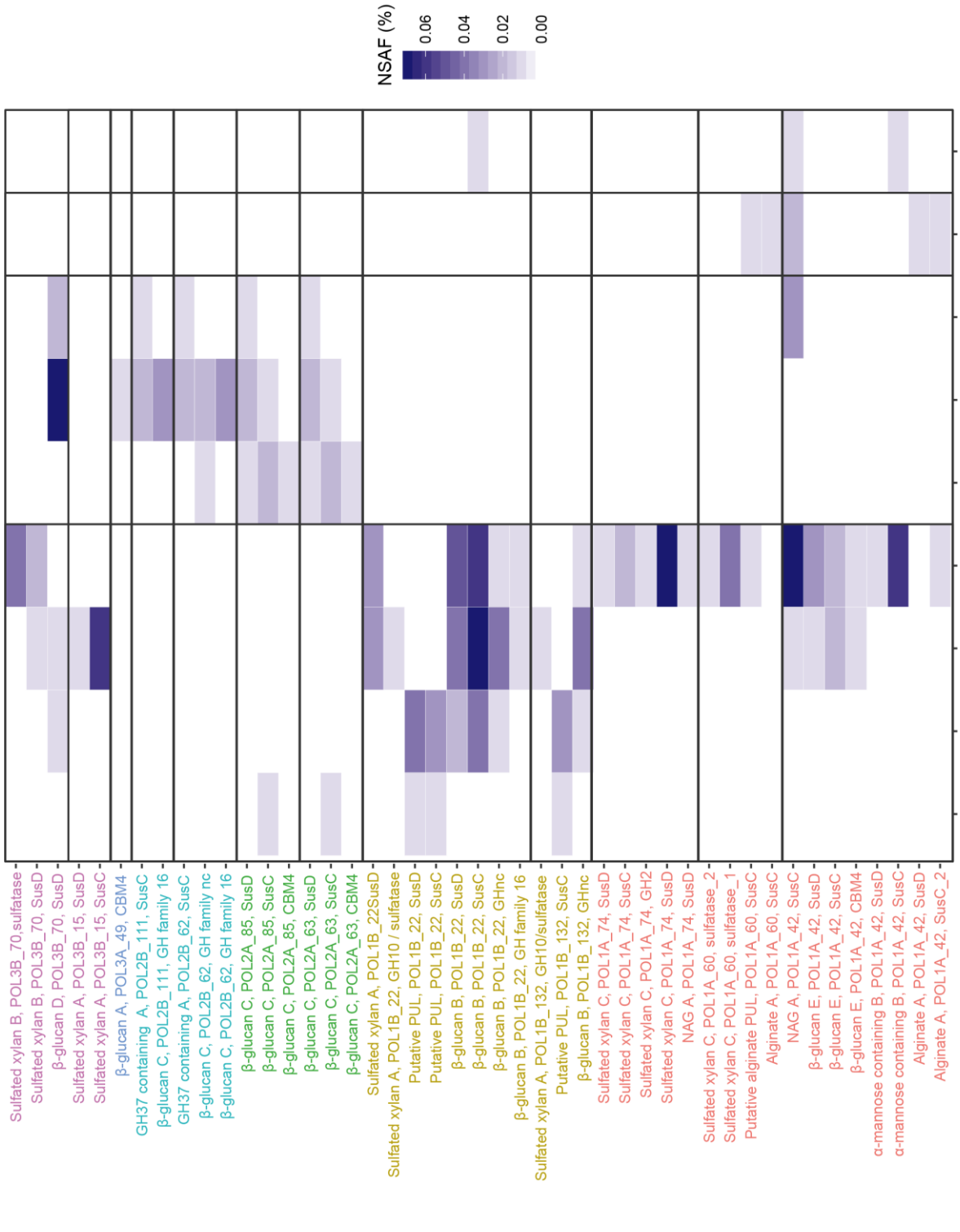


Fig. S6. Expression of PUL-related genes in *Polaribacter* clades based on metaproteome analysis. All expressed genes in *Polaribacter* MAGs/genomes are shown in Table S9. Metaproteome data were retrieved from (14) and (59). Metaproteomes without any detected PUL-related protein are not depicted in the figure. Since sampling dates in 2011 and 2012 were not coincided with the *Polaribacter* peak abundances, comprehensive expression profiles could not be detected. NSAF: normalized spectral abundance factor.

Probe	Sequence (5‘ … 3‘)	Formamide [%]
POL405	**CCCATAGGGCATTCTTCCTACA**	**20**
c21POL405	ACCCATAGGGCATTCTTCCTBCA	20
c13POL405	ACCCATAGGGCAGTCTTCCTACA	20
POL405h1	CGCRGCATGGCKGGATCAGAGTCTC	20
POL405h2	TTCYTCCTGTATAAAAGTAGTTTACAA	20
POL1270	**GTAGATTCGCGCTCTG**	**20**
c14.1POL1270	TTTGTAGATTCGCTCTCTG	20
c14.2POL1270	TTCGTAGATTCGCTCTCTG	20
c5POL1270	TTTGGAGATTCGCGCTCTG	20
POL1270h1	TTGCCAGATGGCTGCTCATTG	20
POL1270h2	TCCGAACTGTGATATGGTTT	20
POL183a	**CTCGATGCCAAGTCTCAA**	**15**
c10POL183a	CTCGATGCCGAGTCTCAATA	15
c15POL183a	CTCGATGCCAAGTCGCAATA	15
c5POL183a	AACTTGATGCCAAGTCTCAA	15
POL183a/h1	TACTATAAGGTATTAATCTTCA	15
POL183a/h2	TAAATCTTTAATTAAAA	15
POL180	**GCCAGTCTATAATACCA**	**10**
c16POL180	GCCAGTCTATAATACTATG	10
c12POL180	ATGCCAGTCTACAATACCA	10
c7POL180	ATGCCAHTCTATAATACCA	10
POL180h1	TAAGGTATTAATCTTCATTTC	10
POL180h2	TCTTTAATTATAAACTGAT	10

Table S1. Newly designed oligonucleotide probes used for CARD-FISH analysis.

Date	Sequencing platform	Total number of reads	Number of quality filtered reads	Source of metagenome
20090331	454 FLX Ti	1,666,535	1,140,320	ENA: ERP001227
20090407	454 FLX Ti	2,109,239	1,537,921	ENA: ERP001227
20090414	454 FLX Ti	4,588,441	3,236,386	ENA: ERP001227
20090616	454 FLX Ti	1,120,072	817,516	ENA: ERP001227
20100303	HiSeq2000	285,509,546	215,734,422	JGI CSP COGITO met01
20100330	HiSeq2500	87,007,870	81,716,370	JGI CSP COGITO mtgs 100330
20100408	HiSeq2000	330,603,192	232,384,184	JGI CSP COGITO met02
20100413	HiSeq2500	85,204,698	79,754,314	JGI CSP COGITO mtgs 100413
20100420	HiSeq2500	83,499,192	78,919,256	JGI CSP COGITO mtgs 100420
20100423	HiSeq2500	79,594,912	75,009,358	JGI CSP COGITO mtgs 100423
20100430	HiSeq2500	72,546,152	68,614,694	JGI CSP COGITO mtgs 100430
20100504	HiSeq2000	274,336,674	224,993,942	JGI CSP COGITO met03
20100511	HiSeq2500	75,567,206	71,051,770	JGI CSP COGITO mtgs 100511
20100518	HiSeq2000	309,971,624	199,665,578	JGI CSP COGITO met04
20110321	HiSeq2500	84,894,030	79,587,610	JGI CSP COGITO mtgs 110321
20110324	HiSeq2000	304,742,176	244,800,088	JGI CSP COGITO met05
20110328	HiSeq2500	75,242,542	69,406,062	JGI CSP COGITO mtgs 110328
20110331	HiSeq2500	74,047,178	69,292,462	JGI CSP COGITO mtgs 110331
20110404	HiSeq2500	79,429,002	73,995,840	JGI CSP COGITO mtgs 110404
20110407	HiSeq2500	88,393,908	82,789,928	JGI CSP COGITO mtgs 110407
20110414	HiSeq2500	76,260,848	71,086,194	JGI CSP COGITO mtgs 110414
20110421	HiSeq2500	81,789,256	76,032,310	JGI CSP COGITO mtgs 110421
20110426	HiSeq2500	84,364,922	78,545,118	JGI CSP COGITO mtgs 110426
20110428	HiSeq2000	295,426,382	235,544,646	JGI CSP COGITO met06
20110506	HiSeq2500	80,085,580	74,970,646	JGI CSP COGITO mtgs 110506
20110509	HiSeq2500	90,174,260	84,434,414	JGI CSP COGITO mtgs 110509
20110512	HiSeq2500	81,683,140	75,835,970	JGI CSP COGITO mtgs 110512
20110516	HiSeq2500	88,557,186	83,034,602	JGI CSP COGITO mtgs 110516
20110519	HiSeq2500	100,457,324	94,028,844	JGI CSP COGITO mtgs 110519
20110523	HiSeq2500	227,207,716	207,300,790	JGI CSP COGITO mtgs 110523
20110526	HiSeq2000	324,219,736	250,091,456	JGI CSP COGITO met07
20110530	HiSeq2500	83,974,950	77,770,314	JGI CSP COGITO mtgs 110530
20120308	HiSeq2000	311,415,662	207,556,978	JGI CSP COGITO met08
20120405	HiSeq2500	173,626,414	161,992,212	JGI CSP COGITO mtgs 120405
20120412	HiSeq2500	175,790,324	164,053,740	JGI CSP COGITO mtgs 120412
20120416	HiSeq2000	258,751,220	210,093,962	JGI CSP COGITO met09
20120426	HiSeq2500	110,390,598	96,979,192	JGI CSP COGITO mtgs 120426
20120503	HiSeq2500	105,224,208	94,343,038	JGI CSP COGITO mtgs 120503
20120510	HiSeq2000	272,267,348	222,320,174	JGI CSP COGITO met10
20120524	HiSeq2500	113,234,474	101,681,308	JGI CSP COGITO mtgs 120524
20120531	HiSeq2500	102,492,998	91,212,794	JGI CSP COGITO mtgs 120531
20120607	HiSeq2500	189,120,228	169,486,964	JGI CSP COGITO mtgs 120607

Table S3. List of the metagenomes used in this study.

MAG/genome	Cluster	Completeness	Contamination	Size (Mbp)
20110509_Bin_30_1	*Polaribacter* 1-a	88.12	1.09	2.33
20110530_Bin_44_1	*Polaribacter* 1-a	89.43	1.89	2.09
20110523_Bin_54_1	*Polaribacter* 1-a	69.21	0.62	1.71
20110519_Bin_30_1	*Polaribacter* 1-a	71.73	0.98	1.67
20120503_Bin_35_1	*Polaribacter* 1-a	79.78	2.49	1.94
20110512_Bin_84_1	***Polaribacter* 1-a**	69.3	0	2.28
20120426_Bin_74_1	***Polaribacter* 1-a**	98.15	0.94	2.69
20100518_Bin_42_1	***Polaribacter* 1-a**	93.24	2.01	2.59
20110512_Bin_60	***Polaribacter* 1-a**	94.51	3.24	2.82
20110523_Bin_132	***Polaribacter* 1-b**	96.85	1.07	2.15
20120503_Bin_22_2	***Polaribacter* 1-b**	92.4	2.53	2.12
20110519_Bin_97	*Polaribacter* 1-b	96.76	0	2.17
20110516_Bin_82	*Polaribacter* 1-b	96.24	0	2.16
20120510_Bin_20_2	*Polaribacter* 1-b	78.56	2.06	1.86
20100330_Bin_63	***Polaribacter* 2-a**	98.32	0	2.18
20100408_Bin_85_1	***Polaribacter* 2-a**	98.32	0	2.18
20110428_Bin_126_1	*Polaribacter* 2-a	98.32	0.5	2.19
20110512_Bin_48	*Polaribacter* 2-a	83.31	1.39	1.76
20110509_Bin_92	*Polaribacter* 2-a	98.01	0	2.11
20120524_Bin_96_2	*Polaribacter* 2-a	95.09	0.17	1.99
20120607_Bin_88_1	*Polaribacter* 2-a	90.77	1.68	1.88
20100420_Bin_25	Polaribacter 2-a	91.18	1.96	2.00
20100413_Bin_64	Polaribacter 2-a	95.96	1.43	2.24
20110506_Bin_86_1	Polaribacter 2-a	97.25	0.11	2.04
20100423_Bin_30	*Polaribacter* 2-a	81.16	3.24	1.83
20100511_Bin_62	*Polaribacter* 2-b	93.6	1.76	2.03
20100518_Bin_111	*Polaribacter* 2-b	94.41	0.34	2.04
20100430_Bin_64	*Polaribacter* 2-b	94.18	0.91	2.05
20100504_Bin_100	*Polaribacter* 2-b	94.41	0.34	2.08
***Polaribacter* Hel1_33_49**	*Polaribacter* 3-a	NA	NA	3.01
***Polaribacter* Hel1_33_78**	*Polaribacter* 3-a	100	NA	3.23
***Polaribacter* Hel1_33_96**	*Polaribacter* 3-a	100	NA	3.10
20100408_Bin_74	*Polaribacter 3-a*	97.53	1.27	2.86
20100420_Bin_62_1	*Polaribacter 3-a*	98.55	0.22	2.89
20100430_Bin_65_1	*Polaribacter* 3-b	91.86	0.95	2.20
20100504_Bin_70_1	***Polaribacter* 3-b**	91.9	1.07	2.15
20100511_Bin_60	*Polaribacter* 3-b	87.58	1.57	2.28
20100518_Bin_59	*Polaribacter* 3-b	87.97	3.22	2.26
20110516_Bin_15_1	***Polaribacter* 3-b**	91.67	3.55	2.30
20110519_Bin_54	*Polaribacter* 3-b	87.96	1.07	2.31

Table S4. Size, contamination and completeness values of spring-bloom associated *Polaribacter* MAGs and genomes.

	MED node 3321	MED node 9198	MED node 6197	MED node 10976	MED node 9011	MED node 9018
Polaribacter 1-a	**0.99**	0.56	0.05	0.16	-0.05	-0.02
Polaribacter 1-b	0.52	**0.98**	-0.05	0.28	0.04	-0.01
Polaribacter 2-a	-0.01	-0.12	**0.74**	-0.01	0.05	-0.1
Polaribacter 2-b	-0.1	-0.07	0	**0.73**	0.15	0.11
Polaribacter 3-a	-0.09	-0.11	0.05	0.03	**0.96**	0.04
Polaribacter 3-b	-0.05	-0.08	0.03	0.33	0.1	**0.88**

	POL405	POL1270	POL183a	POL180
Polaribacter 1-a	**0.88**	-0.33	-0.33	-0.31
Polaribacter 1-b	0.36	-0.3	-0.4	-0.35
Polaribacter 2-a	-0.2	**0.94**	0.01	0.16
Polaribacter 2-b	-0.3	0.73	0.02	0.04
Polaribacter 3-a	-0.2	0.18	**0.61**	0.29
Polaribacter 3-b	0.01	-0.3	0.21	**0.64**

	POL405	POL1270	POL183a	POL180
MED node 3321	**0.96**	-0.32	-0.26	-0.24
MED node 9198	0.93	-0.36	-0.35	-0.31
MED node 6197	-0.09	**0.97**	-0.01	0.08
MED node 10976	-0.33	0.69	-0.28	-0.21
MED node 9011	0.04	-0.12	**0.36**	0.18
MED node 9018	-0.2	-0.35	-0.24	**0.49**

Table S6. Spearman correlation test between *Polaribacter* abundance data obtained in metagenome read recruitment, MED analysis and CARD-FISH counts.

MAG/Genome	Abbreviation	Length (Mbp)	Number of contigs	Number of ORFs	Number of tRNAs	Number of rRNAs
20110512_Bin_84_1	POL1A_84	2.28	83	2,000	29	0
20120426_Bin_74_1	POL1A_74	2.69	14	2,343	33	0
20100508_Bin_42_1	POL1A_42	2.59	131	2,311	30	0
20110512_Bin_60	POL1A_60	2.82	173	2,471	19	0
20110523_Bin_132	POL1B_132	2.14	38	1,922	30	0
20120503_Bin_22_2	POL1B_22	2.12	172	1,969	26	0
20100330_Bin_63	POL2A_63	2.18	10	2,024	34	0
20100408_Bin_85_1	POL2A_85	2.18	14	2,021	33	0
20100511_Bin_62	POL2B_62	2.02	174	1,948	28	0
20100518_Bin_111	POL2B_111	2.04	86	1,900	32	0
Hel1_33_49	PHEL49	3.01	2	2,603	39	9
Hel1_33_78	PHEL78	3.23	1	2,752	39	9
Hel1_33_96	PHEL96	3.01	1	2,637	44	12
20110516_Bin_15_1	POL3B_15	2.30	208	2,141	28	0
20100504_Bin_70_1	POL3B_70	2.15	149	1,984	31	0

Table S7. Statistics about the *Polaribacter* MAGs/genomes.

Polaribacter spring 2009-1						
Polaribacter spring 2010-1	98.2					
Polaribacter spring 2010-2	97.8	98.8				
Polaribacter spring 2009-2	97.6	98.2	97.6			
Polaribacter isolates	97.0	97.2	97.1	98.4		
Polaribacter spring 2009-3	96.5	97.2	96.9	97.8	98.0	
Polaribacter spring 2009-4	96.9	96.8	96.6	97.0	98.0	98.2

Table S10. Mean sequence similarity between North Sea Polaribacter clones and isolates dispersed into seven clusters

Chapter 4

Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms

Published in *The ISME Journal*

doi.org/10.1038/s41396-018-0243-5

Contributions to the manuscript:

Experimental concept and design: 5%

Experimental work and/or acquisition of data: 10%

Data analysis and interpretation: 15%

Preparation of figures and tables: 10%

Drafting of the manuscript: 10%

In this study, I involved in phylogenetic and metagenomic analyses and annotation of *Formosa* A and B genomes. Since this publication represents the collaborative study of twenty authors from different disciplines, my efforts are reflected in low percental contributions.

The ISME Journal
https://doi.org/10.1038/s41396-018-0243-5

ISME

ARTICLE

Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms

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[7,8] · Rudolf I. Amann[2] · Thomas Schweder[1,3]

Received: 20 February 2018 / Revised: 22 June 2018 / Accepted: 30 June 2018

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 a 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.

Introduction

Phytoplankton blooms produce large quantities of beta-glucans, such as laminarin, a soluble β-1,3-glucan with β-1,6 side chains. The breakdown of these polysaccharides by heterotrophic microbes is a central part of the marine carbon cycle. Diatoms alone are estimated to produce ~ 5–15 Gt of laminarin per year as their storage compound, making it a major food resource for heterotrophic marine organisms [1]. Bacterial laminarinase activities are abundant in ocean surface waters, but also within deeper parts of the water column and in sediments [2, 3]. This suggests laminarin-degrading bacteria and their laminarinases are common across the oceans. How bacteria compete for this abundant labile energy substrate is therefore of relevance for a better understanding of the marine carbon cycle. Although partially studied with model organisms in the laboratory [4–7], the enzymes used for laminarin degradation by microbes in the wild remain largely unknown or uncharacterized.

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41396-018-0243-5) contains supplementary material, which is available to authorized users.

✉ Thomas Schweder
schweder@uni-greifswald.de

1 Pharmaceutical Biotechnology, University Greifswald, 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, Greifswald, Germany
6 DSMZ, Braunschweig, Germany
7 National Center of Scientific Research/Pierre and Marie Curie University, Paris, France
8 UMR 7139 Marine Plants and Biomolecules, Station Biologique de Roscoff, Roscoff, Bretagne, France

Published online: 30 July 2018

SPRINGER NATURE

For complete degradation of one polysaccharide, microbes must have an adapted glycolytic pathway that contains multiple enzymes, which individually address each of the different glycosidic linkages and structural compositions present in the macromolecule. The genes of glycan-degrading pathways cluster in operons named polysaccharide utilization loci (PULs). Recent works suggest that each polysaccharide requires a corresponding PUL (for review, see Grondin et al. [8]). Horizontal gene transfer, vertical inheritance, and gene loss distribute PULs asymmetrically among genomes of microbes, creating the molecular basis for polysaccharide resource partitioning [9–11]. This might explain the occurrence of diverse bacterial communities in the human gut [12–15] or in the oceans, whose members rely on different degradation products of the same polysaccharide to co-exist [16–19]. However, it remains unclear whether the degradation of complex carbohydrates is a community effort or mainly driven by highly specialized individual strains. Furthermore, how microbes effectively compete for the same polysaccharide resource, such as the abundant laminarin, is currently unknown.

In previous studies, we reported the high abundance (up to 24% of all bacteria) of the flavobacterial genus *Formosa* during diatom-dominated spring blooms off the North Sea island Helgoland [17, 18]. Furthermore, high laminarin concentrations were measured at the same sampling site [20]. Together, these findings suggested that *Formosa* spp. are prominent candidates for the recycling of laminarin during spring microalgae blooms.

