Polysaccharide utilization by bacteria associated with micro- and macroalgae

Dissertation

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

- Dr. rer. nat. -

dem Fachbereich Biologie/Chemie

der Universität Bremen

vorgelegt von

Fengqing Wang

Bremen, März 2024

Polysaccharide utilization by bacteria associated with micro- and macroalgae

Dissertation

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

- Dr. rer. nat. -

dem Fachbereich Biologie/Chemie

der Universität Bremen

vorgelegt von

Fengqing Wang

Bremen, März 2024

Die vorliegende Doktorarbeit wurde im Rahmen des Programms „International Max Planck Research School of Marine Microbiology (MarMic)“ in der Zeit von November 2019 bis März 2024 am Max Planck Institut für Marine Mikrobiologie angefertigt.

This thesis was prepared under the framework of the International Max Planck Research School of Marine Microbiology (MarMic) at the Max Planck Institute for Marine Microbiology from November 2019 until March 2024.

Gutachter: Prof. Dr. Rudolf Amann

Gutachter: Prof. Dr. Thorsten Brinkhoff

Gutachterin: Dr. Greta Reintjes

Prüfer: Prof. Dr. Michael W. Friedrich

Prüfer: Prof. Dr. Rudolf Amann

Prüfer: Prof. Dr. Thorsten Brinkhoff

Prüferin: Dr. Greta Reintjes

Datum des Promotionskolloquiums: 02.05.2024

也无风雨也无晴。

– 苏轼

Impervious to wind, rain or shine, I'll have my will.

- Translated by Xu Yuan-Zhong

Summary

Algae play a crucial role in marine carbon cycling. This involves various polysaccharides of different functions to which algae convert a significant proportion of their photoassimilated organic carbon. Suitable conditions provided, microalgae (phytoplankton) can proliferate very fast. During such blooms, substantial quantities of dissolved and particulate polysaccharide-rich organic matter are released to the environment, both by exudation of living and by decomposition of dead algal cells. While research on the community and functional dynamics of free-living bacteria during phytoplankton blooms is extensive, knowledge about particle-attached bacteria, especially regarding taxonomic composition and associated gene functions, is limited. Like planktonic microalgae, sessile macroalgae are also rich in polysaccharides and harbor bacteria with significant potential for polysaccharide degradation, particularly for complex polysaccharides. However, research on macroalgal phycosphere microbes so far primarily focused on community composition studies via 16S rRNA gene amplicon sequencing and (meta-)genomic analysis of single algal species or a limited number of samples. This thesis provides a comprehensive exploration of microalgae-associated microbial communities during phytoplankton blooms and of macroalgae-associated phycophere communities, both with a focus on polysaccharide utilization functions.

Chapter I establishes the theoretical foundation of this thesis. It starts out with a review of the critical roles that micro- and macroalgae play within the marine carbon budgets. It then narrows down to explore phytoplankton blooms, key events in marine carbon cycling marked by rapid carbon fixation, remineralization, and release. The studies of marine particles are summarized. Afterwards, common polysaccharides of micro- and macroalgae are introduced, as well as the mechanisms of bacterial polysaccharide utilization. Then research methodologies in microbial ecology are outlined with a focus on the analysis of omics data, which form the core methodologies in the two subsequent chapters, and finally the chapter closes with a summary of the research questions that this thesis aims to address.

In Chapter II, I present a study in which I explored dynamic shifts in community composition and polysaccharide degradation functions of particle-attached bacteria throughout an extensive phytoplankton bloom in contrast to those of free-living bacteria. Bloom progression was captured using microscopic, chlorophyll *a* and 18S rRNA gene amplicon data at high temporal resolution. I identified abundant bacterial clades using corresponding high-resolution 16S rRNA gene amplicon sequencing across three filter size fractions during the bloom. Through metagenome data analysis using both short- and long-read sequencing, I then identified and explored abundant polysaccharide degrading bacteria. Additionally, metaproteome data were analyzed to correlate bacterial proteins with the breakdown of algal glycans. The combined results indicated that, similar to their free-living counterparts, particle-attached polysaccharide-degrading bacteria targeted soluble and structurally simple polysaccharides such as laminarin. However, they also possessed abundant genes dedicated to the degradation of insoluble and structurally complex polysaccharides, setting them apart in their functional capabilities.

As detailed in Chapter III, I had a major part in analyzing the composition and polysaccharide utilization functions of macroalgal phycosphere bacteria on four types of macroalgae across the four seasons. Comparisons were made with microbial communities in surrounding seawater and sediment, revealing fourteen core genera consistently present on all algae. Metagenome analysis focused on polysaccharide degradation and secondary metabolite production, while cultivation techniques yielded pure isolates for draft genome sequencing. A high cultivability of macroalgal phycosphere bacteria enabled sequencing of numerous isolates, offering insights into a manifold of polysaccharide utilization loci and their possible polysaccharide substrates.

Chapter IV presents a comparative analysis of particle-attached bacteria originating from microalgae versus those from macroalgal phycospheres, a pioneering endeavor not previously undertaken. This way, I identified dominant bacteria shared between both studied communities, alongside unique clades specific to each community. Furthermore, the advantages of utilizing PacBio metagenome sequencing, drafting genome sequences of cultivable strains, and the application of multiple databases for functional annotation

are discussed, including insights to enhance the study of particle-attached bacteria and isolated strains. The thesis then closes with an outlook with a proposal for future projects.

Zusammenfassung

Algen spielen eine zentral wichtige Rolle im marinen Kohlenstoffkreislauf. Eine besondere Funktion kommt dabei Polysacchariden zu, da Algen einen erheblichen Teil ihres photoassimilierten organischen Kohlenstoffs in Polysaccharide unterschiedlichster Zusammensetzungen und Funktionen umwandeln. Mikroalgen (Phytoplankton) können sich unter geeigneten Bedingungen rasch vermehren. Während solcher Phytoplanktonblüten werden erhebliche Mengen gelöster und partikulärer polysaccharidreicher organischer Verbindungen freigesetzt, und zwar sowohl durch Exsudation lebender als auch infolge der Zersetzung abgestorbener Algenzellen. Die Dynamik frei lebender Bakterien während Phytoplanktonblüten ist im Hinblick auf die taxonomische Zusammensetzung und die damit einhergehenden Genfunktionen gut untersucht, doch das Wissen über partikelassoziierte Bakterien ist in dieser Hinsicht vergleichsweise begrenzt. Wie planktonische Mikroalgen, so sind auch sessile Makroalgen reich an Polysacchariden. Makroalgen beherbergen auf ihrer Oberfläche eine Vielzahl epiphytischer Bakterien mit erheblichem Potenzial zum Abbau von insbesondere komplexen Polysacchariden. Allerdings konzentrierte sich die Forschung an solchen epiphytischen Bakteriengemeinschaften der Makroalgen-Phykosphäre bislang im Wesentlichen auf die taxonomische Zusammensetzung mittels der Sequenzierung von 16S rRNA-Genamplikons und auf eher punktuelle (Meta-)Genomanalysen von Proben einzelner Algenarten oder einer insgesamt begrenzten Zahl von Proben. In der vorliegende Arbeit werden umfassendere Untersuchung von frei lebenden und partikelassoziierten Bakteriengemeinschaften während Phytoplanktonblüten und von Makroalgen-assoziierten Phykosphären-Bakteriengemeinschaften vorgestellt, und zwar mit besonderem Fokus auf der bakteriellen Nutzung von Polysacchariden.

Kapitel I führt in die theoretischen Grundlagen dieser Arbeit ein. Es beginnt mit einem Überblick über die entscheidende Rolle, die Mikro- und Makroalgen im marinen Kohlenstoffhaushalt spielen. Anschließend werden Phytoplanktonblüten als Schlüsselereignisse im marinen Kohlenstoffkreislauf vorgestellt, welche durch hohe

Raten der Fixierung, Remineralisierung und Freisetzung von Kohlenstoff gekennzeichnet sind. Daraufhin schließt sich eine Übersicht über die Rolle mariner Partikel an. Im Anschluss werden gängige Polysaccharide von Mikro- und Makroalgen sowie die Mechanismen der bakteriellen Polysaccharidverwertung vorgestellt. Schließlich werden relevante Forschungsmethoden in der mikrobiellen Ökologie mit Schwerpunkt auf der Analyse von Omics-Daten skizziert, welche den Kern der Methoden der beiden folgenden Kapiteln bilden. Das Kapitel I endet mit einer Zusammenfassung der Forschungsfragen, welche diese Arbeit zu beantworten sucht.

In Kapitel II stelle ich eine Studie vor, in der ich dynamische Veränderungen in der taxonomischen Zusammensetzung sowie der genetischen Polysaccharidabbau-Funktionen von partikelassoziierten im Vergleich zu frei lebenden Bakterien während einer ausgeprägten Phytoplanktonblüte in der Nordsee untersucht habe. Das Fortschreiten der Blüte wurde mit Hilfe von mikroskopischen Beobachtungen, Chlorophyll *a*-Messungen und 18S rRNA-Genamplikon-Daten in hoher zeitlicher Auflösung erfasst. Mit Hilfe entsprechend zeitlich hochaufgelöster 16S rRNA-Genamplikon-Sequenzierung über drei Filtergrößenfraktionen während der Blüte konnte ich distinkte abundante bakterielle Gruppen identifizieren. Durch die Analyse von Metagenomdaten, bei der sowohl Short- als auch Long-Read-Sequenzierung zum Einsatz kamen, identifizierte und untersuchte ich dann die prominenten Polysaccharid-abbauende Bakterien. Zusätzlich wurden Metaproteomdaten analysiert, um bakterielle Proteine mit dem Abbau von Algenglykanen in Beziehung setzen zu können. Die kombinierten Ergebnisse zeigten, dass die partikelassoziierten, Polysaccharid-abbauenden Bakterien ähnlich wie ihre frei lebenden Gegenstücke auf lösliche und strukturell einfache Polysaccharide wie Laminarin abzielten. Allerdings besaßen sie zudem reichlich Gene für den Abbau unlöslicher und strukturell komplexer Polysaccharide, wodurch sich ihre Funktionsrepertoires deutlich von denen frei lebender Bakterien unterschied.

Wie in Kapitel III beschrieben, war ich maßgeblich an der Analyse der taxonomischen Zusammensetzung und der genetischen Funktionen von Bakterien in den Phykosphären vierer Makroalgenarten während der vier Jahreszeiten beteiligt. Im Vergleich mit mikrobiellen Gemeinschaften des umgebenden Meerwassers und

Sediments wurden vierzehn algenspezifische Kerngattungen identifiziert, welche durchweg auf allen Algen vorkamen. Die Analyse der Metagenomdaten konzentrierte sich auf den Abbau von Polysacchariden und die Produktion von Sekundärmetaboliten. Darüber hinaus erlaubte die hohe Kultivierbarkeit der Phykosphären-Bakterien die Gewinnung zahlreicher Isolate und somit die Sequenzierung einer großen Zahl von Bakteriengenomen, welche zusätzliche Einblicke in eine Vielzahl genetischer Polysaccharidabbau Loci und ihrer möglichen Polysaccharid-Substrate boten.

In Kapitel IV wird eine erste vergleichende Analyse von partikelassoziierten Bakterien während der Blüten von Mikroalgen und von epiphytischen Bakterien aus Phykosphären von Makroalgen vorgestellt. Ein solcher Vergleich, wiewohl sinnvoll, ist in dieser Form bislang noch nicht vorgenommen worden. Mittels dieses Vergleichs konnte ich dominante Bakteriengruppen identifizieren, welche in beiden Bakterien-gemeinschaften vorkamen, sowie solche, die für jede Gemeinschaft spezifisch waren. Darüber hinaus werden in diesem Kapitel die Vorteile von Partialgenomen aus der PacBio-Metagenomsequenzierung sowie der Sequenzierung kultivierbarer Stämme erörtert, und außerdem die Anwendung diverser Datenbanken zur funktionellen Gen-Annotation behandelt. Die Arbeit schließt mit Überlegungen zur Verbesserung der Untersuchung von partikelassoziierten Bakterien inklusive geeigneter isolierter Stämme ab, sowie mit einem Ausblick auf zukünftige Projekte.

Abbreviations

AHL	acyl-homoserine lactone
ASV	amplicon sequence variant
AWI	Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research
CAZyme	carbohydrate-active enzyme
CCS	circular consensus sequencing
CE	carbohydrate esterase
CFU	colony-forming unit
CMIP6	Coupled Model Intercomparison Project v.6
CUT	carbohydrate utilization containing TonB-dependent receptors
DMSP	dimethylsulfoniopropionate
DOC	dissolved organic carbon
DOM	dissolved organic matter
eDNA	environmental DNA
EPS	extracellular polymeric substances
FCSP	fucose-containing sulfated polysaccharide
FL	free-living
GAG	glycosaminoglycan
GH	glycoside hydrolase
Gt	gigatons
GT	glycosyltransferase
HiFi	high-fidelity
HTCS	hybrid two-component system
LTER	long-term ecological research
MAG	metagenome-assembled genome
MFS	major facilitator superfamily
MPI	Max Planck Institute
MS	mass spectrometry
NPP	net primary production
ONT	Oxford Nanopore Technologies
OTU	operational taxonomic unit
PacBio	Pacific Biosciences
PL	polysaccharide lyase
POC	particulate organic carbon
POM	particulate organic matter

PUL	polysaccharide utilization locus
PULDB	PUL database
QS	quorum sensing
SBS	sequencing-by-synthesis
SGBP	surface glycan-binding protein
SSBP	substrate-specific surface sugar-binding protein
T9SS	type IX secretion system
TBDT	TonB-dependent transporter
TEP	transparent extracellular particles
TRAP	tripartite ATP-independent periplasmic
UVP5	Underwater Vision Profiler 5

Contents

Introduction

1.1 The oceanic carbon cycle and its carbon budgets

About 71% of the surface of our planet Earth is covered by oceans. It's estimated that these oceans store about 50-60 times as much carbon as the atmosphere [1-3], namely 140,000 vs. 2,200 Gt (gigatons) (IPCC AR5). Around 0.6% of the oceanic carbon is stored in the surface layer [4, 5]. The annual global photosynthetic net primary production (NPP) amounts to approximately 104.9 Gt of carbon [6]. To this, the oceans contribute a significant 35-65 Gt [6-10], which leads to the annual (net) removal of 2-2.5 Gt of carbon from the atmosphere [3]. The bulk of this NPP is carried out by microscopic planktonic primary producers (phytoplankton) with an estimated 47.5 Gt of fixed carbon per year [6]. This number may even be an underestimation, since satellite-based observations allow only monitoring of chlorophyll in surface waters [11], whereas computational models of phytoplankton productivity are error-prone due to the many variables that need to be considered [12]. In contrast to offshore areas or even the open ocean, sessile macroalgae represent the dominant primary producers in many shallow coastal systems [13-15]. Algal forests have been estimated to fix about 1.3 Gt of carbon per year, which is about equivalent to the NPP of the Amazon rainforest [16], and the total macroalgal NPP has been estimated to reach close to 2 Gt annually [17]. Macroalgal NPP is lower than that of phytoplankton, but likewise also difficult to quantify, as it varies considerably with species and geographic location [18]. For these reasons, macroalgae have been largely excluded in discussions about ocean carbon sinks thus far [17], and only recently have garnered more attention in this respect (e.g., [19]).

Photosynthetic plankters comprise mainly cyanobacteria and microalgae such as unicellular diatoms, dinoflagellates, green algae, and haptophytes (e.g., coccolithophores) [20]. Within the photic zone, they convert dissolved inorganic carbon into organic carbon, subsequently releasing dissolved organic matter (DOM) and particulate organic matter (POM) (Figure 1.1, [21]). About 3-50% of phytoplankton primary production is released as dissolved organic carbon (DOC) ([22] and references therein). The actual number depends on phytoplankton species and environmental conditions such as light and

temperature [23-26]. Healthy phytoplankton cells typically release a small portion of DOC, while weakened, senescent, or dead cells may leak larger amounts or all of their soluble contents [24]. Carbohydrates represent a significant proportion of the soluble compounds that are released by eukaryotic phytoplankton [27, 28]. However, also various osmolytes account for a considerable proportion of the exuded carbon, e.g., dimethylsulfoniopropionate (DMSP) [29] with a global annual production of more than one Gt [30]. DMSP has been estimated to account for 10% of carbon fixed by phytoplankton [31, 32].

Zooplankton is also involved in the production of POC (particulate organic carbon). POC is formed by zooplankton during predation on large algae via fragmentation and by egestion of fecal pellets [33]. These processes are accompanied by the release of additional DOC. For instance, a study on *Arcatia tonsa* copepods has shown that during grazing on diatoms, about 3% of the prey carbon was released as DOC [34]. According to recent estimates, zooplankton and phytoplankton together release about 15.5 Gt C yr^{-1} of POC and 16.9 Gt C yr^{-1} of DOC. Of the POC, 4.4 Gt C yr^{-1} are remineralized, and 2.7 Gt C yr^{-1} are dissolved to DOC [35]. Bacterial cell walls originating, e.g., from viral lysis, represent another source of transient POC [36]. Up to 10-40% of DOC is colloidal, around 10% of which is also transformed into POC [37-40]. Carbohydrates make up 8-10% of the suspended and 3-18% of the sinking POC [41]. The remineralization of POC is mainly carried out by zooplankton and heterotrophic bacteria [21]. Zooplankton thus not only produce POC but also actively participate in POC utilization by incorporating ingested organic carbon into biomass and by releasing carbon dioxide through respiration. It is estimated that small zooplankters on average consume approximately 62% of the daily phytoplankton production [42] (reviewed in [43]). However, a recent study showed that the exact value is uncertain to an extent that corresponds to differences in the scale of Gts of carbon per year, which makes zooplankton grazing one of the largest uncertainties in Coupled Model Intercomparison Project v.6 (CMIP6) marine biogeochemical models [44].

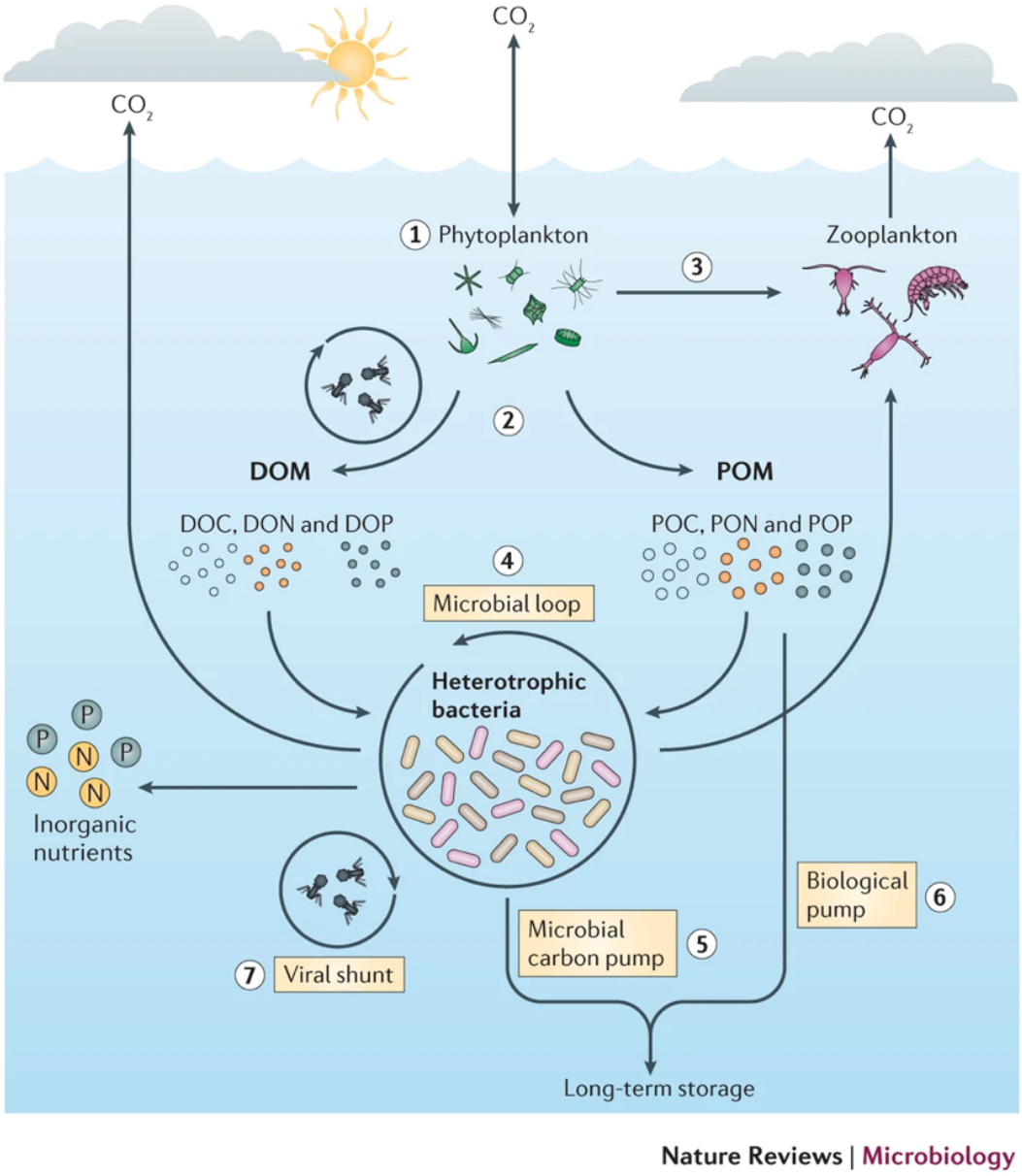


Figure 1.1. The marine carbon cycle. (1) Conversion of inorganic carbon (such as carbon dioxide) to organic carbon by photosynthetic phytoplankton species. (2) Release of both dissolved organic matter (DOM) and particulate organic matter (POM) from phytoplankton. (3) Consumption of phytoplankton biomass by zooplankton grazers. (4) Mineralization and recycling of organic matter by diverse heterotrophic bacteria. (5) Transformation of organic carbon into recalcitrant DOC. (6) Export of phytoplankton-derived POM from surface water to deeper depths via sinking. (7) Contributions of viral-mediated cell lysis to releasing DOM and POM from both the phytoplankton and bacterial pools. (The figure is reproduced from Buchan *et al.* (2014) [21]).

About 75-95% of phytoplankton NPP is utilized by heterotrophic bacteria and zooplankton [3]. The microbial utilization of organic carbon occurs not only in surface waters but extends to deeper waters while particles sink [3]. The DOC that is resistant to microbial decomposition is transported and sequestered by the biological carbon pump [3]. The total carbon sequestered in the oceans amounts to approximately 1,300 Gt [45], with DOC comprising about 700 Gt [46].

1.2 Marine macroalgae and the carbon cycle

Macroalgae are categorized into three phyla: *Rhodophyta* (red algae), *Chlorophyta* (green algae), and *Phaeophyta* (brown algae). A study using hundreds of metagenomes from Tara Oceans [47] and the Malaspina 2010 Circumnavigation [48] revealed that the Atlantic and North Pacific Oceans contain the most diverse macroalgae. They also found macroalgal environmental DNA (eDNA) in water samples at a depth of 4,000 meters and 4,860 km away from the nearest coastline, suggesting that macroalgal tissues can be transported over large distances [19].

Sessile macroalgae and few marine plants (e.g., *Posidonia oceanica* in the Mediterranean) represent the dominant marine vegetation in coastal systems, where they also represent the dominant primary producers [14, 15] and thus play an eminent role in maintaining a high biodiversity [49]. For instance, brown algae have a greater capacity to fix carbon per area than terrestrial forests [50-54]. Macroalgae cover an estimated area of 6.06-7.22 million km^2 with a global NPP of 1.32 Gt C yr^{-1} [16] (range in 1.02-1.96 Gt C yr^{-1} reported by Krause-Jensen *et al.* [17]). These areas and NPP numbers are growing due to steady increases in seaweed aquaculture [55]. Approximately 43% of annual macroalgal NPP contributes to the global carbon flux (around 0.68 Gt C yr^{-1}), ~52% of which is in the form of DOC and ~48% in the form of POC [17]. It has been estimated that macroalgae contribute at least 3% of the global marine NPP, and at least 20% of the coastal NPP [56]. A rough estimate suggests that macroalgae could sequester about 0.17 Gt C yr^{-1} (with a range of 0.06-0.27 Gt C yr^{-1}) globally [17]. Additionally, marine macroalgae can also sequester some of the carbon by producing polysaccharides such

as fucoidan, which are comparably persistent, since they are difficult to utilize by heterotrophic bacteria [54].

1.3 Phytoplankton blooms

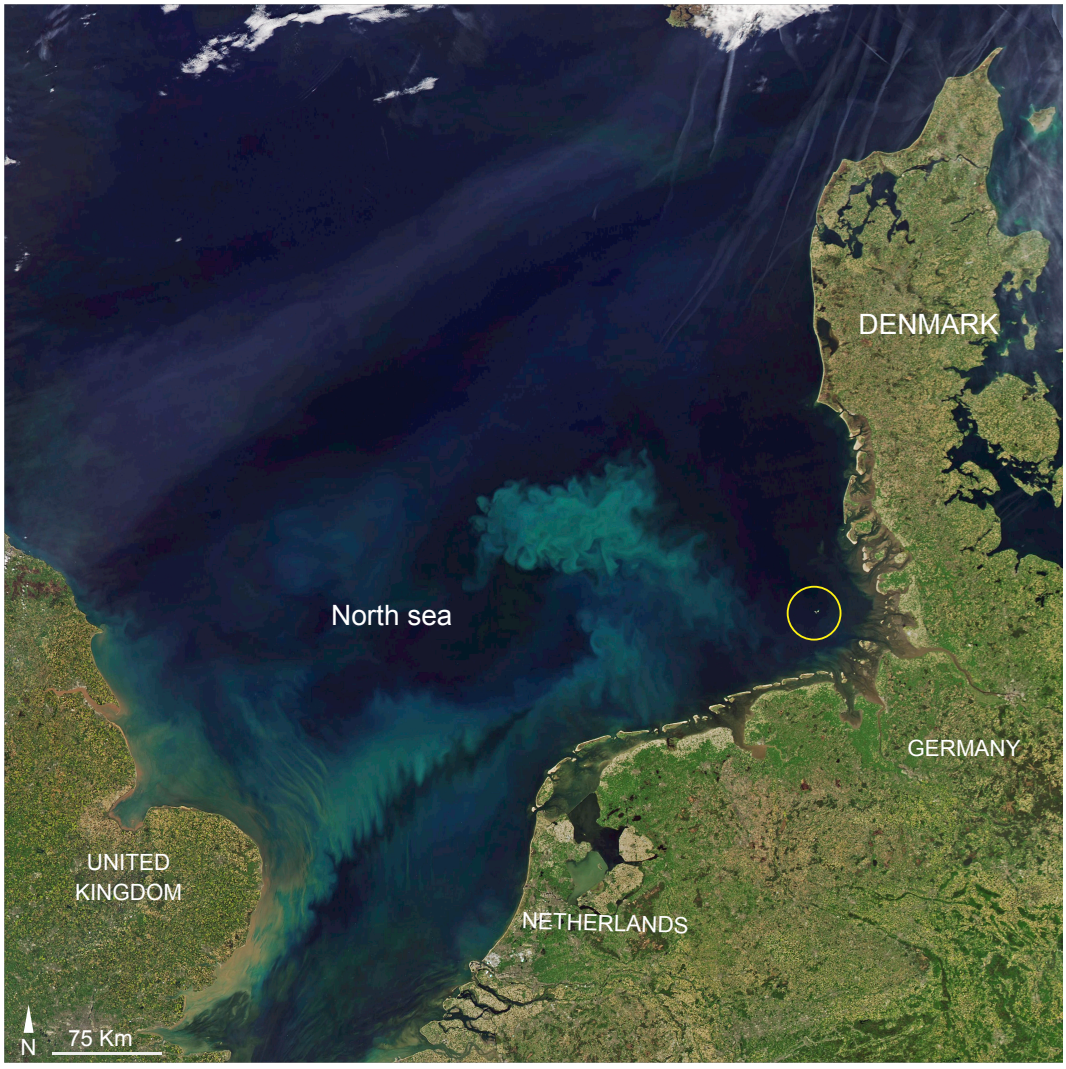


Figure 1.2. Satellite image of a phytoplankton bloom in the German Bight, North Sea in May of 2018. The yellow circle presents the location of Helgoland Island. NASA Earth Observatory images by Joshua Stevens, using Landsat data from the U.S. Geological Survey and MODIS data from LANCE/EOSDIS Rapid Response.

A phytoplankton bloom is characterized by swift proliferation and accumulation of phytoplankton within a short period in the surface layers of both freshwater and seawater systems (e.g., Figure 1.2). Such blooms constitute temporary ecosystem imbalances due to climatic triggers or changes in the food web. Contributing factors comprise sunlight, temperature, and nutrient influxes. The relative balance between the supply and removal of nutrients (including nitrogen, iron, and phosphorus) determines the nutrient that predominantly limits phytoplankton growth. Winds and changes in salinity by influxes of different seawater bodies or rain are also factors that can influence the development of phytoplankton blooms [57]. Nitrogen-rich agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean [58]. An analysis of coastal algal blooms between 2003 and 2020 has revealed significant increases in terms of spatial extents and bloom frequencies [59]. The magnitude of carbon that is fixed during such bloom events acts as the foundation for fueling food webs and ecosystem productivity. High amounts of DOC and POC released into the surrounding water accompany phytoplankton blooms and in particular their decline, thereby supplying substantial amounts of substrates to heterotrophic bacterioplankton.

The water body has a selective influence on which phytoplankton taxa dominate. Diatoms represent the most prominent phytoplankton group and have been estimated to fix up to 20 Gt of carbon annually, contributing about 40% to the POC export [60, 61]. The diatom *Chaetoceros decipiens*, for example, has been shown to release about 21% of its fixed carbon as DOC [23]. Furthermore, datasets generated by the Tara Ocean Foundation from a wide range of oceanic regions have revealed that diatoms have a higher diversity in the open ocean than previously anticipated, suggesting that diatoms may be even more relevant than is generally considered [62].

Photosynthetic dinoflagellates represent the second most important oceanic primary producers. About half of the dinoflagellates are photosynthetic [63], and together with ciliates, they comprise approximately 24% of the total biomass in the northern waters between Ellesmere Island (Canada) and Greenland [64-66]. Some dinoflagellates have symbiotic lifestyles, e.g., as endosymbionts in corals, whereas others are parasitic. Many dinoflagellates wear a protective shell of plates (thecate or lorica) that are mainly

composed of cellulose. These thecate dinoflagellates are believed to be monophyletic, whereas the athecate dinoflagellates may constitute a paraphyletic group [67]. Dinoflagellate blooms are common phenomena, which is a problem since many dinoflagellates are prolific toxin producers [68, 69]. Examples are blooms of *Noctiluca scintillans* in the North Sea [70] and of *Prorocentrum donghaiense* in the East China Sea [71]. Blooms of non-toxic dinoflagellates comprise for example *Protoperidinium steinii* blooms that have been reported for Indian waters [72] and *Gymnodinium impudicum* blooms in the Yellow Sea [73]. It has been shown that the production of dinoflagellate cysts during blooms can significantly contribute to total vertical POC fluxes [74]. Dinoflagellates, while not as abundant as diatoms, contribute significantly to oceanic primary production. For instance, during a *Heterocapsa triquetra* bloom in Puyuhuapi Fjord in Chilean Patagonia, a daily gross primary production has been determined in the range of 0.6 to 1.6 g C m^{-2} d^{-1} [75].

Further abundant and globally distributed phytoplankton taxa include the large group of haptophytes, which include coccolithophorids. The latter have a significant influence on the global carbon cycle through the production and export of particulate inorganic carbon by building calcified shells [76], representing up to 40% of the local primary production in studied habitats [77, 78]. In coastal areas, it has been shown that the haptophyte genus *Phaeocystis*, due to their small cells with high surface-to-volume ratios, can outcompete diatoms under certain conditions [79]. *Phaeocystis* blooms are frequent and have been observed from the tropics to the poles and in coastal and upwelling areas [80-82]. *Phaeocystis* often dominates spring and summer blooms after diatoms have peaked in the coastal North Sea [83], where they can account for up to 65% of the annual primary production [84].

Phytoplankton blooms are transient phenomena that can be terminated by many factors including nutrient depletion, self-shading, grazing (e.g., by copepods), and various infections [85-88], as well as the coagulation of algae and increased sinking of the formed particles [89].

1.4 Marine particles

In the broadest sense, marine particles comprise a diverse range of particulate inorganic and organic matter. The latter comprises matter from marine algae and marine animals, including zooplankton fecal pellets, and organic matter from marine bacteria. A specific type of particle-forming organic matter are the gel-like polysaccharide-rich extracellular polymeric substances (EPS) that numerous microalgae exude, e.g., *Chaetoceros gracilis* [90] and *Phaeocystis globosa* [91]. Many factors affect algal EPS production. For instance, nutrient limitation can increase EPS exudation [92]. A portion of these EPS can further aggregate, whereby denser transparent extracellular particles (TEP) are formed. Similarly, TEP can also originate from substances that are released by bacteria and during the degradation of other organic particles [93]. TEP usually ranges from 0.4 to > 200 µm in diameter and is about two to four orders of magnitude more adhesive than phytoplankton or mineral particles [94, 95]. During phytoplankton blooms, the formation of large amounts of sticky TEP promotes the transformation of DOC to POC [96]. Inorganic components such as mineral dust, sediment, sand, biomineralized shells, suspended clay, and persistent plastic particles further contribute to the composition of aggregated marine particles.

Marine particles play an important role in transporting and sequestering carbon to the seafloor, especially in open oceans. Particles in the upper ocean can form larger aggregates (marine snow) that sink rapidly and contribute to the food web in the dark deep ocean [97, 98]. Sinking rates vary from 0.1 to 100 m d^{-1} depending on particle shape and diameter [99], which can vary over a large range from a few to 100 micrometers or even to centimeter-sized loose aggregates [100]. Kiko *et al.* used several Underwater Vision Profiler 5 (UVP5) camera systems to capture and characterize marine particle abundances and size structure patterns at a global scale [101]. The results showed that particle abundances were higher in areas of high primary productivity and in coastal areas [101]. The proportion of particles with larger sizes increased with increasing water depths [102], e.g., particles with diameters of 3-10 μm were prominent in the mixed layer, accounting for more than half of the total number of particles. At water depths greater than 100 meters, particles with diameters of 5-6 μm were the dominant fraction [102].

Marine particles contain up to 30 wt% of polysaccharides [103]. While particles transport carbon through the seawater column, they are 'hot spots' of intense microbial activity [104]. Throughout the sinking process, approximately 98% of the carbon in particles undergoes extensive biological degradation [103], a process in which PA bacteria play a crucial role. The chemical composition of particles influences the taxonomic composition of both PA and surrounding FL bacterial communities [105]. The abundance of PA bacteria varies across studies, with reported percentages ranging from as low as 0.1% [106] to as high as 20%, and is influenced by factors such as particle size and the sampled marine ecosystem [107]. While generally lower than FL bacterioplankton, PA bacteria can occasionally exhibit locally higher densities [106, 108-110]. Ranked by bacterial density, the majority of PA bacteria have been shown to be in direct contact with TEP particles, algae, or diatom frustules [111]. Many PA bacteria, like members of the genus *Roseobacter*, may have the 'swim-or-stick' lifestyle, attaching to phytoplankton cells using chemotaxis [112, 113] and turning on the 'switch' subsequently [114]. Bižic-Ionescu *et al.* argued that PA bacteria must have free-living lives before encountering particles [115].

1.5 Marine polysaccharides

Polysaccharides are among the most important organic compounds in the marine realm, mainly originating from microalgae, macroalgae, and some bacteria. Photosynthetic algae can fix carbon dioxide in the dark reaction (Calvin Benson cycle) using reduction equivalents gained in the water-splitting light reaction. This initially produces glyceraldehyde 3-phosphate, a triose that can serve either as a source of biochemical energy or as a precursor for various biosynthetic pathways, many of which result in polysaccharides. The polysaccharide content of diatoms, for example, typically ranges from 30 to 65% of the specific dry weight of the cell [116], and can be as high as about 90% in extreme cases [117].

1.5.1 Storage polysaccharides

The abundant marine polysaccharide laminarin acts as the storage molecule in stramenopiles including brown algae [118], diatoms and raphidophytes [119], and in haptophytes (*Prymnesiophyceae*) such as *Phaeocystis* [84, 120]. Diatoms can store substantial proportions of their photo-assimilated carbon in the form of laminarin, for example, up to 70% of the dry organic matter has been demonstrated in *Chaetoceros pseudocurvisetus* [121]. On a global scale, laminarin has been estimated to account for 11 ± 8% of the primary production, which is why laminarin represents a major molecule in the marine carbon cycle [122]. Laminarin production during algal spring blooms in the coastal North Sea is usually driven by various diatom species, haptophytes (mostly *Phaeocystis*) and occasionally raphidophytes (often *Chattonella*), and it has been found that unreleased laminarin contributes significantly to POC in the surface seawater [122]. Laminarin has a simple structure consisting of a β-1,3-linked helical glucose homopolymer backbone with occasional β-1,6-branches. It is water-soluble and rapidly degraded by bacteria. Laminarinase genes are ubiquitous, not only in surface waters [123] but also in deeper waters [124], sediments [125] and macroalgae samples [126]. A very recent study on FL bacteria during a North Sea spring phytoplankton bloom using high temporal resolution metatranscriptomics revealed that laminarin-targeting genes were the highest expressed among all polysaccharide-targeting genes [127].

In the same study, α-glucans represented the second-most prominent class of targeted polysaccharides. This glycan class comprises various polysaccharides with α-1,4 and α-1,6 glycosidic linkages, such as starch and glycogen. Dinoflagellates, red algae (*Rhodophyta*), and green algae (*Chlorophyta*) typically utilize starch as their energy storage form [128], whereas animals, fungi, and bacteria store energy in the form of glycogen [129]. In a recent preprint study, it has been demonstrated that marine *Flavobacteriia* possesses the capability to degrade microalgal laminarin and use the excess glucose to synthesize α-glucan storage polysaccharides [130].

1.5.2 Cell wall polysaccharides

Diatoms have distinctive cell walls that are composed of two silica shells. These shells, termed frustules, consist of two pieces (epi- and hypotheca) that gear into each other like two pieces of a box [131]. Various polysaccharides are encrusted within the frustule silica matrix [132], contributing to the composition of the diatom cell wall. The monosaccharide composition can vary considerably depending on species and physiological state, but in general, mannose is the most abundant monosaccharide [132]. For example, mannose is dominant in *Chaeotoceros socialis* [133], *Phaeodactylum tricornutum* [134], *Stauroneis amphioxys* [135], *Navicula pelliculosa* [136], *Coscinodiscus radiatus* [137], and *Thalassiosira pseudonana* [138]. Fucose is prominent in *Thalassiosira gravida* [133], *Corethron hystrix* [133], and *Nitzschia angularis* [139]. A cell wall polysaccharide containing 3-linked D-mannoses decorated with sulfated glucuronic acid was extracted from *Phaeodactylum tricornutum* [140]. Sulfated glucuronomannan has also been found in other diatom species, e.g., *Navicula pelliculosa* [141], suggesting that it might be a wide-spread constituent of the cell wall in diatoms [142]. Glucose has been found to be dominant in *Thalassiosira weissflogii* [143] and rhamnose in *Chaeotoceros affinis* [133]. For *Thalassiosira pseudonana*, it has been shown that chitin acts as a structural polysaccharide in their frustules [144]. In contrast to many other phytoplankters, raphidophytes and some haptophytes, such as *Phaeocystis*, do not possess outer shells. Flagellated *Phaeocystis* species for instance are enveloped by body scales [80, 145-147], likely comprised of sulfated pectin-like polysaccharides, cellulose and glycopeptides [84] —features commonly found also in other haptophytes (reviewed in [148]). Flagellate cells of *Phaeocystis* form various star-like structures by excreting filaments composed of α-chitin, an N-acetyl-D-glucosamine polymer ([84] and references therein). Many dinophytes on the other hand do feature rigid walls that are known to contain cellulose [149].

Brown algae share common structural polysaccharides with plants (cellulose) and animals (sulfated fucans) [150, 151], but these only account for small proportions of their cell wall polysaccharides [152]. In brown algae, alginates and fucoidans are the unique and dominant cell wall components [151]. Fucoidan, a type of fucose-containing sulfated

polysaccharide (FCSP), was first extracted from kelp by Kylin in 1913 and was named ‘fucoidin’ [153]. It is highly soluble and contains L-fucose and sulfate groups, accompanied by other monosaccharides such as galactose, mannose, xylose, glucuronic acid, and arabinose [154]. Degradation of fucoidan requires many carbohydrate-active enzymes (CAZymes), but only a few bacteria are capable of this process [155, 156]. Consequently, fucoidan has been considered as a way to sequester carbon in oceans [54]. The structural polysaccharides of red algae are of different compositions and contain agar, agarose, agaropectin, carrageenans, cellulose, xylan, sulfated galactans, and porphyran [157-160]. The structural polysaccharides in green algae have not been studied as deeply as in brown and red algae. Ulvans are commonly reported in green algae [161, 162], but cellulose, pectins, xyloglucans, xylans, and sulfated galactans have also been found in green algae [159, 163-165].

1.5.3 Secreted extracellular polysaccharides

Fogg [166] and Hellebust [167] proposed in 1966 and 1974 that phytoplankters secrete various metabolites into the marine environment, a phenomenon that was subsequently confirmed to be common among healthy phytoplankton [168-170]. These metabolites are referred to as extracellular products, with EPS being one major class. Extracellular polysaccharides are the most abundant components in algal-associated EPS, with reports indicating that they constitute 40-95% of the total EPS [171, 172]. Most microalgal extracellular polysaccharides appear in the form of heteropolysaccharides [173] and contain significant amounts of uronic acids and sulfate residues [27, 174]. The composition of extracellular polysaccharides varies among different diatoms [132], for example, the extracellular polysaccharides of *Chaetoceros affinis* [175, 176], *Amphora* rostrate [177] and *Chaetoceros curvisetus* [178] contain a higher amount of fucose and a small amount of galactose and rhamnose, while xylose and glucose dominate in the extracellular polysaccharides of *Cylindrotheca closterium* [179]. Mucopolysaccharides have been shown to contribute 5-60% to the POC during the stationary phase of a bloom [84], and *Phaeocystis* is well known for producing mucopolysaccharides to form colonies. Common monosaccharides in *Phaeocystis globosa* and *Phaeocystis pouchetii* are

arabinose, galactose, mannose and xylose [180, 181] and sialic acid type amino sugars [182].

A recent study has revealed that brown algae release fucoidan into the ocean. This study focused on *Fucus vesiculosus* brown algae along the Baltic Sea coast in southwestern Finland. The findings indicate that *Fucus vesiculosus* secretes fucoidan at a rate of 0.3% of its biomass per day and that fucoidan constitutes no less than 18% to 50% of the total dissolved organic carbon released by brown algae [54]. Similarly, exudates of the red alga *Amansia* contain high amounts of galactose and mannose+xylose and small amounts of glucose, rhamnose, and arabinose [183]. Likewise, exudates of the green alga *Halimeda* have been shown to contain high amounts of galactose, glucose, and mannose+xylose and small amounts of rhamnose and arabinose. Fucose was detected in the exudates of both the red alga *Amansia* and the green alga *Halimeda*, but in very low amounts, which is quite distinct from brown algae [183]. It has also been established that exudates from macroalgae represent another important source of TEP formation [184].

Besides algae, extracellular polysaccharides from marine bacteria and fungi are also an important source that fuel polysaccharide pools in the ocean. Most EPS produced by marine bacteria are heteropolysaccharides [185]. For instance, the main components of extracellular polysaccharides isolated from *Pseudoalteromonas* sp. strain SM20310 from the Arctic Sea were mannose, glucose, galactose, rhamnose, xylose, N-acetylgalactosamine and N-acetylglucosamine [186]. The extracellular polysaccharides of *Polaribacter* strain SM1127, isolated from an Arctic *Laminaria* brown alga, have been shown to consist of N-acetylglucosamine, mannose, glucuronic acid, moderate amounts of galactose and fucose and minor amounts of glucose and rhamnose [187]. Other examples of polysaccharide production by bacteria include polysaccharides such as cellulose, alginate, glucan, and *Vibrio* exopolysaccharides [185, 188-191]. For fungi, it is known that they produce mannans (e.g., [192, 193]). For example, fungi *Lineolata rhizophorae* (present in red mangrove roots) and *Aspergillus fumigatus* (a globally distributed species present in various marine sites) contain galactomannan in the cell wall [192-194]. Extracellular polysaccharides were mainly composed of mannose with minor

amounts of galactose and glucose extracted from the marine fungus *Hansfordia sinuosae* [195].

1.5.4 Host glycans

More than 70% of eukaryotic host cells are decorated with branched surface heteropolysaccharides that are commonly subsumed as host glycans [196, 197]. In the human gut, host glycans include mucin O-linked glycans, N-linked glycoproteins, and highly sulfated glycosaminoglycans (GAGs) [198, 199]. Much of the research on host glycans as of to date is related to mammals. In contrast, the host glycans of algae and invertebrate animals are poorly characterized. N-glycans are known to play a role in interactions between cnidarians and their dinoflagellate symbionts (e.g., [200]), and in general the structures on N-glycans seem to depend on their ecological roles. Also algae are known to feature N-glycans. Studies have revealed that diatoms [201] and red alga *Porphyridium* [202], contain mannose-rich N-glycans, while oligomannosidic type N-glycans have been found in the green microalga *Chlorella vulgaris* [203]. Galactose and rhamnose are also frequent in the glycans of the diatom *Craspedostauros australis* [204]. Sialic acids were found in algae just like in cell surface glycans of invertebrate animals. For instance, galactose associated with sialic acids in N-glycans were found in *Chlamydomonas reinhardtii* green algae [205], and sialic acids have also been reported in dinoflagellates [200]. N-glycans containing xylose were found in *Volvox carteri* green algae [206], and fucose-containing N-glycans have been described for the diatom *Phaeodactylum tricornutum* [207]. High mannose-type N-glycans are predominantly present in many seaweeds [208].

It is known that complex N-glycans serve as a nutrient source for *Bacteroides* species in the human gut [199], in particular, the more conserved core after the removal of decorating monosaccharides and after removal from the protein surface by a GH18 [198]. As hypothesized in Chapter II of this thesis, microalgal host glycans likely represent a prominent polysaccharide substrate in the marine environment. After initial degradation by PA bacteria, large amounts of decorating monosaccharides such as mannose, fucose,

rhamnose, etc., are released to the environment, which are likely substrates for FL bacteria (see Chapter II).

1.6 Algal polysaccharide utilization

In general, polysaccharides of algal origin can be either water-soluble or insoluble. The former tend to become part of the DOM and the latter part of the POM, either as embedded constituents in remnants of algal cells or by aggregation. Heterotrophic bacteria usually take-up substrates in dissolved form, which is why POM must be extracellularly decomposed into soluble components before uptake. This is why simple, soluble polysaccharides (e.g., laminarin) are usually more rapidly remineralized than polysaccharides in particles, especially when they are chemically complex. The speed of degradation of particles is dependent on many factors, but it can be sufficiently slow that POM reaches the seafloor even in the deep sea, where it can become buried and sequestered for extended periods of time, in particular when conditions are anoxic. The rate of bacterial mineralization influences the rate of sequestration, which is why investigations into the degradation of algal polysaccharides by heterotrophic bacteria are of great scientific interest. This is an aspect that also extends beyond the study of how human gut microbiota decompose plant glycans, which is where most of the fundamental mechanisms of bacterial glycan degradation were first discovered [209, 210].

The bacterial utilization of polysaccharides usually relies on specific degradative CAZymes, including glycoside hydrolases (GHs), carbohydrate esterases (CEs), and polysaccharide lyases (PLs). The utilization furthermore typically involves additional proteins and enzymes. This can include substrate-specific surface sugar-binding proteins (SSBPs), outer membrane TonB-dependent transporter (TBDT) complexes including SusD- and SusC-like proteins, sulfatases and deacetylases, transcriptional regulators and inner membrane transporters such as tripartite ATP-independent periplasmic (TRAP) transporters, ATP-binding cassette (ABC) transporters, and major facilitator superfamily (MFS) transporters (e.g., [211-213]). SusE and SusF are two additional types of outer membrane proteins that contribute to starch binding [214]. Also, CAZyme excretion by specific secretion systems can be involved [215].

The genes encoding these uptake proteins are often closely clustered with relatively few intervening genes. The most notable instances of such gene co-localization are the extensively studied polysaccharide utilization loci (sg. polysaccharide utilization locus; abbr. PUL), which are ubiquitous in the genomes of polysaccharide-degrading *Bacteroidota*. The PUL concept was first introduced by Xu *et al*. [216] and first termed by Bjursell *et al*. [210]. Canonical PULs are found most frequently in *Bacteroidota* and, in the original definition, consist of at least one consecutive pair of *susC* and *susD* genes [217]. The *susC* gene encodes a SusC-like TBDT [218], while the *susD* gene encodes a SusD-like protein that serves as a lid capable of opening and closing the TBDT. CAZymes play a crucial role within PULs, with extracellular enzymes responsible for breaking down polysaccharides into oligosaccharides. These oligosaccharides are then transported into the periplasm through SusCD-like TBDTs [219-221]. Subsequently, intracellular enzymes are involved in the hydrolysis of oligosaccharides into monosaccharides. Examples of functional modeling are shown in Figure 1.3. Sulfatase genes are commonly found in PULs of marine polysaccharide-degrading *Bacteroidota* due to the fact that many marine algae produce sulfated polysaccharides [222]. For example, sulfated polysaccharides are in the form of carrageenan, fucoidan and ulvan (reviewed in [223]). Many PULs, particularly those associated with *Flavobacteriaceae*, have been experimentally validated, e.g., laminarin and alginate PULs for 'Gramella forsetii' KT0803 [224], an ulvan PUL for 'Formosa agariphila' KMM 3901^T [225], α- and β-mannan PULs for *Salegentibacter* sp. Hel_I_6 [226], and α-glucan, carrageenan, xylan and fucoidan-containing PULs for *Zobellia galactanivorans* Dsij^T [227]. Currently, PULs that have been experimentally characterized and PULs that have been predicted in 2,065 bacteroidetes species are present in the PUL database (PULDB; http://www.cazy.org/PULDB) [228]. The number of genes contained in different PULs varies widely. One of the longest reported PULs (or PUL cluster) known to date has been predicted in the core macroalgal phycosphere species *Algibacter* sp. 4-1052 genome and comprises around 100 genes [126].

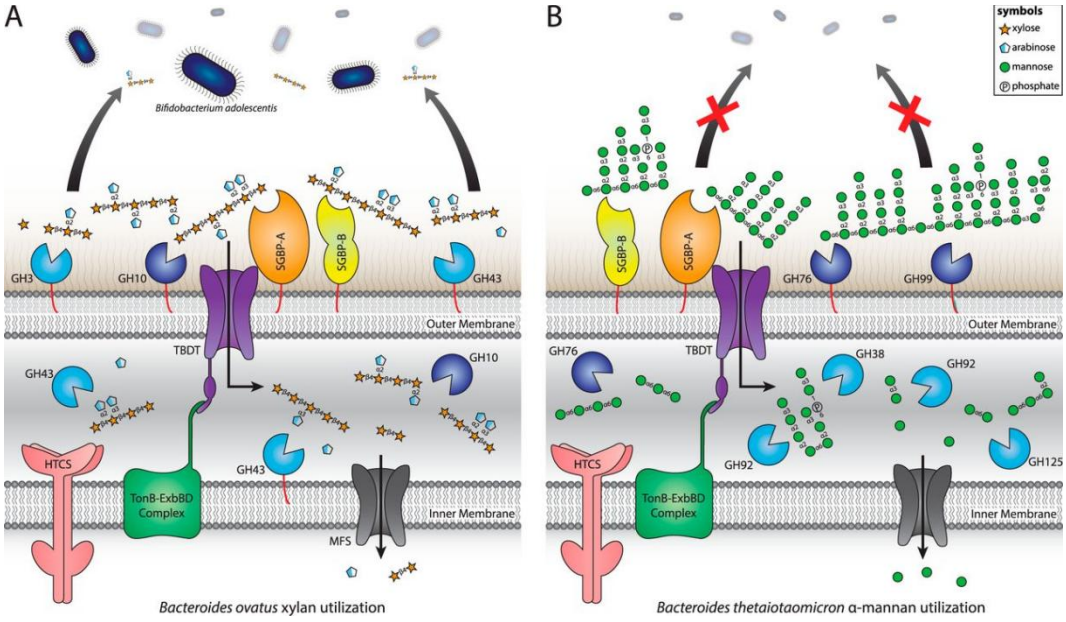


Figure 1.3. Functional modeling of xylan and α-mannan utilization processing in polysaccharide utilization loci (PULs). (A) *Bacteroides ovatus* utilize xylan in a distributive mechanism. (B) *Bacteroides thetaiotaomicron* utilize α-mannan in a 'selfish' mechanism. SGBP: surface glycan-binding protein. TBDT: TonB-dependent transporter. MFS: major facilitator superfamily. HTCS: hybrid two-component system. (The figure is reproduced from Grondin *et al*. (2014) [218]).

It's worth emphasizing that not all TBDTs are SusC-like proteins. Different TBDTs are also present. These TBDTs are most often associated with the uptake of larger compounds, particularly iron siderophore complexes and vitamin B_{12} [229]. Some PUL-like loci lack genes for a SusC-like TBDT and a SusD-like protein. Instead, they feature other TBDTs that have been proposed as carbohydrate utilization containing TonB-dependent receptor (CUT) loci [230]. For instance, *Pseudoalteromonas haloplanktis* ANT/505 utilizes pectin through the action of its CUT locus. The intricate process involves the initial deconstruction of pectin into pectate by the collaborative action of PL1 along with CE8 and PL3 family enzymes. Subsequently, a distinct PL1 family enzyme facilitates the further breakdown of pectate into oligomeric forms, which are transported into the periplasm through a non-SuC-like TBDT and decomposed to galacturonate/unsaturated galacturonate using a GH105 family enzyme [213]. Francis *et al*. suggested that CAZyme-rich gene clusters without TBDTs should also be considered as loci for bacterial

polysaccharide utilization [211]. The prediction of CAZyme genes is mainly based on the Carbohydrate-Active enZYmes Database, which currently encompasses 187 GH families, 43 PL families, and 20 CE families (http://www.cazy.org, data as of 2023/12/06) [231]. Notably, marine algae produce sulfated polysaccharides that are not found in land plants [222]. Genes encoding sulfatase are often found in marine bacteria that can degrade marine polysaccharides, with a particularly high percentage of genes encoding sulfatase in the phyla *Verrucomicrobiota* (e.g., [156, 232, 233]) and *Planctomycetota* (e.g., [234]). SulfAtlas is a database for the annotation of sulfatases, including families and subfamilies, that is maintained by the Marine Glycobiology and ABiMS teams at the Station Biologique de Roscoff (http://abims.sb-roscoff.fr/sulfatlas) [235].

Additionally, there is a case where bacteria utilize polysaccharides, i.e., where *susCD* gene pairs and CAZyme-encoding genes do not exhibit the typical contiguous arrangement, but rather are distantly situated and tend to be regulated by as yet unidentified transcriptional regulators. These are often referred to as 'non-canonical' PULs (e.g., [227, 236]). This form of 'non-canonical' PUL frequently relies on the type IX secretion system (T9SS), a cellular machinery prevalent across the entire *Bacteroidota* phylum that is instrumental for the secretion of specific enzymes and also facilitates gliding motility within this microbial group (reviewed in [215]). T9SS also helps cellulolytic *Cytophaga hutchinsonii* utilize cellulose without TBDTs [237].

Three main mechanisms of polysaccharide uptake have been observed in marine bacteria, named 'selfish', sharing, and cheating/scavenging behaviors [238-242]. The so-called 'selfish' behavior refers to the capture of polysaccharides by bacteria via surface-associated enzymes and their degradation to oligosaccharides, which are then directly introduced into the periplasm using TBDTs (figure B of Figure 1.3). The oligosaccharides are then degraded to monosaccharides and transported to the cytoplasm. Utilizing polysaccharides via 'selfish' behavior is an efficient way and was observed in many bacteroidetes [238]. It was found that selfish bacteria are common throughout the water column of the ocean and widely spread [242-244].

1.7 Interactions between bacteria and micro- and macro-algae beyond polysaccharide utilization

The extracellular substances that algae exude into their surroundings create a layer for interactions with bacteria [113, 245]. This region, which is termed ‘phycosphere’, extends outward from the algal cells to an undetermined distance, and bacterial growth is stimulated by the extracellular substances released by the algal cells [245].

Oxygenic photosynthesis, a pivotal natural process, commenced at least 2.7 billion years ago by primordial cyanobacteria [246], close to 1 billion years after the presumed (and debated) origin of life about 3.4 to 3.5 billion years ago [247]. In the contemporary ocean, eukaryotic phytoplankton predominantly comprises diatoms, dinoflagellates, and coccolithophores [248]. Dinoflagellates first appeared over 650 million years ago [249], diatoms approximately 250 million years ago [250] and coccolithophores are a comparatively young group of around 200 million years of age [251]. Co-evolution between phytoplankton and associated bacteria has hence progressed for millions of years. As a consequence, the relationships between algae and bacteria are manifold and include various mutualistic as well as competitive interactions. For instance, it has been demonstrated that bacteria can provide vitamin B_{12} to algae [252]. Algae lacking nitrogen-fixing capabilities may also require bacteria for the acquisition of reduced organic nitrogen compounds [253]. Bacteria may compete with phytoplankton for inorganic nutrients (e.g. ammonium and nitrate), especially under conditions of mineral nutrient limitation, which could affect the carbon cycle [254].

The symbiotic relationship between algae and bacteria extends beyond microalgae to interactions between macroalgae and bacteria [255]. In the complex marine environment, macroalgae provide a relatively stable and nutrient-rich habitat for bacteria to thrive [256]. Macroalgae in turn regulate the composition of their colonizing bacterial communities, for instance, via releasing quorum sensing (QS) inhibitors [257]. Releasing QS inhibitors is relatively common in the microbial community of macroalgal phycospheres. For example, nearly 40% of bacterial strains isolated from the surface of the brown alga *Fucus vesiculosus* have the ability to degrade acyl-homoserine lactones (AHLs) [258]. Marine macroalgae also participate in microbial community regulation

through the periodic shedding of cells from the biofilm surface layer [259]. Phycosphere bacteria can play a role in regulating the morphological development of their host algae. Studies have revealed that *Roseovarius* sp., *Maribacter* sp., *Algoriphagus* sp., and *Polaribacter* sp. can regulate the morphology of *Fucus vesiculosus* [260]. Some macroalgae-associated bacteria can lead to diseases of their hosts [261]. For instance, one cause of bleaching of the red macroalga *Delisea pulchra* is the release of chemicals by *Ruegeria* sp. R11, resulting in the loss of algal pigments [262, 263]. When large macroalgae become unhealthy, they can also be further damaged by usually symbiotic bacteria that have the ability to degrade macroalgal tissue [264].

1.8 Long-term ecological research (LTER) site 'Kabeltonne' off Helgoland Island

The North Sea is an extension of the North Atlantic and has an area of approximately 574,980 km^2. Helgoland Island (a.k.a. Heligoland in English) is the only German offshore island and is located about 50 km from both the northern shore of Lower Saxony and the western shore of Schleswig-Holstein in the German Bight. The Alfred Wegener Institute (AWI) Helmholtz Centre for Polar and Marine Research operates the BAH ('Biologische Anstalt Helgoland'), a permanent marine station on Helgoland [265], which maintains research at the long-term ecological research (LTER) site 'Kabeltonne' between the main Helgoland island and the minor island Düne (54° 11.3′ N, 7° 54.0′ E) (Figure 1.4, [266]). Water at this LTER site has been sampled daily to determine temperature and salinity since 1873, and microbiological data on colony-forming units (CFUs) have been collected since 1962 [267]. In the following years, new methods were introduced, and since more biological data have been collected. The water body at this site is shallow, with a depth varying between 6 m to 10 m over the tidal cycle, which is why the water body is usually well-mixed [268]. The sediment at the LTER site is sandy [269]. It's very possible to collect suspended sandy particles during filtration of surface water samples, especially during and after strong winds or storms. The dominating wind directions in the German Bight are northwest and southwest [270]. The water around Helgoland Island is transported from the English Channel alongside the Dutch and Frisian coast to Helgoland and is also influenced by coastal water masses and plumes from the Elbe and Weser river estuaries

(~60 to 70 km distance) when northeasterly to easterly winds prevail for some days [127, 266-268]. This leads to a stratification in temperature and salinity and brings in higher concentrations of inorganic nutrients [271].

Data from 1973 to 2006 show that there is an annually recurrent spring phytoplankton bloom at Helgoland Roads [272], typically beginning in early March and ending in late May. These blooms are often dominated by the diatoms *Chaetoceros debilis*, *Chaetoceros minimus*, *Mediopyxis helysia*, *Rhizosolenia styliformis* and *Thalassiosira nordenskioeldii*, the raphidophyte *Chattonella*, the haptophyte *Phaeocystis* and dinoflagellates in terms of cell numbers [123], and by the diatoms *Thalassiosira nordenskioeldii* and *Mediopyxis helysia* and the raphidophyte *Chattonella* when considering algal biomass [123].

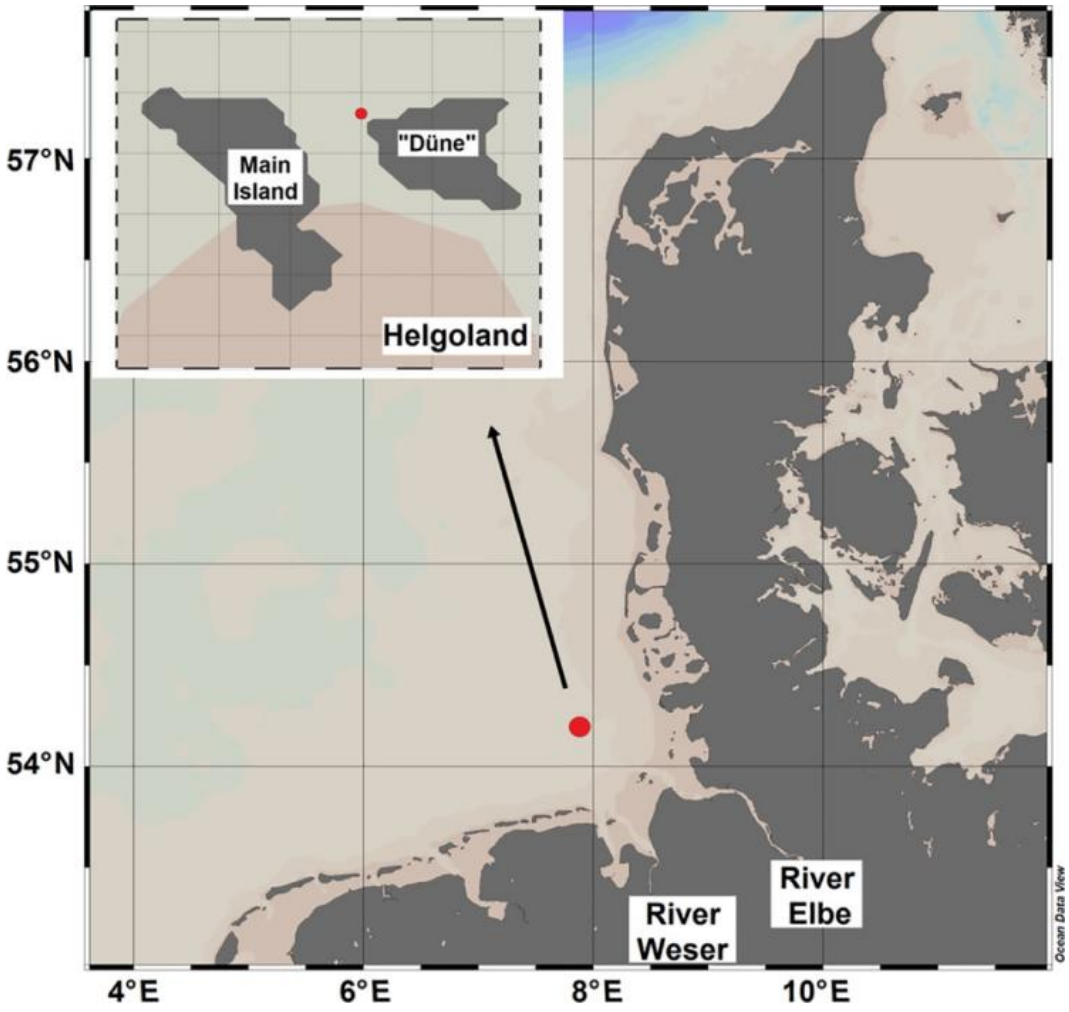


Figure 1.4. Location of the study site Helgoland Roads indicated by the red symbol ('Kabeltonne'; 54° 11.3′ N, 7° 54.0′ E) in the German Bight (North Sea). (The figure is reproduced from Merder *et al*. (2021) [266]).

The Department of Molecular Ecology at the Max Planck Institute (MPI) for Marine Microbiology in Bremen has collected samples during spring blooms at Helgoland Roads since 2009 to study the dynamics of the free-living (FL) bloom-associated bacterial community and their polysaccharide utilization functions *in situ* [123, 127, 211, 273]. These studies have revealed that the associated bacteria are predominantly *Bacteroidetes* (*Flavobacteriia* in particular [123]), *Alphaproteobacteria* (e.g., *Roseobacter* clade members [274]) and *Gammaproteobacteria* (e.g., SAR92 clade members [127]). In *Bacteroidetes,* the bulk of glycan degradation is mediated by few clades, which include 'Formosa', *Cd*. Prosiliicoccus, *Polaribacter*, *Aurantivirga*, *Cd*. Abditibacter, and members of the NS5 marine group [127, 212]. A study on particle-attached (PA) bacterial communities with five time points of each size-fraction during a phytoplankton spring bloom off Helgoland in 2018 revealed that clade BD1-7, *Sulfitobacter*, *Algibacter*, *Rhodococcus*, *Colwellia*, *Psychromonas*, *Winogradskyella* and *Maribacter* were abundant in these communities [111]. However, so far little is known about the temporal dynamics of PA bacterial communities during spring phytoplankton blooms. Likewise, little was known about potential differences in polysaccharide degradation between FL and PA bacteria during such blooms. These questions are addressed in Chapter II of this thesis.

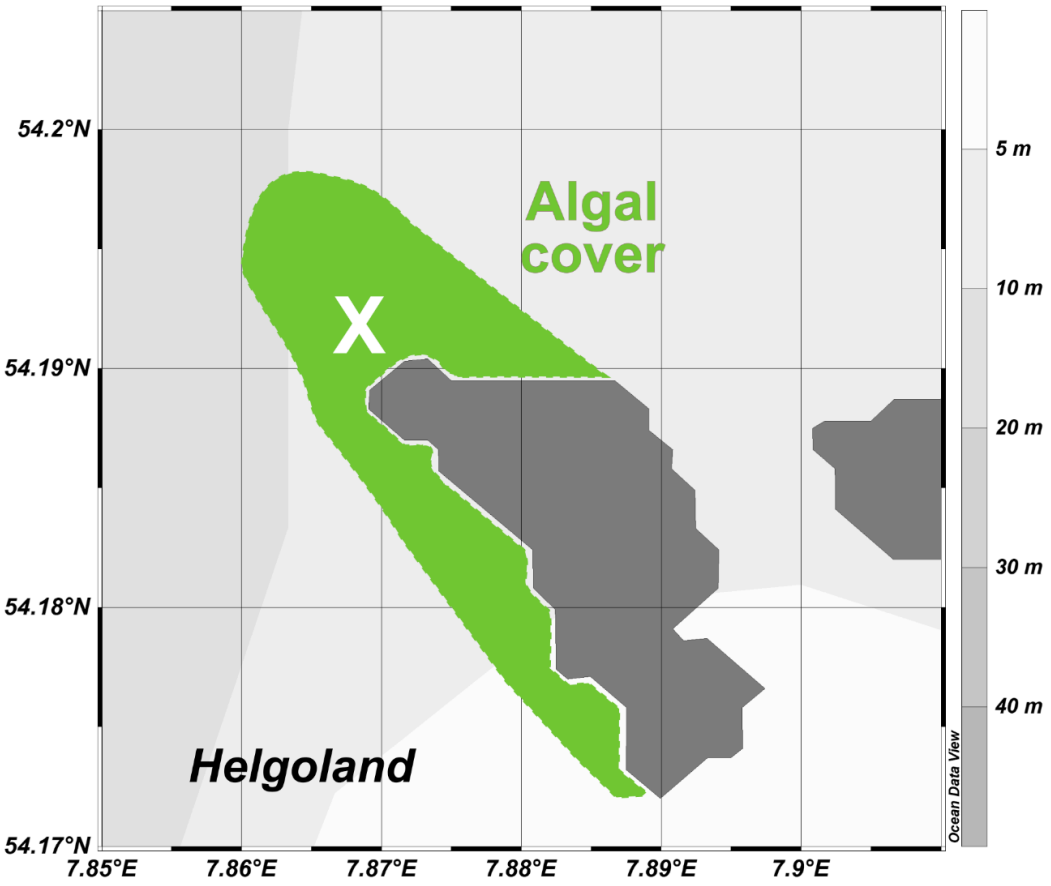


Figure 1.5. The dense macroalgal forests near Helgoland Island are schematically depicted in green. (The figure is reproduced from Bunse *et al.* (2021) [275]).

Additionally, there are considerable amounts of various species of macroalgae near Helgoland Island [268, 275-278], as shown in Figure 1.5. Considering that the distance between the macroalgal forest and the LTER site is not that far, the pool of polysaccharides at the LTER site is also expected to contain certain amounts of macroalgal polysaccharides.

1.9 Research methods in microbial ecology

1.9.1 Cultivation-dependent approaches

Methods for the study of environmental bacteria can be divided into cultivation-dependent and cultivation-independent approaches. The most significant advantage of cultivation-dependent approaches is that the classical isolation and cultivation of bacterial strains

allow for the most reliable and verifiable metabolic analyses. Cultivated strains are also the most accurate way to provide unbiased complete genome sequences that offer precise studies of taxonomic and genetic information [279]. Many insights into ecologically important processes were actually devised on the basis of studies of cultivated bacterial strains. Some more recent examples are studies on rare marine bacteria that degrade complex polysaccharides [156] and the discovery of marine heterotrophic bacteria that can synthesize DMSP [280]. While scientists have put forth diverse strategies to enhance the cultivability of bacteria and have achieved some advancements [281], it remains a challenge, as only a limited fraction of environmental bacteria is cultivable with current techniques, often comprising merely 0.1-1% of the total bacterial population in a given habitat [282].

1.9.2 High-throughput sequencing of 16S rRNA gene amplicons

In recent decades, the exploration of the natural diversity of uncultivated microorganisms has been significantly bolstered by advancements in 16S rRNA gene research. Since Carl Woese's seminal work in 1977, wherein he proposed the utilization of 16S rRNA gene sequencing as markers for investigating bacterial phylogeny [283] (a groundbreaking contribution that revolutionized microbial ecology) substantial progress has been achieved in the technology for sequencing 16S rRNA genes. Notably, the Illumina MiSeq (and, to a lesser extent, MiniSeq) system has emerged as a prominent and widely utilized instrument for sequencing 16S rRNA gene amplicons, firmly establishing itself as the instrument of choice in the contemporary microbial diversity research [284]. For short-amplicon sequencing, the variable regions V1-V2/V3 [285, 286], V3-V4/V5 [287-289], V4-V5 [290] and V4 [291, 292] are the most commonly used targets. A comparative analysis between V3-V4 and V4-V5 primer sets using samples from diverse environments such as sea ice, surface water, marine snow, deep water, and sediment has shown that despite the primer sets exhibiting varying sensitivities to internal diversity, both accurately reflected a similar taxonomic composition down to the rank of genus [290]. The advantage of Illumina sequencing is that it is sufficiently cost-effective and accurate, while one of the disadvantages is that the combined length of paired-end reads is limited to about 500

nucleotides. However, advances in sequencing technology have also enabled the sequence of full-length bacterial 16S rRNA genes directly. Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) are the two prominent technology providers currently spearheading this approach. Both platforms can provide near-complete 16S rRNA gene sequences and have been used in diversity studies (e.g., [293-295]). However, both approaches have the disadvantage of higher sequencing error rates [296]. The accuracy could be increased by using a unique molecular identifier [297] or by post-sequencing error-correction using appropriate software [298, 299].

Taxonomic classification of 16S rRNA gene sequences involves the comparison of each sequence with reference databases [300-302] and clustering-based methods that utilize a sequence similarity threshold to generate operational taxonomic units (OTUs) [303-305]. An alternative approach relies on sequencing error correction algorithms to resolve amplicon sequence variants (ASVs) from 16S rRNA gene amplicon data. This method has demonstrated superior discriminative capabilities in capturing ecological patterns [306-309]. ASV methods have high sensitivity, distinguishing sequence variants differing by even one nucleotide and specificity, separating individual species from others within a genus [310]. DADA2 is a fast and accurate pipeline that infers exact ASVs from amplicon data and reports fewer false positive sequence variants than other methods to report false OTUs [309]. This is why DADA2 was used in Chapters II and III of this thesis.

1.9.3 Meta-omics

The introduction of large-scale sequencing of entire microbial communities (metagenomics) was propelled by the introduction of the first commercial sequencing-by-synthesis (SBS) sequencer, the GS20 that was introduced by 454 Life Sciences in 2005. This marked a significant breakthrough in microbial ecology, enabling a more comprehensive exploration of the taxonomic composition and metabolic potential of uncultivable microorganisms as has been demonstrated in the pioneering Sargasso Sea sequencing project [311]. Presently, the predominant methods for metagenomics include short read sequencing from Illumina (e.g., HiSeq, NovaSeq, and NextSeq platforms) and to a lesser extent BGI (DNBSEQ platforms), long read sequencing from PacBio (Sequel

II and Revio platforms), and sequencing of very long read from ONT (MinION, GridION and PromethION platforms). Illumina short-reads offer advantages, such as a greater sequencing depth and a lower overall error rate [312]. However, the error rate is not uniform but deteriorates with length, and read length on overall is limited to rather short reads of usually 2x250 bp. Long reads, on the other hand, can produce extended sequences but at the price of lower accuracy (Illumina: 99.99%; PacBio: 99.9%; ONT: 99.14% [313, 314]). PacBio circular consensus sequencing (CCS) on the Sequel II platform ensures a low-error rate in high-fidelity (HiFi) reads, albeit at the expense of a shorter length compared to traditional long-read technology [313]. Metagenome-assembled genomes (MAGs) derived from CCS exhibit enhanced integrity with fewer contigs [315], facilitating functional annotation. Additionally, long reads enable the retrieval of a greater number of 16S rRNA genes without the need for primer-based amplification [315]. In this thesis, the quality of MAGs was categorized into high-quality MAGs (>90% intact and <5% contamination), medium-quality MAGs (≥50% intact and <10% contamination), and low-quality MAGs (<50% intact and >10% contamination) with reference to Bowers *et al*. [316].

Metagenome analysis can elucidate the composition of bacterial communities and their potential functions. However, it falls short of revealing which processes are actively expressed. Addressing this gap, metatranscriptomics and metaproteomics represent valuable methods to provide insights into the actively expressed genes and proteins within a community. The integrated analysis of expression and MAG data allows to discern processes that are expressed by individual microorganisms *in situ* and, in longitudinal studies, provides insights into dynamic changes of expression over time.

Metabolomics is a technique for measuring and comparing metabolites produced by cells [317]. These metabolites include various compounds such as sugars, amino acids, lipids, and organic acids. Metabolomics is considered the closest representation of the phenotype [318] and seems to be the endpoint of ‘meta-omics,’ which also includes metagenomics, metatranscriptomics and metaproteomics [319]. Mass spectrometry (MS) is the dominant technology in metabolomics due to its good sensitivity and accuracy [320-

322], but MS also imposes severe limits on the structural complexity and mass of the analytes.

1.10 Aims of this thesis

The primary aim of this thesis was to gain a better understanding of the community composition and ecological roles of marine microalgae and macroalgae-associated microbial communities, specifically of FL and PA bacteria during a diverse spring phytoplankton bloom in the North Sea, and of epiphytic bacteria on the specimen of brown, green and red algae sampled from a coastal reef in Weihai (China). Functional aspects were particularly focused on polysaccharide degradation. These studies and their specific goals can be summarized as follows:

i) Chapter II: Samples were obtained from seawater during the spring phytoplankton bloom of the year 2018 at the LTER site on Helgoland Island, filtered through 0.2-3 μm, 3-10 μm and >10 μm pore-sized filters. Data included physicochemical parameters (e.g., temperature, salinity, nutrients, chlorophyll *a*), microscopic algal biodiversity and biovolume data, total bacterial cell counts for the 0.2-3 μm fraction, 18S rRNA gene amplicon data, 16S rRNA gene amplicon data for FL (0.2-3 μm), PA3 (3-10 μm) and PA10 (>10 μm) bacteria, metagenomics of FL (18 samples), PA3 (16 samples) and PA10 (8 samples) and metaproteomics from 10 selected time points. The goal was to characterize similarities and differences in the structure and dynamics of the FL and PA microbial communities during the spring phytoplankton bloom, dominant taxa during and after bloom, and their potential to degrade polysaccharides.

ii) Chapter III: Samples were obtained from four macroalgal species (red, green, and brown algae), surrounding seawater, and sediments during each of the four seasons from the offshore coast of Weihai, China. The data included 16S rRNA amplicon (92 samples) and metagenome data (23 samples), as well as draft genome sequences of 956 strains (568 candidate new species). The goals were to elucidate the core phycosphere bacteria of the four macroalgae, the correlation of the phycosphere bacteria with seasons and macroalgal species and the

functional characterization of macroalgal phycosphere bacteria focusing on polysaccharide degradation and secondary metabolite production.

iii) Chapter IV: A comparative analysis of the studies in Chapters II and III aims to elucidate the similarities and differences in community structure and function between PA bacteria during microalgal blooms and core phycosphere bacteria of macroalgae.

Chapter II

Particle-attached bacteria act as gatekeepers in the decomposition of complex phytoplankton polysaccharides

Feng-Qing Wang, Daniel Bartosik, Chandni Sidhu, Robin Siebers, De-Chen Lu, Anke Trautwein-Schult, Dörte Becher, Bruno Huettel, Johannes Rick, Inga V. Kirstein, Karen H. Wiltshire, Thomas Schweder, Bernhard M. Fuchs, Mia M. Bengtsson, Hanno Teeling and Rudolf I. Amann

Manuscript published in Microbiome Journal

Contribution of the candidate in % of the total work

Experimental concept and design – 60%

Experimental work/acquisition of experimental data – 0%

Data analysis and interpretation – 75%

Preparation of figures and tables – 90%

Drafting of the manuscript – 60%

Wang *et al. Microbiome* (2024) 12:32
https://doi.org/10.1186/s40168-024-01757-5

Microbiome

RESEARCH

Open Access

Particle-attached bacteria act as gatekeepers in the decomposition of complex phytoplankton polysaccharides

Feng-Qing Wang[1], Daniel Bartosik[2,3†], Chandni Sidhu[1†], Robin Siebers[4], De-Chen Lu[1,5], Anke Trautwein-Schult[4], Dörte Becher[4], Bruno Huettel[6], Johannes Rick[7], Inga V. Kirstein[7], Karen H. Wiltshire[7], Thomas Schweder[2,3], Bernhard M. Fuchs[1], Mia M. Bengtsson[4*], Hanno Teeling[1*] and Rudolf I. Amann[1*]

Abstract

Background Marine microalgae (phytoplankton) mediate almost half of the worldwide photosynthetic carbon dioxide fixation and therefore play a pivotal role in global carbon cycling, most prominently during massive phytoplankton blooms. Phytoplankton biomass consists of considerable proportions of polysaccharides, substantial parts of which are rapidly remineralized by heterotrophic bacteria. We analyzed the diversity, activity, and functional potential of such polysaccharide-degrading bacteria in different size fractions during a diverse spring phytoplankton bloom at Helgoland Roads (southern North Sea) at high temporal resolution using microscopic, physicochemical, biodiversity, metagenome, and metaproteome analyses.

Results Prominent active 0.2–3 µm free-living clades comprised *Aurantivirga*, "Formosa", *Cd*. Prosiliicoccus, NS4, NS5, *Amylibacter*, *Planktomarina*, SAR11 Ia, SAR92, and SAR86, whereas BD1-7, *Stappiaceae*, *Nitrincolaceae*, *Methylophagaceae*, *Sulfitobacter*, NS9, *Polaribacter*, *Lentimonas*, CL500-3, *Algibacter*, and *Glaciecola* dominated 3–10 µm and > 10 µm particles. Particle-attached bacteria were more diverse and exhibited more dynamic adaptive shifts over time in terms of taxonomic composition and repertoires of encoded polysaccharide-targeting enzymes. In total, 305 species-level metagenome-assembled genomes were obtained, including 152 particle-attached bacteria, 100 of which were novel for the sampling site with 76 representing new species. Compared to free-living bacteria, they featured on average larger metagenome-assembled genomes with higher proportions of polysaccharide utilization loci. The latter were predicted to target a broader spectrum of polysaccharide substrates, ranging from readily soluble, simple structured storage polysaccharides (e.g., laminarin, α-glucans) to less soluble, complex structural, or secreted polysaccharides (e.g., xylans, cellulose, pectins). In particular, the potential to target poorly soluble or complex polysaccharides was more widespread among abundant and active particle-attached bacteria.

Conclusions Particle-attached bacteria represented only 1% of all bloom-associated bacteria, yet our data suggest that many abundant active clades played a pivotal gatekeeping role in the solubilization and subsequent degradation

†Daniel Bartosik and Chandni Sidhu contributed equally to this work.

*Correspondence:
Mia M. Bengtsson
mia.bengtsson@uni-greifswald.de
Hanno Teeling
hteeling@mpi-bremen.de
Rudolf I. Amann
ramann@mpi-bremen.de
Full list of author information is available at the end of the article

of numerous important classes of algal glycans. The high diversity of polysaccharide niches among the most active particle-attached clades therefore is a determining factor for the proportion of algal polysaccharides that can be rapidly remineralized during generally short-lived phytoplankton bloom events.

Keywords Algal bloom, Algal polysaccharide, Bacterioplankton, *Bacteroidota*, Carbohydrate-active enzyme, Carbon budget, Carbon cycle, Free-living bacteria, Helgoland Roads LTER, Marine microbes, Particle-attached bacteria, Particulate organic matter, Polysaccharide utilization locus

Background

Global photosynthetic net primary production (NPP) amounts to an estimated 104.9 gigatons of carbon per year [1]. Almost half of this is allotted to algae, in particular, to the small unicellular planktonic algae (phytoplankton) that dominate the world's oceans [2]. Diatoms (*Bacillariophyta*) represent the most prominent phytoplankton group, in particular, in polar and upwelling regions, and have been estimated to fix up to 20 gigatons of carbon annually [3]. It has been suggested that the silicate shells (frustules) of diatoms provide a competitive advantage over other phytoplankton by allowing them to save energy for cytoskeleton maintenance [4]. Further abundant and globally distributed phytoplankton taxa include photosynthetic dinoflagellates and haptophytes (*Haptophyta*), such as coccolithophorids. For the haptophyte genus *Phaeocystis*, it has been shown that their small cells with high surface-to-volume ratios can outcompete diatom productivity under certain conditions [5]. *Phaeocystis* often dominates spring and summer blooms after diatoms have peaked in the coastal North Sea [6], where they can account for up to 65% of the annual primary production [7].

Primary production by marine phytoplankton is not constant but culminates during phytoplankton blooms. Such blooms can be massive, yet they are usually short-lived. Bloom termination is often initiated by nutrient depletion and can be amplified by a number of factors, such as self-shading, grazing (e.g., by copepods), and various infections, e.g., by viruses, algicidal bacteria, parasitic peronosporomycetes (oomycetes), dinoflagellates, and marine fungi [8–11]. Also, the coagulation of algae and increased sinking of the formed particles due to reduced buoyancy can play a role [12].

During phytoplankton blooms, copious amounts of algal organic matter are released as dissolved or particulate organic matter (DOM, POM). Most of this is rapidly remineralized by heterotrophic bacteria and zooplankton, but the exact proportions are a matter of debate. It has been estimated that 62% of the daily phytoplankton production is on average consumed by small zooplankton [13] (reviewed in [14]). Zooplankton sloppy feeding and excretion in turn increase the DOM and POM pools available to bacteria [15], and measurements of bacterial respiration rates have suggested that bacteria remineralize 70–92% of the POM within the mesopelagic zone (−200 to −1000 m) [16]. Only about 1–3% of biological net primary production reaches bathypelagic depths (below −1000 m) [17] via the so-called biological pump, where it can be sequestered for longer periods of time. According to recent estimates, about 10 gigatons of carbon are exported to the deep sea annually, including 1.3 gigatons by the biological pump, 15% of which is phytodetritus [18].

About 95 to > 99% of the epipelagic marine bacteria typically consist of DOM-decomposing free-living (FL) planktonic bacteria (bacterioplankton) and the remainder of POM-decomposing particle-attached (PA) bacteria [19–21]. However, high particle abundances can elevate proportions of PA bacteria, at times possibly even above those of FL bacteria [19]. FL bacteria have been estimated to mediate 53% of the DOM and PA bacteria 50% of the POM fluxes [16]. Likewise, PA bacteria have been shown to exhibit higher per-cell activities (e.g., [20]) and higher proportions of hydrolytic enzymes [22, 23]. However, currently, we have only a poor understanding of the factors that determine the fractions of the organic matter that are remineralized by FL bacteria, PA bacteria, and the fractions that either feed the pool of recalcitrant DOM or sink out to the sea floor.

Depending on the developmental stage and physiological condition, up to 75% [24] or even more [25] of the dry weight of algae can consist of various polysaccharides, e.g., as intracellular stores of biochemical energy and as cell matrix and cell wall components. Many of these polysaccharides have no counterparts in terrestrial plants, in particular, those that are anionic, e.g., due to sulfation. The dominating polysaccharides in marine macroalgae (seaweeds) are well known, such as laminarins, fucoidans, cellulose, and alginates in brown algae (*Phaeophyta*); cellulose, xylans, and ulvans in green algae (*Chlorophyta*); and agars, carrageenans, and galactans (including porphyran and furcellaran) in red algae (*Rhodophyta*). Less is known about microalgal polysaccharides. Brown macroalgae and other stramenopiles, including diatom and raphidophyte phytoplankters, contain laminarin as a store of photoassimilated biochemical energy [26]. Laminarin, which is also used by haptophyte phytoplankters,

is a water-soluble, structurally simple β-1,3-linked helical homopolymer of glucose with occasional β-1,6-branches that typically consist of 20 to 30 monomers [27]. The dry weight of diatoms can consist of up to 35% of laminarin during exponential growth, and even up to 80% has been reported for the stationary phase [28]. Laminarin is therefore one of the most abundant polysaccharides on Earth [29].

The polysaccharides that are encrusted in the siliceous diatom frustules are more heterogeneous, and little is known about their structures. Studies of *Phaeodactylum tricornutum* have identified a sulfated glucuronomannan as a major cell wall component that might be widespread in diatoms [30]. However, monosaccharides other than mannose and glucuronic acid have been identified in frustules, including fucose, galactose, glucose, xylose, rhamnose, and arabinose [31]. Compositions depend on diatom species and physiological state, which would indicate a huge diversity in corresponding structures. Considering the prevalence of diatoms, these polysaccharides are produced in large quantities and play a non-negligible role in global carbon cycling.

Many microalgae also exudate polysaccharide-rich extracellular polymeric substances (EPS). EPS have many functions, e.g., providing a nutritious matrix to attract beneficial bacteria, particularly in the immediate algal phycosphere. EPS also increase cell surface adhesiveness and thereby promote algae aggregation and flocculation [32–34]. Likewise, a portion of the EPS itself can coagulate into more dense transparent extracellular particles (TEP). The amount of EPS that algae produce depends on many factors. It tends to increase when nutrients become limiting, which is commonly interpreted as a mechanism to dispose excess carbon [35] as a substitute for adaptive photosynthesis downregulation. Not much is known about EPS composition, which may vary depending on algal species and physiological conditions, but most EPS seem to contain high proportions of sulfated polysaccharides [36].

Due to the inherent chemical heterogeneity and structural complexity, no bacterium can harbor the genes required to decompose all algal polysaccharides. Instead, bacteria specialize in subsets, which is why the remineralization of algal polysaccharides is a collective endeavor of polysaccharide-degrading bacteria with distinct substrate niches. Genes that code for the polysaccharide degradation machinery in bacterial genomes are often co-located as operons or regulons. Such polysaccharide utilization loci (sg. polysaccharide utilization locus (PUL)) are particularly prominent in the genomes of polysaccharide-degrading *Bacteroidota*, where they typically comprise a *susCD* gene tandem that codes for a SusD-like substrate-binding and for a SusC-like channel protein of a TonB-dependent transporter (TBDT) [37]. These are accompanied by genes coding for degradative carbohydrate-active enzymes (CAZymes), namely glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), and by accessory genes coding for, e.g., surface glycan-binding proteins (e.g., [38]), sulfatases, ABC transporters, and other associated functions. PUL lengths depend on the target substrate and can vary considerably. While a typical laminarin PUL consists of around 20 genes (e.g., [39]), PUL-rich loci can also encompass close to 100 genes (e.g., [40]).

