UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Institut de Ciències del Mar

Uncovering Diversity and Structure of BACTEROIDETES POPULATIONS in Marine Environments

Cristina Díez Vives

Barcelona
Octubre 2013

UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Institut de Ciències del Mar
CSIC
Consejo Superior de Investigaciones Científicas

Uncovering Diversity and Structure of Bacteroidetes Populations in Marine Environments

(Descubriendo la Diversidad y Estructura de las poblaciones de Bacteroidetes en Ambientes Marinos)

Cristina Díez Vives

Tesis Doctoral presentada por Dª Cristina Díez Vives para obtener el grado de Doctor por la Universidad de las Palmas de Gran Canaria.

Directores: Dra. Silvia G. Acinas y Dr. Josep M. Gasol

Universidad de las Palmas de Gran Canarias

Institut de Ciències del Mar (ICM-CSIC)

DOCTORADO EN OCEANOGRAFÍA
Bienio 2010-2012. Con Mención de Calidad en la ANECA.

En Barcelona a, 10 de Octubre de 2013

La Doctoranda	El Director	El Co-director
Cristina Díez Vives	Silvia G. Acinas	Josep M. Gasol

Esta tesis se ha desarollado con el apoyo de una beca predoctoral del CSIC (JAE Intro). El estudio experimental se ha realizado en el Instituto de Ciencias del Mar (ICM-CSIC) y se ha financiado parcialmente por el Ministerio Español de Ciencia e Innovación mediante los proyectos MICRODIVERSITY (CGL2008-00762/BOS) y MicroOcean PANGENOMICS (CGL2011-26848/BOS) conseguidos por SGA.

A mis padres,
a mi abuela,
y a mi hermano.

AGRADECIMIENTOS

Uno siempre teme el momento de escribir los agradecimientos, básicamente por todo lo que ello implica. Significa que sí, que el final ha llegado, que esto parece que se termina, por mucho que lo hayas dudado por el camino. Significa, también, mirar atrás y darse cuenta de cómo han cambiado las cosas, de cuanta gente ha entrado y salido de nuestras vidas en estos años; incluso cómo hemos cambiado nosotros mismos desde que empezamos. El entorno de la ciencia no es estático, sino todo lo contrario, da vértigo recordar todas las experiencias que llegamos a vivir los becarios: los congresos, los viajes, muestreos, estancias, *meetings*, vuelos, vuelos y más vuelos. Simplemente, me siento muy afortunada por ello y aquí quiero expresar mi más sincero agradecimiento a todos los que, de un modo u otro, habéis contribuido a realizar esta aventura.

Primero de todo, agradecer a mis directores de tesis la oportunidad de haber entrado en su grupo y poder formar parte de este "estilo de vida". A Silvia G. Acinas por sus ideas, consejos, correcciones y sobretodo por haberme "transmitido horizontalmente", ya desde el principio, la importancia de la divulgación, de comunicar lo que hacemos pero también de asistir a las presentaciones de otras personas, algo que yo tanto rechazaba al principio y de lo que acabé disfrutando y aprendiendo más adelante. A Pep Gasol, que tanto se resistió al principio a hacerme un poquito de caso, pero que lo ha dado todo en este último "sprint" para cumplir con "mis prisas" por acabar. Muchísimas gracias por todo, y sobretodo por el esfuerzo que hiciste al final con correcciones solapadas y trabajando hasta las tantas, pero sin perder nunca el humor y apoyando hasta el final. Un verdadero gusto haber trabajado contigo.

Gracias también al resto de "jefazos": a Marta R, Rafel C, Albert C, Ramón M, Celia M por horas de conversaciones en la terraza; y especialmente a Carles P por haber participado "en la sombra" en casi todas las discusiones y correcciones de los trabajos.

Luego uno empieza a pensar en la gente a la que quiere agradecer cosas y dejar aquí sus nombres escritos para siempre, y te das cuenta de que sería imposible nombrarlos a todos. En unos pocos años en el ICM uno puede llegar a conocer más gente interesante de la que jamás habría imaginado. El primer día que llegué aquí me esperaban Bea F y Lorena E en el despacho, quienes me acogieron y cuidaron desde el primer momento. Hoy, todavía compartiendo despacho con mi querida Bea (que suerte que haya sido así!), pero también con Elena L, Ana G, Pedro L y Rosana A, me siento muy afortunada de haberos conocido y compartido juntos tantísimas horas de conversaciones, charlas, desesperación, alegrías, etc. Muchas gracias a todos por el agradable ambiente que hemos compartido, y especialmente a Rosana por haber escuchado pacientemente todos mis lloros en este último tiempo, ¡gracias guapa!, me has ayudado muchismo.

También quiero recordar a personas que pasaron por aquí un tiempo, pero dejaron una bonita huella en su paso, a Verónica con la que pasé el mejor primer año de tesis posible (¡qué triste cuando marchaste!). A Sergio, Tina y Oliver, por unos preciosos meses "a la europea". A otros compañeros que terminaron y han seguido sus caminos como Itziar, Clara R, Juancho, Thomas, Martí, Ero. A las técnicos que mantienen el orden seamos 5 o 500 en el laboratorio y con los pedidos, gracias Vane, Clara C, Irene, Paula y Eli.

A los compañeros que seguimos por aquí, siempre ocupados, siempre contando el tiempo que nos queda, incluso desde el primer año que llegamos. A Massimo, que será mi sucesor, a Guillem, Fran C, Ana Maria, Raquel, Ivo, Ramiro, la reciente doctora y compañera de poyata Montse, a Pablo por la ayuda prestada en este final y futuros artículos compartidos. También a compañeros de comidas y fiestas, a Fran A, Mireia, Carolina, Matina, Isabel, Sdena, Encarna, Estela, Rachele y a los más "festeros" Roy y Rodrigo ¡madre mía! qué últimos mesecitos habéis pasado, y he podido compartir a ratos con vosotros (menos de los que hubiera querido). A los compis de mi grupo paralelo de esponjas, Teresa, Johannes, Miguel, Eugeni y Nuria, ¡qué gran equipo!. Y a mis inesperados compañeros de piso, "Peter" y Sergio, gracias por acogerme, ha sido fantástico un cambio de aires con vosotros, y ¡qué bien se está en el pisito, hacedme hueco cuando venga de visita!

Quiero acordarme también de mis roommates Karen y Meghan, quienes convirtieron una estancia en el extranjero en maravillosa al introducirme en el mundo de la escalada. Y que derivó en conocer a mis compañeros de aventuras en Barcelona, los "tolokos", gracias por tantísimas horas de escalada, alpinismo, sky, etc. A Laura, Arantxa, Nil, Mireia, Jordi, Natalia, Ali, Nuria, Javi, a Ferran "eres la caña, chà", y a David por darme la oportunidad de vivir todas esas experiencias.

A mis amigos de Castellón, que inevitablemente se han ido alejando de mi vida, pero que formaron una enorme parte en su momento, como Vicent, Efrén, Julio, Pere, Valentín y, sobre todo, Javi R, por haber estado y seguir siempre preocupándote de mi. También a Isa C, la más bella bailarina (¡y coreógrafa!), y fantástica diseñadora que he conocido ;P Gracias por la portada y el diseño general, pero muchísimas más gracias por haberme ayudado en los peores momentos y sufrirlos conmigo, gracias guapísima.

Y por supuesto, agradecer a las personas que han sido más cercanas a mi durante todos estos años de tesis. A Jose M, quien desde que llegó hizo cambiar mi forma de ver la ciencia a un modo más entusiasta, y quien me hizo sentir incluso más afortunada de pertenecer a este "mundillo". Gracias por infinitos días compartidos, por tus incansables consejos y apoyo, por tu ayuda en el trabajo y, sobretodo, por haberme hecho formar parte de tu grupo, en el que tan a gusto me he encontrado siempre y del que espero

seguir formando parte en el futuro.

A Xavi L ¡aisshh Xavi!, no hay hojas suficientes para agradecértelo todo. Tu incondicional compañía, tus esfuerzos por sacarme de mil y un “marrones” en los que tan fácilmente me meto (y no me preguntes como lo hago). Mil gracias por tus consejos, por aguantarme horas y horas de quejas y llantos, por estar siempre ahí cuando te he necesitado. Eres una persona que todo el mundo nombra en sus tesis porque te llevas bien con todos, pero en mi caso el agradecimiento es muy especial: sin tí, el tiempo en Barcelona no hubiera sido ni parecido. Gracias, gracias, gracias.

Y por último, a la persona más importante de esta tesis: a Oswaldo P. Tú me enseñaste la ciencia cuando yo ni siquiera la consideraba una opción. Ya en la Universidad me mostraste qué era hacer investigación, con su necesario rigor, sacrificio, cansancio, incluso desesperación, pero también la enorme satisfacción de hacer lo que nos gusta. Además, me enseñaste desde el primer momento cómo hacer bien las cosas: no solo a hacerlas, uno no puede ser vago en ésto, tiene que hacerlo lo mejor que pueda. Ya sabes que gracias a ti estoy yo aquí. Sin ti, sin tu orientación en mis primeros pasos llenos de ilusión, sin luego tus consejos de seguir adelante cuando uno quiere tirar varias veces la toalla por el camino, y desde luego sin tu constante y sacrificada ayuda día a día, esto no hubiera sido posible. Qué más puedo decirte que, todo lo que soy y llegue a ser, es gracias a tí, que has sido mi “master” y que eres la persona más importante en mi vida.

Por supuesto, quiero acabar agradeciendo a mi familia toda su comprensión. A mi hermano Alberto, por entenderme y escucharme con tanto entusiasmo cuando le hablaba de mi trabajo, de resultados, proyectos y demás, gracias por esas largas charlas y por estar siempre ahí interesado. A mis padres y abuela, puessss, yo no diría que me hayan apoyado desde el principio y esas cosas, porque más bien han tardado unos añitos en ver que la niña no estaba tan descarrilada como parecía. ¡Madre mía! ¡ cuanto tiempo me ha llevado que consigan valorar que mirar bichitos en el mar también es importante! Pero con gran esfuerzo y paciencia por ambas partes, yo diría que lo han entendido y hasta han acabado estando orgullosos de mi. Un enorme ¡Gracias! a todos por ser quienes sois, y ser mi familia.

CONTENTS

SUMMARY

One of the main questions in microbial ecology is to understand what are the processes that structure and govern the species composition of communities. Answering to this question implies the need to detect bacterial species in environmental samples or at least to define ecologically meaningful "units of differentiation". The dynamics and distribution of different broad taxonomic groups that constitute the bacterioplankton (at the phylum or class level) has been widely studied. The different large groups are known to follow different spatial distributions and seasonal cycles, but important variations also occur at lower levels (i.e. species or ecologically distinct populations). In this thesis we aimed to study some of these aspects for an important group of marine bacteria: the phylum Bacteroidetes.

First at all, we evaluated the existing molecular tools in an updated phylogenetical and methodological context. Strengths and weaknesses of different probes used in different techniques were thoroughly studied and discussed, and alternatives were designed for some specific purposes related with the study of this group in marine environmental samples. Standing on these tools, we studied the phylum at different levels of resolution resulting from assorted molecular approaches. We related entire phylum abundances with environmental parameters as they vary spatially and temporally in the oceanic realm. We identified the most abundant phylotypes forming the Bacteroidetes assemblages and their ecological variability. Both the relative abundance and the changes in the phylotypes constituting the community followed clear spatial and seasonal patterns, which were thoroughly described.

To increase the resolution of our study, we built extensive clone libraries including the 16S-ITS-23S ribosomal operon. The phylogenetic study of these sequences from different environmental samples showed that closely related phylotypes were organized in monophyletic clades of habitat-specific sequences. This confirmed that similar bacteria (in terms of the 16S rRNA) seem to present different physiologies, which would drive their distribution and adaption to specific ecological niches. The study of the distributions of these clades would help to understand the ecology of this group.

Finally, we integrated the above information to design several clade-specific qPCR primer sets targeting putative Bacteroidetes ecotypes. These primers were thoroughly tested as the best possible compromise between specificity and performance for the intended purpose, and they were extensively evaluated *in silico*. An optimization of the qPCR technique workflow and a demonstration of its usefulness was carried out using the NS2b group. These primers offer a toolbox for hierarchical quantitative studies characterizing Bacteroidetes populations in different environments. Such studies should help to explore and corroborate links of certain groups with specific habitats and environmental variables.

General Introduction

1. The marine environment

1.1. A short description of the marine environment

Oceans occupy the vast majority, approximately 70%, of the Earth's surface and they have a profound effect on all biogeochemical cycles and participate in climate regulation. The oceans contain different habitats characterized by depth, light or temperature. The sea bottom (benthic province) is clearly different from the water column (pelagic province), which is subdivided into discrete zones, largely as a function of water depth. Together with the penetration of sunlight, depth partitions water column into several zones. On the basis of light, the water column is separated into a well-illuminated photic zone, a "twilight" zone, and a totally dark aphotic zone [Pinet, 2009]. The water layer in which temperature changes rapidly with depth is known as the thermocline. In temperate regions, seasonal thermoclines establish during the summer time, with higher temperature in the surface layer.

Seawater shows a remarkable stable chemical composition. Beside water molecules (H_2O), the two major ions are Na^+ and Cl^- that confer salinity to seawater. With an average 35‰ by weight, salinity decreases with the incorporation of freshwater (i.e. through rain or snow precipitations, ice melting or river inflow) and increases at the ocean surface through evaporation. The combined effect of temperature, pressure and salinity determines water density, which in turn greatly influences vertical water movements. As a general rule, nutrients are scarce in the open ocean and more abundant in oceanic upwelling regions as well as in coastal areas. With low concentrations of biologically essential elements (i.e. phosphorous (P), nitrogen (N) and silica (Si), including labile organic carbon), the surface seawater can be defined as an oligotrophic (relatively nutrient-poor) environment.

On top of the coastal *vs.* open-ocean division of the pelagic realm, latitude affects the way in which water stratifies or mixes, as a consequence of the amount of solar energy reaching the oceans. Polar oceans show a strong seasonal signal, with winter and summer stratification, and a primary production lasting most of the summer, while tropical oceans might be permanently stratified, and have a relatively low and constant chlorophyll value. Temperate seas stratify in summer and are mixed during winter. A pulse of production in late winter/early summer, with a second smaller pulse in the autumn is a characteristic of this type of ecosystem.

For the above reasons, and even though the sea may seem rather homogeneous

from a superficial point of view, marked physical and chemical discontinuities throughout the vast oceanic extensions determine biotic areas in which diverse microbial communities dwell, highly adapted to their ecological niches defined by light, nutrients, and water physical properties.

1.2. The Mediterranean Sea

The Mediterranean Sea is a large (2,969,000 km^2) and deep (average 1,460 m, maximum 5,267 m) enclosed sea placed in the temperate region. Its basin is almost completely enclosed by land, which affects some of its oceanographic features (e.g., tides are very limited as a result of the narrow connection with the Atlantic Ocean). Evaporation greatly exceeds precipitation and river runoff in the Mediterranean, and is higher in its eastern half, causing the water level to decrease and salinity to increase from west to east. The resulting pressure gradient pushes relatively cold, lower-salinity water from the Atlantic across the Mediterranean basin (Fig. 1). This water warms up to the east, where it becomes saltier and then sinks in the Levantine Sea, before circulating west and exiting through the Strait of Gibraltar [Pinet, 2009]. The Mediterranean has narrow continental shelves and a large area of open sea. Uncommon features of the Mediterranean are the high homothermy from 300–500 m to the bottom, where temperatures vary from 12.8°C – 13.5 °C in the western basin to 13.5 °C – 15.5°C in the eastern, and the high salinity, with values of 37.5 – 39.5 ‰ [Emig & Geistdoerfer, 2005].

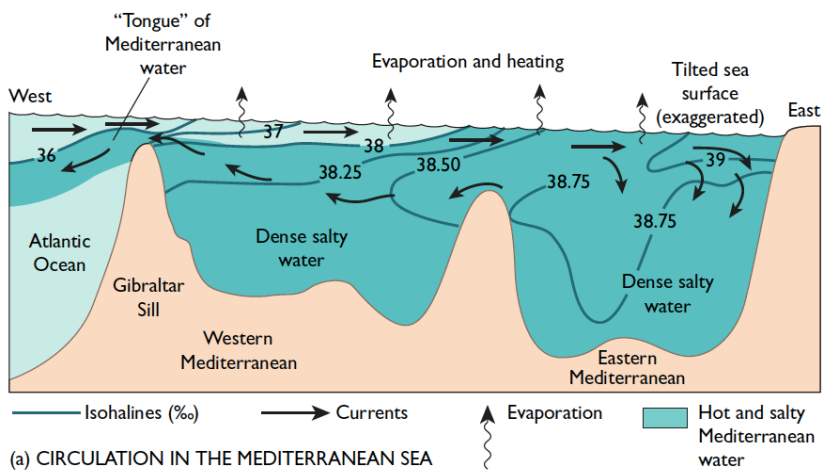


Figure 1. Circulation in the Mediterranean Sea. Warm, salty water sinks because of its high density, filling the deeps and spilling over the Gibraltar sill into the North Atlantic Ocean. This water is replaced by the surface inflow of seawater from the Atlantic Ocean (from Pinet, 2009).

The annual mean sea surface temperature shows high seasonality. Low nutrients and chlorophyll a (Chl *a*) pools rank the basin as oligotrophic to ultraoligotrophic [Krom *et al.,* 1991; Antoine *et al.,* 1995]. This is mainly due to the very low concentration of inorganic phosphorus, which is known to limit primary production [Thingstad & Rassoulzadegan, 1995, 1999; Thingstad *et al.,* 2005]. This limitation may somehow be buffered by local inputs from highly populated coasts and from the atmosphere, particularly by depositions of Saharan dust. Primary production decreases from north to south and west to east [Danovaro *et al.,* 1999].

Low values of chlorophyll (less than 0.2 μg chl *a* l^{-1}) are present over large areas [Siokou-Frangou *et al.,* 2010]. In the Catalan-North Balearic frontal area an exception occurs with the phytoplankton blooms observed throughout the late winter and early spring [Estrada *et al.,* 1985; Lévy *et al.,* 1998]. A deep chlorophyll maximum (DCM), generally not exceeding 1.5 μg *chl* a l^{-1}, is a permanent feature for the whole basin over most of the annual cycle. The strong seasonality ruling the basin favors the alternation of phytoplankton populations dominated by different functional groups and species [Siokou-Frangou *et al.,* 2010].

In this thesis we have mostly worked with samples obtained in a coastal - open ocean transect in the NW Mediterranean area. This transect was sampled during cruise Modivus, in September 2007, and it included 5 sampling points where vertical profiles were studied. At the most offshore station down to 2,000 meters. September is still a summer stratification period and for that reason we obtained samples including many different habitats where we would expect to find different thriving populations: the coast, the open surface waters, the DCM, the meso- and bathypelagic waters, etc. We tried to maximize the range of habitats studied.

Seasonality was targeted with the study of samples from the Blanes Bay microbial observatory (BBMO). This is a coastal site with an average chlorophyll concentration and phytoplankton productivity quite low for a coastal site [Gasol *et al.,* 2012], that shows a very strong seasonal signal reflected in primary production, chlorophyll, seasonality in the development of cyanobacterial populations, archaeal groups [Galand *et al.,* 2010], aerobic anoxygenic phototrophic prokaryotes [Ferrera *et al.,* 2013] and many other bacterial groups [G. Salazar, unpublished]. Samples are routinely taken at this site since year 1997, and we profited from this sampling scheme to obtain samples throughout the year that represent different ecological conditions and, presumably, different niches for the growth of different types of marine bacteria.

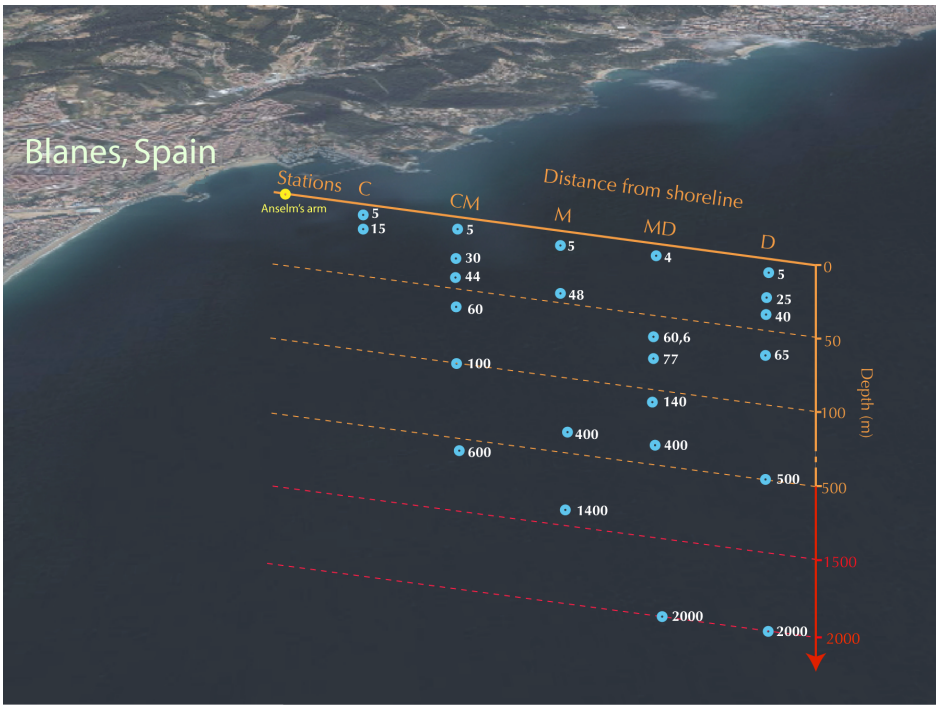


Figure 2. Sampling scheme at the area of study, showing the surface and vertical profiles covered in the Cruise Modivus. The yellow dot represents the coastal surface sampling point for the temporal studies at the BBMO. The blue dots represent the sampled depths in the spatial study (image by Thomas Pommier).

1.3. Marine microorganisms

Microorganisms are not visible to the naked eye. Despite their small size, the total mass of bacteria in the oceans exceeds the combined mass of zooplankton and fish [Pomeroy *et al.,* 2007]. By now, it is clear that microorganisms are the "unseen majority" with $35 \cdot 10^{28}$ cells in the ocean surface [Whitman *et al.,* 1998]. The oceans harbor numerous and diverse groups of viruses, protists, Bacteria and Archaea that carry out the largest fraction of the biological activity in the ocean. Several of these microbes are autotrophs, and some derive energy from photosynthesis. Photosynthetic organisms, eukaryotic phytoplankton and *Cyanobacteria*, convert inorganic carbon (CO_2) to organic carbon, as biomass.

In the open ocean, *Cyanobacteria* are responsible for about 80% of the total oceanic primary production, which represents about half of the planet's total primary production [Falkowski *et al.,* 1998; Field, 1998]. The phytoplankton cells release a large fraction of their daily photosynthesis production as DOM (Dissolved Organic Matter). The substances released include, among others, carbohydrates (mono-, oligo- and polysaccharides), and nitrogenated compounds (amino acids, peptides and proteins) [Hellebust, 1965; Fogg, 1983; Myklestad, 1995]. Processes involving other organisms also have a prominent role in the generation of oceanic DOM [Nagata & Kirchman, 1999], such as sloppy feeding and excretion by protozoan grazers, or virally-induced cell lysis [Nagata, 2000; Suttle, 2005]. 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 released DOM.

DOM supports the growth of heterotrophic microorganisms [Ducklow & Carlson, 1992] that oxidize approximately one-half of the labile, rapidly recycled fraction of the carbon fixed by photosynthesis [Azam *et al.,* 1983; Fenchel, 1988; Robinson & Williams, 2005], although this proportion can be significantly higher in oligotrophic waters [del Giorgio *et al.,* 1997]. As a consequence of heterotroph's growth, carbon dioxide is released to the ocean and essential inorganic nutrients are recycled into the water column. A fraction of the DOM is assimilated by bacteria and re-introduced into the food web through grazing of bacteria by protozoans (i.e. mostly flagellates and ciliates), which are later consumed by larger zooplankton and subsequently consumed by metazoans. This derivation through bacteria and protists limits organic matter loses via sedimentation and deep burial. This trophic pathway of DOM, know as the "microbial loop" [Pomeroy, 1974; Azam *et al.,* 1983; Ducklow*,* 1983], is crucial to the marine trophic web because essential inorganic nutrients are recycled through bacterial metabolism into the water column.

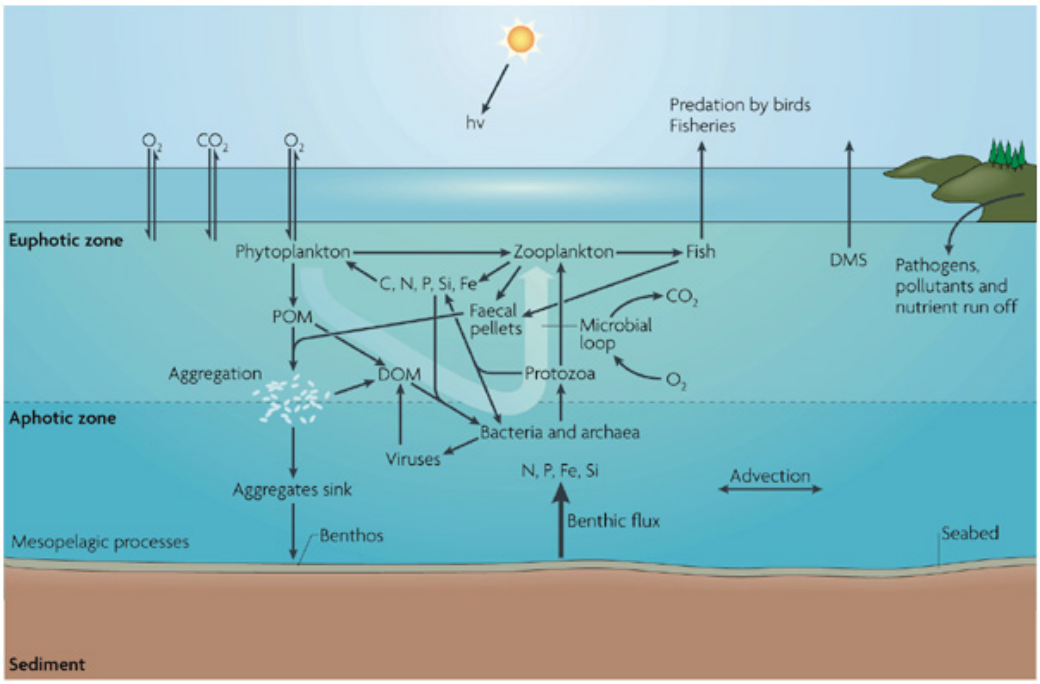


Figure 3. Microbial structuring of the marine ecosystem. A large fraction of the organic matter that is synthesized by primary producers becomes dissolved organic matter (DOM) and is taken up almost exclusively by bacteria. Most of the DOM is respired to carbon dioxide and a fraction is assimilated and re-introduced into the classical food chain (phytoplankton to zooplankton to fish). DMS, dimethylsulphide; hv, light; POM, particulate organic matter (from Azam & Malfatti, 2007).

2. From in vitro-culture approaches to environmental genomics

2.1 Towards a molecular taxonomy

Microbiology has built its foundations on the discovery of microorganisms using the microscope on one side, and on the cultivation of microorganisms on plates on another side. Pioneer microscopist Anton van Leewenhoeck, observed in the XVIIth century living creatures in the rainwater, which he called "animalcules", using a handcrafted microscope. He could notice different morphologies and associations of microorganisms. Recording and describing these different types of bacteria can be said to be the birth of the field of bacterial taxonomy. Profuse observation of different microbial morphologies, the use of cultivation techniques on solidified gelatin media and the development of chemical tests (enzymatic assays) to distinguish among similar morphotypes, helped publish in 1923 the first edition of the "Bergey's manual of Determinative Bacterioloy" [Bergey *et al.,* 1923], which still provides a useful reference tool in the identification of bacteria.