In this study, we explored molecular strategies, which provide competitive advantages to the genus *Formosa* during microalgal blooms in general and for laminarin utilization in particular. We examined two strains, *Formosa* Hel3_A1_48 (referred to as strain A) and *Formosa* Hel1_33_131 (strain B), both of which were isolated from the same sampling location [21], and which are representative of two distinct taxonomical clades found during phytoplankton blooms [22]. The combination of high-resolution metaproteomics and metagenomics of spring bloom water samples with the detailed proteomic and biochemical characterization of the respective PUL in a cultured model strain (*Formosa* B) allowed us to show that a specialized enzyme repertoire represents one of the adaptive mechanisms that provide a competitive advantage in substrate exploitation. Using laminarin as a model substrate, we demonstrate how a microalgal glycan resource can promote the enrichment of individual dominating taxa from an initially diverse microbial community with similar metabolic functions. Our data indicate that *Formosa* B tightly couples glycan utilization with the uptake of nitrogen compounds. This suggests that a balanced carbon and nitrogen diet is required for competitive laminarin utilization during phytoplankton blooms.

Materials and methods

Growth experiments and physiological characterization

The investigated strains *Formosa* sp. Hel1_33_131 (*Formosa* strain B) and *Formosa* sp. Hel3_A1_48 (*Formosa* strain A) were isolated by dilution cultivation during a spring and a summer phytoplankton bloom, respectively, from surface water near the North Sea island Helgoland in the German Bight [21]. Growth experiments were performed in a modified HaHa medium [21] (with 0.1 g L^{-1} peptone, 0.1 g L^{-1} casamino acids, 0.1 g L^{-1} yeast extract, 200 µM NH_4Cl, and 16 µM KH_2PO_4) with defined carbon sources as substrates at 12 °C during gentle shaking at 55 rpm. For the proteome analyses, described below, D-glucose and laminarin (L9634, Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) were used as carbon sources (concentrations: 2 g L^{-1}). In addition, the utilization of chitin (SAFSC9213, VWR) was tested in this medium and these cultures were used as a control condition for the in vitro proteome analyses with glucose and laminarin. All growth experiments were carried out in triplicates. Cells were harvested by centrifugation (15 min; 9500 × *g*; 4 °C), and the resulting pellets and supernatants were stored at − 80 °C until use.

Genome sequencing, assembly, and annotation

For genome sequencing of the strains *Formosa* A (Hel3_A1_48) and B (Hel1_33_131) DNA was extracted according to the protocol of Zhou et al. [23]. Sequencing was performed at LGC Genomics (Berlin, Germany) using the 454 GS FLX Ti platform (454 Life Sciences, Branfort, CT, USA) using standard shotgun libraries. Draft genomes were assembled with Newbler v2.6 for Hel3_A1_48 from 640,093 reads (406,983,286 bp) and for Hel1_33_131 from 636,323 reads (410,253,204 bp), yielding 2,025,184 bp (77 contigs) and 2,727,763 bp (61 contigs), respectively. The remaining gaps were closed by PCR and Sanger sequencing, yielding circular assemblies of 2,016,454 bp for Hel3_A1_48 (*Formosa* A) and 2,735,158 bp for Hel1_33_131 (*Formosa* B). Gene prediction and annotation (including the phylogeny-guided carbohydrate-active enzyme (CAZyme) annotations provided in Supplementary Table S2) were performed as described previously [24]. Further bioinformatic analyses are described in Supplementary Information. Annotated genome sequences were submitted to NCBI's GenBank with the accession numbers

CP017259.1 for *Formosa* sp. Hel3_A1_48 (*Formosa* strain A) and CP017260.1 for *Formosa* sp. Hel1_33_131 (*Formosa* strain B).

Proteome analyses

The soluble intracellular proteome, the enriched membrane-associated proteome, and the soluble extracellular proteome was characterized from exponentially growing cells of *Formosa* strain B. Details of the protein extraction and subproteome enrichment can be found in Supplementary Information.

Peptides were subjected to a reversed phase C18 column chromatography on a nano ACQUITY-UPLC (Waters Corporation, Milford, MA, USA) and separated as described by Otto et al. [25]. Mass spectrometry (MS) and MS/MS data were recorded using an online-coupled LTQ-Orbitrap Classic mass spectrometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). We searched MS spectra against a target-decoy protein sequence database including sequences of *Formosa* B (Hel1_33_131) and of common laboratory contaminants.

Protein searches were performed using MaxQuant with the integrated Andromeda engine [26] with a peptide level FDR (false discovery rate) set to 0.01 (1%). Only proteins that could be detected in at least two out of three replicates were counted as identified. The automatically calculated iBAQ values (intensity-based absolute quantification; i.e., peak area divided by the sum of all theoretical peptides) were used to manually calculate riBAQ values (relative iBAQ; giving the relative protein abundance in % of all proteins in the same sample, [27]) for semiquantitative comparisons between samples from different nutrient conditions. Tests for differential expression were performed using Perseus [28] v. 1.6.1.1 with Welchs *t* test (permutation-based FDR 0.05).

The mass spectrometry proteomics data are available through the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [29] with the dataset identifier PXD007934.

Biochemical enzyme characterizations

The cloning of the FbGH17A gene (locus tag FORMB_24720) and the FaGH17B gene (locus tag FORMB_24740) is described in Supplementary Information. Cloning of the FbGH30 gene is described by Becker et al. [20]. Detailed information on the overexpression, enzyme refolding, and purification of these proteins can be found in Supplementary Information. For enzyme characterizations laminarin from *Laminaria digitata* (0.1% [w/v]; Sigma) was hydrolyzed over the course of 60 min at 37 °C with 100 nM purified enzyme (~5 μg mL^{-1} of FbGH30, FbGH17A, or FbGH17B) in 50 mM MOPS buffer at pH 7. The preparation and purification of debranched laminarin as well as the determination of kinetic parameters of the three enzymes FbGH30, FbGH17A, and FbGH17B acting on native and debranched laminarin is explained in Supplementary Information. High performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) was applied for qualitative product analysis of the enzyme reactions (see Supplementary Information).

Protein crystallization and structure solution

Crystals of FbGH17A were obtained by hanging drop vapor diffusion of the protein with 12.3 mg mL^{-1} mixed 1:1 with a "well solution" (0.03 M $MgCl_2$, 0.1 M MOPS (pH 7), 9% PEG8K) supplemented with 15% ethylene glycol. The crystals were cryoprotected prior to freezing in the "well solution" supplemented with ethylene glycol to a final concentration of 30%. Crystals were frozen by flash freezing in liquid nitrogen in nylon loops. X-ray diffraction data were collected at the DESY P11 beamline. The structure was solved by molecular replacement using PHASER in the phenix suite [30, 31] using the pdb 4wtp [32]. The model was built using BUCCANEER [33] and Coot, refined in REFMAC5 [34], and validated and deposited with pdb code 6FCG.

Results

Genome properties and phylogeny

We sequenced and annotated the genomes of the *Formosa* strains A and B. Both have a single chromosome with a GC content of 36.4 and 36.6%, respectively. With 2,016,454 bp (strain A) and 2,735,158 bp (strain B) they possess small genomes compared with other marine polysaccharide-degrading *Flavobacteriia* [4, 24, 35, 36]. Strain A has 1913 predicted genes including 1866 coding sequences (CDS), 40 tRNAs genes, and 2 rRNA operons (identical 5 S, 16 S, 23 S rRNA genes), whereas the strain B genome encodes 2675 predicted genes with 2628 CDS, 39 tRNAs genes, and 2 rRNA operons (identical 5 S, 16 S, 23 S rRNA genes).

Phylogenetic analyses based on 16 S rRNA gene sequences indicate that the *Formosa* strains A and B are representatives of two previously uncultured clades of the genus *Formosa*. These occur not only in the North Sea, but also in surface waters from coastal and open ocean sites throughout the world (Supplementary Figure S1). Of the 33 full-length *Formosa* 16 S rRNA sequences obtained from 2009 spring bloom bacterioplankton [17], 16 were >99%

identical to *Formosa* sp. Hel1_33_131 (*Formosa* B) (Supplementary Figure S1).

Formosa genomes encode PULs for laminarin degradation

Genome annotation suggested that the *Formosa* strains A and B are specialized polysaccharide degraders, which concentrate their genetic potential on a small set of sugars. *Formosa* A contains seven PULs (Supplementary Figure S2) and 28 glycoside hydrolases (Supplementary Table S1), whereas *Formosa* B contains six PULs (Supplementary Figure S3) and 21 glycoside hydrolases (Supplementary Table S1). This is a very small repertoire, even compared with other marine *Bacteroidetes* isolated from algal blooms [7]. These small CAZyme repertoires contrast particularly with those of generalist polysaccharide degraders isolated from macroalgae, such as *Formosa agariphila* [24], which have broad polysaccharide-degrading capacity. *F. agariphila* has, for example, a genome size of 4.48 Mbp, and 84 glycoside hydrolases in 13 PULs (Supplementary Table S1).

To functionally characterize laminarin-specific PULs of the *Formosa* strains A and B, we searched the genomes for enzymes belonging to known laminarinase-containing families (Supplementary Table S1). We found putative laminarinases of the families GH16 and GH17 but also enzymes of the GH3 and GH30 families as well as a member of the newly described GH149 family [37], located in close proximity to TonB-dependent receptors (TBDR) and SusD-like proteins, which are indicators of PULs [38, 39]. Our results suggest that there are three putative laminarin-specific genomic PULs in both *Formosa* strains (Supplementary Figure S2–S5).

The laminarin PULs 1 and 2 of *Formosa* A and B revealed a high synteny with PULs from other bacteroidetal strains (Supplementary Figure S4) from North Sea surface water [21, 36, 40]. This points to a potential for competition between those groups, but also suggests that this part of the laminarin utilization machinery is highly conserved. However, the *Formosa* B PULs 1 and 2 are enlarged with laminarinases and transporters that are partially not present in the other bacteria. Moreover, the entire PUL 3 of *Formosa* B is missing in these other strains (Supplementary Figure S5). Instead, *Formosa* B's PUL 3 shows synteny to PULs of other marine *Flavobacteriia*, which do, however, not possess the PULs 1 and 2 (Supplementary Figure S5).

Laminarin elicits the expression of specific polysaccharide utilization loci in Formosa

Incubation experiments with fluorescently labeled (FLA) laminarin revealed the ability of *Formosa* B to quickly react and take up laminarin. *Formosa* B accumulated high amounts of FLA-laminarin after just 5 min of incubation (Fig. 1a). Additionally, the halo-like staining pattern showed that the FLA-laminarin was imported into the periplasm of the cells by a "selfish" uptake mechanism [41, 42]. Selfish substrate uptake is dependent on the presence of SusCD-like transporters and secures an enrichment of substrate in the periplasmic space without diffusive loss [42].

To elucidate the metabolism of *Formosa* B on laminarin and to verify whether laminarin specifically controls the expression of the genomically detectable PULs we performed cultivation experiments with this bacterium with purified laminarin as growth substrate. Growth curves of *Formosa* B in HaHa medium with laminarin, glucose and only protein extracts, respectively, are shown in Fig. 1b. We used proteomics to record the global protein expression patterns with these substrates. We investigated (i) the soluble intracellular proteome, (ii) the enriched membrane proteome, and (iii) the extracellular proteome (see Supplementary Information and Supplementary Tables S2A–C). These comparative analyses showed that although glucose is the monomer of laminarin, the utilization of either carbon source led to quite different proteomic signatures in different functional protein categories, such as in nucleotide, lipid, and coenzyme metabolism as well as in carbohydrate metabolism and transport (Supplementary Figure S6). About 100 proteins were significantly higher abundant or only found in laminarin incubations in *Formosa* B (Supplementary Figure S7, Supplementary Table S3). Of all three substrates, laminarin elicited the strongest expression of the three laminarin PULs of *Formosa* B (Fig. 1c), which is indicative of specific and tightly controlled expression. The SusD-like protein (FORMB_10080) of PUL 1 and the GH16 (FORMB_m24690) of PUL 2 were exclusively expressed with laminarin but not with the other substrates. Furthermore, the expression of PUL 3 was exclusively induced by laminarin and not detectable with glucose or only peptone (Fig. 1c). The specific response of the *Formosa* PULs to laminarin and not to glucose implies that the three-dimensional structure of laminarin might be the key to induce the expression of these PULs.

Biochemical analysis of laminarinases expressed by Formosa spp

To functionally characterize PUL-encoded proteins and to map the laminarin degradation pathway, we cloned and biochemically analyzed putative laminarinases the function of which could not be merrily solved by comparative sequence analyses with known enzyme functions. We cloned and examined the genes encoding FbGH17A (locus tag: FORMB_24720), FaGH17B (locus tag: FORMB_24740), and FbGH30 (locus tag:

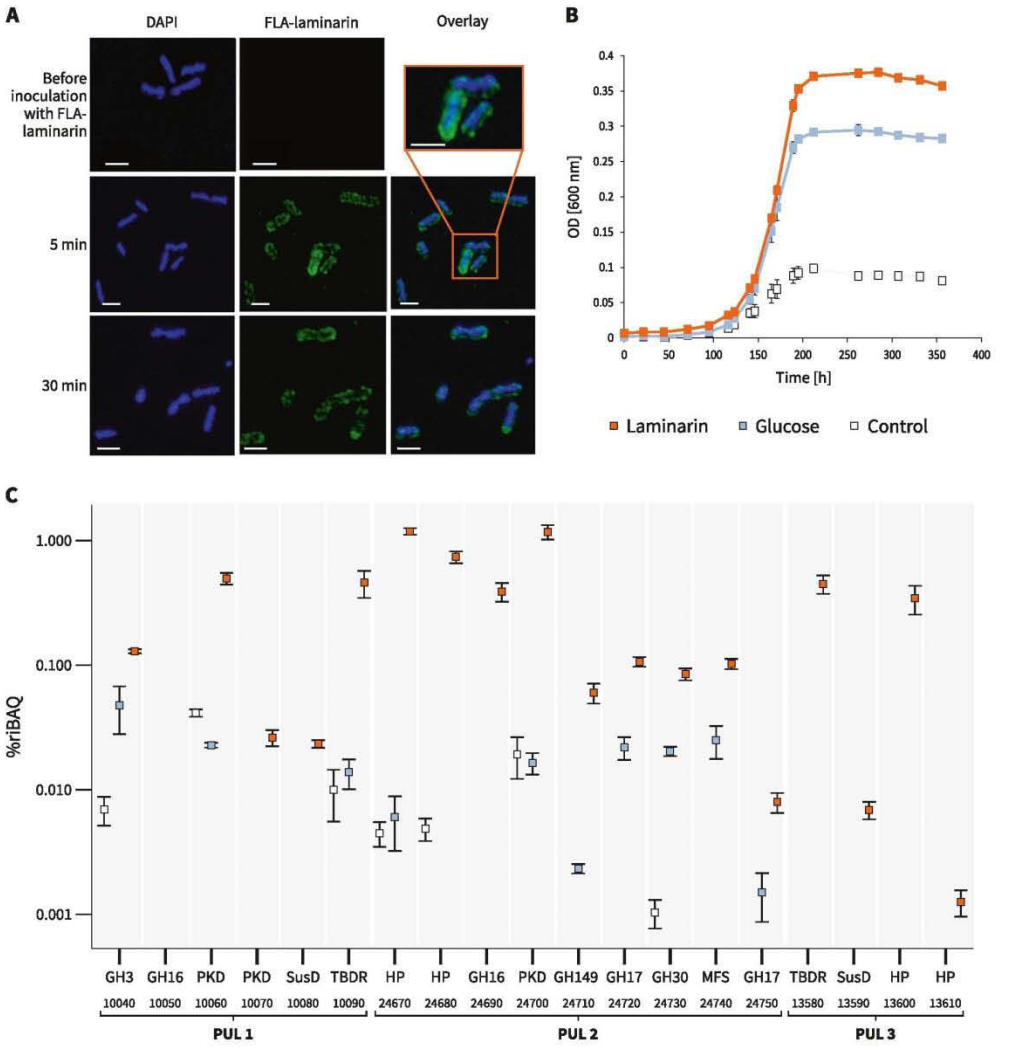


Fig. 1 Laminarin utilization of *Formosa* B. **a** SR-SIM of *Formosa* B cells before inoculation with FLA-laminarin at 5 and 30 min after incubation with FLA-laminarin. Images show cell staining by DAPI (left, blue), FLA-laminarin (middle, green), and an overlay showing both FLA-laminarin staining and DAPI (right). Scale bar = 1 µm **b** Growth curves of three biological replicates at 12 °C in modified HaHa_100V medium [21] with 2 g L^{-1} laminarin or 2 g L^{-1} glucose. The "control" culture contained only 0.1 g L^{-1} peptone, 0.1 g L^{-1} yeast extract, and 0.1 g L^{-1} casamino acids but no additional carbon sources. **c** Expression profile and gene organization of the laminarin utilization PULs 1–3 in *Formosa* B. Relative protein abundances (in % riBAQ) of PUL-encoded proteins detected in the membrane protein fractions of each three independent cultures grown on laminarin (orange), glucose (blue), and chitin (control, gray) are shown (for riBAQ values see Supplementary Tables S2A and S3). Putative protein functions (e.g., GH3) and the respective locus tags (e.g., 10040) are indicated. The squares represent the mean values of the replicates for every protein and each substrate. The error bars refer to the standard error of the mean. Proteins that could be detected in at least two out of three independent biological replicates of each substrate condition are shown (for individual replicate numbers see Supplementary Table S2A). GH, glycoside hydrolase; PKD, PKD-domain containing protein; SusD, SusD-family protein; HP, hypothetical proteins; MFS, major facilitator superfamily; TBDR, TonB-dependent receptor

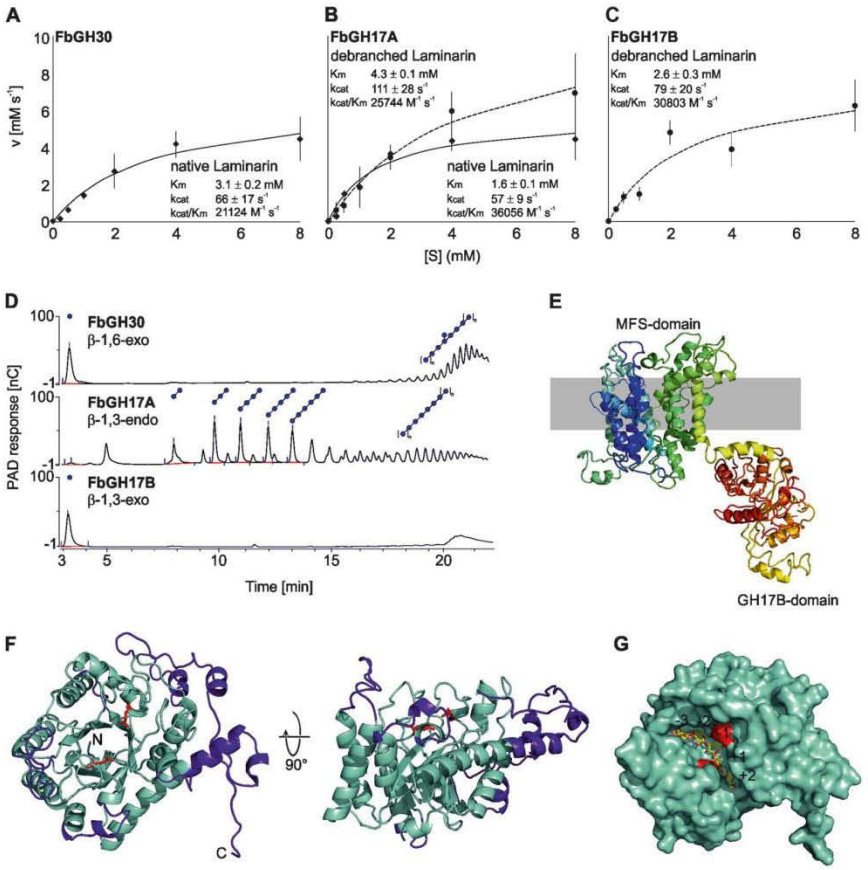


Fig. 2 Biochemical characterization of different laminarinases from *Formosa* B. Michaelis–Menten Kinetic (**a**) of FbGH30 on native laminarin, **b** of FbGH17A on native and debranched laminarin and **c** of FbGH17B on debranched laminarin (native laminarin is illustrated by the solid lines and debranched laminarin by the dashed lines). **d** Visualization of all three enzymatic activities was done using HPAEC-PAD. FbGH30 hydrolyzed native laminarin. After this debranching reaction, the laminarin was purified to remove glucose for the following steps. This debranched laminarin was used in the FbGH17A reaction. FbGH17B hydrolyzed the products of the previous FbGH17A reaction without any further purification in between. **e** 3D structure model of both the MFS-domain and the associated FbGH17B-domain and its potential arrangement within the inner-membrane. The modeling was performed using Phyre2. **f** The overall structure of FbGH17A is displayed in cyan with the additions colored purple. Highlighted in red are the catalytic residues. The N- and C-termini are labeled. **g** A surface view of FbGH17A with a modeled substrate complex shown as sticks in yellow from a GH17 transferase of *Rhizomucor miehei* with laminaritriose and laminaribiose in the −3 to −1 and + 1 to + 2 subsites, respectively

FORMB_24730). As all three proteins are encoded in a single gene cluster, we hypothesized that these enzymes might work together in spatial proximity. To test this hypothesis, we conducted a series of biochemical experiments, which revealed that the FbGH30 enzyme hydrolyzed the β-1,6-linked glucose side chains of laminarin (K_M: 3.1 ± 0.2 mM and K_{cat}/K_M: 21124 $M^{-1}s^{-1}$) (Fig. 2a), whereas it was inactive on the debranched substrate. The enzyme FbGH17A hydrolyzed both the debranched laminarin product of FbGH30 and the native laminarin, although with a markedly higher specific activity on the debranched product (K_M: 1.6 ± 0.1 mM; K_{cat}/K_M: 36056 $M^{-1}s^{-1}$) than on laminarin itself (K_M: 4.3 ± 0.1 mM; K_{cat}/K_M: 25744 $M^{-1}s^{-1}$) (Fig. 2b). Preference for debranched laminarin was even more pronounced with FbGH17B, which only hydrolyzed debranched laminarin (K_M: 2.6 ± 0.3 mM; K_{cat}/K_M: 30803 $M^{-1}s^{-1}$) (Fig. 2c) and was inactive on the branched form.