We have analyzed the microbial response of bacteria to spring phytoplankton blooms in a series of studies at the long-term ecological research (LTER) site "Kabeltonne" off Helgoland Island in the southern North Sea [39, 41–44], in which we focused on the response of FL (0.2–3 µm) bacteria and their associated polysaccharide niches. Recently, we could exemplarily show that abundant bloom-associated FL bacterioplankton clades preferentially consume water-soluble, low-complexity storage polysaccharides such as laminarin and α-glucans, which therefore exert a strong community structuring effect [39]. Also, other polysaccharides, such as alginate or mannose-containing polysaccharides, play a role, albeit in lower quantities [39]. An unknown proportion of the dissolved polysaccharides originate from POM that has been solubilized by PA bacteria and diffused away before uptake. However, so far, little is known about the involved polysaccharide-degrading PA bacteria and their connection to FL bacteria. A recent comparative study of FL and PA metagenome-assembled genomes (MAGs) from different water depths in the North Pacific Subtropical Gyre has shown that PA bacteria are characterized by higher predicted growth efficiencies and, on average, larger genomes with higher proportions of genes for peptidases, CAZymes, secretion, sensing and motility [23].

In this study, we investigated a diverse spring phytoplankton bloom that took place in 2018 off Helgoland Roads at high temporal resolution (51 sampling dates over a 90-day period). We aimed to disentangle the roles of PA bacteria in comparison with FL bacteria with respect to their potential to degrade phytoplankton-derived polysaccharides. We collected microscopic algal biodiversity and biovolume data, eukaryote 18S rRNA gene amplicon data, and 16S rRNA gene amplicon data of bacterial communities from FL (0.2–3 µm), PA3 (3–10 µm), and PA10 (> 10 µm) filter fractions together with a broad range of physicochemical data. In addition, we performed metagenomics of FL (18 samples), PA3 (16 samples), and PA10 (8 samples) bacterial communities, reconstructed MAGs of abundant key players, and compared their polysaccharide degradation potentials. These data were complemented by metaproteomes from

10 selected time points during the bloom to link bacterial protein to the decomposition of algal glycans.

Results

The 2018 Helgoland spring phytoplankton bloom was diverse and polyphasic

Based on microscopically determined phytoplankton taxa, corresponding biovolume estimates (< 0.1 to 1.71 mm^3 L^{-1}, Additional file 1: Table S1), and chlorophyll *a* measurements (~ 2 to 33.8 units, Additional file 1: Table S1), the 2018 Helgoland spring bloom consisted of a pre-bloom phase dominated by lowly abundant diatoms and *Phaeocystis* sp. haptophytes (March 1 to April 9), a diatom-dominated phase (April 10 to May 08) largely overlapping with a notable bloom of *Chattonella* raphidophytes (April 19 to May 11), and a late phase dominated by *Phaeocystis* sp. haptophytes and few *Dinophyceae* (May 09 to May 31) (Fig. 1).

Diatoms comprised various *Chaetoceros* species and *Thalassiosira rotula*. After they went into decline, *Phaeocystis* sp. and *Dinophyceae* numbers increased, with *Phaeocystis* sp. becoming dominant until the first wave of blooming algae ended about 1 week into June. A remarkable correlation was obtained between chlorophyll *a* measurements and estimated biovolumes of photosynthetic plankters (Additional file 2: Fig. S1A). Additional non-photosynthetic plankters comprised in particular dinoflagellates, e.g., *Noctiluca scintillans*. The latter was detected at the end of May and, despite being low in numbers, dominated the biovolume of unicellular eukaryotic plankters due to large cell sizes (Additional file 1: Table S1, Additional file 2: Fig. S1B).

Analysis of the 15 most abundant 18S rRNA gene amplicon sequence variants (ASVs) largely supported microscopic observations (Additional file 2: Fig. S2). For example, the *Chaetoceros* bloom was detected in the PA10, and the *Phaeocystis* and *Noctiluca* blooms in the PA3 fractions (*Phaeocystis* cells are small, and *Noctiluca* cells are fragile and thus broke during filtration). One inconsistency was that *Chattonella* could not be detected, likely because their particularly fragile, large, wall-less cells disintegrated during filtration. In addition, 18S rRNA ASV data revealed a noteworthy peak of *Cryothecomonas* nanoflagellates towards the end of the diatom bloom.

We focused on the period from March 1 to May 31. FL bacterial total cell counts (TCC) increased continuously from 0.5×10^9 L^{-1} on March 1 to a peak of 3.3×10^9 L^{-1} on May 24 (Fig. 1). This increase was gradual during the pre- and main bloom phases and progressed more

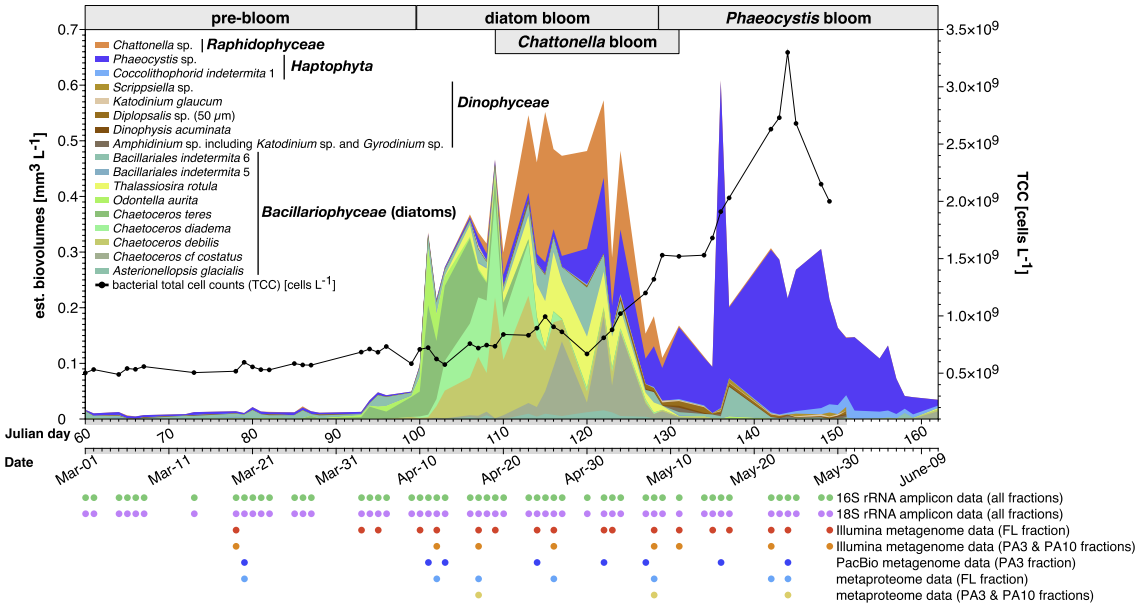


Fig. 1 2018 spring phytoplankton bloom at Helgoland Roads and associated datasets. Estimated biovolumes of abundant phytoplankton taxa (stacked colored areas) as assessed by microscopy and of total cell counts (TCC) of DAPI (4′,6-diamidino-2-phenylindole)-stained bacteria (black line). Complete data was sampled until the end of May. Additional data for June is shown to visualize the end of the bloom but does not include TCC and *Dinophyceae* biovolume information. The core sampling period from March 1 until May 31 is indicated by a gray area on the abscissa. Additional data including abundances of non-photosynthetic plankters are provided in Additional file 2: Fig. S1 and Additional file 3. Associated omics sampling dates are indicated by colored circles at the bottom

rapidly after diatoms peaked at the end of April. Likewise, flagellate numbers increased throughout the bloom, ranging from 2.4×10^6 L^{-1} on March 1 to 1.2×10^7 L^{-1} on May 16 (Additional file 2: Fig. S3A). Flagellate numbers correlated well with Chl *a* and biovolume estimates, indicating that flagellates not only preyed on bacteria but also on microalgae. The remaining zooplankton was dominated by various copepod species with undulating abundances over time that showed no clear correlation to phytoplankton data, possibly due to vertical migration in and out of the sampled surface water (Additional file 1: Table S2).

An influx of nutrient-rich coastal water triggered the onset of the bloom

Physicochemical data indicated an incursion of nutrient-rich coastal waters at the onset of the diatom bloom, as on April 10 nitrate concentrations spiked to 19.0 µM, silicate concentrations spiked to 10.7 µM, and salinity decreased from 33.8 to 32.6 (Additional file 2: Fig. S3B-D). A second influx likely occurred from May 22 to 29 and was accompanied by an increase in silicate concentrations from 1.0 to 4.0 µM and a drop of salinity to 31.7. A spike in phosphate concentrations from 0 to 0.7 µM was also detected during this period (Additional file 2: Fig. S3E).

Wind directional data (Additional file 1: Table S3) supported these influx events, since northeasterly to easterly winds dominated from April 9 to 14 and from May 23 to 30 (Additional file 2: Fig. S4; see [39] for details). Additional rain and sunshine data are provided in Additional file 1: Table S4.

FL and PA bacterial communities exhibited distinct diversities and compositional shifts over the bloom's progression

16S rRNA gene amplicon sequencing of 153 samples from FL and PA fractions yielded 24,356 unique ASVs (Additional file 1: Table S5). Good's coverage (a measure for the proportion of singletons) indicated that this was adequate to capture basically all of the diversity of the FL communities (avg. coverage, ~ 1.0) and most of the diversities of both PA fractions (avg. coverages, 0.96 and 0.91, respectively) (Additional file 2: Fig. S5A). FL bacterial communities had significantly lower alpha diversity indices (Chao1, Simpson's, Shannon) than PA3 and PA10 communities (ANOVA, $p < 0.001$), whereas the difference between both PA communities was less pronounced (Additional file 2: Fig. S5B-D). Shannon indices exhibited distinct patterns over time (Fig. 2A, B), with a high correlation of PA3 and PA10 samples ($p < 0.0001$, Fig. 2A). FL Shannon indices were highest prior to the pre- and early diatom-bloom, decreased and stayed low during the main diatom and early *Phaeocystis* bloom phases, and finally increased again towards the bloom's end in the late *Phaeocystis* bloom. No such trend was observed for PA communities (Fig. 2A, B).

NMDS analyses based on weighted UniFrac distances corroborated that PA3 and PA10 communities were more alike and FL communities more distinct. Pre-bloom communities grouped well and were distinct from main and late-bloom communities (Fig. 2C, left to right). While these differences were less pronounced between the two main bloom phases, they were still detectable. The average distances between the pre-bloom and the main bloom stages were also smaller in the FL than in both PA fractions (Fig. 2C), indicating that the bloom caused a more profound community change in both PA fractions.

Distinct bloom phases selected for distinct genera in all size fractions

Both, FL and PA communities showed clear temporal successions of distinct bacterial clades. PA communities, however, were not only more diverse but also dominated by different taxa and exhibited more dynamic compositional shifts (Fig. 3, Additional file 1: Table S5). While FL communities were dominated by *Alphaproteobacteria*, proportions were lower within PA3 and PA10 communities (Additional file 2: Fig. S6A). In contrast, *Gammaproteobacteria* exhibited particularly high relative abundances in PA3 and PA10 communities but less so in FL communities. Likewise, *Verrucomicrobiota* and *Planctomycetota* exhibited higher relative abundances in PA3 and PA10 than in FL communities, whereas *Bacteroidota* were ubiquitous in all samples (Additional file 2: Fig. S6A). *Flavobacteriaceae* accounted for similar percentages in all fractions before May 4 (FL, 7–25%; PA3, 4–19%; PA10, 5–24%, Fig. 3). Afterwards, *Flavobacteriaceae* proportions increased rapidly during May 4 to May 8 and May 15 to May 29 in the FL (up to 30%) but not in both PA fractions (up to 18%) (Fig. 3). *Cryomorphaceae* relative abundances were higher in FL than in PA communities, while it was the opposite for *Saprospiraceae* (Fig. 3).

As reported for FL bacterioplankton sampled at Helgoland Roads in previous years [41, 42], SAR11 clade Ia, *Planktomarina*, and *Amylibacter* accounted for a substantial fraction of ASVs. These three alphaproteobacterial clades had high relative abundances in all FL samples but exhibited lower relative abundances in PA3 and were even rare in PA10 samples (Additional file 2: Fig. S7A). Members of alphaproteobacterial unclassified *Stappiaceae* were thriving in PA3 during the late *Phaeocystis* bloom, while alphaproteobacterial *Sulfitobacter* simultaneously ramped up in PA10 samples.

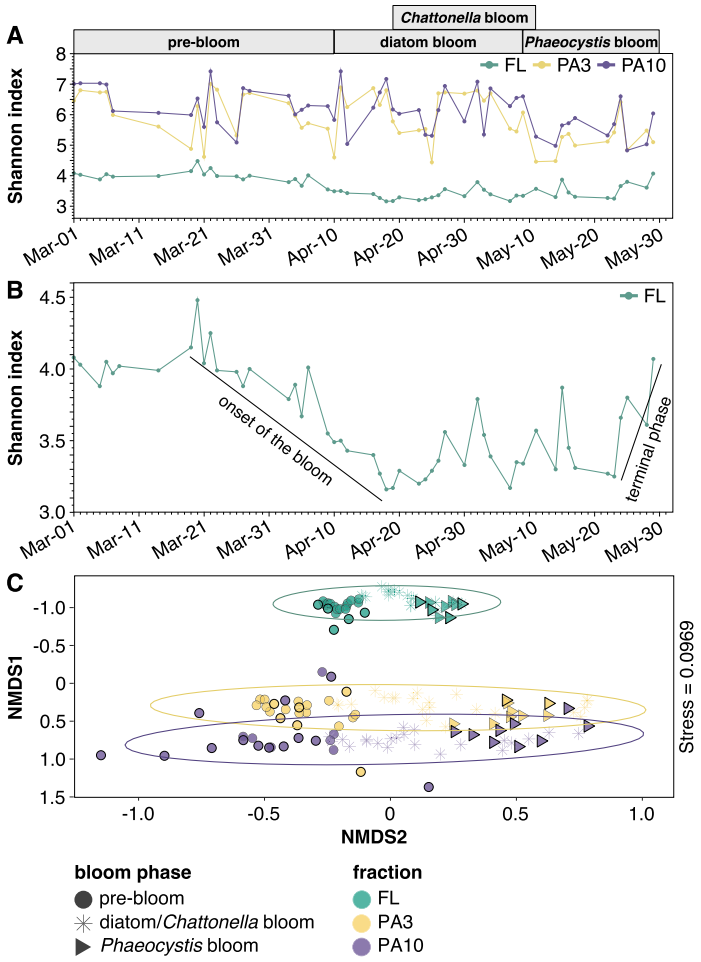


Fig. 2 Diversity as assessed by 16S rRNA gene amplicon data. **A** Shannon index of the sampled bacterial communities over time. Colors correspond to sample fractions (FL, 0.2–3 µm; PA3, 3–10 µm; PA10, > 10 µm). **B** Shannon index of the FL fraction over time. **C** Non-metric multidimensional scaling (NMDS) analysis of the bacterial community of all three fractions. Colors correspond to sample fractions and shapes to bloom phases

During the late diatom and *Phaeocystis* bloom phases, the FL bacterial community consisted primarily of *Bacteroidota*, including *Cd.* Prosiliicoccus [45], *Aurantivirga*, "Formosa", and members of the NS3a and NS5 marine groups, *Gammaproteobacteria* including SAR92 and unclassified *Nitrincolaceae*, as well as a distinct group of *Verrucomicrobiota* including *Lentimonas* (Additional file 2: Fig. S7A-B). Members of *Cd.* Prosiliicoccus, the NS5 marine group, and *Aurantivirga* were also detected in the PA fractions but with lower relative abundances, some of which were probably due to carryover during fractionating filtration (Additional file 2: Fig. S7A). "Formosa" was present with similar low overall maximum relative abundance in FL and PA fractions (Additional file 2: Fig. S7B). *Algibacter* was mostly detected in the PA fractions and increased during the main bloom phases in PA3 communities (Additional file 2: Fig. S7B).

Polaribacter was not as abundant in FL communities as in previous [42] or later [39] years. More *Polaribacter* and unclassified *Saprospiraceae* were detected in both PA than in FL communities. *Maribacter* and *Winogradskyella* [21] were thriving during the late bloom phases but only in PA10 communities (Additional file 2: Fig. S7B).

During the bloom, *Gammaproteobacteria* had higher relative abundances in PA than in FL communities. For instance, in comparison with FL communities, members of the BD1-7 clade and *Colwellia* had higher relative abundances in PA communities during the diatom and *Phaeocystis* bloom phases, while unclassified

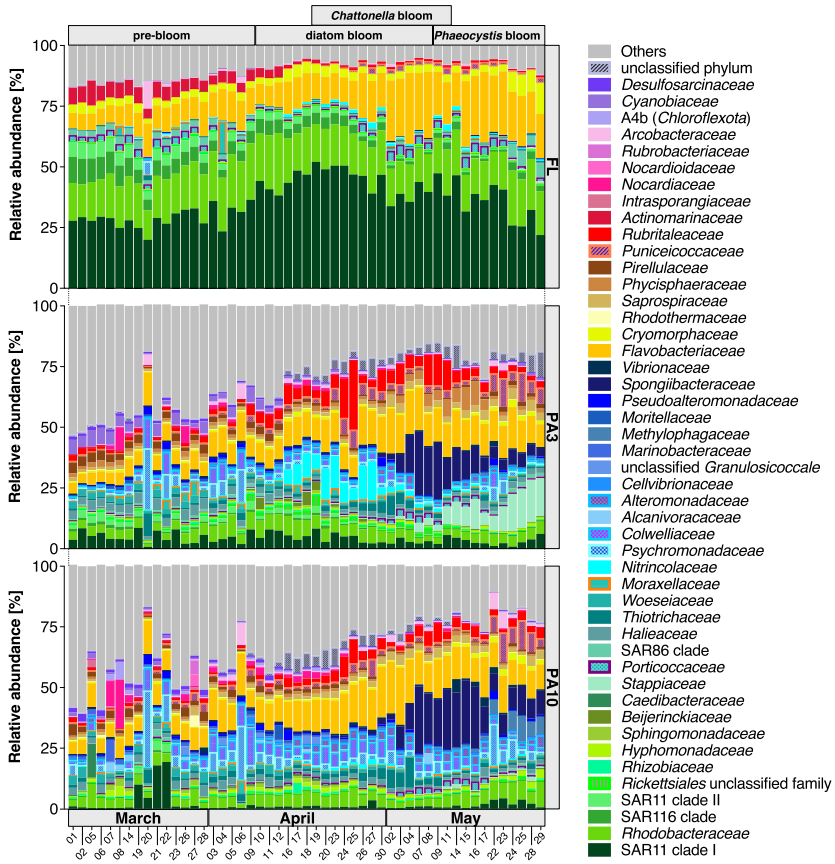


Fig. 3 Compositional differences among fractions as assessed by 16S rRNA gene amplicon data. Relative abundances of families are shown that belong to the topmost five abundant in at least two samples. The remaining families are subsumed as "Others." Details are provided in Additional file 1: Table S5

Nitrincolaceae had higher relative abundances in PA3 communities (Additional file 2: Fig. S7A). Likewise, unclassified *Methylophagaceae* and *Glaciecola* exhibited higher relative abundances in PA10 communities (Additional file 2: Fig. S7A-B).

Furthermore, *Persicirhabdus* (*Verrucomicrobiota*) were proportionally more abundant in PA communities during the diatom bloom, and members of CL500-3 (*Planctomycetota*), known to be also associated with blooming freshwater algae [46], were proportionally more abundant in PA3 communities during the late diatom and *Phaeocystis* bloom phases. Similar bloom-associated temporal dynamics were also discernible in several groups that were not among the selected topmost genera. For instance, *Arenicella* and "Formosa" members were present only during the late diatom and *Phaeocystis* bloom phases in all fractions (Additional file 2: Fig. S7B). Likewise, members of the SUP05 cluster and NS7 marine group were present during the pre-bloom and early diatom bloom phases in FL communities, while at the same time, members of the DEV007 clade (*Verrucomicrobiota*) were present in PA3 communities (Additional file 2: Fig. S7B). Further composition dynamics at the ASV level are provided in Additional file 3.

FL and PA community members exhibited distinct CAZyme composition dynamics

We selected 42 samples for metagenome sequencing, namely 18 FL (Illumina), 16 PA3 (Illumina, 8; PacBio, 8), and 8 PA10 (Illumina) samples (Additional file 1: Table S6). The resulting metagenomes amounted to 1.6 Tbp raw sequences. *K*-mer-based metagenome composition analyses corroborated significant differences between fractions (PERMANOVA, $p = 0.003$; Additional file 2: Fig. S8A). Distances between the pre-bloom and the two bloom periods were closer for FL than for PA data, corroborating 16S rRNA gene amplicon-based NMDS analyses (Fig. 2C). Individual assemblies of all

metagenomes yielded 18.2 Gbp with 2.5 kbp minimum length (Additional file 1: Table S6).

We computed and compared CAZyme relative frequencies in assembled Illumina metagenome data over time (eight samples of each fraction, Additional file 1: Table S7). Overall, genes targeting β-1,3-glucan (laminarin) were most frequent, with peaking relative frequencies towards the end of the diatom bloom (Fig. 4A). Respective genes were dominated by *Bacteroidota* and *Gammaproteobacteria* in all fractions, whereas *Verrucomicrobiota* (more prominent in FL fractions) and *Planctomycetota* (more prominent in PA fractions) contributed only little (Additional file 1: Table S7). These data suggested an overall increase of laminarin-consuming bacteria when the diatom bloom collapsed, most notably in PA3 communities. This corroborates recent data from FL bacteria during the 2020 Helgoland spring bloom, where laminarin PULs were the most frequent and highest expressed of all PULs [39]. In terms of gene compositions, β-glucan PULs comprised the previously described variant-1 [39] coding for GH149, GH17, GH16, GH158 and GH30 enzymes (including variations), variant-2 coding for GH16 or GH17 and GH3 enzymes [39], and a PUL type coding only for GH16 enzymes (Additional file 2: Fig. S9).

Genes targeting α-glucans exhibited the second highest relative frequencies and exhibited no discernible trend (Fig. 4B). For the most part, respective genes were more frequent in PA than in FL communities. More *Gammaproteobacteria* and *Planctomycetota* contained these genes in PA than in FL communities. Four types of α-glucan PULs were present: type I coding for only one or more GH13 enzymes; type II coding for GH13, GH65, and sometimes an additional GH31 enzyme; type III coding for GH13, GH77, and GH57 enzymes; and type IV coding only for GH13 and GH31 enzymes (Additional file 2: Fig. S9). Genes targeting alginate were also

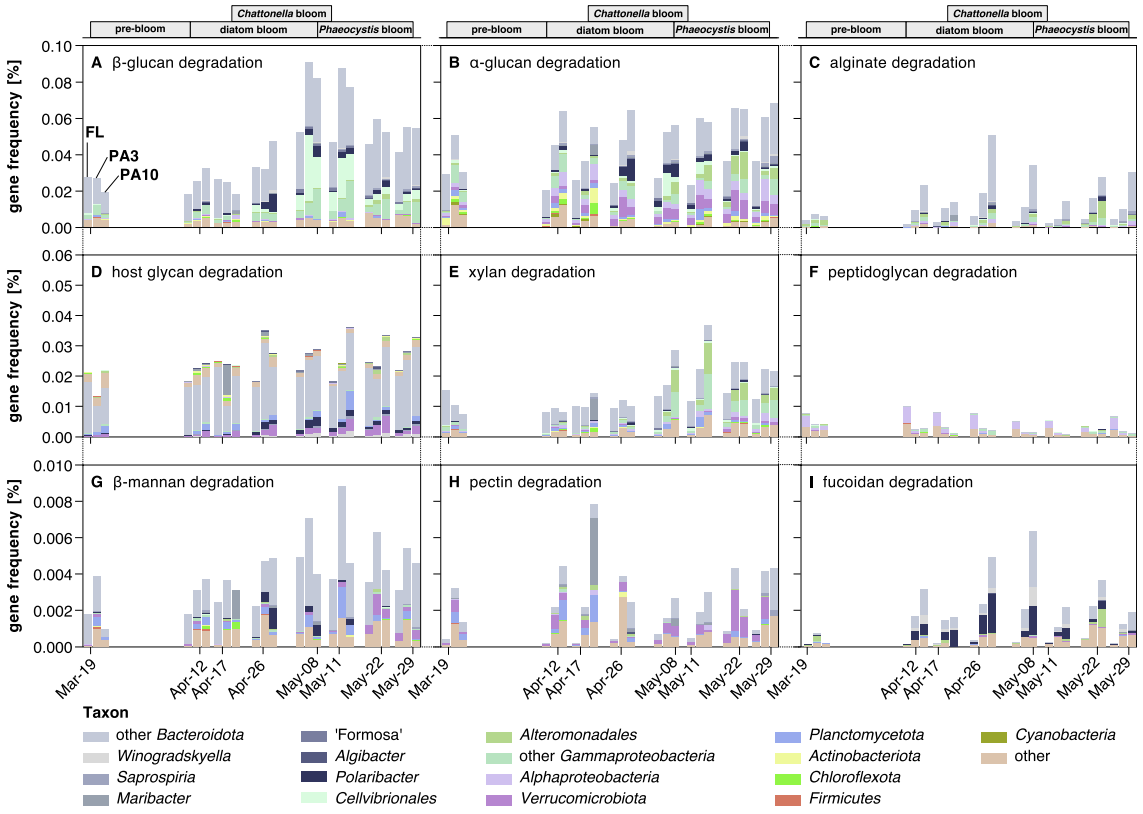


Fig. 4 Frequencies of CAZyme genes attributed to the degradation of specific polysaccharide substrates. CAZyme genes in metagenomes and corresponding substrates were predicted using the dbCAN3-sub database. Gene frequencies were calculated as follows: frequency = Σ(average coverage of target genes) × 100/Σ(average coverage of all genes). Only dates with data for all fractions were plotted as stacked bar charts (left to right: FL, 0.2–3 µm; PA3, 3–10 µm; PA10, > 10 µm). Colors represent dominating taxa. Additional data for α-rhamnosides, chitin, arabinans, α-mannans, cellulose, and sialic acids is depicted in Additional file 2: Fig. S10, and complete data is summarized in Additional file 1: Table S7

frequent, with notable higher proportions in PA communities, particularly in PA10 (Fig. 4C).

Relative frequencies of host glycan degradation genes, e.g., genes targeting eukaryotic N-glycans, showed no trend in FL communities but were for the most part higher in PA communities, where they increased during the diatom and *Phaeocystis* bloom phases (Fig. 4D). Respective genes attributed to unclassified *Bacteroidota*, *Polaribacter*, *Verrucomicrobiota*, and *Planctomycetota*, with the latter preferring PA10 fractions.

Xylan degradation genes were rarer. Their relative frequencies ramped up in PA communities after the diatom bloom abated, in particular, in PA3 with high proportions of *Alteromonadales* and other *Gammaproteobacteria* during the *Phaeocystis* bloom (Fig. 4E). Genes targeting peptidoglycan (murein) were notably more frequent among FL than PA bacteria. Proportions were highest during the pre- and diatom bloom stages, and lower during the late bloom (Fig. 4F).

For α-mannans, there were no consistent differences between FL and PA communities (Additional file 2: Fig. S10D). Frequencies were highest during the early diatom bloom phase and leveled off towards the end of the *Phaeocystis* bloom. In contrast, proportions of genes for β-mannan degradation were often highest among PA3 bacteria, in particular, during the diatom to *Phaeocystis* bloom transition phase (Fig. 4G). A transition in α-mannan degradation from *Flavobacteriales* to other *Bacteroidota* was observed before, during, and after the bloom (Additional file 2: Fig. S10D), whereas bacterial communities harboring β-mannan degradation genes were dominated by *Planctomycetota*, *Verrucomicrobiota*, and *Bacteroidota* during the main and late bloom stages (Fig. 4G).

Proportions of genes targeting pectins (Fig. 4H) were notably higher in both PA fractions and increased during the diatom and late *Phaeocystis* blooms. During the diatom and *Phaeocystis* bloom phases, fucoidan degradation genes had much higher proportions in PA10 than in PA3 or FL communities, which coincided with a notable increase in the proportions of *Polaribacter*, *Maribacter*, and other *Bacteroidota* (Fig. 4I).

Genes for the degradation of sialic acids were more frequent in PA3 communities but ramped up in FL communities after the diatom bloom abated (Additional file 2: Fig. S10F). However, during the *Phaeocystis* bloom, sialic acid degradation potential appeared to have shifted towards PA communities. This shift was characterized by an increase in *Planctomycetota* and unclassified *Bacteroidota*, along with the emergence of *Saprospiria* (Additional file 2: Fig. S10F). Genes for the degradation of α-rhamnosides, chitooligosaccharides, arabinans, and cellulose were notably more frequent in both PA fractions and increased as the bloom progressed (Additional file 2: Fig. S10A-C, E).

Representative PA and FL community MAGs

From all 42 assembled metagenomes, we reconstructed 1944 initial bins (Illumina, 1721; PacBio, 223). Manual refinement retained 146 MAGs that fulfilled the MIMAG high-quality (HQ) criteria (> 90% completeness; < 5% contamination; presence of 23S, 16S, and 5S rRNA genes; and ≥ 18 tRNAs) [47], 964 MAGs (Illumina 895; PacBio, 69) of at least medium quality (MQ) (≥ 50% completeness, < 10% contamination), and 399 Illumina MAGs that did adhere to the "near complete" category by Almeida et al. [48] and were thus treated like HQ MAGs in downstream analyses (Additional file 1: Table S8, Additional file 2: Fig. S11). Dereplication of these in total 1509 MAGs at 95% ANI yielded 305 species-level MAGs of 16 known phyla (Additional file 1: Table S8, Additional file 2: Fig. S12), including 139 (45.7%) HQ MAGs. The average number of recovered MAGs per sample increased with decreasing filter pore size in fractionating filtration (PA10, 20; PA3, 26; FL, 57), due to the decreased capture of eukaryotic biomass [49]. The average size of HQ MAGs was larger in PA than in FL communities (Additional file 2: Figs. S13-14). Sizes of abundant MAGs ranged from 0.8 to 8.6 Mbp in PA and from 0.6 to 2.9 Mbp in FL communities (Fig. 5). Details are provided in Additional file 1: Table S9 and Additional file 3.

Based on 16S rRNA gene amplicon data (Additional file 1: Table S5), we selected 40 high-abundance genera (Additional file 1: Table S10), 39 of which were represented by corresponding MAGs. MAGs from FL community samples included previously identified relevant clades at Helgoland Roads, such as *Aurantivirga* [44], *Polaribacter* [50], "Formosa" species Hel1_33_131 [51], *Cd.* Prosiliicoccus [45], the NS4 and NS5 marine groups [52], *Amylibacter*, and the SAR11 Ia, SAR92, and SAR86 clades. MAG abundances based on read frequencies were also not unusual compared to earlier years with one notable exception. We detected *Polaribacter* clade 2-b [50] (Additional file 2: Fig. S15), but in agreement with ASV data, its maximum relative abundance of < 0.1% was much lower than in previous observations, with recorded maxima of 14.7% (2010), 19.8% (2011), and 34.5% (2012) [53].

More *Planctomycetota* (FL, 2; PA3, 9; PA10, 2) and *Chitinophagales* (FL, 2; PA3, 7; PA10, 5) MAGs were obtained from PA3 and PA10 metagenomes, whereas more alphaproteobacterial MAGs were obtained from FL metagenomes (FL, 31; PA3, 22; PA10, 23). Also, the genus diversity of *Flavobacteriaceae* MAGs was higher in PA than in FL communities (FL, 12; PA3, 17; PA10, 14). Based on ASV and MAG abundance data, we categorized

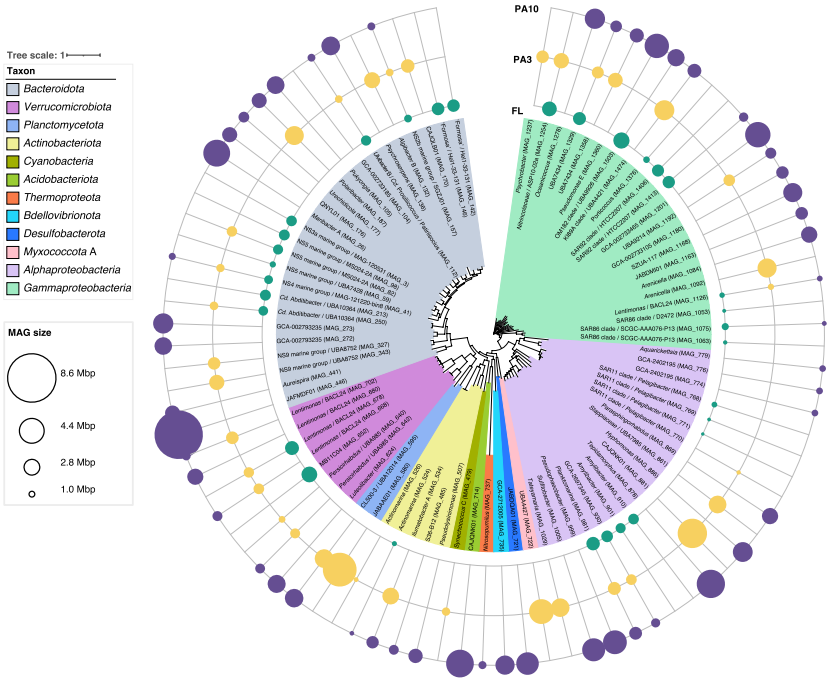


Fig. 5 Sizes of the ten topmost abundant MAGs across each sample ($n = 136$). De-replication was carried out within each fraction, and MAGs belonging to the top ten in terms of abundance (as determined by MAG abundance) were specifically chosen for this analysis. The central maximum-likelihood tree of MAGs was computed in anvi'o v7.1 based on protein sequences of 38 universal single-copy genes. Circle area sizes correspond to MAG sizes with colors representing the fractions from which the MAGs were retrieved

MAGs into those that were most abundant in either FL or PA communities and those that were abundant in both. In total, 152 MAGs belonged to the abundant PA MAGs (Additional file 1: Table S11), including *Polaribacter*. Further details are provided in Additional file 3.

Novel PA community MAGs

Comparison of all MAGs to those obtained for FL bacteria during the 2010 to 2016 spring blooms [43, 44] identified 100 species-level PA MAGs (HQ, 42; MQ, 58) that were uniquely obtained in 2018 (Additional file 1: Table S8, Additional file 2: Fig. S16). According to GTDB r207_v2, 76 of these MAGs represented novel species (Additional file 1: Table S8).

The 100 PA MAGs comprised a high proportion of over 3.5 Mbp (38/100 as compared to 57/305 for all species-level MAGs) (Additional file 2: Fig. S16C). They were dominated by *Paraglaciecola* (MAG_1218), *Pseudoalteromonas* (MAG_1211 and 1212), and UBA12014 (CL500-3, *Planctomycetota*, MAG_591) in PA3 and GCA-002793235 (*Vicingaceae*, MAG_272), GCA-002733465 (*Kangiellaceae*, MAG_1201), *Maribacter* A (MAG_26), *Polaribacter* (MAG_189), and *Pseudolysinimonas* (MAG_507) in PA10 communities (Additional file 2: Fig. S12, Additional file 3). We also obtained HQ MAGs of *Acidobacteriota* and *Chloroflexota*, plus seven of ten *Planctomycetota* MAGs and four *Polaribacter* MAGs representing species that we did not identify at Helgoland Roads before.

Few particularly active but distinct MAGs dominated FL and PA communities

Seven sampled FL metaproteomes yielded 43,750 unique proteins (Additional file 1: Table S12). A total of 15,906 of these proteins were assigned to 177 FL MAGs, with up to 16.3% SusC- and SusD-like proteins and various TBDTs (Additional file 2: Fig. S17, Additional file 3). Such high proportions agree with previous metaproteome studies on bloom-associated bacterial communities [41, 43, 54]. Conversely, only 5018 proteins were obtained from three PA3 and PA10 metaproteome sampling dates, which were dominated by eukaryotic proteins (42.7 to 64.0%). Just 932 of these proteins could be assigned to bacterial MAGs (Additional file 1: Table S12, Additional file 3), which is why we used these data only to pinpoint the most active PA MAGs (Additional file 2: Fig. S18).

Protein abundance data corresponded well with calculated MAG abundances from corresponding Illumina

metagenomes (Additional file 2: Fig. S19). Abundant MAGs with high overall protein expression on all seven FL sampling dates comprised members of the NS4 marine group, *Planktomarina*, and the OM182 and SAR11 clade Ia clades. The highest overall expression was observed in a *Nitrincolaceae* ASP10-02a clade MAG, but only during the diatom bloom phase (Fig. 6). The expression of SusC-like proteins is indicative of oligosaccharide uptake in *Bacteroidota*. MAGs with high SusC-like protein expression during the diatom and *Phaeocystis* bloom phases comprised members of *Cd.* Prosiliicoccus, *Aurantivirga*, the NS3a, NS5, and NS4 marine groups, *Cd.* Abditibacter, and the *Cyclobacteriaceae* clade UBA4465 (Fig. 6). Other MAGs expressed SusC-like proteins only during distinct bloom phases, such as during the diatom bloom phase (other members of the NS5 marine group), the pre- and diatom bloom phases (a member of the NS2b marine group), the late diatom and *Phaeocystis* bloom phases (members of "Formosa"), or only the *Phaeocystis* bloom phase (again a member of the NS5 marine group).

Apart from bacteroidotal SusC-like proteins, TBDTs for the uptake of larger substrates, possibly including oligosaccharides, were predominantly expressed by members of various gammaproteobacterial clades. MAGs of the OM182, SAR92, and SAR86 clades exhibited high TBDT expression during all sampling dates, while others showed such expression mostly during the *Phaeocystis* bloom, e.g., MAGs of the SAR86 clade and *Glaciecola*.

In accordance with ASV data, *Polaribacter* were found to be only lowly abundant and hardly expressed in FL communities but prominent in PA communities (Additional file 2: Fig. S7A). Further data are shown in Additional file 2: Fig. S20. It is noteworthy that *Alphaproteobacteria* in FL communities expressed mostly ABC-type transporters, whereas in PA10 communities, members of the alphaproteobacterial genera *Parasphingopyxis*, *Parasphingorhabdus*, *Maricaulis*, and *Hyphomonas* featured the highest TBDT expressions (Additional file 2: Fig. S19). *Parasphingopyxis* species have been isolated from red macroalgae and *Maricaulis* from dinoflagellate phycospheres [55, 56], while *Parasphingorhabdus* species have been found in mollusk guts [57]. *Maricaulis* and *Hyphomonas* can attach to surfaces via prosthecae and feature complicated life cycles [58, 59]. Further details on active MAGs are provided in Additional file 3.

Active CAZymes, PULs, and PUL-like clusters

Expressed CAZymes in FL community metaproteome data mapped to PULs and PUL-like clusters that were predicted to target host glycans, α-glucans, β-glucans, xyloglucans, fucose, alginate, and chitin (Additional file 1: Table S12). High expression was also observed for α-glucan degradation CAZymes with a peak on April 26, and β-glucan (laminarin) degradation CAZymes, which were particularly expressed during the diatom bloom's end and the second bloom phase (May 8, 22, and 24) (Additional file 2: Fig. S21). On May 8, after the diatom bloom, also few CAZymes targeting fucose-containing polysaccharides were expressed. Complementary

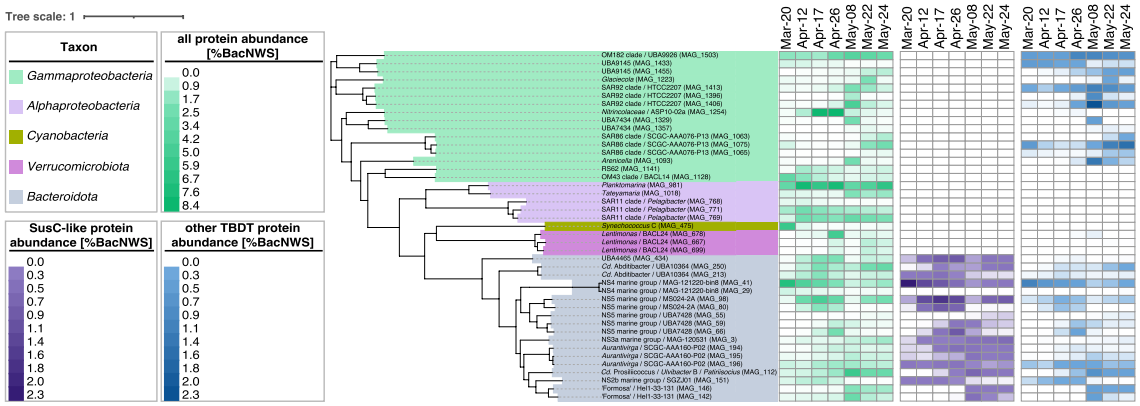


Fig. 6 Topmost expressed bacterial MAGs obtained from sampled FL bacteria. In total, 44 highly expressed bacterial FL MAGs were selected, and a maximum-likelihood tree was computed in anvi'o v7.1 based on protein sequences of 38 universal single-copy genes. Taxonomic affiliations according to GTDB r207_v2 are provided for each MAG with corresponding Silva r138.1 affiliations in parenthesis. Corresponding phyla (class for *Proteobacteria*) are represented by background colors. Protein expression based on summarized protein abundances as a percentage of bacterial normalized weighted spectra (%BacNWS) are represented by heatmaps: (i) overall expression (shades of green), (ii) expression of SusC-like proteins (shades of purple), and (iii) expression of TonB-dependent transporters (TBDTs) except for SusC-like proteins (shades of blue)

information on the PA metaproteome data is provided in Additional file 2: Fig. S18 and Additional file 3.

MAG analyses highlight distinct polysaccharide degradation potentials in abundant FL and PA community members

Out of the 305 species-level MAGs of all fractions, 244 contained degradative CAZymes including 161 candidate PULs (*susCD* gene tandems plus at least 1 degradative CAZyme), 1056 PUL-like clusters (1 *susC*-, *susD*-like or other TBDT gene plus at least 1 degradative CAZyme), and 652 CAZyme-rich gene clusters (at least 3 degradative CAZyme genes) (Additional file 1: Table S13, Additional file 2: Fig. S22).

We linked MAGs with 16S rRNA gene amplicon data to leverage the high temporal resolution amplicon data to uncover variations in MAG abundances (Additional file 1: Table S10, Additional file 2: Fig. S23), for which we selected the 71 most abundant MAGs for in-depth PUL analysis. Nine of these harbored 40 or more CAZyme genes, all of which were prevalent in PA communities (Fig. 7). A description of the most prominent MAGs, their links to ASVs and changes over time as well as their key CAZyme genes and inferred substrates is provided in Additional file 3, whereas a more holistic summary of the main results is provided in the subsequent discussion.

Discussion

The 2018 spring phytoplankton bloom at Helgoland Roads was among the most diverse in terms of phytoplankton species richness that we analyzed since 2009 [42, 43], in particular, compared to that of 2020, where algal biomass was almost entirely dominated by few diatom species during two sharply separated bloom phases [39]. The 2018 spring bloom in contrast was characterized by more complex gradual successions of diatoms, raphidophytes, haptophytes, and—to a lesser extent—photosynthetic dinoflagellates.

In 2018, an influx of nitrate- and silicate-rich freshwater around April 10 was likely instrumental in bolstering the diatom bloom, which resulted in an almost complete consumption of free silicate within a fortnight. A second influx event around May 23 during the late *Phaeocystis* bloom coincided with the emergence of *Noctiluca scintillans*, a heterotrophic giant dinoflagellate (0.2–2 mm diameter) that frequently occurs in Helgoland waters

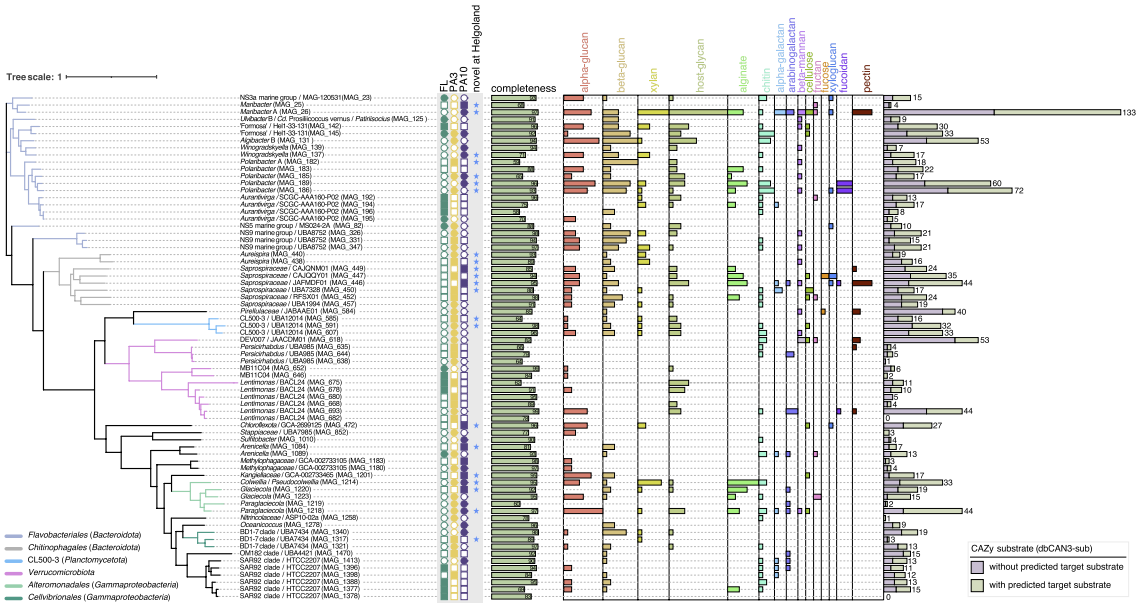


Fig. 7 Predicted substrates of PULs and PUL-like clusters in abundant MAGs. Based on 16S rRNA amplicon and corresponding MAG abundance data, we selected 71 MAGs from bacterial clades that were prominent either during particular bloom phases and/or in particular size fractions. A maximum-likelihood tree (computed in anvi'o v7.1 based on protein sequences of 38 universal single-copy genes) is shown to the left. PULs and PUL-like clusters in these MAGs were annotated and corresponding target substrate classes were using the dbCAN3-sub database. The numbers of involved degradative CAZyme genes (GH, PL, CE) are represented by bar charts. The stacked bars to the very right correspond to the total number of CAZyme genes per MAG with and without predicted substrate specificities. In addition, the fractions are indicated where the MAGs were most abundant (FL, 0.2–3 µm; PA3, 3–10 µm; PA10, > 10 µm), with circles representing data obtained from corresponding ASVs and squares representing read-based MAG abundances. Further details are provided in Additional file 1: Table S7

from June to August [60]. *N. scintillans* was most probably transported with coastal waters to Helgoland, and since *N. scintillans* prey on *Phaeocystis* [61], likely contributed to the *Phaeocystis* bloom's demise. Likewise, *Cryothecomonas* nanoflagellates, detected after the diatom bloom's peak, prey on diatoms [62], and thus likely contributed to the termination of the diatom bloom.

The bacterioplankton responded to the spring bloom with swift successions of distinct clades, which were more dynamic in the more diverse PA communities. Some bacterial clades correlated with distinct phytoplankton bloom phases, e.g., *Polaribacter*, *Winogradskyella*, and unclassified *Nitrincolaceae* with the diatom bloom, and unclassified *Stappiaceae*, *Sulfitobacter*, and unclassified *Methylophagaceae* with the *Phaeocystis* bloom. Other clades were abundant during both, the late diatom and *Phaeocystis* bloom phases, e.g., *Cd*. Prosiliicoccus, "Formosa", *Algibacter*, *Glaciecola*, and the BD1-7 clade.

The positive selection of bloom-adapted bacterial clades resulted in a decline in the diversity of FL bacteria and a size increase of the most abundant MAGs, notably in *Bacteroidota*, *Gammaproteobacteria*, and, to a lesser extent, *Planctomycetota*, *Verrucomicrobiota*, and *Alphaproteobacteria*. Diversity increased again during the collapse of the *Phaeocystis* bloom with the proliferation of more opportunistic generalists, such as members of the SAR86 clade and *Methylophagaceae* [63]. These clades featured smaller genomes, which was reflected in a decrease in the average size of abundant MAGs during the terminal bloom phase. Contrasting patterns were observed in both PA fractions, where diversities did not decrease during the main bloom phases, while the sizes of the most abundant MAGs decreased during the diatom bloom and increased notably during the late *Phaeocystis* bloom towards the bloom's end. This illustrates that different selective forces shaped FL and PA communities.

PA bacteria harbored more genes to degrade hardly soluble and structurally complex polysaccharides

Both abundant FL and PA bacteria featured high proportions of polysaccharide-degrading bacteria, however, with distinct CAZyme and PUL repertoires. Genes for the degradation of laminarins and α-glucans, both abundant, soluble, and structurally simply storage glucans, were the most prominent among FL bacteria, corroborating previous observations at Helgoland Roads [39]. Surprisingly, such genes were proportionally even more abundant among PA bacteria. Owing to sheer numbers, FL bacteria likely decomposed the bulk of laminarins and α-glucans, but the high proportion of respective genes in PA bacteria indicates that they are far from insignificant in this process. PUL analyses of 71 abundant MAGs from all fractions (Fig. 7) substantiated the salient role of storage glucans, as 40 contained β-glucan and 43 α-glucan PULs (Additional file 2: Fig. S9), fortifying the view that these glucans become available to PA bacteria that colonize senescent or dead algae.

Like during the 2020 spring bloom [39], α-glucan PULs were also dominated by type I α-glucan PULs in 2018. In addition to the previously described α-glucan PULs types I, II, and IV [39, 44], we also identified an additional type III comprising GH13, GH57, and GH77 (see [64]) genes in *Gammaproteobacteria*. MAG analyses showed an increase in PA *Gammaproteobacteria* with α- and β-glucan utilization genes after the diatom bloom (Additional file 2: Fig. S9). For example, MAG_1223 (*Glaciecola*) and MAG_1218 (*Paraglaciecola*) contained abundant genes for α-glucan hydrolysis, while MAG_1340 (BD1-7 clade) was rich in β-glucan hydrolytic genes (Fig. 7), suggesting that besides *Bacteroidota* also *Gammaproteobacteria* are significant α- and β-glucan consumers during algae die off phases.

Many marine macroalgae, e.g., *Saccharina* and *Fucus* brown algae, release gel-forming alginate and pectin-like polysaccharides [65]. Besides, alginate biosynthesis genes have been found in bloom-associated marine SAR92 clade *Gammaproteobacteria* [63]. Metatranscriptome analyses have furthermore suggested that alginate is an abundant bacterial substrate during spring blooms at Helgoland Roads [39]. Alginate and pectin degradation gene frequencies were notably more abundant in PA communities, reflecting the low solubilities of both substrates. Alginate gene frequencies in general prevailed over pectin degradation gene frequencies. This corroborates studies on the bacterial colonization of synthetic alginate and pectin particles by Bunse et al., where alginate was the preferred substrate [66]. We found alginate utilization genes predominantly in metagenome sequences attributed to unclassified *Bacteroidetes*, *Polaribacter*, and *Alteromonadales*. This was corroborated by MAG analyses, with alginate PULs present in PA *Saprospiraceae (Bacteroidota)*, *Polaribacter* (*Bacteroidota*), and *Colwellia* (*Alteromonadales*) (Fig. 7). These in situ data also support the in vitro experiments of Bunse et al., who identified *Colwellia* as among the primary colonizers on synthetic alginate particles [66].

Xylan degradation gene frequencies were largely stable among FL bacteria but increased considerably in PA bacteria during the late diatom and *Chattonella* bloom phases, surpassing FL frequencies more than twofold. This likely reflects an increased availability of structural xylans from disintegrating algae as well as poor xylan solubilities. Studies on the diatom *Thalassiosira weissflogii* have shown that its xylans and mannans are primarily found in POM and only little in DOM [67], consistent with functions as cell wall polysaccharides [30, 68]. We

found xylan degradation genes in PA bacteria (e.g., *Colwellia*) and in FL bacteria (e.g., "Formosa"), but overall gene proportions suggest a higher proportion of xylan degrading bacteria among PA bacteria.

The bacterial cell wall polysaccharide peptidoglycan is also hardly soluble. However, peptidoglycan degradation gene [69] frequencies were substantially higher in FL than in PA bacteria, suggesting that peptidoglycan is rapidly solubilized and recycled. This is consistent with the fact that peptidoglycan is not known to significantly accumulate in POM [70].

Conversely, fucoidan utilization gene frequencies were consistently higher in PA than in FL communities. Apart from the well-known *Verrucomicrobiota* [71], we observed such genes also in *Polaribacter*-affiliating metagenome sequences during the diatom bloom. This was confirmed by the presence of fucoidan-targeting gene clusters in various PA MAGs, including *Polaribacter* MAG_186 and MAG_189, *Saprospiraceae* MAG_446 and MAG_449, *Planctomycetota* MAG_584, and *Lentimonas* MAG_693 (Fig. 7). *Polaribacter* MAG_189 was only abundant at the beginning of the diatom bloom, whereas *Polaribacter* MAG_186 prevailed during the diatom bloom, and the *Planctomycetota*, *Saprospiraceae*, and *Lentimonas* MAGs were most abundant during the *Phaeocystis* bloom. The presence of fucoidan throughout the bloom is plausible, since fucoidan-containing polysaccharides secreted by diatoms [67, 72] are rather persistent to bacterial degradation [73].

Genes for host glycan recognition, binding, and degradation were present in 33 of the 71 studied abundant MAGs, in particular in *Bacteroidota*, *Verrucomicrobiota*, and *Planctomycetota* (Fig. 7), and comprised GH92 (α-mannosidase) as well as GH20 and GH109 (β-1,6-N-acetylglucosaminidase) family genes. Host glycans are branched heteropolysaccharides that decorate eukaryotic host cell surfaces, e.g., mucin O-linked glycans, N-linked glycoproteins, and highly sulfated glycosaminoglycans (GAGs) in the human gut [74, 75]. Microalgae are also decorated with host glycans that are known to play a role in algal symbiont interactions with their hosts (e.g., [76]). In particular, mannose-rich N-glycans have been detected in microalgae [77, 78]. Binding to host glycans allows bacteria to initiate colonization of eukaryote surfaces, but host glycans also constitute an important substrate, not only for human gut bacteria but also for PA bacteria during phytoplankton blooms. For human gut *Bacteroides*, it has been shown that problematic antennary monosaccharides are removed from host glycans before uptake [75]. It is likely that such extracellular pre-digestion also occurs among marine bacteria, and if the selectively removed monosaccharides can diffuse away, this would explain the source of soluble sulfated methylpentoses (fucose, rhamnose) that constitute a preferential substrate for recurring small-celled FL *Verrucomicrobioata* at Helgoland Roads [79].

The highly adaptable CAZome

CAZyme repertoires can vary considerably even between species of the same genus (e.g., [39]). For instance, *Winogradskyella* HQ MAG_139 harbored a much lower number of PULs and CAZyme-rich gene clusters than MAG_137, even though the latter was of lesser quality (94% vs 71%, Fig. 7). *Lentimonas* represents another illustrative case with three HQ MAGs, of which only MAG_693 contained abundant CAZyme genes (Fig. 7). CAZyme gene and PUL repertoires thus confer information about the adaptation of a given species towards a specific polysaccharide niche rather than its overall phylogenetic position in the tree of life. This implicates that the process of polysaccharide niche adaptation must considerably outpace the evolution of novel species, possibly by frequent lateral gene transfer [80].

CAZyme-rich MAGs often exhibited higher relative abundances during the bloom than closely related ones with fewer CAZymes, particularly in the PA fractions. This trend was evident for MAGs of *Maribacter*, *Winogradskyella*, *Polaribacter*, *Lentimonas*, and the CL500-3 and BD1-7 clades. *Lentimonas* MAG_693 for example harbored CAZyme genes targeting a variety of polysaccharide substrates (Fig. 7). Corresponding ASV data confirmed that this MAG represented a distinctively PA-associated species, consistent with a previous study on *Lentimonas* [81]. In this study, we found additional preferentially FL *Lentimonas* species, highlighting a broader niche spectrum within members of this genus (Fig. 7). Similarly, members of *Polaribacter* usually exhibit high FL abundances during diatom-dominated blooms at Helgoland Roads [50]. *Polaribacter* MAG_183 in this study corresponds to an abundant FL *Polaribacter* (MAG P_MB288) that we observed during the Helgoland spring bloom in 2020 [39]. However, in contrast to 2020, the dominating *Polaribacter* during the 2018 bloom were distinct and exhibited a clear preference for PA communities. In accordance with general trends, the dominant PA *Polaribacter* featured a larger MAG (MAG_189) with a higher number of CAZyme genes.

Noteworthy clades with low CAZyme gene proportions

Members of the BD1-7 clade, unclassified *Stappiaceae*, and *Nitrincolaceae* featured high abundances but not CAZyme-rich MAGs. The gammaproteobacterial BD1-7 clade belongs to the group of Oligotrophic Marine *Gammaproteobacteria* (OMG) [82]. Members of this clade are known to associate with phytoplankton [83] as with other eukaryotes, such as sponges [84], corals [85], brown algae

[86], and squids [87], where they may exert symbiotic functions. Alphaproteobacterial *Stappiaceae* are related to the abundant *Roseobacteraceae*. The genera include bacteriochlorophyll-producing *Roseibium*, members of which have been isolated from red algae [88], corals [89], oysters [90], and dinoflagellates [91], possibly also in a symbiotic function. *Nitrincolaceae* (formerly *Oceanospirillaceae*), abundant in PA3 communities, are opportunistic *Gammaproteobacteria* that frequently associate with phytoplankton (e.g., [92]), including *Reinekea* species, which we have observed in high abundance at Helgoland Roads before [42, 93]. Members of the *Nitrincolaceae* ASP10-2a clade are known to be diatom-associated [94].

Bloom-associated PA bacteria and global carbon cycling

The bulk of particles during phytoplankton blooms are either formed directly by aggregation of algal necromass or indirectly via the formation and excretion of fecal pellets by grazing of small zooplankton. The latter have been estimated to consume almost two-thirds of the phytoplankton cells on a daily basis [13], which is why fecal pellets constitute high proportions of the POM during phytoplankton blooms. Copepods are abundant zooplankters and have short gut transmit times (30 to 90 min [95, 96]), which is why their fecal pellets contain considerable proportions of only partially degraded microalgae [97]. A substantial part of the captured PA communities in our study thus represent primary or secondary fecal pellet colonizers that consume residual pelleted algal polysaccharides. According to recent estimates, phytoplankton-specific loss rates to zooplankton grazing constitute the greatest uncertainty in CMIP6 marine biogeochemical models used to assess bacterial remineralization versus sequestration rates of algal biomass. These uncertainties range in the gigatons of carbon per year [98], which is substantial considering that recent anthropogenic carbon emissions have been estimated at around 10 gigatons per year (IPCC for the year 2018).

Concluding remarks

Marine bacteria that colonize suspended particles inhabit a much more diverse habitat with ampler niche spaces and closer interactions than free-floating bacteria in the water column. However, to obtain quantitative data on bacterial polysaccharide degradation on particles in situ poses a considerable challenge. This applies in particular to reliable biochemical data on polysaccharide turnover rates, precise cell counts of PA bacteria, and even precise PA bacterial diversities owing to high proportions of chloroplast sequences in corresponding 16S rRNA gene amplicon data. Also, to obtain corresponding sufficiently deep metaproteome data remains a challenge. Finally, our sampling method does neither allow to discriminate different types of particles apart from broad size ranges nor to discriminate between loosely particle-associated and truly particle-attached bacteria—a limitation that we have discussed in detail in a previous study on the diversity, isolation, and cultivation of PA bacteria during the 2018 Helgoland spring bloom [21].

These challenges notwithstanding, we could demonstrate that PA bacterial communities were more diverse and underwent more dynamic changes in response to the 2018 spring phytoplankton bloom at Helgoland Roads than their FL counterparts. PA communities also featured a substantially higher metabolic potential for the degradation of a wide variety of polysaccharides. This was not only evident from assembled metagenome data but also in representative MAGs of abundant and active species. In the aforementioned study [21], we have also shown that PA bacteria represented less than 1% of the total bacterial community during spring 2018 at Helgoland Roads. However, considering that a major proportion of algal necromass passes through the POM pool, these bacteria must act as gatekeepers for the solubilization and subsequent remineralization of significant, yet-to-be-quantified proportions of algal polysaccharides during and after phytoplankton blooms, in spite of being considerably outnumbered by FL bacteria.

Materials and methods

Sampling, physicochemical, and phytoplankton data

Seawater samples were collected during spring 2018 (March 1 to May 29) off the North Sea island Helgoland (German Bight) at the LTER site "Kabeltonne" (54° 11.3′ N, 7° 54.0′ E, DEIMS.iD: https://deims.org/1e96ef9b-0915-4661-849f-b3a72f5aa9b1) by fractionating filtration (FL, 0.2–3 µm; PA3, 3–10 µm; PA10, > 10 µm) as described previously [42] (see Additional file 3 for details).

Wind direction data were obtained from the Climate Data Store of the Copernicus Climate Change Service [99]. Other physicochemical data, such as Secchi depth, water temperature, salinity, chlorophyll *a* content, dissolved inorganic nitrogen (NO_2^-, NO_3^-, NH_4^+), silicate, and phosphate as well as microscopic algae and zooplankton counts and taxonomic classifications were obtained as part of the Helgoland Roads LTER time series [100, 101]. Biovolumes of abundant plankters were determined in the framework of the Sylt Roads time series [102] as described elsewhere [103]. These data are summarized in Additional file 1: Table S1. Both the Helgoland and Sylt Roads time series are conducted by the Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (Bremerhaven, Germany).

16S and 18S rRNA gene amplicon sequencing and analysis

Sequencing of 16S rRNA gene amplicons was performed at the Max Planck Genome Centre Cologne (Germany). DNA from biomass retained on filters was extracted as described elsewhere [21] and amplified using primers 341F and 805R targeting the V3 and V4 regions [104] for the FL samples, and primers 515F and 806R targeting the V4 region [105] for the PA3 and PA10 samples. Sequencing was carried out on an Illumina HiSeq 2500 (Illumina, San Diego, CA, USA) in rapid mode with 2 × 250 bp paired-end reads.

Sequences were analyzed for single nucleotide-resolved amplicon sequence variants (ASVs) using the DADA2 v1.19.2 package [106] with R v4.0.3 (http://www.R-project.org) (Additional file 3). ASVs assigned to chloroplasts, mitochondria, *Eukarya*, *Archaea*, or unclassified sequences were excluded from further analyses (Additional file 2: Fig. S6B). Possible impacts of the different primer sets and the omission of rarefaction are provided in Additional file 3, as well as details on the analysis of the 18S rRNA amplicon data.

Metagenome sequencing and assembly

Metagenomes were sequenced at the Max Planck Genome Centre Cologne, 34 on an Illumina HiSeq 2500 using 2 × 150 bp chemistry, and eight additional PA3 metagenomes on a PacBio Sequel II (Menlo Park, CA, USA) using one SMRT cell per sample in long-read HiFi mode. The quality of Illumina reads was assessed with FastQC v0.11.9 [107].

Quality-filtered reads from FL metagenomes were assembled individually within SPAdes v3.11.1 [108]. Quality-filtered reads from PA3 and PA10 Illumina metagenomes were assembled individually using MEGAHIT v1.2.9 [109]. Assemblies of PA3 PacBio metagenomes were generated using Flye v2.9.1 [110] (Additional file 3). Assembly quality was assessed with QUAST [111]. Contigs below 2.5 kbp were removed using *anvi-script-reformat-fasta* within anvi'o v6.2 [112].

Metagenome-assembled genome (MAG) retrieval and analysis

MAG retrieval was performed as described previously [43] (Additional file 3). MAGs were classified into low-, medium-, and high-quality categories according to the criteria described in Bowers et al. [47] using CheckM v1.1.3 [113]. Only medium- and high-quality MAGs were used in further analyses. Dereplication was done using dRep v3.0.0 [114] with an average nucleotide identity (ANI) > 95%. ANI was calculated using FastANI [115]. MAG abundances were calculated as described previously [116] (Additional file 3). 16S rRNA gene sequences were extracted from MAGs using barrnap v0.9 (https://github.com/tseemann/barrnap) and subsequently classified in the Silva Incremental Aligner (SINA) with Silva SSU 138.1 taxonomy [117]. MAG taxonomies were determined by GTDB-Tk v2.1.0 [118, 119] with GTDB release R207_v2. Differences in the denominations of taxa between Silva and GTDB were resolved as described previously [44]. A phylogenomic tree of dereplicated MAGs was constructed using FastTree [120] from within anvi'o v7.1 and visualized using interactive Tree of Life (iTol) v6.5.6 [121].

Interrelation of 16S rRNA gene amplicon and MAG data

Blastn was used to search all prevalent ASVs from abundant genera within the 16S rRNA gene amplicon dataset against all MAG-derived 16S rRNA gene sequences. For identical hits with 100% coverage, we assumed that changes in ASV relative abundance reflected changes of the corresponding MAG over time. Since not all MAGs contained 16S rRNA genes, we extended our search to all MAGs that we obtained from the Helgoland metagenome samples from 2010 [44], 2012 [44], 2016 [43, 44], and 2020 [39]. MAGs from 2018 without 16S rRNA gene sequence were considered to match MAGs from other sampling years, if both exhibited an ANI of at least 95%. In addition, we included two matching MAGs from the GTDB database. Details are provided in Additional file 1: Table S10.

Metagenome and MAG annotation

For assembled Illumina metagenomes, protein-coding sequences were predicted using Prodigal [122], Aragorn [123], and barrnap (https://github.com/tseemann/barrnap) as implemented in Prokka v1.14.6 [124] (default settings). For PacBio data, FragGeneScan v1.31 [125] was used (setting *-w 1*) due to a higher number of frameshifts. Functional MAG annotations were done as described in Additional file 3.

Gene frequency analyses

Eukaryotic and unclassified reads were removed from unassembled Illumina metagenomes according to Kaiju v1.9.0 [126] annotations. Metagenomes were subsequently assembled with MEGAHIT v1.2.9, and frequencies of genes of interest were computed for each metagenome as follows: gene frequency = (sum of average coverage of target gene(s)) × 100 / (sum of average coverage of all genes) [42]. The average coverages of target genes were determined in SqueezeMeta v1.3.1 [127] using bowtie2 [128] for mapping. CAZymes were predicted as described in Additional file 3.

Prediction of CAZyme-rich gene clusters and PULs

CAZyme-rich gene clusters and PULs were identified in a sliding window approach as described previously [40, 43] with a window length of ten genes. When at least three genes within the window coded for either GHs, PLs, CEs, sulfatase, TBDTs, or SusD-like proteins, we considered this a candidate locus. The resulting CAZyme-rich loci were manually annotated based on a combination of multiple databases (Additional file 3). Putative target substrate classes of PULs, PUL-like, and CAZyme-rich gene clusters in MAGs were predicted using the dbCAN3-sub database [129].

Metaproteome analyses

Metaproteomes were analyzed on seven dates for FL samples (2018/03/20, 2018/04/12, 2018/04/17, 2018/04/26, 2018/05/08, 2018/05/22, 2018/05/24) and on three dates for PA samples (2018/04/17, 2018/05/08, 2018/05/24). Proteins were extracted from filtered biomass and subsequently analyzed as described elsewhere [43, 130, 131] (Additional file 3).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-024-01757-5.

Additional file 1: Supplementary tables.

Additional file 2: Supplementary figures.

Additional file 3: Supplementary text.

Acknowledgements

We acknowledge the Helgoland sampling team, including Lily Franzmeyer, Mirja Meiners, Sabine Kühn, Peter Rücknagel, Jörg Wulf, Nina Heinzmann, Greta Giljan, Jan Brüwer, and Anneke Heins. We also acknowledge the BAH team, including Eva Maria Brodte, Antje Wichels, and FS Aade and FS Uthörn captains and crews for the help with sampling, analyses, logistics, and provision of lab space. Furthermore, we thank our colleagues from the Max Planck Genome Centre Cologne for their assistance in metagenomics. Further thanks go to Lilly Franzmeyer, Thomas Sura, Daniela Zühlke, Doreen Schultz, Katharina Riedel, and Pierre-Alexander Mücke from the University of Greifswald for the technical assistance in MS-based metaproteomics. Finally, we thank our former colleagues Karen Krüger and T. Ben Francis for the assistance in MAG CAZyme analysis and Luis H. Orellana for the MAG abundance calculations. Fengqing Wang is a member of the International Max Planck Research School of Marine Microbiology (MarMic).

Authors' contributions

F.W., C.S., and D.L.: bioinformatics and data integration and analysis. D.B., R.S., A.T.S., and D.B.: metaproteomics. B.H.: sequencing. J.R.: algal biovolumes. I.V.K. and K.H.W.: Helgoland Roads time series data. B.M.F.: sampling logistics. M.M.B.: 16S and 18S rRNA gene amplicon sampling and analysis. F.W., C.S., D.L., and H.T.: compilation of the manuscript. B.M.F., T.S., M.M.B., H.T., and R.I.A.: study design. All authors read, discussed, and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This study was funded by the Max Planck Society and supported by the German Research Foundation (DFG) in the framework of the research unit FOR2406 "Proteogenomics of Marine Polysaccharide Utilization (POMPU)" by grants of D.B. (BE 3869/4-3), B.M.F. (FU 627/2-3), T.S. (SCHW 595/10-3, SCHW 595/11-3), M.M.B. (RI 969/9-2), H.T. (TE 813/2-3), and R.I.A. (AM 73/9-3). The Helgoland time series is supported by the Biological Station Helgoland, Alfred Wegener Institute, and Helmholtz Center for Polar and Marine Research (AWI_BAH_o 1).

Availability of data and materials

Metagenome reads, assemblies, and MAGs were deposited in the European Nucleotide Archive (ENA) under project numbers PRJEB38290 and PRJEB67502. 16S rRNA gene amplicon sequences of FL and PA fractions were deposited in ENA under project numbers PRJEB51721 and PRJEB51816, respectively. Mass spectrometry proteome data were deposited at the ProteomeXchange Consortium via the PRIDE partner repository [132]. Original mass spectrometry proteome data of FL bacteria are accessible as project PXD042676 and data for PA bacteria as project PXD046705.

Declarations

Ethics approval and consent to participate

Ethics approval was not required for the study.

Competing interests

The authors declare no competing interests.

Author details

[1]Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, Germany. [2]Institute of Pharmacy, University of Greifswald, Felix-Hausdorff-Straße 3, 17489 Greifswald, Germany. [3]Institute of Marine Biotechnology, Walther-Rathenau-Straße 49a, 17489 Greifswald, Germany. [4]Institute of Microbiology, University of Greifswald, Felix-Hausdorff-Straße 8, 17489 Greifswald, Germany. [5]Marine College, Shandong University, Weihai 264209, China. [6]Max Planck Genome Centre Cologne, Carl von Linné-Weg 10, 50829 Cologne, Germany. [7]Alfred Wegener Institute for Polar and Marine Research, Biologische Anstalt Helgoland, Helgoland 27483, Germany.

Received: 7 November 2023 Accepted: 4 January 2024
Published online: 20 February 2024

References

1. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. Science. 1998;281(5374):237–40.
2. Falkowski PG, Barber RT, Smetacek VV. Biogeochemical controls and feedbacks on ocean primary production. Science. 1998;281(5374):200–6.
3. Mann DG. The species concept in diatoms: evidence for morphologically distinct, sympatric gamodemes in four epipelic species. Plant Syst Evol. 1989;164:215–37.
4. Inomura K, Karlusich JJP, Dutkiewicz S, Deutsch C, Harrison PJ, Bowler C. High growth rate of diatoms explained by reduced carbon requirement and low energy cost of silica deposition. Microbiol Spectr. 2023:e03311-22.
5. Arrigo KR, Robinson DH, Worthen DL, Dunbar RB, DiTullio GR, VanWoert M, et al. Phytoplankton community structure and the drawdown of nutrients and CO_2 in the southern ocean. Science. 1999;283(5400):365–7.
6. Lancelot C, Gypens N, Billen G, Garnier J, Roubeix V. Testing an integrated river-ocean mathematical tool for linking marine eutrophication to land use: the *Phaeocystis*-dominated Belgian coastal zone (Southern North Sea) over the past 50 years. J Mar Syst. 2007;64:216–28.
7. Alderkamp AC, Buma AGJ, van Rijssel M. The carbohydrates of *Phaeocystis* and their degradation in the microbial food web. Biogeochemistry. 2007;83:99–118.
8. Vincent F, Gralka M, Schleyer G, Schatz D, Cabrera-Brufau M, Kuhlisch C, et al. Viral infection switches the balance between bacterial and

eukaryotic recyclers of organic matter during coccolithophore blooms. Nat Commun. 2023;14(1):510.
9. Zhang ZH, Li DH, Xie RZ, Guo RY, Nair S, Han H, et al. Plastoquinone synthesis inhibition by tetrabromo biphenyldiol as a widespread algicidal mechanism of marine bacteria. ISME J. 2023;17(11):1979–92.
10. Scholz B, Guillou L, Marano AV, Neuhauser S, Sullivan BK, Karsten U, et al. Zoosporic parasites infecting marine diatoms - a black box that needs to be opened. Fungal Ecol. 2016;19:59–76.
11. Garvetto A, Nézan E, Badis Y, Bilien G, Arce P, Bresnan E, et al. Novel widespread marine oomycetes parasitising diatoms, including the toxic genus *Pseudo-nitzschia*: genetic, morphological, and ecological characterisation. Front Microbiol. 2018;9:2918.
12. Fernández-Méndez M, Wenzhöfer F, Peeken I, Sørensen HL, Glud RN, Boetius A. Composition, buoyancy regulation and fate of ice algal aggregates in the Central Arctic Ocean. PLoS ONE. 2014;9(9):e107452.
13. Schmoker C, Hernández-León S, Calbet A. Microzooplankton grazing in the oceans: impacts, data variability, knowledge gaps and future directions. J Plankton Res. 2013;35(4):691–706.
14. Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. Science. 2015;347(6223):1257594.
15. Vargas CA, Cuevas LA, González HE, Daneri G. Bacterial growth response to copepod grazing in aquatic ecosystems. J Mar Biol Assoc UK. 2007;87(3):667–74.
16. Giering SL, Sanders R, Lampitt RS, Anderson TR, Tamburini C, Boutrif M, et al. Reconciliation of the carbon budget in the ocean's twilight zone. Nature. 2014;507(7493):480–3.
17. De La Rocha CL, Passow U. Factors influencing the sinking of POC and the efficiency of the biological carbon pump. Deep Sea Research Part II: Topical Studies in Oceanography. 2007;54(5–7):639–58.
18. Nowicki M, DeVries T, Siegel DA. Quantifying the carbon export and sequestration pathways of the ocean's biological carbon pump. Glob Biogeochem Cycles. 2022;36:e2021GB007083.
19. Simon M, Grossart HP, Schweitzer B, Ploug H. Microbial ecology of organic aggregates in aquatic ecosystems. Aquat Microb Ecol. 2002;28(2):175–211.
20. Turley CM, Stutt ED. Depth-related cell-specific bacterial leucine incorporation rates on particles and its biogeochemical significance in the Northwest Mediterranean. Limnol Oceanogr. 2000;45(2):419–25.
21. Heins A, Reintjes G, Amann RI, Harder J. Particle collection in Imhoff sedimentation cones enriches both motile chemotactic and particle-attached bacteria. Front Microbiol. 2021;12:643730.
22. Azúa I, Unanue M, Ayo B, Artolozaga I, Arrieta JM, Iriberri J. Influence of organic matter quality in the cleavage of polymers by marine bacterial communities. J Plankton Res. 2003;25(12):1451–60.
23. Leu AO, Eppley JM, Burger A, DeLong EF. Diverse genomic traits differentiate sinking-particle-associated versus free-living microbes throughout the oligotrophic open ocean water column. mBio. 2022;13(4):e01569–22.
24. Patel AK, Vadrale AP, Singhania RR, Michaud P, Pandey A, Chen SJ, et al. Algal polysaccharides: current status and future prospects. Phytochem Rev. 2023;22:1167–96.
25. 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.
26. Chen J, Yang J, Du H, Aslam M, Wang W, Chen W, et al. Laminarin, a major polysaccharide in stramenopiles. Mar Drugs. 2021;19(10):576.
27. Zvyagintseva TN, Shevchenko NM, Popivnich IB, Isakov VV, Scobun AS, Sundukova EV, et al. A new procedure for the separation of water-soluble polysaccharides from brown seaweeds. Carbohyd Res. 1999;322(1–2):32–9.
28. Myklestad SM. Production, chemical structure, metabolism, and biological function of the (1→3)-linked, β3-D-glucans in diatoms. Biol Oceanogr. 1989;6(3–4):313–26.
29. Becker S, Tebben J, Coffinet S, Wiltshire K, Iversen MH, Harder T, et al. Laminarin is a major molecule in the marine carbon cycle. Proc Natl Acad Sci USA. 2020;117(12):6599–607.
30. 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.
31. 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.
32. Mühlenbruch M, Grossart HP, Eigemann F, Voss M. Mini-review: Phytoplankton-derived polysaccharides in the marine environment and their interactions with heterotrophic bacteria. Environ Microbiol. 2018;20(8):2671–85.
33. Shnyukova EI, Zolotariova YK. Ecological role of exopolysaccharides of *Bacillariophyta*: a review. Int J Algae. 2017;19(1):5–20.
34. Thornton DCO. Diatom aggregation in the sea: mechanisms and ecological implications. Eur J Phycol. 2002;37(2):149–61.
35. Babiak W, Krzemińska I. Extracellular polymeric substances (EPS) as microalgal bioproducts: a review of factors affecting EPS synthesis and application in flocculation processes. Energies. 2021;14(13):4007.
36. Martino PD. Extracellular polymeric substances, a key element in understanding biofilm phenotype. Aims Microbiol. 2018;4(2):274–88.
37. 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.
38. White JBR, Silale A, Feasey M, Heunis T, Zhu Y, Zheng H, et al. Outer membrane utilisomes mediate glycan uptake in gut Bacteroidetes. Nature. 2023;618(7965):583–9.
39. Sidhu C, Kirstein IV, Meunier CL, Rick J, Fofonova V, Wiltshire KH, et al. Dissolved storage glycans shaped the community composition of abundant bacterioplankton clades during a North Sea spring phytoplankton bloom. Microbiome. 2023;11(1):77.
40. Lu D, Wang F, Amann RI, Teeling H, Du JZ. Epiphytic common core bacteria in the microbiomes of co-located green (*Ulva*), brown (*Saccharina*) and red (*Grateloupia*, *Gelidium*) macroalgae. Microbiome. 2023;11(1):126.
41. 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.
42. 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.
43. Francis TB, Bartosik D, Sura T, Sichert A, Hehemann JH, Markert S, et al. Changing expression patterns of TonB-dependent transporters suggest shifts in polysaccharide consumption over the course of a spring phytoplankton bloom. ISME J. 2021;15(8):2336–50.
44. Krüger K, Chafee M, Francis TB, del Rio TG, Becher D, Schweder T, et al. In marine *Bacteroidetes* the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. ISME J. 2019;13(11):2800–16.
45. Francis TB, Krüger K, Fuchs BM, Teeling H, Amann RI. *Candidatus* Prosiliicoccus vernus, a spring phytoplankton bloom associated member of the *Flavobacteriaceae*. Syst Appl Microbiol. 2019;42(1):41–53.
46. Okazaki Y, Fujinaga S, Tanaka A, Kohzu A, Oyagi H, Nakano S. Ubiquity and quantitative significance of bacterioplankton lineages inhabiting the oxygenated hypolimnion of deep freshwater lakes. ISME J. 2017;11(10):2279–93.
47. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;35(8):725–31.
48. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. Nature. 2019;568(7753):499–504.
49. Kavagutti VS, Bulzu PA, Chiriac CM, Salcher MM, Mukherjee I, Shabarova T, et al. High-resolution metagenomic reconstruction of the freshwater spring bloom. Microbiome. 2023;11(1):15.
50. Avcı B, Krüger K, Fuchs BM, Teeling H, Amann RI. Polysaccharide niche partitioning of distinct *Polaribacter* clades during North Sea spring algal blooms. ISME J. 2020;14(6):1369–83.
51. 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.

52. Alonso C, Warnecke F, Amann R, Pernthaler J. High local and global diversity of *Flavobacteria* in marine plankton. Environ Microbiol. 2007;9(5):1253–66.
53. Chafee M, Fernàndez-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.
54. 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.
55. Jeong SE, Kim KH, Baek K, Jeon CO. *Parasphingopyxis algicola* sp. nov., isolated from a marine red alga *Asparagopsis taxiformis* and emended description of the genus *Parasphingopyxis* Uchida *et al*. 2012. Int J Syst Evol Microbiol. 2017;67(10):3877–81.
56. Zhang XL, Qi M, Li QH, Cui ZD, Yang Q. *Maricaulis alexandrii* sp. nov., a novel active bioflocculants-bearing and dimorphic prosthecate bacterium isolated from marine phycosphere. Antonie Van Leeuwenhoek. 2021;114(8):1195–203.
57. Yoo JH, Han JE, Lee JY, Jeong SW, Jeong YS, Lee JY, et al. *Parasphingorhabdus cellanae* sp. nov., isolated from the gut of a Korean limpet, *Cellana toreuma*. Int J Syst Evol Microbiol. 2022;72(8):005470.
58. Abraham WR, Strömpl C, Meyer H, Lindholst S, Moore ERB, Christ R, et al. Phylogeny and polyphasic taxonomy of *Caulobacter* species. Proposal of *Maricaulis* gen. nov. with *Maricaulis maris* (Poindexter) comb. nov. as the type species, and emended description of the genera *Brevundirnonas* and *Caulobacter*. Int J Syst Evol Microbiol. 1999;49(3):1053–73.
59. Abraham WR, Rohde M. The family *Hyphomonadaceae*. In: Rosenberg E, EF DL, Lory S, Stackebrandt E, Thompson F, editors. The prokaryotes: *Alphaproteobacteria* and *Betaproteobacteria*. Berlin, Heidelberg: Springer Berlin Heidelberg. 2014:p. 283–99.
60. Löder MGJ, Kraberg AC, Aberle N, Peters S, Wiltshire KH. Dinoflagellates and ciliates at Helgoland Roads North Sea. Helgoland Mar Res. 2012;66:11–23.
61. Weisse T, Tande K, Verity P, Hansen F, Gieskes W. The trophic significance of *Phaeocystis* blooms. J Mar Syst. 1994;5(1):67–79.
62. Schnepf E, Kühn SF. Food uptake and fine structure of *Cryothecomonas longipes* sp nov., a marine nanoflagellate incertae sedis feeding phagotrophically on large diatoms. Helgoland Mar Res. 2000;54(1):18–32.
63. Francis TB, Urich T, Mikolasch A, Teeling H, Amann R. North Sea spring bloom-associated *Gammaproteobacteria* fill diverse heterotrophic niches. Environ Microbiome. 2021;16:15.
64. Mareček F, Møller MS, Svensson B, Janeček Š. A putative novel starch-binding domain revealed by in silico analysis of the N-terminal domain in bacterial amylomaltases from the family GH77. 3 Biotech. 2021;11(5):229.
65. Koch H, Dürwald A, Schweder T, Noriega-Ortega B, Vidal-Melgosa S, Hehemann JH, et al. Biphasic cellular adaptations and ecological implications of *Alteromonas macleodii* degrading a mixture of algal polysaccharides. ISME J. 2019;13(1):92–103.
66. Bunse C, Koch H, Breider S, Simon M, Wietz M. Sweet spheres: succession and CAZyme expression of marine bacterial communities colonizing a mix of alginate and pectin particles. Environ Microbiol. 2021;23(6):3130–48.
67. Huang GY, Vidal-Melgosa S, Sichert A, Becker S, Fang Y, Niggemann J, et al. Secretion of sulfated fucans by diatoms may contribute to marine aggregate formation. Limnol Oceanogr. 2021;66(10):3768–82.
68. Hecky RE, Mopper K, Kilham P, Degens ET. The amino acid and sugar composition of diatom cell-walls. Mar Biol. 1973;19(4):323–31.
69. Humann J, Lenz LL. Bacterial peptidoglycan-degrading enzymes and their impact on host muropeptide detection. J Innate Immun. 2009;1(2):88–97.
70. Kitayama K, Hama T, Yanagi K. Bioreactivity of peptidoglycan in seawater. Aquat Microb Ecol. 2007;46:85–93.
71. Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, et al. *Verrucomicrobia* use hundreds of enzymes to digest the algal polysaccharide fucoidan. Nat Microbiol. 2020;5(8):1026–39.
72. Vidal-Melgosa S, Sichert A, Francis TB, Bartosik D, Niggemann J, Wichels A, et al. Diatom fucan polysaccharide precipitates carbon during algal blooms. Nat Commun. 2021;12(1):1150.
73. Bligh M, Nguyen N, Buck-Wiese H, Vidal-Melgosa S, Hehemann JH. Structures and functions of algal glycans shape their capacity to sequester carbon in the ocean. Curr Opin Chem Biol. 2022;71:102204.
74. Brown HA, Koropatkin NM. Host glycan utilization within the *Bacteroidetes* Sus-like paradigm. Glycobiology. 2021;31(6):697–706.
75. Briliūtė J, Urbanowicz PA, Luis AS, Baslé A, Paterson N, Rebello O, et al. Complex N-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. Nat Microbiol. 2019;4(9):1571–81.
76. Tivey TR, Parkinson JE, Mandelare PE, Adpressa DA, Peng W, Dong X, et al. N-linked surface glycan biosynthesis, composition, inhibition, and function in cnidarian-dinoflagellate symbiosis. Microb Ecol. 2020;80(1):223–36.
77. Baïet B, Burel C, Saint-Jean B, Louvet R, Menu-Bouaouiche L, Kiefer-Meyer MC, et al. N-glycans of *Phaeodactylum tricornutum* diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme. J Biol Chem. 2011;286(8):6152–64.
78. Mócsai R, Figl R, Troschl C, Strasser R, Svehla E, Windwarder M, et al. N-glycans of the microalga *Chlorella vulgaris* are of the oligomannosidic type but highly methylated. Sci Rep. 2019;9(1):331.
79. Orellana LH, Francis TB, Ferraro M, Hehemann JH, Fuchs BM, Amann RI. *Verrucomicrobiota* are specialist consumers of sulfated methyl pentoses during diatom blooms. ISME J. 2022;16(3):630–41.
80. 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(7290):908–12.
81. Ren YH, Luo ZH, Liu Q, Wei B, Wu YH, Shu WS, et al. Insights into community assembly mechanisms, biogeography, and metabolic potential of particle-associated and free-living prokaryotes in tropical oligotrophic surface oceans. Front Mar Sci. 2022;9:923295.
82. Cho JC, Giovannoni SJ. Cultivation and growth characteristics of a diverse group of oligotrophic marine *Gammaproteobacteria*. Appl Environ Microbiol. 2004;70(1):432–40.
83. Wemheuer B, Güllert S, Billerbeck S, Giebel HA, 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.
84. Holert J, Cardenas E, Bergstrand LH, Zaikova E, Hahn AS, Hallam SJ, et al. Metagenomes reveal global distribution of bacterial steroid catabolism in natural, engineered, and host environments. mBio. 2018;9(1):e02345–17.
85. Yu XP, Yu KF, Liao ZH, Chen B, Deng CQ, Yu JY, et al. Seasonal fluctuations in symbiotic bacteria and their role in environmental adaptation of the scleractinian coral *Acropora pruinosa* in high-latitude coral reef area of the South China Sea. Sci Total Environ. 2021;792:148438.
86. Paix B, Othmani A, Debroas D, Culioli G, Briand JF. Temporal covariation of epibacterial community and surface metabolome in the Mediterranean seaweed holobiont *Taonia atomaria*. Environ Microbiol. 2019;21(9):3346–63.
87. Hu XJ, Su HC, Zhang P, Chen ZZ, Xu Y, Xu WJ, et al. Microbial community characteristics of the intestine and gills of medium-form populations of *Sthenoteuthis oualaniensis* in the South China Sea. Fishes. 2022;7(4):191.
88. Suzuki T, Muroga Y, Takahama M, Nishimura Y. *Roseigium denhamense* gen. nov., sp. nov. and *Roseibium hemelinense* sp. nov., aerobic bacteriochlorophyll-containing bacteria isolated from the east and west coasts of Australia. Int J Syst Evol Microbiol. 2000;50(6):2151–6.
89. Couceiro JF, Keller-Costa T, Marques M, Kyrpides NC, Woyke T, Whitman WB, et al. The *Roseibium album* (*Labrenzia alba*) genome possesses multiple symbiosis factors possibly underpinning host-microbe relationships in the marine benthos. Microbiol Resour Announc. 2021;10(34):e0032021.
90. Karimi E, Keller-Costa T, Slaby BM, Cox CJ, da Rocha UN, Hentschel U, et al. Genomic blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable *Alphaproteobacteria*. Sci Rep. 2019;9(1):1999.
91. Aguirre EG, Carlson HK, Kenkel CD. Complete genome sequence of *Roseibium* sp. strain Sym1, a bacterial associate of *Symbiodinium linucheae*, the microalgal symbiont of the anemone *Aiptasia*. Microbiol Resour Announc. 2023;12(3):e0111822.