These tools allowed microbiologists to develop the taxonomy of bacterial isolates. Bacteriologists had used widely these protocols until they realized that most bacteria enumerated under the microscope could not be cultivated on agar plates or liquid media. In this sense, marine microbiologists estimated that only 1% of the cells formed colonies on agar plates, and named this issue the "Great plate count anomaly" [Amann *et al.,* 1995]. With the advances in the study of DNA in the mid 1950's, new techniques emerged (i.e. guanine and cytosine ratio, DNA-DNA hybridization). A breakdown in bacterial taxonomy was achieved by Woese and colleagues, who utilized the sequence of ribosomal ribonucleic acids (rRNA) to catalogue related bacterial species [Woese & Fox, 1977]. The choice of ribosomal genes for taxonomic studies was based on the fact that ribosomes are biologically universal molecules with a highly constrained and conserved function (i.e. the translation of messenger RNA, mRNA, into proteins). Since they are essential to all biological metabolisms, ribosomes are present in living organisms since the very early stages in the evolution of life, and have changed relatively little with evolution. Ribosomal genes share a high similarity in their conserved regions (used for gene detection), but vary among different lineages in their less conserved regions (hypervariable regions), which are used for phylogeny classification [Woese, 1987; Doolittle, 1999]. Small subunit ribosomal sequences have been invaluable for uncovering the vast diversity of microbial life [DeLong & Pace, 2001], and for assigning uncultivable organisms as new species [Hugenholts, 1998].

2.2 The bacterial "species" concept in microbial ecology

One of the crucial topics in microbial ecology is the characterization of bacterial diversity (e.g. species richness and evenness), and how to define bacterial species in environmental samples. A microbial species concept is fundamental for any biodiversity study. As stated by Mayr [1982]: 'The species [...] is the basic unit of ecology […] no ecosystem can be fully understood until it has been dissected into its component species and until the mutual interactions of these species are understood'. We must discover how molecular variation is organized into fundamental units to understand bacterial community composition and structure before addressing many other relevant eco-evolutionary questions. For bacterial systematics, as the DNA-based methods became available, they were applied to the problem of species demarcation, but were merely calibrated to correlate with previously established phenotypic clusters. Thus "gold standard molecular cutoffs" were calibrated to yield the species grouping already determined by phenotypic

clustering. For instance, the 70% cutoff for DNA-DNA hybridization was calibrated to yield the phenotypic clusters previously recognized as separate species, and the 3% cutoff for 16S rRNA divergence was calibrated to yield the species previously determined by DNA-DNA hybridization and phenotypic clustering [Stackebrandt & Goebel, 1994]. However, comparative genomics has revealed that species defined at these levels are known to encompass large diversity in genome content, physiology and ecology [Staley, 2006; Ward, 2006] and are expected to underestimate the total diversity present in the samples. Comparative genomics has shown that all pairs of genomes from the same "species" showing ≈94% ANI (Average Nucleotide Identity), correspond to the traditional 70% DNA-DNA similarity and to >98.5% of 16S rRNA gene identity [Konstantinidis & Tiedje, 2005; Goris *et al.,* 2007]. Moreover, a recent study using a clustering analyses with identity cut-offs between 98.7 - 99.1%, based on 40 universal protein coding genes (pMGs), found these to be the best cutoff to accurate delineate prokaryotic species and this cutoff was well correlated with the range 98.5 - 99% in 16S rRNA sequence similarity [Mende *et al.,* 2013].

2.3. Evolutionary processes and bacterial species

These ways of demarcating species remain faithful to the notion that phenotypically defined species are true species. This happens because bacterial systematics is lacking a theory-based concept of species, all we can do is calibrate each new molecular technique to fit the clusters previously determined by phenotypic criteria [Cohan, 2002].

Fred Cohan, an evolutionary biologist at Wesleyan University [Middletown, Connecticut, USA) noted that a named bacterial "species" should be assimilated more to a "genus" than to a "species" concept. He suggested that microbial systematics should be transformed by ecological and evolutionary theory, and he has proposed the concept of "ecotype" to split bacterial species into small, biologically meaningful units [Cohan, 2001, 2002]. Summarizing Cohan's ecotype concept, it states that given enough time, a series of periodic selection events and random neutral mutations that accumulate within each separate population leads to genetically distinct clades of two distinct ecotypes, where the average sequence divergence between ecotypes is much greater than the average sequence divergence within them, for any gene shared by the ecotypes. In addition, each ecotype is expected to be identifiable as a monophyletic group in a phylogeny based on DNA sequence data [Cohan, 2002] (Fig. 4). Indeed, it is widely demonstrated that all living

organisms fall within discrete clusters of closely related individuals as defined by their sequence similarity [Palys *et al.,* 1997]. More recently, a broad-scale study of bacterial diversity in the environment found that more than 50% of the bacterial ribotypes fall into discrete clusters containing less than 1% sequence divergence [Acinas *et al.,* 2004a]. These authors suggested that these microdiverse clusters (this so-called microdiversity) correspond to Cohan´s ecotype concept.

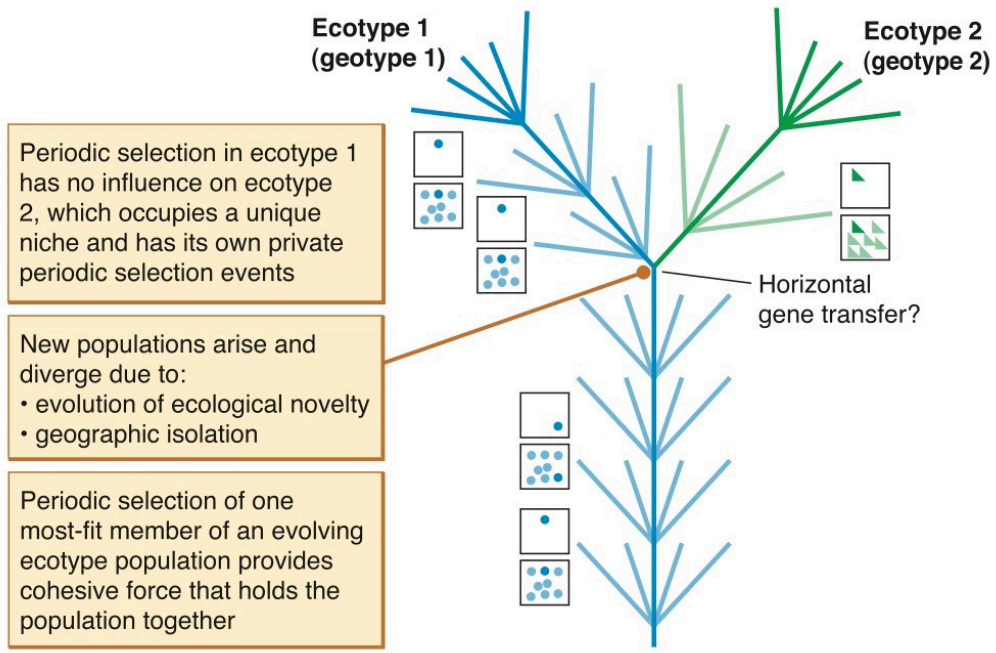


Figure 4. A model of the evolution of distinct ecotypes. Ecologically specialized populations (ecotypes) evolve through a series of events in which the best adapted population outcompetes all others within the ecotype, squashing diversity within the population; subsequently, new diversity arises within the ecotype. From the base of the tree, successive periodic selection events are associated with the evolution of an ecotype. Blue lines forming fans and boxes with many circles indicate diversity within an ecotype that goes extinct (light blue lines and circles), when one better variant (dark blue line and circle) is favored and moves upward to initiate new diversity. Mutations that allow a population member to occupy a new niche lead to the emergence of a new population (green lines) that undergoes periodic selection independent of the parent population; the two populations diverge, each held together by independent periodic selections (from Ward, 2006).

Ecotypes are similar phylogenetic subgroups of bacteria that differ in physiological details that determine niche specificity [Cohan, 2006]. Hence, the term ecotype describe populations with unique distributions along ecological or physical gradients, and can be used to identify basic community units that occupy unique ecological niches or are

found in distinct locations. With this concept, species are no longer merely a cluster of similar organisms; a species would now be viewed as a fundamental unit of ecology and evolution, with certain dynamical properties. According to this view, speciation, and ecological species definitions, must consider bacterial gene-transfer processes, which are erratic and transfer only a small part of the genome [Prosser *et al.,* 2007]. The rates of homologous recombination and horizontal transfer vary widely among different bacterial species [Gogarten & Townsend, 2005; Vos & Didelot, 2009; Didelot & Maiden 2010; Fernández-Gómez *et al.,* 2012]. This genetic plasticity and heterogeneity is one reason why we are still far from a consensus on the nature of bacterial species [Fisher *et al.,* 2006], and perhaps these processes are so relevant that microbial populations do not fit in a traditional species definition [Ward, 2006].

2.4. Using phylogenies to uncover ecotypes

Divergent 16S rRNA sequences may represent different strains and species, whose ribosomal genes have undergone repeated random fixation of mutations and gene conversions during evolutionary divergence [Field *et al.,* 1997]. To test whether genetic variation in ribosomal gene clusters is correlated with long-term evolutionary divergence we could look for niche partitioning among sequence types. If gene clusters recovered from environmental DNA are associated to a specific environmental factor (i.e. depth), it could be suggested that some of the variation observed within gene clusters corresponds to bacterial "speciation" with depth. Several examples provide evidence that the fine structure of phylogenetic 16S rRNA trees can provide useful information about evolution and functional specialization in natural microbial populations. Two famous examples that illustrate this idea are the cases of the marine *Cyanobacteria* (*Prochlorococcus* and *Synechococcus*) and the SAR11 gene cluster.

The phylogenetic tree of cultured marine *Cyanobacteria* has several interesting topological features that correlate with organism physiology. The deepest branch separates *Prochlorococcus* from *Synechococcus*, which differ markedly in their light-harvesting structures and ecology [Urbach *et al.,* 1998]. Two ecotypes associated with different depths in an oceanic water column and with distinct light adaptations have been detected within *Prochlorococcus* based on 16S rRNA clades: Low-Light (LL) adapted ecotypes grow near the bottom of the photic zone, with high Chl b/a_2 ratios; and High-Ligh (HL) adapted ecotypes are found at shallower depths in the water column, with low Cha b/a_2 ratios [Moore *et al.,* 1998; West & Scanlan, 1999]. An increasing number of

studies focused on the abundance of these *Prochlorococcus* populations have revealed the spatial [Zinser *et al.,* 2006] and also the temporal [Malmstrom *et al.,* 2010] dynamics of the different ecotypes.

Another example is the SAR11 clade belonging to Alphaproteobacteria division, the prokaryote cell most abundant in the ocean that forms a clade with rRNA sequence identities as low as 87% [Rappé & Giovanonni, 2003]. The SAR11 clade currently contains only a small number of cultivated representatives, all of which contain nearly identical rRNA gene sequences. The clade is divided into two subclades or ecotypes (based on the small-subunit rRNA gene analysis), which appear to roughly correspond to surface and deep types [Field *et al.,* 1997]. The long internal branches connecting SAR11 subclades in rRNA-based phylograms suggest a lengthy evolutionary history punctuated by the divergence of ecotypes and episodes of periodic selection [Rappé & Giovannoni, 2003]. Furthermore, quantitative analyses have provided data supporting spatial and temporal patterns of lineage distributions [Morris *et al.,* 2005; Carlson *et al.,* 2009].

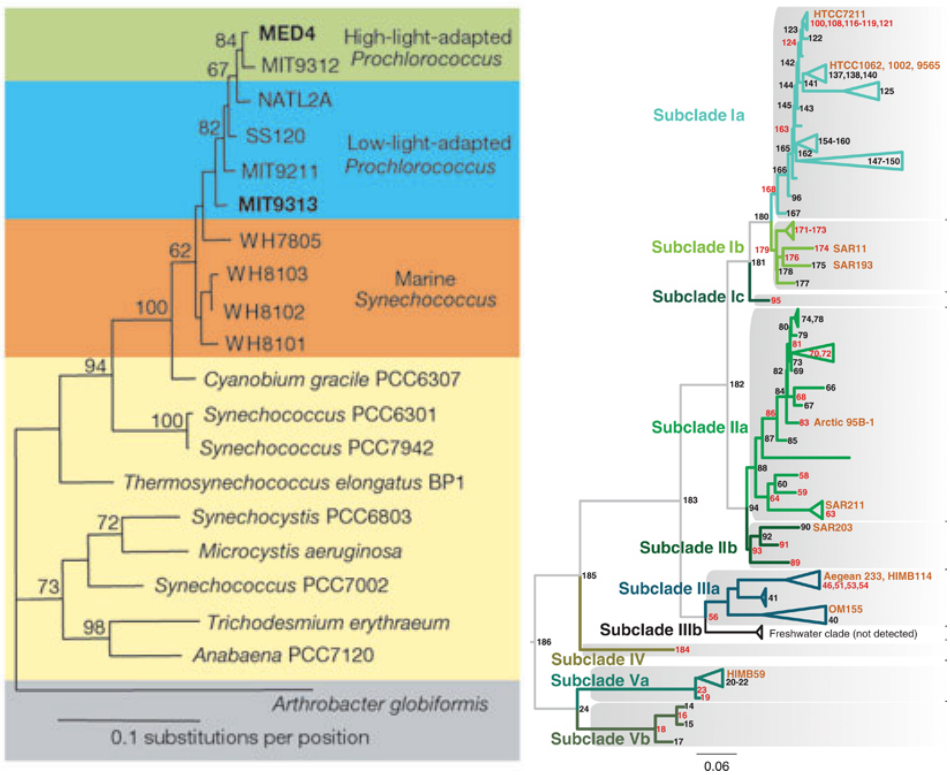


Figure 5. A) Relationships between *Prochlorococcus* and other *Cyanobacteria* inferred using 16S rRNA genes (from Rocap et al., 2003). B) SAR11 subclade dynamics at BATS using 16S rRNA genes (Vergin et al., 2013).

Additional reports support similar conclusions on the association of 16S rRNA-based clades and specific habitats. Distinct hot spring *Synechococcus* 16S rRNA genotypes were found associated with different temperatures and thermal adaptations [Ferris & Ward, 1997; Allewalt *et al.,* 2006]; a subdivision (SAR324) of the Deltaproteobacteria lineage specialized in lower ocean surface layers has been reported as well [Wright *et al.,* 1997]. In addition, a gene cluster (SAR202) of Green Non-sulfur bacteria was most abundant at the lower boundary of the deep chlorophyll maximum [Giovanonni *et al.,* 1996]. However, there are no pre-stablished criteria regarding the level of variation expected for different taxa in prokaryotes [Fox *et al.,* 1992; Stackebrandt & Goebel 1994; Cohan, 2006], and even less so for the level of variation expected among different ecotypes [Palys *et al.,* 1997; Furhman & Campell 1998; Moore *et al.,* 1998].

Several models have been developed to delineate the species or ecotypes boundaries, based on phylogeny and having a common ground on evolutionary theory, such as ecotype simulation [Koeppel *et al.,* 2008] and AdaptML [Hunt *et al.,* 2008] wherein ecotypes are identified as groups of phylogenetically cohesive sequences without relying on a predefined genetic cutoff or phylogenetic relationships to already named species [Cohan & Perry, 2007; Cohan & Koeppel, 2008; Preheim *et al.,* 2011; Szabo *et al.,* 2012). Specifically, in AdaptML such ecologically distinct bacterial populations can be associated to their predicted habitats [Hunt *et al.,* 2008; Preheim *et al.,* 2011]. While we have not yet used them in this PhD thesis work, they are the logical next step and they would be implemented in the near future to the work we report in chapter 4.

2.5 Other markers of microdiversity

Although the advantages of the small subunit (SSU) rRNA as phylogenetic marker are well documented with respect to its information content and the comprehensiveness of the available data set, it is generally accepted that rRNA-based conclusions can reflect evolutionary history only roughly [Ludwig *et al.,* 1998; Ludwig & Klenk, 2001]. This molecule may not be the best marker in all circumstances, particularly when trying to examine at closer inter- or intraspecific relationships. Additional phylogenetic markers may approximate more detailed phylogenies in some scenarios [Sapp, 2005; Roux *et al.,* 2011], thus dividing clades into minor subgroups or ecotypes. Some of the fine-scale analyses of certain groups have been made with protein-coding genes: for instance, *Synechococcus* subgroups have been studied using *cpc* BA-IGS [Roberstson *et al.,* 2001] or ITS-1 and *cpc*BA operon to distinguish *Pseudoanabaena* ecotypes [Acinas *et al.,* 2009].

Identification of deep-sea ecotypes of *Alteromonas macleodii* [López-López *et al.,* 2005] was also attempted using *gry*B and *rpo*B.

The internal transcribed spacer (ITS) of the rRNA operon is increasingly used to assess microdiversity in bacterial population genetics for its high variation in length and sequence [e.g., García-Martínez *et al.,* 1999; Di Meo *et al.,* 2000; Boyer *et al.,* 2001; Hurtado *et al.,* 2003; Vogel *et al.,* 2003; Brown & Fuhrman, 2005; Acinas *et al.,* 2009]. Compared to other molecular markers (16S rRNA or protein-coding genes), the ITS region experiences low selective constraints, evolves rapidly, and provides a high-resolution estimate of gene flow and genetic structuring at the population scale [Gürtler & Stanisich, 1996; Antón *et al.,* 1998; Schloter *et al.,* 2000; Rocap *et al.,* 2002, 2003; Brown & Fuhrman, 2005].

In the same examples of marine bacteria presented above, the use of the ITS regions has revealed new subdivisions of the existing ecotypes among *Cyanobacteria* and SAR11 clades. Among the cyanobacterial lineages, the ITS region has split previously defined ecotypes in new types associated with particular phenotypes and geography [Rocap *et al.,* 2002]. Similarly, variations in the ITS region identified depth-dependent ecotypes among thermophilic *Synechococcus* [Ferris *et al.,* 2003]. Regarding the SAR 11 clade, ecotypes have also been further divided into lineages, related with specific depths and geography, and potentially indicating local adaptations [García-Martínez & Rodríguez-Valera, 2000; Brown & Furhman 2005]. These ecotypes also follow depth-dependent shifts in relative abundances [Ngugi & Stingl, 2012].

3. The Phylum *Bacteroidetes*

3.1. Natural abundances and ecological role of marine Bacteroidetes

Culture-independent approaches such as 16S rRNA gene clone libraries, denaturing gradient gel electrophoresis (DGGE), and fluorescence *in situ* hybridization (FISH), have revealed members of the phylum Bacteroidetes to constitute one of the most abundant bacterioplankton groups [e.g. 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], polar sediments [Ravenschlag *et al.,* 2001], coastal sediments [Llobet-Brossa *et al.,* 1998], coastal waters [Eilers *et al.,* 2001; O'Sullivan *et al.,* 2004], and the open ocean [Simon *et al.,* 1999; Abell & Bowman, 2005b; Schattenhofer *et al.,* 2009]. High abundances of Bacteroidetes have been linked to cold waters [Simon *et al.,* 1999;

Abell & Bowman, 2005b], upwelling systems [Alonso-Sáez *et al.,* 2007; Alonso-Sáez *et al.,* 2012] and during phytoplankton blooms [Simon *et al.,* 1999; O'Sullivan *et al.,* 2004; Pinhassi *et al.,* 2004; Teeling *et al.,* 2012]. This distribution suggests a preference for rather productive environments. Despite this widely reported preference, in some cases Bacteroidetes phylotypes seem also to dominate in conditions when blooms do not occur [e.g. Riemann *et al.,* 2000; Pinhassi *et al.,* 2004; West *et al.,* 2008], suggesting that only certain marine Bacteroidetes clades might be responding primarily to the phytoplankton blooms. In addition, Bacteroidetes have also been detected in oligotrophic marine surface waters, sometimes in significant large numbers [Schattenhofer *et al.,* 2009].

Besides their relative abundances in different environments, Bacteroidetes are shown to play a specialized ecological role in DOM uptake and degradation. Cultured Bacteroidetes isolates are well known to act proficiently in degrading biopolymers such as cellulose and chitin, which are part of the high molecular mass (HMW) fraction of the DOM [Reichenbach & Weeks, 1981; Riemann *et al.,* 2000; Cottrell & Kirchman, 2000b; Pinhassi *et al.,* 2004]. Polysaccharides are one of the main constituents of the DOM released by phytoplankton [Biddanda & Benner, 1997], and thus the microorganisms that initiate polysaccharide degradation have a key role in OM recycling in the oceans. Experimental studies have shown that Bacteroidetes was the dominant clade after incubation with high molecular weight OM, but were not dominant when low molecular weight OM was tested [Covert & Moran, 2001]. Similarly, a protein enrichment incubation experiment triggered mostly the response of Bacteroidetes [Pinhassi *et al.,* 1999].

In natural marine systems, supporting observations for HMW DOM use by uncultured Bacteroidetes *are* derived from studies with combined microautoradiography and FISH techniques. Bacteroidetes was the most abundant group consuming chitin, N-acetylglucosamine, and protein, whereas they were generally underrepresented in assemblage consuming amino acids, usually dominated by members of the α-Proteobacteria [Cottrell & Kirchman, 2000b; Alonso-Saéz *et al.,* 2012]. More recent genomic studies provide additional support for the preferential use of polysaccharides by members of this phylum. Whole genome sequence of some marine Bacteroidetes has revealed that they encode specialized metabolic machinery for he degradation of high molecular weight substrates [Bauer *et al.,* 2006; González *et al.,* 2008; Woyke *et al.,* 2009; Gómez-Pereira *et al.,* 2012; Fernández-Gómez *et al.,* 2013].

Another important feature of Bacteroidetes is the presence of proteorhodopsin (PR) [Gómez-Consarnau *et al.,* 2007; Fernández-Gómez *et al.,* 2013] that functions as

a light-driven H^+ pump. A dual life strategy has been proposed for some Bacteroidetes (i.e. *Polaribacter* sp. MED 152 and *Dockdonia* sp. MED 134). These organisms would grow optimally attached to particles or other surfaces using polymeric substances and, when the organic matter is depleted on the particle, would float passively in the nutrient-poor water in search of fresh particles to colonize [Gónzalez *et al.,* 2008, 2011]. Under the later condition, they would use PR to obtain energy from light providing a survival advantage during periods of low nutrients, but also it could provide additional energy for degradation of complex organic matter or uptake of amino acids and peptides at low concentrations [Gómez-Consarnau *et al.,* 2007; González *et al*., 2008, 2011].

3.2. Phylogeny

The taxonomy of the phylum Bacteroidetes (previously known as *Cytophaga-Flavobacteria-Bacteroides*) has undergone several reclassifications over the past years. The latest scheme comprises four classes well delineated on the basis of their 16S rRNA sequences, and two groups affiliated to the phylum but not readily assigned to any of these clades [Lugwig *et al.,* 2001] (Fig. 6). The advent of the ribosomal rRNA approach and the advances in sequencing technologies have lead to an exponential increase in the number of Bacteroidetes sequences from yet uncultivated microorganisms available in public databases. In the Silva database release 115 [Pruesse *et al.,* 2007] there are a total of 433,933 Bacteroidetes 16S rRNA gene sequences.

Members of the class Bacteroidia are mostly strict anaerobes, usually associated with mammals or other animals, as disease agents or as part of the gut microbiota. Particularly, members of the genus *Bacteroides* are the numerically dominant bacteria in the human large intestine [Salyers, 1984]. Some families of the class Bacteroidia are also found in marine environments (e.g., Marinilabiaceae) [Goffredi, 2009]. The classes Sphingobacteria and Cytophagia have recently been separated [Ludwig *et al.,* 2010]. Both classes include mostly aerobic heterotrophic motile bacteria that are found in a variety of habitats. Members of the class Flavobacteria have widely diverse ecological niches and physiological characteristics. The class includes the families Flavobacteriaceae, Blattabacteriaceae and Cryomorphaceae [Ludwig, 2010]. 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 spread within the other classes. Within the class Flavobacteria distinct “marine clade” has been delimited based on 16S rRNA gene phylogeny. The “marine clade” comprises genera solely isolated from oceanic environments [Bowman, 2006].

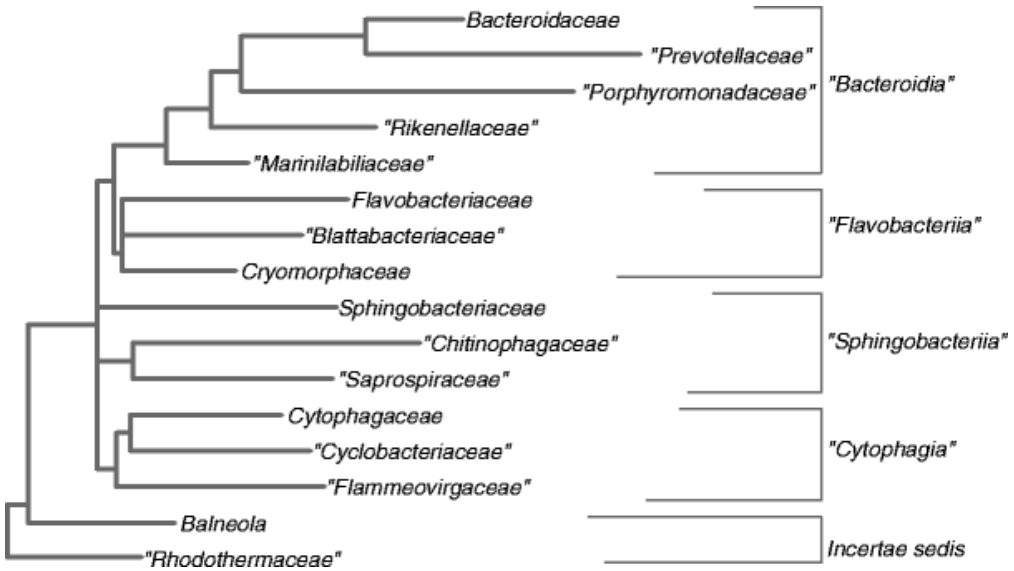


Figure 6. Overview of the phylum "Bacteroidetes". This phylum contains 15 families classified within four classes (from Ludwig et al., 2010).

3.3. Molecular techniques applied to the study of Bacteroidetes

With the widespread use of molecular techniques, some aspects of the abundance, distribution, structure and ecology of marine bacteria has progressed exponentially in recent years [Zinger *et al.,* 2012]. However, such understanding relative to Bacteroidetes has not deeply explored due to some methodological limitations. On one hand, most of the studies of marine bacteria using environmental clone libraries have used universal primers. As a result, Bacteroidetes were not even in the list of heterotrophic bacteria present in marine systems [reviewed in Giovannoni & Rappé, 2000]. Moreover, some studies indicated gross discrepancies between the percentage of sequences that Bacteroidetes represent in clone libraries compared with the abundances when enumerated by FISH with the most commonly used probe [CF319a, Manz *et al.,* 1996] [Cottrel & Kirchman, 2000; Eilers *et al.,* 2000]. A commonly accepted explanation is the apparent bias against this phylum shown by the universal PCR primers commonly used in clone library building [Cottrell & Kirchman, 2000a; Eilers *et al.,* 2000; Kirchman *et al.,* 2003].

On the other hand, studies regarding the abundance of Bacteroidetes also present some not easily reconcilable issues. For instance, the widely accepted Bacteroidetes-specific probe CF319a may not yield accurate quantitative data due to poor group coverage [Weller *et al.,* 2000; Kirchman *et al.,* 2002; O'Sullivan *et al.,* 2002; Chen *et al.,* 2006; Amann & Fuchs, 2008; Amann *et al.,* 2008], leading to underestimations of abundances. At the same time, CF319a has also been shown to present low specificity [Amann *et al.,* 2008], thus maybe artificially overestimating Bacteroidetes numbers. Furthermore, a study comparing two group-specific probes CF319a and CF560 [O'Sullivan *et al.,* 2002] found 15% higher Bacteroidetes abundance values with the latter probe [Cottrell *et al.,* 2005]. In summary, studies on the abundance and distribution of Bacteroidetes are not only scarce, but also largely inaccurate due to the lack of confidence on the molecular probes used for this group.

3.4. A closer view at the different Bacteroidetes clades

Traditionally, Bacteroidetes have been studied as a very broad group, and evidence of differential distribution patterns and detailed analyses of its microdiversity is scarce. Only a few studies have attempted the quantification of distinct clades of Bacteroidetes in marine environments. Examples of subgroups specifically studied are the *Cytophaga marinoflava-latercula* linage [Eilers *et al.,* 2001], clade DE2 [Kirchman *et al.,* 2003], AGG58 cluster [O'Sullivan *et al.,* 2004], *Polaribacter* clade [Malmstrom *et al.,* 2007], NS marine clades [Alonso *et al.,* 2007], and VISON clades [Gómez-Pereira *et al.,* 2010]. These studies are mostly based on using group-specific FISH probes and have shown that different Flavobacteria communities dominate distinct water masses, different seasons, or depth ranges. However, a closer view at the molecular diversity of some of these clusters suggests that they are still phylogenetic too broad, and subgroups of more closely related sequences should be identified before constructing ecologically meaningful categories at the level of populations or ecotypes.

