Marine *Bacteroidetes*: distribution patterns and role in the degradation of organic matter

Dissertation
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
- Dr. rer. nat. -

dem Fachbereich Biologie/Chemie der
Universität Bremen
vorgelegt von

Paola Rocio Gómez Pereira

Bremen, Februar 2010

Die vorliegende Arbeit wurde in der Zeit von April 2007 bis Februar 2010 am Max–Planck–Institut für marine Mikrobiologie in Bremen angefertigt.

1. Gutachter: Prof. Dr. Rudolf Amann
2. Gutachter: Prof. Dr. Victor Smetacek

1. Prüfer: Dr. Bernhard Fuchs
2. Prüfer: Prof. Dr. Ulrich Fischer

Tag des Promotionskolloquiums: 9 April 2010

Para mis padres

Abstract

Oceans occupy two thirds of the Earth's surface, have a key role in biogeochemical cycles, and hold a vast biodiversity. Microorganisms in the world oceans are extremely abundant, their abundance is estimated to be 10^{29}. They have a central role in the recycling of organic matter, therefore they influence the air–sea exchange of carbon dioxide, carbon flux through the food web, and carbon sedimentation by sinking of dead material. *Bacteroidetes* is one of the most abundant bacterial phyla in marine systems and its members are hypothesized to play a pivotal role in the recycling of organic matter. However, most of the evidence about their role is derived from cultivated species. *Bacteroidetes* is a highly diverse phylum and cultured strains represent the minority of the marine bacteroidetal community, hence, our knowledge about their ecological role is largely incomplete.

In this thesis *Bacteroidetes* in open ocean and in coastal seas were investigated by a suite of molecular methods. The diversity and clade–specific abundance of *Bacteroidetes* were analyzed in different oceanic provinces in the North Atlantic Ocean. Comparative sequence analysis of 16S ribosomal RNA (rRNA) gene libraries revealed a high diversity and significant spatial variability. Major bacteroidetal clades were delimited based on 16S rRNA gene sequence identities and further quantified by fluorescence *in situ* hybridization (FISH). Preliminary FISH analysis indicated that certain *Bacteroidetes* clades were present in numbers lower than the detection limit of the method. Therefore, the FISH quantification protocol was modified by increasing the sample volume, which allowed us to reliably quantify populations down to absolute numbers of only 500 cells ml^{-1}.

All bacteroidetal clades evaluated showed pronounced regional distribution patterns. Interestingly, our analysis revealed that the bacteroidetal community composition largely reflects the satellite space–based demarcation of ocean provinces. Large insert fosmid–based metagenomic libraries constructed from two distinct

oceanic provinces in the North Atlantic Ocean and an extensive genome comparison gave first insights into an adaptation of the bacteroidetal community to distinct environmental conditions. Genomic analysis highlighted that marine *Bacteroidetes* have a marked metabolic potential for the degradation of proteins and bacterial cell wall components, particularly in oligotrophic water masses. Our findings also provide evidence of a strong specialization for the degradation of particular polymers. Glycolytic potential was confined to a member of the genus *Polaribacter* which had several glycoside hydrolases, and a higher number of sulfatases than other members of the same genus. This *Polaribacter* flavobacterium might represent an ecotype within the genus with a particular adaptation for the degradation of sulfated polysaccharides, known to be major constituents of phytoplankton cell walls. Supporting the hypothesis of *Bacteroidetes* as particle inhabitants, a specific *in situ* enrichment of certain bacteroidetal clades in the phycosphere of phytoplankton cells was proved in the phytoplankton-rich oceanic province. We could establish a link between *Bacteroidetes* metabolic information, organism identity, and abundance, thus providing hints with respect to the significance of certain metabolic traits. Furthermore, the *Bacteroidetes* clades delimited in the open ocean were also identified in the coastal systems analyzed in this thesis, giving evidence about the ecological relevance of those bacteroidetal clades in different marine settings. In summary, the combination of genomics and diversity studies allowed us to link main aspects about the *Bacteroidetes* community, enabling the better understanding of their ecological role in marine systems.

Zusammenfassung

Ozeane bedecken zwei Drittel der Erdoberfläche, besitzen zentrale Funktionen in den biogeochemischen Stoffkreisläufen und beinhalten eine große Biodiversität. Mikroorganismen sind in den Ozeanen sehr abundant, ihre Zahl wird auf 10^{29} geschätzt. Sie spielen eine zentrale Rolle im Recycling von organischem Material, dabei beeinflussen sie den Austausch des Kohlendioxids zwischen Luft und Meer, den Kohlenstofffluss in der biologischen Nahrungskette und die Ablagerung von Kohlenstoff durch Absinken von totem Material. *Bacteroidetes* sind eines der am häufigsten vorkommenden Bakterien–Phyla in marinen Systemen und es wird angenommen, dass sie Schlüsselfunktionen im Recycling von organischem Material besitzen. Allerdings sind die meisten dieser Funktionen von kultivierten Arten abgeleitet. *Bacteroidetes* sind phylogenetisch ein sehr diverses Phylum, jedoch gibt es nur wenige kultivierte Stämme von *Bacteroidetes* aus marinen Habitaten. Dementsprechend ist unser Wissen über ihre ökologischen Rolle größtenteils unvollständig.

In dieser Arbeit wurden *Bacteroidetes* aus dem offenen Ozean und aus küstennahen Meeren mittels mehrerer molekularbiologischer Methoden untersucht. Es wurde die Diversität und Gruppen–spezifische Häufigkeit von *Bacteroidetes* in verschiedenen ozeanischen Provinzen des Nordatlantischen Ozeans analysiert. Vergleichende Sequenzanalysen von 16S ribosomalen RNA (rRNA) Genbanken zeigte eine große Vielfalt und eine signifikante räumliche Variabilität. Die vorherrschenden *Bacteroidetes*–Gruppen konnten anhand der 16S rRNA Gensequenzierung definiert und durch Fluoreszenz *in situ* Hybridisierung (FISH) quantifiziert werden. Vorläufige FISH Analysen zeigten, dass bestimmte Entwicklungslinien der *Bacteroidetes* in ihrer Häufigkeit unterhalb der Nachweisgrenze der Methode lagen. Demzufolge wurde das Protokoll zur Quantifizierung von FISH–positiven Zellen modifiziert. Eine Erhöhung des filtrierten Probenvolumens ermöglichte es mir, auch Populationen von nur 500 Zellen ml^{-1} zuverlässig zu quantifizieren.

Alle untersuchten *Bacteroidetes*–Gruppen zeigten ausgeprägte regionale Verteilungsmuster. Interessanterweise zeigte unsere Analyse, dass die Anordnung der Gemeinschaft der *Bacteroidetes* größtenteils mit der Satelliten–gestützten Abgrenzung der ozeanischen Provinzen übereinstimmt. Aus zwei verschiedenen ozeanischen Provinzen des Nordatlantiks wurden große Fosmid–basierte Metagenombanken konstruiert und ein umfassender Genomvergleich erlaubte erste Einblicke in die Anpassung der *Bacteroidetes* an verschiedene Umweltbedingungen. Die Genomanalysen zeigten, dass marine *Bacteroidetes* ein deutliches metabolisches Potenzial für den Abbau von Proteinen und Zellwandbestandteilen von Bakterien besitzen, insbesondere in oligotrophen Wassermassen. Meine Ergebnisse erbrachten ebenso den Nachweis einer starken Spezialisierung im Abbau von bestimmten Polymeren. Das glykolytische Potenzial beschränkte sich auf ein Mitglied der Gattung *Polaribacter*, das einige Glycosid–Hydrolasen besaß und eine größere Anzahl von Sulfatasen aufwies, als andere Mitglieder dieser Gattung. Dieses Flavobakterium der Gattung *Polaribacter* könnte innerhalb der Gattung einen Ökotyp darstellen, der sulfathaltige Polysaccharide abbaut. Diese Polysaccharide gehören zu den Hauptbestandteilen der Zellwände des Phytoplankton. Die Hypothese, dass *Bacteroidetes* Partikel besiedeln, konnte in der spezifischen Anreicherung von bestimmten Gruppen in der Phykosphäre von Phytoplanktonzellen aus dem Phytoplankton–reichen Wasser des Nordatlantiks *in situ* bestätigt werden. In dieser Arbeit konnte ich zwischen der metabolischen Information, der Identität der Organismen und der Häufigkeit der Bacteroidetes eine Verbindung im Hinblick auf die Signifikanz bestimmter metabolischer Merkmale etablieren. Zudem konnten die *Bacteroidetes*–Gruppen des offenen Ozeans auch in küstennahen Systemen identifiziert und ihre ökologische Relevanz in verschiedenen marinen Habitaten bestimmt werden. Die Kombination von Genomanalysen und Diversitätsstudien erlaubte es mir, wichtige Aspekte der *Bacteroidetes*-Gemeinschaft zu verknüpfen und ihre ökologische Rolle in marinen Systemen besser verstehen zu lernen.

Contents

Abbreviations

ARCT Arctic province

bp base pairs

BPLR Boreal Polar province

CARD catalyzed reporter deposition fluorescence

Chl a chlorophyll a

DAPI 4'6–diamidino–2–phenylindole

DNA deoxyribonucleic acid

DOM dissolved organic matter

EGC East Greenland Current

FA formamide

FISH fluorescence *in situ* hybridization

G+C guanine + cytosine content

GH glycoside hydrolases

HMW high molecular weight

NADR North Atlantic Drift province

NAST North Atlantic Subtropical province

ORF open reading frame

PCR polymerase chain reaction

POM particular organic matter

PSU practical salinity units

RNA ribonucleic acid

rRNA ribosomal RNA

Sus starch utilization system

UTC universal time

Part I

Combined Presentation of Results

Chapter 1

Introduction

1.1 Marine environments

Oceans in the marine carbon cycle

Oceans occupy the vast majority of the Earth's surface, approximately 70%, and they have a profound effect on biogeochemical cycles and climate. They hold a large and still unknown biodiversity, with organisms ranging in size from a few microns (0.2-2 μm, picoplankton) to several meters (e.g. 30 m, the blue whale) that interact in complex food webs. A considerable fraction of the oceanic food web is driven by the energy that flows through dissolved organic matter (DOM) (Nagata, 2000), which represents one of the largest active reservoirs of organic carbon on Earth (Hedges, 1992). Photosynthetic organisms, mainly eukaryotic phytoplankton and cyanobacteria, convert inorganic carbon (CO_2) to organic carbon, as biomass (Figure 1.1). The phytoplankton cells release a major fraction of their daily photosynthesis production as DOM. The released substances include, among others, carbohydrates (mono-, oligo- and polysaccharides), and nitrogenous compounds (amino acids, peptides and proteins) (Hellebust, 1965; Fogg, 1983; Myklestad, 1995). Polysaccharides are one of the main constituents of the DOM released by phytoplankton (Biddanda and Benner, 1997). Other processes also have a prominent role in the production of oceanic DOM, such as excretion and "sloppy feeding" by grazers; and cell lyses induced by viruses

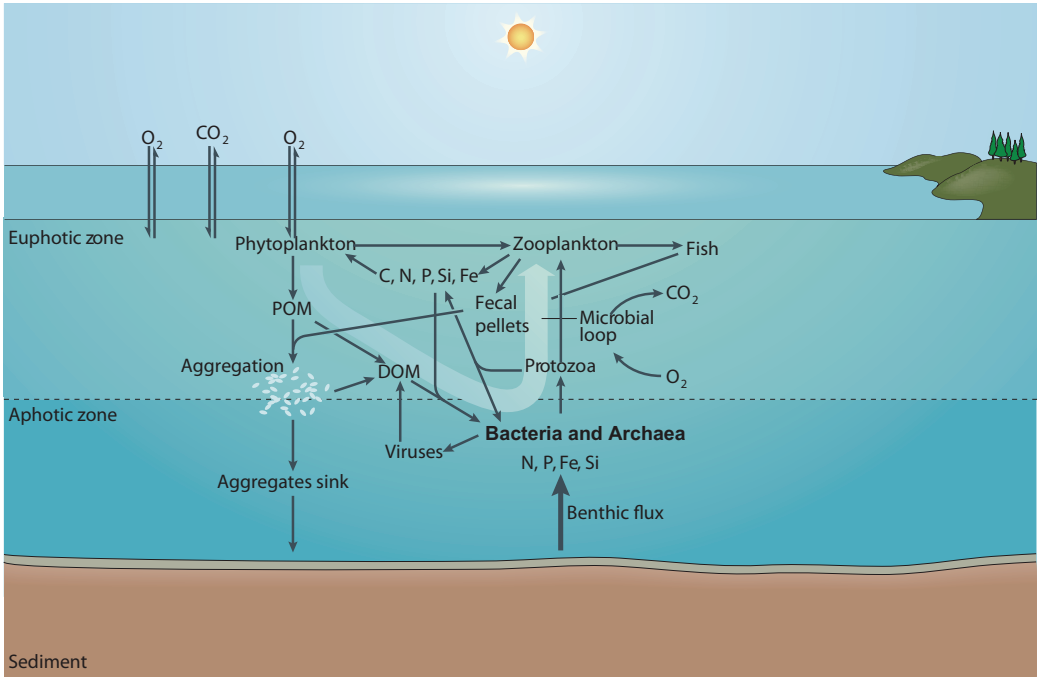


Figure 1.1: Microbial structuring of the marine ecosystem (modified from Azam and Malfatti, 2007). A large fraction of the organic matter that is synthesized by primary producers becomes dissolved organic matter (DOM) and is taken up almost exclusively by *Bacteria* and *Archaea*. Most of the DOM is respired to carbon dioxide and a fraction is assimilated and re–introduced into the classical food chain (phytoplankton to zooplankton to fish). The action of *Bacteria* and *Archaea* on organic matter plays a major part in carbon cycling through DOM. It therefore influences the air–sea exchange of carbon dioxide, carbon storage through sinking and carbon flux to fisheries.

(Nagata, 2000; Suttle, 2005). These processes will result in the release of DOM not only derived from phytoplankton, but also from other organisms. Particularly, structural components of microbial cell walls, such as bacterial and archaeal membrane components (Tanoue et al., 1995) and peptidoglycan (McCarthy et al., 1998), can make up a significant fraction of the DOM released as a result of their death due to protozoan grazing and viral infection (Nagata and Kirchman, 1999).

The DOM released by the different processes supports the growth of heterotrophic microorganisms (Ducklow and Carlson, 1992) and most of it is used and respired by them (Robinson and Williams, 2005). Eventually, it will

cycle through the food web by the microbial loop (Pomeroy, 1974). Moreover, a fraction of the carbon fixed in the form of biomass in the euphotic zone is ultimately exported to the mesopelagic ocean by physical mixing, gravitational settling (e.g. marine snow, fecal pellets), or active transport by zooplankton. These processes draw down significant amounts of carbon and are commonly known as the "biological pump" (reviewed by Raven and Falkowski, 1999). Overall, it results in complex and interconnected cycles in which microorganisms play a pivotal role.

The North Atlantic Ocean

The Atlantic Ocean is the second largest of the Earth's Oceans. It is divided into the South Atlantic, south of the Equator, and the North Atlantic, north of the Equator. The North Atlantic Ocean has several features, which makes it unique among the world's ocean basins.

The northern North Atlantic is strongly influenced by several currents: the East Greenland Current, the Irminger Current, the North Atlantic Current, and the Azores Current (Figure 1.2). The East Greenland Current (EGC) flows southward along the eastern coast of Greenland, from 79°N to 60°N (Woodgate et al., 1999). The EGC transports recirculating Atlantic and Arctic Ocean water masses, and >90% of the ice exported from the Arctic Ocean (Rudels et al., 1999; Woodgate et al., 1999). Therefore the EGC is an important link between the Arctic Ocean and the North Atlantic Ocean. The North Atlantic Current (NAC) originates in the Gulf Stream and it is unique in transporting warm tropical waters towards much higher latitudes than any other western boundary current (Krauss, 1986). The flow of the Gulf Stream diverges northeasterly in the NAC and southesterly to the Subtropical Atlantic and the Azores current (Krauss, 1986).

One of the most remarkable characteristics of the North Atlantic Ocean is that it supports by far the most extensive and strongest spring bloom of all the oceans. The annual spring bloom is known as one of the most conspicuous seasonal events in the world ocean (e.g. Ducklow and Harris, 1993). It starts north of about 40°N latitude and subsequently develops northwards

(Longhurst, 1998). The concentration of chlorophyll a remains high during summer and it gradually returns to winter concentrations (Henson et al., 2006). During summer, there is a shift from dominance by the >20 μm chlorophyll fraction to dominance by the <20 μm chlorophyll fraction (Gifford et al., 1995). During winter, the area between Iceland and the Azores experiences very high levels of wind stress and of heat flux, especially in the zone of 50°N–60°N, which drives a deep winter mixing of the water masses, down to 750–900 m (Hellerma, 1967). A consequence of the deep mixing and of the spring bloom is the formation of a deep nutrient front where, at 1000 m across the ocean at about 30°N, the concentrations of nitrate, phosphate, and silicate reach very low levels that have no counterparts in other ocean basins (Longhurst, 1998).

The northern North Atlantic is also particularly important to the world ocean circulation because it is where the North Atlantic Deep Water (NADW) is formed. On the eastern side of the Fram Strain region in the northern Greenland Sea the relatively warm and saline water of Atlantic origin is transported northwards towards the pole (Paquette et al., 1985). As these waters move north the temperature decreases and they sink beneath the Arctic surface waters and ultimately contribute to the formation of the NADW, one of the dominant deep water masses of the worlds oceans. Therefore, high biological productivity in the northern North Atlantic and subsequent vertical flux could result in high downward fluxes of carbon (Smith et al., 1991), which will be trapped in the deep water currents for thousands of years.

Coastal ecosystems

Littoral ecosystems such as lagoons, estuaries, and salt marshes are among the most productive natural ecosystems on Earth (Abreu et al., 1994; Castel et al., 1996). Phytoplankton make up a larger fraction of the biomass in coastal seas than in ocean gyres, and species differ between coastal and ocean populations. Two examples of coastal environments are coastal lagoons and estuaries, which are transition areas between the sea and rivers. The physical parameters strongly vary along the gradients resulting in highly

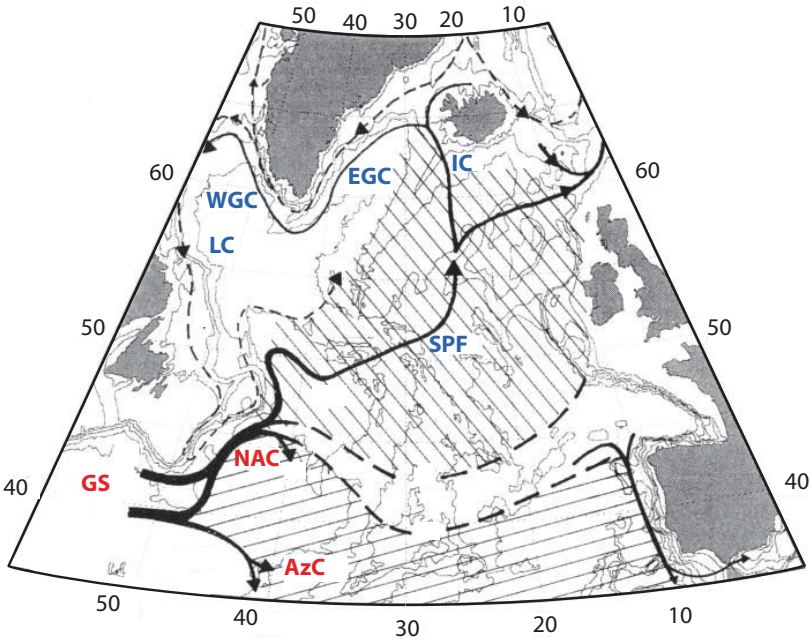


Figure 1.2: Major currents of the northern North Atlantic adapted from Krauss (1986) and Rossby (1996). The left slanted shading indicates the region where the warm waters of the Subpolar Front can be found. The shading in the south represents waters of the subtropical gyre. Abbreviations are IC, Irminger Current; EGC and WGC, East and West Greenland Currents; NAC, North Atlantic Current; LC, Labrador Current; SPF, Subpolar Front; GS, Gulf Stream; AzC, Azores Current

dynamic environments. There is a wide spectrum of estuary types in terms of the degree of connection with the sea. Salt wedge estuaries, for example, are formed where a large river flows into an area with a small tidal range, generating vast areas with intermediate salinities. When the river flow moves seawards, most fine material is carried out to the sea as a well–defined plume, and the estuary itself is dominated by coarse sediments (Little, 2000).

Coastal lagoons are shallow water bodies separated from the ocean by a barrier, connected at least intermittently to the ocean by one or more restricted inlets, and usually oriented shore–parallel (Kjerfve, 1994). Coastal lagoons are strongly influenced by tides and winds which cause large spatial and temporal physicochemical gradients, for example the salinity gradient within a lagoon can vary from freshwater to hypersaline conditions (Kjerfve,

1994). Moreover, coastal lagoons are usually surrounded by extensive areas of marshes that are important sources of DOM (Oliveira and Kjerfve, 1993) and therefore, they have a fundamental role in the synthesis and transformation of organic matter (Kjerfve, 1994; Bonilla et al., 2006).

Because of the high physical variation in coastal lagoons and estuaries, they create a rigorous environment for the biological communities which adapt to these transition zones (Little, 2000). Many of these gradients - including salinity, organic matter, and nutrient concentration- are thought to influence microbial diversity and physiology (Barcina et al., 1997; Giovannoni and Rappe, 2000).

1.2 Marine microorganisms

Revealing the "unseen majority"

Microorganisms are not visible to the naked eye, and could only be studied after the development of the microscope by A. van Leewenhoeck (1677). He observed "little animals in bodies of water" (van Leewenhoeck, 1677). Since then, the advances in marine microbiology have been partly driven by the development of new methods (Azam, 2001). By now, it is clear that microorganisms are the "unseen majority", being estimated that there are a total of 35 x 10^{28} in the world's surface oceans (Whitman et al., 1998). Despite their small size, the total mass of bacteria in the oceans exceeds the combined mass of zooplankton and fish (Pomeroy et al., 2007). The advent of molecular techniques during the last three decades has revealed a great amount of information about main ecological aspects of microorganisms. Knowledge about their (i) diversity, (ii) abundance and, (iii) role in natural environments has been gained making possible the application of theory in microbial ecology (Prosser et al., 2007).

Microbial diversity

Carl Woese's pioneering comparative studies of rRNA sequences proved the universal relatedness of all life and proposed the three domains of life: *Archaea* (formerly archaebacteria), *Bacteria* (eubacteria), and *Eucarya* (eucaryotes) (Woese and Fox, 1977). Phylogenetic sequence comparisons opened the possibility to study yet uncultured microorganisms in their natural environment. In the 1980's Norman Pace and co–workers proposed the "ribosomal RNA approach" by which the 16S rRNA gene of a microorganism is directly obtained by PCR, cloning, and sequencing (Olsen et al., 1986) (Figure 1.3 A). The analysis of 16S rRNA sequences produces numerical identity values that can be used as circumscription limits for taxa (Rossello-Mora and Amann, 2001). A genus is roughly defined by species with a 95% 16S rRNA gene identity (Ludwig et al., 1998), while a 97% identity is usually considered as a boundary for species circumscription, based on the criteria

that strains of the same species have a DNA-DNA genomic similarity higher than 70% (Wayne et al., 1987).

Since the development of the ribosomal approach, an immense diversity has been revealed. Currently, there are 25 phyla with cultured representatives and more than 50 others defined only by the relatedness of 16S rRNA gene sequences (Pace, 2009). The advances in high throughput next generation sequencing technologies during the last decade, as 454–pyrosequencing (Margulies et al., 2005), has lead to the discovery of an immense and yet uncharacterised microbial diversity (Sogin et al., 2006). It has been proposed that the diversity of marine microorganisms might consist of two compartments, the "core taxa", comprising the more abundant clades which are usually retrieved by classical molecular approaches; and a long tail of low abundance rare taxa that have not previously been characterized (Pedros-Alio, 2006). The rare taxa has been called the "rare biosphere", and it is retrieved by high throughput sequencing approaches (Sogin et al., 2006; Huber et al., 2007; Rusch et al., 2007). However, the rare taxa might also be a fundamental constituent of the microbial communities. Despite of their low abundance, rare clades can be responsible for most of the uptake of key nutrients (Musat et al., 2008) and can carry out fundamental ecosystems processes (Kuypers et al., 2005). Due to methodological constraints, so far there has not been a quantitative determination of the "rare biosphere" microbial clades.

Microbial abundance

Why are some species rare and others common? Why does a species occur at low population densities in some places and at high densities in others? Which factors cause fluctuations in species abundance? These are crucial questions in ecology, which do not have simple answers for either macro- or microorganisms (Begon et al., 2006).

In the microbial world, the species definition is rather difficult (Rossello-Mora and Amann, 2001). Instead, natural microbial assemblages could be regarded as more or less loose collections of individual genotypic populations

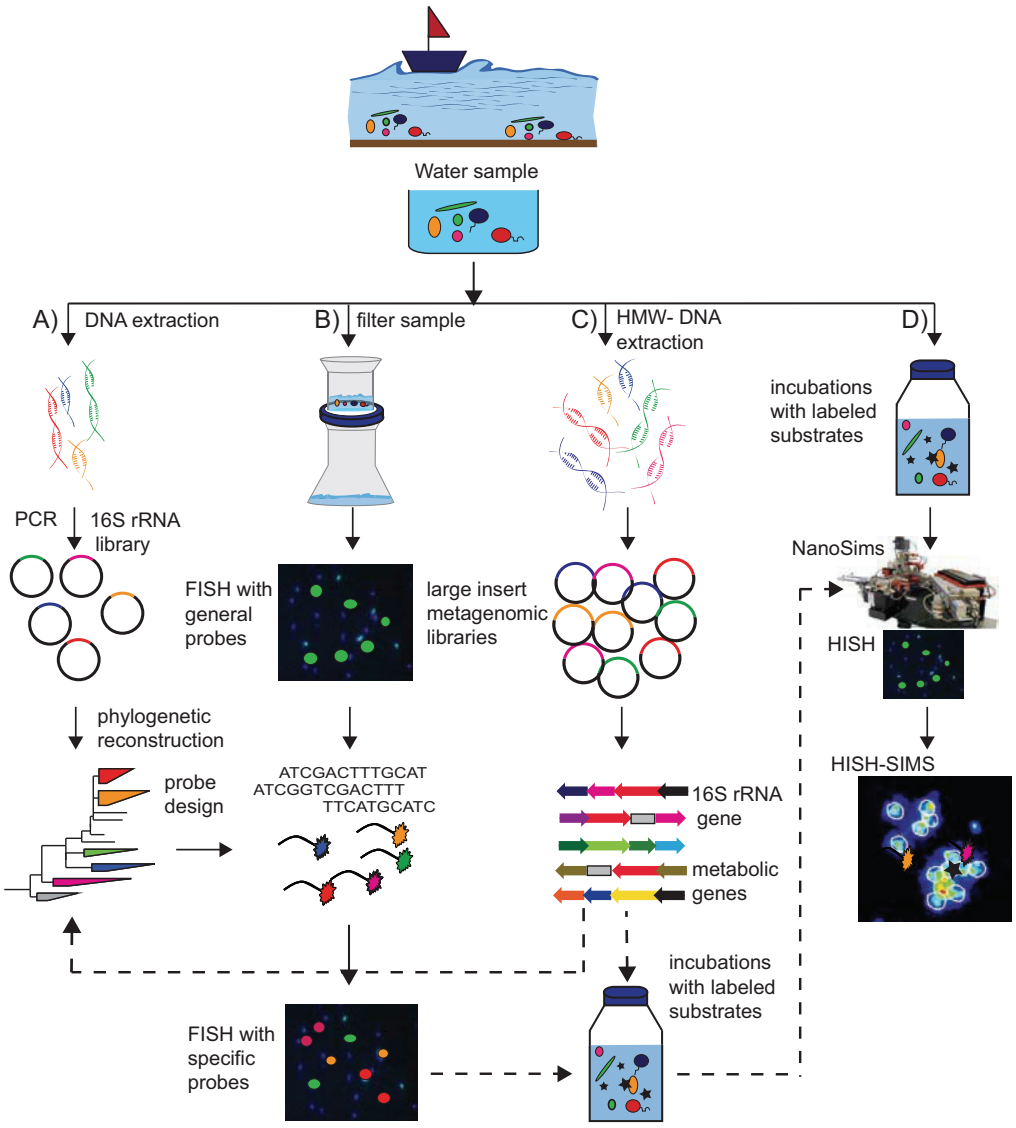


Figure 1.3: Methodological approaches in marine microbiology (A)16S rRNA full cycle, (B) FISH, (C) large insert metagenomic libraries and (D) incubation experiments with labeled substrates analyzed by halogen *in situ* hybridization (HISH) and nanoscale secondary ion mass spectroscopy (NanoSims). Dashed lines indicate connections between methodological approaches

that change over time in features such as growth, mortality, and size (Pernthaler and Amann, 2005). Within this assemblage, populations are often considered as a distinct entity that can be reliably identified and quantified (Pernthaler and Amann, 2005).

One of the major techniques for quantification of defined populations in their natural environment is fluorescence *in situ* hybridization (FISH) with fluorescently labeled rRNA-targeted oligonucleotide probes (Amann et al., 1990) (Figure 1.3 B). Comprehensive rRNA gene sequence databases are necessary for the design of oligonucleotide FISH probes that will exclusively bind to the ribosomes of microorganisms of a defined clade. Its quantification is then possible by epifluoresence microscopy (Amann et al., 1990). Marine microorganism are typically small, approximately 0.5 μm in diameter, and have only a few hundred ribosomes. Therefore, detection with fluorescently mono-labeled oligonucleotides is difficult (Amann and Fuchs, 2008). Catalyzed reporter deposition (CARD)–FISH was further developed to increase the sensitivity of the method. CARD–FISH combines the *in situ* hybridization with horseradish peroxidase labeled oligonucleotide probes and enzymatic signal amplification with fluorescently labeled tyramides, which confers a strong and stable fluorescent signal (Pernthaler et al., 2002). By the application of this method the percentage of microorganisms that could be identified in marine systems increased three times (Pernthaler et al., 2002), and it allows its reliable quantification in the oligotrophic open ocean (Schattenhofer et al., 2009).

Microorganisms' role

Microorganisms are both phylogenetically and metabolically diverse. In general, they participate in all biogeochemical cycles and also in linking trophic webs (Pomeroy, 1974; Azam et al., 1983). It is, therefore, fundamental to understand the role of the major microbial clades in their natural environment in order to answer questions about ecosystem function.

The ultimate approach for characterizing the bacterial physiology and metabolic requirements is to have them in culture collections. However,

only a small proportion of the microorganisms sampled from the environment can be readily cultivated using conventional techniques (Staley and Konopka, 1985; Giovannoni and Stingl, 2007). The development of "dilution to extinction" cultivation techniques has been successful in the isolation of many relevant marine microorganisms (Connon and Giovannoni, 2002; Stingl et al., 2007). The cultivation of resident microorganisms is valuable to describe microbial processes but it is not enough on its own (DeLong, 2009) and cultivation independent molecular surveys are needed to describe their ecological role.

The advent of metagenomics has allowed the direct link between identity and function of yet uncultivated microorganisms in their natural environment. Metagenomics is defined as the analysis of the collective microbial genomes contained in an environmental sample (Riesenfeld et al., 2004). Metagenomic analysis involves isolating high molecular weight (HMW) DNA from environmental samples, cloning the DNA into a suitable vector, transforming the clones into a host bacterium, and screening the resulting transformants (Handelsman, 2004) (Figure 1.3 C). Depending on the vector used, metagenomics can either consist of small insert size (approx. 3 kbp, "shotgun") or of large insert size libraries, where longer genomic fragments (30–100 kbp) are cloned into fosmids or bacterial artificial chromosomes (BACs) vectors. Alternately the isolated environmental DNA can be sequenced without previous cloning by next–generation sequencing technologies such as 454–pyrosequencing (Margulies et al., 2005). The 454–pyrosequencing technology has rapidly evolved from initial average read lengths of 100 bp to 450 bp (DeLong, 2009).

Short reads metagenomics, as 454–pyrosequencing and shotgun libraries, are high–throughput sequencing approaches which could generate billions of sequences reads in a short time. For example, a single study of surface ocean waters has generated 6.3 billion bp (Rusch et al., 2007). On the other hand, large insert size libraries have the advantage that they give access to DNA contigs that often carry a phylogenetic marker associated with genomic information (Beja et al., 2000, 2002). This approach would eventually lead to the identification of metabolic genes or even entire operons for a given

organism or clade (e.g. Martin-Cuadrado et al., 2008).

One of the most remarkable examples of metagenomics in the field of marine microbiology has been the discovery of a widespread and diverse retinal binding rhodopsin protein, called proteorhodopsin (Beja et al., 2000). A gene encoding a protein with characteristic features of rhodopsin–like pigments was originally identified in a BAC clone that carried a 16S rRNA gene of the marine gamma bacterial clade SAR86 (Beja et al., 2000). Proteorhodopsin acts as a light driven proton pump (Beja et al., 2000, 2001) and it can result in photophosphorylation of ADP to ATP (Martinez et al., 2007). Since then, there has been increasing information indicating that proteorhodopsins are widespread in marine environments and among phylogenetic groups (Fuhrman et al., 2008a; Frigaard et al., 2006). The discovery of proteorhodopsin has challenged the notion that solar energy can only enter marine ecosystem by chlorophyll-based photosynthesis (Fuhrman et al., 2008a). However, whether the presence of proteorhodopsin enhances the growth under light conditions is still unclear (Giovannoni et al., 2005; Gomez-Consarnau et al., 2007; Stingl et al., 2007) suggesting that it could have an array of functions in marine systems.

Ultimately, the predicted information gained by metagenomics can be used to design targeted incubation experiments to further evaluate the ecophysiology of defined microbial clades (Figure 1.3 D). Two of the techniques to measure substrate uptake of single cells are microautoradiography (MAR)-FISH and nano-scale secondary-ion mass spectrometry (nanoSIMS). MAR–FISH combines micro–autogradiography to track the uptake of radiolabeled tracers with the identification by FISH (Lee et al., 1999; Ouverney and Fuhrman, 1999). Two of the main drawbacks of MAR–FISH are the limited availability of environmentally relevant radioactively labeled substrates, and that the substrate uptake is not quantitative (Alonso, 2005). NanoSIMS has the advantage that it can quantitatively measure the uptake of any stable isotope as well as any radioisotope with a suitable half–life (Lechene et al., 2006). NanoSIMS can be further combined with *in situ* hybridization using halogen labeled probes (HISH–SIMS) and therefore individual cells can be identified (Musat et al., 2008) (Figure 1.3 D). Preliminary studies that com-

bine FISH with NanoSIMS have allowed the phylogenetic identification of microbial partners and the monitoring of the transfer of metabolites between them (Behrens et al., 2008).

Biogeography of microorganism in the oceans

The study of microbial biogeography goes back to the beginning of the last century with the famous statement of Baas-Becking: "Everything is everywhere, but the environment selects" (Baas-Becking, 1934) which has recently gained renewed interest (Ramette and Tiedje, 2007). It has been proposed that because of their small cell sizes, huge populations, and few geographical barriers, microorganisms might not experience geographic differentiations (Fenchel and Finlay, 2004; Finlay, 2002). However, this idea is challenged by an increasing number of studies (discussed in Pommier et al., 2007). In the marine realm, major picoplankton clades (planktonic microorganisms with a size between 0.2-2 μm) seem to have ubiquitous distributions; however, new clades emerge and clear distribution patterns can be envisioned from studies that focus on unique hydrographic features (Giovannoni and Stingl, 2005) rather indicating that the "environment selects" is a strong force shaping the communities (Pommier et al., 2007).

In surface oceans, less than 20 phylogenetic clades seem to constitute a significant fraction of picoplankton (Giovannoni and Stingl, 2005) and they represent the most abundant clades (Schattenhofer et al., 2009). Among those are members of the *Alphaproteobacteria* (Morris et al., 2002), *Gammaproteobacteria* (Mullins et al., 1995; Rappe et al., 1997) and *Bacteroidetes* phyla (Kirchman 2002), photosynthetic cyanobacteria such as the genera *Prochlorococcus* (Chisholm et al., 1988) and *Synechococcus* (Waterbury et al., 1979), and the archaeal phylum *Crenarchaeota* (Karner et al., 2001). When analyzed at a higher phylogenetic threshold, distinct picoplankton populations show pronounced seasonal (Fuhrman et al., 2006), latitudinal (Fuhrman et al., 2008b) and depth (DeLong et al., 2006) distribution patterns in marine systems. For example, the alphaproteobacterial SAR11 clade is the most abundant picoplankton clade in the world's surface oceans (Morris et al.,

2002; Carlson et al., 2009). However, depth and seasonal patterns of closely related ecotypes exist within this ubiquitous clade (Field et al., 1997; Carlson et al., 2009), which might reflect physiological adaptation to different environmental conditions.

Bacteroidetes is one of the numerically dominant picoplankton in both coastal and open ocean systems; however, no major clades have been consistently investigated within this phylum. Picoplankton of the *Bacteroidetes* phylum is the main focus of this thesis. The current state of knowledge on its phylogeny, distribution patterns, and possible role will be discussed in the following section.

1.3 Phylum *Bacteroidetes*

Phylogeny

The taxonomy of the phylum *Bacteroidetes* has passed through several reclassifications over the past years. Members of what nowadays comprise the phylum *Bacteroidetes* are referred in the literature as *Cytophaga/Flavobacteria*–like bacteria, *Cytophaga–Flavobacteria–Bacteroides*, *Cytophaga–Flexibacter–Bacteroides* and permutations of these names. Members of this phylum are heterotrophic–gram negative bacteria, with various morphologies from small rods to long filaments. It contains a mixture of physiological types, from obligate aerobes to obligate anaerobes. The phylum currently comprises the four classes *Bacteroidia*, *Sphingobacteria*, *Cytophagia* and *Flavobacteria* (Ludwig et al., 2008) (Figure 1.4).

Class *Bacteroidia*

Members of the class *Bacteroidia* are mostly strict anaerobes; they are gram–negative, non–spore forming, non–motile bacteria. *Bacteroidia* are usually associated with mammals or other animals, as disease agents (e.g. *Porphyromonas gingivalis*) or as part of the gut flora (e.g. *Bacteroides thetaiotaomicron*). Particularly, members of the genus *Bacteroides*, are the numerically dominant bacteria in the human large intestine (Salyers, 1984). As members of the intestinal flora, they have a major role in the breakdown of dietary polysaccharides (Smith et al., 2006).

Classes *Sphingobacteria* and *Cytophagia*

The classes *Sphingobacteria* and *Cytophagia* have recently been separated based on 16S rRNA sequence similarities (Ludwig et al., 2008). Both classes include mostly aerobic heterotrophic motile bacteria that are found in a variety of habitats. For example, the genus *Saprospira* of the *Sphingobacteria* class appears to exclusively live in aquatic environments (Reichenbach, 2006); species of this genus have been isolated from both marine (Lewin, 1970) and freshwater environments (Lewin, 1965). Interestingly, the marine *S. grandis*,

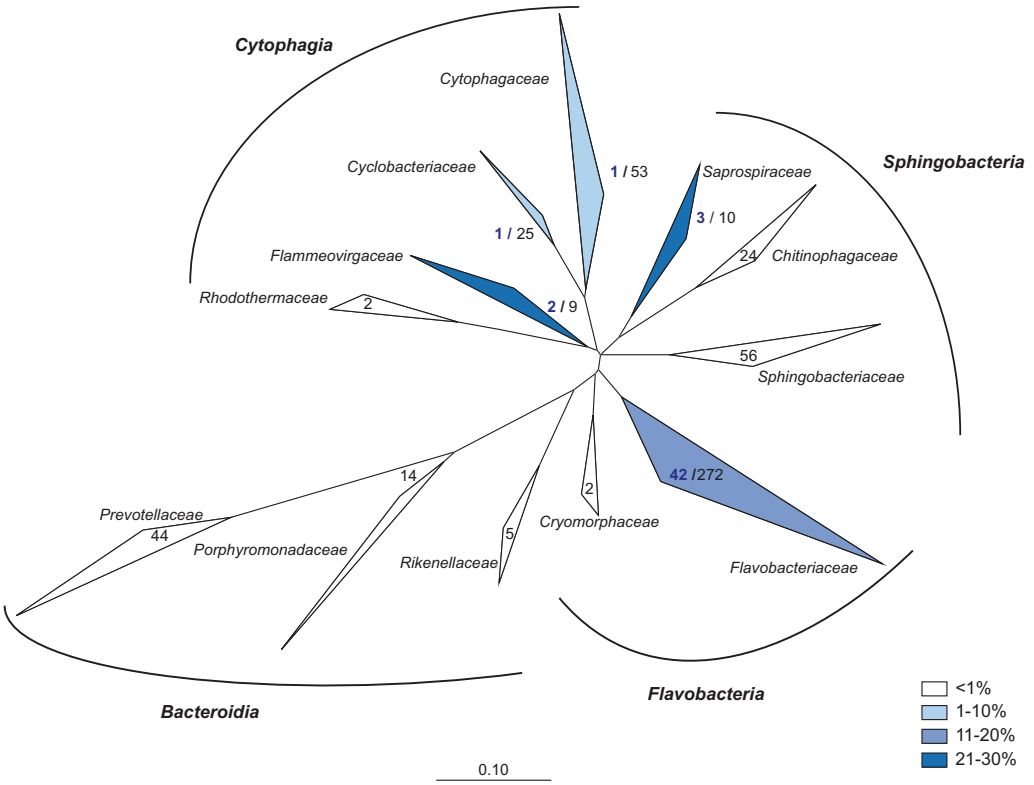


Figure 1.4: 16S rRNA gene based phylogenetic tree of the *Bacteroidetes* phylum type strains depicting classes (bold characters) and families. Color of the clades shown in wedges indicates the fraction of strains isolated from marine environments (bold blue characters). Total number of type strains for each family is also shown. Phylogenetic tree was adapted from the "All-Species Living Tree", release 100 (September 2009); the tree reconstruction was performed by using the maximum likelihood algorithm RAxML (Yarza et al., 2008).

can subsist on other microbes; they can catch motile bacteria by their flagella before killing and digesting them (Lewin, 1997). The genus *Chitinophaga* encompasses long filamentous gliding bacteria that have strong chitinolytic activities in soils (Sangkhobol and Skerman, 1981).

Members of the class *Cytophaga* may be aerobic, microaerophilic, or facultative anaerobic. Cell shape within this class vary from thin long filaments to almost spherical rod cells. Many of them are able to degrade macromolecules as proteins, chitin, pectin, agar, starch, or cellulose (Reichenbach, 2006). They occupy a wide range of environments; for example, the genus *Cytophaga*

mainly inhabits soils and decaying plant material (Nakagawa and Yamasato, 1996), while *Microscilla* spp. (class *Cytophagaceae*, Naganuma et al., 1993) or *Cyclobacterium* spp. (class *Flammeovirgaceae*, Nedashkovskaya et al., 2005) have been isolated from marine environments. Members of the genus *Flexibacter* are freshwater and marine fish pathogens (Soltani et al., 1996).

Class *Flavobacteria*

Members of the class *Flavobacteria* have widely diverse ecological niches and physiological characteristics, and even within some genera individual species might show contrasting characteristics (Bernardet and Nakagawa, 2006). The class includes the families *Flavobacteriaceae*, *Blattabacteriaceae*, and *Cryomorphaceae*, the latter exclusively including strains isolated from polar environments (Bowman et al., 2003). Members of the family *Flavobacteriaceae* have been isolated from a wide range of habitats, even within genera, and this is considered as one of the most intriguing characteristics of the family (Bernardet and Nakagawa, 2006). For example, strains have been isolated from sea ice (e.g. *Aequorivita antarctica*, Bowman and Nichols, 2002), human dental plaque (e.g. *Capnocytophaga granulosa*, Yamamoto et al., 1994), and soil (e.g. *Flavobacterium johnsoniae*, Bernardet et al., 1996).

Most of the marine *Bacteroidetes* type strains are affiliated with the class *Flavobacteria* and within this class a distinct "marine clade" has been delimited based on 16S rRNA gene phylogeny. The "marine clade" comprises genera solely isolated from oceanic environments (Bowman, 2006), it includes species that are highly cold adapted, as *Polaribacter* spp. (Gosink et al., 1998), but also mesophilic species, as *Arenibacter* spp. (Ivanova et al., 2001; Bowman, 2006).

In his famous monograph "Marine Microbiology" Claude ZoBell already reported that "Flavobacterium" were one of the main bacterial clades in the sea (ZoBell, 1946). The advent of the ribosomal rRNA approach (Figure 1.3 A) and the advances in sequencing technologies have lead to an exponential increase of sequences available in public databases from yet uncultivated microorganisms. In the database Silva (release 100, Pruesse et al., 2007) there

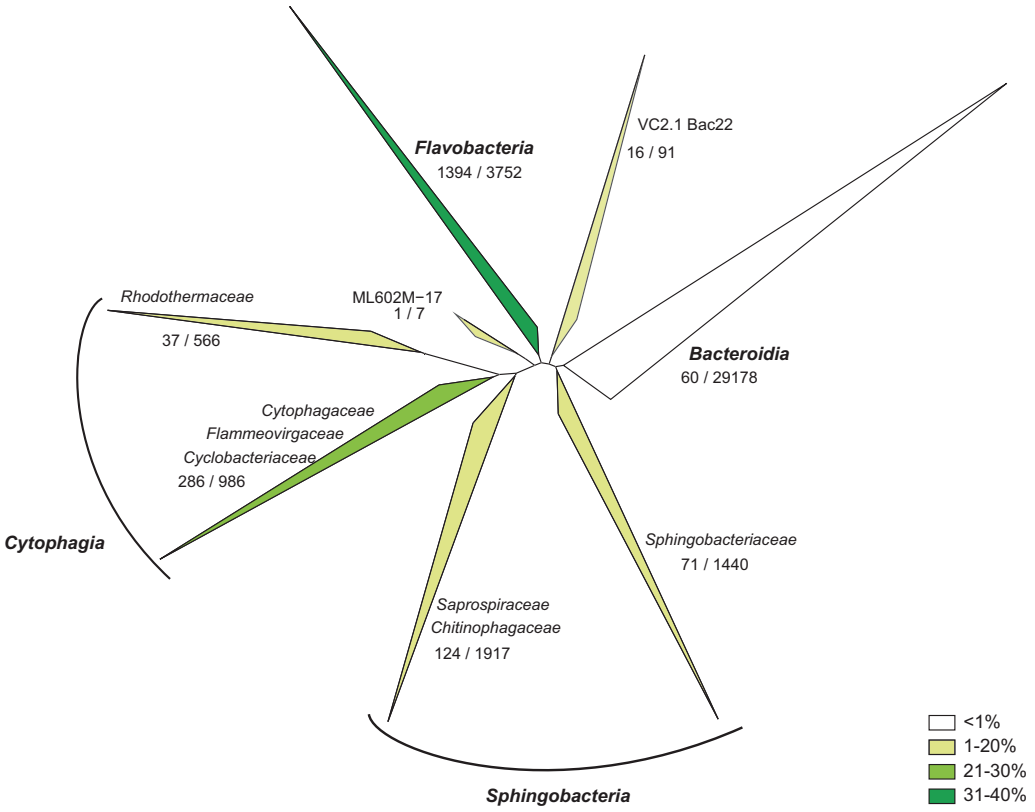


Figure 1.5: Schematic phylogenetic tree of the 16S rRNA gene sequences of *Bacteroidetes* in the database Silva 100 (Pruesse et al., 2007). Color of the clades shown in wedges indicates the fraction of sequences that have been retrieved from marine environments. Total number and total marine 16S rRNA gene sequences are given. Total sequences represents the number of sequences in each clade that in the database contained information about its isolation source.

are a total of 41,700 *Bacteroidetes* 16S rRNA gene sequences. In agreement with the evidence given by cultured strains, more than half of the *Bacteroidetes* sequences that have been retrieved from marine environments are affiliated with the class *Flavobacteria*, while a lower fraction is with *Cytophagia* and *Sphingobacteria* (Figure 1.5). Overall, both culture–based and culture independent approaches highlight the importance of *Flavobacteria* in marine environments.

Natural abundances and distribution patterns of marine *Bacteroidetes*

Culture independent approaches, such as 16S rRNA gene clone libraries, denaturing gradient gel electrophoresis (DGGE), and fluorescence *in situ* hybridization, have revealed that members of the *Bacteroidetes* phylum constitute one of the most abundant picoplankton groups (e.g. Glockner et al., 1999; Kirchman, 2002). *Bacteroidetes* thrive in a variety of marine systems including hydrothermal vents (Sievert et al., 2000; Kormas et al., 2006), polar sediments (Ravenschlag et al., 2001), coastal sediments (Llobet-Brossa et al., 1998) and waters (Eilers et al., 2001; O'Sullivan et al., 2004), and the open ocean (Simon et al., 1999; Abell and Bowman, 2005b; Schattenhofer et al., 2009). There is an increasing body of evidence indicating that in open oceans and in coastal waters most of the marine *Bacteroidetes* can be phylogenetically affiliated with the class *Flavobacteria* (Kirchman et al., 2003; Alonso et al., 2007; Chen et al., 2008) (Figure 1.5). High abundance of *Bacteroidetes* has been linked to cold waters (Simon et al., 1999; Abell and Bowman, 2005b), phytoplankton blooms (Simon et al., 1999), to the photic zone (Schattenhofer et al., 2009) and to upwelling systems (Alonso-Saez et al., 2007). This distribution suggests a preference for more productive conditions. However, *Bacteroidetes* have also been detected in oligotrophic marine surface waters in significant numbers (Schattenhofer et al., 2009).

Cell abundances and distribution patterns of marine *Bacteroidetes* in picoplankton have mostly been obtained by FISH with the probes CF319a (Manz et al., 1996) and CFB560 (O'Sullivan et al., 2002). The specificity of these two probes has recently been reviewed (Amann and Fuchs, 2008). The probe CF319a mostly covers *Flavobacteria* (90%) and *Sphingobacteria–Cytophagia* (90%), while only 30% of *Bacteroidia*. Probe CFB560 has a higher coverage of the phylum, 93% , evenly distributed among the classes. Additionally, CF319a does match several sequences outside the *Bacteroidetes* phylum while this is not the case for probe CFB560 (Amann and Fuchs, 2008). However, hybridizations with probe CFB560 seem not to be reliable. Counts with this probe are either not comparable with CF319a or positive

signals can not be reliably quantified (P. Gómez-Pereira, F. Simonato, C. Alonso; unpublished observations).

Overall, the probe CF319a is rather general, targeting most of the marine members of the phylum and so far evidence of distribution patterns of distinct *Bacteroidetes* clades is sparse. Examples include FISH-based quantification of the clade *Cytophaga marinoflava-latercula*, of the uncultured clades DE2 and AGG58, and of the genus *Polaribacter*. *Cytophaga marinoflava-latercula* was shown to be up to 6% of total picoplankton during spring and summer off the island of Helgoland, located in the German Bight of the North Sea (Eilers et al., 2001). Clade DE2, originally retrieved from the Delaware Estuary, was found to account for up to 10% of picoplankton in this estuary as well as in the Chukchi Sea, Arctic Ocean (Kirchman et al., 2003; Malmstrom et al., 2007). Members of the genus *Polaribacter* have been shown to form prominent fractions in polar oceans (Malmstrom et al., 2007), and up to 30% in Antarctic sea ice (Brinkmeyer et al., 2003). The uncultured clade "AGG58 branch 2" (O'Sullivan et al., 2004) was 3 ± 4% of the picoplankton in the Arctic Ocean (Malmstrom et al., 2007). However, based on an *in silico* evaluation, this probe is not specific for the AGG58 clade and it possibly hybridizes to several other marine *Bacteroidetes* (Appendix A).

The picoplankton composition of lakes and oceans clearly differs (Glockner et al., 1999), and within *Bacteroidetes* there is also evidence of distinct freshwater populations. For example, the LD2 is a cosmopolitan freshwater clade (Zwart et al., 2002). Members of this clade can form filaments under elevated grazing conditions, and these LD2 filaments comprised more than 40% of the total picoplankton biomass in a mesotrophic lake (Pernthaler et al., 2004). Rare freshwater flavobacterial populations might be important players in lakes. The cultivable flavobacterial FLAV2 clade represented a small but highly active population in an alpine lake. Its growth rates doubled during a period when chlorophyll a was maximum (Zeder et al., 2009).

Overall, the data available so far indicate that distinct marine *Bacteroidetes* clades prevail under different scenarios; however there is no consistent information about the environmental conditions that selects for certain clades.

Ecological role of marine *Bacteroidetes*

There has been a long-standing notion that *Bacteroidetes* have a role in the degradation of polysaccharides and other constituents of the organic matter (reviewed in Reichenbach and Weeks, 1981). There are several lines of evidence indicating this potential role in marine systems; however, some are rather indirect.

Affiliation with phytoplankton

One indirect line of evidence about the role of *Bacteroidetes* in marine systems is derived from the observation that their abundance increases towards more productive areas (Alonso-Saez et al., 2007) and they are usually one of the dominant clades during phytoplankton blooms (e.g. Simon et al., 1999; O'Sullivan et al., 2004; Pinhassi et al., 2004). For example, *Bacteroidetes* comprised 70% of the total picoplankton community during a *Phaeocystis* spp. bloom in the Southern Ocean (Figure 1.6 A) (Simon et al., 1999). Specifically, the vast majority of phylotypes retrieved from an algae bloom in the English Channel were affiliated with a single bacteroidetal clade, the AGG58 (O'Sullivan et al., 2004). Interestingly, Pinhassi and co–workers showed that distinct *Bacteroidetes* phylotypes were either present during diatoms or phytoflagellates blooms (Pinhassi et al., 2004). Yet, other bacteroidetal phylotypes dominate in scenarios when blooms do not occur (Riemann et al., 2000; Pinhassi et al., 2004; West et al., 2008). For example, in a natural iron fertilization induced bloom, the clade AGG58 dominated during the phytoplankton bloom while members of the genus *Polaribacter* did in areas outside the bloom (West et al., 2008). These studies indicate that only certain marine *Bacteroidetes* clades might be responding to the phytoplankton blooms.

The link between *Bacteroidetes* and phytoplankton might be happening at the nano– millimetre scale. Bacteroidetal strains (Barbeyron et al., 2001) and 16S rRNA gene sequences (Longford et al., 2007; Sapp et al., 2007) have been retrieved from the surface of algae cells. Moreover, *Bacteroidetes* cells have been shown to be localized in the phycosphere of *Alexandrium* spp. cells (Biegala et al., 2002). *Bacteroidetes* have also been reported as inhabitants of

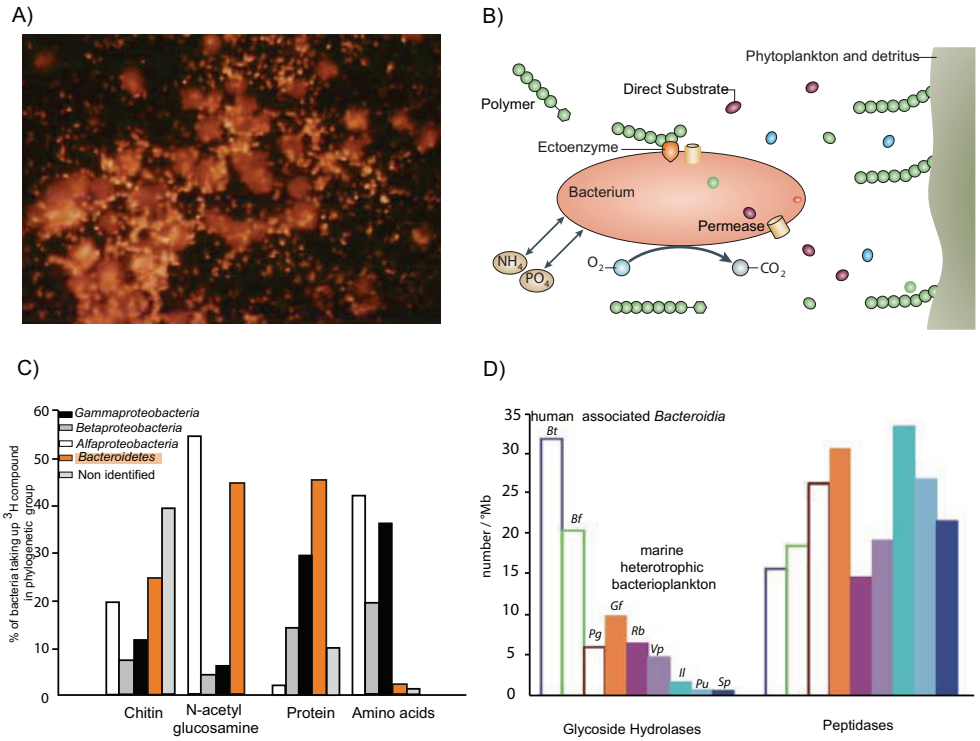


Figure 1.6: Role of *Bacteroidetes* in the degradation of DOM (A) Micrograph depicts *Bacteroidetes* cells, hybridized by FISH with Cy3–labeled oligonucleotide probe CF319a, and *Phaeocystis* cells during a bloom in the Southern Ocean (marginal ice zone; 68°50.57' S, 06°01.09' N) (Simon et al., 1999). (B) Scheme of the possible nanometre–millimetre scale interactions between *Bacteroidetes* and phytoplankton or detritus (modified from Azam and Malfatti, 2007). The microbial strategies that are depicted here include environmental sensing, permeases and cell–surface hydrolases; depicted also the ability to take up, as well as release, of NH_4^+ and PO_4^{3-}. (C) Relative abundance (%) of the major picoplankton groups consuming tritiated chitin, N–acetyl glucosamine, protein, and amino acids in the Roosevelt Inlet (Delaware Bay), assayed by MAR-FISH. *Bacteroidetes* cells were hybridized with the Cy3 labeled oligonucleotide probe CF319a (modified from Cottrell and Kirchman, 2000). (D) Comparison of hydrolytic capabilities between *Gramella forsetii*, other marine heterotrophic picoplankton, and human–associated *Bacteroidia*. The graph depicts the number per megabase of glycoside hydrolases and peptidases encoded in the genomes of Bt, *Bacteroides thetaiotaomicron*; Bf, *Bacteroides fragilis* YCH46; Pg, *Porphyromonas gingivalis*; Gf, '*Gramella forsetii*'; Rb, *Rhodopirellula baltica*; Vp, *Vibrio parahaemolyticus*; Il, *Idiomarina loihiensis*; Pu, *Pelagibacter ubique* HTCC1062; Sp, *Silicibacter pomeroyi* DSS-3

particles (Crump et al., 1999), transparent exopolymeric particles (Pedrotti et al., 2009), marine snow (Woebken et al., 2007a) and phytoplankton detritus aggregates (Abell and Bowman, 2005a). These micro–niches are rich in organic material and they constitute "hot spots" for bacterial growth (Azam and Long, 2001). For example, phytoplankton cells release polymers, mainly polysaccharides, during all growth phases to the phycosphere (Myklestad, 1995). Therefore, *Bacteroidetes* that inhabit these micro–niches could hydrolyze the organic material using cell–surface associated hydrolases (Smith et al., 1992; Azam and Malfatti, 2007) (Figure 1.6 B). However, *Bacteroidetes* lineages might respond differentially to phytoplankton as a consequence of a physiological ability to hydrolyze different compounds released by algae cells.

Incubation experiments

Another line of evidence about the role of *Bacteroidetes* lineages in the degradation of DOM is derived from incubation experiments. Covert and Moran (2001) identified that *Bacteroidetes* was the dominant clade in an incubation with high molecular weight OM, but were not in the low molecular weight OM experiment (Covert and Moran, 2001). More precisely, Cottrell and Kirchman (2000) quantified single bacteroidetal cells that incorporated an array of titrated substrates. The authors showed that *Bacteroidetes* constituted the population that consumed the highest amounts of proteins, N-acetyl-glucosamine (NAG), and chitin (Figure 1.6 C). In agreement with this observation, a protein enriched incubation experiment mostly triggered the response of *Bacteroidetes* (Pinhassi et al., 1999).

Genomics

The degradation of DOM by microorganisms mainly depends on the presence and coordinated expression of hydrolytic enzymes (Arnosti, 2003; Warren, 1996) (Figure 1.6 B). Particularly, the crucial step in the degradation of OM is the extracellular break–down of macromolecules to subunits of less than 600 Da which can then be taken up by membrane transporters (Weiss et al., 1991). Polysaccharides, which are important constituents of the OM

produced by algae in the oceans (Benner et al., 1992), can be degraded by glycoside hydrolases (GH). Glycoside hydrolases hydrolyze the glycosidic bonds between two or more carbohydrate moieties, and a synergistic action of different GHs is necessary for the efficient degradation of polysaccharides (Warren, 1996). Its initial breakdown is the limiting step in the degradation (Arnosti, 2003), therefore the microorganisms that initiate the polysaccharide degradation have a pivotal role in the OM recycling in surface oceans.

The whole genome sequence of marine *Bacteroidetes* strains has revealed that they encode a specialized genetic machinery for the degradation of polysaccharides and other high molecular weight substrates. The first genome of a marine aerobic *Flavobacteria*, *Gramella forsetii*, revealed a great glycolytic potential (Figure 1.6 D) (Bauer et al., 2006). A high number of glycoside hydrolases (10.5 per Mbp) were identified in its genome, interestingly several of them show sequence similarity to enzymes known to hydrolyse polysaccharides that are major constituents of plant and marine algal cell walls (Bauer et al., 2006). Interestingly, in the *G. forsetii* genome, approximately half of the GHs are encoded in the vicinity of genes homologous to constituents of an outer membrane starch utilization system (Sus) from *Bacteroides thetaiotaomicron*.

The Sus system has been characterized in the human–associated *B. thetaiotaomicron*, which has an elaborate equipment for acquiring and hydrolyzing otherwise indigestible dietary polysaccharides (Xu et al., 2003). The Sus operon consists of eight genes (*susA*, *susB*, *susC*, *susD*, *susF*, *susE*, *susG*, *susR*) whose products are involved in binding, transporting, and hydrolyzing starch (e.g. Shipman et al., 2000). The products of two adjacent genes within this operon, *susC* and *susD*, are outer membrane proteins involved in binding starch extracellularly, a prerequisite for its subsequent breakdown (e.g. Reeves et al., 1997; Shipman et al., 2000; Cho and Salyers, 2001) (Figure 1.7). The SusC-like proteins are part of a larger family of TonB receptor-like proteins, which are energy–coupled import systems through which substances of low concentration, and of a size too large to permit diffusion, are translocated across the outer membrane (Braun and Herrmann, 1993).

The soil *Bacteroidetes Flavobacterium johnsoniae* and *Cytophagia hutchin-*

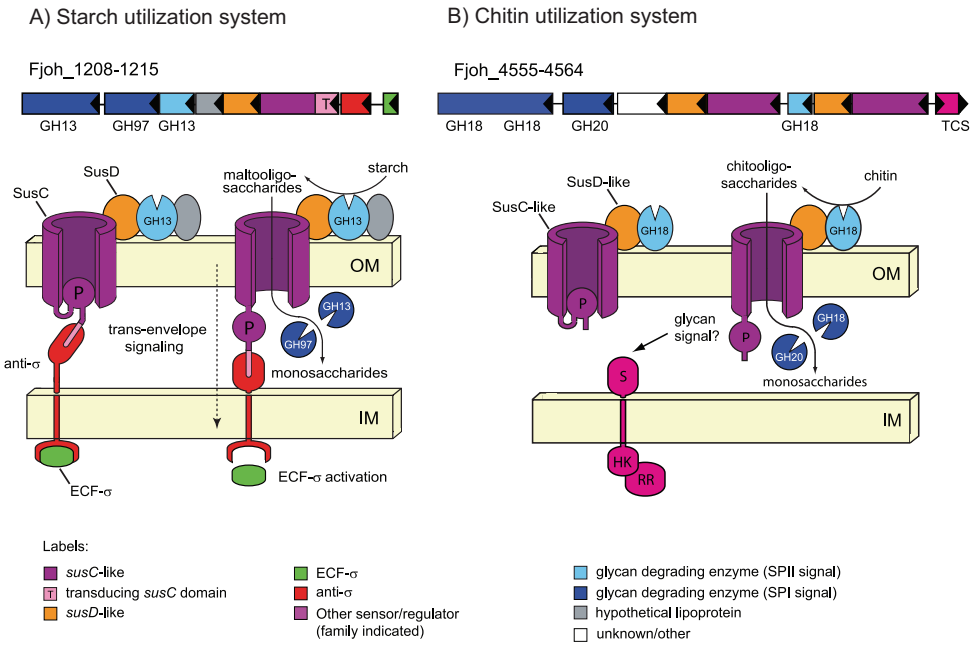


Figure 1.7: Polysaccharide utilization loci (PUL) in *Flavobacterium johnsoniae* (from McBride et al., 2009) (A) Putative PUL with components that are similar to those of the prototypic starch utilization system (Sus) of *B. thetaiotaomicron*. This system includes two GH13 alpha-amylases, and a single GH97 alpha-glucosidase. This system is linked to an extracytoplasmic function sigma (ECF-sigma)-anti-sigma transcriptional regulator pair (green and red, respectively). These regulatory elements function by coupling to a specialized N-terminal "transducing domain"(pink) attached to the SusC-like transporter and together comprise a "transenvelope signaling" pathway spanning both bacterial membranes. (B) Putative chitin utilization PUL which includes three GH predicted to target the beta-1,4-N-acetylglucosamine linkages found in chitin (GH18 and GH20 enzymes). The chitin utilization PUL is associated with a classic two-component regulatory system (dark pink). OM, outer membrane; IM, inner membrane.

sonii might have a similar strategy for the degradation of starch and other polysaccharides (Figure 1.7). The analysis of the *F. johnsoniae* genome suggests that SusC–SusD might also participate in the binding and subsequent hydrolysis of chitin and hemicellulose (Figure 1.7 B). SusD-like cell surface proteins might bind chitin and perform the initial digestion to form soluble oligomers that are transported into the periplasm for further digestion (McBride et al., 2009). In line with the genomic evidence, is the observation

that *F. johnsoniae* cells need to be in contact with insoluble chitin for its efficient utilization (McBride et al., 2003). It has been suggested that in *C. hutchinsonii*, a cellulolytic member of the class *Cytophagia*, SusC and SusD-like proteins may be involved in the binding and utilization of cellulose (Xie et al., 2007).

In the marine realm, there are three more published bacteroidetal genomes: *Polaribacter dokdonensis* MED152 (Gonzalez et al., 2008), and two draft genomes, MS024-2A and MS024-3C, retrieved from single cells by multi displacement amplification (Woyke et al., 2009). Like the genome of *G. forsetii*, also the genomes of *P. dokdonensis* MED152, MS024-2A and MS024-3C encode a high number of GH (7.9-10.4 per Mbp). Moreover, *P. dokdonensis* MED152 encodes five SusC–SusD homologues that are all located in the vicinity of GHs (Gonzalez et al., 2008). It is probable that marine *Flavobacteria* have a similar strategy as *B. thetaiotaomicron* in binding the polysaccharides extracellularly to start its degradation.

Marine *Bacteroidetes* also seem to have a great proteolytic potential. Their genomes encode a high number of genes involved in the degradation of proteins. When the genome of *G. forsetii* was sequenced it had the highest number of peptidases per Mbp (30.5, total 116) after the deep sea gammaproteobacterium *Idiomarina loihiensis* which had been proposed to rely mainly on amino acids for carbon and energy supply (Hou et al., 2004). Other sequenced genomes revealed an even higher number of peptidases, as MS024-2A (38 per Mbp) and MS024-3C (39 per Mbp) (Woyke et al., 2009).

Dissolved proteins, peptides, and aminosugars are an important constituent of the marine OM pool (Nagata, 2000), and may be important sources of energy and nitrogen for marine bacteria (Benner and Kaiser, 2003; Obayashi and Suzuki, 2005). The genomic bacteroidetal proteolytic potential is in line with observations in incubation experiments (Figure 1.6 C), in which blooms of *Bacteroidetes* were triggered by protein amendment. Moreover, it has been reported that several *Tenacibaculum* spp. strains are unable to use nitrate or ammonium as a nitrogen source, indicating that they require an organic nitrogen source for growth (Suzuki et al., 2001; Pinhassi et al., 2004). As well, *G. forsetii* lacks the genes for nitrate or nitrite assimilation, and it

seems to rely on reduced nitrogen sources as organically bound nitrogen in amino acids (Bauer et al., 2006).

Another interesting feature revealed by the genome sequencing of marine strains was the identification of the proteorhodopsin gene in *P. dokdonensis* MED152 (Gonzalez et al., 2008), *Dokdonia donghaensis* MED134 (Gomez-Consarnau et al., 2007), and in both single amplified genomes MS024-2A and MS024-3C (Woyke et al., 2009). Additionally, the growth of *Dokdonia donghaensis* MED134 is enhanced when exposed to light at low or intermediate organic carbon concentrations. It is hypothesized that proteorhodopsin gives *D. donghaensis* MED134 an advantage when the availability of OM or polysaccharides decrease (Gomez-Consarnau et al., 2007). However, *G. forsetii* does not encode the proteorhodopsin gene (Bauer et al., 2006) pointing at distinct adaptations to the marine environment.

Overall, whole genome sequencing indicates that marine *Bacteroidetes*, and particularly *Flavobacteria*, might play an important role in the degradation of the oceanic OM pool. Since *Bacteroidetes* is a phylogenetically highly diverse phylum (Alonso et al., 2007), our knowledge of the genomic potential of the bacteroidetal community in the ocean is still largely incomplete. Culture-independent techniques are necessary in order to better understand the role of the major uncultured marine *Bacteroidetes* clades and to link them with their distribution patterns in distinct marine scenarios.

Chapter 2

Aims

General Objective

The aim of this PhD thesis was to thoroughly characterize the diversity, distribution patterns, and ecological role of marine planktonic *Bacteroidetes* clades by culture-independent approaches.

Specific Objectives and Hypotheses

I In order to quantify specific *Bacteroidetes* populations that constituted a low fraction of the picoplankton, a modification of the standard FISH protocol was needed. One objective of this thesis (Section 3.1) was therefore the development, testing and application of a modified protocol for sampling and counting populations that constitute less than 1% of the picoplankton. We hypothesized that the number of FISH–positive signals will linearly increase with increasing volumes of filtered sample, and hence "rare" populations can be reliably quantified. Manuscript I

II The objective of Section 3.3 was to quantify flavobacterial clades that inhabit contrasting oceanic provinces in the northern North Atlantic Ocean and to identify environmental factors that control their abundance. We hypothesized that: (i) general bacteroidetal abundances are

higher in the areas where temperature is low and nutrient concentration high, and in the euphotic zone in which phytoplankton abundance is high, (ii) the diversity varies locally, and (iii) different flavobacterial clades prevail in different water masses and are correlated with distinct environmental parameters. Manuscript I

III The objective of the study presented in Section 3.4 was to gain insights into the potential function of the various *Bacteroidetes* clades (defined in Section 3.3) in two contrasting oceanic provinces in the North Atlantic Ocean. Particularly, the focus of this study was to search for genes involved in the degradation of polysaccharides, proteins, and adhesion to surfaces or cells. We hypothesized that *Bacteroidetes* in the nutrient richer areas will be enriched in an array of proteins involved in the degradation of phytoplankton derived polysaccharides, surface adhesion and gliding. Manuscript II

IV The objectives of Section 3.5 and 3.6 were to investigate if the *Bacteroidetes* clades defined in an open ocean environment also inhabit coastal areas. More specifically, to investigate *Bacteroidetes* clades population sizes and distribution patters in a river-ocean gradient and within a coastal lagoon. We hypothesized that (i) total *Bacteroidetes* abundance will be higher in coastal areas and will tend to increase where chlorophyll a and organic matter is higher, (ii) the open ocean *Bacteroidetes* clades will be present in coastal settings, but abundances will vary along the gradients. Manuscript III and IV

Chapter 3

Results and Discussion

3.1 Testing a modified fluorescence *in situ* hybridization counting protocol for rare populations

One of the limitations of fluorescence *in situ* hybridization (FISH) is its rather high detection limit. The lower threshold for confident quantification typically ranges between 0.1-2% of total picoplankton counts. Hybridizations performed with 16S rRNA–targeted oligonucleotide probes for narrow phylogenetic clades (i.e. few 16S rRNA gene sequences) might result in less than 1 positive cell per microscopic field. Hence, such infrequent events will decrease the counting precision leading to an unreliable determination of population sizes (Pernthaler et al., 2003).

One part of this thesis was to test a modified protocol for sampling and counting populations that constitute less than 1% of the picoplankton. The experimental idea for the modification of the counting protocol was that the FISH detection limit could be lowered by filtering larger volumes of seawater samples resulting in a higher cell density. We hypothesized that the number of FISH–positive signals will linearly increase with increasing volumes of filtered sample. Therefore, it will result in no significantly different calculated abundances between the filters through which high and low volumes of

seawater were filtered, from here on named high and low volume filters.

For the testing of the quantification protocol, picoplankton populations were enumerated from filters containing increasing volumes of seawater samples. Samples were taken at the German North Sea island Helgoland (54°11.3' N; 07°54.0'E) and were filtered in increasing amounts from 10, 25, 50, 100, 250, to 400 ml, in triplicates for each volume. Hybridizations were performed with probes targeting picoplankton clades that comprise different ranges of abundance. The probes used were ROS537 (Eilers et al., 2001) for the *Roseobacter* clade, NOR5-730 (Eilers et al., 2001) for the gammaproteobacterial NOR5/OM60 clade, and PLA46 (Neef et al., 1998) for the *Planctomycetes* phylum.

In the standard FISH quantification protocol usually 10-50 ml of seawater are filtered on a 47 mm diameter polycarbonate filter. For quantification, first all FISH positive cells and subsequently all cells stained with a general DNA dye, usually DAPI, are counted in the same microscopic field. In the alternative protocol tested, only the probe-conferred fluorescence was counted in twenty microscopic fields for each volume replicate and the absolute abundance of the clade was determined after subtraction of the negative control. The high volume filters resulted in agglomeration of DAPI–stained cells and therefore the total cells counts were done in a low volume filter.

For all the evaluated clades, probe-specific counts proportionally increased with increasing amounts of filtered sample R^2=0.9 (Figure 3.1 A, C, E). Signals in controls for unspecific staining by CARD–FISH stayed low, accounting for only 1-2 signals in 10 microscopic fields, both in the high and low volume filters. Whereas calculated clade abundances were not significantly different between the different volume samples, the standard errors were markedly lower for the high volume filters (Figure 3.1 B, D, F). The upper limit in the amount of filtered sample will depend on the abundance of the targeted clades. For example, in the case of NOR5/OM60 (Figure 3.1 C, D), which was 3% of the total picoplankton, a 400 ml volume sample resulted in agglomeration of the NOR5-730 positive cells and therefore the final abundance would be underestimated. Yet, for rare populations of less than 1% of picoplankton high–volume FISH filters markedly improve the accuracy of the

counts.

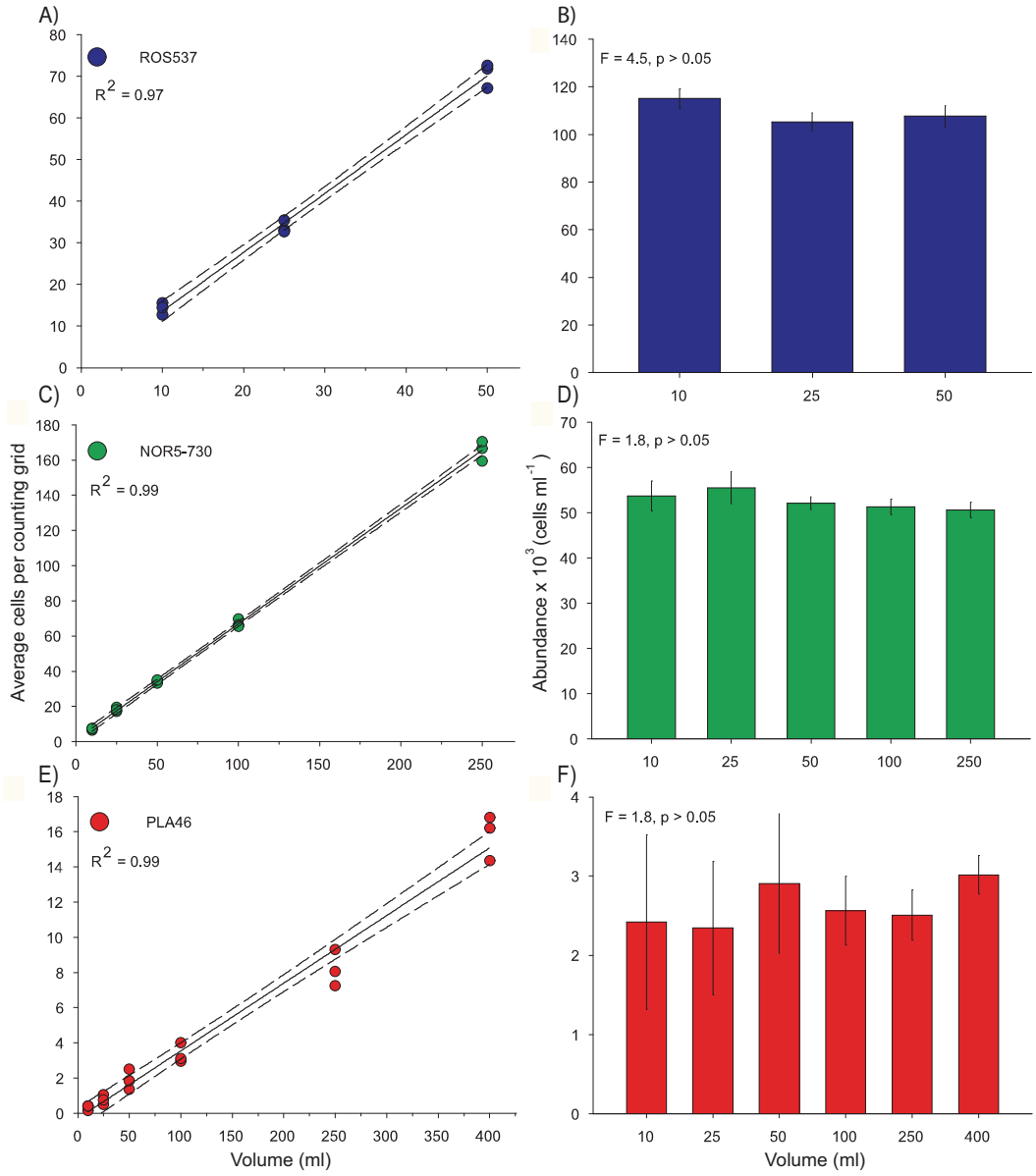


Figure 3.1: Evaluation of high volume water samples with the probes ROS537 (A-B), NOR5-730 (C-D) and PLA46 (E-F). (A), (C) (E): Average cells per counting grid at the different increasing volume samples indicating regression line (solid line), regression coefficient (R^2) and 95% confidence intervals (dashed line) of the 3 volume replicates. (B) (D) (F): Calculated abundance of each clade indicating F statistic and p–value result from ANOVA (B and D) or ANOVA on ranks (F) tests.

We concluded that the use of high volume filters is suitable and reliable for the calculation of the abundance of rare populations. High–volume FISH filters were used for the quantification of *Bacteroidetes* rare clades in the North Atlantic Ocean.

3.2 Characterization of the northern North Atlantic biogeographic provinces

Oceanic biogeographic provinces are categories used for comparing biogeochemical processes and biodiversity between ocean regions. One of the classification systems of the pelagic environment is that proposed by Longhurst and colleagues (Longhurst, 1995), which is based on basin-scale empirically–derived chlorophyll and primary production estimates from remotely sensed measurements of ocean color, as obtained with the Coastal Zone Color Scanner (CZCS). The province concept has provided a framework to spatially aggregate or separate data for comparisons of biogeochemical processes over broad regions of the global ocean. However, it is unknown how province boundaries respond to seasonal and climate forcing (Oliver and Irwin, 2008). To overcome this limitation, Oliver and Irwin (2008) use an objective classification on global remote sensing data to automatically produce time and space resolved ocean provinces.

The samples of the studies presented in Sections 3.3 and 3.4 were taken along a transect in the North Atlantic Ocean from 66°39.27'N; 29°36.65'W (station 2=S2) to 34°24.87'N; 28°28.90' W (S19) during the VISION cruise (diVersIty, Structure and functION) MSM03/01 on board the research vessel Maria S. Merian from September 21 to September 30, 2006 (Figure 3.2). The study area could be subdivided in four oceanic provinces according to the Longhurst classification system (Longhurst, 1998). The boundaries of the oceanic provinces as defined by the Longhurstian classification were in good agreement with the delimitations according to Oliver and Irwin (2008) based on the satellite-observed thermal and optical properties of the study area during September 2006 (Figure 3.2 A). However, the four Longhurstian provinces could be further separated by the latter approach into nine distinct water masses (Figure 3.2 B).

The **Boreal Polar Province (BPLR)** comprises the Arctic Ocean, between North America, Greenland, and Asia, and the coastal Greenland and Labrador Currents. The BPLR is dominated by ice cover that opens significantly during summer in some coastal and archipelagic regions. The seaward

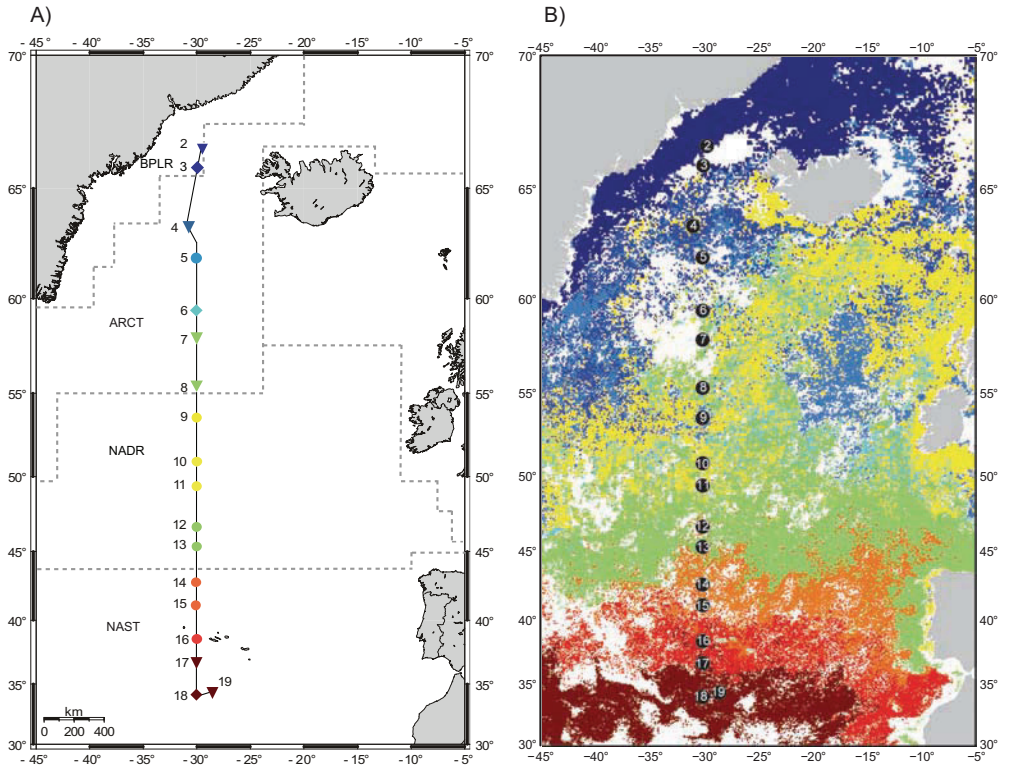


Figure 3.2: (A) Map indicating the VISION cruise track and sampled stations S2-S19 and boundaries of the Boreal Polar (BPLR), Arctic (ARCT), North Atlantic Drift (NADR) and North Atlantic Subtropical (NAST) oceanic provinces. Color of symbols represents different water masses as shown in B. (B) Significant water masses determined by the 2006 September mean of Advanced Very High Resolution Radiometer (AVHRR) sea surface temperature and the Sea- viewing Wide Field-of-view Sensor (SeaWiFS) water leaving radiance according to Oliver and Irwin (2008). The different colors represent different water mass types.

boundary of the East Greenland Current (EGC), the East Greenland Polar Front, defines the outer boundary of the province. At the East Greenland Polar Front the stream of polar surface water meets warmer, saltier Atlantic water. The northernmost stations (S2 and S3) of the VISION cruise track were located at the boundary of BPLR and ARCT (see below) and satellite derived water masses confirmed that they represented a separate area. These stations had an average chlorophyll a concentration of 0.7 μg l^{-1} (Figure 3.3) and surface water temperature below 3°C. Those stations had low salinity

(<33 PSU), which increased in the following stations of the Atlantic Arctic Province (Figure 3.4 A, B).

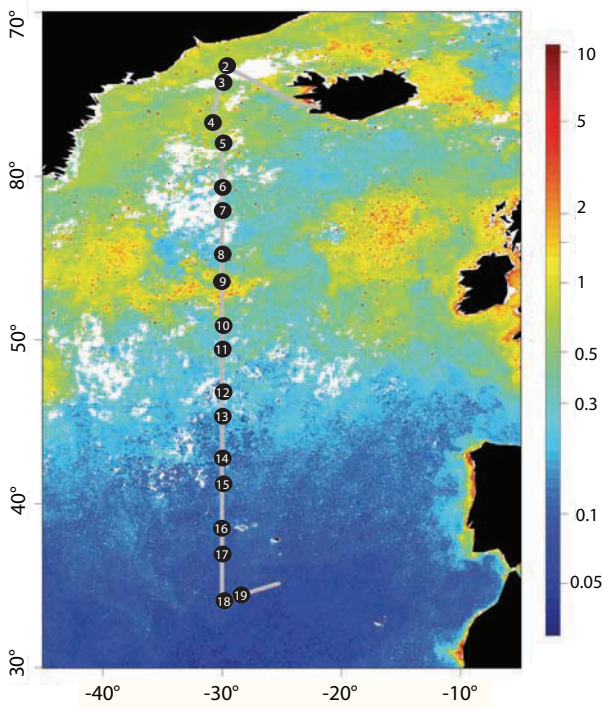


Figure 3.3: September 2006 monthly average chlorophyll a concentration (μg l^{-1}) from the SeaWiFS satellite (http://oceancolor.gsfc.nasa.gov). Missing data is shown as white

The **Atlantic Arctic Province (ARCT)** lies between the East Greenland Polar Front, seawards of the EGC, and the oceanic Polar or Subarctic Front. The limits of the ARCT province are rather variable and cannot often be traced in satellite images. By the end of the summer, to the north of the Oceanic Polar Front (52°N-54°N) the chlorophyll maximum is within the mixed layer and it had maximum values in comparison with the next province (Longhurst, 1998). In our sampling, the surface chlorophyll a concentrations were variable within the province (0.3-1 μg l^{-1}) and four water masses (stations S4, S5, S6, S7+S8) were differentiated. The higher surface chlorophyll a concentration was reached in latitudes 63°N-61°N (0.7-1 μg l^{-1}). The temperature in the upper mix layer ranged between 10-12°C and salinity between

34.9-35.1 PSU (Figure 3.4 A, B). The concentrations of phosphate (PO_4^{-3}), nitrate (NO_3^-), nitrite (NO_2^-) and dissolved silicate (Si) in the upper 50 m were higher in the BPLR and ARCT, ranging between 0.2-0.6 μM, 2.0-6.5 μM, 0.1-0.2 μM and 0.6-3.2 μM, respectively, in comparison with the more southern provinces.

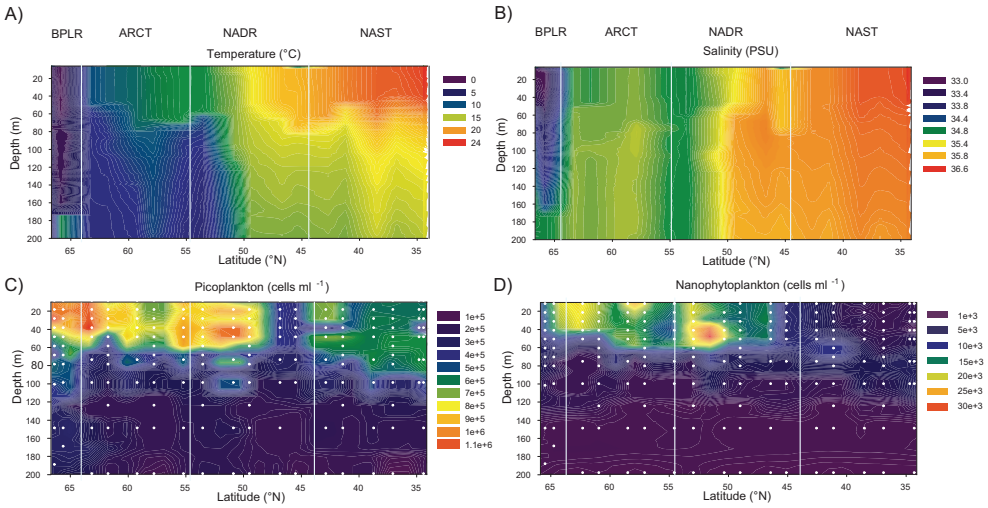


Figure 3.4: Latitudinal depth distribution of (A) Temperature (°C), (B) Salinity (PSU), (C) abundance of picoplankton (cells ml^{-1}) and (D) abundance of nanophytoplankton (cells ml^{-1}). Dots represent the sampling depths. Oceanic provinces as indicated in Figure 3.2

The **North Atlantic Drift Province (NADR)** comprises the North Atlantic Current (NAC), south of the Oceanic Polar Front. The southern boundary of the province lies at approximately 40° N. In this province, at about 45°N, is where the spring bloom starts and then develops through the province. By the end of the summer the transition to oligotrophic conditions had occurred patchily, to the south of 46°N nitrate is depleted and the chlorophyll maximum is significantly deeper than the bottom of the mixed layer (Longhurst, 1998). Along the VISION cruise transect, the stations S9 to S13 were located in the NADR. The surface chlorophyll a concentration ranged between 0.2-0.6 μg l^{-1}, the temperature gradually increased up to 20°C and the concentration of inorganic nutrients decreased along the NADR to very

low values. The abundance of picoplankton was higher in the BPLR, ARCT and NADR provinces (0.4-1.1 $\times$ 10^6 cells ml^{-1}) than in the NAST province (0.2-0.6 $\times$ 10^6 cells ml^{-1}) (Figure 3.4 D). The abundance of nanophytoplankton was also higher in the upper mix layer of the BPLR, ARCT and NADR (12-36 $\times$ 10^3 cells ml^{-1}), and it steeply decreased in the NADR at 46-45°N (8×10^3 cells ml^{-1}) (Figure 3.4 C).

The **North Atlantic Subtropical Gyral Province (NAST)** comprises the central gyre of the North Atlantic polewards of the subtropical convergence. The mid-Atlantic Ridge constrains the main recirculation gyre within the western basin. The NAST is in general an oligotrophic area with low levels of primary production. In the southernmost water body (S17-S19) of the NAST province (S14-S19) the chlorophyll a concentration and nanophytoplankton abundance had the lowest values, <0.1 μg l^{-1} and 2.7-8.2 $\times$ 10^3 cells ml^{-1}. The concentration of nutrients was also very low, with $PO_4{}^{-3}$, $NO_3{}^-$, $NO_2{}^-$ <0.06 μM and Si <0.3 μM. $NH_4{}^+$ concentrations ranged between 0.2-0.6 μM in surface layers, and decreased below detection limit in deeper waters. The temperature and salinity reached maximum values of 24°C and 36.6 PSU, respectively.

In summary, the northernmost stations of the VISION cruise track (BPLR and northern ARCT provinces) had the higher nanophytoplankton abundance, chlorophyll a and nutrient concentrations. Contrastingly, the southernmost stations had very low levels of the above mentioned parameters but high temperature and salinity. This setting of contrasting environmental characteristics was therefore a good testing ground for hypothesis on the biogeography of *Bacteroidetes* clades and on their role in the degradation of organic matter.

3.3 Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean

The major objective of the this study was to quantify flavobacterial clades that inhabit contrasting oceanic provinces in the northern North Atlantic Ocean and to identify environmental factors that control their abundance. Firstly, the distribution of total *Bacteroidetes* was investigated by CARD–FISH with the general probe CF319a along the VISION transect (Figure 3.2) and in the depth profile. The *Bacteroidetes* diversity was characterized by comparative sequence analysis of 16S rRNA gene libraries in three stations of contrasting environmental characteristic (S3, BPLR; S6, ARCT, and S18 NAST) in which *Bacteroidetes* represented a different fraction of the total picoplankton. Based on the phylogenetic analysis of the 16S rRNA gene sequences of flavobacterial clades were delimited. Published and newly designed oligonucleotide probes were used to enumerate eleven flavobacterial clades by CARD–FISH. To further explain the observed distribution patterns correlation analyses were done with environmental data (physicochemical parameters and biological variables). Additionally, we performed a preliminary check regarding whether a correlation detected with nanophytoplankton was caused by a direct attachment of *Flavobacteria* to the phycosphere.