92. Thiele S, Vader A, Thomson S, Saubrekka K, Petelenz E, Armo HR, et al. The summer bacterial and archaeal community composition of the northern Barents Sea. Prog Oceanogr. 2023:103054.
93. 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.
94. Bertrand EM, McCrow JP, Moustafa A, Zheng H, McQuaid JB, Delmont TO, et al. Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. Proc Natl Acad Sci USA. 2015;112(32):9938–43.
95. Mayzaud P, Tirelli V, Bernard JM, Roche-Mayzaud O. The influence of food quality on the nutritional acclimation of the copepod *Acartia clausi*. J Mar Syst. 1998;15(1–4):483–93.
96. Tirelli V, Mayzaud P. Relationship between functional response and gut transit time in the calanoid copepod *Acartia clausi*: role of food quantity and quality. J Plankton Res. 2005;27(6):557–68.
97. Köster M, Sietmann R, Meuche A, Paffenhöfer GA. The ultrastructure of a doliolid and a copepod fecal pellet. J Plankton Res. 2011;33(10):1538–49.
98. Rohr T, Richardson AJ, Lenton A, Chamberlain MA, Shadwick EH. Zooplankton grazing is the largest source of uncertainty for marine carbon cycling in CMIP6 models. Commun Earth Environ. 2023;4(1):212.
99. Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz SJ, et al. ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.adbb2d47.
100. Wiltshire KH, Kraberg A, Bartsch I, Boersma M, Franke HD, Freund J, et al. Helgoland Roads, North Sea: 45 years of change. Estuar Coast. 2010;33(2):295–310.
101. Kraberg A, Kieb U, Peters S, Wiltshire KH. An updated phytoplankton check-list for the Helgoland Roads time series station with eleven new records of diatoms and dinoflagellates. Helgoland Mar Res. 2019;73:9.
102. Armonies W, Asmus H, Buschbaum C, Lackschewitz D, Reise K, Rick J. Microscopic species make the diversity: a checklist of marine flora and fauna around the Island of Sylt in the North Sea. Helgoland Mar Res. 2018;72(1):11.
103. Hillebrand H, Dürselen CD, Kirschtel D, Pollingher U, Zohary T. Biovolume calculation for pelagic and benthic microalgae. J Phycol. 1999;35(2):403–24.
104. Herlemann DPR, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. 2011;5(10):1571–9.
105. Lucas J, Wichels A, Teeling H, Chafee M, Scharfe M, Gerdts G. Annual dynamics of North Sea bacterioplankton: seasonal variability superimposes short-term variation. Fems Microbiol Ecol. 2015;91(9):fiv099.
106. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581–3.
107. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010:Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
108. 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.
109. Li DH, Luo RB, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016;102:3–11.
110. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods. 2020;17(11):1103–10.
111. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5.
112. 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.
113. 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.
114. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11(12):2864–8.
115. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018;9(1):5114.
116. Orellana LH, Francis TB, Krüger K, Teeling H, Müller MC, Fuchs BM, et al. Niche differentiation among annually recurrent coastal Marine Group II Euryarchaeota. ISME J. 2019;13(12):3024–36.
117. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28(14):1823–9.
118. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020;36(6):1925–7.
119. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for *Bacteria* and *Archaea*. Nat Biotechnol. 2020;38(9):1079–86.
120. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5(3):e9490.
121. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6.
122. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 2010;11:119.
123. Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 2004;32(1):11–6.
124. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068–9.
125. Rho M, Tang H, Ye Y. FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res. 2010;38(20):e191.
126. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7(1):11257.
127. Tamames J, Puente-Sánchez F. SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. Front Microbiol. 2018;9:3349.
128. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
129. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023;51(W1):W115–21.
130. Deusch S, Seifert J. Catching the tip of the iceberg - evaluation of sample preparation protocols for metaproteomic studies of the rumen microbiota. Proteomics. 2015;15(20):3590–5.
131. Schultz D, Zühlke D, Bernhardt J, Francis TB, Albrecht D, Hirschfeld C, et al. An optimized metaproteomics protocol for a holistic taxonomic and functional characterization of microbial communities from marine particles. Environ Microbiol Rep. 2020;12(4):367–76.
132. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022;50(D1):D543–52.

Publisher's Note

Supplementary Figures for this manuscript include the following:

Supplementary Fig. S1 to S25.

Supplementary Figures

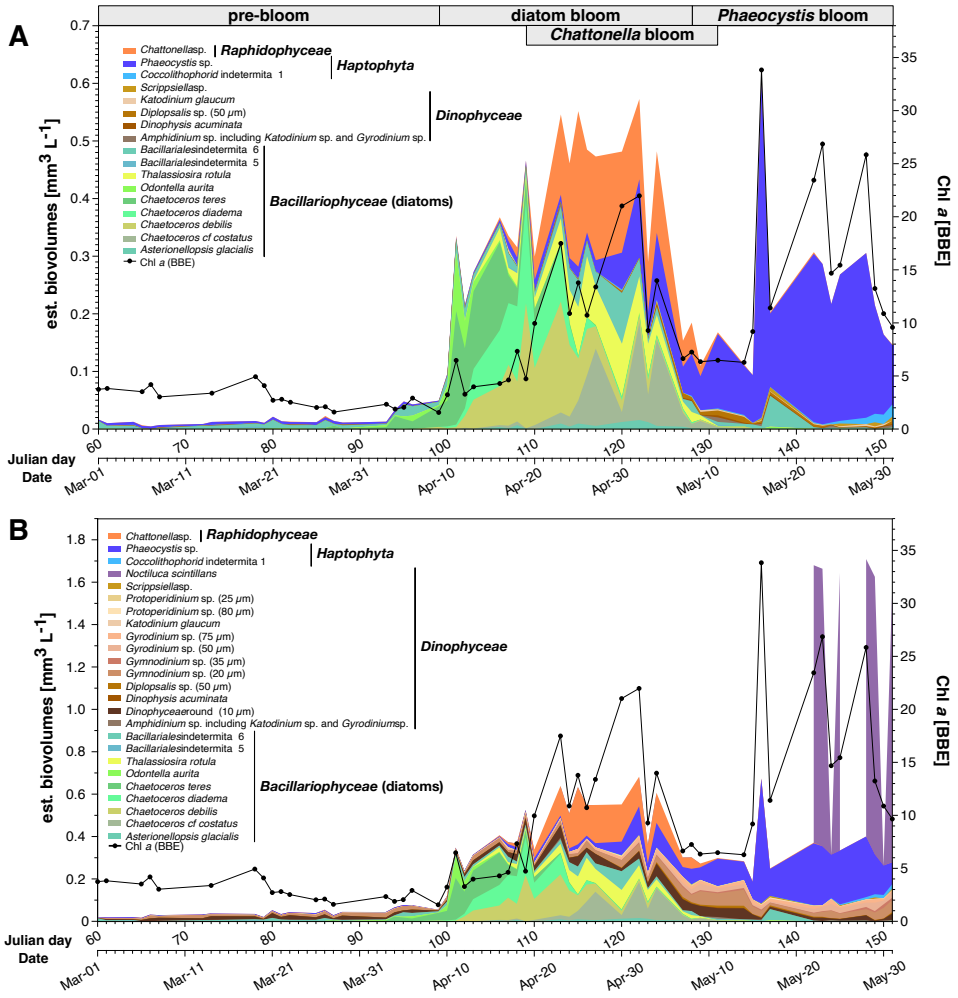


Fig. S1. 2018 spring phytoplankton bloom at Helgoland Roads. Estimated biovolumes of microscopically determined abundant plankton taxa (stacked colored areas) and chlorophyll *a* measurements (black line) as assessed by fluorescence-based algal group analyzer measurements (AlgaeLabAnalyser, BBE Moldaenke GmbH, Schwentinental, Germany). **A** photosynthetic plankters, **B** all plankters.

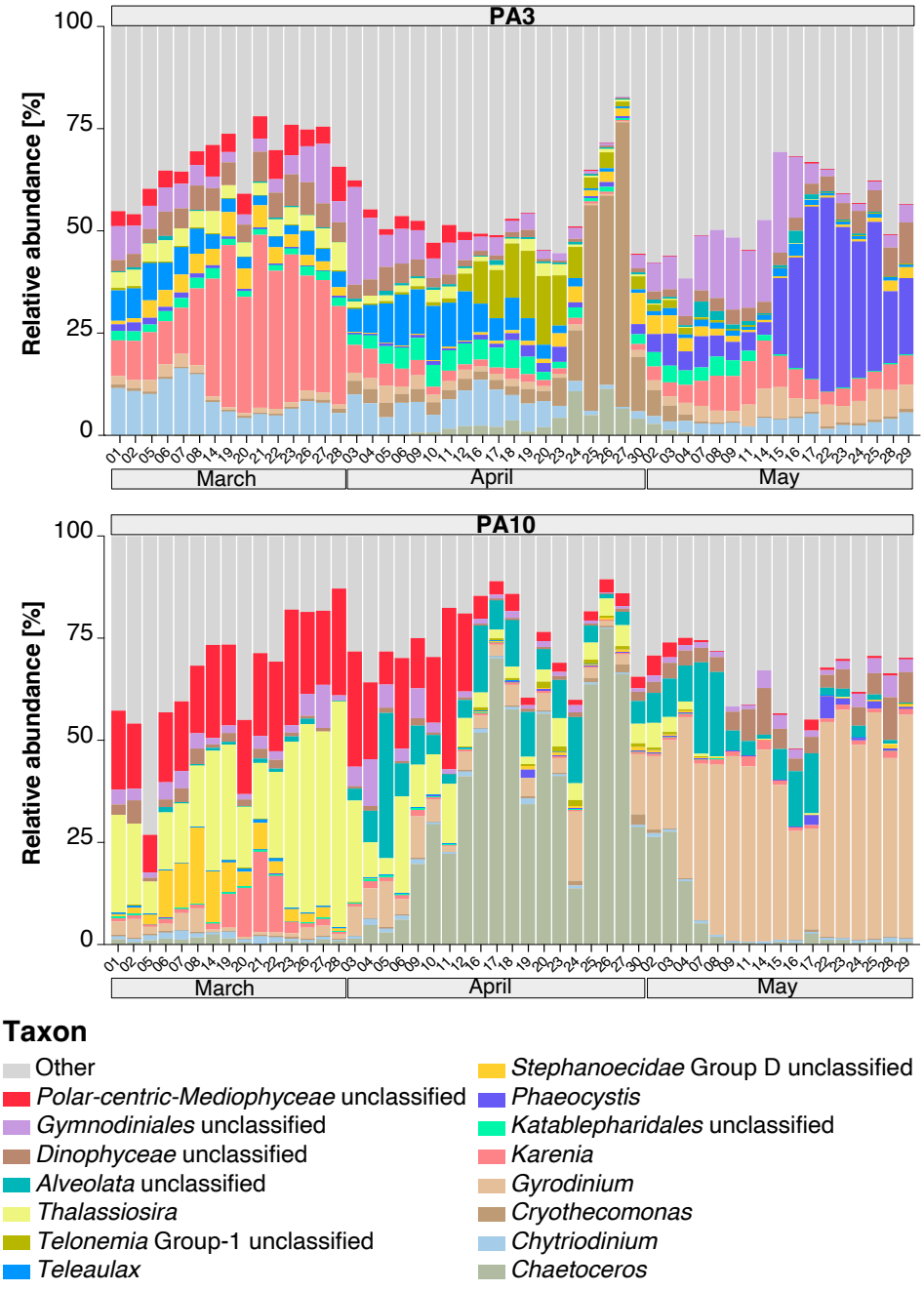


Fig. S2. Abundant 18S rRNA gene ASVs in PA fractions. Shown are the collective relative abundances of the 15 most abundant ASVs in each sample. PA3: 3-10 μm, PA10: >10 μm.

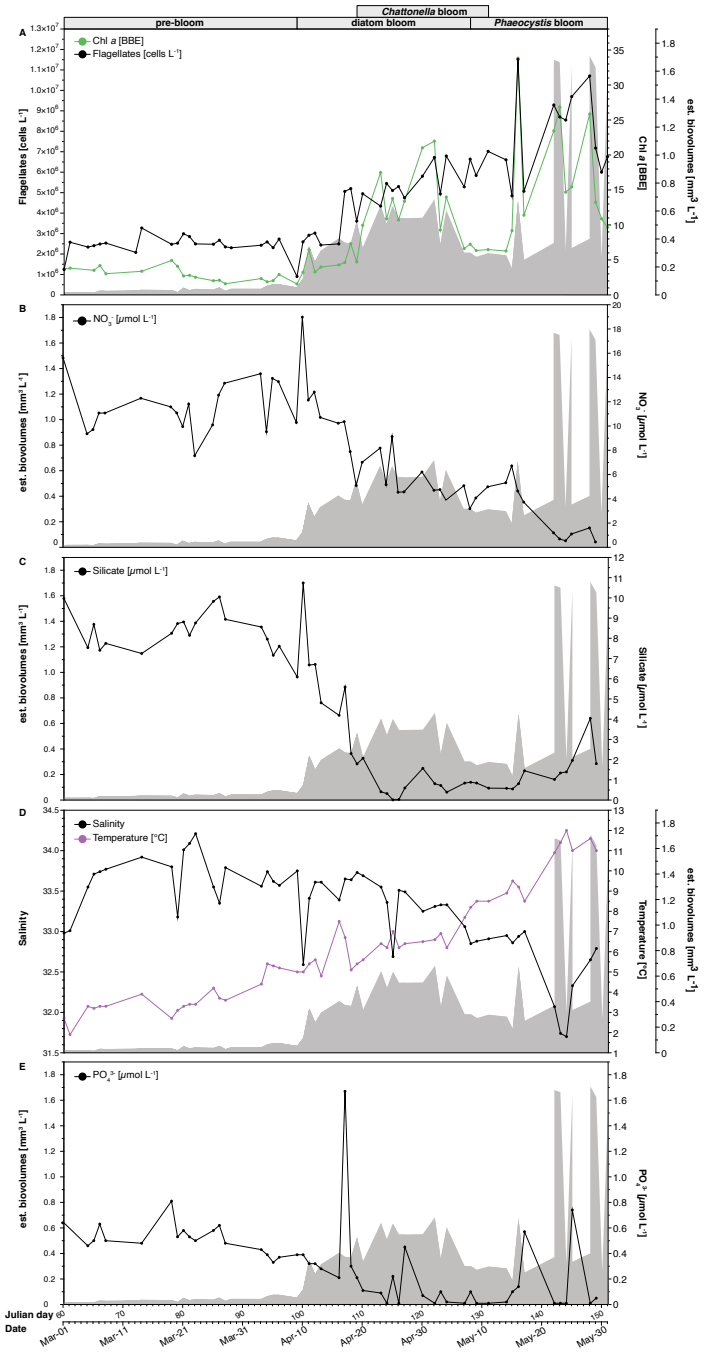


Fig. S3. Flagellate cell counts and physicochemical data. Data are represented by line graphs with biovolume estimates of abundant phytoplankton and non-photosynthetic dinoflagellates as background (see Fig. S1B). The distinct bloom stages are indicated on the top. Corresponding data is provided in Table S1 in Additional file 1.

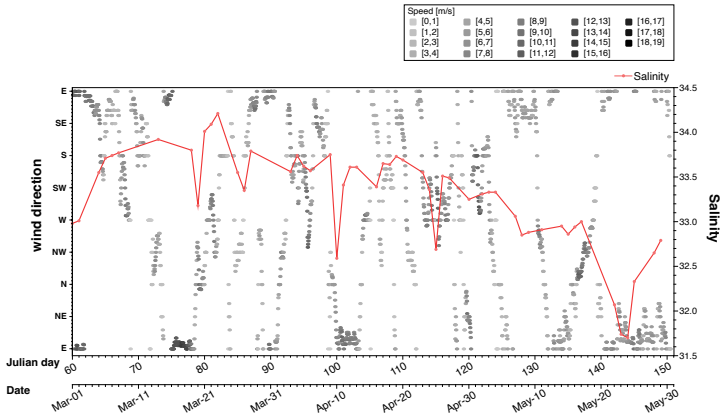


Fig. S4. Wind directional and salinity data at Helgoland Roads during March to May 2018. Wind components at 10 m above the sea surface were obtained from the Climate Data Store of the Copernicus Climate Change Service (ERA 5 product). Corresponding raw data is provided in Table S3 in Additional file 1. Additional weather data is provided in Table S4 in Additional file 1.

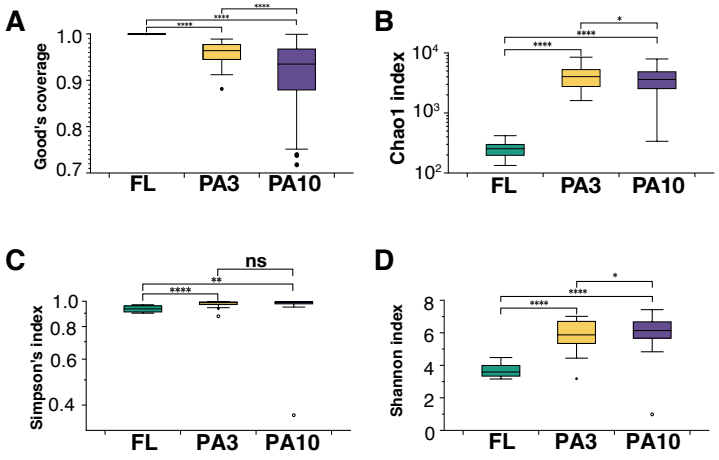


Fig. S5. Sample diversity indices based on 16S rRNA gene ASVs. Sample diversities of the bacterial communities of all size fractions (FL: 0.2-3 μm, PA3: 3-10 μm, PA10: >10 μm) were assessed by **A** Good's coverage, **B** Chao1, **C** Simpson and **D** Shannon indices. Statistical significance was assessed using one-way ANOVA (ns: not significant; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$).

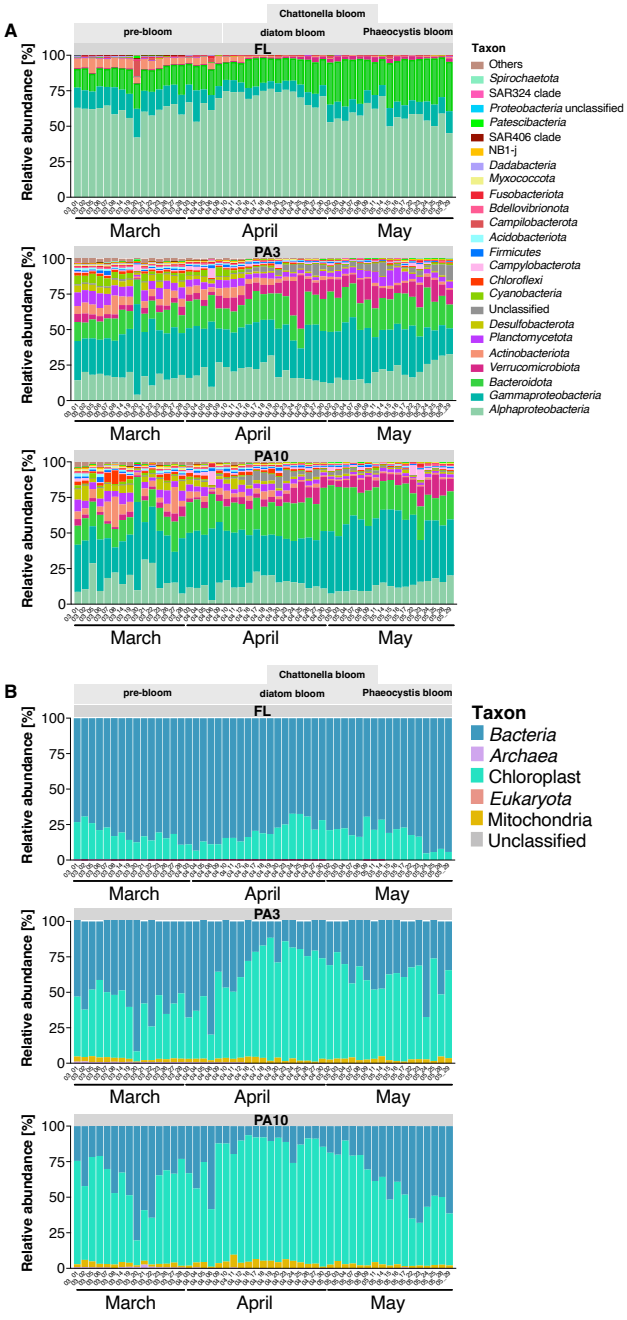


Fig. S6. Community composition as assessed by 16S rRNA gene amplicon data at higher taxonomic levels. A Phylum level taxa with *Proteobacteria* represented by *Alphaproteobacteria* and *Gammaproteobacteria* classes. **B** Domain level taxa. Shown are the collective relative abundances of the ten most abundant phyla in each sample. FL: 0.2-3 μm, PA3: 3-10 μm, PA10: >10 μm.

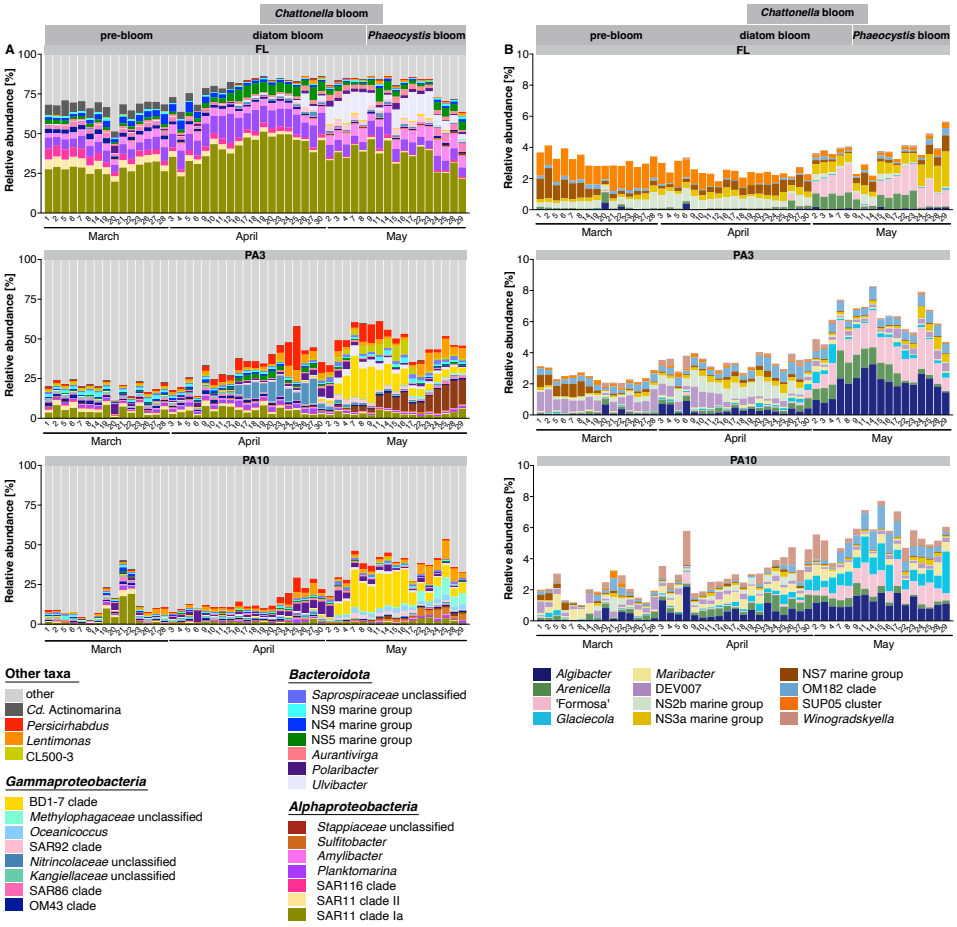


Fig. S7. Compositional differences among fractions as assessed by 16S rRNA gene amplicon data. A Collective relative abundances of the five most abundant genera in each sample for all three fractions (FL: 0.2-3 μm, PA3: 3-10 μm, PA10: >10 μm). Genera with high relative abundances in only few samples were subsumed as “other” in order to reduce complexity. Details are provided in Table S5 in Additional file 1. **B** Rarer genera with either clear preferences for distinct fractions or specific bloom phases.

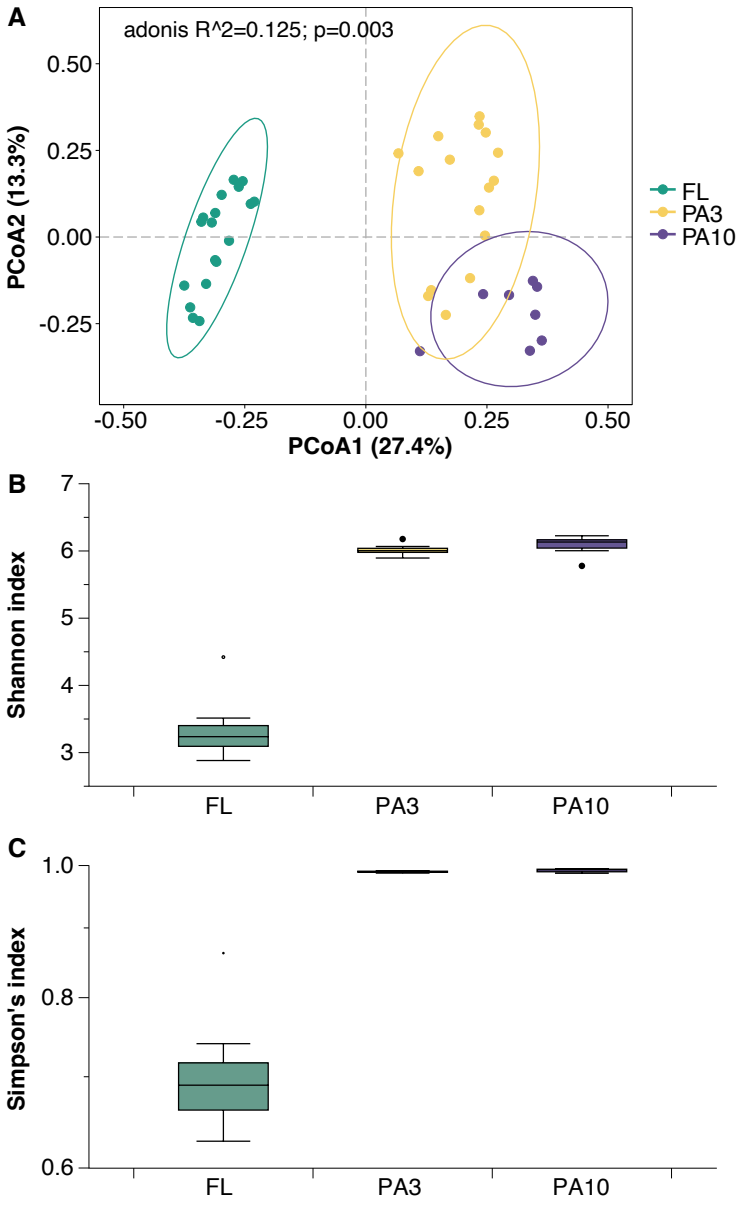


Fig. S8. Diversity analysis of 2018 Helgoland spring bloom metagenomes. **A** Principal component analysis of AB-Jaccard distances computed from k-mer read analysis in Simka v1.5.3 for 24 Illumina metagenomes (eight for each of the 0.2-3 μm FL, 3-10 μm PA3 and >10 μm PA10 size fractions). **B** Shannon and **C** Simpson's diversity indices of the sampled bacterial communities using the same 24 Illumina metagenomes. Corresponding data are provided in Table S14 in Additional file 1 and methodological details are provided in Additional file 3. Colors correspond to sample fractions.

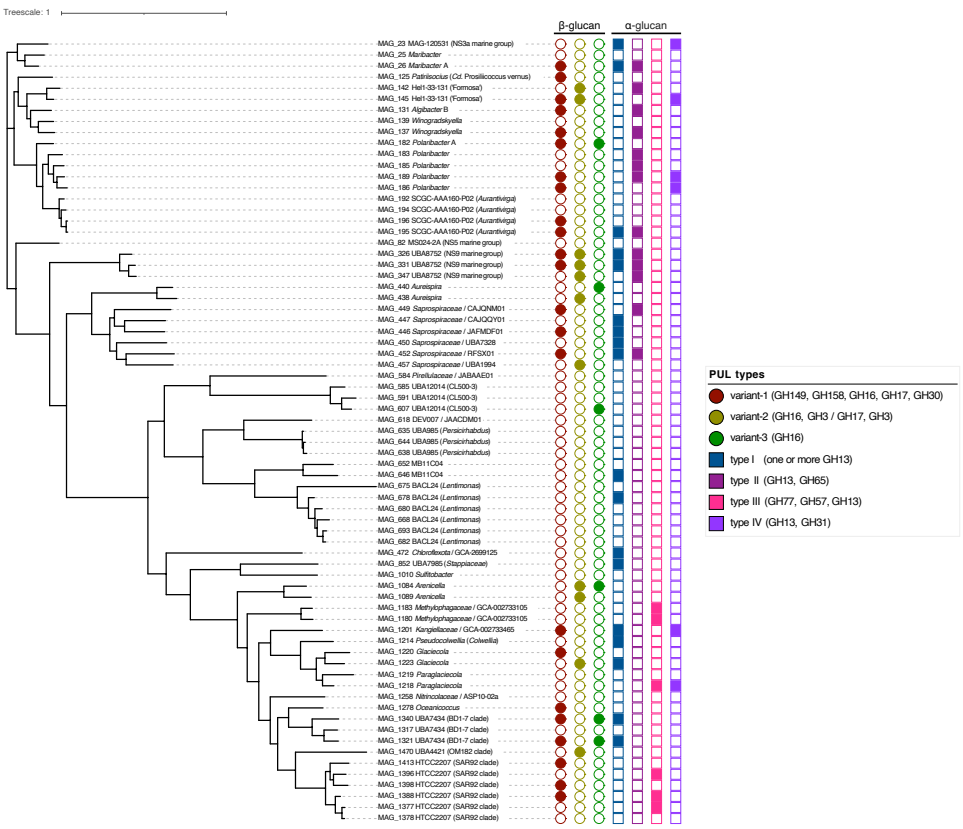


Fig. S9. α- and β-glucan PUL-types & CAZyme clusters in 71 selected abundant MAGs. β-glucan PULs - variant-1: GH149, GH158, GH16, GH17, GH30; variant-2: GH16, GH3/GH17, GH3; variant-3: GH16. α-glucan PULs - type I: one or more GH13; type II: GH13, GH65; type III: GH77, GH57, GH13; type IV: GH13, GH31.

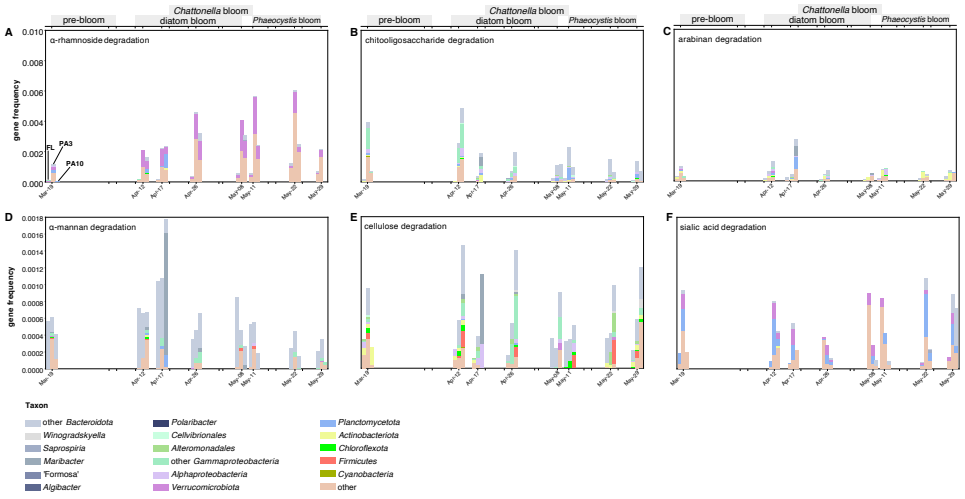


Fig. S10. Frequencies of CAZyme genes attributed to the degradation of A α-rhamnosides, B chitooligosaccharides, C arabinan, D α-mannan, E cellulose and F sialic acids. CAZymes for the degradation of predicted substrates were extracted from all metagenomes (Table S7 in Additional file 1), and gene frequencies were calculated as follows: frequency = Σ(average coverage of target genes) *100 / Σ(average coverage of all genes). Substrate predictions were annotated using the dbCAN3-sub database. Only dates with data for all fractions were plotted as stacked bar charts (left to right: FL: 0.2-3 μm, PA3: 3-10 μm, PA10: >10 μm). Dominating taxa are highlighted by colors.

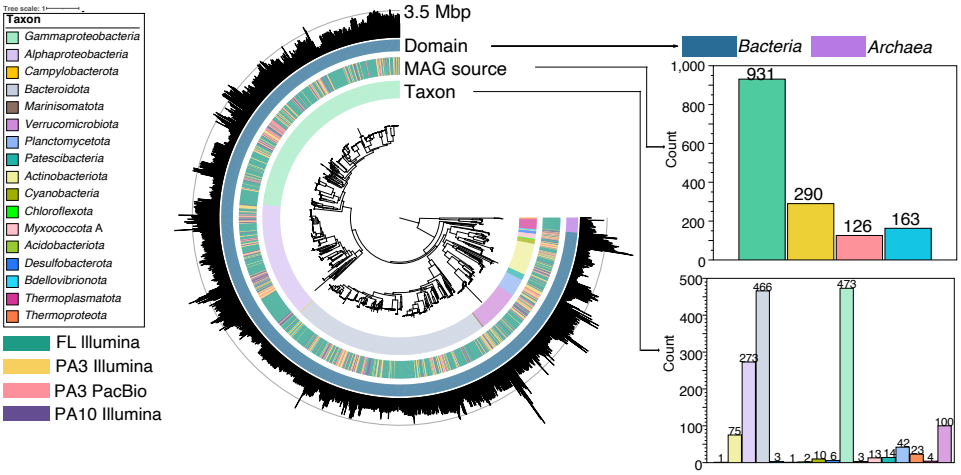


Fig. S11. Overview of the MAGs reconstructed from the metagenome dataset of the 2018 Helgoland spring bloom. Maximum-likelihood tree of 1,509 MAGs calculated in anvi'o v7.1 based on protein sequences of 38 universal single-copy genes with surrounding circles representing (inside to outside): (i) phylum-level taxonomy (class for *Proteobacteria*), (ii) size fraction of origin and sequencing platform (FL: 0.2-3 μm, PA3: 3-10 μm, PA10: >10 μm), (iii) domain-level taxonomy, (iv) circular bar chart representing MAG sizes (dotted line: 3.5 Mbp).

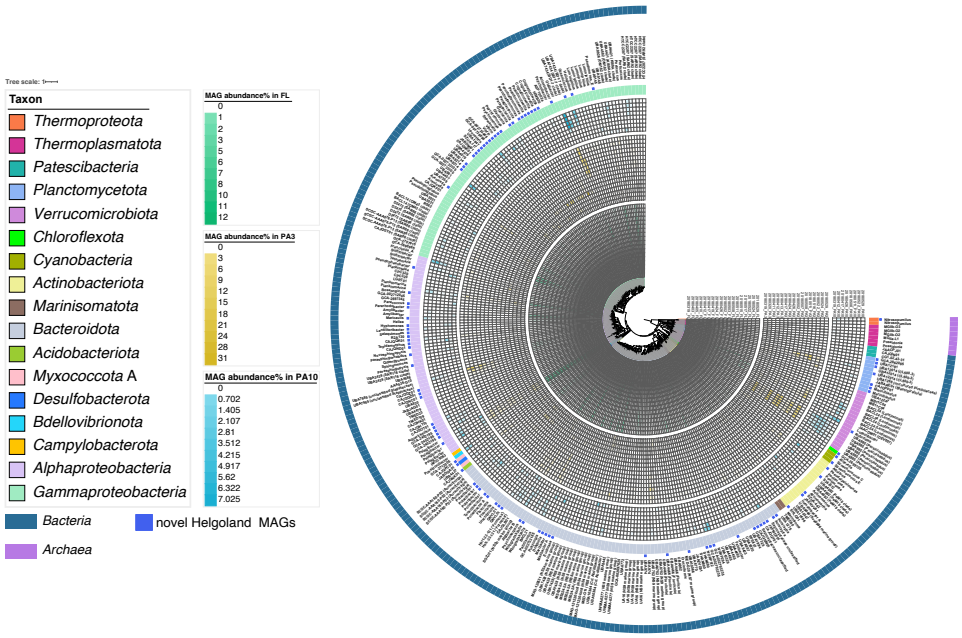


Fig. S12. Overview of dereplicated MAGs and their calculated relative abundances during the 2018 Helgoland spring bloom. Maximum-likelihood tree of 305 dereplicated MAGs calculated in anvi'o v7.1 based on protein sequences of 38 universal single-copy genes with surrounding circles representing (inside to outside): (i) FL MAG abundances (green), (ii) PA3 MAG abundance (yellow), (iii) PA10 MAG abundances (turquoise). (FL: 0.2-3 μm, PA3: 3-10 μm, PA10: >10 μm).

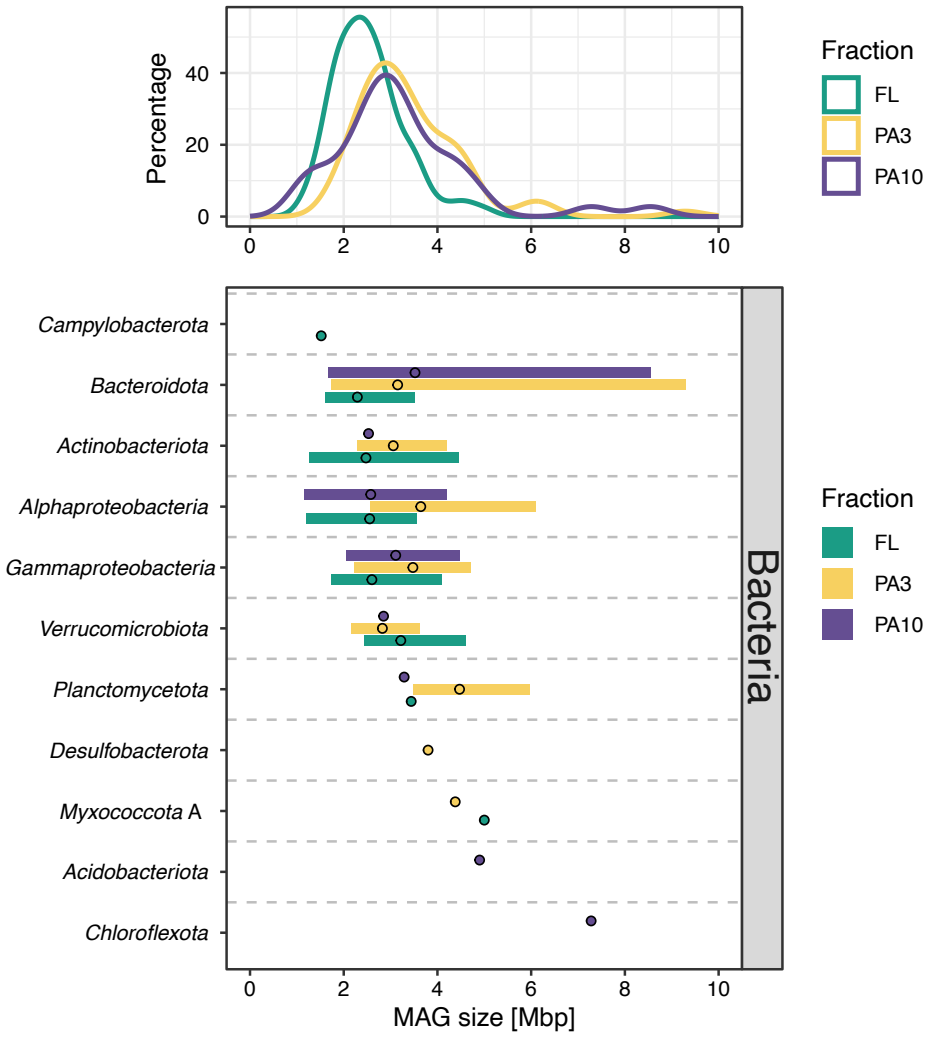


Fig. S13. Sizes of bacterial MAGs retrieved from the 2018 FL, PA3 and PA10 metagenomes. High-quality MAGs (n=186) were dereplicated within each fraction (FL: 0.2-3 μm, PA3: 3-10 μm, PA10: >10 μm) and used for comparison. Percentage in this context signifies the distribution of MAGs across various MAG sizes. Higher density values associated with a particular MAG size indicate a greater abundance of MAGs possessing that specific size.

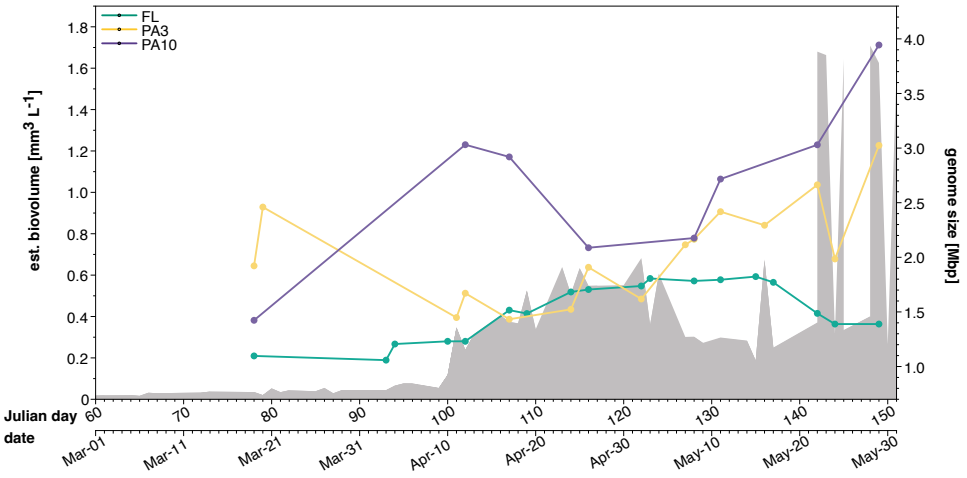


Fig. S14. Average sizes of the ten most abundant MAGs of each sample over time (n=136). De-replication was carried out within each fraction (FL: 0.2-3 μm, PA3: 3-10 μm, PA10: >10 μm), and MAGs belonging to the top ten in terms of abundance (as determined by MAG abundance) were specifically chosen for this analysis. The average value was derived by computing the average size of the ten most abundantly represented MAGs within the specified fraction for that particular date. The gray area in the background represents algal biovolumes including non-photosynthetic plankters (as in Fig. S1B).

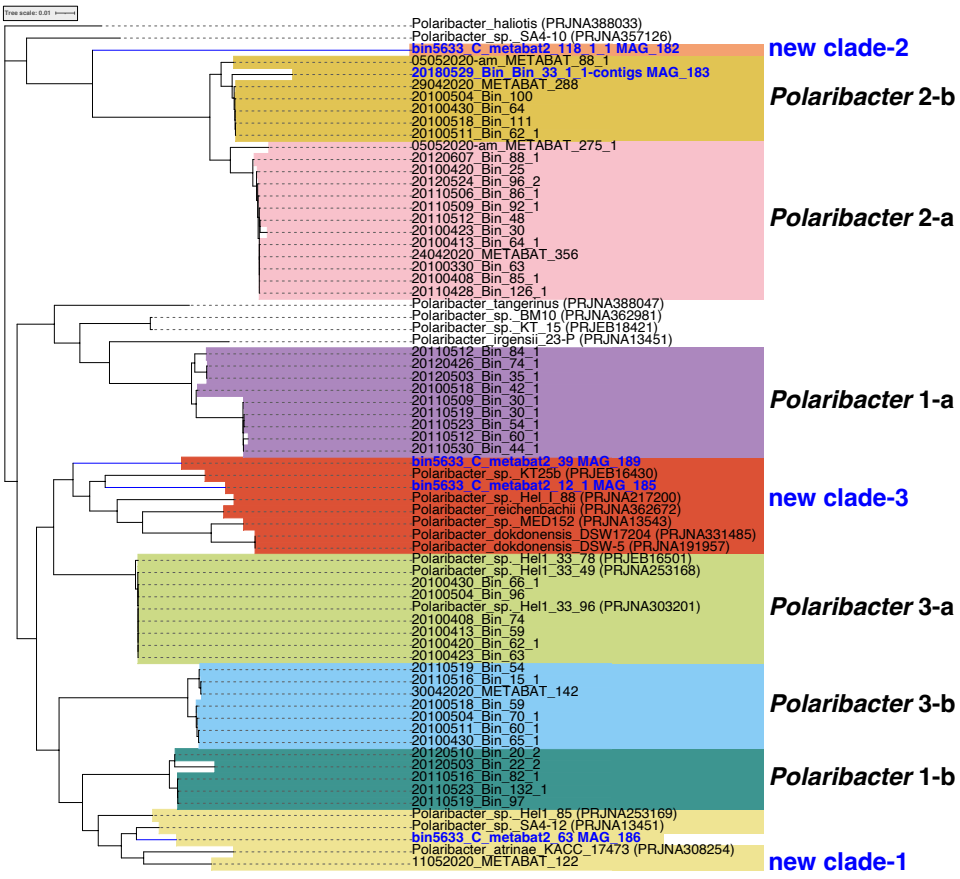


Fig. S15. Phylogenomic tree of *Polaribacter* MAGs from Helgoland Roads. Maximum-likelihood tree of all available *Polaribacter* MAGs from Helgoland Roads (n=52) built in anvi'o v7.1 based on protein sequences of 38 universal single-copy genes. This tree was constructed with data from a previous study as a reference [1]. MAGs from the 2018 datasets are highlighted in blue.

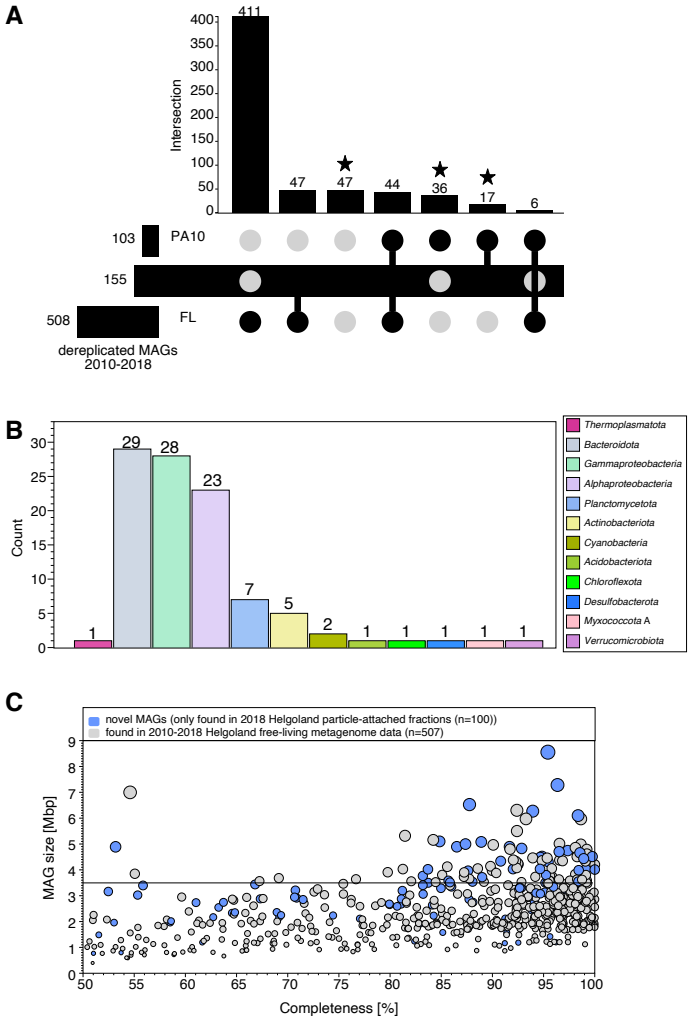


Fig. S16. Novel MAGs from Helgoland Road samples. We collectively analyzed MAGs from metagenomes that we sampled at Helgoland Roads during spring blooms in 2010, 2011, 2012, 2016 and 2018 (n=608). **A** MAG distribution per faction. Asterisks represent MAGs unique to the PA fractions of the 2018 data. **B** Phylum-level taxonomy of the novel 2018 PA MAGs (*Proteobacteria* are represented by the abundant *Alphaproteobacteria* and *Gammaproteobacteria* classes). **C** Sizes vs. completeness of all dereplicated MAGs (n=608). Circle area sizes correspond to MAG sizes. MAGs that were present only in the 2018 PA fractions are highlighted in blue. MAGs that were retrieved from the FL fractions of 2010, 2011, 2012, 2016 and 2018 are depicted in light gray. (FL: 0.2-3 μm, PA3: 3-10 μm, PA10: >10 μm).

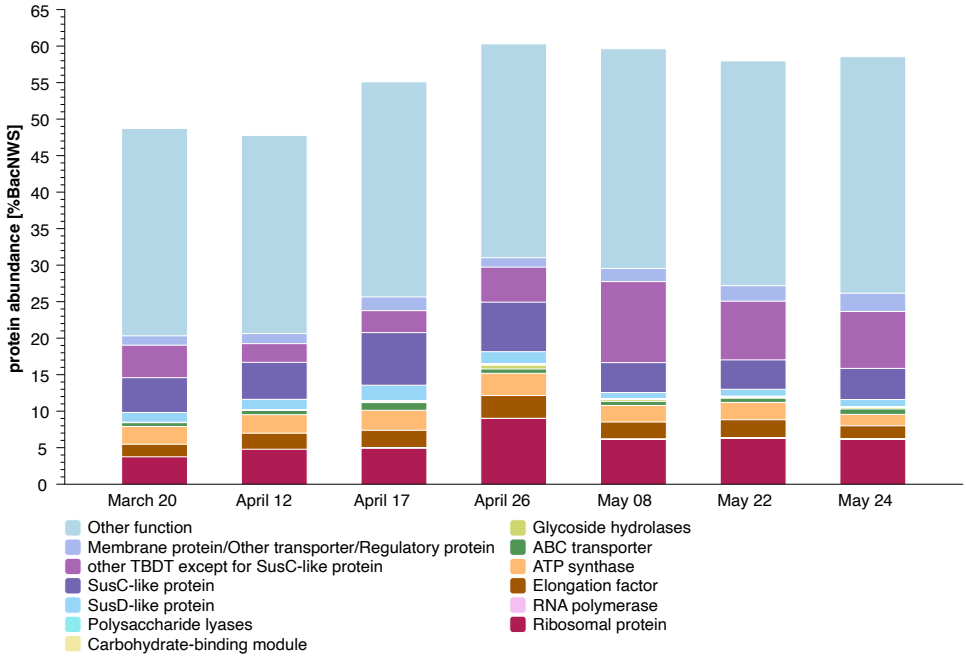


Fig. S17. Abundant expressed proteins in the seven FL metaproteomes sampled during the 2018 spring phytoplankton bloom at Helgoland Roads. Relative protein abundances are expressed as bacterial normalized weighted spectra (%BacNWS): We calculated the percent normalized weighted spectra (%NWS) for each protein group by dividing the 'Quantitative Value' obtained in Scaffold v4.11.1 by the sum of all quantitative values in the sample (average value for all three biological replicates for each sample). If a protein group was not identified in a replicate, we included it as '0' in the calculation. Values were normalized to 100% (%BacNWS) for comparability across samples. Noteworthy are the high proportions of SusC- and SusD-like proteins and other TonB-dependent transporters (TBDTs).

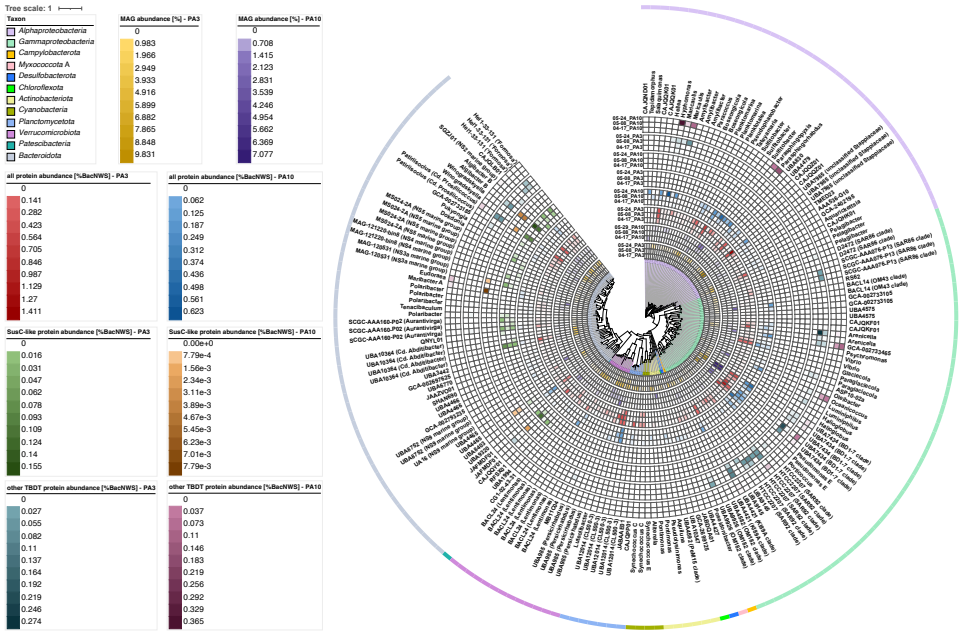


Fig. S18. Expression of bacterial MAGs from PA samples during the 2018 spring phytoplankton bloom at Helgoland Roads. The central maximum-likelihood tree of expressed MAGs obtained from PA3 and PA10 size fractions (n=183) was calculated in anvi'o v7.1 based on protein sequences of 38 universal single-copy genes. Predicted proteins of representative MAGs from all size fractions (FL: 0.2-3 μm, PA3: 3-10 μm, PA10: >10 μm) were used to search for the assignment of the protein fragments from proteome mass spectrometry. Circles represent (inside to outside): (i) PA3 MAG abundances, (ii) PA10 MAG abundances, (iii) PA3 MAG expression, (iv) PA10 MAG expression, (v) PA3 MAG expression of SusC-like proteins, (vi) PA10 MAG expression of SusC-like proteins, (vii) expression of TonB-dependent transporters (TBDTs) except for SusC-like proteins in PA3 MAGs, (viii) expression of TBDTs except for SusC-like proteins in PA10 MAGs, (ix) genus-level MAG taxonomy (GTDB), (x) colored arcs representing phylum-level MAG taxonomy (class for *Proteobacteria*).

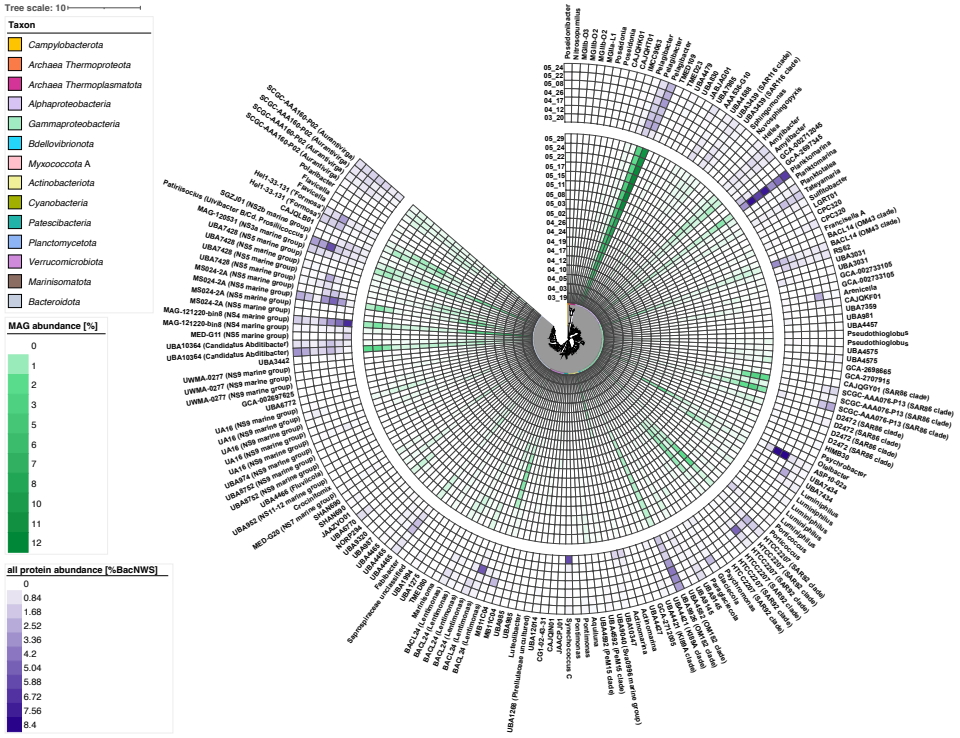


Fig. S19. Comparison of MAG abundance and MAG expression profiles in the FL fraction. The central maximum-likelihood tree of expressed FL (0.2-3 μm) MAGs (n=182) was calculated in anvi'o v7.1 based on protein sequences of 38 universal single-copy genes. Surrounding circles represent (inside to outside): (i) MAG abundance (green), (ii) overall expression based on summarized percentage of bacterial normalized weighted spectra (%BacNWS, purple), (iii) taxonomy as determined by GTDB r207_v2 (taxonomy in parentheses are based on Silva r138.1), (iv) phylum (class for *Proteobacteria*) level taxonomy represented by colored arcs. %BacNWS: We calculated the percent normalized weighted spectra (%NWS) for each protein group by dividing the 'Quantitative Value' obtained in Scaffold v4.11.1 by the sum of all quantitative values in the sample (average value for all three biological replicates for each sample). If a protein group was not identified in a replicate, we included it as '0' in the calculation. Values were normalized to 100% (%BacNWS) for comparability across samples.

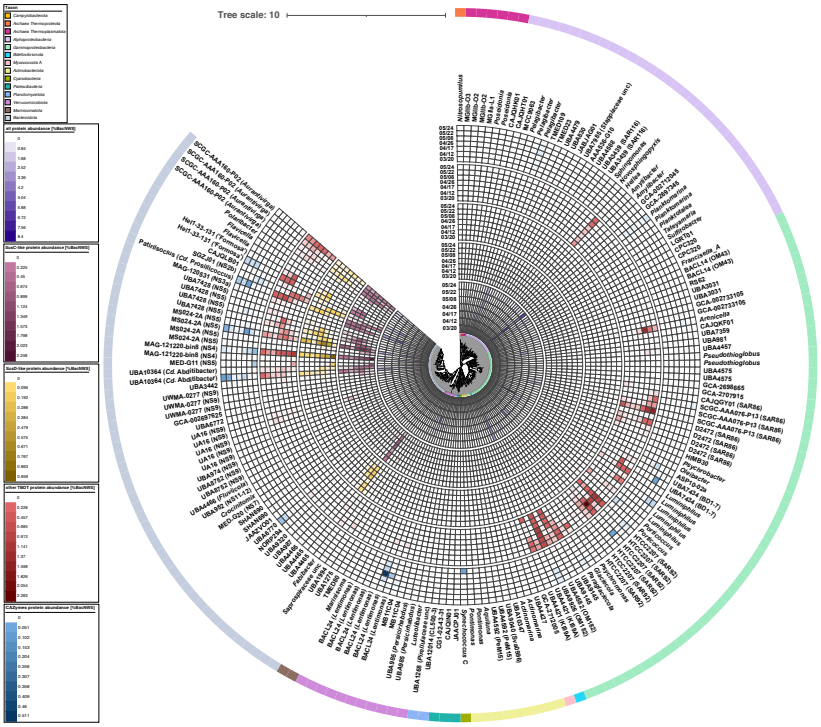


Fig. S20. Expressed bacterial MAGs in the FL fraction. The central maximum-likelihood tree of expressed FL (0.2-3 μm) MAGs (n=182) was calculated in anvio v7.1 based on protein sequences of 38 universal single-copy genes. Surrounding circles represent (inside to outside): (i) overall expression based on the summarized percentage of bacterial normalized weighted spectra (%BacNWS), (ii) expression of SusC-like proteins, (iii) expression of SusD-like proteins, (iv) expression of TonB-dependent transporters (TBDTs) except for SusC-like proteins, (v) expression of degradative CAZymes (GH, CE, PL), (vi) taxonomy as determined by GTDB r207_v2 (taxonomy in parentheses are based on Silva r138.1), (vii) phylum (class for *Proteobacteria*) level taxonomy represented by colored arcs. %BacNWS: We calculated the percent normalized weighted spectra (%NWS) for each protein group by dividing the 'Quantitative Value' obtained in Scaffold v4.11.1 by the sum of all quantitative values in the sample (average value for all three biological replicates for each sample). If a protein group was not identified in a replicate, we included it as '0' in the calculation. Values were normalized to 100% (%BacNWS) for comparability across samples.

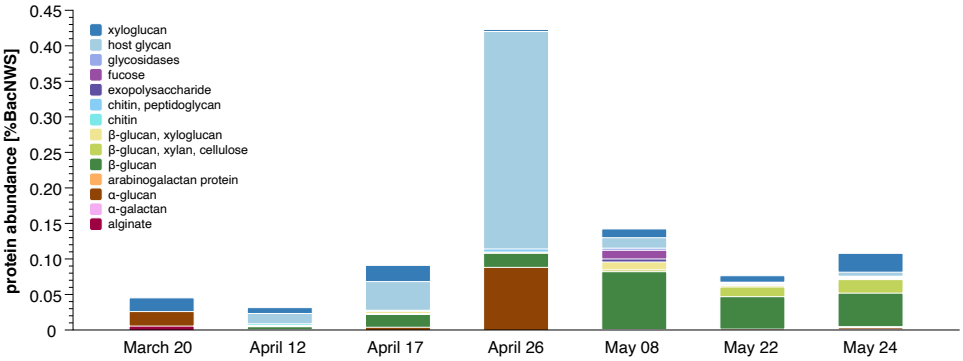


Fig. S21. Abundant expressed CAZymes in the seven FL metaproteomes sampled during the 2018 spring phytoplankton bloom at Helgoland Roads. Relative protein abundances are expressed as bacterial normalized weighted spectra (%BacNWS). Corresponding GHs, PLs and CEs included in CAZyme analyses were annotated with dbCAN3-sub and are listed in Table S12 in Additional file 1. %BacNWS: We calculated the percent normalized weighted spectra (%NWS) for each protein group by dividing the 'Quantitative Value' obtained in Scaffold v4.11.1 by the sum of all quantitative values in the sample (average value for all three biological replicates for each sample). If a protein group was not identified in a replicate, we included it as '0' in the calculation. Values were normalized to 100% (%BacNWS) for comparability across samples.

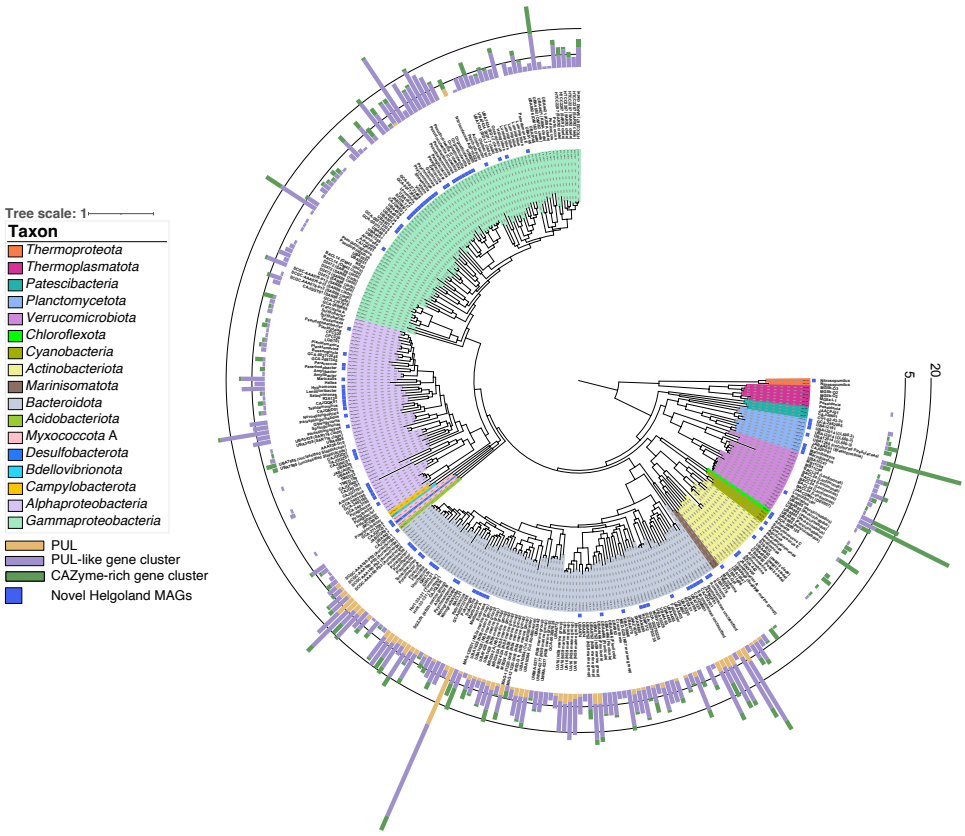


Fig. S22. Numbers of PULs, PUL-like and CAZyme-rich gene clusters of dereplicated MAGs. The central maximum-likelihood tree of 305 de-replicated bacterial MAGs was calculated in anvi'o v7.1 based on protein sequences of 38 universal single-copy genes. Surrounding circles represent (inside to outside): (i) phylum-level taxonomy (class for *Proteobacteria*), (ii) novelty with respect to our previous studies on Helgoland Roads (solid blue squares), (iii) stacked bar plots representing PUL counts, PUL-like clusters and CAZyme-rich gene clusters with black lines representing scales of 5 and 20.

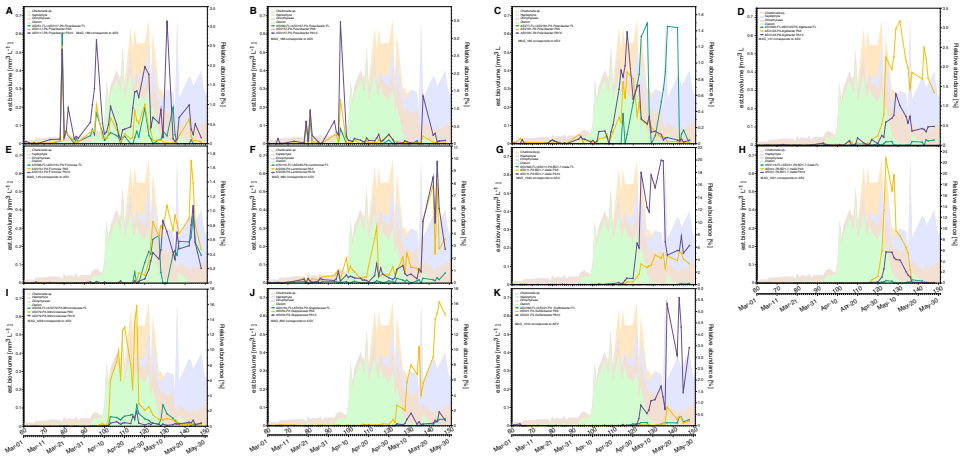


Fig. S23. Relative abundance of ASVs corresponding to MAGs. The four main algal groups are represented as colored areas, and abundant ASVs are represented by line graphs. Identical ASVs in different fractions are indicated in the legend (e.g., "ASV51-FL=ASV117-PA"). The MAG that corresponds to the ASVs in each panel is indicated (e.g., "MAG_186 corresponds to ASV").

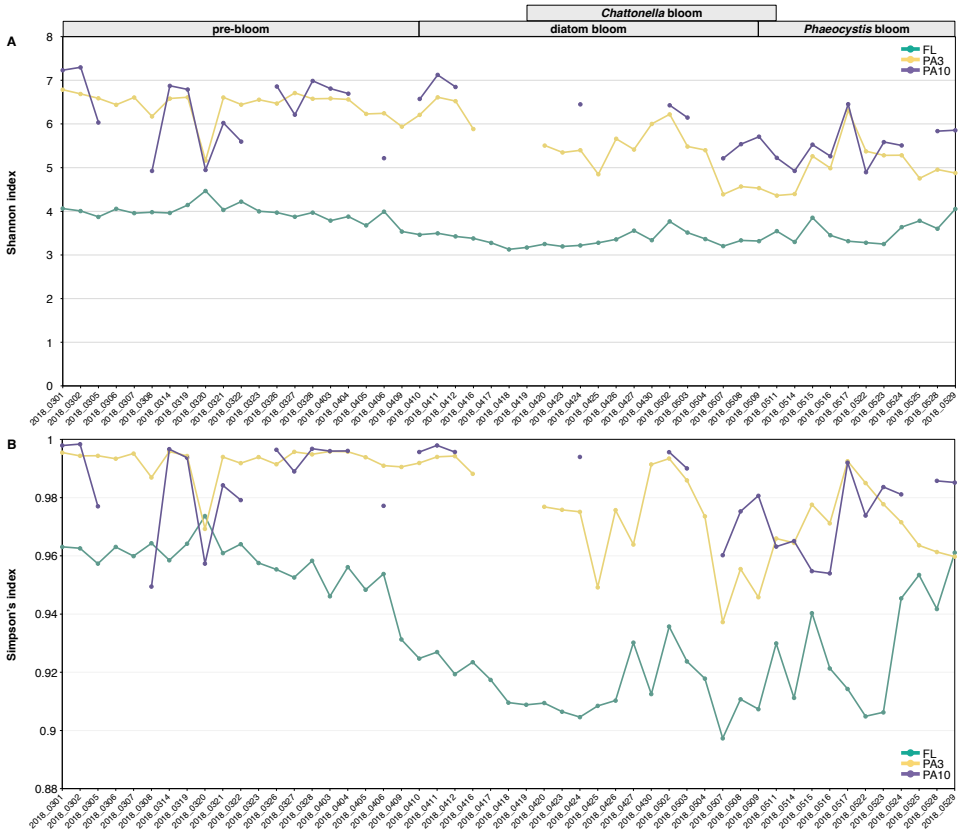


Fig. S24. Sample diversity indices based on 16S rRNA gene ASVs with rarefaction. Sample diversities of the bacterial communities with rarefaction of all size fractions (FL: 0.2-3 μm, PA3: 3-10 μm, PA10: >10 μm) were assessed by **A** Shannon and **B** Simpson's diversity indices. 10,000 ASVs were randomly selected. The "Rarefy" function in the R package GuniFrac [2] was used for rarefaction. Colors correspond to sample fractions.

Fig. S25. Example polysaccharide utilization loci. MAG affiliations for each contig are indicated. Colors of gene representations indicate gene types and corresponding numbers indicate CAZyme family associations. Substrates were predicted with dbCAN3-sub. Some PULs are incomplete. A possible frameshift is marked by an asterisk.

References

1. Avcı B, Krüger K, Fuchs BM, Teeling H, Amann RI. Polysaccharide niche partitioning of distinct *Polaribacter* clades during North Sea spring algal blooms. ISME J. 2020;14(6):1369-83.
2. Chen J, Chen MJ. Package 'GUniFrac'. The Comprehensive R Archive Network (CRAN). 2018.

Microbiome - Additional file 1

Particle-attached bacteria act as gatekeepers in the decomposition of complex phytoplankton polysaccharides

Feng-Qing Wang[1], Daniel Bartosik[2,3†], Chandni Sidhu[1†], Robin Siebers[4], De-Chen Lu[1,5], Anke Trautwein-Schult[4], Dörte Becher[4], Bruno Huettel[6], Johannes Rick[7], Inga V. Kirstein[7], Karen H. Wiltshire[7], Thomas Schweder[2,3], Bernhard M. Fuchs[1], Mia M. Bengtsson[4*], Hanno Teeling[1*], Rudolf I. Amann[1*]

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

[2] Institute of Pharmacy, University of Greifswald, Felix-Hausdorff-Straße 3, 17489 Greifswald, Germany

[3] Institute of Marine Biotechnology, Walther-Rathenau-Straße 49a, 17489 Greifswald, Germany

[4] Institute of Microbiology, University of Greifswald, Felix-Hausdorff-Straße 8, 17489 Greifswald, Germany

[5] Marine College, Shandong University, Weihai 264209, China

[6] Max Planck Genome Centre Cologne, Carl von Linné-Weg 10, 50829 Köln, Germany

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

[†] These authors contributed equally to this study.

* Corresponding authors:

Rudolf I. Amann, Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, e-mail: ramann@mpi-bremen.de, phone: +49 421 2028 9300

Hanno Teeling, Max Planck Institute for Marine Microbiology, Celsiusstraße 1, 28359 Bremen, e-mail: hteeling@mpi-bremen.de, phone: +49 421 2028 9760

Mia M. Bengtsson, University of Greifswald, Felix Hausdorff-Straße 8, 17489 Greifswald, Germany, e-mail: mia.bengtsson@uni-greifswald.de, phone: +49 383 4420 4212

E-mail addresses and telephone numbers of all authors:

Feng-Qing Wang	fwang@mpi-bremen.de	+49 421 2028 9390
Daniel Bartosik	daniel.bartosik@uni-greifswald.de	+49 3834 420 4918
Chandni Sidhu	csidhu@mpi-bremen.de	+49 421 2028 9582
Robin Siebers	siebersr@vodafonemail.de	+49 3834 420 5931
De-Chen Lu	202267000015@sdu.edu.cn	+86 0631 5688303
Anke Trautwein-Schult	anke.trautwein-schult@uni-greifswald.de	+49 3834 420 5922
Dörte Becher	dbecher@uni-greifswald.de	+49 3834 420 5903
Bruno Hüttel	huettel@mpipz.mpg.de	+49 221 5062 828
Johannes Rick	johannes.rick@awi.de	+49 4651 956 4220
Inga V. Kirstein	inga.kirstein@awi.de	+49 4725 819 3153
Karen H. Wiltshire	Karen.Wiltshire@awi.de	+49 4651 956 4112
Thomas Schweder	schweder@uni-greifswald.de	+49 3834 420 4212
Bernhard M. Fuchs	bfuchs@mpi-bremen.de	+49 421 2028 9350
Maria M. Bengtsson	mia.bengtsson@uni-greifswald.de	+49 421 2028 9390
Hanno Teeling	hteeling@mpi-bremen.de	+49 421 2028 9760
Rudolf I. Amann	ramann@mpi-bremen.de	+49 421 2028 9300

This file includes:

Supplementary Results

Phytoplankton bloom characteristics and physicochemical data

During the sampling period, surface water temperatures increased from 2.7 to 12 °C, and the salinity ranged from 31.7 to 34.2 (Fig. S3D in Additional file 2). The concentrations of silicate and nitrate decreased notably with the onset of the main bloom phase. Silicate was almost completely consumed at the peak of the diatom bloom (Fig. S3C in Additional file 2), because diatoms consume silicate to build their frustules (epitheca and hyptheca). In contrast, the decrease in nitrate concentrations was slower and reached its lowest values during the late bloom stage (Fig. S3B in Additional file 2). Since nitrate is the most important nitrogen source for phytoplankton growth, it can be assumed that the bloom went into terminal decline shortly after the end of the sampled time period.

Weather data

Rain was rather sparse during the sampling period. April 29 had the highest precipitation with 34 L/m^2 (Table S4 in Additional file 1), which did not have any detectable effect on salinity or any other of the data. In contrast, the continuous increase in sunshine duration in May notably promoted the phytoplankton bloom (Table S4 in Additional file 1).

Diversity analysis of FL and PA bacterial communities

The 16S rRNA gene amplicon data analysis discussed in the main manuscript was obtained with different primer sets for FL and PA bacteria (see Materials and Methods in the main manuscript). Furthermore, an uneven number of sequences was obtained for FL and PA fractions even after normalization due to high proportions of chloroplast sequences in PA data (Table S5 in Additional file 1). We hence conducted diversity analyses with rarefaction to account for different sample sizes. However, the results were similar to analyses without rarefaction (Figs. 2A-S24 in Additional file 2) but led to a loss of some of the undersampled PA10 points in time. We therefore included the analyses without rarefaction in the main manuscript (Fig. 2), even more so, since rarefaction is a debated process that may induce its own bias [1, 2].

As a second test, we compared the 16S rRNA gene amplicon data with independent

read-based taxonomic classifications (see supplementary Materials and Methods in this document). These data supported the 16S rRNA gene-based analyses, as they also revealed a substantially higher diversity of PA3 and PA10 communities compared to FL communities (Fig. S8B-C in Additional file 2), as well as more similar community compositions between PA3 and PA10 fractions compared to FL fractions (Fig. S8A in Additional File 2). The latter was also supported by the stark differences in the taxonomic affiliations of MAGs that were obtained from PA and FL communities. These results combined indicate that influences of different primer pairs and uneven data sizes were negligible compared to the pronounced differences in the most abundant taxa that we analyzed. Still, we do acknowledge the potential influence that the use of different primers during 16S rRNA gene sequencing may have for the analysis of less abundant or even rare taxa, for which in particular the PA10 data would benefit from additional sequencing depth.

Composition of FL and PA bacterial communities

After removal of chloroplast and mitochondria sequences from merged 16S rRNA gene amplicon data, 2,546 and 18,879 bacterial ASVs were retained for the FL and PA fractions, respectively. FL ASVs represented 27 phyla, 43 classes, 126 orders, 153 families and 263 genera. PA ASVs represented 68 phyla, 149 classes, 361 orders, 498 families and 1,151 genera.

For each size fraction, we investigated ASVs present in at least 80% of the samples with a relative abundance exceeding 0.1% in at least one sample. Such common ASVs represented 11.7-85.2% (FL, 45 ASVs), 47.3-77.1% (PA3, 242 ASVs), and 14.4-84.2% (PA10, 127 ASVs), respectively (Table S5 in Additional file 1).

In FL communities, members of the SAR11 clade II, SAR116 clade, unclassified *Rhodobacteraceae*, OM43 clade, NS4 marine group, and *Cd.* Actinomarina were abundant during the pre-bloom and early diatom-dominated main bloom phases, but their abundances decreased during the diatom bloom's decline and the late surge in haptophytes (Fig. S7A in Additional file 2). Similar trends were observed for members of the SAR11 clade II, SAR116 clade and NS4 marine group with lower relative abundances in both PA fractions. Contrary to the FL fraction, Cd. Actinomarina ASVs were hardly detectable in PA fractions (FL: 0.1-10.3%, PA3: 0-0.4%, PA10: 0-1.2%).

ASV level composition of FL and PA bacterial communities

ASV level taxonomic assignments that are discussed in this section are summarized in Table S5 in Additional file 1.

SAR11 clade Ia

In the FL fraction, there were two very abundant ASVs affiliating with the SAR11 clade Ia (ASV_1, ASV_2). ASV_1 slightly declined during the late diatom and *Phaeocystis* bloom phases, whereas ASV_2 was more abundant during these bloom phases. Notably, no SAR11 clade Ia ASV with the same dynamic change was found in PA fractions. There were also ASVs with high abundances only during the pre-bloom, which declined (relatively) with the onset of the bloom. These ASVs were found in both, FL and PA fractions (ASV_25 and ASV_33 in the FL, and ASV_52 in the PA fractions).

Planktomarina

The genus *Planktomarina* was dominated by a single ASV in all fractions (FL ASV_3 was 100% similar to PA ASV_15).

Amylibacter

Two *Amylibacter* ASVs (ASV_7, ASV_17) had higher abundance in the FL fraction, with ASV_7 exhibiting higher abundance throughout all sampled points in time, and ASV_17 with higher abundance only during the *Phaeocystis* bloom. The latter ASV was not detected in the PA fractions.

BD1-7 clade

There were two dominant BD1-7 clade ASVs in the PA fractions. ASV_41 was more abundant in the PA3 (0-22.7%), while ASV_11 was more abundant in the PA10 (0-20.0%) fraction. Both were abundant during the late diatom and *Phaeocystis* bloom phases.

Ulvibacter

A single ASV (ASV_6) dominated *Ulvibacter* in the FL size fraction (0-17.6%), with

highest relative abundances during the late diatom and early *Phaeocystis* bloom phases. In PA fractions, ASV_46 (100% similarity with FL ASV_6) was dominant in PA3, but with lower relative abundance (0-6.6%).

unclassified Nitrincolaceae

Unclassified *Nitrincolaceae* were dominated by a single ASV (ASV_79) in PA3 fractions with relative abundances of up to 15.8%. This clade was only abundant during the diatom bloom phase.

unclassified Stappiaceae

Unclassified *Stappiaceae* were dominated by ASV_34 in PA3 fractions with relative abundances of 0-16.3%. The highest relative abundances were observed during the *Phaeocystis* bloom phase.

Persicirhabdus (Verrucomicrobiota)

Two ASVs (ASV_29, ASV_32) dominated *Persicirhabdus* in PA fractions. Both were more abundant in the PA3 (0.1-18% and 0.02-7.8%) than in the PA10 fraction (0-6% and 0-2.8%).

CL500-3 (Planktomycetota)

A single ASV (ASV_99) dominated the CL500-3 clade. It reached up to 8.6% relative abundance in PA3 fractions during the late diatom and early *Phaeocystis* bloom phases.

Sulfitobacter

Sulfitobacter was represented by a single dominant ASV (ASV_21) in PA10 size fractions with up to 5.6% relative abundance, specifically during the *Phaeocystis* bloom phase.

'Formosa'

In the FL size fractions, 'Formosa' was represented by two ASVs with similar relative abundances (ASV_79: 0-1.2%, ASV_96: 0-0.9%), in particular during the late diatom

and early *Phaeocystis* bloom phases. Additional distinct two 'Formosa' ASVs were present in the PA3 fractions (ASV_271: 0-1.0%, ASV_151: 0-1.7%).

Algibacter

In the genus *Algibacter* a single ASV (ASV_123) dominated. It was present in PA3 fractions with up to 3.2% relative abundance, specifically during the late diatom and early *Phaeocystis* bloom phases.

Polaribacter

The *Polaribacter* genus was dominated by five ASVs. ASV_117, ASV_218, ASV_26 and ASV_181 were abundant in both PA3 (0-2.5%, 0-2.2%, 0-2.0% and 0-0.9%, respectively) and PA10 fractions (0-3.1%, 0-1.9%, 0-2.0% and 0-1.4%, respectively), whereas ASV_1063 was detected in PA10 (0-1.1%) but barely in PA3 (0-0.4%) fractions. All five *Polaribacter* ASVs exhibited their highest relative abundances during the diatom bloom.

Glaciecola

The genus *Glaciecola* was dominated by a single ASV (ASV_381) with a relative abundance of 0-1.8% in the PA10 size fraction during the late diatom and *Phaeocystis* bloom phases.

Reconstruction of MAGs

MAGs with an ANI ≥95% were combined to species-level clusters and subsequently taxonomically assigned. *Bacteroidota* (96 clusters, 466 MAGs), *Gammaproteobacteria* (84 clusters, 473 MAGs) and *Alphaproteobacteria* (57 clusters, 273 MAGs) represented the most abundant taxa, followed by *Actinobacteriota* (17 clusters, 75 MAGs), *Verrucomicrobiota* (17 clusters, 100 MAGs), *Planctomycetota* (10 clusters, 42 MAGs), *Patescibacteria* (3 clusters, 14 MAGs), *Cyanobacteria* (3 clusters, 10 MAGs), *Myxococcota* (2 clusters, 13 MAGs), *Marinisomatota* (2 clusters, 3 MAGs), *Desulfobacterota* (1 cluster,6 MAGs), *Bdellovibrionota* (1 cluster, 3 MAGs), *Chloroflexota* (1 cluster, 2 MAGs), *Acidobacteriota* (1 cluster, 1 MAG), *Campylobacterota* (1 cluster, 1 MAG), *Thermoplasmatota* (Archaea, 6 clusters, 23

MAGs) and *Thermoproteota* (Archaea, 2 clusters, 4 MAGs).

152 abundant PA MAGs

Based on ASV and MAG abundance data, we categorized MAGs into those that were most abundant in either FL or PA communities and those that were abundant in both. A total 152 MAGs of those that could be taxonomically were most abundant in PA communities. These MAGs comprised *Bacteroidota* (45 MAGs), *Gammaproteobacteria* (37 MAGs), *Alphaproteobacteria* (29 MAGs), *Verrucomicrobiota* (11 MAGs), *Actinobacteriota* (10 MAGs), *Planctomycetota* (10 MAGs), *Cyanobacteria* (3 MAGs), *Myxococcota* A (2 MAGs), *Acidobacteriota* (1 MAG), *Bdellovibrionota* (1 MAG), *Chloroflexota* (1 MAG), *Desulfobacterota* (1 MAG) and *Thermoproteota* (1 MAG).

MAG sizes as a function of size fraction, taxonomy and time

Average HQ MAGs sizes were larger in PA than in FL communities (Fig. S13 in Additional file 2), most notably for *Bacteroidota* and *Gammaproteobacteria*. Exceptions were *Alphaproteobacteria* and *Planctomycetota* with similar sizes in FL and PA10, but larger MAGs in PA3 communities, and *Verrucomicrobiota* with larger average MAG sizes in FL communities. The largest MAG was found in the *Bacteroidota* (*Saprospiraceae* MAG_98, 8.6 Mbp), followed by a MAG from the *Chloroflexota* (*Promineofilaceae* MAG_535, 7.3 Mbp), both from PA10 samples. However, on overall, median MAG sizes of FL community members of all phyla except for *Myxococcota* A (5.0 Mbp) were below the average aquatic genome size of 3.1 Mbp [3].

We observed an increase in size within the FL communities during the diatom bloom (Fig. S14 in Additional file 2). This trend reversed amidst the *Phaeocystis* bloom phase, likely due to increased abundances of SAR86, a clade characterized by relatively small genome sizes (0.7 to 1.5 Mbp in our data). Different trends were observed for both PA communities. Average MAG sizes of PA3 communities increased during the diatom and *Phaeocystis* bloom phases, whereas sizes of PA10 communities first decreased during the diatom bloom and increased during the *Phaeocystis* bloom phases. In general, fluctuations over time were more pronounced

for both PA communities, likely due to the natural variability of particulate matter on the filters themselves.

Categories of expressed proteins of FL communities

Of the combined detected proteins in the metaproteomes of FL bacteria, 15,906 (36.4%) could be assigned to 177 of the 182 dereplicated FL MAGs. Annotated functions comprised 2,565 ribosomal proteins, 718 ABC transporters, 704 TBDTs including 103 SusC-like proteins, 347 CAZymes and 80 SusD-like proteins (Fig. S17 in Additional file 2).

Expression of SusC-like proteins and other TBDTs

In the 177 MAGs with mapped metaproteome data, we identified 168 SusC-like proteins and 2,071 other TBDTs, of which 103 SusC-like proteins and 601 other TBDTs were expressed. Combined SusC-like relative protein abundances peaked on April 17 (7.2%) and April 26 (6.8%) during the diatom-bloom, then decreased on May 8 (4.1%), May 22 (4.0%), and May 24 (4.3%), which was even lower than the expressed SusC-like proteins during the pre-bloom on the March 20 (4.8%) and during the onset of the diatom-bloom on the April 12 (5.1%). SusD-like proteins exhibited a similar expression pattern but with lower relative protein abundances (0-0.06%). The relative expression of other TBDTs decreased from March 20 (4.6%) to April 12 (2.6%) and April 17 (3.0%). Afterwards, TBDT expression started to increase again until pre-bloom levels on April 26 (4.8%), and then continued to its highest relative expression on May 8 (11.1%). After this peak, values decreased again (May 22: 8.0%, May 24: 7.8%) but stayed above pre-bloom levels.

In *Bacteroidota*, SusC-like protein expression corresponded well with overall MAG abundances (Fig. S20 in Additional file 2) with the exception of *Cd.* Abditibacter (MAG_401) which despite high overall relative abundance featured no detectable SusC-like protein expression. The expression of transporter proteins differed during different bloom periods (Fig. S17 in Additional file 2). During the diatom-dominated bloom phase on April 12, 17 and 26, SusC-like proteins were mainly expressed by the NS5 marine group, UBA4465 and UBA7428. On May 8, expressed SusC-like proteins were more from *Aurantivirga*, *Cd.* Prosiliicoccus, and 'Formosa'. SusC-like proteins

from the NS3a marine group and *Cd.* Abditibacter expressed more on May 22 and 24. The shifts in other TBDT expression profiles were more pronounced, with a significant increase on May 8, which was dominated by other TBDTs from the SAR92 clade, *Arenicella* and *Cd.* Prosiliicoccus. On May 22 and 24, the dominating MAGs with expressed TBDT shifted to the SAR86 clade.