The "appropriate phylogenetic level of resolution" that is needed for ecological studies will depend on the specific questions asked. In order to detect correlations with environmental parameters, an experimental design targeted to the study of the total abundance of the group may not find any relevant correlation while different subgroups may be undergoing complex dynamics, responding to different physico-chemical or biotic variations, thus, implying niche specialization.

4. Approaches used in this thesis to study the ecology of microbial lineages

Microbial ecology is the study of microorganism in the environment and their interactions with each other. The study of microorganisms cannot rely on observation of phenotypical characters, and classical microscopy is not, by far, a powerful enough technique for their identification. Culture approaches (when feasible) and molecular techniques are needed to identify the different groups. In this thesis, I have applied a combination of several molecular approaches summarized below (Fig. 7):

A general technique for the identification and quantification of microorganisms is FISH (with the CARD-FISH modification for improved sensitivity, Amann *et al.,* 2008). FISH probes can facilitate an initial assessment of the dominance of certain taxa for large taxonomic groups, such as Archaea, Bacteria and Eukaryotic domains, or even for narrower groups. However, being a direct count, usually conducted manually under the fluorescence microscope, it is a very labor-intensive technique and there is a limited number of probes that can be tested in a single water sample (i.e. filter). In addition, the labeling depends on the specificity and the efficiency of probe binding.

The molecular characterization of natural microbial communities can also be undertaken with the use of the polymerase chain reaction (PCR) [White *et al.,* 1989]. A standard work-flow of this kind consists in the extraction of the nucleic acids of the whole community, and the amplification of rRNA gene fragments. These fragments can be analyzed by fingerprinting techniques such as denaturing gel gradient electrophoresis (DDGE), or they can be cloned into plasmid vectors. PCR-DGGE is a rapid and efficient technique consisting on the separation of DNA amplicons with the same length but with different sequences, which may differ by as little as a single base pair modification [Scheffield *et al.,* 1989; Muyzer *et al.,* 1993, 1998]. A major advantage of this method is its potential to display a visual profile (a “fingerprint”) of the community, and to monitor the changes occurring in different conditions. Cloning methodologies are an easy way to separate unique DNA fragments, previously amplified by PCR, and to introduce them in vectors, which are cloned into live cells. Both DGGE bands and cloned products can be sequenced, deposited in data bases and compared with others available in public databases such as Genbank (at the National Center for Biotechnology Information, NCBI, or resources for quality checked and aligned ribosomal RNA sequence data such a as the Silva database or the Ribosomal Database Project. Public databases are readily accessible for comparative analyses, and most often they also provide softwares (e.g., ARB, Ludwig *et al.,* 2004) or online tools, for such purposes. Following these approaches,

we are now able to analyze unique characters, such as the DNA sequences, directly, to build phylogenetic trees reflecting their relationships, and to design more specific probes or primers to target specific discrete subgroups.

We also made use of high-throughput sequencing approaches to analyze 27 metagenomes from different environmental samples. We chose a 454/Roche sequencing platform for a direct sequencing of extracted DNA [Hingamp *et al.,* 2013]. This next-generation sequencing approach rely on DNA fragmentation and capture via various ligation strategies [Gilbert & Dupont, 2011], thus foregoing traditional cloning strategies for DNA sequencing. These results are not affected by PCR-derived biases since no DNA amplification using specific primers is employed. The 16S rRNA sequences are subsequently extracted from the metagenomes with models that identify reads of RNA genes within the metagenomic reads [Huang *et al.,* 2009].

Finally, quantitative techniques such as qPCR are sensitive, flexible, low cost and extremely throughput. However, as for FISH, the accuracy of the technique largely depends on the design of appropriate primers or probes and it may also be subject to the PCR biases. In contrast, qPCR allows the estimation of PCR amplification efficiency, and it is more sensitive when low abundances of cells may limit the application of FISH in many aquatic habitats. Q-PCR has been shown to be a robust, highly reproducible and sensitive method to quantitatively track gene changes across temporal and spatial scales under varying environmental or experimental conditions [e.g. Fandino *et al.,* 2005; Zinser *et al.,* 2006; Eiler *et al.,* 2007]. Moreover, the quantitative data generated can be used to relate variation in gene abundances in comparison with variation in abiotic or biotic factors and/or biological activities and process rates.

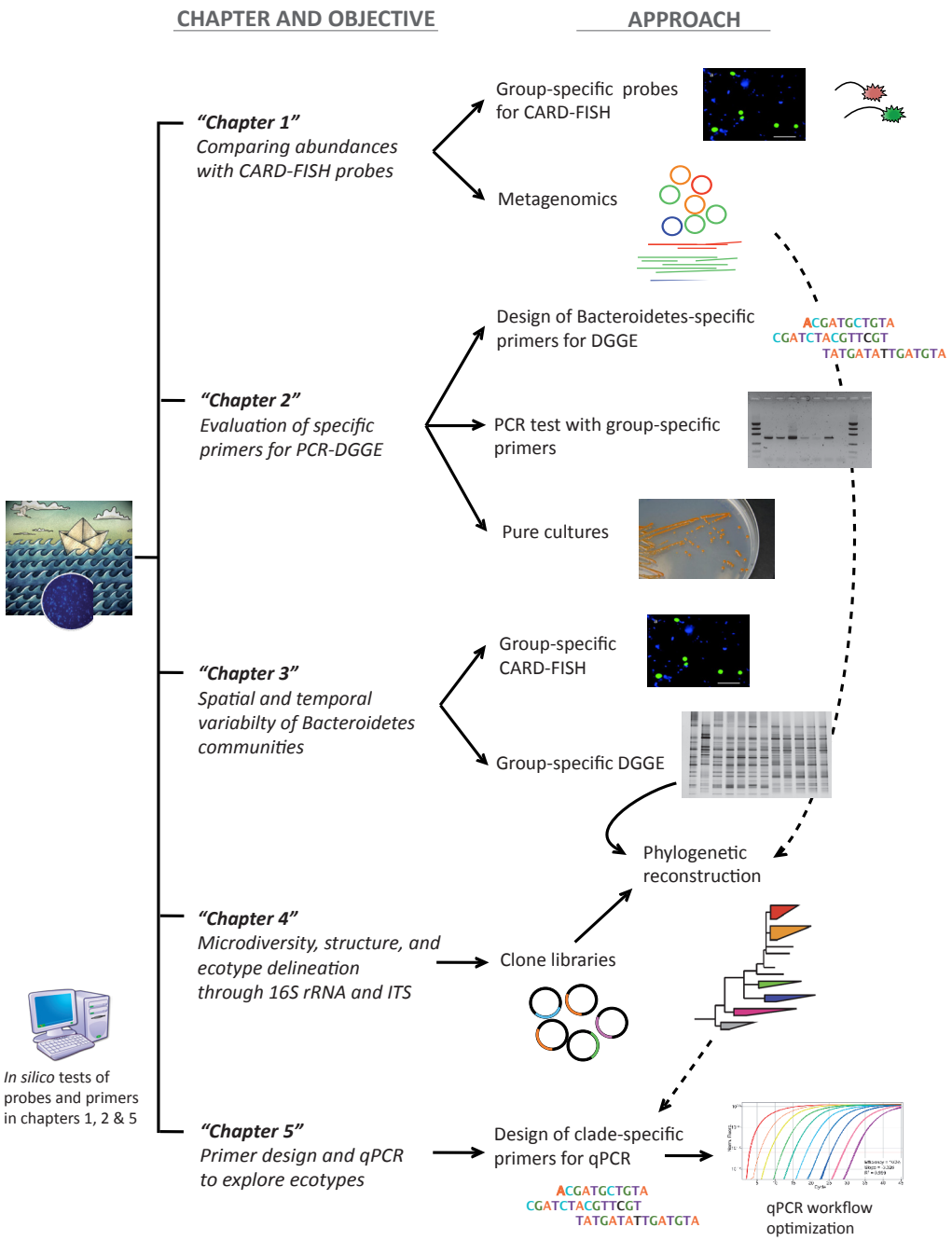


Figure 7. Outline of the different objectives and molecular approaches used in each chapter of this thesis.

AIMS OF THE THESIS

General objective

The overall aim of this work was to generate tools and workflows to characterize the diversity, structure and distribution of phylogenetic distinct marine Bacteroidetes clades by culture-independent approaches.

The Specific objectives of this thesis were:

I. Due to the limited performance of current molecular probes (low coverage and specificity), our first objective was to design and develop improved group-specific probes for reliable quantification of Bacteroidetes populations. This is covered in Chapter 1.

II. In order to avoid the low representation of Bacteroidetes in fingerprinting approaches performed with universal primers, our next goal was to design and test group-specific primers suitable for population analyses by DGGE. This is addressed in Chapter 2.

III. The objective in Chapter 3 was to quantify and to explore the population dynamics of Bacteroidetes in different marine environments. More specifically, to investigate the abundances and population structure and their variation, spatially (in a coastal site and in a deep environment) and temporally (over two years in a surface coastal station) using quantitative FISH and fingerprinting techniques improved after the specific objectives I and II. We hypothesized that Bacteroidetes abundances would show seasonal dynamics and that different populations would inhabit different water depths and in different seasons.

IV. In view of the results of objective III and the limitations observed in the fingerprinting techniques, a more detailed study of Bacteroidetes populations was our objective covered in Chapter 4. In order to thoroughly characterize the diversity of Bacteroidetes in the North Western Mediterranean Sea, we chose selected representative samples from those studied in Chapter 3 to build and analyze

extensive clone libraries of 16S-ITS rRNA gene sequences. We hypothesized that the fine-scale phylogeny of Bacteroidetes would help delineate habitat specific-clusters within Bacteroidetes populations, and that ITS-based phylogenies would allow to further subdivide some clusters into putative ecotypes not detected with the 16S rRNA gene.

V. Finally, we aimed to quantify some of the putative ecotypes identified in Chapter 4 to determine whether they are always associated with a given environment. This would imply that these sequence clusters actually represent ecological different populations (i.e. ecotypes). Moreover, quantitative analyses would set a comprehensive framework for studies seeking meaningful correlations of their abundance in different habitats with the environmental factors. In doing so, we generated a library of primers sets suitable for quantitative PCR analysis. This is developed in Chapter 5.

Chapter 1

Comparative Abundance of Marine Bacteroidetes with paired CARD-FISH probes and metagenomic 16S rRNA sequences

Summary

Catalyzed reporter deposition-fluorescence in situ hybridization (CARD-FISH) is a powerful approach commonly used in microbial ecology to quantify bacterial taxa. We compared the performance of the widely used Bacteroidetes specific CF319a probe to the newly tested CF968 probe. Their accuracy was assessed across 37 globally distributed environmental samples as well as in two consecutive years at the Blanes Bay Microbial Observatory (NW Mediterranean). We also compared the CARD-FISH data with Bacteroidetes16S rRNA gene sequences retrieved from 27 marine metagenomes from the TARA Oceans expedition. Our results demonstrate that (i) CF319a hybridizes with non-Bacteroidetes isolates, specifically with members of Rhodobacter and Alteromonodales; (ii) quantitative estimations with both probes were not highly different at single sampling points but showed significant differences when seasonal dynamics were investigated, and (iii) CF968 detected a seasonal pattern within marine Bacteroidetes that was overlooked by the CF319a probe. Comparison of CARD-FISH and metagenomic data suggests that differences between probes are likely due to a combination of unspecific hybridization, differential target of specific phylotypes, as well as lower coverage by CF319a. We propose the CF968 probe as an alternative to CF319a since it has higher coverage, greater specificity, and provides better discrimination within marine Bacteroidetes.

Comparative abundance of marine Bacteroidetes with paired CARD-FISH probes and metagenomic 16S rRNA sequences. **Silvia G. Acinas1, Isabel Ferrera, Hugo Sarmento, Cristina Díez-Vives, Irene Forn, Clara Ruiz-González, Francisco M. Cornejo-Castillo, Guillem Salazar and Josep M. Gasol** (Environmental Microbiology, in review).

Introduction

Marine Bacteroidetes, mostly belonging to the class Flavobacteria, represent one of the major groups of bacterioplankton in a variety of environments [Kirchman *et al.,* 2003; Chen *et al.,* 2006; Alonso *et al.,* 2007]. They are typically associated to cold waters, upwelling systems and phytoplankton blooms [Simon *et al.,*1999; Pinhassi *et al.,* 2004; Abell & Bowman, 2005b; Schattenhofer *et al.,* 2009; Gómez-Pereira *et al.,* 2010] and usually are found enriched in the particle-attached fraction of the bacterioplankton [DeLong *et al.,* 1993; Crespo *et al.,* 2013]. Although lower abundances of Bacteroidetes are reported in the oligotrophic open ocean and in deeper waters [Schattenhofer *et al.,* 2009; Gómez-Pereira *et al.,* 2010; Lefort and Gasol, 2013; Díez-Vives *et al.,* unpublished], this group represents a cosmopolitan phylum in marine environments. Several quantification approaches have been used to evaluate its abundance in marine environments. CARD-FISH analyses have shown Bacteroidetes abundances between 10-20% of the prokaryotes in diverse marine surface waters [Glockner *et al.,* 1999; Cottrell & Kirchman, 2000a; Alonso-Sáez *et al.,* 2007; Ruiz-González *et al.,* 2012a; Lefort and Gasol, 2013], and quantitative PCR (qPCR) reported abundances ranging from less than 1% up to 53%, being the highest numbers associated to phytoplankton blooms [Abell & Bowman, 2005b; Fandino *et al.,* 2005].

Recently, the relative abundances of marine Bacteroidetes sequences have also been obtained through high-throughput sequencing technologies such as 454 FLX Titanium and Illumina (HiSeq2000/GAIIx) providing similar ranges between 8-20% [Rusch *et al.,* 2007; Kirchman *et al.,* 2010; Campbell *et al.,* 2011; Crespo *et al.,* 2013].

Approaches such as CARD-FISH combined with microautoradiography have unveiled relevant ecological features of the Bacteroidetes phylum, such as their relatively low contribution to the uptake of low molecular weight substances [Cottrell & Kirchman, 2000b; Alonso-Sáez *et al.,* 2012], or their resistance to UVR and PAR exposure [Alonso-Sáez *et al.,* 2006; Ruiz-González *et al.,* 2012b]. Comparative genomics analyses have pointed out the presence of a large number of peptidases, glycoside hydrolases (GHs), glycosyl transferases, adhesion proteins and genes for gliding motility that allow them to grow attached on particle surfaces [Fernández-Gómez *et al.,* 2012; Gómez-Pereira *et al.,* 2012; Fernández-Gómez *et al.,* 2013]. Despite their moderate growth rates [Teira *et al.,* 2009; Ferrera *et al.,* 2011], marine Bacteroidetes are particularly successful in the degradation of organic matter in the ocean [Riemann *et al.,* 2000; Pinhassi *et al.,* 2004; Sarmento & Gasol, 2012]. The capacity to consume proteins and its preferential use

of high-molecular-weight polymers [Cottrell & Kirchman, 2000b; González *et al.,* 2008] and the use of the proteorhodopsin to obtain additional energy from light [González *et al.,* 2008; Gómez-Consarnau *et al.,* 2007] contribute to make the Bacteroidetes an ecological-relevant marine taxa.

The growing interest of this phylum has led to the design through the years of several primers and oligonucleotides FISH probes for their accurate quantification [Manz *et al.,* 1996; Weller *et al.,* 2000; O´Sullivan *et al.,* 2002; Kirchman *et al.,* 2003; Chen *et al.,* 2006; Blümel *et al.,* 2007; Mühling *et al.,* 2008]. However, the probe CF319a, introduced 17 years ago by Manz *et al.* [1996] is still the most commonly used probe to quantify the abundance of marine Bacteroidetes. With the continuing expansion of sequence databases, a first evaluation of the *in silico* coverage of available probes for this taxon was performed in 2008 in which flaws related to a high number of outgroup hits were already pointed out [Amann & Fuchs, 2008]. Additionally, a recent review that considered not only the exponential growth of sequence databases, but also used *in silico* and culture tests, indicated that PCR primer CF968 had the best *in silico* coverage for Flavobacteria [Díez-Vives *et al.,* 2012, chapter 2], although it was not tested as a CARD-FISH probe and, therefore, a comparative quantitative study would be desirable. Similarly to the fact that distinct 16S rRNA gene sequencing primers may result in slightly distinct taxonomic assignation [Pommier *et al.,* 2007], comparative analyses of several specific-group probes with similar *in silico* efficiency and targeting different 16S rRNA gene variables (V) regions should be performed to test the adequacy of the different probes and to corroborate the abundance and dynamics of the different bacterial groups.

Here, we compared the abundance of marine Bacteroidetes populations by CARD-FISH using two specific probes that target different V regions of the 16S rRNA gene. We analyzed their differences with cultures, *in silico* and in situ comparing the most commonly used marine Bacteroidetes probe (CF319a) which targets the V3 region and the newly tested CF968 CARD-FISH probe binding to the V6 region. We compared their abundances in environmental samples globally distributed across the world's ocean and during two consecutive years in the Blanes Bay Microbial Observatory (BBMO; NW Mediterranean) in order to assess temporal dynamics. Additionally, due to the potential of metagenomics as a PCR-free technique for quantitative diversity studies, we conducted a comparative analysis between CARD-FISH counts and 16S rRNA gene sequences of Bacteroidetes retrieved from a set of marine metagenomes from the TARA Oceans global expedition [Karsenti *et al.,* 2011], and analyzed the specific phylogenetic taxa retrieved by each of the probes evaluated here.

Experimental procedures

Collection of environmental CARD-FISH samples

Environmental samples for CARD-FISH counts were collected during four oceanographic expeditions (Fig. 1; Table 1SM): cruises ATOS I and II in the Arctic and Antarctic oceans respectively, cruise MODIVUS, which covered both a horizontal transect and vertical profiles in the Northwestern Mediterranean Sea, and the TARA Oceans expedition which includes a selection of stations from the Western to Eastern Mediterranean Sea, the Red Sea and the North Indian Ocean (Fig. 1).

All seawater samples (200-µm pre-filtered) were fixed (1-12 h) with paraformaldehyde (1% final concentration) at 4℃ in the dark and gently filtered through 0.2 µm polycarbonate filters, rinsed with milli-Q water and stored at -20℃ until processing. In the BBMO samples, 25 ml were filtered onto 47 mm filters whereas for ATOS and MODIVUS samples 10 to 15 ml were filtered through 25 mm filters. In TARA Oceans, 10 ml of surface and DCM and 90 ml of mesopelagic samples were respectively filtered through 25 mm filters.

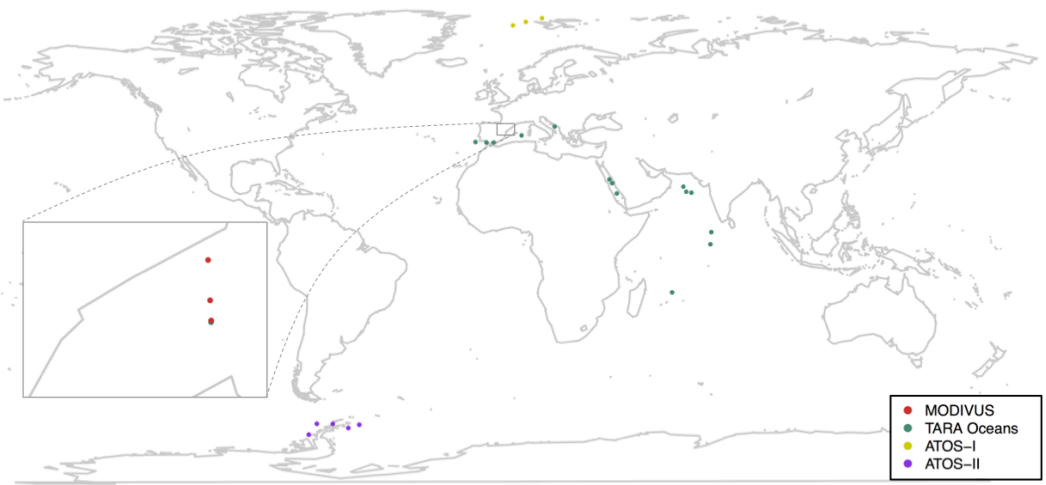


Figure 1. Map showing the geographic location of the samples considered in this study.

Table 1SM. Geographic Locations and sampling depth of the samples used for CARD-FISH comparison from 4 different oceanographic cruises.

St.	Latitude	Longitude	Cruise	Exped.	Ocean	Depth	Sampling Date	T (°C)
1	80°49.57′	013°14.22′	ATOSI-19.1a/ ATOSI-19.1b*	ATOSI	Arctic	SRF (5 m)	2009/09/13	
2	79°30.07′	007°29.64′	ATOSI-12.1a/ ATOSI-12.1b*	ATOSI	Arctic	SRF (5 m)	2009/09/13	
3	78°13.08′	002°51.48′	ATOSI-7.1a/ ATOSI-7.1b*	ATOSI	Arctic	SRF (5 m)	2009/09/13	
4	-63°51.82′	-051°51.94′	ATOSII-5.2a	ATOSII	Antarctic	SRF (5 m)	2009/09/13	
5	-65°01.17′	-055°45.43′	ATOSII-7.2a	ATOSII	Antarctic	SRF (5 m)	2009/09/13	
6	-63°33.97′	-061°16.09′	ATOSII-13.2a	ATOSII	Antarctic	SRF (5 m)	2009/09/13	
7	-65°31.75′	-066°56.41′	ATOSII-16.2a	ATOSII	Antarctic	SRF (5 m)	2009/09/13	
8	-67°22.24′	-069°48.37′	ATOSII-19.2a	ATOSII	Antarctic	SRF (5 m)	2009/09/13	
9	36°43.52′	-010°28.25′	TARA-St.3	TARA Oceans	Atlantic Ocean	SRF (5 m)	2009/09/13	
10	36°33.20′	-006°34.01′	TARA-St.4	TARA Oceans	Atlantic Ocean	SRF (5 m)	2009/09/15	
11	36°31.37′	-004°00.13′	TARA-St.6**	TARA Oceans	Mediterranean Sea	SRF (5 m)/ DCM (35 m)	2009/09/21	17,0
12	39°03.60′	005°55.89′	TARA-St.9**	TARA Oceans	Mediterranean Sea	SRF (5 m)/ DCM (55 m)	2009/09/28	24,5
13	40°38.97′	002°51.49′	TARA-St.10**	TARA Oceans	Mediterranean Sea	SRF (5 m)/ DCM (72 m)	2009/09/30	23,7
14	42°10.46′	017°43.16′	TARA-St.23**	TARA Oceans	Mediterranean Sea	SRF (5 m)/ DCM (56 m)	2009/11/18	17.1/ 16.0
15	41°24.96′	002°48.60′	MDV-CM	Modivus	Mediterranean Sea	SRF (5 m)/ DCM (44 m)	2007/09/21	
16	40°54.72′	002°50.58′	MDV-MD	Modivus	Mediterranean Sea	SRF (5 m)/ DCM (60 m)	2007/09/22	
17	40°39.60′	002°51.60′	MDV-D	Modivus	Mediterranean Sea	SRF (5 m)/ Deep (2000 m)	2007/09/23	
18	23°22.53′	037°14.46′	TARA-St.32	TARA Oceans	Red Sea	SRF (5 m)/ DCM (80 m)	2010/01/11	25,9
19	22°03.10′	038°13.00′	TARA-St.33	TARA Oceans	Red Sea	SRF (5 m)	2010/01/13	27,3
20	18°23.93′	039°51.76′	TARA-St.34	TARA Oceans	Red Sea	SRF (5 m)/ DCM (60 m)	2010/01/20	27,6
21	20°49.05′	063°30.72′	TARA-St.36	TARA Oceans	Arabian Sea	SRF (5 m)	2010/03/12	26,0
22	19°01.23′	064°33.22′	TARA-St.38	TARA Oceans	Arabian Sea	SRF (5 m)/ OMZ (350 m)	2010/03/15	26.3/ 14.7
23	18°40.27′	066°25.15′	TARA-St.39	TARA Oceans	Arabian Sea	SRF (5 m)/ OMZ (270 m)	2010/03/20	27.4/ 15.6
24	04°39.59′	073°29.13′	TARA-St.43	TARA Oceans	North Indian Ocean	SRF (5 m)	2010/04/05	30,0
25	00°20.25′	073°09.65′	TARA-St.46	TARA Oceans	North Indian Ocean	SRF (5 m)	2010/04/15	30,1
26	-16°48.50′	059°30.25′	TARA-St.49	TARA Oceans	North Indian Ocean	SRF (5 m)	2010/04/23	28,3

SRF: Surface; DCM: Deep Chlorophyll Maximum; OMZ: Oxygen Minimum Zone; Deep: Bathypelagic waters; (*) Two replicates were analyzed from these filters.(**) Three replicates were analyzed from these filters.

Catalyzed reporter deposition-fluorescence in situ hybridization (CARD-FISH)

CARD-FISH was carried out following the protocol described by Pernthaler *et al.,*[2004]. Two horseradish peroxidase (HRP)-labeled probes were used to analyze the relative abundance of Bacteroidetes cells in the seawater samples: CF319a [Manz *et al.,* 1996] and CF968 (the new CFB CARD-FISH probe). The CF319a probe was purchased from biomers.net (Ulm, Germany) and CF968 was synthesized with a 5'-aminolink (Thermo Fischer Scientific Ulm, Germany) and subsequently labeled with the HRP (Roche Diagnostic13 Boehringer, Meylan, France) according to Urdea *et al.,* [1988] and Amann *et al.,* [1992]. The filters were first coated with a thin layer of agarose, permeabilized with lysozime (10 mg ml^{-1}, 37℃, 1 h) and achromopeptidase (60 U ml^{-1}, 37℃, 30 min). Hybridizations were carried out on filter sections overnight at 35℃ and specific hybridization conditions were established by addition of formamide to the hybridization buffer (55% formamide for both probes). After hybridization, filter sections were washed for 15 min at 37℃ and the signal was amplified with Alexa 488-labeled tyramide for 15 min at 46℃.

Counterstaining of CARD-FISH preparations was done with 4,6-diamidino-2-phenylindole (DAPI, 1 µg ml^{-1}). Cell enumeration was carried out either manually or using an Olympus BX61 epifluorescence microscope applying a threshold-based macro [details in Massana *et al.,* 1997a] on Image Pro Plus software adapted for CARD-FISH enumeration. The results from the manual and semi-automatic counts were calibrated successfully. Between 500 and 800 DAPI-positive cells were counted manually and 1000 to 2000 in the semi-automatic protocol, in both cases in a minimum of 10 fields. The results are presented as percentage of positive hybridized cells to the total DAPI-stained prokaryotic cells.

CARD-FISH CF968 probe optimization

The CF968 (5'-GGTAAGGTTCCTCGCGTA-3') probe was tested against two Flavobacteriaceae cultures, MED 134 (*Dokdonia donghaensis*) and MED 152 (*Polaribacter dokdonensis*) from our culture collection isolated from the BBMO. During probe testing, temperature was kept constant at 35°C and the stringency varied with different formamide concentrations (from 0 to 60%) in the hybridization buffer. The optimal formamide concentration for the CF968 probe was found to be 55%.

In silico test of the CARD-FISH probes

Probe evaluation was carried out using the ARB software [Ludwig *et al.,* 2004] with the curated SILVA (http://www.arb-silva.de) rDNA database [Pruesse *et al.,* 2007; Quast *et al.,*

2013]. The release used was the "non-redundant" SSU Ref NR 104 dataset, which is a modified version of the full reference release in which redundancies (>99 % identity) were filtered out (about 50 % of the entries). The dataset (February 2011) contained 290,603 SSU rDNA sequences, including 24,319 Bacteroidetes. Target and non-target matches were evaluated with the 0-weighted option in ARB's probe match tool.