Bacteroidetes diversity and distribution

Our data of the transect in the North Atlantic between 65°N and 33°N confirm the preference of *Bacteroidetes* for cold and nutrient rich water masses (Figure 3.5) that had been reported by others (Simon et al., 1999; Brinkmeyer et al., 2003; Kirchman et al., 2003; Abell and Bowman, 2005b). The maximum *Bacteroidetes* abundance was in surface waters of the BPLR province (18.2 ± 2.8%) and it decreased along ARCT and NADR to a minimum of 7.0 ± 1.1% in the NAST province. However, we also demonstrate that flavobacterial clades thrive under oligotrophic conditions in warm waters,

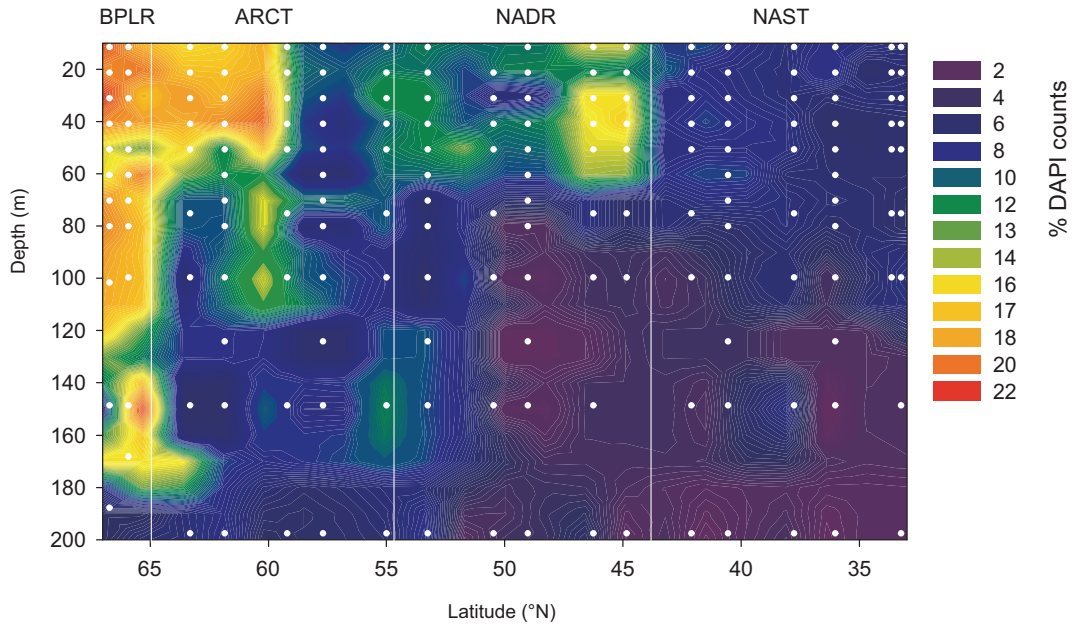


Figure 3.5: Latitudinal depth distribution of *Bacteroidetes* as detected by probe CF319a (relative abundance, % of DAPI counts). Oceanic provinces as indicated in Figure 3.2. Dots represent the sampling depths.

both in the NADR and the NAST, extending data of an Atlantic Meridional Transect cruise between 40°N and 20°S (Schattenhofer et al., 2009). The following question was whether a shift in the *Bacteroidetes* community composition in contrasting provinces occurred. Three 16S rRNA gene libraries were constructed from surface water samples of S3 (BPLR), S6 (ARCT), and S18 (NAST) in which *Bacteroidetes* represented a different fraction of the picoplankton. Because the abundance of *Bacteroidetes* was highest in the upper 50 m, we focused further analysis on the mixed surface water layer.

Comparative sequence analysis showed that sequences affiliated with the class *Flavobacteria* dominated all three libraries (Figure 3.6) despite that the primer pair used for screening the clone libraries had a similar coverage of the classes *Flavobacteria* and *Sphingobacteria* (Chen et al., 2006) including the class *Cytophagia*, adding further evidence for the dominance of class *Flavobacteria* in marine environments. Moreover, a detailed phylogenetic analysis placed most of the flavobacterial sequences within the "marine

clade" (Bowman, 2006), and several of them were affiliated with cultured genera as *Polaribacter* spp. , *Leeuwenhoekiella* spp. , and *Dokdonia* spp. Most of the sequences clustered with yet uncultured clades mostly retrieved from various marine pelagic environments, both coastal and open ocean as well as temperate and polar sites. Sequences from the three stations were affiliated with both the DE2 clade (Kirchman et al., 2003) and with the AGG58 clade (O'Sullivan et al., 2004). We also defined new clades on the basis of sequences from the VISION cruise (VIS): VIS1, VIS2, VIS3, VIS4, VIS5 and VIS6 (Figure 3.6 and FISH oligonucleotide probes were designed in order to quantify them. All marine *Bacteroidetes* FISH probes designed in this thesis are listed in Appendix A.

Distinct distribution patterns of flavobacterial clades

First FISH analyses using oligonucelotide probes for the clades VIS2, VIS3, VIS5 and VIS6 resulted in counts of less than 1 cell per microscopic field, corresponding to relative abundances of 0.5% and lower. Similarly, in the more southern stations, the otherwise abundant clades VIS1, VIS4 and DE2 dropped below 0.5% of total bacteria in abundance. Such infrequent events were below the threshold for confident quantification by FISH. In this framework, we tested the high volume FISH filters described in Section 3.1. Additionally to the tests presented in Figure 3.1, cells of clade DE2 were counted on low– and high–volume filters from S3 and no statistical difference in the counts was detected ($P<0.05$). On the basis of those results, except for *Polaribacter* spp. all flavobacterial clades were quantified on 500 ml filters.

CARD–FISH proved that members of the genus *Polaribacter* were present throughout the transect and the absolute numbers in the surface waters were consistently high (e.g. $72 \pm 45 \times 10^3$ cells ml^{-1} in S4; $5.5 \pm 0.7 \times 10^3$ cells ml^{-1} in S19). The abundance was negatively correlated with temperature (r_s= -0.89) indicating a preference for colder water masses of less than 10°C, which is within the optimal growth temperature of several cultured strains (Gosink et al., 1998). Our data further supports the importance of *Polaribacter* spp. in polar environments (Bano and Hollibaugh, 2002; Brinkmeyer

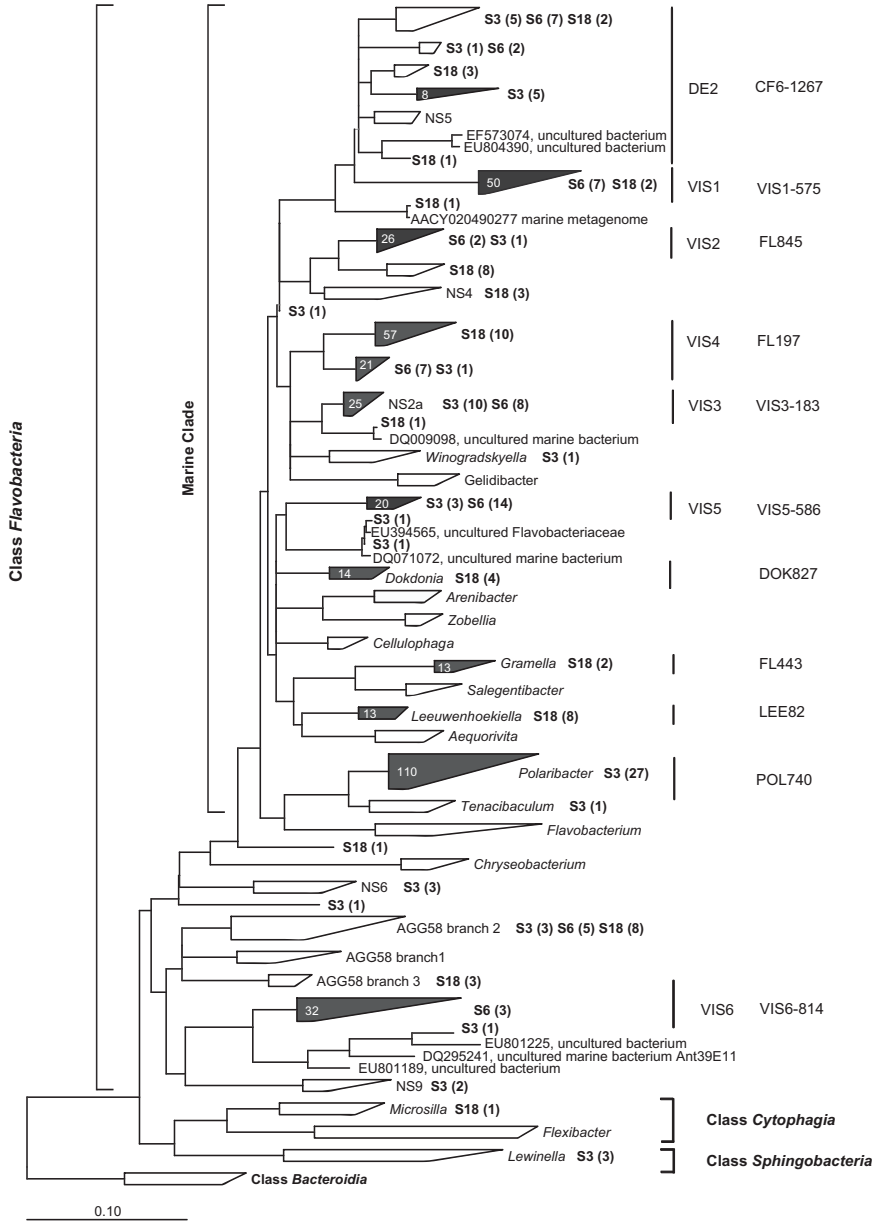


Figure 3.6: Phylogenetic association of sequences retrieved from BPLR (S3), ARCT (S6) and NAST (S18) indicating the number of sequences of each station that contributes to the clades shown in wedges. Grey shadows indicate target groups of flavobacterial clade–specific oligonucleotide probes indicating the number of targeted sequences in the database Silva 96 (Pruesse et al., 2007). Name of the probes are given to the right.

et al., 2003). Although the abundance of *Polaribacter* spp. decreased more than one order of magnitude in NAST, where the temperature was more than 20°C, they were still the largest of the flavobacterial clades in these blue waters. This could be the result of a more cosmopolitan distribution of a single *Polaribacter* lineage (Section 3.6).

Clades VIS1 and VIS4, both encompassing mostly 16S rRNA gene sequences from temperate and subtropical open ocean sites, presented its maximum abundance in the ARCT province. VIS1 was above 1×10^4 cells ml^{-1} throughout ARCT only, a density reached by VIS4 at S4 and S6, but not at S7 and S8. Clade DE2 was again different, with a maximum slightly below 1×10^3 cells ml^{-1} in BPLR. DE2 comprises sequences that have mostly been retrieved from coastal systems as the Delaware Estuary (Kirchman et al., 2003), the North Sea (Eilers et al., 2001; Alonso et al., 2007) and the coast of Plymouth (O'Sullivan et al., 2004). The abundance of DE2 in our open ocean samples was significantly lower than the 10% previously reported e.g. for the Delaware Estuary (Kirchman et al., 2003). On our transect the BPLR stations might have been the ones with the most significant coastal influence.

The VIS2, VIS3, VIS5 and VIS6 clades consistently formed groups of <0.5% relative abundance. In this study, we could quantify these populations down to absolute numbers of only 10^2 cells ml^{-1}. Interestingly, of the rare clades only the abundance of VIS2 was explained by an environmental variable. It was positively correlated with temperature (r_s=0.73). This group consists of sequences retrieved from coastal environments and open ocean, but none from polar origin, pointing at a preference of clade VIS2 for warmer waters. The rare clades VIS3 and VIS6 significantly decreased from S6 to S7, mirroring a change in the optical properties of the water masses. Clade VIS6 comprises phylotypes from the North Atlantic Ocean as well as phylotypes previously retrieved during a phytoplankton bloom in naturally iron fertilized waters (West et al., 2008) but in this study, we could not detect a correlation between abundance of VIS6 and micro– or nanophytoplankton abundance or chlorophyll a fluorescence.

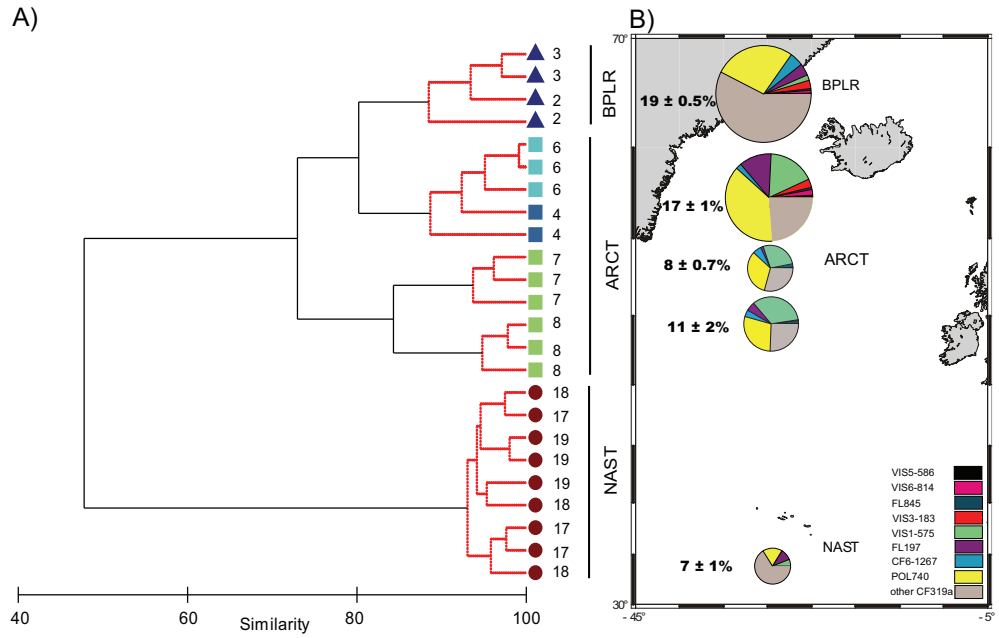


Figure 3.7: Flavobacterial community composition, as analyzed by FISH with clade specific–probes in the BPLR, ARCT and NAST. (A) Dendrogram representing the similarity of the flavobacterial community in the three provinces. Symbols color coded as in Figure 3.2 and numbers indicate station replicates. Black lines in the dendrogram represent significant clusters at the 99% level using SIMPROF (similarity profile permutation test, Clarke et al., 2008); red lines represent not significant clusters at this level. (B) Relative contribution of each flavobacterial subgroup to the total CF319a. Pie chart represents average group sizes of clustering stations in (A); area of pie represents the relative abundance of CF319a (percentage is given to the left).

An analysis of the 16S rRNA gene sequences similarities between the sites revealed that there was only little overlap between the different water bodies. S3 and S18 shared only a 2-5% of its phylotypes (at a 97% identity threshold) while S3 and S6, which were more similar in terms of environmental characteristics, shared 25-40% of flavobacterial phylotypes. CARD-FISH quantification of flavobacterial clades added numbers in support of distinct populations along the transect (Figure 3.7). The shifts in the flavobacterial community coincided with transitions from one water mass to another, indicated by the different color symbols in Figure 3.7. Within the ARCT

province, for example, a sharp decrease in chlorophyll a fluorescence was evident from S6 to S7 which was mirrored by an abrupt reduction of VIS3, VIS4, and VIS6 populations as well as total *Flavobacteria* (Figure 3.7). Overall, we identified marked distribution patterns of flavobacterial clades in contrasting water bodies of the North Atlantic Ocean. Our results are in line with the increasing body of evidence that distinct picoplankton populations show pronounced seasonal (Fuhrman et al., 2006), latitudinal (Fuhrman et al., 2008b) and depth (DeLong et al., 2006) distribution patterns in marine systems.

Phytoplankton association of flavobacterial clades

Links between members of the class *Flavobacteria* and phytoplankton have been frequently reported (Simon et al., 1999; O'Sullivan et al., 2004; Pinhassi et al., 2004; Grossart et al., 2005; Sapp et al., 2007) which has been related with their potential ability to degrade the molecules released by the phytoplankton (Kirchman, 2002). In this study we detected a significantly positive correlation between total *Flavobacteria* as quantified by probe CF319a and nanophytoplankton ($r_s = 0.76$), as well as for the clades DE2 and VIS1 with nanophytoplankton ($r_s = 0.57$ and $r_s = 0.74$, respectively). The high–volume 500 ml filters of S3 were used to do a preliminary check regarding whether the correlation is caused by a direct attachment of *Flavobacteria* to the phycosphere of nanophytoplankton cells. Microscopy showed that *Bacteroidetes* did indeed reside in the phycosphere of nanophytoplankton cells (Figure 3.8). As this observation could be a consequence of random agglomeration of cells during filtration, we compared the localization of CF319a-positive cells with the localization of the abundant alphaproteobacterial SAR11 clade. We quantified the CF319a–positive cells inside and outside the phycosphere of nanophytoplankton (Figure 3.8 A) and compared the ratio with the relative localization of the free–living SAR11 clade (Figure 3.8 B) evaluated with the probe SAR11-441 (Morris et al., 2002). The SAR11 cells were more abundant (27% at S3, 20 m; M Schattenhofer, unpublished results) than *Flavobacteria* (19% at S3, 20 m) in the BPLR province. Assuming a random co–occurrence, more SAR11 cells would be expected per nanophytoplank-

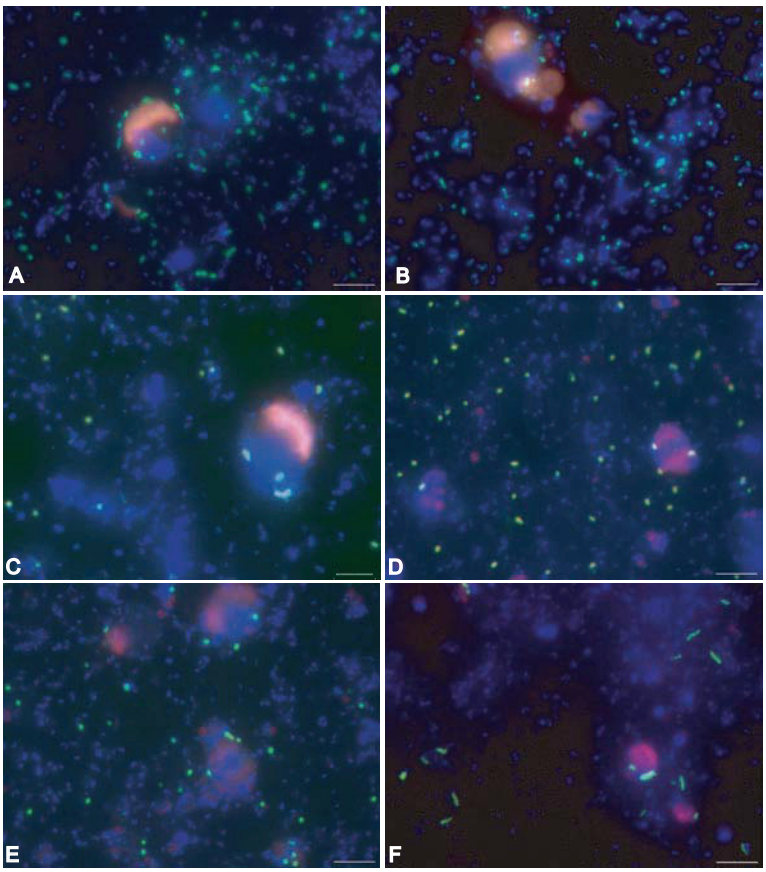


Figure 3.8: Epifluorescence micrographs of nanopytoplankton cells and bacteria in the phycosphere. Red, autofluorescence of nanophytoplankton cells; blue, DAPI signals of DNA containing cells; and green, CARD–FISH positive cells of: (A) *Bacteroidetes*, (B) SAR11 clade, (C) *Polaribacter*, (D) VIS1, (E) VIS4 and (F) DE2. A, B, C, F: S3, 20 m; D, E: S4, 20 m. Bar = 5 μm

ton cell compared with flavobacterial cells. However, the ratio inside and outside the phycosphere was significantly higher ($p<0.01$) for CF319a (2.5 $\pm$ 0.8, n=13) than for SAR11 (0.8$\pm$0.2, n=10). Our results clearly indicate a preference of *Flavobacteria* for attachment to nanophytoplankton at S3 in comparison with a typical free–living bacterium as SAR11 (Giovannoni et al., 1990; Rappe et al., 2002). Moreover, microscopy demonstrated that cells of four of the flavobacterial clades, VIS1, DE2, VIS4 and *Polaribacter*, did also reside in the phycosphere of nanophytoplankton cells (Figure 3.8).

We hypothesize that future studies will demonstrate distinct, and pre-

dictable preferences of specific strains of *Flavobacteria* to particular nanophytoplankton species, further underlining the strict coupling of phytoplankton and picoplankton dynamics in this fascinating group of marine bacteria. Genomic data would yield insights about their metabolic potential and in how they interact with phytoplankton.

3.4 Genomic content of *Bacteroidetes* from contrasting oceanic provinces in the North Atlantic Ocean

In the previous study we identify major *Bacteroidetes* clades in the North Atlantic Ocean and demonstrate that they have distinct distributions. The motivation of the investigation presented in this section was to get first insights into the metabolic potential of the *Bacteroidetes* clades in contrasting oceanic provinces, and to compare them with fully sequenced *Bacteroidetes* genomes. Two large genomic insert metagenomic libraries were constructed from DNA extracted from the BPLR and NAST. Clones for sequencing were selected based on a screening with bacteroidal 16S rRNA gene PCR primers. Screening of approximately 15,500 fosmid clones from S3 and 27,600 clones from S18 yielded 28 *Bacteroidetes* fosmids, which were subsequently sequenced.

Several highly similar fosmids were identified in both libraries and therefore the longer overlapping fragments were merged and considered as one contig (Table 1, indicated with an M, supplementary Table 1 of Manuscript II), while shorter redundant fragments were excluded from further analysis. A total of 21 fosmids roughly equaling 0.9 Mbp of non–redundant sequence information and carrying a total of 695 open reading frames (ORFs) were further analyzed.

All *Bacteroidetes* fosmids carried the 16S rRNA gene as phylogenetic marker. This allowed the establishment of a link between metabolic information and organism identity and abundance, thus providing insights about the significance of certain metabolic traits in different ecological settings.

Table 1: Chracterization of the *Bacteroidetes* fosmids retrieved in this study

		Genomic characteristics			Closest relatives					Abundance		
Fosmid	**Clade**	**Length (kbp)**	**ORFs**	**GC (%)**	**genome-sequenced strain**	**identity (%)**	**VISION clone**	**Accession N°**	**identity (%)**	**Probe**	**Cell ml^{-1}x 10^{2}**	
Province / Station											**BPLR**	**NAST**
BPLR / S3												
S3-843-F5	*Polaribacter*	41.1	32	32.6	*Polaribacter irgensii* 23-P	98	VIS-St3-9	FN433361	97	POL740^{+}	500 ± 150	68 ± 15
S3-860-D3	*Polaribacter*	44.9	27	32.5	*Polaribacter irgensii* 23-P	98	VIS-St3-41	FN433365	99	POL740^{+}	500 ± 150	68 ± 15
S3-805-E4	DE2	39.6	33	35.3	Flavobacteria MS024-2A	94	VIS-St18-44	FN433426	93	F805E4-471	25 ± 1.1	n.d.
S3-861-A8	VIS5	39.2	29	33.1	*Dokdonia donghaensis* MED134	92	VIS-St3-57	FN433381	99	VIS5-586	5.2 ± 1.4	n.d.
S3-933-H4	*Flavobacteria*	27.8	29	35.1	*Psychroflexus torquis* ATCC 700755	85	VIS-St3-103	FN433403	85	-	-	-
S3-979-C6 (M^{1})	VIS6	36.8	27	43.1	*Kordia algicida* OT-1	87	VIS-St6-77	FN433393	96	VIS6-814	17 ± 5.9	n.d.
S3-825-D3 (M^{1})	VIS6	31.9	23 (4)	43.3	*Kordia algicida* OT-1	87	VIS-St6-77	FN433393	96	VIS6-814	17 ± 5.9	n.d.
S3-816-G8	ANT	39.3	28	38.7	Flavobacteria bacterium BAL38	86	VIS-St3-66	FN433355	99	ANT-981 ANT-72	25 ± 10	n.d.
S3-858-E2 (M^{2})	Spingobacteria A	36.7	34 (33)	37.2	*Candidatus* Amoebophilus asiaticus 5a2	81	VIS-St3-37	FN433448	99	SPA-730	9.5	n.d.
S3-892-C7 (M^{2})	Spingobacteria A	38.7	30 (19)	37.3	*Candidatus* Amoebophilus asiaticus 5a2	81	VIS-St3-37	FN433448	99	SPA-730	9.5	n.d.
S3-893-A7 *	Spingobacteria A	36.0	24	37.7	*Candidatus* Amoebophilus asiaticus 5a2	81	VIS-St3-37	FN433448	99	SPA-730	9.5	n.d.
S3-891-H6 *	Spingobacteria A	37.2	25	37.3	*Candidatus* Amoebophilus asiaticus 5a2	81	VIS-St3-37	FN433448	99	SPA-730	9.5	n.d.
S3-972-A4	Sphingobacteria	30.8	32	38.4	*Microscilla marina* ATCC 23134	82	none	-	-	-	-	-

Table 1: Chracterization of the *Bacteroidetes* fosmids retrieved in this study, continued

		Genomic characteristics			Closest relatives					Abundance		
Fosmid	**Clade**	**Length (kbp)**	**ORFs**	**GC (%)**	**genome-sequenced strain**	**identity (%)**	**VISION clone**	**Accession N°**	**identity (%)**	**Probe**	**Cell ml^{-1}x 10^2**	
Province / Station											**BPLR**	**NAST**
NAST / S18												
S18-841-H9	*Dokdonia*	43.1	39	39.8	*Dokdonia donghaensis* MED134	100	VIS-St18-40	FN433430	99	DOK827	n.d.	<1
S18-997-C3	*Leeuwenhoekiella*	37.0	20	36.4	*Leeuwenhoekiella blandensis* MED217	96	VIS-St18-9	FN433425	99	LEE82	n.d.	<1
S18-1087-F8 (M^3)	*Leeuwenhoekiella*	41.5	39	37.6	*Leeuwenhoekiella blandensis* MED217	96	VIS-St18-9	FN433425	99	LEE82	n.d.	<1
S18-1013-C10 (M^3)	*Leeuwenhoekiella*	35.1	36 (19)	35.9	*Leeuwenhoekiella blandensis* MED217	96	VIS-St18-9	FN433425	99	LEE82	n.d.	<1
S18-847-G1 *	*Leeuwenhoekiella*	29.9	27	37.2	*Leeuwenhoekiella blandensis* MED217	96	VIS-St18-9	FN433425	99	LEE82	n.d.	<1
S18-1082-D10	NS4	32.9	28	31.9	*Kordia algicida* OT-1	90	VIS-St18-37	FN433422	99	-	-	-
S18-873-E7	DE2	45.8	34	33.9	Flavobacteria MS024-2A	96	VIS-St18-43	FN433427	100	F8737E-843	n.d.	13 ± 1.1
S18-920-B9	VIS4	46.3	40	31.8	Flavobacteriales ALC-1	91	VIS-St18-31	FN433420	99	FL197	73 ± 15	38 ± 6.2
S18-812-F11	VIS3	40.9	35	39.2	Flavobacteriales ALC-1	93	VIS-St18-59	FN433291	99	-	-	-
S18-848-B2	NS9	41.6	34	57.1	Flavobacteria BBFL7	86	VIS-St3-53	FN433379	95	NS9-664	7.3 ± 2.8	8.2 ± 0.2
S18-906-E12	AGG58 branch 3	43.4	19	37	*Kordia algicida* OT-1	86	VIS-St18-45	FN433411	96	AGG58-742	0.4 ± 0.08	5.4 ± 0.2
S18-870-C2	AGG58 branch 2	39.8	29	31.4	*Gramella forsetii* KT0803	86	VIS-St18-60	FN433292	99	-	-	-
S18-1049-A9 (M^4)	Sphingobacteria C	36.9	31	45.6	*Algoriphagus* sp. PR1	82	none	-	-	SPC-814	n.d.	<1
S18-858-A8 (M^4)	Sphingobacteria C	37.2	30 (3)	43.7	*Algoriphagus* sp. PR1	82	none	-	-	SPC-814	n.d.	<1
S18-1001-F9	*Cytophagia*	37.2	32	38.1	*Algoriphagus* sp. PR1	86	VIS-St18-50	FN433287	99	CYT-734 CYT-1462	17 ± 0.2	5.4 ± 0.6

M, merged fosmids in **M^1**, S3-825-979M; **M^2**, S3-858-892; **M^3**, S18-1087-1013; **M^4**, S18-1049-858; number of ORFs merged given in brackets. +, Malstrom et al., 2007. n.d. no detected

Phylogenetic affiliation

The majority of fosmids in our set originated from *Flavobacteria*, which is in line with the previous study. Marine *Bacteroidetes* belonging to the classes *Sphingobacteria* and *Cytophagia*, seem to be less abundant in marine plankton (Alonso et al., 2007) (Figure 1.4 and 1.5). Our set of genomic fragments contains only one from *Cytophagia* and three from *Sphingobacteria* (Table 1) and their analyses initiates a closer characterization of their potential ecological role in marine plankton. Almost all the *Flavobacteria* fosmids were phylogenetically affiliated with the major clades previously identified, such as VIS3, VIS4, VIS5, VIS6, DE2 (Kirchman et al., 2003), AGG58 (DeLong et al., 1993; O'Sullivan et al., 2004), and NS4 and NS9 (Alonso et al., 2007) (Table 1). Consequently, they reveal metabolic information about key players in the contrasting cold nutrient–rich BPLR and subtropical oligotrophic NAST.

Interestingly, the 16S rRNA gene of a fosmid retrieved from the BPLR (S3-816-G8) was related to that of a fosmid retrieved from Antarctic sea ice, Ant39E11 (Grzymski et al., 2006). The clade was denominated ANT and comprised both fosmids and other sequences retrieved from coastal and open ocean sites. Cells of the ANT clade were targeted with a FISH oligonucleotide probe and were detected in the BPLR (S2-S3) (2.5×10^3 cells ml^{-1}), but not in the NAST (Table 1). It indicates that ANT might inhabit water masses with low temperature. The ANT fosmids presented a surprisingly conserved gene content and order. Those fragments had an amino acid sequence similarity of 78% along a 28.6 kbp stretch (Figure 3.9 A), and 23 out of 28 ORFs in S3-816-G8 exhibited highest pairwise similarity (blastp) with predicted proteins from Ant39E11 (Grzymski et al., 2006). It suggests that the type of environment might have a strong influence in shaping the metabolic potential of this polar marine flavobacteria.

The comparison of metagenomic fragments with whole genomes from closely related isolated organisms opens the possibility to investigate their genetic variability. Of the 17 flavobacterial fosmids in our set, five belonged to the readily cultured genera *Dokdonia* (one from S18), *Leeuwenhoekiella*

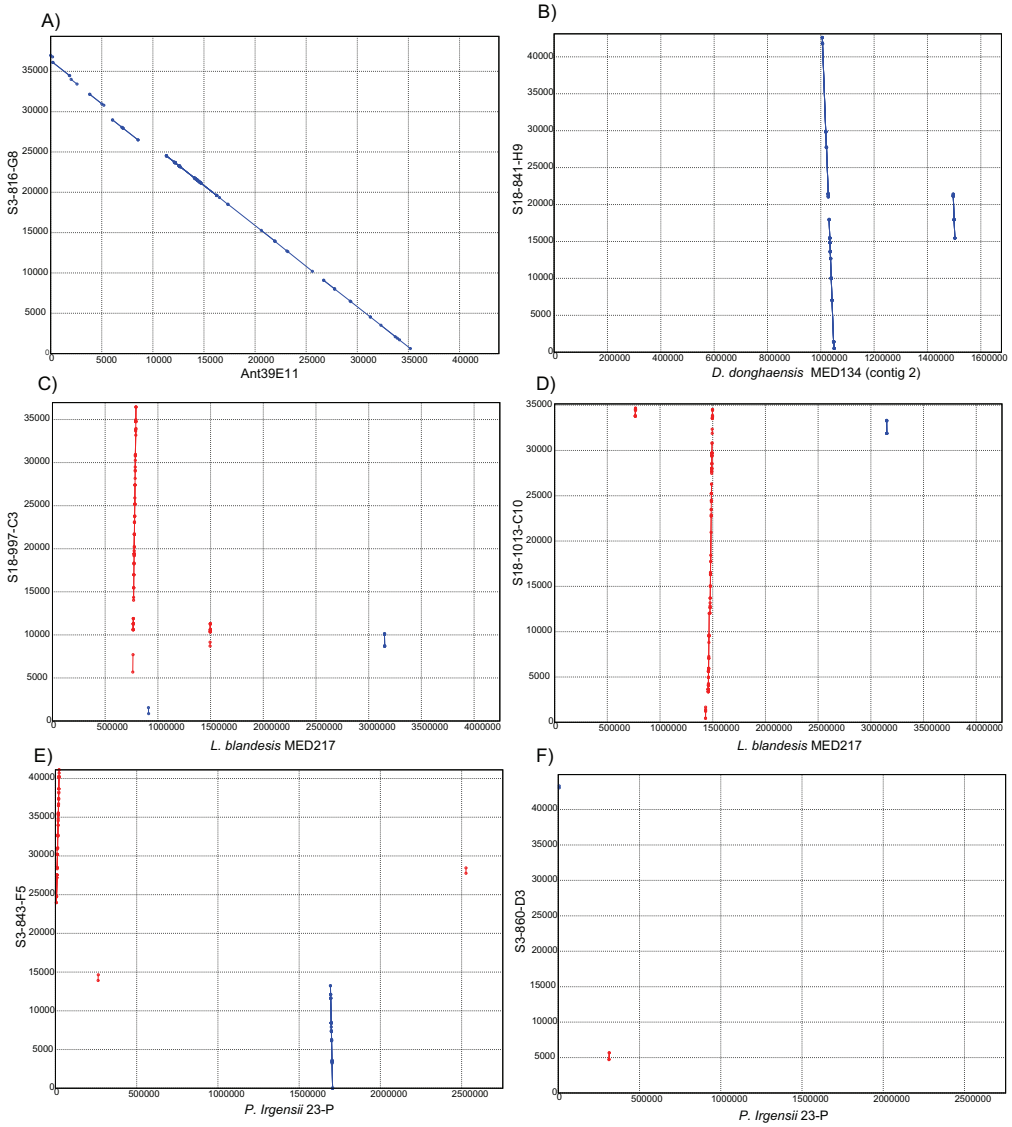


Figure 3.9: Pairwise amino acid sequence alignment with matches mapped back to the original DNA coordinate system of (A) Uncultured genomic fragments Ant39E11 and S18-816-G8; (B) *D. donghanensis* MED134 (contig 2) and fosmid S18-841-H9; *L. blandesis* MED152 with fosmids (C) S18-997-C3 and (D) S18-1013-C10; *P. irgensii* 23-P with fosmids (E) S3-845-F5 and (F) S3-860-D3. Fosmid S18-1013-C10 presented as example of the three *Leuwenhokiella* spp. highly similar fosmids (S18-1087-F8 and S18-847-G1). Aligned regions in the same direction plotted in red and reverse complement in blue.

(two from S18), and *Polaribacter* (two from S3) (Table 1). Their deduced amino acid sequences were found to be highly similar (74-94% similarity) to almost contiguous 40 kbp regions in the genomes of strains from these genera (Figure 3.9 B, C, D, E), with the exception of one *Polaribacter* fosmid (Figure 3.9 F). This suggests low genomic variability between closely related strains, at least in the analyzed region, even if they thrive in different marine environments. The strains *L. blandensis* MED217 and *D. donghaensis* MED134 have been isolated from the Blanes Bay in the Mediterranean Sea (Gomez-Consarnau et al., 2007), which is an inland sea with high organic material input and nutrient concentrations (Lucea et al., 2005) relative to the oligotrophic NAST, from where the fosmids were retrieved. The genome of *D. donghaensis* MED134 contains a proteorhodopsin gene, and its growth is enhanced by light at low concentrations of DOM (Gomez-Consarnau et al., 2007). It has been suggested that this might be a strategy to adapt to scenarios of organic matter depletion (Gomez-Consarnau et al., 2007; Gonzalez et al., 2008) as in the oligotrophic NAST. In our study, however, we did not identify genes related with a light driven metabolism.

Interestingly, one of the *Polaribacter* fosmids did not show such a high amino acid similarity with the *Polaribacter* spp. strains. Despite that both *Polaribacter* fosmids had a high 16S rRNA gene similarity with *P. dokdonensis* MED152 (95%) and *P. irgensii* 23-P (98%), fosmid S3-860-D3 did not share any region with either of the two genomes while fosmid S3-843-F5 had several regions in common with both genomes (Figure 3.9 E, F).

Degradation capabilities

We focused our analysis on genes involved in degradation of organic matter (OM) and in surface adhesion in order to address, for two contrasting marine scenarios, the hypothesized role of marine *Bacteroidetes* as particle–associated biopolymer–degraders. Genes involved in other metabolic processes are not discussed in the context of this study but are listed in the Appendix B.

It has been shown both on isolates and by culture-independent approaches

that aquatic *Bacteroidetes* are capable of degrading proteins, polysaccharides like chitin as well as amino–sugars, like N–acetylglucosamine (Cottrell and Kirchman, 2000). Furthermore, the analysis of genomes from marine *Bacteroidetes* has revealed a generally high number of genes encoding peptidases and glycoside hydrolases (Bauer et al., 2006; Gonzalez et al., 2008; Woyke et al., 2009). In our genomic fragments data set, we identified an overall broadly distributed potential for the degradation of proteins and the amino–sugar containing bacterial cell wall component peptidoglycan (see below), whereas predicted polysaccharide degradation activities were confined to three instances on a single fosmid.

Polysaccharide degradation

In the entire set of predicted proteins encoded by the 21 fosmid sequences, only three glycoside hydrolases (GH) were identified, all of them located on the *Polaribacter* fosmid S3-860-D3 (Figure 3.10). The three glycoside hydrolases belonged to the family 92, members of which have been shown to act as alpha–1,2–mannosidases (Cantarel et al., 2009). GH family 92 can hydrolyze the alpha 1,2–linked side chains of mannan as well as the N–linkage to high mannose content oligosaccharides in glycoproteins. The polysaccharide mannan is one of the components of plant cell walls and also of the cell walls of some green (Frei and Preston, 1968) and red macroalgae where it occurs as sulfated xylomannan (Erra-Basells et al., 2000). Besides the GH, the *Polaribacter* fosmid was found to encode another enzyme of the D–mannose metabolism, a mannose–6–phosphate isomerase, which converts fructose 6–phosphate and mannose–6–phosphate to each other. According to the prediction of certain cellular localization signals one of the GH (ORF0) was exported to the periplasmic space and the other two were secreted and located in the outer membrane (Figure 3.10 B). Extracellular enzymes are required to initiate the remineralization of high molecular weight OM (reviewed Arnosti, 2003). The genomic content of this *Polaribacter* organism indicates that it may be involved in the processing of the extracellular polysaccharide mannan.

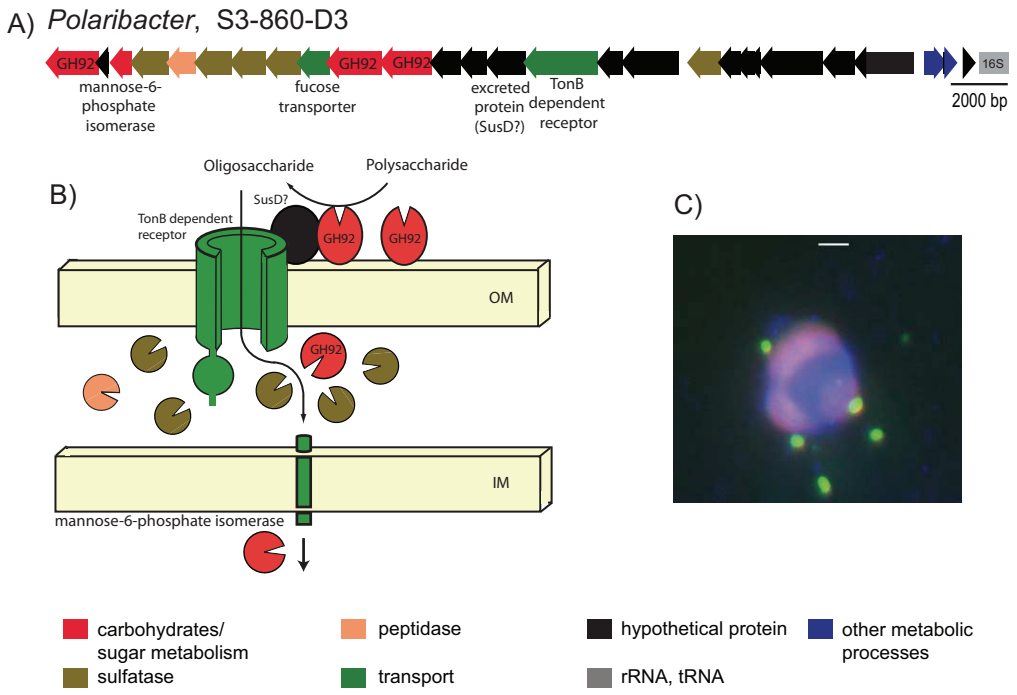


Figure 3.10: *Polaribacter* fosmid S3-860-D3 (A) Genomic content of the entire fosmid and (B) putative cellular localization, based on predicted signals, of proteins involved in polysaccharide degradation. Sketch adapted from McBride et al. (2009). ORF located downstream of the TonB dependent receptor is an outer-membrane protein that could be distantly related with SusD (C) Micrograph of *Polaribacter* cells, hybridized with probe POL740, localized in the phycosphere of a nanophytoplankton cell in S3 20m. Bar = 2 μm.

A comparative genome analysis of marine *Flavobacteria* draft genomes revealed that certain strains contain a higher number of glycoside hydrolase genes than others (Manuscript II, Supplementary Figure 3). This apparent variability within *Flavobacteria* might be further illustrated by the clustering observed here of three GH genes on one single genomic fragment originated from a *Polaribacter* organism (S3-860-D3).

A surprising accumulation of five sulfatase genes was found in the vicinity of the GH genes on the *Polaribacter* S3-860-D3 fosmid (Figure 3.10). This is a higher number on this short DNA–fragment than found in the entire genomes of the flavobacterial strains *P. irgensii* 23-P and *P. dokdo-*

nensis MED152 (three and two sulfatase genes, respectively). In general, marine planktonic *Bacteroidetes* are not particularly enriched in sulfatases (Manuscript II, Supplementary Figure 3), the ones with a higher number are Flavobacteriales bacterium HTCC2170 (6.5 per Mbp, unpublished) and the draft genome retrieved from a single cell by multiple displacement amplification, MS024-2A (3.1 per Mbp) (Woyke et al., 2009). For all of the fosmid-encoded sulfatases a signal-peptide was predicted, thus the proteins are probably exported from the cytoplasm. Moreover, one of the sulfatase genes was found next to a gene encoding an L-fucose transporter. Fucose is the fundamental subunit of fucoidan, a class of fucose-enriched sulfated polysaccharide found in the extracellular matrix of brown algae (Cumashi et al., 2007).

In the marine realm *Planctomycetes* are the bacteria with the highest number of sulfatase–encoding genes (Glockner et al., 2003; Woebken et al., 2007b). It has been hypothesized that they are used to access more effectively the carbon skeleton of sulfated polysaccharides as an energy source (Glockner et al., 2003). Given the genomic neighborhood of the sulfatases on the fosmid S3-860-D3, this might also be the case for the *Polaribacter*–related bacterium. Consistent with the genomic finding, the abundance of the *Polaribacter* genus was shown to be higher in the more productive BPLR and Arctic provinces (Figure 3.7), where the phytoplankton abundance was higher. Additionally *Polaribacter* cells have been observed localized in the phycosphere of nanophytoplankton cells (Figure 3.8 C and Figure 3.10 C). The *Polaribacter* fosmid presented here might indicate that certain marine *Bacteroidetes* lineages or strains are enriched in sulfatases, possibly as a specialization for the degradation of sulfated polysaccharides.

One of the *Bacteroidetes* distinctive features related with the degradation of polysaccharides is the presence in their genomes of several SusD–like outer membrane proteins (Shipman et al., 2000), encoded directly downstream of TonB–dependent outer membrane receptors (SusC–like proteins), and in the vicinity of GHs (Xu et al., 2003; Bauer et al., 2006; McBride et al., 2009) or peptidases (McBride et al., 2009). In *B. thetaiotaomicron*, which is an anaerobic human–associated *Bacteroidia* with an elaborate equipment for

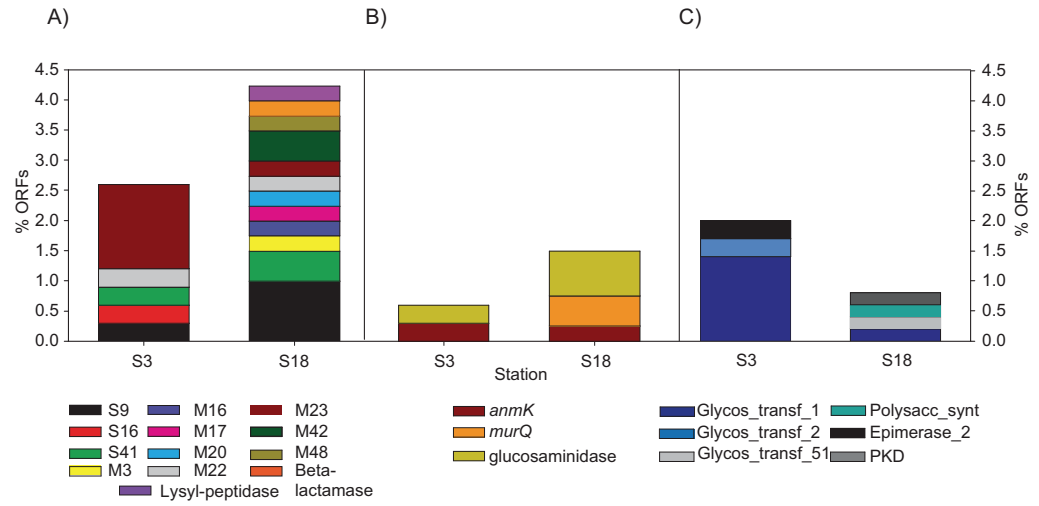


Figure 3.11: Comparison of *Bacteroidetes* metagenome between S3 and S18 of (A) peptidases, (B) peptidoglycan hydrolases and, (C) adhesion proteins expressed as the percentage total ORFs. Abbreviations in legends correspond to Pfam models in A, C and glucosaminidase in B (E value <E-10) or gene name in B. *anmK*: anhydro–N–acetylmuramic acid kinase; *murQ*: N–acetylmuramic acid 6–phosphate etherase.

acquiring and hydrolyzing otherwise indigestible dietary polysaccharides (Xu et al., 2003), SusC and SusD together with other outer membrane proteins and GH form a multicomponent outer membrane starch utilization system (Sus) (Shipman et al., 2000). The Sus system binds starch extracellularly before starting its degradation (Anderson and Salyers, 1989a,b; Cho and Salyers, 2001). In the *Bacteroidetes* fosmids sequenced here only one potential SusD–like protein was encoded in the vicinity of a TonB–dependent outer membrane receptor (on the *Leeuwenhoekiella*-affiliated fosmid S18-997-C3). In the marine *Bacteroidetes* genomes there is a clear increase in the SusCD–like pairs per Mbp with the increasing number of GH per Mbp (Manuscript II, Supplementary Figure 4a). This finding is in line with the evidence that SusCD–like proteins function in conjunction with GH for polysaccharides degradation. Moreover, no relation was found between the SusCD–like ratios and peptidases, which seem to be less often localized in the vicinity of SusCD–like (Bauer et al., 2006; McBride et al., 2009).

Protein degradation and recycling of peptidoglycan

In our data set, *Bacteroidetes* appear to have an extensive proteolytic potential (Figure 3.11 A). In contrast to the glycoside hydrolases, which were concentrated on one fosmid, peptidases were identified on almost all *Bacteroidetes* fosmids. A higher percentage of peptidases was identified on the *Bacteroidetes* fosmids from S18 compared to that of S3 (Figure 3.11 A). The diversity of peptidase families was also higher in S18 (12 different families vs. 5 in S3). Of the S18 peptidases, 30% exhibited either an export (SpI) or secretion (SpII) signal, whereas 50% of the peptidases from S3 showed an export signal (SpI). Exported and secreted peptidases may liberate low molecular weight peptides that are either transported directly into the cell or are hydrolyzed further to even smaller peptides or free amino acids (Cottrell et al., 2005). Dissolved proteins, peptides, and amino–sugars may be important sources of energy and nitrogen for marine bacteria (Benner and Kaiser, 2003; Obayashi and Suzuki, 2005) and particularly, marine *Bacteroidetes* seem to have an important role in the degradation of protein and N–acetyl–glucosamine (Pinhassi et al., 1999; Cottrell and Kirchman, 2000). The most frequent peptidase identified in the fosmids from S3 and generally in the genomes was a peptidase that lyses bacterial cell wall peptidoglycans (family M23) that is used by certain bacteria as a defensive or feeding mechanism (Rawlings et al., 2008) . Additionally, on the flavobacterial fosmids several peptidoglycan hydrolases were identified (Figure 3.11 B), particularly the DE2 and VIS4 fosmids were enriched in peptidases and peptidoglycan hydrolases (Figure 3.12). Peptidoglycan is degraded by lytic transglycosylases and endopeptidases (Holtje, 1998) whereby anhydro–N–acetyl muramic acid (anhMurNAc) is one of the degradation products. By the enzymes anhydro–N–acetylmuramic acid kinase (*anmK*) and N–acetylmuramic acid 6–phosphate etherase (*murQ*), anhMurNAc is then converted to N–acetyl–glucosamine 6–phosphate (GlcNAc 6-P) (Jaeger et al., 2005; Uehara et al., 2006), which can ultimately be shuttled into glycolysis. Both DE2 fosmids encode the M23 peptidase, a glycoprotein acetylglucosaminidase -probably located in the outer membrane- and a muramic acid kinase (*anmK*), the

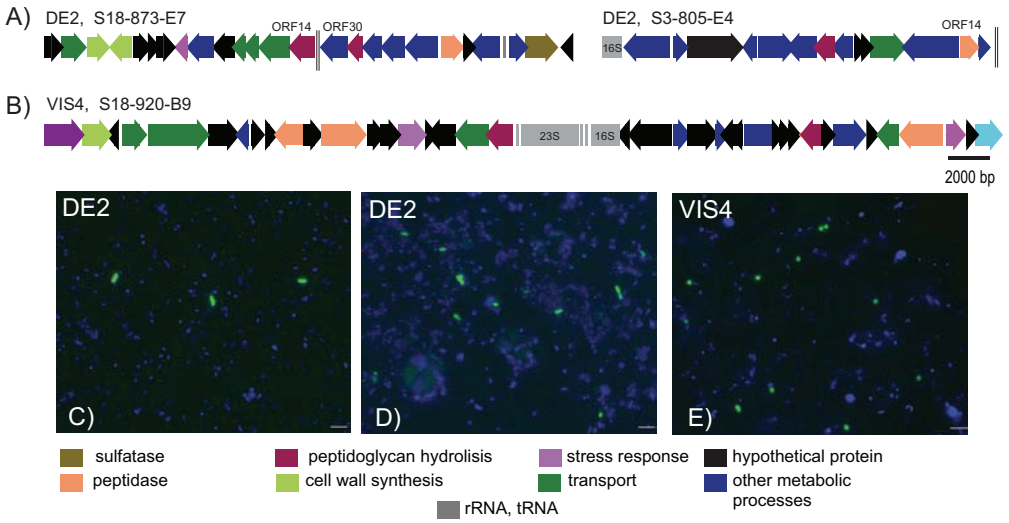


Figure 3.12: Genomic content of (A) DE2 (S18-873-E7, S3-805-E4) and (B) VIS4 fosmids (S18-920-B2). Double lines separating arrows indicate that ORFs are not contiguous. Micrograph of cells hybridized with specific oligonucleotide probes for the rRNA of fosmid (C) S18-873-E7, (D) S3-805-E4 and (E) S18-920-B2. Bar = 2 μm. Picture C and D taken by C. Bennke

latter two also on the VIS4 fosmid. The *murQ* gene was found on the VIS3 fosmid S18-841-H9 and on the *Dokdonia* fosmid, on the later also a a glycoprotein acetylglucosaminidase was found.

It has been shown that *E. coli* can grow in muramic acid, which is an amino–sugar that exclusively occurs in peptidoglycan, as its sole carbon source (Dahl et al., 2004). Muramic acid can be recycled from the own cell wall or obtained from the environment (Uehara et al., 2006). In marine systems, the picoplankton assemblage can also readily hydrolyze and remineralize both polysaccharides and peptide–components of peptidoglycan (Nagata et al., 2003). Structural components of microbial cells including membranes and peptidoglycan can be introduced to seawater as DOM during bacterial decay due to protozoan grazing and viral infection (Nagata and Kirchman, 1999). In our data set *Flavobacteria* are enriched in peptidases and peptidoglycan hydrolases, particularly in the NAST (Figure 3.11 A, B), which might reflect that under conditions of low phytoplankton derived polysaccharide concentrations they could thrive on bacterial derived

organic matter as carbon and nitrogen sources.

Surface adhesion and gliding

One of the prominent characteristics of marine *Bacteroidetes* is their association with phytoplankton. Strains (Barbeyron et al., 2001) and 16S rRNA sequences (Longford et al., 2007; Sapp et al., 2007) of members of *Bacteroidetes* have been retrieved from the surface of algae cells. Moreover, bacteroidetal cells have been observed localized on the algal phycosphere (Biegala et al., 2002) (Figure 3.8). One type of compounds that might help to establish or maintain this physical association are exopolysaccharides (EPS) (Nichols et al., 2005) which are involved in the mediation of cell/cell and cell/surface adhesion (Hall-Stoodley et al., 2004). On the *Bacteroidetes* fosmids we identified several proteins possibly involved in polysaccharide biosynthesis, which were in higher frequencies in S3 (Figure 3.11 C). In the BPLR and ARCT provinces there is a higher phytoplankton biomass in comparison with the NAST (Figure 3.4 C), therefore it is expected that more *Bacteroidetes* will tend to attach to phytoplankton cells.

Particularly, a *Sphingobacteria* fosmid, from a clade denominated Sphingobacteria A, had several predicted proteins related to surface adhesion (Figure 3.13 A). Glycosyl transferases (GT) participate in the biosynthesis of different kinds of polysaccharides, including EPS. Three GT of the family 1 were found on the Sphingobacteria A fosmid (ORF0, ORF1 and ORF5, Figure 3.13 A) that were in a cluster interspersed with genes encoding mannose–1–phosphate guanylyltransferase (ORF4), involved in the synthesis of the capsular polysaccharide colanic acid, and UDP–N–acetylglucosamine 2–epimerase (ORF2) involved in lipopolysaccharide biosynthesis. Additionally, cells of this clade were indeed observed in the phycosphere of nanophytoplankton cells (Figure 3.13 D). Moreover, the abundance of Sphingobacteria A was higher in S3 (9.5×10^2 cells ml^{-1}), where phytoplankton biomass was higher, and could not be detected in the oligotrophic NAST. Moreover, 16S rRNA gene sequences closely related to those of Sphingobacteria A have been retrieved from the surface of the red macroalgae *Delisea pulchra* and the green

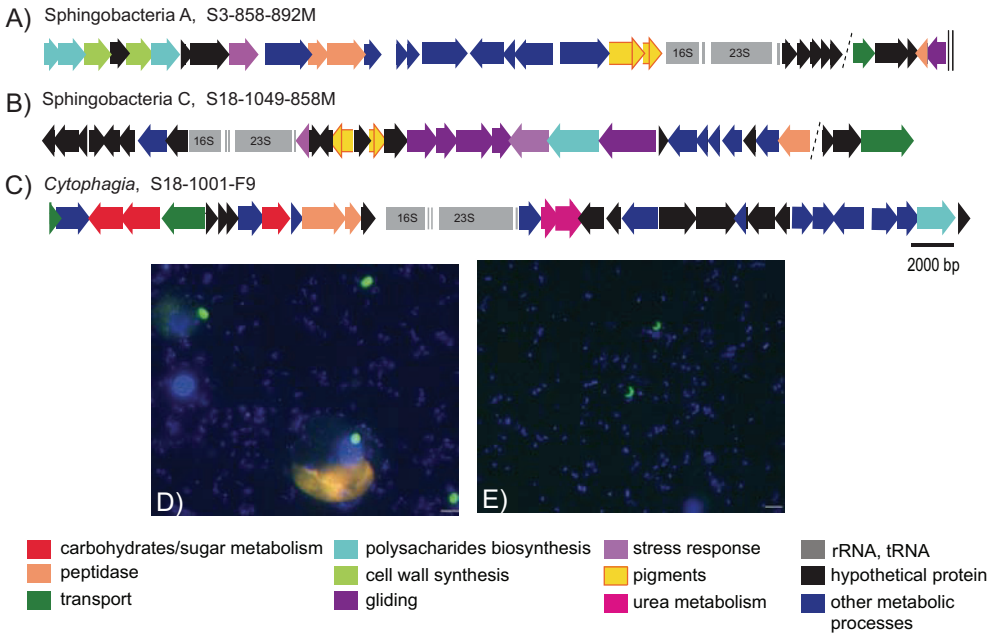


Figure 3.13: Genomic content of (A) Sphingobacteria A (S3-858-892M) (B) Sphingobacteria C (S18-1049-858M) and (C) *Cytophagia* (S18-1001-F9) fosmids. A dashed diagonal line indicates were genomic fragments have been merged. Micrograph of cells hybridized with specific oligonucleotide probes for the rRNA of fosmid (D) S3-858-892M and (E) S18-1001-F9. Bar = 2 μm.

Ulva australis (e.g. DQ269044, Longford et al., 2007). The sphingobacterial fosmids had peptidases of the family M22, which contains endopeptidases that cleave only proteins that are O–sialoglycosylated. A sialoglycoprotein is a combination of sialic acid and glycoprotein. Interestingly, the amino-sugar sialic acid has been identified in the exopolymers produced by *Phaeocystis* cells (Orellana et al., 2003). The Sphingobacteria A cells might thrive attached to phytoplankton cells and degrade some of the released compounds that contain sialic acid.

Another trait of *Bacteroidetes* related with the attachment of cells to surfaces is gliding motility (Hall-Stoodley et al., 2004). A total of 14 genes involved in gliding have been identified and characterized in the soil bacterium *Flavobacterium johnsoniae*, which can attach to and move along surfaces by gliding motility (McBride, 2004). In this study, four gliding genes (*gldK,*

gldM, *gldO* and *gldL*) were detected in an *Sphingobacteria* fosmid -from a clade denominated Sphingobacteria C- (Figure 3.13 B) and one each (*gldJ*) on the flavobacterial fosmids ANT and VIS6. In *F. johnsoniae* disruption in some of the gld genes, including *gldK* and *gdL*, results in complete deficiency in chitin utilization (Braun and McBride, 2005), but the connection between motility and chitin digestion is not clearly understood (Xie et al., 2007). Based on the genomic information, we propose that the marine planktonic *Sphingobacteria* might be able to attach to surfaces and to degrade carbohydrates, sugars or proteins released by the phytoplankton.

Overall, we identified a great proteolytic potential in marine *Bacteroidetes* corroborating previous genomic (Bauer et al., 2006) and experimental (Cottrell and Kirchman, 2000) findings. Glycolytic potential was confined to a member of the genus *Polaribacter*, which revealed a whole armory of proteins that might function in degradation of sulfated polysaccharides. Contrastingly, peptidases were found in several fosmids. The preliminary comparison between both provinces revealed a greater proteolytic potential in the oligotrophic NAST whereas a higher potential for adhesion in the phytoplankton richer BPLR. However, in order to perform a robust *Bacteroidetes* functional comparison, a more extensive sequencing effort is needed (DeLong et al., 2006; Rusch et al., 2007).

Increasing the *Bacteroidetes* sequencing coverage

The fosmids sequenced so far yielded approximately 0.5 Mbp of DNA sequence in each library, which is equivalent to 1/6 of a genome (3 Mbp sized). Despite that the fosmids sequenced were representative of the major *Bacteroidetes* clades, this low sequencing coverage of their genomes only allow us to perform preliminary comparisons. Additionally, the fosmids analyzed so far were selected by a PCR–based screening with *Bacteroidetes* specific primers targeting the 16S rRNA gene. The rRNA is a highly conserved molecule (Ludwig et al., 1998) part of a core of informational genes which are only weakly affected by horizontal gene transfer (Daubin et al., 2003). Therefore the genes identified in the neighborhood often are housekeeping

genes involved in cell maintenance.

An alternative approach for screening metagenomics libraries is the sequencing of the metagenomic clones ends (end-sequencing) what provides approximately 500-800 bp from each clone end. This approach can provide first insights into the metabolic capabilities and the taxonomic groups available in the metagenomic library (Glockner and Meyerdierks, 2006; Feingersch et al., 2010). In the present study a total of 15,615 from S3 and 16,225 from S18 end–sequences with an average length 600 bp were generated. This data set represents 10 Mbp in each station, which is equivalent to approximately 3 bacterial genomes (average 3 Mbp size). The aim was to taxonomically assign the fragments in order to identify clones which carried a *Bacteroidetes* insert. The end-sequences were analyzed by S. Huang (Huang, 2009) with a bioinformatic pipeline that allowed the taxonomic assignment of short DNA sequences. Three tools (DarkHorse, Kirsten, and CARMA) were applied that evaluated the end sequences coding regions at the protein level using external databases as references. The tools DarkHorse (Podell and Gaasterland, 2007) and Kirsten (Huang, 2009) are based on the occurrence of taxonomic terms in BLAST hits; the latter also considers quality values to ensure the prediction accuracy (Huang, 2009). The tool CARMA (Krause et al., 2008) is based on an alignment to Pfam protein families and a phylogenetic analysis based on the multiple alignments. Those analyses yielded a total of 120 in S3 and 91 in S18 fosmids end–sequence pairs assigned to the phylum *Bacteroidetes*. The 211 bacteroidetal fosmids were manually isolated from their original microtiter storage plates and a subset was validated by sequencing their ends. Those fosmids are currently in the sequencing pipeline of the sequencing center "Genoscope" (Evry, France).

Assuming an insert size of 35 kbp and an average genome size of 3 Mbp, the sequencing of this fosmids will result in an approximate equivalent of one *Bacteroidetes* genome from each station. Similarity searches of the fosmids ORFs will be performed against sequence databases (NCBI nr, NCBI nt, SwissProt) and protein family databases (Pfam, InterPro, COG). The results will be retrieved from the databases using the framework Silmaril, which is a C++ application programming interface (H. Teeling, unpublished). Searches

of genes involved in degradation, (e.g. glycoside hydrolases, peptidases, peptidoglycan hydrolases), attachment, and light metabolism will be performed. A preliminary analysis on the end–sequences has already been performed using the framework described above. However, only three *Bacteroidetes* glycoside hydrolases were found (one in S3, two in S18) while no SusD–like protein was found in any of the libraries, further supporting the need of a higher sequencing coverage in order to compare the metabolic capabilities of the *Bacteroidetes* community.

In order to further classify the *Bacteroidetes* fosmids, phylogenetic markers will be searched in the dataset. Alternative markers to the rRNA gene are, for example, the elongation factor EF–Tu, the catalytic subunit of the proton translocating ATPase, or the hsp60 heat shock protein (Ludwig and Klenk, 2005). Additionally, recruitment of fragments based on BLAST-N searches (Rusch et al., 2007) will be performed with the fosmids encoding the 16S rRNA gene (Table 1) and *Bacteroidetes* whole genome sequences as references.

With this data at hand we aim to perform a robust comparison between the degradation capabilities of *Bacteroidetes* from the nutrient richer BPLR and from the oligotrophic NAST.

3.5 *Bacteroidetes* in coastal marine sites

The third section of this thesis summarizes two studies in coastal areas. The bacterial community of the Lagoon of Venice and of the Río de la Plata estuary were investigated by a suite of molecular methods. The objectives were to investigate if the *Bacteroidetes* clades present in open ocean (Section 3.3 and Figure 3.6) also inhabit coastal areas and how their distribution patterns vary along a river–ocean gradient. In the following section, results will be presented concerning (i) diversity, (ii) abundance and (iii) distribution patterns of *Bacteroidetes* clades in both environments.

Characterization of study sites

The Venice lagoon is a shallow water basin located at the coast of the northern Adriatic Sea connected with it by three inlets: Lido, Malaocco and Chioggia. The Lagoon of Venice receives high organic pollution mainly caused by anthropogenic activities (Losso and Volpi Ghirardinia, 2010). The current situation of the lagoon is worrisome in view of its apparent degradation, with signs of anoxia in bottom layers, a decline of zooplankton, and high concentrations of sulfide in the sediments (Sorokin et al., 1996; Facca et al., 2002). However, only recent studies have focused on the seasonal and spatial dynamic of the microbial community. It has been highlighted that there is a strong seasonality of the picoplankton assemblage (Celussi et al., 2009) and also spatial differences between sub–basins (Borin et al., 2009), which might reflect the influence of different pollutants. *Gammaproteobacteria* of the *Vibrio* genus, which presence is related to urban polluted marine areas (Garay et al., 1985), dominated the microbial assemblage in sediment samples close to urban centers (Borin et al., 2009).

The Río de la Plata estuary (Figure 3.14), situated between Argentina and Uruguay, is characterized by a salt–wedge regime formed by the mixing of the Paraná and Uruguay rivers with the Atlantic Ocean (Giordano and Lasta, 2004a,b). Salt–wedge estuaries are formed where a large river flows into an area with a small tidal range. The less dense freshwater floats on the denser sea water; as it moves seaward it spreads out forming a thinner layer

Figure 3.14: Aerial picture of the Río de la Plata Estuary showing the flow of the Río Uruguay and Río Paraná. Capital cities of Uruguay (Montevideo) and of Argentina (Buenos Aires) are shown. Image obtained from www.nasaimages.org.

and the sea water beneath assumes the shape of wedge (Little, 2000). As the Paraná and Uruguay rivers encounter the Atlantic Ocean they generate a vast area with intermediate salinities that is highly variable depending on the rivers flow (frontal zone) (Giordano and Lasta, 2004b). When the river flow moves seawards, riverine sediments are transported to the sea as a well–defined plume generating areas with high turbidity (Giordano and Lasta, 2004a) (Figure 3.14). The position of the turbidity maximum is highly variable according to tide, river discharge, and wind (Framinan and Brown, 1996). The Río de la Plata entails a highly productive area that sustains valuable fisheries of Uruguay, Argentina, and international fleets (Bisbal, 1995; Martinez and Retta, 2001; Acha et al., 2008). The Río de la Plata also constitutes the main point source of freshwater in the South Atlantic (Boltovskoy et al., 1999). Despite its key significance, there is still very limited understanding about basic ecological aspects of the Río de la Plata planktonic communities (Calliari et al., 2009).

Overall, coastal lagoons and estuaries are highly productive systems with strong gradients, therefore it is compelling to investigate how the picoplankton response to the environmental changes. Particularly, as coastal systems receive high inputs of organic matter from its basins and sediments it is of

key relevance to understand the fate of *Bacteroidetes* along those gradients.

In the Río de la Plata samples were taken during late austral summer (9 of March 2008) during the cruise ARTEMISA1 onboard of the R.V. Aldebaran. A transect was followed along the salinity gradient occurring in the outer region of the Río de la Plata (Figure 3.15 A). Three samples were chosen as representatives of the different environments for construction of the 16S rRNA clone libraries (S3, riverine; S5, frontal zone; S9, marine). An additional library was constructed from the marine station S9, from a sample taken in austral autumn 2009 (April). In the Lagoon of Venice surface water samples were taken near the city of Chioggia during autumn (October 17, 2007) and summer (June 10, 2008) in two sites, within the lagoon and in the Southern Chioggia inlet (Figure 3.15 B).

Bacteroidetes diversity and distribution in the Lagoon of Venice and Río de la Plata

Along the Río de la Plata transect, *Bacteroidetes* was the second most abundant group, they constituted on average 18% of all DAPI stained cells (Figure 3.16 A). *Bacteroidetes* sharply increased in their abundance between the riverine surface water and the frontal zone, reaching their highest numbers at S4 and S5, located at the very centre of the frontal zone. The frontal zone was characterized by peaks in organic matter (2.4 mg l^{-1}), chlorophyll a (6.4 μg l^{-1}), and general availability of nutrients (Figure 1 in Manuscript III), indicating that members of this group would be particularly favoured by intermediate salinity and high productivity. High abundances of *Bacteroidetes* at the turbidity maximum of estuarine systems have been previously reported (Bouvier and del Giorgio, 2002). Members of this group have been repeatedly reported as prone to particle attachment (DeLong et al., 1993; Crump et al., 1999; Rink et al., 2008), and as primary colonizers of marine phytoplankton detritus (Pinhassi et al., 2004; Abell and Bowman, 2005a).

In the Lagoon of Venice, *Bacteroidetes* constituted the most abundant phylum in October 2007 at both sites (55 $\pm$ 1% at the Lagoon and 44 $\pm$ 0.8% at the inlet, % of DAPI stained cells), while in June 2008 it was the second

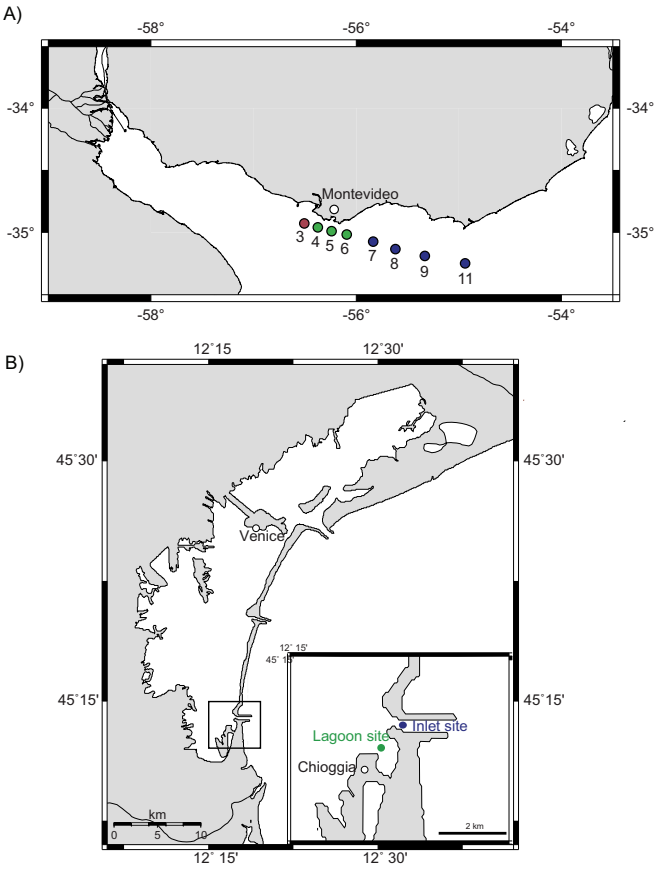


Figure 3.15: (A) Transect of the ARTEMISA1 cruise in the Río de la Plata indicating sampling stations color coded accordingly to riverine (red), frontal zone (green), and marine stations (blue). (B) Map of the Lagoon of Venice indicating the two sampling sites, inlet and lagoon. Maps were created with the online map creation tool, OMC (www.aquarius.ifm-geomar.de).

most abundant group after *Alphaproteobacteria* (34% at both sites) (Figure 3.16 B). In the Lagoon of Venice, members of *Bacteroidetes* were almost twice more abundant than in the Río de la Plata, and generally higher than in other coastal settings (e.g. Eilers et al., 2001; Alonso-Saez et al., 2007; Rink et al., 2008). Interestingly, the maximum *Bacteroidetes* abundance did not correspond with the peak in chlorophyll a concentration (lagoon site in June 2008). Contrarily, the highest *Bacteroidetes* abundance was detected in the Lagoon 2007 sample when the chlorophyll a concentration was rather

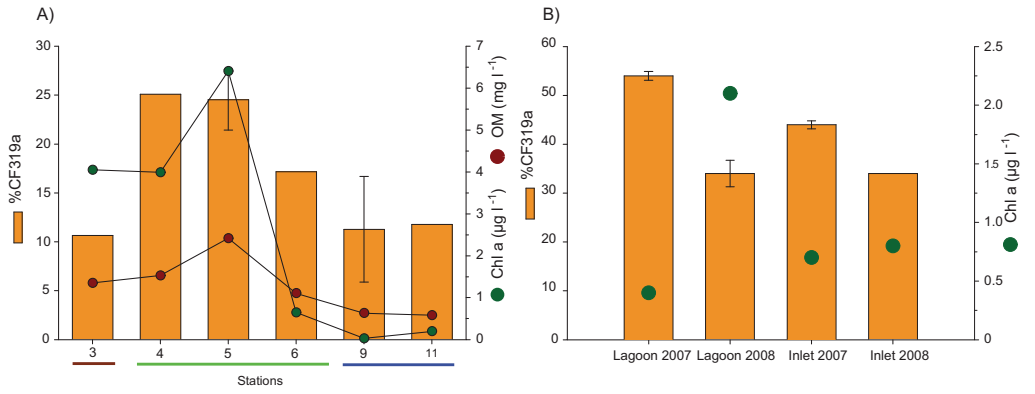


Figure 3.16: Distribution of *Bacteroidetes*as detected by probe CF319a (relative abundance, % of DAPI counts) (A) along the transect in the Río de la Plata estuary and (B) in the lagoon and inlet sites of the Lagoon of Venice. Chlorophyll a (Chl a) and organic matter concentration (OM) (only in A) are also plotted. Colour code of the Río de la Plata stations as in Figure 3.15. Note the different scale

low (Figure 3.16 B). High *Bacteroidetes* abundance and clone frequencies are often related to high values of chlorophyll a and nutrients (Simon et al., 1999; Riemann et al., 2000; Abell and Bowman, 2005b).

Bacteroidetes diversity

Our data from two coastal sites revealed that the vast majority of sequences were phylogenetically affiliated with the class *Flavobacteria* (94% in the Venice Lagoon, n=226; 88% in the Río de la Plata, n=34) (Figure 3.17) giving more evidence that *Flavobacteria* is the dominant class of *Bacteroidetes* in marine picoplankton (e.g. Kirchman et al., 2003; Alonso et al., 2007; Chen et al., 2008). Other sequences were afiliated with the class *Sphingobacteria* and none was with *Cytophagia*. Most of the *Bacteroidetes* clades, with a sequence identity between 90-96%, retrieved from the Lagoon of Venice encompassed sequences from both the lagoon and inlet site, suggesting that the bacteroidetal community was rather similar between sites. However, considering a 97% sequence identity threshold, exclusive phylotypes were identified in the lagoon and inlet sites. At the same identity level, almost all the *Bac-*

teroidetes sequences retrieved in the Río de la Plata were characteristic of the different samples; there was only one phylotype shared between the riverine and the marine libraries (RdeP-St3-275 and RdeP-St9-181) which belonged to the clade NS4. At the 95 % identity threshold, clade NS9 comprised sequences retrieved from two sites: marine and frontal zone.

In the Lagoon of Venice a relative high number of the *Bacteroidetes* sequences were affiliated with cultured genera: 20% with the strain "Aureimarina marisflavi" (EF108215; Yang and Cho, unpublished) and 10% with the genus *Polaribacter* (Figure 3.17). Thus, it provides further evidence that members of the genus *Polaribacter* are a significant constituent of the bacteroidetal community in both polar (Brinkmeyer et al., 2003; Malmstrom et al., 2007) and temperate marine environments, suggesting a widespread distribution of the *Polaribacter* genus in the marine realm (Section 3.6).

Most of the sequences retrieved from the Venice Lagoon and all from the Río de la Plata were affiliated with clades for which so far no cultured representatives have been obtained. Several sequences were affiliated with the VIS clades (VIS2, VIS3, VIS4, VIS5, VIS6), for example 17 Venice Lagoon and three Rio de la Plata sequences were affiliated with the clade VIS3. One of the major VIS clades present in the Lagoon of Venice was VIS6. A fraction of 10% of the *Bacteroidetes* sequences belonged to this clade (14 sequences from the inlet and 10 from the lagoon). VIS6 comprises phylotypes that have been retrieved from the North Atlantic Ocean and during a phytoplankton bloom in naturally iron fertilized waters (West et al., 2008). Only two of these VIS6 sequences were from the Lagoon 2008 sample, when chlorophyll a concentration was the highest (Figure 3.16 B).

Sequences from both study sites were affiliated with clades that have been defined based on sequences retrieved from the North Sea (NS clades): NS6, NS4 and NS9. Particularly, a 10% of the Venice Lagoon sequences (from both sampling sites) were affiliated with the North Sea clade NS6; within the clade a distinct lagoon phylotype could be distinguished. Twenty four percent of the Venice Lagoon 16S rRNA sequences (40 sequences from the inlet, 13 sequences from the lagoon) could be affiliated with the DE2 cluster (Kirchman et al., 2003). The DE2 comprises sequences that have mostly been retrieved

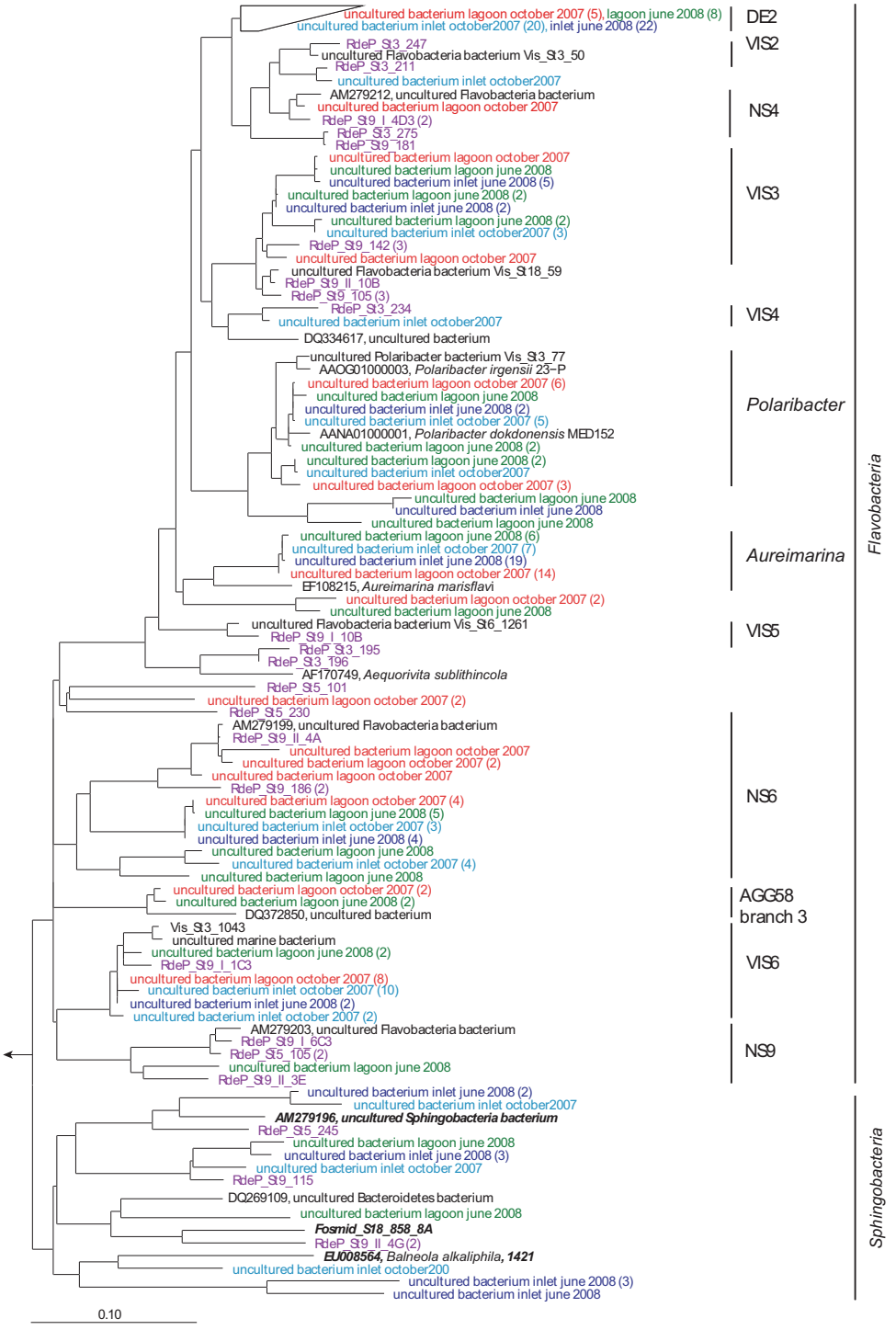


Figure 3.17: Phylogenetic association of 16S rRNA gene sequences retrieved in the Río de la Plata and in the Lagoon of Venice. Red: lagoon sequences retrieved in October 2007; turquoise: inlet sequences of October 2007; green: lagoon sequences of June 2008; blue: inlet sequences of June 2008. Purple: Río de la Plata sequences. Tree was reconstructed with full length sequences with the maximum likelihood algorithm RaxML; partial sequences were added to the existing tree without allowing changes in the overall tree topology.

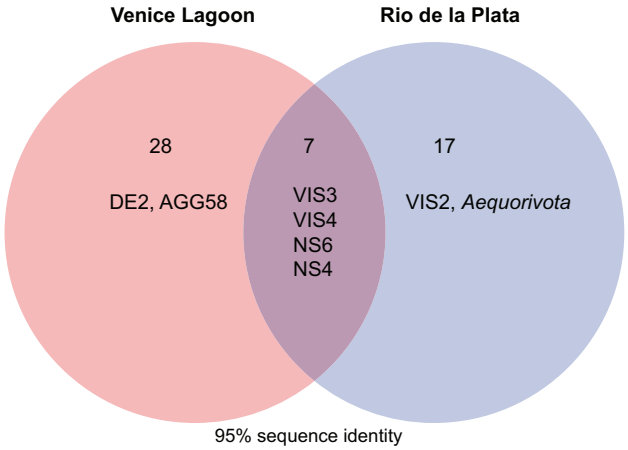


Figure 3.18: Shared and site–specific phylotypes between the Río de la Plata and Lagoon of Venice at 95% sequence identity level.

from coastal systems as the Delaware Estuary (Kirchman et al., 2003), the North Sea (Alonso et al., 2007) and the coast of Plymouth (O'Sullivan et al., 2004). Its high clone frequency in the Venice Lagoon again corroborates the relevance of this uncultured clade in coastal systems. The clade AGG58 has also been considered as a major coastal flavobacterial clade (O'Sullivan et al., 2004), but surprisingly only three sequences from the Venice Lagoon (lagoon site) were affiliated with this clade, and none from the Rio de la Plata.

A comparison between the Lagoon of Venice and the Río de la Plata 16S rRNA gene libraries revealed that, at a 95% sequence identity, seven phylotypes were shared between the libraries (Figure 3.18). The shared phylotypes were affiliated with the flavobacterial clades VIS3, VIS4, NS6, and NS4; and with a sphingobacterial clade. However, if a higher threshold is considered (unique) no sequence was shared between both study sites. It might point to an existence of a high microdiversity within the major *Bacteroidetes* clades as has been previously reported for the picoplankton in other coastal systems (Acinas et al., 2004; Alonso et al., 2007).

Bacteroidetes clade distribution in the Río de la Plata

In the Río de la Plata the abundance of three *Bacteroidetes* clades was evaluated with specific FISH oligonucleotide probes. The abundance of members of the *Polaribacter* genus (probe POL740), and of the uncultured clades DE2 (probe CF6-1267) and VIS1 (probe VIS1-575) were quantified along the transect (Figure 3.19). *Polaribacter*, VIS1 and DE2 increased in abundance in the frontal zone (5.4%, 2.4% and 0.8% on average respectively) which was the site with higher chlorophyll a concentration, providing further hints of a potential link between these flavobacterial clades and the phytoplankton community. Members of these groups have been observed to be especially abundant in the phycosphere of marine phytoplankton (Figure 3.8). In the frontal zone the abundance of VIS1 flavobacteria tended to increase with depth, which might point to a possible input of these bacteria from sediments resuspension. Among the probe–targeted groups, the genus *Polaribacter* was the most abundant, with a maximum of 9% in S6. This finding in the Río de la Plata estuary is in line with the results from the Venice Lagoon, based on sequence retrieval, indicating a widespread distribution of *Polaribacter* in temperate marine environments (Section 4). In the riverine stations the percentage of hybridized cells with the clade specific probes was the lowest what might indicate the presence of freshwater clades as the LD2 (Pernthaler et al., 2004) the CL500-6 (Zwart et al., 2002) and cfIII (Flectobacillus) lineages (Glockner et al., 2000). Clearly, additional oligonucleotides probes are needed to characterize the *Bacteroidetes* community in the Río de la Plata.

Overall, with both studies we confirm the importance of members of the phylum *Bacteroidetes*, and particularly of the class *Flavobacteria*, in marine coastal waters as has been shown by others (e.g. Eilers et al., 2001; Kirchman, 2002). Despite the frequently reported link of *Bacteroidetes* with algae blooms, (e.g. Pinhassi et al., 2004) in the Lagoon of Venice the peak of chlorophyll a concentration and of *Bacteroietes* abundance did not coincide. Contrastingly, in the Río de la Plata the *Bacteroidetes* abundance

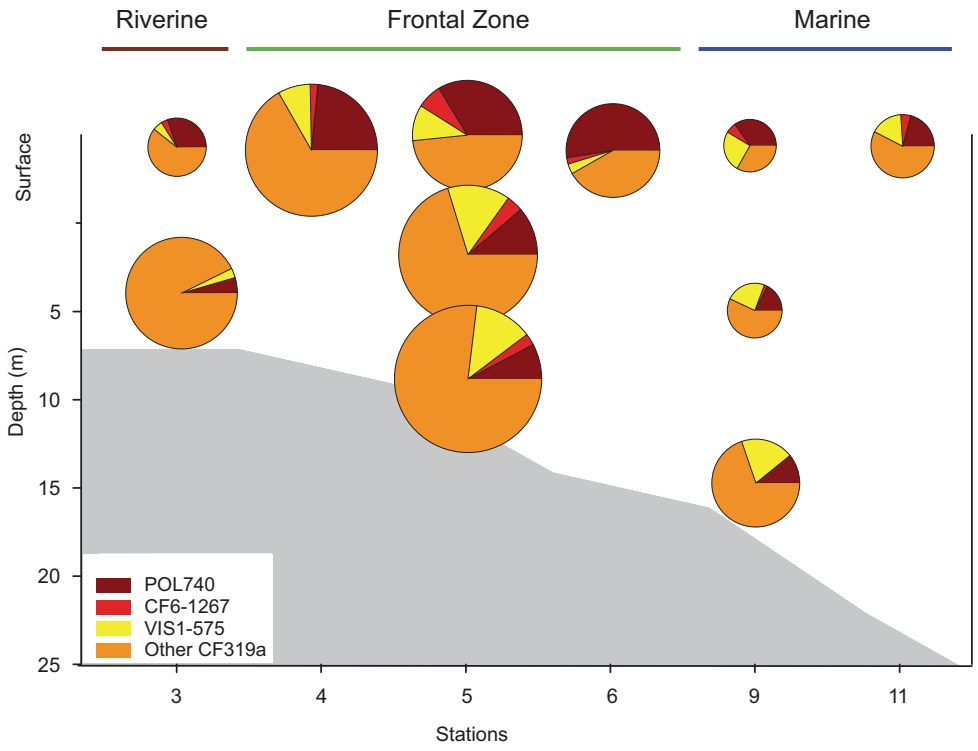


Figure 3.19: *Bacteroidetes* community composition in the Río de la Plata as analyzed by FISH with clade specific–probes. Pie charts represent the relative contribution of the clades *Polaribacter* (POL740), VIS1 (VIS1-575), and DE2 (CF6-1267) to the total *Bacteroidetes* (CF319a). Area of the pie chart represents the relative abundance of CF319a.

was highest in the frontal zone, where chlorophyll a had its maximum. The preliminary sequence analysis suggested that there are differences in the *Bacteroidetes* community composition of the riverine, the frontal zone and the marine areas, reflecting the pattern of the total microbial community (Figure 3 Manuscript III). Further analyses should increase the sequence coverage, which would serve as the basis for the subsequent design of oligonucleotide probes in order to fully characterize the bacteroidetal community.

3.6 Comparison between open ocean and coastal *Bacteroidetes*

Major picoplankton groups are found in both coastal environments and ocean gyres (Giovannoni and Stingl, 2005). However, the abundance of clades within the major groups varies along the gradient from continental margins to open ocean (Alonso-Saez et al., 2007). Examples include clades within the phylum *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria*. The alphaproteobacterial *Roseobacter* clade and the gammaproteobacterial NOR/OM60 are more abundant towards coastal areas in comparison to the open ocean (Alonso-Saez et al., 2007; Yan et al., 2009). Moreover, a phylotype within the *Betaproteobacteria* appears to be confined to coastal areas (Rappe et al., 1997). *Bacteroidetes* are significantly abundant both in coastal seas and in the open ocean, however, a detailed comparison between its diversity and clade abundances in distinct sites has not yet been performed.

A comparative analysis of the 16S rRNA gene clone libraries from the North Atlantic Ocean, the Lagoon of Venice, the Río de la Plata and the North Sea (Alonso et al., 2007) revealed that, at the 95% sequence identity, half of the phylotypes from the North Atlantic Ocean were also present in the coastal sites (Figure 3.20 A). The shared phylotypes were mostly the clades that have been discussed in the different sections of this thesis: DE2, VIS and NS clades, suggesting that these clades might be widespread in different geographic locations with distinct environmental characteristics. Moreover, the *Bacteroidetes* clades evaluated so far by FISH–oligonucleotide probes in a coastal site (Río de la Plata) and in the open ocean (North Atlantic Ocean) were detected in a rather similar percentage range. *Polaribacter* and VIS1 were in both environments between 1-9% and 0.5-4% respectively, while DE2 was slightly higher in the Rio de la Plata (0.5-1.5%) than in the North Atlantic Ocean (0.4-0.8%). Interestingly, despite of the rather similar range in the clades relative abundance between coastal and open ocean sites, their distribution patterns within each habitat varied locally (Figure 3.7 and 3.19) probably in response to environmental gradients.

Major *Bacteroidetes* clades were present in the coastal and open ocean li-

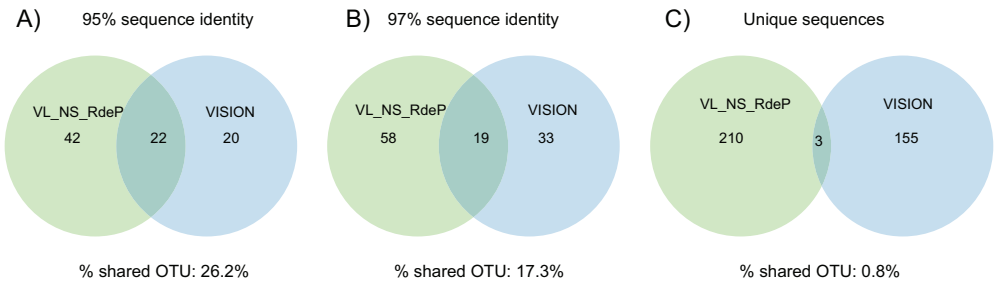


Figure 3.20: Shared and site–specific phylotypes between coastal (Venice Lagoon, VL; North Sea, NS;. Rio de la Plata, RdeP) and open ocean (VISION, North Atlantic Ocean) at the (A) 95% , (B) 97% and (C) unique 16S rRNA gene sequence identity thresholds.

braries, however less than 1% of unique phylotypes were shared between those libraries (Figure 3.20 C). The low similarity between sites at a higher resolution suggests that microdiversity might exist within closely related clades (Acinas et al., 2004; Woebken et al., 2008). At the species level the 16S rRNA gene lacks resolution power (Rossello-Mora and Amann, 2001). For example, there are different species with identical 16S rRNA gene (Probst et al., 1998) or microheterogenity of this gene within single species (Ibrahim et al., 1997). Consequently, analyses of the 23S rRNA gene sequence or of the 16S–23S rRNA internal spacer (ITS) are needed to further characterize habitat–specific *Bacteroidetes* clades. The ITS has an evolutionary rate 10 times faster than that of the 16S rRNA gene (LeblondBourget et al., 1996), and therefore can give insights into diversity with a higher resolution (Woebken et al., 2008).

Acinas and co–workers (2004) suggested that microdiverse ribotype clusters are important units of differentiation in natural bacterial communities. In order to determine whether microdiverse ribotype clusters represent ecotypes, a comparison of their genomic variation is required (Acinas et al., 2004). Genomic comparisons have shown that even almost indistinguishable strains, with an average nucleotide identity of their shared genes higher than 99%, can constitute distinct ecotypes (Konstantinidis and Tiedje, 2005). A clear case of ecotypes in marine systems is that of the unicellular cyanobac-

terium *Prochlorococcus*. In surface oceans, there are co–occurring *Prochlorococcus* cells with highly similar ribomosmal RNA sequences (>97%) that have different sensitivities to trace metals (Mann et al., 2002), nitrogen usage abilities and, light optima (Moore et al., 2002). The genome sequence of low and high light adapted strains are significantly different in size and gene content, possibly reflecting the adaptation of the ecotypes to key environmental variables (Rocap et al., 2003).

Within the *Polaribacter* genus there might be similar ecotypes. In the metagenomics libraries constructed from the BPLR in the North Atlantic Ocean (Section 3.4), two *Polaribacter*–like fosmids with a high 16S rRNA gene identity (99%) were retrieved. These *Polaribacter* flavobacterium seem to represent two distinct ecotypes. Despite that both *Polaribacter* fosmids have a high 16S rRNA gene identity with *P. irgensii* 23-P (98%), fosmid S3-860-D3 did not share any region in its amino acid sequence with the genome of *P. irgensii* 23-P (Figure 3.9 F). Interestingly, the *Polaribacter* fosmid S3-860-D3 revealed a whole armory of proteins that might function in degradation of sulfated polysaccharides. On this short genomic fragment a higher number of sulfatases were found than in the whole genome sequence of other flavobacterial strains.