To elucidate how these enzymes work together in successive laminarin degradation, we used high-performance

liquid chromatography method with photo diode array detection analyses. The data indicated an enzymatic functional cascade in three steps (Fig. 2d): The exo-acting β-1, 6-glucosidase FbGH30 removes the glucose side chains from laminarin (Supplementary Figure S8A). The endo-acting β-1, 3-glucan hydrolase FbGH17A degrades the remaining debranched laminarin into oligosaccharides (Supplementary Figure S8B). The exo-acting β-1, 3-glucosidase FbGH17b processes these oligosaccharides into glucose (Supplementary Figure S8C). FbGH17b is part of a multi-modular protein, which is encoded by a gene that also codes for an N-terminal major facilitator superfamily (MFS) transporter, suggesting that hydrolysis and product uptake might be coupled. The MFS transporter contains 12 transmembrane-spanning helices (as predicted by Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index)), with the last C-terminal helix and the attached GH17 domain (Fig. 2e), which would enable the simultaneous cleavage of oligosaccharides and the sugar transported through the MFS. Blast analysis revealed that this fusion is common among marine *Flavobacteriia*, suggesting that such multi-modular transporter-associated enzyme may be a conserved mechanism for boosting laminarin utilization.

In order to examine the molecular basis of substrate specificity the X-ray crystal structure of GH17A was solved (Fig. 2f, see Supplementary Information). Compared to the monomeric GH17 structures, FbGH17A has significant insertions and is larger. The structure of GH17A allows for the deduction of the molecular basis of substrate specificity for the laminarinase. Based on the GH17 complexes obtained for *R. miehei* [32], a model was generated of a laminarin product involving five monomers bound to the catalytic groove, two on the aglycone side, and three on the glycone side (Fig. 2g). The reducing and non-reducing ends of the modeled glycan are free, suggesting the protein can act in the middle of the chain as expected for an endo-acting glycoside hydrolase (see Supplementary Information). Furthermore, given the conformation of the modeled glycan, 6-O-β-glucose branching would be possible only at subsite + 1 and anything further away (+ 3 or − 4). In other words, within the native polysaccharide the enzyme would need a stretch of at least three free β1,3-glucose moieties to act. This structural data supports the observation that GH17 activity on laminarin is bolstered by the action of the debranching enzyme GH30.

Laminarin stimulates the co-expression of selected peptidases and transporters

Formosa B encodes 69 peptidases in its relatively small genome (2.7 MB). Other laminarin-degrading marine *Bacteroidetes* like *Gramella forsetii* KT0803^T (79 peptidases), *Polaribacter* sp. Hel1_85 (84 peptidases), *Jejuia pallidilutea* (58 peptidases), and *Flaviramulus ichthyoenteri* (63 peptidases) show a comparable number of peptidase genes, although their genomes are around twice as large as that of *Formosa* B. Our proteome analysis of *Formosa* B revealed that 41 peptidases are expressed in the presence of laminarin (Supplementary Table S4). Nine of these peptidases showed a significantly higher protein abundance on laminarin in the enriched membrane proteome, compared with glucose or the control culture, or were exclusively found after incubation with laminarin (Fig. 3a and Supplementary Table S3). In addition, a putative peptide ABC transporter ATP-binding protein (FORMB_10920) and a putative oligopeptide permease ABC transporter protein (OppC; FORMB_20460) were detected, which showed a significantly higher abundance under laminarin conditions (Supplementary Table S3). This indicates a coupling of the peptide metabolism with laminarin utilization in *Formosa* B.

An exceptionally high expression with glucose and laminarin was visible for a putative porin (FORMB_11920, 10% riBAQ; Fig. 3b, c), an outer membrane protein, which was not detectable in the control cultivations with peptone (Supplementary Table S3). The porin-encoding gene is located in an operon with a putative ammonium transporter and clusters with several genes involved in nitrogen metabolism, including two putative glutamate synthase genes and an additional supposed ammonium transporter (Fig. 3c). All nitrogen metabolism-related genes in the direct vicinity of the porin-encoding gene were only found to be expressed with glucose and laminarin in peptone-containing cultures in comparison with the peptone-only control culture without these carbon sources (Fig. 3c).

In situ abundance and relevance of Formosa strain A and B

We investigated the in situ abundance of the *Formosa* strains A and B by recruiting *Formosa* reads from the 44 metagenomes of the years 2009–2012 from Helgoland bacterioplankton samples [18]. At the ≥ 95% average nucleotide identity (ANI) threshold, the strain A and B genomes recruited up to 0.28% and 2.94% of individual metagenomic reads in 2009, 0.04% and 0.99% for 2010, 0.03% and 0.98% for 2011, and 0.02% and 0.21% for 2012 (Supplementary Table S5), respectively. The mapped reads covered up to 91%, 99%, 97%, and 94% of the strain B genome from 2009 to 2012, respectively, and only up to 58% of the strain A genome in 2012 (Fig. 4 and Supplementary Table S5). This suggests that strain B was recurrent and abundant during the spring bloom events, whereas strain A was likely more representative for late summer blooms reaching highest abundances of mapped reads in September 2009. Reads mapped to the *Formosa* strain B genome with 70–93% ANI suggest the presence of other closely related

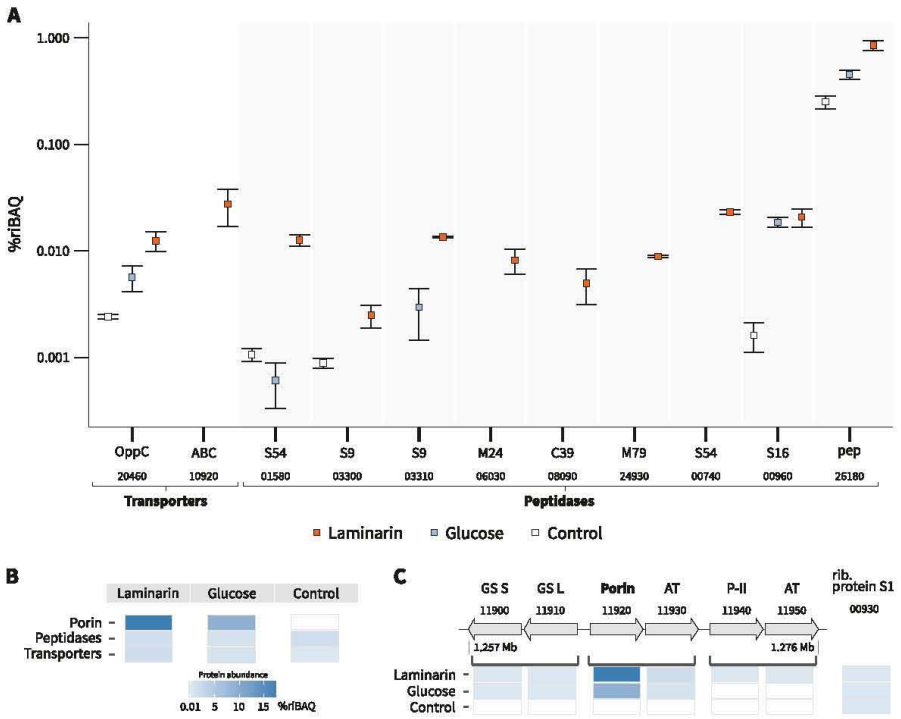


Fig. 3 Abundance of peptidases and putative peptide transporter proteins of *Formosa* B. **a** Proteomic signatures of selected proteases and two putative peptide transporters, which showed an increased expression under laminarin conditions. The two laminarin-induced peptidases shown on the right (00960, 26180) were also detected in the metaproteome of a spring bloom in 2009 (see Supplementary Table S8A). Relative protein abundances are depicted as % riBAQ values. The putative peptidase families and the respective locus tags are indicated. The squares represent the mean values of the three replicates for every protein and each substrate. The error bars refer to the standard error of the mean. Proteins that could be detected in at least two out of three independent biological replicates of each substrate condition are shown (for individual replicate numbers see Supplementary Table S2A). **b** Comparison of total protein abundances (in %riBAQ) of peptidases (see Supplementary Table S4), nitrogen-associated transporters and the porin (FORMB_11920) in *Formosa* B under the three investigated substrate conditions (see Supplementary Table S3). **c** Genomic structure of the porin-encoding cluster and the abundance patterns of the corresponding proteins. Brackets indicate putative operons. Protein functions and the respective locus tags are indicated. GS S: glutamate synthase subunit S, GS L: glutamate synthase subunit L, AT: ammonium transporter, P-II: nitrogen regulatory protein P-II

Formosa spp. during the spring blooms of 2009 to 2012 that reached up to 6.84%, 1.2%, 5.1%, and 1.8% of the metagenome reads, respectively (Fig. 4). Altogether, these results indicate that strain B is one of the representatives of the recurrent *Formosa* clade during North Sea spring microalgae blooms [17, 18].

Identification of Formosa-specific enzymes and transporters during microalgal blooms

All three *Formosa* B PULs were completely covered by metagenomic contigs of the spring bloom in 2009 and 2010 (Fig. 5a and Supplementary Tables S6–S7), and partially covered in the metagenomes of 2011 and 2012 (Supplementary Table S7). This illustrates the strong selection pressure imposed by laminarin on this pathway during four consecutive annual spring phytoplankton blooms in the North Sea.

We examined the presence of polysaccharide degradation- and consumption-related proteins of the *Formosa* strains A and B in the in situ metaproteomes of spring blooms in 2009 and 2010 (Supplementary Table S8). The proteome analysis of the planktonic bacterial fraction sampled during the spring bloom on 7 April 2009 uncovered 46 proteins from *Formosa* strain A and 361 proteins from *Formosa* strain B. Remarkably, several marker proteins from the putative laminarin-specific *Formosa* B PULs were highly abundant (Fig. 5b and Supplementary Table S8A) in

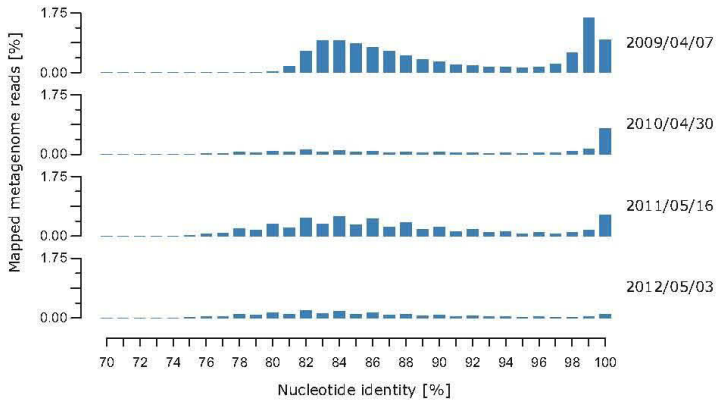


Fig. 4 Relative abundance of *Formosa* strain B and related species during four spring phytoplankton blooms indicated by the percentage of metagenomics reads mapped at different nucleotide identities. Reads recruited at ≤ 93% nucleotide identity represent other *Formosa* spp. that are abundant during the spring bloom events at Helgoland, Germany from 2009 until 2012. Dates on the right indicate the four metagenomes (i.e., time points), which produced highest mapping coverage with the *Formosa* strain B genome in their respective years. For a summary of all 44 metagenomes (up to 18 time points per year) and their mapping results see Supplementary Table S5

the metaproteome samples. This analysis identified 13 proteins of the PULs 1, 2, and 3 (see also Supporting Information) and thus indicated that a significant proportion of *Formosa* B's laminarin PULs were expressed in situ during the spring bloom in 2009. Although the metaproteome analysis of 2010 uncovered fewer proteins from both *Formosa* strains, three marker proteins of PUL 1 from *Formosa* B were detected in the environmental samples (see Supplementary Information and Supplementary Table S8B).

Besides glycoside hydrolases and laminarin-specific transporter proteins, we also identified several *Formosa* B proteins in the environmental metaproteome samples of 2009, which are involved in the central catabolism of the monosaccharide glucose, the product of laminarin hydrolysis (see Fig. 5b, Supplementary Information and Supplementary Table S8A). This includes nearly all glycolytic enzymes as well as a putative glycogen synthase of *Formosa* B. These data indicate that the *Formosa* B strain substantially contributed to laminarin degradation and turnover during a diatom-driven phytoplankton bloom.

In addition, several proteins of *Formosa* B involved in nitrogen metabolism could be detected in the metaproteome analyses of the spring bloom 2009 (Supplementary Table S8A). This includes the putative porin (FORMB_11920), a peptide ABC transporter ATP-binding protein (FORMB_10920), an oligopeptide permease ABC transporter protein (OppC; FORMB_20460), and eight peptidases (Fig. 5b). This underlines a strong coupling of the peptide metabolism with laminarin utilization of *Formosa* B under in situ conditions.

Discussion

This study provides detailed insights in the adaptations which make *Formosa* strains successful competitors in the early breakdown of organic matter during diatom blooms. Combining comparative in vitro and in situ proteogenomics with biochemical enzyme characterization reveals that the key to this process is the sensing and utilization of laminarin. Our data indicate that this polysaccharide is used in two ways: as a major source of energy, and as a signal molecule, which induces transporters and digestive enzymes to use also other compounds released from the lysis of diatom cells.

The two environmentally relevant *Formosa* strains examined in this study feature streamlined genomes, which are significantly smaller than those of many other marine *Flavobacteriia*. With a lower number of total proteins to synthesize, *Formosa* A and B can dedicate a higher relative proportion of their genomic and proteomic resources to the digestion of laminarin. Their CAZyme repertoire is strongly reduced compared to versatile polysaccharide degraders such as *F. agariphila* [24] and *Zobellia galactanivorans* [35], which were isolated from macroalgae. It is, however, similar to another member of North Sea spring bacterioplankton, *Polaribacter* sp. Hel1_33_49 [7]. In contrast to macroalgae-associated laminarin-degrading bacteria, such as *Z. galactanivorans* [43], neither of the *Formosa* strains possesses a mannitol dehydrogenase, which indicates a specialization of *Formosa* A and B to chrysolaminarin. This type of laminarin lacks mannitol residues and is preferentially produced by diatoms.

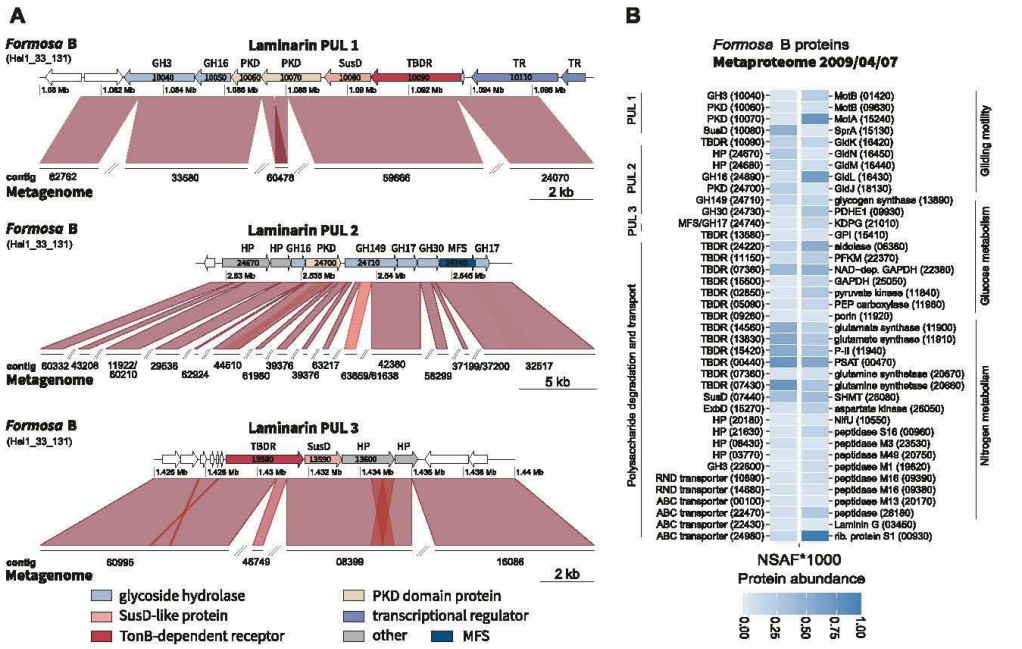


Fig. 5 Detection of *Formosa* strain B laminarin PULs in the Helgoland spring bloom metagenome and metaproteome in 2009 [17]. **a** Synteny between the laminarin PULs of *Formosa* sp. Hel1_33_131 and partial PUL sequences in the metagenomes from 2009/04/07. The sequence comparisons were performed with Bl2seq (BLASTn, *E* value 1e-5). Sequence similarities are depicted by red hues for direct comparisons. Darker colors correspond to higher identities. Gene locus tags are subsequent numbers within PULs and are indicated in the figure for most of the genes (for visibility's sake, gene names of very small genes were omitted). **b** Heatmap of the relative abundance of *Formosa* strain B proteins (displayed as normalized spectral abundance factor values, NSAF*1000) detected in the metaproteome from 07 April 2009. Displayed are selected proteins, which likely play a role in polysaccharide or protein utilization. A highly abundant ribosomal protein of *Formosa* strain B (rib. protein S1, lower right) is also displayed as a reference to illustrate the high abundance of polysaccharide utilization-specific proteins during the bloom condition. Gene locus tag numbers are given in parenthesis. GH, glycoside hydrolase; PKD, PKD-domain containing protein; SusD, SusD-family protein; TBDR, TonB-dependent receptor; HP, hypothetical protein; MFS, major facilitator superfamily; ExbD, subunit of the Ton system for energy transduction; MotB, motor rotation protein; Gld, gliding motility proteins; PDHE, pyruvate dehydrogenase E1 component; KDPG, 2-dehydro-3-deoxyphosphogluconate aldolase; GPI, glucose-6-phosphate isomerase; PKFM, 6-phosphofructokinase; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; P-II, nitrogen regulatory protein P-II; PSAT, phosphoserine aminotransferase; SHMT, serine hydroxymethyltransferase

We found a specific laminarin protein abundance pattern in *Formosa* B, which differs from the protein expression pattern in presence of the sugar monomer of this polysaccharide, glucose. A similar laminarin-specific control of gene expression was suggested for the marine flavobacterium *G. forsetii* [4]. Interestingly, this laminarin-specific proteome signature of *Formosa* B includes not only the proteins required for laminarin uptake and utilization, but also peptidases and transporters for amino-acid utilization. The *Formosa* cells, upon sensing of laminarin, thus appear to react in two ways: First, they enhance the expression of outer membrane proteins to degrade and rapidly transport the energy molecule laminarin into their periplasm, utilizing the selfish polysaccharide uptake mechanism recently demonstrated for marine *Flavobacteriia* [42]. Second, the abundance of amino-acid and nitrogen metabolism-related proteins is increased to boost the recycling of nitrogen building blocks, which are required for rapid growth of *Formosa* bacteria and become available simultaneously with laminarin upon algal lysis.