High SusC-like protein expression in *Flavobacteriaceae* included MAG_146 (NS2b marine group), MAG_142 ('Formosa'), MAG_131 (*Winogradskyella*), MAG_104 (*Algibacter* B), MAG_98 (*Dokdonia*), MAG_41 (NS5 marine group), MAG_183 (*Tenacibaculum*), MAG_195 (*Polaribacter*), MAG_176, MAG_194 and MAG_196 (*Aurantivirga*), and in *Saprospiraceae* MAG_446 (JAFMDF) and MAG_452 (RFSX01). In addition, MAG_26 (*Eudoraea)* featured the highest SusC-like protein expression (0.008%) in PA10 on May 24. MAG_434 (UBA4465) exhibited high SusC-like protein expression on May 8 and May 24 in PA10 and May 24 in PA3.

MAG_24 (NS4 marine group) had some expression of other TBDTs in PA10 on May 24. MAG_680 (*Lentimonas*), and MAG_585 and MAG_607 (*Planctomycetota;* UBA12014) expressed TBDTs in PA10 and PA3 fractions. Further TBDT expression was mostly observed in *Alpha-* and *Gammaproteobacteria*, for example in MAG_1474 (UBA4421 / KI89A clade), MAG_1396, MAG_1406 and MAG_1413 (SAR92), MAG_1278 (*Oceanicoccus*), MAG_1201 (GCA-002733465), MAG_1093 (*Arenicella*), MAG_1075 (SAR86), MAG_865 (*Parasphingopyxis*), MAG_869 (*Parasphingorhabdus*), MAG_887 (*Maricaulis*) and MAG_886 (*Hyphomonas*).

MAG_1005 (*Sulfitobacter*) expressed periplasmic and bacterial extracellular solute binding proteins on May 24 in both PA3 and PA10 communities. MAG_1018 (*Tateyamaria*) expressed porins and bacterial extracellular solute binding proteins on May 8 and 24 in both PA3 and PA10 communities (Fig. S18 in Additional file 2).

Expression of catabolic CAZymes in FL communities

GHs showed the highest relative expression on the April 26 (0.5%), which included the highest relative expression of α-fucose-containing degradation genes (GH95 and GH29), α-mannan-containing (GH92) and α-glucan-containing (GH13_9lCBM48, GH13_3 and GH13_14) substrates (Fig. S20 in Additional file 2). Expressed GH29 was detected in MAGs of the NS5 marine group and BACL24 (*Lentimonas*).

Expressed GH95 and GH92 were only detected in a BACL24 (*Lentimonas*) MAG. Laminarin-targeting GH16_3 and GH149 exhibited particularly high relative protein abundances on May 8 (0.07% each), May 22 (0.04%, 0.03%) and May 24 (0.04% each) when compared to the other four metaproteome sample dates. GH16_3 and GH149 were expressed in several MAGs, not only in *Bacteroidota* but also in *Gammaproteobacteria*. The relative abundance of these MAGs increased during the late diatom and *Phaeocystis* bloom phases. Expressed GH33 was only detected on May 22 (0.02%) and May 24 (0.03%), while expressed GH74 was detected across all samples (0.01-0.1%). Expressed PL22 was also detected on all the samples except for March 20.

MAG analyses highlight distinct polysaccharide degradation potentials in abundant FL and PA communities

We linked MAGs with 16S rRNA gene amplicon data to leverage the high temporal resolution amplicon data to uncover variations in MAG abundances (Table S10 in Additional file 1, Fig. S23 in Additional file 2), for which we selected the 71 most abundant MAGs for in-depth PUL analysis. Nine of these harbored 40 or more CAZyme genes, all of which were prevalent in PA communities (Fig. 7). A holistic summary of the main results is provided in the discussion of the main manuscript. Here, we describe the key CAZyme genes and inferred substrates of the most prominent MAGs.

Order Flavobacteriales

Maribacter (2 MAGs): MAG_26 was abundant in PA10 communities and showed no clear abundance trend, except for a decrease during the *Phaeocystis* bloom. It possessed 24 predicted PULs and 133 CAZyme genes assigned to 18 substrates, including five PULs predicted to target complex host glycans (e.g., Fig. S25). Additional PULs were predicted to target α-glucans, β-glucans, xylans, xyloglucans, cellulose, and arabinogalactans (Fig. S25). In contrast, *Maribacter* MAG_25 featured only a single fructan (GH32) PUL, showcasing a pronounced intra-genus-level variation in PUL repertoires.

Polaribacter (5 MAGs including *Polaribacter* A): Three of the five *Polaribacter*

MAGs correlated with distinct ASVs (Fig. S23A-C in Additional file 2). MAG_185 exhibited no discernible preference for any fraction, whereas MAG_186 and 189 were more prevalent in PA10 communities. *Polaribacter* PA MAGs correlated with the diatom bloom phase, as their abundances increased amidst the diatom phase and decreased when the diatom bloom waned. During the early *Phaeocystis* bloom, MAG abundances increased from their nadir and then rapidly declined again. MAG_186 correlated well with diatom and *Phaeocystis* biovolume estimates, while MAG_189 displayed high relative abundance in a few samples. All *Polaribacter* MAGs featured PULs for α-glucan, β-glucan, and alginate, but MAG_186 (19 predicted PULs and PULs-like clusters) and MAG_189 featured further PULs including putative fucoidan PULs (Fig. S25).

Algibacter (1 MAG): MAG_131 represented another PA clade, whose abundance increased rapidly and peaked during the diatom bloom's collapse, and then decreased during the *Phaeocystis* bloom to 2% abundance (Fig. S23D in Additional file 2). It featured 15 predicted PULs and PULs-like clusters and 53 CAZyme genes, including two predicted PULs for α-glucan and one for laminarin, as well as abundant genes for host glycans.

'Formosa' (2 MAGs): 'Formosa' MAG_145 consistently increased during the diatom bloom's decline and peaked during the *Phaeocystis* bloom in FL and PA communities (Fig. S23E in Additional file 2). In contrast, MAG_142, represented a FL member of the same genus as the previously cultured Helgoland strain Hel1_33_131 [4]. It was abundant during and after the diatom/*Chattonella* bloom, corroborating previously detected recurrence during coastal North Sea spring phytoplankton blooms [5]. Both, MAG_142 and 145 contained similar CAZyme gene numbers (n=30-33), with MAG_142 representing the FL MAG with the highest predicted number of PULs and PULs-like clusters (n=15). Both MAGs featured PULs for α-glucan, β-glucan, host glycans and cellulose, with MAG_142 featuring additional genes to target sulfated xylans and β-mannans.

Order Chitinophagales

CAZyme gene numbers in the eight *Chitinophagales* MAGs varied between 17 and 44 (Fig. 7). The highest number was observed in MAG_446 (*Saprospiraceae*; JAFMDF01)

383 with PULs predicted to target at least eleven polysaccharide substrates, including
384 pectin (GH28, GH106, GH105), alginate, β-glucan (variant-1) and xyloglucan (GH74).
385 The related MAG_447 featured 35 CAZyme genes, including predicted PULs for
386 alginate (PL7), α-glucan (type I), xyloglucan and β-glucan/xylan (CBM6, GH16, GH5,
387 GH3, GH30). A similar PUL was also identified in MAG_461 and MAG_457 (both
388 UBA1994). MAG_449 (24 CAZymes) harbored predicted PULs for alginate,
389 carrageenan (GH127), β-glucan (variant-1) and α-glucan (type II). MAG_452 harbored
390 the same PUL for β- and α-glucans plus an additional predicted fructan PUL (GH32).
391 MAG_450 featured only a single predicted PUL for arabinogalactan (GH42), and the
392 remaining MAGs featured predicted variant-2 and variant-3 β-glucan PULs.
393
394 *Phylum* Verrucomicrobiota
395 The twelve selected *Verrucomicrobiota* MAGs featured null (no CAZyme annotation
396 using dbCAN3-sub) to 44 CAZyme genes, with the highest number present in
397 MAG_693 (*Lentimonas*), including predicted fucoidan- (Fig. S25) and pectin-targeting
398 CAZyme-rich gene clusters. This MAG exhibited significantly higher abundance in PA
399 than in FL communities. Its abundance increased after the diatom bloom and peaked
400 during the *Phaeocystis* bloom (Fig. S23F in Additional file 2). In contrast, MAG_644
401 (UBA985; *Persicirhabdus*) possessed only five CAZyme genes, was more abundant
402 in PA3 communities and peaked amidst the diatom bloom.
403
404 *Genus* CL500-3 (Planctomycetota)
405 The three selected CL500-3 MAGs (MAG_585, MAG_591, MAG_607) all exhibited
406 the highest relative abundances in the PA3 fractions (Fig. 7), two of which represented
407 novel Helgoland MAGs. All these MAGs featured PULs for α-glucan, β-glucan, xylan
408 and host-glycans. MAG_591 and MAG_607 featured also PULs for chitin.
409
410 *Order* Cellvibrionales
411 The BD1-7 clade (3 MAGs) was abundantly represented (Fig. S7 in Additional file 2),
412 with a dominant ASV in each of the PA communities (Fig. S23G-H in Additional file 2).
413 Corresponding MAG_1340 and 1321 had similar CAZyme profiles with 19 and 13
414 genes, respectively, whereas MAG_1317 featured only three such genes. MAG_1321

and 1340 reached significantly higher abundances (13.6% and 7.0%, respectively) than MAG_1317 (0.12%).

Order Alteromonadales

Colwellia (1 MAG)*:* MAG_1214 had the highest relative abundances in the PA fractions and represented a novel Helgoland MAG. The highest relative abundance of the corresponding ASV was detected before the diatom bloom (Table S5 in Additional file 1). MAG_1214 featured PUL-like clusters for α-glucan, β-glucan, xylan, host-glycan, alginate and chitin, and was particularly rich in PUL-like clusters for alginate.

Paraglaciecola (2 MAGs)*:* Both of the two selected *Paraglaciecola* MAGs were proportionally more abundant in the PA fractions. They comprise the more complete (97%) MAG_1218 and the less complete MAG_1219 (67%). The former featured PULs predicted to target α-glucan, xylan, host-glycan, alginate, chitin, α-galactan, arabinogalactan, β-mannan and cellulose, but notably no PULs for β-glucan. MAG_1218 increased during the diatom bloom's decline and decreased before the diatom bloom's decline (Table S5 in Additional file 1).

Other proteobacterial clades

Other noteworthy MAGs included MAG_1258 (unclassified *Nitrincolaceae*; ASP10-02a), MAG_852 (UBA7985; *Stappiaceae*), and MAG_1010 (*Sulfitobacter*), which exhibited high abundances in either PA3 or PA10 communities (Fig. S23I-K in Additional file 2) but did not feature abundant CAZyme genes (Fig. 7).

Supplementary Materials and Methods

Sampling for 16S/18S rRNA gene amplicon and metagenome sequencing

In brief, samples from about 1 m depth were successively filtered through 10, 3, and 0.2 μm pore-sized polycarbonate filters (Millipore, Schwalbach, Germany) to separate the bulk of free-living (0.2-3 μm; FL) and particle-attached bacteria (3-10 μm and >10 μm; PA3 and PA10). Filters were stored at -80 °C until further use.

16S rRNA gene amplicon sequencing and analysis

Adapters were trimmed with cutadapt v1.15 [6]. After removing barcoded primers, the

DADA2 v1.19.2 package was used to analyze the sequences [7]. Forward and reverse fastq files were filtered using *filterAndTrim* with default parameters except for *truncLen=c(220,230)* for FL and *truncLen=c(200,200)* for PA3 and PA10 sequences. Resulting files were subsequently merged using mergePairs with default parameters. The SILVA SSU v138 Ref NR database [8] was used for taxonomic assignment. Relative read abundances were calculated using Phyloseq v1.34.0 [9] in R. Good's coverage and diversity indices (Chao1, Simpson's, Shannon) were calculated using the vegan v2.5-7 R package [10]. One-way analysis of variance (ANOVA) and PERMANOVA were performed in Prism (GraphPad Software, Boston, MA, USA). To determine community composition differences between FL and PA bacteria, non-metric multidimensional scaling (NMDS) with Bray-Curtis dissimilarity was computed using vegan in R.

18S rRNA gene amplicon sequencing and analysis

The V7 region of the 18S rRNA gene was amplified using the primers F-1183mod and R-1443mod [11] coupled to custom adaptor-barcode constructs. PCR amplification and Illumina MiSeq library preparation and sequencing (V3 chemistry) were carried out by LGC Genomics (LGC Genomics, Berlin, Germany). Sequence reads free of the adaptor and primer sequences were processed using DADA2 in R. Resulting ASVs were classified using the Protist Ribosomal Reference database (PR2) [12] (v4.13, *minboot*: 50) for 18S rRNA with the RDP classifier [13] built-in DADA2. Reads classified as metazoan (zooplankton) were removed prior to downstream analysis.

Parameters of metagenome assembly

Parameters used in SPAdes v3.11.1 [14]: *-meta*; k-mer lengths: 21, 33, 55, 77, 99, 127; error correction enabled. Parameters used in MEGAHIT v1.2.9 [15]: kmer length: 21. Parameters used in Flye v2.9.1 [16]: *--meta --pacbio-hifi --keep-haplotypes --hifi-error 0.01*.

Metagenome taxonomic classification

We conducted reads-based taxonomic classifications of Illumina-sequenced metagenomes of each of the three size fractions (FL: 0.2-3 μm, PA3: 3-10 μm,

PA10: >10 μm) from the following eight dates: 2018/03/19, 2018/04/12, 2018/04/17, 2018/04/26, 2018/05/08, 2018/05/11, 2018/05/22, 2018/05/29. Taxonomic classification of these 24 metagenomes was done by sequential use of Kraken 2 [17], Kaiju [18], Centrifuge [19] and CLARK [20] predictions. In brief, we first ran Kraken 2 (k-mers) and then funneled the remaining unassigned reads to Kaiju (protein-coding genes). Reads that were still unassigned were then imported into Centrifuge (Burrows-Wheeler transform plus Ferragina-Manzini indexing) and ultimately into CLARK (k-mers).

Metagenome-assembled genome (MAG) retrieval

Illumina metagenome assemblies were binned individually within Anvi'o v6.2 using CONCOCT [21], MetaBAT2 [22] and MaxBin2 [23]. Assemblies of PacBio metagenomes were binned individually within Anvi'o v7.1 [24] using MetaBAT2 and MaxBin2. Binning results were aggregated in DAS Tool [25] to find an optimized, non-redundant set for each assembly. Bins were manually refined using *anvi-refine* and the Anvi'o interactive interface for the reconstruction of MAGs.

MAG abundances

For Illumina metagenomes BBMap v38.86 (https://sourceforge.net/projects/bbmap) was used to map reads to corresponding dereplicated medium and high-quality MAGs (mode: *fast, idfilter: 97, minid: 99*), and for PacBio metagenomes Minimap2 v2.24 [26] was used in an analogous manner (*map-hifi* preset). Resulting SAM files were converted to BAM files using Samtools [27]. Sequencing depth was determined using genomecov (*-bga* option) from bedtools [28]. The *BedGraph.tad.rb* script (option range 80) from the enveomics collection [29] was then used to estimate the 80% central truncated average of the sequencing depth (TAD). Sequencing depths for all microbial genomes were determined using the MicrobeCensus pipeline [30] as follows: MAG abundance = (TAD of a MAG) / (sum of sequencing depths of all microbial genomes in the corresponding metagenome).

Annotation of predicted genes from metagenomes and MAGs

CAZymes were predicted via the run_dbcan program of dbCAN2 [31], Diamond blastp

searches against the Carbohydrate-Active Enzymes (CAZy) database (http://www.cazy.org/) [32] (as of July 31, 2020) and HMMER v3.3.2 [33] against Pfam HMM models [34] (as of September 2020).

Genes coding for TonB-dependent transporters (TBDTs including SusC-like proteins) were predicted by searching against the corresponding TIGRFAM and PFAM profiles using HMMER. Profiles used for the prediction of TBDTs: TIGR01352, TIGR01778, TIGR01779, TIGR01782, TIGR01783, TIGR01785, TIGR01786, TIGR02796, TIGR02797, TIGR02803, TIGR02804, TIGR02805, TIGR04056, TIGR04057, PF00593 (TonB-dependent receptor), PF07715 (TonB-dependent receptor plug domain), PF13620 (Carboxypeptidase regulatory-like domain), PF01618 (MotA/TolQ/ExbB proton channel family) and PF13715 (CarboxypepD_reg-like domain). SusD-like proteins and sulfatases were annotated by searching against corresponding Pfam profiles (PF07980, PF12741, PF12771, PF14322, PF00884) using HMMER.

Prediction of CAZyme-rich gene clusters and PULs

We searched for candidate genes (genes coding for PLs, GHs, CEs, sulfatases, TBDTs, or SusD-like proteins), and once such a candidate gene was found, ten genes downstream were checked for further candidates. In case an additional candidate gene was found, the window of ten genes was shifted by one gene. The process was repeated until no further candidate genes were found [35].

Substrate class prediction of CAZymes in metagenomes and MAGs

Target substrate classes of CAZymes were assigned using the dbCAN3-sub database [36]. PULDB [37] was also considered as a reference.

Selection of housekeeping genes for phylogenetic analysis

Sequences of the following ribosomal proteins and single copy genes were used for the phylogenomic analysis of MAGs: RBFA, Ribosomal_L1, Ribosomal_L17, Ribosomal_L23, Ribosomal_L3, Ribosomal_L4, Ribosomal_L5, Ribosomal_L6, Ribosomal_S11, Ribosomal_S13, Ribosomal_S15, Ribosomal_S16, Ribosomal_S17, Ribosomal_S19, Ribosomal_S2, Ribosomal_S6, Ribosomal_S7, Ribosomal_S8,

543 Ribosomal_S9, SecE, SecG, SecY, SmpB, ADK, AICARFT_IMPCHas, ATP-synt,
544 ATP-synt_A, EF_TS, Ribosomal_L24, RNA_pol_L, RNA_pol_Rpb6, RRF, RsfS, RuvX,
545 tRNA_m1G_MT, tRNA-synt_1d, TsaE and YajC [24].
546
547 *Metaproteome analyses of the FL fraction – sample preparation*
548 Filters for metaproteome analysis were prepared as previously described [38]. One-
549 eighth of a filter (Millipore Express PLUS Membrane, polyethersulfone, hydrophilic,
550 0.2 μm pore size, diameter 142 mm) was cut into approximately 10 x 10 mm fragments
551 and transferred to 15 mL low binding tubes containing 1 mL resuspension buffer 1
552 (50 mM Tris-HCl (pH 7.5), 0.1 mg mL^{1} chloramphenicol, 1 mM phenylmethylsulfonyl
553 fluoride (PMSF)). After mixing with 1.5 mL resuspension buffer 2 (20 mM Tris-HCl
554 pH 7.5, 2% SDS (w/v)) for 10 min at 60 °C at 1,000 rpm in a thermo-mixer (Eppendorf,
555 Wesseling-Berzdorf, Germany), 5 mL DNAse buffer (20 mM Tris-HCl pH 7.5,
556 0.1 mg mL^{-1} MgCl2, 1 mM PMSF, 1 µg mL^{-1} DNAse I) was added, and cells were lysed
557 by ultra-sonication (amplitude 51-60%; cycle 0.5; 4-times 2 min) on ice. The lysate
558 was incubated in the thermo-mixer for 10 min at 37 °C at 1,000 rpm. After
559 centrifugation for 10 min at 4 °C at 10,000 $\times g$, the supernatant (containing protein
560 extract) was collected and the pelleted filter pieces were stirred and centrifuged again
561 for 1 min at 4 °C at 5,000 $\times g$. The supernatant was added to the previously collected
562 supernatant. Proteins in the supernatant were precipitated by adding pre-cooled
563 trichloroacetic acid (20% TCA (v/v)) and after inverting the tube approximately 10-
564 times, the precipitate was pelleted by centrifugation (30 min, 4 °C, 12,000 $\times g$). The
565 protein pellet was washed 3-times in pre-cooled (-20 °C) acetone (10 min, 4 °C,
566 12,000 $\times g$) and dried at room temperature. The protein pellet was resuspended in 2×
567 SDS sample loading buffer (4% SDS (w/v), 20% glycerine (w/v), 100 mM Tris-HCl
568 pH 6.8, bromphenol blue (tip of a spatula, to add color), 3.6% 2-mercaptoethanol (v/v)
569 (freshly added before use)) by incubation 5 min at 95 °C, 5 min sonication bath before
570 vortexing, and separated by 1D SDS-PAGE (Criterion TG 4-20% Precast Midi Gel,
571 BIO-RAD Laboratories, Inc., USA). After separation, fixation, and staining with
572 Coomassie, each gel lane was cut into 20 pieces as described previously with some
573 modifications [39]. The gel pieces were destained 3-times for 10 min with 1 mL of gel
574 washing buffer (200 mM ammonium bicarbonate in 30% acetonitrile (v/v)) at 37 °C

under vigorous shaking. The destained gel pieces were dehydrated in 1 mL 100% acetonitrile (v/v) for 20 min before drying in a vacuum centrifuge at 30 °C. For reduction and alkylation, the gel pieces were treated with 100 μL 10 mM Dithiothreitol in 25 mM ammonium bicarbonate buffer, incubated for 1 h at 56 °C, mixed with 100 μL 55 mM iodoacetamide in 25 mM ammonium bicarbonate buffer and further incubated in the dark for 45 min at room temperature before the supernatant was removed. The gel pieces were washed with 1 mL 25 mM ammonium bicarbonate buffer (10 min, 1,000 rpm) before the supernatant was removed. Next, the gel pieces were dehydrated with 500 μL 100% acetonitrile for 10 min, and the supernatant was removed before the gel pieces were completely dried in a vacuum centrifuge (20 min) and finally covered with 120 μL trypsin solution (2 μg/mL Trypsin (Promega™). After rehydration for 20 min at room temperature, excess trypsin solution was removed with gel loader tips and incubated in a thermo-mixer for 15 h at 37 °C without shaking. Peptides were eluted with 120 μL solvent A (water MS grade in 0,1% acetic acid (v/v)) by sonication for 15 min. The supernatant was transferred into a new tube. Peptides were eluted again with 120 μL 30% acetonitrile (v/v) by sonication for 15 min. The supernatant was transferred into the same new tube as used before. The sample volume was reduced in a vacuum centrifuge to a maximum of 15 to 20 μL. The peptides were desalted via ZipTips μC18 (Merck Millipore, P10 tip size) according to the manufacturer's protocol. The eluted samples were dried in a vacuum centrifuge and resuspended in 10 μL 0.5x Biognosys™ iRT standard kit in solvent A.

Metaproteome analyses of the FL fraction - LC-MS/MS measurement and data analysis

For measurement, an Easy-nLC1000 (Thermo Fisher Scientific, Waltham, MA, USA) was coupled to an Q Exactive mass spectrometer (Thermo Fisher Scientific). Samples were loaded onto in-house packed capillary columns of 20 cm length and 75 μm inner diameter. Columns were filled with Dr. Maisch ReproSil Pur 120 C18-AQ 1.9 μm (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). Peptides were separated using a 131 min nonlinear binary gradient from 2% to 99% solvent B (99.9% acetonitrile(v/v), 0.1% acetic acid (v/v)) in solvent A at a constant flow rate of 300 nL min^{-1}. The MS1 scan was recorded in the orbitrap with a mass window of 300–1,650 m/z and a

607 resolution of 140,000 at 200 m/z. The 15 most intense precursor ions (ions with an
608 unassigned charge or a charge of 1,7,8, >8 are excluded) were selected for HCD
609 fragmentation with a normalized collision energy of NCE 27. The resulting MS/MS
610 spectra were acquired were recorded with a resolution of 17.500 at 200 m/z. Dynamic
611 exclusion and lock mass correction were enabled.

612 All MS/MS spectra were analyzed using Mascot (version 2.7.0.1; Matrix Science,
613 London, UK). Mascot was set up to search the database containing all protein
614 sequences from the 18 metagenomes obtained during the spring bloom of 2018,
615 assuming the digestion enzyme trypsin. For database construction, redundant
616 proteins from the 18 metagenomic samples were removed using cd-hit [40] with a
617 clustering threshold of 97% identity. The created database was added by a set of
618 common laboratory contaminants and reverse entries, amounting to 81,874,922
619 sequences in the final database.

620 The database search with Mascot [41] was performed with the following
621 parameters: fragment ion mass tolerance and parent ion tolerance of 10 ppm, none
622 missed cleavages, methionine oxidation as a variable modification, and cysteine
623 carbamidomethylation as fixed modification. Scaffold (version 4.11.1; Proteome
624 Software Inc., Portland, OR) was used to merge the search results and to validate
625 MS/MS based peptide and protein identifications. During creation of the Scaffold file,
626 an additional X! Tandem search was performed for validation (version 2017.2.1.4; The
627 GPM, thegpm.org; version X!Tandem Alanine) with default settings (fragment ion mass
628 tolerance and parent ion tolerance of 10 ppm, carbamidomethyl of cysteine as fixed
629 modification, Glu->pyro-Glu of the n-terminus, ammonia-loss of the n-terminus,
630 Gln->pyro-Glu of the n-terminus, oxidation of methionine and carbamidomethyl of
631 cysteine as variable modifications). Peptide identifications were accepted if they could
632 be established at greater than 95% probability by the PeptideProphet algorithm [42]
633 with Scaffold delta-mass correction. Protein identifications were accepted if they could
634 be established at greater than 99% probability and contained at least 2 identified
635 peptides. Protein probabilities were assigned by the ProteinProphet algorithm [43].
636 Proteins that contained similar peptides and could not be differentiated based on
637 MS/MS analysis alone were grouped to satisfy the principles of parsimony.

638 For (semi-)quantitative analysis, the Scaffold's 'Quantitative Value' for normalized,
639 weighted spectra for each protein group was divided by the sum of all quantitative
640 values for the sample to calculate the percent normalized weighted spectra (%NWS).
641 Average values were calculated from all three biological replicates of each sample.
642 Protein groups that were not identified within a replicate were included as '0' in this
643 calculation. In order to make Bacteria-specific %NWS readily comparable across all
644 samples, all bacterial spectra were normalized to 100% (%BacNWS).
645
646 *Assignment of expressed proteins to MAGs for FL metaproteomes*
647 To assign identified expressed protein groups to MAGs, amino acid sequences were
648 aligned to all predicted proteins of all MAGs from the FL fraction 2018 using BLAST
649 v2.11.0 [44]. Alignments with identities >99%, e-values below E-4 and coverages of at
650 least 50% for both the query and the subject were considered correctly assigned.
651
652 *Metaproteome analyses of the PA fractions – sample preparation*
653 Sample preparation for the PA fractions has been described earlier for this dataset [45,
654 46]. In brief, sequential filtration was performed for three selected time points with
655 polycarbonate membrane filters (142 mm diameter, Millipore) of 3 μm and 10 μm pore
656 sizes. Proteins were extracted using the bead-beating protocol described in [45],
657 separated by 1D SDS-PAGE and split into 20 fractions per sample. After washing and
658 in-gel-trypsin digestion, peptides were desalted using ZipTips μC18 (Merck Millipore,
659 P10 tip size) as described for the FL fraction and dried by vacuum centrifugation.
660
661 *Metaproteome analyses of the PA fractions - LC-MS/MS measurement and data*
662 *analysis*
663 LC-MS / MS measurement for the PA fractions was performed in triplicates using an
664 Orbitrap VelosTM mass spectrometer (ThermoFisher Scientific, Waltham, MA, USA).
665 2018 MAGs were included into the database for analysis. For this, 2018 MAGs were
666 added to the metagenome-based database described in [46]. Redundant sequences
667 were removed with cd-hit [40] with a clustering threshold of 97% identity before adding
668 the 2018 MAGs to ensure that all 2018 MAG sequence information would be kept for
669 downstream analysis. The final database used for analysis of the PA metaproteome

fractions contained 15,529,863 entries.

To allow for comparison between the 3-10 μm and the >10 μm metaproteome fractions, samples for both filter sizes and from all three-time points were merged into one dataset in Scaffold. If protein groups contained more than one protein, 2018 MAG sequences were prioritized as representative proteins of that protein group. As described for the FL fraction, (semi-)quantitative values were calculated as the average of three technical replicates for each sample based on the percentage of the normalized weighted spectra value provided by the Scaffold. Protein groups that were not identified in a replicate were included with a value of '0' in this calculation.

Code Availability

Softwares/bioinformatic tools used in this study:

Anvio 6.2 and Anvio 7.1: https://github.com/merenlab/anvio;

Barrnap: https://github.com/tseemann/barrnap;

BBMap: https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbmap-guide/;

BLAST: ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST;

Centrifuge: https://github.com/centrifugal/centrifuge;

CheckM: https://ecogenomics.github.io/CheckM;

CLARK: http://clark.cs.ucr.edu;

CONCOCT: https://github.com/BinPro/CONCOCT;

Cutadapt: https://github.com/marcelm/cutadapt;

DAS Tool: https://github.com/cmks/DAS_Tool;

dbCAN2: http://bcb.unl.edu/dbCAN2/index.php;

DADA2: https://github.com/benjjneb/dada2;

dRep: https://drep.readthedocs.io/en/latest/;

DIAMOND: http://www.diamondsearch.org/index.php;

FastANI: https://github.com/ParBLiSS/FastANI;

FastQC: http://www.bioinformatics.babraham.ac.uk/projects/fastqc;

FastTree: http://www.microbesonline.org/fasttree;

Flye: https://github.com/fenderglass/Flye;

FragGeneScan: https://github.com/gaberoo/FragGeneScan;

GTDB-Tk: https://github.com/Ecogenomics/GTDBTk;

HMMER: http://hmmer.org;

iTOL v6.5.6: https://github.com/iBiology/iTOL;

Kaiju: https://github.com/bioinformatics-centre/kaiju;

Karen2: https://github.com/DerrickWood/kraken2;

MAFFT: https://mafft.cbrc.jp/alignment/software;

MaxBin2: https://sourceforge.net/projects/maxbin2;

MEGAHIT: https://github.com/voutcn/megahit;

MetaBAT2: https://bitbucket.org/berkeleylab/metabat;

MicrobeCensus: https://github.com/snayfach/MicrobeCensus;

Minimap2: https://github.com/lh3/minimap2;

Prodigal: https://github.com/hyattpd/Prodigal;

Prokka: https://github.com/tseemann/prokka;

QUAST: https://github.com/ablab/quast;

R: https://www.r-project.org;

RDP classifier: https://github.com/rdpstaff/classifier;

SAMTools: http://www.htslib.org;

Silvangs: https://ngs.arb-silva.de/silvangs/#;

Simka: https://github.com/GATB/simka;

SPAdes: https://github.com/ablab/spades;

SqueezeMeta: https://github.com/jtamames/SqueezeMeta;

Database Availability

Database used in this study:

Cazy: http://www.cazy.org;

dbCAN-sub: https://bcb.unl.edu/dbCAN_sub/;

Pfam: http://pfam.xfam.org;

Protist Ribosomal Reference database (PR2): https://pr2-database.org;

SILVA SSU v138 Ref NR database: https://www.arb-silva.de/documentation/release-138/;

TIGRFAM: https://www.jcvi.org/research/tigrfams;

734 **References**

735 1. McMurdie PJ, Holmes S. Waste not, want not: Why rarefying microbiome data is
736 inadmissible. PLoS Comput Biol. 2014;10(4):e1003531.
737 2. Hong J, Karaoz U, de Valpine P, Fithian W. To rarefy or not to rarefy: robustness
738 and efficiency trade-offs of rarefying microbiome data. Bioinformatics.
739 2022;38(9):2389-96.
740 3. Rodriguez-Gijon A, Nuy JK, Mehrshad M, Buck M, Schulz F, Woyke T, et al. A
741 genomic perspective across Earth's microbiomes reveals that genome size in
742 Archaea and Bacteria is linked to ecosystem type and trophic strategy. Front
743 Microbiol. 2022;12:761869.
744 4. Hahnke RL, Bennke CM, Fuchs BM, Mann AJ, Rhiel E, Teeling H, et al. Dilution
745 cultivation of marine heterotrophic bacteria abundant after a spring
746 phytoplankton bloom in the North Sea. Environ Microbiol. 2015;17(10):3515-26.
747 5. Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L, et al.
748 Recurring patterns in bacterioplankton dynamics during coastal spring algae
749 blooms. eLife. 2016;5:e11888.
750 6. Martin M. Cutadapt removes adapter sequences from high-throughput
751 sequencing reads. EMBnet journal. 2011;17(1):10-2.
752 7. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP.
753 DADA2: high-resolution sample inference from Illumina amplicon data. Nat
754 Methods. 2016;13(7):581-3.
755 8. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA
756 ribosomal RNA gene database project: improved data processing and web-
757 based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.
758 9. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive
759 analysis and graphics of microbiome census data. Plos One. 2013;8(4):e61217.
760 10. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci.
761 2003;14(6):927-30.
762 11. Ray JL, Althammer J, Skaar KS, Simonelli P, Larsen A, Stoecker D, et al.
763 Metabarcoding and metabolome analyses of copepod grazing reveal feeding
764 preference and linkage to metabolite classes in dynamic microbial plankton
765 communities. Mol Ecol. 2016;25(21):5585-602.

766 12. Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, et al. The Protist
767 Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small
768 sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res.
769 2013;41(Database issue):D597-604.
770 13. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid
771 assignment of rRNA sequences into the new bacterial taxonomy. Applied and
772 environmental microbiology. 2007;73(16):5261-7.
773 14. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
774 SPAdes: a new genome assembly algorithm and its applications to single-cell
775 sequencing. J Comput Biol. 2012;19(5):455-77.
776 15. Li DH, Luo RB, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0:
777 a fast and scalable metagenome assembler driven by advanced methodologies
778 and community practices. Methods. 2016;102:3-11.
779 16. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al.
780 metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat
781 Methods. 2020;17(11):1103-10.
782 17. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification
783 using exact alignments. Genome Biol. 2014;15(3):R46.
784 18. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for
785 metagenomics with Kaiju. Nat Commun. 2016;7(1):11257.
786 19. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive
787 classification of metagenomic sequences. Genome Res. 2016;26(12):1721-9.
788 20. Ounit R, Wanamaker S, Close TJ, Lonardi S. CLARK: fast and accurate
789 classification of metagenomic and genomic sequences using discriminative k-
790 mers. Bmc Genomics. 2015;16(1):236.
791 21. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning
792 metagenomic contigs by coverage and composition. Nat Methods.
793 2014;11(11):1144-6.
794 22. Kang DWD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an
795 adaptive binning algorithm for robust and efficient genome reconstruction from
796 metagenome assemblies. Peerj. 2019;7:e7359.

797 23. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm
798 to recover genomes from multiple metagenomic datasets. Bioinformatics.
799 2016;32(4):605-7.
800 24. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o:
801 an advanced analysis and visualization platform for 'omics data. PeerJ.
802 2015;3:e1319.
803 25. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al.
804 Recovery of genomes from metagenomes via a dereplication, aggregation and
805 scoring strategy. Nat Microbiol. 2018;3(7):836-43.
806 26. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics.
807 2018;34(18):3094-100.
808 27. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve
809 years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
810 28. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
811 features. Bioinformatics. 2010;26(6):841-2.
812 29. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for
813 specialized analyses of microbial genomes and metagenomes. PeerJ Preprints.
814 2016;4:e1900v1.
815 30. Nayfach S, Pollard KS. Average genome size estimation improves comparative
816 metagenomics and sheds light on the functional ecology of the human
817 microbiome. Genome Biol. 2015;16(1):51.
818 31. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta
819 server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res.
820 2018;46(W1):W95-W101.
821 32. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The
822 carbohydrate-active enzyme database: functions and literature. Nucleic Acids
823 Res. 2022;50(D1):D571-D7.
824 33. Eddy SR. A new generation of homology search tools based on probabilistic
825 inference. Genome Inform. 2009;23(1):205-11.
826 34. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL,
827 et al. Pfam: The protein families database in 2021. Nucleic Acids Res.
828 2021;49(D1):D412-D9.

829 35. Lu D, Wang F, Amann RI, Teeling H, Du JZ. Epiphytic common core bacteria in
830 the microbiomes of co-located green (*Ulva*), brown (*Saccharina*) and red
831 (*Grateloupia*, *Gelidium*) macroalgae. Microbiome. 2023;11(1):126.
832 36. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated
833 carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res.
834 2023;51(W1):W115-W21.
835 37. Terrapon N, Lombard V, Drula E, Lapebie P, Al-Masaudi S, Gilbert HJ, et al.
836 PULDB: the expanded database of Polysaccharide Utilization Loci. Nucleic Acids
837 Res. 2018;46(D1):D677-D83.
838 38. Deusch S, Seifert J. Catching the tip of the iceberg - evaluation of sample
839 preparation protocols for metaproteomic studies of the rumen microbiota.
840 Proteomics. 2015;15(20):3590-5.
841 39. Bonn F, Bartel J, Büttner K, Hecker M, Otto A, Becher D. Picking vanished
842 proteins from the void: how to collect and ship/share extremely dilute proteins in
843 a reproducible and highly efficient manner. Anal Chem. 2014;86(15):7421-7.
844 40. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of
845 protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658-9.
846 41. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein
847 identification by searching sequence databases using mass spectrometry data.
848 Electrophoresis. 1999;20(18):3551-67.
849 42. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to
850 estimate the accuracy of peptide identifications made by MS/MS and database
851 search. Anal Chem. 2002;74(20):5383-92.
852 43. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying
853 proteins by tandem mass spectrometry. Anal Chem. 2003;75(17):4646-58.
854 44. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment
855 search tool. J Mol Biol. 1990;215(3):403-10.
856 45. Schultz D, Zuhlke D, Bernhardt J, Francis TB, Albrecht D, Hirschfeld C, et al. An
857 optimized metaproteomics protocol for a holistic taxonomic and functional
858 characterization of microbial communities from marine particles. Environ
859 Microbiol Rep. 2020;12(4):367-76.

Chapter III

Epiphytic common core bacteriain the microbiomes of co-located green (*Ulva*), brown (*Saccharina*) and red (*Grateloupia*, *Gelidium*) macroalgae

Manuscript published in Microbiome Journal

De-Chen Lu, **Feng-Qing Wang**, Rudolf I. Amann, Hanno Teeling and Zong-Jun Du

Contribution of the candidate in % of the total work

Experimental concept and design – 20%

Experimental work/acquisition of experimental data – 10%

Data analysis and interpretation – 40%

Preparation of figures and tables – 40%

Drafting of the manuscript – 30%

Lu *et al. Microbiome* (2023) 11:126
https://doi.org/10.1186/s40168-023-01559-1

Microbiome

RESEARCH

Open Access

Epiphytic common core bacteria in the microbiomes of co-located green (*Ulva*), brown (*Saccharina*) and red (*Grateloupia*, *Gelidium*) macroalgae

De-Chen Lu[1,2,3], Feng-Qing Wang[2], Rudolf I. Amann[2], Hanno Teeling[2*] and Zong-Jun Du[1,3*]

Abstract

Background Macroalgal epiphytic microbial communities constitute a rich resource for novel enzymes and compounds, but studies so far largely focused on tag-based microbial diversity analyses or limited metagenome sequencing of single macroalgal species.

Results We sampled epiphytic bacteria from specimens of *Ulva* sp. (green algae), *Saccharina* sp. (brown algae), *Grateloupia* sp. and *Gelidium* sp. (both red algae) together with seawater and sediment controls from a coastal reef in Weihai, China, during all seasons. Using 16S rRNA amplicon sequencing, we identified 14 core genera (consistently present on all macroalgae), and 14 dominant genera (consistently present on three of the macroalgae). Core genera represented ~0.7% of all genera, yet accounted for on average 51.1% of the bacterial abundances. Plate cultivation from all samples yielded 5,527 strains (macroalgae: 4,426) representing 1,235 species (685 potentially novel). Sequencing of selected strains yielded 820 non-redundant draft genomes (506 potentially novel), and sequencing of 23 sampled metagenomes yielded 1,619 metagenome-assembled genomes (MAGs), representing further 1,183 non-redundant genomes. 230 isolates and 153 genomes were obtained from the 28 core/dominant genera. We analyzed the genomic potential of phycosphere bacteria to degrade algal polysaccharides and to produce bioactive secondary metabolites. We predicted 4,451 polysaccharide utilization loci (PULs) and 8,810 biosynthetic gene clusters (BGCs). These were particularly prevalent in core/dominant genera.

Conclusions Our metabolic annotations and analyses of MAGs and genomes provide new insights into novel species of phycosphere bacteria and their ecological niches for an improved understanding of the macroalgal phycosphere microbiome.

Keywords Algae, Bacteria, Biofilm, Biosynthetic gene cluster, *Gelidium*, *Grateloupia*, Macroalgae, Marine, Phycosphere, Microbiome, Polysaccharide utilization locus, *Saccharina*, *Ulva*

*Correspondence:
Hanno Teeling
hteeling@mpi-bremen.de
Zong-Jun Du
duzongjun@sdu.edu.cn
Full list of author information is available at the end of the article

Background

The term 'macroalgae' subsumes three major lineages: *Rhodophyta* (red algae), *Chlorophyta* (green algae) and *Phaeophyta* (brown algae) comprising approximately 12,000 species [1] that occur in coastal marine ecosystems worldwide. Macroalgae surfaces are colonized by bacteria and macroalgae-associated bacteria have co-evolved with macroalgae for roughly 1.6 billion years [2] with a complex and close relationship [3, 4]. The region of close algae-bacteria interactions is termed 'phycosphere' according to Bell and Mitchell (1972) [5]. The phycosphere microbiome is notably distinct from microbes of the surrounding seawater in terms of composition and functions [3, 4]. It supports the macroalgal host in essential functions, such as the morphological development [6] by the provision of growth factors [7], acclimation to environmental changes [8], release and settlement of algal spores [9], and the provision of vitamins and nutrients [7, 10]. Algal phycospheres also harbor potentially harmful bacteria, such as pathogens [11], or commensal bacteria that can degrade macroalgal tissues [12].

Macroalgae play an eminent role for maintaining high bioproductivity and biodiversity in coastal systems [13] and are thus of huge importance to various aspects of human life [14–16]. Compared to terrestrial plants, macroalgae have the benefits of higher growth rates, higher biomass yields, lower fiber, and higher polysaccharide contents [16]. Their combined biomass equals about 1,521 TgC yr^{-1} (range: 1,020-1,960 TgC yr^{-1}) [17], and their ecological role thus parallels that of terrestrial plants. Macroalgae release 14 to 35% of their photoassimilated net primary production to the environment [18]. Some of this dissolved or aggregated particulate organic matter is rather recalcitrant and thus only slowly and partially degraded by marine bacteria. Such organic matter can sequester carbon for longer periods of time, as has been recently described for algal fucoidan [18]. However, most algal biomass is quickly remineralized by marine bacteria [19] and thereby routed back into the global carbon cycle.

Since macroalgae are usually sessile and predominantly inhabit coastal areas, they are subject to dynamic environmental changes, which in term affect their phycosphere community compositions [20]. Host morphology also plays a role, as has been shown with artificial algae of various shapes [3]. Such abiotic influences notwithstanding, phycosphere communities have shown to be also host-specific in various studies. For example, Lachnit et al. described both, seasonal variations and host specificities in the colonization patterns of three macroalgal species [21]. Different mechanisms have been proposed for host-specific colonization, such as a random occupation of phycosphere ecological niches by species with suitable adaptations, or the selection of functional genes on a community level [22, 23]. However, research is lacking for common core bacteria in different macroalgae in terms of taxonomy, representative genomes and ecophysiological functions.

Members of the following phyla dominate macroalgal phycospheres and are thus believed to be indispensable for proper phycosphere functioning: *Proteobacteria*, *Bacteroidota*, *Verrucomicrobiota*, *Planctomycetota*, *Firmicutes*, *Patescibacteria* and *Cyanobacteria* [3, 4, 10, 20–23]. Much less is known about these phycosphere bacteria than about those associated with terrestrial plants, particularly those of the rhizosphere. However, recent years have witnessed a growing interest in phycosphere bacteria of marine plants and algae that surpasses mere descriptions of microbial community composition, as is exemplified by recent studies of seaweed [24] and kelp microbiomes [10]. In particular the mechanisms that determine and maintain colonization patterns as well as the underlying genetic functions are of interest, not least because such functions bear the potential for useful industrial applications.

Two traits are prevalent among phycosphere bacteria, namely the potentials to degrade various algal polysaccharides and to produce a plethora of secondary metabolites. A substantial part of algal biomass consists of various diverse and complex polysaccharides. The primary polysaccharides in *Phaeophyta* are laminarins, fucoidans, cellulose and alginates [25], in *Chlorophyta* cellulose, xylans and ulvans [26, 27], and in *Rhodophyta* agars, carrageenans and galactans (including porphyran and furcellan) [15]. Many of these polysaccharides are anionic, sulfated and do not have equivalents in terrestrial plants [25]. In bacteria, the genes for the breakdown and take-up of polysaccharides are often co-located in dedicated polysaccharide utilization loci (PULs), in particular in the *Bacteroidota*. The capacity to degrade various land plant polysaccharides has been well studied in human gut *Bacteroidota* [26], and in some marine *Bacteroidota* targeting algal polysaccharides, e.g., alginate [28], laminarin [29, 30] and carrageenan [31]. However, a large-scale, systematic inventory of PULs of macroalgal phycosphere bacteria is as yet missing. Recent analyses have also shed light on the potential of marine bacteria to produce metabolites on a global scale, focusing either on planktonic bacteria [32] or marine biofilm-forming bacteria [33]. However, a comprehensive evaluation of the potential for secondary metabolite production of macroalgal phycosphere bacteria is lacking.

In this study, we investigate phycosphere bacteria of four algal species: *Ulva* sp. (green algae), *Saccharina* sp. (brown algae), *Grateloupia* sp. and *Gelidium* sp. (both red algae). Samples were taken in spring, summer,

winter and autumn together with seawater and sediment controls from a coastal reef at Weihai, China. We used a combination of 16S rRNA tag-based biodiversity analyses, extensive cultivation, as well as genome and deep metagenome sequencing in order to characterize and compare phycosphere communities, and in particular to identify common core genera (Fig. 1). We report a large number of cultured strains including novel core/dominant phycosphere strains, corresponding genomes, and insights into the potential of phycosphere bacteria to degrade algal polysaccharides and to synthesize bioactive secondary metabolites, some of which may control phycosphere community composition. The resulting comprehensive dataset of novel microbial species, their genomes and associated gene functions, represents a significant stepping stone towards a better understanding of

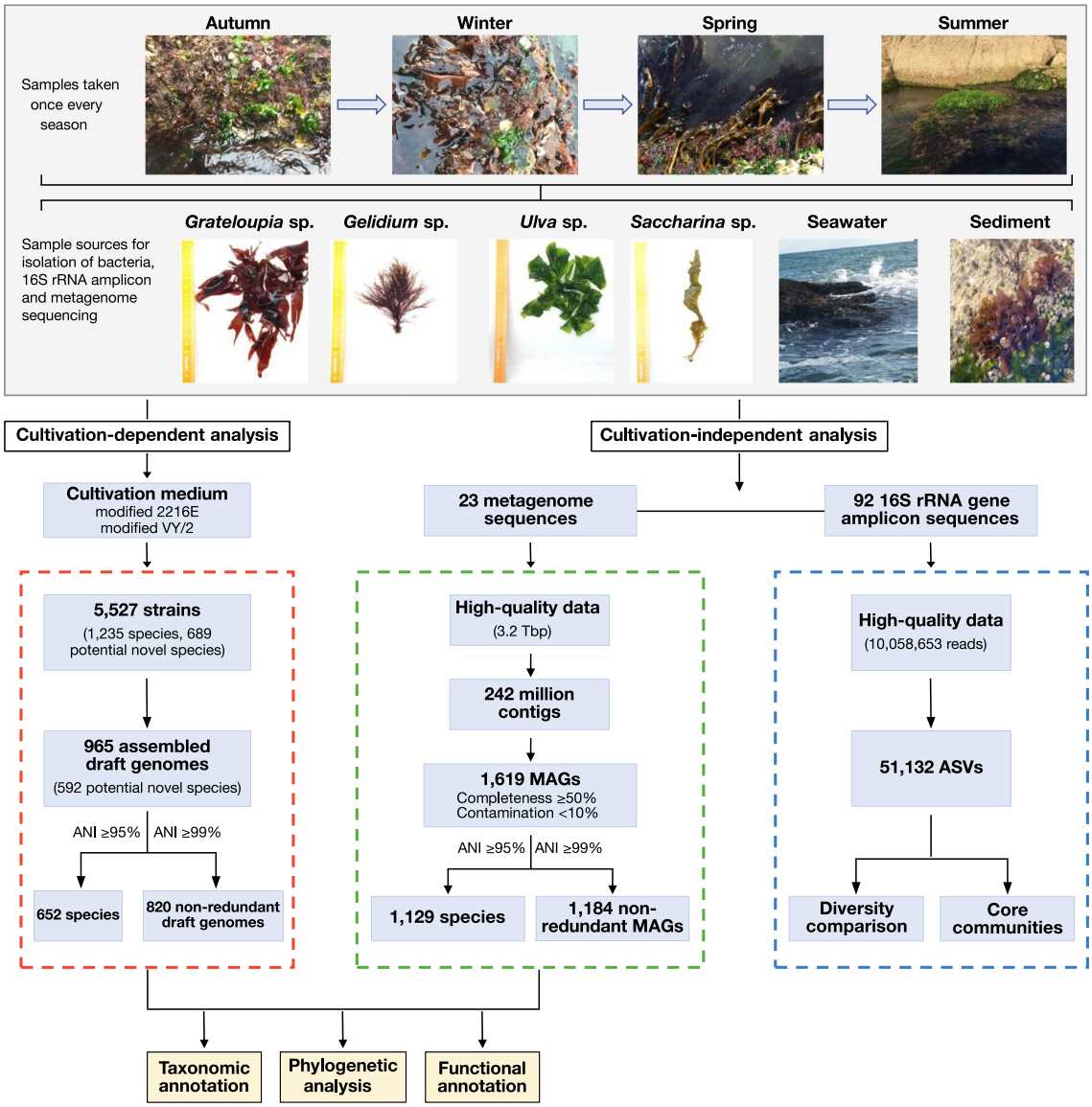


Fig. 1 Study workflow. Samples were taken from a coastal reef in Weihai (China) once during each season. Four macroalgal species were sampled, plus sediment and seawater controls. Data analysis consisted of (i) the 16S rRNA gene tag pipeline (blue box), (ii) cultivation and draft genome sequencing of isolated strains (red box), and (iii) sequencing of community DNA with subsequent reconstruction of MAGs (green box)

the global ocean microbiome in general and macroalgal phycosphere bacteria in particular, and paves the way to functional studies on representative strains.

Results

All algae featured similar yet diverse phycosphere communities with notable seasonalities

Rarefaction curves of the 200 most abundant 16S rRNA ASVs (amplicon sequence variants) plateaued around 90% for most macroalgal and seawater samples. The top 20 ASVs alone accounted for close to 50% of the total abundance of the macroalgal samples, except for the *Saccharina* sp. brown algae summer samples and the two red algae species. The sediment samples were a different matter, as their rarefaction curves did not plateau, indicating higher overall diversities due to much higher numbers of rare taxa (Fig. S1b in Additional file 2).

In ASV α-diversity (richness) analyses, phycosphere samples exhibited similar overall diversities than seawater, but lower diversities than sediment samples, corroborating the rarefaction analyses (Fig. S1a in Additional file 2). Phycospheres were most diverse in summer except for *Gelidium* sp. (Fig. S1a in Additional file 2). Simpson's diversity median values exceeded 0.8 for all habitats apart from *Saccharina* sp. in winter (0.5) due to high *Rubritalea* (*Verrucomicrobiota*) relative abundances (53.1% ± 30.7; see Discussion). Likewise, *Saccharina* sp. phycosphere communities had lower median Shannon diversity values (3.7 ± 1.8) than those from other macroalgae (4.3 ± 0.6) (Fig. S1a in Additional file 2).

Principal coordinate analysis (PCoA) of ASV β-diversity using the Bray–Curtis dissimilarity index revealed clustering by habitat (Fig. 2a), with phycosphere data clearly separated from sediment and seawater controls. Pairwise comparisons of only phycosphere samples, however, did not uncover significant differences, suggesting a considerable degree of shared taxa between the sampled macroalgal species (Fig. 2b). After removal of core taxa ASVs, i.e., of taxa occurring on all macroalgae (see Materials and methods), samples clustered more clearly according to season (Fig. 2c), indicating that non-core taxa contributed more to seasonal variation.

The complete amplicon dataset comprised ASVs of 68 phyla, 56 of which were present on macroalgae (21,381 unique ASVs, Table S1 in Additional file 3). UniFrac UPGMA cluster analysis confirmed significant differences between the sediment, seawater and phycosphere habitats (Figs. 2, S3 in Additional file 2). The relative abundance of *Bacteroidota* in phycosphere samples was generally higher compared to seawater samples, which featured *Bacteroidota* abundances of up to 25.1% only in spring (Fig. S3 in Additional file 2). The sediment samples were even more distinct (Figs. 3, S3 in Additional file 2). Seasonal variations were obvious within all phycosphere communities (Figs. 3, S4 in Additional file 2). Samples from the same macroalgal species clustered for most seasons, particularly in the case of *Ulva* sp., *Grateloupia* sp. and *Gelidium* sp. in spring, suggesting particularly similar phycosphere communities (Figs. 2a, b, S3 in Additional file 2). Though differences among habitats became more apparent at the family and genus levels, there still was considerable consistency across macroalgal phycospheres (Figs. 3, S4 in Additional file 2).

Phycospheres were dominated by few core phycosphere taxa

ASV analyses revealed that the majority of bacterial families in the phycospheres were represented by only one or two genera, while few, such as *Flavobacteriaceae* and *Rhodobacteraceae*, were more broadly represented (Figs. 3, S2 in Additional file 2, Table S1 in Additional file 3). This low overall evenness underscores that phycosphere communities were largely dominated by few abundant clades. Fourteen core genera from eight families (phyla *Proteobacteria*, *Bacteroidota*, *Verrucomicrobiota*, *Actinobacteriota*) were present on all macroalgae with ≥ 1% abundance in at least one of the samples (Fig. 3, Table 1, Table S1 in Additional file 3). *Sphingomonadaceae* and *Arenicellaceae* represented additional, diverse core families without any genus reaching ≥ 1% abundance in any sample (Fig. 3, Table S1 in Additional file 3). Core phycosphere genera comprised, on average, 1.4% of all phycosphere genera (*Gelidium* sp., 14/972, *Grateloupia* sp., 14/1,000, *Ulva* sp., 14/973 and *Saccharina* sp., 14/870), but accounted for on average 43.5% (*Gelidium* sp.), 53.9% (*Grateloupia* sp.), 58.3% (*Ulva* sp.) and 48.8% (*Saccharina* sp.) of all phycosphere bacteria (Table S1 in Additional file 3, Fig. S3, heatmap in Additional file 2). By comparison, the average relative abundances of these core phycosphere genera in seawater and sediment samples were only 5.7% and 1.5%, respectively (Table S1 in Additional file 3, Fig. S3, heatmap in Additional file 2). Fourteen additional genera were abundantly present in three of the four macroalgal species, hereinafter termed dominant genera (Fig. 3, Table 1). The relative abundances of all 28 prevalent genera varied in a similar fashion across seasons on all algae.

Strains of 230 species from 16 abundant core and dominant phycosphere genera

Cultivation yielded in total 5,527 strains (macroalgae: 4,426). Clustering of their 16S rRNA gene sequences revealed that they represent 1,235 species (98.7% identity criterion) from 444 genera (94.5% identity criterion), including 968 species from macroalgae (Table S2 in Additional file 3). Almost two-thirds of the species were

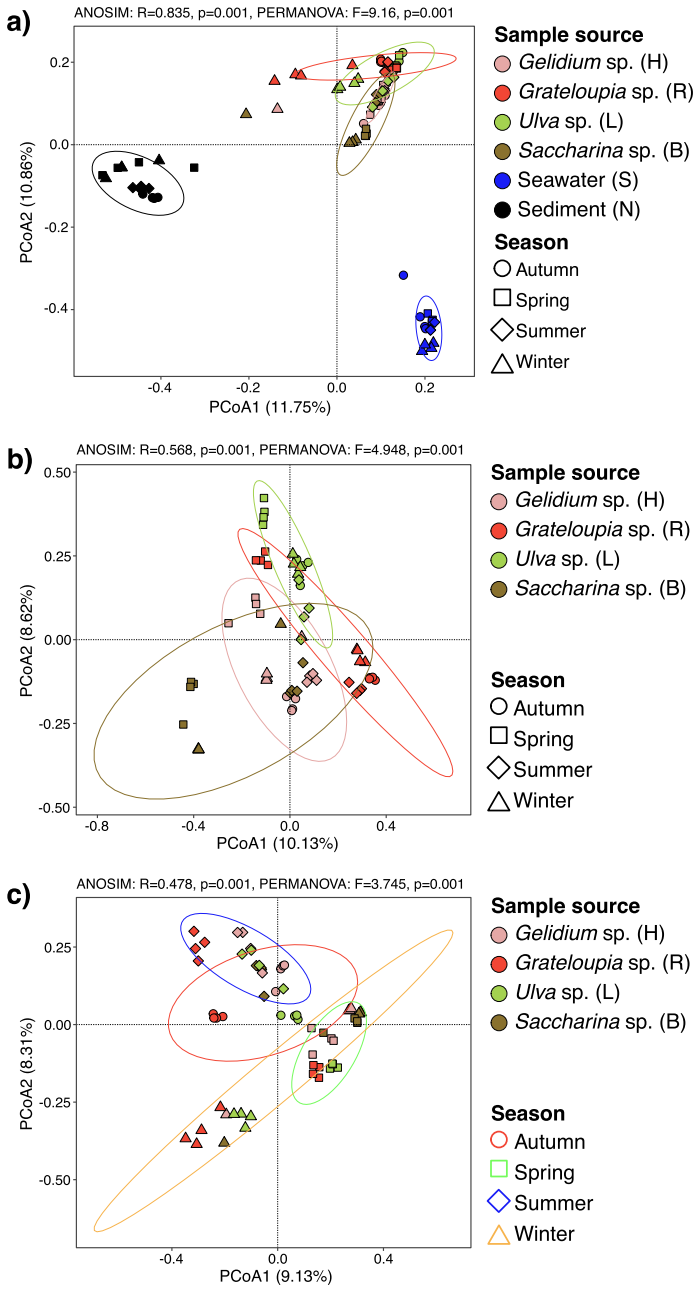


Fig. 2 Principal coordinate analysis (PCoA) plots of Bray–Curtis similarities of samples and seasons calculated using unweighted UniFrac distances (each point corresponds to an individual sample). **a** macroalgal samples ($n=60$), surrounding seawater ($n=15$), and surrounding sediment ($n=17$). **b** only macroalgal samples ($n=60$). **c** only non-core macroalgal samples ($n=60$). Details are provided in Additional file 3

only isolated once (42.1%) or twice (19.3%). According to 16S rRNA amplicon analysis, about half of the macroalgal strains (2,492) exhibited ≥2% abundance in at least one macroalgal sample (Fig. S5 in Additional file 2, Table S2 in Additional file 3). As in 16S rRNA gene amplicon analysis, taxonomy patters of the isolated strains were more similar among macroalgal samples than between these

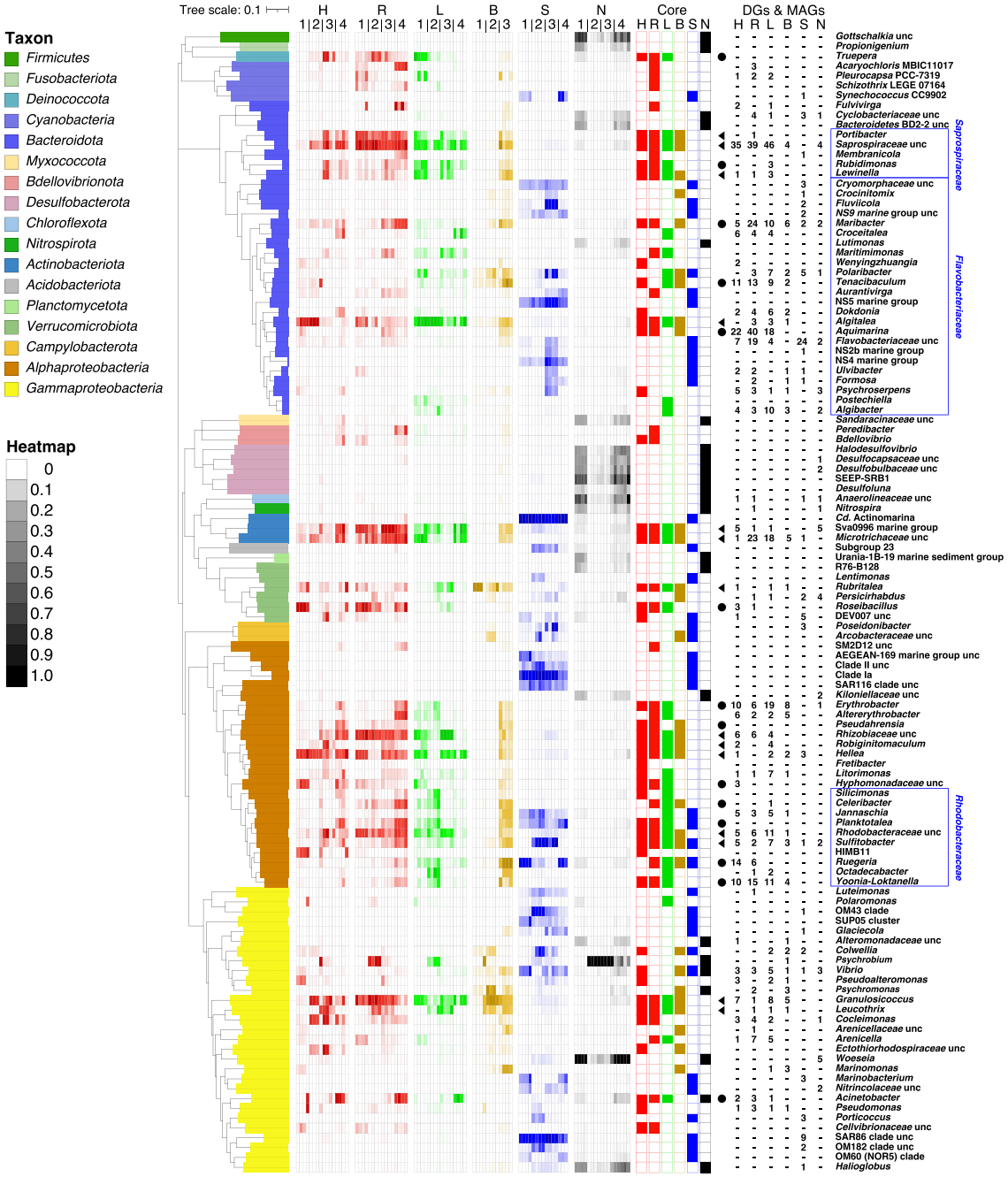


Fig. 3 Phylogeny of 116 genera present in ≥ 85% of the samples of each habitat (four macroalgae plus sediment and seawater controls) with ≥ 1% relative abundance in at least one sample. Phylogenies were calculated using RAxML with 1,000 rapid bootstrap replicates based on similarities of full-length 16S rRNA gene sequences of the corresponding genera from SILVA NR Ref v138. Nomenclature: H = *Gelidium* sp., R = *Grateloupia* sp., L = *Ulva* sp., B = *Saccharina* sp., S = seawater, N = sediment, 1 = autumn, 2 = winter, 3 = spring, 4 = summer. Core phycosphere genera (present on all macroalgae) are highlighted by solid black triangles, and dominant phycosphere genera (present on three macroalgae) by solid black circles. Numbers in the six rightmost columns represent numbers of draft genomes (DGs) and MAGs obtained from all six habitats

Table 1 List of the 14 core and 14 dominant phycosphere genera

Taxa	Relative abundance (AVERAGE ± STDEV) (number of MAGs/draft genomes/cultured strains)					
Genus	***Gelidium* sp. (H)**	***Grateloupia* sp. (R)**	***Ulva* sp. (L)**	***Saccharina* sp. (B)**	**Seawater (S)**	**Sediment (N)**
Saprospiraceae unc[a]	(3.96 ± 3.16) (35/-/-)	(10.15 ± 4.27) (39/-/-)	(4.28 ± 2.39) (47/-/-)	(1.79 ± 2.96) (4/-/-)	(0.08 ± 0.05) (-/-/-)	(0.2 ± 0.14) (4/-/-)
Portibacter[a]	(1 ± 0.96) (-/-/-)	(3.79 ± 1.59) (1/-/-)	(1.21 ± 1.13) (-/-/-)	(0.43 ± 0.72) (-/-/-)	(0.02 ± 0.03) (-/-/-)	(0.05 ± 0.04) (-/-/-)
Lewinella[a]	(0.96 ± 1.36) (1/-/-)	(0.65 ± 0.64) (1/-/-)	(2 ± 2.28) (3/-/-)	(0.88 ± 1.16) (-/-/-)	(0.01 ± 0.02) (-/-/-)	(0.03 ± 0.03) (-/-/-)
Algitalea[a]	(8.9 ± 11.7) (-/-/1)	(1.44 ± 1.17) (-/3/33)	(15.07 ± 13.44) (-/3/42)	(1.36 ± 0.62) (-/1/4)	(0.08 ± 0.06) (-/-/1)	(0.07 ± 0.07) (-/-/-)
Microtrichaceae unc[a]	(2.24 ± 1.95) (1/-/-)	(4.49 ± 2.57) (23/-/-)	(4.89 ± 6.18) (19/-/-)	(0.46 ± 0.58) (5/-/-)	(0.03 ± 0.04) (1/-/-)	(0.11 ± 0.1) (-/-/-)
Sva0996 marine group[a]	(1.97 ± 2.74) (-/-/-)	(9.23 ± 7.87) (5/-/-)	(4.89 ± 6.03) (1/-/-)	(0.64 ± 0.74) (1/-/-)	(0.27 ± 0.16) (5/-/-)	(0.28 ± 0.19) (-/-/-)
Rubritalea[a]	(2.06 ± 4.24) (1/-/-)	(0.62 ± 0.53) (-/-/-)	(1.02 ± 1.51) (1/-/-)	(24.37 ± 31.1) (1/-/-)	(0.13 ± 0.2) (-/-/-)	(0.04 ± 0.06) (-/-/-)
Rhizobiaceae unc[a]	(1.53 ± 1.64) (6/-/-)	(5.97 ± 3.34) (6/-/-)	(9.63 ± 14.93) (4/-/-)	(0.56 ± 0.83) (-/-/-)	(0.1 ± 0.11) (-/-/-)	(0.11 ± 0.13) (-/-/-)
Robiginitomaculum[a]	(1.22 ± 1.97) (2/-/1)	(0.56 ± 0.7) (-/-/-)	(0.32 ± 0.59) (4/-/-)	(0.38 ± 0.49) (-/-/-)	(0.01 ± 0.02) (-/-/-)	(0.01 ± 0.01) (-/-/-)
Hellea[a]	(7.45 ± 5.81) (-/1/2)	(3.08 ± 2.19) (-/-/-)	(4.69 ± 3.82) (2/-/-)	(0.49 ± 0.65) (2/-/-)	(0.11 ± 0.1) (3/-/-)	(0.11 ± 0.13) (-/-/-)
Rhodobacteraceae unc[a]	(2.32 ± 3.34) (9/-/-)	(3.99 ± 1.68) (8/-/-)	(3.45 ± 3.23) (13/-/-)	(0.49 ± 0.55) (1/-/-)	(0.19 ± 0.11) (7/-/-)	(0.1 ± 0.06) (3/1/-)
Sulfitobacter[a]	(0.68 ± 0.7) (2/3/26)	(0.97 ± 0.7) (1/1/18)	(0.64 ± 0.53) (-/7/39)	(1.04 ± 0.97) (-/3/38)	(4.52 ± 5.11) (1/-/2)	(0.07 ± 0.07) (-/2/17)
Granulosicoccus[a]	(7.55 ± 12.59) (5/2/5)	(7.84 ± 7) (1/-/1)	(4.32 ± 3.13) (8/-/-)	(11.65 ± 13.11) (4/1/4)	(0.13 ± 0.08) (-/-/1)	(0.22 ± 0.15) (-/-/-)
Leucothrix[a]	(1.69 ± 2.39) (-/-/-)	(1.13 ± 1.33) (1/-/-)	(2.81 ± 4.56) (1/-/-)	(3.1 ± 3.31) (1/-/2)	(0.04 ± 0.05) (-/-/-)	(0.05 ± 0.08) (-/-/-)
Truepera[b]	(2.12 ± 3.68) (-/-/-)	(1.74 ± 1.9) (-/-/-)	(2.91 ± 5.18) (-/-/-)	(0.15 ± 0.18) (-/-/-)	(0.01 ± 0.01) (-/-/-)	(0.18 ± 0.12) (-/-/-)
Rubidimonas[b]	(0.69 ± 0.82) (-/-/-)	(0.71 ± 0.75) (-/-/-)	(2.08 ± 1.98) (3/-/-)	(0.28 ± 0.36) (-/-/-)	(0.01 ± 0.01) (-/-/-)	(0.01 ± 0.02) (-/-/-)
Maribacter[b]	(0.83 ± 0.65) (4/1/21)	(1.41 ± 1.32) (6/18/81)	(0.31 ± 0.17) (2/8/27)	(1.13 ± 1.48) (1/5/31)	(0.02 ± 0.02) (-/2/2)	(0.12 ± 0.08) (-/2/10)
Tenacibaculum[b]	(0.75 ± 1.76) (-/11/27)	(0.09 ± 0.08) (1/12/46)	(0.31 ± 0.35) (2/7/56)	(3.37 ± 3.01) (1/1/6)	(0.06 ± 0.08) (-/-/1)	(0.03 ± 0.05) (-/-/14)
Aquimarina[b]	(0.29 ± 0.52) (5/17/87)	(0.49 ± 0.7) (3/37/236)	(0.05 ± 0.08) (8/10/38)	(0.37 ± 0.44) (-/-/14)	(0.01 ± 0.01) (-/-/1)	(0.03 ± 0.04) (-/-/1)
Roseibacillus[b]	(6.08 ± 10.53) (3/-/-)	(4.57 ± 4.83) (1/-/-)	(0.35 ± 0.79) (-/-/-)	(0.05 ± 0.06) (-/-/-)	(0.03 ± 0.02) (-/-/-)	(0.07 ± 0.04) (-/-/-)
Erythrobacter[b]	(0.41 ± 0.73) (-/10/60)	(0.47 ± 0.62) (-/6/40)	(0.53 ± 0.73) (5/14/53)	(0.36 ± 0.56) (1/7/22)	(0 ± 0.01) (-/-/1)	(0.03 ± 0.03) (-/1/15)
Pseudahrensia[b]	(0.66 ± 0.82) (-/-/1)	(0.51 ± 0.3) (-/-/-)	(0.27 ± 0.23) (-/-/-)	(0.63 ± 0.86) (-/-/-)	(0.02 ± 0.02) (-/-/1)	(0.09 ± 0.05) (-/-/-)
Hyphomonadaceae unc[b]	(2.41 ± 3.1) (3/-/-)	(0.65 ± 0.47) (-/-/-)	(0.55 ± 0.99) (-/-/-)	(0.25 ± 0.37) (-/-/-)	(0.03 ± 0.04) (-/-/-)	(0.01 ± 0.01) (-/-/-)
Celeribacter[b]	(0.23 ± 0.38) (-/-/-)	(1.43 ± 1) (-/-/-)	(1.21 ± 1.71) (1/-/-)	(0.98 ± 1.13) (-/-/-)	(0.03 ± 0.04) (-/-/7)	(0.02 ± 0.03) (-/-/1)
Planktotalea[b]	(0.61 ± 1.02) (-/-/1)	(1.41 ± 1.22) (-/-/-)	(1.28 ± 1.47) (-/-/-)	(0.72 ± 0.87) (-/-/-)	(1.79 ± 1.83) (-/-/1)	(0.05 ± 0.08) (-/-/1)
Yoonia-Loktanella[b]	(0.17 ± 0.3) (-/10/28)	(0.43 ± 0.39) (2/13/62)	(0.81 ± 1.16) (1/10/69)	(1.01 ± 2.01) (1/3/24)	(0.03 ± 0.03) (-/-/7)	(0 ± 0.02) (-/-/7)
Ruegeria[b]	(0.17 ± 0.25) (-/14/63)	(0.35 ± 0.33) (-/6/56)	(1.04 ± 0.91) (-/-/11)	(3.57 ± 4.87) (-/-/12)	(4.15 ± 4.23) (-/-/1)	(0.18 ± 0.13) (-/-/40)
Acinetobacter[b]	(3.62 ± 6.39) (1/1/1)	(1.78 ± 3.52) (1/2/4)	(5.06 ± 15.95) (1/-/3)	(0.03 ± 0.03) (-/-/2)	(0.12 ± 0.09) (-/-/-)	(0.41 ± 0.31) (-/-/3)

[a] Genus represents a core phycosphere genus

[b] Genus represents a dominant phycosphere genus

and the sediment and seawater samples (Fig. S6 in Additional file 2).

We compared the 16S rRNA sequences of all strains with the 16S rRNA gene amplicon data representing 51,132 bacterial ASV nodes (Table S1 in Additional file 3). At a ≥98.7% identity criterion, 851 of the strains matched 787 ASVs (Table S2 in Additional file 3), with 618 strains matching a single ASVs, and 233 with one-to-many assignments to 169 additional ASVs. At a 97% identity criterion, a mean cultivability of 18.1% was obtained for macroalgal phycosphere species vs. 6.3% and 1.5% for seawater and sediments, respectively. Consequently, CFU numbers obtained from macroalgal samples (5.6 to 5.8×10^5 CFU g^{-1} on average) were two to three orders

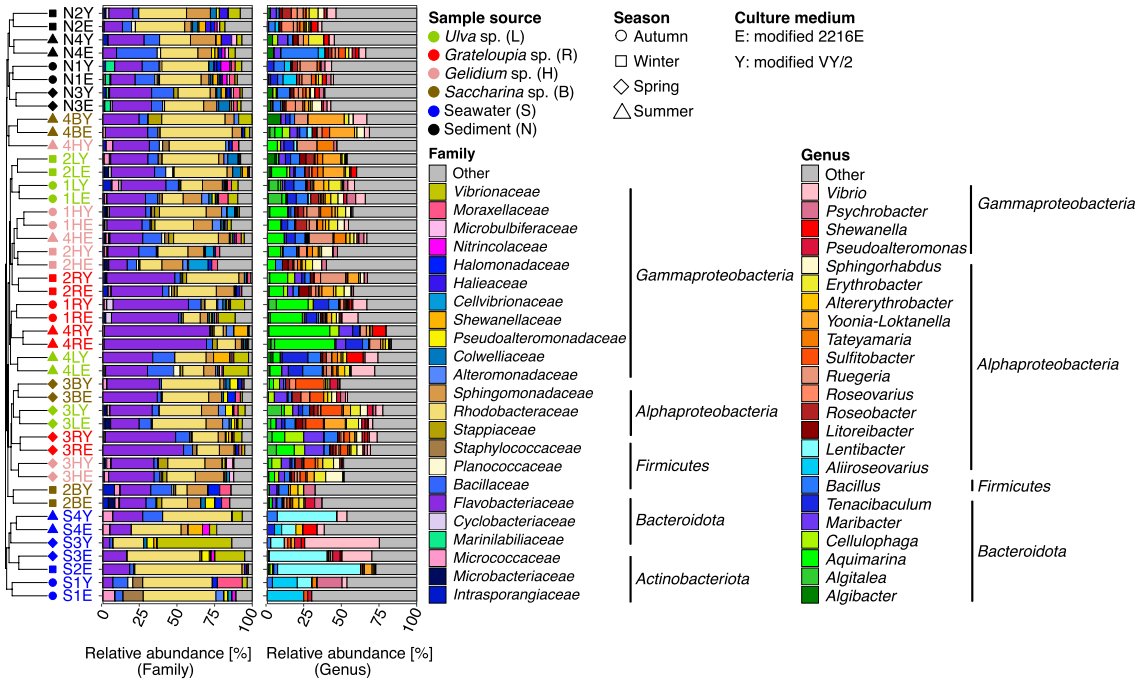


Fig. 4 Cultivable phycosphere bacteria depending on macroalgal host, season and culture medium. Samples were grouped by weighted UniFrac distances using Ward linkage (dendrogram). Mean community compositions of the top 20 taxa are shown for family and genus levels

higher than those from seawater and sediment samples, respectively (Fig. S7 in Additional file 2).

The strains included 735 novel species (577 from macroalgae). Proportions were highest among *Bacteroidota* (62.6%), *Proteobacteria* (53.6%), *Actinobacteriota* (16.1%), *Firmicutes* (7.8%), *Campylobacterota* (100%) and *Verrucomicrobiota* (100%) (Table S2 in Additional file 3). Without consideration of 29 strains with incomplete taxonomies, in total 230 species (1,556 strains) were representatives of 6/14 core and 10/14 dominant phycosphere genera (*Algitalea*, *Granulosicoccus, Hellea, Sulfitobacter*, *Leucothrix*, *Robiginitomaculum,* and *Maribacter*, *Tenacibaculum*, *Aquimarina*, *Erythrobacter*, *Planktotalea*, *Yoonia-Loktanella*, *Ruegeria*, *Acinetobacter*, *Pseudahrensia*, *Celeribacter*) (Fig. 4). In particular, the strains of *Granulosicoccus* (11), *Hellea* (2), *Leucothrix* (2) and *Robiginitomaculum* (1) are noteworthy, since members of these highly abundant phycosphere genera remain difficult to cultivate [6, 10, 12].

Large numbers of draft genomes and MAGs from phycosphere bacteria, including novel species

Based on 16S rRNA sequence similarity, we selected 965 (macroalgae: 864) strains for draft sequencing, including 550 redundant novel species and 42 redundant novel genera (Tables S2, S3 in Additional file 3). Comparisons to 14,131 available published reference genomes [34] revealed that the obtained draft genomes corresponded to 652 species (95% ANI, 65% alignment) represented by 820 non-redundant DGs (99% ANI), including genomes of 399 (macroalgae: 342) novel species, as well as genomes of 246 (macroalgae: 221) species complementing validly described species not yet represented by genomes. From all metagenomes we obtained 1,619 (macroalgae: 936) MAGs with ≥ 50% completeness and < 10% contamination estimates. These corresponded to 1,129 species (95% ANI) represented by 1,184 non-redundant MAGs (99% ANI) (Fig. 1).

In total 961 DGs and 545 MAGs had > 90% completeness and < 5% contamination estimates, but did not fulfill MIMAG 'high-quality' criteria [35] due to 482 lacking complete rRNA gene operons. However, they did adhere to the 'nearly complete' category introduced by Almeida et al. [36]. 82.7% (795/961) of these nearly complete DGs and 88.4% (482/545) of the high-quality MAGs did not affiliate with any described species when using the Genome Taxonomy Database Toolkit (GTDB-Tk) (Fig. S8 in Additional file 2).

In order to determine the total number of species, we also clustered the initial 965 DGs and 1,619 MAGs using a multi-step distance-based approach (95% ANI). This

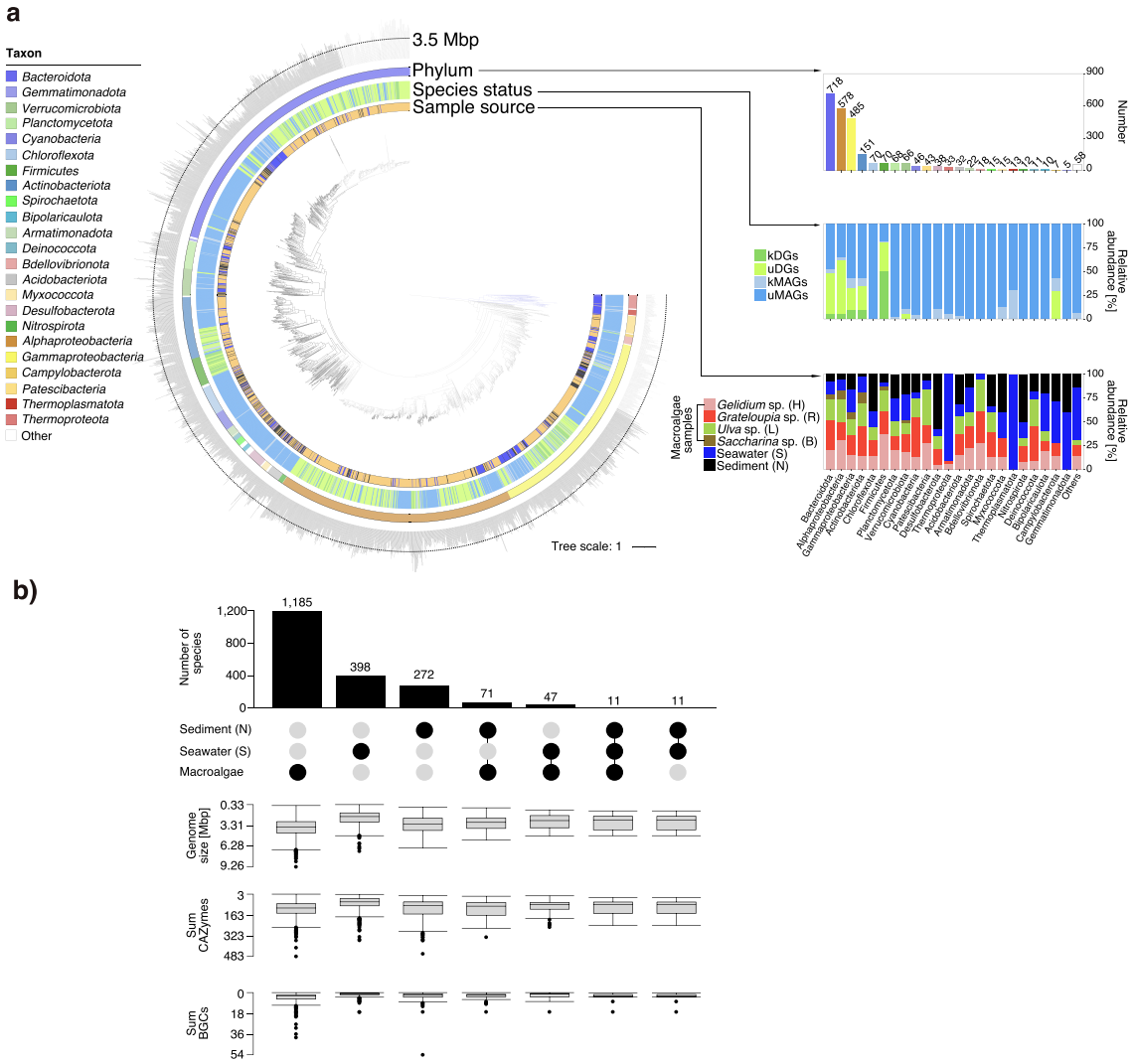


Fig. 5 Metagenome-assembled genomes (MAGs) and draft genomes (DGs). **a** Phylogenomic tree of all 2,584 bacterial MAGs and DGs based on protein sequences of 43 universal single-copy genes with circles representing (inside to outside): (i) sample source and origin of the MAGs and DGs (relative proportions), (ii) known and unknown MAGs and DGs within the most abundant bacteria taxa with ≥ 5 genomes [state: unknown MAGs (uMAGs), known MAGs (kMAGs), unknown draft genomes (uDGs), and known draft genomes (kDGs)], (iii) GTDB phylum classification and absolute (redundant) numbers of MAGs and DGs obtained for each phylum, (iv) genome size (the tree was constructed using anvi'o v6.2 and visualized in iTOL v6.5.6). Total number of genomes from each sample: *Gelidium*: 539; *Grateloupia*: 609; *Saccharina*: 151; *Ulva*: 502; seawater: 469; sediment: 314. **b** Number of species-level MAGs and DGs that were either unique to or shared by sampled habitats. Vertical bars represent numbers of species shared between the study sets indicated by black dots in the lower panel

resulted in 1,781 (macroalgae: 1,185) inferred prokaryotic species, 1,689 *Bacteria* (macroalgae: 1,182) and 49 *Archaea* (macroalgae: 3) (Table S3 in Additional file 3). *Archaea* exhibited only low overall abundances, as did *Firmicutes*. The latter, however, were frequently isolated due to cultivation bias (Fig. 5a).

15/138 species-level genomes of novel core/dominant phycosphere bacteria

We analyzed all genomes representing core/dominant phycosphere genera, consisting of 28/228 (macroalgae: 25/223) DGs and 282/57 (macroalgae: 263/57) MAGs. These included 15 novel core and 138 novel dominant species.

The most frequent core and dominant phycosphere genera comprised *Sulfitobacter*, *Aquimarina*, *Maribacter*, *Tenacibaculum*, *Ruegeria*, *Yoonia-Loktanella*, *Erythrobacter*, *Microtrichaceae* unc., *Saprospiraceae* unc. and *Granulosicoccus* (Fig. 3, Table S3 in Additional file 3). Those represented by high numbers of species exhibited similar abundance patterns on all macroalgae and were hardly found in the control samples. At the family level, an even higher number of isolated strains represented core/dominant phycosphere bacteria (Fig. S4 in Additional file 2).

Phycosphere *Bacteroidota* harbored high proportions of as yet unknown genes

Automatic annotation of DGs and MAGs based on the EggNOG v5, COG (2020) and Pfam (2020) databases resulted in function predictions for on average 80.9%, 75.9% and 77.1% of the genes, respectively (Fig. S9 in Additional file 2). However, when using the more specific UniProtKB and KEGG databases, 46.8% and 75.6% of the genes did not yield any annotations. Among all phyla, the 376 genomes obtained from cultured *Bacteroidota* (305 from macroalgae) had the highest proportion of unknown genes. This exemplifies that macroalgae-colonizing *Bacteroidota* constitute a particularly rich resource of as yet unknown gene functions. Genomes from macroalgal phycosphere bacteria were on average larger than those from sediment and seawater bacteria, with seawater samples featuring the smallest average genome size (Fig. 5b).

It is beyond the scope of this study to interpret the functional potential of all genomes. Instead, we focus on two prevalent traits of phycosphere bacteria, namely their potentials to degrade algal polysaccharides and to synthesize bioactive compounds (Fig. 5b).

Phycosphere *Bacteroidota* dominated the degradation of algal polysaccharides

We searched all DGs and MAGs for carbohydrate-active enzyme (CAZyme) genes and identified 292,848 homologs. *Bacteroidota* (717), *Chloroflexi* (70), *Planctomycetota* (68), *Verrucomicrobiota* (66), *Acidobacteriota* (32) and *Actinobacteriota* (151) genomes encoded the highest proportions of catabolic CAZymes, i.e., glycoside hydrolases (GHs), carbohydrate esterases (CEs), carbohydrate-binding modules (CBMs), auxiliary activities (AAs) and polysaccharide lyases (PLs) (Fig. S10 in Additional file 2). The majority (61.8%) of CAZyme genes were found in *Bacteroidota*, corroborating the pivotal role that members of this phylum play in the degradation of algal polysaccharides [37]. Predicted CAZymes comprised 30.6% GHs, 29.9% glycosyltransferases (GTs), 15.1% CEs, 10.2% CBMs, 5.1% PLs and 5.1% AAs.These proportions were similar across samples (Table S3 in Additional file 3). AAs were more prevalent in macroalgae-associated *Alphaproteobacteria* than in any other phylum (Fig. S10, pie in Additional file 2). Many of the so far described 17 AA families represent lytic polysaccharide monoxygenases, e.g., AA9 acts mainly on cellulose and xyloglucan, AA11 on chitin, AA13 on starch and AA14 on xylan. This suggests a distinct role of *Alphaproteobacteria* in algal polysaccharide degradation.

More than 40% (121,015) of the CAZymes featured signal peptide predictions. Few signal peptides were predicted for GTs (2.4%) and AAs (1.7%), whereas much higher proportions were predicted for PLs (76.5%), GHs (55.6%) and CEs (42.9%), indicating periplasmic or extracellular locations (Table S3 in Additional file 3). These proportions were similar across samples. Surprisingly, the proportion of predicted secreted sulfatases, required for desulfation of sulfated algal polysaccharides [38], were ~11% and ~13% higher in seawater and sediments than in phycosphere bacteria (Table S3 in Additional file 3). In particular, *Planctomycetota* and *Verrucomicrobiota* featured high numbers of CAZyme and sulfatase genes (Fig. S11 in Additional file 2).

We classified candidate loci for polysaccharide degradation into four categories (Fig. S12a in Additional file 2): (i) PULs consisting of CAZyme genes and *susCD* pairs, (ii) PUL-like clusters with CAZyme genes and an encoded TonB-dependent receptor, (iii) CAZyme-rich gene clusters (CGC) consisting solely of CAZymes, and (iv) *susCD* loci without detectable CAZymes. We identified 4,451 PULs, 6,376 PUL-like loci, 19,826 CGCs and 1,699 *susCD* only loci (Table S3 in Additional file 3). The majority were found in DGs (3,461, 3,875, 9,572 and 1,076) (Fig. S13 in Additional file 2, Table S4 in Additional file 3) due to higher overall completeness compared to MAGs. Sulfatase genes were present in 22.3% of the PULs, 5.5% of PUL-like gene clusters, 7.0% of CGCs and 2.9% of *susCD* only loci, underscoring the relevance of polysaccharide sulfation in marine algal polysaccharides (Table S3 in Additional file 3).