Moreover, within the *Polaribacter* genus there are distinct closely related temperate and polar clades (Figure 3.21). All the polaribacter–like sequences retrieved in the Venice Lagoon libraries comprised a separated clade with *P. dokdonensis* strains (97-98% identity). These strains have been isolated from temperate marine environments and have an optimal growth temperature between 25°C–28°C (Yoon et al., 2006). All other sequences within this clade had been retrieved from temperate coastal environments and none of them was from polar origin. Contrarily, all the 16S rRNA gene sequences from the BPRL province of the North Atlantic Ocean clustered in two clades with higher similarities to several *P. irgensii* strains (98%-99%) and other sequences retrieved from low temperature environments (e.g. Southern Ocean, Antarctic ice). The strains *P. irgensii*, *P. filamentus*, *P.franzmannii* have been isolated from Antarctic and Arctic sea water or ice (Gosink et al., 1998) and comprise, together with other sequences retrieved from low temperature

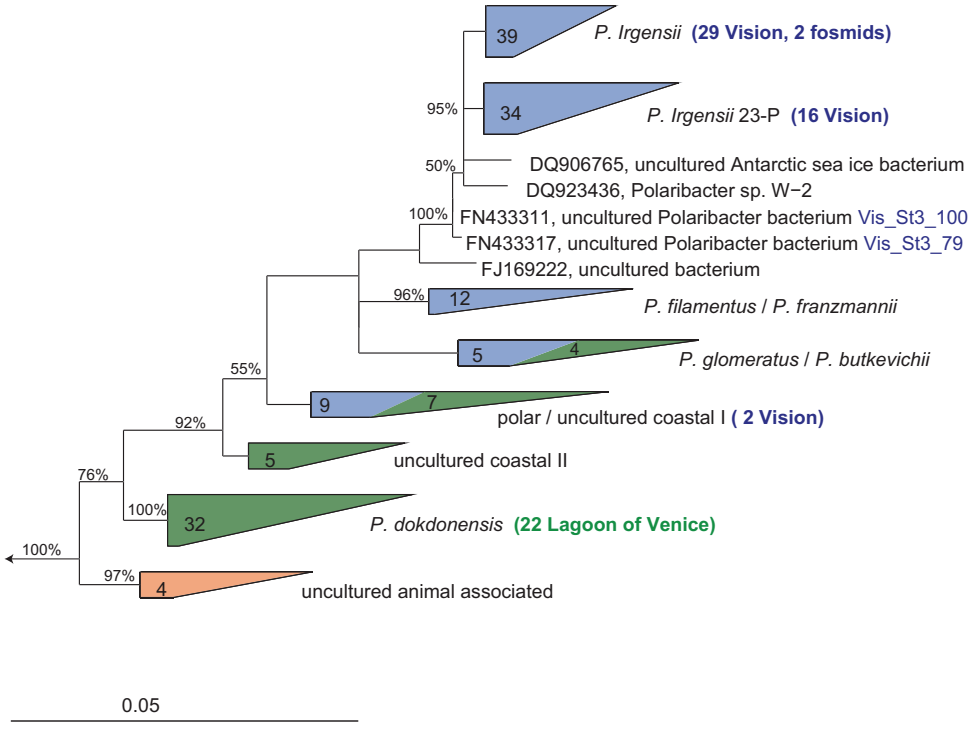


Figure 3.21: Phylogenetic affiliation of *Polaribacter* 16S rRNA gene sequences. Consensus tree was built with full–length sequences in the Silva 100 database (Pruesse et al., 2007) on the basis of neighbour joining and maximum likelihood methods, calculated without and with 30 % *Bacteroidetes* positional conservatory filters. Outgroup was *Tenacibaculum* strains (arrow). Colour of the clades shown in wedges represents the habitat from where the sequences were retrieved: Blue, polar; green, temperate coastal; red, diverse. The number of sequences in each group is indicated. Sequences retrieved from the VISION samples (blue bold) and from the Lagoon of Venice (green bold) are also indicated.

environments, polar clades . The 16S rRNA gene identity between the polar clades of *P. irgensii* and the temperate clade of *P. dokdonensis* is between 95% and 96%. Our preliminary analysis of uncultured *Bacteroidetes* phylotypes suggests that there is a clear separation of at least two clades within *Polaribacter*, i.e. “polar” polaribacters and “temperate” polaribacters. Even within a single closely related *Polaribacter* clade ecotypes might exist.

Considering the high abundance of *Polaribacter* spp. (Figure 3.7, Figure 3.19) and its significant intra–genus diversity both in its physiology and in

16S rRNA identity we recommend that future studies should attempt to subdivide this important genus. We hypothesize that this will reveal uncoupled population dynamics of warm– and cold–water phylotypes with different life styles.

Chapter 4

Final Remarks

High abundance of members of the class *Flavobacteria* in marine waters has already been reported by Claude ZoBell in his famous monography "Marine Microbiology" (ZoBell, 1946). He linked the occurrence of these aerobic heterotrophic bacteria that often form yellow to orange colonies on agar plates to the aerobic degradation of algal polymers. In this thesis we add molecular evidence supporting the general importance of *Bacteroidetes*, and particularly *Flavobacteria*, in marine picoplankton. Our analyses confirm the predominance of the class *Flavobacteria*; yet, as revealed by metagenomics, *Sphingobacteria* and *Cytophagia* might also be important players in marine systems. Their natural abundances are rather unknown, hence, further studies should design FISH oligonucleotide probes for those classes in order to know which is their contribution to the *Bacteroidetes* picoplankton community.

A conceptual framework has been proposed in which biodiversity of an ecosystem is described by two extremes: a set of abundant taxa that are relevant in carbon and energy flow, and a vast majority of rare taxa (Pedros-Alio, 2006). The rare taxa, however, might also carry on fundamental ecosystem processes (Musat et al., 2008). A major outcome of this thesis was a CARD-FISH protocol for the identification and quantification of rare clades. In the studies presented here, a relative high fraction of the *Bacteroidetes* community of both compartments was identified and reliably quantified. This was

achieved by a suite of clade-specific oligonucleotide FISH probes applied on high–volume FISH filters. With CARD–FISH of high volume filters quantitative studies on "the rare biosphere" are now within reach, enabling the subsequent analysis of their role in biogeochemical cycles.

With the application of the methodological approaches mentioned above, the occurrence of distinct *Bacteroidetes* communities in the North Atlantic Ocean was unveiled. Interestingly, our analysis revealed that the bacteroidetal community composition largely reflects the space–based demarcation of ocean provinces, which is mostly based on the chlorophyll content and sea surface temperature of those waters. Nevertheless, rare taxa seemed not to be directly influenced (based on correlation analysis) by physicochemical or biotic parameters, i.e. by "bottom–up" controls. It suggests that other factors might be influencing their population sizes. Picoplantkon can be tightly controlled by protistan grazing and viral lysis (Pernthaler, 2005), known as "top–down" controls. Protists might have distinct prey preferences (Massana et al., 2009), therefore differentially controlling the population sizes of picoplantkon clades (Massana et al., 2009; Pernthaler et al., 2004). Knowledge about species-specific grazing over marine *Bacteroidetes* clades is scarce, hence, it is of fundamental importance to address this topic in order to understand their role in the marine "microbial loop".

Bacteroidetes genomic fragments, mostly from yet uncultured clades, and an extensive genome comparison gave evidence of a strong specialization for the degradation of particular polymers and proteins, corroborating previous genomic (Bauer et al., 2006) and experimental information (Cottrell and Kirchman, 2000). Interestingly, polysaccharide degradation was confined to a single *Polaribacter* bacterium. It might represent a distinct ecotype within the genus with a specialization for the degradation of sulfated polysaccharides, known to be major constituents of algae cells.

Furthermore, a comparison between oceanic provinces gave first insights into an adaptation of the bacteroidetal community for contrasting environmental conditions. However, for robust statistical comparisons a higher sequencing coverage of the *Bacteroidetes* (meta)–genome is now needed. The bacteroidetal community that inhabit oligotrophic water masses showed a

higher potential for proteins and amino sugars degradation, while in the phytoplankon–rich water masses a higher potential for adhesion to cells was identified.

Fluorescence *in situ* hybridization analysis gave supporting evidence about the niche of *Bacteroidetes* as particle inhabitants. We proved a specific enrichment of certain bacteroidetal clades in the phycosphere of nanophytoplankton cells. Distinct taxonomic groups of phytoplankton release different amounts and types of exudates to its phycosphere (e.g. Biddanda and Benner, 1997), which might drive the colonization of particular *Bacteroidetes* phylotypes. Single phytoplankton cells with its inhabitant bacteroidetal cells -identified by FISH- could be separated by fluorescence–activated flow cytometry cell sorting further enabling single–cell genomics (e.g. Woyke et al., 2009). By this approach we seek to unveil the precise genomic potential of the phytoplantkon–associated *Bacteroidetes*.

The microscale in which microorganisms interact with organic matter and phytoplankton is crucial for understanding the cycling of carbon in the oceans (Azam and Malfatti, 2007). The microscale interactions affect ocean–basin–scale processes, including the cycling of carbon, nitrogen and phosphorus and the biogeochemical behavior of organic matter, as sinking for example (Azam and Malfatti, 2007). Follow up studies should elucidate the structure of the *Bacteroidetes*–phytoplankton micro–scale interactions and the identification of the conjoint partners. Interestingly, atomic force microscopy (AFM) has revealed a tight spatial association between bacteria–bacteria and bacteria–*Synechococcus* in the marine plankton (Malfatti and Azam, 2009). AFM can be combined with epifluorescence microscopy (EFM) (Mangold et al., 2008) and it is promising to further allow the identification of the partners by fluorescence *in situ* hybridization.

Further studies should also address the *Bacteroidetes in situ* activity. Particularly, the advent of single cell techniques like microautoradiography (MAR) and nanometer–scale secondary–mass spectrometry (NanoSIMS) combined with FISH (Behrens et al., 2008; Musat et al., 2008) has allowed the phylogenetic identification of microbial partners and to monitor the transfer of metabolites between them (Behrens et al., 2008). We suggest that

tracer incubation experiments with radioactively or isotopically labeled phytoplankton and bacterial derived organic matter should be performed and the incorporation rates of single *Bacteroidetes* cells evaluated.

The combination of genomics, identification, quantification, and incubation experiments will reveal species–specific interactions between marine *Bacteroidetes* and phytoplankton, allowing the better understanding of their ecological role in the marine carbon cycle.

References

Abell, G. C. J. and J. P. Bowman: 2005a, 'Colonization and community dynamics of class *Flavobacteria* on diatom detritus in experimental mesocosms based on Southern Ocean seawater'. *Fems Microbiology Ecology* **53**(3), 379–391.

Abell, G. C. J. and J. P. Bowman: 2005b, 'Ecological and biogeographic relationships of class *Flavobacteria* in the Southern Ocean'. *Fems Microbiology Ecology* **51**(2), 265–277.

Abreu, P. C., C. Odebrecht, and A. Gonzalez: 1994, 'Particulate and dissolved phytoplankton production of the patos lagoon estuary, southern brazil - comparison of methods and influencing factors'. *Journal of Plankton Research* **16**(7), 737–753.

Acha, E. M., H. Mianzan, R. Guerrero, J. Carreto, D. Giberto, N. Montoya, and M. Carignan: 2008, 'An overview of physical and ecological processes in the Rio de la Plata Estuary'. *Continental Shelf Research* **28**(13), 1579–1588.

Acinas, S. G., V. Klepac-Ceraj, D. E. Hunt, C. Pharino, I. Ceraj, D. L. Distel, and M. F. Polz: 2004, 'Fine-scale phylogenetic architecture of a complex bacterial community'. *Nature* **430**(6999), 551–554.

Alonso, C.: 2005, 'Identitiy and activity of marine microbial populations as revealed by single cell techniques'. Phd, Universitat Bremen.

Alonso, C., F. Warnecke, R. Amann, and J. Pernthaler: 2007, 'High local

and global diversity of *Flavobacteria* in marine plankton'. *Environmental Microbiology* **9**(5), 1253–1266.

Alonso-Saez, L., J. Aristegui, J. Pinhassi, L. Gomez-Consarnau, J. M. Gonzalez, D. Vaque, S. Agusti, and J. M. Gasol: 2007, 'Bacterial assemblage structure and carbon metabolism along a productivity gradient in the NE Atlantic Ocean'. *Aquatic Microbial Ecology* **46**(1), 43–53.

Amann, R. and B. M. Fuchs: 2008, 'Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques'. *Nature Reviews Microbiology* **6**(5), 339–348.

Amann, R. I., B. J. Binder, R. J. Olson, S. W. Chisholm, R. Devereux, and D. A. Stahl: 1990, 'Combination of 16s ribosomal RNA targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations'. *Applied and Environmental Microbiology* **56**(6), 1919–1925.

Anderson, K. L. and A. A. Salyers: 1989a, 'Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch binding sites and periplasmic starch degrading enzymes'. *Journal of Bacteriology* **171**(6), 3192–3198.

Anderson, K. L. and A. A. Salyers: 1989b, 'Genetic evidence that outer membrane binding of starch is required for starch utilization by *Bacteroides thetaiotaomicron*'. *Journal of Bacteriology* **171**(6), 3199–3204.

Arnosti, C.: 2003, 'Microbial extracellular enzymes and their role in dissolved organic matter cycling'. In: S. Findlay and R. Sinsabaugh (eds.): *Aquatic ecosystems: interactivity of dissolved organic matter*, Aquatic Ecology Series. San Diego, CA: Academic Press, pp. 316–342.

Azam, F.: 2001, 'Introduction, history, and overview: The 'methods' to our madness'. *Methods in Microbiology, Vol 30* **30**, 1–12.

Azam, F., T. Fenchel, J. G. Field, J. S. Gray, L. A. Meyerreil, and F. Thingstad: 1983, 'The ecological role of water-column microbes in the sea'. *Marine Ecology-Progress Series* **10**(3), 257–263.

Azam, F. and R. A. Long: 2001, 'Oceanography: Sea snow microcosms'. *Nature* **414**(6863), 495–498.

Azam, F. and F. Malfatti: 2007, 'Microbial structuring of marine ecosystems'. *Nature Reviews Microbiology* **5**(10), 782–791.

Baas-Becking, L.: 1934, *Geobiologie of Inleiding Tot de Milieukunde*. Van Stockkum & Zoon, The Hague, The Netherlands: .

Bano, N. and J. T. Hollibaugh: 2002, 'Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean'. *Applied and Environmental Microbiology* **68**(2), 505–518.

Barbeyron, T., S. L'Haridon, E. Corre, B. Kloareg, and P. Potin: 2001, '*Zobellia galactanovorans* gen. nov., sp nov., a marine species of *Flavobacteriaceae* isolated from a red alga, and classification of *Cytophaga uliginosa* (ZoBell and Upham 1944) Reichenbach 1989 as *Zobellia uliginosa* gen. nov., comb. nov.'. *International Journal of Systematic and Evolutionary Microbiology* **51**, 985–997.

Barcina, I., P. Lebaron, and J. VivesRego: 1997, 'Survival of allochthonous bacteria in aquatic systems: A biological approach'. *Fems Microbiology Ecology* **23**(1), 1–9.

Bauer, M., M. Kube, H. Teeling, M. Richter, T. Lombardot, E. Allers, C. A. Wurdemann, C. Quast, H. Kuhl, F. Knaust, D. Woebken, K. Bischof, M. Mussmann, J. V. Choudhuri, F. Meyer, R. Reinhardt, R. I. Amann, and F. O. Glockner: 2006, 'Whole genome analysis of the marine *Bacteroidetes 'Gramella forsetii'* reveals adaptations to degradation of polymeric organic matter'. *Environmental Microbiology* **8**(12), 2201–2213.

Begon, M., C. Townsend, and J. Harper: 2006, *Ecology: From Individuals to Ecosystems*. Oxford: Blackwell Publishing, fourth edition.

Behrens, S., T. Losekann, J. Pett-Ridge, P. K. Weber, W. O. Ng, B. S. Stevenson, I. D. Hutcheon, D. A. Relman, and A. M. Spormann: 2008, 'Linking microbial phylogeny to metabolic activity at the single-cell level

by using enhanced element labeling-catalyzed reporter deposition fluorescence in situ hybridization (EL-FISH) and NanoSIMS'. *Applied and Environmental Microbiology* **74**(10), 3143–3150.

Beja, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich, and E. F. DeLong: 2000, 'Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea'. *Science* **289**(5486), 1902–1906.

Beja, O., E. V. Koonin, L. Aravind, L. T. Taylor, H. Seitz, J. L. Stein, D. C. Bensen, R. A. Feldman, R. V. Swanson, and E. F. DeLong: 2002, 'Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces'. *Applied and Environmental Microbiology* **68**(1), 335–345.

Beja, O., E. N. Spudich, J. L. Spudich, M. Leclerc, and E. F. DeLong: 2001, 'Proteorhodopsin phototrophy in the ocean'. *Nature* **411**(6839), 786–789.

Benner, R. and K. Kaiser: 2003, 'Abundance of amino sugars and peptidoglycan in marine particulate and dissolved organic matter'. *Limnology and Oceanography* **48**(1), 118–128.

Benner, R., J. D. Pakulski, M. Mccarthy, J. I. Hedges, and P. G. Hatcher: 1992, 'Bulk chemical characteristics of dissolved organic-matter in the ocean'. *Science* **255**(5051), 1561–1564.

Bernardet, J. F. and Y. Nakagawa: 2006, 'Introduction to the Family *Flavobacteriacea*'. In: M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer, and E. Stackebrandt (eds.): *The Porkaryotes*, Vol. 7. New York: Springer-Verlag, third edition, pp. 550–580.

Bernardet, J. F., P. Segers, M. Vancanneyt, F. Berthe, K. Kersters, and P. Vandamme: 1996, 'Cutting a gordian knot: Emended classification and description of the genus *Flavobacterium*, emended description of the family *Flavobacteriaceae*, and proposal of *Flavobacterium hydatis* nom nov (basonym, *Cytophaga aquatilis* Strohl and Tait 1978)'. *International Journal of Systematic Bacteriology* **46**(1), 128–148.

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

Biegala, I. C., G. Kennaway, E. Alverca, J. F. Lennon, D. Vaulot, and N. Simon: 2002, 'Identification of bacteria associated with dinoflagellates (*Dinophyceae*) *Alexandrium* spp. using tyramide signal amplification-fluorescent in situ hybridization and confocal microscopy'. *Journal of Phycology* **38**(2), 404–411.

Bisbal, G. A.: 1995, 'The southeast South-american shelf large marine ecosystem - evolution and components'. *Marine Policy* **19**(1), 21–38.

Boltovskoy, D., M. Gibbons, D. Hutchings, and D. Binet: 1999, 'General biological features of South Atlantic'. In: D. Boltovskoy (ed.): *Zooplankton*. Leiden, The Netherlands: Backhyus Publishers, pp. 1–42.

Bonilla, S., D. Conde, L. Aubriot, L. Rodriguez-Gallego, C. Piccini, E. Meerhof, L. Rodriguez-Grana, D. Calliari, P. Gómez, I. Machado, and A. Britos: 2006, 'Procesos estructuradores de las comunidades biológicas en lagunas costeras de Uruguay'. In: R. Menafra, L. Rodriguez-Gallego, F. Scarabino, and D. Conde (eds.): *Bases para la conservación y el manejo de la costa Uruguaya*, Vol. 1. Montevideo: Vida Silvestre Uruguay, pp. 611–630.

Borin, S., L. Brusetti, D. Daffonchio, E. Delaney, and F. Baldi: 2009, 'Biodiversity of prokaryotic communities in sediments of different sub-basins of the Venice lagoon'. *Research in Microbiology* **160**(5), 307–314.

Bouvier, T. C. and P. A. del Giorgio: 2002, 'Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries'. *Limnology and Oceanography* **47**(2), 453–470.

Bowman, J. P.: 2006, 'The Marine Clade of the Family *Flavobacteriaceae*: The Genera *Aequorivita, Arenibacter, Cellulophaga, Croceibacter, Formosa, Gelidibacter, Gillisia, Maribacter, Mesonia, Muricauda, Polaribacter, Psychroflexus, Psychroserpens, Robiginitalea, Salegentibacter,*

Tenacibaculum, Ulvibacter, Vitellibacter and *Zobellia*'. In: M. Dworkin, S. Falkow, E. Rosenberg, K. Schleifer, and E. Stackebrandt (eds.): *The Prokaryotes*, Vol. 7. New York: Springer-Verlag, third edition, pp. 677–694.

Bowman, J. P., C. Mancuso, C. M. Nichols, and J. A. E. Gibson: 2003, '*Algoriphagus ratkowskyi* gen. nov., sp nov., *Brumimicrobium glaciale* gen. nov., sp nov., *Cryomorpha ignava* gen. nov., sp nov and *Crocinitomix* catalasitica gen. nov., sp nov., novel flavobacteria isolated from various polar habitats'. *International Journal of Systematic and Evolutionary Microbiology* **53**, 1343–1355.

Bowman, J. P. and D. S. Nichols: 2002, '*Aequorivita* gen. nov., a member of the family Flavobacteriaceae isolated from terrestrial and marine Antarctic habitats'. *International Journal of Systematic and Evolutionary Microbiology* **52**, 1533–1541.

Braun, T. F. and M. J. McBride: 2005, '*Flavobacterium johnsoniae* GldJ is a lipoprotein that is required for gliding motility'. *Journal of Bacteriology* **187**(8), 2628–2637.

Braun, V. and C. Herrmann: 1993, 'Evolutionary relationship of uptake systems for biopolymers in *Escherichia coli* - cross-complementation between the TonB Exbb Exbd and the Tola Tolq Tolr proteins'. *Molecular Microbiology* **8**(2), 261–268.

Brinkmeyer, R., K. Knittel, J. Jurgens, H. Weyland, R. Amann, and E. Helmke: 2003, 'Diversity and structure of bacterial communities in arctic versus antarctic pack ice'. *Applied and Environmental Microbiology* **69**(11), 6610–6619.

Calliari, D., E. Brugnoli, G. Ferrari, and D. Vizziano: 2009, 'Phytoplankton distribution and production along a wide environmental gradient in the South-West Atlantic off Uruguay'. *Hydrobiologia* **620**, 47–61.

Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat: 2009, 'The Carbohydrate-Active EnZymes database (CAZy):

an expert resource for Glycogenomics'. *Nucleic Acids Research* **37**, D233–D238.

Carlson, C. A., R. Morris, R. Parsons, A. H. Treusch, S. J. Giovannoni, and K. Vergin: 2009, 'Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea'. *Isme Journal* **3**(3), 283–295.

Castel, J., P. Caumette, and R. Herbert: 1996, 'Eutrophication gradients in coastal lagoons as exemplified by the Bassin d'Arcachon and the Etang du Prevost'. *Hydrobiologia* **329**(1-3), R9–R28.

Celussi, M., A. Pugnetti, and P. Del Negro: 2009, 'Structural dynamics of bacterioplankton assemblages in the Lagoon of Venice'. *Estuarine Coastal and Shelf Science* **84**(1), 154–160.

Chen, X. H., Y. Zeng, and N. Jiao: 2008, 'Characterization of *Cytophaga-Flavobacteria* community structure in the bering sea by cluster-specific 16S rRNA gene amplification analysis'. *Journal of Microbiology and Biotechnology* **18**(2), 194–198.

Chen, X. H., Y. H. Zeng, and N. Z. Jiao: 2006, 'Development and evaluation of specific 16S rDNA primers for marine *Cytophaga-Flavobacteria* cluster'. *Molecular Ecology Notes* **6**(4), 1278–1281.

Chisholm, S., R. Olson, R. Zettler, J. Waterbury, R. Goericke, and N. Welschmeyer: 1988, 'A novel free-living prochlorophyte occurs at high cell concentrations in the oceanic euphotic zone'. *Nature* **334**, 340343.

Cho, K. H. and A. A. Salyers: 2001, 'Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*'. *Journal of Bacteriology* **183**(24), 7224–7230.

Clarke, K. R., P. J. Somerfield, and R. N. Gorley: 2008, 'Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-

environment linkage'. *Journal of Experimental Marine Biology and Ecology* **366**(1-2), 56–69.

Connon, S. A. and S. J. Giovannoni: 2002, 'High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates'. *Applied and Environmental Microbiology* **68**(8), 3878–3885.

Cottrell, M. T. and D. L. Kirchman: 2000, 'Natural assemblages of marine proteobacteria and members of the *Cytophaga-Flavobacter* cluster consuming low- and high-molecular-weight dissolved organic matter'. *Applied and Environmental Microbiology* **66**(4), 1692–1697.

Cottrell, M. T., L. Y. Yu, and D. L. Kirchman: 2005, 'Sequence and expression analyses of *Cytophaga*-like hydrolases in a Western arctic metagenornic library and the Sargasso seat'. *Applied and Environmental Microbiology* **71**(12), 8506–8513.

Covert, J. S. and M. A. Moran: 2001, 'Molecular characterization of estuarine bacterial communities that use high- and low-molecular weight fractions of dissolved organic carbon'. *Aquatic Microbial Ecology* **25**(2), 127–139.

Crump, B. C., E. V. Armbrust, and J. A. Baross: 1999, 'Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia river, its estuary, and the adjacent coastal ocean'. *Applied and Environmental Microbiology* **65**(7), 3192–3204.

Cumashi, A., N. A. Ushakova, M. E. Preobrazhenskaya, A. D'Incecco, A. Piccoli, L. Totani, N. Tinari, G. E. Morozevich, A. E. Berman, M. I. Bilan, A. I. Usov, N. E. Ustyuzhanina, A. A. Grachev, C. J. Sanderson, M. Kelly, G. A. Rabinovich, S. Iacobelli, N. E. Nifantiev, and C. I. Nazio: 2007, 'A comparative study of the anti-inflammatory, anticoagulant, antiangiogenic, and antiadhesive activities of nine different fucoidans from brown seaweeds'. *Glycobiology* **17**(5), 541–552.

Dahl, U., T. Jaeger, B. T. Nguyen, J. M. Sattler, and C. Mayer: 2004, 'Identification of a phosphotransferase system of *Escherichia coli* required

for growth on N-acetylmuramic acid'. *Journal of Bacteriology* **186**(8), 2385–2392.

Daubin, V., N. A. Moran, and H. Ochman: 2003, 'Phylogenetics and the cohesion of bacterial genomes'. *Science* **301**(5634), 829–832.

DeLong, E. F.: 2009, 'The microbial ocean from genomes to biomes'. *Nature* **459**(7244), 200–206.

DeLong, E. F., D. G. Franks, and A. L. Alldredge: 1993, 'Phylogenetic diversity of aggregate-attached vs free-living marine bacterial assemblages'. *Limnology and Oceanography* **38**(5), 924–934.

DeLong, E. F., C. M. Preston, T. Mincer, V. Rich, S. J. Hallam, N. U. Frigaard, A. Martinez, M. B. Sullivan, R. Edwards, B. R. Brito, S. W. Chisholm, and D. M. Karl: 2006, 'Community genomics among stratified microbial assemblages in the ocean's interior'. *Science* **311**(5760), 496–503.

Ducklow, H. W. and C. A. Carlson: 1992, 'Oceanic bacterial production'. *Advances in Microbial Ecology* **12**, 113–181.

Ducklow, H. W. and R. P. Harris: 1993, 'Introduction to the Jgofs North-Atlantic bloom experiment'. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **40**(1-2), 1–8.

Eilers, H., J. Pernthaler, J. Peplies, F. O. Glockner, G. Gerdts, and R. Amann: 2001, 'Isolation of novel pelagic bacteria from the German bight and their seasonal contributions to surface picoplankton'. *Applied and Environmental Microbiology* **67**(11), 5134–5142.

Erra-Basells, R., A. A. Kolender, M. C. Matulewicz, H. Nonami, and A. S. Cerezo: 2000, 'Matrix-assisted ultraviolet laser-desorption ionization time-of-flight mass spectrometry of sulfated mannans from the red seaweed *Nothogenia fastigiata*'. *Carbohydrate Research* **329**(1), 157–167.

Facca, C., A. Sfriso, and G. Socal: 2002, 'Changes in abundance and composition of phytoplankton and microphytobenthos due to increased sediment

fluxes in the Venice lagoon, Italy'. *Estuarine Coastal and Shelf Science* **54**(5), 773–792.

Feingersch, R., M. T. Suzuki, M. Shmoish, I. Sharon, G. Sabehi, F. Partensky, and O. Beja: 2010, 'Microbial community genomics in eastern Mediterranean Sea surface waters'. *Isme Journal* **4**(1), 78–87.

Fenchel, T. and B. J. Finlay: 2004, 'The ubiquity of small species: Patterns of local and global diversity'. *Bioscience* **54**(8), 777–784.

Field, K. G., D. Gordon, T. Wright, M. Rappe, E. Urbach, K. Vergin, and S. J. Giovannoni: 1997, 'Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria'. *Applied and Environmental Microbiology* **63**(1), 63–70.

Finlay, B. J.: 2002, 'Global dispersal of free-living microbial eukaryote species'. *Science* **296**(5570), 1061–1063.

Fogg, G. E.: 1983, 'The ecological significance of extracellular products of phytoplankton photosynthesis'. *Botanica Marina* **26**(1), 3–14.

Framinan, M. B. and O. B. Brown: 1996, 'Study of the Rio de la Plata turbidity front .1. Spatial and temporal distribution'. *Continental Shelf Research* **16**(10), 1259–&.

Frei, E. and R. D. Preston: 1968, 'Non-cellulosic structural polysaccharides in algal cell walls , mannan in siphoneous green algae'. *Proceedings of the Royal Society of London Series B-Biological Sciences* **169**(1015), 127–&.

Frigaard, N. U., A. Martinez, T. J. Mincer, and E. F. DeLong: 2006, 'Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea'. *Nature* **439**(7078), 847–850.

Fuhrman, J. A., I. Hewson, M. S. Schwalbach, J. A. Steele, M. V. Brown, and S. Naeem: 2006, 'Annually reoccurring bacterial communities are predictable from ocean conditions'. *Proceedings of the National Academy of Sciences of the United States of America* **103**(35), 13104–13109.

Fuhrman, J. A., M. S. Schwalbach, and U. Stingl: 2008a, 'Opinion - Proteorhodopsins: an array of physiological roles?'. *Nature Reviews Microbiology* **6**(6), 488–494.

Fuhrman, J. A., J. A. Steele, I. Hewson, M. S. Schwalbach, M. V. Brown, J. L. Green, and J. H. Brown: 2008b, 'A latitudinal diversity gradient in planktonic marine bacteria'. *Proceedings of the National Academy of Sciences of the United States of America* **105**(22), 7774–7778.

Garay, E., A. Arnau, and C. Amaro: 1985, 'Incidence of *Vibrio cholerae* and Related *Vibrios* in a Coastal Lagoon and Seawater Influenced by Lake Discharges Along an Annual Cycle'. *Applied and Environmental Microbiology* **50**(2), 426–430.

Gifford, D. J., L. M. Fessenden, P. R. Garrahan, and E. Martin: 1995, 'Grazing by Microzooplankton and Mesozooplankton in the High-Latitude North-Atlantic Ocean - Spring Versus Summer Dynamics'. *Journal of Geophysical Research-Oceans* **100**(C4), 6665–6675.

Giordano, S. and C. Lasta: 2004a, 'Erosión en las costas del Río de la Plata y su Frente Marítimo'.

Giordano, S. and C. Lasta: 2004b, 'Patrones de circulación en el Río de la Plata y su Area de influencia'.

Giovannoni, S. and M. Rappe: 2000, 'Evolution, diversity, and molecular ecology of marine prokaryotes'. In: D. L. Kirchman (ed.): *Microbial ecology of the oceans.* New York: Wiley-Liss, Inc, pp. 47–84.

Giovannoni, S. and U. Stingl: 2007, 'The importance of culturing bacterioplankton in the 'omics' age'. *Nature Reviews Microbiology* **5**(10), 820–826.

Giovannoni, S. J., L. Bibbs, J. C. Cho, M. D. Stapels, R. Desiderio, K. L. Vergin, M. S. Rappe, S. Laney, L. J. Wilhelm, H. J. Tripp, E. J. Mathur, and D. F. Barofsky: 2005, 'Proteorhodopsin in the ubiquitous marine bacterium SAR11'. *Nature* **438**(7064), 82–85.

Giovannoni, S. J., T. B. Britschgi, C. L. Moyer, and K. G. Field: 1990, 'Genetic diversity in sargasso sea bacterioplankton'. *Nature* **345**(6270), 60–63.

Giovannoni, S. J. and U. Stingl: 2005, 'Molecular diversity and ecology of microbial plankton'. *Nature* **437**(7057), 343–348.

Glockner, F. O., B. M. Fuchs, and R. Amann: 1999, 'Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization'. *Applied and Environmental Microbiology* **65**(8), 3721–3726.

Glockner, F. O., M. Kube, M. Bauer, H. Teeling, T. Lombardot, W. Ludwig, D. Gade, A. Beck, K. Borzym, K. Heitmann, R. Rabus, H. Schlesner, R. Amann, and R. Reinhardt: 2003, 'Complete genome sequence of the marine planctomycete *Pirellula* sp strain 1'. *Proceedings of the National Academy of Sciences of the United States of America* **100**(14), 8298–8303.

Glockner, F. O. and A. Meyerdierks: 2006, 'Metagenome Analysis'. In: E. Stackebrandt (ed.): *Molecular identification, systematics and population structure of prokaryotes*. Heilderberg: Springer-Verlag, pp. 261–285.

Glockner, F. O., E. Zaichikov, N. Belkova, L. Denissova, J. Pernthaler, A. Pernthaler, and R. Amann: 2000, 'Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria'. *Applied and Environmental Microbiology* **66**(11), 5053–5065.

Gomez-Consarnau, L., J. M. , M. Coll-Llado, P. Gourdon, T. Pascher, R. Neutze, C. Pedros-Alio, and J. Pinhassi: 2007, 'Light stimulates growth of proteorhodopsin-containing marine *Flavobacteria*'. *Nature* **445**(7124), 210–213.

Gonzalez, J. M., B. Fernandez-Gomez, A. Fernandez-Guerra, L. Gomez-Consarnau, O. Sanchez, M. Coll-Llado, J. del Campo, L. Escudero, R. Rodriguez-Martinez, L. Alonso-Saez, M. Latasa, I. Paulsen, O.

Nedashkovskaya, I. Lekunberri, J. Pinhassi, and C. Pedros-Alio: 2008, 'Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp MED152 (Flavobacteria)'. *Proceedings of the National Academy of Sciences of the United States of America* **105**(25), 8724–8729.

Gosink, J. J., C. R. Woese, and J. T. Staley: 1998, '*Polaribacter* gen. nov., with three new species, *P.irgensii* sp. nov., *P.franzmannii* sp. nov., and *P.filamentus* sp. nov., gas vacuolate polar marine bacteria of the *Cytophaga-Flavobacterium-Bacteroides* group and reclassification of '*Flectobacillus glomeratus*' as *Polaribacter glomeratus* comb. nov.'. *International Journal of Systematic Bacteriology* **48**, 223–235.

Grossart, H. P., F. Levold, M. Allgaier, M. Simon, and T. Brinkhoff: 2005, 'Marine diatom species harbour distinct bacterial communities'. *Environmental Microbiology* **7**(6), 860–873.

Grzymski, J. J., B. J. Carter, E. F. DeLong, R. A. Feldman, A. Ghadiri, and A. E. Murray: 2006, 'Comparative genomics of DNA fragments from six antarctic marine planktonic bacteria'. *Applied and Environmental Microbiology* **72**(2), 1532–1541.

Hall-Stoodley, L., J. W. Costerton, and P. Stoodley: 2004, 'Bacterial biofilms: From the natural environment to infectious diseases'. *Nature Reviews Microbiology* **2**(2), 95–108.

Handelsman, J.: 2004, 'Metagenomics: Application of genomics to uncultured microorganisms'. *Microbiology and Molecular Biology Reviews* **68**(4), 669–685.

Hedges, J. I.: 1992, 'Global Biogeochemical Cycles - Progress and Problems'. *Marine Chemistry* **39**(1-3), 67–93.

Hellebust, J. A.: 1965, 'Excretion of some organic compounds by marine phytoplankton'. *Limnology and Oceanography* **10**, 192–206.

Hellerma, S.: 1967, 'An updated estimate of wind stress on world ocean'. *Monthly Weather Review* **95**(9), 607–&.

Henson, S. A., I. Robinson, J. T. Allen, and J. J. Waniek: 2006, 'Effect of meteorological conditions on interannual variability in timing and magnitude of the spring bloom in the Irminger Basin, North Atlantic'. *Deep-Sea Research Part I-Oceanographic Research Papers* **53**(10), 1601–1615.

Holtje, J. V.: 1998, 'Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*'. *Microbiology and Molecular Biology Reviews* **62**(1), 181–+.

Hou, S. B., J. H. Saw, K. S. Lee, T. A. Freitas, C. Belisle, Y. Kawarabayasi, S. P. Donachie, A. Pikina, M. Y. Galperin, E. V. Koonin, K. S. Makarova, M. V. Omelchenko, A. Sorokin, Y. I. Wolf, Q. X. Li, Y. S. Keum, S. Campbell, J. Denery, S. I. Aizawa, S. Shibata, A. Malahoff, and M. Alam: 2004, 'Genome sequence of the deep-sea gamma-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy'. *Proceedings of the National Academy of Sciences of the United States of America* **101**(52), 18036–18041.

Huang, S.: 2009, 'A bioinformatic pipeline for the taxonomic assignment of metagenome fragments for future analysis of an uncultured anaerobic plantomycete from a Black Sea microbial mat'. Diploma, University Bremen.

Huber, J. A., D. Mark Welch, H. G. Morrison, S. M. Huse, P. R. Neal, D. A. Butterfield, and M. L. Sogin: 2007, 'Microbial population structures in the deep marine biosphere'. *Science* **318**(5847), 97–100.

Ibrahim, A., P. GernerSmidt, and W. Liesack: 1997, 'Phylogenetic relationship of the twenty-one DNA groups of the genus *Acinetobacter* as revealed by 16S ribosomal DNA sequence analysis'. *International Journal of Systematic Bacteriology* **47**(3), 837–841.

Ivanova, E. P., O. I. Nedashkovskaya, J. Chun, A. M. Lysenko, G. M. Frolova, V. I. Svetashev, M. V. Vysotskii, V. V. Mikhailov, A. Huq, and R. R. Colwell: 2001, '*Arenibacter* gen. nov., new genus of the family *Flavobacteriaceae* and description of a new species, *Arenibacter latericius* sp nov.'.

International Journal of Systematic and Evolutionary Microbiology **51**, 1987–1995.

Jaeger, T., M. Arsic, and C. Mayer: 2005, 'Scission of the lactyl ether bond of N-acetylmuramic acid by *Escherichia coli* "etherase"'. *Journal of Biological Chemistry* **280**(34), 30100–30106.

Karner, M. B., E. F. DeLong, and D. M. Karl: 2001, 'Archaeal dominance in the mesopelagic zone of the Pacific Ocean'. *Nature* **409**(6819), 507–510.

Kirchman, D. L.: 2002, 'The ecology of C*ytophaga-Flavobacteria* in aquatic environments'. *Fems Microbiology Ecology* **39**(2), 91–100.

Kirchman, D. L., L. Y. Yu, and M. T. Cottrell: 2003, 'Diversity and abundance of uncultured *Cytophaga*-like bacteria in the Delaware Estuary'. *Applied and Environmental Microbiology* **69**(11), 6587–6596.

Kjerfve, B.: 1994, *Coastal lagoon processes*. Amsterdam: Elsevier.

Konstantinidis, K. T. and J. M. Tiedje: 2005, 'Genomic insights that advance the species definition for prokaryotes'. *Proceedings of the National Academy of Sciences of the United States of America* **102**(7), 2567–2572.

Kormas, K. A., M. K. Tivey, K. Von Damm, and A. Teske: 2006, 'Bacterial and archaeal phylotypes associated with distinct mineralogical layers of a white smoker spire from a deep-sea hydrothermal vent site (9 degrees N, East Pacific Rise)'. *Environmental Microbiology* **8**(5), 909–920.

Krause, L., N. N. Diaz, A. Goesmann, S. Kelley, T. W. Nattkemper, F. Rohwer, R. A. Edwards, and J. Stoye: 2008, 'Phylogenetic classification of short environmental DNA fragments'. *Nucleic Acids Research* **36**(7), 2230–2239.

Krauss, W.: 1986, 'The North-Atlantic Current'. *Journal of Geophysical Research-Oceans* **91**(C4), 5061–5074.

Kuypers, M. M. M., G. Lavik, D. Woebken, M. Schmid, B. M. Fuchs, R. Amann, B. B. Jorgensen, and M. S. M. Jetten: 2005, 'Massive nitrogen

loss from the Benguela upwelling system through anaerobic ammonium oxidation'. *Proceedings of the National Academy of Sciences of the United States of America* **102**(18), 6478–6483.

LeblondBourget, N., H. Philippe, I. Mangin, and B. Decaris: 1996, '16S rRNA and 16S to 23S internal transcribed spacer sequence analyses reveal inter- and intraspecific *Bifidobacterium* phylogeny'. *International Journal of Systematic Bacteriology* **46**(1), 102–111.

Lechene, C., F. Hillion, G. McMahon, D. Benson, A. Kleinfeld, P. Kampf, D. Distel, Y. Luyten, J. Bonventre, D. Hentschel, K. Park, S. Ito, M. Schwartz, G. Benichou, and G. Slodzian: 2006, 'High-resolution quantitative imaging of mammalian and bacterial cells using stable isotope mass spectrometry'. *Journal of Biology* **5**, 20–.

Lee, N., P. H. Nielsen, K. H. Andreasen, S. Juretschko, J. L. Nielsen, K. H. Schleifer, and M. Wagner: 1999, 'Combination of fluorescent in situ hybridization and microautoradiography - a new tool for structure-function analyses in microbial ecology'. *Applied and Environmental Microbiology* **65**(3), 1289–1297.

Lewin, R. A.: 1965, 'Freshwater Species of *Saprospira*'. *Canadian Journal of Microbiology* **11**(2), 135–&.

Lewin, R. A.: 1970, '*Saprospira toviformis* nov-spec *Flexibacterales* from a New-Zealand Seashore'. *Canadian Journal of Microbiology* **16**(6), 507–&.

Lewin, R. A.: 1997, '*Saprospira grandis*: A flexibacterium that can catch bacterial prey by "ixotrophy"'. *Microbial Ecology* **34**(3), 232–236.

Little, C.: 2000, *The biology of soft shores and estuaries*. Oxford: Oxford University Press.

Llobet-Brossa, E., R. Rossello-Mora, and R. Amann: 1998, 'Microbial community composition of Wadden Sea sediments as revealed by fluorescence in situ hybridization'. *Applied and Environmental Microbiology* **64**(7), 2691–2696.

Longford, S. R., N. A. Tujula, G. R. Crocetti, A. J. Holmes, C. Holmstrom, S. Kjelleberg, P. D. Steinberg, and M. W. Taylor: 2007, 'Comparisons of diversity of bacterial communities associated with three sessile marine eukaryotes'. *Aquatic Microbial Ecology* **48**(3), 217–229.

Longhurst, A.: 1995, 'Seasonal cycles of pelagic production and consumption'. *Progress In Oceanography* **36**(2), 77–167.

Longhurst, A.: 1998, *Ecological geography of the sea*. London: Academic Press, first edition.

Losso, C. and A. Volpi Ghirardinia: 2010, 'Overview of ecotoxicological studies performed in the Venice Lagoon (Italy)'. *Environment International* **36**(1), 92–121.

Loy, A., M. Horn, and M. Wagner: 2003, 'probeBase: an online resource for rRNA-targeted oligonucleotide probes'. *Nucleic Acids Research* **31**(1), 514–516.

Lucea, A., C. M. Duarte, S. Agusti, and H. Kennedy: 2005, 'Nutrient dynamics and ecosystem metabolism in the Bay of Blanes (NW Mediterranean)'. *Biogeochemistry* **73**(2), 303–323.

Ludwig, W., J. P. Euzeby, and W. B. Whitman: 2008, 'Draft taxonomic outline of the *Bacteroidetes*, *Planctomycetes, Chlamydiae, Spirochaetes, Fibrobacteres, Fusobacteria, Acidobacteria, Verrucomicrobia, Dictyoglomi*, and *Gemmatimonadetes*'. In: B. Hedlund, N. Krieg, W. Ludwig, B. Paster, J. Staley, N. Ward, and W. B. Whitman (eds.): *Bergey's Manual of Systematic Bacteriology*, Vol. 4. New York: Springer -Verlag, second edition.

Ludwig, W. and H. Klenk: 2005, 'Overview: a phylogenetic backbone and taxonomic framework for procaryotic systematics'. In: D. J. Brenner, N. Krieg, J. T. Staley, and G. Garrity (eds.): *Bergey's Manual of Systematic Bacteriology*. New York: Springer -Verlag, second edition.

Ludwig, W., O. Strunk, S. Klugbauer, N. Klugbauer, M. Weizenegger, J. Neumaier, M. Bachleitner, and K. H. Schleifer: 1998, 'Bacterial phylogeny based on comparative sequence analysis'. *Electrophoresis* **19**(4), 554–568.

Malfatti, F. and F. Azam: 2009, 'Atomic force microscopy reveals microscale networks and possible symbioses among pelagic marine bacteria'. *Aquatic Microbial Ecology* **58**(1), 1–14.

Malmstrom, R. R., T. R. A. Straza, M. T. Cottrell, and D. L. Kirchman: 2007, 'Diversity, abundance, and biomass production of bacterial groups in the western Arctic Ocean'. *Aquatic Microbial Ecology* **47**(1), 45–55.

Mangold, S., K. Harneit, T. Rohwerder, G. Claus, and W. Sand: 2008, 'Novel combination of atomic force microscopy and epifluorescence microscopy for visualization of leaching bacteria on pyrite'. *Applied and Environmental Microbiology* **74**(2), 410–415.

Mann, E. L., N. Ahlgren, J. W. Moffett, and S. W. Chisholm: 2002, 'Copper toxicity and cyanobacteria ecology in the Sargasso Sea'. *Limnology and Oceanography* **47**(4), 976–988.

Manz, W., R. Amann, W. Ludwig, M. Vancanneyt, and K. H. Schleifer: 1996, 'Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum cytophaga-flavobacter-bacteroides in the natural environment'. *Microbiology-Uk* **142**, 1097–1106.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. T. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. G. Yu, R. F.

Begley, and J. M. Rothberg: 2005, 'Genome sequencing in microfabricated high-density picolitre reactors'. *Nature* **437**(7057), 376–380.

Martin-Cuadrado, A. B., F. Rodriguez-Valera, D. Moreira, J. C. Alba, E. Ivars-Martinez, M. R. Henn, E. Talla, and P. Lopez-Garcia: 2008, 'Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions'. *Isme Journal* **2**(8), 865–886.

Martinez, A., A. S. Bradley, J. R. Waldbauer, R. E. Summons, and E. F. DeLong: 2007, 'Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host'. *Proceedings of the National Academy of Sciences of the United States of America* **104**(13), 5590–5595.

Martinez, C. and S. Retta: 2001, 'Caracterizacion de las areas de cria de la corvina *Micropogonias furnieri* en la zona costera uruguaya'. In: D. Vizziano, Puig, Mesones, and G. Nagy (eds.): *El Rio de la Plata. Investigacion para la gestion del Ambiente, los recursos pesqueros y la pesqueria en el frente salino.* Montevideo: Programa Ecoplata, pp. 141–148.

Massana, R., F. Unrein, R. Rodriguez-Martinez, I. Forn, T. Lefort, J. Pinhassi, and F. Not: 2009, 'Grazing rates and functional diversity of uncultured heterotrophic flagellates'. *Isme Journal* **3**(5), 588–596.

McBride, M. J.: 2004, '*Cytophaga flavobacterium* gliding motility'. *Journal of Molecular Microbiology and Biotechnology* **7**(1-2), 63–71.

McBride, M. J., T. F. Braun, and J. L. Brust: 2003, '*Flavobacterium johnsoniae* GldH is a lipoprotein that is required for gliding motility and chitin utilization'. *Journal of Bacteriology* **185**(22), 6648–6657.

McBride, M. J., G. Xie, E. C. Martens, A. Lapidus, B. Henrissat, R. G. Rhodes, E. Goltsman, W. Wang, J. Xu, D. W. Hunnicutt, A. M. Staroscik, T. R. Hoover, Y. Q. Cheng, and J. L. Stein: 2009, 'Novel Features of the Polysaccharide-Digesting Gliding Bacterium *Flavobacterium johnsoniae* as

Revealed by Genome Sequence Analysis'. *Applied and Environmental Microbiology* **75**(21), 6864–6875.

McCarthy, M. D., J. I. Hedges, and R. Benner: 1998, 'Major bacterial contribution to marine dissolved organic nitrogen'. *Science* **281**(5374), 231–234.

Meyer, F., A. Goesmann, A. C. McHardy, D. Bartels, T. Bekel, J. Clausen, J. Kalinowski, B. Linke, O. Rupp, R. Giegerich, and A. Puhler: 2003, 'GenDB - an open source genome annotation system for prokaryote genomes'. *Nucleic Acids Research* **31**(8), 2187–2195.

Moore, L. R., A. F. Post, G. Rocap, and S. W. Chisholm: 2002, 'Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*'. *Limnology and Oceanography* **47**(4), 989–996.

Morris, R. M., M. S. Rappe, S. A. Connon, K. L. Vergin, W. A. Siebold, C. A. Carlson, and S. J. Giovannoni: 2002, 'SAR11 clade dominates ocean surface bacterioplankton communities'. *Nature* **420**(6917), 806–810.

Mullins, T. D., T. B. Britschgi, R. L. Krest, and S. J. Giovannoni: 1995, 'Genetic Comparisons Reveal the Same Unknown Bacterial Lineages in Atlantic and Pacific Bacterioplankton Communities'. *Limnology and Oceanography* **40**(1), 148–158.

Musat, N., H. Halm, B. Winterholler, P. Hoppe, S. Peduzzi, F. Hillion, F. Horreard, R. Amann, B. B. Jorgensen, and M. M. M. Kuypers: 2008, 'A single-cell view on the ecophysiology of anaerobic phototrophic bacteria'. *Proceedings of the National Academy of Sciences of the United States of America* **105**(46), 17861–17866.

Myklestad, S. M.: 1995, 'Release of extracellular products by phytoplankton with special emphasis on polysaccharides'. *Science of the Total Environment* **165**(1-3), 155–164.

Naganuma, T., D. A. Coury, M. Polnefuller, A. Gibor, and K. Horikoshi: 1993, 'Characterization of agarolytic *Microscilla* isolates and their extracellular agarases'. *Systematic and Applied Microbiology* **16**(2), 183–190.

Nagata, T.: 2000, 'Production mechanisms of dissolved organic matter'. In: *Microbial ecology of the oceans*. New York: Wiley-Liss, Inc, pp. 121–152.

Nagata, T. and D. L. Kirchman: 1999, 'Bacterial mortality: A pathway for the formation of refractory DOM?'. In: M. Brylinsky, C. Bell, and P. Johnson-Green (eds.): *New frontiers in microbial ecology: Proceedings of the 8th International Symposium on Microbial Ecology. Atlantic Canada Society for Microbial Ecology*. pp. 153–158.

Nagata, T., B. Meon, and D. L. Kirchman: 2003, 'Microbial degradation of peptidoglycan in seawater'. *Limnology and Oceanography* **48**(2), 745–754.

Nakagawa, Y. and K. Yamasato: 1996, 'Emendation of the genus *Cytophaga* and transfer of *Cytophaga agarovorans* and *Cytophaga salmonicolor* to *Marinilabilia* gen nov: Phylogenetic analysis of the *Flavobacterium-Cytophaga* complex'. *International Journal of Systematic Bacteriology* **46**(2), 599–603.

Nedashkovskaya, O. I., S. B. Kim, M. S. Lee, M. S. Park, K. H. Lee, A. M. Lysenko, H. W. Oh, V. V. Mikhailov, and K. S. Bae: 2005, '*Cyclobacterium amurskyense* sp nov., a novel marine bacterium isolated from sea water'. *International Journal of Systematic and Evolutionary Microbiology* **55**, 2391–2394.

Neef, A., R. Amann, H. Schlesner, and K. H. Schleifer: 1998, 'Monitoring a widespread bacterial group: in situ detection of planctomycetes with 16S rRNA-targeted probes'. *Microbiology-Uk* **144**, 3257–3266.

Nichols, C. M., S. G. Lardiere, J. P. Bowman, P. D. Nichols, J. A. E. Gibson, and J. Guezennec: 2005, 'Chemical characterization of exopolysaccharides from Antarctic marine bacteria'. *Microbial Ecology* **49**(4), 578–589.

Obayashi, Y. and S. Suzuki: 2005, 'Proteolytic enzymes in coastal surface seawater: Significant activity of endopeptidases and exopeptidases'. *Limnology and Oceanography* **50**(2), 722–726.

Oliveira, A. M. and B. Kjerfve: 1993, 'Environmental responses of a tropical coastal lagoon system to hydrological variability Mundau-Manguaba, Brazil'. *Estuarine Coastal and Shelf Science* **37**(6), 575–591.

Oliver, M. J. and A. J. Irwin: 2008, 'Objective global ocean biogeographic provinces'. *Geophysical Research Letters* **35**(15), –.

Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace, and D. A. Stahl: 1986, 'Microbial ecology and evolution a ribosomal RNA approach'. *Annual Review of Microbiology* **40**, 337–365.

Orellana, M. V., E. J. Lessard, E. Dycus, W. C. Chin, M. S. Foy, and P. Verdugo: 2003, 'Tracing the source and fate of biopolymers in seawater: application of an immunological technique'. *Marine Chemistry* **83**(1-2), 89–99.

O'Sullivan, L. A., K. E. Fuller, E. M. Thomas, C. M. Turley, J. C. Fry, and A. J. Weightman: 2004, 'Distribution and culturability of the uncultivated 'AGG58 cluster' of Bacteroidetes phylum in aquatic environments'. *Fems Microbiology Ecology* **47**(3), 359–370.

O'Sullivan, L. A., A. J. Weightman, and J. C. Fry: 2002, 'New degenerate cytophaga-flexibacter-bacteroides-specific 16S ribosomal DNA-targeted oligonucleotide probes reveal high bacterial diversity in river taff epilithon'. *Applied and Environmental Microbiology* **68**(4), 2093–2093.

Ouverney, C. C. and J. A. Fuhrman: 1999, 'Combined microautoradiography-16S rRNA probe technique for determination of radioisotope uptake by specific microbial cell types in situ'. *Applied and Environmental Microbiology* **65**(4), 1746–1752.

Pace, N. R.: 2009, 'Mapping the Tree of Life: Progress and Prospects'. *Microbiology and Molecular Biology Reviews* **73**(4), 565–576.

Paquette, R. G., R. H. Bourke, J. F. Newton, and W. F. Perdue: 1985, 'The East Greenland polar front in autumn'. *Journal of Geophysical Research-Oceans* **90**(Nc3), 4866–4882.

Pedros-Alio, C.: 2006, 'Marine microbial diversity: can it be determined?'. *Trends in Microbiology* **14**(6), 257–263.

Pedrotti, M. L., S. Beauvais, M. E. Kerros, K. Iversen, and F. Peters: 2009, 'Bacterial colonization of transparent exopolymeric particles in mesocosms under different turbulence intensities and nutrient conditions'. *Aquatic Microbial Ecology* **55**(3), 301–312.

Pernthaler, A., J. Pernthaler, and R. Amann: 2002, 'Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria'. *Applied and Environmental Microbiology* **68**(6), 3094–3101.

Pernthaler, J.: 2005, 'Predation on prokaryotes in the water column and its ecological implications'. *Nature Reviews Microbiology* **3**(7), 537–546.

Pernthaler, J. and R. Amann: 2005, 'Fate of heterotrophic microbes in pelagic habitats: Focus on populations'. *Microbiology and Molecular Biology Reviews* **69**(3), 440–461.

Pernthaler, J., A. Pernthaler, and R. Amann: 2003, 'Automated enumeration of groups of marine picoplankton after fluorescence in situ hybridization'. *Applied and Environmental Microbiology* **69**(5), 2631–2637.

Pernthaler, J., E. Zollner, F. Warnecke, and K. Jurgens: 2004, 'Bloom of filamentous bacteria in a mesotrophic lake: Identity and potential controlling mechanism'. *Applied and Environmental Microbiology* **70**(10), 6272–6281.

Pinhassi, J., F. Azam, J. Hemphala, R. A. Long, J. Martinez, U. L. Zweifel, and A. Hagstrom: 1999, 'Coupling between bacterioplankton species composition, population dynamics, and organic matter degradation'. *Aquatic Microbial Ecology* **17**(1), 13–26.

Pinhassi, J., M. M. Sala, H. Havskum, F. Peters, O. Guadayol, A. Malits, and C. L. Marrase: 2004, 'Changes in bacterioplankton composition under different phytoplankton regimens'. *Applied and Environmental Microbiology* **70**(11), 6753–6766.

Podell, S. and T. Gaasterland: 2007, 'DarkHorse: a method for genome-wide prediction of horizontal gene transfer'. *Genome Biology* **8**(2), –.

Pomeroy, L., P. Williams, F. Azam, and F. Hobbie: 2007, 'The Microbial Loop'. *Oceanography* **28-33**.

Pomeroy, L. R.: 1974, 'Oceans Food Web, a Changing Paradigm'. *Bioscience* **24**(9), 499–504.

Pommier, T., B. Canback, L. Riemann, K. H. Bostrom, K. Simu, P. Lundberg, A. Tunlid, and A. Hagstrom: 2007, 'Global patterns of diversity and community structure in marine bacterioplankton'. *Molecular Ecology* **16**(4), 867–880.

Probst, A. J., C. Hertel, L. Richter, L. Wassill, W. Ludwig, and W. P. Hammes: 1998, '*Staphylococcus condimenti* sp. nov., from say sauce mash, and S*taphylococcus carnosus* (Schleifer and Fischer 1982) subsp. utilis subsp. nov.'. *International Journal of Systematic Bacteriology* **48**, 651–658.

Prosser, J. I., B. J. M. Bohannan, T. P. Curtis, R. J. Ellis, M. K. Firestone, R. P. Freckleton, J. L. Green, L. E. Green, K. Killham, J. J. Lennon, A. M. Osborn, M. Solan, C. J. van der Gast, and J. P. W. Young: 2007, 'Essay - The role of ecological theory in microbial ecology'. *Nature Reviews Microbiology* **5**(5), 384–392.

Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. G. Ludwig, J. Peplies, and F. O. Glockner: 2007, 'SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB'. *Nucleic Acids Research* **35**(21), 7188–7196.

Quast, C.: 2006, 'MicHanThi-design and implementation of a system for the prediction of gene functions in genome annotation projects'. Diploma, University of Bremen.

Ramette, A. and J. M. Tiedje: 2007, 'Biogeography: An emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution'. *Microbial Ecology* **53**(2), 197–207.

Rappe, M. S., S. A. Connon, K. L. Vergin, and S. J. Giovannoni: 2002, 'Cultivation of the ubiquitous SAR11 marine bacterioplankton clade'. *Nature* **418**(6898), 630–633.

Rappe, M. S., P. F. Kemp, and S. J. Giovannoni: 1997, 'Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina'. *Limnology and Oceanography* **42**(5), 811–826.

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

Ravenschlag, K., K. Sahm, and R. Amann: 2001, 'Quantitative molecular analysis of the microbial community in marine Arctic sediments (Svalbard)'. *Applied and Environmental Microbiology* **67**(1), 387–395.

Rawlings, N. D., F. R. Morton, C. Y. Kok, J. Kong, and A. J. Barrett: 2008, 'MEROPS: the peptidase database'. *Nucleic Acids Research* **36**, D320–D325.

Reeves, A. R., G. R. Wang, and A. A. Salyers: 1997, 'Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron'. *Journal of Bacteriology* **179**(3), 643–649.

Reichenbach, H.: 2006, 'The Order *Cytophagales*'. In: M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer, and E. Stackebrandt (eds.): *The Prokaryotes*, Vol. 7. New York: Springer-Verlag, pp. 549–590.

Reichenbach, H. and O. Weeks: 1981, *The Flavobacterium-Cytophaga group*. Weinheim, Germany: Verlag Chemie.

Riemann, L., G. F. Steward, and F. Azam: 2000, 'Dynamics of bacterial community composition and activity during a mesocosm diatom bloom'. *Applied and Environmental Microbiology* **66**(2), 578–587.

Riesenfeld, C. S., P. D. Schloss, and J. Handelsman: 2004, 'Metagenomics: Genomic analysis of microbial communities'. *Annual Review of Genetics* **38**, 525–552.

Rink, B., T. Martens, D. Fischer, A. Lemke, H. P. Grossart, M. Simon, and T. Brinkhoff: 2008, 'Short-term dynamics of bacterial communities in a tidally affected coastal ecosystem'. *Fems Microbiology Ecology* **66**(2), 306–319.

Robinson, C. and P. l. Williams: 2005, 'Respiration and its measurement in surface marine waters.'. In: P. A. del Giorgio and P. J. l. B. Williams (eds.): *Respiration in Aquatic Ecosystems*. Oxford: Oxford University Press, pp. 147–180.

Rocap, G., F. W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N. A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W. R. Hess, Z. I. Johnson, M. Land, D. Lindell, A. F. Post, W. Regala, M. Shah, S. L. Shaw, C. Steglich, M. B. Sullivan, C. S. Ting, A. Tolonen, E. A. Webb, E. R. Zinser, and S. W. Chisholm: 2003, 'Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation'. *Nature* **424**(6952), 1042–1047.

Rossby, T.: 1996, 'The North Atlantic Current and surrounding waters: At the crossroads'. *Reviews of Geophysics* **34**(4), 463–481.

Rossello-Mora, R. and R. Amann: 2001, 'The species concept for prokaryotes'. *Fems Microbiology Reviews* **25**(1), 39–67.

Rudels, B., H. J. Friedrich, and D. Quadfasel: 1999, 'The arctic circumpolar boundary current'. *Deep-Sea Research Part II-Topical Studies in Oceanography* **46**(6-7), 1023–1062.

Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Y. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. E. Venter, K. Li, S. Kravitz, J. F. Heidelberg, T. Utterback, Y. H. Rogers, L. I. Falcon, V. Souza, G. Bonilla-Rosso, L. E.

Eguiarte, D. M. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M. R. Ferrari, R. L. Strausberg, K. Nealson, R. Friedman, M. Frazier, and J. C. Venter: 2007, 'The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific'. *Plos Biology* **5**(3), 398–431.

Salyers, A. A.: 1984, 'Bacteroides of the human lower intestinal tract'. *Annual Review of Microbiology* **38**, 293–313.

Sangkhobol, V. and V. B. D. Skerman: 1981, 'Chitinophaga, a New Genus of Chitinolytic Myxobacteria'. *International Journal of Systematic Bacteriology* **31**(3), 285–293.

Sapp, M., A. S. Schwaderer, K. H. Wiltshire, H. G. Hoppe, G. Gerdts, and A. Wichels: 2007, 'Species-specific bacterial communities in the phycosphere of microalgae?'. *Microbial Ecology* **53**(4), 683–699.

Schattenhofer, M., B. M. Fuchs, R. Amann, M. V. Zubkov, G. A. Tarran, and J. Pernthaler: 2009, 'Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean'. *Environmental Microbiology* **11**(8), 2078–2093.

Shipman, J. A., J. E. Berleman, and A. A. Salyers: 2000, 'Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*'. *Journal of Bacteriology* **182**(19), 5365–5372.

Sievert, S. M., J. Kuever, and G. Muyzer: 2000, 'Identification of 16S ribosomal DNA-defined bacterial populations at a shallow submarine hydrothermal vent near Milos Island (Greece)'. *Applied and Environmental Microbiology* **66**(7), 3102–3109.

Simon, M., F. O. Glockner, and R. Amann: 1999, 'Different community structure and temperature optima of heterotrophic picoplankton in various regions of the Southern Ocean'. *Aquatic Microbial Ecology* **18**(3), 275–284.

Smith, C., E. Rocha, and B. Paster: 2006, 'The medically important *Bacteroides* spp. in health disease'. In: M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer, and E. Stackebrandt (eds.): *The Prokaryotes*, Vol. 7. New York: Springer-Verlag, second edition, pp. 381–427.

Smith, D. C., M. Simon, A. L. Alldredge, and F. Azam: 1992, 'Intense Hydrolytic Enzyme-Activity on Marine Aggregates and Implications for Rapid Particle Dissolution'. *Nature* **359**(6391), 139–142.

Smith, W. O., L. A. Codispoti, D. M. Nelson, T. Manley, E. J. Buskey, H. J. Niebauer, and G. F. Cota: 1991, 'Importance of *Phaeocystis* blooms in the high-latitude ocean carbon-cycle'. *Nature* **352**(6335), 514–516.

Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl: 2006, 'Microbial diversity in the deep sea and the underexplored "rare biosphere"'. *Proceedings of the National Academy of Sciences of the United States of America* **103**(32), 12115–12120.

Soltani, M., B. L. Munday, and C. M. Burke: 1996, 'The relative susceptibility of fish to infections by *Flexibacter columnaris* and *Flexibacter maritimus*'. *Aquaculture* **140**(3), 259–264.

Sorokin, Y. I., P. Y. Sorokin, O. Giovanardi, and L. DallaVenezia: 1996, 'Study of the ecosystem of the lagoon of Venice, with emphasis on anthropogenic impact'. *Marine Ecology-Progress Series* **141**(1-3), 247–261.

Staley, J. T. and A. Konopka: 1985, 'Measurement of insitu activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats'. *Annual Review of Microbiology* **39**, 321–346.

Stingl, U., H. J. Tripp, and S. J. Giovannoni: 2007, 'Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site'. *Isme Journal* **1**(4), 361–371.

Suttle, C. A.: 2005, 'Viruses in the sea'. *Nature* **437**(7057), 356–361.

Suzuki, M., Y. Nakagawa, S. Harayama, and S. Yamamoto: 2001, 'Phylogenetic analysis and taxonomic study of marine *Cytophaga*-like bacteria: proposal for *Tenacibaculum* gen. nov with *Tenacibaculum maritimum* comb. nov and *Tenacibaculum ovolyticum* comb. nov., and description of T*enacibaculum mesophilum* sp nov and *Tenacibaculum amylolyticum* sp nov.'. *International Journal of Systematic and Evolutionary Microbiology* **51**, 1639–1652.

Tanoue, E., S. Nishiyama, M. Kamo, and A. Tsugita: 1995, 'Bacterial membranes: possible source of a major dissolved protein in seawater'. *Geochimica Et Cosmochimica Acta* **59**(12), 2643–2648.

Uehara, T., K. Suefuji, T. Jaeger, C. Mayer, and J. T. Park: 2006, 'MurQ etherase is required by Escherichia coli in order to metabolize anhydro-N-acetylmuramic acid obtained either from the environment or from its own cell wall'. *Journal of Bacteriology* **188**(4), 1660–1662.

van Leewenhoeck, A.: 1677, 'Observations, communicated to the publisher by Mr. Antony van Leewenhoeck, in a Dutch letter of the 9th of Octob. 1676. Here English'd: concerning little animals by him observed in rain-well-sea and snow water; as also in water wherein pepper had lain infused'. *Philosophical Transactions* **12**, 821–831.

Warren, R. A. J.: 1996, 'Microbial hydrolysis of polysaccharides'. *Annual Review of Microbiology* **50**, 183–212.

Waterbury, J., S. Watson, R. Guillard, and R. Brand: 1979, 'Wide-spread occurrence of a unicellular, marine planktonic, cyanobacterium'. *Nature* **277**, 293–294.

Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Truper: 1987, 'Report of the Ad hoc committee on reconciliation of approaches to bacterial systematics'. *International Journal of Systematic Bacteriology* **37**(4), 463–464.

Weiss, M. S., U. Abele, J. Weckesser, W. Welte, E. Schiltz, and G. E. Schulz: 1991, 'Molecular architecture and electrostatic properties of a bacterial porin'. *Science* **254**(5038), 1627–1630.

Weller, R., F. O. Glockner, and R. Amann: 2000, '16S rRNA-targeted oligonucleotide probes for the in situ detection of members of the phylum *Cytophaga- Flavobacterim-Bacteroides*'. *Systematic and Applied Microbiology* **23**(1), 107–114.

West, N. J., I. Obernosterer, O. Zemb, and P. Lebaron: 2008, 'Major differences of bacterial diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom in the Southern Ocean'. *Environmental Microbiology* **10**(3), 738–756.

Whitman, W. B., D. C. Coleman, and W. J. Wiebe: 1998, 'Prokaryotes: The unseen majority'. *Proceedings of the National Academy of Sciences of the United States of America* **95**(12), 6578–6583.

Woebken, D., B. A. Fuchs, M. A. A. Kuypers, and R. Amann: 2007a, 'Potential interactions of particle-associated anammox bacteria with bacterial and archaeal partners in the Namibian upwelling system'. *Applied and Environmental Microbiology* **73**(14), 4648–4657.

Woebken, D., P. Lam, M. M. M. Kuypers, S. W. A. Naqvi, B. Kartal, M. Strous, M. S. M. Jetten, B. M. Fuchs, and R. Amann: 2008, 'A microdiversity study of anammox bacteria reveals a novel *Candidatus Scalindua* phylotype in marine oxygen minimum zones'. *Environmental Microbiology* **10**(11), 3106–3119.

Woebken, D., H. Teeling, P. Wecker, A. Dumitriu, I. Kostadinov, E. F. DeLong, R. Amann, and F. O. Glockner: 2007b, 'Fosmids of novel marine Planctomycetes from the Namibian and Oregon coast upwelling systems and their cross-comparison with planctomycete genomes'. *Isme Journal* **1**(5), 419–435.

Woese, C. R. and G. E. Fox: 1977, 'Phylogenetic structure of prokaryotic domain primary kingdoms'. *Proceedings of the National Academy of Sciences of the United States of America* **74**(11), 5088–5090.

Woodgate, R. A., E. Fahrbach, and G. Rohardt: 1999, 'Structure and transports of the East Greenland Current at 75 degrees N from moored current meters'. *Journal of Geophysical Research-Oceans* **104**(C8), 18059–18072.

Woyke, T., G. Xie, A. Copeland, J. M. Gonzalez, C. Han, H. Kiss, J. H. Saw, P. Senin, C. Yang, S. Chatterji, J. F. Cheng, J. A. Eisen, M. E. Sieracki, and R. Stepanauskas: 2009, 'Assembling the marine metagenome, one cell at a time'. *PLoS One* **4**(4), e5299.

Xie, G., D. C. Bruce, J. F. Challacombe, O. Chertkov, J. C. Detter, P. Gilna, C. S. Han, S. Lucas, M. Misra, G. L. Myers, P. Richardson, R. Tapia, N. Thayer, L. S. Thompson, T. S. Brettin, B. Henrissat, D. B. Wilson, and M. J. McBride: 2007, 'Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*'. *Applied and Environmental Microbiology* **73**(11), 3536–3546.

Xu, J., M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper, and J. I. Gordon: 2003, 'A genomic view of the human *Bacteroides thetaiotaomicron* symbiosis'. *Science* **299**(5615), 2074–2076.

Yamamoto, T., S. Kajiura, Y. Hirai, and T. Watanabe: 1994, '*Capnocytophaga haemolytica* sp. nov and *Capnocytophaga granulosa* sp. nov, from Human Dental Plaque'. *International Journal of Systematic Bacteriology* **44**(2), 324–329.

Yan, S., B. M. Fuchs, S. Lenk, J. Harder, J. Wulf, N. Z. Jiao, and R. Arnann: 2009, 'Biogeography and phylogeny of the NOR5/OM60 clade of *Gammaproteobacteria*'. *Systematic and Applied Microbiology* **32**(2), 124–139.

Yarza, P., M. Richter, J. Peplies, J. Euzeby, R. Amann, K. H. Schleifer, W. Ludwig, F. O. Glockner, and R. Rossello-Mora: 2008, 'The All-Species

Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains'. *Systematic and Applied Microbiology* **31**(4), 241–250.

Yoon, J. H., S. J. Kang, and T. K. Oh: 2006, '*Polaribacter dokdonensis* sp nov., isolated from seawater'. *International Journal of Systematic and Evolutionary Microbiology* **56**, 1251–1255.

Zeder, M., S. Peter, T. Shabarova, and J. Pernthaler: 2009, 'A small population of planktonic *Flavobacteria* with disproportionally high growth during the spring phytoplankton bloom in a prealpine lake'. *Environmental Microbiology* **11**(10), 2676–2686.

ZoBell, C. E.: 1946, *Marine microbiology: A monograph on hydrobacteriology*. Waltham, Massachusetts: Chronica Botanica.

Zwart, G., B. C. Crump, M. P. K. V. Agterveld, F. Hagen, and S. K. Han: 2002, 'Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers'. *Aquatic Microbial Ecology* **28**(2), 141–155.

Part II

Manuscripts

List of manuscripts presented in this thesis

Contributions to the manuscripts and publications presented in this thesis

I Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean
Paola R Gómez-Pereira, Bernhard M Fuchs, Cecilia Alonso, Matthew J Oliver, Justus EE van Beusekom, and Rudolf Amann
International Society of Microbial Ecology (ISME) Journal; 2010, 4, 472-487.
P.R.G.P, B.M.F. and R.A. designed research; P.R.G.P. realized experiments; C.A., J.E.E.vB, M.J.O. contributed experimental data; P.R.G.P. wrote the manuscript with editorial assistance of B.M.F. and R.A.

II Genomic content of uncultured *Bacteroidetes* from contrasting oceanic provinces in the North Atlantic Ocean
Paola R Gómez-Pereira, Margarete Schüler, Bernhard M Fuchs, Christin Bennke, Valerie Barbe, and Rudolf Amann
Manuscript in preparation. Intended as article in the ISME journal. P. R.G.P., B.M.F., M.S., and R.A. designed research; P.R.G.P. realized experiments; V.B. and C.B. contributed experimental data; P.R.G.P. wrote the manuscript with editorial assistance of M.S.

III Multilevel diversity analysis of bacterial communities along the environmental gradient Río de la Plata–South Atlantic Ocean

Cecilia Alonso, Paola Gómez-Pereira, Alban Ramette, Leonardo Ortega, Bernhard M. Fuchs, and Rudolf Amann

Manuscript submitted to Aquatic Microbial Ecology

C.A. designed research. P.R.G.P. contributed in experimental design and realized experiments: DNA extraction, 16S rRNA sequencing and quantification of *Bacteroidetes* by FISH. P.R.G.P. contributed in writing the manuscript.

IV Bacterioplankton diversity and community composition in the southern Lagoon of Venice

Francesca Simonato, Paola R. Gómez-Pereira, Bernhard M. Fuchs, and Rudolf Amann

Systematic and Applied Microbiology; 2010,33, 128-138.

F.S. designed and realized experiments. P.R.G.P. contributed experimental data: diversity analysis of 16S rRNA clone libraries and *Bacteroidetes* phylogenetic analysis. P.R.G.P. contributed in writing and revising the manuscript.

Manuscript I

Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean

Paola R Gómez-Pereira, Bernhard M Fuchs, Cecilia Alonso,
Matthew J Oliver, Justus EE van Beusekom, and Rudolf Amann

International Society of Microbial Ecology (ISME) Journal
2010, 4, 472-487

The ISME Journal (2010) 4, 472–487

www.nature.com/ismej

ORIGINAL ARTICLE

Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean

Paola R Gómez-Pereira[1], Bernhard M Fuchs[1], Cecilia Alonso[2], Matthew J Oliver[3], Justus EE van Beusekom[4] and Rudolf Amann[1]
[1]*Department of Molecular Ecology, Max-Planck Institute for Marine Microbiology, Bremen, Germany;* [2]*Departamento de Microbiología, Instituto de Investigaciones Biológicas Clemente Estable, Montevideo, Uruguay;* [3]*College of Earth, Ocean and Environment, University of Delaware, Lewes, DE, USA and* [4]*Coastal Ecology Department, Alfred Wegener Institute for Polar and Marine Research, Hafenstrasse, List/Sylt, Germany*

Members of the class *Flavobacteria* in the phylum *Bacteroidetes* are among the most abundant picoplankton in coastal and polar oceans. Their diversity is high in marine waters. However, quantitative information about distribution patterns of flavobacterial clades is scarce. We analyzed the diversity and clade-specific abundances of individual *Flavobacteria* in different oceanic provinces in the North Atlantic Ocean. Samples were taken along the 30°W meridian between the East Greenland current and the North Atlantic subtropical gyre. Comparative sequence analysis of 16S ribosomal RNA (rRNA) gene libraries revealed high diversity and significant spatial variability within the class *Flavobacteria*. Published and newly designed oligonucleotide probes were used to enumerate eleven flavobacterial clades by catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH). We found that different provinces harbor distinct flavobacterial communities. Clade DE2 accounted for a substantial fraction of total *Flavobacteria* only in the Polar Biome (BPLR), whereas the VISION clades VIS1 and VIS4 significantly increased in the Arctic (ARCT) province. Members of the genus *Polaribacter* were the most abundant clade in all the water masses analyzed, with highest absolute numbers in BPLR and ARCT. We improved the CARD-FISH protocol to quantify the rare clades VIS2, VIS3, VIS5 and VIS6, which were present in abundances below 0.5%. They all showed pronounced regional distribution patterns. Microscopic analysis proved a specific enrichment of *Flavobacteria* in the phycosphere of nanophytoplankton of BPLR and ARCT. Our results suggest that different marine flavobacterial clades have distinct niches and different life strategies.
The ISME Journal (2010) **4,** 472–487; doi:10.1038/ismej.2009.142; published online 7 January 2010
Subject Category: microbial population and community ecology
Keywords: CARD-FISH; bacterioplankton; marine; probe design; rare populations; nanophytoplankton

Introduction

Members of the *Bacteroidetes* phylum, formerly also known as *Cytophaga–Flavobacteria–Bacteroides* cluster, constitute one of the major groups of picoplankton (Glöckner *et al.*, 1999; Kirchman, 2002). *Bacteroidetes* thrive in a variety of marine systems, including hydrothermal vents (Sievert *et al.*, 2000; Kormas *et al.*, 2006), coastal sediments (Llobet-Brossa *et al.*, 1998) and waters (Eilers *et al.*, 2001; O'Sullivan *et al.*, 2004) and in open ocean waters (Simon *et al.*, 1999; Abell and Bowman, 2005; Schattenhofer *et al.*, 2009). Taxonomically, the phylum *Bacteroidetes* consists of the classes *Bacteroidia*, *Flavobacteria, Sphingobacteria* and *Cytophagia*, the latter previously included in the class *Sphingobacteria* (Ludwig *et al.*, 2008). There is more and more evidence, for example, from comparative 16S ribosomal RNA (rRNA) sequence analysis, that *Flavobacteria* is the dominant class of *Bacteroidetes* in marine picoplankton (for example, Kirchman *et al.*, 2003; Alonso *et al.*, 2007; Chen *et al.*, 2008). High abundance has been linked to cold waters (Simon *et al.*, 1999; Abell and Bowman, 2005), phytoplankton blooms (Simon *et al.*, 1999), photic zone (Schattenhofer *et al.*, 2009) and upwelling systems (Alonso-Saez *et al.*, 2007). This distribution suggests a preference for more productive conditions. However, *Flavobacteria* have also been detected in oligotrophic marine surface waters in significant numbers (Schattenhofer *et al.*, 2009).

Sequence-based studies have shown the dominance of particular flavobacterial phylotypes during algal blooms, whereas other clades of *Flavobacteria*

Correspondence: B Fuchs, Department of Molecular Ecology, Max-Planck Institute for Marine Microbiology, Celsiusstr 1, 28359 Bremen, Germany.
E-mail: bfuchs@mpi-bremen.de
Received 24 August 2009; revised 12 November 2009; accepted 22 November 2009; published online 7 January 2010

dominated before the blooms (Riemann *et al.*, 2000; O'Sullivan *et al.*, 2004; Pinhassi *et al.*, 2004). In addition, *Flavobacteria* could be localized in the phycosphere (Rooney-Varga *et al.*, 2005; Sapp *et al.*, 2007), which is the space immediately surrounding an algal cell (Bell and Mitchell, 1972). Moreover, this association might depend on the composition and dynamics of the phytoplankton and flavobacterial communities (Grossart *et al.*, 2005; Rooney-Varga *et al.*, 2005). However, most of these studies were not combined with a quantitative technique and, therefore, there is no information about the clade-specific population sizes.

Cell abundances and distribution patterns of members of the phylum *Bacteroidetes* in marine picoplankton have been obtained by fluorescence *in situ* hybridization (FISH) using probes CF319a (Manz *et al.*, 1996) and CFB560 (O'Sullivan *et al.*, 2002); the specificity of these two probes has recently been reviewed (Amann and Fuchs, 2008). Evidence for distribution patterns of distinct flavobacterial clades, however, is sparse. Examples include FISH-based quantifications of the clade *Cytophaga marinoflava-latercula*, the uncultured cluster DE2 and the genus *Polaribacter*. *Cytophaga marinoflava-latercula* was shown to constitute up to 6% of total picoplankton population during spring and summer, off the island of Helgoland, in the German Bight of the North Sea (Eilers *et al.*, 2001). Clade DE2, originally retrieved from the Delaware Estuary, was found to account for up to 10% of picoplankton population in this estuary as well as in the Chukchi Sea, Arctic Ocean (Kirchman *et al.*, 2003; Malmstrom *et al.*, 2007). Members of the genus *Polaribacter* have been shown to form prominent fractions in polar oceans (Brinkmeyer *et al.*, 2003; Malmstrom *et al.*, 2007).

This study was performed in the North Atlantic between the East Greenland current and the North Atlantic subtropical gyre (NAST), an area characterized by annual spring blooms known as one of the most conspicuous seasonal events in the world ocean (for example, Ducklow and Harris, 1993). The spring bloom starts north of about 40°N latitude and subsequently develops northwards (Longhurst, 1998). The concentration of chlorophyll *a* remains high during summer and it gradually returns to winter concentrations (Henson *et al.*, 2006). During summer, there is a shift from dominance by the $>20\,\mu m$ chlorophyll fraction to dominance by the $<20\,\mu m$ chlorophyll fraction (Gifford *et al.*, 1995). In contrast, the NAST is an oligotrophic area with low levels of primary production (Longhurst, 1995). Thus, the North Atlantic Ocean is characterized by strong gradients that could influence the distribution patterns of picoplankton groups.

The objective of this study is to quantify flavobacterial clades that inhabit these contrasting oceanic provinces in the northern North Atlantic Ocean and to identify environmental factors that control their abundance. We analyzed samples of four oceanic provinces as defined by Longhurst (1995), which encompassed nine significantly different water masses according to the classification of Oliver and Irwin (2008). We hypothesized that: (i) members of the class *Flavobacteria* are present throughout the transect, (ii) general flavobacterial abundances are higher at the more northern stations where temperature is low and nutrient concentration high, and in the euphotic zone in which phytoplankton abundance is high, (iii) the flavobacterial diversity varies locally, and (iv) different flavobacterial clades prevail in different water masses and are correlated with distinct environmental parameters (for example, temperature, nutrient concentration and phytoplankton abundance).

Materials and methods

Samples were taken along a transect in the North Atlantic Ocean from 66°39.27′N; 29°36.65′W (station 2 = S2) to 34°24.87′N; 28°28.90′W (S19) during the VISION cruise (diVersIty, Structure and functIONing) MSM03/01 on board the research vessel Maria S. Merian from September 21 to September 30, 2006 (Figure 1a, Table 1). Along this transect, each station was sampled at 12 depths, from 10 m down to 250 m or 500 m. Samples were collected with a rosette of 20-l Niskin bottles mounted on a conductivity–temperature–density profiler. Chlorophyll *a* fluorescence was measured on board by a flow through fluorometer (microFlu-Chl, TriOS Optical Sensors GmbH, Oldenburg, Germany). Total picoplankton and nanophytoplankton populations were enumerated after fixation with particle-free formaldehyde solution (37% w/v, Fluka, Taufkirchen, Germany; final concentration, 1% v/v) and staining with SYBR Green (1:10 000 dilution of stock; Sigma-Aldrich, Taufkirchen, Germany) (Marie *et al.*, 1997) using a FACScalibur flow cytometer equipped with an air-cooled laser providing blue light at 488 nm as described previously (Tarran *et al.*, 2006). Profiles were grouped according to oceanic provinces defined by two classification systems (Longhurst, 1998; Oliver and Irwin 2008). The Longhurstian provinces used in this study include the Boreal Polar (BPLR), the Arctic (ARCT), the North Atlantic Drift (NADR) and the North Atlantic Subtropical East (NAST) provinces. Oliver and Irwin (2008) delineated provinces by objectively clustering satellite observed thermal and optical properties. The provinces defined by Oliver and Irwin (2008) were computed on the basis of the September 2006 mean of the Advanced Very High Resolution Radiometer (AVHRR) sea surface temperature and the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) water leaving radiance.

Nutrient analysis

Water samples for nutrients analysis were filtered directly after sampling through 0.45-μm in-line

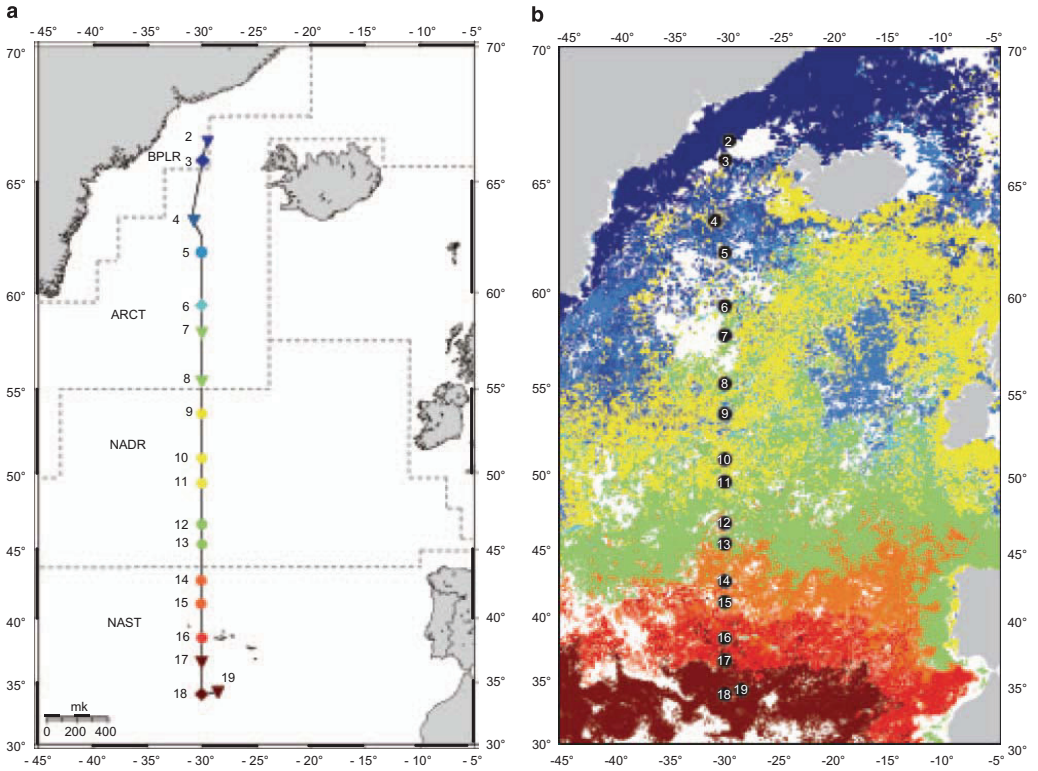


Figure 1 (**a**) Map indicating the VISION cruise track and sampled stations S2–S19 and boundaries of the oceanic provinces Boreal Polar (BPLR), Arctic (ARCT), North Atlantic Drift (NADR) and North Atlantic Subtropical (NAST) provinces. Color of symbols represents different water masses as showed in Figure 1b. Symbol types represent analyses done in the sampled stations: Spheres (○) indicate stations where only CF319a counts were performed; triangles (▼) indicate stations where flavobacterial clade-specific and CF319a counts were performed; and rhomboids (◆) indicate stations where 16S rRNA gene clone libraries and counts were performed. (**b**) Significant water masses determined by the September 2006 mean of AVHRR sea surface temperature and SeaWiFS water leaving radiance according to Oliver and Irwin (2008). The different colors represent different water mass types.

Table 1 Coordinates, date and time, temperature (temp), salinity (sal), dissolved oxygen (DO) in the 20-m samples of those stations (S) from which the 16S rRNA gene libraries were constructed

Province	*S*	*Coordinates*		*Date; time (UTC)*	*Temp (°C)*	*Sal (PSU)*	*DO (ml l^{-1})*
		Latitude	*Longitude*				
BPLR	3	65 °52,64′N	29 °56,54′W	22.09.06; 19:00	0.7	33.0	7.7
ARCT	6	59 °20,87′N	29 °59,89′W	24.09.06; 08:00	10.9	35.0	5.9
NAST	18	34 °04,43′N	30 °00,09′W	30.09.06; 08:00	23.5	36.4	4.7

Abbreviations: ARCT, Arctic; BPLR, Boreal Polar; NAST, North Atlantic subtropical province; UTC, Coordinated Universal Time; PSU, practical salinity units.
Time indicated is the time when sampling procedure started.

filters attached to a 60-ml pre-cleaned syringe into two 12-ml polystyrole tubes. Samples were stored at 4 °C (dissolved silicate) or −80 °C (NH_4^+, PO_4^{3-}, NO_3^- and NO_2^-). The samples were spectrophotometrically measured with a continuous-flow analyzer using standard AA3 methods (Seal Analytical, Norderstedt, Germany) using a variant of the method of Grasshoff *et al.* (1983). The s.d. values were calculated for repeated measurements of five surface samples (20 m) and five deep samples (200 m), and

were for PO_4^{3-} and NO_2^- 0.01 μM at all levels, for NO_{2+3} 0.03 μM at the 1 μM level and 0.07 μM at the 18 μM level, and for NH_4^+ 0.07–0.08 μM at all levels.

DNA extraction and 16S rRNA gene libraries

The 16S rRNA gene libraries were constructed from the 20-m samples at S3, S6 and S18 (Figure 1). Approximately 100-l samples were pooled from five Niskin bottles, pre-filtered through a 10-μM steel mesh cartridge (Wolftechnik, Weil der Stadt, Germany), and microorganisms were collected on cellulose acetate filters (142 mm diameter, 0.2 μM pore size; Sartorius, Goettingen, Germany). Filters were stored at −80 °C. DNA was extracted from the 0.2-μM filters according to procedure used by Zhou *et al.* (1996).

Almost complete bacterial 16S rRNA genes were amplified from the extracted DNA using the primer pairs GM3F, 5′-AGAGTTTGATCMTGGC-3′, and GM4R, 5′-TACCTTGTTACGACTT-3′ (Muyzer *et al.*, 1995). The PCR was run at the following cycling conditions: initial denaturation at 96 °C for 3 min, followed by 20 cycles consisting of denaturation (96 °C for 1 min), annealing (48 °C for 2 min) and extension (72 °C for 3 min) and a final step at 72 °C for 15 min. The PCR products were purified using the QIAquick PCR purification kit (QIAGEN, Hilden, Germany), ligated using the TOPO TA cloning kit (Invitrogen, Karlsruhe, Germany) or pGEM-T-Easy (Promega, Mannheim, Germany) and cloned into high-efficiency competent cells of *Escherichia coli* (JM109). Libraries were screened for *Bacteroidetes* containing clones by PCR with the primer pair CF315-F, 5′-ACKGGYACTGAGAYAC GG-3′; and CF967-R, 5′-GGTAAGGTTCCTCGCGTA-3′ (Chen *et al.*, 2006). The PCR reaction and cycling conditions were as described above but with 30 cycles. Annealing temperature was 61 °C after optimization of the primers to avoid false positives. Plasmids of positive clones were isolated with Montage plasmid $MiniPrep_{96}$ kit (Millipore, Eschborn, Germany) and 50–100 ng of template was used per sequencing reaction performed with ABI BigDye and ABI PRISM3100 Genetic Analyser (Applied Biosystems, Darmstadt, Germany). All sequences were checked for chimera origin using the online tool Bellerophon from Greengenes (Huber *et al.*, 2004; DeSantis *et al.*, 2006) and possible chimera sequences were excluded from further analysis. The 16S rRNA gene sequences of this study were deposited in GeneBank under the accession numbers FN433282–FN433465.

Phylogenetic reconstruction

Sequences were analyzed with the software package ARB (Ludwig *et al.*, 2004) using Silva database release 96, October 2008 (Pruesse *et al.*, 2007). Phylogenetic reconstruction was done following the 'standard operating procedure for phylogenetic inference' (Peplies *et al.*, 2008). Operational taxonomic units were calculated with DOTUR (Distance-Based OTU and Richness; Schloss and Handelsman, 2005) using the furthest neighbor algorithm from a distance matrix generated in ARB. From the DOTUR results, we identified all the sequences that contributed to each operational taxonomic unit. These data were imported to EstimateS (Colwell, 2005) and similarity indexes calculated according to Chao *et al.* (2005). Coverage index of the clone libraries was calculated using the formula $C = (1-n/N) * 100$, where N represents the number of all clones carrying a *Bacteroidetes* insert and n the number of phylotypes at 99% identity appearing only once in the library (Good 1953; Kemp and Aller 2004).

Catalyzed reporter deposition (CARD)-FISH

Samples for CARD-FISH were fixed with particle-free formaldehyde solution (37% w/v, Fluka; final concentration, 1% v/v) for 2 h at room temperature. Water samples were filtered with low pressure Hg onto polycarbonate filters (type GTTP; 0.2 μM pore size; 47 mm diameter; Millipore, Eschborn, Germany) in varying volumes. Typically 15–30 ml was filtered from the surface water layer and 50–150 ml from samples below 100 m depth. In addition, 500-ml aliquots were filtered from the surface water layer. The filters were stored frozen at −20 °C for further analyses. The CARD-FISH procedure was performed according to Pernthaler *et al.* (2004) with modifications as published elsewhere (Schattenhofer *et al.*, 2009). Hybridizations were done with horseradish peroxidase-labeled oligonucleotides probes (Biomers, Ulm, Germany) at varying formamide concentrations depending on the probe used (Table 2). Working solutions of probes and competitors (both at 50 ng μl^{-1}) were mixed with hybridization buffer in a 1:1:300 proportion and hybridization was carried out overnight at 46 °C. The probe-delivered horseradish peroxidase was detected with tyramides that were custom labeled with fluorescein (Molecular Probes, Eugene, OR, USA). Subsequently, all cells were stained with 4,6-diamidino-2-phenylindole (1 μg ml^{-1}). Hybridized and 4,6-diamidino-2-phenylindole-stained cells were quantified on an Axioplan II Imaging epifluorescence microscope (Zeiss, Jena, Germany).

The signals in the negative control (NON338) were low, accounting for less than one cell in 10 fields. Differences in clade abundance between provinces were tested by *t*-test on log (x + 1) transformed data. If data did not pass normality and equal variance test, Mann–Whitney–Wilcoxon test was applied.