Formosa strain B possesses an extended repertoire of laminarin-specific enzymes and transporters, which is larger than that of other laminarin-degrading bacteria such as *Polaribacter* sp. Hel1_33_49 [7] or *G. forsetii* [4]. Our subproteome and bioinformatic analyses indicate that many laminarin-degrading enzymes of *Formosa* B are surface-tethered or localized in the periplasmic space and in the cytoplasmic membrane, respectively. The different TBDR, laminarinases, transporters, and additional enzymes combine complementary activities into an efficient laminarin

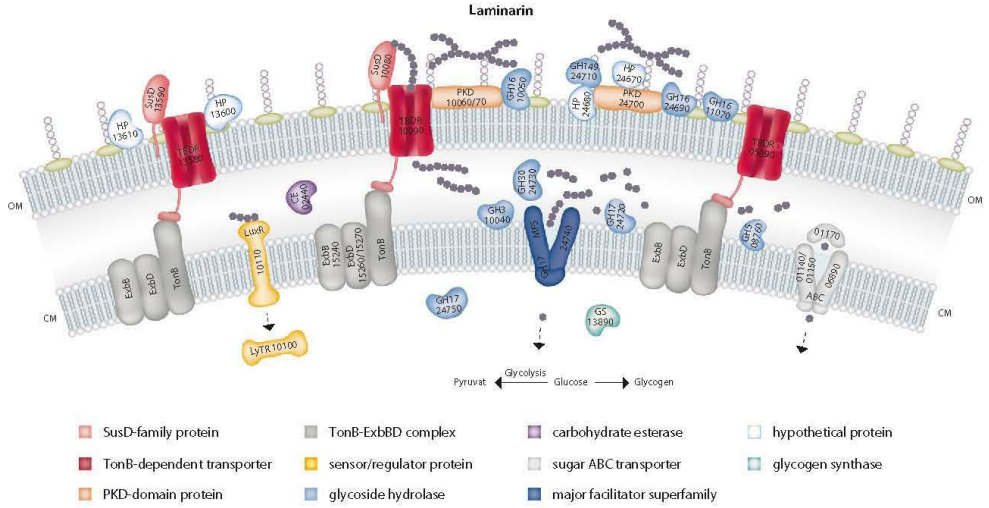


Fig. 6 A tentative model of laminarin utilization pathways in *Formosa* B. Protein localizations were predicted in silico according to Romine [46] and were deduced by subproteome analyses (see also Supplementary Table S3). Additional biochemical experiments are required to ascertain this model

disassembly line for degradation and uptake (Fig. 6). The biochemical experiments presented here support the annotation of the conserved cluster of genes as encoding for a laminarin utilization pathway. Here, two enzymes that are likely residents of the periplasm are shown to work together towards the complete degradation of laminarin in a highly specific manner. The X-ray crystal structure of GH17A reveals the possible molecular determinants of substrate specificity and the propensity of the enzyme to be more active on unbranched laminarin.

The multi-modular protein FORMB_24740 (FbGH17b) combines a glycoside hydrolase (GH17) with a membrane spanning transport protein and may represent an adaptive mechanism for laminarin utilization. The integration of the transport and hydrolysis processes into a single protein could facilitate improved consumption of the sugars by increasing the activity of the fused GH17. Increased activity would also reduce the necessary enzyme copy number, and thereby resource consumption for synthesis of this protein. To our knowledge, such a transporter-CAZyme-fusion has not been described for other bacteria as yet, but its conservation in nature suggests that this could provide a significant benefit.

The exceptionally strong accumulation of a putative porin in glucose- and laminarin-controlled cultures and the co-induction of the 12 surrounding genes of this porin-encoding genomic cluster, all of which play a role in nitrogen metabolism, could indicate a function of this transporter protein in the uptake of peptides as nitrogen and amino-acid source. Endo-acting proteases might degrade proteins released by lysed microalgae into peptides, which are then imported through the porin into the periplasm. An efficient capture of these peptides with a highly abundant porin system might be especially useful in the highly diffusive marine environment. With laminarin and glucose as easily metabolizable carbon sources, such a strategy could be crucial for a balanced carbon and nitrogen diet.

Members of the phylum *Bacteroidetes* are primary degraders of microalgal polysaccharides during phytoplankton blooms, and are therefore key players in marine carbon cycling. However, underlying enzymatic mechanisms and adaptations that drive the specialization of these highly competitive bacteria remain obscure. We reveal and prove in this study the specific activity and ecological niche of two abundant marine *Bacteroidetes* strains in complex microbial communities during diatom-driven phytoplankton blooms. Our results show an extraordinary degree of specialization for the *Formosa* strains A and B, which enable these marine *Bacteroidetes* to successfully compete for laminarin against a multitude of other laminarin-degrading microbes in bloom situations [1, 4, 7, 44, 45]. Our data furthermore indicate that fast growth on beta-glucans such as laminarin requires a balanced diet that also includes nitrogen sources like peptides. The induction of several cell wall-associated peptidases and peptide-specific transporters in *Formosa* B during growth on laminarin suggests that these bacteria pursue a complex uptake strategy, which encompasses both sugars and nitrogen compounds. This

may make marine *Flavobacteriia* so successful in their diffusion-open environment.

Acknowledgements We are grateful to Jana Matulla for technical assistance, Sebastian Grund for mass spectrometry analyses, Carol Arnosti and her laboratory for producing the FLA-labeled laminarin, and Nicolas Terrapon for critical comments on the putative function of selected PUL-encoded proteins. We thank the crew of the Biological Institute Helgoland of the Alfred-Wegener-Institute, especially Antje Wichels, for their excellent logistic support during our sampling campaigns at the Kabeltonne Helgoland from 2009 until 2012. The Federal Ministry of Education and Research (BMBF) funded parts of this work within the "Microbial Interactions in Marine Systems" project (MIMAS project 03F0480A). The work was also financially supported by the DFG in the framework of the research unit FOR2406 "Proteogenomics of Marine Polysaccharide Utilization" (POMPU) by grants of R. Amann (AM 73/9-1), H. Teeling (TE 813/2-1), B. Fuchs (FU 627/2-1), D. Becher (BE 3869/4-1), J.-H. Hehemann (HE 7217/2-1), and T. Schweder (SCHW 595/10-1). F. Unfried was supported by scholarships from the Institute of Marine Biotechnology e.V. and the Ph.D. graduate program of the International Max Planck Research School of Marine Microbiology (MarMic).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Alderkamp AC, van Rijssel M, Bolhuis H. Characterization of marine bacteria and the activity of their enzyme systems involved in degradation of the algal storage glucan laminarin. FEMS Microbiol Ecol. 2007;59:108–17.
2. Keith S, Arnosti C. Extracellular enzyme activity in a river-bay-shelf transect: variations in polysaccharide hydrolysis rates with substrate and size class. Aquat Microb Ecol. 2001;24:243–53.
3. Arnosti C, Durkin S, Jeffrey W. Patterns of extracellular enzyme activities among pelagic marine microbial communities: implications for cycling of dissolved organic carbon. Aquat Microb Ecol. 2005;38:135–45.
4. Kabisch A, Otto A, Konig S, Becher D, Albrecht D, Schuler M, et al. Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes 'Gramella forsetii' KT0803. ISME J. 2014;8:1492–502.
5. Labourel A, Jam M, Jeudy A, Hehemann JH, Czjzek M, Michel G. The beta-glucanase ZgLamA from Zobellia galactanivorans evolved a bent active site adapted for efficient degradation of algal laminarin. J Biol Chem. 2014;289:2027–42.
6. Labourel A, Jam M, Legentil L, Sylla B, Hehemann J-H, Ferrières V, et al. Structural and biochemical characterization of the laminarinase ZgLamCGH16 from Zobellia galactanivorans suggests preferred recognition of branched laminarin. Acta Crystallogr Sect D Biol Crystallogr. 2015;71:173–84.
7. Xing P, Hahnke RL, Unfried F, Markert S, Huang S, Barbeyron T, et al. Niches of two polysaccharide-degrading Polaribacter isolates from the North Sea during a spring diatom bloom. ISME J. 2015;9:1410–22.
8. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. Polysaccharide utilization loci: fuelling microbial communities. J Bacteriol. 2017;199:00860–16.
9. Hehemann JH, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature. 2010;464:908–12.
10. Hehemann JH, Arevalo P, Datta MS, Yu X, Corzett CH, Henschel A, et al. Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. Nat Commun. 2016;7:12860.
11. Hehemann JH, Truong LV, Unfried F, Welsch N, Kabisch J, Heiden SE, et al. Aquatic adaptation of a laterally acquired pectin degradation pathway in marine gammaproteobacteria. Environ Microbiol. 2017;19:2320–33.
12. Cockburn DW, Koropatkin NM. Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease. J Mol Biol. 2016;428:3230–52.
13. El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Microbiol. 2013;11:497–504.
14. Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, et al. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol. 2011;9:e1001221.
15. Ndeh D, Rogowski A, Cartmell A, Luis AS, Basle A, Gray J et al. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. Nature. 2017;544:65–70.
16. Needham DM, Fuhrman JA. Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. Nat Microbiol. 2016;1:16005.
17. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science. 2012;336:608–11.
18. Teeling H, Fuchs BM, Bennke CM, Kruger K, Chafee M, Kappelmann L, et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. Elife. 2016;5:e11888.
19. Buchan A, LeCleir GR, Gulvik CA, Gonzalez JM. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat Rev Microbiol. 2014;12:686–98.
20. Becker S, Scheffel A, Polz MF, Hehemann JH. Accurate quantification of laminarin in marine organic matter with enzymes from marine microbes. Appl Environ Microbiol. 2017;83:pii:e03389–16.
21. Hahnke RL, Bennke CM, Fuchs BM, Mann AJ, Rhiel E, Teeling H, et al. Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. Environ Microbiol. 2015;17:3515–26.
22. Chafee M, Fernandez-Guerra A, Buttigieg PL, Gerdts G, Eren AM, Teeling H et al. Recurrent patterns of microdiversity in a temperate coastal marine environment. ISME J. 2017;12:237–52.
23. Zhou J, Bruns MA, Tiedje JM. DNA recovery from soils of diverse composition. Appl Environ Microbiol. 1996;62:316–22.
24. Mann AJ, Hahnke RL, Huang S, Werner J, Xing P, Barbeyron T, et al. The genome of the alga-associated marine flavobacterium Formosa agariphila KMM 3901T reveals a broad potential for

degradation of algal polysaccharides. Appl Environ Microbiol. 2013;79:6813–22.
25. Otto A, Bernhardt J, Meyer H, Schaffer M, Herbst FA, Siebourg J, et al. Systems-wide temporal proteomic profiling in glucose-starved Bacillus subtilis. Nat Commun. 2010;1:137.
26. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26:1367–72.
27. Shin JB, Krey JF, Hassan A, Metlagel Z, Tauscher AN, Pagana JM, et al. Molecular architecture of the chick vestibular hair bundle. Nat Neurosci. 2013;16:365–74.
28. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016;13:731–40.
29. Vizcaino JA, Cote RG, Csordas A, Dianes JA, Fabregat A, Foster JM, et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 2013;41:D1063–69.
30. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66:213–21.
31. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40:658–74.
32. Qin Z, Yan Q, Lei J, Yang S, Jiang Z, Wu S. The first crystal structure of a glycoside hydrolase family 17 beta-1,3-glucanosyltransferase displays a unique catalytic cleft. Acta Crystallogr D Biol Crystallogr. 2015;71:1714–24.
33. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr. 2006;62:1002–11.
34. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011;67:355–67.
35. Barbeyron T, Thomas F, Barbe V, Teeling H, Schenowitz C, Dossat C, et al. Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: example of the model algae-associated bacterium Zobellia galactanivorans DsijT. Environ Microbiol. 2016;18:4610–27.
36. Bauer M, Kube M, Teeling H, Richter M, Lombardot T, Allers E, et al. Whole genome analysis of the marine Bacteroidetes'-Gramella forsetii' reveals adaptations to degradation of polymeric organic matter. Environ Microbiol. 2006;8:2201–13.
37. Kuhaudomlarp S, Patron NJ, Henrissat B, Rejzek M, Saalbach G, Field RA. Identification of Euglena gracilis beta-1,3-glucan phosphorylase and establishment of a new glycoside hydrolase (GH) family GH149. J Biol Chem. 2018;293:2865–76.
38. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, et al. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. Cell. 2010;141:1241–52.
39. Tang K, Jiao N, Liu K, Zhang Y, Li S. Distribution and functions of TonB-dependent transporters in marine bacteria and environments: implications for dissolved organic matter utilization. PLoS ONE. 2012;7:e41204.
40. Panschin I, Huang S, Meier-Kolthoff JP, Tindall BJ, Rohde M, Verbarg S, et al. Comparing polysaccharide decomposition between the type strains Gramella echinicola MM 6050(T) (DSM 19838(T)) and Gramella portivictoriae UST040801-001(T) (DSM 23547(T)), and emended description of Gramella echinicola Nedashkovskaya et al. 2005 emend. Shahina et al. 2014 and Gramella portivictoriae Lau et al. 2005. Stand Genom Sci. 2016;11:37.
41. Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron E, Pudlo NA, et al. Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. Nature. 2015;517:165–9.
42. Reintjes G, Arnosti C, Fuchs BM, Amann R. An alternative polysaccharide uptake mechanism of marine bacteria. ISME J. 2017;11:1640–50.
43. Groisillier A, Labourel A, Michel G, Tonon T. The mannitol utilization system of the marine bacterium Zobellia galactanivorans. Appl Environ Microbiol. 2015;81:1799–812.
44. Bennke CM, Kruger K, Kappelmann L, Huang S, Gobet A, Schuler M, et al. Polysaccharide utilisation loci of Bacteroidetes from two contrasting open ocean sites in the North Atlantic. Environ Microbiol. 2016;18:4456–70.
45. Cardman Z, Arnosti C, Durbin A, Ziervogel K, Cox C, Steen AD, et al. Verrucomicrobia are candidates for polysaccharide-degrading bacterioplankton in an arctic fjord of Svalbard. Appl Environ Microbiol. 2014;80:3749–56.
46. Romine MF. Genome-wide protein localization prediction strategies for gram negative bacteria. BMC Genomics. 2011;12:S1.

Chapter 5

Discussion

“In so far as a scientific statement speaks about reality, it must be falsifiable; and in so far as it is not falsifiable, it does not speak about reality.

Karl Popper

5) Discussion

5.1) Defining niche components of the marine *Flavobacteriia* and *Gammaproteobacteria*

During the course of this doctoral thesis, I investigated the genomic and metabolic potential of the genera *Reinekea*, *Polaribacter,* and *Formosa* in order to generate hypotheses for their ecological niches. In this general discussion, I first compare the common and variable niche components of these flavobacterial and gammaproteobacterial genera.

As common metabolisms, the studied bacterial clades encode genes for proteorhodopsin and phosphorus acquisition (Fig. 5.1). It was demonstrated that some proteorhodopsin types are used to generate light-induced energy production and provide a mixotrophic lifestyle for heterotrophic bacteria (1). '*Reinekea forsetii*' further extends this energy metabolism with dissimilatory sulfur oxidation (Chapter 2, Figure S2). Mixotrophy is suggested to be prevalent in marine bacterioplankton and to have a profound impact on the carbon cycling in surface oceans (2). This versatility in energy metabolism could provide a competitive advantage, especially during the starvation period. Furthermore, examined bacterial clades have the genetic machinery to synthesize or to hydrolyze polyphosphate (Fig. 5.1). Phosphate concentrations quickly become limiting during spring blooms (3) (4), because phosphate is taken up by both phytoplankton and bacterioplankton. The polyphosphate storage, therefore, could increase the fitness of heterotrophic bacteria during bloom events. *Reinekea* spp. also encode genes to utilize phosphonate (Fig. 5.1), an important component of organic phosphorus in the oceans (5). Accessing this secondary pool could further bolster the selective advantage of *Reinekea* in phosphorus-limited environments.

Close association with algae during spring blooms enables access to the phycosphere, the region surrounding a phytoplankton cell that is enriched by exudates, and provides efficient uptake of released organic matter (6). Likewise, the analyzed *Flavobacteriia* and *Gammaproteobacteria* could associate with algae using different mechanisms (Fig. 5.1). *Reinekea* spp. possess mannose-sensitive haemagglutinin (MSHA)-like pilus genes, which were shown to promote bacterial attachment to surfaces of the green alga

Ulva australis (7). In contrast, *Formosa* and *Polaribacter* spp. encode gliding motility genes that enable the movement of bacteria along surfaces and might facilitate substrate exploration and subsequent colonization of algal particles (8). Overall, these results suggest that (i) generating energy from diverse sources, (ii) survival in the phosphate-limited environment, and (iii) association with algae are likely important commonalities of investigated genera (Fig. 5.1).

The studied *Flavobacteriia* and *Gammaproteobacteria* further differentiate their niches by utilizing diverse high molecular weight compounds with various mechanisms (Fig. 5.1). The genera *Formosa* and *Polaribacter* likely possess a coupled metabolism to utilize polysaccharides and proteins released during algal blooms. Overexpression of 41 peptidase and porin-encoding genes in the presence of laminarin suggested a coupled protein and carbohydrate metabolism for *Formosa* strain B (Chapter 4, Fig. 3). Such a mechanism could also be present in *Polaribacter* 2-a and 2-b. These clades encode putative β-glucan PULs with a combination of glycoside hydrolases (GHs) and peptidases (Chapter 3, Fig. 7). This coupled mechanism could enable an efficient protein and polysaccharide uptake in the highly dilute marine environment.

Formosa and Polaribacter also encode for a large number of PULs that target diverse algal polysaccharides. The analyzed *Formosa* species are specialized on laminarin utilization with an extended repertoire of specific enzymes and transporters for the “selfish” uptake (Chapter 4, Fig.1) and degradation of this algal glycan (Chapter 4, Fig. 6). For instance, a multi-modular GH17 family protein integrates transport and hydrolysis processes (Chapter 4, Fig. 2) and facilitates an adaptive mechanism for improved laminarin utilization (Chapter 4, Fig. 1). In addition to this specialized laminarin operon, PULs putatively targeting α-mannan and β-xylose-containing polysaccharides and chitin were also annotated in the *Formosa* strains A and B (9). On the other hand, six distinct clades of the North Sea *Polaribacter* encode 78 putative PULs targeting a large variety of algal substrates such as α- and β-glucans, mannose-rich sulfated polysaccharides, alginate, and arabinogalactan (Chapter 3, Fig. 7). Orchestrating the binding, uptake, and degradation of particular polysaccharides, this PUL spectrum suggests a broader

carbohydrate utilization metabolism for the North Sea *Polaribacter* clades. Overall, these specialized metabolic and genetic mechanisms provide an ecological advantage for the two examined *Flavobacteriia* in degradation of various high molecular weight compounds (Fig. 5.1).

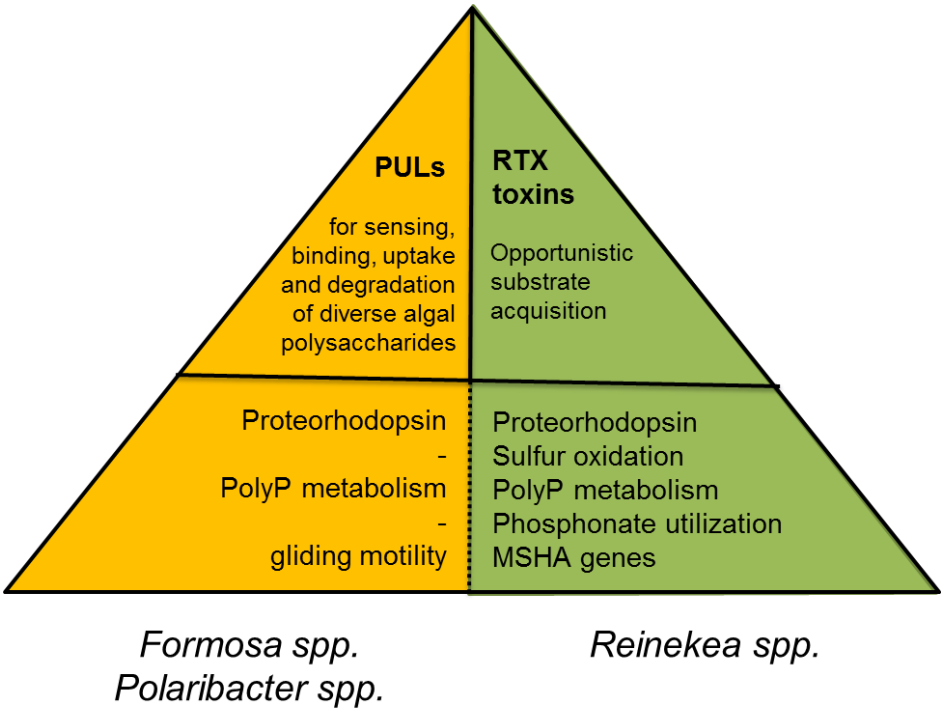


Fig. 5.1. An ecological comparison of the three studied genera based on their genetic repertoires.

The gammaproteobacterial isolate '*Reinekea forsetii*' features an opportunistic substrate acquisition strategy. It does not encode any PULs, but possess high numbers of CAZymes specific for the degradation of α-glucans and peptidoglycan (Fig. 5.1). '*R. forsetii*', encodes fourteen GH13 genes (Chapter 2, Table S5), which are known to degrade α-glucans such as glycogen. Five GH23 and two GH73 genes (Chapter 2, Table S5), predominantly comprising peptidoglycan lyases, were also annotated. Glycogen is a bacterial and animal storage compound (10) and peptidoglycan is released when bacteria disintegrate during bloom events. Moreover, '*R. forsetii*' encodes RTX toxin genes and its crude cell extracts have an inhibitory effect on *Polaribacter* Hel1_88 cultures (Chapter 2, Fig. 4). Therefore, *Reinekea* spp. might not degrade algal polysaccharides, but rather actively acquire substrates from other abundant bacteria that feed on algal biomass during bloom events. This

mechanism suggests an opportunistic lifestyle for this bloom-associated gammaproteobacterium and highlights an important ecological difference to the investigated *Flavobacteriia*, which primarily degrade algal polysaccharides with a specific genetic machinery. Overall, core metabolisms such as phosphorus acquisition and algae association contribute to the selection of the examined *Flavobacteriia* and *Gammaproteobacteria* and distinct polysaccharide utilization strategies likely differentiate their niche space during North Sea spring algal blooms (Fig. 5.1).