Hierarchical clustering according to Bernard [39] with a 100% distance threshold separated the 4,451 PULs into 2,260 clusters. About one-third (763) contained at least two identical PULs, whereas two-thirds were unique. Few PULs were frequent, as only 1.8% (40) of the clusters had more than ten identical instances. Genomes from macroalgae and sediments contained on average more PULs than those from seawater. Compared to seawater, PUL numbers were 1.6 times higher in phycosphere and 2.8 times higher in sediment genomes (Table S5 in Additional file 3). In particular *Bacteroidota* from the phycospheres (*Flavobacteriaceae*) and sediment (*Marinilabiliaceae*) featured more species than seawater samples and higher numbers of more diverse PULs (Fig. 6). In phycospheres,

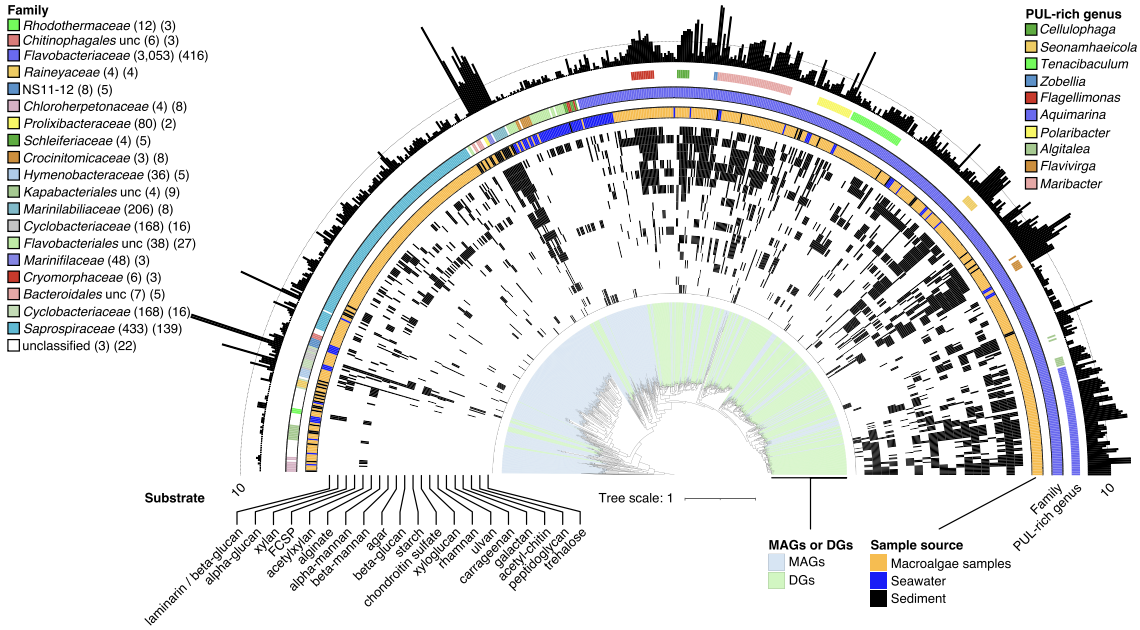


Fig. 6 PUL distribution in metagenome-assembled genomes (MAGs) and draft genomes (DGs). Depicted is a phylogenomic tree for all 741 bacterial MAGs (including 27 unclassified MAGs at the root) and DGs based on protein sequences of 43 universal single-copy genes with circles representing (inside to outside): (i) MAGs or DGs, (ii) predicted polysaccharide degradation capacities based on PUL-associated CAZyme annotations, (iii) sample source, (iv) GTDB family classification, v) highlighting of PUL-rich taxa, (vi) bar chart representing the number of predicted PULs. Numbers in parentheses indicate PUL numbers and genome numbers in the corresponding families, respectively

PUL-rich species mainly belonged to *Zobellia*, *Polaribacter*, *Aquimarina*, *Tenacibaculum*, *Algitalea* and *Maribacter*, representing either core or dominant phycosphere genera (Fig. 6). Additional PUL-rich genera comprised *Cellulophaga*, *Flagellimonas*, *Flavivirga* and *Seonamhaeicola,* which were mainly isolated from macroalgae (Fig. 6). In sediments, *Prolixibacteraceae* and *Marinilabiliaceae* were particularly PUL-rich (both up to 30 PULs), and in seawater *Maribacter* species (up to 24 PULs) [29] (Fig. 6, Table S3 in Additional file 3).

The largest PUL (tandem repeat and hybrid *susCD* PUL) of in total 99 genes (48 CAZyme genes) was found in the core phycosphere species *Algibacter* sp. 4-1052 (*Bacteroidota*; *Flavobacteriaceae*) isolated from *Ulva* sp. (Table S4 in Additional file 3). This PUL, rich in GH29, GH106, PL40, PL25 and sulfatase genes, may target fucoidan, ulvan and/or rhamnogalacturonan (Fig. 7). The largest CGC (93 genes) was found in a *Gaetbulibacter* species (*Bacteroidota*; *Flavobacteriaceae*) isolated from *Grateloupia* sp. and sediment (Table S4 in Additional file 3). *Draconibacterium* sp. X8 (*Bacteroidota*; *Prolixibacteraceae*) isolated from *Gelidium* sp. featured the highest number of PULs (50) (Table S3 in Additional file 3), the third highest number of CAZyme genes (412), and the highest percentage of CAZymes in PULs (85.7%).

Sequence analysis of PUL-encoded SusC and SusD substrate-binding and take-up proteins can provide hints on possible glycan substrates [40]. Hence, we combined phylogenetic SusC/D protein tree and PUL CAZyme composition analyses to infer possible substrate classes (Additional file 1). The complete SusC/D protein tree featured 157 SusD and 159 SusC clusters. Each cluster contained at least five SusC/D protein sequences and represented PULs of similar CAZymes composition (Fig. S14 in Additional file 2, Table S5 in Additional file 3). Examples are GH3/GH16 for β-glucans (including laminarin), GH13/GH65 for α-glucans or PL6/PL7/PL12/PL17 for alginate. The most frequent predicted substrates were xylose-containing polysaccharides (779) (178 PULs containing solely putative acetylxylan esterases of the CE1, CE3 or CE4 families), β-glucans/laminarin (618), α-glucans (482), fucose-containing sulfated polysaccharides (FCSPs) (444), alginates (426), α-mannans (268), β-mannans (220), sulfated α-rhamnose-containing polysaccharides (219), agars (192), chondroitin (158), xyloglucan (133) galactans (128), ulvans (127), starch (114), carrageenans (109), chitin (109), pectin (72), peptidoglycan (69), levans/fructans (36) and porphyran (31) (Fig. S14 in Additional file 2, Table S5 in Additional file 3). In general, a large number of PULs were rich in sulfatase

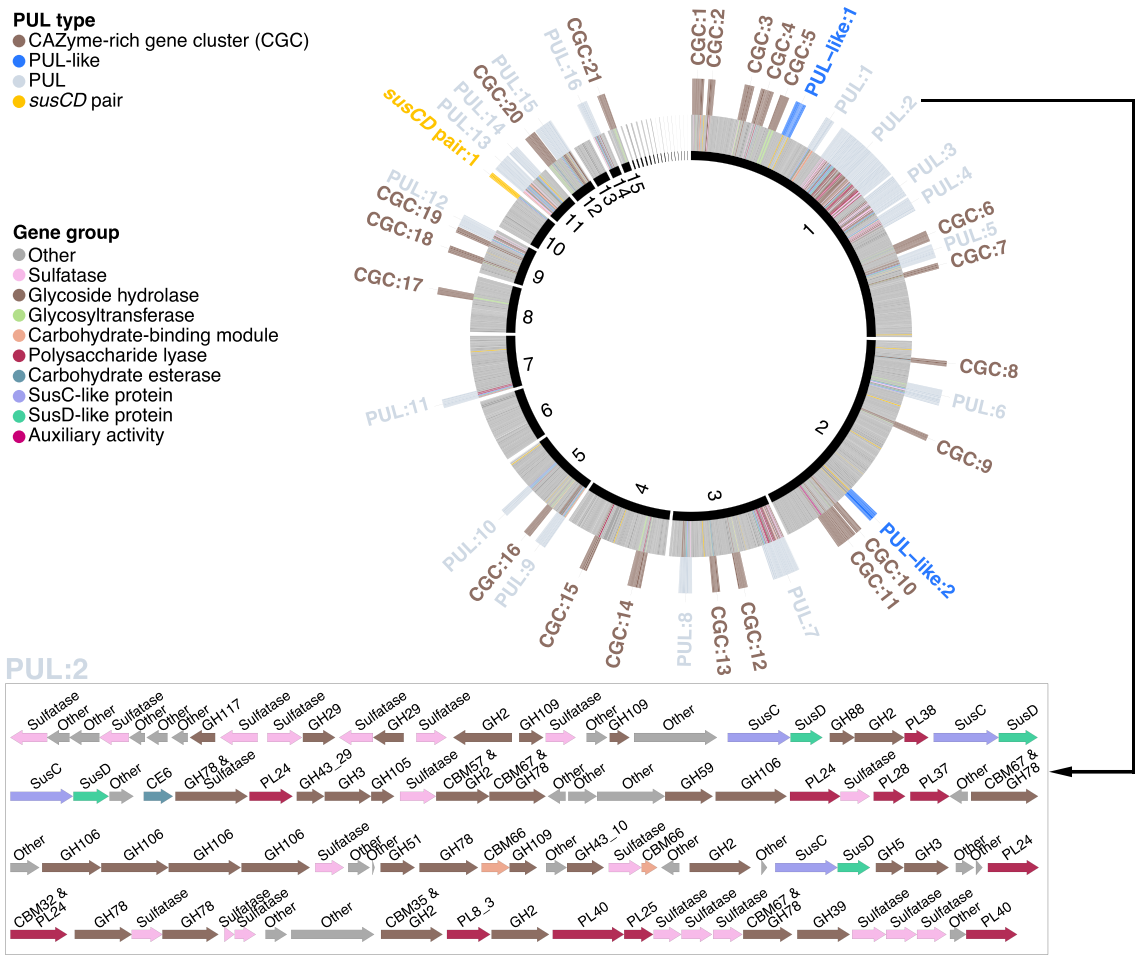


Fig. 7 Overview of the *Algibacter* sp. strain 4–1052 draft genome. From inside to outside: (i) contig ID (sorted by lengths), (ii) CAZyme and sulfatase genes, (iii) positions of loci potentially involved in polysaccharide degradation, (iv) locus type. Inset: Structure of the longest PUL (PUL:2)

or deacetylase genes, suggesting sulfated and acetylated polysaccharide substrate targets (Table S6 in Additional file 3). Of course, PULs with common substrate predictions were not exactly identical due to the extent of variation in PUL compositions (Table S5 in Additional file 3). Consequently, a wide range of as yet undescribed PULs was identified, and some larger PULs were ascribed to multiple polysaccharide substrates (Fig. S14 in Additional file 2, Table S5 in Additional file 3).

Phycosphere taxa, in particular *Bacteroidota*, were surprisingly rich in biosynthetic gene clusters

We identified 8,810 putative BGCs (Table S7 in Additional file 3). Predicted product classes comprised terpenes (28.3%), bacteriocins (12.3%), non-ribosomal peptides (NRPS) (10.5%) and NRPS-like clusters (8.0%), homoserine lactones (7.8%), type III polyketide synthases (7.5%), type I polyketide synthases (5.9%) and beta-lactones (5.4%).

Since DGs were generally more complete than MAGs (Fig. S15 in Additional file 2), they featured lower proportions of incomplete BGCs (Fig. S16 in Additional file 2). 20.1% of the 4,816 BGCs predicted in DGs resided on contig edges and were thus potentially incomplete, while this was the case for 73.2% of the 3,994 BGCs predicted in MAGs. We observed clear distinctions between phyla (Fig. S17a in Additional file 2), but no clear trends were observed for BGC families with respect to habitat (Fig. S17b in Additional file 2). Still, we identified more than 483 BGCs > 50 kbp and 1,561 BGCs > 30 kbp (Table S7 in Additional file 3). The largest was identified in a *Streptomyces* species retrieved from *Gelidium* sp. It coded for no less than 22 PKS and NRPS modules.

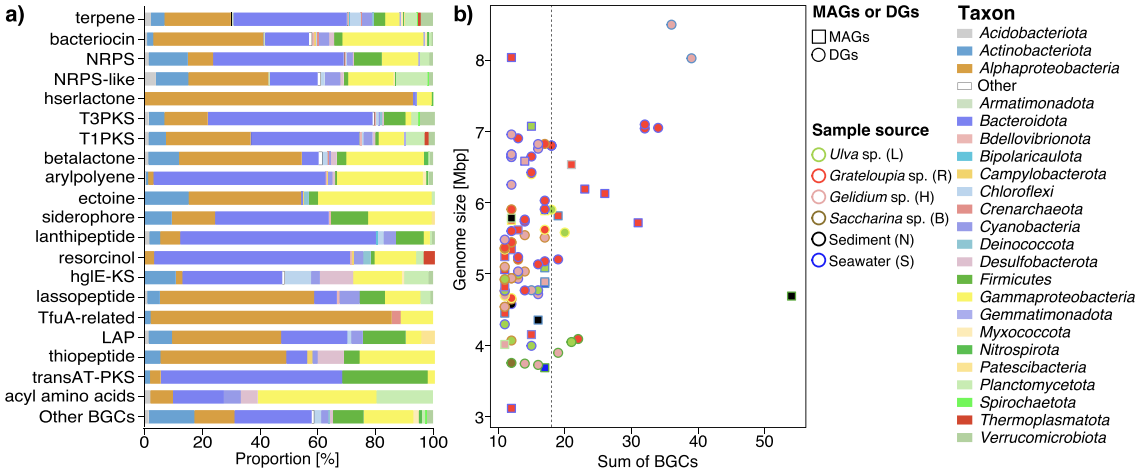


Fig. 8 Biosynthetic gene cluster composition and distribution among 1,619 metagenome-assembled genomes (MAGs) and 965 draft genomes (DGs) from all samples. **a** Proportions of BGC types in MAGs and DGs of different phyla. **b** Top 100 BGCs versus genome sizes with MAGs represented by squares and DGs by circles. Fill colors represent taxonomies, and border colors sample sources. Circle and square sizes correspond to genome sizes. The right side of the dotted line represents the top 20 with the largest number of BGCs, which mainly belong to the *Bacteroidota*. Details are provided in Table S4 in Additional file 3

Ninety-three of the top 100 genomes with the highest number of BGCs belonged to phycosphere bacteria and ten of the top 20 genomes with the highest number of BGCs belonged to phycosphere *Bacteroidota* (Fig. 8b). The latter indicates that the potential for secondary metabolite production in this phylum may as yet have been underestimated. *Bacteroidota* had high proportions of BGCs for terpene and NRPS biosynthesis (Fig. 8a), e.g., the novel core phycosphere species *Aquimarina* sp. 2-328 (Table S7 in Additional file 3).

Most BGCs were identified in *Bacteroidota*, *Alphaproteobaceria*, *Gammaproteobacteria*, *Firmicutes* and *Actinobacteriota* (Figs. 9a, S16 in Additional file 2), all taxa that are rich in core phycosphere bacteria. *Firmicutes* and *Actinobacteriota* are known for abundant secondary metabolite production [33]. We found 559 BGCs in 151 *Actinobacteriota* genomes (including 100 MAGs), covering a broad diversity of predicted products. While the highest number of BGCs (54) was found in a *Firmicutes* MAG from sediment (Fig. 8b), the second (39) and third (36) highest numbers were found in draft genomes of actinobacterial *Streptomyces* strains 3-371 isolated from macroalgae (Fig. 9c). *Alphaproteobacteria* were particularly rich in BGCs, many coding for homoserine lactones, especially the core phycosphere family *Rhodobacteriaceae* (Fig. 9a, b), e.g., the phycosphere species *Roseovarius* sp. 3-342 (*Rhodobacteraceae*) isolated from *Gelidium* sp (Fig. 9b, Table S7 in Additional file 3) contained six related gene clusters.

Discussion

Approximately 40–80% of the *Bacteria* and *Archaea* on Earth reside in biofilms [41]. Selected biofilms have been extensively studied [33], but little is known about the diversities and functions of marine macroalgal biofilms, in particular on a global scale. Algal colonization is influenced by stochastic as well as deterministic processes. While functionally redundant yet taxonomically distinct species can replace each other (stochastics) [4, 22], it has also been shown that phycosphere bacteria share a robust pool of essential genetic functions (determinism) [23]. Both allow for largely varying phycosphere compositions, but more selective processes must be at play, since it has also been reported that phycosphere communities are at least in parts host-specific [21].

We observed surprisingly stable core phycosphere compositions across all four studied algae species on genus and family levels, in particular with respect to dominating members of *Alphaproteobacteria*, *Gammaproteobacteria* and *Bacteroidota*. Core genera, while representing only a minor proportion of the phycosphere diversities, made up a major proportion of the phycosphere abundances, even though their relative proportions fluctuated throughout seasons. This is unlikely a purely biogeographic effect of sampling in close proximity, because some core genera have also been described in other studies [22, 23]. Phycospheres of *Ulva australis* for example feature high abundances of *Lewinella* (*Lewinellaceae*), *Maribacter* (*Flavobacteriaceae*),

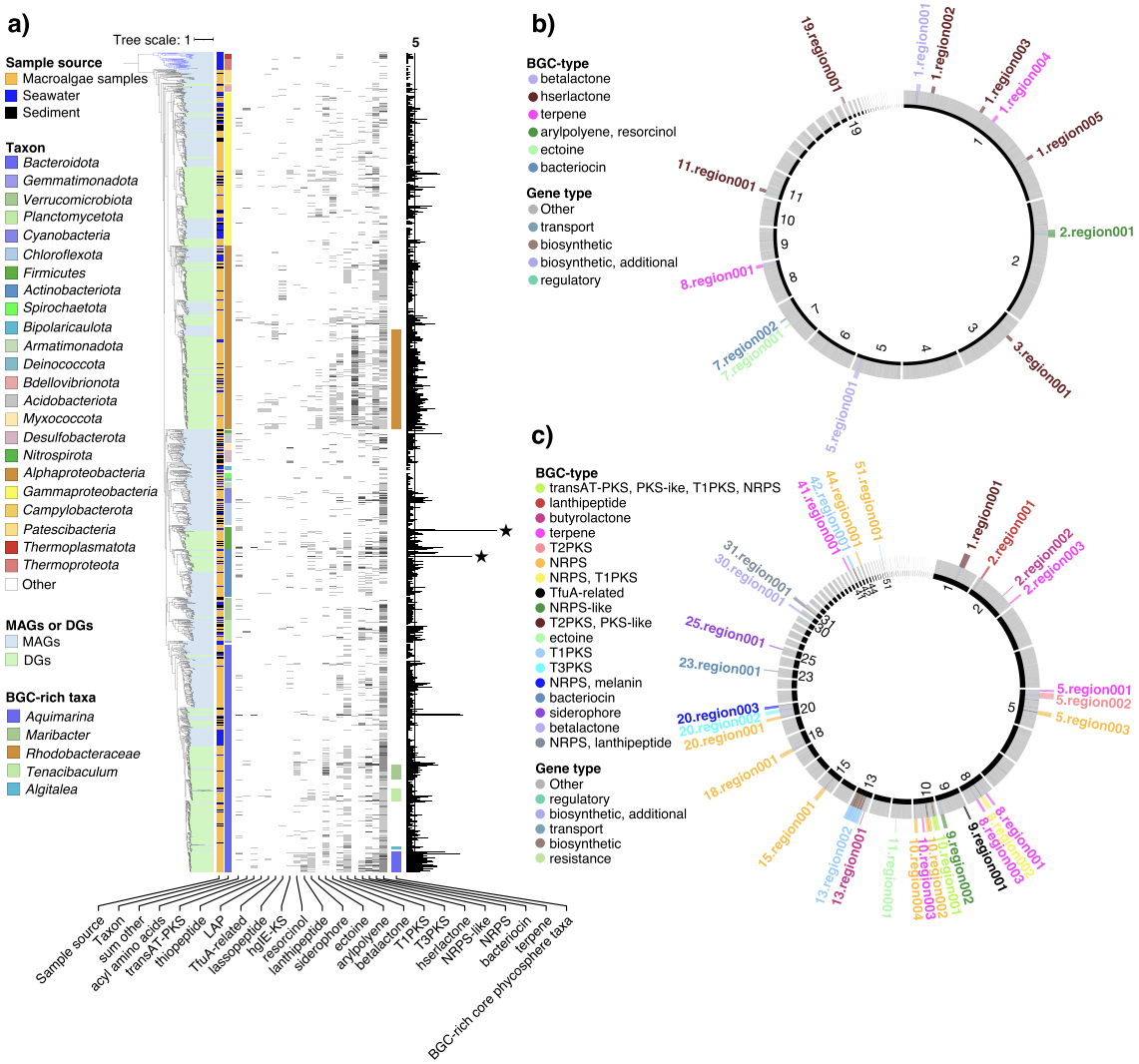


Fig. 9 Overview of biosynthetic gene clusters. **a** Phylogenomic tree for all 2,584 bacterial metagenome-assembled genomes (MAGs) and draft genomes (DGs) based on protein sequences of 43 universal single-copy genes (blue branches represent *Archaea*). From left to right: (i) origin: MAG or DG, (ii) sample source, (iii) GTDB phylum annotation, (iii) the number of various abundant BGCs, (iv) BGC-rich core phycosphere taxa, and (v) the sum of BGCs. The two strains with the most BGCs *Ruminiclostridium* sp. (*Firmicutes*) and *Streptomyces* sp. (*Actinobacteriota*) are marked by asterisks. **b** Overview of BCGs in *Roseovarius* sp. strain 2–342. From inside to outside: (i) contig ID (sorted by lengths), (ii) genes related to BGCs, (iii) BGC type, (iv) BGC identifier. **c** Overview of BCGs in *Streptomyces* sp. strain 3–371. From inside to outside: (i) contig ID (sorted by lengths), (ii) genes related to BGCs, (iii) BGC type, (iv) BGC identifier

Loktanella (*Roseobacteraceae*), *Sulfitobacter* (*Roseobacteraceae*) and *Erythrobacter* (*Erythrobacteraceae*) [4]. Also, *Granulosicoccus* has been shown to dwell on multiple macroalgal species [10].

It seems that the sampled reef harbors a pool of common and widespread potential phycosphere bacteria, some of which are more successful in macroalgal colonization than others, in particular members of the core/dominant genera. Superimposed are host-specific and stochastic phycosphere taxa. To elucidate, whether or not the core/dominant community is stable over longer periods of time, or gradually changing as it is part of a larger pool of suitable bacteria that can functionally replace each other, would require multiple years of consecutive studies and thus remains an open question.

The *Flavobacteriaceae* and *Saprospiraceae* core families are of particular interest. *Flavobacteriaceae* are known to degrade biopolymers [40] and have been found in various marine [42] and terrestrial habits [42, 43], and in association with microalgae [40], macroalgae [3, 4] and marine animals [42]. Symbiotic *Flavobacteriaceae* are also known to produce vital compounds for their hosts [44, 45]. For instance, members of the genus *Zobellia* are known to induce morphogenesis of *Monostroma oxyspermum* green algae [45]. Likewise, *Saprospiraceae* have been isolated from diverse marine habitats, including seawater, particles, sediments and macroalgae such as *Ulva* spp. and *Delisea pulchra* [3, 4, 46]. Members of the *Saprospiraceae* are likely involved in the breakdown of complex organic compounds [47] and in algal endosymbiosis [43].

Verrucomicrobiota are also known to be associated with macroalgae [3]. Members of the *Verrucomicrobiota* and its sister phylum *Planctomycetota* [48] have been suggested as specialists for sulfated algal polysaccharides, since their genomes tend to feature copious sulfatase genes [49]. Verrucomicrobial *Rubritaleaceae* are known to feature biofilm-forming bacteria [50] and were abundantly present on *Saccharina* sp. winter samples. The latter might be a consequence of *Saccharina* sp. being in the seeding stage during this time. Recent studies indicate that some free-living *Verrucomicrobiota* specialize in the degradation of fucose- and rhamnose-rich algal polysaccharides including fucoidan [49, 51].

Expanding the catalog of known algal phycosphere bacterial species

Most bacteria from marine macroalgae resist common cultivation techniques, and those that have been cultured mostly belong to the 'rare biosphere' [52]. In this study, we could culture strains from 367 genera (macroalgae: 302), including six (*Hellea*, *Algitalea*, *Sulfitobacter*, *Granulosicoccus*, *Leucothrix*, *Robiginitomaculum*) core and ten (*Maribacter*, *Tenacibaculum*, *Aquimarina*, *Erythrobacter*, *Planktotalea*, *Yoonia-Loktanella*, *Ruegeria*, *Acinetobacter*, *Pseudahrensia*, *Celeribacter*) dominant phycosphere genera (Fig. 3) (Table S2 in Additional file 3). The cultured core phycosphere species mainly belong to the *Rhodobacteraceae* and *Flavobacteriaceae* families (55.4% of the total). In addition, 29 strains were obtained with either unresolved or incomplete taxonomies. About eight to nine times as many dominant than core species were obtained using cultivation. Conversely, four to five times as many MAGs of core than dominant species were obtained using metagenomics. This illustrates that some core taxa are difficult to cultivate and that a large fraction of the core phycosphere species remains without a cultured representative. However, as exemplified by our study, macroalgal phycospheres also host high numbers of cultivable species that can be readily explored.

As of June 2022, the number of validly published prokaryote species stood at 18,297 with a total of 3,365 genera (names validly published under the ICNP, w/o synonyms; https://lpsn.dsmz.de/text/numbers). These numbers are far from reflecting the existing natural bacterial diversity. Among the so far validly described cultured species, only 203 were obtained from macroalgae. In this study, we isolated 689 novel species, the most prevalent of which need to be validly described. Still, much of the diversity of the macroalgal microbiome remains uncultured, including prevalent clades with important ecophysiological functions.

Polysaccharides and PULs

Variations in chemical structures of macroalgal polysaccharides depend not only on the species, but also on the body parts and developmental stage of the sampled macroalgae, season, and other environmental factors [25]. Bacteria that degrade such polysaccharides require numerous or adaptive, complex PULs to account for these variations. A single PUL often encodes the entire apparatus to degrade a specific glycan, but in the case of chemically complex glycans, it has been shown that multiple PULs can be involved [53]. This might explain, why in *Bacteroidota* we observed not only large numbers of PULs, but also a high diversity of CAZyme genes, in particular in large hybrid *susCD* PULs (Fig. 7).

The current challenge is not to obtain more PUL data, but rather to infer the functions of the plethora of PULs that have already been identified. The PUL gene repertoire and diversity in phycosphere *Bacteroidota* suggest a high level of functional redundancy, which may enable adaptation to various macroalgal hosts. This redundancy might be the result of PUL acquisitions via horizontal gene transfer [23, 54]. Indicative of the latter is that PUL patterns were not always congruent with the 16S phylogeny (Fig. 6).

We found similar collective PUL repertoires in the epiphytic bacteria of all sampled macroalgae, which supports the presence of functional guilds within the macroalgal microbiome with members that can functionally fill in for each other. In particular, *Bacteroidota* in all sampled habitats were rich in PULs, underpinning the exceptional role that *Bacteroidota* play in marine polysaccharide degradation. PULs predicted to target well-defined, structurally simple polysaccharides, such as laminarin, starch and alginate, comprised fewer CAZyme genes and were more conserved than PULs predicted to target more complex polysaccharides, such as carrageenans and ulvans. Some of the larger, complex PULs might

actually address multiple substrates. For example, cluster 27_1 in the SusC/D protein tree comprised carrageenan PULs with a family GH5_2 gene (Table S5 in Additional file 3). The latter might target either xylan (endo-β-1,4-xylanase function) or cellulose (endo-β-1,4-glucanase function), which often coexist with carrageenan in natural habitats. Likewise, PULs predicted to target ulvans and rhamnogalacturonans contained additional endohydrolases seemingly unrelated to the actual substrate. The reason might be that algal sulfated polysaccharides are rarely homogeneous, but mostly complex heterogeneous mixtures of different glycans [55]. Further predicted substrates included sulfated α-rhamnose- and α-galactose-containing polysaccharides, FCSPs, agars, and fructose-rich polysaccharides such as fructans and levans, plus bacterial polysaccharides such as gellan, peptidoglycan, O-antigenic side chains, eukaryotic N-glycans, and common small sugar molecules, such as trehalose and sialic acids (Additional file 1).

Recalcitrant macroalgal polysaccharides eventually end up in the sediment [1], and some sediment taxa with high numbers of CAZymes and PULs, such as bacteroidotal *Marinilabiliaceae* and *Prolixibacteraceae*, have the potential to further degrade such polysaccharides (Figs. 6, S18 in Additional file 2). In our samples, *Marinilabiliaceae* from sediments featured similar PUL numbers than macroalgal core taxa (Fig. 6). We therefore suppose that *Marinilabiliaceae* play an important role in the degradation of macroalgal polysaccharides in marine sediments. *Planctomycetota* and *Verrucomicrobiota* also seem to play such a role in sediments, as they featured more CAZyme genes than those from macroalgal samples, but fewer sulfatases (Fig. S11 in Additional file 2). Interestingly, *Planctomycetota* and *Verrucomicrobiota* in seawater featured more sulfatase genes than those from macroalgae and sediments. This is likely a consequence of different dominating taxa (Fig. S11 in Additional file 2), and might indicate that those in phycospheres seem to preferentially degrade less sulfated and thus more accessible polysaccharides.

Secondary metabolites

Phycosphere bacteria are known to produce secondary metabolites, including antibacterial substances [46, 56]. The latter are crucial for maintaining a specific phycosphere community composition [57].

Phycosphere bacteria in our samples had larger genomes and relatively more BGCs compared to seawater and sediment bacteria (Fig. 5b). There were also notable taxonomic differences (Figs. 3, S4 in Additional file 2). *Flavobacteriaceae* and *Rhodobacteraceae* comprised core/dominant phycosphere taxa with remarkably high BGC proportions (Fig. 9a), for example, members of the genera *Maribacter*, *Algitalea*, *Tenacibaculum*, *Aquimarina*, *Ruegeria* and *Sulfitobacter* (Fig. 9a). Six of the topmost ten abundant phycosphere genomes originated from these two families, which is why respective isolates should be prime targets for the discovery of novel bioactive agents. The *Actinobacteriota* constitute a prime source for the discovery of new drugs. In particular, *Streptomyces* species are prolific producers of antibiotics and other natural agents (Fig. 9a, c) [58]. Due to the depletion of secondary metabolite resources of terrestrial actinomycetes, representatives from marine macroalgal phycospheres, such as *Streptomyces* spp., may become future viable substitutes. For example, actinobacterial *Microtrichaceae* in this study represented a core phycosphere family. While we did not succeed in cultivating a representative species (but did obtain 62 MAGs), macroalgal phycospheres are rich in *Microtrichaceae* and thus a viable resource for the isolation of novel marine actinomycetes (Fig. S4 in Additional file 2). Further non-core/dominant phycosphere genera with members rich in BGCs comprised *Kordiimonas*, *Shewanella*, *Kocuria* and *Bacillus*.

Homoserine lactones, such as N-acyl-L-homoserine lactones (AHLs), act as messenger molecules that enable bacteria to collectively change gene expression, a process known as quorum sensing (QS) [59]. Bacteria isolated from plants [59], macroalgae [60] and animals [61] have been shown to produce AHLs. The first marine phycosphere bacterium for which QS was shown was isolated from the red macroalga *Delisea pulchra,* which appears to have developed natural defense mechanisms to prevent microbial surface fouling [60]. Likewise, almost 40% of the strains isolated from the brown macroalga *Fucus vesiculosus* were able to degrade AHLs [62], suggesting that inhibition of QS could be widespread among algae-associated bacteria. A total of 690 homoserine lactone BGCs were predicted in our study, most in *Rhodobacteraceae*, representing one of the most prevalent core phycosphere families (Figs. S4 in Additional file 2, 9a). *Rhodobacteraceae* could thus play a key role in controlling algae colonization [59].

The bacterial endosymbiont *Cd.* Endobryopsis kahalalidefaciens of *Bryopsis* sp. green algae has abundant and diverse NRP-synthesis BGCs that it uses to produce toxins for the defense of its host [44]. Pure cultures of symbiotic bacteria are usually hard to obtain, whereas epiphytic *Bacteroidota* of macroalgae also have rich NRPS-synthesizing BGCs and are more readily available (Figs. 6a and 9a). Still, the successful translation of NRPS BGCs from phycosphere bacteria via NRPS/PKS megasynthases for drug discovery remains a major challenge for the future.

Terpenes constitute another diverse class of compounds that are mainly produced by plants and fungi

[63]. Also *Cyanobacteria* [32] and *Planctomycetota* [48] are known to feature terpenoid biosynthesis pathways. Both are well represented among the dominant phycosphere taxa, suggesting the production of terpenoid compounds. In addition, we observed the presence of terpene synthesis gene clusters in *Alphaproteobacteria* and in *Bacteroidota* (Figs. 6a and 9a). Most of the predicted BGC products were unclassified (Table S7 in Additional file 3), which reflects our limited knowledge on secondary metabolites and substantiates that phycosphere bacteria represent a rich resource of as yet unexplored biosynthetic functions.

Conclusions

To our knowledge, this dataset represents the largest effort so far on phycosphere bacteria in terms of phylogenetic coverage, cultured isolates and genome data. Our study not only corroborated that all sampled macroalgae were characterized by similar phycosphere communities, but also yielded 689 isolates of novel species. In particular, we succeeded in cultivating a sizable number of strains of core and dominant phycosphere members for future in-depth functional studies. At the same time, we expect that the genome data provided in this study will act as a valuable search space for future metatranscriptome studies of entire macroalgal microbiomes.

As yet, abundant heterotrophic phycosphere bacteria, in particular from the *Planctomycetota*, *Verrucomicrobiota* and *Chloroflexota*, remain uncultured, and thus should be a focus in future studies. Such studies should also include more algal species and multiple sites. Our data represents a stepping stone in this direction and will hopefully serve as a sound basis for further and refined research on the specific adaptations of core phycosphere bacteria.

Materials and methods

Sampling

We sampled a coastal area in Weihai, China (122.12 N, 37.56 E) in 2018/19 on October 15th, January 15th, May 1st, and August 1st. Live *Ulva* sp. (green algae), *Saccharina* sp. (brown algae), *Grateloupia* sp. (red algae), *Gelidium* sp. (red algae), surrounding seawater (-0.1 to -0.5 m) and surface sediment (~5 m depth) were collected in triplicates in sterile plastic bags, kept on ice and transported to the laboratory within 2 h. At each time point, all four macroalgal species were sampled, with the exception of August, where *Saccharina* sp. was decomposed due to summer temperatures. In total, we sequenced 23 metagenomes and 92 16S rRNA gene tag libraries, and isolated 5,527 bacterial strains, 965 of which were draft sequenced (Fig. 1).

Cultured bacteria

Extraction and isolation by dilution of bacteria from phycosphere, seawater and sediment samples are described in Additional file 1. Two media were used for plating, modified 2216E and modified VY/2 medium (Additional file 1). Colonies were selected depending on color, size, and shape. Picked colonies were purified by serial cultivation on plates with identical media. Purified strains were stored at -80 °C in sterile 1% (w/v) saline medium with 15% (v/v) glycerol.

For 16S rRNA gene sequencing the universal bacterial primers 27F and 1492R were used as described elsewhere [64]. PCR products were subsequently Sanger-sequenced by BGI Co. Ltd. (Qingdao, China). Resulting sequences were classified using the EzTaxon server [65] to identify known taxa (≥98.7% similarity to published type strains). Additional taxonomic assignments were done using SILVA v138.1 [66].

Strains of novel species lacking reference genomes in the Type Strains Genome Database [67] and strains present on all macroalgal samples were selected for sequencing. Sequencing was performed by Beijing Novogene Biotechnology (Beijing, China) on a NovaSeq (Illumina, San Diego, CA, USA) with 150 bp PE reads at ≥100×coverage. Reads were quality-filtered and assembled with SPAdes v3.9.1 [68] (–careful –cov-cutoff) with k-mer sizes from 27 to 127 bp and a minimum scaffold length of 200 bp. Further details are provided in Additional file 1.

Environmental 16S rRNA gene tags

We sequenced 16S rRNA gene V3-V4 regions using primers 341F and 806R as described elsewhere [69]. Sequencing was carried out on the Illumina NovaSeq platform using 2×250 bp chemistry at Guangdong Magigene Biotechnology Co., Ltd. (Shanghai, China). Cutadapt v3.0 [70] was used to remove primers and adapters. Reads were trimmed to ≥Q25, and dereplicated using DADA2 [71] (paired-end setting) resulting in tabulated read counts of amplicon sequence variants (ASVs). ASV taxonomies were assigned based on a ≥97% similarity criterion to 16S rRNA sequences in the SILVA v138.1 database, and a 97% similarity threshold was also used for creating OTUs in SILVAngs [72]. Chloroplast and mitochondria sequences were removed from subsequent analyses.

Metagenome-assembled genomes (MAGs)

Library construction and sequencing of metagenomes were performed as presented in Additional file 1. A total of 1.4 Tbp (avg. 65 Gbp per metagenome) were generated (Table S1). Read quality filtering was done with BBDuk v35.14 (http://bbtools.jgi.doe.gov) and verified

with FastQC v0.11. Reads from each sample were subsequently assembled individually using MEGAHIT v1.2.9 [73] with a minimum scaffold length of 2.5 kbp.

BAM files were generated for each metagenome by mapping reads onto assemblies with BBMap v38.86 (minid = 0.99, idfilter = 0.97, fast = t and nodisk = t.) Initial binning was performed from within anvi'o v6.2 [74] using CONCOCT v0.4.0 [75], MaxBin v2.1.1 [76] and MetaBAT v0.2 [77]. Resulting bins were combined with DAS Tool v1.1 [78] in order to find an optimal set. Anvi'o was used for manual bin refinement and CheckM [79] and Prokka v1.13 [80] were used for estimating completeness of MAGs. Genomes were classified into high-, medium-, and low-quality classes according to MIMAG criteria [35].

MAGs were denoted by an initial capital letter specifying the sample (B = *Saccharina*, L = *Ulva*, H = *Grateloupia*, R = *Gelidium*, S = seawater, N = sediment), followed by a number representing the season (1 = autumn, 2 = winter, 3 = spring, 4 = summer), followed by the binning program, and a terminal numeric identifier (Table S3).

Taxonomic inference of MAGs and draft genomes

Initial taxonomic classification of MAGs and draft genomes was done with GTDB-Tk v1.3.0 [81] using the default classify_wf command. In addition, 16S rRNA genes were predicted with Barrnap (https://github.com/tseemann/barrnap) and classified with SILVA v138.1. Inconsistent classifications were resolved by majority rule. For MAGs without 16S rRNA gene, the SILVA taxonomy was taken when both SILVA and GTDB predictions agreed (Fig. S8).

Diversity and core taxa analyses

The methods used for α- and β-diversity analyses are described in Additional file 1. Only genera and families were included that were present in ≥ 85% of a given set of analyzed samples and accounted for ≥ 1% of sequences in at least one sample. For macroalgae, these taxa were categorized as follows: (1) core phycosphere taxa (present on all four macroalgal species), (2) dominant phycosphere taxa (present on three macroalgal species), and (3) host-specific phycosphere taxa (present on one or two macroalgal species). Seawater and sediment core taxa were computed correspondingly.

Phylogenetic analyses and OTU-clustering of MAGs and draft genomes

Phylogenomic analyses of MAGs and draft genomes were executed within anvi'o v6.2 based on concatenated ribosomal protein sequences (Additional file 1). Maximum-likelihood trees were constructed in FastTree v2.1.5 [82] (default settings) and visualized in iTOL v6.5.6 [83]. Draft genome and MAG dereplication were performed using dRep v3.2.0 [84] based on a > 65% alignment and a genome-wide ANI threshold of 95% (-nc 0.65, -sa 0.95). The dRep program was also used to compare these draft genomes to 14,131 published species reference genomes from the GCM [34] and public database (https://www.ncbi.nlm.nih.gov/). Draft genomes exhibiting an ANI < 0.95 were designated as different species.

Functional annotations

Genes were predicted using Prodigal v2.6.3 [85] and annotated with Prokka. Additional annotations were performed using Diamond v0.9.24.125 [86] searches in 'verysensitive' mode against the UniRef100 [87] (as of September 2020) and COG [88] databases, as well as HMMER v3.1b2 [89] searches against the Pfam [90] database (as of September 2020). Further annotations were done by aligning genes to the EggNOG 5.0 [91] database using eggNOG-mapper v2.0.1 [92], peptidases were annotated using BLASTp searches against the MEROPS v9.13 database [93]. Biosynthetic gene clusters were identified using antiSMASH v5.0 [94] with default parameters. Signal peptides were predicted using SignalP v5.0 [95].

Prediction and annotation of PULs and CAZymes-rich gene clusters

Genes coding for carbohydrate-active enzymes (CAZymes) were annotated as described in Krüger et al. [37] using a combination of HMMER searches against the dbCAN v2.0.11 [96] database in conjunction with Diamond v0.9.24.125 searches against the CAZy database [97] as of July 2020. Genes coding for sulfatases, SusC- and SusD-like proteins were predicted using corresponding HMMER and TIGRFAM profiles (Additional file 1). PULs and other CAZyme-rich gene clusters were predicted as described in Francis et al. [98] with a sliding window of ten genes. In addition, we used dbCAN2 [96] to identify such clusters.

Protein phylogenies

Amino acid sequences were aligned using MAFFT v7.313 [99] with L-INS-I and curated manually. RaxML [100] was used to select the best fitting amino acid substitution model, which was subsequently used to generate maximum likelihood trees in FastTree v2.1.5 with default settings. Trees were visualized using iTOL v6.5.6.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-023-01559-1.

Additional file 1. Compilation of supplementary results, supplementary methods and of software tools used in this study.

Additional file 2: Compilation of supplementary figures. **Figure S1.** a) Diversities of macroalgae, seawater and sediment samples as assessed by Shannon and Simpson indices as well as Good's coverage of 16S rRNA ASVs. Statistical significance was assessed using a pairwise Wilcoxon test with Holm *p*-value adjustment for multiple comparisons (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$). b) Rarefaction curves of the top 200 ASVs for all six samples and all four seasons. **Figure S2.** The most abundant taxa as assessed by 16S rRNA gene amplicon data. **Figure S3.** Phycosphere composition as assessed by 16S rRNA gene amplicon data as a function of host species and season. **Figure S4.** Phylogenies and abundances of the 86 most abundant families as assessed by 16S rRNA gene amplicon sequencing. **Figure S5.** 16S rRNA phylogenetic tree reconstruction for 202 genera that were represented by at least three cultured strains. **Figure S6.** Compositional differences of strains depending on sample source and season. **Figure S7.** Numbers of colony forming units (CFUs) per gram of sample depending on habitat and season. **Figure S8.** Workflow for translating GTDB taxonomic classifications to SILVA taxonomic classifications. **Figure S9.** Proportions of genes within 965 metagenome-assembled genomes (MAGs) and 1,618 draft genomes (DGs) with EggNOG, COG (2020), Pfam, UniProtKB, and KEGG annotations, as well as the percentage of genes lacking any functional annotation. **Figure S10.** CAZymes in metagenome-assembled genomes (MAGs) and draft genomes (DGs) of different phyla. **Figure S11.** CAZymes versus sulfatase gene frequencies in prominent phyla and families as assessed in 1,294 metagenome-assembled genomes (MAGs) and 963 draft genomes (DGs) from all six sample sources. **Figure S12.** Categories of loci used to find putative PULs in this study. **Figure S13.** Histograms of the lengths of the four loci described in Fig. S12 in metagenome-assembled genomes (MAGs) and draft genomes (DGs). **Figure S14.** Tree of all 159 clusters derived from 3,769 PUL-associated SusC-like protein sequences from *Bacteroidota* metagenome-assembled genomes (MAGs) and draft genomes (DGs). **Figure S15.** Basic quality metrics of the 1,619 metagenome-assembled genomes (MAGs) and 965 draft genomes (DGs). Box-plots (A-E) show the minimum value, first quartile, median, third quartile and maximum value. **Figure S16.** Biosynthetic gene cluster (BGC) sizes in genomes from distinct phyla. **Figure S17.** Clustering of biosynthetic gene clusters (BGCs) according to sample type and phylogeny. **Figure S18.** Sizes of PULs and PUL-like loci in genomes from distinct *Bacteroidota* families (categories: hybrid *susCD*, single *susCD*, tandem repeat *susCD*, and tandem repeat plus hybrid *susCD* PULs).

Additional file 3: Description of supplementary tables. **Table S1.** Data associated with the 16S rRNA gene amplicon-based community profiling for all six sample sources analyzed in this study. Sequencing, assembly and binning statistics of the 23 metagenome datasets used in this study. These data include the time, season, geographical location, sample, environmental metadata for each sample and library information related to the amplicon sequencing. Furthermore included are summary analyses of the average relative abundances grouped by season and sample type at the genus and family levels, as well as statistical analyses of the proportions of core and dominant taxa in each sample. In addition, this file contains diversity indices, average relative abundances of domain, phylum, family, genus, OTU and ASV levels. **Table S2.** Data associated with the 16S rRNA gene-based community analyses of cultured bacterial strains, including information on sampling time, season, geographical location, source, culture conditions, 16S rRNA sequence information, new species attributes and taxonomic status information. Included are also summary analyses about average relative abundances at phylum, family, genus and OTU levels, as well as core taxa analyses results at the family and genus levels (matched to the 16S amplicon data). In addition, the file contains EZcloud and SILVA 138 sequence alignment results. **Table S3.** Summary data on the 1,619 MAGs and 965 draft genomes, including completeness, contamination, contig number, tRNA number, quality classification, size (Mbp), N50 value, species cluster ID in dRep, and the annotation results from GTDB SR202, EZcloud and SILVA 138 ordered according to their positions on the phylogenetic tree in Fig. 4. **Table S4.** Summary information about the four categories of PULs / PUL-like loci used in this study that were found with sliding window lengths from 1 and 10. The information includes: taxonomic affiliation, length (number of genes), number and type of comprised CAZyme genes, PUL composition (CAZyme genes, *tonB*, *susCD*, sulfatase genes), information on *susCD* genes in classical PULs and the density of CAZyme genes in each PUL. **Table S5.** Information on PULs from this study and published reference PULs, including descriptions of each PUL cluster in the SusC/D protein trees (single *susCD* PULs, hybrid *susCD* PULs, tandem-repeat *susCD* PULs, and tandem-repeat and hybrid *susCD* PULs). Also included is information about the source genome, the source genome type, its taxonomy and habitat as well as PUL ID, cluster number, number of CAZyme genes, composition (CAZymes gene, *susCD*, TonB and sulfatase genes) and genomes, possible substrate. For classical PULs, detailed information of the SusC/D protein tree is provided, including, gene ID, PUL ID, PUL type, PUL composition and potential substrates. **Table S6.** Details on the four categories of PULs and PUL-like loci used in this study in the 1,619 MAGs and 965 draft genomes, including gene composition. gene locus tags and gene annotations from multiple databases (KEGG, CAZy, EggNOG, COG, SignalP, MEROPS and Pfam). **Table S7.** Details on all BGCs predicted in the 1,619 MAGs and 965 draft genomes. This includes overall function predictions and gene function predictions according to KEGG, CAZy, EggNOG, COG, SignalP, MEROPS and Pfam searches. **Table S8.** Annotated putative PUL substrates based on dbCAN-PUL data (dbCAN-PUL is a database of experimentally characterized CAZyme gene clusters and their substrates), and substrate and enzyme cleavage information from the CAZy database (http://www.cazy.org/). These substrates represent automatically derived similarity-based bioinformatic predictions and are thus not as accurate as biochemically characterizations of PUL functions would be.

Acknowledgements

We acknowledge the POMPU consortium (Protegenomics of Marine Polysaccharide Utilization) funded by the German Research Foundation (FOR 2406) for general support. Our special thanks go to Xi Feng and Jin-Yu Zhang for helping in the process of cultivating bacteria. These data of published species reference genomes were produced by GCM 2.0 10K type strain sequencing project supported by WDCM in collaboration with the user community.

Authors' contributions

Study design: DCL, HT and ZJD. Sample collection and processing, isolation and cultivation of bacterial isolates: DCL and ZJD. Initial data analyses and curation, bioinformatic analyses as well as data visualization: DCL. Assistance in PCR, deposition of bacterial strains, bioinformatic analyses, and data visualization: FQW. Supporting input on ecology and bioinformatics: RIA and HT. Drafting of the manuscript: DCL and HT. All authors edited and approved the final manuscript.

Funding

This work was supported by Science & Technology Fundamental Resources Investigation Program (Grant No. 2019FY100700, 2022FY101100) and the National Natural Science Foundation of China (32070002). De-Chen Lu was furthermore supported by a scholarship granted by the China Scholarship Council (CSC) and by a follow-up stipend of the Max Planck Society.

Availability of data and materials

Sequences are available from the European Nucleotide Archive under accessions PRJEB51052 (16S rRNA tags, Table S1), PRJEB50838 (metagenomes and MAGs, Tables S1 and S3), and PRJEB57783 (genomes of cultured bacteria). All deposited strains (Table S2) are available from the Marine Culture Collection of China (MCCC) on request. The presented datasets except for metagenomes are also archived at Zenodo (https://doi.org/10.5281/zenodo.7556438).

Declarations

Ethics approval and consent to participate

Ethics approval was not required for the study.

Competing interests
The authors declare no competing interests.

Author details
[1]Marine College, Shandong University, No. 180, Wenhua Xilu, Weihai, Shandong Province 264209, China. [2]Max Planck Institute for Marine Microbiology, Celsiusstraße 1, Bremen 28359, Germany. [3]State Key Laboratory of Microbial Technology, Institute of Microbial Technology, Shandong University, Qingdao 266237, China.

Received: 11 April 2023 Accepted: 27 April 2023
Published online: 01 June 2023

References

1. Jard G, Marfaing H, Carrère H, Delgenes JP, Steyer JP, Dumas C. French Brittany macroalgae screening: composition and methane potential for potential alternative sources of energy and products. Bioresour Technol. 2013;144:492–8.
2. Bengtson S, Sallstedt T, Belivanova V, Whitehouse M. Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. PLoS Biol. 2017;15:e2000735.
3. Lemay MA, Chen MY, Mazel F, Hind KR, Starko S, Keeling PJ, et al. Morphological complexity affects the diversity of marine microbiomes. ISME J. 2020;15:1372–86.
4. 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;5:590–600.
5. Wayne B, Ralph M. Chemotactic and growth responses of marine bacteria to algal extracellular products. J Chem Inf Model. 2013;53:1689–99.
6. Marshall K, Joint I, Callow ME, Callow JA. Effect of marine bacterial isolates on the growth and morphology of axenic plantlets of the green alga *Ulva linza*. Microb Ecol. 2006;52:302–10.
7. Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. Nature. 2005;438:90–3.
8. Dittami SM, Duboscq-Bidot L, Perennou M, Gobet A, Corre E, Boyen C, et al. Host-microbe interactions as a driver of acclimation to salinity gradients in brown algal cultures. ISME J. 2016;10:51–63.
9. Joint I, Tait K, Wheeler G. Cross-kingdom signalling: exploitation of bacterial quorum sensing molecules by the green seaweed *Ulva*. Philos Trans R Soc B Biol Sci. 2007;362:1223–33.
10 Weigel BL, Miranda KK, Fogarty EC, Watson AR, Pfister CA. Functional insights into the *Kelp* microbiome from metagenome-assembled genomes. mSystems. 2022;7:e01422-21.
11. Case RJ, Longford SR, Campbell AH, Low A, Tujula N, Steinberg PD, et al. Temperature induced bacterial virulence and bleaching disease in a chemically defended marine macroalga. Environ Microbiol. 2011;13:529–37.
12. Martin M, Barbeyron T, Martin R, Portetelle D, Michel G, Vandenbol M. The cultivable surface microbiota of the brown alga *Ascophyllum nodosum* is enriched in macroalgal-polysaccharide-degrading bacteria. Front Microbiol. 2015;6:1–14.
13. Schiel DR, Lilley SA. Gradients of disturbance to an algal canopy and the modification of an intertidal community. Mar Ecol Prog Ser. 2007;339:1–11.
14 Cherry P, O'hara C, Magee PJ, Mcsorley EM, Allsopp PJ. Risks and benefits of consuming edible seaweeds. Nutr Rev. 2019;77:307–29.
15. Ismail MM, Alotaibi BS, EL-Sheekh MM. Therapeutic uses of red macroalgae. Molecules. 2020;25:1–14.
16. Sudhakar K, Mamat R, Samykano M, Azmi WH, Ishak WFW, Yusaf T. An overview of marine macroalgae as bioresource. Renew Sustain Energy Rev. 2018;91:165–79.
17. Krause-Jensen D, Duarte CM. Substantial role of macroalgae in marine carbon sequestration. Nat Geosci. 2016;9:737–42.
18. Buck-Wiese H, Andskog MA, Nguyen NP, Bligh M, Asmala E, Vidal-Melgosa S, et al. Fucoid brown algae inject fucoidan carbon into the ocean. Proc Natl Acad Sci U S A. 2023;120:e2210561119.
19. Brunet M, Le Duff N, Barbeyron T, Thomas F. Consuming fresh macroalgae induces specific catabolic pathways, stress reactions and Type IX secretion in marine flavobacterial pioneer degraders. ISME J. 2022;16:2027–39.
20. Hengst MB, Andrade S, González B, Correa JA. Changes in epiphytic bacterial communities of intertidal seaweeds modulated by host, temporality, and copper enrichment. Microb Ecol. 2010;60:282–90.
21. Lachnit T, Meske D, Wahl M, Harder T, Schmitz R. Epibacterial community patterns on marine macroalgae are host-specific but temporally variable. Environ Microbiol. 2011;13:655–65.
22. Tujula NA, Crocetti GR, Burke C, Thomas T, Holmström C, Kjelleberg S. Variability and abundance of the epiphytic bacterial community associated with a green marine *Ulvacean* alga. ISME J. 2010;4:301–11.
23. Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T. Bacterial community assembly based on functional genes rather than species. Proc Natl Acad Sci U S A. 2011;108:14288–93.
24. Egan S, Harder T, Burke C, Steinberg P, Kjelleberg S, Thomas T. The seaweed holobiont: understanding seaweed-bacteria interactions. FEMS Microbiol Rev. 2013;37:462–76.
25. Skriptsova AV, Shevchenko NM, Tarbeeva DV, Zvyagintseva TN. Comparative study of polysaccharides from reproductive and sterile tissues of five brown seaweeds. Mar Biotechnol. 2012;14:304–11.
26. 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.
27. Robic A, Gaillard C, Sassi JF, Leral Y, Lahaye M. Ultrastructure of *Ulvan*: a polysaccharide from green seaweeds. Biopolymers. 2009;91:652–64.
28. Thomas F, Barbeyron T, Tonon T, Génicot S, Czjzek M, Michel G. Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine *Flavobacteriia* to their independent transfers to marine *Proteobacteria* and human gut *Bacteroides*. Environ Microbiol. 2012;14:2379–94.
29. Kabisch A, Otto A, Ko S, Schu M, Teeling H, Amann RI, et al. Functional characterization of polysaccharide utilization loci in the marine *Bacteroidetes* '*Gramella forsetii*' KT0803. ISME J. 2014;8:1492–502.
30. 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.
31. Ficko-Blean E, Préchoux A, Thomas F, Rochat T, Larocque R, Zhu Y, et al. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. Nat Commun. 2017;8:1685.
32. Paoli L, Ruscheweyh H-J, Forneris CC, Hubrich F, Kautsar S, Bhushan A, et al. Biosynthetic potential of the global ocean microbiome. Nature. 2022;607:111–8.
33. Zhang W, Ding W, Li YX, Tam C, Bougouffa S, Wang R, et al. Marine biofilms constitute a bank of hidden microbial diversity and functional potential. Nat Commun. 2019;10:1–10.
34. Wu L, McCluskey K, Desmeth P, Liu S, Hideaki S, Yin Y, et al. The global catalogue of microorganisms 10K type strain sequencing project: closing the genomic gaps for the validly published prokaryotic and fungi species. Gigascience. 2018;7:1–4.
35. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;35:725–31.
36. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. Nature. 2019;568:499–504.
37. Krüger K, Chafee M, Ben Francis T, Glavina T, Becher D, Schweder T, et al. In marine *Bacteroidetes* the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. ISME J. 2019;13:2800–16.
38. Barbeyron T, Brillet-Guéguen L, Carré W, Carrière C, Caron C, Czjzek M, et al. Matching the diversity of sulfated biomolecules: creation of a classification database for sulfatases reflecting their substrate specificity. PLoS One. 2016;11:1–33.
39. Lapébie P, Lombard V, Drula E, Terrapon N, Henrissat B. *Bacteroidetes* use thousands of enzyme combinations to break down glycans. Nat Commun. 2019;10:2043.
40. Kappelmann L, Krüger K, Harder J, Markert S, Unfried F, Becher D, et al. Polysaccharide utilization loci of North Sea *Flavobacteriia* as

basis for using SusC/D-protein expression for predicting major phytoplankton glycans. ISME J. 2019;13:76–91.
41. Flemming HC, Wuertz S. Bacteria and archaea on Earth and their abundance in biofilms. Nat Rev Microbiol. 2019;17:247–60.
42. Gavriilidou A, Gutleben J, Versluis D, Forgiarini F, Van Passel MWJ, Ingham CJ, et al. Comparative genomic analysis of *Flavobacteriaceae*: insights into carbohydrate metabolism, gliding motility and secondary metabolite biosynthesis. BMC Genomics. 2020;21:1–21.
43. Zozaya-Valdés E, Roth-Schulze AJ, Egan S, Thomas T. Microbial community function in the bleaching disease of the marine macroalgae *Delisea pulchra*. Environ Microbiol. 2017;19:3012–24.
44. Zan J, Li Z, Diarey Tianero M, Davis J, Hill RT, Donia MS. A microbial factory for defensive kahalalides in a tripartite marine symbiosis. Science. 2019;364:6732.
45. Matsuo Y, Suzuki M, Kasai H, Shizuri Y, Harayama S. Isolation and phylogenetic characterization of bacteria capable of inducing differentiation in the green alga Monostroma oxyspermum. Environ Microbiol. 2003;5:25–35.
46. Longford SR, Tujula NA, Crocetti GR, Holmes AJ, Holmström C, Kjelleberg S, et al. Comparisons of diversity of bacterial communities associated with three sessile marine eukaryotes. Aquat Microb Ecol. 2007;48:217–29.
47. Kim NK, Oh S, Liu WT. Enrichment and characterization of microbial consortia degrading soluble microbial products discharged from anaerobic methanogenic bioreactors. Water Res. 2016;90:395–404.
48. Wiegand S, Jogler M, Boedeker C, Pinto D, Vollmers J, Rivas-Marín E, et al. Cultivation and functional characterization of 79 planctomycetes uncovers their unique biology. Nat Microbiol. 2020;5:126–40.
49. Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, et al. *Verrucomicrobia* use hundreds of enzymes to digest the algal polysaccharide fucoidan. Nat Microbiol. 2020;5:1026–39.
50. Chiang E, Schmidt ML, Berry MA, Biddanda BA, Burtner A, Johengen TH, et al. *Verrucomicrobia* are prevalent in north-temperate freshwater lakes and display class-level preferences between lake habitats. PLoS One. 2018;13:1–20.
51. Vidal-Melgosa S, Sichert A, Ben Francis T, Bartosik D, Niggemann J, Wichels A, et al. Diatom fucan polysaccharide precipitates carbon during algal blooms. Nat Commun. 2021;12:1–13.
52. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc Natl Acad Sci U S A. 2006;103:12115–20.
53. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol. 2012;10:323–35.
54. Song W, Wemheuer B, Steinberg PD, Marzinelli EM, Thomas T. Contribution of horizontal gene transfer to the functionality of microbial biofilm on a macroalgae. ISME J. 2021;15:807–17.
55. Pomin VH, Mourão PAS. Structure, biology, evolution, and medical importance of sulfated fucans and galactans. Glycobiology. 2008;18:1016–27.
56. Goecke F, Labes A, Wiese J, Imhoff JF. Phylogenetic analysis and antibiotic activity of bacteria isolated from the surface of two co-occurring macroalgae from the Baltic Sea. Eur J Phycol. 2013;48:47–60.
57. Rao D, Webb JS, Holmström C, Case R, Low A, Steinberg P, et al. Low densities of epiphytic bacteria from the marine alga *Ulva australis* inhibit settlement of fouling organisms. Appl Environ Microbiol. 2007;73:7844–52.
58. Grueneberg J, Engelen AH, Costa R, Wichard T. Macroalgal morphogenesis induced by waterborne compounds and bacteria in coastal seawater. PLoS One. 2016;11:1–22.
59. Ji YY, Zhang B, Zhang P, Chen LC, Si YW, Wan XY, Li C, Wang RH, Tian Y, Zhang Z, Tian CF. Rhizobial migration toward roots mediated by FadL-ExoFQP modulation of extracellular long-chain AHLs. ISME J. 2023;17(3):417–31.
60. Kjelleberg S, Steinberg P, Givskov M, Gram L, Manefield M, De Nys R. Do marine natural products interfere with prokaryotic AHL regulatory systems? Aquat Microb Ecol. 1997;13:85–93.
61. Hughes DT, Terekhova DA, Liou L, Hovde CJ, Sahl JW, Patankar AV, et al. Chemical sensing in mammalian host-bacterial commensal associations. Proc Natl Acad Sci U S A. 2010;107:9831–6.
62. Romero M, Martin-Cuadrado AB, Roca-Rivada A, Cabello AM, Otero A. Quorum quenching in cultivable bacteria from dense marine coastal microbial communities. FEMS Microbiol Ecol. 2011;75:205–17.
63. Yamada Y, Kuzuyama T, Komatsu M, Shin-ya K, Omura S, Cane DE, et al. Terpene synthases are widely distributed in bacteria. Proc Natl Acad Sci U S A. 2015;112:857–62.
64. Mu DS, Liang QY, Wang XM, Lu DC, Shi MJ, Chen GJ, et al. Metatranscriptomic and comparative genomic insights into resuscitation mechanisms during enrichment culturing. Microbiome. 2018;6:1–15.
65. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol. 2017;67:1613–7.
66. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823–9.
67. Wu L, Ma J. The global catalogue of microorganisms (GCM) 10K type strain sequencing project: providing services to taxonomists for standard genome sequencing and annotation. Int J Syst Evol Microbiol. 2019;69:895–8.
68. 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;19:455–77.
69. Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. PLoS One. 2014;9:e105592.
70. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;18:6–9.
71. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.
72. 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;41:590–6.
73. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674–6.
74. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platformfor 'omics data. PeerJ. 2015;3:1–29.
75. 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;11:1144–6.
76. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32:605–7.
77. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:1–15.
78. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018;3:836–43.
79. 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;25:1043–55.
80. Prokka ST. Rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
81. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics. 2020;36:1925–7.
82. Price MN, Dehal PS, Arkin AP. FastTree 2-approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5:e9490.
83. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242–5.
84. Olm MR, Brown CT, Brooks B, Banfield JF. DRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11:2864–8.
85. Hyatt D, Chen GL, LoCascio PF, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:1–11.
86. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2014;12:59–60.

87. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics. 2007;23:1282–8.
88. Galperin MY, Wolf YI, Makarova KS, Alvarez RV, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. 2021;49:D274–81. Oxford University Press.
89. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol. 2011;7:e1002195.
90. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279–85.
91. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:309–14.
92. 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;34:2115–22.
93. Rawlings ND, Barrett AJ, Finn R. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 2016;44:D343–50.
94 Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 2019;47:W81-7. Oxford University Press.
95. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 2019;37:420–3.
96. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. DbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46:W95-101.
97. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:490–5.
98. Ben Francis T, Bartosik D, Sura T, Sichert A, Hehemann JH, Markert S, et al. Changing expression patterns of TonB-dependent transporters suggest shifts in polysaccharide consumption over the course of a spring phytoplankton bloom. ISME J. 2021;15:2336–50.
99. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
100. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

Publisher's Note

Supplementary Figures for this manuscript includes the following:

Supplementary Fig. S1 to S18.

Supplementary Figures

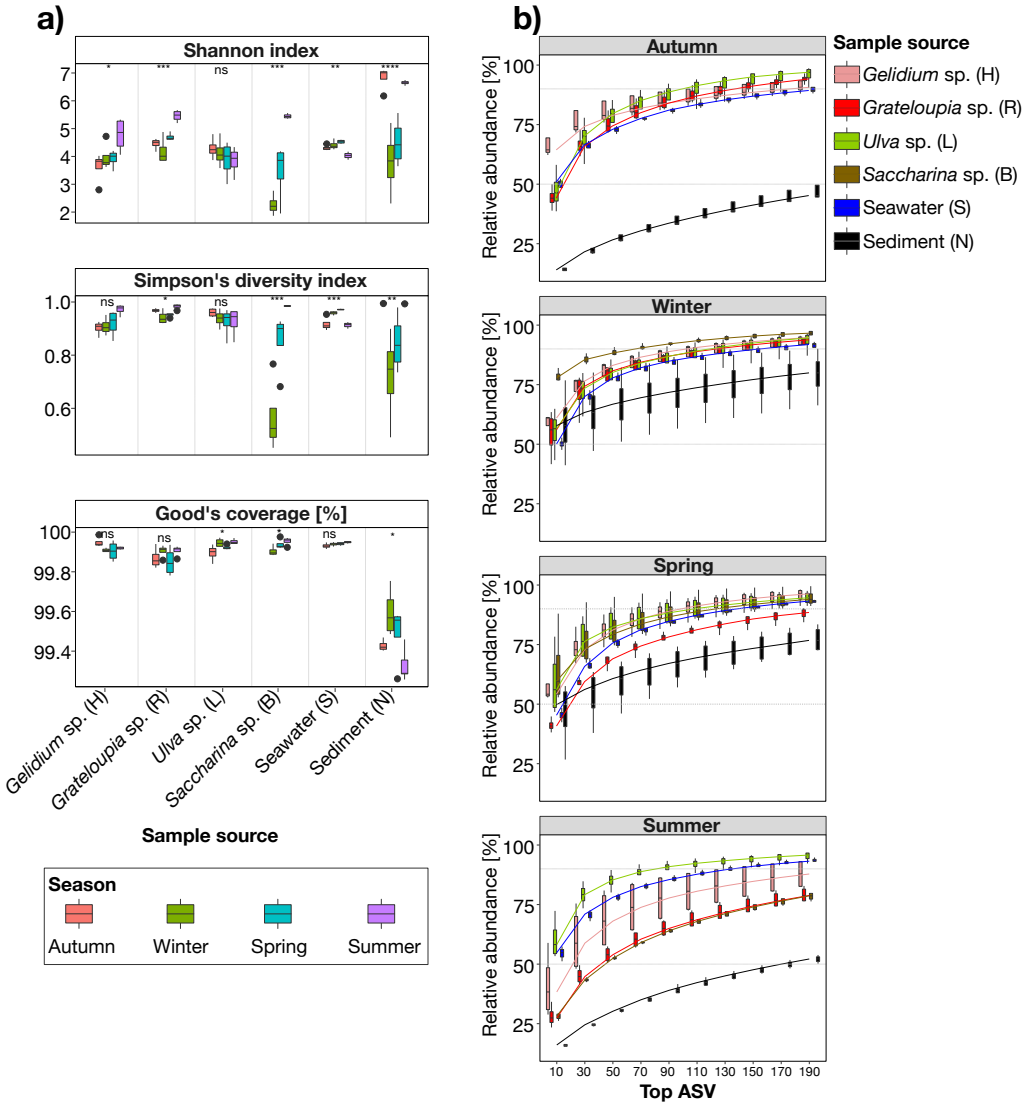


Figure S1. a) Diversities of macroalgae, seawater and sediment samples as assessed by Shannon and Simpson indices as well as Good's coverage of 16S rRNA ASVs. Statistical significance was assessed using a pairwise Wilcoxon test with Holm p-value adjustment for multiple comparisons (*, p <0.05; **, p <0.01; ***, p <0.001). **b)** Rarefaction curves of the top 200 ASVs for all six samples and all four seasons.

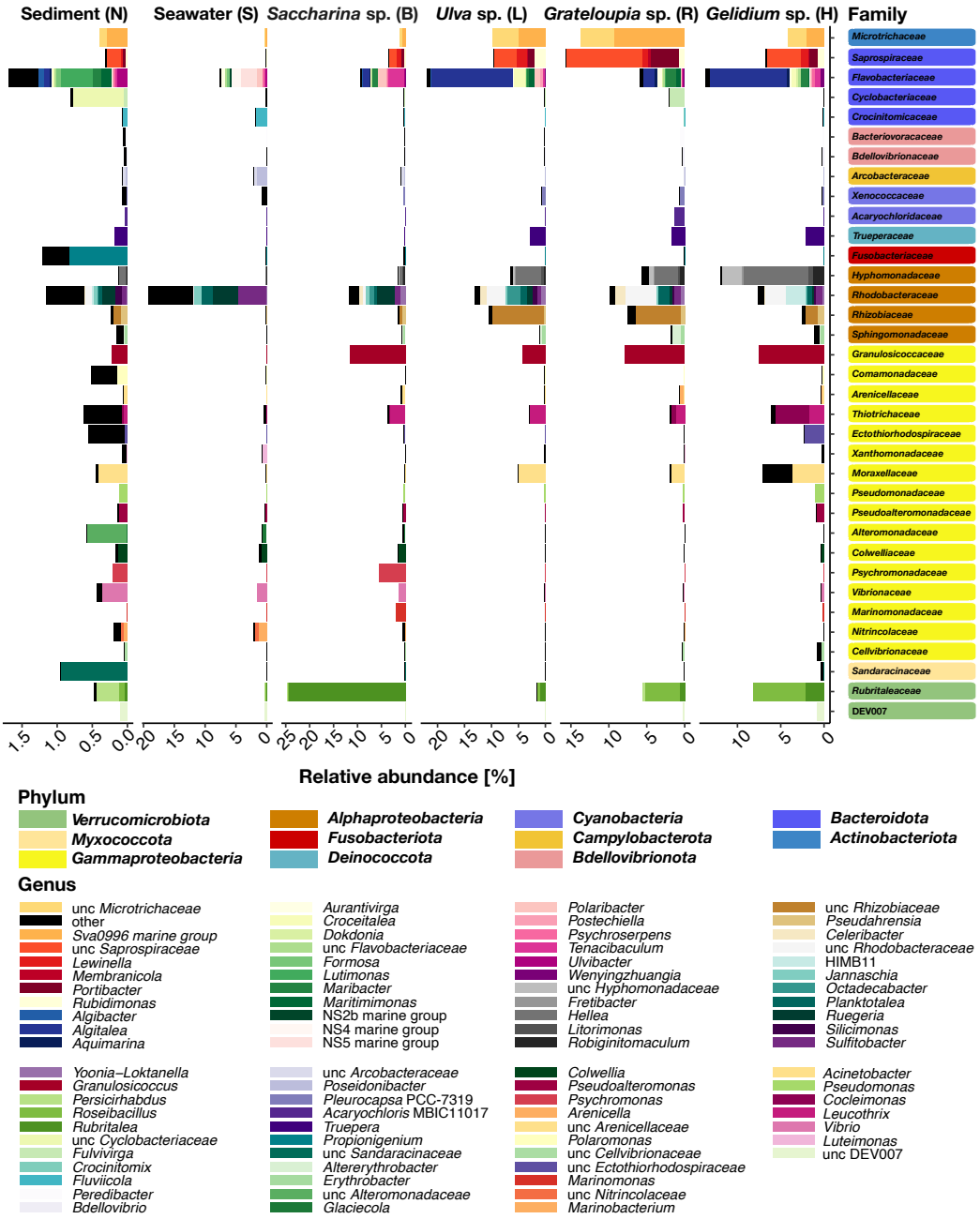


Figure S2. The most abundant taxa as assessed by 16S rRNA gene amplicon data. Bars (left) represent median relative abundances of the most prominent genera. Corresponding families (right) are color-coded according to their respective phyla.

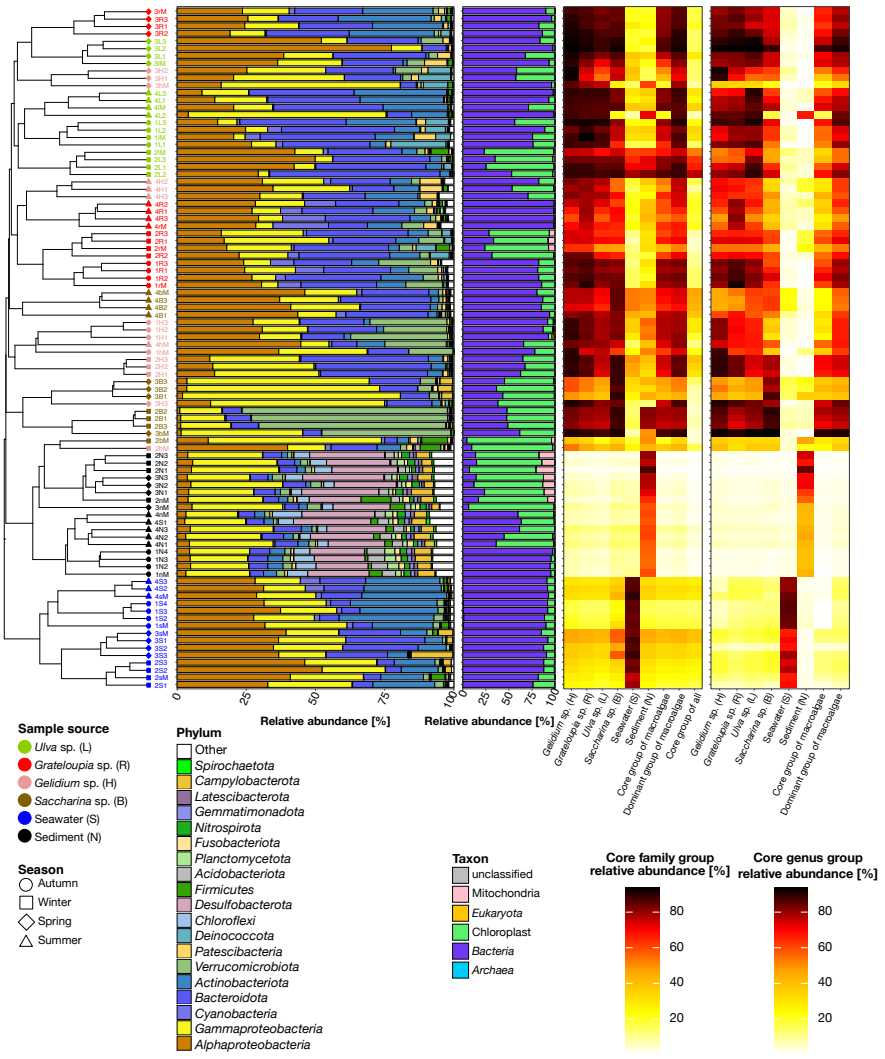


Figure S3. Phycosphere composition as assessed by 16S rRNA gene amplicon data as a function of host species and season. From left to right: (i) Clustering with colors representing sample sources and shapes representing seasons (UPGMA clustering of weighted UniFrac distances). Names consist of the season (1-4), macroalgal species (B=*Saccharina* sp., L=*Ulva* sp., H=*Gelidium* sp., R=*Grateloupia* sp., S=seawater, N=sediment), plus a sample replicate identifier (the numbers 1-4 represent distinct samples, and M represents the same sample that was used for metagenomics). (ii) Relative abundances of bacterial phyla. For *Proteobacteria* the most prominent proteobacterial classes *Alpha-* and *Gammaproteobacteria* are shown. (iii) Relative abundances of top-level taxa. (iv) Heatmaps of habitat-specific core taxa at the family and genus levels.

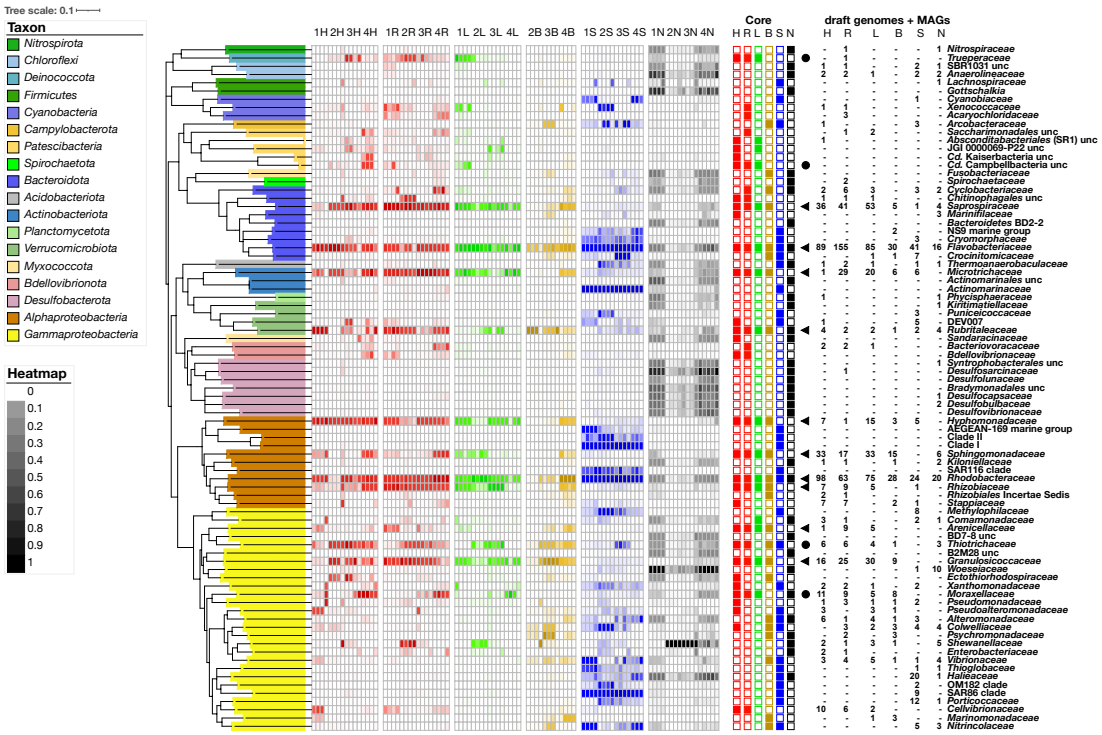


Figure S4. Phylogenies and abundances of the 86 most abundant families as assessed by 16S rRNA gene amplicon sequencing. Phylogeny was computed using RAxML with 1,000 bootstrap replicates. Heatmap values were obtained by converting relative abundances to Max-Min values ranging from 0-1. Samples are denoted by an initial capital letter specifying the sample (H=*Grateloupia* sp., R=*Gelidium* sp., L=*Ulva* sp., B=*Saccharina* sp., S=seawater, N=sediment), followed by a number representing season (1=autumn, 2=winter, 3=spring, 4=summer). Solid squares to the right of the heatmaps represent core families of corresponding samples. Circles and triangles represent overall core and dominant phycosphere families. The following six numerical columns represent the corresponding numbers of genomes and MAGs that were obtained.

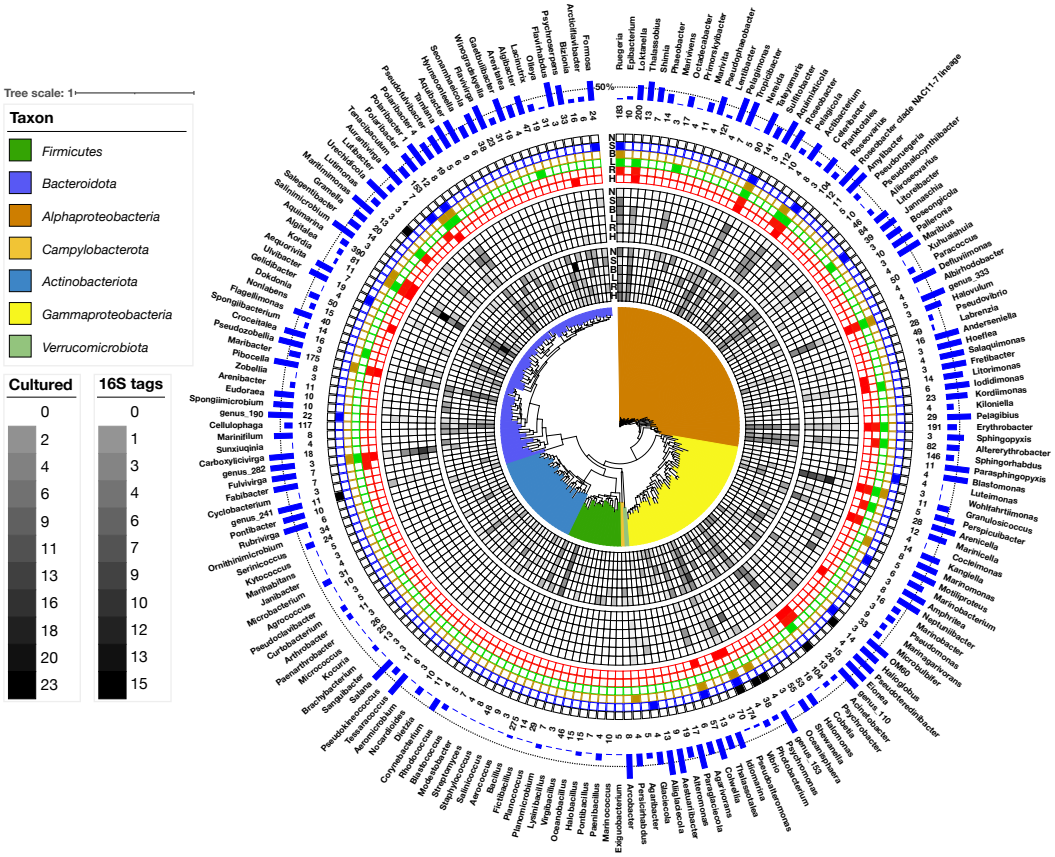


Figure S5. 16S rRNA phylogenetic tree reconstruction for 202 genera that were represented by at least three cultured strains. The tree was calculated using RAxML with 1,000 bootstrap replicates. Representative full lengths 16S rRNA gene sequences for each genus were selected randomly from corresponding cultured strains. From inside to outside: (i) heatmap representing relative abundances of cultured genera in all samples (H=*Gelidium* sp., R=*Grateloupia* sp., L=*Ulva* sp., B=*Saccharina* sp., S=seawater, N=sediment); (ii) heatmap representing abundances derived from 16S rRNA amplicon sequencing; (iii) indication of core community members by solid squares for the four algae (red, green, brown), seawater (blue) and sediments (black); (iv) numbers of obtained cultured strains; (v) bars representing the percentage of new species.

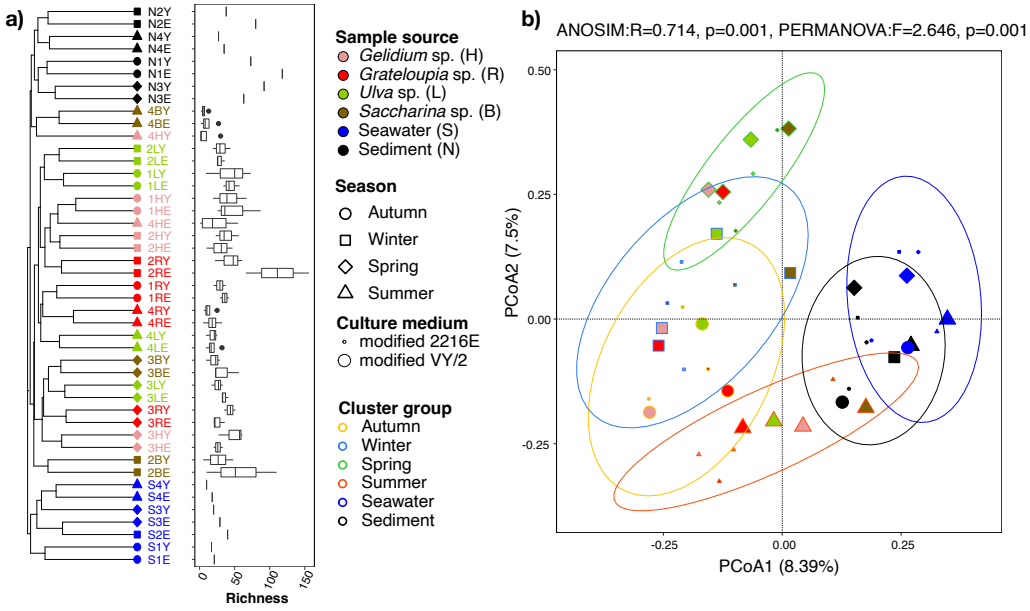


Figure S6. Compositional differences of strains depending on sample source and season. a) (left) samples were grouped by weighted UniFrac distances using Ward linkage (dendrogram) between sample sources, (right) corresponding richness (showing median and inter-quartile ranges, OTU level) for each group. **b)** Principal coordinate analysis plots of Bray-Curtis similarities of samples and seasons calculated from unweighted UniFrac distances for macroalgal samples (n=52), surrounding seawater (n = 7), and surrounding sediment (n = 8). Colors correspond to sample sources and shapes to seasons. Sizes represent the two media that were used. Results of PERMANOVA and ANOSIM tests for significance between groups are shown for each plot.

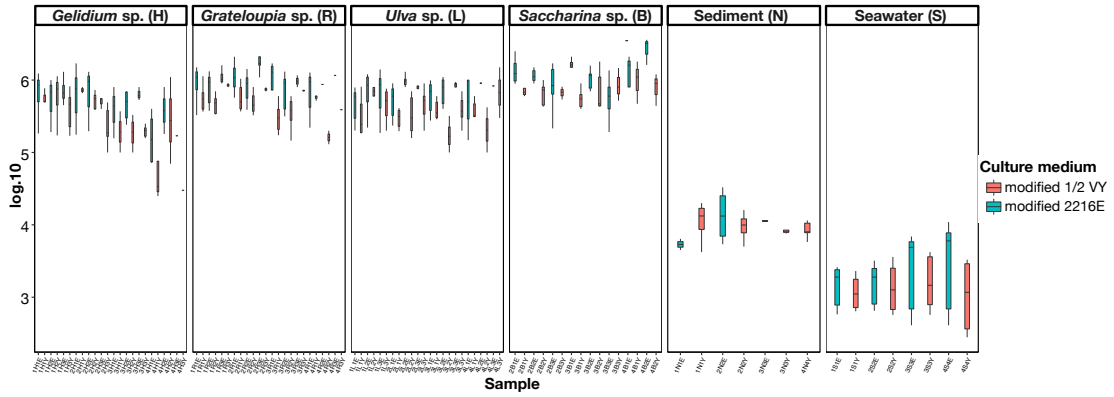


Figure S7. Numbers of colony forming units (CFUs) per gram of sample depending on habitat and season. Depicted are numbers obtained with both media, modified marine modified VY/2 (salmon) and 2216E (turquoise) medium for samples of *Gelidium* sp. (red algae), *Grateloupia* sp. (red algae), *Ulva* sp. (green algae), *Saccharina* sp. (brown algae), sediment (5 meters underwater), and seawater (-0.1 to -0.5 m water depth). The average CFU numbers obtained from macroalgae ranged from 5.5×10^5 CFU g^{-1} (modified VY/2 medium) to 5.8×10^5 CFU g^{-1} (modified 2216E medium). The average CFU numbers obtained from seawater and sediment samples were lower by about three and two orders of magnitude, respectively.

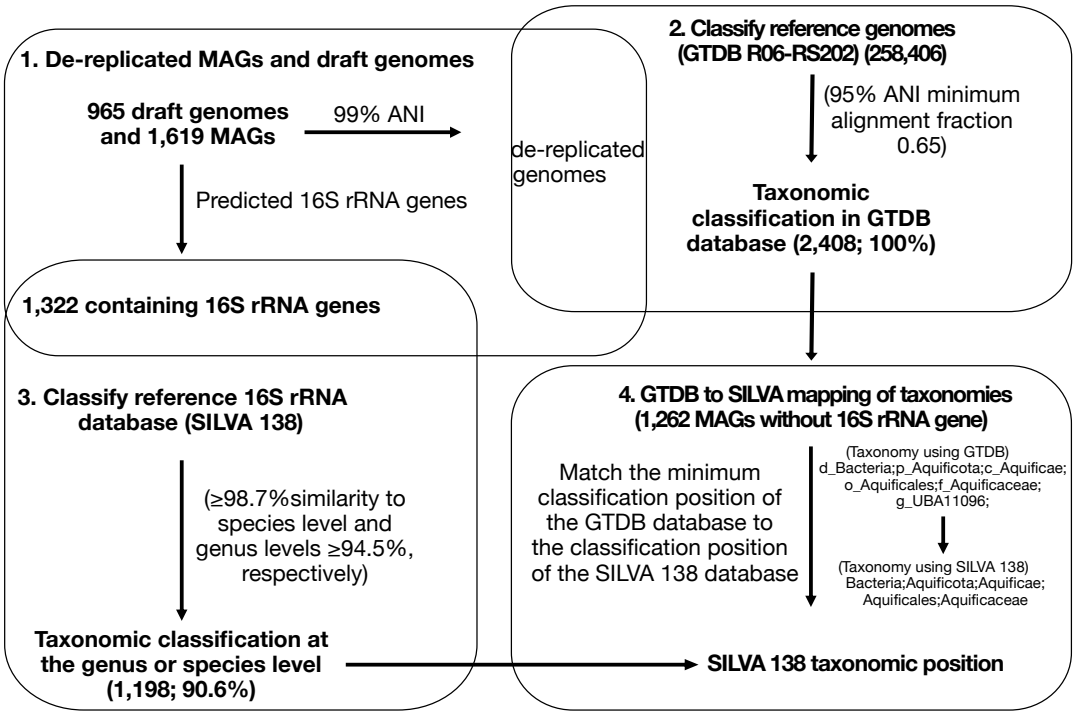


Figure S8. Workflow for translating GTDB taxonomic classifications to SILVA taxonomic classifications.