TARA Oceans collection of samples for metagenomes

Water samples were collected from two depths (Surface and deep chlorophyll maximum (DCM)) using a large peristaltic pump (TECH-POMPES, A40) or a rosette sampler, and subsequently separated into five size fractions (0.2-1.6, 0.8-5, 5-20, 20-180 and 180-2000 mm). For larger fractions (>5 mm) samples were obtained after filtering the collected seawater using the Gravity Plankton Sieving System, a serial of nets of different size fractions [Karsenti *et al.,* 2011]. To obtain the fractions smaller than 5 µm, 100 L of seawater were first filtered through 200 and 20 µm meshes to remove large plankton. Further filtering was done by pumping water serially through 142 mm filters of either 1.6 and 0.2 µm or 5 and 0.8 µm pore size with a peristaltic pump (Masterflex, EW-77410-10). After filtration, filters were flash frozen and kept for about 4 weeks at -20°C on the schooner and then at -80°C in the laboratory until the DNA extraction was performed.

Bacteroidetes 16S rRNA sequences extracted from TARA Oceans metagenomes

Twenty-seven metagenomes from the TARA Oceans expedition were sequenced by 454 FLX-Ti pyrosequencing by the French National Genome Sequencing Center Genoscope, France. All details of DNA extraction, sequencing and accession number for these metagenomes can be found elsewhere [Hingamp *et al.,* 2013]. These selected marine metagenomes correspond to 13 distinct TARA Oceans sampling stations including samples from surface, deep chlorophyll maximum (DCM) and oxygen minimum zones (OMZ). These TARA Ocean metagenomes belonged to different plankton size fractions: 0.2-1.6, 0.8-5, 5-20, 20-180 and 180-2000 mm (details in Table 3SM).

The 16S rRNA gene sequences were extracted from these TARA Ocean metagenomes using a customized version of rna_hmm.py, a python wrapper script that uses hidden Markov models (HMM) to identify and annotate metagenomic reads that contain fragments of rRNA genes [Huang *et al.,* 2009]. Briefly, HMMs for 16S rRNA sequences were generated using hmmbuild [version 3.0; Eddy, 2011] on aligned 16S rRNA gene sequences that were downloaded from the SILVA database [Quast *et al.,* 2013]

(http://www.arb-silva.de/projects/ssu-ref-nr/) (version SSU r104; alignment quality = 100; sequence length >= 1200; sequence quality >=75; restrict search to SILVA). Each candidate read was trimmed (and reverse-complemented if necessary) to segments that were predicted to contain 16S rRNA sequences and a length filter of 250 bp for 454 reads was applied.

Table 3SM. TARA Oceans metagenomes used in this study to extract the 16S rRNA sequences from Bacteroidetes.

TARA ST	Latitude	Longitude	Ocean	Depth*	Sampling Date	Plankton Size Fraction (µm)
TARA-St.3	36°43.52′	-010°28.25′	Atlantic Ocean	SRF (5 m)	2009/09/13	0.22-1.6
TARA-St.4	36°33.20′	-006°34.01′	Atlantic Ocean	SRF (5 m)	2009/09/15	0.22-1.6
TARA-St.6	36°31.37′	-004°00.13′	Mediterranean Sea	SRF (5 m)	2009/09/21	0.22-1.6
TARA-St.7	37°02.32′	-001°56.99′	Mediterranean Sea	SRF (5 m)	2009/09/23	0.22-1.6
TARA-St.7	37°02.32′	-001°56.99′	Mediterranean Sea	SRF (5 m)	2009/09/23	0.8-5
TARA-St.7	37°02.32′	-001°56.99′	Mediterranean Sea	SRF (5 m)	2009/09/23	5- 20
TARA-St.7	37°02.32′	-001°56.99′	Mediterranean Sea	SRF (5 m)	2009/09/23	20-180
TARA-St.7	37°02.32′	-001°56.99′	Mediterranean Sea	SRF (5 m)	2009/09/23	180-2000
TARA-St.7	37°02.32′	-001°56.99′	Mediterranean Sea	DCM (42 m)	2009/09/23	0.22-1.6
TARA-St.7	37°02.32′	-001°56.99′	Mediterranean Sea	DCM (42 m)	2009/09/23	0.8-5
TARA-St.7	37°02.32′	-001°56.99′	Mediterranean Sea	DCM (42 m)	2009/09/23	5- 20
TARA-St.7	37°02.32′	-001°56.99′	Mediterranean Sea	DCM (42 m)	2009/09/23	20-180
TARA-St.7	37°02.32′	-001°56.99′	Mediterranean Sea	DCM (42 m)	2009/09/23	180-2000
TARA-St.23	42°10.46′	017°43.16′	Mediterranean Sea	SRF (5 m)	2009/11/18	0.22-1.6
TARA-St.23	42°10.46′	017°43.16′	Mediterranean Sea	DCM (56 m)	2009/11/18	0.22-1.6
TARA-St.23	42°10.46′	017°43.16′	Mediterranean Sea	Mixed	2009/11/18	0-0.22
TARA-St.30	33°55.07′	032°53.62′	Mediterranean Sea	SRF (5 m)	2009/12/14	0.22-1.6
TARA-St.30	33°55.07′	032°53.62′	Mediterranean Sea	Mixed	2009/12/14	0-0.22
TARA-St.31	27°08.10′	034°48.40′	Red Sea	SRF (5 m)	2010/01/09	0.22-1.6
TARA-St.36	20°49.05′	063°30.72′	Arabian Sea	SRF (5 m)	2010/03/12	0.22-1.6
TARA-St.38	19°01.23′	064°33.22′	Arabian Sea	SRF (5 m)	2010/03/15	0.22-1.6
TARA-St.38	19°01.23′	064°33.22′	Arabian Sea	OMZ (350 m)	2010/03/15	0.22-1.6
TARA-St.39	18°40.27′	066°25.15′	Arabian Sea	SRF (5 m)	2010/03/20	0.22-1.6
TARA-St.39	18°40.27′	066°25.15′	Arabian Sea	OMZ (270 m)	2010/03/20	0.22-1.6
TARA-St.43	04°39.59′	073°29.13′	North Indian Ocean	SRF (5 m)	2010/04/05	0.22-1.6
TARA-St.46	00°20.25′	073°09.65′	North Indian Ocean	SRF (5 m)	2010/04/15	0.22-1.6
TARA-St.49	-16°48.50′	059°30.25′	North Indian Ocean	SRF (5 m)	2010/04/23	0.22-1.6

* Different depths were sampled; SRF: Surface; DCM: Deep Chlorophyll Maximum; Mixed: Surface and DCM together; OMZ: Oxygen Minimum Zone

A total of 7,182 16S rRNA reads were retrieved from the TARA Oceans metagenomes built into a database using ARB software (version 5.1) [Ludwig *et al.,* 2004]. Using the match probe tool of ARB, only the 16S rRNA reads with perfect match (zero mismatches) with either the CF319a or the CF968 probe sequences were enumerated and included into our next analyses. First, the taxonomic assignment of the retrieved 16S rRNA reads was carried out using the classifier tool of the Ribosomal Database Project web (http://rdp.cme.msu.edu/classifier/classifier.jsp) [Cole *et al.,* 2008]. Secondly, a phylogenetic tree was constructed. Sequences were first aligned using the SINA multiple alignment algorithm (http://www.arb-silva.de/aligner/) [Pruesse *et al.,* 2012] and imported into ARB software [Ludwig *et al.,* 2004]. The phylogentic tree was performed by the insertion of sequences into the ARB-formatted SILVA 108 release database using the parsimony algorithm. The 16S rRNA sequences from TARA Oceans metagenomes data used in this study are accessible from the Sequence Read Archive (SRA) of the European Nucleotide Archive (ENA) through accession number ERA155563, ERA155562 and ERA241291.

Statistical Analyses

Paired Student's t-test was used to test the significance of the dissimilarities between the percentages of hybridization of the two probes. Statistical analyses were carried out in the R environment (www.r-project.org) using the Vegan package [Oksanen *et al.,* 2009].

Results and Discussion

In silico comparison of the CF319a and CF968 Bacteroidetes probes.

We compared the usefulness of the most commonly used CFB probe (CF319a) to a new probe (CF968). The CF968 existed previously as a PCR and PCR-DGGE primer [Chen *et al.,* 2006; Muhling *et al.,* 2008; Díez-Vives *et al.,* 2012] and was also reported as qPCR primer [De Gregoris *et al.,* 2011]. However, this CF968 had not been tested as a CARD-FISH probe before. *In silico* testing against the version of the SILVA dataset (dataset February 2011) confirmed that one of the main flaws of probe CF319a is its high number of outgroup hits, mainly distributed within different families of the phylum Crenarchaeota and the class Proteobacteria (Fig. 2C). The CF319a probe presented about 92% coverage of the sequences belonging to the class Flavobacteria, and 34% of the members of the classes Sphingobacteria and Cytophaga [Díez-Vives *et al.,* 2012].

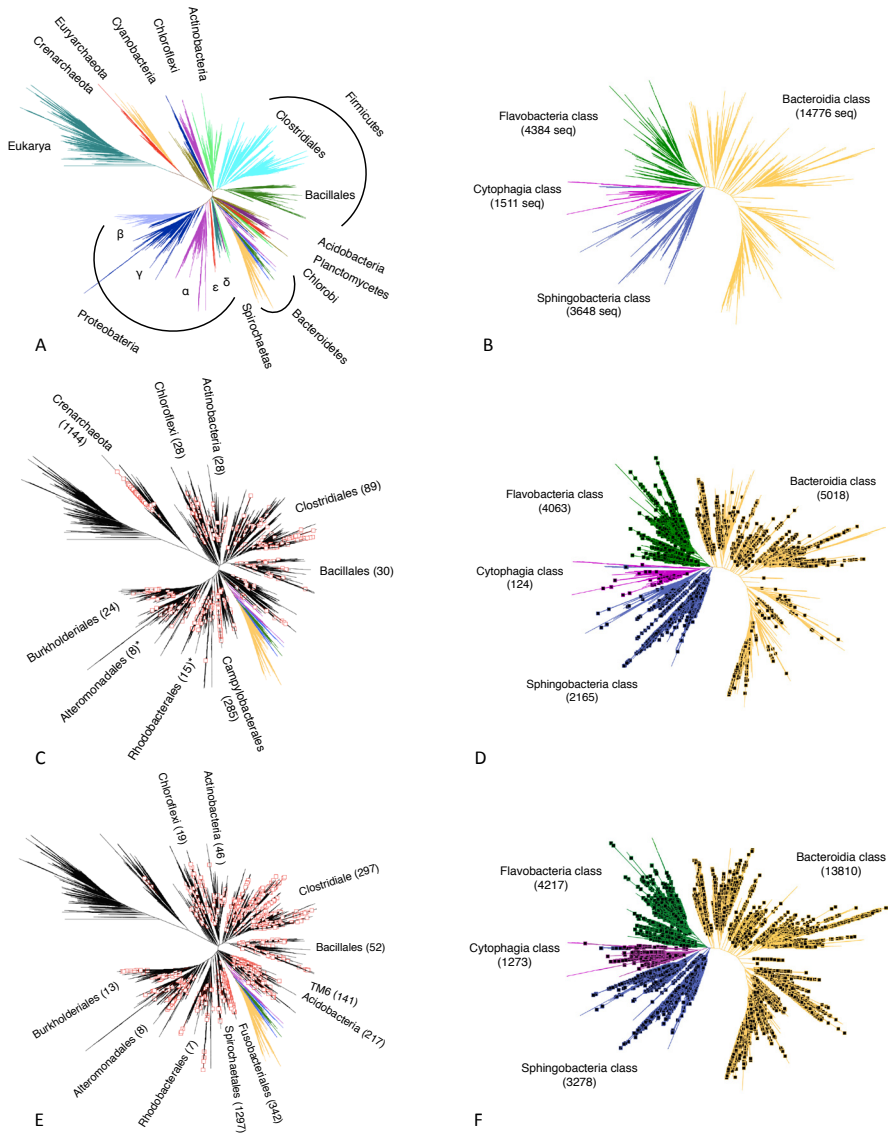


Figure 2. Graphical representation of the sensitivity and specificity of the CF319a and CF968 probes *in silico* tests. A: Phylogenetic tree of the 290,603 16S rRNA sequences of all domains included in the curated SILVA dataset. B: Detail of the Bacteroidetes tree that includes 24,319 sequences classified in the four different Classes. Major groups are identified and branches colored differently. C-D: Predicted hits of the CF319a in the global dataset, classified as outgroup hits (C) and positive (target) hits within Bacteroidetes (D). E-F: Predicted hits of the CF968 in the global dataset, outgroup hits (E) and positive hits within Bacteroidetes (F). Outgroups hits are represented as red squares (C, E) and positive hits as black squares (D, F). Numbers in brackets represent either the outgroup (C and E) or the target group hits (D and F) using the 0.5 weight filter. Taxa with asterisks indicate those bacterial groups from where cultures were used in the CARD-FISH tests presented in Fig. 3.

These *in silico* observations had already been reported by Amann and Fuchs [2008], and clearly show that the probe CF319a does not fully detect all the members of the phylum Bacteroidetes. On the contrary, the new CARD-FISH probe CF968 exhibited better *in silico* performance compared with other published Bacteroidetes probes, resulting in 93% positive hits for Flavobacteria (similar to CF319a) but with better coverage within the Sphingobacteria and Cytophaga classes with 90 and 84% respectively. The *in silico* outgroup hits in the compared probes were found in different taxa (Fig. 2E). The CF319a outgroup hits were within the Crenarchaeota and the Proteobacteria, mostly related to the alphaproteobacterial groups Rhodobacteraceae and the gammaproteobacterial Alteromonadaceae [Díez-Vives *et al.,* 2012]. Instead, the CF968 outgroups were mostly constrained within Spirochaetas, Acidobacteria, and Firmicutes (all of them rare groups in marine bacterioplankton communities) and some Proteobacteria (Fig. 2E) [Díez-Vives *et al.,* 2012].

Although *in silico* analyses of both probes showed similar performance for Flavobacteria, the non-specific probe binding to other taxa, or the low accessibility of the probe to the ribosomal gene might have an effect in the abundance estimations when tested in environmental samples. To first test the latter possibility, we checked whether CF319a and CF968 exhibited different accessibility in the secondary structure of the 16S rRNA gene based on a previous study on relative fluorescence hybridization intensities expressed as different class of brightness levels (from Class I to Class VI) [Behrens *et al.,* 2003]. However, neither the probe CF319a nor the CF968 showed low accessibility (class V) or block positions (class VI) on the 16S rRNA gene structure (data not shown). Both probes classify as Class III, so hypothetical differences in abundance among these probes are not likely due to different accessibility of the probe.

Both probes were also evaluated with a mechanistic FISH model based on the thermodynamics of nucleic acid hybridization as described by Yilmaz *et al.,* [2004]. In this model, the affinity of a probe to its folded nucleic acid target is defined as the overall Gibbs free energy change ($\Delta G^\circ_{overall}$). A lower (more negative) $\Delta G^\circ_{overall}$ represents a greater potential for the formation of a probe-RNA duplex and thus higher brightness. Using the mathFISH webtool (http://mathfish.cee.wisc.edu) [Yilmaz *et al.,* 2011] we found that $\Delta G^\circ_{overall}$ for CF319a is -15.2Kcal mol^{-1} and -16.1 Kcal mol^{-1} for CF968 (this is an average of $\Delta G^\circ_{overall}$ for each probe with 10 different sequences). Considering that a threshold of $\Delta G^\circ_{overall}$ -13.0 Kcal mol^{-1} was proposed in the design of FISH probes to maximize hybridization efficiency without compromising specificity, we can say that both probes seem also equivalent from the thermodynamic point of view.

Unspecific hybridization of the CF319a probe.

Some of the CF319a outgroups hits found *in silico* were related to the Rhodobacteraceae and Alteromonadaceae. In a previous study, we used the PCR primers sets CF319a-907RM and 357F-CF968R to retrieve and sequence bands in denaturing gradient gel electrophoresis [Díez–Vives *et al.,* 2012]. The CF319a and CF968 PCR primers correspond to the same position and sequence than their respective oligonucleotide CARD-FISH probes. Band sequencing resulted in the retrieval of several sequences related to Rhodobacteraceae and Alteromonadaceae when using the primer set CF319a-907RM. In contrast, all the sequences recovered from the set GC357F-CF968R were related to marine Flavobacteria [Díez–Vives *et al.,* 2012]. To further confirm these findings and the *in silico* results, we evaluated the specificity of the CF319a and CF968 probes in five marine isolates of our culture collection (BBMO). In particular, we used Rhodobacter-like ZOCON-B2 and ZOCON-B3, and the Alteromonadales-like isolates ZOCON-C2, ZOCON-D11 and ZOCON-F9 [Salazar *et al.,* unpublished]. We performed CARD-FISH hybridizations with the Bacteroidetes specific probe CF319a and our new CF968 probe and the corresponding positive probes for the cultures tested: ALP968 for Alphaproteobacteria (Rhodobacter-like isolates) and and GAM42a for Gammaproteobacteria (Alteromonadales-like isolates) (Fig. 3A). As suspected, all cultures tested were positively hybridized with the CF319a probe whereas negative hybridizations were observed for CF968 (Fig. 3A). These results confirmed that CF319a hybridizes with several phylotypes of the Rhodobacteraceae and Alteromonadales groups not only *in silico* but also in culture.

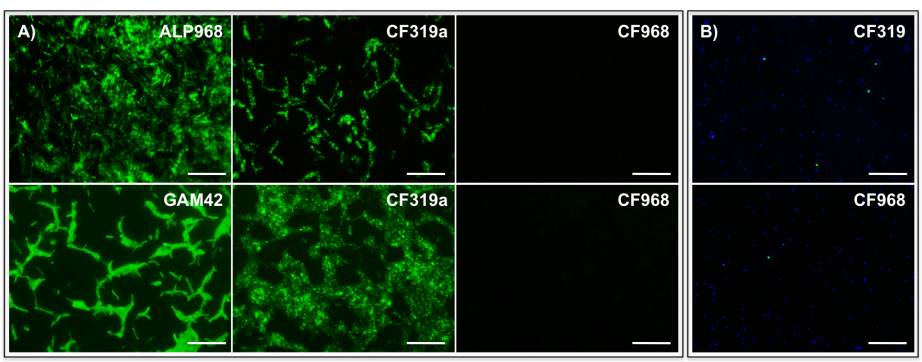


Figure 3. (A) CARD-FISH photomicrographs of alphaproteobacterial isolate ZOCON-B3 (upper panel) and the gammaproteobacterial isolate ZOCON-D11 (lower panel) hybridized with HRP-labeled oligonucleotide probes. For each panel, the left image shows the positive control hybridization, the central one the CF319a probe, and the right panel the CF968 probe. (B) Composed photomicrographs of marine bacteria stained with DAPI and hybridized with HRP-labeled CF319a (upper image) and CF968 (lower image) oligonucleotide probes from a Red Sea environmental sample (St. 32-Surface, TARA Oceans). Scale bar = 5 µm.

These two groups of marine bacteria are relevant in most marine environments. The Rhodobacteraceae can account for up to 20% of total bacterioplankton in coastal areas [Buchan *et al.,* 2005] and in our site BBMO are about 7% [Alonso-Sáez *et al.,* 2007]. The Alteromonadales can also be abundant, usually associated to organic matter aggregates or under certain conditions such as post-algal bloom events [Acinas *et al.,* 1999; Alonso-Sáez *et al.,* 2007; Crespo *et al.,* 2013]. Therefore, the non-specific hybridization detected using CF319a with non-Bacteroidetes phylotypes may blur correct quantification estimates, particularly when these taxa are abundant in the marine environment.

Abundance of marine Bacteroidetes in environmental samples.

We compared the performance of both probes in 37 environmental samples from contrasting marine environments including the Arctic, Antarctic, Northern Indian Ocean, and the Red and Mediterranean Sea (Table 1SM; Fig. 4). Positive cell hybridization was found in all samples with both probes, and one of the environmental samples tested is shown in Fig. 3B. The quantification of Bacteroidetes included samples from deep-chlorophyll maximum (DCM), oxygen minimum zones (OMZ; 270 and 350 m) and bathypelagic waters (2000 m) (Fig. 4). Our data indicate that CF319a detects higher cell numbers in 70% of the marine samples tested (Fig. 4, Fig. 5). Although differences between both probes at single sampling points seem not significant, some clear differences were observed in specific samples from Mediterranean waters (TARA-St.23, MDV-St.CM), Red Sea (TARA-St.32, TARA-St.33), North Indian Ocean (TARA-St.46), Arctic (ATOSI-St.7.1a) and Antarctic (ATOSII-St.5.2a) mostly at surface and DCM waters and with differences accounting for up to 90% (Fig. 4, Fig. 5). Higher abundances with the new probe CF968 compared with CF319a were also detected in one surface Mediterranean sample, MDV-St.CM where 8% and 15% of cells were hybridized with CF319a and CF968 respectively, and in a few other stations with minor differences (Fig. 4).

The highest abundances of Bacteroidetes were found in Antarctic waters, where 43 and 46% of cells hybridized with CF968 and CF319a, respectively (Fig. 4). This large numbers were associated with a bloom of diatoms, specifically of *Thalassiosira* [Ruiz-González *et al.,* 2012c]. A high abundance of marine Bacteroidetes in cold waters associated to algal blooms has been recurrently reported in polar regions with values up to 72%, during a *Phaeocystis* bloom measured with FISH [Simon *et al.,* 1999; Ruiz-González *et al.,* 2012c] and between 17-30% estimated with qPCR [Abell & Bowman, 2005b].

We also compared Bacteroidetes abundances along the Mediterranean Sea from west to east in surface and DCM samples for a set of stations from the TARA Oceans global expedition (Fig. 4, Table 2SM). The main differences between both probes (CF319a *vs.* CF968) were found in coastal stations: at the surface of St.10 with a 12% and 2% respectively, and St.6, St.9 and St.23 at the DCM with differences between 30% and 50% (Fig. 4). Furthermore, in the vertical gradient sampled during the transect MODIVUS in the Mediterranean Sea from the surface down to 2000 m, Bacteroidetes had maximum abundances at the surface (5 m) and DCM (44 m) with on average contributions of 12-14% of total DAPI counts for CF968 and CF319a. In contrast, much lower numbers were found at deeper waters (500 m and 2000 m), with a minimum of 3% and 1% of DAPI counts for probes CF319a and CF968, respectively (Fig. 4).

Figure 4. Histogram showing the percentage of cells hybridized with the CF319a and CF968 probes in 5 different oceanographic regions.

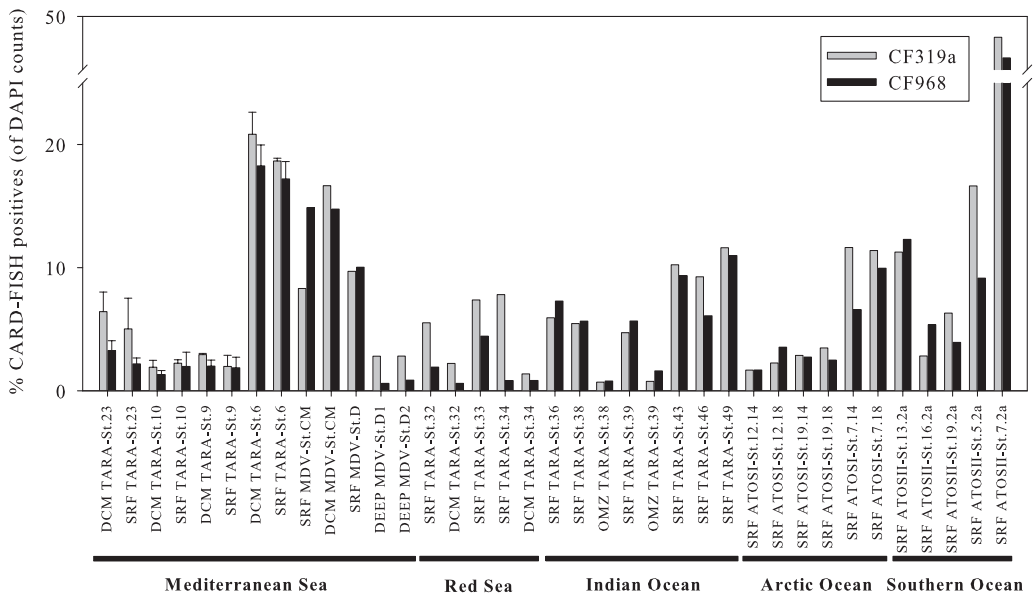


Table 2SM. Comparative abundances of Bacteroidetes (using probes CF319a and CF968), Gammaproteobacteria (probe Gam42a), *Alteromonadaceae* (probe Alt1413) and *Rhodobacteraceae* (probe Ros537) in samples from the Mediterranean Sea collected by the TARA Oceans expedition. Numbers in bold represent the highest abundance of Bacteroidetes that coincided with the highest abundance of *Rhodobacteraceae* (shown in red).

TARA	CF319a (%)	CF968 (%)	Gam42a (%)	Alt1413 (%)	Ros537 (%)
ST3 SRF	1	1	4.2	1	1
ST4 SRF	3.9	0	1	1	1
ST6 SRF	**21.4**	**17.4**	18.5	1	**5.8**
ST6 DCM	**22**	**16.1**	21.9	3.5	**5.1**
ST9 SRF	6.5	1.3	5.3	1.7	1.7
ST9 DCM	4.2	1.1	1.8	1	1
ST10 SRF	12	2.3	1.8	1	
ST10 DCM	1.4	1	1.9	1.2	1
ST 23 SRF	7	2	7	1	3
ST23 DCM	8	3	5	2	2

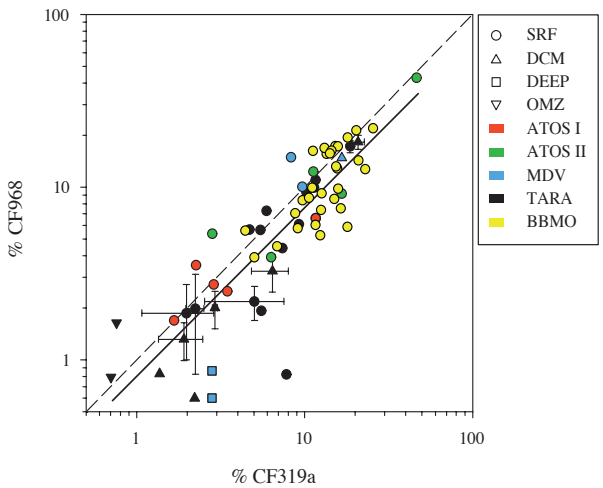


Figure 5. Comparison between counts with probes CF319a and CF968 in the same samples of Figure 4 but also including the two years of temporal dynamics at the BBMO.

Using the metagenomes sequenced by the TARA Oceans expedition, we compared the Bacteroidetes abundance with two PCR-free quantitative approaches: cell abundance counts using both CARD-FISH probes, and the relative abundance of the 16S rRNA gene sequences that belonged to Bacteroidetes for a total of 27 metagenomes from TARA stations (Fig. 1SM). Most relative abundances by both probes were within 3% of the metagenomes 16S rDNA abundances, and the two relationships were almost indistinguishable, with non significantly different slopes, although the correlation was slightly better for probe CF968, and the ordinate was higher for probe CF319a than for probe CF968 (Fig. 1SM), reflecting that both probes provided values close to the metagenome data.

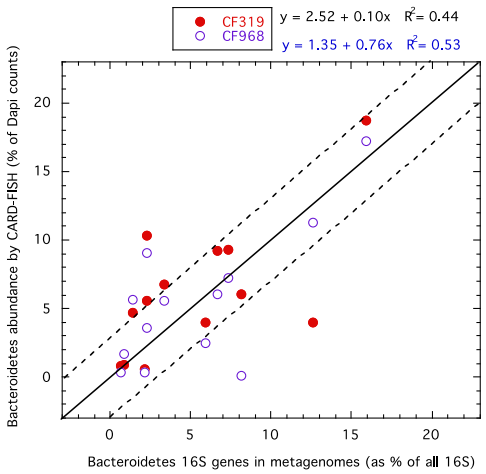


Figure 1SM. Plot comparing the relative abundance of Bacteroidetes between CARD-FISH and 16S rRNA sequences that matched perfectly to each of the sequence probes extracted from the TARA Oceans metagenomes. A linear regression model was applied and 95% confidence intervals were computed for the slope and statistically shown not to be different to a value of 1 (CI: 0.75-1.35 for CF319a and CI: 0.6-1.21 for CF968).