5.2) Effects of deterministic and stochastic forces on microbial community composition

The diversity of bacterial communities in conjunction with the genetic repertoire was explored using a dense sampling over multiple years in the southern North Sea. These efforts yielded a time-series suited for the interpretation of the effect of stochastic and deterministic forces on the microbial species composition. Previous studies indicated selection of a particular group of heterotrophic bacteria with distinct glycan utilization potential (3) (4) and their recurrence during North Sea spring algal blooms off Helgoland (11). My research explores the niches of these bacteria to shed light on the ecological strategies enabling their recurrent selection during bloom events.

Recurrence of members of the genera *Formosa* and *Polaribacter* during the North Sea spring algal blooms suggests that deterministic forces are at play. As one of the representative of recurrent *Formosa* clade, strain B reached relative abundances up to 2.9%, 1%, 1% and 0.2% of the free-living bacterioplankton during 2009-2012 spring algal blooms in the southern North Sea (4). This strain also has a pronounced specialization on laminarin (Chapter 4, Fig. 5) that is coupled to peptide utilization (Chapter 4, Fig. 3). *Polaribacter* 2-a is also recurrently abundant as a first responder in the 2009-2012 blooms (1.1%, 5.9%, 1.2%, and 0.4%) (Chapter 3, Fig. 3) with a β-glucan PUL containing both CAZymes and peptidases (Chapter 3, Fig. 7). This suggests a coupled peptide and carbohydrate metabolisms as well. These commonalities are further supported by small genome sizes and high peptidase contents (Chapter 3, Fig. 6). Such streamlined mechanisms could

enable quick growth of *Formosa* and *Polaribacter* 2-a, particularly during the initial termination of algal blooms, and thus contribute their selection.

Other *Polaribacter* clades were also recurrent during the investigated North Sea spring algal blooms. *Polaribacter* 3-b was detected in the late phase of the 2009-2012 bloom events with relative abundances up to 3.4%, 10.2%, 0.6%, and 3.2% (Chapter 3, Fig.3). This clade has a distinctive sulfated xylan PUL, and the expression of its corresponding *susC* gene was demonstrated (Chapter 3, Fig. 7). Sulfated xylans are likely less abundant and accessible compared to the storage polysaccharide laminarin, but could still provide a dedicated niche, which might explain the recurrent abundance of *Polaribacter* 3-b. Furthermore, the clade *Polaribacter* 1-a also reached high abundances (3%, 16%, 29%) in 2009, 2011 and 2012, most notably following blooms of *Chattonella* algae (Chapter 3, Fig.3). This co-occurrence could suggest a specific interaction or feeding mechanism. Such specialization is also corroborated by large genome sizes and wide polysaccharide utilization capacity of this clade. Therefore, particular polysaccharides released from *Chattonella* might provide a specific niche for *Polaribacter* 1-a. Overall, these observations suggest that the release of high molecular weight compounds during spring algal blooms creates certain niche spaces in a deterministic manner, which recurrently selects for particular clades of adapted *Flavobacteriia* (Fig. 5.2).

The *Reinekea* clade had a varying abundance pattern during the investigated North Sea spring algal blooms. CARD-FISH assessments with the genus-specific probe REI731 showed that *Reinekea* spp. were present in 2009-2012 bloom events albeit with decreasing abundances (~16%, ~5%, ~2%, and ~1%) (Chapter 2, Fig. S1). Additionally, oligotyping analysis of 16S rRNA tag sequences from 2013 and 2016 bloom events revealed read abundances of 1% and 0.001%, respectively (Meghan Chafee, personal communication). In 2017, no *Reinekea* oligotype was detected (T. Ben Francis, personal communication). In support of the postulates of the neutral theory, these observations suggest the random selection of *Reinekea* spp. and indicate the relative role of stochastic processes in addition to niche-based explanations. Such a dual mechanism was also demonstrated to structure microbial

communities in a soil microbiome (12). Furthermore, in a recent study, Rivett and Bell suggested a link between abundance and function of certain taxa in complex microbial communities and concluded that rare phylotypes influence narrow functional measures such as degradation of specific substrates (13). Likewise, '*Reinekea forsetii*' specifically encodes two GH26 genes (Chapter 2, Table S5), which were not detected in twenty closely-related marine *Gammaproteobacteria* genomes (Chapter 2, Table S6). These genes are likely involved in the decomposition of mannan (14), a polysaccharide that is found in the diatom cell walls (15). Moreover, *Reinekea* spp. were abundant only during spring blooms, and their abundances decreased together with that of the diatom *Thalassiosira nordenskioeldii* between 2009 and 2012 (Chapter 2, Fig. S1). Therefore, this specific mannan utilization capacity could create a narrow niche space and sustain North Sea *Reinekea* clade as a rare taxon during North Sea spring algal blooms that occasionally spikes under favorable conditions. These results overall suggest the potential influence of both deterministic and stochastic processes on the ecological selection of the examined gammaproteobacterium during North Sea spring algal blooms (Fig. 5.2).

The fluctuating abundance and substitution of recurrent bacteria by other clades suggest functionally-driven niche spaces during the North Sea spring algal blooms. For example, decreasing abundances of *Reinekea* spp. coincided with increasing proportions of *Balneatrix* and *Alteramonadaceae* clades between the 2009 and 2012 blooms (Fig. 1.9). Likewise, *Polaribacter* 2-a peaked together with *Ulvibacter* spp. (recently reclassified as *Cd*. 'Prosiliicossus') and VIS6 clade in 2010 and was detected in lower abundances during the 2011 and 2012 spring blooms, while *Ulvibacter* and VIS6 had the relatively higher numbers (Fig. 1.9) (Chapter 3, Fig. 3). Similar to *Polaribacter* 2-a, representative of the *Prosiliicoccus* clade, *Cd*. 'P. vernus' also have a streamlined genome with a limited polysaccharide degradation capacity coupled with a wide protein utilization potential (T. Ben Francis, personal communication).

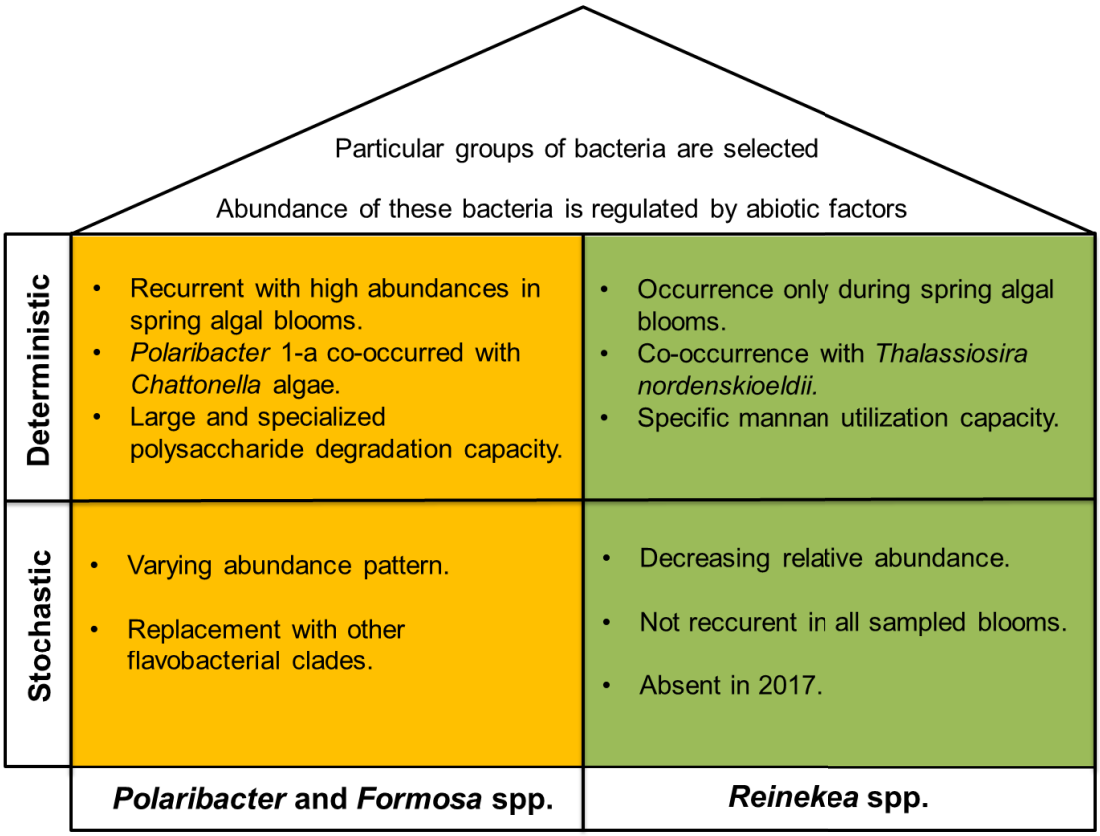


Fig. 5.2. A hypothetical model demonstrating the effect of deterministic and stochastic processes on the assembly of examined *Flavobacteriia* and *Gammaproteobacteria* during North Sea spring algal blooms. Abiotic factors likely regulate the assembly of heterotrophic bacteria selected during spring algal blooms.

The functional and phylogenetic dynamics of recurrent bacteria could be explained by the lottery hypothesis, which is derived from the investigation of coral reef fish communities (Fig. 5.2) (16). This hypothesis states that the initial colonization of an ecological niche is random within a functionally equivalent group of species. A suitable species within this group happens to arrive first occupies and dominates the niche space. Incorporating both neutral and deterministic aspects, lottery hypothesis is also applied in microbial ecology. For example, Burke and colleagues investigated the phylogenetic and functional diversity of bacterial communities on the green alga *Ulva australis* and surrounding seawater (17) (18). They demonstrated that algae-associated communities are highly distinct from the ones in the seawater and highly variable among individual algal specimen. Despite the high phylogenetic variability, the similarity in functional genes present across all algae-associated species was high (70%). These results imply selective mechanisms of bacterial assembly on algal surfaces (niche partitioning) and

high variability between algal hosts suggests random colonization (neutral processes) or effect of abiotic parameters (e.g. temperature and salinity), whose effect on microbial species composition is difficult to measure *in situ*. Likewise, a particular group of heterotrophic bacteria possessing similar polysaccharide catabolisms are recurrent with varying abundances in North Sea spring algal blooms (Fig. 5.2). I thus hypothesize that multiple bacterial clades encoding the genes to utilize a given glycan could share the same polysaccharide niche and their abundance could be regulated by abiotic forces.

5.3) Ecological importance of inter- and intra-genus diversity

In her article entitled “Diversity is the question, not the answer”, Ashley Shade emphasized that diversity without contextual information provides limited insights into the mechanisms driving microbial community composition (19). Therefore, diversity measurements should be coupled with functional comparisons to test specific hypotheses about the influence of diversity on niche composition within a complex microbial community. In this respect, taxonomic and functional analyses of representatives of the genera *Reinekea*, *Polaribacter* and *Formosa* in this thesis allows me to interpret their ecological niches in conjunction with the diversity.

In-depth phylogenetic analyses of *Reinekea*, *Polaribacter* and *Formosa* spp. revealed both clonal as well as diverse populations. The North Sea *Reinekea* clade had a decreasing abundance over years and likely was presented by a single species (Chapter 2, Fig. 2) (Chapter 2, Fig. 3). This quasi-clonal structure is also supported by the dominance of a single oligotype during North Sea spring algal blooms (Chapter 2, Fig. S1). In contrast, members of the flavobacterial genera *Formosa* and *Polaribacter* maintained abundance in relatively higher numbers and represented more diverse communities. Three major sub-clades were identified in the genus *Formosa* (11), and four in *Polaribacter* with the addition of two minor ones (Chapter 3, Fig. 3). Abiotic disturbances may affect the availability of resources and create new niche spaces, which drives replacement or shift in these communities (19). High diversity within a bacterial clade could fulfill the metabolic requirements to adapt to environmental turbulence but a clonal clade with a single species

might not be resilient to tackle these dynamic conditions. Therefore, recurrent selection of diverse *Polaribacter* and *Formosa* spp. and high oscillations in the abundance of clonal *Reinekea* clade could be explained by various levels of inter-genus microdiversity, which is supported by phenotypic plasticity.

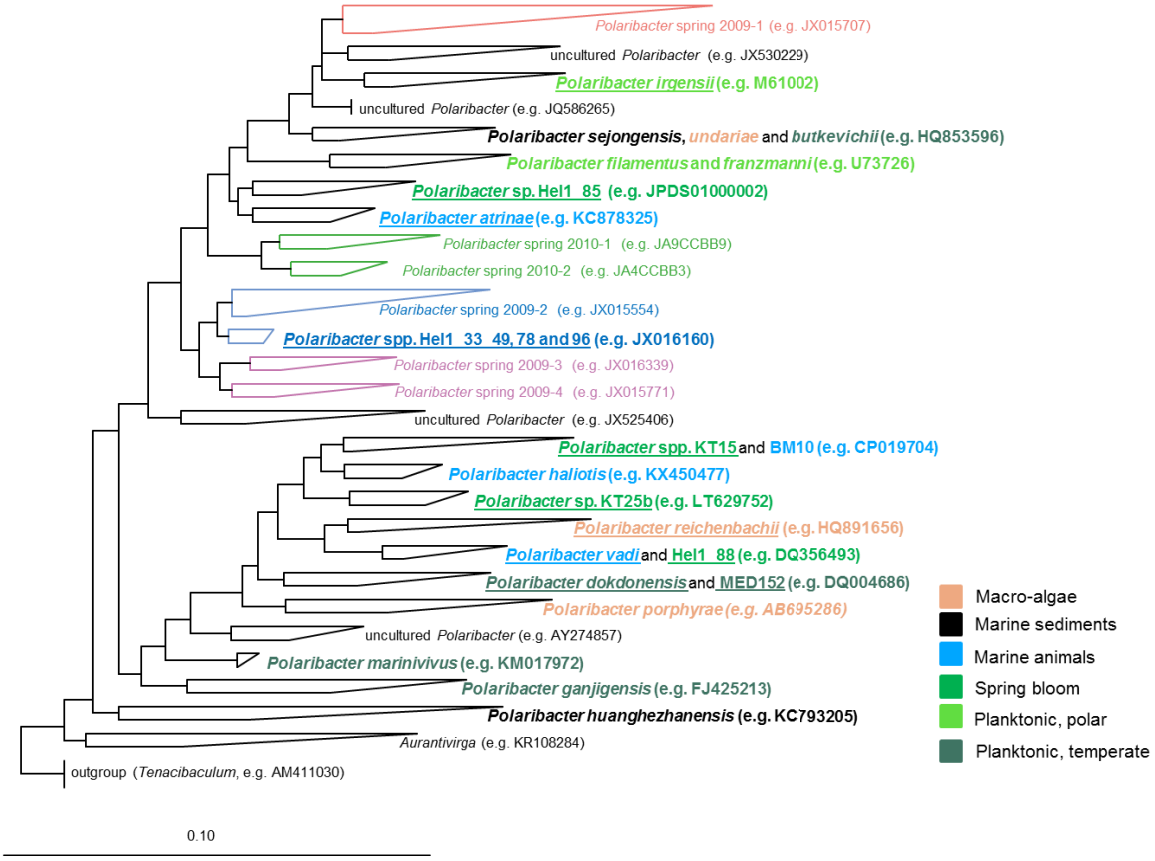


Fig. 5.3. Consensus phylogenetic tree of 16S rRNA sequences in the genus *Polaribacter*. Phylogenetic analysis was performed as described in Chapter 3. The isolation site of the described species and the North Sea isolates are depicted in different colors. 16S rRNA clone sequences from 2009 and 2010 the North Sea spring algal blooms are also shown. The species with sequenced genomes are underlined. *Tenacibaculum*-related sequences were used as outgroup. Bar: 0.1 substitutions per nucleotide position.

Phylogenetic and genomic analyses of members of the genus *Polaribacter* suggested a large geno- and phenotypic plasticity to adapt to diverse environments. 16S rRNA genes of all *Polaribacter*-related sequences in the SILVA SSURef database (v.132) were analyzed, and phylogeny of validly described species together with North Sea *Polaribacter* isolates was investigated. The results did not indicate any monophyletic group, which is related to isolation sites (Fig. 5.3). I also analyzed the polysaccharide niches of the genus *Polaribacter* by PUL analyses (Table 5.1). Macro-algae, marine

animal, and algal bloom associated *Polaribacter* spp., isolated in solid media, have a large PUL spectrum to degrade diverse polysaccharides. As “core PULs”, all of these isolates encode α-glucan, β-glucan, and alginate PULs (Table 5.1). The polysaccharide niche is further differentiated by targeting various polysaccharides. For example, *P. reichenbacchii* possesses putative PULs to utilize α-mannose-containing polysaccharides, pectin, fucose-containing polysaccharides, and agar, while *Polaribacter* sp. KT15 additionally harbors by a putative ulvan PUL (Table 5.1). Furthermore, North Sea *Polaribacter* MAGs also have a diverse PUL spectra but are devoid of some “core PULs” likely due to their lower genome completeness values. On the other hand, temperate seawater and polar isolates (e.g. *P. dokdonensis)* only have “core PULs” (Table 5.1). Therefore, these differences in PUL architecture indicate habitat-related and polysaccharide-dependent niche differentiation in the genus *Polaribacter*.

The large niche space of the environmentally relevant North Sea *Polaribacter* clade was also demonstrated using phylogenetic and (meta)genomic analyses. Six distinct *Polaribacter* sub-clades with distinct polysaccharide degradation capabilities (Fig. 3.7) were shown to be differentially abundant during the investigated bloom events (Fig. 3.3). These results suggest a polysaccharide-driven niche differentiation between North Sea *Polaribacter* clade. Such “division of labor” was recently shown in *Bacillus subtilis* cultures, which separate into different subpopulations to produce and to store acetate (20), This study highlights how closely-related bacteria are metabolically specialized to efficiently exploit environmental sources. Furthermore, Hehemann and colleagues indicated how an adaptive radiation, primarily by horizontal gene transfer, could lead a fine-scale resource partitioning in closely-related marine bacteria. They showed the ecophysiological differentiation in *Vibrionaceae* populations to degrade different forms of an algal glycan (21). These studies overall support the metabolic diversification of North Sea *Polaribacter* clade to degrade a wide spectrum of algal polysaccharides and motivate a further investigation of niche separation in the strain level.

Species	α-glucan	β-glucan	Alginate	NAG	Rhamnan	α-mannose cont. poly.	Pectin	Fucose cont. poly.	Arabino galactan	Xylan	Ulvan	Agar	Iota-carrageenan cont. poly.	Other PULS
P. reichenbachii	+	+	+			+	+	+				+		5
P. vadi	+	+	+			+	+	+	+	+				4
P. atrinae	+	+	+			+			+					4
P. KT25b	+	+	+					+	+				+	1
P. Hel1_88	+	+	+		+				+					
P. Hel1_85	+	+	+					+						
P. KT15	+	+	+								+			
Polaribacter 1-a	+	+	+	+		+			+	+				1
Polaribacter 1-b		+		+						+				
Polaribacter 2-a	+	+												
Polaribacter 2-b		+	+											1
Polaribacter 3-a	+	+		+		+			+					
Polaribacter 3-b		+				+				+				
P.dokdonensis	+	+	+											
P.irgensii		+	+											1

Table 5.1. PUL spectrum of the already sequenced genomes and algal bloom associated MAGs in the genus *Polaribacter*. NAG: N-acetyl-D-glucosamine. The number of other putative PULs are also shown in the table. The PUL annotation of North Sea isolates and MAGs are adapted from (9) and Chapter 3, respectively. PULs in the other *Polaribacter* genomes are provided in Supp. File 1.

Isolation and genomic characterization of three closely-related strains shed light on the pheno- and genotypic plasticity within the genus *Polaribacter* in great detail. *Polaribacter* spp. Hel1_33_49, Hel1_33_78, Hel1_33_96 were isolated at the same sampling site in the 2010 bloom event (22). Phylogenetic analysis revealed that these strains contain identical 16S rRNA sequences (Fig. 3.1). Average nucleotide identity (ANI) value between the genomes of Hel1_33_49 and Hel1_33_96 is 99.8%, and 98.7% for Hel_33_78 (Chapter 3, Fig. S1). Genome annotations suggested that all strains encode PULs targeting β-glucans and mannose-rich sulfated polysaccharides (Fig. 3.7). Strains Hel1_33_49 and Hel1_33_96 also contain α-glucan PULs but another closely related strain, Hel1_33_78, is devoid of such a PUL and rather possesses arabinogalactan and N-acetyl-D-glucosamine PULs (Fig. 3.7). These genomic analyses were also verified by pure culture experiments. All strains grew with laminarin, but growth with glycogen was not detected in Hel1_33_78 cultures (Richard Hahnke, personal communication). Such physiological differences were also demonstrated using a set of *Brevundimonas alba* strains (23). These strains have identical 16S rRNA gene sequences yet each strain utilized a specific combination of 59 carbon

substrates. Overall these analyses highlight the ecological importance of intra-genus microdiversity and suggest genetically determined adaptations between closely-related bacteria to occupy distinct niches.