For the evaluation of *Bacteroidetes* associated with nanophytoplankton cells, hybridizations were performed on 500-ml samples from surface waters using probes, CF319a and SAR11-441, targeting the SAR11 clade (Morris *et al.*, 2002) at 25% formamide. Nanophytoplankton cells were identified by

Table 2 Flavobacterial clade-specific oligonucleotide probes used in this study

Probe	*Target*	*Sequence (5′→3′)*	*FA (%)*	*Reference*
CF319a	*Bacteroidetes*	TGGTCCGTGTCTCAGTAC	35	Manz *et al.* (1996)
POL740	*Polaribacter*	CCCTCAGCGTCAGTACATACGT	35	Malmstrom *et al.* (2007)
CF6-1267	DE2	GAAGATTCGCTCCTCCTC	35	Kirchman *et al.* (2003)
FL197	VIS4	TCATCTCATACCGTAACC	35	This study
Comp FL197_1		TCATCCCCTACCGTAACC		
Comp FL197_2		YCATCTCKTACCGTAACC		
VIS6-814	VIS6	CAGCGAGTGATGATCGTT	15	This study
Comp VIS6-814		CAGCGAGTGATCATCGTT		
VIS5-586	VIS5	CCACTGACTTAGATTCCC	25	This study
Comp VIS5-586_1		CCACTGACTTAGACGCCC		
Comp VIS5-586_2		CCACTGACTTAAATTTCC		
FL845	VIS2	GCTTAGCCACTCATCTAA	15	This study
Comp FL845_1		GCTTAGCCACTCAACTAA		
Comp FL845_2		GCTTAGCCACTCAACCAA		
VIS3-183	VIS3	CTCTAGAAGTGATGCCAC	25	This study
VIS1-575	VIS1	CTTAACAAACAGCCTGCGGACC	35	This study
Comp VIS1-575		CTTAAAAAACAGCCTGCGGACC		
FL443	*Gramella*	TAGATGTGGTTCTTCCCG	65	Alonso unpub
LEE82	*Leeuwenhoenkiella*	ACTCGTCAGCTTCCCGAA	25	This study
Comp LEE82_1		GCTCGCCAGCCTCCCGAA		
Comp LEE82_2		ACTCGTCAGCATCCCGAA		
DOK827	*Dokdonia*	GTCCGAAAACCAAACAACTAG	20	This study
Comp DOK827_1		GTCCGAAAACCACACAGCTAG		
Comp DOK827_2		GUCCGAAAACCGAACAGCTAG		

Abbreviations: Comp, unlabeled competitor oligonucleotides; FA, formamide concentration (v/v) in the hybridization buffer.

epifluorescence microscopy from their pigment fluorescence and an approximate diameter of 5–10 μM. Pictures were taken with a black and white camera (AxioCam MRm; Carl Zeiss, Jena, Germany). The phycosphere, which is defined as the area immediately surrounding nanophytoplankton cells (Bell and Mitchell, 1972), was delimited and measured on the micrographs using the software AxioVision Rel. 4.7.2 (Carl Zeiss). More than 30 nanophytoplankton cells were evaluated in more than 10 images. The CF319a- and SAR11-positive signals were quantified inside the phycosphere, including the nanophytoplankton cell, and outside of it. Ratios of counts per area unit between inside and outside the phycosphere were calculated and used for statistical analysis with the *t*-test.

Testing a modified counting protocol for rare populations

For quantification of rare populations, we enumerated picoplankton populations from filters containing increasing volumes of water samples. Samples were taken at the German North Sea island Helgoland (54°11.3′N; 07°54.0′E) and were filtered on polycarbonate filters (0.2 μM pore size) in increasing amounts from 10, 25, 50, 100, 250 to 400 ml in triplicates for each volume. Hybridizations were performed as described above using probes ROS537, for members of the *Roseobacter* clade (Eilers *et al.*, 2001) at 35% formamide; NOR5-730, for members of the gammaproteobacterial NOR5/OM60 clade (Eilers *et al.*, 2001) at 50% formamide and PLA46, for *Plantomycetes* (Neef *et al.*, 1998) at 30%. The NON338 probe was used as negative control at 35% formamide. Only the probe-conferred fluorescence was counted in 20 microscopic fields for each volume replicate and the absolute abundance of the clade was determined after subtraction of the negative control. To test whether significant differences exit among the different volumes, one-way analysis of variance was applied. When normality or equal variance test were not passed, analysis of variance on ranks was used.

Probe design

Oligonucleotide probes targeting the 16S rRNA of flavobacterial clades were designed with the probe design tool in ARB, as well as unlabeled competitor oligonucleotides (Manz *et al.*, 1992; Table 2; Supplementary Table 1). For optimization of stringency conditions, a series of hybridizations at increasing formamide concentrations (10–50%) were evaluated. The optimal formamide concentration was the highest concentration before signal intensity decreased (Supplementary Table 2). Unlabeled competitor oligonucleotides, which allow a single mismatch discrimination by blocking non-target probe-binding sites (Manz *et al.*, 1992), were used at all times to assure specificity (Supplementary Table 1).

Probe POL740 (Malmstrom *et al.*, 2007) was re-evaluated on the strain *Polaribacter filamentus* (DSMZ13964, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany). The LEE82 and DOK827 probes were tested on *Leeuwenhoekiella* sp. and *Dokdonia* sp. strains,

respectively. The strains were isolated during the cruise and kindly provided by J Harder (unpublished data).

Multi- and univariated statistics on Flavobacteria *abundances*

Multivariate statistical analyses were applied to test the differences in the flavobacterial community among the different water masses. A similarity matrix was calculated with Bray–Curtis coefficient on square root-transformed abundance data. The Bray–Curtis matrix was represented by hierarchical agglomerative cluster analysis using group average as the linkage algorithm. Significance of the grouping in the cluster analysis was tested using similarity profile permutation test at 99% significance level. The similarity profile permutation tests for statistically significant evidence of genuine clusters in an *a priori* unstructured set of samples (Clarke *et al.*, 2008). Multivariate statistics were performed with software Primer v6 (Version 6.1.5, Primer-E, Plymouth, UK; Clarke and Gorley, 2006).

Correlations between environmental variables and flavobacterial abundances were analyzed with Spearman rank correlation test. The test was performed with physicochemical parameters (temperature, salinity, dissolved silicate, NH_4^+ and PO_4^{3-}, which is correlated with NO_3^- and NO_2^-), chlorophyll *a* fluorescence and abundance of nanophytoplankton and of different microphytoplankton classes (dinoflagellates, diatoms and coccolithophores). Microphytoplankton data were kindly provided by J Ufkes and V Smetacek (personal communication). Correlations with counts obtained with probe CF319a were performed with data from surface to 50 m of all stations. The significance level of the correlations was adjusted by the Bonferroni correction for multiple comparisons, according to which the *P*-value is divided by the total number of simultaneous independent comparisons (Legendre and Legendre, 1998). All univariated statistical analyses were performed using the software SigmaStat 3.5 (SYSTAT, Santa Clara, CA, USA).

Results

Characterization of study area

Boundaries of the oceanic provinces as defined by the Longhurstian classification were in good agreement with delimitations according to Oliver and Irwin (2008), based on the satellite-observed thermal and optical properties of the study area during September 2006 (Figure 1b). However, the four Longhurstian provinces could be further separated by the latter approach into nine distinct water masses (Figure 1). The northernmost stations (S2 and S3) were located at the boundary of BPLR and ARCT provinces and satellite-derived water masses confirmed that they represented a separate area. These stations had an average chlorophyll *a* concentration of 0.7 $\mu g\,l^{-1}$, surface water temperature below 3 °C and low salinity (<33 PSU, practical salinity units) (Supplementary Figures 1, 2a and b). Within ARCT province, four water masses (stations S4, S5, S6, S7 + S8) were differentiated with chlorophyll *a* concentrations ranging between 0.3 and 1 $\mu g\,l^{-1}$, and temperature between 10 and 12 °C. Along the NADR province (S9–S13), the temperature gradually increased up to 20 °C. In the southernmost water body (S17–S19) of the NAST province (S14–S19), the temperature reached 24 °C, and chlorophyll *a* had the lowest value of <0.1 $\mu g\,l^{-1}$. The concentrations of phosphate (PO_4^{3-}), nitrate (NO_3^-), nitrite (NO_2^-) and dissolved silicate (Si) in the upper 50 m were higher in the BPLR and ARCT provinces, ranging between 0.2 and 0.6 μM, 2.0 and 6.5 μM, 0.1 and 0.2 μM, and 0.6 and 3.2 μM, respectively. The concentrations decreased along the NADR province to very low concentrations in the NAST with PO_4^{3-}, NO_3^-, NO_2^- <0.06 μM, and Si <0.3 μM. The NH_4^+ concentrations ranged between 0.2 and 0.6 μM in surface layers and decreased below detection limit in deeper waters.

The abundance of picoplankton in the surface mixed layer was higher in the BPLR, ARCT and NADR provinces ($0.4–1.1 \times 10^6$ cells per ml) than in the NAST province ($0.2–0.6 \times 10^6$ cells per ml; Supplementary Figure 2c). The abundance of nanophytoplankton was also higher in the upper mix layer of the BPLR, ARCT and NADR provinces ($4.9–36 \times 10^3$ cells per ml) than in the NAST province ($2.7–8.2 \times 10^3$ cells per ml; Supplementary Figure 2d).

Bacteroidetes diversity and distribution

The abundance of *Bacteroidetes* along the transect was quantified by CARD-FISH using the probe CF319a down to a depth of 500 m (Figure 2; Supplementary Table 3). Maximum counts of 18.2 ± 2.8% of total picoplankton were detected in surface waters of the BPLR province. *Bacteroidetes*

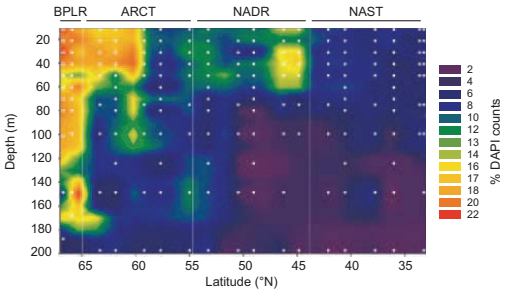


Figure 2 Latitudinal depth distribution of *Bacteroidetes* as detected by probe CF319a (relative abundance, % of 4,6-diamidino-2-phenylindole (DAPI) counts). Oceanic provinces as indicated in Figure 1a. Dots represent the sampling depths.

abundance decreased along ARCT and NADR provinces to a minimum of 7.0 ± 1.1% in the NAST province. Abundances decreased with depth and were between 0.9 and 8% at 200 m. As the abundance of *Bacteroidetes* was highest in the upper 50 m, we focused further analysis on the mixed surface water layer.

As a basis for the subsequent quantification of flavobaterial clades, three 16S rRNA gene libraries were constructed with general bacterial primers from surface water samples of S3 (BPLR), S6 (ARCT) and S18 (NAST) in which *Bacteroidetes* represented 19, 17 and 7% of picoplankton population, respectively. We screened a total of 1632 clones of which 190 carried a *Bacteroidetes* insert. After removal of six putative chimeric sequences, we further analyzed 71 clones from S3 (61 > 1400 nt, 58 unique), 55 from S6 (52 > 1400 nt, 43 unique), and 58 from S18 (55 > 1400 nt, 54 unique). We opted for almost full-length 16S rRNA sequences to facilitate subsequent probe design.

The coverage of the *Bacteroidetes* diversity at 99% sequence identity level was high at S3 and S6 with 79 and 92%, respectively. It was lower at S18 (63%). Moreover, the three libraries had little overlap in their diversity (Supplementary Table 4). The similarity between S3 and S6 was low with indices ranging between 0.1–0.3 and no operational taxonomic units were shared between S3/S6 and S18. Similarity indices at 97% identity confirmed that S3 and S6 were more similar to each other (0.2–0.4) than both were to S18 (0.02–0.05).

Comparative sequence analysis showed that sequences affiliated with the class *Flavobacteria* dominated all three libraries (Figure 3) despite the fact that the primer pair used for screening had a similar coverage of the two classes *Flavobacteria* and *Sphingobacteria* (Chen *et al.*, 2006), including the new class *Cytophagia*. A total of 180 sequences belonged to the class *Flavobacteria*, only three to the class *Sphingobacteria*, one to the class *Cytophagia* and none to the class *Bacteroidia*.

A detailed phylogenetic analysis placed 83% of the flavobacterial sequences (151/180) within the 'marine clade' that comprises genera that have been exclusively isolated from marine habitats (Bowman, 2006). A total of 43 sequences within the marine clade grouped closely with those of cultured genera. For example, S3 contained 27 sequences affiliated with the genus *Polaribacter*. They all had the target site of probe POL740 (Malmstrom *et al.*, 2007). From S18, we obtained eight sequences of *Leeuwenhoekiella* spp. and four of *Dokdonia* spp. For these two genera, probes were designed in this study (LEE82, DOK827; Supplementary Table 5). Another 107 sequences clustered with yet uncultured clades mostly retrieved from various marine pelagic environments, both coastal and open ocean, as well as temperate and polar sites. Sequences related to clade DE2, originally described by sequences retrieved from the Delaware Estuary (Kirchman *et al.*, 2003), were present in S3 (11), S6 (9) and S18 (6). The DE2 clade was subsequently quantified using the probe CF6-1267 (Kirchman *et al.*, 2003).

We also defined new clades on the basis of sequences from the VISION cruise. Clade VIS1 contained 9 and clade VIS4 contained 18 sequences that originated mainly from S6 and S18. The probes VIS1-575 and FL197 were designed to quantify VIS1 and VIS4, respectively. The VIS2 clade includes sequences from both coastal and open ocean environments; this clade was quantified by probe FL845. Clade VIS5, targeted by VIS5-586, contained sequences from S3 and S6, and also sequences from other diverse marine habitats. The VIS3 clade, targeted by VIS3-183, contained 10 sequences from S3 and 8 from S6 that were most closely related to sequences previously retrieved from the North Sea (Alonso *et al.*, 2007). A total of 19 sequences were related to the uncultured lineage 'AGG58' (DeLong *et al.*, 1993; O'Sullivan *et al.*, 2004). Moreover, three sequences from S6 clustered tightly with those of a natural iron fertilization event (West *et al.*, 2008). We named this clade VIS6 and targeted it with probe VIS6-814.

Quantifying rare flavobacterial clades

First FISH analyses using probes for the clades VIS2, VIS3, VIS5 and VIS6 resulted in counts of less than 1 cell per microscopic field, corresponding to relative abundances of 0.5% and lower. Similarly, in the more southern stations, the otherwise abundant clades VIS1, VIS4 and DE2 dropped below 0.5% of total bacteria in abundance. Such infrequent events were below the threshold for confident quantification by FISH. We therefore tested whether this limitation could be overcome by filtering larger volumes resulting in a higher cell density. Probe-specific counts proportionally increased with increasing amounts of filtered sample ($r^2 = 0.9$; Supplementary Figures 3a, c and e). Signals in controls for unspecific staining by CARD-FISH stayed low, accounting for only 1–2 signals in 10 microscopic fields, in both the filters through which high and low volumes of seawater were filtered (from here on named high- and low-volume filters). Although calculated clade abundances were not significantly different, the s.e. values were markedly lower for the high-volume filters (Supplementary Figures 3d and f). Moreover, when cells of clade DE2, detected by probe CF6-1267, were counted on low- and high-volume filters from S3, no statistical difference was detected in the counts ($P > 0.05$). On the basis of these results, except for *Polaribacter* spp. all flavobacterial clades were quantified on 500-ml filters.

On these filters, the average numbers of probe-positive cells per counting grid for the abundant clades were in the range of 20–226 for probe VIS1-575; 5–200 for probe FL197; and 1–78 for probe CF6-1267 along the transect. For the low-abundance clades, the numbers ranged from 1–55 for VIS3-183; 1–38 for VIS6-586; 5–10 for FL845; and 1–9 for VIS5-586. The lower limit for quantification with the

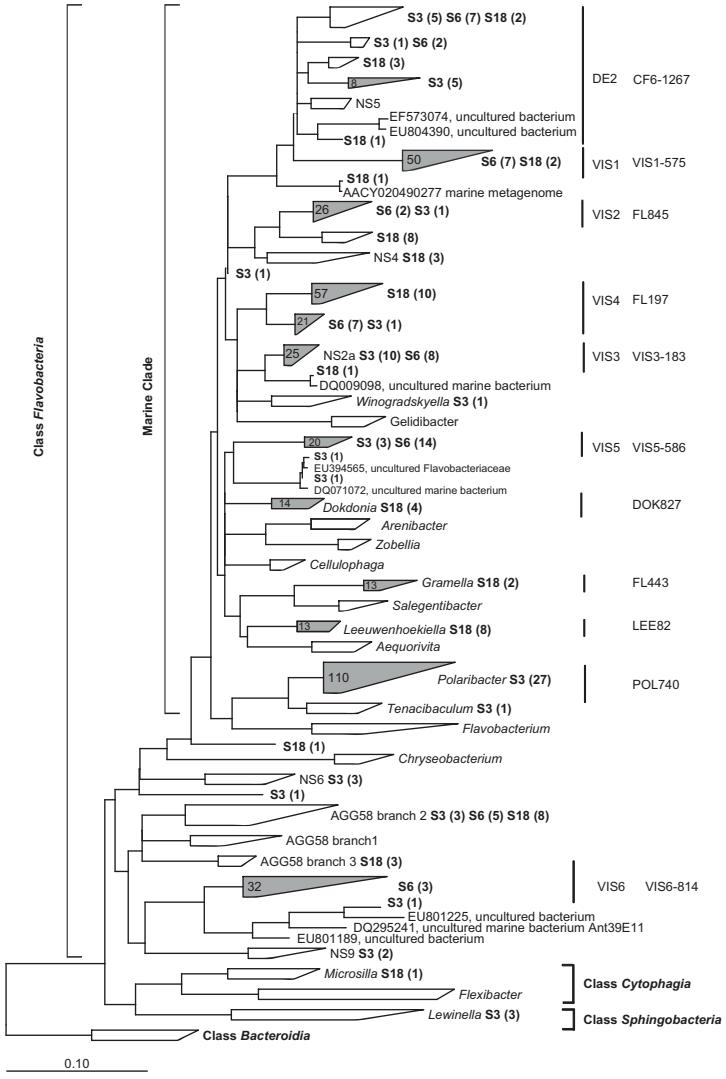


Figure 3 Phylogenetic association of sequences retrieved in this study from Boreal Polar (BPLR) (S3), Arctic (ARCT) (S6) and North Atlantic Subtropical (NAST) (S18), indicating the number of sequences of each station that contributes to the clades shown in wedges. Consensus tree was built with full-length sequences on the basis of neighbor joining and maximum likelihood methods, calculated without and with 50% bacteria positional conservatory and 30% *Bacteroidetes* positional conservatory filters. Gray shadows indicate target clades of flavobacterial clade-specific oligonucleotide probes indicating the number of targeted sequences in the database Silva 96 (Pruesse *et al.*, 2007). Name of the probes are given to the right. VIS: VISION clades.

modified protocol was set to 1 cell per counting grid on average.

Abundance of flavobacterial clades

The genus *Polaribacter* and the clades VIS1, VIS4 and DE2 were detected throughout the transect (Figure 4a; Supplementary Table 6). *Polaribacter* spp. was the most abundant clade at all the stations and depths analyzed, except for S8 in which clade VIS1 was slightly more abundant. Abundances were higher in BPLR and ARCT provinces, with a maximum detected at S4 with $72 \pm 45 \times 10^3$ *Polaribacter* spp. cells per ml that account for $6.7 \pm 3.5\%$ of picoplankton population. Counts decreased toward S7 ($P < 0.01$) and further south to

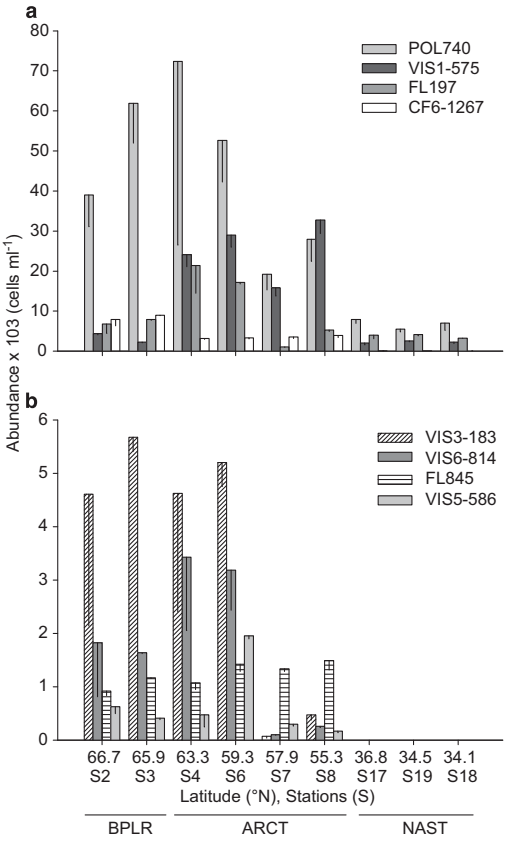


Figure 4 Total abundance of flavobacterial clade-specific probes in the Boreal Polar (BPLR), Arctic (ARCT) and North Atlantic Subtropical (NAST) provinces. (**a**) High-abundance clades and (**b**) low-abundance clades. Error bars represent s.d. values for duplicates (S2–S4; 20 and 40 m) or triplicates (S6–S19, 10, 20 and 30 m), shown as the negative s.d. value. All samples are from the upper mix layer and, therefore, considered as replicates.

S19 ($5.5 \pm 0.7 \times 10^3$ cells per ml; $1.0 \pm 0.1\%$). Members of the clade VIS1 significantly increased from BPLR to ARCT province ($P<0.01$), reaching a peak of $29 \pm 3 \times 10^3$ cells per ml ($3.4 \pm 0.4\%$) at S8. The VIS1 abundance in surface water was again lower at S19 ($2.5 \pm 0.2 \times 10^3$ cells per ml; 0.5 ± 0.05 %). At S4–S6, there was a significant increase in clade VIS4 ($P<0.001$), with a maximum of $21 \pm 7 \times 10^3$ cells per ml or $2.1 \pm 0.3\%$ at S4. This was followed by a steep decrease toward S7 ($1.0 \pm 0.2 \times 10^3$ cells per ml; $0.1 \pm 0.03\%$; $P<0.001$). The DE2 clade was most abundant in BPLR at S3 ($8.9 \pm 0.1 \times 10^3$ cells per ml; $0.9 \pm 0.07\%$); it showed a step-like decrease in the ARCT province ($3.3 \pm 0.2 \times 10^3$ cells per ml; $0.4 \pm 0.03\%$; S4; $P<0.001$) and a second steep step in the NAST stations (105 ± 12 cells per ml; 0.02%; S18; $P<0.001$).

Clades VIS3, VIS6, VIS2 and VIS5 were detected in the BPLR and ARCT provinces, but not in the NAST province (Figure 4b; Supplementary Table 6). The VIS3 and VIS6 clades were more abundant between S2 and S6 with a maximum of VIS3 at S3 with $5.6 \pm 0.2 \times 10^3$ cells per ml ($0.6 \pm 0.02\%$), and one of VIS6 at S4 with $3.4 \pm 1.3 \times 10^3$ cells per ml ($0.3 \pm 0.08\%$). As for VIS4, both groups significantly decreased within the ARCT province from S6 to S7 ($P<0.01$). The VIS2 clade had a relatively uniform distribution in BPLR and ARCT provinces with only a slight increase towards its maximum at S8 ($1.5 \pm 0.2 \times 10^3$ cells per ml; $0.1 \pm 0.03\%$). The group VIS5 represented a low fraction of the total bacterial community in both provinces (maximum at S6: $19.5 \pm 0.5 \times 10^2$ cells per ml; 0.2%).

Flavobacterial community composition in different provinces

Cluster analysis of absolute clade abundances showed that the flavobacterial communities were rather homogeneous within distinct water masses and different between them (Figure 5). Replicate samples of S2 and S3—representing BPLR—clustered with each other and were distinct from the adjacent ARCT stations S4 and S6 (Figure 5a). Within the ARCT province, the S4/S6 cluster was clearly separated from the flavobacterial communities at S7 and S8. The NAST stations S17–S19 were clustered tightly and were different from all other samples. The clustering was generally in good agreement with the water mass classification on the basis of its thermal and optical properties. There are only two minor differences between the satellite-derived classification of water masses and the clustering of flavobacterial communities. The first difference is that S7 and S8 occur in the same satellite-derived water mass (Figure 1), but are statistically distinct according to the cluster analysis (Figure 5a). The second is that S4 and S6 are in distinct water masses according to the satellite-based classification, but are not statistically different with respect to the clustering of flavobacterial clades.

A visualization of flavobacterial community composition in the different water bodies supported the clustering (Figure 5b). In the ARCT samples about 70% of the members of the class *Flavobacteria* could be assigned to specific clades, whereas in BPLR and NAST provinces the overall coverage with the set of clade-specific probes used in this study was <50%. Overall, the *Polaribacter* clade was the most important one accounting on average for about 30% of *Flavobacteria*, a fraction only equaled by VIS1 in S7 and S8. In the NAST province, the community was significantly different with *Polaribacter* and VIS4 accounting for 11 and 10% of *Flavobacteria*, respectively, followed by VIS1 (5.8%).

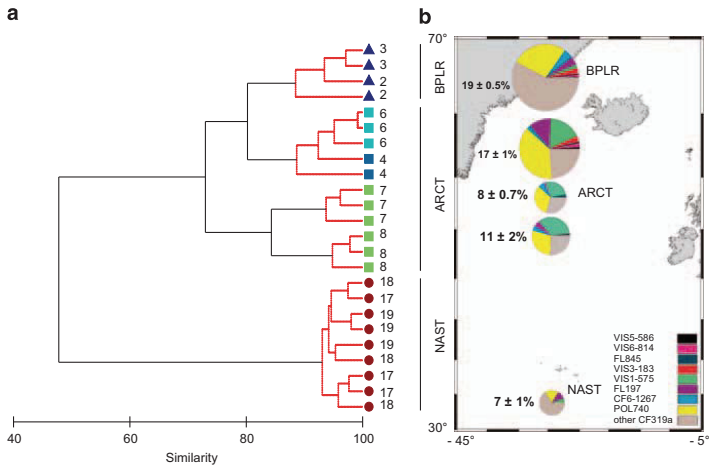


Figure 5 Flavobacterial community composition, as analyzed by fluorescence *in situ* hybridization (FISH) with clade specific-probes in the Boreal Polar (BPLR), Arctic (ARCT) and North Atlantic Subtropical (NAST) provinces. (**a**) Dendrogram representing the similarity of the flavobacterial community in the three provinces. Symbols are color coded as in Figure 1 and numbers indicate station replicates. Black lines in the dendrogram represent significant clusters at the 99% level using SIMPROF (similarity profile permutation test) test; red lines represent non-significant clusters at this level. (**b**) Relative contribution of each flavobacterial subgroup to the total CF319a. Pie chart represents average group sizes of clustering stations in (**a**); area of pie represents the relative abundance of CF319a (percentage is given to the left).

Table 3 Spearman rank correlation between temperature (temp), salinity (sal), phosphate (PO_4^{3-}) and dissolved silicate concentration (Si), chlorophyll *a* fluorescence (Chl *a*), total abundance of microphytoplankton (microphyto), nanophytoplankton (nanophyto) and *Flavobacteria* clade abundances

	CF319a	*POL740*	*CF6-1267*	*FL-845*	*VIS1-575*
Temp	−0.87	−0.89	−0.79	0.73	
Sal	−0.86	−0.82	−0.87		
PO_4^{3-}	0.84	0.87			
Si	0.87	0.84	0.77	−0.78	
Chl *a*	0.66				0.85
Microphyto	0.63				
Nanophyto	0.76		0.57		0.74

r_s values indicated are statistically significant after Bonferroni correction, at the level $P<0.05$. n being the number of simultaneous independent comparisons.

Further correlation analysis of single environmental variables revealed that counts with probe CF319a, as well as abundance of *Polaribacter* spp. and clade DE2 were significantly negatively correlated with temperature with r_s values of −0.87, −0.89 and −0.79, respectively (Table 3). The same applies for salinity with r_s value ranging between −0.82 and −0.87. The correlation with silicate concentration was significantly positive (r_s values: 0.77–0.87). Abundance of clade VIS2 was significantly positively correlated with temperature ($r_s=0.73$) and negatively with silicate concentration ($r_s=-0.78$). Abundances of VIS1, VIS3, VIS4, VIS5 and VIS6 clades did not correlate with any physicochemical parameter.

Taking biological variables into consideration, counts with probe CF319a were significantly correlated with chlorophyll *a* fluorescence ($r_s=0.66$), and with the total abundance of microphytoplankton ($r_s=0.63$) and nanophytoplankton ($r_s=0.76$). A stronger correlation was detected between the group VIS1 and chlorophyll *a* fluorescence ($r_s=0.85$) and nanophytoplankton abundance ($r_s=0.74$). Members of the DE2 clade were significantly correlated with the abundance of nanophytoplankton ($r_s=0.57$). The high-density 500-ml filters of S3 were used to do a preliminary check regarding whether the correlation is caused by a direct attachment of *Flavobacteria* to the phycosphere of nanophytoplankton cells (Figure 6). We quantified the CF319a-positive cells inside and outside the phycosphere of nanophytoplankton (Figure 6a) and compared the ratio with the relative localization of the free-living SAR11 clade (Figure 6b). The SAR11 cells were more abundant (27% at S3, 20 m; M Schattenhofer, unpublished results) than *Flavobacteria* (19% at S3, 20 m) in the BPLR province. Assuming a random co-occurrence, more SAR11 cells would be expected per nanophytoplankton cell compared with flavobacterial cells. However, the ratio inside and outside the phycosphere was 2.5 ± 0.8 ($n=13$) for CF319a and 0.8 ± 0.2 ($n=10$) for SAR11–441. The difference was significant ($P<0.01$), implying that *Flavobacteria* were specifically enriched in the phycosphere compared with SAR11 cells. The clades *Polaribacter*, VIS1, VIS4 and DE2 were identified in the phycosphere of nanophytoplankton (Figures 6c–f), but an exact

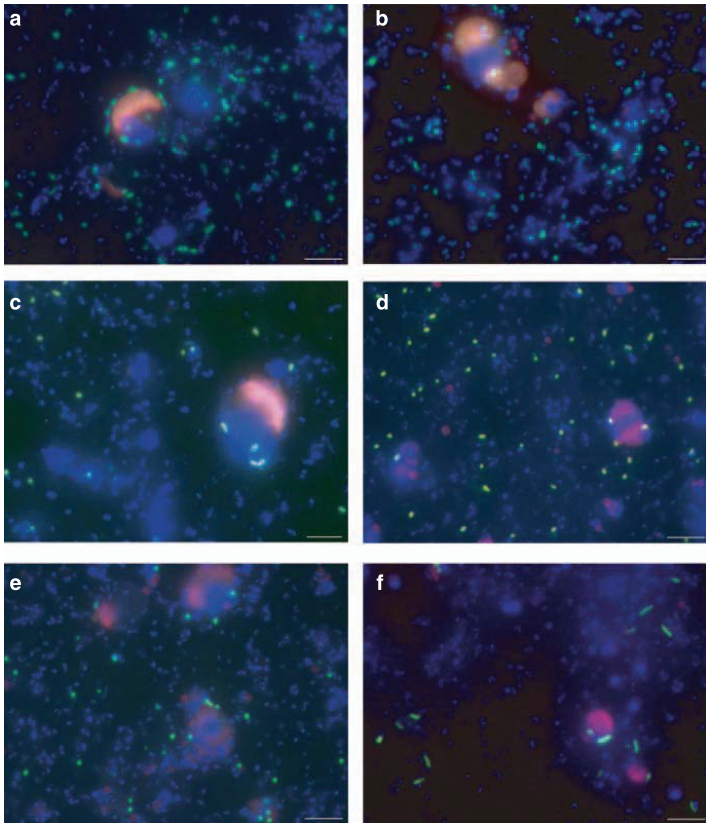


Figure 6 Epifluorescence micrographs of nanopytoplankton cells and bacteria in the phycosphere. Red, autofluorescence of nanophytoplankton cells; blue, DAPI signals of DNA containing cells; and green, signals of catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH)-positive cells using the probes: (**a**) CF319a, (**b**) SAR11-441, (**c**) POL740, (**d**) VIS1-575, (**e**) FL197 and (**f**) CF6-1267. a–c and f: S3, 20 m; d and e: S4, 20 m. Bar = 5 μM.

quantification was not possible. Filters obtained during the VISION cruise had too low numbers of algae with attached flavobacteria for a reliable quantification. Even higher volumes should have been filtered or the nanophytoplankton fraction should have been concentrated before filtration.

Discussion

High abundance of members of the class *Flavobacteria* in marine waters has already been reported by ZoBell (1946) in his famous monography 'Marine Microbiology'. He linked the occurrence of these aerobic heterotrophic bacteria that often form yellow-to-orange colonies on agar plates to the aerobic degradation of algal polymers. We here add molecular evidence supporting (i) the general importance of *Flavobacteria* in marine picoplankton and (ii) the occurrence of distinct flavobacterial communities in different marine water bodies. Our data of a transect in the North Atlantic between 65°N and 33°N confirm the preference of certain flavobacterial clades for cold and nutrient-rich water masses (Figure 2), as it had been reported by others (Simon *et al.*, 1999; Brinkmeyer *et al.*, 2003; Kirchman *et al.*, 2003; Abell and Bowman, 2005). Extending data of an Atlantic meridional transect cruise between 40°N and 20°S (Schattenhofer *et al.*, 2009), we demonstrate that other flavobacterial clades thrive under oligotrophic conditions in warm waters, both in the NADR and the NAST provinces.

Comparative 16S rRNA sequence analysis and high-resolution quantitative CARD-FISH using a newly designed set of clade-specific oligonucleotide probes provide independent evidence for the existence of distinct flavobacterial communities in different water bodies. The coverage of flavobacterial diversity in three representative clone libraries (S3, S6 and S18) was not complete, but

with values between 63% (NAST) and 92% (ARCT) was sufficient to clearly prove that there was only little overlap between the different water bodies. S3 and S18 shared only 2–5% of its phylotypes (Supplementary Table 4), whereas S3 and S6, which were more similar in terms of environmental characteristics, shared 25–40% of flavobacterial phylotypes.

The CARD-FISH quantification added numbers in support of distinct flavobacterial populations along the transect (Figures 4 and 5). These shifts coincided with transitions from one water mass to another (Figure 5a). Within the ARCT province, for example, a sharp decrease in chlorophyll *a* fluorescence was evident from S6 to S7, which was mirrored by an abrupt reduction of VIS3, VIS4 and VIS6 populations as well as of total *Flavobacteria* (Figures 1,2 and 4).

Overall, we identified marked distribution patterns of flavobacterial clades in contrasting water bodies of the North Atlantic Ocean. Our results are in line with the increasing body of evidence that distinct picoplankton populations show pronounced seasonal (Fuhrman *et al.*, 2006), latitudinal (Fuhrman *et al.*, 2008) and depth (DeLong *et al.*, 2006) distribution patterns in marine systems. For example, SAR11 abundances follow distinct annual distribution patterns in the euphotic and upper mesopelagic zones of the Sargasso Sea (Carlson *et al.*, 2009). Different high light- and low light-adapted ecotypes of *Prochlorococcus* show marked depth distributions in the subtropical and tropical Atlantic Ocean (Johnson *et al.*, 2006). Similarly, characteristic latitudinal distribution patterns of *Synechococcus* clades have been identified, with particular clades dominating temperate mesotrophic waters (Zwirglmaier *et al.*, 2007).

Distinct distribution patterns of abundant flavobacterial clades

The CARD-FISH analysis proved that members of the genus *Polaribacter* were present throughout the transect, although related sequences were only retrieved from BPLR province. The absolute numbers in the surface waters were consistently high and ranged between 7×10^4 (S4) and 5×10^3 (S19) cells per ml. The abundance was negatively correlated with temperature indicating a preference for colder water masses of less than 10 °C, which is within the optimal growth temperature of several cultured strains (Gosink *et al.*, 1998). Our data further support the importance of *Polaribacter* spp. in polar environments (Bano and Hollibaugh, 2002; Brinkmeyer *et al.*, 2003). Yet, although abundances of *Polaribacter* spp. decreased more than one order of magnitude in NAST province, in which the temperature was more than 20 °C, they were still the largest of the flavobacterial clades in these blue waters. This could be the result of a more cosmopolitan distribution of a single *Polaribacter* lineage. Alternatively, this could reflect a lack of discrimination, by probe POL740, of several mutually exclusive species in different water bodies. Recently, *P. dokdonensis* has been described to have a growth optimum of 25–28 °C (Yoon *et al.*, 2006). *P. dokdonensis* MED152 can grow in nutrient-poor environments and use proteorhodopsin for generating energy from light (Gonzalez *et al.*, 2008). Considering the high abundance of *Polaribacter* spp. and its significant intra-genus diversity both in its physiology and in 16S rRNA similarity, for example, *P. dokdonensis* shares only 96.2–96.8% with the other *Polaribacter* spp. type strains (Yoon *et al.*, 2006), we recommend that future studies should attempt to subdivide this important genus. We hypothesize that this will reveal uncoupled population dynamics of warm- and cold-water phylotypes with different life styles.

In contrast to *Polaribacter* spp., clades VIS1 and VIS4—both encompassing mostly 16S rRNA gene sequences from temperate and subtropical open ocean sites—had more narrow population maxima. The VIS1 clade was above 1×10^4 cells per ml throughout ARCT province only, a density reached by VIS4 at S4 and S6, but not at S7 and S8. Clade DE2 was again different, with a maximum slightly below 1×10^4 cells per ml in BPLR. The DE2 clade comprises sequences that have mostly been retrieved from coastal systems, such as the Delaware Estuary (Kirchman *et al.*, 2003), the North Sea (Eilers *et al.*, 2001, Alonso *et al.*, 2007) and the coast of Plymouth (O'Sullivan *et al.*, 2004). The abundance of DE2 in our open ocean samples was significantly lower than the 10% previously reported, for example, for the Delaware Estuary (Kirchman *et al.*, 2003). On our transect, the BPLR stations might have been the ones with the most significant coastal influence.

Toward a reliable quantification of rare flavobacterial populations

A major output of this study was a protocol for the identification and quantification of rare clades (Figure 4b). The VIS2, VIS3, VIS5 and VIS6 clades consistently formed groups of $<0.5\%$ relative abundance. In this study, we could quantify these populations down to absolute numbers of only 10^2 cells per ml, which was equivalent to detecting one target cell in 10^4 non-target cells. Interestingly, of the rare clades only the abundance of VIS2 was explained by an environmental variable. It was positively correlated with temperature ($r_s = 0.73$). This group consists of sequences retrieved from coastal environments and open ocean, but none from polar origin pointing at a preference for warmer waters.

As mentioned above, the rare clades, VIS3 and VIS6, significantly decreased from S6 to S7 (Figure 4b; Supplementary Table 6), mirroring a change in the optical properties of the water masses

(Figure1b and Figure 5a). Clade VIS6 comprises also phylotypes previously retrieved during a phytoplankton bloom in naturally iron-fertilized waters (West *et al.*, 2008); West *et al.* (2008) had reported a dominance of these phylotypes during the bloom. In this study, we could not detect a correlation between abundance of VIS6 and micro- or nanophytoplankton abundance or chlorophyll *a* fluorescence.

Except for *Polaribacter* spp. readily culturable flavobacteria were rarely detected. Members of the genus *Gramella* could not be detected in any of the samples, whereas those of genera *Dokdonia* and *Leeuwenhoekiella* were only observed in samples from the NAST province. Their abundance was too low to be quantified, but their conspicuously large morphology allowed for a confident identification in all three last stations (S17–S19). Genome analysis of *Dokdonia* spp. MED134 revealed that this strain contains a gene for proteorhodopsin and its growth is enhanced when exposed to light under low or intermediate labile organic matter concentrations (Gomez-Consarnau *et al.*, 2007), which could point to an adaptation of this group to oligotrophic environments.

In summary, the rare flavobacterial clades also showed differential distribution patterns. These clades, except for one case, did not correlate with other biological or environmental variables. With CARD-FISH of high volume filters, quantitative studies on 'the rare biosphere' (Sogin *et al.*, 2006) are now within reach. Moreover, rare clades can be responsible for most of the uptake of key nutrients (Musat *et al.*, 2008) and can carry out fundamental ecosystem processes (Kuypers *et al.*, 2005). Additional studies are now required to find out whether the rare flavobacterial clades are always present in low numbers, or if they bloom under particular conditions. Their exact ecosystem functions are yet to be elucidated.

Phytoplankton association of flavobacterial clades

Links between members of the class *Flavobacteria* and phytoplankton have been frequently reported (Simon *et al.*, 1999; O'Sullivan *et al.*, 2004; Pinhassi *et al.*, 2004; Grossart *et al.*, 2005; Sapp *et al.*, 2007). Phytoplankton cells release polymers, mainly polysaccharides, to the phycosphere during all the phases of their growth (Myklestad, 1995). Therefore, the association of *Flavobacteria* with algae has been related to their potential ability to degrade molecules released by phytoplankton (Kirchman, 2002). During phytoplankton blooms certain flavobacterial clades develop short-term maxima, whereas others dominate before the blooms (Riemann *et al.*, 2000; Pinhassi *et al.*, 2004; West *et al.*, 2008). Flavobacterial lineages might respond differentially to phytoplankton as a consequence of a physiological ability to take up different compounds released by algal cells. For example, the use of complex macromolecules varies among cultured genera of the class *Flavobacteria* (Bernardet and Nakagawa, 2006). Furthermore, genomic data point to a specialization of distinct flavobacterial clades for degradation of polysaccharides and proteins (Bauer *et al.*, 2006; Woyke *et al.*, 2009).

In our study, we detected a significantly positive correlation between total *Flavobacteria*, as quantified by probe CF319a, and nanophytoplankton, as well as for the clades VIS1 and DE2 with nanophytoplankton (Table 3). Microscopy demonstrated that cells of four flavobacterial clades: VIS1, DE2, VIS4 and *Polaribacter*, did indeed reside in the phycosphere of nanophytoplankton cells (Figures 6c–f). As this observation could be a consequence of random agglomeration of cells during filtration, we compared the localization of CF319a-positive cells with the localization of the abundant alphaproteobacterial SAR11 clade. Our results clearly indicate a preference of *Flavobacteria* for attachment to nanophytoplankton at S3 in comparison with a typical free-living bacterium like SAR11 (Giovannoni *et al.*, 1990; Rappe *et al.*, 2002). Follow-up studies should simultaneously identify nanophytoplankton species and flavobacterial clades to investigate the presence of distinct and predictable species-specific interactions. We are confident that the combination of genomic data on metabolic potential, substrate incubation experiments and information on *in situ* co-localization will soon yield detailed insight in how exactly this group of 'old' marine bacteria is interacting with phytoplankton.

Acknowledgements

We thank the Captain, Officers and Crew of the FS Maria S Merian (cruise MSM03/01) for their help during the cruise. We also thank Mrs M Schattenhofer for her help during sampling. We acknowledge the help of J Harder, J Ufkes and V Smetacek for providing unpublished data. We thank G Gerdts for advices on statistical analysis. Satellite analysis was funded by NASA Research Opportunities in Space and Earth Science, Terrestrial Ecology and Biodiversity Program NNH05ZDA001N. This study was funded by the Max Planck Society.

References

Abell GCJ, Bowman JP. (2005). Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbiol Ecol* **51**: 265–277.

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

Alonso-Saez L, Aristegui J, Pinhassi J, Gomez-Consarnau L, Gonzalez JM, Vaque D *et al.* (2007). Bacterial assemblage structure and carbon metabolism along a productivity gradient in the NE Atlantic Ocean. *Aquat Microb Ecol* **46**: 43–53.

Amann R, Fuchs BM. (2008). Single-cell identification in microbial communities by improved fluorescence *in situ* hybridization techniques. *Nat Rev Microbiol* **6**: 339–348.

Bano N, Hollibaugh JT. (2002). Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. *Appl Environ Microb* **68**: 505–518.

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

Bell W, Mitchell R. (1972). Chemotactic and growth tesponses of marine bacteria to algal extracellular products. *Biol Bull* **143**: 265.

Bernardet JF, Nakagawa Y. (2006). Introduction to the Family *Flavobacteriacea*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds). *The Porkaryotes*, 3rd edn. Springer-Verlag: New York. pp 550–580.

Bowman JP. (2006). The marine clade of the family Flavobacteriaceae: the genera Aequorivita, Arenibacter, Cellulophaga, Croceibacter, Formosa, Gelidibacter, Gillisia, Maribacter, Mesonia, Muricauda, Polaribacter, Psychroexus, Psychroserpens, Robiginitalea, Salegentibacter, Tenacibaculum, Ulvibacter, Vitellibacter and Zobellia. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds). *The Prokaryotes*, 3rd edn. Springer-Verlag: New York. pp 677–694.

Brinkmeyer R, Knittel K, Jurgens J, Weyland H, Amann R, Helmke E. (2003). Diversity and structure of bacterial communities in arctic versus antarctic pack ice. *Appl Environ Microb* **69**: 6610–6619.

Carlson CA, Morris R, Parsons R, Treusch AH, Giovannoni SJ, Vergin K. (2009). Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J* **3**: 283–295.

Chao A, Chazdon RL, Colwell RK, Shen TJ. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol Lett* **8**: 148–159.

Chen XH, Zeng Y, Jiao N. (2008). Characterization of Cytophaga–Flavobacteria community structure in the Bering Sea by cluster-specific 16S rRNA gene amplification analysis. *J Microbiol Biotechnol* **18**: 194–198.

Chen XH, Zeng YH, Jiao NZ. (2006). Development and evaluation of specific 16S rDNA primers for marine Cytophaga–Flavobacteria cluster. *Mol Ecol Notes* **6**: 1278–1281.

Clarke KR, Gorley RN. (2006). *PRIMER v6: User manual/tutorial*. PRIMER-E: Plymouth, UK.

Clarke KR, Somerfield PJ, Gorley RN. (2008). Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *J Exp Mar Biol Ecol* **366**: 56–69.

Colwell RK. (2005). *EstimateS: Statistical estimation of species richness and shared species from samples*. Version 7.5.User's Guide and application published at:http://purl.oclc.org/estimates.

DeLong EF, Franks DG, Alldredge AL. (1993). Phylogenetic diversity of aggregate-attached vs free-living marine bacterial assemblages. *Limnol Oceanogr* **38**: 924–934.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microb* **72**: 5069–5072.

Ducklow HW, Harris RP. (1993). Introduction to the JGOFS north Atlantic bloom experiment. *Deep-Sea Res Pt Ii* **40**: 1–8.

Eilers H, Pernthaler J, Peplies J, Glockner FO, Gerdts G, Amann R. (2001). Isolation of novel pelagic bacteria from the German bight and their seasonal contributions to surface picoplankton. *Appl Environ Microbiol* **67**: 5134–5142.

Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* **103**: 13104–13109.

Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL *et al.* (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci USA* **105**: 7774–7778.

Gifford DJ, Fessenden LM, Garrahan PR, Martin E. (1995). Grazing by microzooplankton and mesozooplankton in the high-latitude North Atlantic Ocean spring versus summer dynamics. *J Geophys Res-Oceans* **100**: 6665–6675.

Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**: 60–63.

Glöckner FO, Fuchs BM, Amann R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence *in situ* hybridization. *Appl Environ Microbiol* **65**: 3721–3726.

Gomez-Consarnau L, Gonzalez JM, Coll-Llado M, Gourdon P, Pascher T, Neutze R *et al.* (2007). Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445**: 210–213.

Gonzalez JM, Fernandez-Gomez B, Fernandez-Guerra A, Gomez-Consarnau L, Sanchez O, Coll-Llado M *et al.* (2008). Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp MED152 (Flavobacteria). *Proc Natl Acad Sci USA* **105**: 8724–8729.

Good IJ. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**: 237–264.

Gosink JJ, Woese CR, Staley JT. (1998). *Polaribacter* gen. nov., with three new species, *P. irgensii* sp. nov., *P. franzmannii* sp. nov., and *P. filamentus* sp. nov., gas vacuolate polar marine bacteria of the Cytophaga–Flavobacterium–Bacteroides group and reclassification of '*Flectobacillus glomeratus*' as *Polaribacter glomeratus* comb. nov. *Int J Syst Bacteriol* **48**: 223–235.

Grasshoff K, Kremling K, Ehrhardt M. (1983). *Methods of Seawater Analysis*. Wiley-VHC: Berlin, Germany, p 419.

Grossart HP, Levold F, Allgaier M, Simon M, Brinkhoff T. (2005). Marine diatom species harbour distinct bacterial communities. *Environ Microbiol* **7**: 860–873.

Henson SA, Robinson I, Allen JT, Waniek JJ. (2006). Effect of meteorological conditions on interannual variability in timing and magnitude of the spring bloom in the Irminger Basin, North Atlantic. *Deep-Sea Res Pt I* **53**: 1601–1615.

Huber T, Faulkner G, Hugenholtz P. (2004). Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**: 2317–2319.

Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW. (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**: 1737–1740.

Kemp PF, Aller JY. (2004). Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. *FEMS Microbiol Ecol* **47**: 161–177.

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

Kirchman DL, Yu LY, Cottrell MT. (2003). Diversity and abundance of uncultured Cytophaga-like bacteria in the Delaware Estuary. *Appl Environ Microbiol* **69**: 6587–6596.

Kormas KA, Tivey MK, Von Damm K, Teske A. (2006). Bacterial and archaeal phylotypes associated with distinct mineralogical layers of a white smoker spire from a deep-sea hydrothermal vent site (9 degrees N, East Pacific Rise). *Environ Microbiol* **8**: 909–920.

Kuypers MMM, Lavik G, Woebken D, Schmid M, Fuchs BM, Amann R *et al.* (2005). Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation. *Proc Natl Acad Sci USA* **102**: 6478–6483.

Legendre P, Legendre L. (1998). *Numerical Ecology*, Second English edn. Elsevier: Amsterdam, the Netherlands.

Llobet-Brossa E, Rossello-Mora R, Amann R. (1998). Microbial community composition of Wadden Sea sediments as revealed by fluorescence *in situ* hybridization. *Appl Environ Microb* **64**: 2691–2696.

Longhurst A. (1995). Seasonal cycles of pelagic production and consumption. *Prog Oceanogr* **36**: 77–167.

Longhurst A. (1998). *Ecological geography of the sea*, 1st edn. Academic Press: London.

Ludwig W, Euzeby JP, Whitman WB. (2008). Draft Taxonomic outline of the Bacteroidetes, Planctomycetes, Chlamydiae, Spirochaetes, Fibrobacteres, Fusobacteria, Acidobacteria, Verrucomicrobia, Dictyoglomi, and Gemmatimonadetes. In: Hedlund B, Krieg NR, Ludwig W, Paster BJ, Staley JT, Ward N *et al.* (eds). *Bergey's Manual of Systematic Bacteriology*, 2nd edn. Springer -Verlag: New York, pp 5–10.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

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

Manz W, Amann R, Ludwig W, Vancanneyt M, Schleifer KH. (1996). Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum cytophaga–flavobacter–bacteroides in the natural environment. *Microbiol* **142**: 1097–1106.

Manz W, Amann R, Ludwig W, Wagner M, Schleifer KH. (1992). Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria—problems and solutions. *Syst Appl Microbiol* **15**: 593–600.

Marie D, Partensky F, Jacquet S, Vaulot D. (1997). Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl Environ Microb* **63**: 186–193.

Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA *et al.* (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**: 806–810.

Musat N, Halm H, Winterholler B, Hoppe P, Peduzzi S, Hillion F *et al.* (2008). A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proc Natl Acad Sci USA* **105**: 17861–17866.

Muyzer G, Teske A, Wirsen CO, Jannasch HW. (1995). Phylogenetic-relationships of Thiomicrospira species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel-electrophoresis of 16S rDNA fragments. *Arch Microbiol* **164**: 165–172.

Myklestad SM. (1995). Release of extracellular products by phytoplankton with special emphasis on polysaccharides. *Sci Total Environ* **165**: 155–164.

Neef A, Amann R, Schlesner H, Schleifer KH. (1998). Monitoring a widespread bacterial group: *in situ* detection of planctomycetes with 16S rRNA-targeted probes. *Microbiology* **144**: 3257–3266.

O'Sullivan LA, Fuller KE, Thomas EM, Turley CM, Fry JC, Weightman AJ. (2004). Distribution and culturability of the uncultivated 'AGG58 cluster' of Bacteroidetes phylum in aquatic environments. *FEMS Microbiol Ecol* **47**: 359–370.

O'Sullivan LA, Weightman AJ, Fry JC. (2002). New degenerate cytophaga-flexibacter-bacteroides-specific 16S ribosomal DNA-targeted oligonucleotide probes reveal high bacterial diversity in river taff epilithon. *Appl Environ Microb* **68**: 2093–2093.

Oliver MJ, Irwin AJ. (2008). Objective global ocean biogeographic provinces. *Geophys Res Lett* **35**: L15601 doi:10.1029/2008GL034238.

Peplies J, Kottmann R, Ludwig W, Glockner FO. (2008). A standard operating procedure for phylogenetic inference (SOPPI) using (rRNA) marker genes. *Syst Appl Microbiol* **31**: 251–257.

Pernthaler A, Pernthaler J, Amann R. (2004). Sensitive multicolor fluorescence *in situ* hybridization for the identification of environmental microorganisms. In: Kowalchuk G, de Bruijn FJ, Head IM, Akkermans ADL, van Elsas JD (eds). *Molecular Microbial Ecology Manual*, 2nd edn. Kluwer Academic Publishers: Dordrecht, the Netherlands. pp 711–726.

Pinhassi J, Sala MM, Havskum H, Peters F, Guadayol O, Malits A *et al.* (2004). Changes in bacterioplankton composition under different phytoplankton regimens. *Appl Environ Microb* **70**: 6753–6766.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J *et al.* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.

Rappe MS, Connon SA, Vergin KL, Giovannoni SJ. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**: 630–633.

Riemann L, Steward GF, Azam F. (2000). Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. *Appl Environ Microb* **66**: 578–587.

Rooney-Varga JN, Giewat MW, Savin MC, Sood S, LeGresley M, Martin JL. (2005). Links between phytoplankton and bacterial community dynamics in a coastal marine environment. *Microbial Ecol* **49**: 163–175.

Sapp M, Schwaderer AS, Wiltshire KH, Hoppe HG, Gerdts G, Wichels A. (2007). Species-specific bacterial

communities in the phycosphere of microalgae? *Microbial Ecol* **53**: 683–699.

Schattenhofer M, Fuchs BM, Amann R, Zubkov MV, Tarran GA, Pernthaler J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ Microbiol* **11**: 2078–2093.

Schloss PD, Handelsman J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microb* **71**: 1501–1506.

Sievert SM, Kuever J, Muyzer G. (2000). Identification of 16S ribosomal DNA-defined bacterial populations at a shallow submarine hydrothermal vent near Milos Island (Greece). *Appl Environ Microb* **66**: 3102–3109.

Simon M, Glockner FO, Amann R. (1999). Different community structure and temperature optima of heterotrophic picoplankton in various regions of the Southern Ocean. *Aquat Microb Ecol* **18**: 275–284.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103**: 12115–12120.

Tarran GA, Heywood JL, Zubkov MV. (2006). Latitudinal changes in the standing stocks of nano- and picoeukaryotic phytoplankton in the Atlantic Ocean. *Deep-Sea Res Pt Ii* **53**: 1516–1529.

West NJ, Obernosterer I, Zemb O, Lebaron P. (2008). Major differences of bacterial diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom in the Southern Ocean. *Environ Microbiol* **10**: 738–756.

Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, Kiss H *et al.* (2009). Assembling the marine metagenome, one cell at a time. *PloS ONE* **4**: e5299.

Yoon JH, Kang SJ, Oh TK. (2006). *Polaribacter dokdonensis* sp nov., isolated from seawater. *Int J Syst Evol Micr* **56**: 1251–1255.

Zhou JZ, Bruns MA, Tiedje JM. (1996). DNA recovery from soils of diverse composition. *Appl Environ Microb* **62**: 316–322.

ZoBell CE. (1946). *Marine Microbiology: A monograph on Hydrobacteriology*. Chronica Botanica: Waltham, Massachusetts.

Zwirglmaier K, Heywood JL, Chamberlain K, Woodward EM, Zubkov MV, Scanlan DJ. (2007). Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. *Environ Microbiol* **9**: 1278–1290.

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

Supplementary material

Supplementary Table 1: Sequences of oligonucleotide probes and competitors specific for the flavobacterial clades designed in this study. The table shows the target sequences of the probes with their respective target clade and the closely related non-target sequences with less than 1.5 weighted mismatches (Ludwig *et al.,* 2004). Dots indicate full identity to the probe target sequence; only mismatches to the target are stated. The competitor sequences were designed to fully match the non-target sequences. They are identical to the probe sequence except for the nucleotides marked in bold-red, to ensure a full binding to the non-target sequences. Probe VIS3-183 is not shown because no competitors were designed.

FL197	Probe (5' - 3')	TCATCTCATACCGTAACC
VIS4	Target (3' - 5')	....
	Non-Target (3' - 5')	.....G.G..........
	Non-Target (3' - 5')	R......M..........
comp FL197_1	*Competitor (5' - 3')*	TCATC**C**C**C**TACCGTAACC
comp FL197_2	*Competitor (5' - 3')*	**Y**CATCTC**K**TACCGTAACC
VIS6-814	Probe (5' - 3')	CAGCGAGTGATGATCGTT
VIS6	Target (3' - 5')	....
	Non-Target (3' - 5')	...........G......
comp VIS6-814	*Competitor (5' - 3')*	CAGCGAGTGAT**C**ATCGTT
VIS5-586	Probe (5' - 3')	CCACTGACTTAGATTCCC
VIS5	Target (3' - 5')	....
	Non-Target (3' - 5')	............GC...
	Non-Target (3' - 5')	...........T......
comp VIS5-586_1	*Competitor (5' - 3')*	CCACTGACTTAGA**C**G**C**CC
comp VIS5-586_2	*Competitor (5' - 3')*	CCACTGACTTA**A**ATTTCC
FL845	Probe (5' - 3')	GCTTAGCCACTCATCTAA
VIS2	Target (3' - 5')	....
	Non-Target (3' - 5')	............T....
	Non-Target (3' - 5')	............T.G..
comp FL845_1	*Competitor (5' - 3')*	GCTTAGCCACTCA**A**CTAA
comp FL845_2	*Competitor*	GCTTAGCCACTCA**A**C**C**AA
VIS1-575	Probe (5' - 3')	CTTAACAAACAGCCTGCGGACC
VIS1	Target (3' - 5')	....
	Non-Target (3' - 5')	.....T................
comp VIS1-575	*Competitor (5' - 3')*	CTTAA**A**AAACAGCCTGCGGACC
LEE82	Probe (5' - 3')	ACTCGTCAGCTTCCCGAA
Leeuwenhoenkiella	Target (3' - 5')	....
	Non-Target (3' - 5')	C....G....G.......
	Non-Target (3' - 5')	..........T.......
comp LEE82_1	*Competitor (5' - 3')*	**G**CTCG**C**CAGC**C**TCCCGAA
comp LEE82_2	*Competitor (5' - 3')*	ACTCGTCAGC**A**TCCCGAA
DOK827	Probe (5' - 3')	GTCCGAAAACCAAACAACTAG
Dokdonia	Target (3' - 5')	....
	Non-Target (3' - 5')	............G...C....
	Non-Target (3' - 5')	............C....C....
comp DOK827_1	*Competitor (5' - 3')*	GTCCGAAAACCA**C**ACA**G**CTAG
comp DOK827_2	*Competitor (5' - 3')*	GUCCGAAAACC**G**AACA**G**CTAG

Supplementary Table 2: Formamide series used in the optimization of newly designed FISH probes that target uncultured clades indicating signal intensity (I) from very bright (+ + + +) to very low brightness (-) and the average number (N°) of probe-positive cells per counting grid (n=20). Shaded cells indicate the optimal formamide concentration used for further hybridizations. n.d.: not done.

Probe	VIS5-586		VIS2-845		VIS6-814		VIS3-183		VIS1-575		VIS4-197	
FA (%)	I	N°	I	N°	I	N°	I	N°	I	N°	I	N°
10	+ +++	n.d.	+ +++	350	+ +++	n.d.	+ +++	n.d.	+ +++	n.d.	+ +++	n.d.
15	n.d.		+ +++	300	+ +++	468	n.d.		n.d.		n.d.	
20	+ +++	37	+ ++	273	+ ++	357	+ +++	24	+ +++	n.d.	+ +++	273
25	+ +++	34	++	n.d	+ +	n.d.	+ +++	22	n.d.		n.d.	
30	+ ++	32	++	n.d	+ +	n.d	+ + +	14	+ +++	377	+ +++	256
35	+ +	22	+	n.d	n.d		+	n.d.	+ +++	349	+ +++	267
40	-	n.d.	-	n.d.	-	n.d	-	n.d.	+++	391	+ +++	270
45	n.d.		n.d.		n.d.		n.d.		+ +	389	+ +	160
50	-	n.d.	-	n.d.	-	n.d	-	n.d.	+	n.d.	-	n.d.

Supplementary Table 3: Total counts of CF319a expressed as DAPI% obtained in this study.

Station	Lat °N	Lon ° W	Depth	CF319a
2	66.39	-29.37	10	21.4
2	66.39	-29.37	20	19.1
2	66.39	-29.37	30	22.1
2	66.39	-29.37	40	19.3
2	66.39	-29.37	50	14.7
2	66.39	-29.37	60	15.0
2	66.39	-29.37	70	18.7
2	66.39	-29.37	80	19.4
2	66.39	-29.37	100	17.9
2	66.39	-29.37	150	7.2
2	66.39	-29.37	190	4.7
2	66.39	-29.37	275	1.5
3	65.53	-29.57	10	17.3
3	65.53	-29.57	20	19.7
3	65.53	-29.57	30	16.7
3	65.53	-29.57	40	18.2
3	65.53	-29.57	50	13.1
3	65.53	-29.57	60	19.5
3	65.53	-29.57	70	16.3
3	65.53	-29.57	80	--
3	65.53	-29.57	100	17.0
3	65.53	-29.57	150	21.8
3	65.53	-29.57	200	--
4	63.20	-30.53	10	15.8
4	63.20	-30.53	20	17.0
4	63.20	-30.53	30	18.6
4	63.20	-30.53	40	17.8
4	63.20	-30.53	50	16.2
4	63.20	-30.53	75	9.3
4	63.20	-30.53	100	6.0
4	63.20	-30.53	150	4.9
4	63.20	-30.53	200	6.5
4	63.20	-30.53	300	7.5
4	63.20	-30.53	400	2.3
4	63.20	-30.53	500	3.5
5	61.55	-30.03	10	16.2
5	61.55	-30.03	20	16.9
5	61.55	-30.03	30	17.3
5	61.55	-30.03	40	18.8
5	61.55	-30.03	50	11.4
5	61.55	-30.03	60	11.9
5	61.55	-30.03	70	9.3
5	61.55	-30.03	80	9.1
5	61.55	-30.03	100	11.0
5	61.55	-30.03	125	8.1
5	61.55	-30.03	150	5.3
5	61.55	-30.03	200	7.9
6	59.21	-29.60	10	17.7
6	59.21	-29.60	20	17.5
6	59.21	-29.60	30	19.4
6	59.21	-29.60	40	20.6

6	59.21	-29.60	50	17.2
6	59.21	-29.60	75	16.1
6	59.21	-29.60	100	14.8
6	59.21	-29.60	150	9.8
6	59.21	-29.60	200	5.0
6	59.21	-29.60	300	16.1
6	59.21	-29.60	400	3.6
6	59.21	-29.60	500	--
7	57.52	-30.01	10	8.1
7	57.52	-30.01	20	9.4
7	57.52	-30.01	30	8.1
7	57.52	-30.01	40	6.6
7	57.52	-30.01	50	7.8
7	57.52	-30.01	60	5.5
7	57.52	-30.01	70	10.5
7	57.52	-30.01	80	6.2
7	57.52	-30.01	100	8.8
7	57.52	-30.01	125	5.8
7	57.52	-30.01	150	7.0
7	57.52	-30.01	200	5.6
8	55.15	-30.02	10	9.5
8	55.15	-30.02	20	11.0
8	55.15	-30.02	30	13.3
8	55.15	-30.02	40	10.0
8	55.15	-30.02	50	10.9
8	55.15	-30.02	75	9.6
8	55.15	-30.02	100	7.7
8	55.15	-30.02	150	11.5
8	55.15	-30.02	200	6.0
8	55.15	-30.02	300	5.9
8	55.15	-30.02	400	4.3
8	55.15	-30.02	500	3.9
9	53.33	-30.01	10	11.3
9	53.33	-30.01	20	10.6
9	53.33	-30.01	30	12.8
9	53.33	-30.01	40	11.8
9	53.33	-30.01	50	11.7
9	53.33	-30.01	60	10.3
9	53.33	-30.01	70	5.5
9	53.33	-30.01	80	6.0
9	53.33	-30.01	100	5.3
9	53.33	-30.01	125	9.2
9	53.33	-30.01	150	8.9
9	53.33	-30.01	200	8.1
10	50.51	-29.60	20	8.6
10	50.51	-29.60	30	9.2
10	50.51	-29.60	40	13.2
10	50.51	-29.60	50	8.2
10	50.51	-29.60	75	9.0
10	50.51	-29.60	100	7.4
10	50.51	-29.60	150	2.3
10	50.51	-29.60	200	5.6
10	50.51	-29.60	300	--
10	50.51	-29.60	400	--
10	50.51	-29.60	500	--
11	49.26	-30.01	10	10.8
11	49.26	-30.01	20	11.5
11	49.26	-30.01	30	6.1
11	49.26	-30.01	40	11.0

11	49.26	-30.01	50	9.8
11	49.26	-30.01	60	8.9
11	49.26	-30.01	70	6.0
11	49.26	-30.01	80	2.2
11	49.26	-30.01	100	1.5
11	49.26	-30.01	125	1.4
11	49.26	-30.01	150	1.5
11	49.26	-30.01	200	3.7
12	46.44	-30.02	10	35.5
12	46.44	-30.02	20	30.6
12	46.44	-30.02	30	15.7
12	46.44	-30.02	40	15.0
12	46.44	-30.02	50	14.7
12	46.44	-30.02	75	7.6
12	46.44	-30.02	100	3.4
12	46.44	-30.02	150	3.4
12	46.44	-30.02	200	3.2
12	46.44	-30.02	300	5.4
12	46.44	-30.02	400	6.2
12	46.44	-30.02	500	2.2
13	45.22	-30.01	10	14.9
13	45.22	-30.01	20	11.0
13	45.22	-30.01	30	15.7
13	45.22	-30.01	40	16.3
13	45.22	-30.01	50	14.7
13	45.22	-30.01	75	5.3
13	45.22	-30.01	200	3.8
13	45.22	-30.01	400	2.0
14	42.42	-30.02	10	8.1
14	42.42	-30.02	20	9.5
14	42.42	-30.02	30	7.1
14	42.42	-30.02	40	7.8
14	42.42	-30.02	50	7.0
14	42.42	-30.02	75	6.7
14	42.42	-30.02	100	3.0
14	42.42	-30.02	150	3.5
14	42.42	-30.02	200	2.9
14	42.42	-30.02	300	2.2
14	42.42	-30.02	400	1.2
14	42.42	-30.02	500	1.2
15	41.12	-29.60	10	9.0
15	41.12	-29.60	20	7.6
15	41.12	-29.60	30	8.1
15	41.12	-29.60	40	8.7
15	41.12	-29.60	50	7.6
15	41.12	-29.60	60	8.9
15	41.12	-29.60	70	7.5
15	41.12	-29.60	80	5.6
15	41.12	-29.60	100	3.8
15	41.12	-29.60	125	3.8
15	41.12	-29.60	150	2.8
15	41.12	-29.60	200	1.3
16	38.29	-29.60	10	6.8
16	38.29	-29.60	20	6.7
16	38.29	-29.60	30	6.7
16	38.29	-29.60	40	7.4
16	38.29	-29.60	50	7.6
16	38.29	-29.60	75	7.0
16	38.29	-29.60	100	7.1

16	38.29	-29.60	150	--	
16	38.29	-29.60	200		3.2
16	38.29	-29.60	300		1.4
16	38.29	-29.60	400		1.5
16	38.29	-29.60	500		2.2
17	36.47	-29.60	10		8.0
17	36.47	-29.60	20		9.0
17	36.47	-29.60	30		6.3
17	36.47	-29.60	40		5.2
17	36.47	-29.60	50		5.9
17	36.47	-29.60	60		6.0
17	36.47	-29.60	70		6.3
17	36.47	-29.60	80		5.8
17	36.47	-29.60	100		3.0
17	36.47	-29.60	125		2.6
17	36.47	-29.60	150		1.2
17	36.47	-29.60	200		1.5
18	34.44	-30.01	10		7.0
18	34.44	-30.01	20		6.1
18	34.44	-30.01	30		6.8
18	34.44	-30.01	40		5.9
18	34.44	-30.01	50		6.0
18	34.44	-30.01	75		5.1
18	34.44	-30.01	100		7.7
18	34.44	-30.01	150		2.9
18	34.44	-30.01	200		2.2
18	34.44	-30.01	300		0.9
18	34.44	-30.01	400		1.2
18	34.44	-30.01	500		1.5
19	34.25	-28.29	10		7.4
19	34.25	-28.29	20		5.4
19	34.25	-28.29	30		6.7
19	34.25	-28.29	40		5.5
19	34.25	-28.29	50		5.5
19	34.25	-28.29	75		6.5
19	34.25	-28.29	100		5.7
19	34.25	-28.29	250		1.5
19	34.25	-28.29	500		0.9

Supplementary Table 4: Shared diversity indexes at 99% and 97% identity between the three 16S rRNA gene clone libraries. J_{clas}: Jaccard classic; L_{clas}: Sorensen classic; J_{abd}: Chao-Jaccard-Raw Abundance-based; L_{abd}: Sorensen-Jaccard-Raw Abundance-based according to Chao *et al*., (2005).

	N° Shared OTU	J_{clas}	L_{clas}	J_{abd}	L_{abd}
97%					
S3, S6	6	0.25	0.40	0.25	0.40
S3, S18	1	0.03	0.05	0.02	0.03
S6,S18	4	0.15	0.27	0.11	0.20
99%					
S3, S6	5	0.18	0.30	0.18	0.31

Supplementary Table 5: Number of sequences retrieved in this study targeted by the flavobacterial clade-specific oligonucleotide probes. Includes [(a)] four and [(b)] one partial sequences not covering the target site, but closely related (>99% identity) to the targeted sequences.

Target group	Probe	S3	S6	S18
DE2	CF6-1267	5	0	0
VIS1	VIS1-575	0	7	2
VIS2	FL845	1	2	0
VIS4	FL197	1	7	10
VIS3	VIS3-183	10	8 [(a)]	0
VIS5	VIS5-586	3	14	0
Dokdonia	DOK827	0	0	4
Gramella	FL443	0	0	2
Leeuwenhoekiella	LEE82	0	0	8 [(b)]
Polaribacter	POL740	27	0	0
VIS6	VIS6-814	0	3	0
	total targeted	47	41	26
	total sequences	71	55	58
	% targeted	**65**	**75**	**45**

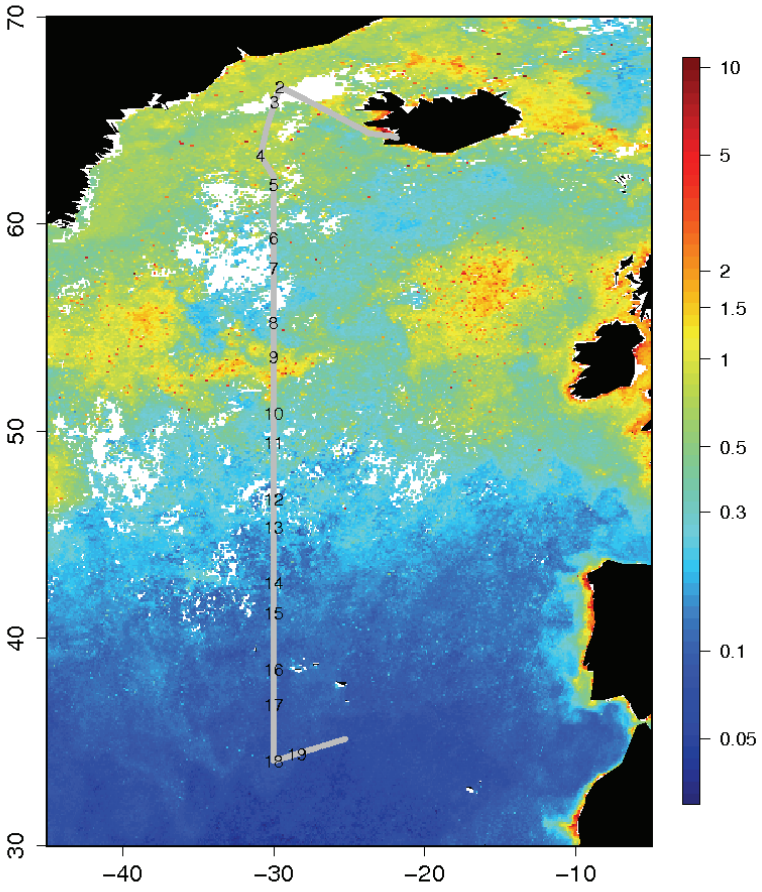
Supplementary Table 6: Abundance of *Flavobacteria* clades in the stations analyzed expressed as average and standard deviation of duplicates (S2-S4) or triplicates (S6-S19). n.d.: no detected.

		VIS5-586	FL845	VIS6-814	VIS3-183	CF6-1267	FL197	VIS1-575	POL740
Province	S	cells ml^{-1} x 10^{2}				cells ml^{-1} x 10^{3}			
BPLR	2	6.2 ± 1.3	9.2 ± 8	18 ± 10	46 ± 25	7.9 ± 1.6	6.9 ± 2.3	4.4 ± 0.3	39 ± 7.8
BPLR	3	4.1 ± 0.3	12 ± 0.1	16 ± 0.1	57 ± 2.5	8.9 ± 0.1	7.9 ± 0.2	2.2 ± 0.1	62 ± 10
ARCT	4	4.7 ± 2.2	11 ± 1.2	34 ± 14	46 ± 22	3.2 ± 0.2	21 ± 6.9	24 ± 3.0	72 ± 46
ARCT	6	20 ± 0. 5	14 ± 1.3	32 ± 7.5	52 ± 4.4	3.3 ± 0.2	17 ± 0.3	29 ± 3.0	53 ± 10
ARCT	7	2.9 ± 0.3	13 ± 0.4	1.0 ± 0.7	0.7± 0.3	3.5 ± 0.4	1.0 ± 0.2	16 ± 2.0	19 ± 3.8
ARCT	8	1.7 ± 0.3	15 ± 1.8	2.6 ± 0.2	4.7 ± 0.9	3.4 ± 0.5	5.2 ± 0.2	33 ± 3.3	28 ± 5.5
NAST	17	n.d.				0.1 ± 0.03	4.0 ± 0.8	2.0 ± 4.5	7.9 ± 0.9
NAST	18	n.d.				0.1 ± 0.01	3.3 ± 0.1	2.2 ± 0.3	7.0 ± 1.8
NAST	19	n.d.				1.1 ± 0.01	4.2 ± 0.3	2.5 ± 0.2	5.5 ± 0.7

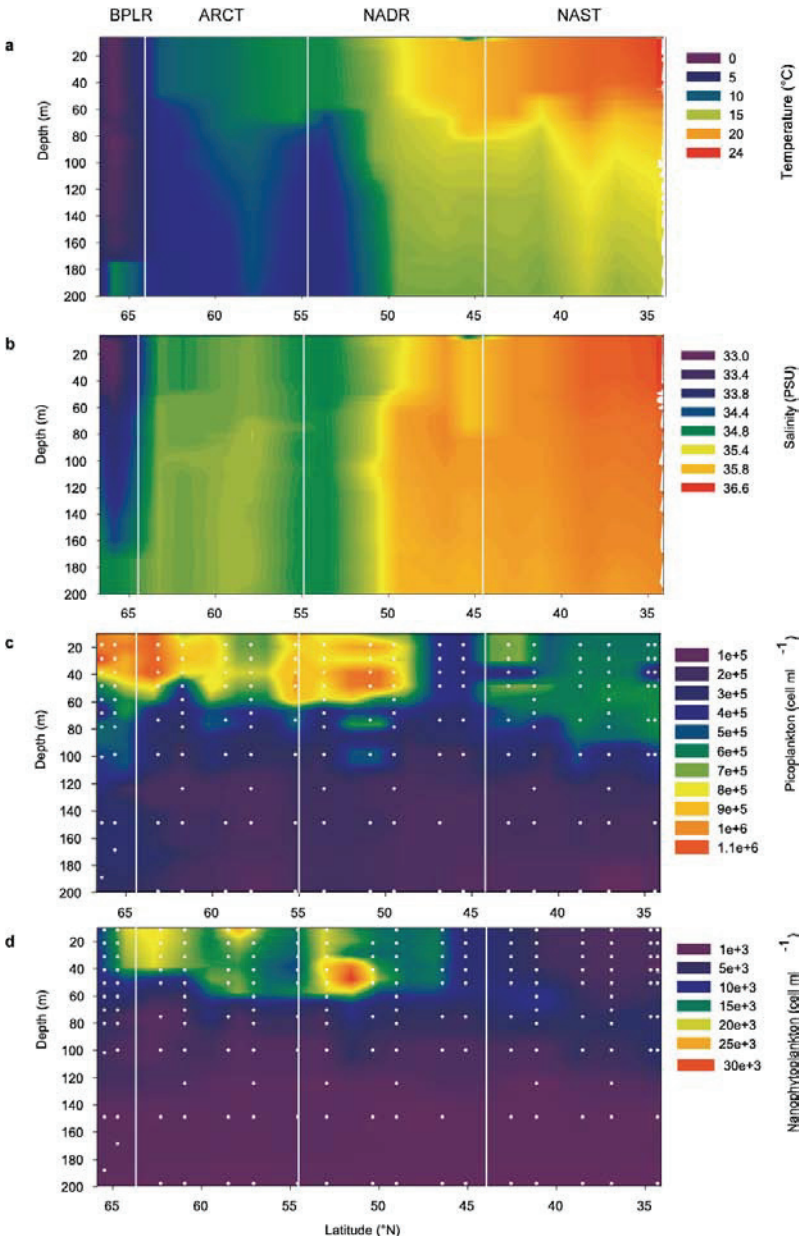
Supplementary Figure 1: (a) September 2006 monthly average Chlorophyll *a* concentration (µg l^{-1}) from the SeaWiFS satellite (http://oceancolor.gsfc.nasa.gov/). Missing data is shown as white.

Supplementary Figure 2: Latitudinal depth distribution of **(a)** Temperature (°C), **(b)** Salinity (PSU), **(c)** abundance of picoplankton (cells ml^{-1}) and **(d)** abundance of nanophytoplankton (cells ml^{-1}). Dots represent the sampling depths.

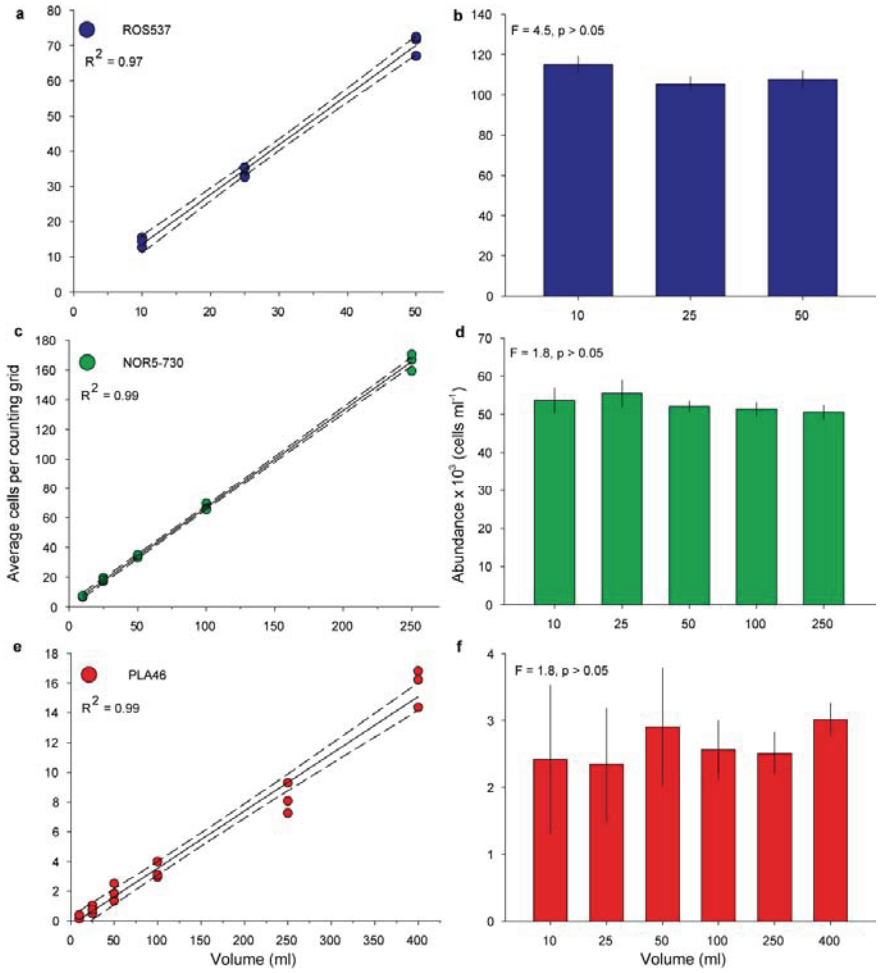
Supplementary Figure 3: Evaluation of high volume water samples with the probes ROS537 (a-b), NOR5-730 (c-d) and PLA46 (e-f). **(a) (c) (d)**: Average cells per counting grid at the different increasing volume samples indicating regression line (solid line), regression coefficient (R^2) and 95% confidence intervals (dashed line) of the 3 volume replicates. **(b) (d) (f)**: Calculated abundance of each clade indicating F statistic and p-value result from ANOVA (b and d) or ANOVA on ranks (f) tests.



Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3

Manuscript II

Genomic content of uncultured *Bacteroidetes* from contrasting oceanic provinces in the North Atlantic Ocean

Paola R Gómez-Pereira, Margarete Schüler, Bernhard M Fuchs,
Christin Bennke, Valerie Barbe, and Rudolf Amann

Manuscript in preparation. Intended as article in the ISME journal.

Genomic content of uncultured *Bacteroidetes* from contrasting oceanic provinces in the North Atlantic Ocean

Paola R Gómez-Pereira[1], Margarete Schüler[1,2*], Bernhard M Fuchs[1], Christin Bennke[1], Valerie Barbe[3], and Rudolf Amann[1*]

[1] Max Planck Institute for Marine Microbiology, Molecular Ecology Department, Bremen, Germany.

[2] Max Planck Institute for Biochemistry, Department of Molecular Structural Biology, Martinsried Germany.

[3] Genoscope, Evry, France

* Joint corresponding authors

M. Schüler; Address: Am Klopferspitz 18 ; E-mail: schueler@biochem.mpg.de

R. Amann; Address: Celssiustrasse 1; E-mail: ramann@mpi-bremen.de

Running title: Genomic potential of marine *Bacteroidetes*

Intended as article in the ISME journal

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

Summary

Bacteroidetes are widespread in marine systems where they are hypothesized to have a remarkable role in the recycling of organic matter. Whole genome sequenced strains have revealed a great glycolytic and proteolytic potential. The aim of this study was to compare the genomic information of pure cultures with metagenomics fragments of distinct *Bacteroidetes* clades in contrasting oceanic provinces in the North Atlantic Ocean, the polar biome (BPLR) and the North Atlantic Subtropical (NAST). We present here the analysis of 21 *Bacteroidetes* fosmids carrying the 16S rRNA gene as phylogenetic marker, and its comparison with complete *Bacteroidetes* genomes. Almost all genomic fragments belonged to clades previously identified in the same water masses, revealing genomic information about *Bacteroidetes* key players. The majority of the fosmids, based on the 16S rRNA gene phylogeny, were affiliated with *Flavobacteria*. We also present novel information about *Cytophagia* and *Sphingobacteria;* the latter genomic fragment encoding several proteins potentially involved in the attachment to other cells or surfaces. In our genomic fragments data set, we identified an overall broadly distributed potential for the degradation of proteins and the amino-sugar containing bacterial cell wall component peptidoglycan, particularly in higher frequency in the oligotrophic NAST. Polysaccharide degradation was confined to a single fosmid of the *Polaribacter* genus. It showed a whole armory of enzymes that might function in degradation of sulfated polysaccharides known to be major constituents of phytoplankton cell walls. This study provides support for a strong specialization of marine *Bacteroidetes* for the degradation of particular polymers.

Key words: Metagenomics, *Flavobacteria,* picoplankton, CARD-FISH, fosmids, genome comparison, hydrolases, sulfatases.

47 **Introduction**

48 *Bacteroidetes* is one of the major bacterial lineages in marine systems; they
49 thrive in a variety of environments at high abundance, e.g. in coastal waters (Kirchman
50 et al 2003), anoxic marine sediments (Llobet-Brossa et al 1998), and in open ocean
51 (Schattenhofer et al 2009), suggesting a great adaptation to diverse environmental
52 conditions. In surface open ocean they can comprise up to 70% of the total picoplankton
53 during phytoplankton blooms (Simon et al 1999), but they are also present in
54 oligtotrophic oceanic areas (Gómez-Pereira et al 2010, Schattenhofer et al 2009).
55 Moreover, there is evidence that *Bacteroidetes* have an pivotal role in the degradation of
56 organic matter (Kirchman 2002).

57 Organic matter (OM) in the oceans is one of the largest reservoirs of organic
58 carbon on Earth (Hedges 1992). Dissolved OM in the surface ocean is mainly composed
59 of polysaccharides (Benner et al 1992), proteins, lipids (Ogawa and Tanoue 2003) and
60 bacterial cell wall components such as peptidoglycan (McCarthy et al 1998) and
61 membrane components (Tanoue et al 1995). Those substrates can support heterotrophic
62 growth (Keil and Kirchman 1991, Nagata et al 2003), particularly of *Bacteroidetes*,
63 which metabolize high molecular weight compounds such as proteins and chitin
64 (Cottrell and Kirchman 2000). The initial breakdown of OM is the limiting step in its
65 degradation (e.g. (Arnosti 2003), hence *Bacteroidetes* are hypothesized to play a key
66 role in this process in the oceans.

67 Taxonomically, the phylum *Bacteroidetes* consists of the classes *Bacteroidia*,
68 *Sphingobacteria, Cytophagia* and *Flavobacteria,* and the latter seems to be the
69 dominant in marine picoplankton (Alonso et al 2007, Chen et al 2008). Whole genome
70 sequencing of marine *Bacteroidetes* strains, most of them belonging to the class
71 *Flavobacteria*, has revealed a great adaptation for polymeric OM degradation (Bauer et
72 al 2006, Gonzalez et al 2008, Woyke et al 2009). A high number of peptidases,

glycoside hydrolases, and starch binding domains, homologous to the ones characterized in the human-associated polysaccharide degrader *Bacteroides thetaiotaomicron*, have been detected in the genomes of *Gramella forsetii* (Bauer et al 2006), *Polaribacter dokdonensis* MED152 (Gonzalez et al 2008), and the single cell amplified genomes of uncultured flavobacteria MS024-2A and MS024-3C (Woyke et al 2009).

Marine *Bacteroidetes* inhabit different kinds of marine particles (Crump et al 1999, Pedrotti et al 2009, Woebken et al 2007a) and phytoplankton cells (Grossart et al 2005) and this association has been related with their potential role in the degradation of OM (Kirchman 2002). Whole genome sequenced strains gives support to this conception. The genomes of *P. dokdonensis* MED152 (Gonzalez et al 2008) and *G. forsetii* (Bauer et al 2006) encode an extensive potential for surface adhesion and gliding. Moreover gliding motility, a feature related to the attachment of cells to cells or to surfaces (Hall-Stoodley et al 2004), is known in several *Bacteroidetes* strains.

The presence of light driven metabolism has been described in *Bacteroidetes* genomes (Gomez-Consarnau et al 2007, Gonzalez et al 2008), but not in all of them (Bauer et al 2006). This point at distinct adaptations to the marine environment. Given the great flavobacterial diversity (Alonso et al 2007) whole genome sequenced strains still represent a minority of the flavobacterial clades in the ocean, and might only constitute a small fraction of the community (Woyke et al 2009). Therefore a metagenomic approach is desired to unveil the possible role in the degradation of OM of yet uncultured bacteroidetal clades.

This study was performed in two contrasting oceanic provinces of the northern North Atlantic Ocean (Supplementary Figure 1). The Polar Biome, BPLR, had higher concentration of chlorophyll *a* (0.7 $\mu g\ l^{-1}$), phytoplankton biomass and inorganic nutrients, and low temperature of (0.7°C). The North Atlantic subtropical province

99 (NAST) had very low concentrations of chlorophyll *a* (<0.1 µg l^{-1}), nutrients, and a
100 higher temperature 20°C (Table 1). The *Bacteroidetes* diversity in these two sites had
101 little overlap: at the 97% 16S rRNA gene identity level only 2-3% of the phylotypes
102 were shared between both provinces (Gómez-Pereira et al 2010). CARD-FISH
103 quantification of flavobacterial clades confirmed that the BPLR and NAST were inhabit
104 by distinct populations (Gómez-Pereira et al 2010). The aim of this study was to use
105 fosmid-based comparative (meta)genomics to gain insights about the potential
106 functions of the various bacteroidetal clades. Particularly, we focused on genes
107 involved in the degradation of organic matter and in the adhesion to surfaces or cells
108 and an extensive comparison with *Bacteroidetes* genomes.
109
110 **Materials and Methods**
111 *Study site*
112 Samples were taken in the North Atlantic Ocean during the VISION cruise in
113 September 2006 (Supplementary Figure 1; Table 1). Metagenomics libraries were
114 constructed from two contrasting oceanic provinces: S3 was located in BPLR and S18
115 in the NAST, according to the delimitation proposed by (Longhurst 1995).
116 Environmental characteristics are given in Table 1 and were previously described by
117 Gomez-Pereira et al. (2010).
118 *Library construction*
119 Samples were collected with a rosette of 20 l Niskin bottles mounted on a
120 conductivity-temperature-density (CTD) profiler at 20 m depth. Approximately 100 liter
121 samples were pooled from 5 Niskin bottles, pre-filtered through a 10 µm steel mesh
122 cartridge (Wolftechnik, Weil der Stadt, Germany), and microorganisms were collected
123 on cellulose acetate filters (142 mm diameter, 0.2 µm pore size; Sartorius, Goettingen,
124 Germany) and stored at -80°C for further processing. Genomic DNA was extracted

125 from the 0.2 µm filters according to (Zhou et al 1996) and library construction was done
126 immediately after the overnight resuspension of the DNA. Fosmids libraries were
127 constructed with the Copy Control™ Fosmid Library Production Kit (EPICENTRE
128 Biotechnologies, Madison, WI, USA) following instructions of the manufacturer and as
129 previously described by (Woebken et al 2007b) (for details see Supplementary
130 Information).

131 *Fosmid library screening*

132 Clones were pooled according to the row-column-plate scheme (Asakawa et al
133 1997) after induction for high copy number according to the instruction of the Control
134 Copy Control™ Fosmid Library Production Kit (EPICENTRE Biotechnologies). The
135 DNA of the pooled clones was subsequently extracted using the 96-well Montage
136 Plasmid MiniPrep96 kit (Millipore, Eschborn, Germany). Fosmids libraries were PCR-
137 screened with the bacteroidetal specific primer pair CF315 and CF967 (Chen et al 2006)
138 as described previously (Gómez-Pereira et al 2010). The pooling scheme helped to
139 traced back the positive PCR products to single clones of the fosmids library.

140 Sequencing of selected *Bacteroidetes* fosmids was done at Genoscope (Paris,
141 France). Fosmids from S3 were sequenced by a transposon bomb strategy. *In vitro*
142 transposition was done with HyperMu™ Transposons Kit Kan-1 (EPICENTRE
143 Biotechnologies). A total of 384 clones per fosmid were end-sequenced by dye-
144 terminator chemistry with an ABI3730 sequencer (Applied Biosystems, Foster City,
145 USA) resulting in approximately 14-fold coverage. Sequences were assembled using
146 Phred/phrap/Consed software (www.phrap.org). Fosmids from S18 were sequenced by
147 454-pyrosequencing, with a 454 GS flx / Titanium (454 Life Sciences, Roche
148 Diagnostics, Branford, CT, USA) at a 20-fold coverage. Sequences were assembled
149 with the software Newbler (454 Life Sciences, Roche Diagnostics) and validated with

the Consed interface (Gordon et al 1998). Finishing was done by primer walking on fosmid clones and PCRs products. All fosmids were assembled in a single contig.

Gene prediction and annotation

Gene prediction was done with an in-house pipeline (porfs) (J. Waldman, unpublished) which combines the outputs of Glimmer v3.02 (Salzberg et al 1998), Zcurve v1.02 (Zhang and Wang 2000), Metagene v1.0 (Noguchi et al 2006) and MED v2.0 (Zhu et al 2007). Results of those tools were evaluated using Artemis (Rutherford et al 2000). Annotation was performed with GenDB v2 system (Meyer et al 2003) using various tools for each predicted gene. Similarity searches were performed against sequence databases (NCBI nr, NCBI nt, SwissProt) and protein family databases (Pfam, InterPro, COG). From these predictions, an automatic annotation was generated using the fuzzy logic-based auto-annotation tool MicHanThi (Quast 2006). High-quality annotations were generated by manual revision of each gene annotation. Protein localizations were predicted with SignalP v2.0 (Nielsen et al 1997) and PSORTb (Gardy et al 2003); lipoproteins were identified using LipoP (Juncker et al 2003) and according to (Seydel et al 1999). Transmembrane helix predictions were done with TMHMM v2.0 (Krogh et al 2001). Information on predicted peptidases was obtained from the MEROPS peptidase database (Rawlings et al 2008) and on predicted glycoside hydrolases from the Carbohydrate Active Enzyme database (CAZy) (Cantarel et al 2009).