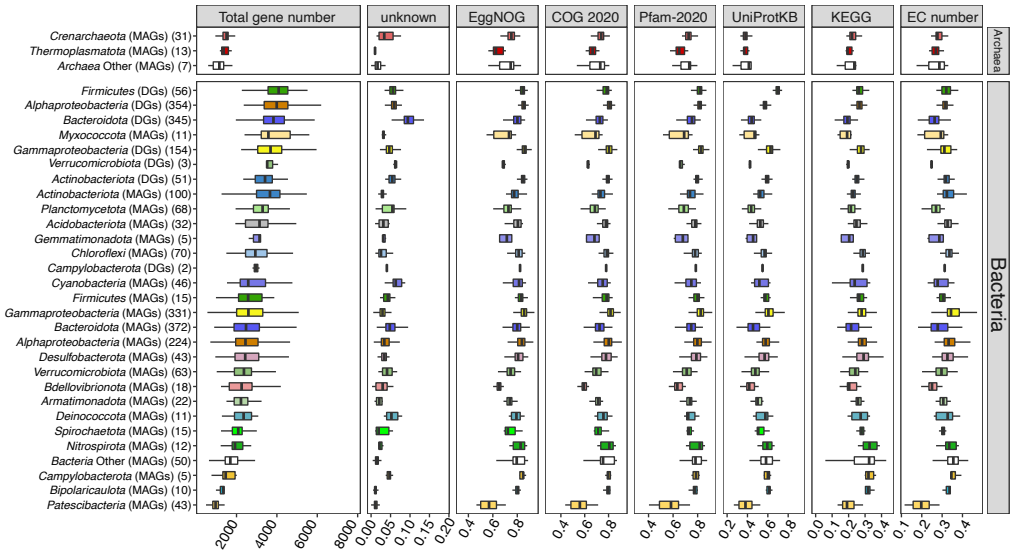


Figure S9. Proportions of genes within 965 MAGs and 1,618 draft genomes (DGs) with EggNOG, COG (2020), Pfam, UniProtKB, and KEGG annotations, as well as the percentage of genes lacking any functional annotation.

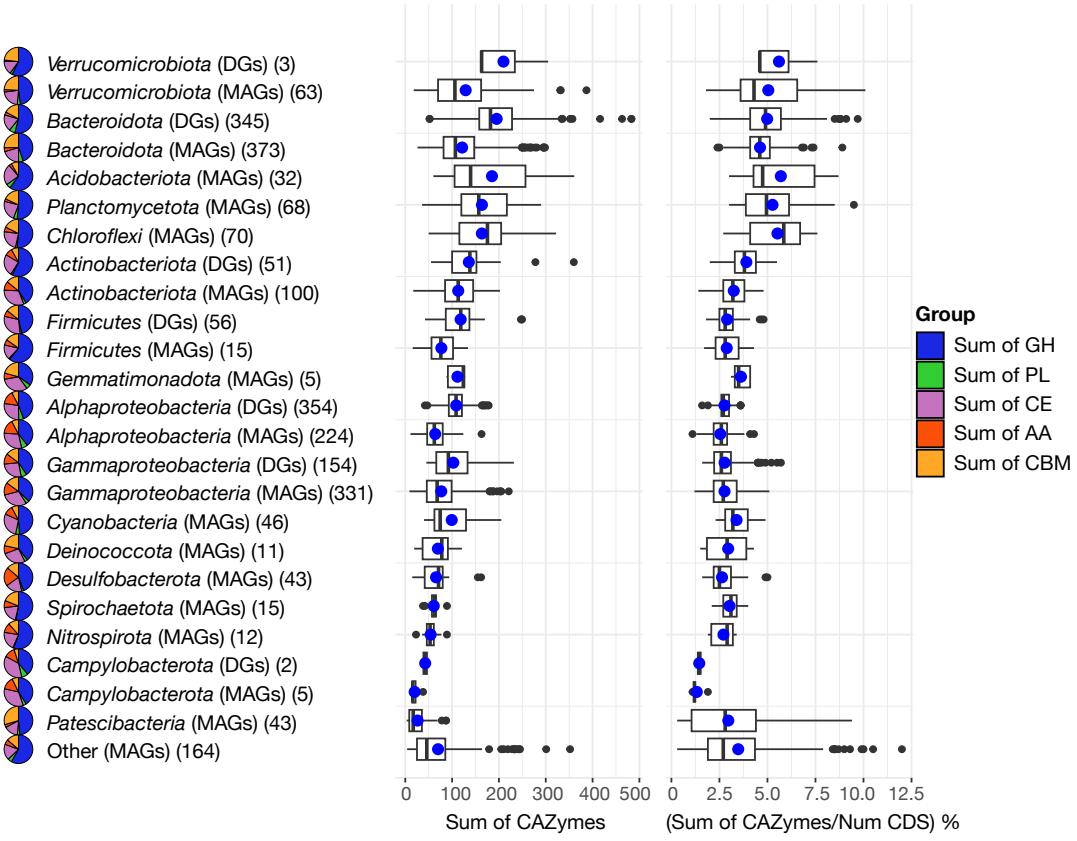


Figure S10. CAZymes in MAGs and draft genomes (DGs) of different phyla. Pie charts represent relative proportions of CAZyme classes except for GTs (glycosyltransferases) - GHs (glycoside hydrolases), PLs (polysaccharide lyases), CEs (carbohydrate esterases), AAs (auxiliary activities), and CBMs (carbohydrate-binding modules). The numbers of MAGs and DGs are indicated in parentheses. Box plots represent distributions of absolute (left) and relative (right) CAZyme numbers. Averages are represented as blue dots.

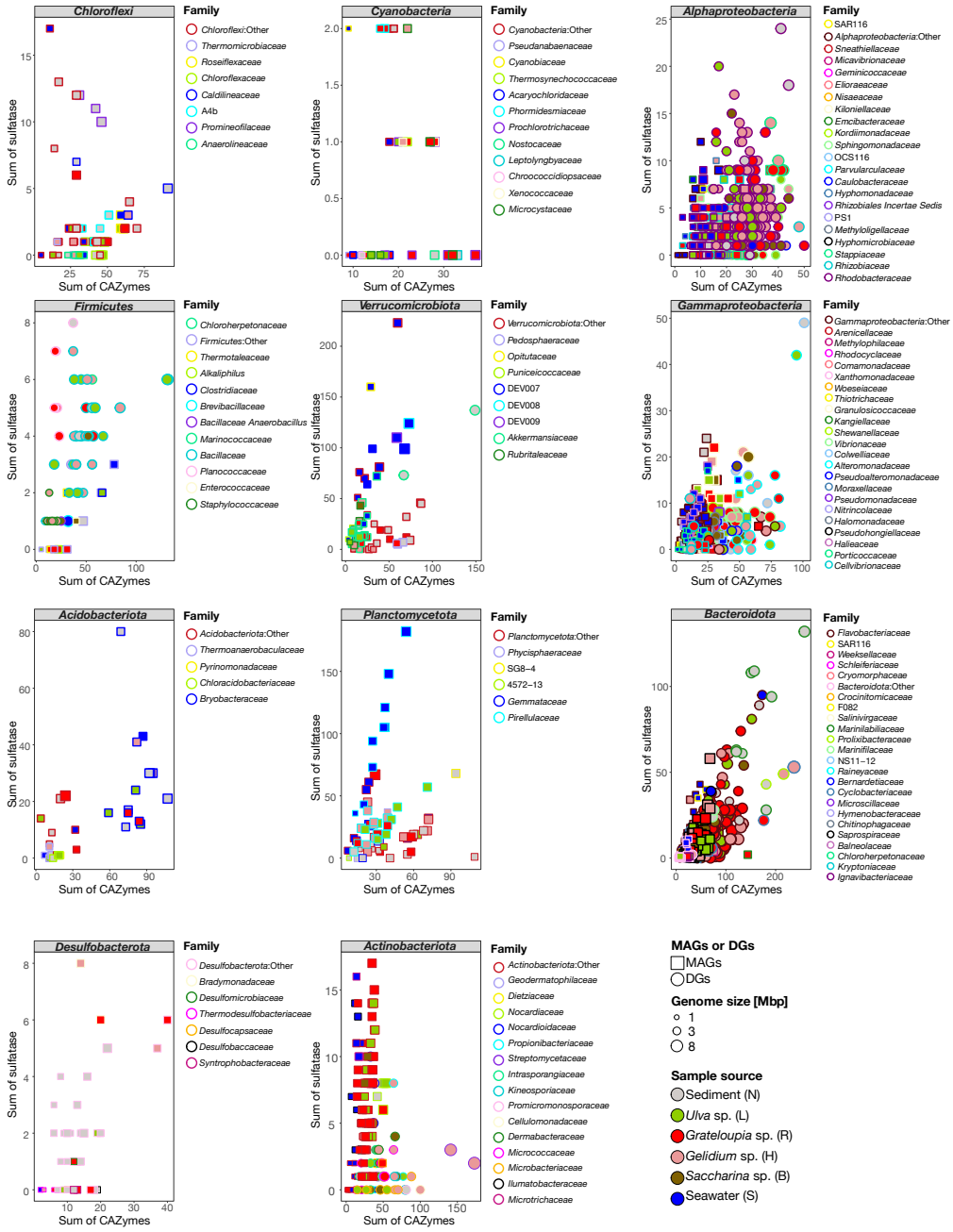
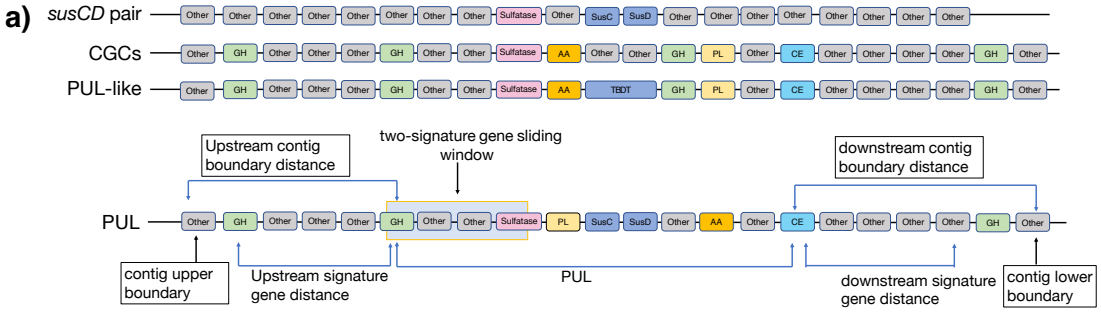
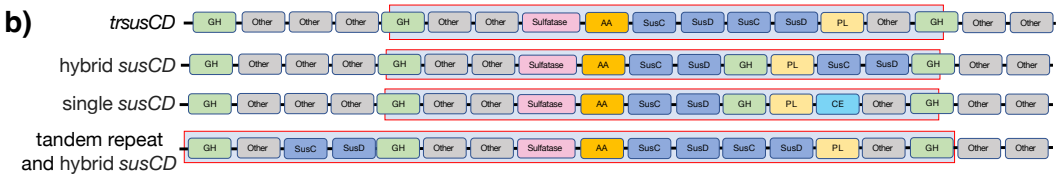


Figure S11. **CAZymes versus sulfatase gene frequencies in prominent phyla and families as assessed in 1,294 MAGs and 963 draft genomes (DGs) from all six sample sources.** MAGs are represented by squares and DGs by circles, with border colors representing families and fill colors representing sample types. Circle sizes correspond to genome sizes. Detailed information is provided in Table S3 in Additional file 3.



Potential PULs were extracted by finding all loci, where at least three predicted PUL genes (encoding sulfatases, CAZymes, SusC- or SusD-like proteins, TonB-dependent transporters (TBDTs) were less than a settable distance signal genes apart (in the example above, it is 2, and 10 is used in this study).
1. *susCD* pair: *susCD* loci without known CAZymes;
2. CAZyme gene clusters (CGCs): CAZyme gene clusters without *susCD*-like gene pairs;
3. PUL-like: CAZyme gene clusters with TBDTs;
4. PUL: CAZyme gene clusters with *susCD*-like gene pairs or anyone of *susCD*-like gene pairs.



Four categories (hybrid *susCD*, single *susCD*, tandem repeat *susCD* and tandem repeat plus hybrid *susCD* PULs) based on the distance between different PULs.
1. *trsusCD* (tandem repeat susC/D): PULs containing two *susCD* gene pairs in tandem;
2. hybrid *susCD*: PULs containing at least two non-tandem *susCD* gene pairs;
3. single *susCD*: PULs containing only one *susCD* gene pairs;
4. Tandem repeat and hybrid *susCD*: PULs containing two *susCD* gene pairs in tandem and at least one non-tandem *susCD* gene pairs.

Figure S12. Categories of loci used to find putative PULs in this study. a) locus with only a *susCD* gene pair, CAZyme-rich gene cluster, PUL-like locus, conventional PUL. **b)** Sub-classification of PULs according to *susCD* gene arrangements.

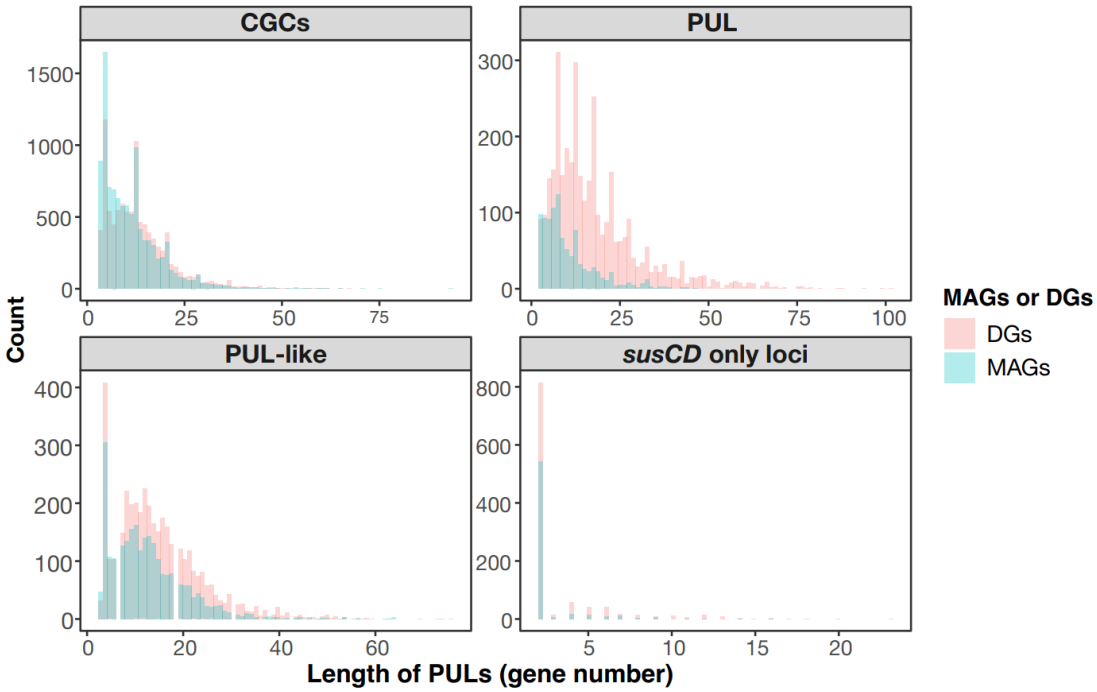


Figure S13. Histograms of the lengths of the four loci described in Fig. S12 in MAGs and draft genomes (DGs).

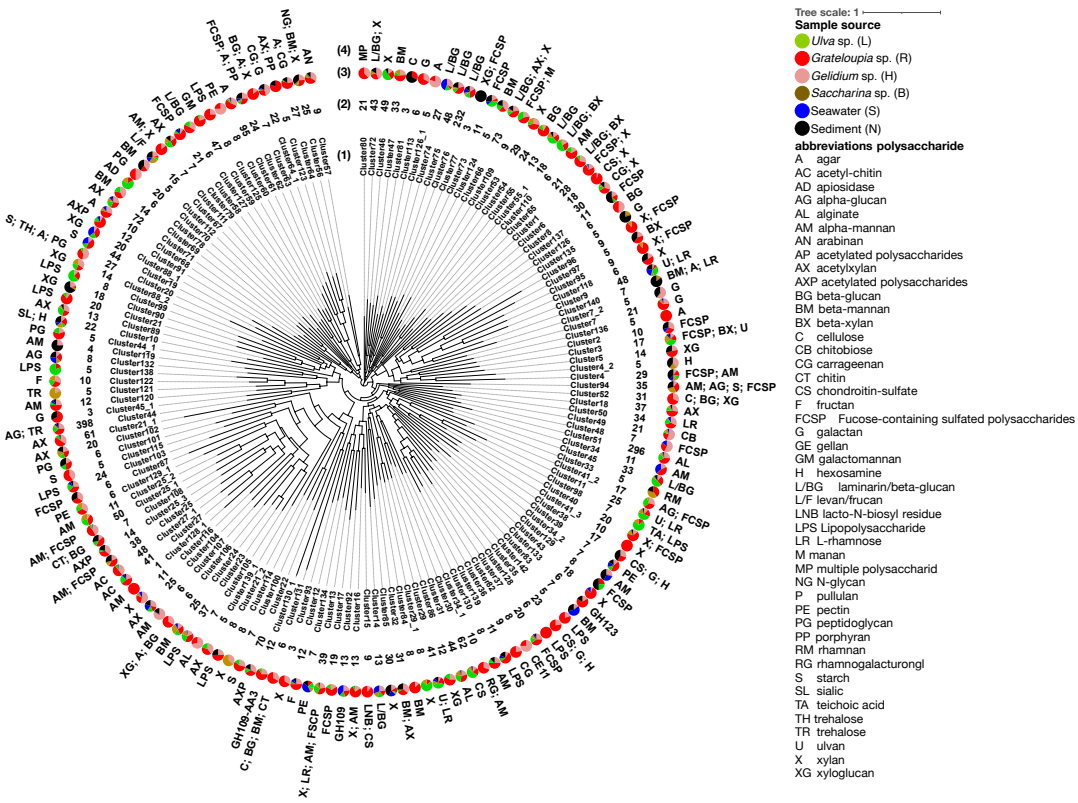


Figure S14. Tree of all 159 clusters derived from 3,769 PUL-associated SusC-like protein sequences from *Bacteroidota* MAGs and draft genomes (inside to outside): (i) cluster identifier, (ii) the number of SusC-like proteins, (iii) pie charts of corresponding samples sources, (iv) substrate predictions (short form).

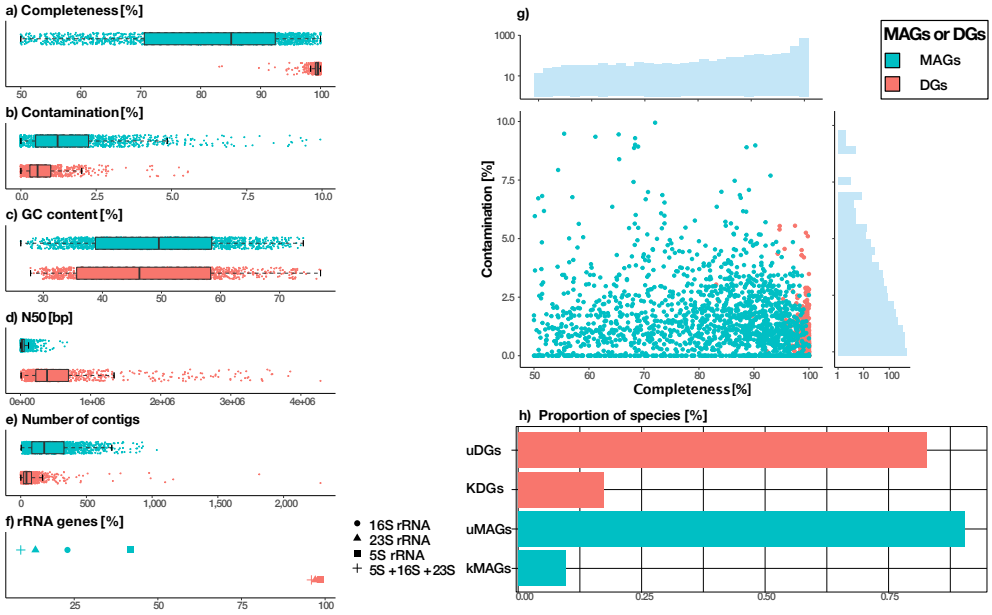


Figure S15. Basic quality metrics of the 1,619 MAGs and 965 draft genomes (DGs). Box-plots (A-E) show the minimum value, first quartile, median, third quartile and maximum value. Of the in total 2,584 genomes, 1,506 had >90% completeness and <5% contamination, 1,023 of which contained complete rRNA operons plus at least 18 of the standard tRNAs. Details are provided in Table S3 in Additional file 3.

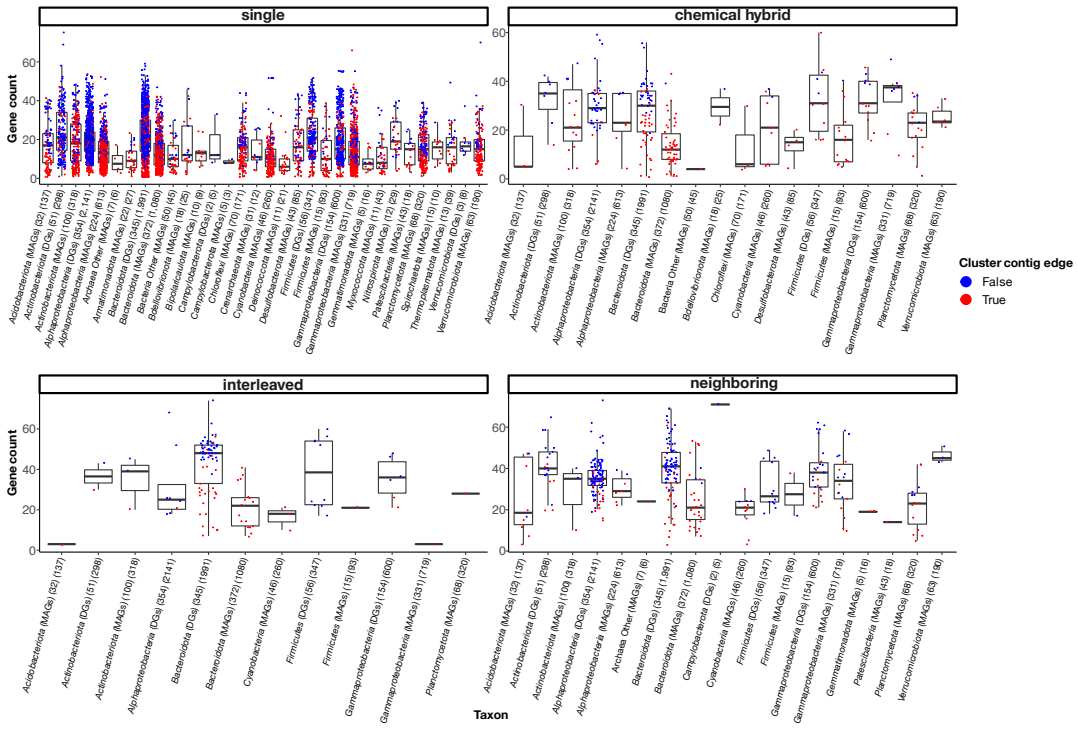


Figure S16. Biosynthetic gene cluster (BGC) sizes in genomes from distinct phyla. Four BGC categories were defined based on the distance between BGCs (see below). Each point in the box-plots represents one BGC. Red dots represent BGCs at contig edges, while blue dots represent BGCs within contigs. Numbers in parentheses refer to the number of genomes, and the corresponding total number of BGCs. Single: candidate cluster with only one BGC; chemical hybrid: candidate cluster containing BGCs sharing cluster-defining CDS/genes/gene products; interleaved: candidate clusters containing BGCs which do not share cluster-defining CDS features, but whose core locations overlap; neighboring: candidate clusters containing BGCs that do not match either chemical or interleaved variants, but transitively overlap in their neighborhoods.

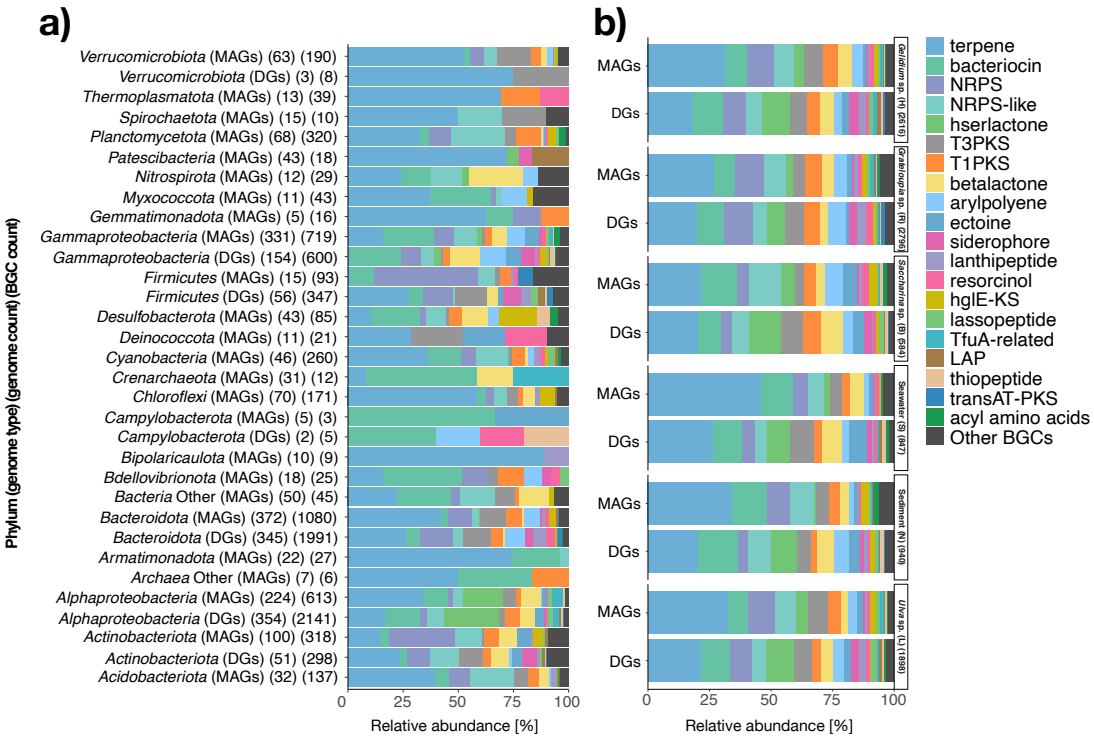


Figure S17. Clustering of biosynthetic gene clusters (BGCs) according to sample type and phylogeny. a) BGC types across habitats for MAGs and draft genomes (DGs). Numbers in the parentheses represents the numbers of MAGs or DGs and the corresponding sum of BGCs. **b)** Proportions of BGCs in MAGs and DGs obtained from all six sample sources. Numbers in parentheses correspond to the total numbers of BGCs.

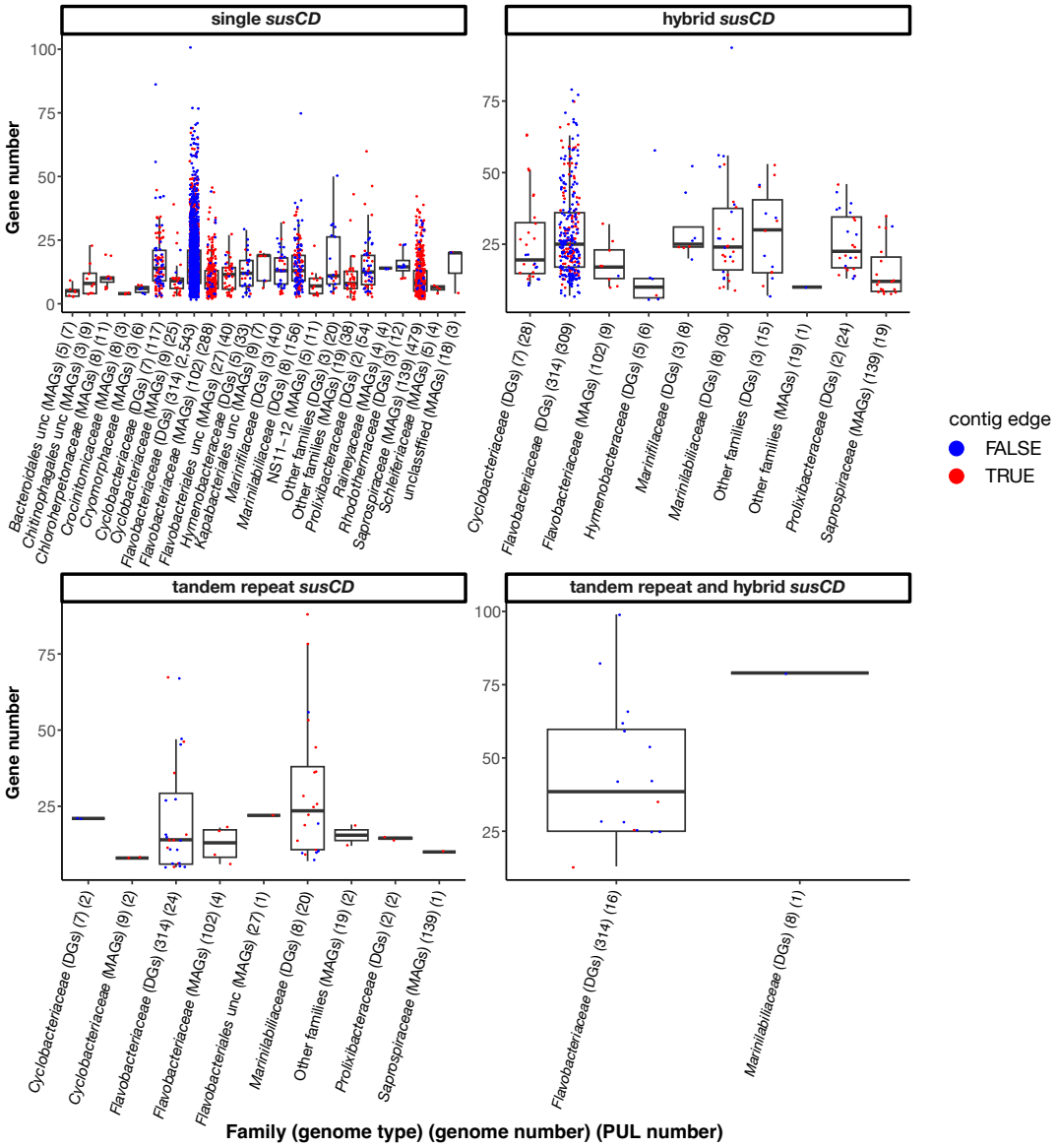


Figure S18. Sizes of PULs and PUL-like loci in genomes from distinct *Bacteroidota* families (categories: hybrid *susCD*, single *susCD*, tandem repeat *susCD*, and tandem repeat plus hybrid *susCD* PULs). Each point in the box-plots represents one PUL. Red dots represent loci at contig edges, whereas blue dots represent loci within contigs. Numbers in parentheses correspond to the number of genomes and the total number of corresponding PULs. DGs: draft genomes.

Microbiome - Additional file 1

Epiphytic common core bacteria in the microbiomes of co-located green (*Ulva*), brown (*Saccharina*) and red (*Grateloupia*, *Gelidium*) macroalgae

De-Chen Lu[1,2,3], Feng-Qing Wang[2], Rudolf I. Amann[2], Hanno Teeling[2*], Zong-Jun Du[1,3*]

[1] Marine College, Shandong University, Weihai 264209, China

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

[3] State Key Laboratory of Microbial Technology, Institute of Microbial Technology, Shandong University, Qingdao 266237, China

[*] Corresponding authors:

Zong-Jun Du, Marine College, Shandong University, No. 180, Wenhua xilu, Weihai, Shandong Province, 264209 P. R. China, e-mail: duzongjun@sdu.edu.cn, phone: +86 0631 5688303

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

Running title: Microbiomes of marine macroalgae

E-mail addresses and telephone numbers of all authors:

De-Chen Lu	202267000015@sdu.edu.cn	**+86 0631 5688303**
Feng-Qing Wang	fwang@mpi-bremen.de	**+49 421 2028 9390**
Rudolf I. Amann	ramann@mpi-bremen.de	**+49 421 2028 9300**
Hanno Teeling	hteeling@mpi-bremen.de	**+49 421 2028 9760**
Zong-Jun Du	duzongjun@sdu.edu.cn	**+86 0631 5688303**

Conflict of interest

The authors declare no conflict of interest.

This file includes:

Supplementary Results

Supplementary Methods

Supplementary Code availability

Supplementary References

Supplementary results

Sequencing and cultivation – basic metrics

Sequencing of 16S rRNA V3-V4 regions was done for 92 samples: 16x *Ulva* sp. (green algae), 16x *Grateloupia* sp. (red algae), 16x *Gelidium* sp. (red algae), 12x *Saccharina* sp. (brown algae), 15x surrounding seawater (-0.1 to -0.5 m), and 17x surrounding sediment. 10,058,653 amplicons of around 450 bp were obtained (average=109,333, max=282,978, min=24,866, SD=51,366), and clustered into 51,132 ASVs (Table S1 in Additional file 3).

Metagenomes were sequenced of 23 samples: 4x *Ulva* sp., 4x *Grateloupia* sp., 4x *Gelidium* sp., 3x *Saccharina* sp., 4x surrounding seawater, and 4x surrounding sediment. In total 1.3 Tbp of high-quality metagenome data with an average of 14.1 Gbp per sample were generated (Table S1 in Additional file 3). Assemblies ≥2.5 kbp ranged from 0.3 to 0.9 Gbp.

Cultivation of bacteria yielded 5,527 pure cultures (phycospheres: 4,426) representing eight phyla, 444 genera, and 1,235 species (phycospheres: 879), including 52 potential novel genera and 637 potential novel species (Table S2 in Additional file 3). Sequencing of selected strains yielded 965 draft genomes comprising 3.9 Gbp in total (Table S3 in Additional file 3).

Phycosphere core taxa

Based on 16S rRNA gene ASVs, *Bacteroidota*, *Actinobacteriota*, *Verrucomicrobia*, *Cyanobacteria*, *Deinococcus*, and *Patescibacteria* relative abundances were notably higher in phycospheres than in surrounding seawater and sediments (Fig. S3 in Additional file 2). *Proteobacteria* and *Actinobacteria* relative abundances were higher in phycospheres and seawater, whereas *Chloroflexi*, *Desulfobacterota* and *Campylobacterota* relative abundances were higher in sediments (Fig. S3 in Additional file 2). We studied seasonal changes in the phycosphere communities of all four macroalgae at the family and genus levels. While seasonal changes were observed, the core community as such was remarkably robust (Fig. 3, S4 in Additional file 2). The following 14 genera from eight families were identified as core phycosphere genera: *Saprospiraceae* unc. (*Saprospiraceae*), *Portibacter* (*Saprospiraceae*), *Lewinella* (*Saprospiraceae*), *Algitalea* (*Flavobacteriaceae*), *Microtrichaceae* unc. (*Microtrichaceae*), Sva0996 marine group (*Microtrichaceae*), *Rubritalea* (*Rubritaleaceae*), *Rhizobiaceae* unc. (*Rhizobiaceae*), *Robiginitomaculum* (*Hyphomonadaceae*), *Hellea* (*Hyphomonadaceae*), *Rhodobacteraceae* unc. (*Rhodobacteraceae*), *Sulfitobacter* (*Rhodobacteraceae*), *Granulosicoccus* (*Granulosicoccaceae*) and *Leucothrix* (*Thiotrichaceae*). All core phycosphere genera combined accounted for more than 70% of the diversity of all macroalgae samples, but only for 20% of the corresponding seawater and sediment samples (Table S1 in Additional file 3, Fig. 3).

Core taxa - average percentages on all sampled algae

Core families comprised on average 6.1% (*Gelidium* sp., 29/472), 4.6% (*Grateloupia* sp., 22/478), 3.6% (*Ulva* sp., 17/470) and 5.1% (*Saccharina* sp., 23/452) of all families (Table

S1 in Additional file 3). At the same time, they accounted for 89.1% (*Gelidium sp.*, ±9.3), 85.2% (*Grateloupia* sp., ±7.5), 88.0% (*Ulva* sp., ±6.0) and 85.9% (*Saccharina* sp., ±15.3) of the phycosphere bacterial relative abundances. Core genera comprised, on average, 4.3% (*Gelidium* sp., 42/972), 4.1% (*Grateloupia* sp., 41/1,000), 3.8% (*Ulva* sp., 37/973) and 3.5% (*Saccharina* sp., 30/870) of all phycosphere genera (14). At the same time, these genera accounted for 76.6% (*Gelidium* sp., ±16.5), 79.5% (*Grateloupia* sp., ±9.2), 85.5% (*Ulva* sp., ±5.3) and 72.8% (*Saccharina sp.*, ±17.6) of all phycosphere bacteria (Table S1 in Additional file 3).

Taxa with discernible host-specific and seasonal patterns

The complete 16S rRNA amplicon dataset contained 18,434 ASVs representing genus-level host-specific phycosphere core taxa (total 51,132). Core phycosphere communities were on overall remarkably stable, whereas host-specific phycosphere taxa showed a more pronounced seasonal variability (Figs. 2C, 3, S4 in Additional file 2).

Most other phycosphere taxa appeared only in individual samples, either during single or multiple seasons. Many belonged to rarer phyla, such as *Bdellovibrionota* (*Bacteriovoracaceae* and *Bdellovibrionaceae*), which reached relative abundances of up to 1.1% (*Bdellovibrionaceae*, ±0.9) and 2.1% (*Bacteriovoracaceae*, ±1.3) on *Gelidium* and *Grateloupia* red macroalgal species in summer, respectively. Some clades of the *Patescibacteria,* such as *Cd.* Kaiserbacteria unc., *Saccharimonadales* unc., JGI 0000069-P22 unc. and *Absconditabacteriales* (SR1) unc. exhibited relative abundances of up to 5.0% on *Ulva* sp. green macroalgae in spring (Table S1 in Additional file 3). Further non-core families exhibited high relative abundances only in particular seasons. For example, *Shewanellaceae* were always abundant on *Ulva* sp. and *Gelidium* sp. in winter and on *Grateloupia* sp. and *Saccharina* sp. in spring. Likewise, seawater and sediment exhibited the highest relative abundances of *Shewanellaceae* in the winter and spring (Fig. S4 in Additional file 2).

Moraxellaceae, *Trueperaceae* and *Thiotrichaceae* belonged to the dominant phycosphere community, i.e., they were consistently present on three of the four sampled algal species. *Moraxellaceae* were consistently more abundant in spring and summer and

had low relative abundances on *Saccharina* sp. brown macroalgae (Fig. S4 in Additional file 2). Conversely, *Trueperaceae* were more abundant in autumn on *Ulva* sp. green algae, and in winter to spring on *Gelidium* and *Grateloupia* red algae. Gammaproteobacterial *Thiotrichaceae* were abundant on all macroalgae except *Ulva*, where they were mostly condined to winter and spring. Also in seawater *Thiotrichaceae* abundances were highest in spring (Fig. S4 in Additional file 2).

Diversity of cultured bacteria

In terms of diversity, there was no significant difference between strains obtained from different macroalgal samples. Families exceeding 1% sequence abundance in at least one sample were *Flavobacteriaceae* (32.8%), *Rhodobacteraceae* (24.1%), *Sphingomonadaceae* (8.1%), *Bacillaceae* (6.8%), *Vibrionaceae* (3.1%), *Moraxellaceae* (2.0%), *Alteromonadaceae* (1.7%), *Micrococcaceae* (1.5%), *Stappiaceae* (1.5%), *Halomonadaceae* (1.2%), *Pseudoalteromonadaceae* (1.2%), *Cellvibrionaceae* (1.1%), *Staphylococcaceae* (1.1%), *Shewanellaceae* (1.1%), *Intrasporangiaceae* (1.1%) and *Colwelliaceae* (1.1%) (Table S2 in Additional file 3). More species were found on algae that did not occur in the sediment and seawater controls than vice versa (Fig. S5 in Additional file 2), including in particular members of the *Bacteroidota* (genera *Aquimarina*, *Pibocella*, *Cellulophaga*, *Dokdonia*, *Flagellimonas*, *Tenacibaculum*, *Algibacter*, *Winogradskyella*, *Maribacter*, *Croceitalea*) and *Proteobacteria* (genera *Paraglaciecola*, *Psychrobacter*, *Colwellia*, *Acinetobacter*, *Cobetia*, *Halomonas*, *Sulfitobacter*, *Sphingorhabdus*, *Labrenzia*, *Erythrobacter*, *Jannaschia*, *Tateyamaria*, *Pelagibius*, *Pseudophaeobacter*, *Altererythrobacter*, *Roseovarius*). Other algae-associated species affiliated with *Actinobacteria* (genera *Kocuria*, *Microbacterium*, *Serinicoccus*, *Ornithinimicrobium*, *Arthrobacter*, *Dietzia*) and *Firmicutes* (genera *Planococcus*, *Bacillus*, *Staphylococcus*).

Compositional differences between macroalgal samples, seawater and sediment controls were confirmed by clustering using UniFrac distances (Fig. S6a in Additional file 2) and by PCoA using Bray-Curtis distances (Fig. S6b in Additional file 2). No clear pattern was discernible with respect to algal species, as samples from different algae tended to

cluster according to season rather than species. Also, the two cultivation media that were used seemed to have a negligible influence (Fig. 4). When we conducted a PCoA of 16S rRNA sequences of cultured strains together with corresponding ASVs from 16S rRNA amplicon sequencing (Fig. S2 in Additional file 2) (ANOSIM: R=0.835, p=0.001, PERMANOVA: F=9.16, p=0.001), we obtained seasonal patterns for each algae that were similar as in a PCoA of ASVs from exclusively non-core taxa (Fig. 2c) (ANOSIM: R=0.478, p=0.001, PERMANOVA: F=3.745, p=0.001). This indicates that the majority of cultured species belonged to the more diverse non-core taxa.

MAGs and genomes of family level core taxa

At the family level, 645/30 (macroalgae: 589/29) draft genomes and 548/24 (macroalgae: 474/22) MAGs represented core/dominant phycosphere families (Fig. S4 in Additional file 2, Table S3 in Additional file 3). Some of the families consisted solely of genera below the 2% abundance threshold, which however collectively amounted to more than 2% of the sequences. In particular *Flavobacteriaceae*, *Sphingomonadaceae*, *Rhodobacteraceae*, and *Granulosicoccaceae* were well represented, while *Trueperaceae* and *Cd* Campbellbacteria unc. were only represented by few draft genomes and MAGs, respectively. In addition, 155 genomes representing core seawater and 72 genomes representing core sediment families were obtained.

Clustering of draft genomes and MAGs by habitat

In a corresponding phylogenomic tree (Fig. 5a), the majority of draft genomes and MAGs from macroalgal phycospheres were interspersed by smaller clusters of draft genomes and MAGs from seawater and, to a lesser extent, sediment samples (Fig. 5a). Based on GTDB taxonomy, the tree represented in total 42 known phyla, 78 known classes and 162 known orders (Table S3 in Additional file 3). At the genus (Fig. 3) and family (Fig. S4 in Additional file 2) levels, many taxa covered a wider range of samples and thus habitats. However, on the species level, it became apparent that most originated from a single habitat and that only a minor fraction of the species had a broad habitat range (Table S3 in Additional file 3, Fig. 5b). Clades that consisted exclusively of draft genomes tended to

represent less abundant non-core taxa. Conversely, the more abundant clades tended to be represented by MAGs. For instance, 17 of the 24 most abundant phyla were exclusively represented by MAGs (Fig. 5a). These MAGs comprised 708 uncultured species (95% ANI) that were represented by a single MAGs each, 54.5% of which originated from one of the macroalgal samples (Table S3 in Additional file 3).

Hybrid PULs

As described in the method section futher below, we split artificially predicted long hybrid *susCD* PULs into smaller PULs (Fig. S12 in Additional file 2). 241 hybrid *susCD* PULs were initially predicted and subsequently spit into 440 single *susCD* PULs plus 103 remaining hybrid *susCD* PULs. In addition, eleven tandem repeat and hybrid *susCD* PULs were divided into eleven single *susCD* PULs, eight tandem repeat *susCD* PULs and five tandem repeat and hybrid *susCD* PULs. On overall about 22.5% of PULs were located five genes or less from a contig boundary, and 10.3% even from both boundaries (Table S4 in Additional file 3). This affected MAGs to a larger extent (56.9%) than draft genomes (13.2%). Hence, PUL predictions in genomes of cultured bacteria were more complete and reliable than those in MAGs. Consequently, the average lengths of PULs from cultured bacteria was longer than from MAGs (Figs. S15, S18 in Additional file 2).

Incomplete susCD-like PULs

Most predicted PULs contained the characteristic *susCD*-like gene pair along with co-occurring CAZymes. There were also PULs that lacked any obvious *susD*-like gene, such as predicted digeneaside (α-D-mannopyranosyl-(1→2)-D-glycerate) PULs [1]. We identified in total 110 PULs missing *susC*-like and 169 missing *susD*-like genes. On overall, clustering based protein sequences of *susC*-only PULs was more successful than for *susD*-only PULs, with the former clustering in a single branch of the SusC tree that featured PULs rich in CE1, CE3, or CE4 possibly targeting xylose-containing polysaccharides. Some PULs have been reported to lack *sus* genes, e.g., those targeting trehalose [1]. This may be due to corresponding *susCD* gene pairs and CAZymes genes

not being encoded in one canonical locus (a limitation of the PUL prediction method) [2] [3][4].

PUL classification based on susCD presence

We obtained in total 718 redundant draft genomes and MAGs from *Bacteroidota* (phycosphere: 566; seawater: 98; sediment: 54, with some occurring in multiple habitats). In these genomes, 4,451 PULs were predicted (Table S6 in Additional file 3). The majority featured a single *susCD* gene pair (3,670, 82.5%), followed by hybrid *susCD* PULs (449, 10.1%), tandem *susCD* (58) and tandem/hybrid *susCD* PULs (17). Hybrid *susCD* PULs might have originated from PUL fusions, as these PULs tended to be among the longest predicted PULs (Fig. 7) harboring functionally diverse CAZyme genes. Hybrid PULs were prevalent using our prediction algorithm, especially within draft genomes from cultured strains (Fig. S18 in Additional file 2). Hybrid *susCD* PULs are relatively common among *Bacteroidota*. Even if the window width for PUL detection was adjusted down to one, still 173 hybrid *susCD* PULs were predicted (Table S4 in Additional file 3). In order to estimate the diversity of targeted glycans, we compared PUL CAZyme gene compositions harnessing substrate specificities of CAZyme families and sub-families [5]. From a CAZyme gene presence/absence matrix, we then calculated Jaccard distances [6] representing pairwise CAZymes gene and sulfatase composition dissimilarities. Based on these distances, we finally partitioned the data into 2,257 clusters (100% similarity). Out of 4,451 PULs, 68% belonged to 786 clusters of at least two PULs of identical enzyme composition. while 32% had a unique composition (Table S5 in Additional file 3).

Candidates for GT10 family enzymes

Functional annotations revealed that all draft genomes from sequenced strains coded for CAZymes, with on average 113 genes and an astounding maximum of 483 genes in the novel species *Fulvivirga* sp. (strains 361 and M361-1). Besides typical degradative CAZmyes (GHs: glycoside hydrolases, CEs: carbohydrate esterases, PLs: polysaccharide lyases), we also looked specifically at glycosyltransferases (GTs). GT family 10 for example includes alpha-(1,3)-fucosyltransferases (FucT) that are of

particular biotechnological interest [7][8]. *Muricauda* sp. (*Flavobacteriaceae*) encoded the highest number of GT10 genes (n = 3). This and four other species with GT10 genes (n=2) were all isolated from macroalgae. All represented novel species of prevalent phycosphere bacteria. While a detailed biochemical characterization of the encoded GT10 enzymes requires additional work, the identification of such candidate genes demonstrates the functional potential preserved in our strain collection.

PUL complexity assessment

Prior to glycan backbone depolymerization accessory structures must be removed, which explains the complexity of some PULs. We found that the numbers and complexities of PULs derived from macroalgae and sediments were higher than those from seawater samples. In addition, the proportion of predicted secreted CAZymes, were ~4/4.6% and ~4.1/5.4% (GH/PL) higher in phycosphere and sediment bacteria than in seawater, respectively (Table S3 in Additional file 3). This suggests more synergistic degradation using shared exo-enzymes in phycosphere and sediment communities than in the water column.

The lengths and CAZyme gene richness of hybrid PULs, tandem repeat, and hybrid *susCD* PULs and tandem repeat *susCD* PULs were higher than those of single *susCD* PULs (Table S5 in Additional file 3). Complex PULs (higher CAZymes richness) mainly originated from macroalgal phycosphere and sediment bacteria. In particular, a large number of complex hybrid PULs was obtained that likely target more complex and diverse macroalgal polysaccharide substrates. It was evident that long and complex PULs are more common among tandem repeat and hybrid *susCD* PULs and tandem repeat *susCD* PULs.

Sulfatases and acetylases in PULs

In contrast to terrestrial plants, which contain many acetylated polysaccharides [9], glycans of marine algae are often sulfated [10] and thus require desulfating sulfatases for breakdown. However, the most frequent PULs targeting laminarin, alpha-glucans and alginate rarely contained sulfatase genes, reflecting the fact that these substrate classes

are mostly unsulfated. Only 7.1% of the laminarin, 12.7% of the alpha-glucan (except for clusters 66 and 72), and 6.7% of the alginate PULs contained sulfatase genes. The abundance of laminarin and alpha-glucan PULs can be explained by their function as energy storage molecules that requires a simple, easily accessible, unsulfated structure allowing for quick mobilization of glucose monomers.

Well-known sulfated polysaccharides include galactans (agars and carrageenans) from red algae, ulvans from green algae, and fucans and fucoidans from brown algae. In particular PULs predicted to target ulvans, rhamnans, carrageenans and fucose-containing sulfated polysaccharides (FCSPs) mostly contained sulfatase genes (100%, 83.1%, 78.9%, and 78.8%, respectively). Most of these polysaccharides have persistent structural functions and thus are more complex and thioesterified [10]. As expected, the percentage of PULs with deacetylase genes (2.0%) was much lower than that with sulfatase genes (22.3%).

Horizontal gene transfer of PULs

We assessed the phylogeny of SusC and SusD sequences and associated PUL structures spanning at least 27 families and 70 genera of the *Bacteroidota*. PUL conservation was not always congruent with the 16S phylogeny, which may indicate frequent horizontal transfer (HGT) of PULs (Table S5). This was not only evident for laminarin PULs (Table S5 in Additional file 3, Fig. 6), but also for PULs corresponding to other polysaccharides. We also observed that CAZyme genes close to *susCD* gene pairs tended to be more conserved than those farther away. This is likely a functional effect of coupling recognition, binding and transport of oligosaccharides across the outer membrane by *susCD* with dedicated CAZymes. It might explain, why it seemed that in particular these core CAZymes are preserved in PUL LGT events. Such putative LGT was mostly predicted between members of the same class or order. We consider such events as an important driving force for microbial evolution and niche adaptation within macroalgal ephiphytic microbial communities [11].

Phaeophyta

- Laminarins

Laminarins are short chain β-1,3-linked glucans with occasional β-1,6-branches [12]. Enzymes from the GH families 5, 8, 9, 16, 17, 30, 64, 81, and 157 are involved in laminarin degradation. The backbone is usually broken down by GH16_3 endo-glucanases. PULs targeting laminarin were the most frequent of all predicted PULs (618 PULs, Table S5 in Additional file 3). This reflects the fact that laminarin is one of the most abundant macromolecules on Earth, as it acts as storage compound in brown algae and diatoms [13][14]. A large number of such PULs were also found in *Bacteroidota* genomes isolated from red and green macroalgae, which do not contain laminarin. For example, most *Aquimarina* strains had the potential to degrade laminarin regardless of origin (Table S5 in Additional file 3). Results were highly dependent on bacterial taxonomy instead of host origin.

- Alginates

Alginates are common in brown algae [15][16]. Alginates are linear co-polymers consisting of homopolymeric blocks of (1→4)-linked β-D-mannuronate and α-L-guluronate residues that are covalently linked in alternating sequences or blocks [1]. Alginates are anionic and bind sodium and calcium ions. Alginates were the fifth most frequently predicted PUL substrates in our dataset. They were predicted in 287 of the draft genomes, amounting to a total of 426 alginate-specific PULs. Alginate PULs encode PL6, 7, 14,15, and 17 family alginate lyases [11]. Such PULs have for example been described in numerous *Flavobacteriaceae*, such as members of the NS5 marine group, *Polaribacter*, *Aurantivirga*, and *Gramella* [1][17]. Previous studies on kelp biofilms yielded bacterial MAGs enriched in alginate degradation genes [11]. In our study, *Flagellimonas*, *Aquimarina,* and *Saprospiraceae* revealed strong degradation potentials for alginates, notably without host-specificity towards just *Saccharina* sp. brown algae.

- Fucose-containing sulfated polysaccharides (FCSPs)

Fucoidans occur in brown macroalgae, such as *Fucus vesiculosus*, *Laminaria* spp. (kelp) and *Macrocystis* spp. [18]. The main monomer is sulfated L-fucose, but the chemical

composition of fucoidans is often complex and contains other monosaccharides (mannose, galactose, glucose, xylose, etc.), uronic acids, acetyl groups and even proteins [19]. Known fucosidases are present in the GH29, 95, 107, 141 and 151 families [19]. In a study on *Verrucomicrobia*, it was shown that FCSP degradation requires sulfatases to remove sulfate moieties before GH107 endo-fucanases can cleave the backbone into oligosaccharides. Exo-fucosidases of the GH29, 95, and 141 families subsequently hydrolyze these into fucose and other monomers [19]. It is noteworthy in this context that also MAGs obtained from kelp-colonizing bacteria were rich in enzymes for the degradation of fucoidans (CBM47, GH29, GH95) [11][15][16].

Fucoidan degradation activity has been reported in *Gammaproteobacteria* [20], *Rhodobacteraceae* [21], *Flavobacteriaceae* (*Mariniflexile fucanivorans*) [22][23], and *Verrucomicrobiota* [19][24]. We found FSCP-targeting PULs in members of the genera *Algibacter*, *Maribacter*, *Polaribacter*, *Aquimarina*, *Fulvivirga*, *Flagellimonas*, *Marinilabiliaceae*, and *Prolixibacteraceae* (Fig. 6, Table S5 in Additional file 3).

Chlorophyta

- Beta-xylose-containing substrates

Xylan is a major component of many plant cell walls. It is one of the most structurally variable polysaccharides and occurs for example as arabinoxylan (cereal grains), glucuronoxylan, acetylated and sulfated xyloglucan (marine *Chlorophyta*), or as glucuronoxylan (fruit and vegetables) [1][25][26]. Xylan-targeting PULs can contain GH8, 10, 30, 43, 51, 67, 74, or 115 families. PULs containing GH67 and 115 may target glucuronoxylan, while PULs containing GH51 may target arabinoxylan. Apart from a separate locus coding for a multi-modular GH10 xylanase, predicted xylan degradation genes in this study constituted a single cluster containing GH3, 8, 43 and 115 family genes (Table S5 in Additional file 3). In addition to GHs, there are also a large number of PULs that contain CEs. Such as, family CE7 and 17 carbohydrate esterases have been primarily characterized in terrestrial plants. They act as glucuronyl esterases and acetyl-xylan esterases that degrade lignocellulose and remove acetyl groups from hemicelluloses [27]. CE7 and 15 family enzymes from marine microbes have been

characterized and their activities on xylans have been demonstrated [27][28]. In our dataset, a significant proportion of PULs encoded carbohydrate esterases of families CE1, 3, 6, 7, and 15, known to be capable of removing acetyl groups. These PULs thus may target oligosaccharides containing acetylated xylose, suggesting the presence of xylan acetylation in marine macroalgae. In addition, studies have shown that utilization of some sugars might be facilitated by acetyl xylan esterases (CE6) that are shared within macroalgal phyosphere communities [11]. These bacteria show the potential to utilize xylan. A large number of PULs targeting xylan substrates were found in bacteria in this study, as 344 isolates featured in total of 779 putative xylan-specific PULs.

- Mannose-rich substrates

There are two main classes of mannans, α-mannans and β-mannans with α-1,6- or β-1,4/1,3-mannan backbones, respectively. Beta-mannans have been reported in red and various green macroalga [29], where linear β-mannan seems to replace cellulose as the main cell wall glycan. Such β-mannans are for example targeted by GH26 and 130 CAZymes. Additional GH2, 3, and 88 family genes that all represent diverse functions have been reported in predicted β-mannan-targeting PULs as well [1]. In contrast, predicted α-mannan-targeting PULs have been found to feature GH38, 76, 92, 99, and 125 family CAZymes. They might be targeting α-glucomannans, such as glucuronomannan, a polysaccharide that has been reported for brown algae [15][16][30]. Most α-mannan-targeting PULs code for multiple GH92 family exo-mannosidases and at least one GH76 family endo-α-1,6-mannanase. GH92-enriched PULs are thought to potentially target α-mannose-rich N-glycosylated glycoproteins that are widespread in eukaryotes, including algae [1]. In this study, 238 genomes featured a total of 488 PULs targeting mannose-rich substrates.

- Pectic glycans

Pectins occur in plant cell walls and are abundant in fruits and vegetables, but have also been shown to be also a substrate of marine bacteria [31]. The two major pectins are homogalacturonan (HG) and rhamnogalacturonan-I (RGI). HG consists of α-1,4-linked D-

galacturonic acid (D-GalA), while the backbone of RGI consists of repeating units of the disaccharide α-1,2-L-rhamnose (Rha)-α-1,4-D-GalA. The backbones of HG and RGI are covalently linked [32]. Members of PL1, 2, 3, 9, 10, 22, and GH28, GH105, CE8, CE12, CE13, and CE19 enzymes are involved in microbial pectin degradation [33]. In our dataset, we identified 72 PULs in 37 genomes (Table S5 in Additional file 3) featuring GH28 α-1,4-polygalacturonases, GH88 unsaturated β-glucuronyl hydrolases, family PL1, 9, 22 and 10 pectate lyases, CE8 carbohydrate esterases, and CE12 pectin methylesterases.

- Ulvans

Ulvans are acidic polysaccharides with a structural function in cell walls of green algae [34]. They are highly sulfated and essentially composed of rhamnose-3-sulfate, xylose, xylose-2-sulfate, glucuronic acid, and iduronic acid monomers [34]. Ulvan lyases occur in families PL24, 225, 28, 37, and 40. In our dataset 68 genomes featured in total 127 predicted ulvan PULs.

Rhodophyta

- Carrageenans

Carrageenans are sulfated galactans in marine red algae and in seagrass [35][36]. Carrageenans are mainly composed of alternating 3-linked β-D-galactopyranose (G-units) and 4-linked α-D-galactopyranose (D-units) or 4-linked 3,6-anhydro-α-D-galactopyranose (DA-units), forming a repeating disaccharide unit [37]. GH16, 82, 167, 150 and 127 family enzymes are involved in microbial carrageenan degradation [3].

Our data contained 113 PULs with characteristic carrageenase, GH110 family α-galactosidase, and sulfatase genes that likely target sulfated, galactose-rich substrates (Table S5 in Additional file 3). Some of these PULs coded for additional GH16_17, 82, 127, or 167 family CAZymes as well as predicted GH2 family β-galactosidases (Table S5 in Additional file 3). In addition we identified sulfatases specific for carrageenans in the flavobacterial genera *Zobellia*, *Aquimarina*, *Tenacibaculum*, *Flavivirga*, and *Algibacter*.

- Agars

Agars are jelly-like polysaccharides present in the cell walls of some red algal species [38]. Agars consists of a mixture of mainly agarose and to a lesser extent agaropectin. Agarose forms a linear polymer of repeating units of agarobiose (disaccharide of D-galactose and 3,6-anhydro-L-galactopyranose). Agaropectin consists of alternating units of D-galactose and L-galactose that are heavily modified by sulfate and pyruvate groups [38]. GH16, 50, 86, 96, 117 and 118 family enzymes are involved in microbial agar degradation. In our dataset, a total of 106 genomes featured 192 predicated agar-specific PULs.

- Porphyrans

Porphyrans are present in red algae. They are sulfated galactans composed of alternating 1,4-linked α-L-galactopyranose-6-sulfate (L6S) and 1,3-linked β-D-galactopyranose (G) [39]. GH16 and 86 family enzymes are involved in microbial porphyran degradation. Our dataset featured a total of 31 genomes with 31 predicted porphyran-specific PULs.

Other polysaccharide substrates

- Alpha-glucans

PULs predicted to target α-1,4-glucans, such as starch (α-1,4- and α-1,6-glucan), pullulan (α-1,4- and α-1,6-glucans), dextran (α-1,6-glucan), glycogen (α-1,4- and α-1,6-glucans) and amylose (α-1,4-glucans) were highly abundant (482 PULs, Table S5 in Additional file 3). Starch is widely found in terrestrial plants, microalgae [40], macroalgae and animals, and thus plays a central role in life as a principal store of chemical energy [41][42]. Enzymes of families GH13, 15, 31, 57, 65, 70, 71, 77, and 87 are involved in α-glucan degradation. Some PULs featured only a single GH13 and GH65 gene. These PULs likely target was structurally simple, non-branched α-1,4-glucans such as maltodextrin or amylose [1]. Our dataset featured with 596 predicted α-glucans-specific PULs.

- Chitin

Chitin is structurally similar to cellulose, except that the C_2-hydroxyl (-OH) group is replaced by an acetamide group (NH-CO-CH_3). It is thus a homopolymer of (1→4)-β-linked N-acetylglucosamine (GlcNAc) and belongs to the most abundant polysaccharides in marine habitats. Chitinases are for example present in CAZyme families GH18, 19, 20 and 23. Such chitinases frequently possess chitin-binding domains, e.g., of families CBM5, 12, and 50. Recently, chitinase-encoding genes have been identified in *Aquimarina* strains from marine sponges, corals, sediments, and seawater [43]. We found that macroalgae-associated bacteria of the genera *Aquimarina*, *Tenacibaculum*, *Carboxylicivirga* and *Formosa* contained putative chitin-PULs (Table S5 in Additional file 3), highlighting species of *Flavobacteriaceae* as source of putative novel chitinolytic enzymes. Diverse and abundant lineages within these major bacterial taxa were indeed present in our macroalgal microbiomes (Figs. 3, 6, S4 in Additional file 2).

- Sulfated α-rhamnose-containing polysaccharides

Sulfated rhamnans have been reported for green macroalgae [34]. Our dataset comprised 109 genomes containing a total of 219 PULs predicted to target sulfated α-rhamnose-containing polysaccharides. These PULs featured predicted GH78 family α-L-rhamnosidase genes often accompanied by GH28 family rhamnogalacturonase genes (Table S5 in Additional file 3). Some of these rhamnose-targeting PULs coded for additional predicted GH105 family rhamnogalacturonyl hydrolases, which cleave rhamnose from uronic acids. As GH78 family enzymes have also been shown to act on rhamnogalacturonans, it seems likely that these GH105-containing rhamnose-PULs target rhamnogalacturonans.

- Peptidoglycan

Peptidoglycan (PG) is an essential macromolecule of most eubacterial cell walls [44]. It is composed of a repeating backbone of N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM), connected by β-1→4 glycosidic bonds. Members of glycoside hydrolase families GH18, 23 (lytic transglycosylases), and 73 (β-N-acetylglucosaminidases) are involved in peptidoglycan degradation. Our dataset

contained 66 genomes harboring 69 PULs predicted to target PG. These included members of *Aquimarina*, *Formosa*, *Carboxylicivirga,* and the *Saprospiraceae* that had GH23 and 73 family genes.

- N-glycans

Macroalgal glycans represent a significant nutrient source, and access to these host molecules appears to be important for microbial phycosphere colonization. Among these polysaccharides are eukaryotic N-glycans. The GH18 and GH63 families include endo-β-N-acetylglucosaminidases with a range of different specificities for different types of N-glycans. Some strains of *Tenacibaculum* and *Carboxylicivirga* were predicted to utilize N-glycans via a single PUL targeting high mannose N-glycans (HMNG) (Table S5 in Additional file 3). Our dataset comprised 23 genomes containing a total of 23 PULs predicted to target N-glycans.

- Sialic acids

GHs acting on sialic acids (GH33) and glycosaminoglycans (GH88) were enriched in macroalgae-associated genomes in this study. In total 16 genomes harboring 19 PULs could code for the capacity to tackle sialic acid oligosaccharides via PUL-associated GH33 family sialidases (Table S5 in Additional file 3).

Supplementary Methods

Isolation and plate cultivation of bacteria

Phycosphere bacteria: Complete macroalgae were cut into 9 cm^2 pieces and rinsed thrice with sterile seawater. Afterwards 10 g of these pieces were washed with 10 ml sterile seawater (rotary shaker, 170 rpm., 30 min, 25 °C). One milliliter aliquots were diluted stepwise to 1:100,000 (sterile seawater), and 100 µl were subsequently plated and incubated (21 d, 28 °C).

Seawater samples: One milliliter seawater was diluted stepwise with 9 ml sterile seawater to 1:1,000. Aliquots of 100 µl were then plated and incubated as described above.

Sediment samples: Sediment samples of 1 g were thoroughly mixed with 9 ml sterile seawater (rotary shaker, 170 rpm., 30 min, 25 °C) and then diluted in 1:10 steps with sterile seawater to 1:10,000 [45]. Afterwards, 100 µl aliquots were plated as described [45].

Two media were used for plating, (a) modified 2216E: 18 g sea salt (Sigma-Aldrich, St. Louis, MO, USA), 1.5 g peptone, 0.3 g yeast extract, 0.3 g sodium pyruvate, 0.3 g glucose, 15 g agar, 2 g alginate, 2 g starch, 2 g carrageenan, 2 g cellulose and 0.5 mg vitamin B_{12}; and (b) modified VY/2: 18 g sea salt (Sigma-Aldrich), 1 g $CaCl_2$, 5 g active yeast, 0.3 g sodium pyruvate, 0.3 g glucose, 15 g agar, 2 g alginate, 2g starch, 2 g carrageenan, 2 g cellulose, and 0.5 mg vitamin B_{12} (all amounts per 0.5 L distilled water plus 0.5 L old seawater). Plating was completed within 2 h after sampling. After incubation, colony-forming units were counted with numbers ranging from 10 to 500. Colonies were selected depending on color, size, and shape. Picked colonies were purified by serial cultivation on plates with identical media. Purified strains were stored at -80 °C in sterile 1% (w/v) saline medium with 15% (v/v) glycerol.

DNA extraction from cultured strains and environmental samples

Cultured strains: DNA was extracted using the MiniBEST Bacteria Genomic Extraction kit v3.0 (TaKaRa Bio, Kusatsu, Shiga pref., Japan) and stored at -80 °C until use. Quality was checked by running the samples on 1% sodium boric acid agarose gels and measuring DNA concentrations using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) based on 260/280 nm and 260/230 nm absorbance ratios. Beijing Novogene Biotechnology (Beijing, China) performed sequencing using 150 bp PE technology on an Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA) with a coverage exceeding 100x. Reads were filtered as follows: (i) reads with ≥40% low-quality bases (Q ≤20) were removed, (ii) reads with ≥10% ambiguous bases were removed.

Environmental samples: Loosely attached microorganisms were removed by washing whole macroalgae thrice in flasks that were about 80% filled with sterile seawater on a shaker and collecting the washing suspensions in sterile bottles. Uncut algae fronds were

then put in sterile water and the remaining bacteria were removed by ultrasonication (2x 60 s, 50-60 kHz). The bacterial suspensions were subsequently pooled with the previously obtained washing suspensions for each sample. Per sample, about 60 L of suspended bacteria were filtered through 0.2 µm pore size polycarbonate membrane filters (Millipore, Billerica, MA, USA). DNA was extracted from these filters as described elsewhere [46] with adjustments, by means of the ADX1120 Advanced Water DNA Kit (Guangdong Magigene Biotechnology Co. Ltd., Shanghai, China). For seawater, 1 L of the sample was filtered on a 0.2 µm pore size polycarbonate membrane filter (Millipore). DNA was subsequently extracted from filters using the ADX1120 Advanced Water DNA Kit (Guangdong Magigene Biotechnology), and quantified using the Invitrogen Quant-iT PicoGreen dsDNA reagent (Thermo Fisher Scientific). DNA concentrations ranged from <1 to 20 mg/mL. For sediments, DNA was extracted from 300 mg sub-samples using the Power soil DNA isolation kit (MO BIO, Carlsbad, CA, USA).

Metagenome sequencing

We sequenced 23 bacterial metagenomes, namely 4x *Ulva* sp., 4x *Grateloupia* sp., 4x *Gelidium* sp., 3x *Saccharina* sp., 4x surrounding seawater and 4x surrounding sediment. After DNA extraction (see previous section), metagenome libraries were prepared by sonicating the DNA to a 350 bp insert size range. DNA fragments were subsequently sequenced in paired-end mode (2 × 150 bp) on the Illumina NovaSeq 6000 platform according to the manufacturer's instructions for metagenome analyses at Guangdong Magigene Biotechnology. A total of 1.4 Tbp (avg: 65 Gbp per metagenome) were generated (Table S1 in Additional file 3).

Diversity analyses

Alpha diversity metrics were calculated after rarefying the samples to 12,460 reads per sample. OTUs were clustered at ≥97% similarity. ASV richness, Simpson and Shannon's indices were calculated in R v3.5.1 (http://www.r-project.org) using the vegan, ggplot2, ggpubr, and dplyr packages. Jaccard distances based on the presence/absence of taxa were generated using pairwise comparisons of macroalgae, seawater and sediment

OTUs. Jaccard distances were obtained from principal coordinate analysis in R using the vegan, ggpubr, reshape2 and ggsci packages.

Selection of housekeeping genes for phylogenetic analysis

Sequences of the following ribosomal proteins were used for phylogenomic analysis of draft genomes and MAGs: S12/S23, L1, L13, L14, L16, L17, L18p, L19, L2, L20, L21p, L22, L23, L27, L27A, L28, L29, L3, L32p, L35p, L4, L5, L6, L9_C, S10, S11, S13, S15, S16, S17, S19, S2, S20p, S3_C, S6, S7, S8 and S9.

Annotation of genes coding for sulfatases, SusC- and SusD-like proteins

Genes coding for sulfatases, SusC- and SusD-like proteins were predicted using HMMER with the Pfam profiles PF00884 (sulfatase), PF07715, PF07980, PF12741, PF14322, and PF12771 (SusD-like), and the TIGRFAM profiles TIGR04056 TIGR01352, TIGR01778, TIGR01779, TIGR01782, TIGR01783, TIGR01785, TIGR01786, TIGR02796, TIGR02797, TIGR02803, TIGR02804, TIGR02805, TIGR04057, PF00593 (TonB-dependent receptor) with a search cutoff of E-10.

PUL definitions used in this study

Catabolic CAZymes often cluster in polysaccharide utilization loci (PULs). Most published and experimentally verified PULs have been found in *Bacteroidota*, where they usually contain a characteristic *susCD* gene tandem. We used a broader definition and searched loci with at least three genes coding for either CAZymes, sulfatases, SusC- or SusD-like proteins, or other TonB-dependent transporters. These loci were divided into four categories (Fig. S12a in Additional file 2): (i) PULs consisting of CAZyme genes and a *susCD* pair, (ii) CAZyme-rich gene clusters (CGC) consisting solely of CAZymes, (iii) PUL-like clusters with CAZyme genes and an encoded TonB-dependent receptor, and (iv) *susCD* loci without detectable CAZymes.In some cases, the sequence similarity of a TBDT was too low to be considered as a SusC-like or SusD-like protein. Corresponding loci were still considered as incomplete PULs [1][2].

We divided PULs into four additional categories (Fig. S12b in Additional file 2): (i) tandem repeat *susCD* PULs containing at least one tandem *susCD* pair, (ii) hybrid PULs containing multiple non-adjacent *susCD* pairs, (iii) PULs containing a single *susCD* pair, and (iv) tandem and hybrid PULs containing both, multiple non-adjacent *susCD* gene pairs with at least one tandem *susCD* gene pair.

The number of hybrid PULs increased with increasing sliding window size, whereas the number of single *susCD* gene pair PULs was only little affected. In order to minimize detection of artificial hybrid PULs, such PULs were split when five or more contiguous non-marker genes were present. The split was made amidst the non-marker genes, ensuring that each resulting part retained at least one *susCD* gene pair.

Prediction of PUL substrates

PULs, at least in *Bacteroidota*, are typically defined as a co-regulated operon or regulon containing a *susC*-like and *susD*-like gene pair that encode the outer-membrane glycan-import machinery, and various CAZymes that are grouped into sequence-based families in the CAZy database [47], as well as accessory proteins (e.g., sulfatases, ABC transporters, etc.). Of course, polysaccharide utilization is not always encoded in one canonical locus, but is sometimes spread out in multiple co-regulated loci [3][4]. This might be exemplified by the larger number of CAZyme-rich gene clusters (CGCs) that we found in our analyses (Fig. S13 in Additional file 2). The number of CAZymes in a PUL may correlate with the complexity of the target polysaccharide: A single CAZyme is usually responsible for the hydrolysis of a specific type of linkage. Thus, a substantial enzymatic arsenal is required to hydrolyze highly complex polysaccharides [19][48]. Moreover, the apparent redundancy in terms of CAZyme families in some PULs may relate to targeting different similar but distinct substructures of a polysaccharide. Predicting specificity based on family (sub-family) classification, however, is challenging, since glycoside hydrolase families are often polyspecific.

We used a three-step procedure to predict possible PUL substrates. First of all, we compared CAZymes gene compositions down to the sub-family level with those in the dbCAN-PUL reference database [49] and with a rough PUL classification used by

Kappelmann *et al.* [1] (Table S8 in Additional file 3). When CAZyme gene compositions were identical to known PULs, the target substrate was also considered to be the same. In a second step, we analyzed endo- and exo-acting CAZymes within PULs (Table S8 in Additional file 3) and enumerated the possible substrates (Table S5 in Additional file 3), with substrates corresponding to endo-acting enzymes as possible polysaccharide backbone, and substrates corresponding to the exo-acting enzyme as possible branched chain oligosaccharides. The third step was to construct a phylogenetic tree using the SusC/D protein sequences of all *susCD* genes in a PUL, together with those published by Kappelmann [1] and Krüger [17] and those in the dbCAN-PUL database. This tree contained sequences from at least 27 bacteroidetal families from draft genomes and MAGs of this study plus their closest relatives (Table S5 in Additional file 3). We used the clustering of these sequences to infer PUL substrates. We also observed that CAZymes close to the *susCD* gene pair tended to be more conserved than those that were located farther away. Hence, we also used such conserved CAZymes in order to improve PUL clustering. Finally, we built a consensus from the three above-mentioned methods to infer the most probable substrate class.

Code Availability

Software used in this study:

AntiSMASH 5.0: https://github.com/antismash/antismash;

Anvio 6.2: https://github.com/merenlab/anvio;

Barrnap: https://github.com/tseemann/barrnap;

BBDuk v35.14: https://github.com/BioInfoTools/BBMap;

BBTools: https://jgi.doe.gov/data-and-tools/bbtools;

BLAST: ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST;

CD-HIT: https://github.com/weizhongli/cdhit;

CheckM: https://ecogenomics.github.io/CheckM;

CONCOCT: https://github.com/BinPro/CONCOCT;

dbCAN2: http://bcb.unl.edu/dbCAN2/index.php;

DADA2: https://github.com/benjjneb/dada2;

dRep : https://drep.readthedocs.io/en/latest/;

DIAMOND:http://www.diamondsearch.org/index.php;

eggNOG-mapper:https://github.com/eggnogdb/eggnog-mapper;

FastANI:https://github.com/ParBLiSS/FastANI;

FastQC:http://www.bioinformatics.babraham.ac.uk/projects/fastqc;

FastTree:http://www.microbesonline.org/fasttree;

FeGenie:https://github.com/Arkadiy-Garber/FeGenie;

GTDB-Tk: https://github.com/Ecogenomics/GTDBTk;

HMMER: http://hmmer.org;

iTOL v6.5.6: https://github.com/iBiology/iTOL;

MAFFT: https://mafft.cbrc.jp/alignment/software;

MaxBin2: https://sourceforge.net/projects/maxbin2;

MEGAHIT: https://github.com/voutcn/megahit;

MetaBAT2: https://bitbucket.org/berkeleylab/metabat;

Prodigal: https://github.com/hyattpd/Prodigal;

Prokka: https://github.com/tseemann/prokka;

R: https://www.r-project.org;

SAMTools: http://www.htslib.org;

SPAdes: https://github.com/ablab/spades;

SignalP v5.0: https://github.com/nextgenusfs/funannotate;

vsearch: https://github.com/torognes/vsearch

References

1. Kappelmann L, Krüger K, Harder J, Markert S, Unfried F, Becher D, et al. Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. ISME J. 2019;13:76–91.
2. Hemsworth GR, Déjean G, Davies GJ, Brumer H. Learning from microbial strategies for polysaccharide degradation. Biochem Soc Trans. 2016;44:94–108.
3. Ficko-Blean E, Préchoux A, Thomas F, Rochat T, Larocque R, Zhu Y, et al. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nat Commun.* 2017;8:1685.
4. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. Polysaccharide utilization loci: Fueling microbial communities. J Bacteriol. 2017;199:1–15.
5. Lapébie P, Lombard V, Drula E, Terrapon N. *Bacteroidetes* use thousands of enzyme combinations to break down glycans. *Nat Commun*. 2019;10: 2043.
6. Finch H. Comparison of Distance Measures in Cluster Analysis with Dichotomous Data. J Data Sci. 2021;3:85–100.
7. Petschacher B, Nidetzky B. Biotechnological production of fucosylated human milk oligosaccharides: Prokaryotic fucosyltransferases and their use in biocatalytic cascades or whole cell conversion systems. J Biotechnol. 2016;235:61–83.
8. Jost F, de Vries T, Knegtel RMA, Macher BA. Mutation of amino acids in the alpha 1,3-fucosyltransferase motif affects enzyme activity and Km for donor and acceptor substrates. Glycobiology. 2005;15:165–75.
9. Biely P. Microbial carbohydrate esterases deacetylating plant polysaccharides. Biotechnol Ad. 2012;30:1575–88.
10. Barbeyron T, Brillet-Guéguen L, Carré W, Carrière C, Caron C, Czjzek M, et al. Matching the diversity of sulfated biomolecules: Creation of a classification database for sulfatases reflecting their substrate specificity. PLoS One. 2016;11:1–33.
11. Song W, Wemheuer B, Steinberg PD, Marzinelli EM, Thomas T. Contribution of horizontal gene transfer to the functionality of microbial biofilm on a macroalgae. ISME J. 2021;15:807–17.

12. Vidal-Melgosa S, Sichert A, Francis T Ben, Bartosik D, Niggemann J, Wichels A, et al. Diatom fucan polysaccharide precipitates carbon during algal blooms. Nat Commun. 2021;12:1–13.
13. Becker S, Tebben J, Coffinet S, Wiltshire K, Iversen MH, Harder T, et al. Laminarin is a major molecule in the marine carbon cycle. Proc Natl Acad Sci. 2020;117:6599–607.
14. 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.
15. Deniaud-Bouët E, Hardouin K, Potin P, Kloareg B, Hervé C. A review about brown algal cell walls and fucose-containing sulfated polysaccharides: Cell wall context, biomedical properties and key research challenges. Carbohydr Polym. 2017;175:395–408.
16. Bilan MI, Grachev AA, Shashkov AS, Nifantiev NE, Usov AI. Structure of a fucoidan from the brown seaweed *Fucus serratus*. *Carbohydr Res.* 2006;341:238–45.
17. Krüger K, Chafee M, Ben Francis T, Glavina del Rio T, Becher D, Schweder T, et al. In marine *Bacteroidetes* the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. ISME J. 2019;13:2800–16.
18. Deniaud-Bouët E, Kervarec N, Michel G, Tonon T, Kloareg B, Hervé C. Chemical and enzymatic fractionation of cell walls from *Fucales*: Insights into the structure of the extracellular matrix of brown algae. Ann Bot. 2014;114:1203–16.
19. Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, et al. *Verrucomicrobia* use hundreds of enzymes to digest the algal polysaccharide fucoidan. Nat Microbiol. 2020;5:1026–39.
20. Bakunina IY, Shevchenko LS, Nedashkovskaya OI, Shevchenko NM, Alekseeva SA, Mikhailov V V., et al. Screening of marine bacteria for fucoidanases. *Microbiology*. 2000;69:303–8.
21. Bengtsson MM, Sjøtun K, Storesund JE, Øvreas L. Utilization of kelp-derived carbon sources by kelp surface-associated bacteria. Aquat Microb Ecol. 2011;62:191–9.

22. Sakai T, Ishizuka K, Kato I. Isolation and Characterization of a Fucoidan-Degrading Marine Bacterium. Mar Biotechnol. 2003;5:409–16.
23. Barbeyron T, L'Haridon S, Michel G, Czjzek M. *Mariniflexile fucanivorans* sp. nov., a marine member of the *Flavobacteriaceae* that degrades sulphated fucans from brown algae. Int J Syst Evol Microbiol. 2008;58:2107–13.
24. Orellana LH, Francis T Ben, Ferraro M, Hehemann J-H, Fuchs BM, Amann RI. *Verrucomicrobiota* are specialist consumers of sulfated methyl pentoses during diatom blooms. ISME J. 2021;1–12.
25. Selvendran RR. Chemistry of plant cell walls and dietary fibre. Scand J Gastroenterol. 1987;22:33–41.
26. Bobin-Dubigeon C, Lahaye M, Guillon F, Barry JL, Gallant DJ. Factors limiting the biodegradation of *Ulva* sp cell-wall polysaccharides. J Sci Food Agric. 1997;75:341–51.
27. De Santi C, Willassen NP, Williamson A. Biochemical characterization of a family 15 carbohydrate esterase from a bacterial marine Arctic metagenome. PLoS One. 2016;11:1–22.
28. Hettiarachchi SA, Kwon YK, Lee Y, Jo E, Eom TY, Kang YH, et al. Characterization of an acetyl xylan esterase from the marine bacterium *Ochrovirga pacifica* and its synergism with xylanase on beechwood xylan. Microb Cell Fact. 2019;18:1–10.
29. Moreira LRS, Filho EXF. An overview of mannan structure and mannan-degrading enzyme systems. Appl Microbiol Biotechnol. 2008;79:165–78.
30. Wu J, Lv Y, Liu X, Zhao X, Jiao G, Tai W, et al. Structural Study of Sulfated Fuco-Oligosaccharide Branched Glucuronomannan from *Kjellmaniella crassifolia* by ESI-CID-MS/MS. J Carbohydr Chem. 2015;34:303–17.
31. Bunse C, Koch H, Breider S, Simon M, Wietz M. Sweet spheres: succession and CAZyme expression of marine bacterial communities colonizing a mix of alginate and pectin particles. Environ Microbiol. 2021;23:3130–48.
32. Coenen GJ, Bakx EJ, Verhoef RP, Schols HA, Voragen AGJ. Identification of the connecting linkage between homo- or xylogalacturonan and rhamnogalacturonan type I. Carbohydr Polym. 2007;70:224–35.

33. Luis AS, Briggs J, Zhang X, Farnell B, Ndeh D, Labourel A, et al. Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides*. Nat Microbiol. 2018;3:210–9.
34. Ciancia M, Fernández PV, Leliaert F. Diversity of Sulfated Polysaccharides From Cell Walls of Coenocytic Green Algae and Their Structural Relationships in View of Green Algal Evolution. *Front Plant Sci.* 2020;11:1–15.
35. Aquino RS, Landeira-Fernandez AM, Valente AP, Andrade LR, Mourão PAS. Occurrence of sulfated galactans in marine angiosperms: Evolutionary implications. *Glycobiology.* 2005;15:11–20.
36. Michel G, Chantalat L, Fanchon E, Henrissat B, Kloareg B, Dideberg O. The ι-carrageenase of *Alteromonas* fortis: A β-helix fold-containing enzyme for the degradation of a highly polyanionic polysaccharide. *J Biol Chem.* 2001;276:40202–9.
37. Campo VL, Kawano DF, Silva DB da, Carvalho I. Carrageenans: Biological properties, chemical modifications and structural analysis-A review. Carbohydr Polym. 2009;77:167–80.
38. Zhang Y, Fu X, Duan D, Xu J, Gao X. Preparation and characterization of agar, agarose, and agaropectin from the red alga *Ahnfeltia plicata*. J Oceanol Limnol. 2019;37:815–24.
39. Qiu Y, Jiang H, Fu L, Ci F, Mao X. Porphyran and oligo-porphyran originating from red algae *Porphyra*: Preparation, biological activities, and potential applications. Food Che. 2021;349:129209.
40. Ran W, Wang H, Liu Y, Qi M, Xiang Q, Yao C, et al. Bioresource Technology Storage of starch and lipids in microalgae : Biosynthesis and manipulation by nutrients. Bioresour Technol. 2019;291:121894.
41. Øverland M, Mydland LT, Skrede A. Marine macroalgae as sources of protein and bioactive compounds in feed for monogastric animals. J Sci Food Agric. 2019;99:13–24.

42. Brányiková I, Maršálková B, Doucha J, Brányik T, Bišová K, Zachleder V, et al. Microalgae-novel highly efficient starch producers. *Biotechnol Bioeng.* 2011;108:766–76.
43. Raimundo I, Silva R, Meunier L, Valente SM, Lago-Lestón A, Keller-Costa T, et al. Functional metagenomics reveals differential chitin degradation and utilization features across free-living and host-associated marine microbiomes. Microbiome. 2021;9:1–18.
44. Sun Y, Debeljak P, Obernosterer I. Microbial iron and carbon metabolism as revealed by taxonomy-specific functional diversity in the Southern Ocean. ISME J. 2021;15:2933–46.
45. Mu DS, Liang QY, Wang XM, Lu DC, Shi MJ, Chen GJ, et al. Metatranscriptomic and comparative genomic insights into resuscitation mechanisms during enrichment culturing. Microbiome.2018;6:1-15.
46. Burke C, Kjelleberg S, Thomas T. Selective extraction of bacterial DNA from the surfaces of macroalgae. *Appl Environ Microbiol*. 2009;75:252–6.
47. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:490–5.
48. Reisky L, Préchoux A, Zühlke MK, Bäumgen M, Robb CS, Gerlach N, et al. A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan. Nat Chem Biol. 2019;15:803–12.
49. Ausland C, Zheng J, Yi H, Yang B, Li T, Feng X, et al. dbCAN-PUL: A database of experimentally characterized CAZyme gene clusters and their substrates. Nucleic Acids Res. 2021;49:D523–8.

Chapter IV

Discussion

Discussion

In Chapter II of this thesis, I explored the diversity, dynamics, and functional potential of polysaccharide utilizing bacteria across three filter size fractions during a diverse spring phytoplankton bloom at Helgoland Roads in the southern North Sea. The analysis aimed at a comparison between FL and PA bacteria in a pelagic environment predominantly influenced by microalgal phytoplankton. In Chapter III of this thesis, I then analyzed bacterial communities and their functionalities in the phycospheres of marine macroalgae that were sampled in the intertidal zone of a reef near the city of Weihai in the Chinese province of Shandong. Here, the focus was to explore phycosphere bacterial communities in comparison with bacterial communities in the surrounding (unfiltered) seawater and sediments. While the microalgae-associated PA bacterial communities from the Helgoland coast and the macroalgal phycosphere communities from the Weihai coast share similarities, such as attached lifestyles and an abundance of algal polysaccharides, differences exist, including variations in the types of algal polysaccharides. After a brief recapitulation of some aspects of both studies, I here aim at a more holistic view on both communities by comparing the dominant bacterial taxa and their functions.