5.4) Direct link for identity and function: fluorescence-activated cell sorting and sequencing

Recent advances in computational methods enable to reconstruct microbial genomes from metagenomes (24), yet this approach often does not yield genomes with 16S rRNA genes due to problems in the assembly of this rather conserved gene (24). For example, metagenome assembled genomes (MAGs) of *Polaribacter* clades do not contain any 16S rRNA gene, and a sequence-based link between MED nodes, CARD-FISH counts, and MAGs could not be constructed. Fluorescence-activated cell sorting (FACS) could solve this problem by enabling a direct link between the identity and function of microorganisms. Cell sorting with flow cytometry after *in situ* hybridization of 16S rRNA genes and subsequent sequencing provide a targeted approach to enrich a given microbial population and to reveal its genetic potential (Fig. 5.4). Therefore, the establishment of a taxonomy-driven cell sorting approach is particularly important in the age of omics.

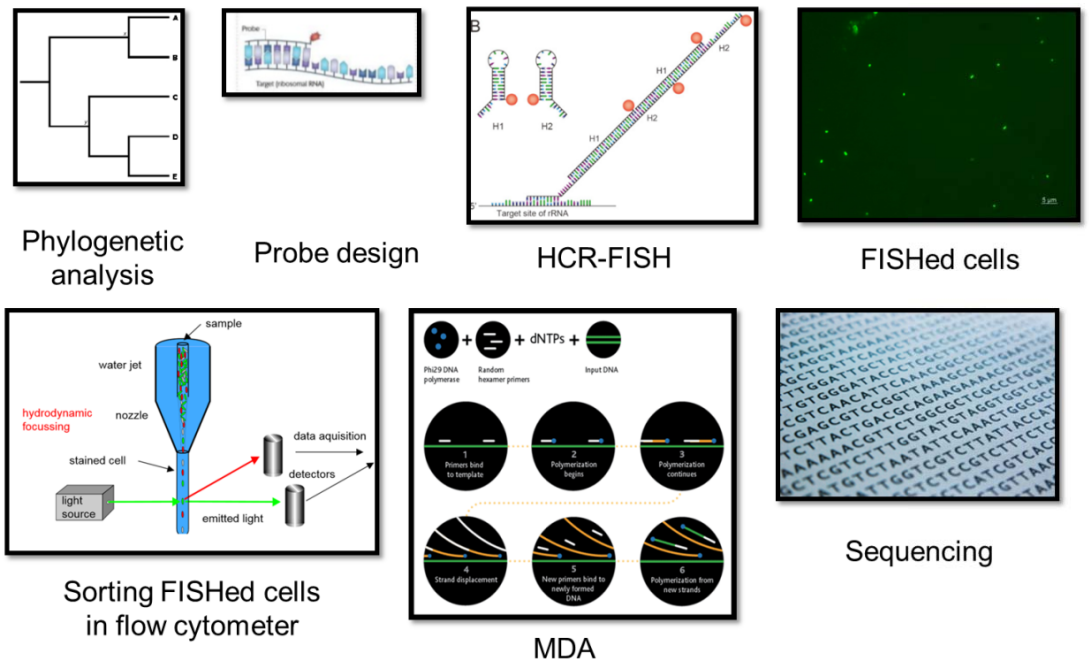


Fig. 5.4. Workflow for the fluorescence activated cell sorting (FACS) and sequencing protocol included in this thesis.

In the scope of this thesis, I also worked on a protocol for fluorescence-activated cell sorting and sequencing of *Polaribacter* clades (Fig. 5.4). Briefly, the unfixed seawater samples were filtered on a 0.2 µm polycarbonate (PC) filters and hybridization chain reaction fluorescence *in situ* hybridization (HCR-FISH) (25) was performed with the probe POL1270 to label the *Polaribacter* clade 2-a. The biomass with the hybridized cells on the PC filter was removed by sonication. This resuspension was stained with DAPI (4′,6-diamidino-2-phenylindole) and used for flow cytometric analysis. Since DAPI-positive cells did not have sufficient brightness, the sorting gate was selected with the green fluorescent (530 ± 20 nm) and forward scatter light (FSC) and compared with the negative control (NON338) to discriminate background noise. A certain number of events were sorted with a purity of >90%. The genetic material of the sorted cells was obtained using multiple displacement amplification (MDA) using 1,500 events. The 16S rRNA gene of amplified genomes was acquired with polymerase chain reaction (PCR) and afterwards sequenced. The 16S rRNA sequence similarity between the sorted cells and the *Polaribacter*-related clones obtained during the 2010 spring bloom was >99%, suggesting the successful genomic enrichment of the targeted *Polaribacter* clade. The sorted cells were sent to the Joint Genome Institute (JGI) for sequencing, and the results will be available in the near future. A detailed protocol is provided in the appendix. Overall, with further modifications, this 16S rRNA-centric approach has a promising potential for linking the identity and function of specific microbial clades in complex environments.

5.5) Algal polysaccharides and heterotrophic bacteria: the food connection

Energy harvesting is a central component of cellular metabolism and represents a dominant factor determining the proliferation of microorganisms. For example, Leigh Monahan and Elizabeth Harry demonstrated the importance of nutrient availability for bacterial survival in the wild and stated that “You are what you eat” (26). Therefore, as the main energy source of heterotrophic bacteria, diverse polysaccharides released during spring algal blooms are most likely the key force driving the niche differentiation between diverse *Polaribacter* clades.

The relationship between the specific substrate and a particular bacterial clade should be demonstrated to verify the hypothesis generated from omics data. Yet traditional methods do not yield structural information and cannot be used for *in situ* quantification of environmentally relevant algal polysaccharides, and thus the marine polysaccharide composition remains largely unknown. Our colleagues from the Marine Glycobiology Group at the MPI Bremen are developing new methods to overcome this problem. For example, Stefan Becker used CAZymes purified from marine bacteria for the selective digestion of a given polysaccharide and proposed laminarin as an abundant algal glycan in the world oceans (Stefan Becker, personal communication). Likewise, Silvia Vidal-Melgosa applied specific monoclonal antibodies to detect and to characterize the structures of polysaccharides during a spring algal bloom off Helgoland and demonstrated the fluctuations in the abundance of diverse algal glycans such as laminarin, mannan, and xylan (Silvia Vidal-Melgosa, personal communication). These results support the view that algal polysaccharides are successively available to heterotrophic bacteria during spring blooms.

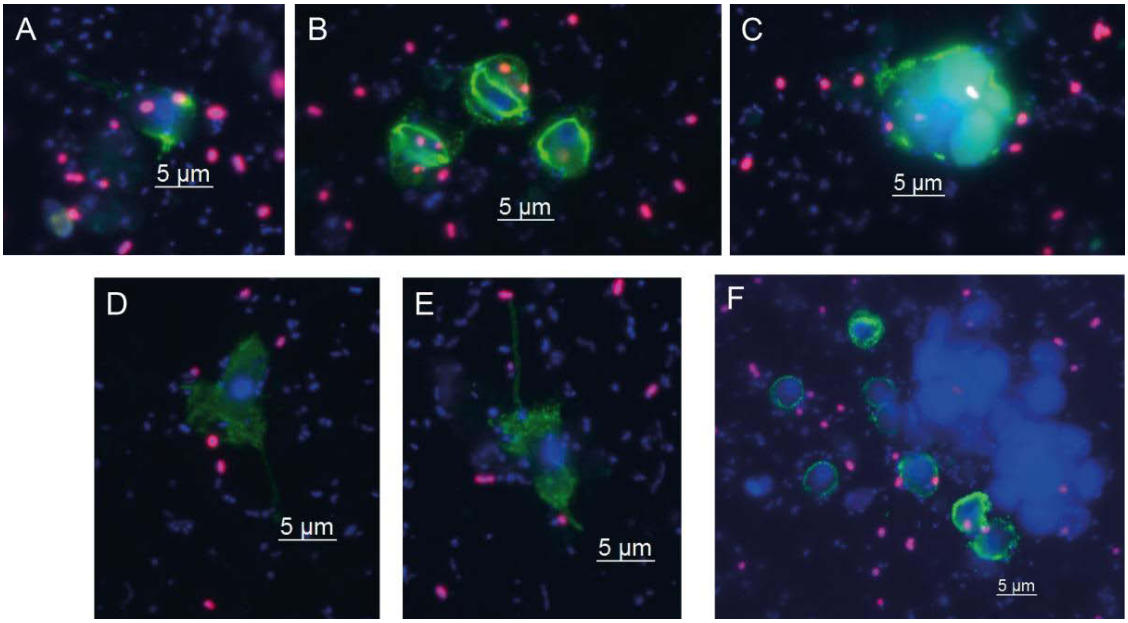


Fig. 5.5. Polysaccharide immuno-labeling and CARD-FISH to visualize xylan (A, B, C, F) and arabinogalactan (D and E) together with *Polaribacter* clade 1-a (probe POL405). FISH signals are red, while green signals belong to the above-mentioned polysaccharides. CARD-FISH was performed as stated in Chapter 3. Seawater samples were filtered on 0.2 µm polycarbonate filters without any pre-filtration step on 20120426 (xylan) and 20110519 (arabinogalactan). Polysaccharide immune-labeling protocol is provided in the Supp. File 2. Scale bar is shown for each image.

I also used the monoclonal antibodies in conjunction with CARD-FISH to investigate the relationship between arabinogalactan and xylan and the clade *Polaribacter* 1-a (Fig. 5.5), which encodes putative PULs to utilize both compounds (Chapter 3, Fig. 7). After staining with clade-specific probe POL405, polysaccharides immuno-labeling with the antibodies LM11 (specific to 1,4-linked xylan (27)) and LM13 (specific to 1,5-linked arabinan (28)) was performed. The xylan antibody stained diverse microalgae-like morphologies, while arabinogalactan was detected only in a microalga structure with a flagellum (Fig. 5.5). These stainings were sometimes adjacent to POL405 signals (Fig. 5.5), indicating (i) the presence of xylan and arabinogalactan during bloom events and (ii) potential binding of POL405 positive cells to algal polysaccharides. Yet any specific relationship between POL405 and these polysaccharides could not be established. There were also many free-living cells without any association with the labeled polysaccharides. Future experiments with fluorescently labeled polysaccharides (29) are needed to visualize and to quantify the uptake by heterotrophic bacteria with a reliable sampling strategy.

5.6) Effect of other environmental parameters on the bacterial niches

Polaribacter clade	**Temperature** (C^{o})	**Salinity** ppm	**Silicate** µmol/l	**Phosphate** µmol/l	**Nitrite** µmol/l	**Nitrate** µmol/l	**Ammonium** µmol/l
1-a	0.58	0.2	-0.55	-0.58	-0.55	-0.66	-0.24
1-b	0.46	0.23	-0.42	-0.52	-0.63	-0.64	-0.07
2-a	0.2	-0.19	-0.8	-0.62	-0.17	-0.26	-0.45
2-b	0.12	-0.13	-0.7	-0.51	-0.31	-0.25	-0.42
3-a	0	0.1	-0.72	-0.52	-0.16	-0.22	-0.36
3-b	0.44	0.01	-0.62	-0.5	-0.72	-0.63	-0.12

Table 5.2. Spearman correlation values between North Sea *Polaribacter* clades and the physiochemical parameters.

The relationship between the abundance of different *Polaribacter* clades and physiochemical parameters such as temperature, salinity, silicate, phosphate, nitrite, nitrate and ammonium (4) was assessed using the Spearman correlation test. A moderate positive correlation between temperature and *Polaribacter* 1-a, 1-b and 3-b was found, yet such a connection was not

detected for salinity (Table 5.2). This moderate relationship could be characteristic of spring seasons due to the increasing seawater temperature and elevated bacterial abundance during algal blooms. On the other hand, silicate and phosphate were negatively correlated with all clades. Likewise, a negative relationship was also observed for nitrite, nitrate, and ammonium (Table 5.2). Such negative correlations between all clades and nutrients are typical since these compounds are taken up by phytoplankton and heterotrophic bacteria and detected in low concentrations during the blooms. Overall these results did not demonstrate any significant difference among *Polaribacter* clades concerning abiotic factors. Therefore, I hypothesize that measured physiochemical parameters are not likely the major driving force for a niche differentiation between diverse *Polaribacter* clades during North Sea spring algal blooms.

5.7) *Aurantivirga*: a recurrent key player with a new name

Phylogenetic analysis of North Sea *Polaribacter* spp. also led to a revision of the taxonomy of other *Flavobacteriia,* which were abundant during the investigated North Sea spring algal blooms. Taxonomy of some 16S rRNA clones (Supp. File 3) obtained during the bloom events has been updated in collaboration with members of the SILVA team. These sequences were classified as *Polaribacter* 4 in the previous version of the SILVA taxonomy and now have been re-named as *Aurantivirga* based on monophyletic branching with high sequence similarity (>95%) to the type species of this genus (Fig. 5.3). *Aurantivirga* is a recently described flavobacterial genus with the type species *Aurantivirga profunda*, which was isolated from deep seawater of the Pacific Ocean (30). This species is a rod-shaped, proteorhodopsin-containing, and aerobic chemoheterotrophic bacterium similar to other *Flavobacteriia* yet does not exhibit gliding motility.

The genus *Aurantivirga* is also recurrently abundant during the studied North Sea spring algal blooms. The genus-specific oligonucleotide probe AUR452 (Supp. File 4) detected *Aurantivirga* abundances up to 6%, 13%, 5%, and 3% from 2009 to 2012 spring algal blooms off Helgoland (Fig. 5.6). This abundance pattern was also corroborated by MED analysis of 16S rRNA tag sequences (11) and read recruitment on metagenome assembled genomes

(MAGs) of the *Aurantivirga* clade (Supp. File 5) (Fig. 5.6). A first glimpse on the genetic repertoire of this clade revealed a putative laminarin PUL, proteorhodopsin, and gliding motility-related genes (Supp. File 6). Strain Hel1_33_7, which was isolated during the 2010 bloom (22), is representative of the North Sea *Aurantivirga* clade based on 16S rRNA gene comparisons (accession number: KF023505). However, the Hel1_33_7 cultures are not growing in the current medium, and thus sufficient biomass for genome sequencing and physiological experiments have not been obtained. Therefore, *Aurantivirga* represents a potential target for future cultivation studies and MAG analyses.

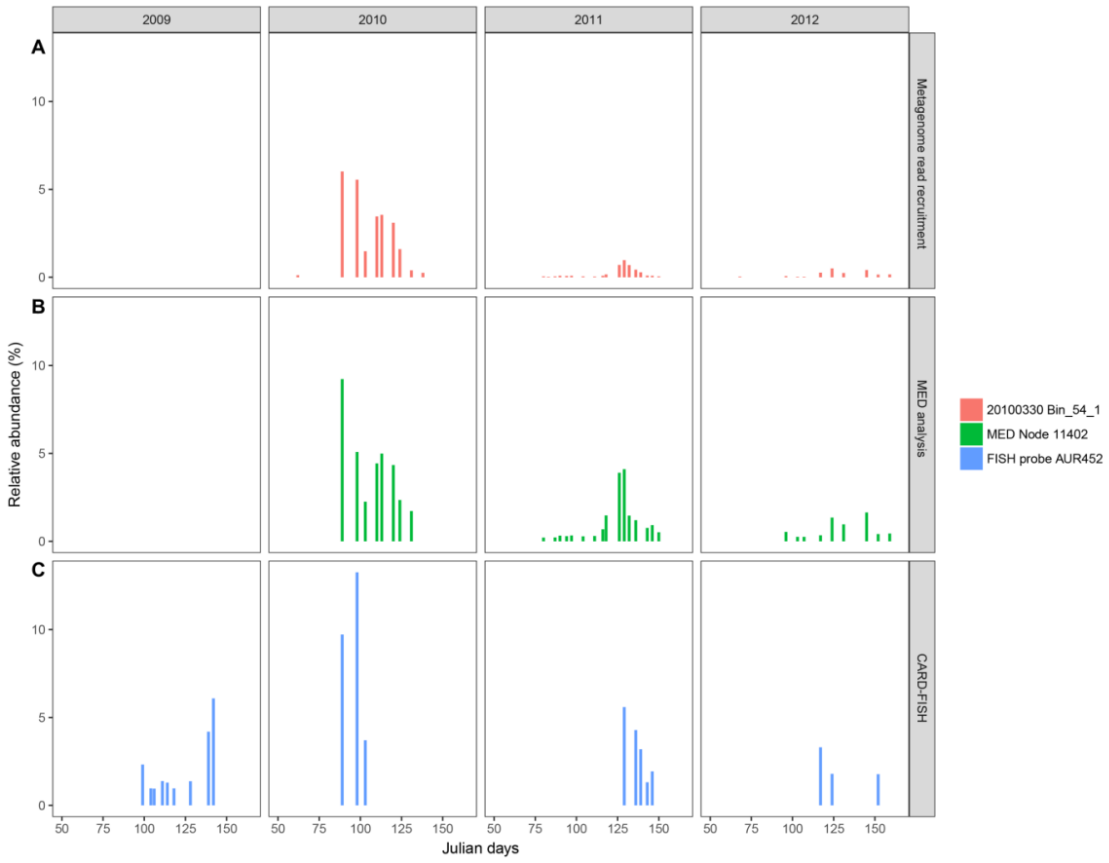


Fig. 5.6. Relative abundance and recurrence of *Aurantivirga* clade during the North Sea spring algal blooms as revealed by (A) metagenome read recruitment, (B) MED analysis of 16S rRNA tag sequences, and (C) CARD-FISH.

5.8) Final remarks and outlook

In this thesis, I investigated the niches of selected heterotrophic bacteria that are recurrently abundant during North Sea spring algal blooms. Genomic and physiological analyses of the genera *Reinekea*, *Polaribacter,* and *Formosa*

demonstrated that polysaccharide utilization is the central niche component of studied bloom-associated bacteria and shapes the microbial community composition during algal blooms.

Collaborative efforts should be continued in the future to holistically understand the ecological and evolutionary mechanisms driving the microbial community structure and function during algal blooms. Based on the results of my thesis, I am proposing the following studies:

- Sampling at the Kabeltonne should continue to further detect the recurrence and functional replacement of diverse *Flavobacteriia* and *Gammaproteobacteria* in future blooms.
- Binning of the time-series metagenomes should be extended to other abundant heterotrophic bacteria to holistically reveal the genomic potential of these bacteria during the bloom events. Furthermore, high-resolution metatranscriptomics and metaproteomics are needed to investigate *in situ* gene expression.
- Targeted cultivation efforts should be made to test the hypotheses generated from omics data. This will enable to determine the substrate-specificity of putative PULs proposed in this thesis. This proven polysaccharide utilization capacity will also guide the design of media with relevant algal substrates that are suited to isolate the representatives of abundant heterotrophic bacteria during bloom events. For example, *Chattonella* cell extracts could enable the successful isolation of the *Polaribacter* clade 1-a, which would also shed light on the probable algal-bacterial interaction suggested in my thesis.
- Vitamin addition to currently used media could potentially enable the growth of strain Hel1_33_7. The revival of this culture will reveal the ecophysiology of an abundant *Aurantivirga* clade during North Sea spring algal blooms.
- Brighter FISH and DAPI signals are needed for a better FACS protocol. Counterstaining with TO-PRO and additional dyes to the connector probes used in HCR FISH would improve the signal brightness. The genome amplification method could be also replaced. For example, the

Illumina Nextera DNA library preparation kit would fully utilize the genetic material from sorted cells without any amplification step. Such approaches would yield targeted sorting and sequencing of specific bacterioplankton clades.

- The sampling strategy and immuno-labeling protocol should be optimized for the reliable staining of algal polysaccharides and associated bacteria. A new sampling scheme, for example, filtration with 3 µm and 10 µm pore-size filters, could be applied for a better spatial resolution. Furthermore, the staining protocol should be optimized to minimize cell loss during washing steps. These efforts would help to elucidate a specific relationship between a given algal polysaccharide and a specific bacterial clade.
- Fluorescently labeled polysaccharide incubations should be conducted to quantify and to visualize the polysaccharide uptake by particular heterotrophic bacteria. In combination with HCR-FISH, these bacteria could be also sorted by flow cytometry.
- A correlative microscopy approach to simultaneously detect the identity, function, and substrate is essential. Polysaccharide immune-labeling or mass spectrometry imaging (31) would complement the identity-function relationship revealed by geneFISH and holistically visualize the “the big picture in one picture”.

5.9) Supplementary information

Supplementary information is available in the CD-ROM, which is provided with the thesis

5.10) References

1. Eiler A. Evidence for the ubiquity of mixotrophic bacteria in upper ocean: implications and consequences. Appl Environ Microb. 2006;72(12):7431-7.

2. Stoecker DK, Hansen PJ, Caron DA, Mitra A. Mixotrophy in the marine plankton. Annu Rev Mar Sci. 2017;9(1):311-35.

3. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-controlled succession of marine bacterioplankton

populations induced by a phytoplankton bloom. Science. 2012;336(6081):608-11.

4. Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L, et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. eLIFE. 2016;5:e11888.

5. Kolowith LC, Ingall ED, Benner R. Composition and cycling of marine organic phosphorus. Limnol Oceanogr. 2001;46(2):309-20.

6. Seymour JR, Amin SA, Raina JB, Stocker R. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. Nat Microbiol. 2017;2:17065.

7. Dalisay DS, Webb JS, Scheffel A, Svenson C, James S, Holmström C, et al. A mannose-sensitive haemagglutinin (MSHA)-like pilus promotes attachment of *Pseudoalteromonas tunicata* cells to the surface of the green alga *Ulva australis*. Microbiology. 2006;152(10):2875-83.

8. Kirchman DL. The ecology of *Cytophaga–Flavobacteria* in aquatic environments. FEMS Microbiol Ecol. 2002;39(2):91-100.

9. Kappelmann L, Krüger K, Hehemann J-H, Harder J, Markert S, Unfried F, et al. Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. ISME J. 2018;doi:10.1038/s41396-018-0242-6.

10. Wilson WA, Roach PJ, Montero M, Baroja-Fernández E, Muñoz FJ, Eydallin G, et al. Regulation of glycogen metabolism in yeast and bacteria. FEMS Microbiol Rev. 2010;34(6):952-85.

11. Chafee M, Fernandez-Guerra A, Buttigieg PL, Gerdts G, Eren AM, Teeling H, et al. Recurrent patterns of microdiversity in a temperate coastal marine environment. ISME J. 2018;12(1):237-52.

12. Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH. Relative roles of niche and neutral processes in structuring a soil microbial community. ISME J. 2010;4(8):337-45.

13. Rivett DW, Bell T. Abundance determines the functional role of bacterial phylotypes in complex communities. Nat Microbiol. 2018;3(7):767-72.

14. Araki T, Kitamikado M. Purification and characterization of a novel exo-β-mannanase from *Aeromonas* sp. F-25. J Biochem. 1982;91(4):1181-6.

15. Chiovitti A, Harper RE, Willis A, Bacic A, Mulvaney P, Wetherbee R. Variations in the substituted 3-linked mannans closely associated with the silicified walls of diatoms. J Phycol. 2005;41(6):1154-61.

16. Sale PF. Reef Fish Lottery. Nat Hist. 1976;85:60-5.

17 Burke C, Thomas T, Lewis M, Steinberg P, Kjelleberg S. Composition, uniqueness and variability of the epiphytic bacterial community of the green alga *Ulva australis*. ISME J. 2011;5(4):590-600.

18. Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T. Bacterial community assembly based on functional genes rather than species. Proc Natl Acad Sci USA. 2011;108(34):14288-93.

19. Shade A. Diversity is the question, not the answer. ISME J. 2017;11:1-6.

20. Rosenthal AZ, Qi Y, Hormoz S, Park J, Li SHJ, Elowitz MB. Metabolic interactions between dynamic bacterial subpopulations. eLIFE. 2018;7:e33099.

21. Hehemann J-H, Arevalo P, Datta MS, Yu X, Corzett CH, Henschel A, et al. Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. Nat Commun. 2016;7:12860.

22. Hahnke RL, Bennke CM, Fuchs BM, Mann AJ, Rhiel E, Teeling H, et al. Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. Environ Microbiol. 2015;17(10):3515-26.

23. Jaspers E, Overmann J. Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. Appl Environ Microb. 2004;70(8):4831-9.

24. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2(11):1533-42.

25. Yamaguchi T, Kawakami S, Hatamoto M, Imachi H, Takahashi M, Araki N, et al. In situ DNA-hybridization chain reaction (HCR): a facilitated in situ HCR system for the detection of environmental microorganisms. Environ Microbiol. 2015;17(7):2532-41.

26. Monahan LG, Harry EJ. You are what you eat: metabolic control of bacterial division. Trends Microbiol. 2016;24(3):181-9.

27. McCartney L, Marcus SE, Knox JP. Monoclonal antibodies to plant Cell wall xylans and arabinoxylans. J Histochem Cytochem. 2005;53(4):543-6.

28. Verhertbruggen Y, Marcus SE, Haeger A, Verhoef R, Schols HA, McMleary BV, et al. Developmental complexity of arabinan polysaccharides and their processing in plant cell walls. Plant J. 2009;59(3):413-25.

29. Reintjes G, Arnosti C, Fuchs BM, Amann R. An alternative polysaccharide uptake mechanism of marine bacteria. ISME J. 2017;11(7):1640-50.

30. Song J, Choi A, Im M, Joung Y, Yoshizawa S, Cho J-C, et al. *Aurantivirga profunda* gen. nov., sp. nov., isolated from deep-seawater, a novel member of the family *Flavobacteriaceae*. Int J Syst Evol Micr. 2015;65(12):4850-6.

31. Kaltenpoth M, Strupat K, Svatoš A. Linking metabolite production to taxonomic identity in environmental samples by (MA)LDI-FISH. ISME J. 2016;10(2):527-31.

Appendix A

Fluorescence-activated sorting of *Polaribacter* 2-a cells from an environmental sample

A.1) Aim and scope

This appendix aims to explain the enrichment of *Polaribacter* 2-a cells from an environmental sample with a fluorescent-activated cell sorting (FACS) approach, which includes:

(i) Hybridization chain reaction fluorescence *in situ* hybridization (HCR-FISH) to target the cells in *Polaribacter* 2-a using the probe POL1270.
(ii) FACS to separate labeled cells in an environmental sample.
(iii) Multiple displacement amplification (MDA) to get the genomic content of sorted cells.
(iv) Amplification and sequencing of the 16S rRNA gene using polymerase chain reaction (PCR) to identify the phylogenetic affiliation of MDA products.

A.2) Sampling

A seawater sample was taken from Kabeltonne, Helgoland (54°11.3′ N, 7°54.0′ E) on April 8, 2010. 10 L of water sample was filtered onto on polyethersulfone (PES) filter (0.2 µm pore size, 142 mm diameter) without using any fixatives and was frozen at -80 °C. The biomass on the PES filter was removed with a cell scraper and suspended in 10 ml 1x PBS on ice. Subsequently, this resuspension was filtered on the 0.2 µm pore size polycarbonate (PC) filter, which was kept in the freezer until the hybridization step.

A.3) Hybridization chain reaction fluorescence *in situ* hybridization (HCR-FISH)

The following protocol was applied for *in situ* hybridization of 16S rRNA genes:

Hybridization with initiator probe and washing

- The hybridization solution was prepared using a buffer with 20% formamide (Table S1) and the initiator probe POL1270 together with helper and competitor probes (Table S2). Final probe concentration is 1 µM.
- Filter sections were dipped into a hybridization solution and put on a petri dish. Some extra hybridization solution was also dispensed on the filter sections.

- A blotting paper was soaked with 6 ml of 20% formamide/water mix, and place in the hybridization chamber.
- Filter sections were put in the hybridization chamber and incubated at 46 °C for 2 hours.
- After hybridization, filter sections were transferred to 50 ml pre-warmed washing buffer (1 ml Tris-HCl, 0.5 ml 0.5 M EDTA, 2.15 ml 5 M NaCl, 0.05 ml 10% SDS, 46.3 ml distilled water) and incubated at 48 °C for 30 minutes.

Second hybridization

- For the second hybridization, one volume of amplifier probes H1 and H2 (4xAtto488, 10 µM) were mixed with one volume of second hybridization buffer (50 mM Na_2HPO_4, 0.9 M NaCl, 0.01% SDS, 1% blocking reagent, 10% dextran sulfate) separately.
- H1 and H2 solutions were incubated at 95 °C for 1.5 min and at 25 °C for 1 min and kept at room temperature until needed.
- A blotting paper was soaked with 6 ml distilled water mix and placed in the hybridization chamber.
- Filter sections were taken from washing buffer and were immediately proceed to the next step.
- H1 and H2 solutions were pooled and immediately spread on the filter sections (final concentration of H1 and H2: 2.5 µM). Some extra hybridization solution was also dispensed on the filter sections.
- Filter sections were put in the hybridization chamber and incubated at 37 °C for 30 minutes.

Second washing

- After the second hybridization, filter sections were transferred to 1x PBS solution and incubated at 4 °C for 5 min.
- The filter sections were transferred to another 1x PBS solution and the same procedure was repeated.
- The sections were incubated in distilled water at 4 °C for 30 seconds.
- The filter sections were washed in 96% EtOH at room temperature for 30 seconds and completely dried on a Whatman paper.

Counterstaining and microscopy

- For counterstaining, small parts were cut from the filter sections and incubated in 10 µl of DAPI solution (1 µg/ml) for 3 minutes.
- Afterwards, filters were washed in distilled water and 80% EtOH solution for 30 seconds.
- The filter sections were dried completely and mounted in Citifluor solution.
- Signals were examined in a fluorescence microscope. A microscopy image showing POL1270 signals is provided in Fig. 1.

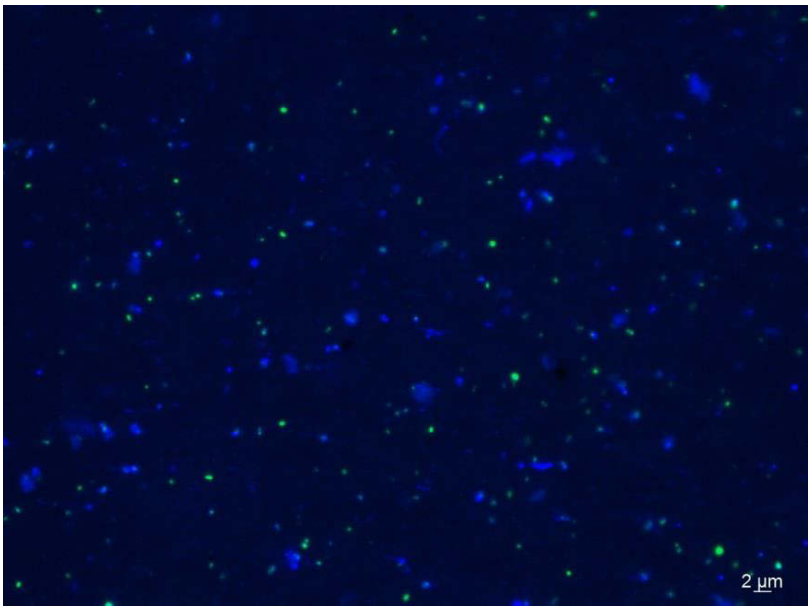


Fig. 1. Microscopy image showing POL1270 (green) and DAPI (blue) signals.

A.4) Cell detachment from the filters

Sonication was used to detach the cells from the filter sections.

- The filter section was cut in four small pieces (ca. 1 cm^2) and put in 500 µl of 1x PBS.
- This suspension was sonicated for 45 seconds (power: 70 watts, pulse: 20%) using a Bandelin Sonoplus sonicator.
- Afterwards, the filter sections were removed from 1x PBS and a filter section was stained with DAPI to visualize the residual cells on the filter.
- 10 µl of the solution was spotted on 0.2 µm pore size polycarbonate filter and stained with DAPI (Fig. 2).

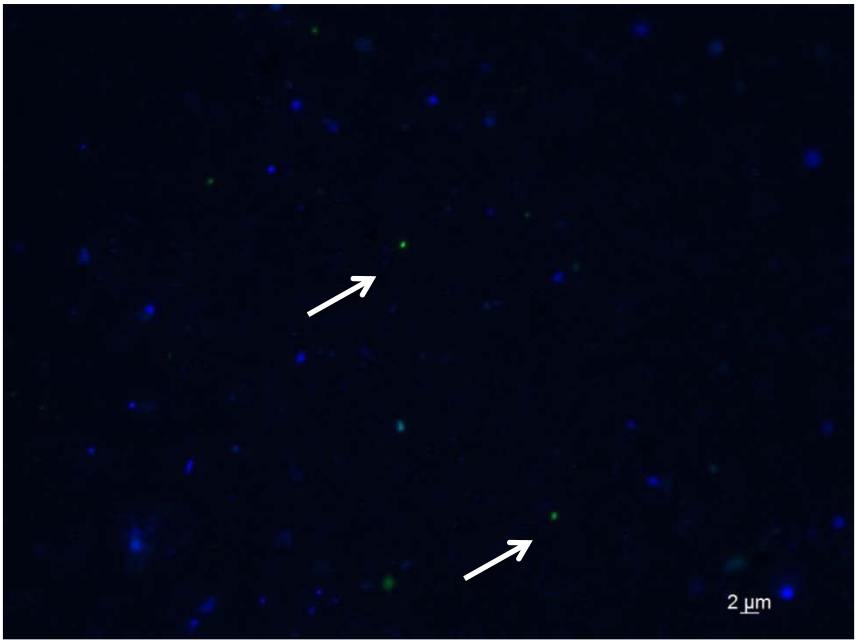


Fig. 2. Microscopy images showing detached DAPI (blue) POL1270 (green) signals from the filter sections using the vortex treatment.

A.5) Fluorescence-activated cell sorting (FACS)

After removal of hybridized cells from the filters, all cell preparations were counterstained with DAPI (1 µg / ml). After the sterilization with 0.15% NaCl solution, cell sorting was done with an Influx flow cytometer (BD Biosciences). Green fluorescence from hybridized cells was detected by using a 530 ± 20 nm band-pass filter while DAPI fluorescence was measured with a 460 ± 25 nm band-pass filter. The system threshold was set in forward scatter light (FSC). Prior to measurements, the instrument was aligned by using fluorescent particles. In the density diagram of FSC versus DAPI, some DAPI-positive cells were gated and plotted. However, DAPI signals with sufficient probe brightness were not observed. Therefore, the sorting was performed using FSC versus green fluorescence using NON338 probe a negative control to determine the background (Fig. 3). For subsequent analyses, a defined number of events (Table 1) was sorted from the specific populations from four samples (Fig. 3).

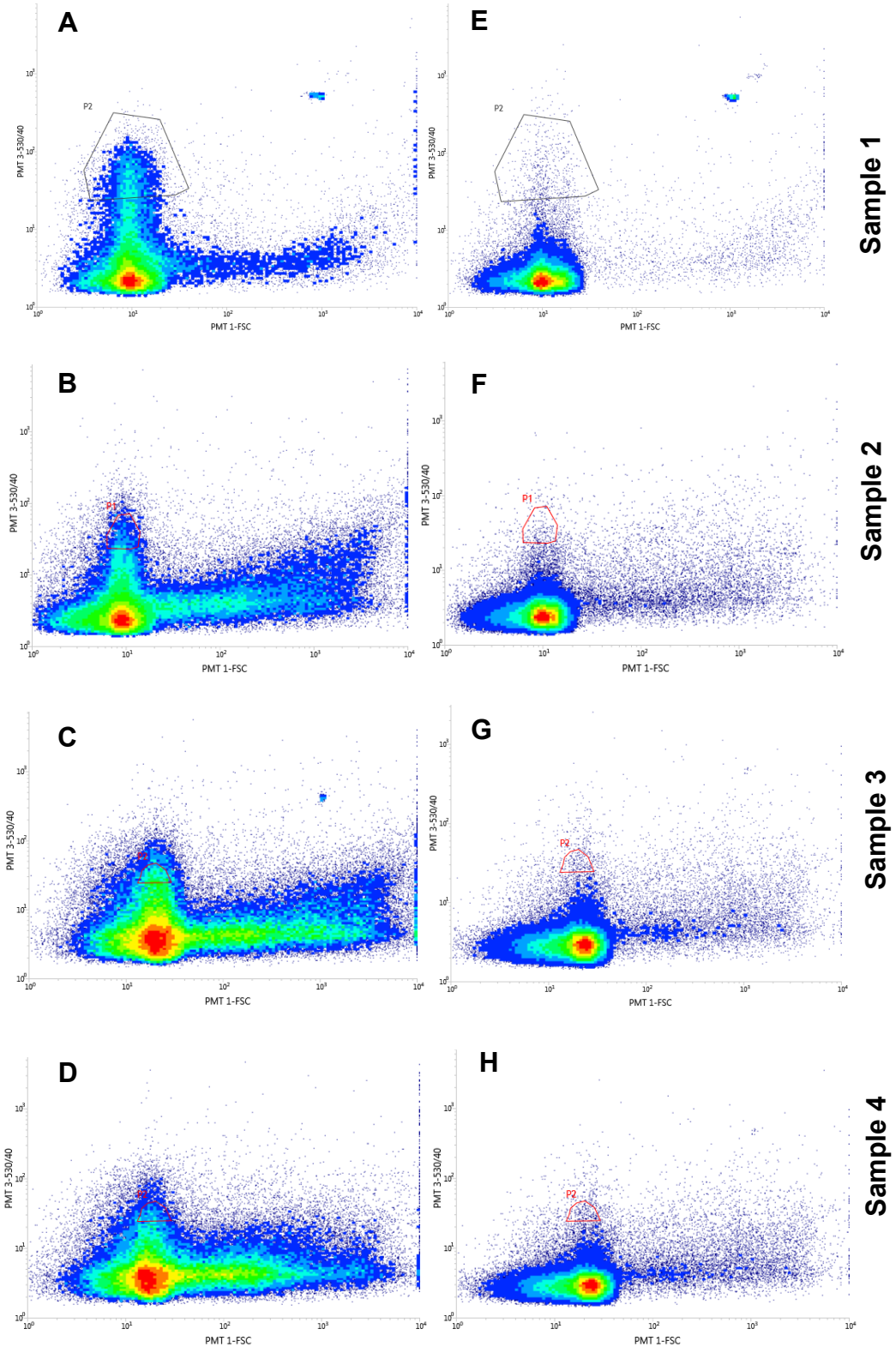


Fig. 3. Density plots of green fluorescence (PMT 3) vs forward scatter light (FSC) POL1270 (A,B,C,D) and NON338 (E,F,G,H). Sorting gates are specified in each sample. 100,000 (A and E) and 50,000 (B, F, C, G, D, H) events were recorded to construct the plots.

Probe	Sample ID	Sorting ID	Event number	ΔP	Event rate (1/s)	Sort Rate (1/s)
POL1270	Sample 1	1-1	1,000	0.7	1000	10
POL1270	Sample 1	1-2	1,000	0.7	1000	10
POL1270	Sample 1	1-3	5,000	0.7	1000	10
POL1270	Sample 2	2-1	1,000	0.7	1000	15
POL1270	Sample 2	2-2	5,000	0.7	1000	15
POL1270	Sample 2	2-3	5,000	0.7	1000	15
POL1270	Sample 3	3-1	5,000	0.7	600	10
POL1270	Sample 4	4-1	1,000	0.7	1200	15
POL1270	Sample 4	4-2	1,000	0.7	1200	15
POL1270	Sample 4	4-3	1,000	0.7	1200	15
POL1270	Sample 4	4-4	5,000	0.7	1200	15
POl1270	Sample 4	4-5	5,000	0.7	1200	15

Table 1. Parameters and event numbers of sorted POL1270 cells from four samples.

To check the purity of sorted cells, 2 µl aliquot from each sorting gate were spotted on 0.2 µm pore size polycarbonate filter and stained with DAPI (Fig.4). The purity for the sorted cells was ≥ 90%.

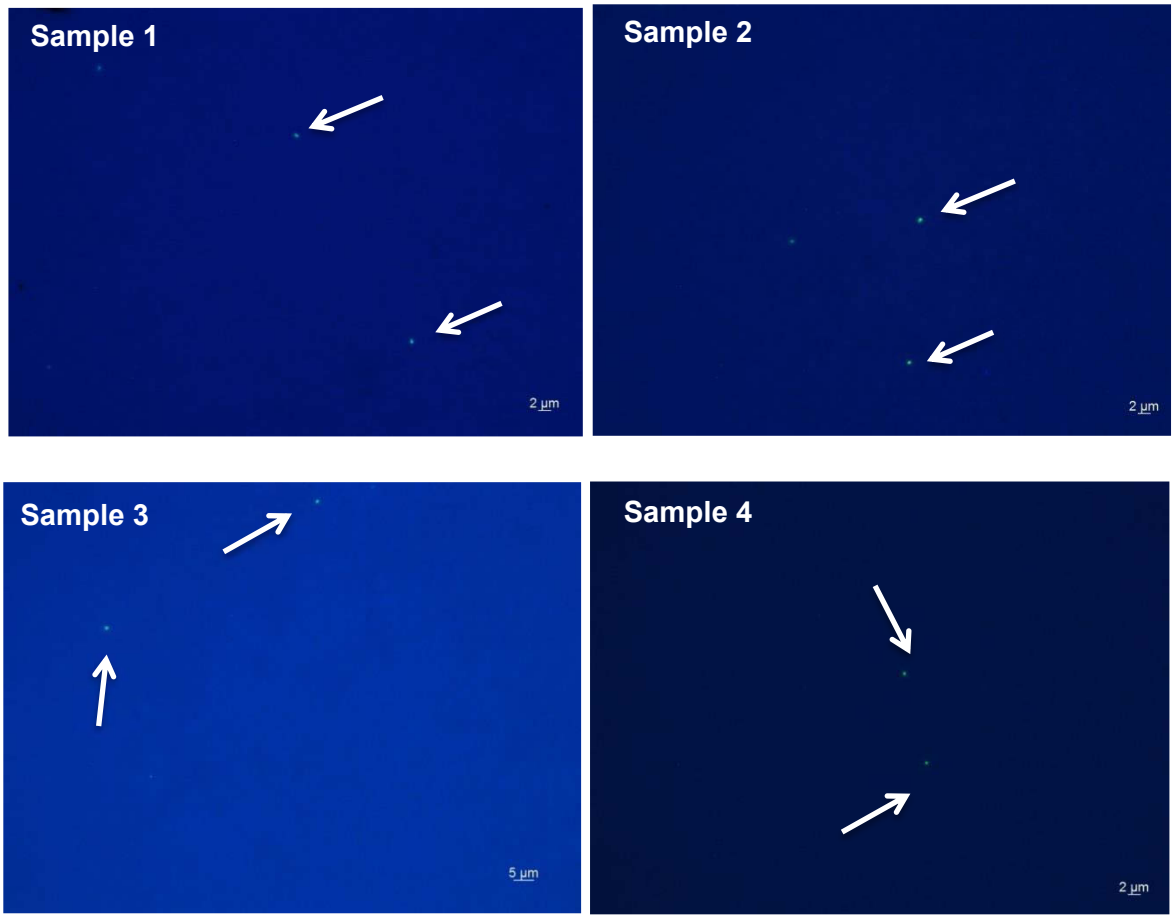


Fig. 4. Microscopic inspection of sorted cells from four samples. Sorted POL1270 cells are shown with arrows.