Temporal dynamics of Marine Bacteroidetes at the BBMO

We also investigated the abundance of marine Bacteroidetes with probes CF319a and CF968 during a two-year temporal series at the BBMO. Interestingly, it was at this temporal scale where we noticed pronounced differences between both Bacteroidetes probes (Fig. 6). While the quantification of Bacteroidetes with the CF319a probe overlooked any seasonality displaying very similar numbers through the year with an average of 16% for 2008 and 10% for 2009 over the two years, the CF968 probe unveiled clear seasonal patterns with a significant decrease in the number of hybridized cells for the group of samples between August to January in 2008 and at least from June to November in 2009 (Fig. 6). In fact, the hybridization rates of the two probes were significantly different for this pool of samples (paired Student's t-test, $p<0.001$).

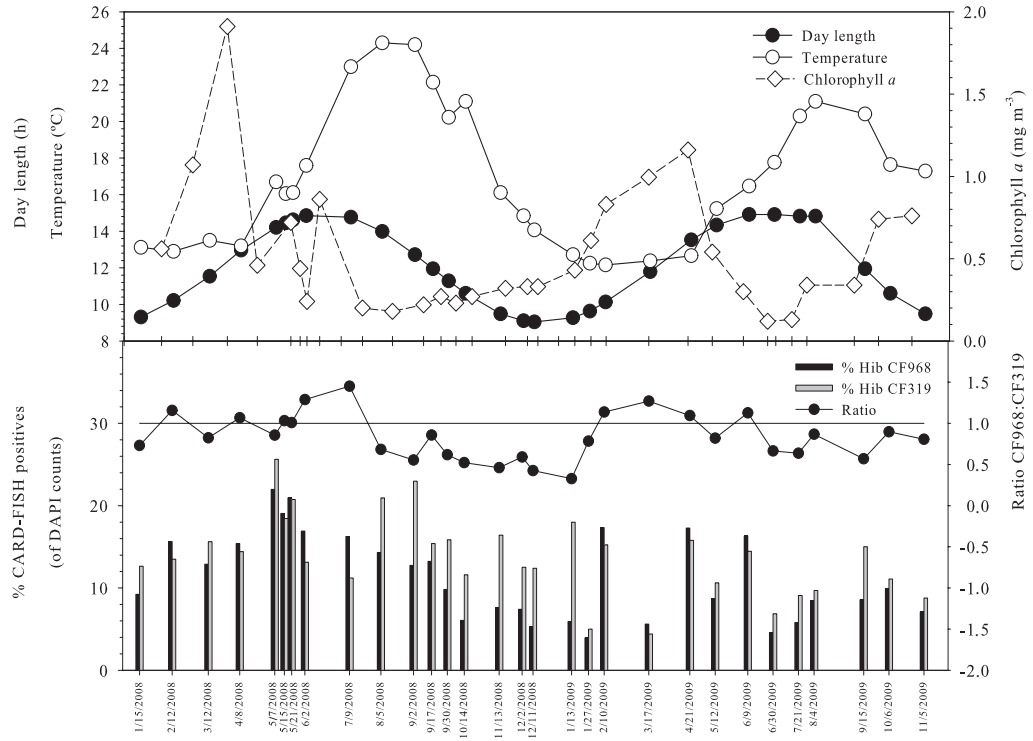


Figure 6. Upper panel: seasonal evolution of temperature, day length and chlorophyll a in Blanes Bay between January 2008 and November 2009. Lower panels: Histograms showing the seasonal variation in the percentage of cells hybridized with the CF319a and CF968 probes during two years in BBMO. On top, the evolution of the ratio between counts with probe CF968 and those with probe CF319a.

We postulated the hypothesis that perhaps the differences between both probes during these samples coincided with the abundance of some members of the Rhodobacteraceae and Alteromonadales since we had observed unspecific hybridization of CF319a towards cultures belonging to these taxa. During the two years survey at BBMO, Gammaproteobacteria (which include Alteromonadales) ranged between 5-28 % (GAM42a probe), and Rhodobacteraceae (Ros537 probe) comprised between 1-17% of total DAPI counts, and thus, these groups were relevant components in this coastal site (Fig. 2SM) [Ruiz-González *et al.,* 2012a]. However, we did not find any clear patterns between the abundances of these groups and those of Bacteroidetes with the BBMO data. In contrast, we found some interesting results in two of the Mediterranean samples from TARA Oceans (TARA St.6 surface and DCM, located close to the Gibraltar strait). Specifically, we found that the highest difference between the two probes (21.4 -22%

with one probe CF319a, and 16.1-17.4% with probe CF986) co-occurred with significant (relatively high) contributions of Rhodobacteraceae (5.1 and 5.8%) (Table 2SM). These results suggest that the discrepancy between both Bacteroidetes probe estimations (up to 6%) in these samples could partially be due to the unspecific hybridization of the CF319a probe to members of Rhodobacteraceae that were highly represented in those samples. Therefore, the unspecific hybridization of CF319a may have an impact on the enumeration of marine Bacteroidetes in some samples.

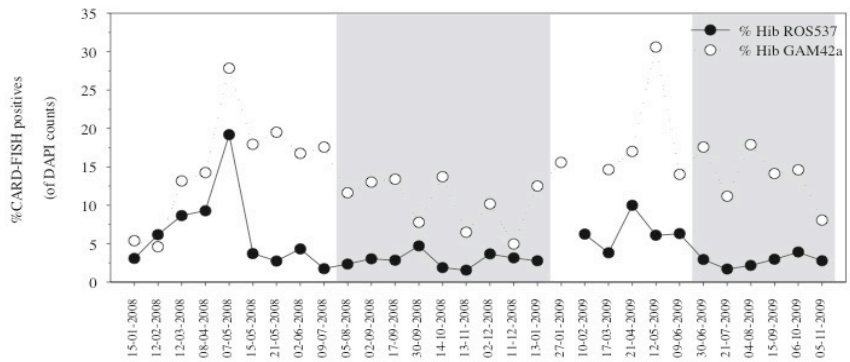


Figure 2SM. Quantification of members of Rhodobacteraceae (probe Ros537) and Gammaproteobacteria (probe GAM42a) during two years in the BBMO. Grey sections in the graph represent the periods of summer-fall in 2008 and part of the 2009 in which the Bacteroidetes CF319a and CF968 probes were significantly different.

CF319a and CF968 probes target different Bacteroidetes phylotypes

A clear discrepancy between the CF319a and CF968 probes was observed when comparing CARD-FISH counts for a variety of marine sampling points at spatial and temporal scales (Fig. 5). Our findings suggest that CF319a might overestimate Bacteroidetes abundance by unspecific hybridization (Figs. 3, 1SM). Nevertheless, the difference between both probes could also be due to the differential target of phylotypes. In fact, we also observed that in some cases CF319a might also underestimate Bacteroidetes compared with CF968 (Fig. 5) perhaps due to a lower target diversity coverage range of the probe CF319a.

To test this possibility, we first built a 16S rRNA gene dataset with a total of 7182 Bacteroidetes sequences retrieved from the 27 TARA Oceans metagenomes (Table 3SM). We then conducted an *in silico* test with these 16S rRNA metagenomic sequences (Fig. 3SM) using the match probe tool of ARB to recruit sequences matching perfectly (zero mismatches) with either the CF319a or the CF968 probe sequences.

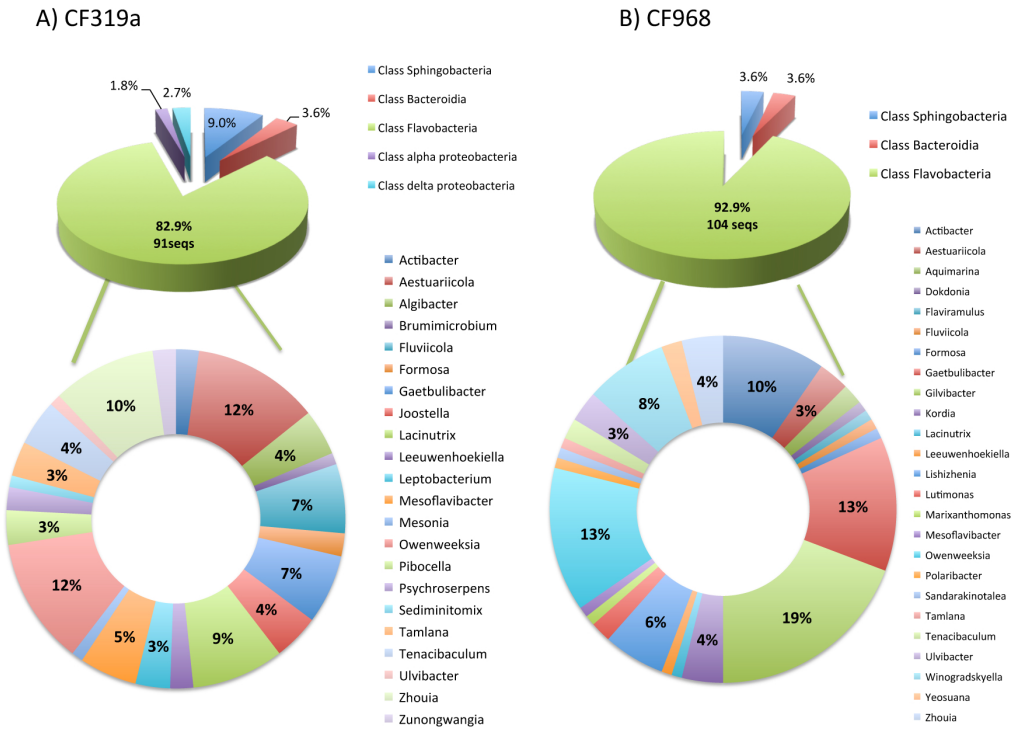


Figure 3SM. A and B) The upper pie chart represents the RDP taxonomy classification of the 16S rRNA sequences retrieved by the CF319a probe (111 sequences) and the CF968 probe (112 sequences) from 27 TARA Oceans metagenomes. The ring in the bottom shows the main genera of Flavobacteria (represented in green in the upper pie) captured by each probe. The percentage numbers in the ring represent the relative abundances, and only those with %>4 are shown in the graphic.

We retrieved a total of 111 and 112 sequences that matched perfectly to CF319a and CF968 respectively. The RDP taxonomic assignment of those sequences targeting both probes confirmed that: (i) CF319a hybridized nonspecifically with members of the Alphaproteobacteria (Rhodospirillaceae) and Deltaproteobacteria (Desulfobacterales, Fig. 3ASM) and (ii) the two probes had different targets in environmental samples within the Bacteroidetes. The CF319a probe matched 22 different genera while the probe CF968 targeted a total of 25 genera (Fig. 7). Although the RDP taxonomic assignation at the genus level has to be taken with caution, it is notorious that only 38% of the Bacteroidetes genera were retrieved by both probes, and that many of the shared genera were detected in different frequency by each probe (Fig. 7). Moreover, while 9 genera

were detected only by the CF391a probe, 12 were covered exclusively by the CF968 probe (Fig. 7). Among the former, genera with abundances > 3% were *Algibacter* (4%), *Joostella* (4%), *Leptobacterium* (3%) and *Pibocella* (3%). On the other hand, the genera detected exclusively by CF968 with contributions >3% were *Gilvibacter* (12%), *Kordia* (4%), *Lishizhenia* (6%), and *Winograsdkyella* (8%).

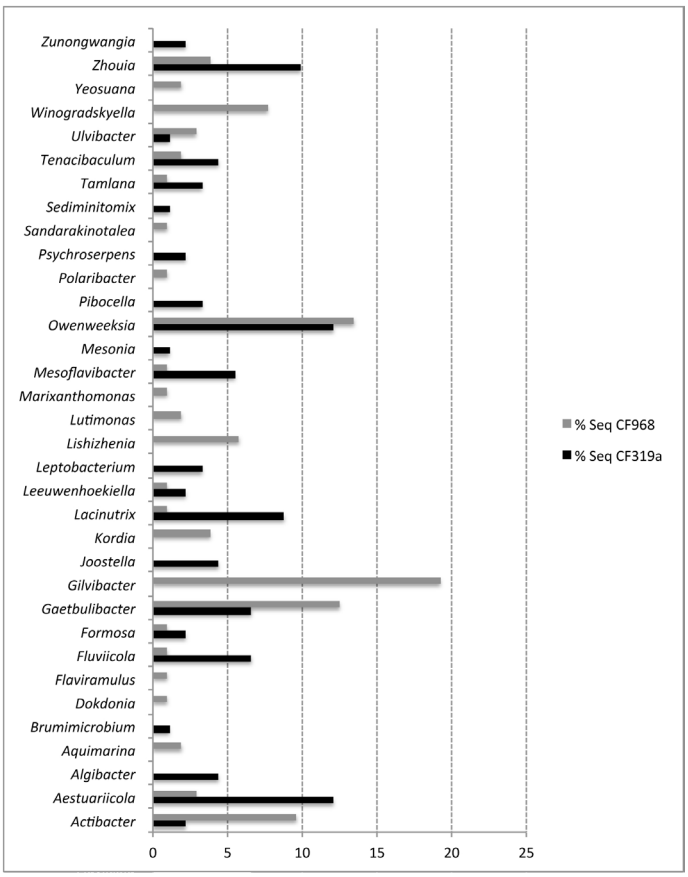


Figure 7. RDP-based taxonomy distribution at the genera level of the Bacteroidetes 16S rRNA sequences retrieved from 27 TARA Oceans metagenomes with perfect match on the CF319a or to the CF968 probe sequences.

To further investigate the specific characteristics of the phylotypes retrieved by each probe, we constructed a phylogenetic tree by mapping the specific 16S rDNA Bacteroidetes sequences (111 and 112 sequences retrieved by CF319a and CF968, respectively) on a SILVA reference tree (Fig. 8). The Bacteroidetes sequences belonged to marine metagenomes globally distributed from 3 depths (SRF, DCM and OMZ) and three distinct size fractions (0.22-1.6 µm; 5-20 µm, and 180-2000 µm) out of the five size fractions available. Although many of the monophyletic clusters were composed by phylotypes targeted by both probes, some clear differences also emerged. Interestingly,

there were more monophyletic clusters among the sequences detected only by CF968 (clusters A, C, D, and E, in Fig. 8) than among those retrieved only by CF319a (cluster B). Cluster A was found in surface waters from the Mediterranean St.7 (close to Gibraltar strait) and within the size fraction 5-20 µm. However clusters C, D and E were more cosmopolitan, being found in surface waters and at the DCM (cluster C) and widely distributed across several TARA stations (Fig. 8). Therefore, these data confirm the different targets of the CF319a and CF968 probes and the lower diversity coverage of the CF319a.

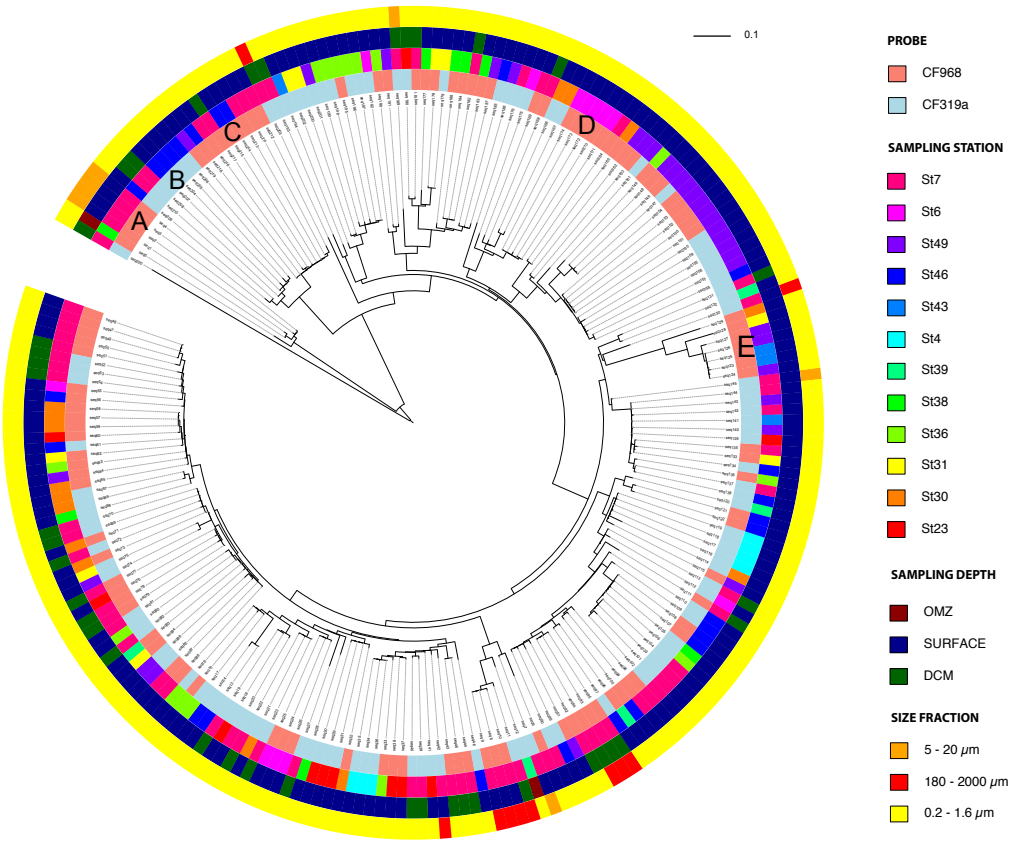


Figure 8. Phylogenetic tree of the Bacteroidetes using the 16S rRNA sequences from the TARA Oceans metagenomes that were targeted by each probe (light blue CF319 and light red CF968 inner ring). The second ring shows each TARA station analyzed. The third ring represents the three depths sampled at TARA St (Surface, DCM and OMZ). Finally, the outer ring displays the three different size fractions with positive hits (0.22-1.6 µm; 5-20 µm, and 180-2000 µm).

Conclusions

We tested the CF968 probe as a new CARD-FISH probe and the results show that it exhibits higher specificity for marine Bacteroidetes and better non-specific hybridization for other groups than probe CF319a. Contrarily, we found unspecific hybridizations of the CF319a probe in cultures of Alphaproteobacteria (Rhodobacteraceae) and Gammaproteobacteria (Alteromonadaceae) and the analyses of 16S rRNA metagenomic sequences targeting this probe sequence revealed other potential outgroup hits for CF319a related to Alphaproteobacteria (Rhodospirillaceae) and Deltaproteobacteria (Desulfobacterales). We also demonstrate that both probes target different phylotypes leading to inconsistent estimates when temporal studies are performed. In view of our results, we suggest the use of CF968 as an alternative CARD FISH probe for the enumeration of marine Bacteroidetes as it provides higher coverage and better specificity than the CF319a probe. Finally, testing multiple probes is a recommended strategy in order to discard potential inaccuracies in cell detection, but also to increase resolution in the discrimination of patterns within bacterial populations.

Acknowledgments

We thank the TARA Oceans expedition (http: www.taraoceans.org) for collecting some of the samples used in these analyses. This is contribution no. XXX of the Tara Oceans Expedition 2009-2012. This work was supported by grants, BACTERIOMICS (CTM2010-12317-E) and MicroOcean PANGENOMICS (CGL2011-26848/BOS) to SGA and HotMix (CTM2011-30010/MAR) to JMG from the Spanish Ministry of Science and Innovation (MICINN). SGA was supported by a RyC contract Spanish Government of Science and Innovation (MICINN) and FP7-OCEAN-2011 (MicroB3) from the EU. Financial support was provided by Ph.D. JAE-Predoc fellowships from CSIC to CDV, CRG and GS. FMCC was supported by a FPI fellowship (MICINN). Postdoctoral Juan de la Cierva research contracts also from MICINN were held by IF and HS. HS also benefited from a fellowship from the Brazilian 'Ciências sem Fronteiras' Program from CAPES (BJT 013/2012).

Chapter 2

Evaluation of marine *Bacteroidetes*-specific primers for microbial diversity and dynamics studies

Abstract

Assumptions on the matching specificity of group-specific bacterial primers may bias the interpretation of environmental microbial studies. As available sequence data continue growing, the performance of primers and probes needs to be reevaluated. Here, we present an evaluation of several commonly used and one newly designed Bacteroidetes-specific primer (CF418). First, we revised the *in silico* primer coverage and specificity with the current SILVA and RDP databases. We found minor differences with previous studies, which could be explained by the chosen databases, taxonomies, and matching criteria. We selected eight commonly used Bacteroidetes primers and tested them with a collection of assorted marine bacterial isolates. We also used the denaturing gradient gel electrophoresis (DGGE) approach in environmental samples to evaluate their ability to yield clear and diverse band patterns corresponding to Bacteroidetes phylotypes. Among the primers tested, CF968 did not provide satisfactory results in DGGE, although it exhibited the highest *in silico* coverage for Flavobacteria. Primers CFB560 and CFB555 presented undesirable features, such as requiring nested protocols or presence of degeneracies. Finally, the new primer CF418 and primer CF319a were used to explore the Bacteroidetes dynamics throughout a 1-year cycle in Mediterranean coastal waters (Blanes Bay Microbial Observatory). Both primers provided clear and diverse banding patterns, but the low specificity of CF319a was evidenced by 83.3 % of the bands sequenced corresponding to nontarget taxa. The satisfactory DGGE banding patterns and the wide diversity of sequences retrieved from DGGE bands with primer CF418 prove it to be a valuable alternative for the study of Bacteroidetes communities, recovering a wide range of phylotypes within the group.

Evaluation of marine Bacteroidetes-specific primers for microbial diversity and dynamics studies. **Cristina Díez-Vives, Josep M. Gasol and Silvia G. Acinas**. 2012. Microbial Ecology 64: 1047-1055

Introduction

The phylum Bacteroidetes or *Cytophaga–Flavobacteria–Bacteroidetes* (CFB) cluster is one of the most abundant heterotrophic bacterial groups in marine environments, representing 10–40 % of total bacteria as estimated by fluorescence in situ hybridization (FISH) counts [Alonso-Sáez *et al.,* 2007; Glöckner *et al.,* 1999; Kirchman, 2002]. This group plays a key role in the degradation of particulate organic matter in the ocean [Kirchman, 2002]. Current classification of phylum Bacteroidetes, based on 16S ribosomal DNA data, considers four different classes: Bacteroidia (strictly anaerobic intestinal bacteria), Flavobacteria, Sphingobacteria, and Cytophagia (the last three mostly aerobic heterotrophic bacteria, which are present in both aquatic and terrestrial habitats) [Ludwig *et al.,* 2010].

Several group-specific probes and primers have been designed over the years for the study of this phylum. The first probe described, CF319a, was presented in 1996 from the analysis of a set of 89 Bacteroidetes 16S rDNA sequences [Manz *et al.,* 1996], and to date, it is the most popular probe used to target Bacteroidetes. Facing the dramatic expansion of gene databases, several studies have designed new probes [Eilers *et al.,* 2001; Gómez-Pereira *et al.,* 2010; Malmstrom *et al.,* 2007; O´Sullivan *et al.,* 2004; Weller *et al.,* 2000] or have modified the available ones [Chen *et al.,* 2006; Kirchman *et al.,* 2003; Suzuki *et al.,* 2001b] to improve the coverage of this phylum and, particularly, to avoid nonspecific binding to other groups. Since the number of sequences available keeps growing exponentially, the choice of the best primers and probes to investigate this bacterial group needs to be periodically revised.

The most widely used applications for Bacteroidetes-specific probes and primers include taxa-specific counts by FISH, generation of rDNA libraries by PCR, and qPCR quantifications [Chen *et al.,* 2006; Kirchman *et al.,* 2003; Suzuki *et al.,* 2001b]. However, there are relatively few studies with fingerprinting techniques, such as the denaturing gradient gel electrophoresis (DGGE) analysis [Dorador *et al.,* 2009; Jaspers *et al.,* 2001; Rink *et al.,* 2007; Santos *et al.,* 2010]. DGGE analyses, using universal bacterial primers, have been applied widely in microbial environmental studies for more than a decade [Alonso-Sáez *et al.,* 2007; Ghiglione *et al.,* 2005; Schauer *et al.,* 2003; Schauer *et al.,* 2000]. However, several studies have emphasized a significant bias of most common universal bacterial primers against specific phyla, particularly Bacteroidetes [Alonso-Sáez *et al.,* 2007; Kirchman, 2002]. Mismatches in the binding region of universal bacterial primers have been pointed as a possible explanation [Glöckner *et al.,* 1999; Suzuki *et*

al., 2001b]. Furthermore, DGGE analyses of samples amplified with universal bacterial primers may mask the less abundant phylotypes [Alonso-Sáez *et al.,* 2007; Castle and Kirchman, 2004; Sánchez *et al.,* 2009] or those generating overlapping bands, and thus, important changes in the bacterial communities may remain undetected [Jaspers *et al.,* 2001; Rink *et al.,* 2007]. A much higher number of Bacteroidetes phylotypes have been resolved with this technique only when using group-specific primers [Abell & Bowman, 2005a; Mühling *et al.,* 2008; Rink *et al.,* 2008].

Although some *in silico* revisions have been made to accommodate the increasing number of probes and primers available for marine Bacteroidetes [e.g., Blümel *et al.,* 2007], few studies have validated them with cultured isolates. In this work, we reevaluate the *in silico* performance of widely used Bacteroidetes-specific primers and one newly designed primer (CF418) using updated sequence databases. We also verify the specificity of selected Bacteroidetes-specific primers with a set of cultured bacterial isolates and test their suitability as primers for DGGE. Finally, we use the CF418 and CF319a primers to explore the Bacteroidetes population dynamics during 1 year in Mediterranean coastal waters (Blanes Bay Microbial Observatory (BBMO)).

Materials and Methods

In Silico Evaluation and Primer Design

Probe design and evaluation was carried out using the ARB software [Ludwig *et al.,* 2004] with the curated SILVA (http://www.arb-silva.de) rDNA database [Pruesse *et al.,* 2007]. The used release was the "non-redundant" SSU Ref NR 104 dataset, which is a modified version of the full reference release in which redundancies (>99 % identity) were filtered out (about 50 % of the entries). The dataset (February 2011) contained 290,603 SSU rDNA sequences, including 24,319 Bacteroidetes. Target and nontarget matches were evaluated, defining a hit as less than 0.5-weighted mismatch in ARB's probe match tool. This value allows some noncanonical base pairing or nondiscriminatory mismatches at the ends of the probe target region. Additional tests were made with the ribosomal database project (RDP) (http://rdp.cme.msu.edu) [Cole *et al.,* 2009], which included (September 2011) 1,921,179 total sequences (82,783 Bacteroidetes). Primer evaluation against the RDP database was restricted to high-quality, >1,200 bp, sequences having data in the annealing region. Since the probe match function available within RDP does not allow evaluation of mismatching primers, comparative results with the SILVA database were

obtained using ARB's probe match with 0-weighted option. Primer design was carried out using the "Probe design" tool in ARB, with all Bacteroidetes sequences of the SSU Ref NR 104 dataset as target hits. Features and performance of candidate primers were evaluated following commonly accepted criteria for primer design and avoiding any degenerate primers [Dieffenbach *et al.,* 1993].

Bacterial Isolates and Environmental DNA Preparation

A set of 35 cultured bacterial isolates (including nine marine Bacteroidetes) was obtained from Northwestern Mediterranean Sea surface waters, collected 1 km off the coast (41° 40'N, 2°48'E) at the Blanes Bay Microbial Observatory (http://www.icm.csic.es/bio/projects/icmicrobis/bbmo/). Details of the sampling area are presented elsewhere [Alonso-Sáez *et al.,* 2007]. Bacteria were isolated on Marine Agar 2216 (Difco) and used to test the specificity of six Bacteroidetes primer sets (details in Table 3SM). Environmental samples were taken monthly at the BBMO from September 2007 to December 2008. Water samples were filtered through 200-µm mesh and kept in 25-l plastic carboys for less than 2 h until processing. Bacterioplankton biomass was retrieved from 8 to 10 l of seawater filtered sequentially through polycarbonate 3 µm (Whatman-7) and Sterivex 0.22 µm pore size. The filters were placed in microtubes with 1.8 ml lysis buffer (40 mM EDTA, 50 mM Tris-HCl, 0.75 M sucrose) or directly flash frozen in N_2. They were stored at −80 °C until processed. Total DNA was extracted from these filters by a standard protocol using phenol/chloroform as described previously [Schauer et al., 2003].