Genome comparisons

Comparative analyses of fosmids and whole genomes from *Bacteroidetes* were performed using the JAVA tool Jcoast v1.5 (Richter et al 2008). For comparative analysis the custom made database GenomesDB was used (Richter et al 2008), which contained 621 Bacteria and Archaea genomes (January 2008) and was updated with publicly available *Bacteroidetes* draft genomes, then amounting to a total of 36

176 genomes, 8 from *Bacteroidia*, 21 from *Flavobacteria* and 7 from *Cytophagia*. The
177 GenomesDB database was used for exploring the taxonomic distribution of genes with
178 BLAST at a cutoff-E-value of E-15.

179 Pairwise comparisons at the amino acid level between fosmids and genomes
180 from organisms of the same genus were performed with the "Promer" tool of the
181 software Mummer v3.0 (Kurtz et al 2004). The nucleotide sequences of fosmids
182 carrying 16S rRNA genes 99% identical were compared with the "Nucmer" tool of
183 Mummer V3.0 (Kurtz et al 2004).

184 *Phylogenetic reconstruction, probe design and catalyzed reporter deposition (CARD)-*
185 *fluorescence in situ hybridization (FISH)*

186 Phylogenetic reconstruction of the 16S rRNA genes of the fosmids was done
187 with the ARB package (Ludwig et al 2004) using the Silva database version 100
188 (Pruesse et al 2007). Oligonucleotide probes targeting the 16S rRNA of the fosmids
189 were designed with the PROBE_DESIGN tool in ARB, as well as unlabeled competitor
190 oligonucleotides (Manz et al 1992) (Supplementary Table 1). Optimization of the
191 probes stringency conditions were evaluated as described before (Gómez-Pereira et al
192 2010). CARD-FISH was performed as previously described on 500 ml filter sections
193 (Gómez-Pereira et al 2010) in samples from the BPLR (S2 and S3) and from the NAST
194 (S17, S18 and S19) provinces.

195

196 **Results**

197 Screening of approximately 15,500 fosmid clones constructed from DNA
198 extracted from station 3 (S3) and 27,600 clones from S18 for 16S rRNA genes yielded
199 28 *Bacteroidetes* fosmids (13 from S3 and 15 from S18), which were subsequently
200 sequenced and manually annotated.

201 During this process, several highly similar fosmids were identified in both
202 libraries, with 99.9-100% 16S rRNA gene identity. Nucleotide pairwise comparisons
203 identified long overlapping regions (16 - 20 kbp) with a 97.3 - 100% nucleotide
204 sequence identity, with the exception of S18-997-C3. The longer overlapping genomic
205 fragments were merged (M) while shorter redundant fragments were excluded from
206 further analysis (Table 2, Supplementary Table 2). A total of 21 fosmids (9 from S3 and
207 12 from S18) roughly equaling 0.9 Mbp of non-redundant sequence information and
208 carrying a total of 695 open reading frames (ORFs) were further analyzed (Table 2).
209
210 ***Phylogenetic affiliation of the 16S rRNA genes on the fosmids***
211 Based on comparative 16S rRNA gene sequence analysis, 17 of the 21 fosmids
212 were affiliated with the class *Flavobacteria* (7 of the 9 fosmids from S3, 10 of the 12
213 fosmids from S18), three with the class *Sphingobacteria* (two from S3, one from S18),
214 and one (S18) with the class *Cytophagia* (Table 2, Figure 1). Analysis of the 16S rRNA
215 gene sequences revealed that 19 of the 21 fosmids in our set (exceptions:
216 sphingobacterial fosmids S3-972-A4 and S18-1049-858M) were affiliated with clades
217 entirely defined by sequences retrieved in PCR clone libraries from marine samples,
218 (e.g. (Alonso et al 2007, Malmstrom et al 2007, O'Sullivan et al 2004, Shaw et al 2008).
219 The 16S rRNA gene sequences from 14 fosmids had between 100 and 97% identity, and
220 those found on four other fosmids had between 96 and 93% identity with sequences
221 retrieved by PCR from the same samples (Gómez-Pereira et al 2010) (Table 2, Figure 1).
222 Within the sphingobacterial fosmids, the 16S rRNA gene sequence of fosmid
223 S3-892-858M showed highest similarity to sequences previously retrieved from marine
224 sites. This clade was denominated Sphingobacteria A (Figure 1) and its abundance was
225 9.5 x 10^2 cell ml^{-1} in the BPLR (S2-S3) (Table 2, Figure 2a). Sphingobacteria A could
226 not be detected in the NAST (S18-19) by CARD-FISH. The closest relatives of the two

227 other sphingobacterial fosmids were sequences from diverse habitats, but not sequences
228 previously retrieved from the same sampling site. The closest relatives of the
229 cytophagial fosmid S18-1001-F9 were sequences retrieved from open ocean sites, and
230 the abundance of the respective cytophagial clade was slightly higher in S3 (17 ± 0.2 x
231 10^2 cell ml^{-1}) than in S18 (5.9 ± 0.4 x 10^2 cell ml^{-1}) (Table 2, Figure 2b).
232 Within the set of flavobacterial fosmids, several belonged to clades that have
233 recently been delineated based on 16S rRNA sequences from the same study site
234 (Gomez-Pereira et al., 2010) (Figure 1, clades labeled "VIS"), e.g. VIS3 (S18-812-F11),
235 VIS4 (S18-920-B9), VIS5 (S3-861-A8) and VIS6 (S3-979-825M). One fosmid from
236 each station (S3-805-E4 and S18-873-E7) belonged to the uncultured DE2 clade
237 retrieved from the Delaware Estuary (Kirchman et al 2003) (95% 16S rRNA gene
238 identity with AY274860). Congruently, the abundance of the flavobacteria from which
239 the fosmid S3-808-E4 derived was higher in S3 (25 ± 1.1 x 10^2 cell ml^{-1}) and was not
240 detected in S18 (Figure 2 c), while the abundance of the flavobacterium from which
241 S18-873-E7 was retrieved was higher in S18 (13 ± 1.1 x 10^2 cell ml^{-1}, Figure 2 d) and
242 not detected in S3. Two of the fosmids from S18, S18-906-E12 and S18-870-C2, were
243 affiliated with the AGG58 clade (DeLong et al 1993, O'Sullivan et al 2004). The closest
244 relatives of the fosmids S18-848-B2 and S18-1082-D10 were sequences previously
245 retrieved from the North Sea (Alonso et al 2007), clades NS9 (e.g AM279203) (Figure f)
246 and NS4 (e.g. AM279212), respectively (Figure 1).
247 The 16S rRNA gene of fosmid S3-816-G8 was related with that of a fosmid
248 retrieved from Antarctic sea ice, Ant39E11 (Grzymski et al 2006). The clade was
249 denominated ANT and comprised both fosmids and other sequences retrieved from
250 coastal and open ocean sites. Cells of the ANT clade were targeted with a FISH
251 oligonucleotide probe and were detected in the BPLR (S2-S3) (25 ± 10 x 10^2 cell ml^{-1}),
252 but not in S18 (Table 2, Figure 2e). Despite belonging to different genera as indicated

253 by a 16S rRNA gene sequence identity of 92%, the ANT fosmids had an amino acid
254 sequence similarity of 78% along a 28.6 kb stretch (Supplementary Figure 2a), and 23
255 out of 28 ORFs in S3-816-G8 exhibited highest pairwise similarity (blastp) with
256 predicted proteins from Ant39E11 (Grzymski et al 2006).

257 Of the 17 flavobacterial fosmids in our set, five belonged to the readily cultured
258 genera *Dokdonia* (one from S18), *Leeuwenhoekiella* (two from S18), and *Polaribacter*
259 (two from S3) (Table 2, Figure 1). The *Dokdonia* fosmid S18-841-H9 showed 94.3%
260 amino acid similarity to a region of the *D. donghaensis* MED143 genome, with 91%
261 coverage of its 43.1 kb sequence (Supplementary Figure 2b). The *Leeuwenhoekiella*
262 fosmids S18-1087-1013M and S18-997-C3 shared between 79.6 and 80.6% amino acid
263 similarity at a coverage of 76 - 87% of their lengths with regions of the *L. blandensis*
264 MED217 genome (Supplementary Figure 2 c, d). The 16S rRNA genes of the two
265 fosmids belonging to the *Polaribacter* genus (S3-860-D3 and S3-843-F5) were 99%
266 identical, but no other similar region between the two was detected. The 16S rRNA
267 identity with *P. irgensii* 23-P was 97.8% for S3-860-D3 and 98.2% for S3-843-F5 while
268 with *P. dokdonensis* MED152 the similarity was lower (95.4% and 94.5% respectively).
269 A comparison of the two *Polaribacter* fosmids with the genomes of these organisms on
270 the amino acid level revealed that fosmid S3-860-D3 did not share any region with
271 either of the two genomes while fosmid S3-843-F5 had several regions in common with
272 the genomes of *P. dokdonensis* MED152 (75% similarity and 66% coverage) and *P.*
273 *irgensii* 23-P (75% similarity and 65% coverage) (Supplementary Figure 2 e, f).
274 Moreover, 30/32 ORFs of S3-843-F5 showed highest pairwise similarity (blastp) to *P.*
275 *dokdonensis* MED152 or *P. irgensii* 23-P proteins, in contrast to only 3/27 of S3-860-
276 D3.

277 The analysis of the genomic content on the *Bacteroidetes* fosmids focused on the
278 search of genes involved in the degradation of marine organic matter, the uptake of its

279 monomers, and the adhesion to other cells. Genes involved in other metabolic processes
280 are not discussed in the context of this study.
281
282 ***Genomic content of Bacteroidetes fosmids***
283 *Polysaccharide degradation*
284 In the entire set of predicted proteins encoded by the 21 fosmid sequences, only
285 three glycoside hydrolases (GH) were identified, all of them located on the *Polaribacter*
286 fosmid S3-860-D3 (Figure 3a, Supplementary Table 3). The three glycoside hydrolases
287 belonged to the CAZy family 92, members of which have been shown to act as alpha-
288 1,2-mannosidases (Cantarel et al 2009). According to the prediction of certain cellular
289 localization signals (Supplementary Table 3), one of the GH (ORF0) was exported to
290 the periplasmic space and the other two were secreted and located in the outer
291 membrane. Besides the three putative alpha-1,2-mannosidases, the *Polaribacter* fosmid
292 was found to encode another enzyme of the D-mannose metabolism, a mannose-6-
293 phosphate isomerase, which converts fructose 6-phosphate and mannose-6-phosphate
294 into each other. A conspicuous accumulation of five sulfatase genes was found on this
295 *Polaribacter* fosmid (Figure 3a, Supplementary Table 4). For all of the fosmid-encoded
296 sulfatases a signal-peptide was predicted, thus the proteins are probably exported from
297 the cytoplasm. Three of the sulfatases were encoded in consecutive ORFs and one of the
298 sulfatase genes was found next to a gene encoding an L-fucose transporter (ORF8). In
299 addition to the sulfatase genes on fosmid S3-860-D3, a sulfatase gene was identified on
300 one of the two DE2-clade fosmids (S18-873-E7, ORF41). The other DE2-clade fosmid
301 in our set, S3-805-E4, harbors a sulfate transporter (ORF11, Figure 3b).
302 *Protein degradation and recycling of peptidoglycan*
303 In contrast to GH, genes encoding peptidases were detected in almost all of the
304 fosmids. A higher percentage of peptidases was identified on the *Bacteroidetes* fosmids

set from S18 compared to that of S3 (Figure 4a). The diversity of peptidase families was also higher in S18 (12 different families vs. 5 in S3). Of the S18 peptidases, 30% exhibited either an export (SpI) or secretion (SpII) signal, whereas 50% of the peptidases from S3 showed an export signal (SpI) (Supplementary Table 5).

Overall, the most frequently found peptidases were from the family M23 (5 ORFs), which are endopeptidases that lyse bacterial cell wall peptidoglycans (Rawlings et al 2008). Members of this family were identified on the two DE2 fosmids (S18-873-E7 ORF35 and S3-805-E4 ORF13), the VIS5 fosmid (S3-861-A8, ORF17, exported), the ANT fosmid (S3-816-G8, ORF2) and on the Sphingobacteria A fosmid (S3-858-892M, ORF38). On the VIS5 fosmid an ATP-dependent protease La2 (S16 family) that degrades unfolded proteins was additionally found. Three different peptidases were detected on the VIS4 fosmid (S18-920-B2, Figure 3c). One of them belonged to the family S41 (ORF43) and contained a PDZ domain that is found in signaling proteins; additionally this peptidase had a lipoprotein signal (SpII) and was probably directed to the outer membrane. We also found a transmembrane peptidase of the family M48 (ORF10), possibly involved in the degradation of abnormal proteins, next to an oligopeptidase (ORF12 and 13, active domain) and in the vicinity of an oligopeptide transporter (ORF41).

A glutamyl aminopeptidase of family M42 (metalloaminopeptidases) was found on the NS4 fosmid (S18-1082-D10, ORF20) and on the NS9 fosmid (S18-848-B2, ORF26) from S18, on the latter located next to a two-component signal transduction system (ORF28 and ORF29, two-component response regulator; ORF30 histidine kinase). These systems enable bacteria to sense, respond, and adapt to a wide range of environments, stressors, and growth conditions. Thus, the NS9-fosmid host is likely capable of responding to the presence of certain proteins in its environment. Additionally, in the NS4 fosmid a Sec-independent protein translocase TatC transport

331 system was identified in the vicinity of the peptidase M42, which might participate in
332 the export of the peptidase. Consistently, a signal peptide was not detected in its amino
333 acid sequence. A glycoprotease of the family M22 was detected on the Sphingobacteria
334 A fosmid S3-858-892M (ORF12) and on the cytophagial S18-1001-F9 fosmid (ORF12)
335 but not on any flavobacterial fosmid. The M22 family contains endopeptidases that
336 exclusively degrade O-sialoglycoproteins.

337 In addition to peptidases of the family M23, several proteins involved in the
338 recycling or degradation of peptidoglycan were found on flavobacterial, but not on
339 sphingobacterial or cytophagial fosmids (Supplementary Table 6). Peptidoglycan, a
340 principal component of bacterial cell walls, is degraded by lytic transglycosylases and
341 endopeptidases (Holtje 1998) whereby anhydro-N-acetyl muramic acid (anhMurNAc) is
342 one of the degradation products. By the enzymes anhydro-N-acetylmuramic acid kinase
343 (*anmK*) and N-acetylmuramic acid 6-phosphate etherase (*murQ*), anhMurNAc is then
344 converted to N-acetyl-glucosamine 6-phosphate (GlcNAc 6-P) (Jaeger et al 2005,
345 Uehara et al 2006), which can ultimately be shuttled into glycolysis. The gene *anmK*
346 was found on the VIS4 fosmid (S18-920-B2, ORF22) (Figure 3c), one of the two DE2
347 fosmids (S18-873-E7, ORF14, Figure 3b) as well as on the VIS6 fosmid (S3-979-825M,
348 ORF20). The *murQ* gene was found on the VIS3 fosmid S18-841-H9 and on the
349 *Dokdonia* fosmid. Additionally, a mannosyl-glycoprotein-endo-beta-N-
350 acetylglucosaminidase was found on fosmids encoding *anmK* (VIS4 and DE2, Figure
351 3b, c) or *murQ* (*Dokdonia*) as well as on the second DE2-fosmid S18-873-E7 (Figure 3b;
352 Supplementary Table 5).

353 A *murF* gene was found on the VIS4 (S18-920-B2), VIS6 (S3-979-825M) and
354 ANT (S3-816-G8) fosmids, which is involved in bacterial cell wall formation
355 (biosynthesis of the murein precursor muramoyl-pentapeptide). In all three cases, the
356 *murF* gene was located downstream of a gliding motility gene (*gldJ*), an organization

found so far almost exclusively in flavobacterial genomes (19 of 21 genomes, data not shown).

Surface adhesion and gliding

A higher percentage of proteins possibly involved in attachment were identified on S3 fosmids compared to S18 fosmids (Figure 4c, Supplementary Table 7). Most of these proteins contained a glycosyl transferase (GT) family 1 domain. Proteins with this domain transfer activated sugars to a variety of substrates, including glycogen, fructose-6-phosphate and lipopolysaccharides; they are involved in various biosynthetic processes including exopolysaccharide and lipopolysaccharide core biosynthesis. Exopolysaccharides can mediate cell/cell and cell/surface adhesion (Hall-Stoodley et al 2004). Three GT1 were found on the Sphingobacteria A fosmid S3-892-858M (ORF0, ORF1 and ORF5, Figure 3d) that were in a cluster interspersed with genes encoding mannose-1-phosphate guanylyltransferase (ORF4), involved in the synthesis of the capsular polysaccharide colanic acid, and UDP-N-acetylglucosamine 2-epimerase (ORF2) involved in lipopolysaccharide biosynthesis.

A protein containing a repeated polycystic kidney disease (PKD) domain was found on the AGG58 branch 3 fosmid (S18-906-E12). Interestingly, this protein had highest pairwise similarity (blastp) to archaeal cell surface layer proteins (e.g. YP_305014). It has been proposed that archaeal proteins containing this domain might be involved in cell-to-cell adhesion (Jing et al 2002). This fosmid also carried an ORF (ORF0) with weak similarity to a GT1 domain protein.

In the entire fosmid set, eight genes encoding proteins associated with gliding motility (McBride 2004) were identified, which had homologues exclusively in almost all *Bacteroidetes* genomes. Four of the gliding proteins (*gldL, gldK, gldO* and *gldM*) and a protein containing a tetratricopeptide repeat (TPR) domain (*sprE*), which has been denominated as a possible gliding motility protein, were found in an operon-like

383 organization on the Sphingobacteria C fosmid S18-1049-858M and were contiguous
384 with a gene encoding a GT51 protein (Figure 3e). Additionally, a *sprE*-homologue is
385 also harbored by the Sphingobacteria A fosmid (S3-892-858M).
386 *Transporters*
387 A similar number of transporters were identified on the fosmid sets from both
388 geographic locations, however, the transporter sets differed with respect to family
389 composition and frequency of members. A total of five TonB-dependent outer
390 membrane receptors were identified on the flavobacterial fosmids of the genera
391 *Polaribacter* (S3-860-D3, ORF14) and *Leeuwenhoekiella* (S18-997-C3, ORF0), and of
392 the clades VIS3 (S18-812-F11, ORF20), AGG58 branch 2 (S18-870-C2, ORF20) and
393 branch 3 (S18-906-E12, ORF22). Two distantly related member of this family were
394 found on the VIS5 (S3-861-A8, ORF18, plug domain) and on the Sphingobacteria C
395 (S18-1049-858, ORF38) fosmid. Additionally, a MotA/ExbB and an ExbD-like protein,
396 possibly proton-transporting components of the TonB-dependent transduction complex,
397 were found on the fosmids of the clades DE2 (S18-873-E7, ORF11 and ORF12) and
398 VIS6 (S3-825-979M, ORF23 and ORF24). In the *Bacteroidetes* genomes analyzed so
399 far, certain TonB-dependent receptors are often found encoded next to SusD-like
400 polysaccharide-binding outer membrane proteins (e.g. (Xu et al 2003). In our fosmid set,
401 we found only one of the five TonB-dependent receptors (S18-997-C3, ORF0) encoded
402 next to an outer membrane SusD family protein on a *Leeuwenhoekiella* fosmid (S18-
403 997-C3, ORF1).
404 High numbers of transporter genes were found on *Leeuwenhoekiella* as well as
405 on AGG58 fosmids. Of the former, S18-997-C3 had an oligopeptide transporter
406 (ORF26) next to an exported dipeptidyl peptidase (ORF24, peptidase family S9) and a
407 beta-lactamase family peptidase (ORF22). On the fosmid S18-1087-1013M three ion-
408 and metal-transporters were identified; ORF16 was an ion channel, ORF22 (S18-1087-

409 C3) a sodium/solute iodide symporter, and ORF22 a zinc/metal transporter (ZIP,
410 PF02535), which might participate in the import of metals to the cells. A ZIP-family
411 metal transporter was also detected on the VIS5 fosmid and on S3-933-H4. In addition,
412 a ferrous iron transport system (FeoAB) was found on fosmid S3-816-G8, like on its
413 close relative Ant39E11.

414 Considering the AGG58 fosmids, one on S18-906-E12 and five ORFs on S18-
415 870-C2 transporters were identified. On S18-906-E12 a TonB-dependent outer
416 membrane receptor was found. On fosmid S18-870-C2 a Na+/glucose co-transporter
417 (ORF2) and a major facilitator superfamily protein family 1 (ORF12) were detected.
418 The latter are single-polypeptide secondary carriers capable of transporting only small
419 solutes (Pao et al 1998). On this fosmid also an ABC transport system (permease and
420 ATP binding subunits, ORF24 and ORF33) was identified.

421 *Pigments*

422 Several genes encoding pigment biosynthesis proteins were identified on the
423 sphingobacterial fosmids. For example, the Sphingobacteria A fosmid S3-892-858M
424 (Figure 3d) had two ORFs involved in the biosynthesis of carotenoids: phytoene
425 dehydrogenase, *crtI* (ORF20) and phytoene synthase, *crtB* (ORF21). The
426 Sphingobacteria C fosmid S18-1049-858M (Figure 3e) had two genes that take part in
427 porphyrin/chlorophyll-metabolism: ORF19 encoded an uroporphyrinogen-III synthase
428 (*hemD*). ORF17 was identified as an oxygen-independent coproporphyrinogen III
429 oxidase (*hemN*). A protein belonging to the magnesium chelatase subunit ChlI family
430 (Pfam profile PF01078) was found on the VIS6 fosmid (S3-979-825M, ORF39).
431 Magnesium-chelatase is a three-component enzyme that catalyses the insertion of Mg^{2+}
432 into protoporphyrin IX, which is the first unique step in the synthesis of
433 bacteriochlorophyll. Members of this family are predicted throughout the genomes of
434 *Bacteroidetes*, even though other genes of this pathway have not been reported. Possibly,

435 these genes encode proteins that are involved in a related process of introducing
436 catalytic metal cores into apoproteins. In our dataset, we did not identify genes part of
437 light dependent pathways.

438 *Special carbon compound conversion/utilization*

439 Two ORFs encoding proteins involved in the catabolism of phenylacetic acid
440 were found on the AGG58 fosmid S18-906-E12. Phenylacetic acid is an aromatic
441 compound used as a carbon and energy source by a wide variety of microorganisms
442 (Ferrandez et al 1998). The gene enoyl CoA hydratase, *paaG*, (ORF11) was also
443 identified in 7/22 *Flavobacteria* and in 2/8 *Bacteroidia* genomes, while the phenylacetic
444 acid degradation oxidoreductase, paaE, (ORF10) was present in all *Flavobacteria*
445 genomes, in 4/7 *Cytophagia,* and 2/8 *Bacteroidia* genomes. Thus, the utilization of this
446 compound might be widespread in *Bacteroidetes*.

447 On the cytophagial fosmid S18-1001-F9 (Figure 4f) two genes involved in the
448 xylulose metabolism (xylose isomerase and xylulose kinase) were detected, and, in
449 addition, a carbohydrate kinase (ORF9), which participates in the activation of
450 carbohydrates. We also identified on this fosmid both units of the nitrogen metabolism
451 enzyme allophanate hydrolase (subunit 1 (ORF20) and 2 (ORF21). Allophanate
452 hydrolase catalyzes the second reaction in the ATP-dependent two-step degradation of
453 urea to ammonia (Kanamori et al 2005).

454

455 **Discussion**

456 The motivation of the present study was to get first insights into the metabolic
457 potential of distinct, mostly uncultured *Bacteroidetes* clades in contrasting oceanic
458 provinces in the North Atlantic Ocean, the polar biome (BPLR) and the North Atlantic
459 Subtropical (NAST), and to compare it with the information available for cultivated
460 strains. We manually annotated 21 *Bacteroidetes* fosmids that had an average insert size

461 of 38 kbp and carried the 16S rRNA gene as phylogenetic marker. This allowed the
462 establishment of a link between metabolic information and organism identity and
463 abundance, thus providing hints with respect to the significance of certain metabolic
464 traits in different ecological settings.

465 The majority of fosmids in our set originated from *Flavobacteria*, which is in line
466 with the emerging evidence that *Flavobacteria* is the predominant class of the phylum
467 *Bacteroidetes* in marine plankton e.g. (Alonso et al 2007, Bowman 2006, Chen et al
468 2008). Marine *Bacteroidetes* belonging to the classes *Sphingobacteria* seem to be less
469 abundant in marine plankton (Alonso et al 2007). Our set of genomic fragments
470 contains only one from *Cytophagia* and three from *Sphingobacteria*. Yet, their analyses
471 initiate a closer characterization of their potential ecological role in marine plankton.

472 Almost all the *Flavobacteria* fosmids were phylogenetically affiliated with the
473 major clades previously identified at the same study sites, such as VIS3, VIS4, VIS5,
474 VIS6 (Gómez-Pereira et al 2010), DE2 (Kirchman et al 2003), and AGG58 (DeLong et
475 al 1993, O'Sullivan et al 2004) (Table 2, Figure 1). Interestingly, none of the fosmids
476 clades were shared between both stations. Consequently, they reveal metabolic
477 information about key players in the contrasting cold nutrient-rich BPLR and
478 subtropical oligotrophic NAST.

479 In our dataset we identified congruence between certain flavobacterial genomic
480 fragments (Supplementary Figure 2). The fosmid S3-816-G8 retrieved from the BPLR
481 and the fosmid Ant39E11 from Antarctic sea ice (Grzymski et al 2006) presented a
482 surprisingly conserved gene content and order, suggesting that the type of environment
483 might have a strong influence in shaping the metabolic potential of polar marine
484 flavobacteria.

485 The comparison of meta-genomic fragments with whole genomes from closely
486 related isolated organisms opens the possibility to investigate their genetic variability.

487 Interestingly, the deduced amino acid sequences of the *Dokdonia*, *Leeuwenhoekiella*
488 and one *Polaribacter* fosmids were found to be highly similar (74-94% similarity) to
489 almost contiguous 40 kbp regions in the genomes of strains from these genera
490 (Supplementary Figure 2). This suggests low genomic variability between closely
491 related strains, at least in the analyzed region, even if they thrive in different marine
492 environments. The strains *L. blandensis* MED217 and *D. donghaensis* MED134 have
493 been isolated from the Blanes Bay in the Mediterranean Sea (Gomez-Consarnau et al
494 2007), which is an inland sea with high organic material input and nutrient
495 concentrations (Lucea et al 2005) relative to the oligotrophic NAST from were the
496 fosmids where retrieved. The genome of *D. donghaensis* MED134 contains a
497 proteorhodopsin gene, and its growth is enhanced by light at low concentrations of
498 DOM (Gomez-Consarnau et al 2007). It has been suggested that this might be a strategy
499 to adapt to scenarios of organic matter depletion (Gomez-Consarnau et al 2007,
500 Gonzalez et al 2008) as the oligotrophic NAST. In our study, we did not identify genes
501 part of light driven metabolisms.
502
503 ***Degradation capabilities***
504 We focused our analysis on genes involved in degradation of organic matter (OM)
505 and in surface adhesion in order to address, for two contrasting marine scenarios, the
506 hypothesized role of marine *Bacteroidetes* as particle-associated biopolymer-degraders.
507 The comparison between two contrasting water masses gave first hints for a metabolic
508 specialization of bacteroidetal populations in the cold nutrient richer BPLR and in the
509 subtropical oligotrophic NAST (Figure 4). However, for robust statistical comparisons a
510 higher sequencing coverage of the *Bacteroidetes* (meta)-genome is now needed.
511 It has been shown both on isolates and by culture-independent approaches that
512 aquatic *Bacteroidetes* are capable of degrading proteins, polysaccharides like chitin as

well as amino-sugars like N-acetylglucosamine (Cottrell and Kirchman 2000). Furthermore, the analysis of genomes from marine *Bacteroidetes* has revealed a generally high number of genes encoding peptidases and glycoside hydrolases (Supplementary Figure 3; (Bauer et al 2006, Gonzalez et al 2008, Woyke et al 2009). In our genomic fragments data set, we identified an overall broadly distributed potential for the degradation of proteins and the amino-sugar containing bacterial cell wall component peptidoglycan (see below), whereas predicted polysaccharide degradation activities were confined to three instances on a single fosmid.

Comparative genome analysis of marine *Flavobacteria* reveals that certain strains contain a higher number of glycoside hydrolase genes than others (Supplementary Figure 3). This apparent variability within *Flavobacteria* might be further illustrated by the clustering observed here of three GH genes on one genomic fragment originated from a *Polaribacter* organism (S3-860-D3) (Figure 3a). All three GH belong to a family of potential alpha-1,2-mannosidases (GH family 92). These enzymes can hydrolyze the alpha 1,2-linked side chains of mannan as well as the N-linkage to high mannose content oligosaccharides in glycoproteins. Members of the GH family 92 were identified in several flavobacterial strains but not in *P. irgensii* P-23, and in *P. dokdonensis* MED152 in one copy only (data not shown).

Extracellular enzymes are required to initiate the remineralization of high molecular weight OM (Arnosti 2003). Two of the mannosidases on fosmid S3-860-D3 are predicted to be secreted to the outer membrane, which implies that this *Polaribacter* strain may be involved in the processing of the extracellular polysaccharide mannan. This polysaccharide is one of the components of plant cell walls and also of the cell walls of some green (Frei and Preston 1968) and red algae where it occurs as sulfated xylomannan (Erra-Basells et al 2000).

538 A surprising accumulation of five sulfatase genes was found in the vicinity of
539 the GH genes on the *Polaribacter* S3-860-D3 fosmid (Figure 3a, Supplementary Table
540 4). This is a higher number on this short DNA-fragment than found in the entire
541 genomes of the *Flavobacteria P. irgensii* 23-P and *P. dokdonensis* MED152 (three and
542 two sulfatase genes, respectively). In general, marine planktonic *Bacteroidetes* are not
543 particularly enriched in sulfatases (Supplementary Figure 4), the ones with a higher
544 number are Flavobacteriales bacterium HTCC2170 (6.5 per Mbp, unpublished) and the
545 uncultured single amplified genome MS024-2A (3.1 per Mbp) (Woyke et al 2009). For
546 all of the fosmid-encoded sulfatases a signal-peptide was predicted, thus the proteins are
547 probably exported from the cytoplasm. Moreover, one of the sulfatase genes was found
548 next to a gene encoding an L-fucose transporter. Fucose is the fundamental subunit of
549 fucoidan, a class of fucose-enriched sulfated polysaccharide found in the extracellular
550 matrix of brown algae (Cumashi et al 2007).

551 In the marine realm *Planctomycetes* are the bacteria with the highest number of
552 sulfatase-encoding genes (Glockner et al 2003, Woebken et al 2007a). It has been
553 hypothesized that they are used to access more effectively the carbon skeleton of
554 sulfated polysaccharides as an energy source (Glockner et al 2003); given the genomic
555 neighborhood of the sulfatases on the fosmid S3-860-D3, this might also be the case for
556 the *Polaribacter*-related bacterium. Consistent with that, the abundance of the
557 *Polaribacter* genus was shown to be higher in the more productive BPLR and Arctic
558 provinces, where the phytoplankton abundance was higher (Gómez-Pereira et al 2010).
559 The *Polaribacter* fosmid presented here might indicate that certain marine
560 *Bacteroidetes* lineages or strains are enriched in sulfatases, possibly as a specialization
561 for the degradation of sulfated polysaccharides.

562 One of the *Bacteroidetes* distinctive features related with the degradation of
563 polysaccharides is a multicomponent outer membrane starch utilization system (Sus)

564 (Shipman et al 2000). The Sus system consist of several SusD-like outer membrane
565 proteins, encoded directly downstream of TonB-dependent outer membrane receptors
566 (SusC-like proteins), other outer membrane proteins (Shipman et al 2000); and in the
567 vicinity of glycoside hydrolases (Bauer et al 2006, McBride et al 2009, Xu et al 2003)
568 or peptidases (McBride et al 2009). In *B. thetaiotaomicron,* which is an anaerobic
569 human-associated *Bacteroidia* with an elaborate equipment for acquiring and
570 hydrolyzing otherwise indigestible dietary polysaccharides (Xu et al 2003), the Sus
571 outer membrane multicomponent system binds starch extracellularly before starting its
572 degradation (Anderson and Salyers 1989a, Anderson and Salyers 1989b, Cho and
573 Salyers 2001). *B. thetaiotaomicron* also encodes a high number of sulfatases (total 31),
574 and several of them are located contiguously with GHs and SusC-SusD pairs (eg. ORFs
575 BT1619 - BT1636). In the *Bacteroidetes* fosmids sequenced here only one potential
576 SusD-like protein was encoded in the vicinity of a TonB- dependent outer membrane
577 receptor (on the *Leeuwenhoekiella*-affiliated fosmid S18-997-C3). In the marine
578 *Bacteroidetes* genomes there is a clear increase in the SusCD-like pairs per Mbp with
579 the increasing number of GH per Mbp (Supplementary Figure 4a). This finding is in
580 line with the evidence that SusCD-like proteins function in conjunction with GH for
581 polysaccharides degradation. Moreover, no relation was found between the SusCD-like
582 ratios and peptidases (Supplementary Figure 4b), which seem to be less often localized
583 in the vicinity of SusCD-like (Bauer et al 2006, McBride et al 2009).

584 In our data set, *Bacteroidetes* appear to have an extensive proteolytic potential.
585 In contrast to the glycoside hydrolases, which were concentrated in one fosmid,
586 peptidases of a variety of families were identified in almost all *Bacteroidetes* fosmids
587 (Figure 4a) and several of them were predicted to be exported to the periplasmic space
588 or directed to the outer membrane (Supplementary Table 5). Exported and secreted
589 peptidases may liberate low molecular weight peptides that are either transported

590 directly into the cell or are hydrolyzed further to even smaller peptides or free amino
591 acids (Cottrell et al 2005). Dissolved proteins, peptides and amino-sugars may be
592 important sources of energy and nitrogen for marine bacteria (Benner and Kaiser 2003,
593 Obayashi and Suzuki 2005) and particularly, marine *Bacteroidetes* seem to have an
594 important role in the degradation of protein and N-acetyl-glucosamine (Cottrell and
595 Kirchman 2000, Pinhassi et al 1999). The most frequent peptidase identified in the
596 fosmids from S3 and generally in the genomes (data not shown) is a peptidase that lyses
597 bacterial cell wall peptidoglycans (family M23) that is used by certain bacteria as a
598 defensive or feeding mechanism. Additionally, on the flavobacterial fosmids several
599 peptidoglycan hydrolases were identified (Figure 3b), particularly the DE2 and VIS4
600 fosmids were enriched in peptidases and peptidoglycan hydrolases (Figure 3b, c). Both
601 DE2 fosmids encode the M23 peptidase, a glycopoprotein acetylglucosaminidase –
602 probably located in the outer membrane- and a muramic acid kinase (*anmK*) (only on
603 S18-873-7E), the latter two also on the VIS4 fosmid. It has been shown that *E. coli* can
604 grow in muramic acid, which is an amino-sugar that exclusively occurs in
605 peptidoglycan, as its sole carbon source (Dahl et al 2004). Muramic acid can be
606 recycled from the own cell wall or obtained from the environment (Uehara et al 2006)
607 and ultimately be shuttled into glycolysis (Jaeger et al 2005, Uehara et al 2006). In
608 marine systems, the bacterial assemblage can also readily hydrolyze and remineralize
609 both polysaccharide- and peptide-components of peptidoglycan (Nagata et al 2003).
610 Structural components of bacterial cells including membranes and peptidoglycan can be
611 introduced to seawater as DOM during bacterial decay due to protozoan grazing and
612 viral infection (Nagata and Kirchman 1999). In our data set *Flavobacteria* are
613 particularly enriched in peptidases and peptidoglycan hydrolases (Figure 4a, b), which
614 might reflect that under conditions of low phytoplankton derived polysaccharide

615 concentrations, as the NAST, they could thrive on bacterial derived organic matter as
616 carbon and nitrogen sources.

617 One of the prominent characteristics of marine *Bacteroidetes* is their association
618 with phytoplankton. Strains (Barbeyron et al 2001) and 16S rRNA gene sequences
619 (Longford et al 2007, Sapp et al 2007) of members of *Bacteroidetes* have been retrieved
620 from the surface of algae cells. Moreover, bacteroidetal cells have been shown to be
621 localized on the algal phycosphere (Biegala et al 2002, Gómez-Pereira et al 2010). One
622 type of compounds that might help to establish or maintain this physical association are
623 exopolysaccharides (EPS) (Nichols et al 2005) which are involved in the mediation of
624 cell/cell and cell/surface adhesion (Hall-Stoodley et al 2004). Additionally, EPS in
625 marine environments can be acting as ligands for trace metals or providing
626 cryoprotection for growth at low temperatures (Nichols et al 2005). Glycosyl
627 transferases participate in the biosynthesis of different kinds of polysaccharides,
628 including EPS. On the *Bacteroidetes* fosmids we identified several proteins possibly
629 involved in polysaccharide biosynthesis, which were in higher frequencies in S3 (Figure
630 4c). In the BPLR and ARCT provinces there is a higher phytoplankton biomass in
631 comparison with the NAST, therefore it is expected that more *Bacteroidetes* will tend to
632 attach to phytoplankton cells.

633 Particularly, the Sphingobacteria A fosmid had several predicted proteins related to
634 surface adhesion (Figure 3d). As additional evidence, cells of this clade were indeed
635 observed in the phycosphere of nanophytoplankton cells (Figure 2a). Moreover, 16S
636 rRNA gene sequences closely related to those of Sphingobacteria A had been retrieved
637 from the surface of the red macroalgae *Delisea pulchra* and the green algae *Ulva*
638 *australis* (e.g. DQ269044, (Longford et al 2007). The sphingobacterial fosmids have
639 peptidases of the family M22, which contains endopeptidases that cleave only proteins
640 that are *O*-sialoglycosylated. A sialoglycoprotein is a combination of sialic acid and

641 glycoprotein. Interestingly, the amino-sugar sialic acid has been identified in the
642 exopolymers produced by *Phaeocystis* cells (Orellana et al 2003). The Sphingobacteria
643 A cells might thrive attached to phytoplankton cells and degrade some of the released
644 compounds that contain sialic acid.

645 Another trait of *Bacteroidetes* related with the attachment of cells to surfaces is
646 gliding motility (Hall-Stoodley et al 2004). A total of 14 genes involved in gliding have
647 been identified and characterized in the soil bacterium *Flavobacterium johnsoniae,*
648 which can attach to and move along surfaces by gliding motility (McBride 2004). In this
649 study, four gliding genes were detected in the Sphingobacteria C fosmid (*gldK*, *gldM*,
650 *gldO* and *gldL*) and one each (*gldJ*) on the flavobacterial fosmids ANT and VIS6. In *F.*
651 *johnsoniae* disruption in some of the *gld* genes, including *gldK* and *gdL*, results in
652 complete deficiency in chitin utilization (Braun and McBride 2005), but the connection
653 between motility and chitin digestion is not clearly understood (Xie et al 2007). Based
654 on the genomic information, we propose that the marine planktonic *Sphingobacteria*
655 might be able to attach to surfaces and to degrade carbohydrates, sugars or proteins
656 released by the phytoplankton.
657
658 **Final remarks and outlook**

659 We identified a large proteolytic potential in marine *Bacteroidetes* corroborating
660 previous genomic (Bauer et al 2006) and experimental information (Cottrell and
661 Kirchman 2000). The comparison between the BPLR and NAST provinces revealed a
662 larger proteolytic potential in the oligotrophic NAST whereas a higher potential for
663 adhesion in the phytoplankton richer BPLR. Glycolytic potential was confined to a
664 member of the genus *Polaribacter*, which revealed a whole armory of proteins that
665 might function in degradation of sulfated polysaccharides. Our findings indicate a
666 pronounced metabolic specialization of the different bacteroidetal clades for

667 degradation of organic matter. The advent of single cell techniques like nanometer-scale
668 secondary-mass spectrometry (NanoSIMS) combined with FISH (Behrens et al 2008,
669 Musat et al 2008) have allowed the phylogenetic identification of microbial partners and
670 the monitoring of the transfer of metabolites between them (Behrens et al 2008). We
671 now suggest that tracer incubation experiments with radioactively or isotopicaly labeled
672 phytoplankton and bacterial derived organic matter should be performed and the
673 incorporation rates of single *Bacteroidetes* cells evaluated in order to address their role
674 in the marine carbon cycle more precisely.
675
676 **Acknowledgment**
677 We thank the Captain, Officers and Crew of the FS Maria S. Merian (cruise
678 MSM03/01) for their help during the cruise. We would like to thank J. Wulf and I.
679 Kostadinov for their help during sampling. We thank A. Meyerdierks for advices on
680 metagenomic libraries; and M Kube at the Max Planck Institute for Molecular Genetics
681 for access to their facilities. We acknowledge the Microbial Genomics Group at the
682 Max Planck Institute for Marine Microbiology for bioinformatic support, in particular to
683 H. Teeling, C. Quast, J. Waldman and M. Richter. This study was funded by the Max
684 Planck Society and by the NoE Marine Genomics Europe.
685
686 Supplementary information is available at the The ISME Journal website

687 **References**

688 Alonso C, Warnecke F, Amann R, Pernthaler J (2007). High local and global diversity
689 of Flavobacteria in marine plankton. *Environ Microbiol* **9:** 1253-1266.
690
691 Anderson KL, Salyers AA (1989a). Genetic evidence that outer membrane binding of
692 starch is required for starch utilization by *Bacteroides thetaiotaomicron*. *J Bacteriol* **171:**
693 3199-3204.
694
695 Anderson KL, Salyers AA (1989b). Biochemical evidence that starch breakdown by
696 Bacteroides thetaiotaomicron involves outer-membrane starch-binding sites and
697 periplasmic starch degrading enzymes. *J Bacteriol* **171:** 3192-3198.
698
699 Arnosti C (2003). Microbial extracellular enzymes and their role in dissolved organic
700 matter cycling. In: Findlay S, Sinsabaugh RL (eds). *Aquatic ecosystems: interactivity of*
701 *dissolved organic matter*. Academic Press: San Diego, CA. pp 316-342.
702
703 Asakawa S, Abe I, Kudoh Y, Kishi N, Wang YM, Kubota R *et al* (1997). Human BAC
704 library: Construction and rapid screening. *Gene* **191:** 69-79.
705
706 Barbeyron T, L'Haridon S, Corre E, Kloareg B, Potin P (2001). *Zobellia*
707 *galactanovorans* gen. nov., sp nov., a marine species of *Flavobacteriaceae* isolated
708 from a red alga, and classification of *Cytophaga uliginosa* (ZoBell and Upham 1944)
709 Reichenbach 1989 as *Zobellia uliginosa* gen. nov., comb. nov. *Int J Syst Evol Micr* **51:**
710 985-997.
711
712 Bauer M, Kube M, Teeling H, Richter M, Lombardot T, Allers E *et al* (2006). Whole
713 genome analysis of the marine *Bacteroidetes 'Gramella forsetii'* reveals adaptations to
714 degradation of polymeric organic matter. *Environ Microbiol* **8:** 2201-2213.
715
716 Behrens S, Losekann T, Pett-Ridge J, Weber PK, Ng WO, Stevenson BS *et al* (2008).
717 Linking microbial phylogeny to metabolic activity at the single-cell level by using
718 enhanced element labeling-catalyzed reporter deposition fluorescence in situ
719 hybridization (EL-FISH) and NanoSIMS. *Appl Environ Microb* **74:** 3143-3150.
720
721 Benner R, Pakulski JD, Mccarthy M, Hedges JI, Hatcher PG (1992). Bulk Chemical
722 Characteristics of Dissolved Organic-Matter in the Ocean. *Science* **255:** 1561-1564.
723
724 Benner R, Kaiser K (2003). Abundance of amino sugars and peptidoglycan in marine
725 particulate and dissolved organic matter. *Limnol Oceanogr* **48:** 118-128.
726
727 Biegala IC, Kennaway G, Alverca E, Lennon JF, Vaulot D, Simon N (2002).
728 Identification of bacteria associated with dinoflagellates (*Dinophyceae*) *Alexandrium*
729 spp. using tyramide signal amplification-fluorescent in situ hybridization and confocal
730 microscopy. *Journal of Phycology* **38:** 404-411.
731
732 Bowman JP (2006). The Marine Clade of the Family Flavobacteriaceae: The Genera
733 *Aequorivita, Arenibacter, Cellulophaga, Croceibacter, Formosa, Gelidibacter, Gillisia,*
734 *Maribacter, Mesonia, Muricauda, Polaribacter, Psychroflexus*, *Psychroserpens,*
735 *Robiginitalea, Salegentibacter, Tenacibaculum, Ulvibacter, Vitellibacter* and *Zobellia*.

736 In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds). *The*
737 *Prokaryotes*, third edn. Springer-Verlag: New York. pp 677-694.
738
739 Braun TF, McBride MJ (2005). Flavobacterium johnsoniae GldJ is a lipoprotein that is
740 required for gliding motility. *J Bacteriol* **187:** 2628-2637.
741
742 Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009).
743 The Carbohydrate-Active EnZymes database (CAZy): an expert resource for
744 Glycogenomics. *Nucleic Acids Res* **37:** D233-D238.
745
746 Chen XH, Zeng YH, Jiao NZ (2006). Development and evaluation of specific 16S
747 rDNA primers for marine *Cytophaga-Flavobacteria* cluster. *Mol Ecol Notes* **6:** 1278-
748 1281.
749
750 Chen XH, Zeng Y, Jiao N (2008). Characterization of *Cytophaga-Flavobacteria*
751 community structure in the bering sea by cluster-specific 16S rRNA gene amplification
752 analysis. *J Microbiol Biotechn* **18:** 194-198.
753
754 Cho KH, Salyers AA (2001). Biochemical analysis of interactions between outer
755 membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*.
756 *J Bacteriol* **183:** 7224-7230.
757
758 Cottrell MT, Kirchman DL (2000). Natural assemblages of marine proteobacteria and
759 members of the *Cytophaga-Flavobacter* cluster consuming low- and high-molecular-
760 weight dissolved organic matter. *Appl Environ Microb* **66:** 1692-1697.
761
762 Cottrell MT, Yu LY, Kirchman DL (2005). Sequence and expression analyses of
763 *Cytophaga*-like hydrolases in a Western arctic metagenornic library and the Sargasso
764 seat. *Appl Environ Microb* **71:** 8506-8513.
765
766 Crump BC, Armbrust EV, Baross JA (1999). Phylogenetic analysis of particle-attached
767 and free-living bacterial communities in the Columbia river, its estuary, and the adjacent
768 coastal ocean. *Appl Environ Microb* **65:** 3192-3204.
769
770 Cumashi A, Ushakova NA, Preobrazhenskaya ME, D'Incecco A, Piccoli A, Totani L *et*
771 *al* (2007). A comparative study of the anti-inflammatory, anticoagulant, antiangiogenic,
772 and antiadhesive activities of nine different fucoidans from brown seaweeds.
773 *Glycobiology* **17:** 541-552.
774
775 Dahl U, Jaeger T, Nguyen BT, Sattler JM, Mayer C (2004). Identification of a
776 phosphotransferase system of *Escherichia coli* required for growth on N-acetylmuramic
777 acid. *J Bacteriol* **186:** 2385-2392.
778
779 DeLong EF, Franks DG, Alldredge AL (1993). Phylogenetic diversity of aggregate-
780 attached vs free-living marine bacterial assemblages. *Limnol Oceanogr* **38:** 924-934.
781
782 Erra-Basells R, Kolender AA, Matulewicz MC, Nonami H, Cerezo AS (2000). Matrix-
783 assisted ultraviolet laser-desorption ionization time-of-flight mass spectrometry of
784 sulfated mannans from the red seaweed Nothogenia fastigiata. *Carbohyd Res* **329:** 157-
785 167.
786

787 Ferrandez A, Minambres B, Garcia B, Olivera ER, Luengo JM, Garcia JL *et al* (1998).
788 Catabolism of phenylacetic acid in *Escherichia coli* - Characterization of a new aerobic
789 hybrid pathway. *J Biol Chem* **273:** 25974-25986.
790
791 Frei E, Preston RD (1968). Non-cellulosic structural polysaccharides in algal cell walls.
792 Mannan in siphoneous green algae. *Proceedings of the Royal Society of London Series*
793 *B-Biological Sciences* **169:** 127-&.
794
795 Gardy JL, Spencer C, Wang K, Ester M, Tusnady GE, Simon I *et al* (2003). PSORT-B:
796 improving protein subcellular localization prediction for Gram-negative bacteria.
797 *Nucleic Acids Res* **31:** 3613-3617.
798
799 Glockner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W *et al* (2003).
800 Complete genome sequence of the marine planctomycete Pirellula sp strain 1. *P Natl*
801 *Acad Sci USA* **100:** 8298-8303.
802
803 Gomez-Consarnau L, Gonzalez JM, Coll-Llado M, Gourdon P, Pascher T, Neutze R *et*
804 *al* (2007). Light stimulates growth of proteorhodopsin-containing marine Flavobacteria.
805 *Nature* **445:** 210-213.
806
807 Gómez-Pereira PR, Fuchs B, Alonso C, Oliver MJ, van Beusekom JEE, Amann R
808 (2010). Distinct flavobacterial communities in contrasting water masses of the North
809 Atlantic Ocean. *Isme J* In press, doi:10.1038/ismej.2009.142.
810
811 Gonzalez JM, Fernandez-Gomez B, Fernandez-Guerra A, Gomez-Consarnau L,
812 Sanchez O, Coll-Llado M *et al* (2008). Genome analysis of the proteorhodopsin-
813 containing marine bacterium *Polaribacter* sp MED152 (*Flavobacteria*). *P Natl Acad Sci*
814 *USA* **105:** 8724-8729.
815
816 Gordon D, Abajian C, Green P (1998). Consed: A graphical tool for sequence finishing.
817 *Genome Research* **8:** 195-202.
818
819 Grossart HP, Levold F, Allgaier M, Simon M, Brinkhoff T (2005). Marine diatom
820 species harbour distinct bacterial communities. *Environ Microbiol* **7:** 860-873.
821
822 Grzymski JJ, Carter BJ, DeLong EF, Feldman RA, Ghadiri A, Murray AE (2006).
823 Comparative genomics of DNA fragments from six antarctic marine planktonic bacteria.
824 *Appl Environ Microb* **72:** 1532-1541.
825
826 Hall-Stoodley L, Costerton JW, Stoodley P (2004). Bacterial biofilms: From the natural
827 environment to infectious diseases. *Nat Rev Microbiol* **2:** 95-108.
828
829 Hedges JI (1992). Global biogeochemical cycles - progress and problems. *Mar Chem* **39:**
830 67-93.
831
832 Holtje JV (1998). Growth of the stress-bearing and shape-maintaining murein sacculus
833 of Escherichia coli. *Microbiol Mol Biol R* **62:** 181-.
834
835 Jaeger T, Arsic M, Mayer C (2005). Scission of the lactyl ether bond of N-
836 acetylmuramic acid by Escherichia coli "etherase". *J Biol Chem* **280:** 30100-30106.
837

838 Jing H, Takagi J, Liu JH, Lindgren S, Zhang RG, Joachimiak A *et al* (2002). Archaeal
839 surface layer proteins contain beta propeller, PKD, and beta helix domains and are
840 related to metazoan cell surface proteins. *Structure* **10:** 1453-1464.
841
842 Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A (2003).
843 Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Science* **12:**
844 1652-1662.
845
846 Kanamori T, Kanou N, Kusakabe S, Atomi H, Imanaka T (2005). Allophanate
847 hydrolase of Oleomonas sagaranensis involved in an ATP-dependent degradation
848 pathway specific to urea. *Fems Microbiology Letters* **245:** 61-65.
849
850 Keil RG, Kirchman DL (1991). Contribution of dissolved free amino-acids and
851 Aammonium to the nitrogen requirements of heterotrophic bacterioplankton. *Mar Ecol-*
852 *Prog Ser* **73:** 1-10.
853
854 Kirchman DL (2002). The ecology of *Cytophaga-Flavobacteria* in aquatic
855 environments. *Fems Microbiol Ecol* **39:** 91-100.
856
857 Kirchman DL, Yu LY, Cottrell MT (2003). Diversity and abundance of uncultured
858 Cytophaga-like bacteria in the Delaware Estuary. *Appl Environ Microb* **69:** 6587-6596.
859
860 Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001). Predicting
861 transmembrane protein topology with a hidden Markov model: Application to complete
862 genomes. *J Mol Biol* **305:** 567-580.
863
864 Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C *et al* (2004).
865 Versatile and open software for comparing large genomes. *Genome Biology* **5:** -.
866
867 Llobet-Brossa E, Rossello-Mora R, Amann R (1998). Microbial community
868 composition of Wadden Sea sediments as revealed by fluorescence in situ hybridization.
869 *Appl Environ Microb* **64:** 2691-2696.
870
871 Longford SR, Tujula NA, Crocetti GR, Holmes AJ, Holmstrom C, Kjelleberg S *et al*
872 (2007). Comparisons of diversity of bacterial communities associated with three sessile
873 marine eukaryotes. *Aquat Microb Ecol* **48:** 217-229.
874
875 Longhurst A (1995). Seasonal cycles of pelagic production and consumption. *Progress*
876 *In Oceanography* **36:** 77-167.
877
878 Lucea A, Duarte CM, Agusti S, Kennedy H (2005). Nutrient dynamics and ecosystem
879 metabolism in the Bay of Blanes (NW Mediterranean). *Biogeochemistry* **73:** 303-323.
880
881 Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al* (2004). ARB:
882 a software environment for sequence data. *Nucleic Acids Res* **32:** 1363-1371.
883
884 Malmstrom RR, Straza TRA, Cottrell MT, Kirchman DL (2007). Diversity, abundance,
885 and biomass production of bacterial groups in the western Arctic Ocean. *Aquat Microb*
886 *Ecol* **47:** 45-55.
887

888 Manz W, Amann R, Ludwig W, Wagner M, Schleifer KH (1992). Phylogenetic
889 Oligodeoxynucleotide Probes for the Major Subclasses of Proteobacteria - Problems and
890 Solutions. *Syst Appl Microbiol* **15:** 593-600.
891
892 McBride MJ (2004). Cytophaga-flavobacterium gliding motility. *Journal of Molecular*
893 *Microbiology and Biotechnology* **7:** 63-71.
894
895 McBride MJ, Xie G, Martens EC, Lapidus A, Henrissat B, Rhodes RG *et al* (2009).
896 Novel features of the polysaccharide digesting gliding bacterium *Flavobacterium*
897 *johnsoniae* as revealed by genome sequence analysis. *Appl Environ Microb* **75:** 6864-
898 6875.
899
900 McCarthy MD, Hedges JI, Benner R (1998). Major bacterial contribution to marine
901 dissolved organic nitrogen. *Science* **281:** 231-234.
902
903 Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J *et al* (2003).
904 GenDB - an open source genome annotation system for prokaryote genomes. *Nucleic*
905 *Acids Res* **31:** 2187-2195.
906
907 Musat N, Halm H, Winterholler B, Hoppe P, Peduzzi S, Hillion F *et al* (2008). A single-
908 cell view on the ecophysiology of anaerobic phototrophic bacteria. *P Natl Acad Sci USA*
909 **105:** 17861-17866.
910
911 Nagata T, Kirchman DL (1999). Bacterial mortality: A pathway for the formation of
912 refractory DOM? In: Brylinsky M, Bell C, Johnson-Green P (eds). *New frontiers in*
913 *microbial ecology: Proceedings of the 8th International Symposium on Microbial*
914 *Ecology. Atlantic Canada Society for Microbial Ecology*
915 pp 153-158.
916
917 Nagata T, Meon B, Kirchman DL (2003). Microbial degradation of peptidoglycan in
918 seawater. *Limnol Oceanogr* **48:** 745-754.
919
920 Nichols CM, Lardiere SG, Bowman JP, Nichols PD, Gibson JAE, Guezennec J (2005).
921 Chemical characterization of exopolysaccharides from Antarctic marine bacteria.
922 *Microbial Ecol* **49:** 578-589.
923
924 Nielsen H, Engelbrecht J, Brunak S, vonHeijne G (1997). Identification of prokaryotic
925 and eukaryotic signal peptides and prediction of their cleavage sites. *Protein*
926 *Engineering* **10:** 1-6.
927
928 Noguchi H, Park J, Takagi T (2006). MetaGene: prokaryotic gene finding from
929 environmental genome shotgun sequences. *Nucleic Acids Res* **34:** 5623-5630.
930
931 O'Sullivan LA, Fuller KE, Thomas EM, Turley CM, Fry JC, Weightman AJ (2004).
932 Distribution and culturability of the uncultivated 'AGG58 cluster' of Bacteroidetes
933 phylum in aquatic environments. *Fems Microbiol Ecol* **47:** 359-370.
934
935 Obayashi Y, Suzuki S (2005). Proteolytic enzymes in coastal surface seawater:
936 Significant activity of endopeptidases and exopeptidases. *Limnol Oceanogr* **50:** 722-726.
937

938 Ogawa H, Tanoue E (2003). Dissolved organic matter in oceanic waters. *J Oceanogr* **59:**
939 129-147.
940
941 Orellana MV, Lessard EJ, Dycus E, Chin WC, Foy MS, Verdugo P (2003). Tracing the
942 source and fate of biopolymers in seawater: application of an immunological technique.
943 *Mar Chem* **83:** 89-99.
944
945 Pao SS, Paulsen IT, Saier MH (1998). Major facilitator superfamily. *Microbiol Mol Biol*
946 *R* **62:** 1-.
947
948 Pedrotti ML, Beauvais S, Kerros ME, Iversen K, Peters F (2009). Bacterial colonization
949 of transparent exopolymeric particles in mesocosms under different turbulence
950 intensities and nutrient conditions. *Aquat Microb Ecol* **55:** 301-312.
951
952 Peplies J, Kottmann R, Ludwig W, Glockner FO (2008). A standard operating
953 procedure for phylogenetic inference (SOPPI) using (rRNA) marker genes. *Syst Appl*
954 *Microbiol* **31:** 251-257.
955
956 Pinhassi J, Azam F, Hemphala J, Long RA, Martinez J, Zweifel UL *et al* (1999).
957 Coupling between bacterioplankton species composition, population dynamics, and
958 organic matter degradation. *Aquat Microb Ecol* **17:** 13-26.
959
960 Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J *et al* (2007). SILVA:
961 a comprehensive online resource for quality checked and aligned ribosomal RNA
962 sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188-7196.
963
964 Quast C (2006). MicHanThi-design and implementation of a system for the prediction
965 of gene functions in genome annotation projects, University of Bremen, Bremen.
966
967 Rawlings ND, Morton FR, Kok CY, Kong J, Barrett AJ (2008). MEROPS: the
968 peptidase database. *Nucleic Acids Res* **36:** D320-D325.
969
970 Richter M, Lombardot T, Kostadinov I, Kottmann R, Duhaime MB, Peplies J *et al*
971 (2008). JCoast - A biologist-centric software tool for data mining and comparison of
972 prokaryotic (meta) genomes. *Bmc Bioinformatics* **9:** -.
973
974 Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA *et al* (2000).
975 Artemis: sequence visualization and annotation. *Bioinformatics* **16:** 944-945.
976
977 Salzberg SL, Delcher AL, Kasif S, White O (1998). Microbial gene identification using
978 interpolated Markov models. *Nucleic Acids Res* **26:** 544-548.
979
980 Sapp M, Schwaderer AS, Wiltshire KH, Hoppe HG, Gerdts G, Wichels A (2007).
981 Species-specific bacterial communities in the phycosphere of microalgae? *Microbial*
982 *Ecol* **53:** 683-699.
983
984 Schattenhofer M, Fuchs BM, Amann R, Zubkov MV, Tarran GA, Pernthaler J (2009).
985 Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean.
986 *Environ Microbiol* **11:** 2078-2093.
987

988 Seydel A, Gounon P, Pugsley AP (1999). Testing the '+2 rule' for lipoprotein sorting in
989 the Escherichia coli cell envelope with a new genetic selection. *Mol Microbiol* **34:** 810-
990 821.
991
992 Shaw AK, Halpern AL, Beeson K, Tran B, Venter JC, Martiny JBH (2008). It's all
993 relative: ranking the diversity of aquatic bacterial communities. *Environ Microbiol* **10:**
994 2200-2210.
995
996 Shipman JA, Berleman JE, Salyers AA (2000). Characterization of four outer
997 membrane proteins involved in binding starch to the cell surface of *Bacteroides*
998 *thetaiotaomicron*. *J Bacteriol* **182:** 5365-5372.
999
1000 Simon M, Glockner FO, Amann R (1999). Different community structure and
1001 temperature optima of heterotrophic picoplankton in various regions of the Southern
1002 Ocean. *Aquat Microb Ecol* **18:** 275-284.
1003
1004 Tanoue E, Nishiyama S, Kamo M, Tsugita A (1995). Bacterial membranes - possible
1005 source of a major dissolved protein in seawater. *Geochim Cosmochim Ac* **59:** 2643-2648.
1006
1007 Uehara T, Suefuji K, Jaeger T, Mayer C, Park JT (2006). MurQ etherase is required by
1008 *Escherichia coli* in order to metabolize anhydro-N-acetylmuramic acid obtained either
1009 from the environment or from its own cell wall. *J Bacteriol* **188:** 1660-1662.
1010
1011 Woebken D, Fuchs BA, Kuypers MAA, Amann R (2007a). Potential interactions of
1012 particle-associated anammox bacteria with bacterial and archaeal partners in the
1013 Namibian upwelling system. *Appl Environ Microb* **73:** 4648-4657.
1014
1015 Woebken D, Teeling H, Wecker P, Dumitriu A, Kostadinov I, DeLong EF *et al* (2007b).
1016 Fosmids of novel marine *Planctomycetes* from the Namibian and Oregon coast
1017 upwelling systems and their cross-comparison with planctomycete genomes. *Isme J* **1:**
1018 419-435.
1019
1020 Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, Kiss H *et al* (2009). Assembling
1021 the marine metagenome, one cell at a time. *PloS one* **4:** e5299.
1022
1023 Xie G, Bruce DC, Challacombe JF, Chertkov O, Detter JC, Gilna P *et al* (2007).
1024 Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Appl*
1025 *Environ Microb* **73:** 3536-3546.
1026
1027 Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC *et al* (2003). A
1028 genomic view of the human *Bacteroides thetaiotaomicron* symbiosis. *Science* **299:**
1029 2074-2076.
1030
1031 Zhang CT, Wang J (2000). Recognition of protein coding genes in the yeast genome at
1032 better than 95% accuracy based an the Z curve. *Nucleic Acids Res* **28:** 2804-2814.
1033
1034 Zhou JZ, Bruns MA, Tiedje JM (1996). DNA recovery from soils of diverse
1035 composition. *Appl Environ Microb* **62:** 316-322.
1036
1037 Zhu HQ, Hu GQ, Yang YF, Wang J, She ZS (2007). MED: a new non-supervised gene
1038 prediction algorithm for bacterial and archaeal genomes. *Bmc Bioinformatics* **8:**

Titles and legends to figures

Figure 1: Phylogenetic association of the fosmids 16S rRNA gene sequences retrieved in this study (S3 in green and S18 in blue) and of fully genome sequenced *Bacteroidetes* strains (orange). Uncultured clade names given to the right. Consensus tree was built according to (Peplies et al 2008) with full length sequences based on neighbor joining and maximum likelihood methods, calculated without and with 50% bacteria positional conservatory and 30% *Bacteroidetes* positional conservatory filters. Maximum likelihood was constructed with the RaxML algorithm and bootstrap.

Figure 2: Epifluorescence micrographs of cells hybridized with oligonucleotide probes for the rRNA of fosmids **(a)** S3-892-858-M; **(b)** S18-1001-F9; **(c)** S3-805-E4; **(d)** S18-873-E7**; (e)** S3-816-G8; **(f)** S18-848-B2. Blue, DAPI signals of DNA containing cells; and green, signals of CARD-FISH positive cells with the probes: a, c, e, S3, 20 m (green margin); b, d, f: S18, 20 m (blue margin). Bar = 2 µm.

Figure 3: Genomic content of selected fosmids, **(a)** *Polaribacter*, **(b)** DE2, **(c)** VIS4, **(d)** Sphingobacteria A, **(e)** Sphingobacteria C, and **(f)** *Cytophagia* showing ORFs involved in degradation of OM, attachment, transport and pigments discussed in text indicating ORF number. Double lines separating arrows indicate that ORFs are not contiguous. A dashed diagonal line indicates were fragments have been merged.

Figure 4: Comparisons of *Bacteroidetes* metagenome between S3 and S18 of **(a)** peptidases, **(b)** peptidoglycan hydrolases and, **(c)** adhesion proteins expressed as the percentage total ORFs. Abbreviations in legends correspond to Pfam models in **a**, **c** and glucosaminidase in **b** (E value < E-10, ***** E value > E-10) or gene name in **b**. *anmK*: anhydro-N-acetylmuramic acid kinase; *murQ:* N-acetylmuramic acid 6-phosphate etherase.

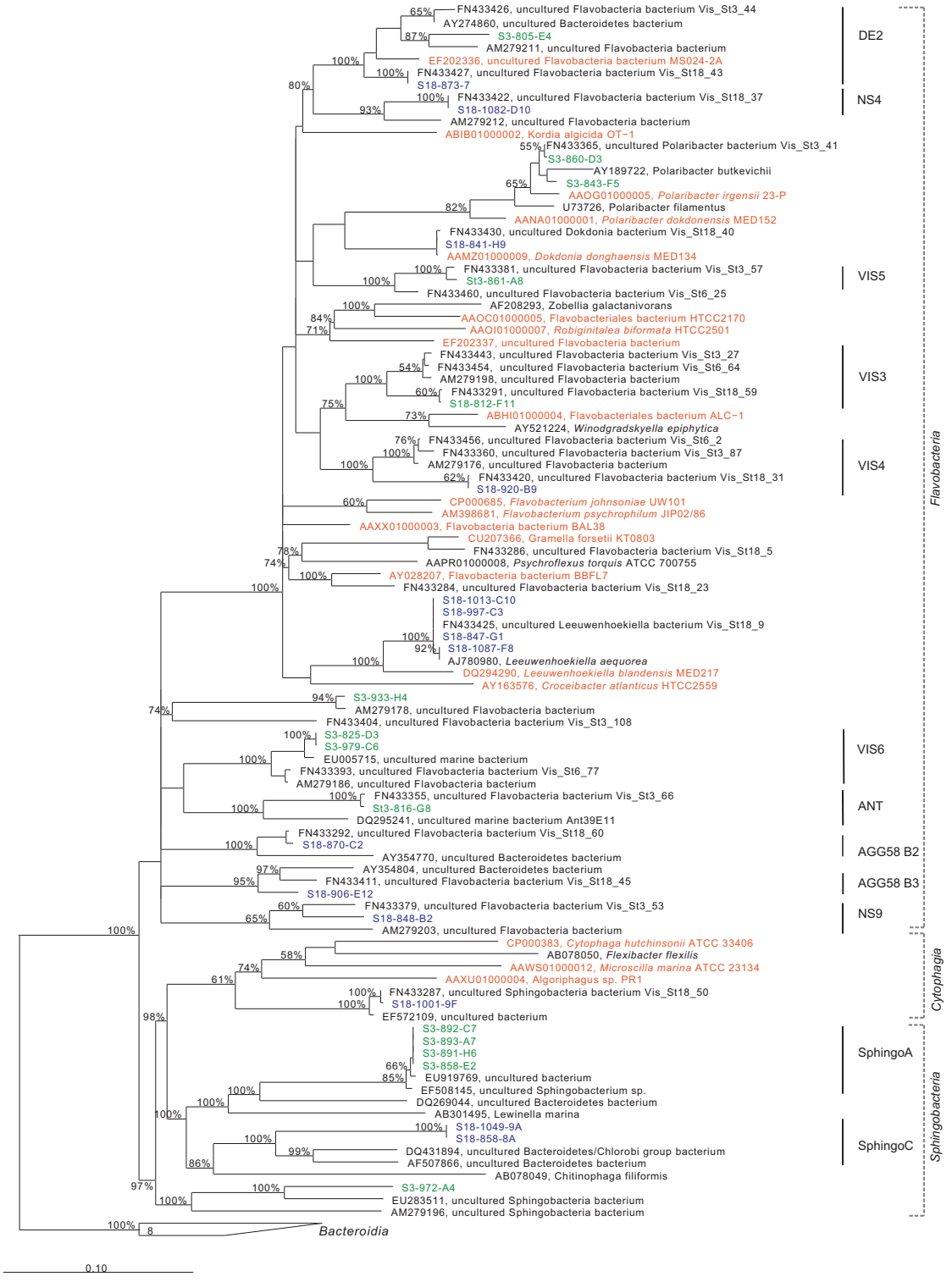

FN433426, uncultured Flavobacteria bacterium Vis_St3_44
AY274860, uncultured Bacteroidetes bacterium
S3-805-E4
AM279211, uncultured Flavobacteria bacterium
EF202336, uncultured Flavobacteria bacterium MS024-2A
FN433427, uncultured Flavobacteria bacterium Vis_St18_43
S18-873-7
FN433422, uncultured Flavobacteria bacterium Vis_St18_37
S18-1082-D10
AM279212, uncultured Flavobacteria bacterium
ABIB01000002, Kordia algicida OT-1
FN433365, uncultured Polaribacter bacterium Vis_St3_41
S3-860-D3
AY189722, Polaribacter butkevichii
S3-843-F5
AAOG01000005, Polaribacter irgensii 23-P
U73726, Polaribacter filamentus
AANA01000001, Polaribacter dokdonensis MED152
FN433430, uncultured Dokdonia bacterium Vis_St18_40
S18-841-H9
AAMZ01000009, Dokdonia donghaensis MED134
FN433381, uncultured Flavobacteria bacterium Vis_St3_57
St3-861-A8
FN433460, uncultured Flavobacteria bacterium Vis_St6_25
AF208293, Zobellia galactanivorans
AAOC01000005, Flavobacteriales bacterium HTCC2170
AAOI01000007, Robiginitalea biformata HTCC2501
EF202337, uncultured Flavobacteria bacterium
FN433443, uncultured Flavobacteria bacterium Vis_St3_27
FN433454, uncultured Flavobacteria bacterium Vis_St6_64
AM279198, uncultured Flavobacteria bacterium
FN433291, uncultured Flavobacteria bacterium Vis_St18_59
S18-812-F11
ABHI01000004, Flavobacteriales bacterium ALC-1
AY521224, Winodgradskyella epiphytica
FN433456, uncultured Flavobacteria bacterium Vis_St6_2
FN433360, uncultured Flavobacteria bacterium Vis_St3_87
AM279176, uncultured Flavobacteria bacterium
FN433420, uncultured Flavobacteria bacterium Vis_St18_31
S18-920-B9
CP000685, Flavobacterium johnsoniae UW101
AM398681, Flavobacterium psychrophilum JIP02/86
AAXX01000003, Flavobacteria bacterium BAL38
CU207366, Gramella forsetii KT0803
FN433286, uncultured Flavobacteria bacterium Vis_St18_5
AAPR01000008, Psychroflexus torquis ATCC 700755
AY028207, Flavobacteria bacterium BBFL7
FN433284, uncultured Flavobacteria bacterium Vis_St18_23
S18-1013-C10
S18-997-C3
FN433425, uncultured Leeuwenhoekiella bacterium Vis_St18_9
S18-847-G1
S18-1087-F8
AJ780980, Leeuwenhoekiella aequorea
DQ294290, Leeuwenhoekiella blandensis MED217
AY163576, Croceibacter atlanticus HTCC2559
S3-933-H4
AM279178, uncultured Flavobacteria bacterium
FN433404, uncultured Flavobacteria bacterium Vis_St3_108
S3-825-D3
S3-979-C6
EU005715, uncultured marine bacterium
FN433393, uncultured Flavobacteria bacterium Vis_St6_77
AM279186, uncultured Flavobacteria bacterium
FN433355, uncultured Flavobacteria bacterium Vis_St3_66
St3-816-G8
DQ295241, uncultured marine bacterium Ant39E11
FN433292, uncultured Flavobacteria bacterium Vis_St18_60
S18-870-C2
AY354770, uncultured Bacteroidetes bacterium
AY354804, uncultured Bacteroidetes bacterium
FN433411, uncultured Flavobacteria bacterium Vis_St18_45
S18-906-E12
FN433379, uncultured Flavobacteria bacterium Vis_St3_53
S18-848-B2
AM279203, uncultured Flavobacteria bacterium
CP000383, Cytophaga hutchinsonii ATCC 33406
AB078050, Flexibacter flexilis
AAWS01000012, Microscilla marina ATCC 23134
AAXU01000004, Algoriphagus sp. PR1
FN433287, uncultured Sphingobacteria bacterium Vis_St18_50
S18-1001-9F
EF572109, uncultured bacterium
S3-892-C7
S3-893-A7
S3-891-H6
S3-858-E2
EU919769, uncultured bacterium
EF508145, uncultured Sphingobacterium sp.
DQ269044, uncultured Bacteroidetes bacterium
AB301495, Lewinella marina
S18-1049-9A
S18-858-8A
DQ431894, uncultured Bacteroidetes/Chlorobi group bacterium
AF507866, uncultured Bacteroidetes bacterium
AB078049, Chitinophaga filiformis
S3-972-A4
EU283511, uncultured Sphingobacteria bacterium
AM279196, uncultured Sphingobacteria bacterium
Bacteroidia
0.10
DE2
NS4
VIS5
VIS3
VIS4
VIS6
ANT
AGG58 B2
AGG58 B3
NS9
SphingoA
SphingoC
Flavobacteria
Cytophagia
Sphingobacteria

Figure 1

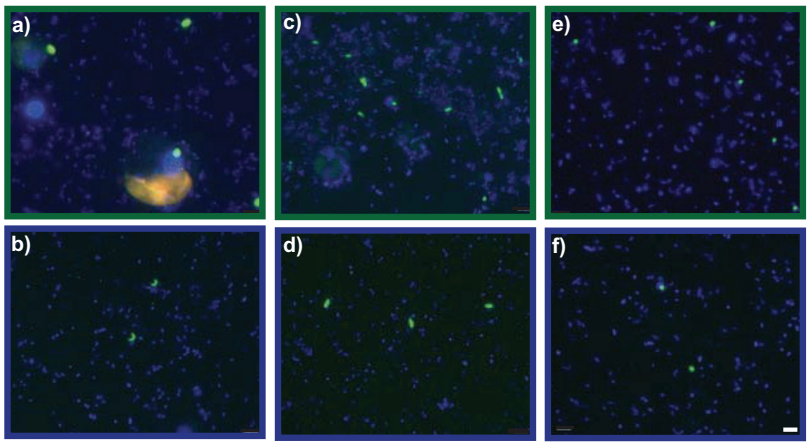


Figure 2

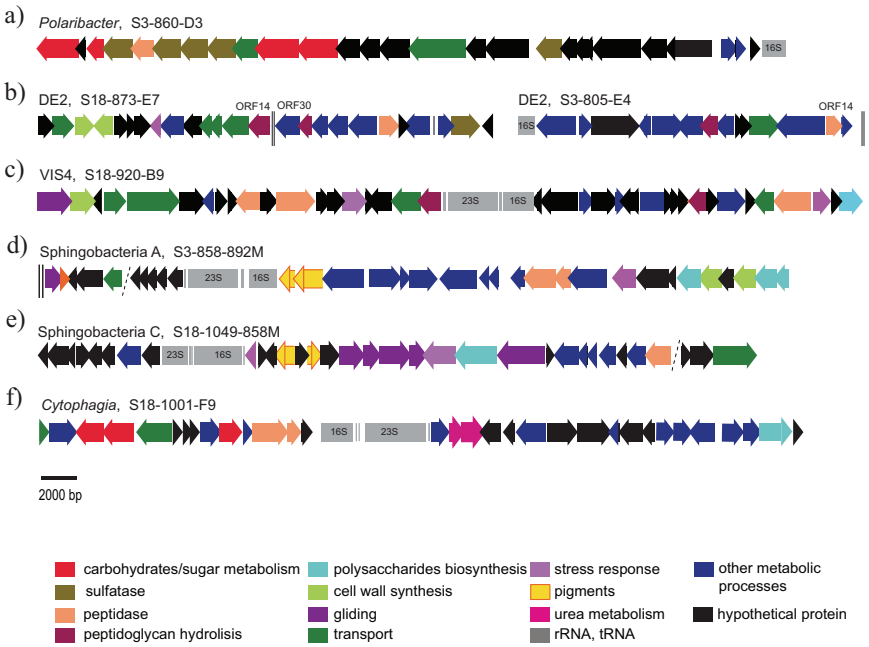

a)
Polaribacter, S3-860-D3
16S
b)
DE2, S18-873-E7
ORF14
ORF30
DE2, S3-805-E4
16S
ORF14
c)
VIS4, S18-920-B9
23S
16S
d)
Sphingobacteria A, S3-858-892M
23S
16S
e)
Sphingobacteria C, S18-1049-858M
23S
16S
f)
Cytophagia, S18-1001-F9
16S
23S
2000 bp
carbohydrates/sugar metabolism
sulfatase
peptidase
peptidoglycan hydrolisis
polysaccharides biosynthesis
cell wall synthesis
gliding
transport
stress response
pigments
urea metabolism
rRNA, tRNA
other metabolic processes
hypothetical protein

Figure 3

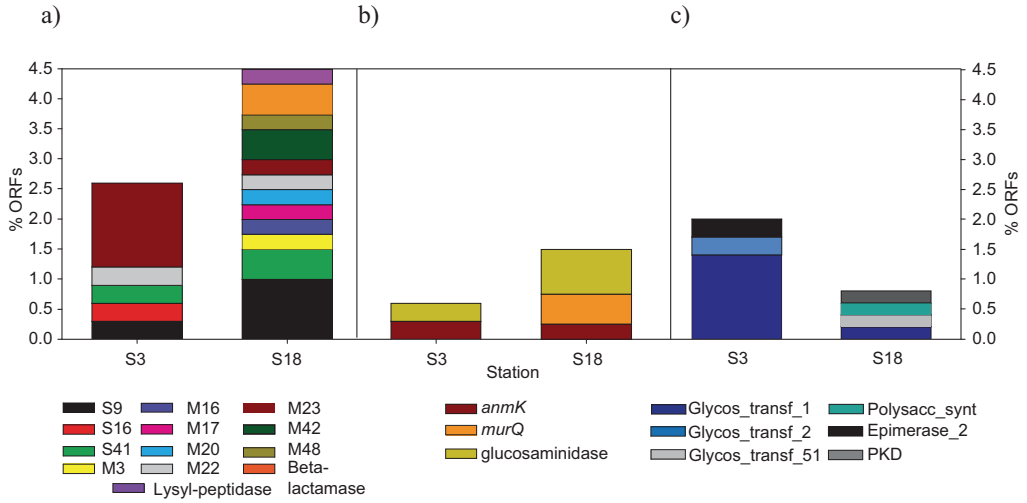

a)
b)
c)
% ORFs
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
S3
S18
Station
S9
S16
S41
M3
M16
M17
M20
M22
M23
M42
M48
Beta-lactamase
Lysyl-peptidase
anmK
murQ
glucosaminidase
Glycos_transf_1
Glycos_transf_2
Glycos_transf_51
Polysacc_synt
Epimerase_2
PKD

Figure 4

Table 1: Oceanic province, coordinates, date and time (UTC), temperature (temp), salinity (sal), dissolved oxygen (DO), inorganic nutrient concentrations (PO_4^{3-} , NO_3^- +NO_2^-, Si, NH_4^+) in the stations from which the metagenomic libraries were constructed indicating volume of sample filtered, library size, number of screened and sequenced clones (in brackets number of analyzed clones after merging). Time indicated is the time when sampling procedure started. UTC: Coordinated Universal Time; PSU, practical salinity units.

	S3	**S18**
Province	BPLR	NAST
Coordinates		
latitude	65°52,64'N	34°04,43'N
longitude	29°56,54'W	30°00,09'W
Date; time (UCT)	22.09.06; 19:00	30.09.06; 08:00
Temp (°C)	0.7	23.5
Sal (PSU)	33.0	36.4
DO (ml l^{-1})	7.7	4.7
PO_4^{3-} (µM)	0.4	0.01
NO_3^-+NO_2^- (µM)	3.0	0.03
NH_4^+ (µM)	0.29	0.23
Si (µM)	2.9	0.3
Volume (L)	90	87
Library size	35,000	50,000
Fosmids screened	15,552	27,648
Fosmids sequenced	13 (9)	16 (12)

Table 2: Chracterization of the *Bacteroidetes* fosmids retrieved in this study

		Genomic characteristics			Closest relatives					Abundance		
Fosmid	**Clade [1]**	**Length (kbp)**	**ORFs**	**GC (%)**	**genome-sequenced strain [2]**	**identity (%)**	**VISION clone**	**Accession N°**	**identity (%)**	**Probe**	**Cell ml^{-1} x 10^{2}**	
Province / Station											**BPLR**	**NAST**
BPLR / S3												
S3-843-F5	*Polaribacter*	41.1	32	32.6	*Polaribacter irgensii* 23-P	98	VIS-St3-9	FN433361	97	POL740$^{+,3}$	500 ± 150	68 ± 15
S3-860-D3	*Polaribacter*	44.9	27	32.5	*Polaribacter irgensii* 23-P	98	VIS-St3-41	FN433365	99	POL740$^{+,3}$	500 ± 150	68 ± 15
S3-805-E4	DE2	39.6	33	35.3	Flavobacteria MS024-2A	94	VIS-St18-44	FN433426	93	F805E4-471	25 ± 1.1	n.d.
S3-861-A8	VIS5	39.2	29	33.1	*Dokdonia donghaensis* MED134	92	VIS-St3-57	FN433381	99	VIS5-586$^{+,4}$	5.2 ± 1.4	n.d.
S3-933-H4	*Flavobacteria*	27.8	29	35.1	*Psychroflexus torquis* ATCC 700755	85	VIS-St3-103	FN433403	85	-	-	-
S3-979-C6 (M^{1})	VIS6	36.8	27	43.1	*Kordia algicida* OT-1	87	VIS-St6-77	FN433393	96	VIS6-814$^{+,4}$	17 ± 5.9	n.d.
S3-825-D3 (M^{1})	VIS6	31.9	23 (4)	43.3	*Kordia algicida* OT-1	87	VIS-St6-77	FN433393	96	VIS6-814$^{+,4}$	17 ± 5.9	n.d.
S3-816-G8	ANT	39.3	28	38.7	Flavobacteria bacterium BAL38	86	VIS-St3-66	FN433355	99	ANT-981 ANT-72	25 ± 10	n.d.
S3-858-E2 (M^{2})	Spingobacteria A	36.7	34 (33)	37.2	*Candidatus* Amoebophilus asiaticus 5a2	81	VIS-St3-37	FN433448	99	SPA-730	9.5	n.d.
S3-892-C7 (M^{2})	Spingobacteria A	38.7	30 (19)	37.3	*Candidatus* Amoebophilus asiaticus 5a2	81	VIS-St3-37	FN433448	99	SPA-730	9.5	n.d.
S3-893-A7 *	Spingobacteria A	36.0	24	37.7	*Candidatus* Amoebophilus asiaticus 5a2	81	VIS-St3-37	FN433448	99	SPA-730	9.5	n.d.
S3-891-H6 *	Spingobacteria A	37.2	25	37.3	*Candidatus* Amoebophilus asiaticus 5a2	81	VIS-St3-37	FN433448	99	SPA-730	9.5	n.d.
S3-972-A4	Sphingobacteria	30.8	32	38.4	*Microscilla marina* ATCC 23134	82	none	-	-	-	-	-

Table 2: Chracterization of the *Bacteroidetes* fosmids retrieved in this study, continued

		Genomic characteristics			Closest relatives					Abundance		
Fosmid	**Clade [1]**	**Length (kbp)**	**ORFs**	**GC (%)**	**genome-sequenced strain [2]**	**identity (%)**	**VISION clone**	**Accession N°**	**identity (%)**	**Probe**	**Cell ml^{-1}x 10^{2}**	
Province / Station											**BPLR**	**NAST**
NAST / S18												
S18-841-H9	*Dokdonia*	43.1	39	39.8	*Dokdonia donghaensis* MED134	100	VIS-St18-40	FN433430	99	DOK827$^{+,4}$	n.d.	<1
S18-997-C3	*Leeuwenhoekiella*	37.0	20	36.4	*Leeuwenhoekiella blandensis* MED217	96	VIS-St18-9	FN433425	99	LEE82$^{+,4}$	n.d.	<1
S18-1087-F8 (M^{3})	*Leeuwenhoekiella*	41.5	39	37.6	*Leeuwenhoekiella blandensis* MED217	96	VIS-St18-9	FN433425	99	LEE82$^{+,4}$	n.d.	<1
S18-1013-C10 (M^{3})	*Leeuwenhoekiella*	35.1	36 (19)	35.9	Leeuwenhoekiella blandensis MED217	96	VIS-St18-9	FN433425	99	LEE82$^{+,4}$	n.d.	<1
S18-847-G1 *	*Leeuwenhoekiella*	29.9	27	37.2	*Leeuwenhoekiella blandensis* MED217	96	VIS-St18-9	FN433425	99	LEE82$^{+,4}$	n.d.	<1
S18-1082-D10	NS4	32.9	28	31.9	*Kordia algicida* OT-1	90	VIS-St18-37	FN433422	99	-	-	-
S18-873-E7	DE2	45.8	34	33.9	Flavobacteria MS024-2A	96	VIS-St18-43	FN433427	100	F8737E-843	n.d.	13 ± 1.1
S18-920-B9	VIS4	46.3	40	31.8	Flavobacteriales ALC-1	91	VIS-St18-31	FN433420	99	FL197$^{+,4}$	73 ± 15	38 ± 6.2
S18-812-F11	VIS3	40.9	35	39.2	Flavobacteriales ALC-1	93	VIS-St18-59	FN433291	99	-	-	-
S18-848-B2	NS9	41.6	34	57.1	Flavobacteria BBFL7	86	VIS-St3-53	FN433379	95	NS9-664	7.3 ± 2.8	8.2 ± 0.2
S18-906-E12	AGG58 branch 3	43.4	19	37	*Kordia algicida* OT-1	86	VIS-St18-45	FN433411	96	AGG58-742	0.4 ± 0.08	5.4 ± 0.2
S18-870-C2	AGG58 branch 2	39.8	29	31.4	*Gramella forsetii* KT0803	86	VIS-St18-60	FN433292	99	-	-	-
S18-1049-A9 (M^{4})	Sphingobacteria C	36.9	31	45.6	*Algoriphagus* sp. PR1	82	none	-	-	SPC-814	n.d.	<1
S18-858-A8 (M^{4})	Sphingobacteria C	37.2	30 (3)	43.7	*Algoriphagus* sp. PR1	82	none	-	-	SPC-814	n.d.	<1
S18-1001-F9	*Cytophagia*	37.2	32	38.1	*Algoriphagus* sp. PR1	86	VIS-St18-50	FN433287	99	CYT-734 CYT-1462	17 ± 0.2	5.4 ± 0.6

1, defined in Gómez-Pereira et al., 2010 and indicated in Fig 1;**2**, accession numbers given in Fig 1;**3**, Malmstrom et al., 2007; **4**, Gómez-Pereira et al., 2010; +, counts published in Gómez-Pereira et al., 2010.

M, merged fosmids in **M^{1}**, S3-825-979M; **M^{2}**, S3-858-892; **M^{3}**, S18-1087-1013; **M^{4}**, S18-1049-858; number of ORFs merged given in brackets

Supplementary Information S1

Material and Methods

DNA extraction

Microorganisms were collected on cellulose acetate filters (142 mm diameter, 0.2 µm pore size; Sartorius, Goettingen, Germany) and stored at -80°C for further processing. Genomic DNA was extracted from the 0.2 µm filters according to (Zhou et al 1996). Half of the filter was incubated at 37°C in 18 ml of extraction buffer containing 100 mM Tris-HCl (pH 8.0), 100 mM EDTA (pH 8.0), 100 mM sodium-phosphate (pH 8.0), 1.5 M NaCl and 1% hexadecyltrimethylammonium -bromide, and 70 µl of proteinase K (600 mAnson U ml^{-1}, Merck, Darmstadt, Germany). Two percent of sodium dodecyl sulfate was added and incubated at 65°C for 2 hours. The supernatant was collected by centrifugation and DNA was extracted for a second time. The lysate was extracted with an equal volume of chloroform/isoamylalcohol (24:1, v/v), repeated three times. DNA was precipitated with 2-propanol at room temperature for one hour, collected by centrifugation and resuspended in 1x TE buffer (10 mM Tris-HCl, pH 8 and 1m EDTA, pH 8) overnight at 4°C. All steps were performed with cell saver tips to avoid shearing the DNA.

Library construction

Fosmids libraries were constructed with the Copy Control Fosmid Library Production Kit (EPICENTRE Biotechnologies, Madison, WI, USA) following instructions of the manufacturer and as previously described by (Woebken et al 2007). Briefly, library construction was done immediately after the overnight resuspension of the DNA. DNA ends were blunt-ended according to the protocol of the Copy Control Fosmid Library Production Kit. DNA of 30-45 kb was selected for ligation by pulsed-field gel electrophoresis (PFGE) with the following conditions: gel of 1% low melting

point agarose (SeaPlaque GTG, Lonza, Rockland, ME, USA) in 0.5x TBE and PFGE program of 12 h running time at 14°C angle of 1201; 6 V/cm; initial switch time 1 s, final switch time 10 s. The gel was stained with SYBR Green I (Invitrogen, Eugene, OR, USA) and visualized with a dark reader trans iluminator (Clare Chemical Research, Dolores, CO, USA). Band of the desired size was excised from the gel and DNA equilibrated three times with 1 x TE (30 min each) and agarose digested with beta agarase (1000 U/ml; 1 ul/100 mg agarose gel) and incubated for 2 hrs. DNA was concentrated and washed with 1 x TE using Microcon tube (Millipore, Billerica, MA, USA) and eluted in PCR water.

DNA was ligated into CopyControl pCC1FOS vectors (EPICENTRE Biotechnologies, Madison) at 4°C for 2 days. The vectors were packaged into the MaxPlax lambda phage (EPICENTRE Biotechnologies). *Escherichia coli* EPI300-T1R (EPICENTRE Biotechnologies) were infected with the phages upon reaching an OD of 0.8–1.0 and plated onto LB-chloramphenicol plates (12.5 μg ml^{-1}). Grown colonies were transferred into LB medium containing chloramphenicol (12.5 μg ml^{-1}), MgSO4 (10 mM), and glycerol (8%) and stored at -80°C. Insert size of 20 random clones was checked in a 1% agarose gel using BAC tracker (EPICENTRE Biotechnologies).

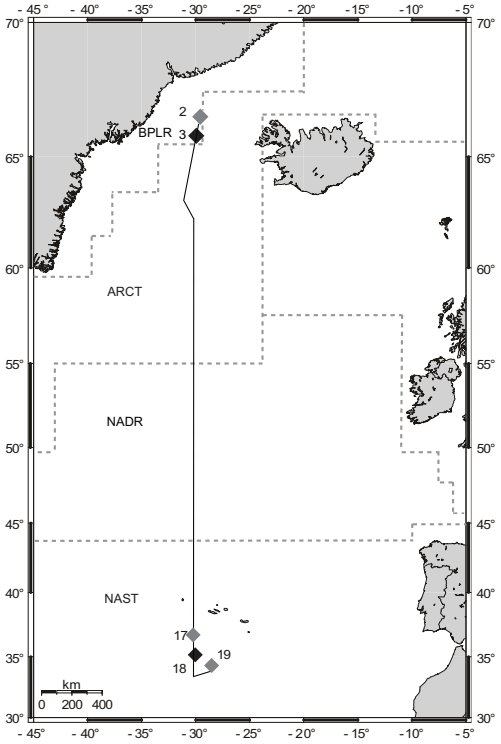
Supplementary Figures

Supplementary Figure 1: Map indicating the VISION cruise track and boundaries of the oceanic provinces Boreal Polar (BPLR), Arctic (ARCT), North Atlantic Drift (NADR) and North Atlantic Subtropical (NAST) provinces. Symbol colours represent analyses done in the sampled stations: grey indicate stations FISH counts were performed; and black indicate stations where metagenomics fosmids libraries were constructed

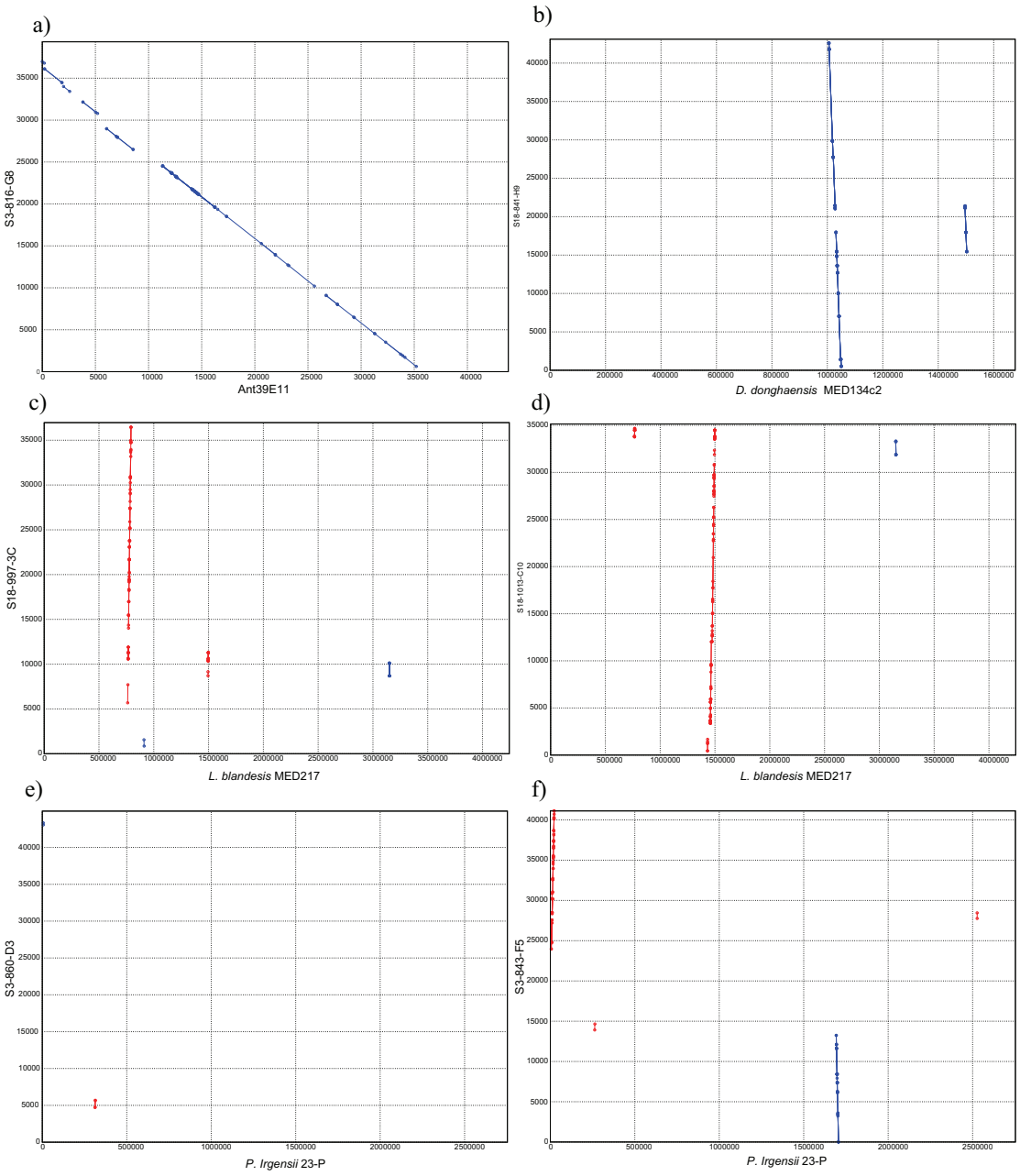
Supplementary Figure 2: Pairwise amino acid sequence alignment of **(a)** Uncultured genomic fragments Ant39E11 and S18-816-G8; **(b)** *Dokdonia donghanensis* MED134 (contig 2) and fosmid S18-841-H9; *L. blandesis* MED152 with fosmids **(c)** S18-997-C3 and **(d)** S18-1013-C10; *Polaribacter irgensii* 23-P with fosmids **(e)** S3-860-D3 and **(f)** S3-845-F5. Fosmid S18-1013-C10 presented an example of the three highly similar fosmids to *Leuwenhokiella* (S18-1087-F8 and S18-847-G1). Aligned regions in the same direction plotted in red and reverse complement in blue.