A.6) Multiple Displacement Amplification (MDA)

Genomic content of the sorted cells was amplified using the Illustra GenomiPhi V3 DNA Amplification Kit (GE Healthcare). The reaction was set in a UV-treated PCR workstation with a dedicated pipette set to avoid any contamination. A bacterial culture (HaHa 1_5) and was used as positive controls, while 3 µl of PCR water and sheath fluid were added in the negative control. The following protocol was applied:

- The samples were exposed to freeze/thaw cycles (2 min in room temperature/ 2 min -20 °C, 3 times).
- 3 µl of the sample (1,500 events) was mixed with 3 µl chemical denaturation solution (400 mM KOH, 10 mM EDTA) and incubated at room temperature for 3 minutes.
- 3 µl of neutralization buffer (400 mM HCl, 600 mM Tris-HCl, pH= 7.5) was added.
- 1 µl of PCR water was added to complete the total volume to 10 µl.
- 10 µl of 150 mM KCl solution was added.
- A total of 20 µl reaction volume was added to ready-to-go GenomiPhi V3 cake contains all the components required for DNA amplification.
- The samples were incubated in the therm cycler using the following program:

  30 °C for 180 minutes
  65 °C for 10 minutes
  Then cool to 4 °C
- Quality and quantity of MDA product were checked using gel electrophoresis. 3 µl of loading dye was mixed with 3 µl of the sample and loaded in 0.8% agarose gel. The gel was exposed to 75 V electric current for 1 h and stained in 1% ethidium bromide bath for 30 min (Fig. 5).

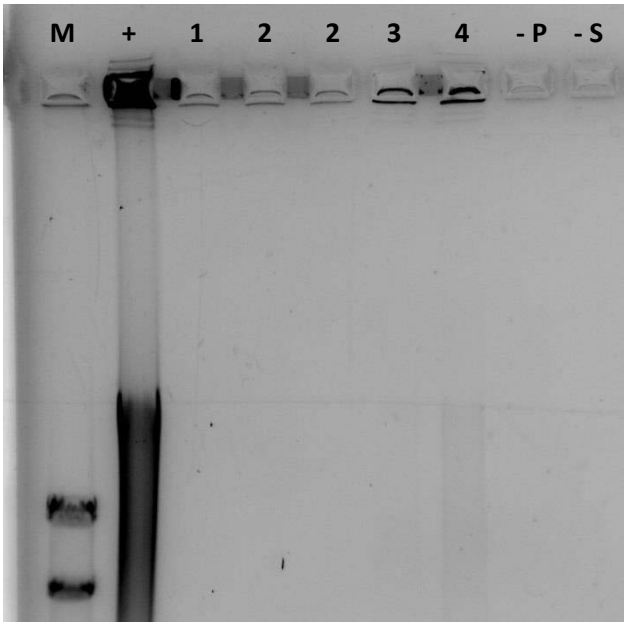


Fig. 5. Agarose gel electrophoresis image of MDA products. M: λ-Hind III marker (size of bands: 23130 bp, 9416 bp); (+): positive control: bacterial culture HaHa 1_5; (-P): negative control; PCR water (-S) negative control sheath fluid. Numbers refer to sample IDs (Table 1).

A.7) Polymerase chain reaction (PCR)

16S rRNA gene was amplified using the polymerase chain reaction (PCR) to identify the phylogenetic affiliation of MDA products. *E.coli* gDNA and PCR water were used as positive and negative controls, respectively. GM3Fand GM4R primer set (Table S3) was selected to amplify the 16S rRNA gene. The following protocol was applied:

- PCR master mix was prepared for 10 reactions (total volume 190 µl):

20 µl BSA (3 mg/ml)
20 µl PCR buffer (10x)
16 µl dNTPs (4x 2.5 mM)
2 µl GM3 (50 µM)
2 µl GM4 (50 µM)
0.4 µl Taq polymerase (5 U/µl)
129.6 µl PCR water

- 19 µl of PCR master mix was mixed with 1 µl of the sample.
- PCR reactions were incubated in the thermocycler using the following program:

94°C	5 min	
94°C	1 min	30x
48°C	1 min	
72°C	2 min	
72°C	10 min	
12°C	hold	

- Quality and quantity of PCR products were checked using gel electrophoresis. 2 µl of loading dye was mixed with 3 µl of the sample and loaded in 1 % agarose gel. The gel was exposed to 100 V electric current for 45 h and stained in 1% ethidium bromide bath for 30 min (Fig. 6).
- The PCR products were purified with the QIAquick PCR purification kit (Qiagen) using the manufacturer's instructions. In the final step, 30µl of the elution buffer was used instead of 50 µl.

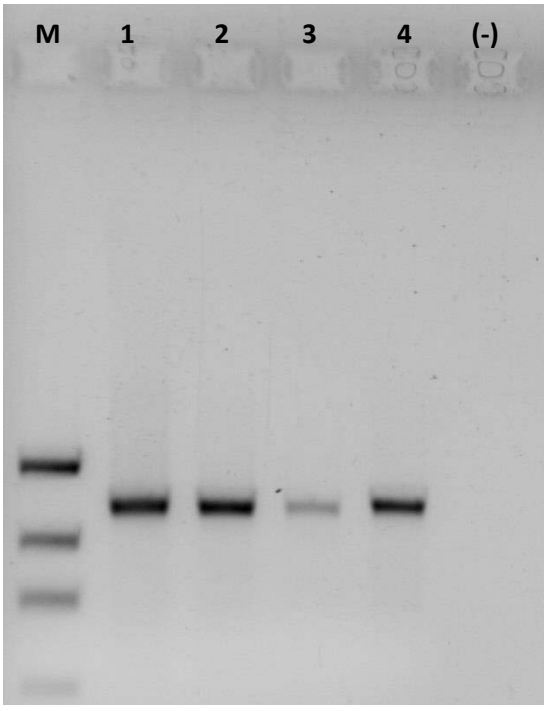


Fig. 6. Agarose gel electrophoresis image of PCR products. M: low DNA mass ladder (size of bands: 2000 bp, 1200 bp, 800 bp); numbers refer to sample IDs (Table 1). (-): negative control: PCR water.

A.8) Cycle sequencing

Cycle sequencing of PCR products was conducted using four different primers: GM3F, GM4R, GM1F, and GM1R (Table S3). The following protocol was applied:

- The reaction mixture was prepared:

1µl Big Dye v 3.1
1 µl sequencing buffer (2.5x)
1 µl primer (5 µM)
1 µl template DNA
1 µl distilled water

- The reactions were incubated in the therm cycler using the following program:

96°C	1 min	
96°C	10 sec	60x
57°C	5 sec	
60°C	4 min	
15°C	hold	

- The products were purified using Sephadex plate and submitted to capillary electrophoresis.

A.9) Sequencing results

Quality assessment of the raw sequences in the software Sequencher suggested that the sorting sample 4 yielded a high-quality 16S rRNA sequence with a length of 1450 bp (Supplementary information). The 16S rRNA sequence similarity between this sample and the *Polaribacter* clone (GQ347886) obtained during the 2010 bloom on April 8, 2010 is 99.8%. Other samples were unclassified or assigned to *Paraburkholderia*, suggesting possible contamination during cell sorting or subsequent analyses.

Sample ID	Sequence length (bp)	Classification based on SILVA taxonomy	Identity (%)	Accession number
1	1433	Betaproteobacteria;Burkholderiales; Burkholderiaceae;Burkholderia-Paraburkholderia;	99.93	AB091191
2	884	Betaproteobacteria;Burkholderiales; Burkholderiaceae;Burkholderia-Paraburkholderia;	100	AB477995
3	735	Unclassified;	-	-
4	1450	Bacteria;Bacteroidetes;Flavobacteriia; Flavobacteriales;Flavobacteriaceae; Polaribacter 2;	99.8	GQ347886

Table 2. Taxonomic classification of the sorted cell 16S rRNA sequences.

A.10) Supplementary information

> 16S_rRNA_sequences_Sample4_1450bp

ACCAGTTTTACCCTAGGCGGCTCCTTGCGGTGACCGACTTCAGGCACCCCCAGCTTCCAT
GGCTTGACGGGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGGATCATGGCTGATATC
CGATTACTAGCGATTCCAGCTTCACGGAGTCGAGTTGCAGACTCCGATCCGAACTGTGAT
ATGGTTTGTAGATTCGCGCTCTGTTGCCAGATGGCTGCTCATTGTCCATACCATTGTAGC
ACGTGTGTGGCCCAGGACGTAAGGGCCGTGATGATTTGACGTCATCCCCACCTTCCTCTC
TACTTGCGTAGGCAGTCTCGTTAGAGTCCCCATCTTTACATGCTGGCAACTAACGACAGG
GGTTGCGCTCGTTATAGGACTTAACCTGACACCTCACGGCACGAGCTGACGACAACCATG
CAGCACCTTGTAATCTGTCCGAAGAAAACTCTATCTCTAAAGCTGTCAGACTACATTTAA
GCCCTGGTAAGGTTCCTCGCGTATCATCGAATTAAACCACATGCTCCACCGCTTGTGCGG
GCCCCCGTCAATTCCTTTGAGTTTCAGTCTTGCGACCGTACTCCCCAGGTGGGATACTTA
TCACTTTCGCTTAGTCACTGAGCTAATGCCCAACAACTAGTATCCATCGTTTACGGCGTG
GACTACCAGGGTATCTAATCCTGTTCGCTCCCCACGCTTTCGTCCCTCAGCGTCAGTACA
TACGTAGTAGACTGCCTTCGCAATCGGTATTCTGTGTAATATCTATGCATTTCACCGCTA
CACTACACATTCTATCTACTTCCATATGACTCAAGTCAACCAGTATCAAAGGCAGTTCCA
TAGTTAAGCTATGGGATTTCACCTCTGACTTAATTGACCGCCTGCGGACCCTTTAAACCC
AATGATTCCGGATAACGCTCGGACCCTCCGTATTACCGCGGCTGCTGGCACGGAGTTAGC
CGGTCCTTATTCTTACAGTACCGTCAAGCTCGTACACGTACGAGTGTTTCTTCCTGTATA
AAAGCAGTTTACAACCCATAGGGCCGTCTTCCTGCACGCGGCATGGCTGGGTCAGAGTTG
CCTCCATTGCCCAATATTCCTCACTGCTGCCTCCCGTAGGAGTCTGGTCCGTGTCTCAGT

ACCAGTGTGGGGGATCTCCCTCTCAGGACCCCTACCTATCGTCGCCATGGTAAGCCGTTA
CCTTACCATCTAGCTAATAGGACGCATAGCCATCTTTTACCGATAAATCTTTAATTATAA
ATTCATGCGAATCTATAATACCATAAGGTATTAATCTTCATTTCTAAAGGCTATCCCTCT
GTAAAAGGTAGGTTCTATACGCGTTACGCACCCGTGCGCCGGTCGTCAGCGGAAGCAAGC
TCCCCTGTTACCCCTCGACTTGCATGTGTTAAGCCTGCCGCTAGCGTTCATCCTGAGCCA
TGATCAAAAT

Compound	**Volume**
5 M NaCl	3.6 ml
1 M Tris HCl	4 ml
20% formamide	4 ml
Blocking reagent	2 ml
MQ water	Add to 20 ml
10% SDS	40 µl

Table S1. The composition of the hybridization buffer used in HCR-FISH.

Probe	**Sequence (5' -> 3')**	**[Formamide]**
POL1270	TTTGTAGATTCGCTCTCTG	20
c14.1POL1270	TTCGTAGATTCGCTCTCTG	20
c14.2POL1270	TTTGGAGATTCGCGCTCTG	20
c5POL1270	TTGCCAGATGGCTGCTCATTG	20
POL1270h1	TCCGAACTGTGATATGGTTT	20
POL1270h2	TCCGAACTGTGATATGGTTT	20

Table S2. The sequences of FISH probes used in this study (modified from Chapter 3, Table S1).

Primer	**Sequence (5' -> 3')**	**Annealing temperature (°C)**
GM3F	AGAGTTTGATCMTGGC	48
GM4R	TACCTTGTTACGACTT	48
GM1F	CCAGCAGCCGCGGTAAT	57
GM1R	ATTACCGCGGCTGCTGG	57

Table S3. The sequences of PCR and cycle sequences primers used in this study.

Appendix B

Visualization of glycoside hydrolase genes of family 92 in the environment using direct-geneFISH and super-resolution microscopy

Visualization of glycoside hydrolase genes of family 92 in the environment using direct-geneFISH and super-resolution microscopy

Laura Zeugner[1], Jimena Barrero Canosa[2], Tanja Fischer[1], **Burak Avcı[1]**, Jens Harder[1],
Harder[1], Hanno Teeling[1], Bernhard M. Fuchs[1], Rudolf Amann[1]

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

[2] Technical University of Berlin, Institute of Environmental Technology, Environmental Microbiology, Germany

Phytoplankton blooms occur as seasonal events in many coastal areas. They grow within weeks to high biomasses and their decay is followed by an annually recurring succession of specialized heterotrophic bacteria. Genomic and proteomic analyses revealed that this succession is substrate-controlled and that certain bacterial strains have specific enzymatic repertoires for algal derived polysaccharide uptake and utilization. *Formosa* strain B (Hel1_33_131) and *Polaribacter* (Hel1_33_49), two *Flavobacteriia* occurring during the algae spring bloom in the German Bight, have specialized polysaccharide utilization loci for the degradation of mannose-rich sulfated polysaccharides that are common compounds of red algae cell walls. Within these loci, we wanted to detect and visualize specific glycoside hydrolase genes of family 92 that are coding for α-mannosidase. We used direct-geneFISH to simultaneously detect the ribosomal RNA for phylogenetic affiliation and the gene of interest in the target organisms. Structured illumination microscopy enabled a subcellular localization of the gene signals in situ. Our experiments showed that it is possible to simultaneously visualize functional genes and their host's rRNA on a single cell level not only in pure cultures but also in environmental samples containing populations of the flavobacterial species. By linking the presence of key genes involved in polysaccharide degradation with the cell identity on environmental samples, direct-geneFISH provides a quantitative insight into niche adaption, e.g. the ability to directly compete with other bacterioplankton clades during and after spring blooms. This method will prove especially valuable for investigating bacterioplankton populations based on their functional traits throughout the ocean.

Electronic Supplementary Information

Chapter 3

Polysaccharide-driven niche differentiation between distinct *Polaribacter* clades during North Sea spring algal blooms

Table S2. Relative abundances of *Polaribacter* clades based on (i) metagenome read recruitment (ii) MED analysis, and (iii) CARD-FISH counts.

Table S5. Metagenome read recruitment on *Polaribacter* MAGs and genomes. For the read recruitment using 2010-2012 metagenomes, most complete and least contaminated MAGs in each cluster, which has < 99% ANI value to other MAGs in this cluster was selected. For 2009 metagenomes, the most complete MAG in each clade was used. Read recruitment setting for each MAG is shown in the table.

Table S8. Genes detected in core and pan genome of four major *Polaribacter* clades

Table S9. MAG/genome-wide expression profile of *Polaribacter* clades. Subject refers to protein sequences detected by metaproteome analysis. Gene hits in blast searches are shown with query cover, e-value, percent identity, and bit score values.

Table S11. Number of peptidase and degradative CAZyme genes annotated in *Polaribacter* MAGs and genomes

Discussion

Supplementary File 1. Annotation of the polysaccharide utilization locus (PUL) genes in the already-sequenced genomes of the genus *Polaribacter.*

Supplementary File 2. Polysaccharide immuno-labeling protocol.

Supplementary File 3. *Aurantivirga*-related 16S rRNA gene sequences.

Supplementary File 4. Details about the oligonucleotide probe AUR452.

Supplementary File 5. Metagenome assembled genome (MAG) sequence of the North Sea *Aurantivirga* clade.

Supplementary File 6. Gene annotation of the *Aurantivirga* MAG.

Acknowledgements

I would like to deeply thank

Prof. Dr. Rudolf Amann for his generous support and continuous encouragement since 2010 and driving my “directed evolution” towards being a good microbial ecologist.

Dr. Hanno Teeling for being an excellent supervisor and mentor with honest criticism and constructive feedback.

PD Dr. Bernhard Fuchs for his invaluable scientific input to my thesis and always having an open door for my spontaneous questions.

Prof. Dr. Thorsten Brinkhoff for accepting to be the second reviewer of my thesis.

Prof. Dr. Rita Groß-Hardt and Dr. Jan-Hendrik Hehemann for participating to my defense committee as examiners.

Bledina Dede and Taylor Priest for involving my defense committee as student members.

Prof. Dr. Jens Harder for introducing me to “the art of cultivation”

Ben Francis, Meghan Chafee, Karen Krüger and Lennart Kappelmann for fruitful discussions and helpful hints to solve my daily bioinformatic problems.

Jörg Wulf, Dolma Michellold, Vanessa Kracke for their contributions to my lab work.

Greta Reintjes, Brandon Seah, Clara Martinez Perez, Benedikt Geier, Oliver Jäckle, Judith Klatt and Merle Ücker for proofreading of my thesis.

All other members of the Department of Molecular Ecology for keeping the group spirit strong.

David Probandt for sharing and enjoying the moment together. I will miss you Davido!

Endosymbionts Benedikt Geier, Miguel Angel Gonzalez Porras, and Oliver Jäckle for accepting me as an “ectosymbiont” to their group and “laughing loudly” together to relax my mind during “the spontaneous” breaks.

The MarMic Class 2016 (aka “garlic class”), especially Josephine Rapp, Rebecca Ansorge, Jenny Bachmann, Edinson Puentes Cala, Anny Cardenas Barbosa, Katie Harding and Sebastian Haas for their genuine friendships.

Bach and Scarlatti for composing vivid and lyric music, which enters to my ears, pass through my mind and reach on my fingers to motivate my writing skills.

Ahmet Hamdi Tanpınar for his masterpiece novel “Huzur”, which shapes my mindset and teaches philosophy of writing and aesthetics.

My “lifelong teachers” at the Technical University of Istanbul, Prof. Dr. A.M. Celâl Şengör, Prof. Dr. Cengiz Dökmeci, Prof. Dr. A. Nihat Berker, and Prof. Dr. İlhan Talınlı for introducing me the fascinating world of science and being excellent role models with solid scientific and intellectual backgrounds.

Son olarak…

Başta üst kat komşum Boran Kartal olmak üzere, Teknik Üniversite ruhunu koruyan ve yaşatan tüm mezunlar! Selam olsun sizlere…

Arpat Özgül! Bremen’deki “gönüllü” sürgünümde “uzaktaki yakınım” oldun. Sağol, varol.

Bremen Türk çetesinden Mert Süzen, Akın Şipal, İsmail Tsavous, Ayhan Feraşet ve Atakan Koç! Gözlerinizden öperim, eyvallah!

Annem Aysen Avcı, babam Süreyya Avcı, ağabeyim Barış Avcı ve diğer aile mensupları! Seneler boyunca bana verdiğiniz destek ve gösterdiğiniz sevgi için ne kadar teşekkür etsem az. Benim bugün buralara kadar gelebilmem sizin sayenizdedir…

Bu tez, doktora öğrenimim sırasında hayatlarını kaybeden dedelerim Hasan Avcı, Hüseyin Üver ve babaannem Nazmiye Avcı’nin aziz hatıralarına adanmıştır.

Ben giderim tezim kalır

Dostlar beni hatırlasın…

Avcı, Burak

Bregenzerstr. 21 28215 Bremen

Bremen, 27th August 2018

ERKLÄRUNG

Hiermit erkläre ich, dass ich die Doktorarbeit mit dem Titel:

Niches of abundant heterotrophic bacteria during North Sea spring algal blooms

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)