PCR and PCR-DGGE

Different combinations of primers were tested following two strategies: (1) using both Bacteroidetes-specific forward (CF319a, CF315, CFB560, CFB555, and CF418) and reverse (CF968R) primers and (2) using a universal bacteria primer, either the forward 357F [Muyzer *et al.,* 1993] or the reverse 907RM [Muyzer *et al.,* 1998], combined with a Bacteroidetes-specific primer. Optimization of the template DNA amount (1, 4, 8, and 10 ng DNA per reaction) was necessary to improve PCR yields with environmental samples, depending on the levels of inhibitory residues present in the sample. PCR reactions (50-µl volume) contained 200 µM each deoxynucleoside triphosphates, 0.2 µM each primer, 2 mM $MgCl_2$, 1× PCR buffer, and 1 U Taq DNA polymerase (Invitrogen). PCR cycles consisted of an initial denaturation step (94 °C, 5 min); 30 cycles of

denaturation (94 °C, 30 s), annealing (55–63 °C, 30 s), and extension (72 °C, 1 min); and a final extension (72 °C, 10 min). PCR products were analyzed and quantified by agarose gel electrophoresis using standards (Low DNA Mass Ladder, GIBCO BRL).

PCRs for subsequent DGGE analyses used versions of the same primer pairs modified with a 5'-CG-clamp (CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG G) in the forward primer, under identical cycling conditions. A single-round PCR was used in all cases except with primer CG-CFB560, which required a nested PCR to generate enough products for subsequent DGGE analysis. In this case, the product of a first reaction using CFB560 and CF968R (Tm 55 °C) was used as template for a second PCR with CG-CFB560 and 907RM (Tm 55 °C). To reduce potential random biases associated with individual reactions, several PCRs of the same sample were run and pooled prior to their DGGE analysis.

DGGE Bands and Sequencing Analyses

DGGE analyses were performed as previously described [Muyzer *et al.,* 1998]. PCR products were loaded on 6 % polyacrylamide gels cast in a continuous DNA-denaturing gradient. Several gradients between 40 and 80 % denaturing agent (100 % denaturant is 7 M urea plus 40 % deionized formamide) were tested for the different primer sets. The optimum gradient chosen for Bacteroidetes-specific gels was 40-70%. Gels were run at 100V for 16h at 60° C in TAE running buffer (40 mM Tris pH 7.4, 20 mM sodium acetate, 1 mM EDTA) using a DGGE-2000 system (CBS Scientific, Del Mar, CA, USA). The gels were stained with SYBR-Gold (Molecular Probes, Eugene, OR, USA) for 15 min and documented with a Fluor-S MultiImager system (Bio-Rad, Hercules, CA, USA). Gel images were analyzed using the QuantityOne software (Bio-Rad). DGGE bands were excised, placed into 20 µl Milli-Q water, and incubated over-night at 4 °C. The supernatants (2 µl), containing the band DNA, were used as template for reamplification with the original primer set (without GC-clamp), and the products were submitted for automated DNA sequencing (Macrogen Inc., Amsterdam, Netherlands). The sequences from DGGE bands were compared against GenBank, National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) using BLAST [Altschul *et al.,* 1990] to determine their most likely phylogenetic affiliation. Gene sequences were deposited in GenBank under accession numbers JN935205–JN935267.

Results and Discussion

Primers Coverage and Specificity

Forty Bacteroidetes-specific primers and probes were compiled in a literature survey (Table 1SM), and the eight most commonly used were analyzed, checking for the number of positive hits inside and outside of the target group (Table 1).

Primers CFB560 and CF968R provided the highest coverage of the overall phylum Bacteroidetes (93.1 and 92.8 %, respectively). Primer CF315 presented 56.6 % coverage due to poor recovery of class Bacteroidia (34.4 %) although it was much higher (90.9 %), considering only the Bacteroidetes classes common from aquatic habitats (Flavobacteria, Sphingobacteria, and Cytophagia). Primer CF319a matched 66.6 % of sequences within the three environmental clades, with Flavobacteria being the best-covered class (92.7 %). Primer Cyt1020R showed low coverage (14 %) mostly restricted to the class Flavobacteria, and the newly designed primer CF418 covered about 43 % of the phylum. Although this new primer CF418 missed entire genera, such as Zobellia and Arenibacter, and some sequences from Maribacter, Psycroflexus, and Robiginatalea (details not shown), these genera are not common from plankton environments. In addition, CF418 matched a significant percentage of class Sphingobacteria sequences (35.4 %).

Optimum primers should present a maximum coverage, but also need to be specific, presenting a minimum number of out-group hits, particularly among other abundant marine taxa. The specificity of the different primers is summarized and illustrated in Table 1 and Fig. 1. Primers CF319a, CF316, and CF315 were quite nonspecific, with a high ratio of out-group/positive hits (between 1:100 and 1:10, depending on the database used). Multiple hits were observed in Crenarchaeota and Proteobacteria (mainly Epsilonproteobacteria, but also Alphaproteobacteria). Primer CF968R exhibited over 2,000 out-group hits; half of them matched Spirochaeta, a rare microorganism in marine bacterioplankton communities, but numerous hits were also observed within the phylum Proteobacteria (Table 1). Primers CFB560, CFB555, and Cyt1020R and the new primer CF418 presented roughly ten times less out-group hits than CF319a, CF316, CF315, and CF968R (Table 1; Fig. 1). Although the RDP database included more Bacteroidetes sequences than the Silva 104 NR and this resulted in a higher number of positive hits, the relative differences in specificity between all primers tested were similar with both databases.

Table 1 *In silico* evaluation of selected Bacteroidetes-specific primers. Number of target hits and coverage (%C) for the overall Phylum Bacteroidetes and for each separate Bacteroidetes class, using the Silva NR104 database. Hits were predicted by ARB's Probe Match Tool using 0.5 weighted mismatch.

Primer	Phylum Bacteroidetes 24319 seq. Hits (%C)	class Bacteroidia 14776 seq. Hits (%C)	class Flavobacteria 4384 seq. Hits (%C)	class Sphingobacteria 3648 seq. Hits (%C)	class Cytophagia 1511 seq. Hits (%C)	Outgroup hits	Main out-groups (Number of positive hits)
CF319a	11361 (46.72)	5018 (33.96)	4063 (92.68)	2165 (59.35)	124 (8.21)	1902	Crenarchaeota (1144) Proteobacteria (470)
CF316	11307 (46.49)	4985 (33.74)	4054 (92.47)	2145 (58.80)	123 (8.14)	1725	Crenarchaeota (1138) Proteobacteria (443)
CF315	13767 (56.61)	5090 (34.45)	4152 (94.71)	3226 (88.43)	1299 (85.97)	2489	Crenarchaeota (1145) Proteobacteria (513) Nitrospirae (158) Chloroflexi (221)
CFB560	22632 (93.06)	13877 (93.92)	4179 (95.32)	3340 (91.56)	1236 (81.80)	284	Firmicutes (143) Proteobacteria (42)
CFB555	20492 (84.26)	13597 (92.02)	4110 (93.75)	1636 (44.85)	1149 (76.04)	160	Firmicutes (118) Proteobacteria (27)
CF968	22578 (92.84)	13810 (93.46)	4217 (96.19)	3278 (89.86)	1273 (84.25)	2863	Spirochaetes (1301) Fusobacteria (346) Acidobacterias (217) Firmicutes (371) Proteobacteria (196)
Cyt1020	3414 (14.04)	0 (0)	3380 (77.10)	30 (0.82)	4 (0.26)	131	Proteobacteria (117)
CF418	10423 (42.86)	6740 (45.61)	2359 (53.81)	1291 (35.39)	33 (2.18)	210	Firmicutes (69) *Cyanobacteria* (33)

Table 1SM Summary of primers and probes described for marine Bacteroidetes (by date of publication)

Name	Specificity (reported)	Applications	Sequence	Formamide % or combing primer	E. coli position	Ref.
CF319a	Cytophaga-Flavobacterium group and genus *Porphyromonas*	DBT, PCR	TGGTCCGTGTCTCAGTAC	35	319-336	27
CF319b	Cytophaga-Flavobacterium group and genus *Porphyromonas*	DBT, PCR	TGGTCCGTATCTCAGTAC	35	319-336	27
BAC303	Genera *Prevotella* and *Bacteroides*	DBT, PCR	CCAATGTGGGGGACCTT	0	303-319	27
FFE8b	*Flavobacterium ferrugineum*	DBT, PCR	CAGCCGCACACCCGTCTT	20	225-242	27
FC	Flavobacterium-Cytophaga	FISH	AGGTACCCCCAGCTTCCATGGCT	30	1408-1430	20
CFB286	Members of the CFB-phylum except F. *Ferrugineum*	FISH	TCCTCTCAGAACCCCTAC	50	286-304	46
CFB563	Members of the CFB-phylum except F. *Ferrugineum*	FISH	GGACCCTTTAAACCCAAT	20	563-580	46
CFB719	Members of the CFB-phylum except F. *ferrugineum*	FISH	AGCTGCCTTCGCAATCGG	30	719-736	46
CFB972	Members of the CFB-phylum except F. *Ferrugineum*	FISH	CCTTGGTAAGGTTCCTCG	20	972-989	46
CFB1082	Members of the CFB-phylum except F. *Ferrugineum*	FISH	TGGCACTTAAGCCGACAC	N.A.	1082-1100	46
CYT1448	*Cytophaga spp.*	FISH	CTAGGCCGCTCCTTACGG	30	1448–1465	14
CYT1438	*Cytophaga spp.*	FISH	CCGCTCCTTACGGTGACG	30	1438-1455	14
CYT191F	Cytophaga group	qPCR	GGGTCCTGAGAGGGRGAT	CYT536R	290-308	44
TM311F	Flexibacter, Bacteroidetes and Cytophaga phylum	qPCR	CCACACTGGTACTGAGACACGGAC	Fluorogenic probe	311-335	44
CFB560	CFB division	SDT	WCCCTTTAAACCCART	30	560–575	33
CFB562	CFBs excluding Cytophaga group	SDT	TACGYWCCCTTTAAACCCA	30	562-580	33
CFB376	CFBs excluding Cytophaga and Bacteroidetes groups	SDT	TGMCCAATATTCCYTACTG	30	376-394	33
CF316	Memebers of the Cytophaga-Flavobacterium group	PCR	CTGGTACTGAGACACGGA	EubA-R	316-333	22
CF6-1267	CF subgroup (DE cluster 2)	FISH	GAAGATTCGCTCCTCCTC	20	1268-1287	22
AGG58-A	AGG58 cluster,but also some other Bacteroidetes	SDT	CTACATGWCAYATTCCGCC	40	670-689	32
AGG58-A-F	AGG58 cluster,but also some other Bacteroidetes	PCR	GGCGGAATRTGWCATGTAG	AGG58-B/C-R	670-689	32
AGG58-B	AGG58 cluster branches 1 and 3	SDT	CGTTMTGCATTTAAGMCC	30	987-1004	32
GG58-B-R	AGG58 cluster branches 1 and 3	PCR	CGTTMTGCATTTAAGMCC	AGG58-A-F	987-1004	32
AGG58-C	AGG58 cluster branch 2	SDT	CACAGGCAGTRTTTCTAGAG	35	N.A.	32

AGG58-C-R	AGG58 cluster branch 2	PCR	CACAGGCAGTRTTTCTAGAG	AGG58-A-F	N.A	32
CF315	Cytophaga-Flavobacteria cluster	PCR	ACKGGYACTGAGAYACGG	CF968R	315-332	9
CF968*	Cytophaga-Flavobacteria cluster/Bacteroidetes	PCR-DGGE	GGTAAGGTTCCTCGCGTA	CF351F	968–985	9, 29
CFB555f	Bacteroidetes	PCR-DGGE	CCGGAWTYATTGGGTTTAAAGGG	CFB968r/907r	555–577	29
Cyt1020R	Bacteroidetes	PCR-DGGE	CATTTAAGCCTTGGTAAGG	27F	978–995	6
Polaribacter-740	*Polaribacter* cluster	FISH	CCCTCAGCGTCAGTACATACGT	30	741-763	26
Polaribacter-218	*Polaribacter* cluster	FISH	GACGCATAGCCATCTTTT ACCG	30	219-241	26
FL197	VIS4 clade	FISH	TCATCTCATACCGTAACC	35	198-232	18
VIS6-814	VIS6 clade	FISH	CAGCGAGTGATGATCGTT	15	815-833	18
VIS5-586	VIS5 clade	FISH	CCACTGACTTAGATTCCC	25	587-605	18
FL845	VIS2 clade	FISH	GCTTAGCCACTCATCTAA	15	846-863	18
VIS3-183	VIS3 clade	FISH	CTCTAGAAGTGATGCCAC	25	184-194	18
VIS1-575	VIS1 clade	FISH	CTTAACAAACAGCCTGCGGACC	35	576-598	18
LEE82	*Leeuwenhoekiella*	FISH	ACTCGTCAGCTTCCCGAA	25	83-104	18
DOK827	*Dokdonia*	FISH	GTCCGAAAACCAAACAACTAG	20	826-850	18
CF418	Bacteroidetes	PCR-DGGE	CCCTATGGGTTGTAAACT	907RM	418- 435	This study

Primers selected in this study for further in silico analyses and tests with isolates are underlined. Abbreviations: DBT= Probe for dot-blot hybridization; SBT= Probe for slot-blot hybridization; FISH= Fluorescence in situ hybridization; DGGE= Denaturing Gradient Gel Electrophoresis; N.A.= Not available.
*CF968 was described in two different papers with identical sequence but using different names (CF967 and CF968).

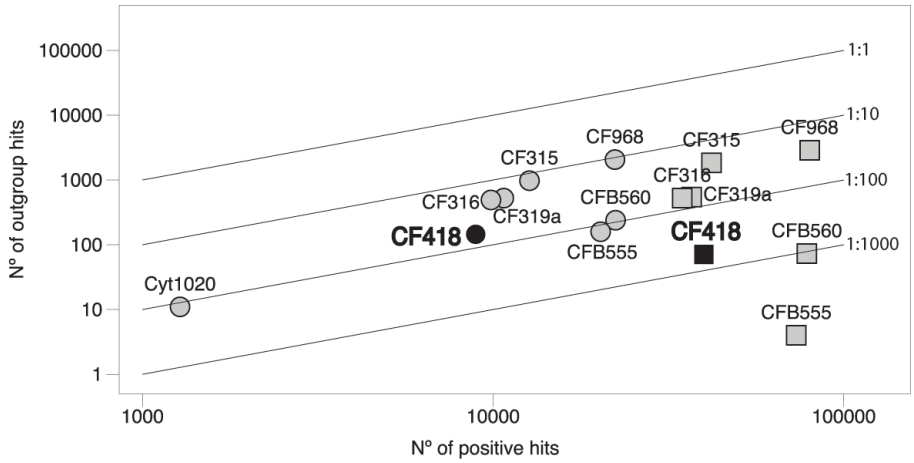


Figure 1 Relationship between target and out-group hits for the 8 selected Bacteroidetes-specific primers (logarithmic scale). The new primer (CF418) is in bold type. Results using both databases are represented as follows: Circles= Silva 104 NR database; Squares= RDP database.

Effect of Databases Growth and In Silico Testing Parameters

We compared our results with those of a recent review of group-specific probes by Amann and Fuchs [Amann & Fuchs, 2008], after which the number of sequences available has grown by roughly 50 %. Overall, differences in primer coverage were minor for primers CF319a and CFB560. However, a relevant difference was noticed in the coverage of class Sphingobacteria (currently split as class Sphingobacteria and class Cytophagia) by primers CF319a (90 % in Amann and Fuchs [Amann & Fuchs, 2008] *vs.* 44.2 % in our study) and CFB560 (100 *vs.* 88.7 %). When the probe match testing was reproduced with the same dataset release used by these authors (Silva release 91), the high coverage values reported among class Sphingobacteria (90 and 100 %) were found to concentrate on taxa classified as Sphingobacteria in that release. This, however, excluded many uncultured and ambiguously classified taxa, which have, afterwards, been grouped within Sphingobacteria in later database releases. When sequences of release 91 were sorted into the three currently considered Bacteroidetes classes, Sphingobacteria coverage was 43 %, rather similar to the value obtained in our work (44.2 %) for primer CF319a.

Other previous studies have used mostly online tools, such as the RDP database. Hence, we also compared the coverage values in these reports with our results using the current RDP release. The values reported by Chen *et al.* [Chen *et al.,* 2006] for coverage of class Sphingobacteria are similar to those obtained with the current release

for primers CF319a (25 *vs.* 20 %), CF315 (85 *vs.* 88.2 %), and CF968R (89.6 *vs.* 88.9 %). Similarly, the coverage values of phylum Bacteroidetes in other reports [Matthew *et al.,* 2011; Mühling *et al.,* 2008] differed slightly from ours for primers CFB555 (83.9 and 87.4 %, respectively, *vs.* 88.4 %) or CF968R (90.5 and 94.6 %, respectively, *vs.* 96.5 %). These discrepancies, using different RDP database releases, were smaller than those obtained using two different updated databases (Table 2SM). For instance, results from the current RDP database and Silva 104 NR (using 0-weighted option to make results comparable) showed discrepancies in coverage of phylum Bacteroidetes by primer CF418 (48.2 *vs.* 36.7 %), with differences within specific classes being even larger (66.7 *vs.* 48.1 % for class Flavobacteria). Our analyses confirmed that differences in primer coverage values appeared more dependent on the choice of database and matching criteria than on the number of sequences available in a given dataset. In general, for a given database and using the same testing parameters, coverage percentages remain rather stable over time despite the exponential increase of available sequences.

Table 2SM In silico evaluation of selected Bacteroidetes-specific primers. Comparison of coverage and specificity within phylum Bacteroidetes was analyzed using the Silva and RDP databases. Primers tested with bacterial isolates in this study are underlined. The Silva database hits were predicted by ARB's Probe Match Tool using 0 weighted mismatch.

Primer	Positive hits in Phylum Bacteroidetes		Coverage (%)		Out-group hits	
	Silva 24319 seq.	RDP 82768 seq.	Silva	RDP	Silva	RDP
CF319a	10726	36815	44.11	44.48	525	545
CF316	9859	34606	40.54	41.81	492	525
CF315	12696	41965	52.20	50.70	976	1844
CFB560	2235	78482	91.90	94.80	239	73
CFB555	20258	73212	83.30	88.44	159	4
CF968	22252	79914	91.50	96.54	2067	2877
Cyt1020	1280	3248	5.26	3.92	11	0
CF418	8920	39914	36.68	48.21	145	71

Specificity Tests Using Cultured Bacteria

We tested six primer sets with DNA from diverse marine bacteria isolates (Table 3SM) to further evaluate their feasibility for microbial diversity studies. In the specific amplification of marine Bacteroidetes and non-Bacteroidetes isolates, primers CFB560, CFB555, CF968R, and CF315 were 100 % effective with all Bacteroidetes tested, while CF319a and CF418 (both with lower coverage of class Cytophagia) did not amplify *Microscilla* sp. (Table 3SM), which presents three and four mismatches, respectively, in the primer binding region (Fig. 1SM). This is in agreement with O'Sullivan *et al.* [2002]. Regarding amplification of nontarget isolates, CF319a (with a maximum of two mismatches against non-Bacteroidetes sequences) and CF315 (with a single mismatch) amplified many non-Bacteroidetes isolates (Table 3SM). Primer CF968R did not amplify any out-group isolates. This primer contains two mismatches, one of them at the 3' end (Fig. 1SM), which partly explains its higher specificity. Primers CF418, CFB560, and CFB555 contain at least four mismatches with nontarget sequences, and CF418 was particularly designed in a variable region, presenting deletions in other bacterial sequences (Fig. 1SM). As expected, this primer did not amplify any out-group. The results of our tests with bacterial isolates suggest that primers were effective, amplifying every sequence with two or less mismatches in the binding region, if they are not located at the 3'-end. Some nonspecific amplifications (in nontarget groups) might be avoided using more stringent PCR conditions, but these would also affect the performance of the primers in environmental samples, where target sequences may be in low abundance in a complex pool of bacterial DNA.

Priming Efficiency and Amplification Biases

Detailed analysis of the primer-binding regions on target and nontarget sequences allowed predicting their performance to some extent, but priming efficiency depends on many factors [von Wintzingerode *et al.,* 1997], and the effect of mismatching nucleotides on the performance of primers cannot be predicted accurately. However, only a few studies have explored the actual performance of newly designed primers using bacterial isolates, and some examples of discrepant performance can be found in the literature [Bru *et al.,* 2008; García-Martínez *et al.,* 2002, Kenzaka *et al.,* 1998]. Within the phylum Bacteroidetes, amplification of out-group isolates, presenting few mismatches, was previously reported [Ludwig *et al.,* 2010] using primer CF319a. Caution must be

Table 3SM Specificity tests of Bacteroidetes-specific primer sets against 35 bacterial isolates.

Bacterial Isolate (ACC #)	Phylogeny	Primer Pair Set					
		CF319aF 907RM	CF315F 907RM	CF555F 907RM	CF560F 907RM	CF418F 907RM	358F CF968R
Bacteroidetes							
Dokdonia sp. MED134 (DQ481462)	Flavobacteriaceae	+	+	+	+	+	+
Polaribacter sp. MED152 (DQ481463)	Flavobacteriaceae	+	+	+	+	+	+
Leeuwenhoekiella blandensis MED217 (DQ294291)	Flavobacteriaceae	+	+	+	+	+	+
Salegentibacter sp. MED220 (DQ681156)	Flavobacteriaceae	+	+	+	+	+	+
Gramella sp. MED367 (EU253577)	Flavobacteriaceae	+	+	+	+	+	+
Maribacter sp. MED381 (EU253579)	Flavobacteriaceae	+	+	+	+	+	+
Leeuwenhoekiella sp. MED495 (DQ681176)	Flavobacteriaceae	+	+	+	+	+	+
Salegentibacter sp. MED532 (EU253596)	Flavobacteriaceae	+	+	+	+	+	+
Microscilla sp. MED571 (EU253600)	Cytophagaceae	-	+	+	+	-	+
α-Proteobacteria							
Phaeobacter sp. MED193 (DQ681150)	Rhodobacteraceae	+	-	-	-	-	-
Erythrobacter sp. MED442 (DQ681170)	Erythrobacteraceae	+	+	-	-	-	-
Methylarcula sp. MED464 (EU253587)	Rhodobacteraceae	+	+	-	-	-	-
Nereida sp. MED128 (DQ681139)	Rhodobacteraceae	+	+	-	-	-	-
Erythrobacter sp. MED458 (DQ681173)	Erythrobacteraceae	+	+	-	-	-	-
Nereida sp. MED132 (DQ681140)	Rhodobacteraceae	+	+	-	-	-	-
Roseovarius sp. MED149 (DQ681144)	Rhodobacteraceae	+	+	-	-	-	-
Phaeobacter sp. MED165 (DQ681146)	Rhodobacteraceae	+	+	-	-	-	-
Paracoccus sp. MED441 (EU253582)	Rhodobacteraceae	+	+	-	-	-	-
Palleronia sp. MED447 (EU253585)	Rhodobacteraceae	+	+	-	-	-	-
Thalassobacter sp. MED588 (DQ681183)	Rhodobacteraceae	+	+	-	-	-	-
β-Proteobacteria							
Limnobacter sp. MED513 (EU253593)	Burkholderiaceae	+	+	-	-	-	-
γ-Proteobacteria							
Alteromonas sp. MED169 (DQ681147)	Alteromonadaceae	+	+	-	-	-	-
Marinobacter sp. MED203 (DQ681151)	Alteromonadaceae	-	+	-	-	-	-
Enterovibrio sp. MED126 (DQ681138)	Vibrionaceae	+	+	-	-	-	-
Saccharophagus sp. MED266 (DQ681163)	Osceanospirillales	+	+	-	-	-	-
Halomonas sp. MED543 (DQ681182)	Halomonadaceae	+	+	-	-	-	-
Pseudoalteromonas sp. MED107 (DQ681131)	Pseudoalteromonadaceae	+	+	-	-	-	-
Idiomarina sp. MED206 (DQ681152)	Idiomarinaceae	+	+	-	-	-	-
Pseudoalteromonas sp. MED246 (DQ681161)	Pseudoalteromonadaceae	+	+	-	-	-	-
Marinobacter sp. MED503 (DQ681178)	Alteromonadaceae	-	+	-	-	-	-
Vibrio sp. MED535 (DQ681180)	Vibrionaceae	+	+	-	-	-	-
Marinomonas sp. MED269 (DQ681164)	Halomonadaceae	+	+	-	-	-	-
Actinobacteria							
Kocuria sp. MED219 (DQ681155)	Micrococcaceae	+	+	-	-	-	-
Firmicutes							
Planococcus sp. MED216 (DQ681154)	Planococcaceae	+	+	-	-	-	-
Thalassobacillus sp. MED146 (DQ681143)	Bacillaceae	-	-	-	-	-	-

Positive (+) or negative (-) PCR amplification of DNA from each isolate. Reactions were carried out at Tm=55 °C.

exercised when primers presenting few mismatches with nontarget sequences are used, and tests with selected isolates should complement *in silico* studies. From our combined analysis, the use of primers, such as CFB560, CFB555, and CF418, presenting at least four mismatches with nontarget groups (Fig. 1SM), is more reliable and allows to run less stringent reactions amplifying relatively scarce target sequences from environmental samples while maintaining good specificity (Table 1).

In addition, the primers ability to produce unbiased amplification of all members of the target group should also be considered. Degenerate primers may cause PCR selection against some templates with different G/C content in the priming regions [Polz & Cavanaugh, 1998; von Wintzingerode *et al.,* 1997]. The reverse primer 907RM used includes one mixed A/C base, but Bacteroidetes sequences do not present variability in this matching region, and unbiased binding of this primer should be expected. However, primers CFB555, CFB560, and CF315 contain degeneracies in which Bacteroidetes present variability, and their use for quantitative or semiquantitative estimates of Bacteroidetes populations from DGGE bands might introduce unexpected biases in the analyses.

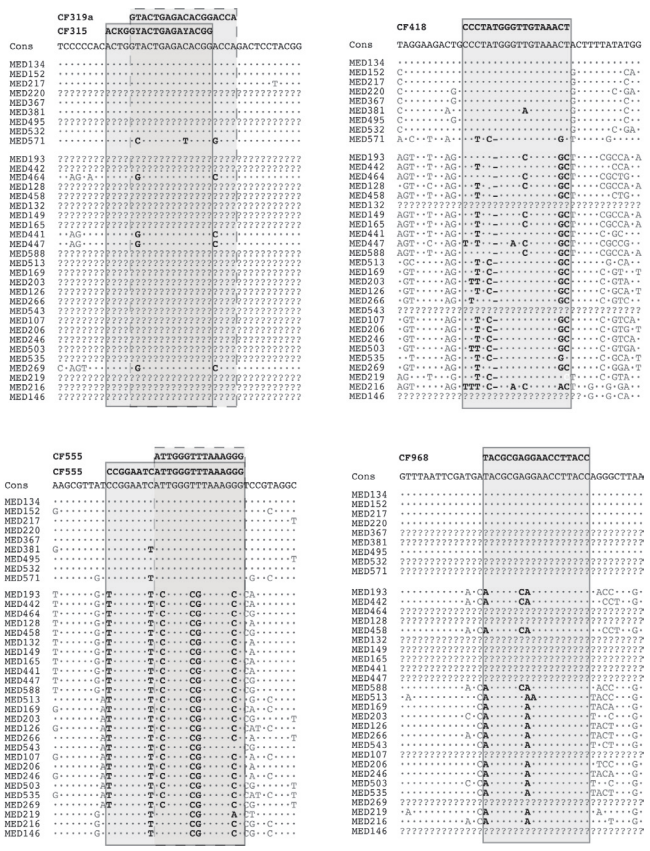


Figure 1SM 16S rDNA alignment of 35 isolates showing the consensus sequence (Cons.) and the binding regions of Bacteroidetes-specific primers (shaded boxes). For primer CF968R, its reverse and complementary sequence is shown to facilitate comparison. Alignment conventions: "?" Missing data; "." Conserved site"; "Capital letters" Variable positions representing primer mismatches.