Supplementary Figure 3: Number per Mbp of glycoside hydrolases, peptidases and sulfatases in *Bacteroidetes* genomes from different habitats. Based on Pfam profiles searches and CAZy family members.

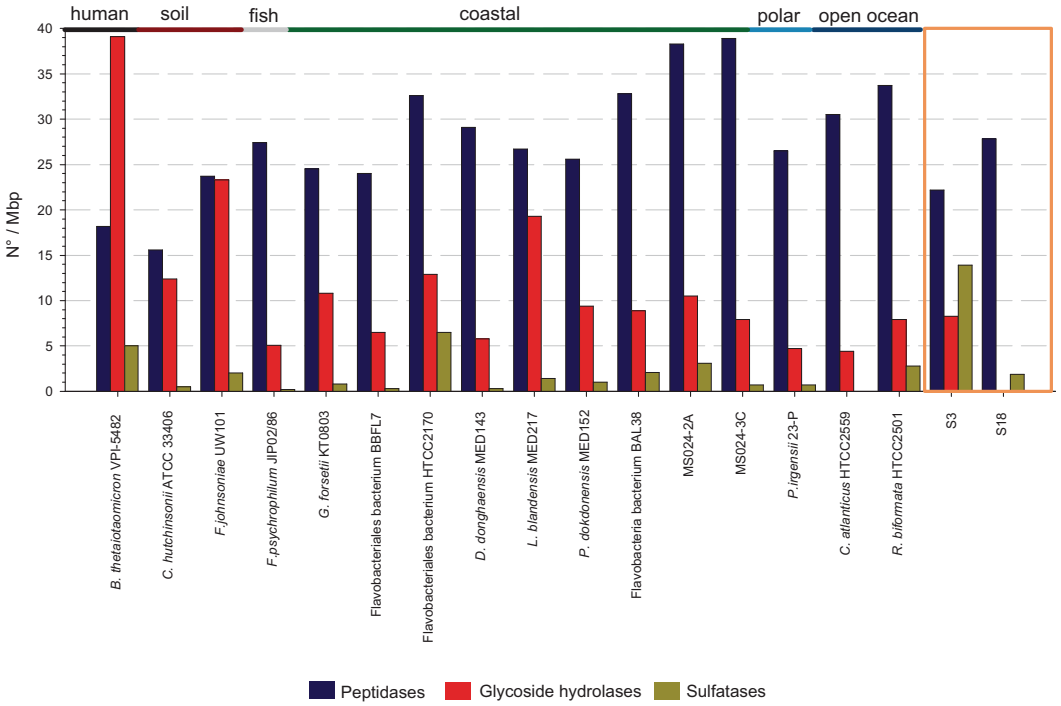
Supplementary Figure 4: Number per Mbp of SusCD-like versus the number per Mbp of **(a)** glycoside hydrolases and **(b)** peptidases in marine *Bacteroidetes* genomes (coastal and open ocean strains shown in Supplementary Figure 3).



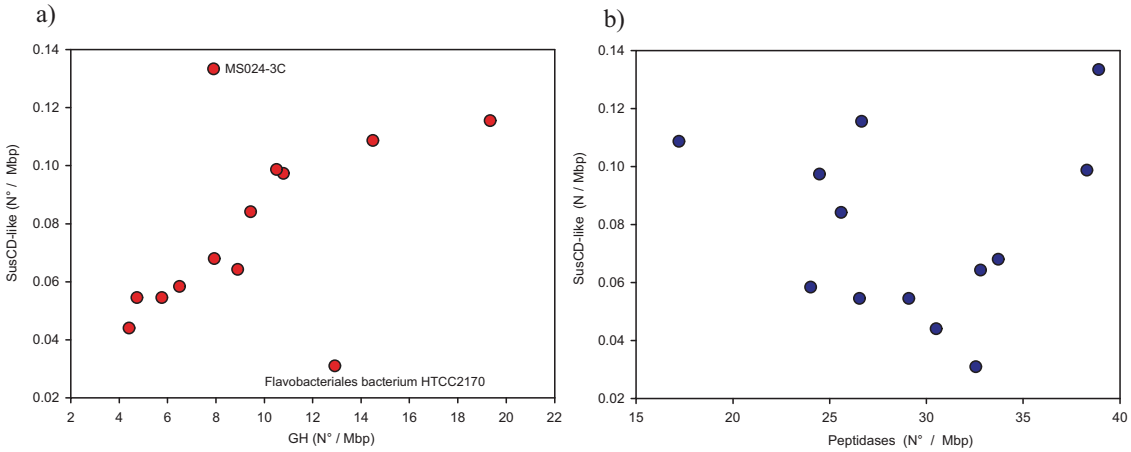
Supplementary Figure 1



Supplementary Figure 2



Suplementary Figure 3



Suplementary Figure 4

Supplementary Table 1: *Bacteroidetes* clade-specific oligonucleotide probes designed in this study. Comp: unlabeled competitor oligonucleotides. FA: formamide concentration (v/v) in the hybridization buffer. (*), (+): probes applied together

Probe	Target clade	Target fosmid	Sequence (5' --> 3')	FA (%)
F805E4-471	DE2 / VIS-805	S3-805-E4	GTAAGTAGGTTTCTTCCTGTAT	30
comp F805E4-471_1			GTAAGTAGGTTTCTTCCTGTAGAAAA	
comp F805E4-471_2			GTAAGTAGGTTTCTTCCTATAT	
F8737E-843	DE2 / VIS-873	S18-873-E7	CGCTTAACCACTCAATCAACCA	20
comp F8737E-843 _1			CGCTTAGCCACTCAACCAACCA	
NS9-664	NS9	S18-848-2B	ACATGACCTATTCCGCCAACTT	35
ANT-981 (*)	ANT	S3-816-G8	TTCCCATTTAAACCCTGGTA	20
comp ANT-981_1			TTCCCATTTAAACCCAGGTA	
comp ANT-981_2			TTCCCATTTAAACCCGGGTA	
ANT-72 (*)	ANT	S3-816-G8	TCAAACATCCGAAGACATTC	20
AGG58-742	AGG58 branch 3	S18-906-12E	CCCTTAGTGTCAATACTGAC	35
comp AGG58-742_1			GCCTCAGCGTCAATACTGAC	
SPA-730	Spingobacteria A	S3-858-892-M	ATATGCCTAGCTAGCTGC	35
SPC-814	Sphingobacteria C	S18-1049-858-M	AAGATCGAGTAAACATCGTT	n.d.
CYT-734 (+)	*Cytophagia*	S18-1001-9F	CAGTTTCTGCCTAGTAAG	25
CYT-1462 (+)	*Cytophagia*		CTAGCTCTACCCTAAACA	25

Supplementary Table 2: Pairwise nucleotide sequence comparisons between fosmids with a 16S rRNA gene 99-100% identical. * genomics fragments that were merged.

			Coverage	
Reference	**Querry**	**Identity (%)**	**Reference**	**Querry**
S3-979-C6*	S3-825-D3*	99.8	80.0	92.4
S3-858-E2*	S3-892-C7*	99.4	41.7	44.0
S3-858-E2	S3-891-H6	99.9	64.6	63.7
S3-893-A7	S3-891-H6	99.1	78.4	75.9
S3-891-H6	S3-892-C7	100	76.6	79.6
S3-858-E2	S3-893-A7	89.9	85.9	87.4
S3-893-A7	S3-892-C7	99.2	53.5	57.4
S18-1087-F8	S18-847-G1	98.6	52.7	72.9
S18-1087-F8*	S18-1013-C10*	99.0	47.9	56.6
S18-1013-C10	S18-847-G1	100	76.9	90.0
S18-1049-A9*	S18-858-A8*	100	88.9	89.4

Supplementary Table 3: Predicted glycoside hydrolases in *Bacteroidetes* metagenome in station (S) 3. Pfam models with E value < E -15. Localization predicted with PSORTb. Lipoproteins identified with LipoP. C: cytoplasm; CM: cytoplasmic membrane; P: periplasmic space; OM, outer membrane. SpI: Signal peptide I; SpII: Signal peptide II; Pos+2=: Amino acid in the +2 position. If Pos+2=D, the lipotrotein is anchored to the cytoplasmic membrane (Seydel et al 1999)

S	Clade	Locus	Description	Pfam model	Signal	Putative Localization
S3	*Polaribacter*	S3-860-D3-2	mannose-6-phosphate isomerase, class I	PMI_typeI	CYT	C
	Polaribacter	S3-860-D3-0	alpha-1,2-mannosidase	Glyco_hydro_92	SpI	P
	Polaribacter	S3-860-D3-9	alpha-1,2-mannosidase	Glyco_hydro_92	SpII (Pos+2=K)	OM
	Polaribacter	S3-860-D3-10	alpha-1,2-mannosidase	Glyco_hydro_92	SpII (Pos+2=S)	OM

Supplementary Table 4: Predicted sulfatases in *Bacteroidetes* metagenome in both stations (S). Pfam model with E value < E-15. Localization predicted with PSORTb. Lipoproteins identified with LipoP. Abbreviations as in Supplementary Table 3.

S	Clade	Locus	Pfam model	Signal	Localization
S3	*Polaribacter*	S3-860-D3-3	sulfatase	SpI	P
	Polaribacter	S3-860-D3-5	sulfatase	SpI	P
	Polaribacter	S3-860-D3-orf6-z	sulfatase	SpI	P
	Polaribacter	S3-860-D3-7	sulfatase	SpI	P
	Polaribacter	S3-860-D3-17	sulfatase	SpI	P
S18	DE2	S18-873-7E-41	sulfatase	SpI	P

Supplementary Table 5: Predicted peptidases in Bacteroidetes metagenome in both stations (S). Pfam model with E value < E-15. Localization predicted with PSORTb. Lipoproteins identified with LipoP. Abbreviations as in Supplementary Table 3. -- Unknown

S	Clade	Locus	Description	Pfam model	Signal	Putative Localization	Putative Function (+)
S3	VIS5	S3-861-A8-1	ATP-dependent protease La	Lon_C (S16)	SpI	P	Turnover of intracellular proteins
	DE2	S3-805-E4-13	Peptidase family M23	M23	CYT	C	Endopeptidase that lyse bacterial cell wall peptidoglycans.
	ANT	S3-816-G8-2	Peptidase family M23	M23	CYT	C	Endopeptidase that lyse bacterial cell wall peptidoglycans.
	Sphingo A	S3-858-892-12	Glycoprotease faimily M22	M22	CYT	C	Endopeptidase that cleaves only proteins that are O-sialoglycosylated
	Sphingo A	S3-858-892-8	Peptidase family M23	M23	CYT	C	Endopeptidase that lyse bacterial cell wall peptidoglycans.
	Polaribacter	S3-860-D3-4	Oligopeptidase	Abhydrolase_3// S9	SpI	P	Biologically active peptides // common to a number of hydrolytic enzymes
	VIS5	S3-861-A8-17	Peptidase family M23	M23	SpI	P	Endopeptidase that lyse bacterial cell wall peptidoglycans.
	Sphingo A	S3-858-892-11	Carboxy-terminal processing protease	S41 // PDZ*	SpI	P	Incorrectly synthesized proteins // PDZ domains are found in diverse signaling proteins.
S18	VIS4	S18-920-B9-13	Oligopeptidase	S9_N	--	--	Degradation of biologically active peptides
	AGG58	S18-870-C2-29	Protein with similarity to peptidase M16	M16 // insulinase*	--	--	M16 domain might be innactive
	VIS4	S18-920-B9-10	Transmembrane metalloprotease	M48	TMH	CM	Degradation of abnormal proteins
	Dokdonia	S18-841-H9-4	Acetylornithine deacetylase	M20	CYT	C	Exopeptidase, hydrolyse the late products of protein degradation so as to complete the conversion of proteins to free amino acids.
	DE2	S18-873-E7-35	Peptidase family M23	M23	CYT	C	Endopeptidase that lyse bacterial cell wall peptidoglycans.
	VIS4	S18-920-B9-12	Oligopeptidase	S9	CYT	C	Degradation of biologically active peptides
	NS9	S18-848-B2-26	Glutamyl aminopeptidase family M42	M42	CYT	C	Metalloaminopeptidase some of which also have acylaminoacylpeptidase activity
	NS4	S18-1082-D10-20	Glutamyl aminopeptidase family M42	M42	CYT	C	Metalloaminopeptidase some of which also have acylaminoacylpeptidase activity
	AGG58	S18-870-C2-11	Oligoendopeptidase family M3	M3	CYT	C	Intracellular degradation of oligopeptides
	Sphingo C	S18-858-1049-35	Aminopeptidase family M17	M17	CYT	C	Exopeptidase processing and regular turnover of intracellular proteins
	Cytophagia	S18-1001-F9-12	Glycoprotease faimily M22	M22	CYT	C	Endopeptidase that cleaves only proteins that are O-sialoglycosylated.
	Cytophagia	S18-1001-F9-11	Carboxy-terminal processing protease	S41 // PDZ*	CYT	C	Incorrectly synthesized proteins // PDZ domains are found in diverse signaling proteins.
	Leeuw.	S18-997-C3-24	Dipeptidyl peptidase	S9 // DPPIV_N	SpI	P	Degradation of biologically active peptides
	AGG58	S18-906-E12-21	Dipeptidylpeptidase	DPPIV_N // S9	SpI	P	Degradation of biologically active peptides
	VIS4	S18-920-B9-43	Carboxy-terminal processing protease	S41 // PDZ	SpII (Pos+2=S)	OM	Incorrectly synthesized proteins // PDZ domains are found in diverse signaling proteins.
	Leeuw.	S18-997-C3-22	Beta-lactamase	Beta-lactamase	SpI	P	Catalyses the opening and hydrolysis of the beta-lactam ring of beta-lactam antibiotics such as penicillins and cephalosporins
	AGG58	S18-906-E12-9	Lysyl-peptidase	--	SpI	P	Highly specific, hydrolyzes lysyl bonds including the Lys-Pro bond.

* E Value > E-15, + information from MEROPS database (http://merops.sanger.ac.uk/)

Supplementary Table 6: Predicted proteins involved in the recycling of peptidoglycan in *Bacteroidetes* metagenome in both stations (S). Pfam model with E value < E-15. Localization predicted with PSORTb. Lipoproteins identified with LipoP. Abbreviations as in Supplementary Table 3.

S	Clade	Locus	Gene name	Putative Function	Pfam model	Signal	Putative Localization
S3	VIS6	S3-979-C6-20	*anmK*	Anhydro-N-acetylmuramic acid kinase	UPF0075	CYT	C
	DE2	S3-805-E4-7		Mannosyl-glycoprotein endo-beta-N-acetylglucosamidase	Glucosaminidase	SpII (Pos+2=S)	OM
S18	Dok..	S18-841-9H-16	*murQ*	N-acetylmuramic acid 6-phosphate etherase	--	CYT	C
	VIS3	S18-812-11F-22	*murQ*	N-acetylmuramic acid 6-phosphate etherase	--	CYT	C
	VIS4	S18-920-9B-22	*anmK*	Anhydro-N-acetylmuramic acid kinase	UPF0075	CYT	C
	DE2	S18-873-E7-14	*anmK*	Anhydro-N-acetylmuramic acid kinase	UPF0075	CYT	C
	Dok.	S18-841-H9-41		Mannosyl-glycoprotein endo-beta-N-acetylglucosamidase	Glucosaminidase	SpII (Pos+2=G)	OM-CM *
	VIS4	S18-920-B9-38		Mannosyl-glycoprotein endo-beta-N-acetylglucosamidase	Glucosaminidase	SpII (Pos+2=G)	OM-CM *
	DE2	S18-873-E7-32		Mannosyl-glycoprotein endo-beta-N-acetylglucosamidase	Glucosaminidase	CYT	C

* according to Seydel et al. (1999) if G in +2 position, lipoproteins could be anchored to either OM or CM

Supplementary Table 7: Predicted proteins possibly involved in surface adhesion in *Bacteroidetes* metagenome in both stations (S). Pfam model with E value < E -15. Localization predicted with PSORTb. Lipoproteins identified with LipoP. Abbreviations as in Supplementary Table 3. – Unknown

S	Clade	Locus	Description	Pfam model	Signal	Putative Localization
S3	ANT	S3-816-G8-26	protein with similarity to glycosyl transferase family 2	Glycos-transf-2 *	TMH	CM
	Flavobacteria	S3-933-H4-19	Glycosyl transferases group 1	Glycos-transf-1	CYT	C
	Sphingo A	S3-892-858-5	Glycosyl transferases group 1	Glycos-transf-1	CYT	C
	Sphingo A	S3-892-858-1	Glycosyl transferases group 1	Glycos-transf-1	CYT	C
	Sphingo A	S3-892-858-0	Glycosyl transferases group 1	Glycos-transf-1	CYT	C
	Sphingo A	S3-892-858-2	UDP-N-acetylglucosamine 2-epimerase	Epimerase -2	CYT	C
S18	Sphingo C	S18-1049-858-26	transglycosylase/transpeptidase penicillin-binding protein	Transgly // Glycos-transf51	--	--
	Cytophagia	S18-1001-F9-35	polysaccharide biosynthesis protein	Polysacc-synt *	TMH	C
	AGG58	S18-906-E12-0	protein with similarity to glycosyl transferase family 1	--	CYT	C
	NS9	S18-1082-D10-3	Glycosyl transferases group 1	Glycos-transf-1 *	CYT	C
	AGG58	S18-906-E12-19	protein containing PKD domain	PKD	SPI	P

* E Value > E-15

Manuscript III

Multilevel diversity analysis of bacterial communities along the environmental gradient Río de la Plata–South Atlantic Ocean

Cecilia Alonso, Paola Gómez-Pereira, Alban Ramette, Leonardo Ortega, Bernhard M. Fuchs, and Rudolf Amann

Manuscript submitted to Aquatic Microbial Ecology

Multilevel diversity analysis of bacterial communities along the environmental gradient Río de la Plata-South Atlantic Ocean

Cecilia Alonso[1*], Paola Gómez-Pereira[2], Alban Ramette[2], Leonardo Ortega[3], Bernhard M. Fuchs[2] & Rudolf Amann[2]

[1]Instituto de Investigaciones Biológicas Clemente Estable, Av. Italia 3318, CP11600, Montevideo, Uruguay.

[2]Max Planck Institute for Marine Microbiology, Celsiusstr. 1 28359, Bremen, Germany.

[3]Dirección Nacional de Recursos Acuáticos, Constituyente 1497, CP 11200, Montevideo, Uruguay.

Intended as article in *Aquatic Microbial Ecology*

Running head: Bacterial diversity of *Río de la Plata*

* Corresponding author:
IIBCE, Avenida Italia 3318
CP. 11600. Montevideo,Uruguay.
Phone: ++598-2-487-1616 ext. 141
Fax: ++598 2 487-5548
Email: calonso@iibce.edu.uy

Abstract

The Río de la Plata is the world's largest estuary, draining the second largest basin of South America. In spite of its key significance at the local and global scale, this is the first attempt made so far to characterize the bacterial diversity of this system. We employed a suite of molecular methods to analyze the bacterial diversity at different phylogenetic depths, both in terms of richness and evenness, and we related these findings with several physicochemical parameters. A clear environmental gradient was observed with three distinctive zones (riverine, frontal, and marine). In turn, the bacterial communities appeared highly diverse and strongly structured by the environment. A comparably high number of different OTUs (325) were identified by Automated Ribosomal Intergenic Spacer Analysis (ARISA). The ARISA results suggested that the composition of bacterioplankton communities was strongly related to physicochemical water characteristics, and that the bacterioplankton had little overlap with the bacterial community of the sediments. Catalyzed Reporter Deposition-Fluorescence *in situ* hybridisation (CARD-FISH) indicated that *Betaproteobacteria* and *Actinobacteria* were more abundant in the riverine zone, whereas *Gammaproteobacteria* and *Planctomycetes* preferred the marine waters. The relative abundance of *Bacteroidetes* was peaking in the frontal zone, and *Alphaproteobacteria* were the numerically dominant phylum throughout the estuary. The analysis of about 300 almost full length sequences pointed to the existence of ubiquitous phylotypes, but also suggested habitat preferences among certain groups. This first look into the bacterial community of the Río de la Plata demonstrated marked diversity patterns, providing further insights into the ecology of globally distributed groups of aquatic bacteria.

Key words: Estuary, Bacteria, Río de la Plata, Diversity, FISH, ARISA, Cloning

Introduction

The Río de la Plata estuary, situated between Argentina and Uruguay, covers an area of approximately 38.800 km^2 and drains the second largest basin of the South American continent. This system entails a highly productive area, sustaining valuable fisheries of Uruguay, Argentina, and international fleets (Martinez & Retta 2001, Bisbal 1995, Acha et al. 2008). The Río de la Plata also constitutes the main point source of freshwater in the South Atlantic (Boltovskoy et al., 1999). This area is particularly relevant for the global carbon budget, acting as a CO_2 sink spot, especially during spring and summer (Feely et al. 2001, Bianchi et al. 2005).

In spite of its key significance, there is still very limited understanding of basic ecological aspects of the planktonic communities of the Río de la Plata (Calliari et al. 2009). In particular, no study has been performed so far on the microbial assemblages at the transition Río de la Plata-Atlantic Ocean. Due to the existence of strong environmental gradients, yet lacking physical barriers, these transitional systems are ideal for studies on the dynamics of the bacterial community composition (BCC).

A number of authors have analyzed the changes in BCC in transitional aquatic environments employing molecular techniques. In their pioneer work (Crump et al. 1999) used 16S rRNA gene clone libraries to characterize microbial communities from freshwater and marine portions of the Columbia River estuary. Since then,

49 several researchers have employed fingerprinting methods, predominantly
50 Denaturing Gradient Gel Electrophoresis (DGGE), to obtain patterns of BCC in
51 different estuarine systems e.g. (Murray et al. 1996, Hollibaugh et al. 2000,
52 Troussellier et al. 2002, Crump et al. 2004). More recently, Hewson & Fuhrman
53 (2004) used automated ribosomal intergenic spacer analysis (ARISA) to study the
54 diversity of the microbial communities at Moreton Bay, Australia. Fluorescence in
55 situ hybridisation (FISH) has been also used to determine the abundance of major
56 bacterial groups in a number of estuaries (Bouvier & del Giorgio 2002, Cottrell &
57 Kirchman 2003, Zhang et al. 2006).
58
59 Although these studies have provided clear insights into the microbial ecology of
60 transitional environments, there is still an obvious need of data on subtropical
61 systems; so far only two of such systems have been targeted (Hewson & Fuhrman
62 2004, Zhang et al. 2006). In addition, there are very few examples of studies in
63 aquatic microbial ecology combining the three methodological approaches
64 mentioned above e.g. (Anton et al. 2000, Watanabe et al. 2000, Simek et al. 2001,
65 Alonso-Sáez et al. 2007). This strategy is desirable: there is still need to address
66 variations in BCC at different phylogenetic levels, and to link this information with
67 appropriate environmental parameters, in order to gain deeper understanding on
68 the ecology of aquatic bacteria.
69
70 Although a few ecophysiological aspects appear to be well established (e.g.
71 negative correlation of natural abundances of *Betaproteobacteria* and salinity)
72 (Bouvier & del Giorgio 2002), there is still considerable lack of information about

the factors controlling the distribution of bacterial populations. Furthermore, the major phylogenetic groups are harbouring members with quite different ecological strategies (Alonso & Pernthaler 2006, Alonso-Saez & Gasol 2007), particularly in habitats characterized by sharp environmental transitions (Alonso et al. 2009).

The aim of this study was to employ a suite of molecular methods to analyze the bacterial diversity along a transect in a subtropical estuary, and to relate diversity patterns with environmentally relevant data.

Experimental procedures

Study area. The Río de la Plata is a large-scale estuary characterized by a salt-wedge regime, low tidal amplitude (<1 m), a broad and permanent connection to the sea, and high susceptibility to atmospheric forcing due to its large extension and shallow water depth (Acha et al. 2008, and references therein). The position of the turbidity maximum is highly variable according to tide, river discharge and wind (Framiñan & Brown 1996). The resulting outflow is on average 22,000 m^3s^{-1} (Framiñan & Brown 1996) showing seasonal variability with maxima in March-June and September-October, and minimum in December-March (Guerrero et al. 1997, Nagy et al. 2002).

Sampling. The cruise ARTEMISA1 was carried out during late austral summer (9 March) of 2008 onboard R.V. Aldebaran (DINARA). A transect with 9 sampling stations (S3 to S11) was followed along the salinity gradient occurring in the outer region of the Río de la Plata; the distance between stations was 10 nautical miles

98 (Fig.1). The bottom depth covered a range from a minimum of 5 m to a maximum
99 of 30 m.
100
101 At each station, a conductivity, temperature, and depth (CTD) cast (SBE-19)
102 profiled the water column from the surface to the bottom. According to the CTD
103 data, six stations were chosen for sampling bacterial communities. Surface water
104 samples were taken using a bucket; sub-surface water samples were taken with
105 Niskin bottles. All sampling devices were acid-washed and seawater-rinsed prior to
106 use. A Smith-McIntyre bottom grab sampler was used for sampling sediment.
107
108 **Physico-chemical characterization.** Water samples were filtered through glass
109 fiber filters (Whatman GF/F, 47 mm) to assess total chlorophyll a and suspended
110 particulate matter. Filters were stored dried and frozen for subsequent analysis in
111 laboratory. Chlorophyll was extracted with 90% acetone and analyzed with a
112 Shimadzu UV–2101 PC, UV-VIS Scanning spectrophotometer. Chlorophyll a
113 concentration was calculated according to Jeffrey & Humphrey (1975) with
114 correction for phaeopigments (Lorenzen 1967). Suspended particulate matter was
115 estimated by the gravimetric method of the particulate material (Strickland &
116 Parsons 1972).
117
118 Water samples for dissolved nutrient analysis were filtered through a Whatman
119 GF/C glass fibre filter. Inorganic nutrient concentrations measured at each station
120 include total nitrogen, total phosphorus (Valderrama 1981), and dissolved fractions
121 of nitrite (Strickland & Parsons 1972), nitrate (Eaton et al. 2005), ammonium

(Koroleff 1976), and soluble reactive phosphorus (Murphy & Riley 1962). Nutrient concentrations were determined colorimetrically employing a Shimadzu UV–2101 PC, UV-VIS Scanning spectrophotometer.

Sample collection for the determination of the *in situ* abundance of different bacterial populations. Water samples were fixed with freshly prepared buffered paraformaldehyde solution (PFA) at a final concentration of 1% for 24 h. Portions of 10 ml were then filtered through polycarbonate filters (type GTTP, pore size, 0.2 µm, diameter, 47 mm, Millipore, Eschborn, Germany). The filters were rinsed twice with sterile PBS and stored at -20°C until further analysis.

DNA collection and extraction. Water sample portions of ten litres were pre-filtered through 50 and 10 µm and then filtered onto cellulose acetate filters (142 mm diameter, 0.2 µm pore size; Sartorius, Goettingen, Germany). Filters were stored at −20°C until further processing. For sediment samples, the first centimetre layer was collected using a sterile Petri dish, and the DNA extraction was performed from a 10 g sub-sample. DNA was extracted following the protocol by (Zhou et al. 1996).

Genetic fingerprinting of the bacterial communities. The standard ARISA protocol by Ramette (2009) was applied. For the amplification of the intergenic spacer the forward universal primer ITSF and the eubacterial ITSReub reverse primer were employed (Cardinale et al. 2004). The reverse primer was labeled with the phosphor-ramidite dye HEX. PCR conditions, purification and sample

146 preparation prior to submitting to capillary electrophoresis were done as described
147 by Ramette (2009). The electropherograms obtained from fragment analysis were
148 examined with GeneMapper software v 3.7 (Applied Biosystems) employing the
149 criteria described by Ramette (2009) The output tables from GeneMapper were
150 analyzed using custom R binning scripts. Triplicates samples were analyzed for
151 replicated fragments. Further analysis was based on the fragments that appeared
152 in at least two of the replicates generating an average profile for each sample.
153
154 Several similarity indexes among samples were calculated based on incidence
155 data of the ARISA fragments using the software EstimateS (Colwell 2005)
156 (http://viceroy.eeb.uconn.edu/EstimateS). Cluster analysis was performed with the
157 Relative Fluorescence Intensity (RFI) data. Similarity matrix was calculated with
158 the Bray-Curtis coefficient and the group average was used as the linkage
159 algorithm. Significance of the grouping in the cluster analysis was tested using
160 similarity profile permutation test (SIMPROF) at 99% significance level. SIMPROF
161 tests for statistically significant evidence of genuine clusters in an *a priori*
162 unstructured set of samples (Clarke et al. 2008). Analysis was performed with the
163 software Primer v6 (Version 6.1.5, Primer-E Ltd, Plymouth, UK) (Clarke & Gorley
164 2006).
165
166 **Construction of clone libraries.** Almost complete bacterial 16S rRNA genes were
167 amplified via Polymerase Chain Reaction (PCR) using the primers 27f-YM (Frank
168 et al. 2008) and GM4R (Muyzer et al. 1993). PCR products were purified using the
169 QIAquick PCR purification kit (Qiagen, Hilden, Germany) and cloned using the

pGEM®-T easy system (Invitrogen, Groningen, Netherlands). Competent *E. coli* cells Top10 (house made) were transformed and plated on Luria Bertani (LB) agar plates containing 50 µg ml^{-1} of ampicillin. Single colonies were transferred into microtiter plates (MTP) containing ampicillin-amended LB medium (50 µg ml-1) and cultivated overnight at 37°C. Glycerol (50% final concentration) was added to each well, and the MTPs were stored at −20°C until further processing.

Sequencing. Plasmids were isolated from clones with Montage plasmid MiniPrep96 kit (Millipore, Eschborn, Germany), and 50-100 ng of template was used per sequencing reaction. Sequencing reactions were performed using the ABI BigDye® chemistry and an ABI 3100 genetic analyser (Applied Biosystems, Foster City, U.S.A.) according to the manufacturer's instructions. M13F and M13R primers were used to obtain 298 16S rRNA gene sequences. Partial sequences were assembled and manually corrected using the software Sequencher (Gene Codes, Michigan, U.S.A.). Chimeric sequences were identified using the Mallard and Pintail programs (Ashelford et al. 2006) and 5 sequences were removed from further analysis. The 16S rRNA gene sequences produced during this study were deposited in GenBank under the accession numbers GU230187 to GU230468.

Diversity analysis. All almost complete sequences were subsequently analyzed using the software DOTUR (Schloss & Handelsman 2005) which allows to group sequences according to a series of defined levels of similarity. The grouping criteria were set to compare sequence stretches of at least 800 bp including the highly variable regions V3 to V7 (Neefs et al. 1990). Data are presented for

194 similarity levels of 97% and 99%, as the standard values to approximate bacterial
195 species (Stackebrandt & Goebel 1994), and to set a conservative threshold for
196 unique sequence types.
197
198 **Population analysis by CARD-FISH.** The percentages of different microbial taxa
199 in the water samples were determined by FISH with horseradish peroxidase-
200 labeled oligonucleotide probes and catalyzed reporter deposition (CARD-FISH)
201 (Pernthaler et al. 2002). The following probes were used to characterize the
202 microbial community: EUB338 I-III (most Bacteria) (Daims et al. 1999), ALF968
203 (most *Alphaproteobacteria*) (Neef 1997), BET42a (most *Betaproteobacteria*) (Manz
204 et al. 1992), GAM42a (most *Gammaproteobacteria*) (Manz et al. 1992), CF319a
205 (many groups of *Bacteroidetes*) (Manz et al. 1996), HGC69a (most *Actinobacteria*)
206 (Roller et al. 1994), PLA46 (Neef et al. 1998), SAR11-441 (members of the
207 Alphaproteobacterial SAR11 clade) (Morris et al. 2002), ROS537 (members of the
208 *Roseobacter-Sulfitobacter-Silicibacter* clade) (Eilers et al. 2001), NOR5-730
209 (members of the gammaproteobacterial NOR5/OM60 clade) (Eilers et al. 2001),
210 SAR86-1245 (members of the gammaproteobacterial SAR86 clade) (Eilers et al.
211 2000), POL740 (members of the genus *Polaribacter* of *Bacteroidetes*) (Malmstrom
212 et al. 2007), CF6-1267 (members of DE cluster 2 of the *Flavobacteria-*
213 *Sphingobacteria* group of the phylum *Bacteroidetes)* (Kirchman et al. 2003) and
214 VIS1-575 (uncultured marine VIS1 clade of the *Flavobacteria-Sphingobacteria*
215 group of the phylum *Bacteroidetes)* (Gomez-Pereira et al. 2009).
216

All probes were purchased from Biomers.net (Ulm, Germany). Signal amplification was performed with FITC labelled tyramides (Molecular Probes, Eugene, Oreg.). CARD-FISH preparations were counterstained with 4',6-diamidino-2-phenylindole (DAPI) at a final concentration of 1 µg ml^{-1}. DAPI- and CARD-FISH-stained cells were counted manually, achieving a minimum of 1000 cells per filter.

Statistical analyses. ARISA profiles and CARD-FISH abundance datasets were Hellinger transformed prior to applying linear multivariate methods (Legendre & Gallagher 2001, Ramette 2007). In order to relate variation in microbial data to variation in spatial and environmental parameters, redundancy analyses (RDA) were performed on quantitative variables that were standardized to unit variance and zero mean. Spatial coordinates (latitude and longitude) were converted to Cartesian spatial coordinates (x, y) using the R package gmt. A forward selection procedure was performed to retain only the environmental variables that significantly explained variation in bacterial data (Legendre & Legendre 1998). Following this procedure, the selected terms were then analyzed in concert with the other contextual parameters using the R package vegan (http://cc.oulu.fi/~jarioksa/softhelp/vegan.html). Graphical display of the RDA models was done using CANOCO (version 4.5. Microcomputer Power, Ithaca, NY).

Results

The vertical distribution of temperature (Figure 2A) showed a seaward decrease (i.e. from S3 to S11). The highest temperatures, concomitant with a low thermal stratification, were detected from S3 to S7 (24.5 to 22.5ºC, respectively). In the

following stations S8 to S11 lower temperatures and higher stratification were found (22ºC at surface and 17ºC in the bottom of S11). The vertical distribution of salinity (Figure 2B) presented a steep seaward gradient. The lower salinity values (3 PSU) were observed at the riverine station S3. The frontal zone involved stations S4, S5 and S6, where the steepest salinity gradient was observed, as surface salinity increased sharply from station S3 to S6 from 3 PSU to 29 PSU. From station S7 on, a practically homogenous marine water column was found.

The surface concentration of suspended particular matter showed a clear decline towards the Atlantic Ocean, with S3 exhibiting the highest concentration (5.4 mg l^{-1}). Organic matter, chlorophyll a, and phaeopigments increased sharply from S3 to a maximum at S5, and then decreased with the oceanic influence (Fig. 2C). Chlorophyll a concentration ranged from 0.03 µg l^{-1} at S9 to 6.4 µg l^{-1} at S5 (Fig. 2C).

In general, higher concentrations of most nutrients were found in the riverine part (TN, TP, SRP, and nitrate). Nutrient levels decreased towards the marine stations e.g. TN decreased from 290 µg l^{-1} to 130 µg l^{-1}, and TP from 94 µg l^{-1} to 38 µg l^{-1} (Fig 2D). Nevertheless, a local peak of nutrients was found at S5 at a depth of 8 m, where the concentrations of TN and TP were the highest found along the transect, 293 µg l^{-1} and 151 µg l^{-1}, respectively. Ammonium was undetectable along the whole transect with the exception of the surface water at S9, where ammonium concentration was 9 µg/l (Fig. 2D).

265 **Bacterial diversity patterns along the environmental gradient.**

266 325 different OTUs were identified by ARISA analysis (suppl. Table 1). The highest
267 richness was found in the riverine surface water (S3). In contrast, the lowest
268 richness occurred at the sediment sample of the same station (suppl. Table 2). In
269 the water column, on an average base, the highest richness was found in the
270 riverine station (108 OTUs), followed by the marine stations (93 OTUs), and finally
271 by the central station of the frontal zone station (81 OTUs) (suppl. Table 2).

272

273 Roughly half of the OTUs were shared between two habitats (suppl. Table 1),
274 whereas 65 OTUs were common to all of them (suppl. Table 1). Several OTUs
275 were distinct to habitats: the number of unique OTUs was 19, 39 and 36, for
276 riverine, frontal zone, and marine stations, respectively (suppl. Table 1). The
277 dominant OTU in the frontal zone – a restriction fragment of a length of
278 approximately 300 nucleotides - was also present in the marine zone (suppl. Table
279 1). In the marine and riverine samples, the dominant restriction fragments had
280 amplicon lengths of 334 bp and 218 bp, respectively, and none of them was habitat
281 exclusive (suppl. Table 1).

282

283 Although each sample displayed a unique community fingerprint, a series of
284 diversity patterns was revealed, both in the longitudinal and in the depth
285 dimensions.

286

287 The sediment samples were more similar to each other than to the water samples
288 of the same station (Fig. 3, suppl. Table 2). The sediment communities appeared

289 relatively similar to the water column assemblage from the riverine station S3 (Fig.
290 3). Although water samples from the surface and 4 m depth of station S3 clustered
291 together, they exhibited lower similarity indexes than comparable samples taken at
292 the marine stations (suppl. Table 2).
293
294 All water samples from fully marine stations (S9-S11) clustered closely together,
295 and depth was not a decisive factor for the clustering (Fig. 3). For example, the
296 water sample from S9 taken at 15 m depth had a similarity Bray-Curtis index of
297 0.74 with the surface water from S11 (suppl. Table 2). The marine stations shared
298 more OTUs between them, at any depth, than with any other sample (suppl. Table
299 2).
300
301 On the contrary, in the frontal zone the samples clustered according to depth rather
302 station, i.e. water samples from S5 at 4 m and 8 m depth were more closely related
303 with the marine samples than with the surface water at the same station (Fig. 3).
304 Interestingly, the surface sample of the centre of the frontal zone was clearly
305 distinct from all other samples (Fig. 3).
306
307 **Bacterial community composition as revealed by FISH and clone libraries.**
308 Average cell detection with the EUB338 I-III probe was 77.8% (SD = 8.7%; n = 11)
309 of all DAPI stained objects. The sum of counts with the probes targeting main
310 bacterial groups (*Alphaproteobacteria, Betaproteobacteria*, *Gammaproteobacteria,*
311 *Bacteroidetes, Actinobacteria,* and *Planctomycetes*) in average accounted for
312 virtually all bacterial cells (86-112% of cells hybrizing to EUB338 I-III).

313 *Alphaproteobacteria* dominated the community in all water samples, representing
314 between 36-57% of all bacterial cells (Fig. 4). Accordingly, most of the sequences
315 retrieved from the water column libraries belonged to members of this group
316 (Fig.5). Virtually all *Alphaproteobacteria* cells in the Río de la Plata estuary could
317 be assigned to either the SAR11 or the *Roseobacter* clades by probes SAR11-441
318 and ROS537, respectively (Fig.4). The many SAR11-affiliated sequences retrieved
319 in the clone libraries were diverse (suppl. Table 3). Most of them were targeted by
320 the SAR11-441 probe. All of *Roseobacter*-related sequences recovered in the
321 three surface water libraries were targeted by probe ROS537. Beside these two
322 main alphaproteobacterial clades, few sequences related to the order *Rickettsiales*
323 were also retrieved from the water column and the sediment.
324
325 Identical 16S rRNA gene sequences were found in all three surface water libraries
326 for SAR11 (GU230202-GU230233 and GU230257-GU230260) (suppl. Table 3).
327 Other sequence types were present in the riverine and frontal zone (GU230246-
328 GU230251 and GU230195-GU230199) (suppl. Table 3). A cluster of SAR11-
329 related sequences was exclusively recovered from the marine library (GU230234-
330 GU230245) (suppl. Table 3).
331
332 *Bacteroidetes* was the second most abundant group. Cells hybridizing with probe
333 CF319a constituted on average 23% of all bacteria (Fig. 4). Members of this group
334 were particularly abundant at station S5, in the very centre of the frontal zone (Fig.
335 2). In this station, *Bacteroidetes* accounted for about one third of all *Bacteria*. Their
336 abundance decreased towards the marine stations and the habitats with low

337 salinity (Fig. 4). Nevertheless, in the clone libraries more *Bacteroidetes* sequences
338 were retrieved in the marine station than in the frontal zone station (Fig. 5). The
339 great majority of sequences retrieved from all stations were affiliated to the class
340 *Flavobacteria* (suppl. Table 3). The specific probes POL740, CF6-1267 and VIS1-
341 575, targeting different groups of *Flavobacteria* accounted on average for one third
342 of the cells targeted by CF319a (Fig.4). Among them, cells targeted by POL740
343 were the most abundant ones, particularly in the riverine and frontal zones, and its
344 abundance decreased with depth (Fig. 4). Abundance of the flavobacterial clade
345 targeted by VIS1-575 was higher in the frontal and marine zones, and for a given
346 station increased with depth (Fig. 4). CF6-1267 targeted cells were the least
347 abundant of the *Bacteroidetes* targeted groups and reached their maximum in the
348 frontal zone (Fig. 4). All the sequences retrieved in this phylum were characteristic
349 of the different samples; none of them was shared between libraries.
350
351 *Gammaproteobacteria* were generally the third group in abundance. On average,
352 they were 19% of all bacterial cells, with the exception of S9 where, depending on
353 the depth, they constituted 24-47% of all cells (Fig.4). Gammaproteobacterial
354 sequences account for about one third of all sequences obtained. Most of the
355 sequences obtained from the water column were affiliated to *Oceanospirillales* and
356 *Alteromonadales* (suppl. Table 3). Several sequences affiliated to *Oceanospirillales*
357 were targeted by the probe SAR86-1245. Hybridizations with this probe showed
358 that members of SAR86 clade constituted in average 6% of *Gammaproteobacteria*
359 (Fig. 4). Other ubiquitous members of this phylum were members of the
360 OM60/NOR5 clade as detected by probe NOR5-730. They represented on average

361 26% of *Gammaproteobacteria*, being particularly abundant in S5 at 4 meters depth,
362 and in the surface of S6 (Fig. 4). Sequences targeted by the NOR5-730 and by the
363 SAR86-1245 probes were retrieved from all libraries.
364
365 Some sequences affiliated to *Gammaproteobacteria* were shared by all three
366 surface water libraries, e.g GU230330-GU230335, which were related to the
367 OM60/NOR5 clade, and GU230311-GU230318 which were related to the SAR86
368 clade (suppl. Table 3). However, there were also diverse gammaproteobacterial
369 sequences related to OM60/NOR5, Oceanospirillales, *Alteromonadales*, and the
370 genera *Endogubula*, *Acinetobacter* and *Psychrobacter* which were found only at
371 the marine station S9 (suppl. Table 3).
372
373 *Betaproteobacteria* were minor components of the community (<5% of all bacteria),
374 except for the surface water samples at the stations with lower salinity where they
375 accounted for 16% and 14% at S3 and S4, respectively (Fig. 4). They were
376 represented by one sequence affiliated with the genera *Methylophilus* in the clone
377 library from the frontal zone (suppl. Table 3).
378
379 *Actinobacteria* were much more abundant at low salinity sites (up to >20% of all
380 bacteria in the surface water of station S3) and showed a very sharp decrease
381 towards marine stations where they average around 1% of all bacteria (Fig. 4).
382 Only few actinobacterial sequences were retrieved from the water samples of the
383 riverine and frontal zone stations, and from the marine sediment samples (Fig. 5).
384 Most of these sequences affiliated with uncultured *Acidimicrobiaceae* and

385 *Microbacteriaceae* previously found in estuarine and marine habitats (suppl. Table
386 3).
387
388 *Planctomycetes* were more abundant (up to 6%) at the marine stations (Fig. 4). In
389 the clone libraries, they were only retrieved from the marine sediment sample (Fig.
390 5). The sequences retrieved were closely related to sequences from uncultured
391 *Planctomycetes* obtained from marine sediments or soil (suppl. Table 3).
392
393 A few sequences affiliated to *Verrucomicrobia* were recovered from all clone
394 libraries. The sequences from riverine and frontal zone water column affiliated to
395 sequences known from *Epixenosomes*, ectosymbiotic *Verrucomicrobia*, whereas
396 the sequences from the marine water and sediment samples affiliated to
397 sequences known from the genus *Prosthecobacter.* Their closest neighbours have
398 been previously isolated from marine environments (suppl. Table 3).
399
400 Finally, a few sequences affiliated with the cyanobacterial genera *Synechococcus*
401 and *Calothrix* were retrieved from the water column libraries at the riverine and
402 marine sites (suppl. Table 3).
403
404 In marked contrast with the marine water samples, the marine sediment was
405 characterized by the virtual absence of *Alphaproteobacteria* and the dominant
406 retrieval of gammaproteobacterial and deltaproteobacterial sequences (Fig.5).
407 Moreover, not a single sequence was identical (even at 97% similarity level)
408 among the libraries from water column and sediment at the marine station (suppl.

409 Table 3).

410

411 Gammaproteobacterial sequences found in the sediment were very diverse and
412 always closely related to sequences retrieved from sediment samples from a
413 variety of systems (e.g. Wadden Sea, Yellow Sea, mid-atlantic ridge) (suppl. Table
414 3). Deltaproteobacterial sequences were affiliated to *Desulfobulbaceae* and
415 *Desulfobacteraceae,* and to *Haliangiaceae*. They were related to uncultured clades
416 previously found in aquatic sediments (suppl. Table 3). Sequences affiliated to
417 *Acidobacteria* were only recovered from the marine sediment sample (Fig. 5). All of
418 them were closely related to sequences from uncultured organisms recovered from
419 marine sediments (suppl. Table 3).

420

421 **Discussion**

422 Our methodological approach allowed for the recovery of complementary
423 information on the bacterial diversity in the gradient Río de la Plata-South Atlantic
424 Ocean. The ARISA analysis yielded an overview of the bacterial diversity at the
425 different sites, guiding the construction of clone libraries. These clone libraries,
426 together with the FISH evaluation, gave hints on the identity and abundance of the
427 individual groups responsible for the global pattern of bacterial diversity in this
428 environment. Altogether, the three techniques provided a multilevel view on
429 bacterial biodiversity ranging from phyla (FISH) to species (ARISA, sequencing).

430

431 **Physico-chemical parameters structure the bacterial communities**

432 The bacterial community appeared strongly structured by the habitat

433 characteristics on all levels of diversity targeted in this work. More specifically, the
434 statistical analysis revealed that the bacterial community composition in the water
435 column was strongly related to the environmental characteristics (Table 1). The
436 geographical factors appeared significant only for ARISA results, indicating that the
437 influence of geography over diversity could be evidenced only at the lower
438 phylogenetic levels (species/strains) (Table 1). Among the environmental
439 parameters, salinity and temperature were the main components explaining the
440 general observed patterns in the bacterial diversity for both ARISA and CARD-
441 FISH results.
442
443 **Insights into the ecology of particular bacterial groups**
444 *Alphaproteobacteria* dominated the bacterial community throughout the Río de la
445 Plata estuary (Fig.4). This finding is different from previous reports on comparable
446 estuarine systems where members of this group were dominant only beyond
447 higher salinity values (≥16 PSU) (Cottrell & Kirchman 2003, Zhang et al. 2006).
448 This is surprising at first glance only. We speculate that due to the massive marine
449 intrusion at the time of sampling (Fig. 2) open-ocean *Alphaproteobacteria*, most
450 notably members of the clade SAR11, were massively transported to the upper
451 part of the estuary.
452
453 Members of the SAR11 clade were abundant along the gradient with a tendency to
454 decline towards the riverine station (Fig. 4). SAR11 has been shown to thrive in
455 coastal and open ocean waters, where it frequently contributes to a high extent to
456 the total bacterial abundance (Morris et al. 2002, Alonso & Pernthaler 2006, Vila-

457 Costa et al. 2007)). In particular, SAR11 was found to constitute the large majority
458 of *Alphaproteobacteria* cells in a comparable system, the Delaware estuary
459 (Kirchman et al. 2005).
460
461 The high diversity found among sequences from SAR11 in our study, points to the
462 existence of ecotypes with different habitat preferences. Moreover, the abundance
463 of SAR11 cells was not related to salinity, but rather with the longitudinal axis of
464 sampling, providing further indication of ecotype differentiation along the gradient
465 (suppl. Fig 1). Habitat preferences in this cluster have been previously observed in
466 relation to water column depth (Garcia-Martinez & Rodriguez-Valera 2000, Carlson
467 et al. 2008).
468
469 Members of the *Roseobacter* clade exhibited a marked variation in abundance,
470 which their highest numbers at the surface of the frontal zone station and their
471 lowest abundance at the surface of the riverine station (Fig. 4). Members of this
472 clade have been frequently found in marine coastal and open waters (González &
473 Moran 1997, Alonso & Pernthaler 2006), and have been isolated from marine or
474 high salt environments (Moran et al. 2003, Pinhassi & Berman 2003). A decline in
475 *Roseobacter* abundance in direct proportion to freshwater influence has been
476 previously reported (González & Moran 1997)). However, in our study *Roseobacter*
477 did not appear strongly co-varying with salinity, indicating that other environmental
478 factors may be relevant (e.g. depth). Their highest abundance was found in S5
479 coincident with the peaks in chlorophyll *a* and nutrient concentrations. Similar
480 results have been reported based on the seasonal abundance of this group in the

Mediterranean Sea (Alonso-Sáez et al. 2007). Furthermore, a number of studies have found a correlation between *Roseobacter* abundance with either chlorophyll a (González et al. 2000), or nutrients concentration (Brown et al. 2005).

Interestingly, *Roseobacter* and SAR11 appeared opposite to each other in the RDA plot (suppl. Fig 1) indicating that these two alphaproteobacterial clades may be controlled by different factors. A previous study showed that both groups displayed an opposite pattern of seasonal abundance in the Mediterranean Sea (Alonso-Sáez et al. 2007)

Are *Betaproteobacteria* and *Actinobacteria* controlled by the same factors?

The abundances of *Betaproteobacteria* and *Actinobacteria* followed a similar pattern along the environmental gradient, indicating that these two groups are dependent on similar environmental factors. Their higher abundances were found in the samples with the lower salinity and higher nutrients values (surface waters from S3, S4 and S5; Figs 2 and 4). Moreover, clearly both groups were negatively co-varying with salinity (suppl. Fig 1).

Betaproteobacteria and *Actinobacteria* have been shown to display their highest abundances -often being the dominant members- in a variety of freshwater environments (Glöckner et al. 1999, Burkert et al. 2003, Hahn et al. 2005), Warnecke et al. 2005, Alonso et al. 2009). High *in situ* abundances of one or both groups have been also found in the freshwater parts of transitional environments, like estuaries (Bouvier & del Giorgio 2002, Kirchman et al. 2005, Zhang et al.

505 2006), or coastal lagoons (Piccini et al. 2006).

506

507 Nevertheless, it has been shown that in the Delaware estuary, the abundance
508 patterns of both groups were not solely explained by salinity (Kirchman et al.
509 2005). Glöckner et al (2000) hypothetized that niche differentiation between
510 *Betaproteobacteria* and *Actinobacteria* in aquatic habitats could be based in a
511 differential use of DOM components (Glöckner et al. 2000). In this line of argument,
512 (Bouvier & del Giorgio 2002) concluded that a highly significant relationship
513 between the abundance of *Betaproteobacteria* and dissolved organic carbon
514 (DOC) in two temperate rivers was not strictly related to the DOC concentration but
515 rather to compositional changes in the DOC pool, which are commonly reported in
516 estuaries (Bouvier & del Giorgio 2002). In the present study, an attempt of
517 characterizing the DOC pool was made via spectral analysis of the water samples
518 (data not shown), however, the statistical analysis indicated that salinity was the
519 main factor structuring community.

520

521 ***Gammaproteobacteria* peak at the marine station 9**

522 The abundance of *Gammaproteobacteria* appeared negatively correlated with
523 temperature (suppl. Fig 1). It did not follow a tendency along the environmental
524 gradient, but exhibited a pronounced peak at S9 (Fig. 4). Previous studies also
525 reported local peaks in gammaproteobacterial abundance in estuarine
526 environments (Bouvier & del Giorgio 2002, Cottrell & Kirchman 2003, Zhang et al.
527 2006). Such high abundances of *Gammaproteobacteria*, as found in our marine
528 station S9, have been previously reported for marine environments like the

529 Mediterranean Sea (Alonso-Sáez et al. 2007), and the North Atlantic ocean
530 (Schattenhofer et al. 2009). Bouvier and del Giorgio (2002) associated local spikes
531 of gammaproteobacterial abundances at the Choptank estuary to point sources of
532 nutrients probably derived from wastewater treatment facilities (Bouvier & del
533 Giorgio 2002). In our study, this peak in gammaproteobacterial abundance was
534 coincident with a pronounced peak in ammonium concentration (Fig. 2).
535 Interestingly, an unusual bloom of a *Gammaproteobacteria* (*Stenotrophomonas*
536 *maltophilia)* was reported associated to a peak in ammonium concentration in a
537 nearby coastal lagoon (Laguna de Rocha) (Piccini et al. 2006). Peaks in
538 ammonium concentration have been associated to high hydrolytic activities, and
539 *Gammaproteobacteria* are potentially one of the main groups of hydrolytic
540 microbes in marine environments (Woebken et al. 2007).
541
542 The analysis of the sequences retrieved from the clone library at S9 did not give
543 strong hints about the identity of the main contributors to the disproportionate
544 gammaproteobacterial abundance. Most of the gammaproteobacterial sequences
545 retrieved in all surface water libraries, including S9, affiliated to the SAR86 cluster
546 within *Oceanospirillales*, and to the NOR5/OM60 cluster within the
547 *Alteromonadales* (Fig. 5). These groups represented on average 31% of the
548 *Gammaproteobateria* in the environment, and were not particularly abundant at S9
549 (Fig. 4). Furthermore, hybridizations with available probes targeting other groups of
550 marine *Gammaproteobacteria* -some of them previously reported to bloom in
551 marine or coastal lagoon samples- (*Acinetobacter*, *Alteromonas, Pseudo-*

552 *alteromonas, Oceanospirillum, Stenotrophomonas maltophilia* and *Vibrio*), did not
553 yield any positive result (data not shown).
554
555 ***Bacteroidetes* and the frontal zone**
556 Members of *Bacteroidetes* sharply increased in their abundance between the
557 riverine surface water and the frontal zone, reaching their highest numbers at the
558 S5, located at the very centre of the frontal zone (Fig. 4). Particularly high
559 abundances of *Bacteroidetes* at the turbidity maximum of estuarine systems have
560 been previously reported (Bouvier & del Giorgio 2002). The frontal zone was
561 characterized by peaks in organic matter, chlorophyll a, and general availability of
562 nutrients, indicating that members of this group would be particularly favoured by
563 intermediate salinity and high productivity, characteristics often found in coastal
564 environments.
565
566 In this system, most of the sequences retrieved belonged to members of the class
567 *Flavobacteria* (suppl. Table 3), as in most estuarine and marine systems (Alonso et
568 al. 2007). Increased numbers and sequence retrieval of *Flavobacteria* have been
569 related to high values of chlorophyll a and nutrients (Abell & Bowman 2005b),
570 (Pommier et al. 2007). Moreover, members of this group have been repeatedly
571 reported as prone to particle attachment (DeLong et al. 1993, Crump et al. 1999),
572 and as primary colonizers of marine phytoplankton detritus (Pinhassi et al. 2004,
573 Abell & Bowman 2005a). In addition, they have been frequently found in
574 association to natural and induced phytoplankton blooms (Simon et al. 1999,

575 Riemann et al. 2000, O'Sullivan et al. 2004, Grossart et al. 2005, Fandino et al.
576 2005).
577
578 The specific probes (POL740, CF6-1267 and VIS1-575) represented on average
579 30% of the cells targeted by the general probe CF319a, and did not target any
580 sequence found in our study, indicating need to further explore diversity in this
581 group to find out most prominent groups in this environment. Moreover, all
582 sequences retrieved were habitat exclusive, pointing to a potentially high local
583 diversity of members of this phylum, as previously observed for other coastal
584 environments (Alonso et al. 2007).
585
586 Among the probe-targeted groups, the genus *Polaribacter* was the most abundant,
587 particularly in the riverine and frontal zones. DE2 and VIS 1 targeted bacteria also
588 exhibited their higher numbers in the frontal zone (Fig. 4). Members of these three
589 groups have been previously observed to be especially abundant in the
590 phycosphere of marine phytoplankton (Gomez-Pereira et al. 2009). In our study,
591 members of these groups were particularly abundant at the samples with higher
592 chlorophyll a values, providing further hints of a potential link between these
593 Flavobacterial clades and phytoplankton abundance. In addition, abundance of
594 VIS1 bacteria also seemed to increase with depth, mostly in the frontal zone, but
595 also in the marine stations (Fig. 5, suppl. Fig. 1).
596
597 **Water and sediment host separate bacterial communities**
598 The degree of similarity between the sediment samples and the water samples

599 taken at the same stations were very low for all three habitats, as indicated by
600 several similarity indexes (suppl. Table 2). This pattern, initially derived from the
601 ARISA results, was further confirmed by comparing the sequences retrieved at the
602 marine S9 from the surface water and the sediment. The proportion of sequences
603 of the main groups retrieved in these two libraries, suggests a likely explanation of
604 the differences in the fingerprint. At a first glance, this comparison shows a virtual
605 absence of sequences affiliated to *Alphaproteobacteria* together with an increased
606 number of sequences affiliated to *Deltaproteobacteria* found in the sediment, in
607 clear contrast to the water column. Moreover, sequences affiliated to *Acidobacteria*
608 appeared in the sediment sample, and were not retrieved in any other library.
609 Furthermore, a closer look into the sequences of two major groups which appeared
610 in both clone libraries, *Gammaproteobacteria* and *Actinobacteria*, revealed that
611 none of those sequences was shared by the water surface libraries and the
612 sediment library. In all cases, the sequences derived from the water surface had
613 their closest relatives among sequences found in other water column libraries,
614 whereas the closest neighbours of sequences found in the sediment were also
615 previously retrieved from marine sediments and ocean crust (suppl. Table 3).
616 Altogether, our findings suggest that despite the shallowness of this system, there
617 is a marked niche specialization of the bacterial groups inhabiting the sediment and
618 the water column.
619
620 **Which portion of the gradient is more diverse?**
621 Although the diversity indexes for the ARISA results did not revealed a strong
622 pattern, there were still some indications about differences in the diversity found

623 along the gradient. ARISA showed the lower richness and evenness, together with
624 the maximal number of unique OTUs, at the frontal zone. This may indicate that
625 less phylotypes would be able to live in the frontal zone, and that a substantial
626 portion of them would be habitat specific. Crump (2004), using DGGE, also found a
627 lower diversity in the estuarine region when compared to the riverine and marine
628 regions in the Plume Island sound (Crump et al. 2004). This observation is also in
629 agreement with the tendency observed for other planktonic communities in the
630 transition Río de la Plata-Atlantic Ocean, e.g. mesozooplankters also have shown
631 increasing diversity towards the marine portions and a decrease in the frontal zone
632 of the Río de la Plata estuary (Cervetto et al. 2006).
633
634 Interestingly, both, the ARISA amplicons and the sequences which were common
635 to all habitats, were retrieved in the highest proportions, whereas the peaks and
636 sequences habitat unique were always a minor fraction of the total RFI or total
637 number of sequences (suppl.Table 1, suppl.Table 3) indicating that the most
638 widespread members of the community were also the most abundant. This finding
639 is in agreement with the results of a global marine survey based on clone libraries
640 (Pommier et al. 2007). However, neither ARISA, nor clone libraries yield
641 quantitative information: Thus, further data are needed to determine actual
642 abundances at the level of single microbial populations.
643
644 **Conclusions**
645 Our multi-level study indicates that the Río de la Plata hosts highly diverse and
646 strongly environmentally structured bacterial communities. Some groups appeared

647 widely distributed, while others seemed to be rather restricted, even in this relative
648 small spatial scale. In the future, it would be interesting to quantify both types of
649 populations with respect to the environmental factors which may determine these
650 spatial patterns.
651
652 **Acknowledgments**
653 The authors would like to thank Pilar Castro and Angel Segura for assistance
654 during the sampling campaign. Germán Pérez is acknowledged for valuable help
655 during nutrient analysis. The Crew and accompanying scientific staff from the
656 Aldebaran are acknowledged for providing an excellent working atmosphere. This
657 study was funded by the Max Planck Society; the sampling campaign was
658 supported by DINARA.
659
660
661
662
663
664
665
666
667

668 References

669 Abell GCJ, Bowman JP (2005a) Colonization and community dynamics of class
670 Flavobacteria on diatom detritus in experimental mesocosms based on Southern
671 Ocean seawater. FEMS Microbiology Ecology 53:379-391
672 Abell GCJ, Bowman JP (2005b) Ecological and biogeographic relationships of class
673 Flavobacteria in the Southern Ocean. FEMS Microbiology Ecology 51:265-277
674 Acha EM, Mianzan H, Guerrero R, Carreto J, Giberto D, Montoya N, Carignan M (2008)
675 An overview of physical and ecological processes in the Rio de la Plata Estuary.
676 Continental Shelf Research 28:1579- 1588
677 Alonso-Sáez L, Balagué V, Sà E, L. , Sánchez O, González J, M. , Pinhassi J, Massana R,
678 Pernthaler J, Pedrós-Alió C, Gasol J, M. (2007) Seasonality in bacterial diversity in
679 north-west Mediterranean coastal waters: assessment through clone libraries,
680 fingerprinting and FISH. FEMS Microbiology Ecology 60:98-112
681 Alonso-Saez L, Gasol JM (2007) Seasonal Variations in the Contributions of Different
682 Bacterial Groups to the Uptake of Low-Molecular-Weight Compounds in
683 Northwestern Mediterranean Coastal Waters. Applied and Environmental
684 Microbiology 73:3528-3535
685 Alonso C, Pernthaler J (2006) Roseobacter and SAR11 dominate microbial glucose uptake
686 in coastal North Sea waters. Environmental Microbiology 8:2022-2030
687 Alonso C, Warnecke F, Amann R, Pernthaler J (2007) High local and global diversity of
688 Flavobacteria in marine plankton. Environmental Microbiology 9:1253-1266
689 Alonso C, Zeder M, Piccini C, Conde D, Pernthaler J (2009) Ecophysiological differences
690 of betaproteobacterial populations in two hydrochemically distinct compartments of
691 a subtropical lagoon. Environmental Microbiology
692 Anton J, Rossello-Mora R, Rodriguez-Valera F, Amann R (2000) Extremely Halophilic
693 Bacteria in Crystallizer Ponds from Solar Salterns. Appl. Environ. Microbiol.
694 66:3052-3057
695 Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2006) New Screening
696 Software Shows that Most Recent Large 16S rRNA Gene Clone Libraries Contain
697 Chimeras. Appl. Environ. Microbiol. 72:5734-5741
698 Bianchi AA, Bianucci L, Piola AR, Ruiz Pino D, Schloss I, Poisson A, Balestrini CF
699 (2005) Vertical stratification and air-sea CO2 fluxes in the Patagonian shelf Journal
700 of Geophysical Research 110:C07003
701 Bisbal GA (1995) The Southeast South American shelf large marine ecosystem: Evolution
702 and components. Marine Policy 19:21-38
703 Bouvier TC, del Giorgio PA (2002) Compositional changes in free-living bacterial
704 communities along a salinity gradient in two temperate estuaries. Limnology and
705 Oceanography 47:453-470
706 Brown MV, Schwalbach MS, Hewson I, Fuhrman JA (2005) Coupling 16S-ITS rDNA
707 clone libraries and automated ribosomal intergenic spacer analysis to show marine
708 microbial diversity: development and application to a time series. Environmental
709 Microbiology 7:1466-1479
710 Burkert U, Warnecke F, H.-D. B, Zwirnmann E, Pernthaler J (2003) Members of a readily
711 enriched beta-proteobacterial clade are common in the surface waters of a humic
712 lake. Applied and Environmental Microbiology 69:6550-6559

713 Calliari D, Brugnoli E, Ferrari G, Vizziano D (2009) Phytoplankton distribution and
714 production along a wide environmental gradient in the South-West Atlantic off
715 Uruguay. Hydrobiologia 620:47-61
716 Cardinale M, Brusetti L, Quatrini P, Borin S, Puglia AM, Rizzi A, Zanardini E, Sorlini C,
717 Corselli C, Daffonchio D (2004) Comparison of Different Primer Sets for Use in
718 Automated Ribosomal Intergenic Spacer Analysis of Complex Bacterial
719 Communities. Appied and Environmental Microbiology 70:6147-6156
720 Carlson CA, Morris R, Parsons R, Treusch AH, Giovannoni SJ, Vergin K (2008) Seasonal
721 dynamics of SAR11 populations in the euphotic and mesopelagic zones of the
722 northwestern Sargasso Sea. ISME J 3:283-295
723 Cervetto G, Calliari D, Rodriguez-Graña L, Lacerot G, Castiglioni R (2006) Zooplankton
724 de ambientes costeros de Uruguay: añadiendo piezas al rompecabezas. In: Menafra
725 R, Rodriguez-Gallego L, Scarabino F, Conde D (eds) Bases para la conservación y
726 el manejo de la costa uruguaya. Vida Silvestre, Montevideo, p 105-112
727 Clarke KR, Gorley RN (2006) PRIMER v6: User manual/tutorial. PRIMER-E: Plymouth,
728 UK.
729 Clarke KR, Somerfield PJ, Gorley RN (2008) Testing of null hypotheses in exploratory
730 community analyses: similarity profiles and biota-environment linkage. J Exp Mar
731 Biol Ecol 366:56-69
732 Colwell RK (2005) EstimateS: Statistical estimation of species richness and shared species
733 from samples. Version 7.5. User's Guide and application.
734 http://purl.oclc.org/estimates
735 Cottrell MT, Kirchman DL (2003) Contribution of major bacterial groups to bacterial
736 biomass production (thymidine and leucine incorporation) in the Delaware estuary.
737 Limnology and Oceanography 48:168-178
738 Crump BC, Armbrust EV, Baross JA (1999) Phylogenetic analysis of particle-attached and
739 free-living bacterial communities in the Columbia River, its estuary, and the
740 adjacent coastal ocean. Applied and Environmental Microbiology 65:3192-3204
741 Crump BC, Hopkinson CS, Sogin ML, Hobbie JE (2004) Microbial biogeography along an
742 estuarine salinity gradient: Combined influences of bacterial growth and residence
743 time. Applied and Environmental Microbiology 70:1494-1505
744 Daims H, Bruhl A, Amann R, Schleifer KH, Wagner M (1999) The domain-specific probe
745 EUB338 is insufficient for the detection of all *Bacteria*: Development and
746 evaluation of a more comprehensive probe set. Systematic and Applied
747 Microbiology. 22:434-444
748 DeLong EF, Franks DG, Alldredge AL (1993) Phylogenetic diversity of aggregate-attached
749 vs. free-living marine bacterial assemblages. Limnology and Oceanography 38:924-
750 934
751 Eaton AD, Clesceri LS, Rice EW, Greenberg AE, Franson MAH (eds) (2005) Standard
752 Methods for the Examination of Water and Wastewater: Centennial Edition. , Vol.
753 American Public Health Association. , Washington, D.C.
754 Eilers H, Pernthaler J, Glöckner FO, Amann R (2000) Culturability and in situ abundance
755 of pelagic bacteria from the North Sea. Applied and Environmental Microbiology
756 66:3044-3051
757 Eilers H, Pernthaler J, Peplies J, Glöckner FO, Gerdts G, Amann R (2001) Isolation of
758 novel pelagic bacteria from the German Bight and their seasonal contribution to
759 surface picoplankton. Applied and Environmental Microbiology 67:5134-5142

760 Fandino LB, Riemann L, Steward GF, Azam F (2005) Population dynamics of Cytophaga-
761 Flavobacteria during marine phytoplankton blooms analyzed by real-time
762 quantitative PCR. Aquatic Microbial Ecology 40:251-257
763 Feely RA, Sabine CL, Takahashi T, Wanninkhof R (2001) Uptake and Storage of Carbon
764 Dioxide in the Ocean: The Global C02 Survey. Oceanography 14:18-32
765 Framiñan MB, Brown OB (1996) Study of the Río de la Plata turbidity front Part I: Spatial
766 and temporal distribution. Continental Shelf Research 16:1259-1282
767 Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ (2008) Critical
768 Evaluation of Two Primers Commonly Used for Amplification of Bacterial 16S
769 rRNA Genes. Appied and Environmental Microbiology 74:2461-2470
770 Garcia-Martinez J, Rodriguez-Valera F (2000) Microdiversity of uncultured marine
771 prokaryotes: the SAR11 cluster and the marine Archaea of Group I. Molecular
772 Ecology 9:935-948
773 Glöckner F-O, Zaichikov E, Belkova N, Denissova L, Pernthaler J, Pernthaler A, Amann R
774 (2000) Comparative 16S rRNA analysis of lake bacterioplankton reveals globally
775 distributed phylogenetic clusters including an abundant group of Actinobacteria.
776 Applied and Environmental Microbiology 66:5053-5065
777 Glöckner FO, Fuchs BM, Amann R (1999) Bacterioplankton compositions of lakes and
778 oceans: a first comparison based on fluorescence in situ hybridization. Applied and
779 Environmental Microbiology 65:3721-3726
780 Gomez-Pereira PR, Fuchs BM, Alonso C, Oliver M, van Beusekom J, Amann R (2009)
781 Distribution patterns and diversity of planktonic Flavobacterial clades in contrasting
782 water masses of the North Atlantic Ocean. ISME Journal in press
783 González JM, Moran MA (1997) Numerical dominance of a group of marine bacteria in the
784 alpha-subclass of the class *Proteobacteria* in coastal seawater. Applied and
785 Environmental Microbiology 63:4237-4242
786 González JM, Simo R, Massana R, Covert JS, Casamayor EO, Pedros-Alio C, Moran MA
787 (2000) Bacterial community structure associated with a
788 dimethylsulfoniopropionate-producing North Atlantic algal bloom. Applied &
789 Environmental Microbiology 66:4237-4246
790 Grossart HP, Levold F, Allgaier M, Simon M, Brinkhoff T (2005) Marine diatom species
791 harbour distinct bacterial communities. Environmental Microbiology 7:860-873
792 Guerrero RA, Acha EM, Framiñan MB, Lasta CA (1997) Physical Oceanography of the
793 Río de la Plata estuary, Argentina. . Continental Shelf Research 17
794 Hahn M, Pöckl M, Wu Q (2005) Low Intraspecific Diversity in a Polynucleobacter
795 Subcluster Population Numerically Dominating Bacterioplankton of a Freshwater
796 Pond. Applied and Environmental Microbiology 71:4539-4547
797 Hewson I, Fuhrman JA (2004) Richness and Diversity of Bacterioplankton Species along
798 an Estuarine Gradient in Moreton Bay, Australia. Applied and Environmental
799 Microbiology 70:3425-3433
800 Hollibaugh JT, Wong PS, Murrell MC (2000) Similarity of particle-associated and free-
801 living bacterial communities in northern San Francisco Bay, California. Aquatic and
802 Microbial Ecology 21:103-114
803 Jeffrey SW, Humphrey GF (1975) New Spectrophotometric equations for determining
804 chlorophylls a, b, c1 and c2 in higher plants, algae and natural phytoplankton.
805 Biochemie und Physiologie Planzen 167:191-194

806 Kirchman DL, A. I. Dittel RR, Malmstrom, Cottrell. MT (2005) Biogeography of major
807 bacterial groups in the Delaware estuary. Limnol. Oceanogr. 50:1697-1706.
808 Kirchman DL, Yu LY, Cottrell MT (2003) Diversity and abundance of uncultured
809 Cytophaga-like bacteria in the Delaware Estuary. Applied and Environmental
810 Microbiology 69:6587-6596
811 Koroleff F (1976) Determination of ammonia. In: Grasshoft K (ed) Methods of Seawater
812 Analysis. Verlag Chemie, p 126-133
813 Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination
814 of species data. Oecologia 129:271-280
815 Legendre P, Legendre L (1998) Numerical Ecology, 2nd English ed.edn.|. Publisher|, City
816 Lorenzen CJ (1967) Determination of chlorophyll and pheopigments: Spectrophotometric
817 equations. Limnology and Oceanography 12:343-346
818 Malmstrom RR, Straza TRA, Cottrell MT, Kirchman DL (2007) Diversity, abundance, and
819 biomass production of bacterial groups in the western Arctic Ocean. Aquatic
820 Microbial Ecology 47:45-55
821 Manz W, Amann R, Ludwig W, Vancanneyt M, Schleifer K-H (1996) Application of a
822 suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria
823 of the phylum *Cytophaga-Flavobacter-Bacteroides* in the natural environment.
824 Microbiology 142:1097-1106
825 Manz W, Amann R, Ludwig W, Wagner M, Schleifer K-H (1992) Phylogenetic
826 oligodeoxynucleotide probes for the major subclasses of *Proteobacteria*: Problems
827 and solutions. Systematic and Applied Microbiology 15:593-600
828 Martinez CM, Retta S (2001) Caracterizacion de las areas de cria de la corvina
829 (Micropogonias furnieri) en la zona costera uruguaya. In: Vizziano D, Puig,
830 Mesones, Nagy, G. (ed) El Rio de la Plata. Investigacion para la gestion del
831 Ambiente, los recursos pesqueros y la pesqueria en el frente salino. Programa
832 Ecoplata, Montevideo, p 141-148:
833 Moran MA, González JM, Kiene RP (2003) Linking a bacterial taxon to sulfur cycling in
834 the sea: Studies of the marine Roseobacter group. Geomicrobiology Journal 20:375-
835 388
836 Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ
837 (2002) SAR11 clade dominates ocean surface bacterioplankton communities.
838 Nature 420:806-810
839 Murphy J, Riley JP (1962) A modified single solution method for the determination of
840 phosphate in natural waters. Anal. Chim. Acta 27
841 Murray AE, Hollibaugh JT, Orrego C (1996) Phylogenetic compositions of
842 bacterioplankton from two California estuaries compared by denaturing gradient gel
843 electrophoresis of 16S rDNA fragments. Appied and Environmental Microbiology
844 62:2676-2680
845 Muyzer G, de Waal EC, Uitterlinden AG (1993) Profiling of complex microbial
846 populations by denaturing gradient gel electrophoresis analysis of polymerase chain
847 reaction-amplified genes coding for 16S rRNA. Applied and Environmental
848 Microbiology 59:695-700
849 Nagy G, Gómez M, López C, Perdomo A (2002) Distribution patterns of nutrients and
850 symptoms of eutrophication in the Río de la Plata Estuary System. Hydrobiologia
851 475/476:125-139

852 Neef A (1997) Anwendung der *in situ*-Einzelzell-Identifizierung von Bakterien zur
853 Populationsanalyse in komplexen mikrobiellen Biozönosen. Ph.D. thesis, Techical
854 University, Munich, Germany, München
855 Neef A, Amann R, Schlesner H, Schleifer K-H (1998) Monitoring a widespread bacterial
856 group: in situ detection of planctomycetes with 16S rRNA-targeted probes.
857 Microbiology 144:3257-3266
858 Neefs JM, Vandepeer Y, Hendriks L, Dewachter R (1990) Compilation of small ribosomal-
859 subunit RNA sequences. Nucleic Acids Research 18:2237-2317
860 O'Sullivan LA, Fuller KE, Thomas EM, Turley CM, Fry JC, Weightman AJ (2004)
861 Distribution and culturability of the uncultivated 'AGG58 cluster' of Bacteroidetes
862 phylum in aquatic environments. Fems Microbiology Ecology 47:359-370
863 Pernthaler A, Pernthaler J, Amann R (2002) Fluorescence in situ hybridization and
864 catalyzed reporter deposition for the identification of marine bacteria. Applied and
865 Environmental Microbiology 68:3094-3101
866 Piccini C, Conde D, Alonso C, Sommaruga R, Pernthaler J (2006) Blooms of Single
867 Bacterial Species in a Coastal Lagoon of the Southwestern Atlantic Ocean. Applied
868 and Environmental Microbiology 72:6560-6568
869 Pinhassi J, Berman T (2003) Differential growth response of colony-forming alpha- and
870 gamma-proteobacteria in dilution culture and nutrient addition experiments from
871 Lake Kinneret (Israel), the eastern Mediterranean Sea, and the Gulf of Eilat.
872 Applied and Environmental Microbiology 69:199-211
873 Pinhassi J, Sala MM, Havskum H, Peters F, Guadayol O, Malits A, Marrase CL (2004)
874 Changes in bacterioplankton composition under different phytoplankton regimens.
875 Applied and Environmental Microbiology 70:6753-6766
876 Pommier T, CanbÄCk B, Riemann L, BostrÖM KH, Simu K, Lundberg P, Tunlid A,
877 HagstrÖM Å (2007) Global patterns of diversity and community structure in marine
878 bacterioplankton. Molecular Ecology 16:867-880
879 Ramette A (2007) Multivariate analyses in microbial ecology. FEMS Microbiology
880 Ecology 62:142-160
881 Ramette A (2009) Quantitative community fingerprinting methods for estimating the
882 abundance of operational taxonomic units in natural microbial communities.
883 Applied and Environmental Microbiology:AEM.02409-02408
884 Riemann L, Steward GF, Azam F (2000) Dynamics of bacterial community composition
885 and activity during a mesocosm diatom bloom. Applied and Environmental
886 Microbiology 66:578-587
887 Roller C, Wagner M, Amann R, Ludwig W, Schleifer K-H (1994) *In situ* probing of Gram-
888 positive bacteria with high DNA G+C content using 23S rRNA-targeted
889 oligonucleotides. Microbiology 140:2849-2858
890 Schattenhofer M, Fuchs B, M. , Amann R, Zubkov M, V. , Tarran G, A. , Pernthaler J
891 (2009) Latitudinal distribution of prokaryotic picoplankton populations in the
892 Atlantic Ocean. Environmental Microbiology 11:2078-2093
893 Schloss PD, Handelsman J (2005) Introducing DOTUR, a Computer Program for Defining
894 Operational Taxonomic Units and Estimating Species Richness. Appl. Environ.
895 Microbiol. 71:1501-1506
896 Simek K, Pernthaler J, Weinbauer M, Hornak K, Dolan J, Nedoma J, Masin M, Amann R
897 (2001) Changes in bacterial community composition, dynamics and viral mortality

898 rates associated with enhanced flagellate grazing in a mesoeutrophic reservoir.
899 Applied and Environmental Microbiology 67:2723-2733
900 Simon M, Glöckner FO, Amann R (1999) Different community structure and temperature
901 optima of heterotrophic picoplankton in various regions of the Southern Ocean.
902 Aquatic Microbial Ecology 18:275-284
903 Stackebrandt E, Goebel B (1994) Taxonomic note: a place for DNA-DNA reassociation
904 and 16S rRNA sequence analysis in the present species definition in bacteriology.
905 Int J Syst Bacteriol 44:846-849
906 Strickland JDH, Parsons TR (1972) A practical handbook of sea water analysis. Fish. Res.
907 Bd. Can. Bull. 167:207-211
908 Troussellier M, Schafer H, Batailler N, Bernard L, Courties C, Lebaron P, Muyzer G,
909 Servais P, Vives-Rego J (2002) Bacterial activity and genetic richness along an
910 estuarine gradient (Rhone River plume, France). Aquatic Microbial Ecology 28:13-
911 24
912 Valderrama JC (1981) The simultaneous analysis of total N and P in natural waters. Mar.
913 Chem 10:1009-1022
914 Vila-Costa M, Pinhassi J, Alonso C, Pernthaler J, Sim, Rafel (2007) An annual cycle of
915 dimethylsulfoniopropionate-sulfur and leucine assimilating bacterioplankton in the
916 coastal NW Mediterranean. Environmental Microbiology 9:2451-2463
917 Warnecke F, Sommaruga R, Hofer JS, Pernthaler J (2005) Abundances, identity and growth
918 state of Actinobacteria in mountain lakes of different transparency. Applied and
919 Environmental Microbiology 71:5551-5559
920 Watanabe K, Watanabe K, Kodama Y, Syutsubo K, Harayama S (2000) Molecular
921 Characterization of Bacterial Populations in Petroleum-Contaminated Groundwater
922 Discharged from Underground Crude Oil Storage Cavities. Appl. Environ.
923 Microbiol. 66:4803-4809
924 Woebken D, Fuchs BM, Kuypers MMM, Amann R (2007) Potential Interactions of
925 Particle-Associated Anammox Bacteria with Bacterial and Archaeal Partners in the
926 Namibian Upwelling System. Appl. Environ. Microbiol. 73:4648-4657
927 Zhang Y, Jiao N, Cottrell MT, Kirchman DL (2006) Contribution of major bacterial groups
928 to bacterial biomass production along a salinity gradient in the South China Sea.
929 Aquatic Microbial Ecology 43:233-241
930 Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition.
931 Appied and Environmental Microbiology 62:316-322
932
933

934 Tables and Figures

Table 1. Effects of environment and space on bacterial community structure and abundance.

	ARISA		CARD-FISH	
	F-ratio[a]	*P*[b]	*F*-ratio	*P*
ENV + Space 2D	3.93	0.020*	4.40	0.004**
ENV + Space 3D	3.31	0.008**	4.83	0.005**
ENV	3.77	0.002**	7.97	0.001***
Space 3D	1.84	0.041*	2.87	0.030*
Space 2D	2.33	0.020*	3.56	0.012*
ENV removing space 3D	3.14	0.020*	4.09	0.012*
ENV removing space 2D	3.55	0.001***	3.24	0.014*
Space 3D removing ENV	1.88	0.078 ns	1.58	0.178 ns
Space 2D removing ENV	2.36	0.020*	1.41	0.289 ns

[a]Redundancy analysis (RDA) was applied to Hellinger-transformed ARISA community data or to CARD-FISH-based bacterial abundance under simple constrained models or partial models (i.e. by removing the effects of certain factors whose effects have to be accounted for) and the corresponding F-ratio are indicated. The different factors were environment (ENV) as represented by salinity and temperature, space 2D (i.e. longitude and latitude) and space 3D (i.e. longitude, latitude and water depth).
[b]The significance of each model was assessed using 1,000 permutations of the model residuals at $P<0.05$. Different symbols represent the level of significance with * ($P\leq0.05$), ** ($P\leq0.01$), *** ($P\leq0.001$), and ns (not significant; $P>0.05$).

935
936

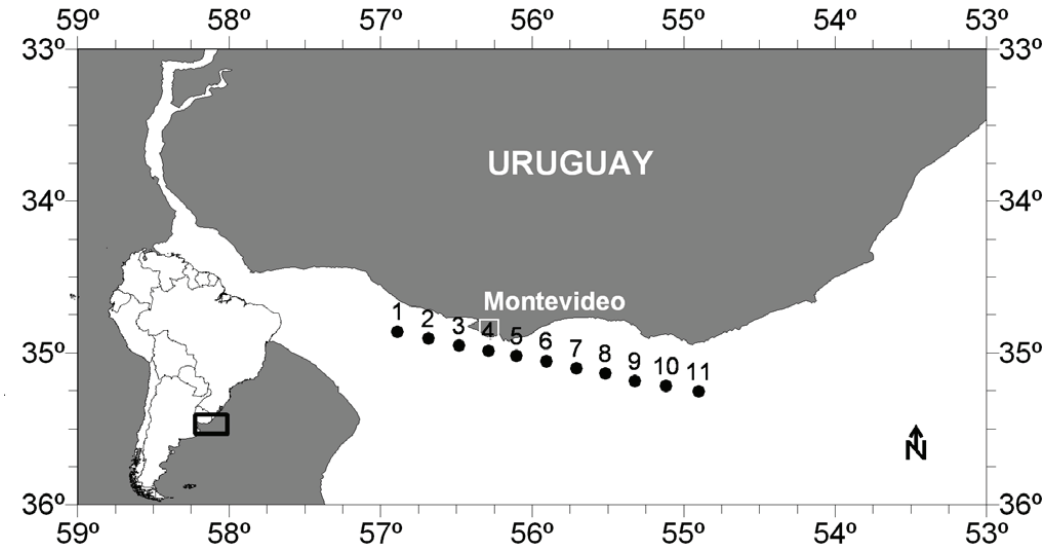


Fig. 1. Transect of the ARTEMISA1 cruise indicating sampling stations.

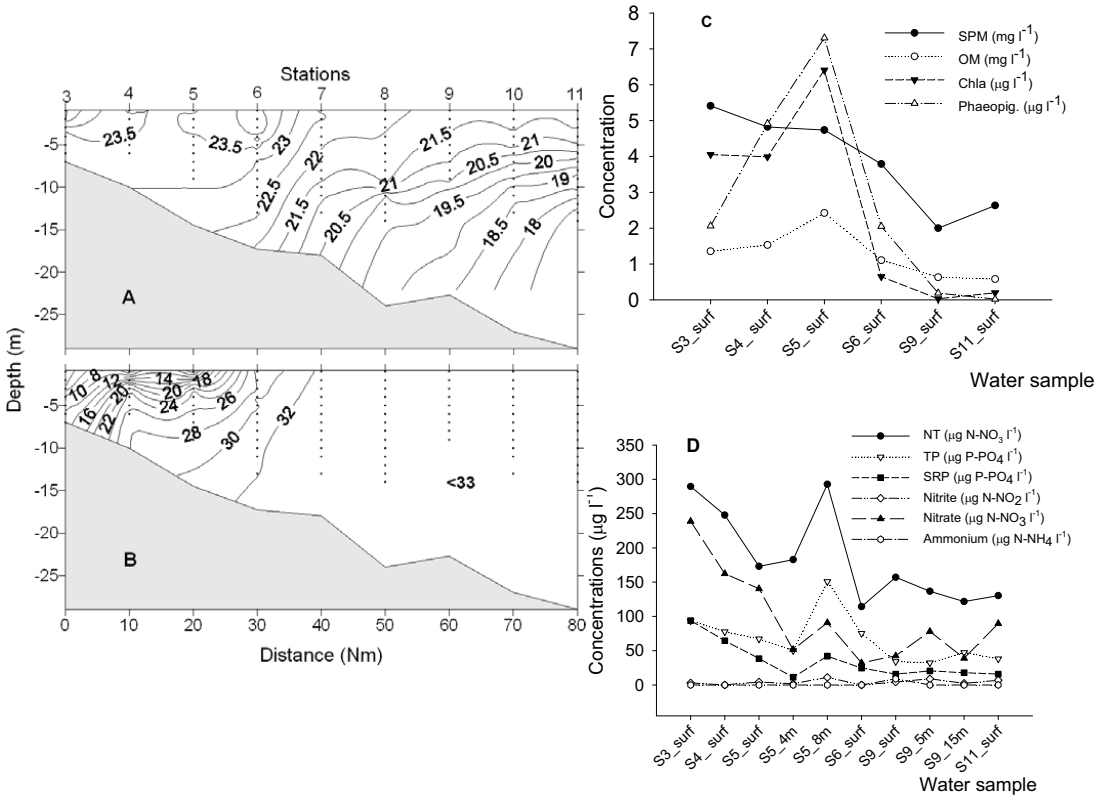


Fig. 2. Distribution of (A) Temperature, (B) Salinity, (C) Suspended particulate matter (SPM), organic matter (OM), chlorophyll a (Chla), phaeopigments (Phaeop.), and (D) total nitrogen (TN), total phosphorus (TP), nitrite, nitrate, ammonium, and soluble reactive phosphorus (SRP) along the sampling transect.

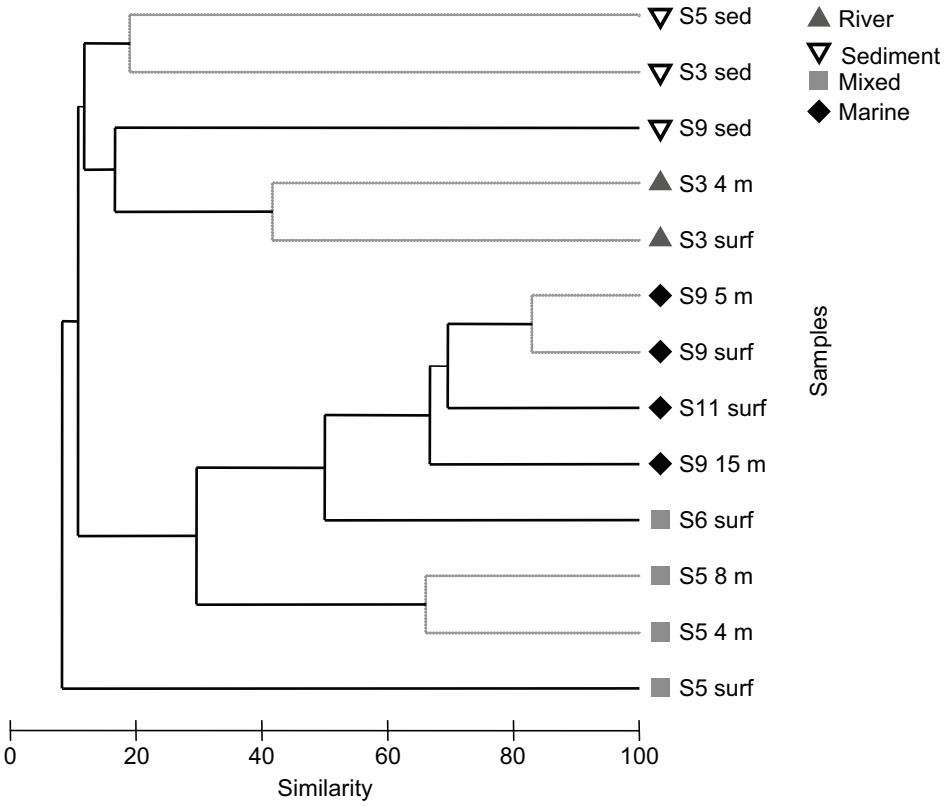


Fig.3 Sample clustering based on ARISA. Cluster analysis was performed using group average based on RFI using Bray Curtis similarity index. Black lines in the dendrogram represent significant clusters at the 99% level using SIMPROF test; grey dashed lines represent not significant clusters at this level.

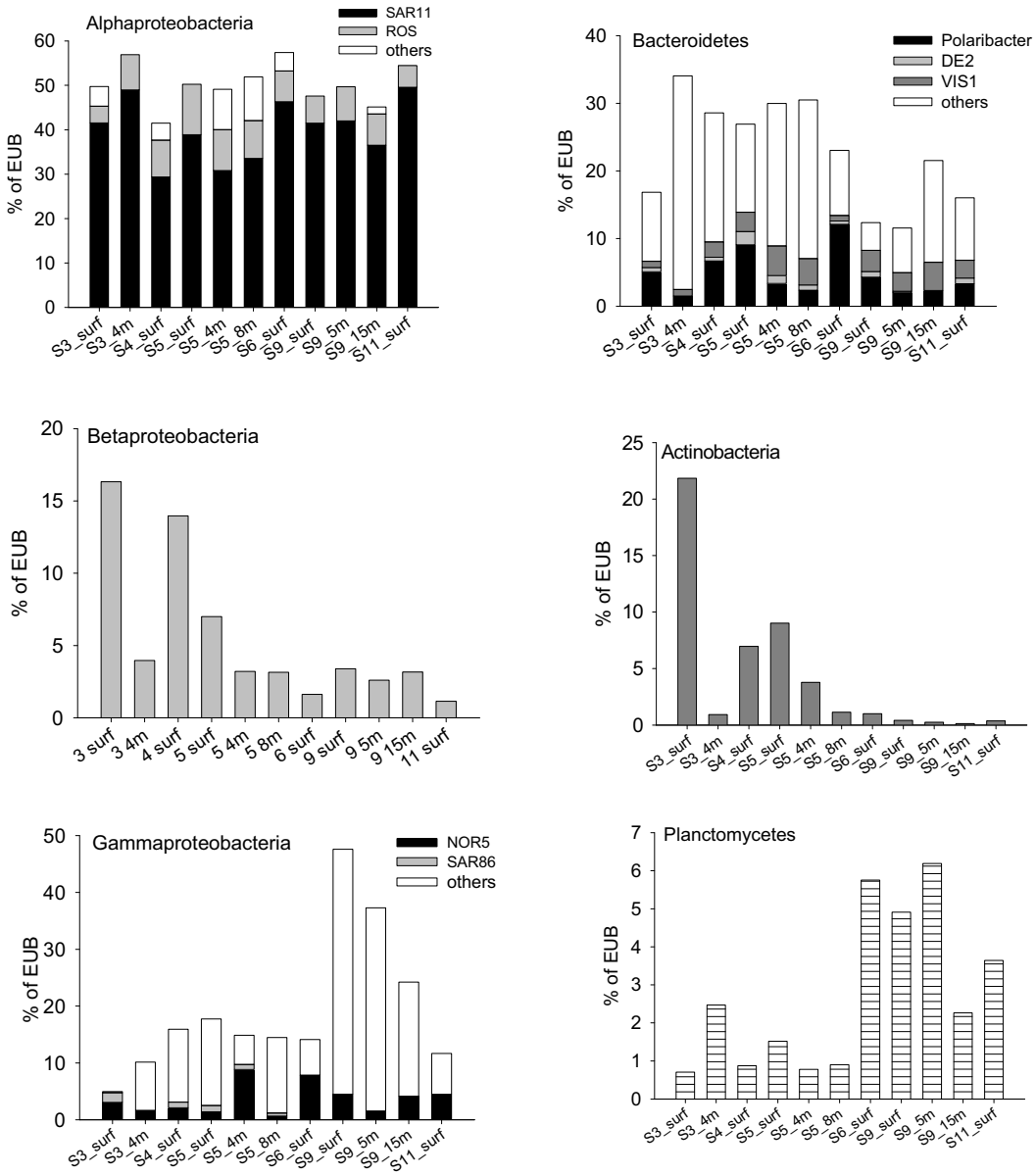


Fig. 4. Abundance of the main bacterial groups as determined by CARD-FISH. Values expressed as percentage of cells hybridised with the EUB I_III probe. Others indicate the difference in counts between the sum of the specific probes and the general probe used to target a given bacterial clade.