The 2018 spring phytoplankton bloom at Helgoland Roads consisted of distinct phases that were dominated by different phytoplankton taxa. It commenced with a diatom-dominated first phase at the end of which *Chattonella* raphidophytes emerged for a comparably short time. A significant second bloom phase was largely dominated by *Phaeocystis* sp. haptophytes. Photosynthetic dinoflagellates, although relatively low in total biovolume, were present throughout the bloom, in particular during the *Phaeocystis* bloom phase. This *Phaeocystis* bloom reached a peak density of 8.98×10^6 cells/L (Chapter II), which was notably higher than peaks levels observed in 2009 (1.09×10^6 cells/L) [123], 2010 (1.69×10^6 cells/L) [123], 2011 (1.67×10^6 cells/L) [123], and 2012 (8.89×10^5 cells/L) [123] as well as 2016 [211] and 2020 [127] without noteworthy *Phaeocystis* numbers. Another difference was the high diversity and complexity of the phytoplankton composition during the 2018 spring bloom as compared to those in the

aforementioned other years, especially the 2020 spring bloom, where the total algal biovolume was almost entirely dominated by few large-celled diatom species [127].

Marine macroalgae play a crucial role as primary producers in shallow nearshore marine ecosystems. Associated bacteria, including epiphytic bacteria as well as bacteria in the surrounding seawater and sediments, are involved in algal biomass recycling. Currently, there is a limited amount of research regarding these bacteria, particularly in terms of functionality. In the study described in Chapter III, samples were collected once during all four seasons from, in total, four species of representative marine red (*Grateloupia*, *Gelidium*), brown (*Saccharina*), and green (*Ulva*) macroalgae. The sampling site resides on a marine peninsula that is bordered by the Yellow Sea to the north. There are no river estuaries, and the site is distant from urban areas, so there are no immediate human influences, such as sewage discharges. Macroalgal samples were collected from rocks close to the shoreline in this sea area. These rocks are submerged by seawater during high tide and partially emerge above sea level at low tide, thus experiencing significant exposure to ultraviolet radiation. The area is characterized by frequent water exchanges and experiences wave action against the rocks.

In the following sections, I will discuss some commonalities and differences between the attached bacterial communities that were sampled during the studied spring phytoplankton bloom and from the four studied macroalgal species.

4.1 Comparison of the microbiota between the studies of Chapters II and III

4.1.1 Prominent bacteria in phytoplankton bloom-associated particle-attached and marine macroalgal phycosphere bacterial communities

Having 16S rRNA gene amplicon data at high temporal resolution from three filter fractions during an almost complete phytoplankton bloom enabled to study differences between free-living (0.2-3 μm) and particle-attached (3-10 μm, >10 μm) bacteria over time. For a first comparison of these data with the data obtained from macroalgal phycospheres, I selected data of PA bacteria from the diatom and *Phaeocystis* bloom periods for analysis (April 10 to May 31). I considered the top 15 genera for at least one sample, resulting in 55 genera from the PA3 fraction (3-10 μm) and 86 genera from the

PA10 fraction (>10 μm). I disregarded seasonal factors for the macroalgal phycosphere bacteria and analyzed combined data from all macroalgal phycosphere samples. Similarly, I selected the top 15 genera based on relative abundances for each sample. This resulted in a total of 110 genera from macroalgal phycosphere samples. A comparison of these two datasets of 107 prominent microalgae-associated PA genera and 110 prominent macroalgal phycosphere genera revealed 26 shared genera (Table 1). The relative abundances are shown in Figure 4.1 of this thesis.

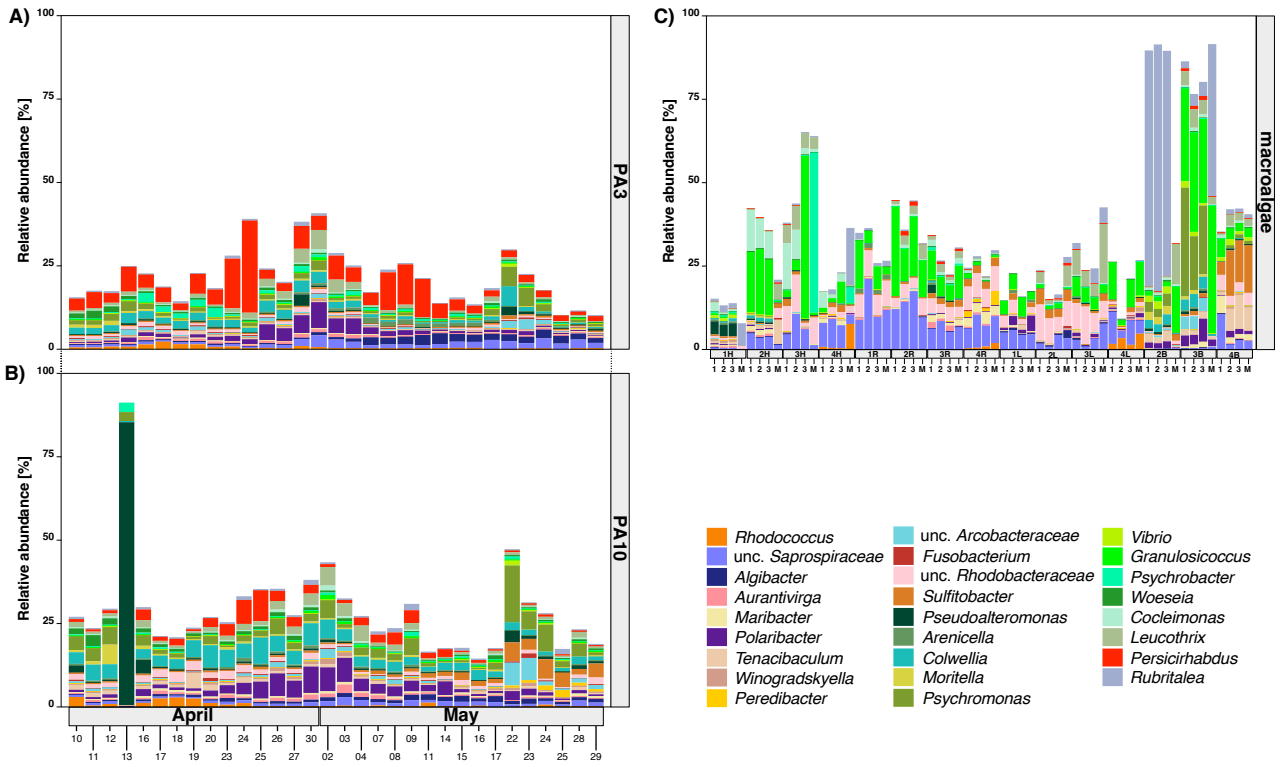


Figure 4.1. Relative abundances of 26 common genera among all prominent genera of the studied microalgae-associated and macroalgal phycosphere communities. **A)** Relative abundances of the 26 common genera in the studied microalgae-associated PA3 fractions (3-10 μm). **B)** Relative abundances of the 26 common genera in the studied microalgae-associated PA10 fractions (>10 μm). **C)** Relative abundances of the 26 common genera in the studied macroalgal phycospheres.

4.1.1.1 *Flavobacteriaceae* were the most prominent bacteria in both communities

The *Flavobacteriaceae* (class *Flavobacteriia*) constitute one of the largest families within the phylum *Bacteroidota*, and many family members are known for a rich repertoire of CAZymes and PULs (e.g., [212]). These members are renowned polysaccharide degraders that play a crucial role in marine carbon and other biogeochemical cycles. Many members of the *Flavobacteriaceae* establish close connections with both microalgae and macroalgae. During phytoplankton blooms, *Flavobacteriaceae* are typical among the first responders and progressively increase in relative abundance, since these bacteria actively participate in the degradation of phytoplankton polysaccharides [123]. In macroalgal phycospheres, members of the *Flavobacteriaceae* represented a significant proportion of the bacterial community and included some of the most prominent polysaccharide degraders (Chapter III). A manuscript (currently under review) to which I contributed corroborates these findings (Appendix - Additional co-author publication 1). It identified novel laminarin-binding CBMs in the model strain *Christiangramia forsetii* KT0803^T [323] (formerly 'Gramella forsetii' KT0803 [224]). The study further illustrated that these newly discovered CBMs were highly expressed during phytoplankton blooms, highlighting their significant role in the degradation of laminarin. Common prominent flavobacterial genera among the communities studied in Chapters II and III include *Algibacter*, *Aurantivirga*, *Maribacter*, *Polaribacter*, *Tenacibaculum*, and *Winogradskyella* (Table 1). *Algibacter* is a genus that is rich in CAZymes [324]. Members have been isolated from various marine environments, including seawater, sediments, and various marine macroalgae, including red algae, green algae, and brown algae [325-329]. Both, a study from Heins *et al*. [111] and the study presented in Chapter II have shown that during the 2018 spring bloom at Helgoland Roads, *Algibacter* species were part of the PA bacterial communities and were particularly abundant towards the end of the diatom and throughout the *Phaeocystis* bloom phases. Both are indicative of a strong correlation with the presence of microalgae. As shown in Chapter III, the relative abundance of *Algibacter* was significantly higher in the studied green algae species compared to the red and brown algae species. Akin to *Flavobacteriaceae* strains that have been reported to possess complete PULs, such as *Zobellia galactanivorans* DsiJ^T [330, 331], *Christiangramia*

forsetii KT0803^T [224, 332], 'Formosa agariphila' KMM 3901^T [333] and *Polaribacter* sp. strains Hel1_33_49 and Hel1_85 [334], the draft genome of *Algibacter* sp. 4-1052, isolated from the studied green macroalgal species, showed a preference for complex algal polysaccharides. This *Algibacter* strain possesses a very long and complete PUL or PUL-rich region with approximately 100 CAZymes that, based on annotations, is assumed to target complex macroalgal polysaccharides of as yet unknown composition. In total, 15 complete PULs were predicted for this *Algibacter* strain. In comparison to other bacteria with rich polysaccharide degradation potential, research on *Algibacter* is rather sparse. Thus, the cultivated *Algibacter* sp. strain 4-1052 deserves more in-depth and comprehensive research to elucidate the roles of its large number of encoded CAZmyes.

Aurantivirga is not only recurrent in the FL fraction during phytoplankton blooms at Helgoland Roads [123, 127, 212] but also among responders to phytoplankton blooms in polar waters [269, 335]. It is known that *Aurantivirga* members are active polysaccharide degraders [127, 212]. *Aurantivirga* members have also been found in enrichments from samples related to macroalgae and biofilms [336, 337]. This is consistent with the results that I obtained in Chapters II and III. There is limited research on the functionality of these *Aurantivirga* members apart from polysaccharide degradation [127]. Currently, there is no study on the lifestyle of any *Aurantivirga* species. Based on the results from Chapters II and III, I can only deduce that *Aurantivirga* likely comprises species with either preferred FL or PA lifestyles.

Many studies support the notion that members of *Maribacter* adopt a mainly surface-attached lifestyle. Some *Maribacter* species have been isolated from microalgae-derived particles [338] and marine macroalgae, including red (e.g., [339, 340]), green (e.g., [341]) and brown algae (e.g., [264]). *Maribacter* species are either non-motile (e.g., [342]) or rely on characteristic bacteroidetal gliding for motility (e.g., [341]). This agrees with a surface-attached lifestyle [343, 344]. *Maribacter* species have been reported to invade PA communities on chitin particles [345]. Members of *Maribacter* are known to be rich in CAZymes [346, 347] and are known for active polysaccharide-degradation on the surfaces of macroalgae [264]. As described in Chapter III, *Maribacter* had higher relative abundances on red and brown algae than on green algae. *Maribacter* members are also

known for engaging in intricate interactions with both other bacteria and with their algal hosts. Cross-feeding interactions were observed between *Maribacter* species and primary consumers such as *Vibrio* and *Pseudoalteromonas* during chitin consumption [345]. The relationship between *Maribacter* and macroalgae is multifaceted and encompasses the consumption of macroalgal polysaccharides [264] and the induction of *Ulva* morphogenesis [348-350]. *Maribacter* may employ a sophisticated mechanism for utilizing complex polysaccharides, possibly involving collaborative interactions with other bacteria. This is a compelling area for research, as studies often focus on the pathway of individual bacterial strain in polysaccharide degradation. However, it is crucial to recognize that the utilization of polysaccharides frequently involves synergistic cooperation among multiple bacteria. *Maribacter* species could serve as model microorganisms for investigating this topic.

Polaribacter is a recurrent FL bacterial clade during phytoplankton blooms at Helgoland Roads that often reaches double-digit relative abundances during and after spring bloom periods [123, 127]. However, in the data from 2018, I did not observe such high *Polaribacter* relative abundances. This suggests that *Polaribacter* is not irreplaceable and that its position may have been taken over by other bacterial groups with similar functions during the studied 2018 spring bloom (Chapter II). Instead, I observed the same *Polaribacter* at lower abundances in both the FL and PA fractions. This means that some *Polaribacter* members can live both FL and PA lifestyles, but the preferential mode of the genus as a whole has not yet been determined [110]. There could be a switching pattern between PA or FL lifestyles among some *Polaribacter* members depending on environmental conditions [351, 352]. *Polaribacter* has been reported to be associated with diatoms (e.g., [123, 353]) and are members of diatom phycospheres (Xiaoyu Yang, Shantou University, pers. commun.). Certain bacterial species can promote diatoms, and specifically, *Polaribacter* has been reported to possibly enhance the sexual reproduction of diatoms [354]. A notable discovery of the PA *Polaribacter* MAG in Chapter II is the identification of complete PULs targeting fucoidan, a feature uncommonly observed within *Bacteroidota*.

It's known that many *Tenacibaculum* members are opportunistic pathogens in fish, as the genus encompasses 32 different species with eight being linked to fish diseases [355], in particular ulcerative tenacibaculosis. *Tenacibaculum* members are also positively correlated to phytoplankton blooms [123, 346, 356] and have been considered as PA bacteria during such booms [111, 357]. This is corroborated by the results presented in Chapter II, in which I found that *Tenacibaculum* members affiliated with PA bacterial communities. *Tenacibaculum* species have also been isolated from macroalgal samples (e.g., [358]) and can be enriched from decaying macroalgae [359, 360]. Some *Tenacibaculum* species are closely related to polysaccharide degradation [275, 346]. Only non-pathogenic *Tenacibaculum* members seem to possess this potential, whereas pathogenic *Tenacibaculum* species cannot degrade polysaccharides even when they carry CAZyme genes [361, 362]. In addition to the potential for utilizing polysaccharides, experiments have demonstrated that members of *Tenacibaculum* can exploit public goods generated by other bacteria when consuming fresh macroalgae [363]. This is probably one of the reasons why *Tenacibaculum* was abundant within the studied macroalgal phycosphere bacteria. In Chapter III, *Tenacibaculum* exhibited a preference for specific algal species, as evidenced by its higher relative 16S rRNA gene sequence abundance in the phycosphere microbiota of brown *Saccharina* algae compared to *Gelidium* sp. red algae. In contrast, negligible *Tenacibaculum* sequence numbers were detected in the phycospheres of *Grateloupia* sp. red algae and *Ulva* sp. green algae. Furthermore, a seasonal preference could be observed. Although seasonality is not a focus of this discussion, I observed that the *Tenacibaculum* relative abundance on *Saccharina* brown algae was notably higher during the summer season compared to all other seasons. It is worth noting that the brown algae samples collected during the summer were already in a state of decay, confirming the aforementioned enrichments of *Tenacibaculum* from decaying macroalgae [359, 360].

Winogradskyella species have been frequently isolated from samples associated with algae, such as microalgae-derived particles [338], macroalgae-derived seawater [364] and macroalgae themselves [365]. The genomes of *Winogradskyella* species are rich in CAZyme and peptidase genes and contain predicted PULs [366]. I also found some

species of *Winogradskyella* rich in sulfatase genes (Chapter III). Notably, different species within the *Winogradskyella* genus exhibit variations in their PUL composition [366], again confirming that members of the same genus can vary considerably in terms of PUL repertoires (see Chapter II). A *Winogradskyella* strain isolated from *Ulva* has been shown to possess ulvan lyase activity, showcasing its capability to degrade green algal ulvan [367]. Additionally, *Winogradskyella* strains have been shown to possess agarolytic activity [368], indicating the potential to degrade agar, a polysaccharide derived from, in particular, red macroalgae. Experimental evidence further supports the notion that *Winogradskyella* strains can be algicidal, as it has been shown that *Winogradskyella* strain B2901 can cause the collapse of microalgal *Ostreococcus* green algae cultures [369]. The results from Chapters II and III confirmed a surface-associated lifestyle for *Winogradskyella*. In the study of Chapter III, PULs targeting laminarin, starch, and alginate were predicted in the genomes of *Winogradskyella* strains originating from red, brown, and green algae. However, PUL-like structures targeting sulfated fucoidan were only predicted in the genomes of *Winogradskyella* strains derived from brown algae. This is logical, considering that among macroalgae, fucoidan is a polysaccharide unique to brown algae.

4.1.1.2 Potential isoprene degraders of the genus *Rhodococcus* were prominent in both PA and macroalgal phycosphere communities

Rhodococcus species are actinobacteria that are frequently detected in or isolated from samples associated with macroalgae, and, in some instances, they were found in high abundance [370-372]. *Rhodococcus* sp. strain AD45 has been extensively studied for its role as a prolific isoprene degrader [373]. This strain was isolated from a freshwater sediment and, as of to date, represents the best-documented isoprene degrader [374]. Isoprene (2-methyl-1,3-butadiene) is one of the most significant volatile organic compounds entering the atmosphere, with fluxes comparable to those of methane [375, 376]. Terrestrial plants are a major source of isoprene that contribute approximately 90% of the total release [377]. The remaining 10% include contributions from marine phytoplankton and macroalgae [378, 379] and some bacteria [380, 381]. While on a global

scale, only a small amount of isoprene comes from marine micro- and macroalgae, the relevance of isoprene in the local habitats that algae-associated bacteria inhabit should not be underestimated. *Rhodococcus* members have been detected in both micro- and macroalgae-related samples, indicating the potential release of isoprene from both. This corroborates previous reports of isoprene production by red and green algae, whereas there are so for no such reports for brown algae [382]. This is strongly supported by our findings of notable *Rhodococcus* relative abundances only in the phycosphere communities of *Grateloupia* and *Gelidium* red and *Ulva* green algae but not in *Saccharina* brown algae. Likewise, I observed *Rhodococcus* only during the diatom bloom phase of the 2018 spring phytoplankton bloom at Helgoland Roads, and the *Rhodococcus* members disappeared with the diatom bloom's decline. This agrees with a report that the diatom species *Chaetoceros calcitrans* produces isoprene [383].

4.1.1.3 The alphaproteobacterial *Rhodobacteraceae* member *Sulfitobacter* was prominent in both communities

Rhodobacteraceae have been found in various marine habitats, such as seawater (e.g., [384]), sediments (e.g., [385]), algae (e.g., [386]) and algae-associated biofilms [387]. *Rhodobacteraceae* members are primary colonizers of particles [388, 389] and have been identified as core bacteria in the diatom phycospheres [390] and macroalgae [391]. Interactions of *Rhodobacteraceae* with both microalgae (e.g., diatoms) [392, 393] and macroalgae (e.g., brown algae) [394, 395] have been reported. Members of the *Rhodobacteraceae* genus *Sulfitobacter* are known for coexisting as well as pathogenic relationships with algae. One well-studied example is *Sulfitobacter* sp. strain D7 and the wide-spread coccolithophore *Emiliania huxleyi* [396], a major producer of DMSP. Analysis of bacterial transcriptomes has validated a signaling role of algal DMSP in mediating the transition of bacteria towards pathogenicity [396]. During coexistence, *Emiliania huxleyi* provides benzoate and other growth substrates to *Sulfitobacter* sp. strain D7. Under nutrient-limiting conditions, algae released more DMSP [397], which may signal the deterioration of the physiological state of the algal host to surrounding bacteria. In response, *Sulfitobacter* sp. strain D7 upregulated its flagellar movement and various

transport systems and turned pathogenic against the compromised *Emiliania huxleyi* population. This relationship could extend to *Phaeocystis*, as it is also known to release DMSP [398].

I discovered that CAZyme-rich gene clusters targeting starch are prevalent in the genomes of *Rhodobacteraceae*. For instance, among the 44 genera of *Rhodobacteraceae* generated in Chapter III, 25 genera had genomes or MAGs containing starch-degrading CAZyme-rich gene clusters. This agrees with the fact that starch is the storage polysaccharide in red and green algae. However, an abundance of CAZyme-rich gene clusters for starch degradation was not detected in *Rhodobacteraceae* in Chapter II. This could be due to the significant species variation within *Rhodobacteraceae* across the two communities, with only macroalgal-associated environments being enriched with *Rhodobacteraceae* capable of starch degradation. On the other hand, it may also be related to the quality of the MAGs, as many (233/305) of the genomes of *Rhodobacteraceae* in Chapter III are from draft-sequenced cultivable strains. In the genomes of *Sulfitobacter*, I detected PUL-like gene clusters for peptidoglycan. Additionally, the presence of PUL-like gene clusters containing PL22 was widespread in *Rhodobacteraceae* (42/44), which may be associated with a scavenging lifestyle of bacterial remnants. However, further research is needed to substantiate this hypothesis, also considering that some genera within the *Rhodobacteraceae*, including *Sulfitobacter*, may actually be polyphyletic [399].

4.1.1.4 *Gammaproteobacteria* contributed the highest number of genera to both communities

Gammaproteobacterial genera contributed the highest numbers of common genera among the prominent genera in both microalgae-associated and macroalgal phycosphere communities (11 genera), even ahead of the *Bacteroidota* (7 genera). This agrees with the high diversity of gammaproteobacterial genera in the PA communities that was demonstrated in Chapter II. These genera comprised *Pseudoalteromonas*, *Arenicella*, *Colwellia*, *Moritella*, *Vibrio*, *Granulosicoccus*, *Psychromonas*, *Psychrobacter*, *Woeseia*, *Cocleimonas*, and *Leucothrix*. This is consistent with previous reports [111, 400-410]. The

presence of *Pseudoalteromonas* [338], *Arenicella* [411], and *Colwellia* [338] during phytoplankton blooms has been described before, possibly due to their ability to degrade polysaccharides.

Pseudoalteromonas is a genus with 49 validly named species to date (https://lpsn.dsmz.de/genus/pseudoalteromonas). Species of marine *Pseudoalteromonas* are commonly associated with marine algae with diverse and complex relationships that include bacteriolytic, agarolytic and algicidal activities [412]. Members of marine *Pseudoalteromonas* have been experimentally verified to degrade pectin and alginate [413, 414]. Pectin is present in red, green and brown macro- as well as in microalgae, whereas alginate is most common in brown macroalgae. The research presented in Chapter III corroborates this aspect. In the genomes examined in Chapter III, I identified PUL-like gene clusters targeting alginate, agar, fucoidan, xylan, and starch. Nevertheless, high relative abundances of *Pseudoalteromonas* were only found in few samples during the studies of Chapters II and III. The reasons for this outcome are currently not known, but it seems that *Pseudoalteromonas* is for the most part outcompeted, possibly by members of the *Flavobacteriia*. Members of *Pseudoalteromonas* are prolific producers of extracellular products that are involved in biofilm formation [415] and the promotion of settlement of invertebrate species [416]. While an increase in phytoplankton mortality has been reported due to *Pseudoalteromonas* spp. [417, 418], this would be challenging to confirm based on the *in situ* data presented in Chapter II.

Few studies describe *Arenicella* members as PA bacteria*,* while others have described *Arenicella* members as macroalgae-associated core bacteria [404]. The latter have the potential to degrade alginate and structural polysaccharides of many macroalgae and to synthesize vitamin B_{12} [404]. Vitamin B_{12} biosynthesis may be part of a mutually beneficial relationship (symbiosis) with macroalgae, since many macroalgae are auxotrophic in terms of vitamin B_{12}*.* Genome analyses of *Arenicella* in Chapters II and III affirmed the capability for starch degradation. However, only the genomes and MAGs of *Arenicella* in Chapter III exhibited PUL-like gene clusters for alginate degradation, which is logical considering alginate is a cell wall polysaccharide of macroalgae. Members

of *Colwellia* have also been confirmed as alginate and pectin degraders *in vitro* [275]. This was corroborated by the *in situ* data presented in Chapter II. In contrast, there are no dedicated reports that would indicate the capability of *Moritella* to degrade specific macroalgal compounds, but research has shown that the addition of macroalgae promotes the enrichment of *Moritella* [405]. Members of the genus *Vibrio* are widespread in marine environments. Vibrio species are known to play an important role in the carbon cycle by using polysaccharides [419]. Both FL and PA *Vibrio* were detected *in situ* [420]. It has been reported that members of *Vibrio* can switch between FL and PA lifestyles [419]. Abundant *Vibrio* have been detected in brown algae microbiota, which was confirmed by the data presented in Chapter III, where *Vibrio* were also only abundant in the phycosphere communities of *Saccharina* brown algae. *Granulosicoccus* is a macroalgal symbiont that has been suggested to have the potential to promote the growth of its host by vitamin B_{12} synthesis [404]. Only CAZyme-rich gene clusters targeting starch were detected in genomes and MAGs in Chapter III. No MAG of *Granulosicoccus* was retrieved in the study presented in Chapter II and only low abundances of *Granulosicoccus* were reported in PA fractions. However, one should not forget that no less than 322 macroalgal species have been identified on Helgoland rocky shores [276]. Therefore, there will be a background of particles from macroalgae during phytoplankton blooms, which could also explain why *Granulosicoccus* members were detected in the data obtained at Helgoland Roads. This would explain, why *Granulosicoccus* were mainly found in the >10 μm fraction. Members of *Psychromonas* are known to perform well in both PA and macroalgal phycosphere bacterial communities [401, 421], which was corroborated by my own studies. In the study of Chapter III, the presence of PULs or CAZyme-rich gene clusters, essential for alginate degradation was also predicted for genomes and MAGs of *Psychromonas*. This is consistent with previous reports [275, 421]. *Psychrobacter* members have also been found in association with North Sea phytoplankton bloom events in PA filter fractions [111, 422]. This is consistent with my results in Chapter II. *Psychrobacter arcticus* strain 273-4 can form biofilms, which agrees with the PA lifestyle [423]. In the study of Chapter III, few CAZyme-rich gene clusters were identified in *Psychrobacter*. However, a relatively higher number of glycosyltransferase

(GT) gene clusters was discovered. The possible role of *Psychrobacter* in macroalgal phycospheres is not well established. *Psychrobacter* was only found among the topmost 15 genera in a single sample in the data presented in Chapter III. It cannot be excluded that these bacteria were brought in from surrounding particles. *Woeseia* members are globally abundant in sea sediments [424, 425], where they take part in the decomposition of detrital proteins [424]. As discussed, the Helgoland sampling site is shallow and susceptible to sediment influence. A previous study has shown that *Woeseia* was abundant in sediments around Helgoland [269], which supports the notion that the PA *Woeseia* described in the study of Chapter II may indeed originate from the sediment. *Woeseia* has been reported to thrive during seaweed cultivation [336], which is consistent with the results in the study of Chapter III. However, the water depth at the sampling site off Weihai is also relatively shallow, and waves occasionally mix the seawater with surface sediments. In Chapter III, *Cocleimonas* was predominantly abundant on *Gelidium* red algae and rarely detected in the other algae samples. However, abundant *Cocleimonas* have been detected in *Ulva*-associated bacterial communities [426]. This indicates that the selection of algal species by *Cocleimonas* may be associated with the geographical location, the exact species, or the physiological state of the algal host. In Helgoland PA communities, *Cocleimonas* appeared among the topmost 15 genera in the PA10 fraction of several samples. As discussed before, this fraction may contain non-negligible amounts of particles from macroalgae. Studies have revealed sulfur-oxidation in *Cocleimonas* species, which contributes to marine sulfur cycling [427, 428]. As for *Leucothrix*, numerous studies have shown that *Leucothrix* is consistently found in the epiphytic microbiota of macroalgae [256, 409, 410, 429]. *Leucothrix* species constitute a group of large filamentous, sulfur-oxidizing, chemoheterotrophic *Gammaproteobacteria* [430]. Thomas *et al.* have assumed that *Leucothrix* species benefited from alginate degradation by other bacteria by scavenging intermediary degradation products [431]. *Leucothrix* has the capability to form thick biofilms that are sometimes visible to the naked human eye [432, 433]. Such biofilms can even form on fish eggs and may also have implications for fish reproduction.

4.1.1.5 Two clades of *Verrucomicrobiota* were prominent in both communities

Members of the *Verrucomicrobiota* are well-known for their capability to degrade complex polysaccharides and the presence of abundant sulfatases [156, 233, 434]. It is therefore not surprising that *Verrucomicrobiota* have been found in co-enrichments of both PA and macroalgal phycosphere communities [400, 435]. A study from 2012 on the global distribution and diversity of *Verrucomicrobia* revealed that in the data available at that time the highest proportions of *Verrucomicrobia* were detected in samples from PA3 (3-10 μm) filter fractions obtained near Helgoland island [435]. The findings in Chapter II of my research corroborate this insofar as *Verrucomicrobia* were more abundant in the PA than the FL filter fractions, specifically with respect to the genera *Persicirhabdus* and *Rubritalea*. However, while both of these genera belong to the *Rubritaleaceae* family, which is known for biofilm-forming bacteria [436], the relative abundances of both genera differed significantly in the studied PA and macroalgal phycosphere communities. In PA bacterial communities, *Persicirhabdus* exhibited much higher relative abundances than *Rubritalea*, whereas in macroalgal phycosphere communities, it was vice versa. In addition, relative abundances of *Rubritalea* in macroalgal phycospheres notably depended on host species and season, which could not be observed for *Persicirhabdus*. Winter samples of *Saccharina* brown algae were particularly rich in *Rubritalea*, whereas *Persicirhabdus* relative abundances showed only little variation among the sampled red, green, and brown algae. While the data obtained thus far is still too sparse for generalizations, this could indicate that *Persicirhabdus* members preferentially colonize particles derived from microalgae, while *Rubritalea* members preferentially colonize surfaces of, in particular, brown macroalgae. As we discussed above, the background of particles from macroalgae during phytoplankton blooms (also supported by abundantly expressed alginate lyases during such bloom) could also explain why, in particular, *Rubritalea* members were detected in the data obtained at Helgoland Roads.

The conclusion that members of *Persicirhabdus* are PA bacteria agrees with prior studies [111, 435, 437]. In line with the results in Chapter III, *Rubritalea* members are frequently identified as dominant community members on brown algae [438, 439]. Likewise, *Rubritalea* has also been recognized as one of the dominant community

members on *Ulva* sp. green algae [440, 441]. This is relevant, since *Rubritalea* members have a demonstrated potential to promote macroalgal growth via the production of squalene [442, 443], a precursor to steroids and D vitamins [444]. Both steroids and D vitamins can stimulate macroalgal growth [445]. The genome of *Persicirhabdus* is rich in sulfatase and CAZyme genes, in contrast to *Rubritalea*, which lacks such genes. This discrepancy further highlights the distinct ecological roles of the two in macroalgal-associated environments.

Table 1 List of shared genera from the sum of the topmost 15 genera present in all bloom-associated samples of PA bacterial communities and macroalgal bacterial communities of the studies presented in Chapters II and III.

Taxa	**Genus**
Actinobacteriota	*Rhodococcus*
Bacteroidota	unclassified *Saprospiraceae*
	Algibacter
	Aurantivirga
	Maribacter
	Polaribacter
	Tenacibaculum
	Winogradskyella
Bdellovibrionota	*Peredibacter*
Campylobacterota	unclassified *Arcobacteraceae*
Fusobacteriota	*Fusobacterium*
Alphaproteobacteria	unclassified *Rhodobacteraceae*
	Sulfitobacter
Gammaproteobacteria	*Pseudoalteromonas*

	Arenicella
	Colwellia
	Moritella
	Vibrio
	Granulosicoccus
	Psychromonas
	Psychrobacte
	Woeseia
	Cocleimonas
	Leucothrix
Verrucomicrobiota	*Persicirhabdus*
	Rubritalea

4.1.2 Clades that were prominent in macroalgal phycosphere but not in PA communities

The comparison of prominent phytoplankton-associated PA genera and prominent macroalgal phycosphere genera (see Section 4.1.1) revealed that 83 genera were exclusive to macroalgal phycospheres. Among these, 18 genera belonged to the phylum *Bacteroidetes*, 23 to *Alphaproteobacteria*, and 18 to *Gammaproteobacteria*. The remaining details are summarized in Table 1 of the Appendix. Among those 83 genera, *Algitalea* (*Flavobacteriaceae*), an unidentified genus of the *Rhizobiaceae*, two gammaproteobacterial genera including *Psychrobium* and *Acinetobacter*, and *Roseibacillus* (*Verrucomicrobiota*) stood out as the most prominent. Despite not being very prominent in relative abundance, members of the *Saprospiraceae* are also closely associated with macroalgae. While one unclassified *Saprospiraceae* genus was shared between both datasets (Table 1), the other three genera—*Portibacter*, *Lewinella*, and *Rubidimonas*—were not only more abundant but also exclusive to macroalgal phycospheres. This underscores the prominent status of *Saprospiraceae* in the macroalgal phycosphere bacterial community, as illustrated in Figure 3 of Chapter III.

Significantly, within the *Alphaproteobacteria*, the *Sphingomonadaceae* family encompasses one unidentified genus along with three known genera—*Altererythrobacter*, *Erythrobacter*, and *Sphingomonas*. Despite their modest relative abundance in macroalgal samples, these genera are distinctly identified as macroalgal phycosphere-prominent genera. The *Sphingomonadaceae* family does not have a substantial presence in PA bacterial communities. In the subsequent subsections, I will delve into a detailed discussion of the genera mentioned above.

4.1.2.1 Five dominant genera in macroalga-phycosphere communities

Members of the genus of *Algitalea* were initially discovered and described on the surface of the green algae *Ulva pertusa* [446]. *Algitalea* is abundant in the epiphytic microbiota of red, green, and brown algae [409], which was confirmed by the results in the study of Chapter III. It shows that, compared to brown algae, members of *Algitalea* have higher relative abundances in the bacterial communities of green and red algae, reaching up to about 50%. Currently, there are no studies on the functions and ecological role of *Algitalea* members. The *Algitalea* strains and genomes isolated from marine macroalgae in Chapter III offer potential for further research. I discovered that *Algitalea* possesses an abundance of complete PULs and CAZyme-rich gene clusters for polysaccharide degradation. For instance, PULs targeting starch are commonly found in *Algitalea*. PULs for degrading alginate, sulfated rhamnose, ulvan, and agar were also identified in the genomes of *Algitalea* strains. The functionality of these PULs awaits further experimental validation.

Many members of the *Rhizobiaceae* have symbiotic relationships with plants but are also found in macroalgae epiphytic bacteria and may have a beneficial effect on the healthy growth of macroalgae [447]. *Rhizobiaceae* members were detected as epiphytic bacteria which agrees with the study in Chapter III, where I found that unclassified members of *Rhizobiaceae* showed high relative abundance in macroalgae red and green phycosphere communities. *Psychrobium* species are known to utilize alginate [275, 421], a polysaccharide chiefly associated with brown algae. This utilization is consistent with their notably high relative abundances in brown algal phycosphere bacterial communities,

as documented in Chapter III. The presence of PULs or CAZyme-rich gene clusters, essential for alginate degradation, was also predicted in a MAG of *Psychrobium* (Chapter III). While *Acinetobacter* may not be abundant in CAZyme genes (results in the study of Chapter III), it has been previously reported to possess UV-resistance [448], a trait that could offer a significant survival advantage in the specific environmental conditions of the sampling sites described in Chapter III. *Roseibacillus* members were previously documented as prevalent on brown algae [439]. However, in the research of Chapter III, I observed that *Roseibacillus* members exhibited high relative abundances in red and green algae, with a rare presence in brown algae. Additionally, species of *Roseibacillus* are not characterized by a high content of CAZyme genes, as noted in Chapter III.

4.1.2.2 The family *Saprospiraceae* holds particular interest in macroalgal phycosphere communities

The *Saprospiraceae* are a family of *Bacteroidota*, whose members exhibit diverse morphologies, including filamentous and rod-shaped forms [449]. In marine habitats, *Saprospiraceae* are typically found in surface-attached communities [450], such as biofilms [451, 452]. Members of *Saprospiraceae* have also been identified as PA bacteria [453] and are frequently observed during algal blooms [454], where they have reported interactions with diatoms [455]. My results confirm these observations. In Chapter II, some *Saprospiraceae* members were identified as PA bacteria and showed increased abundance during the late diatom and *Phaeocystis* bloom phases. Certain genera within this family have been reported to possess algicidal or microalgal predation capabilities [456, 457]. This reveals that, compared to their presence in PA bacterial communities, *Saprospiraceae* hold a more prominent position within macroalgal phycosphere communities. This prominence is especially notable in the three genera mentioned above, *Portibacter*, *Lewinella*, and *Rubidimonas*, which were reported as dominant groups in epiphytic bacteria of macroalgae previously [458-460]. They are likely important in the hydrolysis and utilization of complex carbon sources [449]. The related genome analysis in Chapter III corroborates this conclusion. I discovered that the MAGs of these three genera are rich in CAZyme genes. For instance, this particular *Lewinella* MAG

encompasses 27 CAZyme-rich clusters within its genome, nine of which are complete PULs.

In Chapter III, *Saprospiraceae* exhibited high relative abundances in the phycosphere communities of *Grateloupia* and *Gelidium* red and *Ulva* green macroalgae. Minor relative abundances of *Saprospiraceae* were also detected in *Saccharina* brown algal phycosphere communities. This agrees with previous reports that *Saprospiraceae* are among the core microbial symbionts associated with macroalgae [459, 461]. Members of the *Saprospiraceae* are also known for high relative abundances in sludge wastewater treatment systems [449, 462]. They may play a vital role in the breakdown of complex carbon sources [449, 462, 463]. There is limited research on the degradation of polysaccharides by *Saprospiraceae*, particularly those originating from marine environments. In Chapters II and III, I observed that the genomes of *Saprospiraceae* from microalgae bloom-associated particles and from macroalgal phycospheres are often larger (8.6 Mbp) and contain abundant CAZyme genes, suggesting a significant potential for polysaccharide degradation. Hence, members of *Saprospiraceae* represent a promising subject for more in-depth investigations, especially with respect to polysaccharide degradation capabilities and interactions between bacteria and algae. In the study of Chapter II, high-quality MAGs of *Saprospiraceae* were acquired, and in the study of Chapter III, both high-quality MAGs and draft genomes of strains were obtained, providing a solid foundation for such future analyses.

4.1.2.3 *Sphingomonadaceae* species bear the potential for antioxidant production and polysaccharide degradation

Numerous *Sphingomonadaceae* members are associated with the green algae *Ulva* sp. [464]. Notably, *Erythrobacter* emerged as the predominant phycosphere genus for both red and green macroalgae, while *Sphingomonas* was observed in association with red, green, and brown macroalgae in Chapter III. Additionally, studies have identified members of the *Sphingomonadaceae* in bacterioplankton and on biofilms on microplastics [465, 466]. The production of Reactive Oxygen Species (ROS) is a recognized phenomenon in habitats dominated by macroalgae [467]. Antioxidants play a

crucial role in mitigating oxidative damage by ROS and UV radiation, thereby supporting macroalgal growth. Within the *Sphingomonadaceae* family, certain members are noted for their antioxidant production. Specifically, *Altererythrobacter ishigakiensis* NBRC 107699 is recognized for producing astaxanthin [468], while various *Erythrobacter* species are known to synthesize an array of carotenoids [469]. Both astaxanthin and carotenoids are acknowledged antioxidants that are commercially used.

Moreover, *Sphingomonadaceae* members are involved in the degradation of algal-associated polysaccharides. For instance, the genome of *Erythrobacter* sp. 3-20A1M includes not only pectinase genes but also features genes for cellulases, which play a role in decomposing plant and algal cell walls [470]. *Sphingomonas* sp. strain A1 has been identified as an alginolytic strain capable of degrading alginate [471], a primary cell wall component of brown macroalgae. Furthermore, gene clusters responsible for the degradation of polysaccharides were discovered in the genomes in Chapter III. For instance, I found PUL-like gene clusters targeting alginate, xylan, and starch in the genomes of *Altererythrobacter* and *Erythrobacter*. This demonstrates the role of alphaproteobacterial bacteria in the degradation of marine macroalgae-derived polysaccharides.

4.1.2.4 Interesting findings regarding bacteria not among the topmost 15 clades

I have discovered some intriguing results regarding bacteria not among the topmost 15 clades. An example is the DEV007 clade from the phylum *Verrucomicrobiota*, which had high abundances exclusively on the sampled red algae and in the surrounding seawater. In contrast, DEV007 was nearly undetectable in the phycosphere microbial communities of green and brown algae. While DEV007 was detected in the PA fractions in Chapter II, its relative abundance was found to be low. Gene analysis of DEV007 MAGs revealed the absence of gene clusters specifically targeting the degradation of alginate and fucoidan, which are polysaccharides in the cell walls of brown algae [151]. However, gene clusters associated with the degradation of starch were detected, with starch being the storage polysaccharide in red and green algae [128]. Consequently, high-quality MAGs were obtained exclusively from the red algal metagenomes in the study of Chapter III.

In Chapter III, Figure 3 demonstrated that gammaprotebacterial *Marinomonas* were present at a higher relative abundance in the phycosphere communities of brown algae but had lower relative abundances in red algae phycospheres and were almost undetectable in green algae phycosopheres, unfiltered seawater, and sediment samples. This observation is consistent with reports suggesting that *Marinomonas* species are specifically associated with brown algae [264]. In Chapter II, it was observed that the relative abundances were exceedingly low. The presence of gene clusters for alginate degradation in the genome of *Marinomonas* from Chapter III further substantiates a preference of *Marinomonas* for brown algae.

4.1.3 Clades that were prominent in PA communities but not in macroalgal phycospheres

The comparative results in Section 4.1.1 also revealed 80 genera and genus-level clades that were predominantly present in bloom-associated PA communities (refer to Appendix Table 2). Those with higher abundances include the CL500-3 clade of the phylum *Planctomycetota*, *Paraglaciecola*, *Oceanicoccus*, the BD1-7 clade, and unclassified *Methylophagaceae* and *Nitrincolaceae* of the class *Gammaproteobacteria*. The BD1-7 clade and unclassified *Methylophagaceae* and *Nitrincolaceae* have been discussed in the discussion of Chapter II. They were among the more abundant bacterial groups within the PA bacterial community, particularly the BD1-7 clade, which was dominant within the PA bacterial communities. Planctomycetes, including the CL500-3 clade, were proportionally more abundant in PA3 communities during the late diatom and *Phaeocystis* bloom phases in the study of Chapter II. The abundance of planctomycetes in Chapter II is different from that in Chapter III. Five genera were recognized as prominent among the planctomycetes within the PA communities (Chapter II), yet none of these planctomycetes genera were prominent in the macroalgal phycospheres (Chapter III). Originally identified as planktonic bacteria, planctomycetes are hypothesized to lead predominantly attached lifestyle, potentially facilitated by the presence of holdfast structures and stalks [472]. *Paraglaciecola* is associated with diatoms, consistent with the findings of the study in Chapter II [473]. It has also been isolated within macroalgal samples [474]. However, its

abundance in macroalgal phycosphere bacterial communities was relatively low, not ranking among the topmost 15 in relative abundance in any of the samples (Chapter III). Members of *Paraglaciecola* possess an extensive array of CAZyme genes [474]. In Chapter II, *Paraglaciecola* is noted for harboring a rich set of genes for degrading α-glucan and alginate. MAGs of *Oceanicoccus* have also been identified in the FL bacterial communities of the Helgoland blooms in previous years [211]. However, *Oceanicoccus* exhibited a higher relative abundance in PA communities (Chapter II), a finding corroborated by metaproteomic studies [475]. *Oceanicoccus* has a number of TBDTs comparable to that of *Bacteroidota* [476], suggesting a potential role in the degradation of high-molecular-weight substances, although this has not yet been experimentally validated.

4.2 Enhanced analytical depth through the introduction of PacBio metagenome sequencing and draft genome sequencing of cultivable strains

The advancement of sequencing technologies and the associated reduction in cost per base continue to foster the collection of metagenomic data, enabling a deeper understanding of microbial communities in increasingly diverse environments. There have been extensive oftentimes rather technical discussions about the advantages and disadvantages of long-read sequencing in recent years, which are beyond the scope of this thesis. Instead, I will share some more general observations. Long-read sequencing offers higher quality sequences, allowing more in-depth analysis of metagenomes and genomes (including MAGs) and providing additional insights that are not attainable through Illumina sequencing. In the study of Chapter II, I performed long-read sequencing on eight samples from the PA3 (3-10 μm) fraction using the PacBio Sequel II and IIe platforms, resulting in Hi-Fi reads. The comparison of MAGs retrieved in Illumina and PacBio data are shown in Figure 4.2. This indeed improved the analysis in the study of Chapter II in the following aspects. 1), It allowed me to extract complete or nearly-complete 16S rRNA genes for describing the microbial community composition of the samples and to validate these data by comparison with 16S rRNA amplicon data. 2), It facilitated the assembly of high-quality, less fragmented MAGs, significantly enhancing

the precision of genome annotation. This improvement was particularly evident with respect to the accurate prediction of the gene structure of PULs, illustrating the profound impact of reduced fragmentation on the reliability of genomic insights. Many significant findings in the study of Chapter II regarding PULs relied on PacBio MAGs. For example, I predicted a comprehensive potential PUL capable of degrading fucoidan in a *Polaribacter* MAG, which was not retrieved in corresponding Illumina data. The ability to degrade fucoidan is observed in *Verrucomicrobiota*, rendering the identification in PA *Polaribacter* notably fascinating. 3), I could extract complete or near-complete 16S rRNA sequences from PacBio MAGs for taxonomic confirmation and correlate the taxonomic positions of ASV obtained from the Silva SSU 138.1 taxonomy with MAG positions obtained from the GTDB release R207_v2. This is crucial as it bridged ASV and MAG data, linking bacterial relative abundance with functionality. There are discrepancies between the taxonomic descriptions in the GTDB and Silva databases, and failing to reconcile these two can lead to overlooking research on certain critical taxa. For instance, planctomycetotal CL500-3 was prominent during the late diatom and *Phaeocystis* bloom phases in Chapter II research, yet CL500-3 does not have a taxonomic position in the GTDB database. Sequence alignment of the 16S RNA gene of MAGs with the Silva database identified UBA12014 as CL500-3, enabling the analysis of CL500-3's polysaccharide degradation functional potential. The probability of obtaining 16S rRNA sequences from Illumina MAGs is far lower than for PacBio MAGs, and the length of 16S rRNA sequences obtained from Illumina MAGs is shorter. A detailed comparison between Illumina and PacBio MAGs has been recently published by a colleague [315]. 4), Thanks to the higher likelihood of obtaining high-quality 16S rRNA sequences from PacBio MAGs, I was able to associate abundant ASVs with MAGs. For instance, the BD1-7 clade was most prominent in PA fractions during the late diatom and *Phaeocystis* bloom phases. Through the comparison of ASVs with 16S rRNA sequences of MAGs, I could identify species consistent with the most prominent ASVs and could thereby link *in situ* abundances and specific polysaccharide degradation functions. 5), Also, by comparing ASVs with 16S rRNA sequences of MAGs, I found that certain species within the same genus had higher abundance in FL, while others had higher abundance in PA fractions.

Those with higher abundance in PA fractions had more polysaccharide hydrolases, like *Lentimonas* (see Figure 7 in Chapter II). This comparison also allowed me to assess whether MAGs tend to be associated with the FL or PA fractions (see Figure 7 in Chapter II). Considering that our high temporal resolution 16S rRNA amplicon data (51 samples per fraction) contrasted with the comparatively sparse metagenomic data (FL: 18, PA3: 16, PA10: 8 samples), utilizing 16S rRNA amplicon data to examine bacterial dynamics across the bloom effectively mitigated the limitations encountered in analyzing MAG dynamics throughout the entire sampling period.

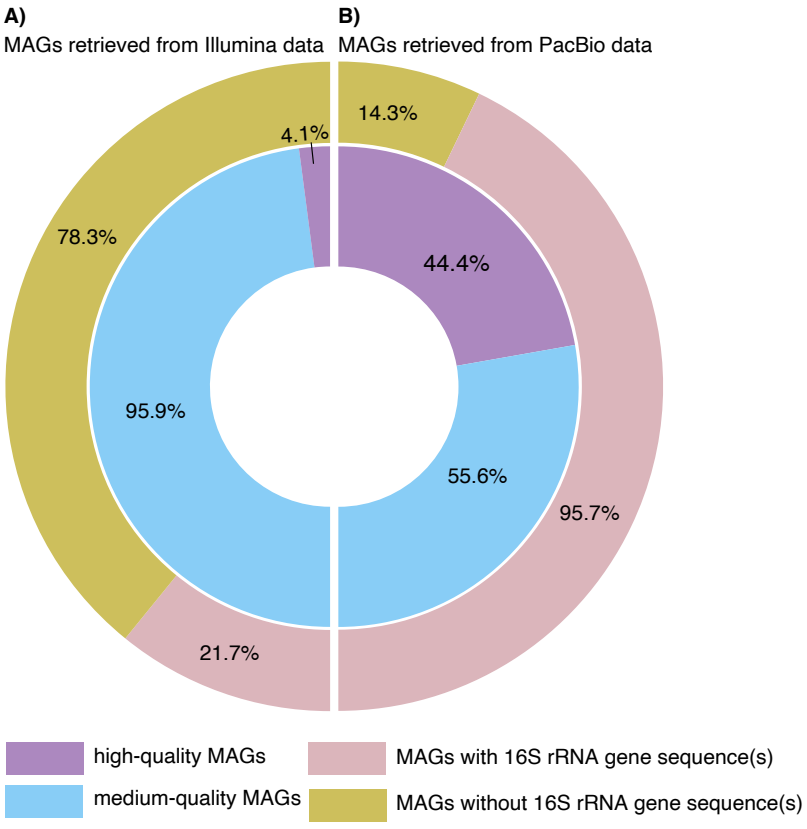


Figure 4.2. Comparison of the metagenome-assembled genomes (MAGs) retrieved from Illumina and PacBio sequence data. The quality of metagenome-assembled genomes (high-quality: >90% completeness, <5% contamination, presence of 23S, 16S and 5S rRNA genes, and ≥18 tRNAs; medium-quality: ≥50% completeness, <10% contamination [316]) and detection of 16S rRNA gene sequences) were compared. Panel A) represents MAGs retrieved from Illumina data, and panel B) represents MAGs retrieved from PacBio data.

Cultivable bacterial strains are the most reliable for validating functionalities. However, currently, on average less than 1% of marine bacteria are cultivable on plates [282], a figure that varies considerably with the environment. As confirmed in the research presented in Chapter III, a much higher percentage of cultivable bacteria exists among macroalgal phycosphere microbiota. In the study of Chapter III, a total of 5,527 bacterial strains were cultivable on plates, 4,426 of which were isolated from macroalgal

psychosphere samples, representing 1,235 species (including 689 potentially new species). Out of these, 965 strains were selected for genome sequencing, 865 of which originated from macroalgal phycosphere samples. This implies that the research in Chapter III provides a wealth of phycosphere bacterial model strains and genomic information, crucial for studying polysaccharide degradation and interactions between macroalgae and bacteria.

In summary, the introduction of high-quality sequencing technologies has made genome function prediction more comprehensive. The acquisition of cultivable strains not only yields more complete genomic information but also allows hypotheses to be experimentally validated.

4.3 Multiple databases were referenced for functional annotation

Beyond genome quality, the most critical factor in functional annotation is the selection of appropriate databases. For instance, annotating the polysaccharide degradation capabilities of a genome relies on multiple databases, including the Carbohydrate-Active enZYmes Database [231], TIGRFAM [477], Pfam [478], and dbCAN [479]. If a detailed classification of sulfatase genes down to subfamilies is required, the SulfAtlas database [235] is also necessary. To maximize the proportion of functional annotations, I incorporated several additional databases for gene annotation, including COG [480] and EggNOG [481]. However, the annotation rates of these databases vary. The results from Chapter III (see Supplementary Figure 9 in Chapter III) showed that the annotation rates for COG and EggNOG are 75.9% and 80.9%, respectively, which are considerably high. Predicting polysaccharide substrates has always been one of the challenges in studying PULs and CAZyme genes. To enhance the prediction of polysaccharide substrates for CAZyme genes and PULs, PULDB [228] and the recently published dbCAN3-sub [482] were introduced (Chapter II). The substrates listed in dbCAN3-sub are published and experimentally verified. The inclusion of these two databases has improved substrate prediction considerably. Nevertheless, many genomes still have at least half of their CAZyme genes unmatchable to any polysaccharide substrates and remain unknown, as illustrated in Figure 7 of Chapter II.

Outlook

The oceans influence the global climate in multiple ways (e.g., convective temperature transport by water and winds including water evaporation and precipitation). In recent years, increasing attention has been paid to processes of marine carbon fixation and remineralization, since the oceans represent one of Earth's largest carbon repositories. Phytoplankton and various forms of marine vegetation, including macroalgae, play a critical role by assimilating atmospheric carbon dioxide through photosynthesis. A considerable fraction of this sequestered carbon is converted to various polysaccharides. Previous research has unveiled a connection between bacteria and the remineralization of these polysaccharides (e.g., [123]). Understanding the storage of carbon in the ocean thus necessitates a thorough examination of the remineralization of fixed carbon, where heterotrophic bacteria play a role as principal degraders in carbon remineralization processes. Therefore, understanding how marine bacteria utilize polysaccharides represents an essential piece in the intricate puzzle of the oceanic carbon cycle.

In the study presented in Chapter II, the LTER site at Helgoland Roads served as a sampling location for the analysis of dynamic shifts in bacterial communities and their functionalities during and after phytoplankton blooms, with a specific emphasis on particle-attached bacteria. Chapter III delves into the community structures and functions of bacteria associated with four types of marine macroalgae. In the discussion of this thesis, I broadly categorize the particle-attached bacteria discussed in Chapter II as microalgal-associated bacteria and undertook a comparative analysis of their community compositions and functionalities versus the macroalgal-associated bacteria detailed in Chapter III. This comparison yielded many insightful conclusions and hypotheses that merit further exploration. Future investigations should precisely define the concept of algae phycosphere bacterial communities and collect specific microalgal phycosphere bacteria for comparative studies alongside macroalgal phycosphere bacteria.

Merely sequencing particle-associated bacteria during phytoplankton blooms remains limiting. The absence of metatranscriptome and sufficiently deep metaproteome data hampered a verification of the functional dynamics of certain particle-attached

bacteria in Chapter II. The prevalence of eukaryotic algae and algal material in particles during phytoplankton blooms also poses methodological challenges for particle metaproteomics and metatranscriptomics. A study to which I contributed (see Appendix - Additional co-author publication 2 for details) employed the same three samples for metaproteomic analyses as those discussed in Chapter II of this thesis. The results of this study indicated that eukaryotic proteins accounted for 70-80% of the proteins identified in the sampled particles. Addressing these challenges and securing high-quality, comprehensive metaproteome and metatranscriptome data should be taken into consideration for future related research. Recent developments in sequencing may help in this regard, as the recently introduced Pacific Biosciences Revio instrument (https://www.pacb.com/revio) promises to at least reduce sequencing cost for long-read sequencing, and Ultima Genomics UG100 (https://www.ultimagenomics.com) has the potential to finally compete with Illumina in short-read sequencing with a possibly better price per base, higher accuracy, and simpler sample preparation.

This thesis also postulates that certain bacteria, such as *Polaribacter*, possess the ability to transition between free-living and particle-attached lifestyles, although the mechanisms for this adaptability remain elusive. Future studies could select appropriate model strains to investigate the mechanisms behind these lifestyle transitions.

After draft genome sequencing of isolated strains or community DNA sequencing and MAG reconstruction for entire microbial communities, gene prediction marks the initial phase of studying the degradation of polysaccharides by bacteria. This is primarily a hypothesis-generating process that enables the pinpointing of the research focus. However, a comprehensive understanding of metabolic pathways when utilizing a specific polysaccharide requires experimental substantiation. Improving the utility of cultivated bacterial strains and exploring their metabolic pathways is therefore essential, such as the suggested study of the fucoidan degradation pathway in *Polaribacter* and the metabolic pathway in *Algibacter* sp. 4-1052.

Sustaining research continuity over time is vital. While Chapter II offers new perspectives, it solely encompasses the spring phytoplankton bloom of a single year. To ascertain the general characteristics of particle-attached bacterial communities and their

interrelations with phytoplankton, it's imperative to conduct studies on bloom-related particle-attached fractions from additional years. Likewise, although Chapter III introduces novel breakthroughs in the realm of macroalgal-associated bacteria, it also restricts its data collection to samples from four seasons within a single year. To probe the stable configurations of macroalgal-associated bacterial communities, ongoing, multi-year studies at consistent sampling locations and during identical seasons with uniform macroalgal species are essential.

The symbiotic interplay between bacteria and algae constitutes a pivotal research subject. For instance, in the study of Chapter II, I discovered that the *Nitrincolaceae* ASP10-02a clade exhibited a profound correlation exclusively with diatoms and stands out as the most prominently expressed group in the free-living metaproteome, albeit with its highly expressed proteins not directly implicated in polysaccharide degradation. This observation sets the stage for my subsequent project. ASP10-02a members may be capable of synthesizing vitamin B_{12}, which fosters the growth of diatoms [483]. The 2020 Helgoland spring phytoplankton bloom was characterized by a simple composition predominantly dominated by diatoms, accompanied by high-quality, high temporal resolution data, including PacBio metagenome data and Illumina metatranscriptome data [127]. We have also acquired additional data for the PA3 (3-10 μm) fraction of the 2020 bloom, which will provide a robust dataset for the investigation of my hypothesis. I will explore the correlation between ASP10-02a and diatoms and the expression during diatom blooms. Moreover, an investigation into how seasonal variations affect the abundance of ASP10-02a and its association with diatoms will be conducted, using phytoplankton bloom data from 2022-2023. I will integrate published global data on the abundance and distribution of diatoms and bacteria to examine the global geographical distribution of ASP10-02a and reveal the correlation between ASP10-02a and diatoms on a global scale. Additionally, pan-genomic analyses will be conducted to uncover the functional characteristics of the ASP10-02a clade and elucidate the mechanisms associated with its interaction with diatoms.

References

1. Caldeira K, Akai M, Brewer P, Chen B, Haugan P, Iwama T, et al. IPCC special report on carbon dioxide capture and storage. Cambridge: Cambridge University Press; 2005.
2. DeVries T. The ocean carbon cycle. *Annu Rev Environ Resour*. 2022;47:317-41.
3. Iversen MH. Carbon export in the ocean: a biologist's perspective. *Ann Rev Mar Sci*. 2023;15:357-81.
4. Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, et al. IPCC. Climate Change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge, United Kingdom and New York, NY: Cambridge University Press; 2013. p. 1535.
5. Lee C, Jiang H, Dasgupta R, Torres M. A framework for understanding whole Earth carbon cycling. In Orcutt B, Daniel I, Dasgupta R, eds. Deep Carbon: Past to present. Cambridge, UK: Cambridge University Press; 2019. pp. 313-57.
6. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*. 1998;281(5374):237-40.
7. Longhurst A, Sathyendranath S, Platt T, Caverhill C. An estimate of global primary production in the ocean from satellite radiometer data. *Plankton Res*. 1995;17(6):1245-71.
8. Ducklow HW. Ocean biogeochemical fluxes: new production and export of organic matter from the upper ocean. *Rev Geophys*. 1995;33(S2):1271-6.
9. Balkanski Y, Monfray P, Batle M, Heimann M. Ocean primary production derived from satellite data: an evaluation with atmospheric oxygen measurements. *Glob Biogeochem Cycles*. 1999;13:257-71.
10. Morel A, Antoine D. Small critters-big effects. *Science*. 2002;296(5575):1980-82

11. Reinart A, Kutser T. Comparison of different satellite sensors in detecting cyanobacterial bloom events in the Baltic Sea. *Remote Sens Environ*. 2006;201(1-2):74-85.
12. Baumert HZ, Petzoldt T. The role of temperature, cellular quota and nutrient concentrations for photosynthesis, growth and light-dark acclimation in phytoplankton. *Limnologica*. 2008;38(3-4):313-26.
13. Duarte CM, Cebrián J. The fate of marine autotrophic production. *Limnol Oceanogr*. 1996;41(8):1758-66.
14. Duarte CM, Middelburg JJ, Caraco N. Major role of marine vegetation on the oceanic carbon cycle. *Biogeosci*. 2005;2:1-8.
15. Smith SV. Marine macrophytes as a global carbon sink. *Science*. 1981;211(4484):838-40.
16. Duarte CM, Gattuso J-P, Hancke K, Gundersen H, Filbee-Dexter K, Pedersen MF, et al. Global estimates of the extent and production of macroalgal forests. *Glob Ecol Biogeogr*. 2022;31(7):1422-39.
17. Krause-Jensen D, Duarte CM. Substantial role of macroalgae in marine carbon sequestration. *Nat Geosci*. 2016;9(10):737-42.
18. Pessarrodona A, Filbee-Dexter K, Krumhansl KA, Pedersen MF, Moore PJ, Wernberg T. A global dataset of seaweed net primary productivity. *Sci Data*. 2022;9(1):484.
19. Ortega A, Geraldi NR, Alam I, Kamau AA, Acinas SG, Logares R, et al. Important contribution of macroalgae to oceanic carbon sequestration. *Nat Geosci*. 2019;12(9):748-54.
20. Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, et al. The evolution of modern eukaryotic phytoplankton. *Science*. 2004;305(5682):354-60.
21. Buchan A, LeCleir GR, Gulvik CA, González JM. Master recyclers: Features and functions of bacteria associated with phytoplankton blooms. *Nat Rev Microbiol*. 2014;12(10):686-98.
22. Moran MA, Ferrer-Gonzalez FX, Fu H, Nowinski B, Olofsson M, Powers MA, et al. The Ocean's labile DOC supply chain. *Limnol Oceanogr*. 2022;67(5):1007-21.

23. Wetz MS, Wheeler PA. Release of dissolved organic matter by coastal diatoms. *Limnol Oceanogr*. 2007;52(2):798-807.
24. Myklestad SM. Dissolved organic carbon from phytoplankton. In: Wangersky PJ, ed. Marine Chemistry. The Handbook of Environmental Chemistry, vol 5D. Springer; 2000.
25. Hellebust JA. Excretion of some organic compounds by marine phytoplankton[1]. *Limnol Oceanogr*. 1965;10(2):192-206.
26. Zlotnik I, Dubinsky Z. The effect of light and temperature on DOC excretion by phytoplankton. *Limnol Oceanogr*. 1989;34(5):831-9.
27. Myklestad SM. Release of extracellular products by phytoplankton with special emphasis on polysaccharides. *Sci Total Environ*. 1995;65(1-3):155-64.
28. Baines SB, Pace ML. The production of dissolved organic matter by phytoplankton and its importance to bacteria: patterns across marine and freshwater systems. *Limnol Oceanogr*. 1991;36(6):1078-90.
29. Yancey PH, Heppenstall M, Ly S, Andrell RM, Gates RD, Carter VL, et al. Betaines and dimethylsulfoniopropionate as major osmolytes in cnidaria with endosymbiotic dinoflagellates. *Physiol Biochem Zool*. 2010;83(1):167-73.
30. Raina JB, Clode PL, Cheong S, Bougoure J, Kilburn MR, Reeder A, et al. Subcellular tracking reveals the location of dimethylsulfoniopropionate in microalgae and visualises its uptake by marine bacteria. *eLife*. 2017;6:e23008.
31. Archer SD, Widdicombe CE, Tarran GA, Rees AP, Burkill PH. Production and turnover of particulate dimethylsulphoniopropionate during a coccolithophore bloom in the northern North Sea. *Aquat Microb Ecol*. 2001; 24(3):225-41.
32. Simó R, D. Archer S, Pedrós-Alió C, Gilpin L, Stelfox-Widdicombe CE. Coupled dynamics of dimethylsulfoniopropionate and dimethylsulfide cycling and themicrobial food web in surface waters of the North Atlantic. *Limnol Oceanogr*. 2022;47(1):53-61.
33. Carlson CA, Hansell DA. Chapter 3 - DOM sources, sinks, reactivity, and budgets. In: Hansell DA, Carlson CA, editors. Biogeochemistry of marine dissolved organic matter. 2nd ed. Boston: Academic Press; 2015. pp. 65-126.

34. Saba GK, Steinberg DK, Bronk DA. The relative importance of sloppy feeding, excretion, and fecal pellet leaching in the release of dissolved carbon and nitrogen by *Acartia tonsa* copepods. *J Exp Mar Biol Ecol*. 2011;404:47-56.
35. Nowicki M, DeVries T, Siegel DA. Quantifying the carbon export and sequestration pathways of the ocean's biological carbon pump. *Glob Biogeochem Cycles*. 2022;36(3):e2021GB007083.
36. Suttle CA. Marine viruses-major players in the global ecosystem. *Nat Rev Microbiol*. 2007;5(10):801-12.
37. Wells ML, Goldberg ED. Marine sub-micron particles. *Mar Chem*. 1992;40:5-18.
38. Kepkay PE. Particle aggregation and the biological reactivity of colloids. *Mar Ecol Prog Ser*. 1994;109:293-304.
39. Guo L, Santschi PH. Composition and cycling of colloids in marine environments. *Rev Geophys*. 1997;35:17-40.
40. Stordal MC, Santschi PH, Gill GA. Colloidal pumping: evidence for the coagulation process using natural colloids tagged with ^{230}Hg. *Environ Sci Technol*. 1996;30(11):3335-40.
41. Panagiotopoulos C, Sempéré R. Analytical methods for the determination of sugars in marine samples: a historical perspective and future directions. *Limnol Oceanogr: Methods*. 2005;3(10):419-54.
42. Schmoker C, Hernández-León S, Calbet A. Microzooplankton grazing in the oceans: impacts, data variability, knowledge gaps and future directions. *J Plankton Res*. 2013;35(4):691-706.
43. Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science*. 2015;347(6223):1257594.
44. Rohr T, Richardson AJ, Lenton A, Chamberlain MA, Shadwick EH. Zooplankton grazing is the largest source of uncertainty for marine carbon cycling in CMIP6 models. *Commun Earth Environ*. 2023;4:212.

45. Nowicki M, DeVries T, Siegel DA. Quantifying the carbon export and sequestration pathways of the ocean's biological carbon pump. *Glob Biogeochem Cycles*. 2022;36(3):e2021GB007083.
46. Toggweiler JR. Catalytic conversions. *Nature*. 1992;356:665-6.
47. Karsenti E, Acinas SG, Bork P, Bowler C, De Vargas C, Raes J, et al. A holistic approach to marine eco-systems biology. *PLoS Biol*. 2011;9(10):e1001177.
48. Duarte CM. Seafaring in the 21st century: the Malaspina 2010 circumnavigation expedition. *Limnol Oceanogr Bull*. 2015;24(1):11-4.
49. Schiel DR, Lilley SA. Gradients of disturbance to an algal canopy and the modifcation of an intertidal community. *Mar Ecol Prog Ser*. 2007;339:1-11.
50. Gao K, McKinley KR. Use of macroalgae for marine biomass production and CO_2 remediation: a review. *J Appl Phycol*. 1994;6:45-60.
51. Keenan TF, Williams CA. The terrestrial carbon sink. *Annu Rev Environ Resour*. 2018;43:219-43.
52. Watanabe K, Yoshida G, Hori M, Umezawa Y, Moki H, Kuwae T. Macroalgal metabolism and lateral carbon flows can create significant carbon sinks. *Biogeosci*. 2020;17(9):2425-40.
53. Tsai DD-W, Chen PH, Ramaraj R. The potential of carbon dioxide capture and sequestration with algae. *Ecol Eng*. 2017;98:17-23.
54. Buck-Wiese H, Andskog MA, Nguyen NP, Bligh M, Asmala E, Vidal-Melgosa S, et al. Fucoid brown algae inject fucoidan carbon into the ocean. *Proc Natl Acad Sci U S A*. 2022;120(1):e2210561119.
55. Food & agriculture organization of the united nations. The state of world fisheries and aquaculture. Sustainability in action. 2020.
56. Dunne JP, Sarmiento JL, Gnanadesikan A. A synthesis of global particle export from the surface ocean and cycling through the ocean interior and on the seafloor. *Glob Biogeochem Cycles*. 2007;21(4):1-16.
57. Trombetta T, Vidussi F, Mas S, Parin D, Simier M, Mostajir B. Water temperature drives phytoplankton blooms in coastal waters. *PLoS ONE*. 2019;14(4):e0214933.

58. Beman JM, Arrigo KR, Matson PA. Agricultural runoff fuels large phytoplankton blooms in vulnerable areas of the ocean. *Nature*. 2005;434:211-4.
59. Dai Y, Yang S, Zhao D, Hu C, Xu W, Anderson DM, et al. Coastal phytoplankton blooms expand and intensify in the 21st century. *Nature*. 2023;615:280-4.
60. Mann DG. The species concept in diatoms: evidence for morphologically distinct, sympatric gamodemes in four epipelic species. *Plant Syst Evol*. 1989;164:215-37.
61. Jin X, Gruber N, Dunne JP, Sarmiento JL, Armstrong RA. Diagnosing the contribution of phytoplankton functional groups to the production and export of particulate organic carbon, $CaCO_3$, and opal from global nutrient and alkalinity distributions. *Glob Biogeochem Cycles*. 2006;20(2):1-17.
62. Malviya S, Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, et al. Insights into global diatom distribution and diversity in the world's ocean. *Proc Natl Acad Sci U S A*. 2016;113(11):E1516-25.
63. Taylor FJR, Hoppenrath M, Saldarriaga JF. Dinoflagellate diversity and distribution. *Biodivers Conserv*. 2008;17:407-18.
64. Saldarriaga JF, Taylor FJR. Dinoflagellata. In: Archibald J, et al., editors. Handbook of the Protists. Springer; 2017.
65. Lovejoy C, Legendre L, Martineau M-J, Bâcle J, von Quillfeldt CH. Distribution of phytoplankton and other protists in the North Water. *Deep Sea Res Part II Top Stud Oceanogr*. 2002;49(22-23):5027-47.
66. Vidussi F, Roy S, Lovejoy C, Gammelgaard M, Thomsen HA, Booth B, et al. Spatial and temporal variability of the phytoplankton community structure in the North Water Polynya, investigated using pigment biomarkers. *Can J Fish Aquat Sci*. 2004;61(11):2038-52.
67. Janouškovec J, Gavelis GS, Burki F, Dinh D, Bachvaroff TR, Gornik SG, et al. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proc Natl Acad Sci U S A*. 2017;114(2):E171-E80.
68. Danchenko S, Dodge JD, Icely JD, Newton A. Dinoflagellate assemblages in the West Iberian upwelling region (Sagres, Portugal) during 1994-2001. *Front Mar Sci*. 2022;9:591759.

69. Cembella AD. Chemical ecology of eukaryotic microalgae in marine ecosystems. *Phycologia*. 2003;42(4):420-47.
70. Ollevier A, Mortelmans J, Aubert A, Deneudt K, Vandegehuchte MB. *Noctiluca scintillans*: dynamics, size measurements and relationships with small soft-bodied plankton in the Belgian Part of the North Sea. *Front Mar Sci*. 2021;8:777999.
71. Hu Z, Liu Y, Deng Y, Tang YZ. The notorious harmful algal blooms-forming dinoflagellate *Prorocentrum donghaiense* produces sexual resting cysts, which widely distribute along the coastal marine sediment of China. *Front Mar Sci*. 2022;9:826736.
72. Sathishkumar RS, Sahu G, Mohanty AK, Arunachalam KD, Venkatesan R. First report of *Protoperidinium steinii* (*Dinophyceae*) bloom from the coastal marine ecosystem – an observation from tropical Indian waters. *Oceanologia*. 2021;63(3):391-402.
73. Wang J, Kong F, Wang Y, Ji N, Song M, Hu Z, et al. Newly recorded bloom-forming dinoflagellate *Gymnodinium impudicum* in Haizhou Bay, Yellow Sea, China. *J Ocean Limnol*. 2022;40:2430–45.
74. Heiskanen AS. Mass encystment and sinking of dinoflagellates during a spring bloom. *Mar Biol*. 1993;116:161-7.
75. Montero P, Pérez-Santos I, Daneri G, Gutiérrez MH, Igor G, Seguel R, et al. A winter dinoflagellate bloom drives high rates of primary production in a Patagonian fjord ecosystem. *Estuar Coast Shelf Sci*. 2017;199:105-16.
76. Hopkins J, Henson SA, Painter SC, Tyrrell T, Poulton AJ. Phenological characteristics of global coccolithophore blooms. *Glob Biogeochem Cycles*. 2015;29(2):239-53.
77. Poulton AJ, Adey TR, Balch WM, Holligan PM. Relating coccolithophore calcification rates to phytoplankton community dynamics: Regional differences and implications for carbon export. *Deep Sea Res Part II Top Stud Oceanogr*. 2007;54(5-7):538-57.
78. Poulton AJ, Painter SC, Young JR, Bates NR, Bowler B, Drapeau D, et al. The 2008 *Emiliania huxleyi* bloom along the Patagonian Shelf: ecology,

biogeochemistry, and cellular calcification. *Glob Biogeochem Cycles*. 2013;27(4):1023-33.

79. Arrigo KR, Robinson DH, Worthen DL, Dunbar RB, DiTullio GR, VanWoert M, et al. Phytoplankton community structure and the drawdown of nutrients and CO_2 in the southern ocean. *Science*. 1999;283(5400):365-7.
80. Baumann MEM, Lancelot C, Brandidni FP, Sakshaugh E, John DM. The taxonomic identity of the cosmopolitan prymnesiophyte *Phaeocystis*: a morphological and ecophysiological approach. *J Mar Syst*. 1994;5(1):5-22.
81. Lancelot C, Keller MD, Rousseau V, Smith WO, Mathot S. Autoecology of the marine haptophyte *Phaeocystis* sp. In: Anderson DM, Cembella AD, Hallegraeff GM, editors. Physiological ecology of harmful algal blooms. Springer-Verlag; 1998. pp. 209-24.
82. Schoemann V, Becquevort S, Stefels J, Rousseau V, Lancelot C. *Phaeocystis* blooms in the global ocean and their controlling mechanisms: a review. *J Sea Res*. 2005;53:43-66.
83. Lancelot C, Gypens N, Billen G, Garnier J, Roubeix V. Testing an integrated river-ocean mathematical tool for linking marine eutrophication to land use: the *Phaeocystis*-dominated Belgian coastal zone (Southern North Sea) over the past 50 years. *J Mar Syst*. 2007;64(1-4):216-28.
84. Alderkamp AC, Buma AG, van Rijssel M. The carbohydrates of *Phaeocystis* and their degradation in the microbial food web. *Biogeochemistry*. 2007;83:99-118.
85. Vincent F, Gralka M, Schleyer G, Schatz D, Cabrera-Brufau M, Kuhlisch C, et al. Viral infection switches the balance between bacterial and eukaryotic recyclers of organic matter during coccolithophore blooms. *Nat Commun*. 2023;14:510.
86. Zhang ZH, Li DH, Xie RZ, Guo RY, Nair S, Han H, et al. Plastoquinone synthesis inhibition by tetrabromo biphenyldiol as a widespread algicidal mechanism of marine bacteria. *ISME J*. 2023;17(11):1-14.
87. Scholz B, Guillou L, Marano AV, Neuhauser S, Sullivan BK, Karsten U, et al. Zoosporic parasites infecting marine diatoms - a black box that needs to be opened. *Fungal Ecol*. 2016;19:59-76.

88. Garvetto A, Nézan E, Badis Y, Bilien G, Arce P, Bresnan E, et al. Novel widespread marine oomycetes parasitising diatoms, including the toxic genus *Pseudo-nitzschia*: genetic, morphological, and ecological characterisation. *Front Microbiol*. 2018;9:2918.

89. Fernández-Méndez M, Wenzhöfer F, Peeken I, Sørensen HL, Glud RN, Boetius A. Composition, buoyancy regulation and fate of ice algal aggregates in the Central Arctic Ocean. *PLoS ONE*. 2014;9(9):e107452.

90. Alldredge AL, Passow U, Logan BE. The abundance and significance of a class of large, transparent organic particles in the ocean. *Deep Sea Res Part I Oceanogr Res Pap*. 1993;40(6):1131-40.

91. Dutz J, Breteler WCMK, Kramer G. Inhibition of copepod feeding by exudates and transparent exopolymer particles (TEP) derived from a *Phaeocystis globosa* dominated phytoplankton community. *Harmful algae*. 2005;4(5):929-40.

92. Babiak W, Krzemińska I. Extracellular polymeric substances (EPS) as microalgal bioproducts: A review of factors affecting EPS synthesis and application in flocculation processes. *Energies*. 2021;14(13):4007.

93. Berman T, Viner-Mozzini Y. Abundance and characteristics of polysaccharide and proteinaceous particles in Lake Kinneret. *Aquat Microb Ecol*. 2001;24(3):255-64.

94. Engel A. The role of transparent exopolymer particles (TEP) in the increase in apparent particle stickiness (α) during the decline of a diatom bloom. *J Plankton Res*. 2000;22(3):485-97.

95. Rochelle-Newall EJ, Mari X, Pringault O. Sticking properties of transparent exopolymeric particles (TEP) during aging and biodegradation. *J Plankton Res*. 2010;32(10):1433-42.

96. Käse L, Geuer JK. Phytoplankton responses to marine climate change-an introduction. In: YOUMARES 8-Oceans Across Boundaries: Learning from each other: Proceedings of the 2017 conference for YOUng MARine RESearchers in Kiel, Germany. Springer International Publishing; 2018.

97. Kiko R, Biastoch A, Brandt P, Cravatte S, Hauss H, Hummels R, et al. Biological and physical influences on marine snowfall at the equator. *Nat Geosci*. 2017;10(11):852-8.
98. Volk T, Hoffert MI. Volk T, Hoffert MI. Ocean carbon pumps: analysis of relative strengths and efficiencies in ocean-driven atmospheric CO_2 changes. In: Sundquist E, Broecker W, eds. The Carbon Cycle and Atmospheric CO_2: Natural Variations Archean to Present. Washington, DC: American Geophysical Union; 1985:99-110.
99. Omand MM, Govindarajan R, He J, Mahadevan A. Sinking flux of particulate organic matter in the oceans: sensitivity to particle characteristics. *Sci Rep*. 2020;10(1):5582.
100. Lombard F, Boss E, Waite AM, Vogt M, Uitz J, Stemmann L, et al. Globally consistent quantitative observations of planktonic ecosystems. *Front Mar Sci*. 2019;6:196.
101. Kiko R, Picheral M, Antoine D, Babin M, Berline L, Biard T, et al. A global marine particle size distribution dataset obtained with the Underwater Vision Profiler 5. *Earth Syst Sci Data*. 2022;14(9):4315-37.
102. Barone B, Bidigare RR, Church MJ, Karl DM, Letelier RM, White AE. Particle distributions and dynamics in the euphotic zone of the North Pacific Subtropical Gyre. *J Geophys Res: Oceans*. 2015;120(5):3229-47.
103. 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(6822):801-4.
104. Azam F. Microbial control of oceanic carbon flux: the plot thickens. *Science*. 1998;280(5364):694-6.
105. Zhang Y, Xiao W, Jiao NZ. Linking biochemical properties of particles to particle-attached and free-living bacterial community structure along the particle density gradient from freshwater to open ocean. *J Geophys Res: Biogeosciences*. 2016;121(8):2261-74.

106. Alldredge AL, Cole JJ, Caron DA. Production of heterotrophic bacteria inhabiting macroscopic organic aggregates (marine snow) from surface waters. *Limnol Oceanogr*. 1986;31(1):68-78.
107. Azam F, Fenchel T, Field JG, Gray JS, Meyer-Reil LA, Thingstad F. The ecological role of water-column microbes in the sea*. *Mar Ecol Prog Ser*. 1983;10(3):257-63.
108. Caron DA, Davis PG, Madin LP, Sieburth JM. Heterotrophic bacteria and bacteriovorous protozoa in oceanic macroaggregates. *Science*. 1982;218(4574):795-7.
109. Simon M, Grossart HP, Schweitzer B, Ploug H. Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat Microb Ecol*. 2002;28(2):175-211.
110. Fernández-Gómez B, Richter M, Schüler M, Pinhassi J, Acinas SG, González JM, et al. Ecology of marine *Bacteroidetes*: a comparative genomics approach. *ISME J*. 2013;7(5):1026-37.
111. Heins A, Reintjes G, Amann RI, Harder J. Particle collection in Imhoff sedimentation cones enriches both motile chemotactic and particle-attached bacteria. *Front Microbiol*. 2021;12:643730.
112. Michael V, Frank O, Bartling P, Scheuner C, Göker M, Brinkmann H, et al. Biofilm plasmids with a rhamnose operon are widely distributed determinants of the 'swim-or-stick' lifestyle in roseobacters. *ISME J*. 2016;10(10):2498-513.
113. 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.
114. Geng H, Belas R. Molecular mechanisms underlying *Roseobacter*-phytoplankton symbioses. *Curr Opin Biotechnol*. 2010;21(3):332-8.
115. Bižić-Ionescu M, Zeder M, Ionescu D, Orlić S, Fuchs BM, Grossart HP, et al. Comparison of bacterial communities on limnic versus coastal marine particles reveals profound differences in colonization. *Environ Microbiol*. 2015;17(10):3500-14.
116. 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.

117. 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.
118. Rupérez P, Ahrazem O, Leal JA. Potential antioxidant capacity of sulfated polysaccharides from the edible marine brown seaweed *Fucus vesiculosus*. *J Agric Food Chem*. 2002;50(4):840-5.
119. Chen J, Yang J, Du H, Aslam M, Wang W, Chen W, et al. Laminarin, a major polysaccharide in stramenopiles. *Mar Drugs*. 2021;19(10):576.
120. Janse I, van Rijssel M, van Hall PJ, Gerwig GJ, Gottschal JC, Prins RA. The storage glucan of *Phaeocystis globosa* (*Prymnesiophyceae*) cells. *J Phycol*. 1996;32(3):382-7.
121. Myklestad SM, Granum E. Biology of (1, 3)-β-glucans and related glucans in protozoans and chromistans, in chemistry, biochemistry, and biology of 1-3 beta glucans and related polysaccharides. Elsevier; 2009. pp. 353-85.
122. Becker S, Tebben J, Coffinet S, Wiltshire K, Iversen MH, Harder T, et al. Laminarin is a major molecule in the marine carbon cycle. *Proc Natl Acad Sci U S A*. 2020;117(12):6599-607.
123. 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.
124. Keith SC, C. A. 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.
125. 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.

126. Lu DC, Wang FQ, Amann RI, Teeling H, Du JZ. Epiphytic common core bacteria in the microbiomes of co-located green (*Ulva*), brown (*Saccharina*) and red (*Grateloupia*, *Gelidium*) macroalgae. *Microbiome*. 2023;11(1):126.
127. Sidhu C, Kirstein IV, Meunier CL, Rick J, Fofonova V, Wiltshire KH, et al. Dissolved storage glycans shaped the community composition of abundant bacterioplankton clades during a North Sea spring phytoplankton bloom. *Microbiome*. 2023;11(1):77.
128. Suzuki E, Suzuki R. Variation of storage polysaccharides in phototrophic microorganisms. *J Appl Glycosci*. 2013;60:21-7.
129. Ball SG, Morell MK. From bacterial glycogen to starch: Understanding the biogenesis of the plant starch granule. *Annu Rev Plant Biol*. 2003;54:207-33.
130. Beidler I, Steinke N, Schulze T, Sidhu C, Bartosik D, Krull J, et al. Alpha-glucans from bacterial necromass indicate an intra-population loop within the marine carbon cycle. *preprint*. 2023. doi: https://doi.org/10.21203/rs.3.rs-3205445/v1.
131. Sims PA, Mann DG, Medlin LK. Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia*. 2006;45(4):361-402.
132. Gügi B, Le Costaouec 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.
133. Haug A, Myklestad S. Polysaccharides of marine diatoms with special reference to *Chaetoceros* species. *Mar Biol*. 1976;34(3):217-22.
134. Abdullahi AS, Underwood GJC, Gretz MR. Extracellular matrix assembly in diatoms (*Bacillariophyceae*). V. Environmental effects on polysaccharide synthesis in the model diatom, *Phaeodactylum tricornutum*[1]. *J Phycol*. 2006;42(2):363-78.
135. McConville MJ, Wetherbee R, Bacic A. Subcellular location and composition of the wall and secreted extracellular sulphated polysaccharides/proteoglycans of the diatom *Stauroneis amphioxys* Gregory. *Protoplasma*. 1999;206:188-200.
136. Hecky RE, Mopper K, Kilham P, Degens ET. The amino acid and sugar composition of diatom cell-walls. *Mar Biol*. 1973;19:323-31.

137. Tesson B, Hildebrand M. Characterization and localization of insoluble organic matrices associated with diatom cell walls: insight into their roles during cell wall formation. *PLoS ONE*. 2013;8(4):e61675.
138. 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: substituted mannans of diatoms. *J Phycol*. 2005;41(6):1154-61.
139. Allan GG, Lewin J, Johnson PG. Marine polymers. IV Diatom polysaccharides. *Bot Mar*. 1972;15(2):102-8.
140. 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.
141. Volcani BE. Volcani BE. Cell wall formation in diatoms: morphogenesis and biochemistry. In: Simpson TL, Volcani BE, editors. Silicon and Siliceous Structures in Biological Systems. New York: Springer-Verlag; 1981. pp. 157-200.
142. Percival E, MeDowell RH. Chemistry and enzymology of marine algal polysaccharides. London and New York: Academic Press; 1967. p. 188.
143. Cowie GL, Hedges JI. Digestion and alteration of the biochemical constituents of a diatom (*Thalassiosira weisflogii*) ingested by an herbivorous zooplankton (*Calanus pacificus*). *Limnol Oceanogr*. 1996;41:581-94.
144. Tesson B, Masse S, Laurent G, Maquet J, Livage J, Martin-Jézéquel V, et al. Contribution of multi-nuclear solid state NMR to the characterization of the *Thalassiosira pseudonana* diatom cell wall. *Anal Bioanal Chem*. 2008;390:1889-98.
145. Parke M, Green JC, Manton I. Observations on the fine structure of zoids of the genus *Phaeocystis* (*Haptophyceae*). *J Mar Biol Assoc UK*. 1971;51(4):927-41.
146. Hallegraeff GM. Scale-bearing and loricate nanoplankton from the East Australian Current. *Bot Mar*. 1983;26(11):493-516.
147. Moestrup Ø. Identification by electron microscopy of marine nanaoplankton from New Zealand, including the description of four new species. *N Z J Bot*. 1979;17(1):61-95.

148. Leadbeater BSC. Cell coverings. In: Green JC, Leadbeater BSC (eds). The haptophyte algae. Oxford: Clarendon Press; 1994. pp. 23-46.
149. Ilya P, Sergei S. Dinoflagellate amphiesma at different stages of the life cycle. *Protistology*. 2012;7:108-15.
150. Michel G, Tonon T, Scornet D, Cock JM, Kloareg B. The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes. *New Phytol*. 2010;188(1):82-97.
151. Kloareg B, Quatrano RS. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanogr Mar Biol Annu Rev*. 1988;26:259-315.
152. Cronshaw J, Myers A, Preston RD. A chemical and physical investigation of the cell walls of some marine algae. *BBA*. 1958;27:89-103.
153. Kylin H. Biochemistry of sea algae. *H Z Physiol Chem*. 1913;83(3):171-97.
154. Hoagland DR, Lieb LL. The complex carbohydrates and forms of sulphurin marinealgae of the pacificoast. *J Biol Chem*. 1915;23(1):287-97.
155. Vidal-Melgosa S, Sichert A, Francis TB, Bartosik D, Niggemann J, Wichels A, et al. Diatom fucan polysaccharide precipitates carbon during algal blooms. *Nat Commun*. 2021;12:1150.
156. Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, et al. *Verrucomicrobia* use hundreds of enzymes to digest the algal polysaccharide fucoidan. *Nat Microbiol*. 2020;5(8):1026-39.
157. Pei Y, Yang S, Xiao Z, Zhou C, Hong P, Qian Z-J. Structural characterization of sulfated polysaccharide isolated from red algae (*Gelidium crinale*) and antioxidant and anti-Inflammatory effects in macrophage cells. *Front Bioeng Biotechnol*. 2021;9:794818.
158. Fernando IPS, Kim D, Nah J-W, Jeon Y-J. Advances in functionalizing fucoidans and alginates (bio)polymers by structural modifications: a review. *Chem Eng J*. 2019;355:33-48.