Choosing the Best Primer Set for DGGE

The suitability of Bacteroidetes-specific primers for DGGE analysis was evaluated taking into account features such as efficient amplification using single-round PCR, diversity and sharpness of the resolved bands, successful reamplication of excised bands, length of the obtained sequences, and diversity of Bacteroidetes phylotypes retrieved (Table 2). Since primer CF968R presented the highest *in silico* coverage for class Flavobacteria and showed specific amplification of marine Bacteroidetes isolates, it was considered a good candidate for DGGE analyses. Thus, we tested its performance in DGGE when combined with a group-specific primer (GC-CF418) on marine Bacteroidetes isolates (Fig. 2SM). Additional combinations of these group-specific (GC-CF418 and CF968R) and universal bacterial (GC-357F and 907RM) primers were equally assayed (Fig. 2SM). While sharp bands were obtained using the universal reverse primer 907RM (regardless of the forward primer used), fuzzy bands were observed when Bacteroidetes-specific reverse CF968R primer was included (Fig. 2SM). In order to verify the presence of fuzzy bands using GC-357F and CF968R with naturally occurring bacteria, we conducted DGGE analyses of environmental samples. Although the gel showed some sharp bands, the general banding pattern profile was unsatisfactory, preventing the use of primer CF968R for further comparative analyses (Fig. 3SM). Despite this poor resolution in DGGE gels, primer CF968R allowed to identify 17 bands, all of them matching uncultured Bacteroidetes sequences. Nevertheless, upon excluding uncultured/environmental sequences from the database, 14 of these bands matched the same sequence, *Owenweeksia hongkongensis*, indicating poor diversity of retrieved phylotypes (Table 4SM).

The available alternative Bacteroidetes-specific reverse primer for DGGE, Cyt1020R [Blümel *et al.,* 2007], was disregarded due to its limited *in silico* coverage (Table 1). Thus, Bacteroidetes-specific forward primers (CF319, CFB555, CFB560, and CF418) were subsequently combined with the universal bacterial reverse primer 907RM, which has been recommended for all-Bacteria PCR-DGGE analyses with Mediterranean samples [Sánchez *et al.,* 2009].

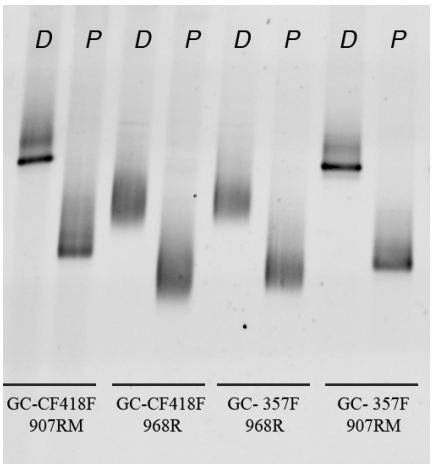


Figure 2SM DGGE analysis of Bacteroidetes isolates Dokdonia sp. MED134 (DQ481462) lanes D and 7; and *Polaribacter* sp. MED152 (DQ481463) lanes P, using different primer combinations where either the forward or reverse was a Bacteroidetes-specific primer (noted at the bottom).

Figure 3SM DGGE profile analysis of environmental samples using GC-357F and CF968R primers. Note the poor DGGE profile obtained with CF968R despite presenting the highest in silico coverage. Numbered dots indicate bands that were excised and sequenced. Lanes are labeled after the sampling months, from September 2007 to September 2008.

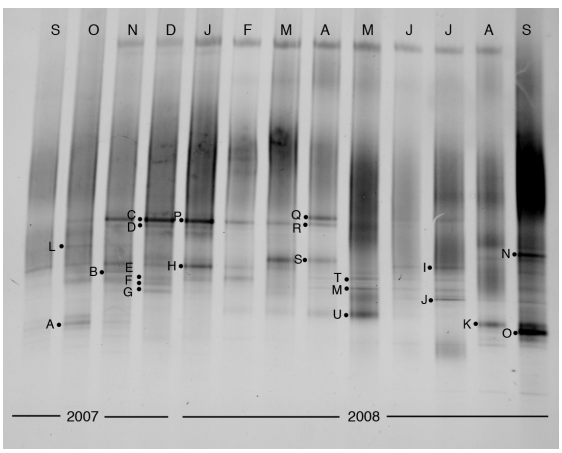


Table 2 Summary of main features and analyses carried out on selected Bacteroidetes-specific primers. Undesired characteristics are remarked in bold type.

Primer	CF319a	CF315	CFB560	CFB555	CF968	CF418	Detailed in
% Coverage Flavobacteria	92.7%	94.7%	95.3%	93.8%	96.2%	**53.8%**	Table 1
Out-group hits	**1902**	**2489**	284	160	**2863**	210	Table 1
Specificity on isolates	**Poor**	**Poor**	Good	Good	Good	Good	Table 3SM
Degenerancies	0	**3**	**2**	**2**	0	0	Table 1SM
Nested PCR	No	-	**Yes**	No	No	No	Text
DGGE resolution	Good	-	Good	Good	**Poor**	Good	Fig. 2, 2SM, 3SM, 4SM
Maximun no. unique bands	34	-	22	17	17	17	Text
Reamplification of bands	12	-	**not achieved**	4	17	5	Text
% Unique Blast hits	100%	-	-	**50%**	71%	100%	Table 4SM
Specificity (sequenced bands)	**Poor**	-	-	Good	Good	Good	

Performance of Primer Sets in DGGE

Comparison of the performance of four Bacteroidetes-specific primers combined with this universal bacterial reverse primer (907RM) was made using three environmental samples (Fig. 4SM). All combinations showed satisfactory banding patterns in terms of well-defined bands and band diversity among samples. Primer CF319a yielded the most diverse banding patterns resolving up to 34 unique bands, compared to 17 bands with CFB555 and CF418 and 22 with CFB560.

Use of the primer CF319a for PCR-DGGE analysis has been previously reported [Jaspers *et al.,* 2001; Rink *et al.,* 2007] using single-round PCR. Under similar PCR-DGGE conditions, we obtained comparable results with good amplifications and satisfactory patterns (Fig. 4SM). Different performance was noted for primer CFB555 compared with previous reports. This primer was originally described to require a seminested protocol because a single-round PCR did not provide reproducible results. In addition, the band profile with the seminested protocol was considered unsatisfactory by the authors [Mühling *et al.,* 2008]. Here, we used this primer (CFB555) in a single-round PCR approach and found good product yield and clear banding profiles, which made the seminested protocol unnecessary (Fig. 4SM). Four bands were identified and excised from samples with primer CFB555, and all sequences (317 bp maximum) were identified as Bacteroidetes, three of them matching the same primary Blast hit (*Owenweeksia hongkongensis*) (Table 4SM). This is in agreement with a previous study in which three phylotypes could be discriminated from five sequenced bands [Mühling *et al.,* 2008]. The use of CFB560 as DGGE primer had not been reported previously. This primer performed well in PCR, but very low yields were obtained once the GC-clamp was included for PCR-DGGE. This was the only primer that required a nested protocol in our study. Despite this, the banding profile was good when amplification could be achieved by this nested protocol. However, no sequences could be obtained from ten bands excised from a DGGE with CFB560 because reamplification attempts failed in all cases. The new CF418 primer yielded good amplification of samples with a single-round PCR.

Although nested PCR is routinely used in many studies to generate enough product for downstream fingerprinting studies or to improve detection sensitivity [Dar *et al.,* 2005; Mühling *et al.,* 2008], it can easily obscure differences between samples when more than 20 cycles are used in the first PCR round, while those differences are preserved in a single-round PCR-DGGE up to 35 cycles [Park & Crowley, 2010]. In addition, the potential for chimera formation and amplification errors increases in a nested protocol,

which often generates secondary, nonspecific bands. Finally, because it is recommended to pool several PCR products on each DGGE lane to account for unpredictable variation between replicates [Polz & Cavanaugh, 1998], nested PCRs considerably increases the cost and labor of the study. Since primer CFB560 needed a nested protocol, its use cannot be recommended in semiquantitative fingerprinting studies. The Bacteroidetes-specific primer combination GC-CF418-907RM performed well in a single-round PCR, thus simplifying the PCR-DGGE workflow as much as possible.

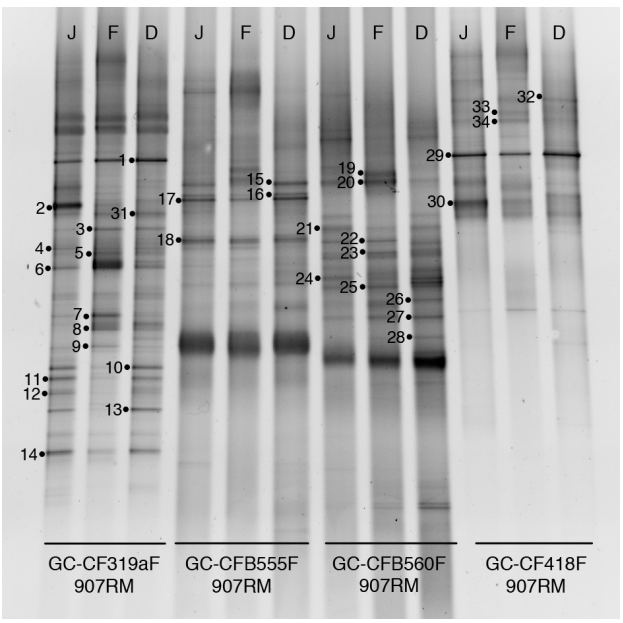


Figure 4SM DGGE analysis of 16S rRNA amplicons with four different marine Bacteroidetes primer sets. The samples correspond to the small fraction (<3 μm pore size filters) of the Blanes Bay Microbial Observatory taken in January (lanes J), February (lanes F) and December (lanes D) 2008. Numbered dots mark bands excised and sequenced (presented in Table 1SM).

Application to 1-Year Time Series

The performance of primers CF319a and CF418 in DGGE analyses was further tested using environmental samples from a 1-year temporal series study in the Mediterranean. With both primer sets, we obtained well-resolved banding patterns, which were variable over the year, both in the number of bands and in their relative intensity (Fig. 2). The number of bands resolved with GC-CF418-907RM ranged from 11 to 22 per lane, whereas 20–35 bands were observed with GC-CF319a-907RM. With both primer sets, summer and winter banding patterns were rather different, mainly defined by a few recurring bands

which were present from November to March and from May to September (Fig. 2). April and October represented transitional periods in which typical winter bands, but also bands characteristic of summer samples, were observed. This pattern of summer–winter shifts observed with Bacteroidetes-specific primer sets paralleled patterns previously described in the overall bacterial communities [e.g., Sánchez *et al.,* 2009].

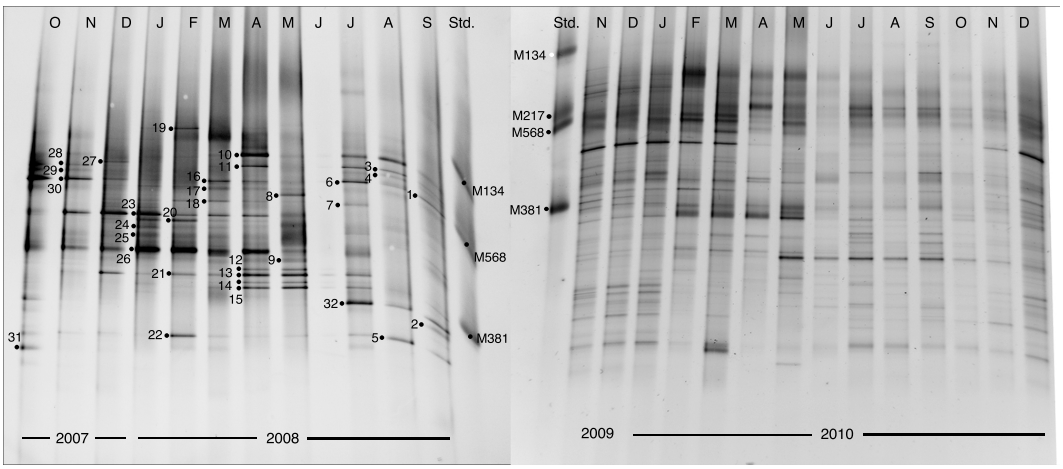


Figure 2 DGGE analysis of 16S rRNA amplicons using Bacteroidetesspecific primers. Left monthly samples of the large fraction (>3-µm pore size filters) using primers GC-CF418-907RM. Right monthly samples of the small fraction (<3-µm pore size filters) using primers GC-CF319a-907RM. Numbered dotsmark bands excised and sequenced (Table 1SM). Std standard containing DNA from three (left gel) or four (right gel) marine Bacteroidetes isolates: Dokdonia sp. MED134 (DQ481462), Leeuwenhoekiella blandensis MED217 (DQ294291), ***Polaribacter*** sp. MED568 (EU253599), and Maribacter sp. MED381 (EU253579).

Notwithstanding the good DGGE visual banding profile and resolution obtained with primer CF319a (Fig. 4SM), 10 out of 12 bands sequences were identified as non-Bacteroidetes, indicating a strong nonspecific binding (Table 4SM). Previous studies with CF319a reported a high specificity for Bacteroidetes, with 100 % of DGGE bands sequences belonging to this group [Rink *et al.,* 2008; Rink *et al.,* 2007]. Nevertheless, under similar conditions, the majority of our sequences were related to class Alphaproteobacteria, in agreement with the numerous hits in this group predicted by the *in silico* tests (Table 1). Moreover, unspecific CARD-FISH hybridization of probe CF319a has been found with members of *Rhodobacteraceae* and *Alteromonadaceae* [Acinas *et al.,* in preparation, chapter 1]. These data discouraged the use of primer CF319a for Bacteroidetes-specific dynamics analyses despite its apparent good DGGE resolution.

Using the new DGGE primer CF418, sequences ranging from 358 to 457 bp were obtained in the comparative gel among primer sets (Fig. 4SM). These were all (5/5) identified as Bacteroidetes, matching different unique closest hits (Table 4SM). In addition, 32 bands were retrieved from the 1-year temporal series gel (Fig. 2), and informative sequences were obtained from 25 of them. All sequences were affiliated to the Bacteroidetes group (Table 4SM) matching 16 different Bacteroidetes taxa. This further confirmed the specificity of this new primer and its power to retrieve a high diversity of Bacteroidetes sequences from environmental samples.

Conclusion

Group-specific primers and probes are essential tools for diversity and dynamics studies, and it is important to have a realistic understanding of their theoretical and experimental performance. Thus, due to the exponential growth of sequence databases and consequent updates of microbial taxonomy, it is necessary to reevaluate them periodically. In this study, minor changes in primer coverage were found with previous studies, and most differences could be explained by the use of different databases, taxonomies, or matching criteria. *In silico* evaluation of primers may provide a general idea on the weaknesses of some primers in terms of coverage, specificity, and unbiased priming efficiency, but a parallel evaluation of their specificity and suitability should be performed for any given approach. We designed a new Bacteroidetes-specific primer (CF418) that presents a high specificity for Bacteroidetes without containing any degeneracy, and whose amplification efficiency allows single-round PCR-DGGE analysis to be carried out. Although this primer presents lower *in silico* coverage than others, none of the six tested primers excelled under all criteria for a DGGE approach (Table 2), and we consider the primer CF418 as a good alternative for seasonal studies and exploration of Bacteroidetes populations by DGGE.

Acknowledgments

We thank Carlos Pedrós-Alió for critical reading of the manuscript. CDV was a recipient of a JAE-Predoc fellowship from CSIC, and SGA was supported by a RyC contract from the Spanish Ministry of Science and Innovation and CONES 2010-0036 from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR). This research was supported by grants: MICRODIVERSITY (CGL2008-00762/BOS) to SGA and STORM (CTM2009-09352/MAR) from the Spanish Ministry of Science and Innovation (MICINN).

Table 4SM Sequences obtained from DGGE bands retrieved using different primer sets.

Phylotype	Acc	Primer	L (nt)	Closest relative in Genbank (environmental or culture) (Acc)	S (%)	Taxonomy (Order > Family)
BBMO_Dec08-SF-1	JN935205	CF319a	529	Uncultured *Flavobacteria bacterium* clone Vis_St6_66 (FN433385)	99	
				Owenweeksia hongkongensis (AB125062)	89	Flavo > Cryomorphaceae
BBMO_Jan08-SF-2	JN935206	CF319a	540	Uncultured phototrophic eukaryote, clone: TKTMrw-B35 (AB611591)	99	
				Micromonas pusilla plastid, strain M1681 (FN563097)	99	Chloro > Mamiellophyceae
BBMO_Dec08-SF-31	JN935207	CF319a	437	Uncultured *Bacteroidetes bacterium* partial clone 3 (AM748178)	99	
				Lishizhenia caseinilytica (AB176674)	92	Flavo > Cryomorphaceae
BBMO_Feb08-SF-5	JN935208	CF319a	528	Uncultured Rhodobacteraceae bacterium clone M7LBP3_10H03 (HQ242603)	99	
				Roseovarius sp. HDW-9 (GQ243422)	96	Alpha >Rhodobacteraceae
BBMO_Jan08-SF-6	JN935209	CF319a	426	Uncultured alpha proteobacterium clone D13W_66 (HM057760)	97	
				Roseovarius sp. L-13 (HQ677209)	96	Alpha > Rhodobacteraceae
BBMO_Feb08-SF-7	JN935210	CF319a	429	Uncultured Rhodobacteraceae bacterium (EF018060)	99	
				Ruegeria sp. 3X/A02/236 (AY576770)	97	Alpha > Rhodobacteraceae
BBMO_Feb08-SF-9	JN935211	CF319a	413	Uncultured bacterium isolate DGGE gel band B16 (EF221654)	99	
				Phaeobacter caeruleus strain UDC410 (HM031996)	97	Alpha > Rhodobacteraceae
BBMO_Dec08-SF-10	JN935212	CF319a	470	Uncultured *Oceanospirillales bacterium* clone HF130_32E15 (EU361654)	99	
				Microbulbifer sp. GB02-2C (GQ118708)	92	Gamma > Alteromonadaceae
BBMO_Jan08-SF-11	JN935213	CF319a	511	Uncultured marine bacterium clone MOLA_MAY07-5m-183 (GU204822)	99	
				Sneathiella glossodoripedis (AB289439)	89	Alpha > Sneathiellaceae
BBMO_Jan08-SF-12	JN935214	CF319a	437	Uncultured alpha proteobacterium clone SHZW504 (HQ163197)	99	
				Brucella sp. HJ114 (DQ167235)	92	Alpha > Brucellaceae
BBMO_Dec08-SF-13	JN935215	CF319a	558	Uncultured cyanobacterium clone DUNE-t8D2-95 (HM117462)	99	
				Prochlorococcus marinus MED4 complete genome (BX548174)	99	Prochlo > Prochlorococcaceae
BBMO_Jan08-SF-14	JN935216	CF319a	470	Uncultured alpha proteobacterium clone HF130_15B09 (EU361386)	100	
				Nisaea sp. DG1596 (EU052763)	94	Alpha > Rhodospirillaceae
BBMO_Dec08-SF-15	JN935217	CFB555	317	Uncultured *Flavobacterium* sp. clone ARTE4_230 (GU230411)	99	
				Owenweeksia hongkongensis (AB125062)	92	Flavo > Cryomorphaceae

BBMO_Dec08-SF-16	JN935218	CFB555	279	Uncultured *Flavobacterium* sp. clone ARTE4_230 (GU230411)	99	
				Owenweeksia hongkongensis (AB125062)	93	Flavo > Cryomorphaceae
BBMO_Dec08-SF-17	JN935219	CFB555	281	Uncultured *Flavobacterium* sp. clone ARTE4_230 (GU230411)	99	
				Owenweeksia hongkongensis (AB125062)	92	Flavo > Cryomorphaceae
BBMO_Jan08-SF-18	JN935220	CFB555	213	Uncultured bacterium clone 051011_T3S4_W_T_SDP_120 (FJ349919)	99	
				Brumimicrobium sp. P99 (EU195945)	94	Flavo > Cryomorphaceae
BBMO_Dec08-SF-32	JN935221	CF418	365	Uncultured *Bacteroidetes bacterium* (GU350221)	99	
				Subsaxibacter broadyi strain P7 (AY693999)	93	Flavo > Flavobacteriaceae
BBMO_Feb08-SF-33	JN935222	CF418	358	Uncultured *Flavobacterium* sp. clone ARTE1_211 (GU230407)	99	
				Kordia algicida (FJ015036)	94	Flavo > Flavobacteriaceae
BBMO_Feb08-SF-34	JN935223	CF418	386	Uncultured marine bacterium clone A6-3-51 (FJ826359)	100	
				Kordia sp. IMCC1412 (GU233518)	91	Flavo > Flavobacteriaceae
BBMO_Jan08-SF-29	JN935224	CF418	457	Uncultured *Flavobacteria bacterium* clone Vis_St6_66 (FN433385)	99	
				Owenweeksia hongkongensis (AB125062)	91	Flavo > Cryomorphaceae
BBMO_Jan08-SF-30	JN935225	CF418	433	Uncultured *Flavobacterium* sp. clone IMS2D23 (JN233066)	99	
				Fluviicola sp. NBRC 101268 (AB517714)	93	Flavo > Cryomorphaceae
BBMO_Oct07-SF-A	JN935226	CF968	520	Uncultured *Bacteroidetes bacterium* clone SW59 (HM474983)	99	
				Owenweeksia hongkongensis (AB125062)	89	Flavo > Cryomorphaceae
BBMO_Dec07-SF-C	JN935227	CF968	526	Uncultured *Flavobacteria bacterium* clone Vis_St6_66 (FN433385)	99	
				Owenweeksia hongkongensis (AB125062)	90	Flavo > Cryomorphaceae
BBMO_Dec07-SF-F	JN935228	CF968	405	Uncultured *Flavobacteria bacterium* clone 211-17-40 (HQ433385)	100	
				Owenweeksia hongkongensis (AB125062)	89	Flavo > Cryomorphaceae
BBMO_Dec07-SF-G	JN935229	CF968	407	Uncultured *Flavobacteria bacterium* clone 211-17-40 (HQ433385)	100	
				Owenweeksia hongkongensis (AB125062)	89	Flavo > Cryomorphaceae
BBMO_Jan08F-SF-H	JN935230	CF968	433	Uncultured c clone SHAZ748 (GQ350099)	99	
				Fluviicola sp. NBRC 101268 (AB517714)	92	Flavo > Cryomorphaceae
BBMO_Jul08-SF-I	JN935231	CF968	469	Uncultured *Bacteroidetes bacterium* clone SHBC1031 (GQ350134)	96	
				Coccinistipes vermicola strain IMCC1411 (EF108212)	91	Flavo > Flavobacteriaceae
BBMO_Jul08-SF-J	JN935232	CF968	554	Uncultured *Saprospiraceae bacterium* clone PEACE2006/81_P3 (EU394567)	99	
				Lewinella cohaerens partial type strain ATCC 23123T (AM295254)	90	Sphingo > Saprospiraceae

BBMO_Aug08-SF-K	JN935233	CF968	556	Uncultured *Bacteroidetes bacterium* clone SW59 (HM474983)	99	
				Owenweeksia hongkongensis (AB125062)	90	Flavo > Cryomorphaceae
BBMO_May08-SF-M	JN935234	CF968	560	Uncultured bacterium clone 1301APX_G09 (GU189020)	98	
				Owenweeksia hongkongensis (AB125062)	88	Flavo > Cryomorphaceae
BBMO_Sep08-SF-N	JN935235	CF968	358	Uncultured *Bacteroidetes bacterium* clone T42_55 (DQ436765)	98	
				Owenweeksia hongkongensis (AB125062)	94	Flavo > Cryomorphaceae
BBMO_Sep08-SF-O	JN935236	CF968	527	Uncultured *Bacteroidetes bacterium* clone SW59 (HM474983)	99	
				Owenweeksia hongkongensis (AB125062)	90	Flavo > Cryomorphaceae
BBMO_Jan08-SF-P	JN935237	CF968	554	Uncultured *Flavobacteria bacterium* clone Vis_St6_66 (FN433385)	100	
				Owenweeksia hongkongensis (AB125062)	90	Flavo > Cryomorphaceae
BBMO_Apr08-SF-Q	JN935238	CF968	551	Uncultured *Bacteroidetes bacterium* clone SHAW696 (GQ349807)	100	
				Owenweeksia hongkongensis (AB125062)	90	Flavo > Cryomorphaceae
BBMO_May08-SF-T	JN935239	CF968	434	Uncultured *Bacteroidetes bacterium*, clone: TKTMmvp-B38 (AB611551)	96	
				Owenweeksia hongkongensis (AB125062)	90	Flavo > Cryomorphaceae
BBMO_May08-SF-U	JN935240	CF968	574	Uncultured *Bacteroidetes bacterium* clone SW59 (HM474983)	99	
				Owenweeksia hongkongensis (AB125062)	90	Flavo > Cryomorphaceae
BBMO_Apr08-SF-S	JN935241	CF968	566	Uncultured bacterium clone S25_281 (EF573937)	98	
				Owenweeksia hongkongensis (AB125062)	92	Flavo > Cryomorphaceae
BBMO_Apr08-SF-R	JN935242	CF968	565	Uncultured *Flavobacteria bacterium* clone Vis_St6_66 (FN433385)	99	
				Owenweeksia hongkongensis (AB125062)	91	Flavo > Cryomorphaceae
BBMO_Sep08-SA-2	JN935243	CF418	406	Uncultured bacterium clone PROA25S_33 (GQ916124)	100	
				Lewinella nigricans partial type strain ATCC 23147T (AM295255)	91	Sphingo > Saprospiraceae
BBMO_Aug08-SA-4	JN935244	CF418	336	Uncultured *Flavobacterium* sp. clone ARTE1_211 (GU230407)	99	
				Kordia algicida (FJ015036)	94	Flavo > Flavobacteriaceae
BBMO_Aug08-SA-5	JN935245	CF418	412	Uncultured *Flavobacterium* sp. clone ARTE4_105 (GU230412)	99	
				Coccinistipes vermicola strain IMCC1411 (EF108212)	88	Flavo > Flavobacteriaceae
BBMO_May08-SA-8	JN935246	CF418	419	Uncultured *Bacteroidetes bacterium*, DGGE band: HB03-36 (AB266020)	99	
				Sediminicola luteus, strain: PMAOS-27 (AB206959)	92	Flavo > Flavobacteriaceae
BBMO_May08-SA-9	JN935247	CF418	413	Uncultured bacterium isolate DGGE gel band SD8 (FJ406506)	100	
				Lewinella nigricans partial type strain ATCC 23147T (AM295255)	92	Sphingo > Saprospiraceae

BBMO_Apr08-SA-10	JN935248	CF418	411	Uncultured bacterium clone PROA61B_11 (GQ916469)	99	
				Flexibacter echinicida strain F11 (AY006470)	96	Cyto > Cytophagaceae
BBMO_Apr08-SA-12	JN935249	CF418	440	Uncultured bacterium isolate DGGE gel band SD8 (FJ406506)	99	
				Lewinella nigricans partial type strain ATCC 23147T (AM295255)	91	Sphingo > Saprospiraceae
BBMO_Apr08-SA-13	JN935250	CF418	423	Uncultured bacterium isolate DGGE gel band SD8 (FJ406506)	99	
				Lewinella nigricans partial type strain ATCC 23147T (AM295255)	92	Sphingo > Saprospiraceae
BBMO_Apr08-SA-14	JN935251	CF418	372	Uncultured bacterium isolate DGGE gel band SD8 (FJ406506)	99	
				Lewinella nigricans partial type strain ATCC 23147T (AM295255)	91	Sphingo > Saprospiraceae
BBMO_Apr08-SA-15	JN935252	CF418	413	Uncultured bacterium isolate DGGE gel band SD8 (FJ406506)	99	
				Lewinella nigricans partial type strain ATCC 23147T (AM295255)	91	Sphingo > Saprospiraceae
BBMO_Mar08-SA-16	JN935253	CF418	329	Uncultured *Flavobacterium* sp. clone SHWN_night2_16S_587 (FJ745097)	97	
				Gelidibacter gilvus strain Z20 (HQ538744)	95	Flavo > Flavobacteriaceae
BBMO_Mar08-SA-17	JN935254	CF418	332	Uncultured *Bacteroidetes bacterium* clone HNA_I_100608-58 (HM037547)	98	
				Gelidibacter gilvus strain Z20 (HQ538744)	93	Flavo > Flavobacteriaceae
BBMO_Mar08-SA-18	JN935255	CF418	327	Uncultured *Flavobacteria bacterium* clone Vis_St6_26 (FN433461)	99	
				Kordia algicida (FJ015036)	94	Flavo > Flavobacteriaceae
BBMO_Feb08-SA-19	JN935256	CF418	351	Uncultured Flavobacteriaceae bacterium clone S.o-5 (HM031429)	98	
				Tenacibaculum sp. MED341 (EU253574)	98	Flavo > Flavobacteriaceae
BBMO_Feb08-SA-20	JN935257	CF418	363	Uncultured bacterium, note: sequence type NZ-BA-3 (AB239760)	98	
				Cytophaga sp. N05VI partial isolate N05VI (AJ786088)	93	Cyto > Cytophagaceae
BBMO_Feb08-SA-20	JN935258	CF418	286	Uncultured *Flavobacteria bacterium* clone 1_G10 (EU600576)	98	
				Coccinistipes vermicola strain IMCC1411 (EF108212)	88	Flavo > Flavobacteriaceae
BBMO_Jan08-SA-23	JN935259	CF418	354	Uncultured *Flavobacterium* sp. clone ARTE4_230 (GU230411)	100	
				Owenweeksia hongkongensis strain UST20020801 (NR_040990)	91	Flavo > Cryomorphaceae
BBMO_Jan08-SA-24	JN935260	CF418	367	Uncultured *Flavobacterium* sp. clone IMS2D23 (JN233066)	98	
				Fluviicola sp. NBRC 101268 (AB517714)	92	Flavo > Cryomorphaceae
BBMO_Jan08-SA-25	JN935261	CF418	411	Uncultured *Flavobacterium* sp. clone IMS2D23 (JN233066)	100	
				Fluviicola sp. NBRC 101268 (AB517714)	94	Flavo > Cryomorphaceae
BBMO_Jan08-SA-26	JN935262	CF418	297	Uncultured *Flavobacterium* sp. clone IMS2D23 (JN233066)	96	
				Lishizhenia caseinilytica strain UST040201-001 (NR_041043)	90	Flavo > Cryomorphaceae

BBMO_Dec07-SA-27	JN935263	CF418	425	Uncultured bacterium clone PROA45B_72 (GQ916039)	99	
				Formosa crassostrea strain KO3 (EU328155)	96	Flavo > Flavobacteriaceae
BBMO_Nov07-SA-28	JN935264	CF418	357	Uncultured *Bacteroidetes bacterium* clone PRE-S47 (DQ656277)	96	
				Winogradskyella echinorum strain KMM 6211 (EU727254)	95	Flavo > Flavobacteriaceae
BBMO_Nov07-SA-29	JN935265	CF418	262	Uncultured bacterium clone ProA23S_72 (GQ916085)	99	
				Cellulophaga fucicola strain WH10-2-3 (HM473184)	93	Flavo > Flavobacteriaceae
BBMO_Oct07-SA-31	JN935266	CF418	414	Uncultured *Flavobacterium* sp. clone ARTE4_105 (GU230412)	99	
				Coccinistipes vermicola strain IMCC1411 (EF108212)	88	Flavo > Flavobacteriaceae
BBMO_Jul08-SA-32	JN935267	CF418	429	Uncultured bacterium clone PROA25S_33 (GQ916124)	99	
				Lewinella nigricans partial type strain ATCC 23147T (AM295255)	91	Sphingo > Saprospiraceae

Names are coded after the sampling location (BBMO), sampling date, sampling depth (S=surface), filter fraction (F= small-fraction, < 3 mm; A=large fraction, > 3 mm) and band reference. The correspondence of sequences to bands marked in the figures is as follows: JN935205 - JN935225 (Fig. 1); JN935226 - JN935242 (Fig. 2SMb); JN935243 - JN935267 (Fig. 2). For each sequence, the top Blast hits including or excluding "uncultured/environmental" sequences is given. Abbreviations: L= Length; S= Similarity; Chloro=Chlorophyta; Alpha= Alphaproteobacteria; Gamma= Gammaproteobacteria; Procho= Prochlorales; Flavo= Flavobacteria; Sphingo= Sphingobacteria; Cyto= Cytophagia.