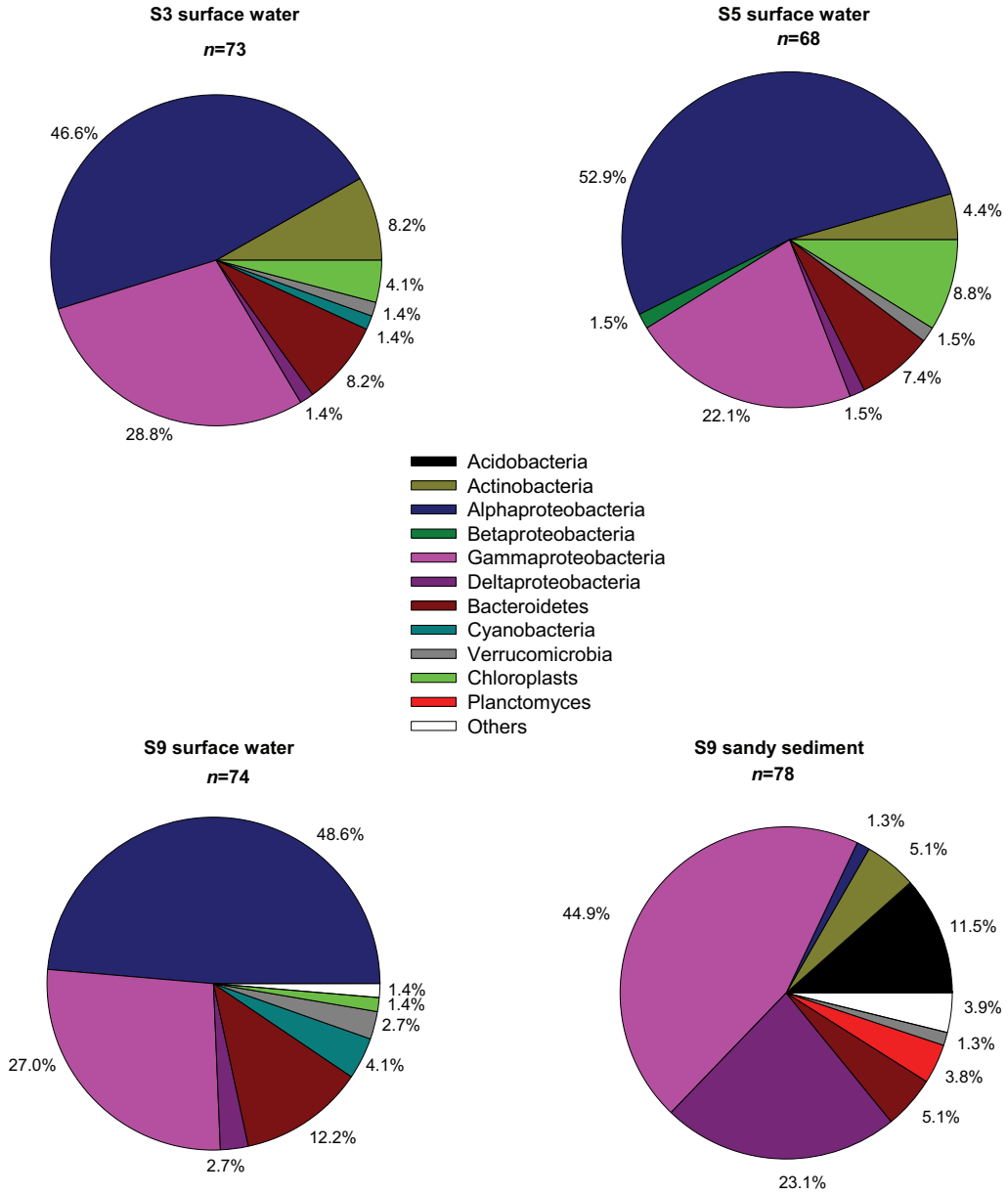
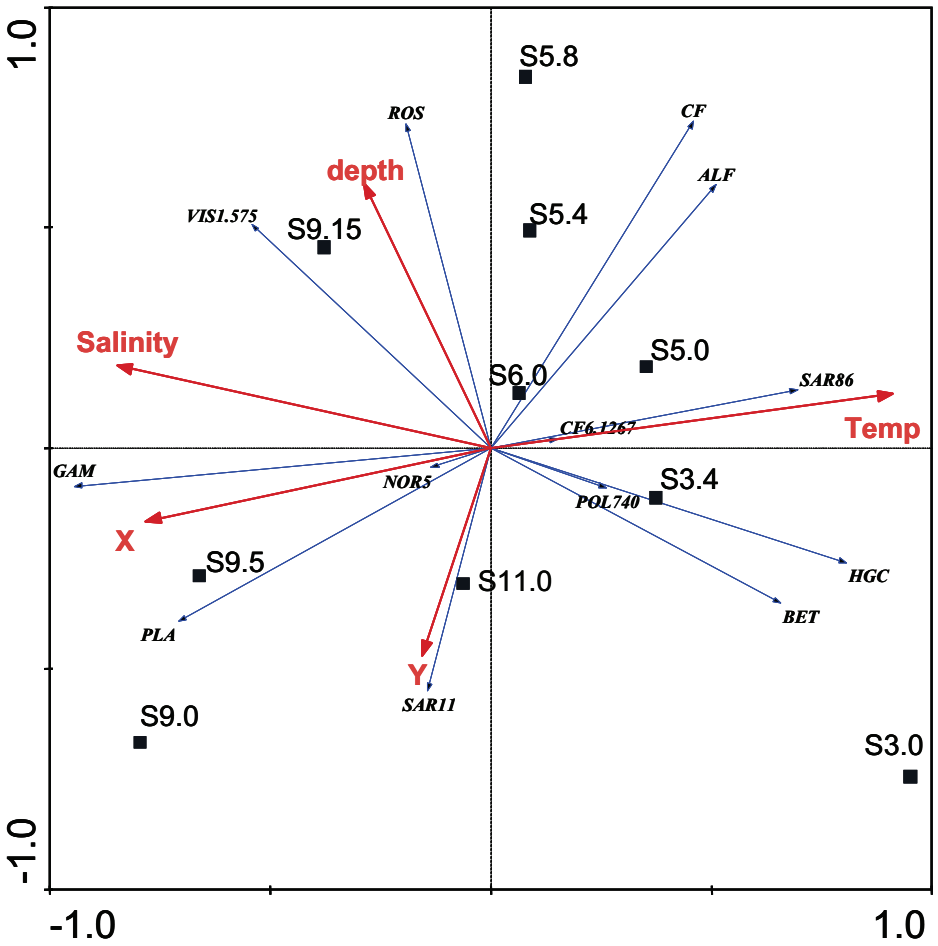
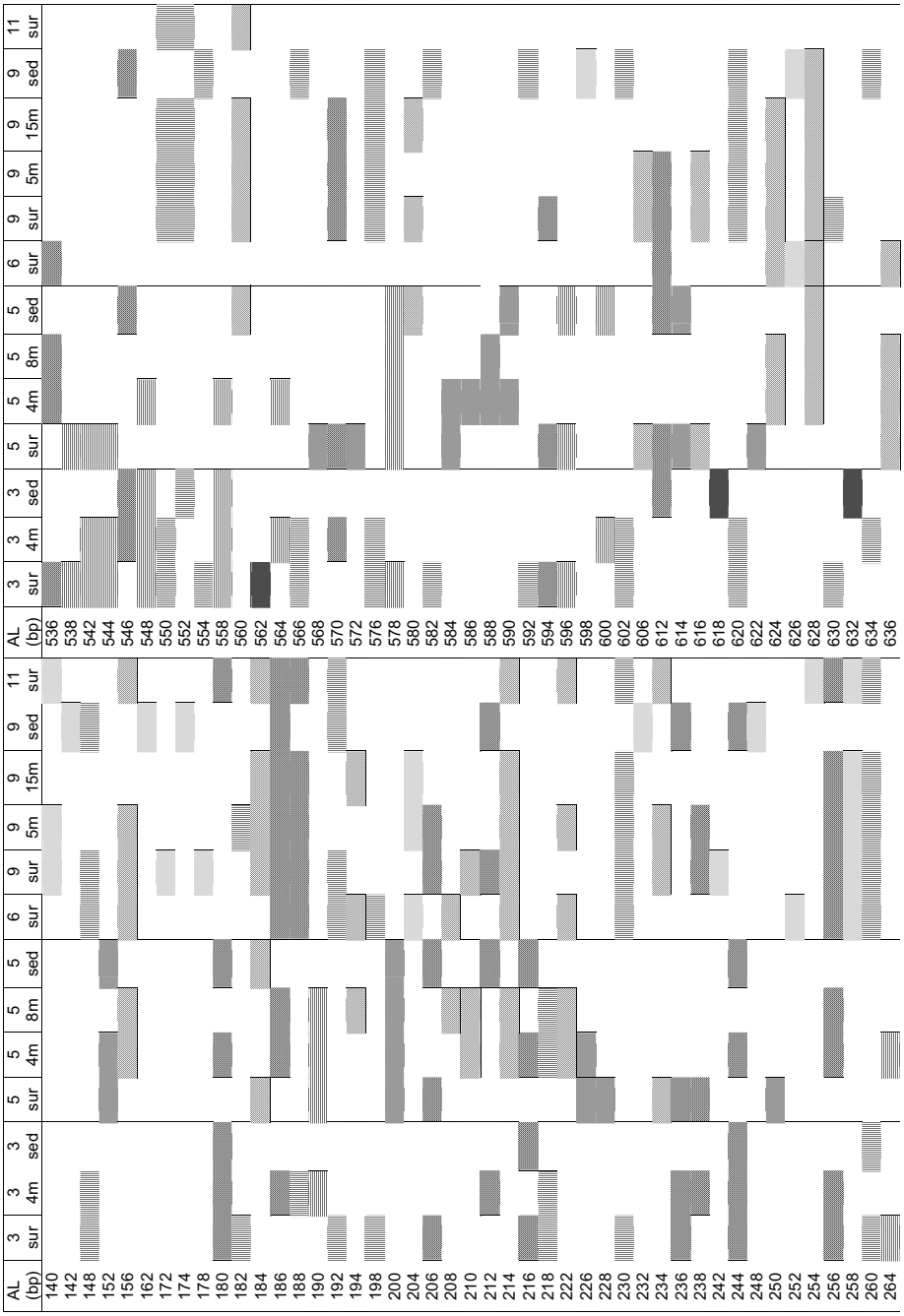


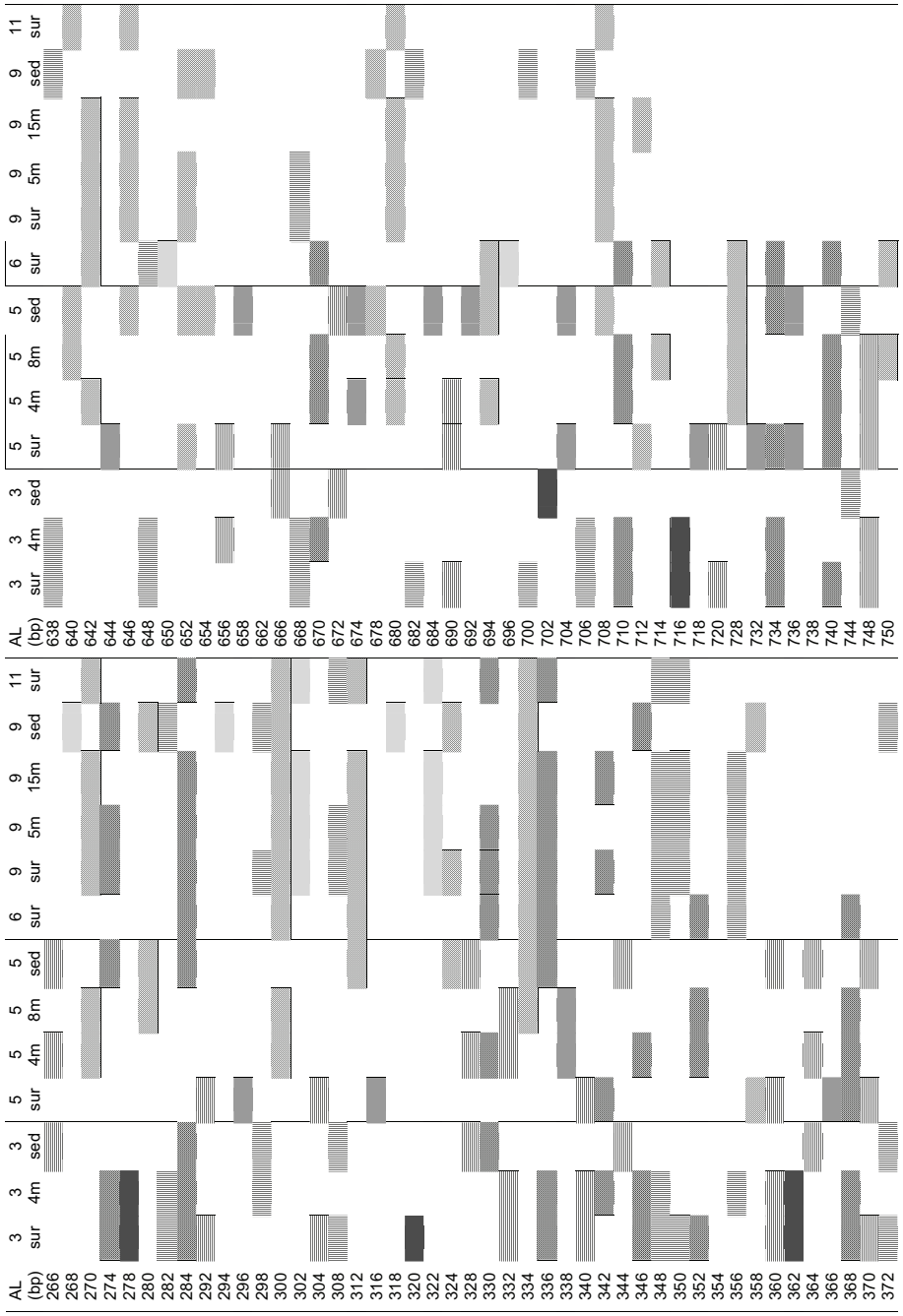
Fig. 5. Clone libraries of the selected riverine, frontal, and marine samples. The percentages indicate the contribution of a given taxonomic group to the total of sequences retrieved in each library.

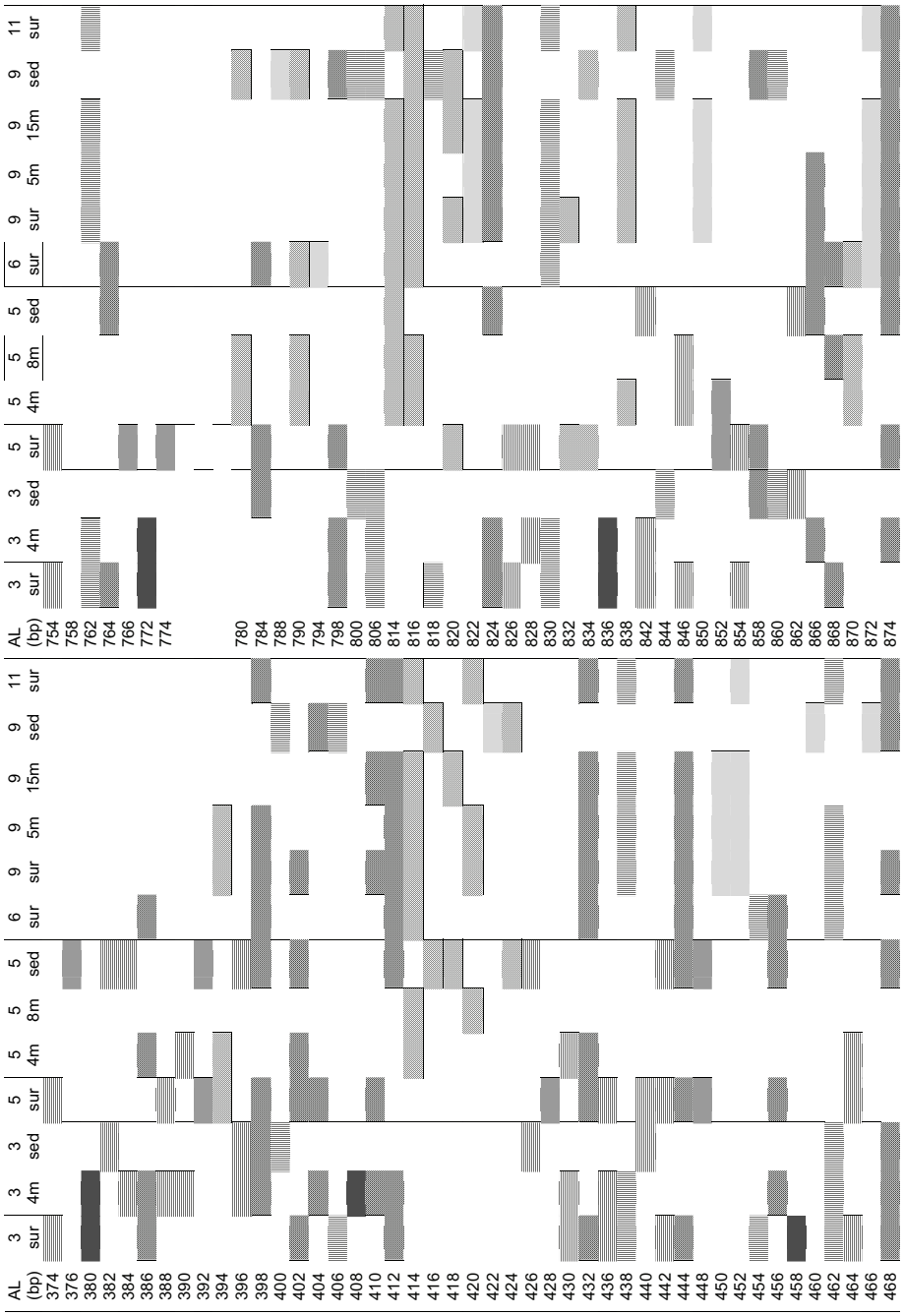


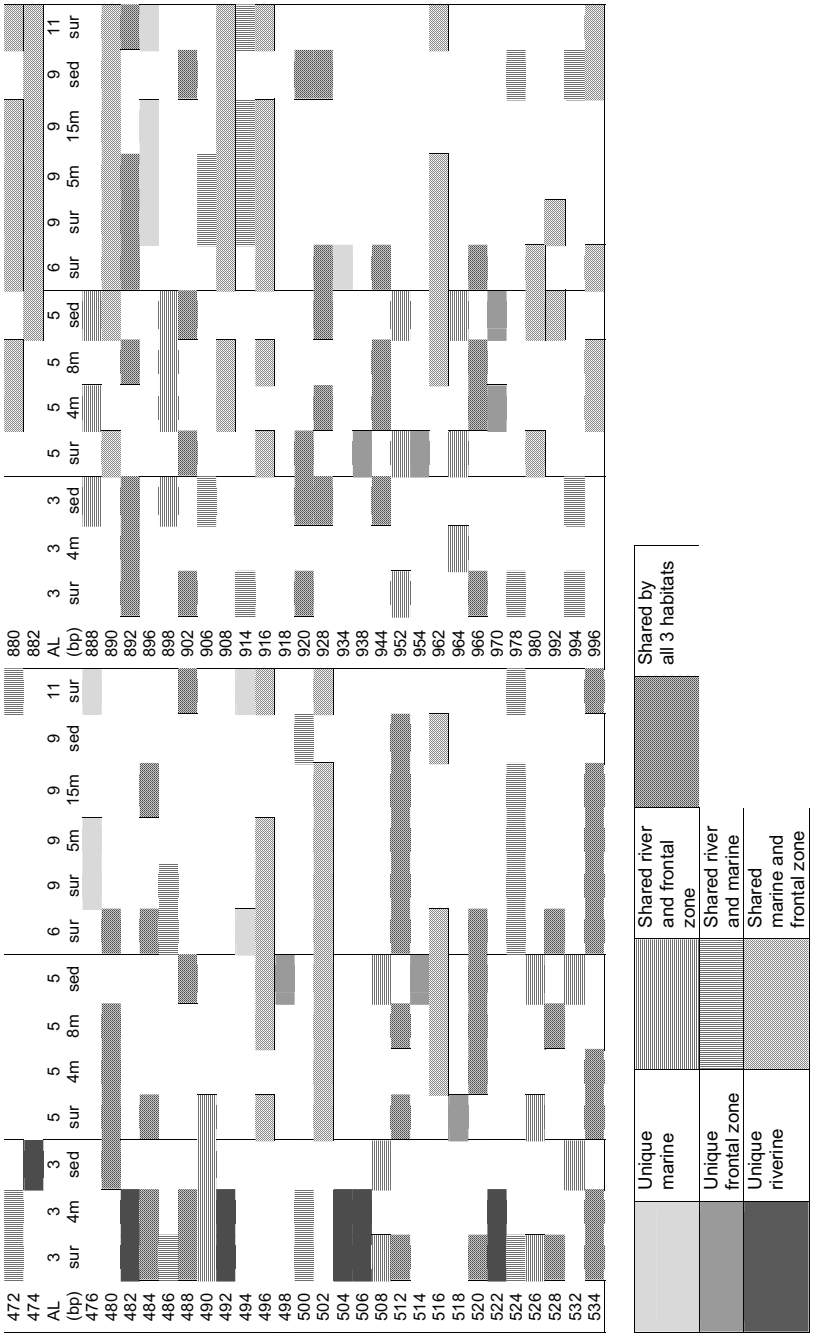
Suppl. Fig. 1. RDA plot for the abundance of bacterial groups as defined by CARD-FISH.

Suppl. Table 1. ARISA fragments. AL means amplicon length. The different shadings indicate the habitat/s of retrieval




AL (bp)
3 sur
3 4m
3 sed
5 sur
5 4m
5 8m
5 sed
6 sur
9 sur
9 5m
9 15m
9 sed
11 sur
266
268
270
274
278
280
282
284
292
294
296
298
300
302
304
308
312
316
318
320
322
324
328
330
332
334
336
338
340
342
344
346
348
350
352
354
356
358
360
362
364
366
368
370
372
638
640
642
644
646
648
650
652
654
656
658
662
666
668
670
672
674
678
680
682
684
690
692
694
696
700
702
704
706
708
710
712
714
716
718
720
728
732
734
736
738
740
744
748
750


AL (bp)
3 sur
3 4m
3 sed
5 sur
5 4m
5 8m
5 sed
6 sur
9 sur
9 5m
9 15m
9 sed
11 sur


AL (bp)
3 sur
3 4m
3 sed
5 sur
5 4m
5 8m
5 sed
6 sur
9 5m
9 15m
9 sed
11 sur
AL (bp)
3 sur
3 4m
3 sed
5 sur
5 4m
5 8m
5 sed
6 sur
9 5m
9 15m
9 sed
11 sur
Unique marine
Unique frontal zone
Unique riverine
Shared river and frontal zone
Shared river and marine
Shared marine and frontal zone
Shared by all 3 habitats

Suppl.Table 2. Sample similarity indexes based on incidence data of the ARISA fragments calculated with the software EstimateS.

First Sample	Second Sample	Sobs First Sample	Sobs Second Sample	Shared Species Observed	Jaccard	Sorensen/ MorisitaHorn/ Bray-Curtis
3 surface	3 4m	124	92	63	0.412	0.583
3 surface	3 sediment	124	54	17	0.106	0.191
3 surface	5 surface	124	103	38	0.201	0.335
3 surface	5 4m	124	79	27	0.153	0.266
3 surface	5 8m	124	60	18	0.108	0.196
3 surface	5 sediment	124	94	26	0.135	0.239
3 surface	6 surface	124	90	34	0.189	0.318
3 surface	9 surface	124	114	38	0.19	0.319
3 surface	9 5m	124	96	32	0.17	0.291
3 surface	9 15m	124	73	23	0.132	0.234
3 surface	9 sediment	124	83	31	0.176	0.3
3 surface	11 surface	124	94	33	0.178	0.303
3 4m	3 sediment	92	54	14	0.106	0.192
3 4m	5 surface	92	103	26	0.154	0.267
3 4m	5 4m	92	79	19	0.125	0.222
3 4m	5 8m	92	60	11	0.078	0.145
3 4m	5 sediment	92	94	22	0.134	0.237
3 4m	6 surface	92	90	24	0.152	0.264
3 4m	9 surface	92	114	33	0.191	0.32
3 4m	9 5m	92	96	27	0.168	0.287
3 4m	9 15m	92	73	21	0.146	0.255
3 4m	9 sediment	92	83	24	0.159	0.274
3 4m	11 surface	92	94	30	0.192	0.323
3 sediment	5 surface	54	103	11	0.075	0.14
3 sediment	5 4m	54	79	14	0.118	0.211
3 sediment	5 8m	54	60	4	0.036	0.07
3 sediment	5 sediment	54	94	23	0.184	0.311
3 sediment	6 surface	54	90	11	0.083	0.153
3 sediment	9 surface	54	114	12	0.077	0.143
3 sediment	9 5m	54	96	10	0.071	0.133
3 sediment	9 15m	54	73	3	0.024	0.047
3 sediment	9 sediment	54	83	16	0.132	0.234
3 sediment	11 surface	54	94	11	0.08	0.149
5 surface	5 4m	103	79	20	0.123	0.22
5 surface	5 8m	103	60	12	0.079	0.147
5 surface	5 sediment	103	94	31	0.187	0.315
5 surface	6 surface	103	90	20	0.116	0.207
5 surface	9 surface	103	114	29	0.154	0.267
5 surface	9 5m	103	96	23	0.131	0.231
5 surface	9 15m	103	73	15	0.093	0.17
5 surface	9 sediment	103	83	14	0.081	0.151
5 surface	11 surface	103	94	16	0.088	0.162
5 4m	5 8m	79	60	45	0.479	0.647
5 4m	5 sediment	79	94	23	0.153	0.266

5 4m	6 surface	79	90	38	0.29	0.45
5 4m	9 surface	79	114	28	0.17	0.29
5 4m	9 5m	79	96	27	0.182	0.309
5 4m	9 15m	79	73	21	0.16	0.276
5 4m	9 sediment	79	83	13	0.087	0.16
5 4m	11 surface	79	94	28	0.193	0.324
5 8m	5 sediment	60	94	14	0.1	0.182
5 8m	6 surface	60	90	43	0.402	0.573
5 8m	9 surface	60	114	28	0.192	0.322
5 8m	9 5m	60	96	28	0.219	0.359
5 8m	9 15m	60	73	21	0.188	0.316
5 8m	9 sediment	60	83	12	0.092	0.168
5 8m	11 surface	60	94	29	0.232	0.377
5 sediment	6 surface	94	90	26	0.165	0.283
5 sediment	9 surface	94	114	35	0.202	0.337
5 sediment	9 5m	94	96	29	0.18	0.305
5 sediment	9 15m	94	73	21	0.144	0.251
5 sediment	9 sediment	94	83	23	0.149	0.26
5 sediment	11 surface	94	94	30	0.19	0.319
6 surface	9 surface	90	114	56	0.378	0.549
6 surface	9 5m	90	96	55	0.42	0.591
6 surface	9 15m	90	73	42	0.347	0.515
6 surface	9 sediment	90	83	18	0.116	0.208
6 surface	11 surface	90	94	55	0.426	0.598
9 surface	9 5m	114	96	93	0.795	0.886
9 surface	9 15m	114	73	68	0.571	0.727
9 surface	9 sediment	114	83	23	0.132	0.234
9 surface	11 surface	114	94	85	0.691	0.817
9 5m	9 15m	96	73	65	0.625	0.769
9 5m	9 sediment	96	83	16	0.098	0.179
9 5m	11 surface	96	94	81	0.743	0.853
9 15m	9 sediment	73	83	15	0.106	0.192
9 15m	11 surface	73	94	62	0.59	0.743
9 sediment	11 surface	83	94	16	0.099	0.181

Suppl. Table 3. Unique full sequences 16S rRNA gene sequences types (sequence similarity ≥99%) recovered during this study, their taxonomic affiliation according to the RDP classifier system, and their closest neighbors in the GenBank database. Contiguous shadow rows indicate 97% similarity among sequences

Accession numbers	Taxonomic Affiliation (RDP/ARB)	Sample	Closest neighbour (BLAST) Accession Number, Clone name, % of sequence similarity
GU230187	Alphaproteobacteria Rickettsiales	Riverine water	EU800863. Uncultured bacterium clone 2C229098 99%
GU230188 - GU230190	Alphaproteobacteria uncultured	Frontal zone water	EU802131. Uncultured bacterium clone 3C003543 99%
GU230191	Alphaproteobacteria uncultured	Frontal zone water	EU803783. Uncultured bacterium clone 5C231412 99%
GU230192	Alphaproteobacteria Rickettsiales	Marine water	EU799453. Uncultured bacterium clone 1C227055 99%
GU230193- GU230194	Alphaproteobacteria Octadecabacter	Marine water	EU804911. Uncultured bacterium clone 6C232887 99%
GU230195 - GU230199	Alphaproteobacteria uncultured	Riverine and frontal zone waters	EU802244. Uncultured bacterium clone 3C003683 99%
GU230200	Alphaproteobacteria Thalssobacter et rel uncultured	Riverine water	DQ234107.2 Uncultured Rhodobacteraceae bacterium clone DS023 99%
GU230201	Alphaproteobacteria Sulfitobacterium	Marine water	EU804384. Uncultured bacterium clone 6C232296 100%
GU230202- GU230233	Alphaproteobacteria SAR11	Riverine, frontal zone and marine waters	Candidatus *Pelagibacter ubique* HTCC1062, complete genome 98-99%
GU230234- GU230245	Alphaproteobacteria SAR11	Marine water	EU804517. Uncultured bacterium clone 6C232455 99%
GU230246- GU230251	Alphaproteobacteria Uncultured	Riverine, frontal zone and marine waters	FJ545630. Uncultured bacterium clone W2-55 99%
GU230252- GU230254	Alphaproteobacteria uncultured	Marine water	EU804407. Uncultured bacterium clone 6C232323 99%
GU230255- GU230256	Alphaproteobacteria uncultured	Marine water	DQ009168. Uncultured marine bacterium clone SPOTSOCT00_5m28 99%
GU230257- GU230260	Alphaproteobacteria Rhodospirillaceae	Riverine, frontal zone and marine waters	DQ009262. Uncultured marine bacterium clone SPOTSAUG01_5m94 99%
GU230261	Alphaproteobacteria Tistrella et rel	Sandy sediment	AM176885. Uncultured bacterium partial 16S rRNA gene, clone SZB60 94%
GU230294- GU230295	Gammaproteobacteria Oceanospirillales	Riverine water	EF574686. Uncultured bacterium clone S25_1030 99%
GU230296	Gammaproteobacteria uncultured	Riverine water	DQ146989. Vibrio sp. V639 90%
GU230297	Gammaproteobacteria NOR5 related	Riverine water	EU010152. Uncultured marine bacterium clone B81 99%
GU230298- GU230299	Gammaproteobacteria NOR5 related	Marine water	U70696. Unidentified gamma proteobacterium OM60 99%
GU230300	Gammaproteobacteria Oceanospirillales	Marine water	FM242300. Uncultured gamma proteobacterium 95%
GU230301	Gammaproteobacteria uncultured	Marine water	EU394558. Uncultured gamma proteobacterium clone PEACE2006/192_P2 99%
GU230302- GU230303	Gammaproteobacteria Alteromonadales Endobugula et rel	Marine water	DQ009136. Uncultured marine bacterium clone SPOTSAUG01_5m75 98%

GU230304-GU230305	Gammaproteobacteria Acinetobacter	Marine water	FJ193150. Uncultured Acinetobacter sp. clone GI6-10b-C06 99%
GU230306	Gammaproteobacteria Psychrobacter	Marine water	DQ200849. Hyphomicrobium sp. YAAJ-4 99%
GU230307-GU230310	Gammaproteobacteria SAR86 related	Riverine water	EU803167. Uncultured bacterium clone 4C230601 97-98%
GU230311-GU230318	Gammaproteobacteria SAR86 related	Riverine, frontal zone and marine waters	EF107104.Uncultured marine bacterium EB80_02D08 98-99%
GU230319-GU230320	Gammaproteobacteria SAR86 related	Riverine water	Uncultured gamma proteobacterium clone PEACE2006/168_P2 99%
GU230321	Gammaproteobacteria SAR86 related	Frontal zone water	EU799251. Uncultured bacterium clone 1C226829 99%
GU230322	Gammaproteobacteria Oceanospirillales	Frontal zone water	EF572116. Uncultured bacterium clone S23_215 99%
GU230323	Gammaproteobacteria Oceanospirillales	Riverine water	EU803061. Uncultured bacterium clone 4C230470 97%
GU230324-GU230325	Gammaproteobacteria Oceanospirillales	Riverine and frontal zone waters	EF471638. Uncultured gamma proteobacterium clone CB01E06 99%
GU230326-GU230329	Gammaproteobacteria NOR5 related	Frontal zone and marine waters	EU799180. Uncultured bacterium clone 1C226751 99%
GU230330-GU230335	Gammaproteobacteria NOR5 related	Riverine, frontal zone and marine waters	EF379865. Uncultured bacterium clone PC-FL10-45 99%
GU230336-GU230337	Gammaproteobacteria Nevskia Hydrogenophaga	Sandy sediment	DQ395002. Uncultured Pseudomonas sp. clone VHS-B4-54 99%
GU230338	Gammaproteobacteria Nevskia Hydrogenophaga	Sandy sediment	EU491400. clone P0X3b5H10 98%
GU230339	Gammaproteobacteria Nevskia Hydrogenophaga	Sandy sediment	EU287133. Uncultured bacterium clone P13-40 99%
GU230340	Gammaproteobacteria Nevskia Hydrogenophaga	Sandy sediment	EU491380. clone P0X3b1H06 98%
GU230341-GU230342	Gammaproteobacteria Nevskia Hydrogenophaga	Sandy sediment	DQ351778. clone Belgica 2005/10-140-5 99%
GU230343	Gammaproteobacteria uncultured	Sandy sediment	EU799754. Uncultured bacterium clone 1C227399 95%
GU230344	Gammaproteobacteria uncultured	Sandy sediment	EU735694. Uncultured bacterium clone SN146 90%
GU230345	Gammaproteobacteria NOR5 related	Sandy sediment	DQ351759. Uncultured gamma proteobacterium clone Belgica2005/10-130-14 97%
GU230346	Gammaproteobacteria Leucothrix	Sandy sediment	AM040128. Uncultured gamma proteobacterium, clone Sylt 32 99%
GU230347	Gammaproteobacteria Francisella et rel	Sandy sediment	DQ337099. Uncultured bacterium clone EV818SWSAP88 94%
GU230348	Gammaproteobacteria Nevskia Hydrogenophaga	Sandy sediment	DQ351802. clone Belgica2005/10-ZG-6 97%
GU230349	Gammaproteobacteria Alteromonadales	Sandy sediment	EF575144. Uncultured bacterium clone S25_1488 93%
GU230350	Gammaproteobacteria Nevskia Hydrogenophaga	Sandy sediment	EU287158. Uncultured bacterium clone P13-65 95%
GU230351	Gammaproteobacteria Thalassomonas	Sandy sediment	AJ294747. Thalassomonas viridans 16S rRNA gene, strain XOM5, CECT 5082 96%
GU230352	Gammaproteobacteria Ifremeria	Sandy sediment	DQ351743. Uncultured gamma proteobacterium clone Belgica2005/10-120-8 99%
GU230353	Gammaproteobacteria Nitrosococcus et rel	Sandy sediment	AF287298. Nitrosococcus halophilus Nc4 87%
GU230354	Gammaproteobacteria Nevskia Hydrogenophaga	Sandy sediment	EU652512. Uncultured bacterium clone D8s-65 96%

GU230355	Gammaproteobacteria Francisella et rel	Sandy sediment	DQ337099. Uncultured bacterium clone EV818SWSAP88 93%
GU230356	Gammaproteobacteria No esta en arb????	Sandy sediment	EF125401. Uncultured bacterium clone MSB-1D8 94%
GU230357	Gammaproteobacteria uncultured	Sandy sediment	FJ545485. Uncultured bacterium clone S1-58 99%
GU230358	Gammaproteobacteria uncultured	Sandy sediment	DQ351757. Uncultured gamma proteobacterium clone Belgica2005/10-130-10 96%
GU230359	Gammaproteobacteria Alteromonadales	Sandy sediment	EF575144. Uncultured bacterium clone S25_1488 94%
GU230360	Gammaproteobacteria uncultured	Sandy sediment	EU652540. Uncultured bacterium clone B8S-21 98%
GU230385	Deltaproteobacteria Bacteriovorax	Frontal zone water	FJ203142. Uncultured bacterium clone SHFG544 94%
GU230386	Deltaproteobacteria	Marine water	EU328012. Uncultured Pelobacter sp. clone B392 87%
GU230387	Deltaproteobacteria Haliangium	Sandy sediment	EU652608. Uncultured bacterium clone B8S-115 91%
GU230388	Deltaproteobacteria uncultured	Sandy sediment	AM039962. Uncultured delta proteobacterium, clone 4aFS 99%
GU230389	Deltaproteobacteria Nannocystis et rel	Sandy sediment	FJ516990. Uncultured Haliangiaceae bacterium clone TDNP_Wbc97_101_1_8 91%
GU230390	Deltaproteobacteria uncutured	Sandy sediment	EU734953. Uncultured bacterium clone 045B58 95%
GU230391	Deltaproteobacteria uncultured	Sandy sediment	EU617842. Uncultured bacterium clone D13S-27 99%
GU230392	Deltaproteobacteria uncultured	Sandy sediment	DQ811828. Uncultured delta proteobacterium clone MSB-5C5 94%
GU230393	Deltaproteobacteria Chondromyces-Polyangium - Nannocystis	Sandy sediment	DQ811828. Uncultured delta proteobacterium clone MSB-5C5 94%
GU230394	Deltaproteobacteria Desulfobacteraceae	Sandy sediment	EU487993. Uncultured bacterium clone CK_1C3_15 95%
GU230395-GU230396	Deltaproteobacteria Uncultured	Sandy sediment	FJ264659. Uncultured bacterium clone OrigSedB19 95%
GU230397	Deltaproteobacteria	Sandy sediment	DQ351798. Uncultured delta proteobacterium clone Belgica2005/10-ZG-2 99%
GU230398	Deltaproteobacteria uncultured	Sandy sediment	EU734950. Uncultured bacterium clone 042C45 98%
GU230399	Deltaproteobacteria Syntrophobacteraceae	Sandy sediment	FJ264586. Uncultured bacterium clone Mn3b-B19 99%
GU230400	Deltaproteobacteria Chondromyces-Polyangium - Nannocystis	Sandy sediment	AB243821. Uncultured bacterium, clone: Niigata-10 86%
GU230401	Deltaproteobacteria Desulfobacteraceae	Sandy sediment	DQ811817. Uncultured delta proteobacterium clone MSB-4G8 97%
GU230407	Bacteroidetes Polaribacter-Tenacibaculum	Riverine water	EF575134. Uncultured bacterium clone S25_1478 99%
GU230408	Bacteroidetes Polaribacter-Tenacibaculum	Riverine water	FJ545595. Uncultured bacterium clone W2-10 99%
GU230409	Bacteroidetes Aequorivita	Riverine water	EU391659.Bacteroidetes bacterium LE20 98%
GU230410	Bacteroidetes Sphingobacteriaceae	Frontal zone water	FJ612395.Uncultured bacterium clone DP10.3.91 99%
GU230411	Bacteroidetes uncultured	Frontal zone water	AY907777. Uncultured bacterium clone A712004 99%
GU230412	Bacteroidetes uncultured	Frontal zone water	EF574855. Uncultured bacterium clone S25_1199 99%

GU230413-GU230414	Bacteroidetes Pseudocytophaga	Marine water	DQ009099. Uncultured marine bacterium clone SPOTSAUG01_5m55 98%
GU230415-GU230417	Bacteroidetes Pseudocytophaga	Marine water	AM279187. Uncultured Flavobacteria bacterium 16S rRNA gene, clone NorSea39 97%
GU230418	Bacteroidetes Flexibacter aggregans	Marine water	EF573775. Uncultured bacterium clone S25_119 99%
GU230419	Bacteroidetes Sphingobacteriaceae	Marine water	AM279161. Uncultured Bacteroidetes bacterium 16S rRNA gene, clone NorSea83 98%
GU230420	Bacteroidetes Gelidibacter-Psychroserpens	Riverine and marine waters	EF574795. Uncultured bacterium clone S25_1139 94%
GU230421	Bacteroidetes Pseudocytophaga	Sandy sediment	DQ334617. Uncultured bacterium clone HB1-48-23 98%
GU230422	Bacteroidetes Lewinella	Sandy sediment	DQ269109. Uncultured Bacteroidetes bacterium clone DPC179 96%
GU230431	Acidobacteria /Gp22	Sandy sediment	EU373931. Uncultured Acidobacterium sp. clone HCM3MC78_3D_FL
GU230432	Acidobacteria uncultured	Sandy sediment	Uncultured Acidobacteria bacterium clone 3B1820-07 94%
GU230433	Acidobacteria /Gp5	Sandy sediment	EU925922. Uncultured bacterium clone 160I37 98%
GU230434-GU230435	Acidobacteria /Gp22	Sandy sediment	EU287178. Uncultured bacterium clone P13-85 96%
GU230436	Acidobacteria /Gp10	Sandy sediment	DQ351805. Uncultured Acidobacteriales bacterium clone Belgica2005/10-ZG-11 97%
GU230437	Acidobacteria	Sandy sediment	EU287102. Uncultured bacterium clone P13-9 98%
GU230438	Acidobacteria /Gp22	Sandy sediment	EU617725. Uncultured bacterium clone C13S-2 98%
GU230440-GU230442	Actinobacteria Acidimicrobiaceae	Riverine water	EF471463. Uncultured actinobacterium clone CB11F03 99%
GU230443	Actinobacteria Acidimicrobiaceae	Riverine water	EU800855. Uncultured bacterium clone 2C229085 100%
GU230444-GU230445	Actinobacteria Microthrix	Frontal zone water	EF471700. Uncultured actinobacterium clone CB01B09 99%
GU230446	Actinomycetales Acidimicrobiaceae	Sandy sediment	FJ624300. Uncultured bacterium clone 115SZ4 98%
GU230447	Actinobacteria Acidimicrobiaceae	Sandy sediment	EU491398. Uncultured bacterium clone P0X3b5H07 96%
GU230453	Planctomycetaceae uncultured	Sandy sediment	EU735003. Uncultured bacterium clone 097B67 98%
GU230454	Planctomycetaceae uncultured	Sandy sediment	EU386084. Uncultured bacterium clone MD2900-B9 90%
GU230455	Planctomycetaceae uncultured	Sandy sediment	EU135167. Uncultured bacterium clone FFCH5957 92%
GU230456	Verrucomicrobia Epixenosomes et rel	Riverine water	EU010227. Uncultured marine bacterium clone B67 98%
GU230457	Verrucomicrobia Epixenosomes et rel	Frontal zone water	FJ826388. Uncultured marine bacterium clone A6-3-80 98%
GU230458-GU230459	Verrucomicrobia Prosthecobacter-Verrucomicrobium	Marine water	DQ513026. Uncultured bacterium clone FS142-18B-02 99%
GU230460	Verrucomicrobia Prosthecobacter-Verrucomicrobium	Sandy sediment	EU919768. Uncultured bacterium clone b28 16S 95%
GU230461	Cyanobacteria Prochlorococcus et rel	Riverine water	EF471456. Uncultured Synechococcus sp. clone CB11C04 99%
GU230462	Cyanobacteria Synechococcus	Marine water	CP000110. Synechococcus sp. CC9605 99%

GU230463-GU230464	Cyanobacteria/GpI	Marine water	AM230685. Calothrix sp. BECID30 95%
GU230465	Betaproteobacteria Methylophilaceae	Frontal zone water	EU799255. Uncultured bacterium clone 1C226833 99%
GU230466	Lentisphaera	Sandy sediment	AY390429. Lentisphaera araneosa strain HTCC2160 98%
GU230467	OP3	Sandy sediment	EU491137. Uncultured bacterium clone P9X2b3F04 91%

Manuscript IV

Bacterioplankton diversity and community composition in the southern Lagoon of Venice

Francesca Simonato, Paola R. Gómez-Pereira, Bernhard M. Fuchs, and Rudolf Amann

Systematic and Applied Microbiology
2010, 33, 128-138

Systematic and Applied Microbiology 33 (2010) 128–138

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Bacterioplankton diversity and community composition in the Southern Lagoon of Venice

Francesca Simonato, Paola R. Gómez-Pereira, Bernhard M. Fuchs, Rudolf Amann *

Max Planck Institute for Marine Microbiology (MPIMM), Celsiusstrasse 1, D-28359 Bremen, Germany

ARTICLE INFO

Article history:
Received 1 October 2009

Keywords:
Lagoon of Venice
Bacterioplankton
16S rRNA
CARD-FISH
Flavobacteria

ABSTRACT

The Lagoon of Venice is a large water basin that exchanges water with the Northern Adriatic Sea through three large inlets. In this study, the 16S rRNA approach was used to investigate the bacterial diversity and community composition within the southern basin of the Lagoon of Venice and at one inlet in October 2007 and June 2008. Comparative sequence analysis of 645 mostly partial 16S rRNA gene sequences indicated high diversity and dominance of *Alphaproteobacteria*, *Gammaproteobacteria* and *Bacteroidetes* at the lagoon as well as at the inlet station, therefore pointing to significant mixing. Many of these sequences were close to the 16S rRNA of marine, often coastal, bacterioplankton, such as the *Roseobacter* clade, the family *Vibrionaceae*, and class *Flavobacteria*. Sequences of *Actinobacteria* were indicators of a freshwater input. The composition of the bacterioplankton was quantified by catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) with a set of rRNA-targeted oligonucleotide probes. CARD-FISH counts corroborated the dominance of members of the phyla *Alphaproteobacteria*, *Gammaproteobacteria* and *Bacteroidetes*. When assessed by a probe set for the quantification of selected clades within *Alphaproteobacteria* and *Gammaproteobacteria*, bacterioplankton composition differed between October 2007 and June 2008, and also between the inlet and the lagoon. In particular, members of the readily culturable copiotrophic gammaproteobacterial genera *Vibrio*, *Alteromonas* and *Pseudoalteromonas* were enriched in the southern basin of the Lagoon of Venice. Interestingly, the alphaproteobacterial SAR11 clade and related clusters were also present in high abundances at the inlet and within the lagoon, which was indicative of inflow of water from the open sea.

Introduction

Littoral ecosystems such as lagoons, estuaries, and salt marshes are among the most productive natural ecosystems on Earth [1,14]. The ecology of coastal lagoons is strongly influenced by the physical and chemical environment, such as the fluxes caused by currents and tides, the import of marine sediment and water, and terrestrial inputs by rivers and groundwater. The sporadic mixing of fresh and marine waters results in a high physico-chemical variability both in space and time [50]. The relatively closed lagoon ecosystems are strongly exposed to nutrient input from the surrounding area, including anthropogenic pollution [9]. Coastal lagoons are usually surrounded by extensive areas of freshwater and salt marshes which are important sources of dissolved organic matter (DOM) [50]. The Venice lagoon is a shallow water basin located on the coast of the Northern Adriatic Sea, just north of the River Po delta. The lagoon is connected to the sea by three 400–900 m wide and 15–50 m deep inlets which divide it into three hydrological basins [58] called the Northern Lagoon or Basin of Lido, the Central Lagoon or Basin of Malamocco, and the Southern Lagoon or Basin of Chioggia. The high water exchange with the sea, which mainly depends on a tidal regime, is one of the most important factors for the quality of waters, sediments and biota. This exchange with the sea also determines the dispersion and reduction of pollutant levels [65]. In the last decade, the Lagoon of Venice has been declared UNESCO patrimony, and as a consequence several studies have focused on the pollution in this ecosystem [7]. The current situation of the lagoon is alarming in view of its apparent degradation, with signs of anoxia in bottom layers in summer, a decline of zooplankton, and a build-up of sulphide in the upper layers of bottom sediments accompanied by a shift in bottom flora composition [59]. Several studies have proved the presence of a wide range of contaminants in its different basins [39] [10] but little is known about the microbial community. Only recent studies have focused on the seasonal and spatial dynamics of the

Abbreviations: CARD, Catalyzed reporter composition; FISH, Fluorescence *in situ* hybridization; FA, formamide.

* Corresponding author.
E-mail address: ramann@mpi-bremen.de (R. Amann).

0723-2020/$ - see front matter © 2010 Elsevier GmbH. All rights reserved.
doi:10.1016/j.syapm.2009.12.006

microbial community of the Lagoon of Venice, and they have highlighted that there is a strong seasonality of the planktonic bacterial assemblage [16], as well as spatial differences in community composition between sub-basins [12].

The aim of this study was to examine the bacterial diversity and community composition in surface waters of the southern basin of the Lagoon of Venice near the city of Chiogga. Here, we present data obtained by the rRNA approach to microbial ecology [6] at the Chioggia inlet and within the lagoon in two different seasons. Insights into bacterioplankton diversity were obtained by comparative sequence analysis of several 16S rRNA gene clone libraries. Furthermore, the quantitative bacterial community composition was investigated by catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) with rRNA-targeted oligonucleotide probes. We hypothesized that the high availability of inorganic and organic nutrients in the lagoon should result in high cell numbers, and shift the taxonomic composition of the bacterial community towards copiotrophic taxa. A particular focus of the study was on members of the class *Flavobacteria* of the phylum *Bacteroidetes* for which we hypothesized high diversity and abundance in this nutrient-rich coastal setting of the Basin of Chioggia based on recent reports [4].

Materials and methods

Study site and sampling procedure

Surface water samples were taken around 9 am near the city of Chioggia, Italy, on October 17, 2007 and June 10, 2008 at two stations (Fig. 1): (i) within the lagoon [45°13′24″ N, 12°17′5″E] (in front of the Hydrobiological Station of Chioggia), and (ii) in the Southern Chioggia inlet [45°13′58″N, 12°17′30″E]. Water depth at the inlet station was around 20 m, whereas the lagoon station only had a depth of 4 m. On October 17, 2007 the general direction of water flow at the inlet was into the lagoon, while on June 10, 2008 the flow was directed outwards. Water samples were collected in replicates directly in acid-rinsed plastic bottles, and immediately transported to the laboratory. For DNA extraction, 1 L aliquots were directly filtered onto 0.2 μm pore size polycarbonate membrane filters (GTTP, 47 mm diameter, Millipore, Eschborn, Germany). Filters were immediately frozen, and kept at −20 °C until processed for nucleic acid extraction.

For CARD-FISH, 10 mL aliquots were collected and fixed with paraformaldehyde (EM Grade, Hatfield, PA) for 1–24 h at 4 °C (1% final concentration). The fixed samples were subsequently filtered onto 0.2 μm pore size polycarbonate filters (GTTP, 47 mm diameter, Millipore) and stored at −20 °C.

Determination of chlorophyll a and b

For determination of chlorophyll concentrations, cells were collected on glass fibre filters (nominal pore size 0.7 μm; Whatman, Dassel, Germany). Samples were immediately frozen at −20 °C. Frozen samples were placed in 15 mL centrifuge tubes, 5 mL of acetone were added, and the tubes were sonicated on ice for 10 × 1 min. Pigments were extracted from the mixture at −20 °C, overnight. Samples were then filtered (Acrodisc CR syringe filter, 0.45 μm pore size, PALL, Gellman Laboratory; Dreieich, Germany), and the filtrate was diluted with water to a final concentration of 70% acetone. Photo-pigments were separated on an HPLC 2695 separation module (Waters, Eschborn, Germany), equipped with a Reprosil-Pur-120 C18, 5 μm 250 × 4.6 mm (Trentec Analysentechnik, Gerlingen, Germany), according to the method given in [63]. Absorption spectra were measured on a Waters 996 photo diode array detector, and the pigments were quantified and identified by comparison to pigment standards (DHI Waters and Environment, Horsholm, Denmark).

CARD-FISH and cell counting

CARD-FISH was performed according to Pernthaler et al. [51] with slight modifications. Briefly, cells were immobilized on GTTP filters by embedding in 0.2% agarose (Metaphor, Cambrex Bio Science, Rockland, USA), and they were subsequently permeabilized by treatment with 10 mg mL^{-1} lysozyme in 50 mM EDTA, 100 mM Tris/HCl for 1 h at 37 °C. Filter sections were cut and hybridized with horseradish peroxidase (HRP)-labelled oligonucleotide probes (Biomers, Ulm, Germany) for 2.5 h at 46 °C. Probes targeted members of the *Bacteria* (EUB338 I-III, 35% FA, [24]), the alphaproteobacterial SAR11 clade (SAR11-441, 25% FA, [45]) and the *Roseobacter* clade (Ros537, 35% FA, [26]), *Alphaproteobacteria* general probe (ALF968, 35% FA, [47]), the gammaproteobacteria affiliated to the family *Vibrionaceae* (GV, 30% FA, [30]), the genus *Pseudoalteromonas* (PSA184, 30% FA, [26]), the NOR5/OM60 clade (mix of probes NOR5-730, [27] and NOR5-1238, 50% FA, [64]), the *Alteromonas/Glaciecola* clade (ALT1413, 40% FA, [26]), *Alteromonas macleodii* (Amac83, 35% FA, [2]), the *Gammaproteobacteria* general probe (GAM42at 35% FA, with competitors according to Manz et al., [43]), the *Betaproteobacteria* (BET42a 35% FA, with competitors according to Manz et al., [43]), the *Cytophaga/Flavobacterium* cluster of *Bacteroidetes* (CF319a, 35% FA, [42]), and the *Planctomycetes* (PLA46, 30% FA, [48]). The hybridized and washed filter sections were subsequently embedded in a mounting mixture containing 1 μg mL^{-1} DAPI [51]. The fraction of DAPI-stained objects detected by a particular probe was determined by epifluorescence microscopy (AXIOPLAN 2, Zeiss, Jena, Germany). Controls with the antisense probe NON338 [60] were always negative. At least 1000 DAPI-stained cells were manually counted.

DNA extraction and PCR amplification

Total nucleic acids were extracted from the 0.2 μm pore size GTTP filters, as described previously [66]. Nearly complete 16S rRNA genes were amplified by PCR using the general bacterial primers GM3 and GM4 [46]. The 50 μL PCR reactions contained 0.3 mg mL^{-1} BSA, 2.5 mM total dNTPs, 1 × PCR buffer including 1.5 mM of Mg chloride, 0.5 mM of each primer, and 1 U of Master Taq Polymerase (Eppendorf, Hamburg, Germany). The amplification was carried out with a Thermocycler Mastercycler (Eppendorf) as follows: an initial denaturation step at 95 °C for 3 min, followed by 25 cycles of 1 min denaturation at 95 °C, 2 min annealing at 44 °C, and 2 min elongation at 72 °C. There was a final extension step of 15 min at 72 °C.

Clone library generation, sequencing and phylogenetic analysis

In order to avoid cloning of small and non-specific PCR products, fragments of the correct size of ∼1.5 kb were excised from 1.5% agarose gel, and subsequently extracted and purified using the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany). Cleaned PCR products were cloned either into the pGEM-T Easy vector (Promega, Madison, WI) or the pCR4 TOPO vector (Invitrogen, Groningen, the Netherlands), according to the manufacturers' instructions. The recombinant vectors were transformed into chemically competent *E. coli* cells, DH10B strain (Invitrogen, Groningen, the Netherlands), plated, and 192 clones from each library were picked. Clone inserts were amplified by PCR using vector primers M13F (5′-GTAAAACGACGGCCAG-3′) and

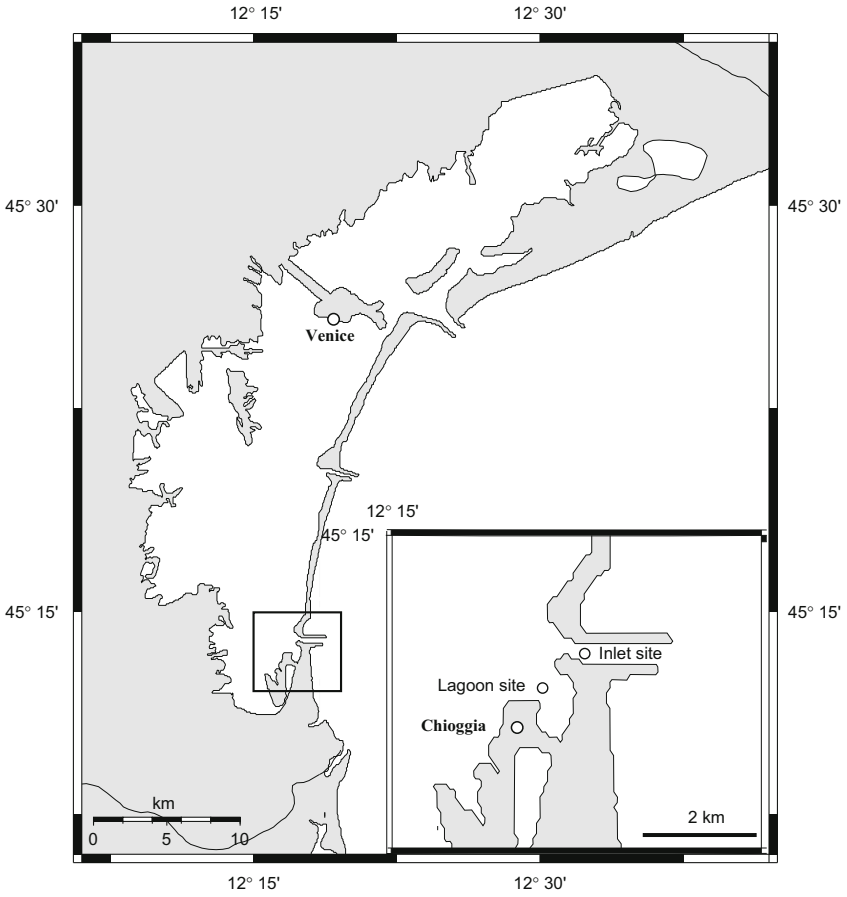


Fig. 1. Map of the Lagoon of Venice indicating the two sampling sites. Maps were created with the online map creation tool, OMC (www.aquarius.ifm-geomar.de).

M13R (5′-CAGGAAACAGCTATGAC-3′). All the PCR products were purified by using the MultiScreen-HV plate (Millipore).

A total of 645 partial sequences were obtained using the GM1F primer (5′-CCAGCAGCCGCGGTAAT-3′; modified from [46]) with an average length of 700 bp. Sequencing reactions were performed using the ABI BigDye® chemistry and an ABI 3100 genetic analyser (Applied Biosystems, Foster City, USA), according to the manufacturer's instructions. Partial sequences were manually corrected using the software Sequencher (Gene Codes, MI, USA).

All sequences were checked for chimera origin with the online tool Bellerophon from Greengenes [25,34] and possible chimera sequences were excluded from further analysis. Phylogenetic analyses were performed using the ARB software package [40] using Silva database release 93 [53]. All the 16 S rRNA partial gene sequences of this study were aligned in ARB and the alignment was subsequently improved manually considering the secondary structure of the rRNA molecule. Phylogenetic reconstruction was carried out with full-length sequences (> 1400 nt) of the closest relatives, using the maximum likelihood algorithm in ARB with 50% position variability filters. Partial sequences were added by parsimony without allowing changes in the overall tree topology. Sequences of this study were deposited at EMBL with the accession numbers FN434610-FN435254.

Operational taxonomic units (OTUs) and the richness estimator indexes (Chao1 [17], abundance based coverage (ACE) [19,20] and Jackknife [13]) were calculated with DOTUR [56] using the furthest neighbour algorithm from a distance matrix generated in ARB. Only sequences longer than 600 bp were considered for the analysis. From the DOTUR results, we identified all the sequences that contributed to each OTU, at a 97% threshold level, and this data was imported into EstimateS [22]. The similarity of the OTUs composition between the inlet and the lagoon was calculated according to Chao et al. [18] with the program EstimateS. Coverage index of the clone libraries was calculated with the formula $C=(1-n/N)\times 100$, where N represents the number of all sequences (longer than 600 bp) and n the number of phylotypes at 97% identity appearing only once in the library [33,35].

Results and discussion

Water characteristics within the Lagoon of Venice were quite variable (http://chioggia.scienze.unipd.it/ITA/ParametriLaguna.html).

In this study, we examined two adjacent sites within the Southern Basin and at the Chioggia inlet in autumn 2007 and summer 2008 (Table 1). A pilot study in June 2007 on a surface water sample from Chioggia with a rather high salinity of 36.9 PSU had revealed a conspicuous bloom of CF319a-positive cells likely affiliated with the *Cytophaga/Flavobacteria* cluster of *Bacteroidetes*. CF319a counts were between 1.0×10^7 and 2.4×10^6 cells mL^{-1}. These flavobacterial abundances were one to two orders of magnitude higher than in other marine surface waters. Moreover, 10% of CF319a-positive cells were filamentous (Fig. 2a and b) indicating that the bacterioplankton was probably subject to protistan predation [52]. However, this phenomenon did not occur on the two sampling dates for which we undertook a more detailed investigation.

The salinity of the June 2008 water was lower at 31–32 PSU than in the October 2007 samples (34.4 and 35.2 PSU at the lagoon and inlet station, respectively). In the lagoon sample of June 2008, concentrations of chlorophyll *a* and *b* were significantly higher. This could be attributed to a bloom of unicellular algae - mostly diatoms (data provided by the Hydrobiological Station of Chioggia) - that might have been caused by high run-off of nutrients into the lagoon from surrounding agricultural fields because of the frequent rainfall. Total bacterioplankton abundances, as determined by DAPI counting, were stable around 3×10^6 cells mL^{-1} in all samples, and they were lower than in June 2007 but still permanently elevated compared to bacterioplankton abundances in the Adriatic Sea. The nutrient measurements revealed a permanent eutrophication in the lagoon at both sites, especially during June 2008 when concentrations of phosphate, nitrite/nitrate and chlorophyll *a* were a factor of 3–5 times higher than in October 2007 (Table 1). The oxygen concentration was between 5.6 and 5.9 mL L^{-1}, except for the June 2008 lagoon station that was slightly suboxic (3.9 mL L^{-1}), which also pointed to eutrophic conditions.

Table 1
Physico-chemical characteristics of surface water samples.

	October 17, 2007		June 10, 2008	
	Lagoon	Inlet	Lagoon	Inlet
Temperature (°C)	17.5	17	23.7	23.4
pH	8.2	8.2	8.4	8.3
Salinity (PSU)	34.4	35.2	31.7	30.5
Chlorophyll *a* (μg L^{-1})	0.4	0.7	2.1	0.8
Chlorophyll *b* (μg L^{-1})	0.1	0.1	0.5	0.1
NH_4 (μmol L^{-1})	2.8	2.3	0.7	0.9
PO_4 (μmol L^{-1})	0.7	0.5	2	1.9
NO_2+NO_3 (μmol L^{-1})	4.9	5.4	28.2	26.1
O_2 (mL L^{-1})	5.6	5.9	3.9	5.8
DAPI-stained cells (cells mL^{-1})	1×10^6	3.2×10^6	3.3×10^6	3.4×10^6

Overall bacterioplankton diversity and composition

With a general bacterial primer set, four 16S rDNA libraries were constructed from water samples taken at the Chioggia inlet and in the lagoon in October 2007 and in June 2008 after a period with heavy rain. We partially sequenced the 16S rRNA region between 500 and 1200 bp of 192 clones from each library.

Overall, the lagoon samples had higher phylotype richness. All the libraries had a similar number of OTUs at a 16S rRNA similarity level of 97%, however, the number of sequences detected only once (singletons) was higher at the lagoon site (Table 2). The richness indexes predicted a higher total number of OTUs at the lagoon site, with 128 (95% confidence interval CI 89–219) for Chao1, 139 (95% CI 101–176) for Jackknife, and 158 (95% CI 108–260) for ACE indexes, whereas at the inlet station the predicted richness was 73 (95% CI 62–103) for Chao1, 77 (95% CI 65–89) for Jackknife and 79 (95% CI 67–110) for ACE. In particular, richness indexes were higher from the lagoon site in June 2008 (S1), which might indicate that a higher number of phylotypes can

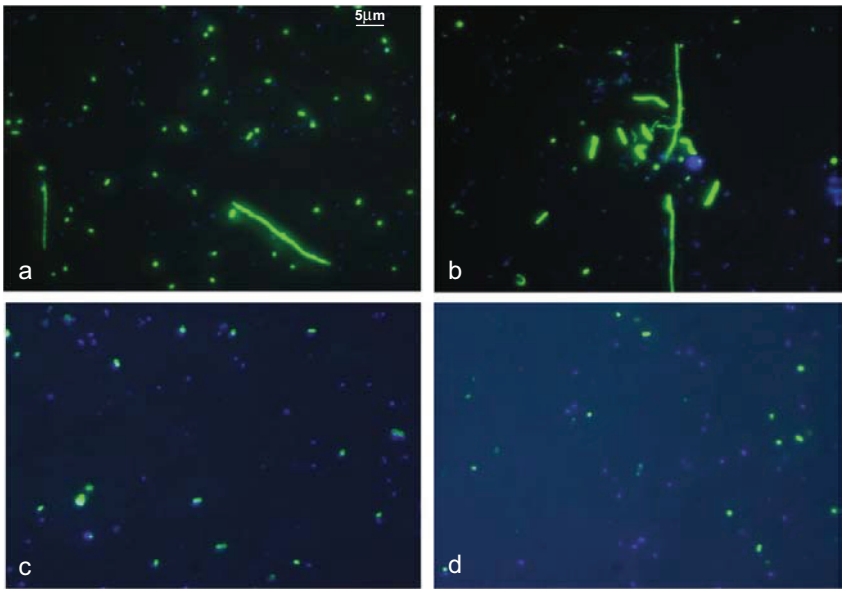


Fig. 2. Epifluorescence micrographs of bacteria in water samples from the Southern Venice lagoon. Blue, DAPI signals of DNA containing cells, and green signals of CARD-FISH positive cells with the probe CF319a. (a) surface water from the lagoon station, June 2007; (b) inlet, June 2007; (c) lagoon October 2007; (d) inlet, October 2007. The bar indicates 5 μm. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

Table 2
Affiliation of 16S rRNA gene sequences.

Bacterial group	October 17, 2007		June 10, 2008	
	Lagoon	Inlet	Lagoon	Inlet
Alphaproteobacteria	10 (6%)	55 (33%)	31 (17%)	40 (27%)
Betaproteobacteria	1 (0.5%)	0%	14 (7.6%)	16 (11%)
Gammaproteobacteria	77 (49%)	19 (13%)	92 (50%)	15 (10%)
Bacteroidetes	61 (38%)	62 (37%)	36 (20%)	67 (45%)
Firmicutes	1 (0.5%)	0%	2 (1.1%)	0%
Actinobacteria	10 (6%)	19 (1.2%)	2 (1.1%)	0%
No. of analyzed clones	159	168	179	149
No. of sequences >600 bp	143	161	122	161
No. of OTUs (97% similarity)	36	41	49	30
No. of singletons	20	16	30	9
Coverage	86	90.1	75.4	92.7

Table 3
Quantification of bacterioplankton groups by CARD-FISH (% of DAPI).

Clade	October 17, 2007		June 10, 2008	
	Lagoon Average $\pm$ SD	Inlet Average $\pm$ SD	Lagoon Average $\pm$ SD	Inlet Average $\pm$ SD
Bacteria	92 ± 0.1	87 ± 0.1	86 ± 2.3	79 ± 1.5
Alphaproteobacteria	19 ± 7.7	34 ± 1.5	48 ± 4.7	48 ± 2.5
Roseobacter	6 ± 0.3	5 ± 0.1	4 ± 0.0	6 ± 0.1
SAR-11	26 ± 0.2	30 ± 0.4	37 ± 5.0	28 ± 0.5
Betaproteobacteria	5 ± 0.4	3 ± 0.3	4 ± 0.2	4 ± 1.2
Gammaproteobacteria	22 ± 2.5	20 ± 0.9	19 ± 0.6	22 ± 1.6
Pseudoalteromonas	1 ± 0.1	0.4 ± 0	1 ± 0.2	0.2 ± 0.1
Vibrionaceae	2 ± 0.1	1 ± 0.1	1 ± 0.1	0.3 ± 0.1
Alteromonas macleodii	1 ± 0	3 ± 0.2	2 ± 0.1	2 ± 0.6
Alteromonas/Glaciecola	5 ± 0.1	3 ± 0.1	5 ± 0.5	2 ± 0.2
NOR5/OM60	1 ± 0.1	2 ± 0.4	1 ± 0.2	2 ± 0.4
Bacteroidetes	55 ± 1.0	44 ± 0.8	34 ± 2.8	34 ± 1.2
Planctomycetes	1 ± 0.1	1 ± 0.1	0.4 ± 0.1	0.3 ± 0
DAPI-stained cells (cells mL^{-1})	1×10^6	3.2×10^6	3.3×10^6	3.4×10^6

thrive under more productive conditions (Table 1). At the 97% similarity level, the phylotypes retrieved from the lagoon and the inlet had little overlap. Twenty-eight OTUs were shared between both sites, which represented a similarity of 0.27, 0.29 or 0.56 for Morisita-Horn, Jaccard and Chao's Jaccard indexes, respectively. The similarity values indicated that despite the mixing of the water masses between the lagoon and the inlet, the two sites harboured unique phylotypes which were not shared.

Overall, the four libraries were dominated by three bacterial groups (Table 2): *Bacteroidetes*, *Alphaproteobacteria*, and *Gammaproteobacteria*. These three groups ought to be expected in marine surface water [31,55], whereas some of the actinobacterial sequences were affiliated to sequences retrieved from lakes and therefore indicated freshwater input.

On average, $85 \pm 7\%$ of the DAPI-stained cells were identified as bacteria with the general bacterial probe mixture EUB338 I-III. CARD-FISH counts with group-specific probes confirmed the dominance of *Bacteroidetes* (CF319a), *Alphaproteobacteria* (ALF968), and *Gammaproteobacteria* (GAM42a) (Table 3). CARD-FISH showed that *Betaproteobacteria* and *Planctomycetes* were minor components of the bacterioplankton in the Lagoon of Venice with relative abundances of 3–5% and 0.3–1%, respectively. None of the sequences retrieved in our 16S rRNA gene libraries were related to *Planctomycetes*.

Bacteroidetes

With relative abundances of 54% in the lagoon and 44% at the inlet, *Bacteroidetes* was the most abundant phylum in October 2007, as detected by CARD-FISH (Table 3). The absolute cell numbers counted with probe CF319a ranged between 1.0 and 1.6×10^6 cells mL^{-1}. The specificity of this probe has recently been evaluated [5]. It covers 90% of the 16S rRNA sequences of the classes *Flavobacteria* and *Sphingobacteria*, but also detects some members of class *Bacteroidia* and some sequences outside this phylum. In the June 2008 samples, probe CF319a detected about one third of the bacterioplankton both at the inlet and at the lagoon stations. Members of *Bacteroidetes* were at that time the second most abundant group after the *Alphaproteobacteria*. High abundances of CF319a-positive cells have already been shown in other coastal marine sites, and have been linked to algal blooms and higher substrate concentrations [27,36,37]. However, in the Lagoon of Venice, the highest *Bacteroidetes* abundance (54% in the Lagoon, October 2007) did not correspond to the sampling point with higher nutrients and chlorophyll *a* concentrations (Lagoon, June 2008), indicating that other factors might be controlling its abundance.

In support of this data, comparative sequence analyses also revealed many phylotypes affiliated with *Bacteroidetes* (Fig. 3). The frequencies of clones affiliated with *Bacteroidetes* were lower than the CARD-FISH count, confirming that this group tends to be underrepresented in 16S rRNA gene clone libraries [23].

Most of the sequences were affiliated with the class *Flavobacteria*, which represented 96% of the total *Bacteroidetes* sequences, providing more evidence that *Flavobacteria* is the dominant class of *Bacteroidetes* in marine bacterioplankton [4,21,37]. Highly similar sequence clusters with identities $>98.5\%$ were retrieved from all four libraries, suggesting that

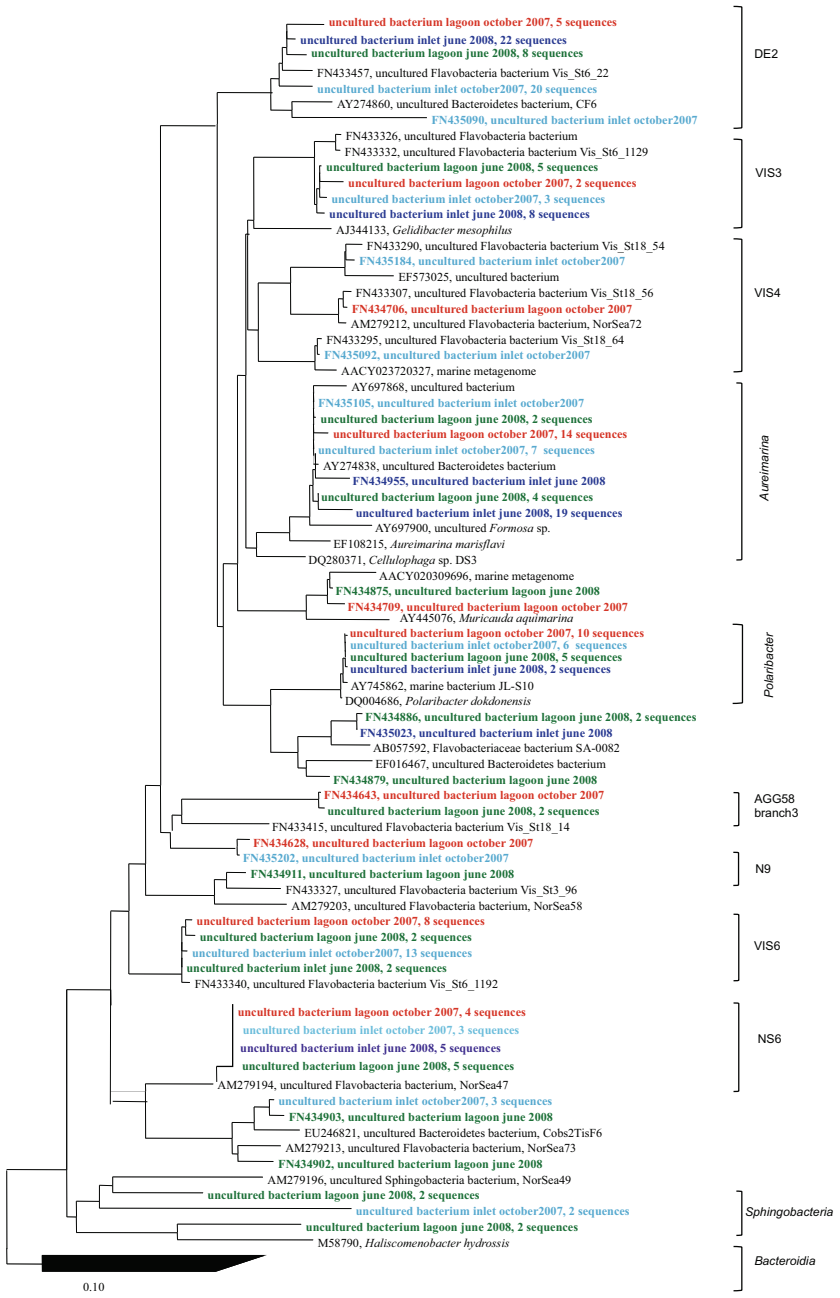


Fig. 3. Phylogenetic tree reconstructed based on 16S rRNA of members of the phylum *Bacteroidetes* from the Southern Venice lagoon with selected reference sequences. The scale bar represents 10% estimated sequence divergence. Red: lagoon sequences retrieved in October 2007; turquoise: inlet sequences of October 2007; green: lagoon sequences of June 2008; blue: inlet sequences of June 2008. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

these clades were permanently present in the southern lagoon and at the inlet (Fig. 3). This included clusters related to the cultured genera *Polaribacter* and *Aureimarina*, which represented the most abundant phylotype in the inlet, as well as clusters for which so far no cultured representatives have been obtained. In our study, a relatively high number of the *Bacteroidetes* sequences

were affiliated with cultured genera, 10% with *Polaribacter* and 20% with *Aureimarina*, which is in contrast to previous findings [4]. However, there is also an evidence that *Polaribacter* is a widespread and abundant flavobacterial genus in different marine environments [32,41]. Overall, it might indicate that the Venice lagoon is enriched by readily cultured flavobacterial clades.

Twenty four percent of the sequences were affiliated with the DE2 clade [37] were retrieved from both sampling sites (Fig. 3). DE2 comprises sequences that have mostly been retrieved from coastal systems, such as the Delaware Estuary [37], the North Sea [4,27] and the coast around Plymouth [49]. Its high clone frequency here also shows the relevance of this uncultured clade in coastal systems. Thirteen sequences from the inlet and eight from the lagoon were affiliated with the clades VIS4 and VIS3, which have been defined from sequences retrieved in the North Atlantic Ocean [32] and they also comprise sequences retrieved from the North Sea [4]. Moreover, we obtained three sequences affiliated with one of the branches of the uncultured clade AGG58 [49] and 25 with the clade VIS6. VIS6 also comprises phylotypes that have been retrieved from the North Atlantic Ocean [32] as well as during a phytoplankton bloom in naturally iron-fertilized waters [62].

The comparative sequence analysis of the four coastal marine libraries indicates a high local and seasonal diversity of *Flavobacteria*. The Lagoon of Venice would certainly be a rich source of new flavobacterial species since members of the class are often readily culturable.

Gammaproteobacteria

One half of all sequences retrieved from the two lagoon samples were affiliated with *Gammaproteobacteria* (Table 2), whereas in both inlet libraries this group accounted for only 10% and 13%. The sequences mostly clustered with those of the readily culturable marine coastal bacteria of the family *Vibrionaceae*, the genus *Pseudoalteromonas*, and the *Alteromonas*/*Glaciecola* group (Fig. 4). The most abundant OTUs (102 sequences) belonged to the genus *Vibrio*, which encompassed only 6 sequences from the inlet samples. Approximately half of the gammaproteobacterial OTUs were present in the 2007 and the 2008 samples, suggesting that these phylotypes are permanently present in the lagoon in significant numbers. Additionally, no major differences in the clone frequencies (Table 2) and in the CARD-FISH counts (Table 3)

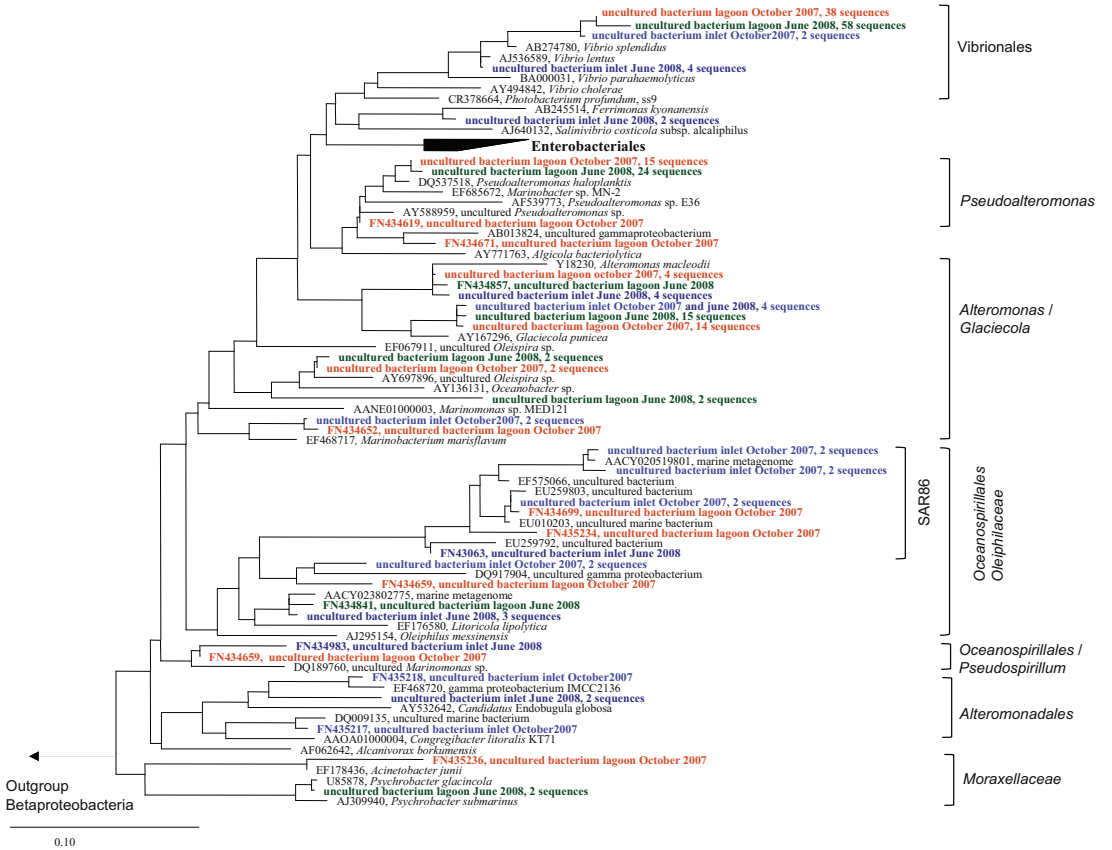


Fig. 4. Phylogenetic tree reconstructed based on 16S rRNA of members of the phylum *Gammaproteobacteria* from the Southern Venice lagoon with selected reference sequences. The scale bar represents 10% estimated sequence divergence. Red: lagoon sequences retrieved in October 2007; turquoise: inlet sequences of October 2007; green: lagoon sequences of June 2008; blue: inlet sequences of June 2008. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

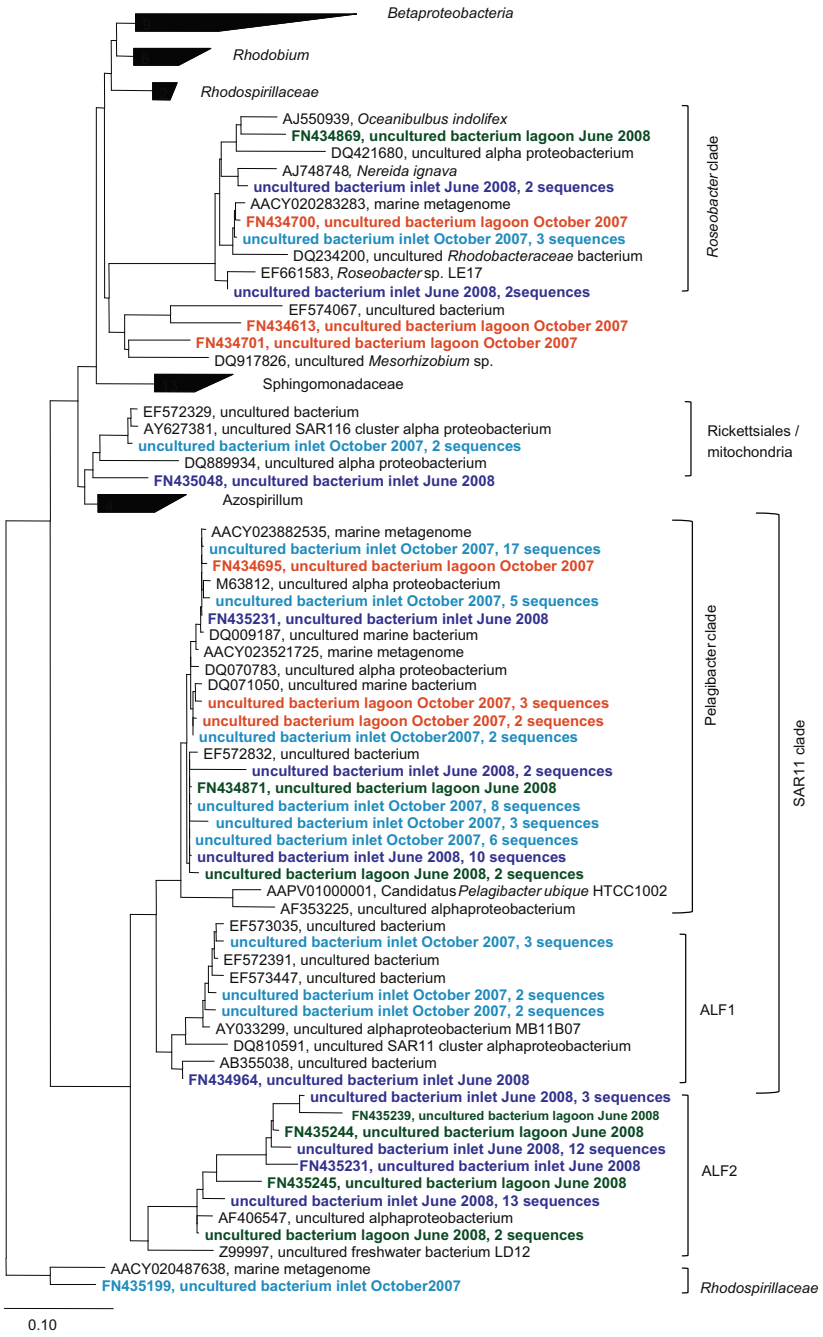


Fig. 5. Phylogenetic tree reconstructed based on 16S rRNA of members of the phylum *Alphaproteobacteria* from the Southern Venice lagoon with selected reference sequences. The scale bar represents 10% estimated sequence divergence. Red: lagoon sequences retrieved in October 2007; turquoise: inlet sequences of October 2007; green: lagoon sequences of June 2008; blue: inlet sequences of June 2008. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.

were detected between sampling times from either the lagoon or the inlet, suggesting that the Lagoon of Venice harbours a relatively stable gammaproteobacterial community.

The discrepancy between the rather high frequency of gammaproteobacterial clones, (i.e., 50% in the lagoon samples; Table 2) and the much lower CARD-FISH counts (around 20% in the lagoon; Table 3) can best be explained by the presence of *Gammaproteobacteria* with genomes containing significantly more rRNA operons than average marine bacteria. It is well documented that members of the genus *Vibrio* have between 8 and 13 copies of 16S rRNA [38]. In a comparative analysis of several molecular methods, *Gammaproteobacteria* clone frequencies and CARD-FISH counts mostly agreed [3], however, in contrast to our study, *Vibrio* sequences were not retrieved.

Sequences related to *Pseudoalteromonas* spp. were only retrieved from the lagoon sites, and not from the inlet (Fig. 4). CARD-FISH indicated that members of the genus *Pseudoalteromonas* were also present at the inlet station although in lower numbers (0.2–0.4%) than in the lagoon (0.9–1.3%) (Table 3). Also, the *Vibrionaceae* and members of the *Alteromonas/Glaciecola* clade were 2–3 times more abundant in the lagoon than in the inlet. CARD-FISH indicated that the *Alteromonas macleodii* probe Amac83 explained all of the *Alteromonas* counts in the two inlet samples, but only about one third of the counts in the lagoon. This indicates a shift from more oligotrophic *Alteromonas* spp., such as *A. macleodii* [2], to more copiotrophic species. Members of the marine NOR5/OM60 clade were detected in all samples by CARD-FISH in abundance typical of coastal settings [64].

The class *Gammaproteobacteria* is known to encompass many copiotrophic clades which dwell under nutrient-rich conditions [15]. In this study, we could clearly demonstrate that some of these copiotrophs were strongly enriched within the lagoon. A high proportion of the gammaprotebacteria sequences from the lagoon samples belonged to the *Vibrionaceae* family, whose presence is related to urban polluted marine areas [28]. Moreover, a recent study in the Lagoon of Venice also highlighted the dominance of gammaptroteobacteria, and particularly of *Vibrio*, in sediment samples from the lagoon areas close to urban centres [12]. The high abundance of not yet identified members of the genus *Vibrio* in the Lagoon of Venice is even more alarming since it has recently been shown that *Vibrio cholerae*, *V. vulnificus* and *V. parahaemolyticus* could acquire virulence genes by the uptake of extracellular DNA in the presence of the abundant polysaccharide chitin [8,11]. Consequently, it would be important to focus further studies on the *Gammaproteobacteria* of the Lagoon of Venice.

Alphaproteobacteria

In contrast to the *Gammaproteobacteria*, the frequencies of sequences affiliated with *Alphaproteobacteria* were higher in the inlet libraries (33% in 2007, 27% in 2008) than in the lagoon libraries (6.3% in 2007, 17% in 2008). Most of the alphaproteobacterial sequences in all four libraries were closely related to those of two typical marine clades, SAR11 and *Roseobacter* (Fig. 5). In the latter clade, three sequences (FN434700, FN435180, FN435182) were $>96\%$ identical to the 16S rRNA sequence of *Donghicola eburneus* isolated from coastal seawater of Korea. One sequence (FN435014) retrieved from the inlet station in June 2008 was $>98\%$ identical to that of another strictly aerobic chemoheterotrophic species, *Nereida ignava*, recently isolated by Pujalte et al. [54] from coastal Mediterranean seawater collected near Valencia. CARD-FISH indicated that the relative abundances of these two clades were rather stable in all samples investigated, with SAR11 accounting for 26–37% and *Roseobacter* for 4–6%.

The most abundant alphaproteobacterial OTU, encompassing 50 sequences from the inlet (mostly from the October 2007 library) and eight sequences from lagoon libraries, belonged to the SAR11 clade and it clustered with "*Candidatus* Pelagibacter ubique" (98% average identity). A particular cluster named ALF1 (Fig. 5), with four phylotypes belonging to the SAR11 clade, was present only in the inlet in 2007 and 2008. The ALF1 sequences had an identity of 88–93% with "*Candidatus* P. ubique". Our sequencing data showed that the ALF1 cluster and the Pelagibacter clade had a different spatial distribution in the Lagoon of Venice. While sequences affiliated with "*Candidatus* P. ubique" were retrieved from the inlet and lagoon sites, the ALF1 cluster was exclusively retrieved from the inlet samples, pointing to a different physiological adaptation of these clades.

Another conspicious cluster of alphaproteobacterial sequences (ALF2) branching even more deeply from SAR11 (86–87% identity with 16S rRNA of "*Candidatus* Pelagibacter ubique") was retrieved only in the June 2008 samples from both stations (closely related to sequence AF406547 retrieved from marine bacterioplankton) (Fig. 5). We determined almost full-length 16S rRNA sequences of these ALF2 clones demonstrating that they were not covered by probe SAR11-441 nor by Ros537.

In the June 2008 sample, counts with probes Ros537 and SAR11-441 together could not explain all counts obtained with probe ALF968, and this might reflect the presence of the ALF2 clades which were not targeted by these probes. In contrast, in the October 2007 samples from which ALF2 was not retrieved, counts with probes Ros537 and SAR11-441 together were higher than counts for ALF968. As shown recently [5], this general alphaproteobacterial probe covers only 81% of the *Alphaproteobacteria*. Therefore, we suggest that in future studies of the bacterioplankton composition of nutrient-rich lagoons additonal probes specific for alphaproteobacterial clades should be applied.

Overall, we could assign the majority of *Alphaproteobacteria* to the SAR11 clade, which is typical of oligotrophic open ocean waters [45]. Its high abundance in all our samples is indicative of a high water exchange between the Lagoon of Venice and the Adriatic Sea (Table 3).

Other bacterial groups

Thirty one *Betaproteobacteria* sequences were retrieved in this study (Table 2), two of which (e.g. FN435040) were affiliated with limnic strains, such as *Limnobacter* sp. MED105. Two sequences (FN434621 and FN435039) were more than $>97\%$ identical to the methylotrophic coastal marine isolate HTCC2181 [29], which belongs to the OM43 clade. This *Betaproteobacteria* clade is commonly found in productive marine coastal ecosystems [57] and is linked to phytoplankton blooms [44].

Additionally, thirty one sequences belonged to *Actinobacteria*. Twenty-eight of them belonged to the genus *Cryobacterium*, most of them (18) were between 96% and 97% identical to sequences retrieved from freshwater systems, for example, with the cluster AcIIA from Lake Saelenvannet, Norway [61]. The actinobacterial sequences could be an indicator of the freshwater input into the Lagoon of Venice.

Conclusions

Our data of the first 16S rRNA-based studies on the bacterioplankton in the southern basin of the Lagoon of Venice provide insights into a highly complex marine coastal site. Comparative sequence analyses and CARD-FISH experiments showed that most representative microbial groups were *Bacteroidetes*, *Gammaproteobacteria* and *Alphaproteobacteria*, which are typical for marine settings. We could confirm our hypotheses of high abundance of *Flavobacteria* and copiotrophic clades of

Gammaproteobacteria in this nutrient-rich aquatic habitat. The 16S rRNA sequence data could be the basis for the future design of specific oligonucleotide probes for closer monitoring of bacterioplankton populations in the context of, for instance, algal blooms, oxygen content, temperature, precipitation and hydrology, as well as anthropogenic pollution.