159. Thiviya P, Gamage A, Liyanapathiranage A, Makehelwala M, Dassanayake RS, Manamperi A, et al. Algal polysaccharides: structure, preparation and applications in food packaging. *Food Chem*. 2023;405(Part A):134903.
160. Xie XT, Cheong KL. Recent advances in marine algae oligosaccharides: structure, analysis, and potential prebiotic activities. *Crit Rev Food Sci Nutr*. 2022;62(28):7703-17.
161. Liu D, Ouyang Y, Chen R, Wang M, Ai C, El-Seedi HR, et al. Nutraceutical potentials of algal ulvan for healthy aging. *Int J Biol Macromol*. 2022;194:422-34.
162. Cheong KL, Zhang Y, Li Z, Li T, Ou Y, Shen J, et al. Role of polysaccharides from marine seaweed as feed additives for methane mitigation in ruminants: a critical review. *Polymers (Basel)*. 2023;15(15):3153.
163. Robic A, Gaillard C, Sassi JF, Lerat Y, Lahaye M. Ultrastructure of ulvan: a polysaccharide from green seaweeds. *Biopolymers*. 2009;91(8):652-64.
164. Wahlström N, Edlund U, Pavia H, Toth G, Jaworski A, Pell AJ, et al. Cellulose from the green macroalgae *Ulva lactuca*: isolation, characterization, optotracing, and production of cellulose nanofibrils. *Cellulose*. 2020;27:3707-25.
165. Domozych DS, Ciancia M, Fangel JU, Mikkelsen MD, Ulvskov P, Willats WG. The cell walls of green algae: a journey through evolution and diversity. *Front Plant Sci*. 2012;3:82.
166. Fogg GE. The extracellular products of algae. *Oceanogr Mar Bioi Annu Rev*. 1966;4:195-212.
167. Hellebust JA. Extracellular products. In: Steward ND, editor. Algal Physiology and Biochemistry. Berkeley, CA: University of California Press; 1974. pp. 838-63.
168. Mague TH, Friberg E, Hughes DJ, Morris I. Extracellular release of carbon by marine phytoplankton; a physiological approach. *Limnol Oceanogr*. 1980;25(2):262-79.
169. Fogg GE. The ecological significance of extracellular products of phytoplankton photosynthesis. *Botanica Marina*. 1983;26(1):3-14.
170. Bjørrisen PK. Phytoplankton exudation of organic matter: Why do healthy cells do it? *Limnol Oceanogr*. 1988;33(1):151-4.

171. Flemming HC, Wingender J. Relevance of microbial extracellular polymeric substances (EPSs) - Part I: structural and ecological aspects. *Water Sci Technol*. 2001;43(6):1-8.
172. Myklestad SM. Release of extracellular products by phytoplankton special emphasis on polysaccharides. *Sci Total Environ*. 1995;165:155-64.
173. Xiao R, Zheng Y. Overview of microalgal extracellular polymeric substances (EPS) and their applications. *Biotechnol Adv*. 2016;34(7):1225-44.
174. Urbani R, Magaletti E, Sist P, Cicero AM. Extracellular carbohydrates released by the marine diatoms *Cylindrotheca closterium*, *Thalassiosira pseudonana* and *Skeletonema costatum*: effect of P-depletion and growth status. *Sci Total Environ*. 2005;353(1-3):300-6.
175. Myklestad S, Haug A, Larsen B. Production of carbohydrates by the marine diatom *Chaetoceros affinis* var. willei (Gran) Hustedt. II. Preliminary investigation of the extracellular polysaccharide. *J Exp Mar Biol Ecol*. 1972;9(2):137-44.
176. Smestad B, Haug A, Myklestad S. Production of carbohydrate by the marine diatom *Chaetoceros affinis* var. willei (Gran) Hustedt. III. Structural studies of the extracellular polysaccharide. *Acta Chem Scand*. 1974;B(28):662-6.
177. Khandeparker RD, Bhosle NB. Extracellular polymeric substances of the marine fouling diatom *Amphora rostrata* Wm.Sm. *Biofouling*. 2001;17(2):117-27.
178. Smestad B, Haug A, Myklestad S. Structural studies of the extracellular polysaccharide produced by the diatom *Chaeotoceros curvisetus* Cleve. *Ser B Org Chem Biochem*. 1975;29(3):337-40.
179. Staats N, de Winder B, Stal L, Mur L. Isolation and characterization of extracellular polysaccharides from the epipelic diatoms *Cylindrotheca closterium* and *Navicula salinarum*. *Eur J Phycol*. 1999;34(2):161-9.
180. van Rijssel M, Janse I, Noordkamp DJB, Gieskes WWC. An inventory of factors that affect polysaccharide production by *Phaeocystis globosa*. *J Sea Res*. 2000;43(3-4):297-306.
181. Janse I, van Rijssel M, Ottema A, Gottschal JC. Microbial breakdown of *Phaeocystis* mucopolysaccharides. *Limnol Oceanogr*. 1999;44(6):1447-57.

182. Orellana MV, Verdugo P. Ultraviolet radiation blocks the organic carbon exchange between the dissolved phase and the gel phase in the ocean. *Limnol Oceanogr*. 2003;48(4):1618-23.
183. Nelson CE, Goldberg SJ, Wegley Kelly L, Haas AF, Smith JE, Rohwer F, et al. Coral and macroalgal exudates vary in neutral sugar composition and differentially enrich reef bacterioplankton lineages. *ISME J*. 2013;7(5):962-79.
184. Ramaiah N, Yoshikawa T, Furuya K. Temporal variations in transparent exopolymer particles (TEP) associated with a diatom spring bloom in a subarctic ria in Japan. *Mar Ecol Prog Ser*. 2001;212:79-88.
185. Decho AW. Microbial exopolymer secretions in ocean environments: their role(s) in food webs and marine processes. *Oceanogr Mar Biol Ann Rev*. 1990;28:73-153.
186. Liu SB, Chen XL, He HL, Zhang XY, Xie BB, Yu Y, et al. Structure and ecological roles of a novel exopolysaccharide from the arctic sea ice bacterium *Pseudoalteromonas* sp. strain SM20310. *Appl Environ Microbiol*. 2013;79(1):224-30.
187. Sun ML, Zhao F, Shi M, Zhang XY, Zhou BC, Zhang YZ, et al. Characterization and biotechnological potential analysis of a new exopolysaccharide from the Arctic marine bacterium *Polaribacter* sp. SM1127. *Sci Rep*. 2015;5:18435.
188. Wotton RS. The ubiquity and many roles of exopolymers (EPS) in aquatic systems. *Oceanogr Mar Biol Annu Revs*. 2004;68(S1):12-21.
189. Serra DO, Richter AM, Hengge R. Cellulose as an architectural element in spatially structured *Escherichia coli* biofilms. *J Bact*. 2013;195(24):5540-54.
190. Hobley L, Harkins C, MacPhee CE, Stanley-Wall NR. Giving structure to the biofilm matrix: and overview of individual strategies and emerging common themes. *FEMS Microbiol Rev*. 2015;39(5):649-69.
191. Rehm BHA, Valla S. Bacterial alginates: biosynthesis and applications. *Appl Microbiol Biotechnol*. 1997;48:281-8.
192. Giménez-Abián MI, Bernabé M, Leal JA, Jiménez-Barbero J, Prieto A. Structure of a galactomannan isolated from the cell wall of the fungus *Lineolata rhizophorae*. *Carbohydr Res*. 2007;342(17):2599-603.

193. Chen Y, Mao WJ, Yan MX, Liu X, Wang SY, Xia Z, et al. Purification, chemical characterization, and bioactivity of an extracellular polysaccharide produced by the marine sponge endogenous fungus *Alternaria* sp. SP-32. *Mar Biotechnol*. 2016;18(3):301-13.

194. Fontaine T, Latgé JP. Galactomannan produced by *Aspergillus fumigatus*: An update on the structure, biosynthesis and biological functions of an emblematic fungal biomarker. *J Fungi*. 2020;6(4):283.

195. Li H, Cao K, Cong P, Liu Y, Cui H, Xue C. Structure characterization and antitumor activity of the extracellular polysaccharide from the marine fungus *Hansfordia sinuosae*. *Carbohydr Polym*. 2018;190(15):87-94.

196. Dell A, Galadari A, Sastre F, Hitchen P. Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int J Microbiol*. 2010;2010:148178.

197. Varki A. Biological roles of glycans. *Glycobiology*. 2017;27(1):3-49.

198. Brown HA, Koropatkin NM. Host glycan utilization within the *Bacteroidetes* Sus-like paradigm. *Glycobiology*. 2021;31(6):697-706.

199. Briliūtė J, Urbanowicz PA, Luis AS, Baslé A, Paterson N, Rebello O, et al. Complex N-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat Microbiol*. 2019;4(9):1571-81.

200. Tivey TR, Parkinson JE, Mandelare PE, Adpressa DA, Peng W, Dong X, et al. N-linked surface glycan biosynthesis, composition, inhibition, and function in cnidarian-dinoflagellate symbiosis. *Microb Ecol*. 2020;80(1):223-36.

201. Baiet B, Burel C, Saint-Jean B, Louvet R, Menu-Bouaouiche L, Kiefer-Meyer MC, et al. N-glycans of *Phaeodactylum tricornutum* diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme. *J Biol Chem*. 2011;286(8):6152-64.

202. Levy-Ontman O, Arad SM, Harvey DJ, Parsons TB, Fairbanks A, Tekoah Y. Unique N-glycan moieties of the 66-kDa cell wall glycoprotein from the red microalga *Porphyridium* sp. *J Biol Chem*. 2011;286(24):21340-52.

203. Mocsai R, Figl R, Troschl C, Strasser R, Svehla E, Windwarder M, et al. N-glycans of the microalga *Chlorella vulgaris* are of the oligomannosidic type but highly methylated. *Sci Rep*. 2019;9(1):331.
204. Bacic A, Burke J, Wetherbee R. Heterogeneous xylose-rich glycans are associated with extracellular glycoproteins from the biofouling diatom *Craspedostauros australis* (*Bacillariophyceae*). *Eur J Phycol*. 2003;38(4):351-60.
205. Mamedov T, Yusibov V. Green algae *Chlamydomonas reinhardtii* possess endogenous sialylated N-glycans. *FEBS Open Bio*. 2011;1:15-22.
206. Mathieu-Rivet E, Mati-Baouche N, Walet-Balieu ML, Lerouge P, Bardor M. N- and O-Glycosylation pathways in the microalgae polyphyletic group. *Front Plant Sci*. 2020;11:609993.
207. Zhang P, Burel C, Plasson C, Kiefer-Meyer MC, Ovide C, Gugi B, et al. Characterization of a GDP-fucose transporter and a fucosyltransferase involved in the fucosylation of glycoproteins in the diatom *Phaeodactylum tricornutum*. *Front Plant Sci*. 2019;10:610.
208. Yoshiie T, Maeda M, Kimura M, Hama Y, Uchida M, Kimura Y. Structural features of N-glycans of seaweed glycoproteins: Predominant occurrence of high-mannose type N-glycans in marine plants. *Biosci Biotechnol Biochem*. 2012;76(10):1996-8.
209. Salyers AA, Vercellotti JR, West SE, Wilkins TD. Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. *Appl Environ Microbiol*. 1977;33(2):319-22.
210. Bjursell MK, Martens EC, Gordon JI. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period*. *J Biol Chem*. 2006;281(47):36269-79.
211. Francis TB, Bartosik D, Sura T, Sichert A, Hehemann JH, Markert S, et al. Changing expression patterns of TonB-dependent transporters suggest shifts in polysaccharide consumption over the course of a spring phytoplankton bloom. *ISME J*. 2021;15(8):2336-50.
212. Krüger K, Chafee M, Francis TB, del Rio TG, Becher D, Schweder T, et al. In marine *Bacteroidetes* the bulk of glycan degradation during algae blooms is

mediated by few clades using a restricted set of genes. *ISME J*. 2019;13(11):2800-16.

213. 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(6):2320-33.
214. 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.
215. McKee LS, Rosa SLL, Westereng B, Eijsink VG, Pope PB, Larsbrink J. Polysaccharide degradation by the *Bacteroidetes*: mechanisms and nomenclature. *Environ Microbiol Rep*. 2021;13(5):559-81.
216. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, et al. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science*. 2003;299(5615):2074-6.
217. 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.
218. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. Polysaccharide utilization loci: fueling microbial communities. *J Bacteriol*. 2017;199(15):e00860-16.
219. Glenwright AJ, Pothula KR, Bhamidimarri SP, Chorev DS, Basle A, Firbank SJ, et al. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature*. 2017;541:407-11.
220. Bolam DN, van den Berg B. TonB-dependent transport by the gut microbiota: novel aspects of an old problem. *Curr Opin Struct Biol*. 2018;51:35-43.
221. Gray DA, White JBR, Oluwole AO, Rath P, Glenwright AJ, Mazur A, et al. Insights into SusCD-mediated glycan import by a prominent gut symbiont. *Nat Commun*. 2021;12:44.
222. Ficko-Blean E, Hervé C, Michel G. Sweet and sour sugars from the sea: the biosynthesis and remodeling of sulfated cell wall polysaccharides from marine macroalgae. *Perspect Phycol*. 2015;2:51-64.

223. Muthukumar J, Chidambaram R, Sukumaran S. Sulfated polysaccharides and its commercial applications in food industries—A review. *J Food Sci Technol*. 2021;58(7):2453-66.

224. 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.

225. Reisky L, Stanetty C, Mihovilovic MD, Schweder T, Hehemann JH, Bornscheuer UT. Biochemical characterization of an ulvan lyase from the marine flavobacterium *Formosa agariphila* KMM 3901^{T}. *Appl Microbiol Biotechnol*. 2018;102(16):6987-96.

226. Chen J, Robb CS, Unfried F, Kappelmann L, Markert S, Song T, et al. Alpha- and beta-mannan utilization by marine *Bacteroidetes*. *Environ Microbiol*. 2018;20(11):4127-40.

227. Ficko-Blean E, Préchoux A, Thomas F, Rochat T, Larocque R, Zhu Y, et al. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nat Commun*. 2017;8:1685.

228. 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):D677-D83.

229. Krewulak KD, Vogel HJ. TonB or not TonB: is that the question? *Biochem Cell Biol*. 2011;89(2):87-97.

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

231. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res*. 2022;50(D1):D571-D7.

232. van Vliet DM, Palakawong Na Ayudthaya S, Diop S, Villanueva L, Stams AJM, Sánchez-Andrea I. Anaerobic degradation of sulfated polysaccharides by two novel *Kiritimatiellales* strains isolated from Black Sea sediment. *Front Microbiol*. 2019;10:253.

233. Orellana LH, Francis TB, Ferraro M, Hehemann JH, Fuchs BM, Amann RI. *Verrucomicrobiota* are specialist consumers of sulfated methyl pentoses during diatom blooms. *ISME J*. 2022;16(3):630-41.

234. Glöckner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W, et al. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc Natl Acad Sci U S A*. 2003;100(14):8298-303.

235. Stam M, Lelièvre P, Hoebeke M, Corre E, Barbeyron T, Michel G. SulfAtlas, the sulfatase database: state of the art and new developments. *Nucleic Acids Res*. 2023;51(D1):D647-D53.

236. Hehemann J-H, Correc G, Thomas F, Bernard T, Barbeyron T, Jam M, et al. Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. *J Biol Chem*. 2012;287:30571-84.

237. Zhu Y, McBride MJ. Deletion of the *Cytophaga hutchinsonii* type IX secretion system gene sprP results in defects in gliding motility and cellulose utilization. *Appl Microbiol Biotechnol*. 2014;98:763-75.

238. Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron EA, Pudlo NA, et al. Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature*. 2015;517:165-9.

239. Allison SD. Cheaters, diffusion and nutrients constrain decomposition by microbial enzymes in spatially structured environments. *Ecol Lett*. 2005;8(6):626-35.

240. Dunny GM, Brickman TJ, Dworkin M. Multicellular behavior in bacteria: communication, cooperation, competition and cheating. *BioEssays*. 2008; 30(4):296-8.

241. Arnosti C, Reintjes G, Amann R. A mechanistic microbial underpinning for the size-reactivity continuum of DOC degradation. *Mar Chem*. 2018;206(20):93-9.

242. Reintjes G, Arnosti C, Fuchs B, Amann R. Selfish, sharing and scavenging bacteria in the Atlantic Ocean: a biogeographical study of bacterial substrate utilisation. *ISME J*. 2019;13(5):1119-32.

243. Reintjes G, Fuchs BM, Amann R, Arnosti C. Extensive microbial processing of polysaccharides in the South Pacific Gyre via selfish uptake and extracellular hydrolysis. *Front Microbiol*. 2020;11:1-14.
244. Reintjes G, Fuchs BM, Scharfe M, Wiltshire KH, Amann R, Arnosti C. Short-term changes in polysaccharide utilization mechanisms of marine bacterioplankton during a spring phytoplankton bloom. *Environ Microbiol Rep*. 2020;22(5):1884-900.
245. Bell W, Mitchell R. Chemotactic and growth responses of marine bacteria to algal extracellular products. *Biol Bull*. 1972;143(2):265-77.
246. Fischer WW, Hemp J, Johnso JE. Evolution of oxygenic photosynthesis. *Annu Rev Earth Planet Sci*. 2016;44:647-83.
247. Schopf JW, editor. Life's Origin: The Beginnings of Biological Evolution. Oakland, CA: 2002. Available from: California Scholarship Online; 2012 Mar 22.
248. Knoll AH, Summons RE, Waldbauer JR, Zumberge JE. The geological succession of primary producers in the oceans. In: Falkowski PG, Knoll AH, eds. Evolution of primary producers in the sea. Burlington, MA: Elsevier; 2007. pp. 133-63.
249. Riding JB, Fensome RA, Soyer-Gobillard M-O, Medlin LK. A review of the dinoflagellates and their evolution from fossils to modern. *J Mar Sci Eng*. 2023;11(1):1.
250. Benoiston A-S, Ibarbalz FM, Bittner L, Guidi L, Jahn O, Dutkiewicz S, et al. The evolution of diatoms and their biogeochemical functions. *Philos Trans R Soc Lond B Biol Sci*. 2017;372(1728):20160397.
251. Claxton LM, McClelland HLO, Hermoso M, Rickaby REM. Eocene emergence of highly calcifying coccolithophores despite declining atmospheric CO_2. *Nat Geosci*. 2022;15:826-31.
252. Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. Algae acquire vitamin B_{12} through a symbiotic relationship with bacteria. *Nature*. 2005;438:90-3.
253. Gonzalez LE, Bashan Y. Increased growth of the microalga *Chlorella vulgaris* when coimmobilized and cocultured in alginate beads with the plant-growth-promoting bacterium *Azospirillum brasilense*. *Appl Environ Microbiol*. 2000;66(4):1527-31.

254. Løvdal T, Eichner C, Grossart H-P, Carbonnel V, Chou L, Martin-Jézéquel V, et al. Competition for inorganic and organic forms of nitrogen and phosphorous between phytoplankton and bacteria during an *Emiliania huxleyi* spring bloom. *Biogeosci*. 2008;5(2):371-83.

255. Hollants J, Leroux O, Leliaert F, Decleyre H, De Clerck O, Willems A. Who is in there? Exploration of endophytic bacteria within the siphonous green seaweed Bryopsis (*Bryopsidales*, *Chlorophyta*). *PLoS ONE*. 2011;6(10):e26458.

256. Goecke F, Labes A, Wiese J, Imhoff JF. Chemical interactions between marine macroalgae and bacteria. *Mar Ecol Prog Ser*. 2010;409:267-99.

257. González JE, Keshavan ND. Messing with bacterial quorum sensing. *Microbiol Mol Biol Rev*. 2006;70(4):859-75.

258. Romero M, Martin-Cuadrado AB, Roca-Rivada A, Cabello AM, Otero A. Quorum quenching in cultivable bacteria from dense marine coastal microbial communities. *FEMS Microbiol Ecol*. 2011;75(2):205-17.

259. Nylund GM, Pavia H. Chemical versus mechanical inhibition of fouling in the red alga *Dilsea carnosa*. *Mar Ecol Prog Ser*. 2005;299:111-21.

260. Grueneberg J, Engelen AH, Costa R, Wichard T. Macroalgal morphogenesis induced by waterborne compounds and bacteria in coastal seawater. *PLoS ONE*. 2016;11(1):e0146307.

261. Liang Z, Liu F, Wang W, Zhang P, Sun X, Wang F, et al. High-throughput sequencing revealed differences of microbial community structure and diversity between healthy and diseased *Caulerpa lentillifera*. *BMC Microbiology*. 2019;19(1):1-15.

262. Case RJ, Longford SR, Campbell AH, Low A, Tujula N, Steinberg PD, et al. Temperature induced bacterial virulence and bleaching disease in a chemically defended marine macroalga. *Environ Microbiol*. 2011;13(2):529-37.

263. Wright JT, Nys RD, Steinberg PD. Geographic variation in halogenated furanones from the red alga *Delisea pulchra* and associated herbivores and epiphytes. *Mar Ecol Prog Ser*. 2000;207:227-41.

264. Martin M, Barbeyron T, Martin R, Portetelle D, Michel G, Vandenbol M. The cultivable surface microbiota of the brown alga *Ascophyllum nodosum* is enriched in macroalgal-polysaccharide-degrading bacteria. *Front Microbiol*. 2015;6:1487.
265. Dummermuth A, Wiltshire KH, Kirstein I, Brodte EM, Wichels A, Shama L, et al. Marine Stations Helgoland and Sylt operated by the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research. *Journal of large-scale research facilities JLSRF*. 2023;8(1):A184.
266. Merder J, Röder H, Dittmar T, Feudel U, Freund JA, Gerdts G, et al. Dissolved organic compounds with synchronous dynamics share chemical properties and origin. *Limnol Oceanogr*. 2021;66(11): 4001-16.
267. Gerdts G, Wichels A, Döpke H, Klings K-W, Gunkel W, Schütt C. 40-year long-term study of microbial parameters near Helgoland (German Bight, North Sea): historical view and future perspectives. *Helgoland Mar Res*. 2004;58(4):230-42.
268. Wiltshire KH, Kraberg A, Bartsch I, Boersma M, Franke HD, Freund J, et al. Helgoland Roads, North Sea: 45 years of change. *Estuar Coast*. 2010;33(2):295-310.
269. Miksch S, Meiners M, Meyerdierks A, Probandt D, Wegener G, Titschack J, et al. Bacterial communities in temperate and polar coastal sands are seasonally stable. *ISME Commun*. 2021;1(1):29.
270. Rubinetti S, Fofonova V, Arnone E, Wiltshire KH. A complete 60-year catalog of wind events in the German Bight (North Sea) derived from ERA5 reanalysis data. *Earth Space Sci*. 2023;10(10):e2023EA003020.
271. Schrum C. Thermohaline stratification and instabilities at tidal mixing fronts: results of an eddy resolving model for the German Bight. *Cont Shelf Res*. 1997;17(6):689-716.
272. Wiltshire KH, Malzahn AM, Wirtz K, Greve W, Janisch S, Mangelsdorf P, et al. Resilience of North Sea phytoplankton spring bloom dynamics: An analysis of long-term data at Helgoland Roads. *Limnol Oceanogr*. 2008;53(4):1294-302.

273. 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.

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

275. Bunse C, Koch H, Breider S, Simon M, Wietz M. Sweet spheres: succession and CAZyme expression of marine bacterial communities colonizing a mix of alginate and pectin particles. *Environ Microbiol*. 2021;23(6):3130-48.

276. Bartsch I, Kuhlenkamp R. The marine macroalgae of Helgoland (North Sea): an annotated list of records between 1845 and 1999. *Helgol Mar Res*. 2000;54:160-89.

277. Uhl F, Bartsch I, Oppelt N. Submerged kelp detection with hyperspectral data. *Remote Sens Environ*. 2016;8(6):487.

278. Koch H, Durwald A, Schweder T, Noriega-Ortega B, Vidal-Melgosa S, Hehemann JH, et al. Biphasic cellular adaptations and ecological implications of *Alteromonas macleodii* degrading a mixture of algal polysaccharides. *ISME J*. 2019;13(1):92-103.

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

280. Curson AR, Liu J, Martínez AB, Green RT, Chan Y, Carrión O, et al. Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process. *Nat Microbiol*. 2017;2:17009.

281. Lewis WH, Tahon G, Geesink P, Sousa DZ, Ettema TJG. Innovations to culturing the uncultured microbial majority. *Nat Rev Microbiol*. 2021;19(4):225-40.

282. 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.

283. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*. 1977;74(11):5088-90.

284. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012;6(8):1621-4.

285. Hoffmann C, Hill DA, Minkah N, Kirn T, Troy A, Artis D, et al. Community-wide response of the gut microbiota to enteropathogenic *Citrobacter rodentium* infection revealed by deep sequencing. *Infect Immun*. 2009;77(10):4668-78.

286. Alcon-Giner C, Caim S, Mitra S, Ketskemety J, Wegmann U, Wain J, et al. Optimisation of 16S rRNA gut microbiota profiling of extremely low birth weight infants. *BMC Genomics*. 2017;18(1):841.

287. Nossa CW, Oberdorf WE, Yang L, Aas JA, Paster BJ, Desantis TZ, et al. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. *World J Gastroenterol*. 2010;16(33):4135-44.

288. Jumpstart Consortium Human Microbiome Project Data Generation Working Group. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS ONE*. 2012;7(6):e39315.

289. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature*. 2012;486(7402):215-21.

290. Fadeev E, Cardozo-Mino MG, Rapp JZ, Bienhold C, Salter I, Salman-Carvalho V, et al. Comparison of two 16S rRNA primers (V3-V4 and V4-V5) for studies of arctic microbial communities. *Front Microbiol*. 2021;12:637526.

291. Claesson MJ, O'Sullivan O, Wang Q, Nikkilä J, Marchesi JR, Smidt H, et al. Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS ONE*. 2009;4(8):e6669.

292. Ghyselinck J, Pfeiffer S, Heylen K, Sessitsch A, De Vos P. The effect of primer choice and short read sequences on the outcome of 16S rRNA gene based diversity studies. *PLoS ONE*. 2013;8(8):e71360.

293. Babauta JT, Atci E, Ha PT, Lindemann SR, Ewing T, Call DR, et al. Localized electron transfer rates and microelectrode-based enrichment of microbial communities within a phototrophic microbial mat. *Front Microbiol*. 2014;5:11.

294. Singer E, Bushnell B, Coleman-Derr D, Bowman B, Bowers RM, Levy A, et al. High-resolution phylogenetic microbial community profiling. *ISME J*. 2016;10(8):2020-32.

295. Zhang T, Li H, Ma S, Cao J, Liao H, Huang Q, et al. The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. *Appl Environ Microbiol*. 2023;89(10):e0060523.

296. Workman RE, Tang AD, Tang PS, Jain M, Tyson JR, Razaghi R, et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat Methods*. 2019;16(12):1297-305.

297. Karst SM, Ziels RM, Kirkegaard RH, Sørensen EA, McDonald D, Zhu Q, et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat Methods*. 2021;18(2):165-9.

298. Rodríguez-Pérez H, Ciuffreda L, Flores C. NanoCLUST: a species-level analysis of 16S rRNA nanopore sequencing data. *Bioinformatics*. 2021;37(11):1600-1.

299. Curry KD, Wang Q, Nute MG, Tyshaieva A, Reeves E, Soriano S, et al. Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat Methods*. 2022;19(7):845-53.

300. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73(16):5261-7.

301. Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet*. 2008;4(11):e1000255.

302. Liu Z, DeSantis TZ, Andersen GL, Knight R. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res*. 2008;36(18):e120.

303. Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol*. 2005;71(3):1501-6.

304. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75(23):7537-41.

305. Huse SM, Welch DM, Morrison HG, Sogin ML. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol*. 2010;12(7):1889-98.

306. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, et al. Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol*. 2013;4(12):1111-9.

307. Tikhonov M, Leach RW, Wingreen NS. Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. *ISME J*. 2015;9(1):68-80.

308. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J*. 2015;9(4):968-79.

309. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581-3.

310. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J*. 2017;11(12):2639-43.

311. 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.

312. Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. *Hum Mol Genet*. 2018;27(R2):R234-R41.

313. Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol*. 2019;37(10):1155-62.

314. Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sorensen EA, Wollenberg RD, et al. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods*. 2022;19:823-6.
315. Orellana LH, Krüger K, Sidhu C, Amann R. Comparing genomes recovered from time-series metagenomes using long- and short-read sequencing technologies. *Microbiome*. 2023;11:105.
316. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol*. 2017;35:725-31.
317. Manchester M, Anand A. Chapter two - Metabolomics: strategies to define the role of metabolism in virus infection and pathogenesis. *Adv Virus Res*. 2017;98:57-81.
318. Bauermeister A, Mannochio-Russo H, Costa-Lotufo LV, Jarmusch AK, Dorrestein PC. Mass spectrometry-based metabolomics in microbiome investigations. *Nat Rev Microbiol*. 2022;20:143-60.
319. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev*. 2007;26:51-78.
320. Gowda GA, Djukovic D. Overview of mass spectrometry-based metabolomics: Opportunities and challenges. *Methods Mol Biol*. 2014;1198:3-12.
321. Aksenov AA, da Silva R, Knight R, Lopes NP, Dorrestein PC. Global chemical analysis of biology by mass spectrometry. *Nat Rev Chem*. 2017;1:0054.
322. Misra BB. The connection and disconnection between microbiome and metabolome: A critical appraisal in clinical research. *Biol Res Nurs*. 2020;22(4):561-76.
323. Panschin I, Becher M, Verbarg S, Sproer C, Rohde M, Schuler M, et al. Description of *Gramella forsetii* sp. nov., a marine *Flavobacteriaceae* isolated from North Sea water, and emended description of *Gramella gaetbulicola* Cho et al. 2011. *Int J Syst Evol Microbiol*. 2017;67(3):697-703.

324. Sun C, Fu GY, Zhang CY, Hu J, Xu L, Wang RJ, et al. Isolation and complete genome sequence of *Algibacter alginolytica* sp. nov., a novel seaweed-degrading *Bacteroidetes* bacterium with diverse putative polysaccharide utilization loci. *Appl Environ Microbiol*. 2016;82(10):2975-87.

325. Wong SK, Park S, Lee JS, Lee KC, Ogura Y, Hayashi T, et al. *Algibacter aquaticus* sp. nov., a slightly alkaliphilic marine *Flavobacterium* isolated from coastal surface water. *Int J Syst Evol Microbiol*. 2017;67(7):2199-204.

326. Jung YJ, Lee YM, Baek K, Hwang CY, Cho Y, Hong SG, et al. *Algibacter psychrophilus* sp. nov., a psychrophilic bacterium isolated from marine sediment. *Int J Syst Evol Microbiol*. 2015;65(Pt_6):1735-40.

327. Zhong XC, Xu W, Zhang Y, Zhang QQ, Du ZJ. *Algibacter marinivivus* sp. nov., isolated from the surface of a marine red alga. *Int J Syst Evol Microbiol*. 2020;70(9):5048-53.

328. Nedashkovskaya OI, Kim SB, Han SK, Rhee MS, Lysenko AM, Rohde M, et al. *Algibacter lectus* gen. nov., sp. nov., a novel member of the family *Flavobacteriaceae* isolated from green algae. *Int J Syst Evol Microbiol*. 2004;54(4):1257-61.

329. Yoon JH, Park S. *Algibacter wandonensis* sp. nov., isolated from sediment around a brown algae (*Undaria pinnatifida*) reservoir. *Int J Syst Evol Microbiol*. 2013;63(Pt_12):4771-6.

330. Thomas F, Barbeyron T, Tonon T, Génicot S, Czjzek M, Michel G. Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine *Flavobacteriia* to their independent transfers to marine *Proteobacteria* and human gut *Bacteroides*. *Environ Microbiol*. 2012;14(9):2379-94.

331. Groisillier A, Labourel A, Michel G, Tonon T. The mannitol utilization system of the marine bacterium *Zobellia galactanivorans*. *Appl Environ Microbiol*. 2015;81(5):1799-812.

332. 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(12):2201-13.

333. Mann AJ, Hahnke RL, Huang SX, Werner J, Xing P, Barbeyron T, et al. The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901^T reveals a broad potential for degradation of algal polysaccharides. *Appl Environ Microbiol*. 2013;79:6813-22.
334. 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.
335. Liu Y, Blain S, Crispi O, Rembauville M, Obernosterer I. Seasonal dynamics of prokaryotes and their associations with diatoms in the Southern Ocean as revealed by an autonomous sampler. *Environ Microbiol*. 2020;22(9):3968-84.
336. Xu N, Wang W, Xu K, Xu Y, Ji D, Chen C, et al. Cultivation of different seaweed species and seasonal changes cause divergence of the microbial community in coastal seawaters. *Front Microbiol*. 2022;13:988743.
337. Naik A, Smithers M, Moisander PH. Impacts of UV-C irradiation on marine biofilm community succession. *Appl Environ Microbiol*. 2022;88(4):e0229821.
338. Heins A, Amann RI, Harder J. Cultivation of particle-associated heterotrophic bacteria during a spring phytoplankton bloom in the North Sea. *Syst Appl Microbiol*. 2021;44(5):126232.
339. Khan SA, Jeong SE, Baek JH, Jeon CO. *Maribacter algicola* sp. nov., isolated from a marine red alga, *Porphyridium marinum*, and transfer of *Maripseudobacter aurantiacus* Chen et al. 2017 to the genus *Maribacter* as *Maribacter aurantiacus* comb. nov.. *Int J Syst Evol Microbiol*. 2020;70(2):797-804.
340. Zhang JY, Xia Y, Feng X, Mu DS, Du ZJ. *Maribacter algarum* sp. nov., a new member of the family *Flavobacteriaceae* isolated from the red alga *Gelidium amansii*. *Int J Syst Evol Microbiol*. 2020;70(6):3679-85.
341. Nedashkovskaya OI, Kim SB, Han SK, Lysenko AM, Rohde M, Rhee MS, et al. *Maribacter* gen. nov., a new member of the family *Flavobacteriaceae*, isolated from marine habitats, containing the species *Maribacter sedimenticola* sp. nov.,

Maribacter aquivivus sp. nov., *Maribacter orientalis* sp. nov. and *Maribacter ulvicola* sp. nov.. *Int J Syst Evol Microbiol*. 2004;54(4):1017-23.

342. Cho KH, Hong SG, Cho HH, Lee YK, Chun J, Lee HK. *Maribacter arcticus* sp. nov., isolated from Arctic marine sediment. *Int J Syst Evol Microbiol*. 2008;58(6):1300-3.
343. Burchard RP. Gliding motility of bacteria. *BioScience*. 1980;30:157-62.
344. Liao C, Liang X, Soupir ML, Jarboe LR. Cellular, particle and environmental parameters influencing attachment in surface waters: a review. *J Appl Microbiol*. 2015;119(2):315-30.
345. Enke TN, Leventhal GE, Metzger M, Saavedra JT, Cordero OX. Microscale ecology regulates particulate organic matter turnover in model marine microbial communities. *Nat Commun*. 2018;9:2743.
346. Kappelmann L, Krüger K, Hehemann JH, Harder J, Markert S, Unfried F, et al. Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J*. 2019;13(1):76-91.
347. Wolter LA, Mitulla M, Kalem J, Daniel R, Simon M, Wietz M. CAZymes in *Maribacter dokdonensis* 62-1 from the patagonian shelf: genomics and Physiology compared to related Flavobacteria and a co-occurring *Alteromonas* strain. *Front Microbiol*. 2021;12:628055.
348. Ghaderiardakani F, Coates JC, Wichard T. Bacteria-induced morphogenesis of *Ulva intestinalis* and *Ulva mutabilis* (Chlorophyta): a contribution to the lottery theory. *FEMS Microbiol Ecol*. 2017;93(8).
349. Alsufyani T, Califano G, Deicke M, Grueneberg J, Weiss A, Engelen AH, et al. Macroalgal-bacterial interactions: identification and role of thallusin in morphogenesis of the seaweed *Ulva* (Chlorophyta). *J Exp Bot*. 2020;71(11):3340-49.
350. Ulrich JF, Gräfe MS, Dhiman S, Wienecke P, Arndt HD, Wichard T. Thallusin quantification in marine bacteria and algae cultures. *Mar Drugs*. 2022;20(11):690.

351. Jain A, Krishnan KP. Differences in free-living and particle-associated bacterial communities and their spatial variation in Kongsfjorden, Arctic. *J Basic Microbiol*. 2017;57(10):827-38.
352. Kirchman DL. The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiol Ecol*. 2002;39(2):91-100.
353. Sterling AR, Holland LZ, Bundy RM, Burns SM, Buck KN, Chappell PD, et al. Potential interactions between diatoms and bacteria are shaped by trace element gradients in the Southern Ocean. *Front Mar Sci*. 2023;9:876830.
354. Di Costanzo F, Di Dato V, Romano G. Diatom-bacteria interactions in the marine environment: complexity, heterogeneity, and potential for biotechnological applications. *Microorganisms*. 2023;11(12):2967.
355. Nowlan JP, Lumsden JS, Russell S. Advancements in characterizing *Tenacibaculum* infections in Canada. *Pathogens*. 2020;9(12):1029.
356. González JM, Simó R, Massana R, Covert JS, Casamayor EO, Pedrós-Alió C, et al. Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. *Appl Environ Microbiol*. 2000;66(10):4237-46.
357. Riemann L, Steward GF, Azam F. Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl Environ Microbiol*. 2000;66(2):578-87.
358. Kristyanto S, Kim KR, Jung J, Kim HM, Kim K, Jeon CO. *Tenacibaculum aquimarinum* sp. nov., isolated from a marine alga and seawater. *Int J Syst Evol Microbiol*. 2022;72(8):005477.
359. Fernandes N, Steinberg P, Rusch D, Kjelleberg S, Thomas T. Community structure and functional gene profile of bacteria on healthy and diseased thalli of the red seaweed *Delisea pulchra*. *PLoS ONE*. 2012;7(12):e50854.
360. Zhu Y, Thomas F, Larocque R, Li N, Duffieux D, Cladière L, et al. Genetic analyses unravel the crucial role of a horizontally acquired alginate lyase for brown algal biomass degradation by *Zobellia galactanivorans*. *Environ Microbiol*. 2017;19(6):2164-81.

361. Pérez-Pascual D, Lunazzi A, Magdelenat G, Rouy Z, Roulet A, Lopez-Roques C, et al. The complete genome sequence of the fish pathogen *Tenacibaculum maritimum* provides insights into virulence mechanisms. *Front Microbiol*. 2017;8:1542.

362. Tsertou MI, Triga A, Droubogiannis S, Kokkari C, Anasi G, Katharios P. Isolation and characterization of a novel *Tenacibaculum* species and a corresponding bacteriophage from a Mediterranean fish hatchery: description of *Tenacibaculum larymnensis* sp. nov. and *Tenacibaculum* phage Larrie. *Front Microbiol*. 2023;14:1078669.

363. Brunet M, Le Duff N, Barbeyron T, Thomas F. Consuming fresh macroalgae induces specific catabolic pathways, stress reactions and Type IX secretion in marine flavobacterial pioneer degraders. *ISME J*. 2022;16(8):2027-39.

364. Alejandre-Colomo C, Viver T, Urdiain M, Francis B, Harder J, Kämpfer P, et al. Taxonomic study of nine new *Winogradskyella* species occurring in the shallow waters of Helgoland Roads, North Sea. Proposal of *Winogradskyella schleiferi* sp. nov., *Winogradskyella costae* sp. nov., *Winogradskyella helgolandensis* sp. nov., *Winogradskyella vidalii* sp. nov., *Winogradskyella forsetii* sp. nov., *Winogradskyella ludwigii* sp. nov., *Winogradskyella ursingii* sp. nov., *Winogradskyella wichelsiae* sp. nov., and Candidatus "Winogradskyella atlantica" sp. nov.. *Syst Appl Microbiol*. 2020;43(6):126128.

365. Xu Y, Li J, Hu Y, Li H, Peng T, Zhong M, et al. *Winogradskyella endarachnes* sp. nov., a marine bacterium isolated from the brown alga *Endarachne binghamiae*. *Int J Syst Evol Microbiol*. 2021;71(1):004577.

366. Alejandre-Colomo C, Francis B, Viver T, Harder J, Fuchs BM, Rossello-Mora R, et al. Cultivable *Winogradskyella* species are genomically distinct from the sympatric abundant candidate species. *ISME Commun*. 2021;1(1):51.

367. Rodrigues V, Onime L, Huws S, Odaneth A, Lali A. Diversity of ulvan and cellulose depolymerizing bacteria associated with the green macroalgae *Ulva* spp. *J Appl Biotechnol Bioeng*. 2017;2(4):136-42.

368. Hinojosa VS, Asenjo J, Leiva S. Agarolytic culturable bacteria associated with three antarctic subtidal macroalgae. *World J Microbiol Biotechnol*. 2018;34(6):73.

369. Vacant S, Benites LF, Salmeron C, Intertaglia L, Norest M, Cadoudal A, et al. Long-term stability of bacterial associations in a microcosm of *Ostreococcus tauri* (*Chlorophyta*, *Mamiellophyceae*). *Front Plant Sci*. 2022;13:814386.

370. Leiva S, Alvarado P, Huang Y, Wang J, Garrido I. Diversity of pigmented Gram-positive bacteria associated with marine macroalgae from Antarctica. *FEMS Microbiol Lett*. 2015;362(24):fnv206.

371. Alvarado P, Huang Y, Wang J, Garrido I, Leiva S. Phylogeny and bioactivity of epiphytic Gram-positive bacteria isolated from three co-occurring antarctic macroalgae. *Antonie Van Leeuwenhoek*. 2018;111(9):1543-55.

372. Tourneroche A, Lami R, Hubas C, Blanchet E, Vallet M, Escoubeyrou K, et al. Bacterial-fungal interactions in the kelp endomicrobiota drive autoinducer-2 quorum sensing. *Front Microbiol*. 2019;10:1693.

373. Sims LP, Lockwood CWJ, Crombie AT, Bradley JM, Le Brun NE, Murrell JC. Purification and characterization of the isoprene monooxygenase from *Rhodococcus* sp. strain AD45. *Appl Environ Microbiol*. 2022;88(7):e0002922.

374. van Hylckama Vlieg JE, Kingma J, van den Wijngaard AJ, Janssen DB. A glutathione S-transferase with activity towards cis-1, 2-dichloroepoxyethane is involved in isoprene utilization by *Rhodococcus* sp. strain AD45. *Appl Environ Microbiol*. 1998;64(8):2800-5.

375. Guenther AB, Jiang X, Heald CL, Sakulyanontvittaya T, Duhl T, Emmons LK, et al. The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. *Geosci Model Dev*. 2012;5:1471-92.

376. Kirschke S, Bousquet P, Ciais P, Saunois M, Canadell JG, Dlugokencky EJ, et al. Three decades of global methane sources and sinks. *Nat Geosci*. 2013;6:813-23.

377. Guenther A. A global model of natural volatile organic compound emissions. *J Geophys Res*. 1995;100(D5):8873-92.

378. Broadgate WJ, Malin G, Küpper FC, Thompson A, Liss PS. Isoprene and other non-methane hydrocarbons from seaweeds: a source of reactive hydrocarbons to the atmosphere. *Mar Chem*. 2004;88(1-2):61-73.

379. Exton DA, McGenity TJ, Steinke M, Smith DJ, Suggett DJ. Uncovering the volatile nature of tropical coastal marine ecosystems in a changing world. *Glob Chang Biol*. 2015;21(4):1383-94.

380. Kuzma J, Nemecek-Marshall M, Pollock WH, Fall R. Bacteria produce the volatile hydrocarbon isoprene. *Curr Microbiol*. 1995;30(2):97-103.

381. Julsing MK, Rijpkema M, Woerdenbag HJ, Quax WJ, Kayser O. Functional analysis of genes involved in the biosynthesis of isoprene in *Bacillus subtilis*. *Appl Microbiol Biotechnol*. 2007;75(6):1377-84.

382. Shaw SL, Gantt B, Meskhidze N. Production and emissions of marine isoprene and monoterpenes: a review. *Adv Meteorol*. 2010:Article ID 408696. doi: https://doi.org/10.1155/2010/408696.

383. Booge D, Schlundt C, Bracher A, Endres S, Zäncker B, Marandino CA. Marine isoprene production and consumption in the mixed layer of the surface ocean—a field study over two oceanic regions. *Biogeosci*. 2018;15(2):649-67.

384. Liang HY, Sun C, Xu L, Xu XW, Cheng H, Wan Y. *Pontibrevibacter nitratireducens* gen. nov., sp. nov., a member of the family *Rhodobacteraceae* isolated from seawater of the Indian Ocean and intertidal zone. *Int J Syst Evol Microbiol*. 2022;72(5):005341.

385. Pohlner M, Dlugosch L, Wemheuer B, Mills H, Engelen B, Reese BK. The majority of active *Rhodobacteraceae* in marine sediments belong to uncultured genera: A molecular approach to link their distribution to environmental conditions. *Front Microbiol*. 2019;10:659.

386. Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M, et al. Phylogenomics of *Rhodobacteraceae* reveals evolutionary adaptation to marine and non-marine habitats. *ISME J*. 2017;11(6):1483-99.

387. Elifantz H, Horn G, Ayon M, Cohen Y, Minz D. *Rhodobacteraceae* are the key members of the microbial community of the initial biofilm formed in Eastern Mediterranean coastal seawater. *FEMS Microbiol Ecol*. 2013;85(2):348-57.

388. Preston CM, Durkin CA, Yamahara KM. DNA metabarcoding reveals organisms contributing to particulate matter flux to abyssal depths in the North East Pacific ocean. *Deep Sea Res Part II Top Stud Oceanogr*. 2020;173:104708.

389. Baumas CMJ, Ababou F-E, Garel M, Bizic M, Ionescu D, Puzenat A, et al. A novel method to sample individual marine snow particles for downstream molecular analyses. *Limnol Oceanogr: Methods*. 2024;22(1):34-46.

390. Mönnich J, Tebben J, Bergemann J, Case R, Wohlrab S, Harder T. Niche-based assembly of bacterial consortia on the diatom *Thalassiosira rotula* is stable and reproducible. *ISME J*. 2020;14(6):1614-25.

391. Florez JZ, Camus C, Hengst MB, Buschmann AH. A functional perspective analysis of macroalgae and epiphytic bacterial community interaction. *Front Microbiol*. 2017;8:2561.

392. Zecher K, Hayes KR, Philipp B. Evidence of interdomain ammonium cross-feeding from methylamine-and glycine betaine-degrading *Rhodobacteraceae* to diatoms as a widespread interaction in the marine phycosphere. *Front Microbiol*. 2020;11:533894.

393. Arandia-Gorostidi N, Krabberød AK, Logares R, Deutschmann IM, Scharek R, Morán XAG, et al. Novel interactions between phytoplankton and bacteria shape microbial seasonal dynamics in coastal ocean waters. *Front Mar Sci*. 2022;9:901201.

394. Dogs M, Wemheuer B, Wolter L, Bergen N, Daniel R, Simon M, et al. *Rhodobacteraceae* on the marine brown alga *Fucus spiralis* are abundant and show physiological adaptation to an epiphytic lifestyle. *Syst Appl Microbiol*. 2017;40(6):370-82.

395. Ashen JB, Goff LJ. Molecular and ecological evidence for species specificity and coevolution in a group of marine algal-bacterial symbioses. *Appl Environ Microbiol*. 2000;66:3024-30.

396. Barak-Gavish N, Dassa B, Kuhlisch C, Nussbaum I, Brandis A, Rosenberg G, et al. Bacterial lifestyle switch in response to algal metabolites. *eLife*. 2023;12:e84400.
397. Barak-Gavish N, Frada MJ, Ku C, Lee PA, DiTullio GR, Malitsky S, et al. Bacterial virulence against an oceanic bloom-forming phytoplankter is mediated by algal DMSP. *Sci Adv*. 2018;4(10):eaau5716.
398. Liss PS, Malin G, Turner SM, Holligan PM. Dimethyl sulphide and *Phaeocystis*: a review. *J Mar Syst*. 1994;5(1):41-53.
399. Liang KYH, Orata FD, Boucher YF, Case RJ. Roseobacters in a sea of poly- and paraphyly: Whole genome-based taxonomy of the family *Rhodobacteraceae* and the proposal for the split of the "Roseobacter Clade" into a novel family, *Roseobacteraceae* fam. nov.. *Front Microbiol*. 2021;12:683109.
400. Lachnit T, Meske D, Wahl M, Harder T, Schmitz R. Epibacterial community patterns on marine macroalgae are host-specific but temporally variable. *Environ Microbiol*. 2011;13(3):655-65.
401. Heins A, Harder J. Particle-associated bacteria in seawater dominate the colony-forming microbiome on ZoBell marine agar. *FEMS Microbiol Ecol*. 2022;99(1):fiac151.
402. Reintjes G, Heins A, Wang C, Amann R. Abundance and composition of particles and their attached microbiomes along an Atlantic Meridional Transect. *Front Mar Sci*. 2023;10:1051510.
403. Wood G, Steinberg PD, Campbell AH, Vergés A, Coleman MA, Marzinelli EM. Host genetics, phenotype and geography structure the microbiome of a foundational seaweed. *Mol Ecol*. 2022;31(7):2189-206.
404. Weigel BL, Miranda KK, Fogarty EC, Watson AR, Pfistere CA. Functional insights into the kelp microbiome from metagenome-assembled genomes. *mSystems*. 2022;7(3):e01422-21.
405. Aromokeye DA, Willis-Poratti G, Wunder LC, Yin X, Wendt J, Richter-Heitmann T, et al. Macroalgae degradation promotes microbial iron reduction via electron shuttling in coastal Antarctic sediments. *Environ Int*. 2021;156:106602.

406. Fontanez KM, Eppley JM, Samo TJ, Karl DM, DeLong EF. Microbial community structure and function on sinking particles in the North Pacific Subtropical Gyre. *Front Microbiol*. 2015;6:469.

407. Longford SR, Tujula NA, Crocetti GR, Holmes AJ, Holmström C, Kjelleberg S, et al. Comparisons of diversity of bacterial communities associated with three sessile marine eukaryotes. *Aquat Microb Ecol*. 2007;48(3):217-29.

408. Pushpakumara BLDU, Tandon K, Willis A, Verbruggen H. Unravelling microalgal-bacterial interactions in aquatic ecosystems through 16S rRNA gene-based co-occurrence networks. *Sci Rep*. 2023;13:2743.

409. Chen J, Zang Y, Yang Z, Qu T, Sun T, Liang S, et al. Composition and functional diversity of epiphytic bacterial and fungal communities on marine macrophytes in an intertidal zone. *Front Microbiol*. 2022;13:839465.

410. Bland JA, Brock TD. The marine bacterium *Leucothrix mucor* as an algal epiphyte. *Mar Biol*. 1973;23:283-92.

411. Thiele S, Storesund JE, Fernández-Méndez M, Assmy P, Øvreås L. A winter-to-summer transition of bacterial and archaeal communities in Arctic sea ice. *Microorganisms*. 2022;10(8):1618.

412. Holmström C, Kjelleberg S. Marine *Pseudoalteromonas* species are associated with higher organisms and produce biologically active extracellular agents. *FEMS Microbiol Ecol*. 1999;30(4):285-93.

413. Van Truong L, Tuyen H, Helmke E, Binh LT, Schweder T. Cloning of two pectate lyase genes from the marine Antarctic bacterium *Pseudoalteromonas haloplanktis* strain ANT/505 and characterization of the enzymes. *Extremophiles*. 2001;5(1):35-44.

414. Xu F, Cha QQ, Zhang YZ, Chen XL. Degradation and utilization of alginate by marine *Pseudoalteromonas*: a review. *Appl Environ Microbiol*. 2021;87(17):e0036821.

415. Iijima S, Washio K, Okahara R, Morikawa M. Biofilm formation and proteolytic activities of *Pseudoalteromonas* bacteria that were isolated from fish farm sediments. *Microb Biotechnol*. 2009;2(3):361-69.

416. Offret C, Desriac F, Le Chevalier P, Mounier J, Jégou C, Fleury Y. Spotlight on antimicrobial metabolites from the marine bacteria *Pseudoalteromonas*: chemodiversity and ecological significance. *Mar Drugs*. 2016;14(7):129.
417. Skerratt JH, Bowman JP, Hallegraeff G, James S, Nichols PD. Algicidal bacteria associated with blooms of a toxic dinoflagellate in a temperate Australian estuary. *Mar Ecol Prog Ser*. 2002;244:1-15.
418. Amaro AM, Fuentes MS, Ogalde SR, Venegas JA, Suárez-Isla BA. Identification and characterization of potentially algal-lytic marine bacteria strongly associated with the toxic dinoflagellate *Alexandrium catenella*. *J Eukaryot Microbiol*. 2005;52(3):191-200.
419. Takemura AF, Chien DM, Polz MF. Associations and dynamics of *Vibrionaceae* in the environment, from the genus to the population level. *Front Microbiol*. 2014;5:38.
420. Zhu S, Wang X, Zhao W, Zhang Y, Song D, Cheng H, et al. Vertical dynamics of free-living and particle-associated vibrio communities in the eastern tropical Indian Ocean. *Front Microbiol*. 2023;14:1285670.
421. Dong S, Yang J, Zhang XY, Shi M, Song XY, Chen XL, et al. Cultivable alginate lyase-excreting bacteria associated with the Arctic brown alga *Laminaria*. *Mar Drugs*. 2012;10(11):2481-91.
422. Alejandre-Colomo C, Harder J, Fuchs BM, Rosselló-Móra R, Amann R. High-throughput cultivation of heterotrophic bacteria during a spring phytoplankton bloom in the North Sea. *Syst Appl Microbiol*. 2020;43(2):126066.
423. Hinsa-Leasure SM, Koid C, Tiedje JM, Schultzhaus JN. Biofilm formation by *Psychrobacter arcticus* and the role of a large adhesin in attachment to surfaces. *Appl Environ Microbiol*. 2013;79(13):3967-73.
424. Hoffmann K, Bienhold C, Buttigieg PL, Knittel K, Laso-Pérez R, Rapp JZ, et al. Diversity and metabolism of *Woeseiales* bacteria, global members of marine sediment communities. *ISME J*. 2020;14(4):1042-56.
425. Suzzi AL, Stat M, Gaston TF, Huggett MJ. Spatial patterns in host-associated and free-living bacterial communities across six temperate estuaries. *FEMS Microbiol Ecol*. 2023;99(7):fiad061.

426. de Jager K. Microorganisms associated with Ulva grown in abalone effluent water: implications for biosecurity. [Master's thesis]. University of Cape Town; 2019.
427. Tanaka N, Romanenko LA, Iino T, Frolova GM, Mikhailov VV. *Cocleimonas flava* gen. nov., sp. nov., a gammaproteobacterium isolated from sand snail (*Umbonium costatum*). *Int J Syst Evol Microbiol*. 2011;61(2):412-6.
428. Leprich DJ, Flood BE, Schroedl PR, Ricci E, Marlow JJ, Girguis PR, et al. Sulfur bacteria promote dissolution of authigenic carbonates at marine methane seeps. *ISME J*. 2021;15(7):2043-56.
429. Goecke F, Labes A, Wiese J, F. IJ. Chemical interactions between marine macroalgae and bacteria. Mar Ecol Prog Ser. 2010;409:267-300.
430. Brock TD. The genus *Leucothrix*. In: Starr MP, Stolp H, Trüper HG, Balows A, Schlegel HG, editors. *The prokaryotes*. A handbook on habitats, isolation and identification of bacteria, vol. I. Berlin: Springer-Verlag; 1981. pp. 400-8.
431. Thomas F, Le Duff N, Wu TD, Cébron A, Uroz S, Riera P, et al. Isotopic tracing reveals single-cell assimilation of a macroalgal polysaccharide by a few marine *Flavobacteria* and *Gammaproteobacteria*. *ISME J*. 2021;15(10):3062-75.
432. Sadusky TJ, Bullis RA. Experimental disinfection of lobster eggs infected with *Leucothrix mucor*. *Biol Bull*. 1994;187(2):254-55.
433. Roalkvam I, Drønen K, Dahle H, Wergeland HI. Microbial communities in a flow-through fish farm for lumpfish (*Cyclopterus lumpus* L.) during healthy rearing conditions. *Front Microbiol*. 2019;10:1594.
434. 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(12):3749-56.
435. Freitas S, Hatosy S, Fuhrman JA, Huse SM, Welch DB, Sogin ML, et al. Global distribution and diversity of marine *Verrucomicrobia*. *ISME J*. 2012;6(8):1499-505.
436. Chiang E, Schmidt ML, Berry MA, Biddanda BA, Burtner A, Johengen TH, et al. *Verrucomicrobia* are prevalent in north-temperate freshwater lakes and display class-level preferences between lake habitats. *PLoS ONE*. 2018;13(3):e0195112.

437. Rahlff J, Stolle C, Giebel HA, Mustaffa NIH, Wurl O, Herlemann PRD. Sea foams are ephemeral hotspots for distinctive bacterial communities contrasting sea-surface microlayer and underlying surface water. *FEMS Microbiol Ecol*. 2021;97(4):fiab035.

438. Niu J, Xie JJ, Guo TY, Fang HH, Zhang YM, Liao SY, et al. Comparison and evaluation of four species of macro-algaes as dietary ingredients in *Litopenaeus vannamei* under normal rearing and WSSV challenge conditions: effect on growth, Immune response, and intestinal microbiota. *Front Physiol*. 2018;9:1880.

439. Park J, Davis K, Lajoie G, Parfrey LW. Alternative approaches to identify core bacteria in *Fucus distichus* microbiome and assess their distribution and host-specificity. *Environ Microbiome*. 2022;17(1):55.

440. Qu T, Hou C, Zhao X, Zhong Y, Guan C, Lin Z, et al. Bacteria associated with *Ulva prolifera*: a vital role in green tide formation and migration. *Harmful Algae*. 2021;108:102104.

441. Elizondo-González R, Quiroz-Guzmán E, Howe A, Yang F, Flater J, Gemin M, et al. Changes on the intestinal bacterial community of white shrimp *Penaeus vannamei* fed with green seaweeds. *J Appl Phycol*. 2020;32:2061-70.

442. Kasai H, Katsuta A, Sekiguchi H, Matsuda S, Adachi K, Shindo K, et al. *Rubritalea squalenifaciens* sp. nov., a squalene-producing marine bacterium belonging to subdivision 1 of the phylum 'Verrucomicrobia'. *Int J Syst Evol Microbiol*. 2007;57(7):1630-34.

443. Yoon J, Matsuo Y, Matsuda S, Adachi K, Kasai H, Yokota A. *Rubritalea sabuli* sp. nov., a carotenoid- and squalene-producing member of the family *Verrucomicrobiaceae*, isolated from marine sediment. *Int J Syst Evol Microbiol*. 2008;58(4):992-7.

444. Bloch KE. Sterol structure and membrane function. *Crit Rev Biochem*. 1983;14(1):47-92.

445. Fries L. D-vitamins and their precursors as growth regulators in axenically cultivated marine macroalgae[1]. *J Phycol*. 1984;20(1):62-6.

446. Yoon J, Adachi K, Kasai H. Isolation and characterization of a novel marine *Bacteroidetes* as *Algitalea ulvae* gen. nov., sp. nov., isolated from the green alga *Ulva pertusa*. *Antonie Van Leeuwenhoek*. 2015;108(2):505-13.

447. Kopprio GA, Cuong LH, Luyen ND, Du TM, Ha TH, Huong LM, et al. Carrageenophyte-attached and planktonic bacterial communities in two distinct bays of Vietnam: eutrophication indicators and insights on ice-ice disease. *Ecol Indic*. 2021;121:107067.

448. Di Capua C, Bortolotti A, Farías ME, Cortez N. UV-resistant *Acinetobacter* sp. isolates from Andean wetlands display high catalase activity. *FEMS Microbiol Lett*. 2011;317(2):181-9.

449. McIlroy SJ, Nielsen PH. The family *Saprospiraceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, eds. The Prokaryotes. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. pp. 863-89.

450. Xia Y, Kong Y, Thomsen TR, Halkjær Nielsen P. Identification and ecophysiological characterization of epiphytic protein-hydrolyzing *Saprospiraceae* ("Candidatus Epiflobacter" spp.) in activated sludge. *Appl Environ Microbiol*. 2008;74(7):2229-38.

451. Sack EL, van der Wielen PW, van der Kooij D. Polysaccharides and proteins added to flowing drinking water at microgram-per-liter levels promote the formation of biofilms predominated by *Bacteroidetes* and *Proteobacteria*. *Appl Environ Microbiol*. 2014;80(8):2360-71.

452. Remple KL, Silbiger NJ, Quinlan ZA, Fox MD, Kelly LW, Donahue MJ, et al. Coral reef biofilm bacterial diversity and successional trajectories are structured by reef benthic organisms and shift under chronic nutrient enrichment. *npj Biofilms Microbiomes*. 2021;7(1):84.

453. Pavlovska M, Prekrasna I, Dykyi E, Zotov A, Dzhulai A, Frolova A, et al. Niche partitioning of bacterial communities along the stratified water column in the Black Sea. *Microbiologyopen*. 2021;10(3):e1195.

454. Baltar F, Palovaara J, Unrein F, Catala P, Horňák K, Šimek K, et al. Marine bacterial community structure resilience to changes in protist predation under phytoplankton bloom conditions. *ISME J*. 2016;10(3):568-81.
455. Kimbrel JA, Samo TJ, Ward C, Nilson D, Thelen MP, Siccardi A, et al. Host selection and stochastic effects influence bacterial community assembly on the microalgal phycosphere. *Algal Research*. 2019;40:101489.
456. Furusawa G, Yoshikawa T, Yasuda A, Sakata T. Algicidal activity and gliding motility of *Saprospira* sp. SS98-5. *Can J Microbiol*. 2003;49(2):92-100.
457. Mavromatis K, Chertkov O, Lapidus A, Nolan M, Lucas S, Tice H, et al. Permanent draft genome sequence of the gliding predator *Saprospira grandis* strain Sa g1 (= HR1). *Stand Genomic Sci*. 2012;6(2):210-9.
458. Sun T, Yang Z, Chen J, Li Y, Wang J, Wang X, et al. Effects of water loss stress under tidal effects on the epiphytic bacterial community of *Sargassum thunbergii* in the intertidal zone. *mSphere*. 2022;7(5):e0030722.
459. Miranda LN, Hutchison K, Grossman AR, Brawley SH. Diversity and abundance of the bacterial community of the red macroalga *Porphyra umbilicalis*: Did bacterial farmers produce macroalgae? *PLoS ONE*. 2013;8:e58269.
460. Korlević M, Markovski M, Zhao Z, Herndl GJ, Najdek M. Seasonal dynamics of epiphytic microbial communities on marine macrophyte surfaces. *Front Microbiol*. 2021;12:671342.
461. Weigel BL, Pfister CA. Successional dynamics and seascape-level patterns of microbial communities on the canopy-forming kelps *Nereocystis luetkeana* and *Macrocystis pyrifera*. *Front Microbiol*. 2019;10:346.
462. Johnston J, LaPara T, Behrens S. Composition and dynamics of the activated sludge microbiome during seasonal nitrification failure. *Sci Rep*. 2019;9(1):4565.
463. Kim NK, Oh S, Liu WT. Enrichment and characterization of microbial consortia degrading soluble microbial products discharged from anaerobic methanogenic bioreactors. *Water Res*. 2016;90:395-404.

464. 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.

465. Yan D, Xia P, Song X, Lin T, Cao H. Community structure and functional diversity of epiphytic bacteria and planktonic bacteria on submerged macrophytes in Caohai Lake, southwest of China. *Ann Microbiol*. 2019;69:933-44.

466. Kesy K, Oberbeckmann S, Kreikemeyer B, Labrenz M. Spatial environmental heterogeneity determines young biofilm assemblages on microplastics in Baltic Sea mesocosms. *Front Microbiol*. 2019;10:1665.

467. Hansel CM, Diaz JM. Production of extracellular reactive oxygen species by marine biota. *Ann Rev Mar Sci*. 2021;13:177-200.

468. Shi XL, Wu YH, Cheng H, Zhang XQ, Wang CS, Xu XW. Complete genome sequence of astaxanthin-producing bacterium *Altererythrobacter ishigakiensis* NBRC 107699. *Mar Genomics*. 2016;30:77-9.

469. Takaichi S, Shimada K, Ishidsu JI. Carotenoids from the aerobic photosynthetic bacterium, *Erythrobacter longus*: β-Carotene and its hydroxyl derivatives. *Arch Microbiol*. 1990;153:118-22.

470. Cho SH, Jeong Y, Lee E, Ko SR, Ahn CY, Oh HM, et al. Assessment of *Erythrobacter* species diversity through pan-genome analysis with newly isolated *Erythrobacter* sp. 3-20A1M. *J Microbiol Biotechnol*. 2021;31(4):601-9.

471. Yonemoto Y, Tanaka H, Hisano T, Sakaguchi K, Abe S, Yamashita T, et al. Bacterial alginate lyase gene: nucleotide sequence and molecular route for generation of alginate lyase species. *J Ferment Bioeng*. 1993;75(5):336-42.

472. Fuerst JA. The planctomycetes: emerging models for microbial ecology, evolution and cell biology. *Microbiol Spectr*. 1995;141:1493-506.

473. Tisserand L, Dadaglio L, Intertaglia L, Catala P, Panagiotopoulos C, Obernosterer I, et al. Use of organic exudates from two polar diatoms by bacterial isolates from the Arctic Ocean. *Philos Trans A Math Phys Eng Sci*. 2020;378(2181):20190356.

474. Schultz-Johansen M, Bech PK, Hennessy RC, Glaring MA, Barbeyron T, Czjzek M, et al. A novel enzyme portfolio for red algal polysaccharide degradation in the

marine bacterium *Paraglaciecola hydrolytica* S66^T encoded in a sizeable polysaccharide utilization locus. *Front Microbiol*. 2018;9:839.

475. Schultz D, Zühlke D, Bernhardt J, Francis TB, Albrecht D, Hirschfeld C, et al. An optimized metaproteomics protocol for a holistic taxonomic and functional characterization of microbial communities from marine particles. *Environ Microbiol Rep*. 2020;12(4):367-76.
476. Francis TB, Urich T, Mikolasch A, Teeling H, Amann R. North Sea spring bloom-associated *Gammaproteobacteria* fill diverse heterotrophic niches. *Environ Microbiome*. 2021;16:15.
477. Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, et al. TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res*. 2001;29(1):41-3.
478. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. *Nucleic Acids Res*. 2021;49(D1):D412-D9.
479. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46(W1):W95-W101.
480. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res*. 2021;49(D1):D274-D81.
481. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47(D1):D309-14.
482. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res*. 2023;51(W1):W115-W21.

483. Bertrand EM, McCrow JP, Moustafa A, Zheng H, McQuaid JB, Delmont TO, et al. Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc Natl Acad Sci U S A*. 2015;112(32):9938-43.

Appendix

Additional co-author publications

Novel laminarin-binding CBMs in multimodular proteins from marine *Bacteroidota* feature prominently in phytoplankton blooms

Marie-Katherin Zühlke[1,2*], Elizabeth Ficko-Blean[3], Daniel Bartosik[1,2], Nicolas Terrapon[4], Alexandra Jeudy[3], Murielle Jam[3], **Fengqing Wang**[5], Norma Welsch[1,2], Alexandra Dürwald[1,6], Laura Torres Martin[1], Robert Larocque[3], Diane Jouanneau[3], Tom Eisenack[1], François Thomas[3], Anke Trautwein-Schult[7], Hanno Teeling[5], Dörte Becher[7], Thomas Schweder[1,2], Mirjam Czjzek[3*]

[1] Pharmaceutical Biotechnology, Institute of Pharmacy, University Greifswald, Greifswald, 17487, Germany

[2] Institute of Marine Biotechnology, 17487 Greifswald, Germany

[3] Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), Sorbonne Université, CNRS, 29680 Roscoff, Bretagne, France

[4] Architecture et Fonction des Macromolécules Biologiques (AFMB), Aix-Marseille Université (AMU, UMR7257), CNRS, 13288 Marseille cedex 9, Marseille, France

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

[6] Helmholtz Institute for One Health, Helmholtz Centre for Infection Research HZI, Greifswald, Germany

[7] Microbial Proteomics, Institute of Microbiology, University Greifswald, 17487 Greifswald, Germany

[*]corresponding authors

Appendix

Mirjam Czjzek, e-mail: czjzek@sb-roscoff.fr

Marie-Katherin Zühlke, e-mail: marie-katherin.zuehlke@uni-greifswald.de

Abstract

The β(1,3)-glucan laminarin functions as storage polysaccharide in marine microalgae such as diatoms. Laminarin is abundant, water-soluble and structured simply, making it an attractive substrate for marine bacteria. As a consequence, many marine bacteria have developed competitive strategies to scavenge and decompose laminarin, which involves carbohydrate-binding modules (CBMs) as key players. We therefore functionally and structurally characterized two yet unassigned domains as laminarin-binding CBMs in multimodular proteins from *Christiangramia forsetii* KT0803^T, hereby unveiling the novel laminarin-binding CBM families CBMxx and CBMyy. We discovered four CBMxx repeats in a surface glycan-binding protein (SGBP) and a single CBMyy combined with a glycoside hydrolase module from family 16 (GH16_3). Our analyses revealed that both modular proteins feature an elongated shape, and that GH16_3 displayed a higher flexibility than SGBP. While flexibility of both polypeptide chains may facilitate recognition and/or degradation of laminarin, constraints in the SGBP may support docking of laminarin onto the bacterial surface. Exploration of bacterial metagenome-assembled genomes (MAGs) from phytoplankton blooms in the North Sea revealed that both laminarin-binding CBM families are widely distributed among marine *Bacteroidota*. High protein abundance of CBMxx- and CBMyy-containing proteins during phytoplankton blooms further underpins their relevance in marine laminarin usage.

Manuscript in Revision in Environmental Microbiology

Marine particle microbiomes during a spring diatom bloom contain active sulfate-reducing bacteria

Siebers, R.[1], Schultz, D.[1], Farza, M. S.[1], Brauer, A.[1], Zühlke, D.[1], Mücke, P. A[1], **Wang, F**.[2], Bernhardt, J.[1], Teeling, H.[2], Becher, D.[1], Riedel, K.[1,3], Kirstein, I. V.[4], Wiltshire, K. H.[4], Hoff, K.J.[5], Schweder, T.[3,6], Urich, T.[1,3], Bengtsson, M. M.[1,3,4*]

[1]Institute of Microbiology, University of Greifswald, Greifswald, Germany

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

[3]Institute of Marine Biotechnology, Greifswald, Germany

[4]Alfred-Wegener-Institute Helmholtz Centre for Polar and Marine Research, Biologische Anstalt Helgoland, Helgoland, Germany

[5]Institute of Mathematics and Computer Science, University of Greifswald, Greifswald, Germany

[6]Institute of Pharmacy, University of Greifswald, Greifswald, Germany

*** Correspondence:**

Mia M. Bengtsson

Institute of Microbiology, University of Greifswald

Felix-Hausdorff-Straße 8

17489 Greifswald, Germany

mia.bengtsson@uni-greifswald.de

Keywords: phytoplankton$_1$, diatoms$_2$, *Desulfobacterota*$_3$, networks$_4$, North Sea$_5$.

Abstract

Phytoplankton blooms fuel marine food webs with labile dissolved carbon and also lead to the formation of particulate organic matter composed of living and dead algal cells. These particles contribute to carbon sequestration and are sites of intense algal-bacterial

interactions, providing diverse niches for microbes to thrive. We analyzed 16S and 18S ribosomal RNA gene amplicon sequences obtained from 51 time points and metaproteomes from 3 time points during a spring phytoplankton bloom in a shallow location (6-10 m depth) in the North Sea. Particulate fractions larger than 10 μm diameter were collected at near daily intervals between early March and late May in 2018. Network analysis identified two major modules representing bacteria co-occurring with diatoms and with dinoflagellates, respectively. The diatom network module included known sulfate-reducing *Desulfobacterota* as well as potentially sulfur-oxidizing *Ectothiorhodospiraceae*. Metaproteome analyses confirmed presence of key enzymes involved in dissimilatory sulfate reduction, a process known to occur in sinking particles at greater depths and in sediments. Our results indicate the presence of sufficiently anoxic niches in the particle fraction of an active phytoplankton bloom to sustain sulfate reduction, and an important role of benthic-pelagic coupling for microbiomes in shallow environments. Our findings may have implications for the understanding of algal-bacterial interactions and carbon export during blooms in shallow-water coastal areas.

Manuscript in Revision in FEMS Microbiology Ecology

Marine Life Science & Technology (2023) 5:400–414
https://doi.org/10.1007/s42995-023-00187-w

RESEARCH PAPER

Short-chain fatty acids (SCFAs) as potential resuscitation factors that promote the isolation and culture of uncultured bacteria in marine sediments

Chun-Shui Sun[1,2,3] · Liu-Yan Zhou[4] · Qi-Yun Liang[1] · Xiao-Man Wang[5] · Yi-Xuan Lei[1] · Zhen-Xing Xu[6] · Feng-Qing Wang[7] · Guan-Jun Chen[1,2] · Zong-Jun Du[1,2,3] · Da-Shuai Mu[1,2,3]

Received: 1 February 2023 / Accepted: 20 June 2023 / Published online: 29 July 2023

Abstract

Many marine bacteria are difficult to culture because they are dormant, rare or found in low-abundances. Enrichment culturing has been widely tested as an important strategy to isolate rare or dormant microbes. However, many more mechanisms remain uncertain. Here, based on 16S rRNA gene high-throughput sequencing and metabolomics technology, it was found that the short-chain fatty acids (SCFAs) in metabolites were significantly correlated with uncultured bacterial groups during enrichment cultures. A pure culture analysis showed that the addition of SCFAs to media also resulted in high efficiency for the isolation of uncultured strains from marine sediments. As a result, 238 strains belonging to 10 phyla, 26 families and 82 species were successfully isolated. Some uncultured rare taxa within *Chlorobi* and *Kiritimatiellaeota* were successfully cultured. Amongst the newly isolated uncultured microbes, most genomes, e.g. bacteria, possess SCFA oxidative degradation genes, and these features might aid these microbes in better adapting to the culture media. A further resuscitation analysis of a viable but non-culturable (VBNC) *Marinilabiliales* strain verified that the addition of SCFAs could break the dormancy of *Marinilabiliales* in 5 days, and the growth curve test showed that the SCFAs could shorten the lag phase and increase the growth rate. Overall, this study provides new insights into SCFAs, which were first studied as resuscitation factors in uncultured marine bacteria. Thus, this study can help improve the utilisation and excavation of marine microbial resources, especially for the most-wanted or key players.

Keywords Marine bacteria · SCFAs · VBNC · Resuscitation culture

Introduction

Bacteria, which are ubiquitous in the marine environment, provide a valuable resource that is still rarely explored or utilised. Through molecular methods, especially 16S rRNA

Edited by Chengchao Chen.

✉ Zong-Jun Du
duzongjun@sdu.edu.cn

✉ Da-Shuai Mu
dashuai.mu@sdu.edu.cn

[1] Marine College, Shandong University, Weihai 264209, China

[2] State Key Laboratory of Microbial Technology, Institute of Microbial Technology, Shandong University, Qingdao 266237, China

[3] Weihai Research Institute of Industrial Technology of Shandong University, Weihai 264209, China

[4] Institute of Microbiology Applications, Xinjiang Academy of Agricultural Sciences, Urumqi 830000, China

[5] Tancheng County Inspection and Testing Center, Tancheng 276100, China

[6] Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo-Ku, Tokyo 113-8657, Japan

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

Appendix

Appendix Table

Appendix Table 1 Clades that were prominent in macroalgal phycospheres but not in PA communities.

Taxon
Bacteria;Acidobacteriota;AT-s3-28;NA;NA;NA
Bacteria;Actinobacteriota;Acidimicrobiia;Microtrichales;Microtrichaceae;NA
Bacteria;Bacteroidota;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides
Bacteria;Bacteroidota;Bacteroidia;Bacteroidales;Marinifilaceae;NA
Bacteria;Bacteroidota;Bacteroidia;Chitinophagales;NA;NA
Bacteria;Bacteroidota;Bacteroidia;Chitinophagales;Saprospiraceae;Aureispira
Bacteria;Bacteroidota;Bacteroidia;Chitinophagales;Saprospiraceae;Lewinella
Bacteria;Bacteroidota;Bacteroidia;Chitinophagales;Saprospiraceae;Portibacter
Bacteria;Bacteroidota;Bacteroidia;Chitinophagales;Saprospiraceae;Rubidimonas
Bacteria;Bacteroidota;Bacteroidia;Cytophagales;Cyclobacteriaceae;Fulvivirga
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Algitalea
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Aquimarina
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Croceitalea
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Dokdonia
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Hyunsoonleella
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Maritimimonas
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Pibocella
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Postechiella
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Wenyingzhuangia
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Weeksellaceae;Chryseobacterium
Bacteria;Bdellovibrionota;Bdellovibrionia;Bdellovibrionales;Bdellovibrionaceae;Bdellovibrio
Bacteria;Cyanobacteria;Cyanobacteriia;Cyanobacteriales;Xenococcaceae;Chroococcidiopsis PCC-6712
Bacteria;Cyanobacteria;Cyanobacteriia;Cyanobacteriales;Xenococcaceae;Pleurocapsa PCC-7319
Bacteria;Cyanobacteria;Cyanobacteriia;Synechococcales;Synechococcales Incertae Sedis;Schizothrix LEGE 07164
Bacteria;Deinococcota;Deinococci;Deinococcales;Trueperaceae;Truepera
Bacteria;Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Streptococcus
Bacteria;Firmicutes;Bacilli;Mycoplasmatales;Mycoplasmataceae;Candidatus Bacilloplasma
Bacteria;Firmicutes;Clostridia;Oscillospirales;Ruminococcaceae;Faecalibacterium
Bacteria;Firmicutes;Negativicutes;Veillonellales-Selenomonadales;Veillonellaceae;Veillonella

Appendix

Bacteria;Fusobacteriota;Fusobacteriia;Fusobacteriales;Fusobacteriaceae;Psychrilyobacter
Bacteria;Myxococcota;Polyangia;Polyangiales;Sandaracinaceae;Sandaracinus
Bacteria;Patescibacteria;Gracilibacteria;Absconditabacteriales (SR1);NA;NA
Bacteria;Patescibacteria;Gracilibacteria;JGI 0000069-P22;NA;NA
Bacteria;Patescibacteria;Parcubacteria;Candidatus Campbellbacteria;NA;NA
Bacteria;Patescibacteria;Parcubacteria;Candidatus Kaiserbacteria;NA;NA
Bacteria;Patescibacteria;Parcubacteria;Candidatus Moranbacteria;NA;NA
Bacteria;Patescibacteria;Parcubacteria;NA;NA;NA
Bacteria;Patescibacteria;Saccharimonadia;Saccharimonadales;NA;NA
Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacterales;Hyphomonadaceae;Hellea
Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacterales;Hyphomonadaceae;Litorimonas
Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacterales;Hyphomonadaceae;NA
Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacterales;Hyphomonadaceae;Robiginitomaculum
Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacterales;Parvularculaceae;Marinicaulis
Bacteria;Proteobacteria;Alphaproteobacteria;Micavibrionales;NA;NA
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;NA
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Paenochrobactrum
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Pseudahrensia
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Stappiaceae;Labrenzia
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Stappiaceae;Pseudovibrio
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Amylibacter
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Ascidiaceihabitans
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;HIMB11
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Lentibacter
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Marivita
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Octadecabacter
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Silicimonas
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Yoonia-Loktanella
Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Altererythrobacter
Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Erythrobacter
Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;NA
Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Sphingomonas
Bacteria;Proteobacteria;Gammaproteobacteria;Aeromonadales;Aeromonadaceae;Oceanisphaera

Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Marinobacteraceae;Marinobacter
Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Psychromonadaceae;Psychromonas
Bacteria;Proteobacteria;Gammaproteobacteria;Arenicellales;Arenicellaceae;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Burkholderiales;Comamonadaceae;Acidovorax
Bacteria;Proteobacteria;Gammaproteobacteria;Burkholderiales;Comamonadaceae;Delftia
Bacteria;Proteobacteria;Gammaproteobacteria;Burkholderiales;Neisseriaceae;Neisseria
Bacteria;Proteobacteria;Gammaproteobacteria;Cellvibrionales;Cellvibrionaceae;Marinagarivorans
Bacteria;Proteobacteria;Gammaproteobacteria;Ectothiorhodospirales;Ectothiorhodospiraceae;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Klebsiella
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Pasteurellaceae;Aggregatibacter
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Shewanellaceae;Psychrobium
Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Nitrincolaceae;Neptunomonas
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Cellvibrionaceae;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Marinomonadaceae;Marinomonas
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Moraxellaceae;Acinetobacter
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae;Pseudomonas
Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;Stenotrophomonas
Bacteria;SAR324 clade(Marine group B);NA;NA;NA;NA
Bacteria;Verrucomicrobiota;Verrucomicrobiae;Verrucomicrobiales;Akkermansiaceae;Akkermansia
Bacteria;Verrucomicrobiota;Verrucomicrobiae;Verrucomicrobiales;DEV007;NA
Bacteria;Verrucomicrobiota;Verrucomicrobiae;Verrucomicrobiales;Rubritaleaceae;Roseibacillus

Appendix Table 2 Clades that were prominent in PA communities but not in macroalgal phycospheres.

Taxon
Bacteria;Actinobacteriota;Acidimicrobiia;Microtrichales;Ilumatobacteraceae;Ilumatobacter
Bacteria;Bacteroidota;Bacteroidia;Bacteroidales;Bacteroidetes BD2-2;NA
Bacteria;Bacteroidota;Bacteroidia;Cytophagales;Cyclobacteriaceae;NA
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Crocinitomicaceae;Crocinitomix
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Cryomorphaceae;NA
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Cryomorphaceae;Vicingus
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Arcticiflavibacter
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Bizionia
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Cellulophaga
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Flavobacterium
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Formosa
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Lutibacter
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;NA
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Nonlabens
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;NS2b marine group
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;NS4 marine group
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;NS5 marine group
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;Flavobacteriaceae;Ulvibacter
Bacteria;Bacteroidota;Bacteroidia;Flavobacteriales;NS9 marine group;NA
Bacteria;Bdellovibrionota;Bdellovibrionia;Bdellovibrionales;Bdellovibrionaceae;OM27 clade
Bacteria;Campylobacterota;Campylobacteria;Campylobacterales;Sulfurovaceae;Sulfurovum
Bacteria;Chloroflexi;Anaerolineae;SBR1031;A4b;NA
Bacteria;Cyanobacteria;Cyanobacteriia;Cyanobacteriales;Chroococcidiopsaceae;Aliterella
Bacteria;Cyanobacteria;Cyanobacteriia;Synechococcales;Cyanobiaceae;Synechococcus CC9902
Bacteria;Desulfobacterota;Desulfobacteria;Desulfobacterales;Desulfosarcinaceae;Sva0081 sediment group
Bacteria;Desulfobacterota;Desulfobulbia;Desulfobulbales;Desulfocapsaceae;NA
Bacteria;Planctomycetota;OM190;NA;NA;NA
Bacteria;Planctomycetota;Phycisphaerae;Phycisphaerales;Phycisphaeraceae;CL500-3
Bacteria;Planctomycetota;Phycisphaerae;Phycisphaerales;Phycisphaeraceae;NA

Bacteria;Planctomycetota;Planctomycetes;Pirellulales;Pirellulaceae;Blastopirellula
Bacteria;Planctomycetota;Planctomycetes;Pirellulales;Pirellulaceae;NA
Bacteria;Proteobacteria;Alphaproteobacteria;Acetobacterales;Acetobacteraceae;Acidiphilium
Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacterales;Hyphomonadaceae;Hyphomonas
Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacterales;Hyphomonadaceae;Maricaulis
Bacteria;Proteobacteria;Alphaproteobacteria;Kiloniellales;Kiloniellaceae;Kiloniella
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Beijerinckiaceae;Methylobacterium-Methylorubrum
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Filomicrobium
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Stappiaceae;NA
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Planktomarina
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Pseudophaeobacter
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Tateyamaria
Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Fokiniaceae;MD3-55
Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;NA;NA
Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Candidatus Megaira
Bacteria;Proteobacteria;Alphaproteobacteria;SAR11 clade;Clade I;Clade Ia
Bacteria;Proteobacteria;Alphaproteobacteria;SAR11 clade;Clade II;NA
Bacteria;Proteobacteria;Alphaproteobacteria;Sneathiellales;Sneathiellaceae;Sneathiella
Bacteria;Proteobacteria;Gammaproteobacteria;Burkholderiales;Comamonadaceae;RS62 marine group
Bacteria;Proteobacteria;Gammaproteobacteria;Burkholderiales;Methylophilaceae;OM43 clade
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Alteromonadaceae;Glaciecola
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Alteromonadaceae;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Alteromonadaceae;Paraglaciecola
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Colwelliaceae;Thalassomonas
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Kangiellaceae;Aliikangiella
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Kangiellaceae;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Psychromonadaceae;Psychromonas
Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Vibrionaceae;Enterovibrio
Bacteria;Proteobacteria;Gammaproteobacteria;Francisellales;Francisellaceae;Francisella
Bacteria;Proteobacteria;Gammaproteobacteria;Gammaproteobacteria Incertae Sedis;Unknown Family;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Granulosicoccales;NA;NA
Bacteria;Proteobacteria;Gammaproteobacteria;NA;NA;NA

Appendix

Bacteria;Proteobacteria;Gammaproteobacteria;Nitrosococcales;Methylophagaceae;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Nitrincolaceae;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Alcanivoracaceae;Ketobacter
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Halieaceae;Halioglobus
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;KI89A clade;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;NA;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;OM182 clade;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Porticoccaceae;SAR92 clade
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;SAR86 clade;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Spongiibacteraceae;BD1-7 clade
Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Spongiibacteraceae;Oceanicoccus
Bacteria;Proteobacteria;Gammaproteobacteria;Thiotrichales;Thiotrichaceae;NA
Bacteria;Proteobacteria;Gammaproteobacteria;Thiotrichales;Thiotrichaceae;Thiothrix
Bacteria;Proteobacteria;NA;NA;NA;NA
Bacteria;Verrucomicrobiota;Lentisphaeria;NA;NA;NA
Bacteria;Verrucomicrobiota;Lentisphaeria;P.palmC41;NA;NA
Bacteria;Verrucomicrobiota;Verrucomicrobiae;Opitutales;Puniceicoccaceae;Lentimonas
Bacteria;Verrucomicrobiota;Verrucomicrobiae;Verrucomicrobiales;Rubritaleaceae;Luteolibacter

Acknowledgments

Firstly, I would like to express my sincere thanks to Prof. Dr. Rudolf Amann, Prof. Dr. Thorsten Brinkhoff, and Dr. Greta Reintjes for agreeing to review my thesis and for being members of my examination board. A special acknowledgment goes to Prof. Dr. Michael W. Friedrich for chairing my committee. I would like to thank Mahum Farhan and Gaétane Sallard for agreeing to be part of my examination board.

My heartful appreciation is extended to Rudi for offering me the opportunity to do doctoral research here and for the abundant support and valuable comments provided. I am particularly grateful to my supervisor, Dr. Hanno Teeling, who accepted my application, brought me here, has consistently guided me, and has been readily available to address my inquiries, whether related to research topics, writing, or life. I am equally grateful for his concern about my life and mental health, which has been extremely helpful for me to face PhD life positively. I express my acknowledgment to Dr. Chandni Sidhu, not only for her meaningful guidance in the learning of bioinformatics but also for her support and collaboration with my projects and for being a good friend in my personal life.

I am incredibly thankful to everyone in the Molecular Ecology Department for their support and for providing a very nice working environment. Special thanks to PD Dr. Bernhard Fuchs, Dr. Chandni Sidhu, Dr. Luis Humberto Orellana Retamal, Prof. Dr. Jens Harder, and Prof. Dr. Jan-Hendrik Hehemann, who were members of my thesis advisory committee meetings, providing valuable advice and comments for my projects.

I deeply appreciate the collaboration and communication with the researchers from the University of Greifswald and AWI (Helgoland). I am also grateful to Prof. Dr. Zong-Jun Du and Dr. De-Chen Lu from Shandong University in China. Particularly, I would like to thank De-Chen, who offered essential support and invaluable discussion during my scientific research and personal life. I would like to thank my former colleagues, Dr. Ben Francis, Dr. Karen Krüger, and Dr. Taylor Priest, for their invaluable help and communication on my study of bioinformatics and projects. Further, I would also like to thank Dr. Taylor Priest for thesis proof-reading. I sincerely appreciate the members

behind the IMPRS MarMic program. Special thanks go to the service groups, with particular appreciation for the IT Department.

My heartfelt thanks go to my family, who always support and love me, allowing me the freedom to pursue my dreams. I am also grateful to my friends, some in China and others in Germany, for their love and support. Whenever I felt down, someone always encouraged me and cheered me up. Last but not least, my thanks extend to Zheng-Yu Chen (MARUM of Uni Bremen) for guidance in Python and help with plotting. I'm very happy to have you, who brings sunshine and love into my life.

Universität Bremen

Ort, Datum: _____

Versicherung an Eides Statt

Ich, Fengqing Wang

.

versichere an Eides Statt durch meine Unterschrift, dass ich die vorstehende Arbeit selbständig und ohne fremde Hilfe angefertigt und alle Stellen, die ich wörtlich dem Sinne nach aus Veröffentlichungen entnommen habe, als solche kenntlich gemacht habe, mich auch keiner anderen als der angegebenen Literatur oder sonstiger Hilfsmittel bedient habe.

Ich versichere an Eides Statt, dass ich die vorgenannten Angaben nach bestem Wissen und Gewissen gemacht habe und dass die Angaben der Wahrheit entsprechen und ich nichts verschwiegen habe.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

_____

Ort, Datum / Unterschrift