Acknowledgements

We acknowledge Vito Sabia, Maria Berica Rasotto, and Giorgio Valle for their generous support, as well as the staff of the Hydrobiological Station Chioggia, and Ilaria Pizzetti from the Max Planck Institute for Marine Microbiology for assistance during sampling. We thank Andrew Bissett for helping with chlorophyll data measurements. This study was funded by the Max Planck Society.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.syapm.2009.12.006.

References

[1] Abreu, P.C., Odebrecht, C., Gonzalez, A., 1994. Particulate and dissolved phytoplankton production of the Patos Lagoon estuary, Southern Brazil – comparison of methods and influencing factors. J. Plankton Res. 16, 737–753.
[2] Allers, E., Niesner, C., Wild, C., Pernthaler, J., 2008. Microbes enriched in seawater after addition of coral mucus. Appl. Environ. Microbiol. 74, 3274–3278.
[3] Alonso-Saez, L., Aristegui, J., Pinhassi, J., Gomez-Consarnau, L., Gonzalez, J.M., Vaque, D., Agusti, S., Gasol, J.M., 2007. Bacterial assemblage structure and carbon metabolism along a productivity gradient in the NE Atlantic Ocean. Aquat. Microb. Ecol. 46, 43–53.
[4] Alonso, C., Warnecke, F., Amann, R., Pernthaler, J., 2007. High local and global diversity of *Flavobacteria* in marine plankton. Environ. Microbiol. 9, 1253–1266.
[5] Amann, R., Fuchs, B.M., 2008. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. Nat. Rev. Microbiol. 6, 339–348.
[6] Amann, R.I., Ludwig, W., Schleifer, K.H., 1995. Phylogenetic identification and *in situ* detection of individual microbial-cells without cultivation. Microbiol. Rev. 59, 143–169.
[7] Ammerman, A.J., McClennen, C.E., 2000. Saving Venice. Science 289, 1301–1302.
[8] Bartlett, D.H., Azam, F., 2005. Chitin, cholera, and competence. Science 310, 1775–1777.
[9] Benlloch, S., RodriguezValera, F., MartinezMurcia, A.J., 1995. Bacterial diversity in two coastal lagoons deduced from 16S rDNA PCR amplification and partial sequencing. FEMS Microbiol. Ecol. 18, 267–279.
[10] Berto, D., Giani, M., Boscolo, R., Covelli, S., Giovanardi, O., Massironi, M., Grassia, L., 2007. Organotins (TBT and DBT) in water, sediments, and gastropods of the southern Venice lagoon (Italy). Mar. Pollut. Bull. 55, 425–435.
[11] Blokesch, M., Schoolnik, G.K., 2008. The extracellular nuclease Dns and its role in natural transformation of *Vibrio cholerae*. J. Bacteriol. 190, 7232–7240.
[12] Borin, S., Brusetti, L., Daffonchio, D., Delaney, E., Baldi, F., 2009. Biodiversity of prokaryotic communities in sediments of different sub-basins of the Venice lagoon. Res. Microbiol. 160, 307–314.
[13] Borneman, J., Triplett, E.W., 1997. Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. Appl. Environ. Microbiol. 63, 2647–2653.
[14] Castel, J., Caumette, P., Herbert, R., 1996. Eutrophication gradients in coastal lagoons as exemplified by the Bassin d'Arcachon and the Etang du Prevost. Hydrobiologia 329, R9–R28.
[15] Cavicchioli, R., Ostrowski, M., Fegatella, F., Goodchild, A., Guixa-Boixereu, N., 2003. Life under nutrient limitation in oligotrophic marine environments: an eco/physiological perspective of *Sphingopyxis alaskensis* (formerly *Sphingomonas alaskensis*). Microb. Ecol. 45, 203–217.
[16] Celussi, M., Pugnetti, A., Del Negro, P., 2009. Structural dynamics of bacterioplankton assemblages in the Lagoon of Venice. Estuarine Coastal Shelf Sci. 84, 154–160.
[17] Chao, A., 1984. Non-parametric estimation of the number of classes in a population. Scand. J. Stat. 11, 265–270.
[18] Chao, A., Chazdon, R.L., Colwell, R.K., Shen, T.J., 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. Ecol. Lett. 8, 148–159.
[19] Chao, A., Lee, S.M., 1992. Estimating the number of classes via sample coverage. J. Am. Stat. Assoc. 87, 210–217.
[20] Chao, A., Ma, M.C., Yang, M.C.K., 1993. Stopping rules and estimation for recapture debugging with unequal failure rates. Biometrika 80, 193–201.
[21] Chen, X.H., Zeng, Y., Jiao, N., 2008. Characterization of *Cytophaga-Flavobacteria* community structure in the Bering Sea by cluster-specific 16S rRNA gene amplification analysis. J. Microbiol. Biotechnol. 18, 194–198.
[22] Colwell, R.K., 2005 Statistical Estimation of Species Richness and Shared Species from Samples. User's Guide and Application. 7.5 ed.
[23] Cottrell, M.T., Kirchman, D.L., 2000. Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence *in situ* hybridization. Appl. Environ. Microbiol. 66, 5116–5122.
[24] Daims, H., Bruhl, A., Amann, R., Schleifer, K.H., Wagner, M., 1999. The domain-specific probe EUB338 is insufficient for the detection of all *Bacteria*: development and evaluation of a more comprehensive probe set. Syst. Appl. Microbiol. 22, 434–444.
[25] DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072.
[26] Eilers, H., Pernthaler, J., Glockner, F.O., Amann, R., 2000. Culturability and *in situ* abundance of pelagic bacteria from the North Sea. Appl. Environ. Microbiol. 66, 3044–3051.
[27] Eilers, H., Pernthaler, J., Peplies, J., Glockner, F.O., Gerdts, G., Amann, R., 2001. Isolation of novel pelagic bacteria from the German Bight and their seasonal contributions to surface picoplankton. Appl. Environ. Microbiol. 67, 5134–5142.
[28] Garay, E., Arnau, A., Amaro, C., 1985. Incidence of *Vibrio cholerae* and related vibrios in a coastal lagoon and seawater influenced by lake discharges along an annual cycle. Appl. Environ. Microbiol. 50, 426–430.
[29] Giovannoni, S.J., Hayakawa, D.H., Tripp, H.J., Stingl, U., Givan, S.A., Cho, J.C., Oh, H.M., Kitner, J.B., Vergin, K.L., Rappe, M.S., 2008. The small genome of an abundant coastal ocean methylotroph. Environ. Microbiol. 10, 1771–1782.
[30] Giuliano, L., De Domenico, M., De Domenico, E., Hofle, M.G., Yakimov, M.M., 1999. Identification of culturable oligotrophic bacteria within naturally occurring bacterioplankton communities of the Ligurian Sea by 16S rRNA sequencing and probing. Microbial. Ecol. 37, 77–85.
[31] Glockner, F.O., Fuchs, B.M., Amann, R., 1999. Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence *in situ* hybridization. Appl. Environ. Microbiol. 65, 3721–3726.
[32] Gómez-Pereira, P.R., Fuchs, B.M., Alonso, C., Oliver, M.J., van Beusekom, J.E.E., Amann, R., 2010. Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean, ISME J., in press, doi:10.1038/ismej.2009.142.
[33] Good, I.J., 1953. The population frequencies of species and the estimation of population parameters. Biometrika 40, 237–264.
[34] Huber, T., Faulkner, G., Hugenholtz, P., 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20, 2317–2319.
[35] Kemp, P.F., Aller, J.Y., 2004. Bacterial diversity in aquatic and other environments: what 16S rDNA libraries can tell us. FEMS Microbiol. Ecol. 47, 161–177.
[36] Kirchman, D.L., 2002. The ecology of *Cytophaga-Flavobacteria* in aquatic environments. FEMS Microbiol. Ecol. 39, 91–100.
[37] Kirchman, D.L., Yu, L.Y., Cottrell, M.T., 2003. Diversity and abundance of uncultured *Cytophaga*-like bacteria in the Delaware Estuary. Appl. Environ. Microbiol. 69, 6587–6596.
[38] Lee, N., Nielsen, P.H., Andreasen, K.H., Juretschko, S., Nielsen, J.L., Schleifer, K.H., Wagner, M., 1999. Combination of fluorescent *in situ* hybridization and microautoradiography – a new tool for structure-function analyses in microbial ecology. Appl. Environ. Microbiol. 65, 1289–1297.
[39] Losso, C., Volpi Ghirardinia, A., 2010. Overview of ecotoxicological studies performed in the Venice Lagoon (Italy). Environ. Int. 36, 92–121.
[40] Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., Schleifer, K.H., 2004. ARB: a software environment for sequence data. Nucleic Acids Res. 32, 1363–1371.
[41] Malmstrom, R.R., Straza, T.R.A., Cottrell, M.T., Kirchman, D.L., 2007. Diversity, abundance, and biomass production of bacterial groups in the western Arctic Ocean. Aquat. Microb. Ecol. 47, 45–55.
[42] Manz, W., Amann, R., Ludwig, W., Vancanneyt, M., Schleifer, K.H., 1996. Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum *Cytophaga-Flavobacter-Bacteroides* in the natural environment. Microbiol.-UK 142, 1097–1106.
[43] Manz, W., Amann, R., Ludwig, W., Wagner, M., Schleifer, K.H., 1992. Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria – problems and solutions. Syst. Appl. Microbiol. 15, 593–600.

[44] Morris, R.M., Longnecker, K., Giovannoni, S.J., 2006. *Pirellula* and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. Environ. Microbiol. 8, 1361–1370.
[45] Morris, R.M., Rappe, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., Giovannoni, S.J., 2002. SAR11 clade dominates ocean surface bacterioplankton communities. Nature 420, 806–810.
[46] Muyzer, G., Dewaal, E.C., Uitterlinden, A.G., 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S ribosomal-RNA. Appl. Environ. Microbiol. 59, 695–700.
[47] Neef, A., Anwendung der in situ Einzelzell-Identifizierung von Bakterien zur Populationsanalyse in komplexen mikrobiellen Biozönosen. Technische Universität München, München, 1997.
[48] Neef, A., Amann, R., Schlesner, H., Schleifer, K.H., 1998. Monitoring a widespread bacterial group: *in situ* detection of *Planctomycetes* with 16S rRNA-targeted probes. Microbiol.-UK 144, 3257–3266.
[49] O'Sullivan, L.A., Fuller, K.E., Thomas, E.M., Turley, C.M., Fry, J.C., Weightman, A.J., 2004. Distribution and culturability of the uncultivated 'AGG58 cluster' of *Bacteroidetes* phylum in aquatic environments. FEMS Microbiol. Ecol. 47, 359–370.
[50] Oliveira, A.M., Kjerfve, B., 1993. Environmental responses of a tropical coastal lagoon system to hydrological variability – Mundau-Manguaba, Brazil. Estuarine Coastal Shelf Sci. 37, 575–591.
[51] Pernthaler, A., Pernthaler, J., Amann, R., 2004. Sensitive multicolor fluorescence in situ hybridization for the identification of environmental microorganisms. In: Kowalchuk, G., de Bruijn, F.J., Head, I.M., Akkermans, A.D.L., van Elsas, J.D. (Eds.), Molecular Microbial Ecology Manual, second ed. Kluwer Academic Publishers, Dordrecht, the Netherlands. pp. 711–726.
[52] Pernthaler, J., 2005. Predation on prokaryotes in the water column and its ecological implications. Nat. Rev. Microbiol. 3, 537–546.
[53] Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W.G., Peplies, J., Glockner, F.O., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35, 7188–7196.
[54] Pujalte, M.J., Macian, M.C., Arahal, D.R., Ludwig, W., Schleifer, K.H., Garay, E., 2005. *Nereida ignava* gen. nov., sp nov., a novel aerobic marine alphaproteobacterium that is closely related to uncultured *Prionitis* (alga) gall symbionts. Int. J. Syst. Evol. Microbiol. 55, 631–636.
[55] Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M.V., Tarran, G.A., Pernthaler, J., 2009. Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. Environ. Microbiol. 11, 2078–2093.
[56] Schloss, P.D., Handelsman, J., 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. 71, 1501–1506.
[57] Sekar, R., Fuchs, B.M., Amann, R., Pernthaler, J., 2004. Flow sorting of marine bacterioplankton after fluorescence *in situ* hybridization. Appl. Environ. Microbiol. 70, 6210–6219.
[58] Sfriso, A., Facca, C., 2007. Distribution and production of macrophytes and phytoplankton in the Lagoon of Venice: comparison of actual and past situation. Hydrobiologia 577, 71–85.
[59] Sorokin, Y.I., Sorokin, P.Y., Giovanardi, O., DallaVenezia, L., 1996. Study of the ecosystem of the Lagoon of Venice, with emphasis on anthropogenic impact. Mar. Ecol.-Prog. Ser. 141, 247–261.
[60] Wallner, G., Amann, R., Beisker, W., 1993. Optimizing fluorescent *in situ* hybridization with ribosomal-RNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. Cytometry 14, 136–143.
[61] Warnecke, F., Amann, R., Pernthaler, J., 2004. Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. Environ. Microbiol. 6, 242–253.
[62] West, N.J., Obernosterer, I., Zemb, O., Lebaron, P., 2008. Major differences of bacterial diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom in the Southern Ocean. Environ. Microbiol. 10, 738–756.
[63] Wright, S.W., Jeffrey, S.W., Mantoura, R.F.C., Llewellyn, C.A., Bjornland, T., Repeta, D., Welschmeyer, N., 1991. Improved HPLC method for the analysis of chlorophylls and carotenoids from marine phytoplankton. Mar. Ecol.-Prog. Ser. 77, 183–196.
[64] Yan, S., Fuchs, B.M., Lenk, S., Harder, J., Wulf, J., Jiao, N.Z., Arnann, R., 2009. Biogeography and phylogeny of the NOR5/OM60 clade of *Gammaproteobacteria*. Syst. Appl. Microbiol. 32, 124–139.
[65] Zanon, F., Rado, N., Centanni, E., Zharova, N., Pavoni, B., 2009. Time trend of butyl- and phenyl-tin contamination in organisms of the Lagoon of Venice (1999–2003). Environmental Monitoring and Assessment 152, 35–45.
[66] Zhou, J.Z., Bruns, M.A., Tiedje, J.M., 1996. DNA recovery from soils of diverse composition. Appl. Environ. Microbiol. 62, 316–322.

Supplementary Table S1: Observed OTUs and predicted richness indexes of 16S rRNA sequences at the 97% level identity for the Lagoon and Inlet clone libraries (October and June) indicating average and 95% confidence intervals (CI)

		CHAO			ACE			JACK		
Library	OTU	average	95%lci	95%hci	average	95%lci	95%hci	average	95%lci	95%hci
Lagoon October	36	60	44	107	83	55	156	58	44	71
Lagoon June	49	103	70	190	115	78	198	105	77	133
Total Lagoon	**61**	**128**	**89**	**219**	**158**	**108**	**260**	**139**	**101**	**176**
Inlet October	41	54	45	85	58	47	85	57	46	68
Inlet June	30	35	31	51	38	33	57	39	31	47
Total Inlet	**57**	**73**	**62**	**103**	**79**	**67**	**110**	**77**	**65**	**89**

Appendices

Appendix A

Marine *Bacteroidetes* FISH oligonucleotide probes

In this chapter I will summarize the marine *Bacteroidetes* oligonucleotide probes available to date based on a literature survey, listed in probeBase (Loy et al., 2003), and those designed in the context of this thesis. The specificity of the oligonucleotide probes was evaluated using a comprehensive and quality checked 16S rRNA gene sequences database (Silva, Pruesse et al., 2007). The Silva database release 100 contains 409,907 sequences longer than 1200 nt for *Bacteria* and longer than 900 nt for *Archaea*.

Table A shows the specificities of the *Bacteroidetes* FISH probes. "Total hits" indicate the total number of sequences in the database that had the reverse complementary target site of the probes. The "Number of probe hits in a clade" was considered as the hits of the probe with the original sequences for which the probe was designed and its close relatives. Close relatives were those sequences in the same phylogenetic branch as the original target sequence, based on an ARB parsimony phylogenetic tree that includes all of the sequences from the Silva 100 database. Oligonucleotide probes can also produce false-positive identifications by hybridizing to organisms that are outside the target group, and they are referred as outgroup hits (Amann and Fuchs, 2008). When the outgroup hits were more than 50% of the hits in a clade the probe was considered as unspecific.

A suite of probes targeting *Bacteroidetes* that are found in particular habitats were designed by Weller et al. (2000). The authors, however, concluded that it is difficult to find habitat specific probes for members of the *Bacteroidetes* phylum and that the design of probes for monophyletic groups should remain the standard approach (Weller et al., 2000). However, an evaluation of the probe CF563, which was designed targeting mostly marine isolates, revealed that it indeed targets a monophyletic group. This probe covers 90% of the *Flavobacteria* entries and it has few outgroup hits (350 sequences). Additionally, this probe was tested in different marine environments and its counts are comparable with the counts obtained with the general CF319a probe (Gómez-Pereira, unpublished observations). There is, however, a clear lack of a general *Sphingobacteria* and *Cytophagia* probe and its numbers in marine plankton are rather unknown. Further studies should attempt to delimit marine clades whithin those classes and to design general sphingobacterial and cytophagial oligonucleotide probes.

Some of the evaluated probes, however, are not specific for any monophyletic clade. For example, the probes targeting the AGG58 clade have hits in the entire *Bacteroidetes* phylum. The probes reported by O'Sullivan and colleagues were originally designed with the objective of screening freshwater and marine DNA extracts and isolated strains (O'Sullivan et al., 2004), and this should remain as its application. We developed a probe for few sequences within the AGG58 clade branch 3; however, we were unable to design a probe that specifically targets all the members of each branch. One of the reasons might be that the clade is phylogenetically not yet resolved. The position of the AGG58 branches in the phylogenetic tree is rather unstable, and most of the sequences used to define those branches were partial 16S rRNA gene sequences (300-900 nt). A higher number of good quality full length sequences are needed to unveil the phylogeny of this clade facilitating then its division and further probe design.

Other probes found to lack specificity were CYT448 and CYT438 (Eilers et al., 2001). Both of them targeted several strains of the marine clade genera e.g. *Aquimarina*, *Zobellia*, *Maribacter*, *Leeuwenhoekiella* among others. Therefore, those probes could be applied for the screening of marine isolates,

but not for quantification in environmental samples.

Three probes targeting members of the genus *Polaribacter* are available. Each of them covers more than 80% of the sequences affiliated with the *Polaribacter* genus. Based on the *in silico* analysis the probe POL740 had the lowest number of hits outside the *Polaribacter* genus. However, none of the three *Polaribacter* probes targets the temperate clade of *P. dokdonensis* (Figure 3.21). Therefore for assessing the abundance of this temperate *Polaribacter* clade new oligonucleotide probes are needed. Our analyses have revealed that significant abundances of *Polaribacter* are obtained with probe POL740 in temperate marine systems (Figure 3.19), which could be hybridizations with cells of the branches "uncultured coastal" I and II indicated in Figure 3.21. A division of this genera is necesarry in order to differentiate *Polaribacter* phylotypes that inhabit distinct environments.

Probe name	Target clade	Probe sequence (5->3)	Total hits	Number of probes hits in a target clade	FA (%)	Environment of target clade	Max %	Reference [1]
CF563	*Flavobacteria*	GGACCCTTTAAACCCAAT	5035	4686	20	mostly marine	-	Weller, 2000
CFB562	CFBs excluding *Cytophaga* group	TACGYWCCCTTTAAACCCA	26606	unspecific	30	diverse environments	-	O´Sullivan, 2002
CFB376	CFBs excluding *Cytophaga* and *Bacteroides* groups	TGMCCAATATTCCYTACTG	48363	unspecific	30	diverse environments	-	O´Sullivan, 2002
Flavobacteria (cultured genera)								
LEE82 *	*Leeuwenhoenkiella*	ACTCGTCAGCTTCCCGAA	21	15	25	coastal / ocean	< 0.01%	Gómez-Pereira, 2010
FL443	*Gramella*	TAGATGTGGTTCTTCCCG	12	12	65	coastal	n.d.	Gómez-Pereira, 2010
CYT448	*C. marinoflava - latercula*	CTAGGCCGCTCCTTACGG	1261	unspecific	30	coastal	5%	Eilers, 2001
CYT438	*C. marinoflava - latercula*	CCGCTCCTTACGGTGACG	322	unspecific	30	coastal	5%	Eilers, 2001
Polaribacter								
POL218* [a]	*Polaribacter*	GACGCATAGCCATCTTTTACCG	194	146	30	polar	11%	Malmstrom , 2007
POL740 [a]	*Polaribacter*	CCCTCAGCGTCAGTACATACGT	132	130	35	polar	11%	Malmstrom , 2007
PB223*	*Polaribacter*	GGACGCATAGCCATCTTT	235	147	15	polar	35%	Brinkmeyer, 2003
POL-173	*Polaribacter*	TCTACAATACCATGAGGC	24	23	n.d.	polar	-	Gómez-Pereira, in prep
Dokdonia								
DOK196	*Dokdonia*	TCTTATACCGCCGAAACT	41	21	30	coastal / ocean	<0.3%	Massasna, 2009
DOK827*	*Dokdonia*	GTCCGAAAACCAAACAACTAG	17	15	20	coastal / ocean	< 0.01%	Gómez-Pereira, 2010

Probe name	Target clade	Probe sequence (5->3)	Total hits	Number of probes hits in a target clade	FA (%)	Environment of target clade	Max %	Reference [1]
Flavobacteria (uncultured clades)								
FL440	NSi / VIS6	GACAAGTTTCTTCCTCTAT	83	77	45	coastal / ocean	n.d.	Alonso, unpublished
FL441	NSd	AGAGTGTTTCTTCCTACA	18	18	35	marine / polar	n.d.	Alonso, unpublished
FL 659	NSm	CATTCCGCCTACTCCAACT	31	unspecific	65	coastal / ocean	n.d.	Alonso, unpublished
FL197*	VIS4	TCATCTCATACCGTAACC	108	97	35	coastal / ocean	2%	Gómez-Pereira, 2010
VIS6-814*	VIS6	CAGCGAGTGATGATCGTT	150	145	15	coastal / ocean	0.5%	Gómez-Pereira, 2010
VIS5-586*	VIS5	CCACTGACTTAGATTCCC	40	36	25	deep sea / ocean	0.2%	Gómez-Pereira, 2010
FL845*	VIS2	GCTTAGCCACTCATCTAA	50	48	15	coastal / ocean	0.2%	Gómez-Pereira, 2010
VIS3-183	VIS3	CTCTAGAAGTGATGCCAC	29	28	25	coastal / ocean	0.7%	Gómez-Pereira, 2010
VIS1-575*	VIS1	CTTAACAAACAGCCTGCGGACC	69	68	35	coastal / ocean	4%	Gómez-Pereira, 2010
ANT-981* [b]	ANT	TTCCCATTTAAACCCTGGTA	47	43	20	coastal	2%	Gómez-Pereira, in prep
ANT-72 [b]	ANT	TCAAACATCCGAAGACATTC	28	26	20	coastal / polar	2%	Gómez-Pereira, in prep
NS9-664	NS9	ACATGACCTATTCCGCCAACTT	22	22	35	coastal / ocean	2%	Gómez-Pereira, in prep
DE clade								
CF6-1267	DE2 cluster	GAAGATTCGCTCCTCCTC	16	15	35	coastal / ocean	10%	Kirchman, 2003
F8737E-843*	DE2 / FOS-873	CGCTTAACCACTCAATCAACCA	13	12	20	ocean	4%	Gómez-Pereira, in prep
F805E4-471*	DE2 / FOS-805	GTAAGTAGGTTTCTTCCTGTAT	32[+]	13	30	coastal / ocean / sediments	1.5%	Gómez-Pereira, in prep

Probe name	Target clade	Probe sequence (5->3)	Total hits	Number of probes hits in a target clade	FA (%)	Environment of target clade	Max %	Reference [1]
AGG58								
AGG58-A*	AGG58	CTACATGWCAYATTCCGCC	2147	unspecific	40	coastal	n.d.	O´Sullivan, 2004
AGG58-B	AGG58 branches 1 and 3	CGTTMTGCATTTAAGMCC	302	unspecific	30	coastal	n.d.	O´Sullivan, 2004
AGG58-C	AGG58 branch 2	CACAGGCAGTRTTTCTAGAG	20	unspecific	35	coastal	n.d.	O´Sullivan, 2004
AGG58-145	AGG58 branch 2	CCACATTTCTATGGGCTATCCC	110	unspecific	30	polar / coastal	7%	Malmstrom, 2007
AGG58-742 *	AGG58 branch 3	CCCTTAGTGTCAATACTGAC	4	4	35	deep sea / ocean	2%	Gómez-Pereira, in prep
Sphingobacteria								
FL1261	NS11-NS12	GGTTTGCGCCTTGTTACCAAG	6	6	55	coastal / ocean / sludge / lake	n.d.	Alonso, unpublished
SPA-730	Sphingobacteria A	ATATGCCTAGCTAGCTGC	14	14	35	ocean / macroalgae	0.1%	Gómez-Pereira, in prep
SPC-814	Sphingobacteria C	AAGATCGAGTAAACATCGTT	5	3	n.d.	sediment / ocean	< 0.01%	Gómez-Pereira, in prep
Cytophagia								
CYT-734 [c]	*Cytophagia*	CAGTTTCTGCCTAGTAAG	6	6	25	ocean	2%	Gómez-Pereira, in prep
CYT-1462 [c]	*Cytophagia*	CTAGCTCTACCCTAAACA	76	76	25	ocean	2%	Gómez-Pereira, in prep

1: full references given in chapter "References"; *: applied with competitor probe; +: weak mismtches resolved with the competitor probes; a: probes used together in Malstrom et al., 2007; b, c: probes used together in Gómez-Pereira, in prep. FA: formamide concentration in hybridization buffer; Max %: maximum percentage of total picoplankton reported; n.d. not determined.

Appendix B

Bacteroidetes fosmids gene content

The following tables list all ORFs in the *Bacteroidetes* fosmids from S3 and S18 in the North Atlantic Ocean and its putative function. Annotation was performed with GenDB v2 system (Meyer et al., 2003) using various tools for each predicted gene. Similarity searches were performed against sequence databases (NCBI nr, NCBI nt, SwissProt) and protein family databases (Pfam, InterPro, COG). From these predictions, an automatic annotation was generated using the fuzzy logic–based auto–annotation tool MicHanThi (Quast, 2006). High-quality annotations were generated by manual revision of each gene annotation.

Appendix B

Genomic content of *Bacteroidetes* fosmids retrieved from S3

***Flavobacteria* , VIS5, S3-861-A8**

ORF	AA	Nuc	Product	Gene	EC
S3_861_A8_1	817	2454	ATP-dependent protease La 2	lon	3.4.21.53
S3_861_A8_2	190	573	conserved hypothetical protein		
S3_861_A8_3	465	1398	conserved hypothetical protein, secreted		
S3_861_A8_4	437	1314	major facilitator superfamily permease		
S3_861_A8_5	164	495	conserved hypothetical protein, secreted		
S3_861_A8_6	163	492	methionine sulphoxide reductase B		1.8.4.-
S3_861_A8_7	133	402	methionine sulfoxide reductase B		1.8.4.-
S3_861_A8_8	458	1377	dihydrolipoamide dehydrogenase		1.8.1.4
S3_861_A8_9	137	414	dehydroquinase, class II		4.2.1.10
S3_861_A8_10	224	675	conserved hypothetical protein		
S3_861_A8_11	194	585	conserved hypothetical protein, secreted		
S3_861_A8_12	187	564	conserved hypothetical protein, secreted		
S3_861_A8_13	323	972	Tyrosine recombinase	xerD	
S3_861_A8_14	521	1566	phosphohydrolase		
S3_861_A8_15	94	285	conserved hypothetical protein		
S3_861_A8_16	96	291	conserved hypothetical protein		
S3_861_A8_17	561	1686	Peptidase M23B		
S3_861_A8_18	821	2466	TonB-dependent receptor, plug		
S3_861_A8_19	381	1146	cysteine desulfhydrase	iscS	2.8.1.7
S3_861_A8_20	294	885	conserved hypothetical protein		
S3_861_A8_21	182	549	hypothetical protein		
S3_861_A8_22	413	1242	Ornithine aminotransferase		
S3_861_A8_23	484	1455	RNA methyltransferase		2.1.1.-
S3_861_A8_24	178	537	conserved hypothetical protein		
S3_861_A8_25	234	705	conserved hypothetical protein, secreted		
S3_861_A8_26	385	1158	protein with similarity to RNA methylase family UPF0020		
S3_861_A8_27	243	732	putative methyltransferase		2.1.1.-
S3_861_A8_28	222	669	conserved hypothetical protein		
S3_861_A8_29		107	5S		
S3_861_A8_30		2826	23S		
S3_861_A8_31		77	t-RNA		
S3_861_A8_32		74	t-RNA		
S3_861_A8_33		1516	16S		

***Flavobacteria, Polaribacter* , S3-843-F5**

ORF	AA	Nuc	Product	Gene	EC
S3_843_F5_0	848	2547	DNA gyrase subunit A	gyrA	5.99.1.3
S3_843_F5_1	848	2547	ATP-dependent Clp protease ATP-binding subunit clpC		
S3_843_F5_2	340	1023	pretein containing thioredoxin-like fold		
S3_843_F5_3	317	954	acetyl-CoA carboxylase, alpha subunit		
S3_843_F5_4	304	915	protein of unknown function DUF6, transmembrane		
S3_843_F5_5	363	1092	predicted permease YjgP/YjgQ		
S3_843_F5_6	376	1131	queuine tRNA-ribosyltransferase		2.4.2.29
S3_843_F5_7	105	318	protein with similarity to pseudouridine synthase		
S3_843_F5_8	164	495	ribosomal large subunit pseudouridine synthase F	rluF	5.4.99.-
S3_843_F5_9	200	603	2OG-Fe(II) oxygenase		1.14.11.2
S3_843_F5_10	135	408	metal transport related protein		
S3_843_F5_11	245	738	protein containing heavy-metal-associated domain		
S3_843_F5_12		109	5S		
S3_843_F5_13		2884	23S		
S3_843_F5_14		77	t-RNA		
S3_843_F5_15		77	t-RNA		
S3_843_F5_16		1510	16S		
S3_843_F5_17	552	1659	hypothetical protein		
S3_843_F5_18	449	1350	conserved hypothetical protein		
S3_843_F5_19	249	750	conserved hypothetical protein		
S3_843_F5_20	665	1998	NAD-dependent DNA ligase	ligA	6.5.1.2
S3_843_F5_21	220	663	conserved hypothetical protein		
S3_843_F5_22	213	642	protein of unknown function DUF422, transmembrane		
S3_843_F5_23	233	702	spheroidene monooxygenase	crtA	
S3_843_F5_24	309	930	conserved hypothetical protein		
S3_843_F5_25	189	570	conserved hypothetical protein, secreted		
S3_843_F5_26	480	1443	glyceraldehyde-3-phosphate dehydrogenase, type I		1.2.1.12
S3_843_F5_27	462	1389	periplasmic trypsin-like serine protease		
S3_843_F5_28	256	771	Diaminopimelate epimerase	dapF	5.1.1.7
S3_843_F5_29	173	522	acetyltransferase		2.3.-.-
S3_843_F5_30	333	1002	aminodeoxychorismate lyase family protein		
S3_843_F5_31	157	474	conserved hypothetical protein		
S3_843_F5_32	234	705	tetrapyrrole methylase family protein		
S3_843_F5_33	145	438	low molecular weight phosphotyrosine protein phosphatase		3.1.3-
S3_843_F5_34	475	1428	chromosomal replication initiation protein	dnaA	
S3_843_F5_35	132	399	thioesterase superfamily protein		
S3_843_F5_36	121	366	conserved hypothetical protein		

Appendix B

***Flavobacteria, Polaribacter,* S3-860-D3**

ORF	AA	Nuc	Product	Gene	EC
S3_860_D3_0	746	2241	glycoside hydrolase family 92		
S3_860_D3_1	109	330	hypothetical protein		
S3_860_D3_2	323	972	mannose-6-phosphate isomerase, class I		5.3.1.8
S3_860_D3_3	533	1602	sulfatase		
S3_860_D3_4	390	1173	oligopeptidase family protein, S9		
S3_860_D3_5	546	1641	N-acetylglucosamine-6-sulfatase		
S3_860_D3_6	498	1497	sulfatase		
S3_860_D3_7	482	1449	N-acetylgalactosamine-6-sulfatase		
S3_860_D3_8	438	1317	L-fucose permease	fucP	
S3_860_D3_9	760	2283	glycoside hydrolase family 92		
S3_860_D3_10	749	2250	glycoside hydrolase family 92		
S3_860_D3_11	425	1278	conserved hypothetical protein		
S3_860_D3_12	358	1077	conserved hypothetical protein		
S3_860_D3_13	578	1737	conserved hypothetical protein, secreted		
S3_860_D3_14	1077	3234	TonB-dependent receptor		
S3_860_D3_15	388	1167	conserved hypothetical protein, secreted		
S3_860_D3_16	818	2457	hypothetical protein		
S3_860_D3_17	457	1374	sulfatase		
S3_860_D3_18	279	840	hypothetical protein		
S3_860_D3_19	204	615	hypothetical protein		
S3_860_D3_20	58	177	conserved hypothetical protein		
S3_860_D3_21	902	2709	conserved hypothetical protein, secreted		
S3_860_D3_22	437	1314	conserved hypothetical protein		
S3_860_D3_23	847	2544	conserved hypothetical protein		
S3_860_D3_24	278	837	oxidoreductase domain protein		
S3_860_D3_25	163	492	oxidoreductase domain protein		
S3_860_D3_26	52	159	hypothetical protein		
S3_860_D3_27		1356	16S rRNA		

Flavobacteria, **DE2, S3-805-E4**

ORF	AA	Nuc	Product	Gene	EC
S3_805_E4_0		978	16S		
S3_805_E4_1	659	1980	conserved hypothetical protein		
S3_805_E4_2	194	585	Ribonuclease HII	rnhB	3.1.26.4
S3_805_E4_3	237	714	octanoyltransferase	lipB	2.3.1.181
S3_805_E4_4	845	2538	hypothetical protein		
S3_805_E4_5	563	1692	lysyl-tRNA synthetase	lysS	6.1.1.6
S3_805_E4_6	428	1287	glutamate-1-semialdehyde 2,1-aminomutase	hemL	5.4.3.8
S3_805_E4_7	277	834	mannosyl-glycoprotein endo-beta-N-acetylglucosamidase		
S3_805_E4_8	302	909	pyridoxal phosphate-dependent enzyme, beta subunit		
S3_805_E4_9	49	150	conserved hypothetical protein		
S3_805_E4_10	177	534	hypothetical protein		
S3_805_E4_11	537	1614	sulfate transporter family protein		
S3_805_E4_12	870	2613	alanyl-tRNA synthetase	alaS	6.1.1.7
S3_805_E4_13	325	978	peptidase, family M23		3.4.24.-
S3_805_E4_14	109	330	putative DNA binding		
S3_805_E4_15	434	1305	GTP-binding protein	engA	
S3_805_E4_16	301	906	GTP-binding protein era homolog	era	
S3_805_E4_17		76	tRNA		
S3_805_E4_18	203	612	[similarity to] alkylhydroperoxidase like protein, AhpD		
S3_805_E4_19	148	447	putative transcription regulators		
S3_805_E4_20	138	417	membrane protein containing DUF1200		
S3_805_E4_21	295	888	esterase		
S3_805_E4_22	48	147	hypothetical protein		
S3_805_E4_23	50	153	hypothetical protein		
S3_805_E4_24	263	792	secreted protein		
S3_805_E4_25	228	687	secreted protein		
S3_805_E4_26	323	972	conserved hypothetical protein		
S3_805_E4_27	454	1365	membrane protein		
S3_805_E4_28	304	915	conserved hypothetical protein		
S3_805_E4_29	902	2709	conserved hypothetical protein, secreted		
S3_805_E4_30	172	519	hypothetical protein		
S3_805_E4_31	120	363	conserved hypothetical protein, secreted		
S3_805_E4_32	721	2166	membrane protein, probable OmpA family		
S3_805_E4_33	308	927	conserved hypothetical protein, secreted		
S3_805_E4_34	425	1278	hypothetical protein		

Appendix B

***Flavobacteria,* ANT, S3-816-G8**

ORF	AA	Nuc	Product	Gene	EC
S3_816_G8_0	122	369	hypothetical protein		
S3_816_G8_1	113	342	MerR family transcriptional regulator		
S3_816_G8_2	325	978	M23 family peptidase		
S3_816_G8_3	834	2505	Alanyl-tRNA synthetase	alaS	6.1.1.7
S3_816_G8_4	218	657	conserved hypothetical protein		
S3_816_G8_5	420	1263	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine liga	murF	6.3.2.10
S3_816_G8_6	526	1581	Gliding motility lipoprotein GldJ	gldJ	
S3_816_G8_7	330	993	conserved hypothetical protein, secreted		
S3_816_G8_8	1153	3462	conserved hypothetical protein		
S3_816_G8_9	407	1224	conserved hypothetical protein		
S3_816_G8_10	331	996	pyruvate dehydrogenase E1 component, alpha subunit		1.2.4.1
S3_816_G8_11	414	1245	Dihydrolipoamide acetyltransferase pyruvate dehydrogenase complex		
S3_816_G8_12	81	246	Ferrous iron transport protein A		
S3_816_G8_13	671	2016	Ferrous iron transport protein B		
S3_816_G8_14	221	666	Short-chain dehydrogenase/reductase family protein		
S3_816_G8_15		1511	16S		
S3_816_G8_16		77	t-RNA		
S3_816_G8_17		77	t-RNA		
S3_816_G8_18		2870	23S		
S3_816_G8_19		110	5S		
S3_816_G8_20	587	1764	conserved hypothetical protein		
S3_816_G8_21	501	1506	replicative DNA helicase	dnaC	3.6.1.-
S3_816_G8_22	318	957	Acetyl-coenzyme A carboxylase carboxyl transferase subun	accA	6.4.1.2
S3_816_G8_23	312	939	Protein of unknown function DUF6, transmembrane		
S3_816_G8_24	359	1080	Predicted permease YjgP/YjgQ, transmembrane		
S3_816_G8_25	354	1065	queuine tRNA-ribosyltransferase		2.4.2.29
S3_816_G8_26	376	1131	protein with similarity to glycosyl transferase family 2		
S3_816_G8_27	207	624	glucose-inhibited division protein B		
S3_816_G8_28	541	1626	aldehyde dehydrogenase family protein		
S3_816_G8_29	213	642	protein containing metallophosphoesterase domain		
S3_816_G8_30	373	1122	DNA polymerase III, beta chain		2.7.7.7
S3_816_G8_31	332	999	hypothetical protein		
S3_816_G8_32	70	213	hypothetical protein		

***Sphingobacteria*, S3-972-A4**

ORF	AA	Nuc	Product	Gene	EC
S3_972_A4_0	1001	3006	hypothetical protein		
S3_972_A4_1	496	1491	formyltransferase/IMP cyclohydrolase		
S3_972_A4_2		1516	16s rRNA		
S3_972_A4_3		74	tRNA		
S3_972_A4_4		74	tRNA		
S3_972_A4_5		69	tRNA		
S3_972_A4_6		2816	23s rRNA		
S3_972_A4_7		111	5s rRNA		
S3_972_A4_8	44	135	hypothetical protein		
S3_972_A4_9	253	762	protein containing methyltransferase domain		
S3_972_A4_10	311	936	lipoyl synthase	lipA	2.8.1.8
S3_972_A4_11	749	2250	hypothetical protein		
S3_972_A4_12	197	594	DNA repair protein		
S3_972_A4_13	179	540	hypothetical protein		
S3_972_A4_14	302	909	LmbE family protein		
S3_972_A4_15	237	714	hypothetical protein		
S3_972_A4_16	156	471	Ribosomal protein S7		
S3_972_A4_17	701	2106	Protein synthesis factor, GTP-binding		
S3_972_A4_18	102	309	Ribosomal protein S10		
S3_972_A4_19	205	618	50S ribosomal protein L3	rplC	
S3_972_A4_20	208	627	Ribosomal protein L4		
S3_972_A4_21	96	291	50S ribosomal protein L12	rplW	
S3_972_A4_22	273	822	50S ribosomal protein L2	rplB	
S3_972_A4_23	87	264	Ribosomal protein S19	rpsS	
S3_972_A4_24	127	384	Ribosomal protein L22		
S3_972_A4_25	247	744	ribosomal protein S3	rpsC	
S3_972_A4_26	139	420	Ribosomal protein L16		
S3_972_A4_27	67	204	Ribosomal protein L29		
S3_972_A4_28	84	255	Ribosomal protein S17		
S3_972_A4_29	122	369	Ribosomal protein L14		
S3_972_A4_30	112	339	Ribosomal protein L14	rplN	
S3_972_A4_31	181	546	50S ribosomal protein L5	rplE	
S3_972_A4_32	89	270	Ribosomal protein S14		
S3_972_A4_33	135	408	Ribosomal protein S8		
S3_972_A4_34	182	549	Ribosomal protein L6		
S3_972_A4_35	114	345	Ribosomal protein L18		
S3_972_A4_36	172	519	Ribosomal protein S5		
S3_972_A4_37	59	180	Ribosomal protein L30		

Appendix B

***Flavobacteria,* S3-933-H4**

ORF	AA	Nuc	Product	Gene	EC
S3_933_H4_0		2859	23S		
S3_933_H4_1		1517	16S		
S3_933_H4_2	421	1266	sulfate adenylyltransferase subunit 1	cysN	2.7.7.4
S3_933_H4_3	300	903	sulfate adenylyltransferase subunit 2	cysD	2.7.7.4
S3_933_H4_4	198	597	Adenylyl-sulfate kinase	cysC	2.7.1.25
S3_933_H4_5	258	777	3'-phosphoadenosine 5'-phosphate phosphatase	cysQ	3.1.3.7
S3_933_H4_6	378	1137	protein containing glycosyl transferase, group 1 domain		
S3_933_H4_7	658	1977	conserved hypothetical protein		
S3_933_H4_8	248	747	3-oxoacyl-[acyl-carrier-protein] reductase	fabG	1.1.1.100
S3_933_H4_9	261	786	protein containing MazG nucleotide pyrophosphohydrolase domain		
S3_933_H4_10	249	750	hypothetical protein		
S3_933_H4_11	258	777	secreted protein		
S3_933_H4_12	373	1122	hypothetical protein		
S3_933_H4_13	124	375	hypothetical protein		
S3_933_H4_14	56	171	hypothetical protein		
S3_933_H4_15	77	234	RNA polymerase sigma factor		
S3_933_H4_16	90	273	RNA polymerase sigma factor		
S3_933_H4_17	927	2784	UvrABC system protein A	uvrA	
S3_933_H4_18	127	384	conserved hypothetical protein		
S3_933_H4_19	186	561	conserved hypothetical protein		
S3_933_H4_20	243	732	protein containing methyltransferase domain		
S3_933_H4_21	241	726	ZIP Zinc transporter family protein		
S3_933_H4_22	209	630	Thioredoxin		
S3_933_H4_23	126	381	hypothetical protein		
S3_933_H4_24	115	348	protein containing histidine kinase		
S3_933_H4_25	195	588	RNA pseudouridine synthase family protein		
S3_933_H4_26	266	801	[weak similarity to] outer membrane protein, OmpA family		
S3_933_H4_27	292	879	[weak similarity to] outer membrane protein, OmpA family		

S3-979-825M	***Flavobacteria,* VIS6 , S3-979-C6**					
ORF merged	**ORF**	**AA**	**Nuc**	**Product**	**Gene**	**EC**
S3_979_825M_0	S3_979_C6_0	374	1125	predicted permease YjgP/YjgQ		
S3_979_825M_1	S3_979_C6_1	312	939	protein of unknown function DUF6, transmembrane		
S3_979_825M_2	S3_979_C6_2	317	954	acetyl-coenzyme A carboxylase carboxyl transferase subuni	accA	6.4.1.2
S3_979_825M_3	S3_979_C6_3	510	1533	replicative DNA helicase	dnaC	3.6.1.-
S3_979_825M_4	S3_979_C6_4	636	1911	hypothetical protein		
S3_979_825M_5	S3_979_C6_5	397	1194	conserved hypothetical protein		
S3_979_825M_6	S3_979_C6_6		107	5S		
S3_979_825M_7	S3_979_C6_7		2866	23S		
S3_979_825M_8	S3_979_C6_8		77	t-RNA		
S3_979_825M_9	S3_979_C6_9		77	t-RNA		
S3_979_825M_10	S3_979_C6_10		1514	16S		
S3_979_825M_11	S3_979_C6_11	228	687	short-chain dehydrogenase/reductase family protein		
S3_979_825M_12	S3_979_C6_12	424	1275	dihydrolipoamide acyltransferases		
S3_979_825M_13	S3_979_C6_13	331	996	pyruvate dehydrogenase E1 component subunit alpha		1.2.4.1
S3_979_825M_14	S3_979_C6_14	371	1116	conserved hypothetical protein		
S3_979_825M_15	S3_979_C6_15	1115	3348	conserved hypothetical protein, secreted		
S3_979_825M_16	S3_979_C6_16	328	987	conserved hypothetical protein, secreted		
S3_979_825M_17	S3_979_C6_17	527	1584	Gliding motility lipoprotein GldJ	gldJ	
S3_979_825M_18	S3_979_C6_18	423	1272	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine liga	murF	
S3_979_825M_19	S3_979_C6_19	381	1146	Acyl-CoA dehydrogenase	acdA	1.3.99.-
S3_979_825M_20	S3_979_C6_20	352	1059	Anhydro-N-acetylmuramic acid kinase	anmK	2.7.1.n1
S3_979_825M_21	S3_979_C6_21	408	1227	protein belonging to Glu/Leu/Phe/Val dehydrogenase family		
S3_979_825M_22	S3_979_C6_22	436	1311	Na+/H+ antiporter NhaD		
S3_979_825M_23	S3_979_C6_23	232	699	MotA/TolQ/ExbB family biopolymer transport protein		
S3_979_825M_24	S3_979_C6_24	133	402	ExbD/TolR family biopolymer transport protein		
S3_979_825M_25	S3_979_C6_25	284	855	conserved hypothetical protein		
S3_979_825M_26	S3_979_C6_26	284	855	conserved hypothetical protein		
S3_979_825M_27	S3_979_C6_27	400	1203	folylpolyglutamate synthase		6.3.2.-
S3_979_825M_28	S3_979_C6_28	195	588	hypothetical protein, secreted		
S3_979_825M_29	S3_979_C6_29	539	1620	Mg chelatase, subunit ChlI		
S3_979_825M_30	S3_979_C6_30	349	1050	radical SAM superfamily protein, UPF0063		
S3_979_825M_31	S3_979_C6_31	154	465	protein containing Pirin, C-terminal domain		
S3_979_825M_32	S3_825_D3_20	285	858	protein containing Pirin, C-terminal domain		
S3_979_825M_33	S3_825_D3_21	231	696	haloacid dehalogenase-like hydrolase superfamily		
S3_979_825M_34	S3_825_D3_22	309	930	conserved hypothetical protein		
S3_979_825M_35	S3_825_D3_23	137	414	conserved hypothetical protein		

Appendix B

S3-892-858M	Sphingobacteria A, S38_892_C7					
ORF merged	**ORF**	**AA**	**Nuc**	**Product**	**Gene**	**EC**
S3_892_858M_0	S3_892_C7_34	208	627	Glycosyl transferases group 1		
S3_892_858M_1	S3_892_C7_33	367	1104	Glycosyl transferases group 1		
S3_892_858M_2	S3_892_C7_32	363	1092	UDP-N-acetylglucosamine 2-epimerase		5.1.3.14
S3_892_858M_3	S3_892_C7_31	230	693	hypothetical protein		
S3_892_858M_4	S3_892_C7_30	356	1071	mannose-1-phosphate guanylyltransferase	manC	2.7.7.13
S3_892_858M_5	S3_892_C7_29	379	1140	Glycosyl transferases group 1		
S3_892_858M_6	S3_892_C7_28	36	111	hypothetical protein		
S3_892_858M_7	S3_892_C7_27	563	1692	hypothetical protein, secreted		
S3_892_858M_8	S3_892_C7_26	392	1179	hypothetical protein		
S3_892_858M_9	S3_892_C7_25	672	2019	UvrABC system protein B	uvrB	
S3_892_858M_10	S3_892_C7_24	255	768	alpha/beta superfamily hydrolase		
S3_892_858M_11	S3_892_C7_23	548	1647	carboxy-terminal processing protease		
S3_892_858M_12	S3_892_C7_22	231	696	glycoprotease, peptidase M22		
S3_892_858M_13	S3_892_C7_21	112	339	regulatory protein		
S3_892_858M_14	S3_892_C7_20	105	318	regulatory protein		
S3_892_858M_15	S3_892_C7_19	669	2010	2-oxoisovalerate dehydrogenase E1 component subunits alpha and beta		1.2.4.4
S3_892_858M_16	S3_892_C7_18	494	1485	aldehyde dehydrogenase family protein		
S3_892_858M_17	S3_892_C7_17	138	417	endoribonuclease L-PSP		
S3_892_858M_18	S3_892_C7_16	567	1704	long-chain fatty-acid-CoA ligase	fadD	6.2.1.3
	S3_858_A8					
S3_892_858M_19	S3_858_A8_1	740	2223	30S ribosomal protein S1	rpsA	
S3_892_858M_20	S3_858_A8_2	521	1566	phytoene dehydrogenase	crtI	1.14.99.-
S3_892_858M_21	S3_858_A8_3	282	849	phytoene synthase	crtB	2.5.1.-
S3_892_858M_22	S3_858_A8_4	34	105	hypothetical protein		
S3_892_858M_23	S3_858_A8_5		1516	16S		
S3_892_858M_24	S3_858_A8_6		77	tRNA		
S3_892_858M_25	S3_858_A8_7		74	tRNA		
S3_892_858M_26	S3_858_A8_8		2852	23S		
S3_892_858M_27	S3_858_A8_9		105	5S		
S3_892_858M_28	S3_858_A8_10	244	735	protein with similarity to polynucleotidyl transferase		
S3_892_858M_29	S3_858_A8_11	195	588	hypothetical protein		
S3_892_858M_30	S3_858_A8_12	182	549	uncharacterised protein family UPF0093		
S3_892_858M_31	S3_858_A8_13	52	159	hypothetical protein		
S3_892_858M_32	S3_858_A8_14	72	219	hypothetical protein		
S3_892_858M_33	S3_858_A8_15	45	138	hypothetical protein		
S3_892_858M_34	S3_858_A8_16	296	891	membrane protein containing DUF6		
S3_892_858M_35	S3_858_A8_17	503	1512	F0F1 ATP synthase subunit beta		
S3_892_858M_36	S3_858_A8_18	82	249	ATPase, F1 complex, delta/epsilon subunit		3.6.3.14
S3_892_858M_37	S3_858_A8_19	140	423	conserved hypothetical protein		
S3_892_858M_38	S3_858_A8_20	272	819	Peptidase M23		
S3_892_858M_39	S3_858_A8_21	1032	3099	protein containing tetratricopeptide repeat domain		
S3_892_858M_40	S3_858_A8_22	409	1230	Alanine dehydrogenase		
S3_892_858M_41	S3_858_A8_23	53	162	hypothetical protein		
S3_892_858M_42	S3_858_A8_24	510	1533	aldehyde dehydrogenase		
S3_892_858M_43	S3_858_A8_25	76	231	hypothetical protein		
S3_892_858M_44	S3_858_A8_26	71	216	conserved hypothetical protein		
S3_892_858M_45	S3_858_A8_27	438	1317	S-adenosyl-L-homocysteine hydrolase	ahcY	3.3.1.1
S3_892_858M_46	S3_858_A8_28	49	150	hypothetical protein		
S3_892_858M_47	S3_858_A8_29	73	222	protein with similarity to twin-arginine translocation protein		
S3_892_858M_48	S3_858_A8_30	317	954	oxygenase superfamily protein		
S3_892_858M_49	S3_858_A8_31	378	1137	protein containing rhodanese domain		
S3_892_858M_50	S3_858_A8_32	230	693	ribosomal large subunit pseudouridine synthase B		
S3_892_858M_51	S3_858_A8_33	515	1548	DNA helicase, DnaB type		3.6.1.-
S3_892_858M_52	S3_858_A8_34	217	654	secreted protein		
S3_892_858M_53	S3_858_A8_35	35	108	hypothetical protein		
S3_892_858M_54	S3_858_A8_36	40	123	hypothetical protein		
S3_892_858M_55	S3_858_A8_37	120	363	hypothetical protein		
S3_892_858M_56	S3_858_A8_38	47	144	hypothetical protein		

Genomic content of *Bacteroidetes* fosmids retrieved from S18

***Flavobacteria, Leeuwenhoekiella,* S18-997-C3**

ORF	AA	Nuc	Product	Gene	EC
S18_997_C3_0	1084	3255	TonB-dependent outer membrane receptor		
S18_997_C3_1	503	1512	SusD/RagB family protein, outer membrane		
S18_997_C3_2	648	1947	conserved hypothetical protein, secreted		
S18_997_C3_3	35	108	hypothetical protein		
S18_997_C3_4		1514	16S		
S18_997_C3_5		74	tRNA		
S18_997_C3_6		77	tRNA		
S18_997_C3_7		2833	23S		
S18_997_C3_8		108	5S		
S18_997_C3_9	1013	802	DNA polymerase III	dnaE	2.7.7.7
S18_997_C3_10	406	1221	DNA polymerase IV	dinB	2.7.7.7
S18_997_C3_11	283	852	conserved hypothetical protein		
S18_997_C3_12	35	108	hypothetical protein		
S18_997_C3_13		76	tRNA		
S18_997_C3_14	249	750	Type III pantothenate kinase	coaX	2.7.1.33
S18_997_C3_15	418	1257	conserved hypothetical protein, secreted		
S18_997_C3_16	457	1374	TPR domain protein		
S18_997_C3_17	196	591	conserved hypothetical protein, secreted		
S18_997_C3_18	40	123	hypothetical protein		
S18_997_C3_19	428	1287	protein containing CBS and DUF21 domains		
S18_997_C3_20	706	2121	protein containing peptidyl-prolyl cis-trans isomerase		
S18_997_C3_21	305	918	conserved hypothetical protein		
S18_997_C3_22	362	1089	Beta-lactamase		
S18_997_C3_23	454	1365	Hydroxymethylglutaryl-CoA reductase	hmgA	1.1.1.88
S18_997_C3_24	723	2172	Dipeptidyl peptidase IV, secreted		
S18_997_C3_25	521	1566	Amino acid/peptide transporter		
S18_997_C3_26	577	1734	Amino acid/peptide transporter		

Appendix B

***Flavobacteria, Dokdonia* , S18-841-H9**

ORF	AA	Nuc	Product	Gene	EC
S18_841_H9_0	121	366	hypothetical protein		
S18_841_H9_2	93	282	hypothetical protein		
S18_841_H9_1	276	831	3-isopropylmalate dehydrogenase	leuB	1.1.1.85
S18_841_H9_3	427	1284	Argininosuccinate lyase	argH	4.3.2.1
S18_841_H9_4	356	1071	Acetylornithine deacetylase	argE	3.5.1.16
S18_841_H9_5	261	786	Acetylglutamate kinase	argB	2.7.2.8
S18_841_H9_6	321	966	N-acetylornithine carbamoyltransferase	argF	2.1.3.3
S18_841_H9_7	374	1125	Acetylornithine aminotransferase	argD	5.4.3.8
S18_841_H9_8	264	795	Pyrroline-5-carboxylate reductase	pycr1	1.5.1.2
S18_841_H9_9	324	975	N-acetyl-gamma-glutamyl-phosphate reductase	argC	1.2.1.38
S18_841_H9_10	393	1182	Argininosuccinate synthase	argG	6.3.4.5
S18_841_H9_11	205	618	conserved hypothetical protein		
S18_841_H9_12	36	111	hypothetical protein		
S18_841_H9_13	521	1566	HD superfamily hydrolase		
S18_841_H9_14	96	291	conserved hypothetical protein		
S18_841_H9_15	96	291	conserved hypothetical protein		
S18_841_H9_16	285	858	N-acetylmuramic acid 6-phosphate etherase	murQ	4.2.-.-
S18_841_H9_17	73	222	conserved hypothetical protein, membrane		
S18_841_H9_18	175	528	conserved hypothetical protein		
S18_841_H9_19		107	5S		
S18_841_H9_20		2829	23S		
S18_841_H9_21		74	tRNA		
S18_841_H9_22		74	tRNA		
S18_841_H9_23		1516	16S		
S18_841_H9_24	92	279	hypothetical protein		
S18_841_H9_25	668	2007	conserved hypothetical protein, secreted		
S18_841_H9_26	212	639	Ribonuclease HII	rnhB	3.1.26.4
S18_841_H9_27	818	2457	conserved hypothetical protein		
S18_841_H9_28	238	717	lipoyltransferase	lipB	2.3.1.181
S18_841_H9_29	49	150	conserved hypothetical protein		
S18_841_H9_30	564	1695	Lysyl-tRNA synthetase	lysS	6.1.1.6
S18_841_H9_31	136	411	hypothetical protein		
S18_841_H9_32	222	669	conserved hypothetical protein		
S18_841_H9_33	387	1164	beta-lactamase		
S18_841_H9_34	267	804	transcriptional regulator		
S18_841_H9_35	128	387	conserved hypothetical protein DUF393		
S18_841_H9_36	153	462	protein containing YiaAB two helix domain		
S18_841_H9_37	490	1473	conserved hypothetical protein		
S18_841_H9_38	825	2478	conserved hypothetical protein, secreted		
S18_841_H9_39	163	492	conserved hypothetical protein, secreted		
S18_841_H9_40	428	1287	Glutamate-1-semialdehyde 2,1-aminomutase	hemL	5.4.3.8
S18_841_H9_41	277	834	mannosyl-glycoprotein endo-beta-N-acetylglucosamidase family protein		3.2.1.-
S18_841_H9_42	302	909	1-aminocyclopropane-1-carboxylate deaminase		
S18_841_H9_43	108	327	conserved hypothetical protein		

***Flavobacteria,* VIS4, S18-920-B9**

ORF	AA	Nuc	Product	Gene	EC
S18_920_B9_0	568	1707	gliding motility-related protein possible serine kinase	gldJ	
S18_920_B9_1	427	1284	UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopim	murF	6.3.2.10
S18_920_B9_2	135	408	hypothetical protein UPF0047		
S18_920_B9_3	372	1119	HlyD family secretion protein		
S18_920_B9_4	1077	3234	AcrB/AcrD/AcrF family membrane transport protein		
S18_920_B9_5	439	1320	putative outer membrane efflux protein TolC		
S18_920_B9_6	163	492	conserved hypothetical protein, secreted		
S18_920_B9_7		78	tRNA		
S18_920_B9_8	163	492	Asparaginase		3.5.1.1
S18_920_B9_9	153	462	conserved hypothetical protein		
S18_920_B9_10	431	1296	transmembrane metalloprotease, peptidase family M48		
S18_920_B9_11	306	921	membrane protein containing DUF1295		
S18_920_B9_12	403	1212	Peptidase S9A		
S18_920_B9_13	249	750	Peptidase S9, prolyl oligopeptidase active site region		
S18_920_B9_14	211	636	hypothetical protein		
S18_920_B9_15	35	108	hypothetical protein		
S18_920_B9_16	485	1458	deoxyribodipyrimidine photolyase		4.1.99.3
S18_920_B9_17	49	150	hypothetical protein		
S18_920_B9_18	124	375	membrane protein		
S18_920_B9_19		77	t-RNA		
S18_920_B9_20	399	1200	Folylpolyglutamate synthase	folC	6.3.2.17
S18_920_B9_21	457	1374	citrate transporter		
S18_920_B9_22	371	1116	Anhydro-N-acetylmuramic acid kinase	anmK	2.7.1.-
S18_920_B9_23		107	5S		
S18_920_B9_24		2819	23S		
S18_920_B9_25		77	t-RNA		
S18_920_B9_26		77	t-RNA		
S18_920_B9_27		1514	16S		
S18_920_B9_28	59	180	hypothetical protein		
S18_920_B9_29	665	1998	conserved hypothetical protein, secreted		
S18_920_B9_30	199	600	Ribonuclease HII	rnhB	3.1.26.4
S18_920_B9_31	460	1383	hypothetical protein		
S18_920_B9_32	232	699	lipoyltransferase	lipB	2.3.1.181
S18_920_B9_33	162	489	hypothetical protein		
S18_920_B9_34	52	159	hypothetical protein		
S18_920_B9_35	563	1692	Lysyl-tRNA synthetase	lysS	6.1.1.6
S18_920_B9_36	106	321	secreted protein		
S18_920_B9_37	76	231	hypothetical protein		
S18_920_B9_38	253	762	mannosyl-glycoprotein endo-beta-N-acetylglucosamidase family protein		
S18_920_B9_39	44	135	conserved hypothetical protein		
S18_920_B9_40	660	1983	urocanate hydratase	uroc1	4.2.1.49
S18_920_B9_41	302	909	oligopeptide ABC transporter, permease protein		
S18_920_B9_42	176	531	hypothetical protein		
S18_920_B9_43	703	2112	periplasmic tail-specific proteinase, peptidase family S41		
S18_920_B9_44	254	765	stationary phase survival protein	surE	3.1.3.5
S18_920_B9_45	68	207	membrane protein		
S18_920_B9_46	368	1107	lipid-A-disaccharide synthase, glycosyl transferase family 19		2.4.1.182

Appendix B

***Flavobacteria,* AGG58 branch 2, S18-870-C2**

ORF	AA	Nuc	Product	Gene	EC
St18_870_C2_0	380	1143	RNA methylase family UPF0020		2.1.1.-
St18_870_C2_1	392	1179	hypothetical protein		
St18_870_C2_2	293	882	hypothetical protein		
St18_870_C2_3	578	1737	sodium/glucose cotransporter		
St18_870_C2_4	161	486	Phosphoribosylaminoimidazole carboxylase	purE	4.1.1.21
St18_870_C2_5	89	270	hypothetical protein		
St18_870_C2_6		1516	16S		
St18_870_C2_7		77	tRNA		
St18_870_C2_8		77	tRNA		
St18_870_C2_9		2874	23S		
St18_870_C2_10		111	5S		
St18_870_C2_11	560	1683	oligoendopeptidase, M3 family		
St18_870_C2_12	405	1218	major facilitator superfamily transporter		
St18_870_C2_13	429	1290	hypothetical protein		
St18_870_C2_14	397	1194	dehydrogenase		
St18_870_C2_15	445	1338	Adenylosuccinate lyase	purB	4.3.2.2
St18_870_C2_16	183	552	hypothetical protein		
St18_870_C2_17	153	462	membrane protein containing DUF456		
St18_870_C2_18	564	1695	inner membrane protein translocase component YidC		
St18_870_C2_19	528	1587	CTP synthetase		6.3.4.2
St18_870_C2_20	749	2250	protein containing TonB dependent receptor domain		
St18_870_C2_21	265	798	DNA polymerase III, delta		2.7.7.7
St18_870_C2_22	923	2772	aconitate hydratase	acnB	4.2.1.3
St18_870_C2_23	153	462	conserved hypothetical protein		
St18_870_C2_24	149	450	Ribosomal protein L9		
St18_870_C2_25	406	1221	ABC transporter permease		
St18_870_C2_26	89	270	Ribosomal protein S18		
St18_870_C2_27	111	336	Ribosomal protein S6		
St18_870_C2_28	544	1635	DNA polymerase III, gamma and tau subunits		
St18_870_C2_29	980	2943	protein with similarity to peptidase M16		
St18_870_C2_30	288	867	UDP-N-acetylenolpyruvoylglucosamine reductase	murB	1.1.1.158
St18_870_C2_31	409	1230	isocitrate dehydrogenase (NADP+)		
St18_870_C2_32	42	129	hypothetical protein		
St18_870_C2_33	222	669	ABC transporter ATP-binding protein		

***Flavobacteria,* NS4, S18-1082-D10**

ORF	AA	Nuc	Product	Gene	EC
S18_1082_D10_0	99	300	hypothetical protein		
S18_1082_D10_1	213	642	Peptide methionine sulfoxide reductase A	msrA	1.8.4.11
S18_1082_D10_2	122	369	Phosphatidic acid phosphatase type 2 family protein		3.1.3.-
S18_1082_D10_3	169	510	glycosyl transferase group 1		
S18_1082_D10_4	160	483	hypothetical protein		
S18_1082_D10_5	1120	3363	protein translocase, SecA subunit	secA	
S18_1082_D10_6	73	222	conserved hypothetical protein		
S18_1082_D10_7	180	543	Cobalamin adenosyltransferase		
S18_1082_D10_8	243	732	ABC transporter		
S18_1082_D10_9	216	651	CDP-alcohol phosphatidyltransferase family protein		
S18_1082_D10_10	266	801	Sec-independent protein translocase, TatC subunit	tatC	
S18_1082_D10_11	321	966	sugar phosphate isomerase, KpsF/GutQ family protein		
S18_1082_D10_12	739	2220	ATP-dependent DNA helicase RecQ		3.6.1.-
S18_1082_D10_13	185	558	protein containing thioredoxin-like domain		1.8.4.-
S18_1082_D10_14	565	1698	ABC transporter, multidrug resistance-like ATP-binding protein		
S18_1082_D10_15	307	924	antitermination protein NusB	nusB	
S18_1082_D10_16	149	450	hypothetical protein containing DUF1573		
S18_1082_D10_17	77	234	hypothetical protein		
S18_1082_D10_18	339	1020	Dihydroorotate dehydrogenase, class 2	pyrD	1.3.3.1
S18_1082_D10_19	632	1899	acetyl-CoA synthetase	acsA	6.2.1.1
S18_1082_D10_20	360	1083	M42 family peptidase, glutamyl aminopeptidase		
S18_1082_D10_21	194	585	conserved hypothetical protein		
S18_1082_D10_22		1514	16S		
S18_1082_D10_23		77	t-RNA		
S18_1082_D10_24		74	t-RNA		
S18_1082_D10_25		2787	23S		
S18_1082_D10_26		104	5S		
S18_1082_D10_27	247	744	protein containing methyltransferase domain		2.1.1.-
S18_1082_D10_28	389	1170	Putative RNA methylase		2.1.1.-
S18_1082_D10_29	233	702	conserved hypothetical protein		
S18_1082_D10_30	157	474	conserved hypothetical protein		
S18_1082_D10_31	469	1410	23S rRNA methyltransferas		2.1.1.-
S18_1082_D10_32	431	1296	conserved hypothetical protein, secreted		

Appendix B

Flavobacteria, **DE2, S18-873-E7**

ORF	AA	Nuc	Product	Gene	EC
S18_873_E7_0	95	288	hypothetical protein		
S18_873_E7_1	342	1029	transporter, NRAMP family protein		
S18_873_E7_2	339	1020	Tetraacyldisaccharide 4'-kinase	lpxK	2.7.1.130
S18_873_E7_3	291	876	lipoic acid synthetase	lipA	2.8.1.8
S18_873_E7_4	277	834	hypothetical protein		
S18_873_E7_5	43	132	hypothetical protein		
S18_873_E7_6	258	777	protein containing DUF164		
S18_873_E7_7	143	432	protein containing stress responsive alpha-beta barrel domain		
S18_873_E7_8		75	tRNA		
S18_873_E7_9	400	1203	Folylpolyglutamate synthase	folC	6.3.2.17
S18_873_E7_10	266	801	conserved hypothetical protein		
S18_873_E7_11	131	396	ExbD/TolR family biopolymer transport protein		
S18_873_E7_12	224	675	MotA/TolQ/ExbB family biopolymer transport protein		
S18_873_E7_13	443	1332	citrate transporter		
S18_873_E7_14	342	1029	Anhydro-N-acetylmuramic acid kinase	anmK	2.7.1.-
S18_873_E7_15	283	852	glutamate synthase (NADPH) large subunit	gltA	1.4.1.13
S18_873_E7_16	1219	3660	glutamate synthase subunit alpha	gltB	1.4.7.1
S18_873_E7_17	488	1467	Glutamate synthase, NADH/NADPH, small subunit 1	gltD	
S18_873_E7_18	202	609	hypothetical protein		
S18_873_E7_19	105		5S		
S18_873_E7_20	2835		23S		
S18_873_E7_21	77		t-RNA		
S18_873_E7_22	74		t-RNA		
S18_873_E7_23	1522		16S		
S18_873_E7_24	715	2148	conserved hypothetical protein		
S18_873_E7_25	191	576	Ribonuclease HII	rnhB	3.1.26.4
S18_873_E7_26	808	2427	Ribonuclease HII		
S18_873_E7_27	235	708	lipoyltransferase	lipB	2.3.1.181
S18_873_E7_28	64	195	hypothetical protein		
S18_873_E7_29	567	1704	Lysyl-tRNA synthetase, class II		6.1.1.6
S18_873_E7_30	428	1287	Glutamate-1-semialdehyde 2,1-aminomutase	hemL	5.4.3.8
S18_873_E7_31	257	774	Mannosyl-glycoprotein endo-beta-N-acetylglucosamidase		3.2.1.-
S18_873_E7_32	309	930	Pyridoxal phosphate-dependent enzyme		
S18_873_E7_33	873	2622	Alanyl-tRNA synthetase	alaRS	6.1.1.7
S18_873_E7_34	176	531	conserved hypothetical protein		
S18_873_E7_35	325	978	M23 family peptidase		
S18_873_E7_36	36	111	hypothetical protein		
S18_873_E7_37	53	162	hypothetical protein		
S18_873_E7_38	447	1344	GTP-binding protein	engA	
S18_873_E7_39		72	tRNA		
S18_873_E7_40	295	888	GTP-binding protein	era	
S18_873_E7_41	502	1509	sulfatase		
S18_873_E7_42	80	243	hypothetical protein		
S18_873_E7_43	71	216	hypothetical protein		

***Flavobacteria,* VIS3, S18-812-F11**

ORF	AA	Nuc	Product	Gene	EC
S18_812_F11_0	257	774	Dihydrodipicolinate synthetase		4.2.1.-
S18_812_F11_1	172	519	conserved hypothetical protein		
S18_812_F11_2	253	762	5'-Nucleotidase		
S18_812_F11_3	663	1992	DNA ligase, NAD-dependent	ligA	6.5.1.2
S18_812_F11_4	302	909	5'-Nucleotidase		
S18_812_F11_5	273	822	modification methylase HemK		2.1.1.-
S18_812_F11_6	162	489	Acyl-CoA N-acyltransferase		
S18_812_F11_7	310	933	Riboflavin biosynthesis protein	ribD	1.1.1.193
S18_812_F11_8	115	348	thioesterase superfamily protein		
S18_812_F11_9	475	1428	chromosomal replication initiation protein	dnaA	
S18_812_F11_10	147	444	Protein-tyrosine phosphatase, low molecular weight		3.1.3.48
S18_812_F11_11	218	657	conserved hypothetical protein		
S18_812_F11_12	236	711	tetrapyrrole methylase family protein		2.1.1.-
S18_812_F11_13	347	1044	aminodeoxychorismate lyase		
S18_812_F11_14	261	786	Diaminopimelate epimerase	dapF	5.1.1.7
S18_812_F11_15	465	1398	protease DegQ precursor	degQ	3.4.21.-
S18_812_F11_16	190	573	glyceraldehyde 3-phosphate dehydrogenase	gap	1.2.1.12
S18_812_F11_17	225	678	tRNA (guanine-N(1)-)-methyltransferase	trmD	2.1.1.31
S18_812_F11_18	116	351	Ribosomal protein L19		
S18_812_F11_19	737	2214	isocitrate dehydrogenase, NADP-dependent	icd	1.1.1.42
S18_812_F11_20	786	2361	TonB-dependent receptor, plug		
S18_812_F11_21	268	807	conserved hypothetical protein, secreted		
S18_812_F11_22	272	819	N-acetylmuramic acid 6-phosphate etherase	murQ	4.2.-.-
S18_812_F11_23	75	228	conserved hypothetical protein, membrane		
S18_812_F11_24		107	5S		
S18_812_F11_25		2821	23S		
S18_812_F11_26		77	t-RNA		
S18_812_F11_27		77	t-RNA		
S18_812_F11_28		1516	16S		
S18_812_F11_29	646	1941	Threonyl-tRNA synthetase	thrS	6.1.1.3
S18_812_F11_30	173	522	Initiation factor 3		
S18_812_F11_31	65	198	Ribosomal protein L35		
S18_812_F11_32	114	345	Ribosomal protein L20		
S18_812_F11_33	420	1263	conserved hypothetical protein		
S18_812_F11_34	512	1539	Replicative DNA helicase	dnaC	3.6.1.-
S18_812_F11_35	317	954	Acetyl-CoA carboxylase carboxyltransferase	accA	6.4.1.2
S18_812_F11_36	294	885	hypothetical protein, transmembrane DUF6		
S18_812_F11_37	357	1074	Predicted permease YjgP/YjgQ		
S18_812_F11_38	376	1131	Queuine tRNA-ribosyltransferase		2.4.2.29
S18_812_F11_39	169	510	protein containing transketolase		

Appendix B

***Flavobacteria,* AGG58 branch 2, S18-906-E12**

ORF	AA	Nuc	Product	Gene	EC
S18_906_E12_0	293	882	glycosyl transferase group 1 domain protein		
S18_906_E12_1	942	2829	conserved hypothetical protein, membrane		
S18_906_E12_2	938	2817	carbamoyl phosphate synthase large subunit	carB	6.3.5.5
S18_906_E12_3		111	5S		
S18_906_E12_4		2867	23S		
S18_906_E12_5		77	tRNA		
S18_906_E12_6		74	tRNA		
S18_906_E12_7		1516	16S		
S18_906_E12_8	227	684	short-chain dehydrogenase/reductase family protein		
S18_906_E12_9	770	2313	lysyl endopeptidase		
S18_906_E12_10	357	1074	phenylacetic acid degradation protein	paaE	
S18_906_E12_11	262	789	enoyl-CoA hydratase	paaG	4.2.1.17
S18_906_E12_12	718	2157	conserved hypothetical proteins		
S18_906_E12_13	234	705	membrane protein		
S18_906_E12_14	381	1146	protein with similarity to RNA methylase		
S18_906_E12_15	498	1497	serine/threonine protein kinases		
S18_906_E12_16	109	330	hypothetical protein		
S18_906_E12_17	552	1659	membrane protein containing DUF819		
S18_906_E12_18	1156	3471	hypothetical protein		
S18_906_E12_19	1697	5094	secreted protein containing PKD domains		
S18_906_E12_20	565	1698	conserved hypothetical protein		
S18_906_E12_21	710	2133	dipeptidylpeptidase, peptidase S9		
S18_906_E12_22	779	2340	TonB dependent receptor domain-containing protein		
S18_906_E12_23	554	1665	CTP synthetase		

Flavobacteria, **DE2, S18-848-B2**

ORF	AA	Nuc	Product	Gene	EC
S18_848_B2_0	251	756	conserved hypothetical protein		
S18_848_B2_1	68	207	hypothetical protein		
S18_848_B2_2	195	588	conserved hypothetical protein		
S18_848_B2_3	423	1272	conserved hypothetical protein		
S18_848_B2_4	509	1530	DNA helicase, DnaB type		3.6.1.-
S18_848_B2_5	238	717	hypothetical protein		
S18_848_B2_6	420	1263	S-adenosylmethionine synthetase	metK	2.5.1.6
S18_848_B2_7	435	1308	GTP-binding protein engA	engA	
S18_848_B2_8		107	5S		
S18_848_B2_9		2769	23S		
S18_848_B2_10		74	tRNA		
S18_848_B2_11		77	tRNA		
S18_848_B2_12		1511	16S		
S18_848_B2_13	218	657	protein containing thioredoxin fold		
S18_848_B2_14	550	1653	amidohydrolase family protein		3.5.-.-
S18_848_B2_15	228	687	oxidoreductase, short chain dehydrogenase/reductase family		
S18_848_B2_16	384	1155	Putative RNA methyltransferase		
S18_848_B2_17	119	360	Translation initiation factor 1, prokaryotic		
S18_848_B2_18	239	720	O-methyltransferase, family 3		2.1.1.-
S18_848_B2_19	403	1212	Succinyl-CoA synthetase beta chain	sucC	6.2.1.5
S18_848_B2_20	216	651	Lipoprotein-releasing system ATP-binding protein; ABC tra	lolD	3.6.3.-
S18_848_B2_21	157	474	hypothetical protein		
S18_848_B2_22	160	483	putative transcriptional regulator transcriptional regulator		
S18_848_B2_23	792	2379	3-hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase/isomerase		
S18_848_B2_24	391	1176	acetyl-CoA acetyltransferase		2.3.1.9
S18_848_B2_25	599	1800	acyl-CoA dehydrogenase domain protein		1.3.99.-
S18_848_B2_26	361	1086	cellulose M, peptidase M42 family protein		
S18_848_B2_27	280	843	protein containing pirin domain		
S18_848_B2_28	126	381	two-component response regulator		
S18_848_B2_29	183	552	two-component response regulator, receiver domain		
S18_848_B2_30	732	2199	signal transduction histidine-protein kinase		
S18_848_B2_31	230	693	conserved hypothetical protein		
S18_848_B2_32	356	1071	protein containing ThiF		
S18_848_B2_33	176	531	conserved hypothetical protein		
S18_848_B2_34	211	636	conserved hypothetical protein		
S18_848_B2_35	77	234	hypothetical protein		
S18_848_B2_36	415	1248	conserved hypothetical protein		
S18_848_B2_37	852	2559	TonB-dependent receptor domain protein		
S18_848_B2_38	295	888	conserved hypothetical protein		

Appendix B

***Cytophagia,* S18-1001-F9**

ORF	AA	Nuc	Product	Gene	EC
S18_1001_F9_0	66	201	SusD/RagB family protein		
S18_1001_F9_1	440	1323	Phosphoesterase, PA-phosphatase related protein		
S18_1001_F9_2	441	1326	Xylose isomerase	xylA	5.3.1.5
S18_1001_F9_3	496	1491	Xylulose kinase	xylB	2.7.1.17
S18_1001_F9_4	560	1683	Na+/glucose symporter		
S18_1001_F9_5	53	162	hypothetical protein		
S18_1001_F9_6	35	108	hypothetical protein		
S18_1001_F9_7	41	126	hypothetical protein		
S18_1001_F9_8	307	924	Glycerophosphoryl diester phosphodiesterase	glpQ	3.1.4.46
S18_1001_F9_9	335	1008	carbohydrate kinase, PfkB family		2.7.1.4
S18_1001_F9_10	138	417	ribonuclease P protein component		
S18_1001_F9_11	545	1638	Peptidase S41A, C-terminal protease		3.4.21.102
S18_1001_F9_12	227	684	glycoprotease family, peptidase M22		
S18_1001_F9_13	171	516	conserved hypothetical protein		
S18_1001_F9_14		1513	16S		
S18_1001_F9_15		74	tRNA		
S18_1001_F9_16		77	tRNA		
S18_1001_F9_17		2871	23S		
S18_1001_F9_18		111	5S		
S18_1001_F9_19	264	795	protein with similarity to LamB		
S18_1001_F9_20	222	669	Allophanate hydrolase subunit 1		
S18_1001_F9_21	305	918	Allophanate hydrolase subunit 2		6.3.4.6
S18_1001_F9_22	337	1014	Protein of unknown function DUF1432		
S18_1001_F9_23	161	486	hypothetical protein		
S18_1001_F9_24	491	1476	Prolyl-tRNA synthetase	proS	6.1.1.15
S18_1001_F9_25	462	1389	hypothetical protein		
S18_1001_F9_26	169	510	Shikimate kinase	aroK	2.7.1.71
S18_1001_F9_27	519	1560	hypothetical protein		
S18_1001_F9_28	361	1086	conserved hypothetical protein, membrane		
S18_1001_F9_29	179	540	protein containing DUF1599		
S18_1001_F9_30	257	774	Dihydropteroate synthase	folP	2.5.1.15
S18_1001_F9_31	277	834	Conserved hypothetical protein CHP00159		
S18_1001_F9_32	397	1194	2-amino-3-ketobutyrate coenzyme A ligase	kbl	2.3.1.29
S18_1001_F9_33	311	936	L-threonine dehydrogenase		
S18_1001_F9_34	260	783	Enoyl-CoA hydratase/isomerase		
S18_1001_F9_35	316	951	protein with similarity to polysaccharide biosynthesis protein		
S18_1001_F9_36	81	246	hypothetical protein		

S18-1049-858M	**Sphingobacteria C, S18-1049-A9**					
ORF merged	**ORF**	**AA**	**Nuc**	**Product**	**Gene**	**EC**
S18_1049_858M_0	S18_1049_A9_0	99	300	hypothetical protein		
S18_1049_858M_1	S18_1049_A9_1	422	1269	fumarylacetoacetase	fahA	3.7.1.2
S18_1049_858M_2	S18_1049_A9_2	100	303	hypothetical protein		
S18_1049_858M_3	S18_1049_A9_3	260	783	conserved hypothetical protein CHP02757		
S18_1049_858M_4	S18_1049_A9_4	239	720	conserved hypothetical protein		
S18_1049_858M_5	S18_1049_A9_5	236	711	conserved hypothetical protein, membrane		
S18_1049_858M_6	S18_1049_A9_6	415	1248	conserved hypothetical protein UCP033563		
S18_1049_858M_7	S18_1049_A9_7	351	1056	Phosphoserine aminotransferase	serC	2.6.1.52
S18_1049_858M_8	S18_1049_A9_8	58	177	hypothetical protein		
S18_1049_858M_9	S18_1049_A9_9		1505	16S		
S18_1049_858M_10	S18_1049_A9_10		74	tRNA		
S18_1049_858M_11	S18_1049_A9_11		77	tRNA		
S18_1049_858M_12	S18_1049_A9_12		2734	23S		
S18_1049_858M_13	S18_1049_A9_13		107	5S		
S18_1049_858M_14	S18_1049_A9_14	156	471	conserved hypothetical protein DUF458, RNase H-like		
S18_1049_858M_15	S18_1049_A9_15	115	348	Ribosomal protein L31		
S18_1049_858M_16	S18_1049_A9_16	248	747	protein similar to formamidase		
S18_1049_858M_17	S18_1049_A9_17	375	1128	oxygen-independent coproporphyrinogen III oxidase	hemF	1.3.99.-
S18_1049_858M_18	S18_1049_A9_18	220	663	membrane protein		
S18_1049_858M_19	S18_1049_A9_19	252	759	uroporphyrinogen-III synthase	hemD	4.2.1.75
S18_1049_858M_20	S18_1049_A9_20	330	993	hypothetical protein		
S18_1049_858M_21	S18_1049_A9_21	439	1320	gliding motility protein GldK	gldK	
S18_1049_858M_22	S18_1049_A9_22	274	825	gliding motility-related protein GldL	gldL	
S18_1049_858M_23	S18_1049_A9_23	585	1758	conserved hypothetical protein, with similarity to GldM		
S18_1049_858M_24	S18_1049_A9_24	290	873	conserved hypothetical protein, with similarity to GldO		
S18_1049_858M_25	S18_1049_A9_25	590	1773	UvrABC system protein C	uvrC	
S18_1049_858M_26	S18_1049_A9_26	772	2319	transglycosylase/transpeptidase penicillin-binding protein		2.4.2.-
S18_1049_858M_27	S18_1049_A9_27	856	2571	protein containing tetratricopeptide repeat domain		
S18_1049_858M_28	S18_1049_A9_28	57	174	hypothetical protein		
S18_1049_858M_29	S18_1049_A9_29	455	1368	acetyl-CoA carboxylase biotin carboxylase subunit	accC	6.3.4.14
S18_1049_858M_30	S18_1049_A9_30	164	495	Acetyl-CoA biotin carboxyl carrier		6.4.1.2
S18_1049_858M_31	S18_1049_A9_31	188	567	translation elongation factor P		
S18_1049_858M_32	S18_1049_A9_32	340	1023	3-oxoacyl-(acyl carrier protein) synthase III	fabH	2.3.1.41
S18_1049_858M_33	S18_1049_A9_33	178	537	conserved hypothetical protein		
S18_1049_858M_34	S18_1049_A9_34	343	1032	pyridoxal phosphate biosynthetic protein	pdxA	1.1.1.262
S18_1049_858M_35	S18_1049_A9_35	477	1434	leucyl aminopeptidase, peptidase M17		
	S18-858-A8					
S18_1049_858M_36	S18_858_A8_32	260	783	dimethyladenosine transferase		2.1.1.-
S18_1049_858M_37	S18_858_A8_33	314	945	phosphoglycerate dehydrogenase	serA	1.1.1.95
S18_1049_858M_38	S18_858_A8_34	816	2451	TonB-dependent receptor, plug domain		

Appendix B

S18-1013-1087M ***Flavobacteria, Leeuwenhoekiella,*** **S18-1013-C10**

ORF merged	ORF	AA	Nuc	Product	Gene	EC
S18_1013_1087M_0	S18_1013_C10_0	113	342	transglutaminase-like domain protein		
S18_1013_1087M_1	S18_1013_C10_1	152	459	acetyltransferase		
S18_1013_1087M_2	S18_1013_C10_2	77	234	hypothetical protein		
S18_1013_1087M_3	S18_1013_C10_3	382	1149	MscS Mechanosensitive ion channel		
S18_1013_1087M_4	S18_1013_C10_4	130	393	conserved hypothetical protein		
S18_1013_1087M_5	S18_1013_C10_5	279	840	branched-chain-amino-acid aminotransferase	ilvE	2.6.1.42
S18_1013_1087M_6	S18_1013_C10_6	186	561	hypothetical protein		
S18_1013_1087M_7	S18_1013_C10_7	90	273	Histone-like bacterial DNA-binding protein		
S18_1013_1087M_8	S18_1013_C10_8	319	960	Methionyl-tRNA formyltransferase	fmt	2.1.2.9
S18_1013_1087M_9	S18_1013_C10_9	635	1908	ATP-dependent DNA helicase recQ	recQ	3.6.1.-
S18_1013_1087M_10	S18_1013_C10_10	180	543	conserved hypothetical protein		
S18_1013_1087M_11	S18_1013_C10_11	96	291	conserved hypothetical protein		
S18_1013_1087M_12	S18_1013_C10_12	224	675	conserved hypothetical protein		
S18_1013_1087M_13	S18_1013_C10_13	437	1314	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	murA	2.5.1.7
S18_1013_1087M_14	S18_1013_C10_14	173	522	5'(3')-deoxyribonucleotidase		
S18_1013_1087M_15	S18_1013_C10_15	99	300	conserved hypothetical protein, membrane		
S18_1013_1087M_16	S18_1013_C10_16	228	687	ABC transporter		
S18_1013_1087M_17	S18_1013_C10_17	434	1305	deoxyribodipyrimidine photolyase		
S18_1013_1087M_18	S18_1013_C10_18	232	699	short-chain dehydrogenase/reductase SDR		

S18-1013-1087M	*Flavobacteria, Leeuwenhoekiella,* S18-1013-C10 Continued					
ORF merged	**ORF**	**AA**	**Nuc**	**Product**	**Gene**	**EC**
	S18-1013-F8					
S18_1013_1087M_19	S18_1087_F8_45	169	510	conserved hypothetical protein		
S18_1013_1087M_20	S18_1087_F8_44	279	840	conserved hypothetical protein, secreted		
S18_1013_1087M_21	S18_1087_F8_43	323	972	protein containing oxidoreductase domain		
S18_1013_1087M_22	S18_1087_F8_42	575	1728	sodium iodide symporter		
S18_1013_1087M_23	S18_1087_F8_41	832	2499	conserved hypothetical protein		
S18_1013_1087M_24	S18_1087_F8_40	193	582	Maf-like protein		
S18_1013_1087M_25	S18_1087_F8_39	303	912	hydroxybenzoate octaprenyltransferase		
S18_1013_1087M_26	S18_1087_F8_38	244	735	conserved hypothetical protein		
S18_1013_1087M_27	S18_1087_F8_37	577	1734	membrane protein containing PspC domain		
S18_1013_1087M_28	S18_1087_F8_36	114	345	transcriptional regulator PadR-like		
S18_1013_1087M_29	S18_1087_F8_35	169	510	conserved hypothetical protein, membrane		
S18_1013_1087M_30	S18_1087_F8_34	143	432	conserved hypothetical protein		
S18_1013_1087M_31	S18_1087_F8_33	266	801	protein containing DUF124		
S18_1013_1087M_32	S18_1087_F8_32	65	198	conserved hypothetical protein		
S18_1013_1087M_33	S18_1087_F8_31	316	951	transcriptional regulator, LysR family protein		
S18_1013_1087M_34	S18_1087_F8_30	54	165	hypothetical protein		
S18_1013_1087M_35	S18_1087_F8_29		108	16S		
S18_1013_1087M_36	S18_1087_F8_28		74	tRNA		
S18_1013_1087M_37	S18_1087_F8_27		77	tRNA		
S18_1013_1087M_38	S18_1087_F8_26		2834	23S		
S18_1013_1087M_39	S18_1087_F8_25		1514	5S		
S18_1013_1087M_40	S18_1087_F8_24	142	429	conserved hypothetical protein		
S18_1013_1087M_41	S18_1087_F8_23	208	627	hemolysin III family family protein, transmembrane		
S18_1013_1087M_42	S18_1087_F8_22	222	669	ZIP family metal transporter		
S18_1013_1087M_43	S18_1087_F8_21	388	1167	RNA methylase family UPF0020		
S18_1013_1087M_44	S18_1087_F8_20	249	750	dipeptidyl peptidase VI		
S18_1013_1087M_45	S18_1087_F8_19	391	1176	acetoacetyl-CoA thiolase		2.3.1.9
S18_1013_1087M_46	S18_1087_F8_18	681	2046	transmembrane HD family protein		
S18_1013_1087M_47	S18_1087_F8_17		78	t-RNA		
S18_1013_1087M_48	S18_1087_F8_16		87	t-RNA		
S18_1013_1087M_49	S18_1087_F8_15	417	1254	putative transposase		
S18_1013_1087M_50	S18_1087_F8_14	143	432	hypothetical protein		
S18_1013_1087M_51	S18_1087_F8_13	278	837	conserved hypothetical protein		
S18_1013_1087M_52	S18_1087_F8_12	424	1275	conserved hypothetical protein		
S18_1013_1087M_53	S18_1087_F8_11	51	156	hypothetical protein		
S18_1013_1087M_54	S18_1087_F8_10	46	141	hypothetical protein		
S18_1013_1087M_55	S18_1087_F8_9	45	138	hypothetical protein		
S18_1013_1087M_56	S18_1087_F8_8	118	357	single-strand binding protein		
S18_1013_1087M_57	S18_1087_F8_7	126	381	conserved hypothetical protein		
S18_1013_1087M_58	S18_1087_F8_6	65	198	conserved hypothetical protein		
S18_1013_1087M_59	S18_1087_F8_5	50	153	hypothetical protein		
S18_1013_1087M_60	S18_1087_F8_4	331	996	transposase IS116/IS110/IS902 family protein		
S18_1013_1087M_61	S18_1087_F8_3	126	381	conserved hypothetical protein		
S18_1013_1087M_62	S18_1087_F8_2	343	1032	protein with similarity to mobilisation protein B		
S18_1013_1087M_63	S18_1087_F8_1	401	1206	protein with similarity to mobilisation protein C		
S18_1013_1087M_64	S18_1087_F8_0	99	300	hypothetical protein		

Acknowledgements

And now is time to say thanks to all the people that have accompanied and supported me during this journey....

Thanks to...

Rudi Amann for giving me the opportunity to join MarMic and to do this work in your group. I also want to thank you for always finding some time for discussions.

Bernhard Fuchs for your daily supervision. Thanks you for your good mood, for always have your door open for discussions, and for encourage me. I have learned a lot from you and it has been a great time!

Victor Smetacek thanks for accepting reviewing this thesis. A very special thanks for inspiring me with your "big picture" views since the first MarMic lectures.

Thanks Prof. Fischer for accepting being part of my thesis committee.

Marga Schüler for all the long phone conversations about *Flavobacteria* and a very big thank for your endless good spirit!

Cecilia Alonso for letting me working with you in the Río de la Plata project, but a much bigger thanks for always telling me that this was possible and I was going to be alright.

Valerie Barbe and her colleagues at Genoscope for the *Bacteroidetes* metagenome sequencing.

Matt Oliver, for trying to explain me what all the satellite data means. Gunnar

Gerdts for the statistical advices. Fransesca Simonato for letting me join you in the Venice Lagoon project.

The Molecular Ecology group, I had a great time working here!. Very specially, I want to thank the help of Jörg Wulf and Anke Meyerdierks. Daniel and Marc for trying to correct my "spanglish" and Jill for the Latex tips. A big thank to all the bioinformatic guys, specially to Sixing Huang, Christian Quast, Michael Richter, and Hanno Teeling for all the support.

My students Abdul Sheik, Cyrielle Viaut, Andreas Krupke, Alessandra Gallina, and Christin Bennke. Thanks all of you for your hard work, but more important for teaching me how to be patient!

Christiane Glöckner, since the first MarMic day you have been taking great care of me.

My friends, you have been a fundamental part of the last years. To all my MarMic class for the great company since those early days in Sylt. Alex, Astrid, Ivo, Joanna, Lars, Luciana, Mel, Mohamed, Sandra (the marmics 11); Amelia, Alberto, Angie, Daniel, Fra, Ila, Marina, Pablo, Sergio, thank for your friendship!!!

Ευχαριστώ πάρα πολύ Στέφανε

Y gracias a mi familia y a mis amigas, quienes a pesar de la gran distancia de un Océano, y sin entender a veces que hago exactamente, me han apoyado incondicionalmente y siempre, siempre están ahí cuando las necesito.

Mama y Papa, sin su apoyo y su creencia en mi no seria nada y jamás hubiese llegado hasta aquí...

Thanks! Danke! Gracias!

Erklärung

Gemäß §6 (5) Nr. 1 - 3 Promotionsordnung erkläre ich hiermit, dass ich die Arbeit mit dem Titel:

Marine *Bacteroidetes*: distribution patterns and role in the degradation of organic matter

I ohne unerlaubte fremde Hilfe angefertigt habe,

II keine anderen, als die von mir angegebenen Quellen und Hilfsmittel benutzt habe,

III die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

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

Paola Rocio Gómez Pereira

Bremen,