Chapter 3

Spatial and temporal variability among marine Bacteroidetes populations in the NW Mediterranean Sea

Abstract

The abundance and structure of Bacteroidetes populations at diverse temporal and spatial scales were investigated in the Northwestern Mediterranean Sea. At a temporal scale, their relative abundance exhibited a marked seasonality, being higher in spring and decreasing in winter. Similarly, Bacteroidetes community structure encompassed three main groups (winter, spring and summer-fall), which mimicked global bacterioplankton seasonality. At the spatial scale, relative abundances were similar in all surface samples along an inshore-offshore transect, but they decreased with depth. Analysis of the community structure identified four markedly different groups mostly related to different depths. Interestingly, seasonal changes in abundance and community structure were not synchronized. Furthermore, richness was higher when Bacteroidetes were less abundant. The variability of Bacteroidetes contributions to community structure in the temporal and spatial scales was correlated with different environmental factors: day length was the most important factor at the temporal scale, and salinity at the spatial scale. The community composition in terms of phylotypes changed significantly over time and along the depth gradients, but season or depth-specific phylogenetic clusters were not identified. Delineation of coherent Bacteroidetes sub-clusters should help to uncover higher resolution patterns within Bacteroidetes and explore associations with environmental and biological variables.

Spatial and temporal variability among marine Bacteroidetes populations in the NW Mediterranean Sea. **Cristina Díez-Vives, Josep M. Gasol and Silvia G. Acinas.** Systematic and Applied Microbiology, in press; doi 10.1016/j.syapm.2013.08.006

Introduction

Members of the Phylum Bacteroidetes are particularly common in the oceans, representing in some sites up to 40% of total bacteria counts determined by in situ hybridization [Glöckner *et al.,* 1999; Cottrell & Kirchman 2000a; Kirchman 2000; Alonso *et al.,* 2007]. Bacteroidetes often constitute the most abundant bacterial group in polar marine environments [Simon *et al.,* 1999; Kirchman *et al.,* 2003, Wells, 2003], coastal waters [Cottrell & Kirchman, 2000a; Eilers *et al.,* 2001; O'Sullivan *et al.,* 2004; Alonso-Sáez *et al.,* 2007] as well as in upwelling systems [Salat, 1999], particularly during phytoplankton blooms [Simon *et al.,* 1999; Eilers at al., 2001; Pinhassi *et al.,* 2004; Abell & Bowman 2005b; Alderkamp *et al.,* 2006; Mary *et al.,* 2006], indicating certain preference for productive habitats. Their ability to degrade polysaccharides and proteins [Cottrell & Kirchman, 2000b; Kirchman, 2002; Fernández-Gómez at al., 2013] suggests that they may play a key role as consumers of algae-derived metabolites. Some studies have also suggested that they occupy a "particle-specialist" niche [Delong *et al.,* 1993; Crump *et al.,* 1999; Crespo *et al.,* 2013]. However, others have reported their presence in both, attached and free-living niches, with some differences between the populations occurring in both fractions [Fandino et al., 2001; Rink *et al.,* 2007; Crespo *et al.,* 2013].

Several studies have focused on the quantification of total marine Bacteroidetes (e.g. [Fandino *et al.,* 2005; Alonso *et al.,* 2007]), and some have differentiated distinct Flavobacteria clusters [Eilers *et al.,* 2001; Kirchman *et al.,* 2003; Brinkmeyer *et al.,* 2003, Malmstrom *et al.,* 2007, Gómez-Pereira *et al.,* 2010; Teeling *et al.,* 2012]. In addition, a few group-specific fingerprinting studies have described the dominant populations in a variety of habitats [Jaspers *et al.,* 2001; Riemann 2001; Rink *et al.,* 2007; Rink *et al.,* 2008; Díez-Vives *et al.,* 2012]. However, comprehensive studies of specific populations along spatial and temporal scales are largely unexplored (e.g., [Bümel *et al.,* 2007]).

Some data on Bacteroidetes community dynamics have been derived from the study of bacterioplankton communities using DGGE with universal primers, but these usually recover few Bacteroidetes bands (e.g., [Casamayor *et al.,* 2002; Schauer *et al.,* 2003; Kan *et al.,* 2006; Alonso *et al.,* 2007; Rink *et al.,* 2007]), or even none (e.g. [Rieman *et al.,* 1999]). In addition, biases of universal bacterial primers against Bacteroidetes have been widely reported [Cottell & Kirchman 2000a; Eilers *et al.,* 2000; Kirchman *et al.,* 2003]. Group-specific primers are likely to provide a more detailed view of specific groups (e.g. [Rink *et al.,* 2007]), but even group-specific oligonucleotide probes such as

the highly used CF319a may not yield accurate quantitative data [Amann and Fuchs 2008; Weller *et al.,* 2000; Cottrell *et al.,* 2005] due to limited group coverage and specificity [Díez-Vives *et al.,* 2013].

In this work, we explored Bacteroidetes dynamics and structure at temporal and spatial scales, analyzing the influence of environmental and biological factors on the abundance and composition of this group. We focused our study in the Northwestern Mediterranean Sea, sampling monthly during a 2-year time series and in a spatial transect including 5 stations with vertical profiles that include the main hydrographical features of the inshore-offshore gradient.

Material and Methods

Sampling

The seasonal study was conducted at the Blanes Bay Microbial Observatory (BBMO, www.icm.csic.es/bio/projects/icmicrobis/bbmo/) (41°40'N, 2°48'E), from September 2006 to September 2008. This site is relatively oligotrophic [Gasol *et al.,* 2012], just 1 km offshore. Surface (0.5 m depth) waters were sampled monthly, filtered through a 200 µm mesh net and transported to the lab (within 1.5 h) under dim light in 25 L polycarbonate carboys.

We also carried out samplings along an inshore-offshore transect comprising five stations in which we sampled vertical profiles (Fig. S1). This transect was sampled from on board "BO García del Cid" during the "Modivus" cruise (September 20 to 23, 2007). This area has been repeatedly studied for physical oceanography [Salat, 1996], water chemistry [Doval *et al.,* 2001], phytoplankton [Estrada and Salat, 1989; Estrada *et al.,* 1993; Estarda *et al.,* 1999], bacteria [Gasol *et al.,* 1995; Pedrós-Alió *et al.,* 1999; Pommier *et al.,* 2010], or viruses [Guixa-Boixereu *et al.,* 1999]. Samples were collected with Niskin bottles mounted on a rosette equipped with a SAIV A/S model SD204 CTD. Water was prefiltered through a 200 µm mesh and processed immediately on board.

Figure S1. Bathymetric map of the NW Mediterranean area showing the Blanes Bay Microbial Observatory (BBMO), and the sampling stations of the spatial transect sampled in the MODIVUS cruise (C, CM, M, MD and D). Sampling depth is noted between brackets at each point.

Environmental and biological variables

Methods for measurements of water temperature, chlorophyll a (Chl *a*) concentration, fluorescence, and dissolved inorganic nutrient concentrations were standard, and have been previously described (e.g. [Alonso-Sáez *et al.,* 2008]). *Synechococcus*, *Prochlorococcus* and photosynthetic picoeukaryotes were enumerated by flow cytometry and distinguished by their different sizes and pigment properties in unstained samples following common procedures (i.e. [Marie *et al.,* 1997]). Heterotrophic prokaryotes were also counted by flow cytometry [Gasol & Del Giorgio, 2000] after staining with SybrGreen I. Chl *a* concentration and photosyntethic picoplankton were not measured at depths below 140 meters.

Catalyzed reporter deposition-fluorescence in situ hybridization (CARD-FISH).

For determination of Bacteroidetes abundance by CARD-FISH [Pernthaler *et al.,* 2002], 25 mL seawater samples were fixed with 2% paraformaldehyde (final concentration), overnight at 4 °C in the dark and gently filtered through 0.2 µm polycarbonate filters (OSMONICS INC., 47 mm), which were rinsed with Milli-Q water and stored at -20 ℃ until processing. The labeling procedure was carried out using a Bacteroidetes-specific

oligonucleotide probe CF968 (5'-GGTAAGGTTCCTCGCGTA-3'). This probe has been previously published as PCR and QPCR primer [Chen *et al.,* 2006; De Gregoris *et al.,* 2011]. In a recent revision, coverage of CF968 for the phylum Bacteroidetes was 92.8%,with a high specificity [Díez-Vives *et al.,* 2012]. Recently, this probe has been successfully applied in the CARD-FISH protocol [Acinas *et al.,* submitted]. Briefly, CF968 was synthesized with a 5'-aminolink (Thermo Fischer Scientific Ulm, Germany) and subsequently labeled with HRP (Roche Diagnostics, Barcelona, Spain) according to Urdea *et al.* [1988] and Amann & Fuchs [2008]. The optimal formamide concentration for the CF968 probe was found to be 55% [Acinas *et al.,* submitted]. Counts were obtained from a minimum of 18 microscope fields representing more than 1000 DAPI-positive cells on each filter. The results (mean and standard deviation) are expressed as the percentage of positive CARD-FISH cells relative to the total DAPI- stained cells (Table S1).

DNA extraction and PCR-DGGE

To collect bacterioplankton biomass, seawater (8-10 L per sample) was filtered using a peristaltic pump sequentially through GF/A filters (47 mm, Whatman) and polycarbonate 0.22 µm pore size filters in the spatial study (Modivus). For samples taken at the coastal surface station (BBMO) we used 47 mm diameter, 3 µm pore polycarbonate filters (Whatman) and 0.22 µm pore Sterivex filters (Durapore, Millipore). The Sterivex units were treated with 1.8 mL of lysis buffer (50 mM Tris- HCl, 40 mM EDTA, 0.75 M sucrose, pH 8.0) and kept at -80°C, whereas the Modivus filters were directly stored at -80 ℃ after flash freezing in liquid N2. Microbial biomass was digested with lysozyme, proteinase K and sodium dodecyl sulfate, and the nucleic acids were extracted with phenol and concentrated in a Centricon-100 (Millipore), as described in Massana *et al.* [1997b].

DNA samples were amplified for DGGE analyses using a Bacteroidetes-specific forward primer (GC-CF418) with a universal bacterial reverse primer (907RM), as described previously [Díez-Vives *et al.,* 2012]. In order to compare Bacteroidetes with bacterioplankton community fingerprints, additional DGGE analyses were carried out with samples from a 1-year time series (from September 2007 to September 2008) using universal bacterial primers (GC-358F and 907RM) [Muyzer *et al.,* 1998], under similar PCR conditions. In both cases, the forward primer included a 5'-CG-clamp (CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG G). Optimization of the template DNA amount (1 ng, 4 ng, 8 ng and 10 ng DNA per reaction) was necessary to improve PCR yields

with environmental samples, depending on the levels of inhibitory residues present. PCR reactions (50 µL final volume) contained 200 µM each deoxynucleoside triphosphates, 0.2 µM each primer, 2 mM $MgCl_2$, 1x PCR buffer, and 1 U Taq DNA polymerase (Invitrogen). The amplification cycle consisted of an initial denaturation step (94 °C, 5 min); 30 cycles of denaturation (94 °C, 30 s), annealing (55–63 °C, 30 s), and extension (72 °C, 1 min); and a final extension (72 °C, 10 min). PCR products were analyzed and quantified by agarose gel electrophoresis using standards (Low DNA Mass Ladder, GIBCO BRL). To mitigate random biases associated with individual reactions, three PCRs of the same template were run and pooled prior to the DGGE analyses. Similar amounts of the PCR products were loaded on 6% polyacrylamide gels cast in a continuous DNA-denaturing gradient (40-70% for Bacteroidetes-specific and 40-80% for universal bacterial gels; 100% denaturant is 7 M urea in 40% deionized formamide). Gels were run at 100 V for 16 h at 60°C in TAE running buffer (40 mM Tris pH 7.4, 20 mM sodium acetate, 1 mM EDTA) using a DGGE-2000 system (CBS Scientific, Del Mar, CA, USA). The gels were stained with SYBR-Gold (Molecular Probes, Eugene, OR, USA) for 15 min and documented with a Fluor-S MultiImager system (Bio-Rad, Hercules, CA, USA).

Relative quantification of DGGE fingerprints

Gel images were analyzed using the Quantity-One software (Bio-Rad). The bands occupying the same position in different lanes of the gels were identified and refined by visual inspection. A data matrix was constructed considering the presence or absence of individual bands in each lane and their relative contribution (in percentage) to the total absorbance of the lane. This dataset was analyzed using the software PRIMER 6 (Primer-E Ltd.). A distance matrix was calculated using the Bray-Curtis dissimilarity, and a dendrogram was obtained from the resemblance matrix by unweighted pair-group method with arithmetic mean (UPGMA). “Similarity profile” (SIMPROF) permutations tests were computed over the dendrograms, seeking statistically significant evidence for the existence of genuine clusters in samples that were a priori unstructured. These tests were done at each node, to verify whether the subgroups had significant internal structure. The distance matrix was also used for a similarity analysis by one-way ANOSIM with all possible permutations.

Sequencing and phylogenetic analysis

The DGGE bands were excised, placed in 20 µl Milli-Q water, and incubated overnight at 4 ℃. The supernatants (2 µl) were used as template DNA for re-amplification with the original primer set (without the GC-clamp) and the products submitted for automated Sanger sequencing (Macrogen Inc., Amsterdam, Netherlands). Several bands occupying the same position in different gel lanes were excised and sequenced to confirm correct matching of the bands. Due to the large amount of bands per lane and their proximity in the universal bacteria DGGE, the excised bands were re-amplified and re-run into a new DGGE to verify their position and uniqueness.

The sequences obtained from the DGGE bands, were compared against Genbank, National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) using BLAST [Altschul *et al.,* 1990] to determine their phylogenetic affiliation. Sequences were aligned with homologous 16S rRNA sequences from SILVA SSU Reference NR database release 106, using the automated aligning tool contained in ARB software package (http://www.arb-home.de) [Ludwig *et al.,* 2004]. The aligned sequences were added to the guide tree using the "quick add" parsimony tool, which does not alter the starting tree topology. The resulting tree was pruned to retain only the relevant closest taxa. Average pairwise percent similarity of overall 16S rDNA sequences, as well as average similarity within and between groups was estimated from an alignment of 54 Bacteroidetes sequences using the MEGA5 program [Tamura *et al.,* 2011]. Gene sequences were deposited in Genbank under accession numbers JX865538-JX865592.

Statistics

We performed two different statistical analyses. Firstly, in order to explore globally the pairwise correlations using raw values, we used Spearman´s rank correlation tests between Bacteroidetes relative contribution to the total abundance of Bacteria (referred to as Bacteroidetes percentage throughout the paper) and the biological and environmental variables included in Table 1. Correlations were analyzed for significance ($p<0.05$ or $p<0.01$) by a Student's 2-tailed t-test. Secondly, we used multiple regression analyses with Bacteroidetes percentage as dependent variable and the biological/environmental variables as predictors. This allowed us to determine the combination of these variables that explained most of the variability of Bacteroidetes percentage observed. We used forward stepwise multiple regression where the independent variables are individually added to the model at each step of the regression until the "best" regression model is

obtained. The variable entered at each step is the one that produces the largest increase in R^2 i.e. the one with the largest Pearson's partial correlation with the Bacteroidetes percentage. This procedure tests the significance of all the variables in the model, and those that are not significant are excluded before the next forward selection step (only intercepts and variables with significant ($P < 0.05$) contributions to the regression model are retained) [Sokal & Rohlf, 1995]. We tested for multicollinearity between the predictor variables. In doing so, we explored correlations among predictor variables, and calculated tolerances for each variable [Sokal & Rohlf, 1995]. Statistical analyses were performed using SPSS v.19.0.

Results

Temporal dynamics in BBMO.

The BBMO site is characterized by a clear seasonality (Fig. S2A), with temperature values ranging from 12.9 to 26.2 °C and salinity from 37 to 39 psu. Summers are characterized by high water temperature (20-26 °C) and low concentration of inorganic nutrients (e.g. PO_4= 0.1 µM) and Chl *a* (ca. 0.3 µg l^{-1}). In contrast, winters present water temperatures around 13-16 °C and higher concentration of nutrients (e.g. PO_4= 1.1 µM) and Chl *a* (1-2 µg l^{-1}). December 2006 and January 2007 displayed slightly warmer temperatures than the following year. Bacterial abundance values were quite stable ranging between 0.43 and $1.6 \cdot 10^6$ cells ml^{-1}. *Synechococcus* abundance was typically high in summer and contributing up to 13.8% to total prokaryote numbers. *Prochlorococcus* values were usually lower (up to 4.5%) and increased in the fall and winter, when *Synechococcus* started to decline. Photosyntethic picoeukaryotes were more abundant during late winter and spring (up to $1.7 \cdot 10^4$ cells ml^{-1}), declining during summer (details not shown).

Abundance of Bacteroidetes and their environmental and biological predictors.

The abundance of Bacteroidetes was quantified by CARD-FISH during the 2-year time series at the BBMO station. The percentage contribution of Bacteroidetes to total community was variable over time accounting from 8 to 22 % of total DAPI counts (Table S1). A seasonal pattern was observed, with the highest values measured from April to May in 2007 and from April until August in 2008, with a marked descent after these periods (Fig. 1A). In terms of absolute abundances, Bacteroidetes counts were significantly higher ($p<0.001$, ANOVA $F(1, 23) = 24.94$) from April to August ($10.0 \pm 1.0 \cdot 10^4$ cells ml^{-1}) than from September to March ($7.8 \pm 1.5 \cdot 10^4$ cells ml^{-1}) in both years (Table S1).

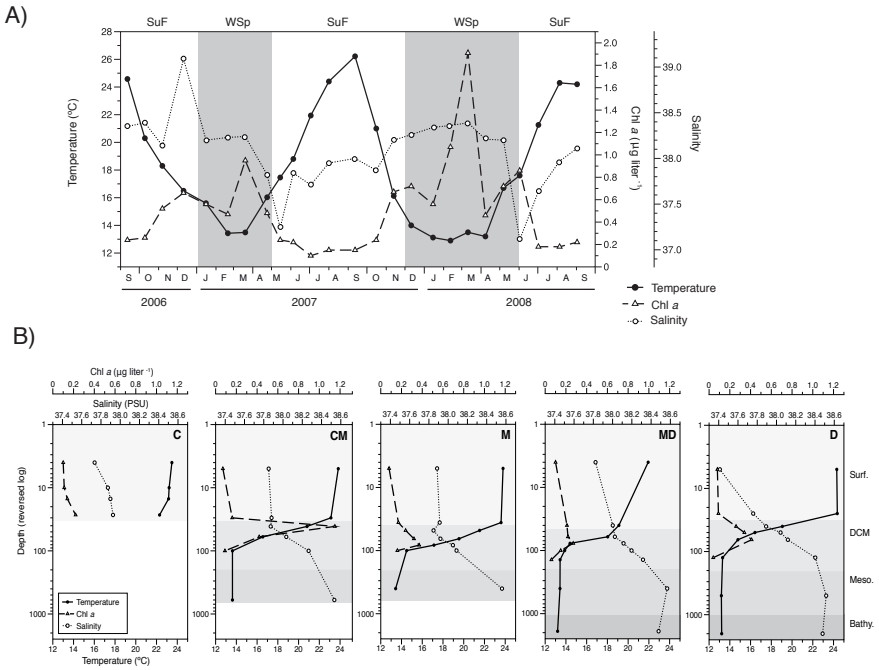


Figure S2. A) Monthly measurements of selected parameters taken during the period September 2006–September 2008, including the division of "Winter-Spring" and "Summer-Fall" groups based on the clustering analyses (see main text). B) Vertical profiles of the same parameters measured at the five stations sampled during the Modivus cruise, including a distinction of surface, DCM, Mesopelagic and Bathypelagic zones.

In the bivariate correlations analyses, only 2 physical and 2 biological variables showed a significant correlation with Bacteroidetes percentage in the temporal series: positive with daylength and negative with ammonia, *Prochlorococcus* and total Bacteria abundance (Table 1). Multiple regression analyses of the temporal series explained a relatively large fraction of the variability of Bacteroidetes percentage contribution to community structure with only a few environmental/biological variables. The best model explained 57.5% of variability of Bacteroidetes percentage with only 3 variables in the following order and sign: day length (+), total bacterial abundance (-) and NH_4 (-) (see Table 2 for the cumulative R^2 explained by each variable). This model did not include Prochlorococus counts, the variable with the largest Spearman's rank correlation value. The predictor variables were not correlated among them and hence no collinearity test was needed. We further tested model residuals and they were normally distributed, as required by this procedure. We also performed the same multiple regression analyses using Bacteroidetes abundance, and we obtained similar results (Table S2).

Table S1. Bacterial abundance in the 2-year time series (left) and in the spatial transects (right) sampled in the study. Total heterotrophic bacteria were enumerated by flow cytometry (column "Bacteria"), while Bacteroidetes relative abundance (column "% CARD-FISH") is expressed as the percentage of cells labelled by CARD-FISH with a Bacteroidetes specific probe, relative to the total DAPI counts. The Bacteroidetes absolute abundance values inferred from these values are also presented (column "Bacteroidetes"). Monthly samples were taken at BBMO from September 2006 to September 2008 and spatial samples belong to the Modivus cruise, in which C, CM, M, MD and D denote the sampled stations along the horizontal transect, followed by the depth at each sampling point ("St-depth").

Temporal (dd-mm-yy)	Bacteria (10^4 cells ml^{-1})	% CARD-FISH	Bacteroidetes (10^4 cells ml^{-1})	Spatial (St-depth)	Bacteria (10^4 cells ml^{-1})	% CARD-FISH	Bacteroidetes (10^4 cells ml^{-1})
12-09-06	81.97	11.74 ± 4.6	9.62	C-4	75.45	10.01 ± 3.95	7.55
10-10-06	92.75	10.27 ± 2.94	9.52	C-15	85.22	2.77 ± 1.85	2.36
07-11-06	56.91	12.69 ± 5.25	7.22	CM-5	65.68	8.05 ± 2.61	5.29
11-12-06	89.83	10.14 ± 3.28	9.10	CM-30	84.81	3.73 ± 1.89	3.16
16-01-07	43.74	12.18 ± 5.43	5.33	CM-41	96.16	13.67 ± 3.13	13.15
20-02-07	48.62	10.66 ± 3.75	5.18	CM-60	88.77	3.78 ± 2.4	3.36
20-03-07	81.85	8.06 ± 4.82	6.60	CM-100	44.39	2.53 ± 1.24	1.12
24-04-07	52.41	19.18 ± 6.5	10.05	CM-600	8.97	0.71 ± 2.53	0.06
15-05-07	64.00	17.98 ± 4.94	11.51	M-5	38.26	5.3 ± 3.26	2.03
05-06-07	78.94	10.12 ± 5.72	7.99	M-48	52.95	3.83 ± 2.74	2.03
03-07-07	96.45	11.29 ± 4.07	10.89	M-400	13.25	1.21 ± 2.51	0.16
01-08-07	88.70	11.09 ± 5.05	9.83	MD-4	52.97	2.61 ± 1.83	1.38
12-09-07	63.49	11.38 ± 4.42	7.22	MD-60	58.08	6.2 ± 1.86	3.60
16-10-07	70.13	11.65 ± 3.08	8.17	MD-77	51.29	2.78 ± 2.43	1.43
13-11-07	66.85	10.33 ± 3.67	6.90	MD-140	23.55	2.61 ± 1.82	0.61
11-12-07	65.49	11.23 ± 4.64	7.35	MD-400	12.18	1.56 ± 3.05	0.19
15-01-08	82.21	9.24 ± 3.02	7.60	MD-1900	5.11	0.66 ± 2.69	0.03
12-02-08	48.33	15.64 ± 8.11	7.56	D-5	39.29	15.03 ± 3.02	5.90
11-03-08	76.39	12.89 ± 2.66	9.85	D-25	49.24	11.12 ± 2.52	5.48
08-04-08	88.48	15.4 ± 3.15	13.62	D-40	51.97	9.49 ± 3.00	4.93
07-05-08	57.05	21.98 ± 4.5	12.54	D-65	54.92	6.06 ± 3.12	3.33
02-06-08	62.67	16.92 ± 2.67	10.60	D-500	9.91	0.6 ± 2.79	0.06
02-07-08	65.93	16.27 ± 5.31	10.73	D-2000	4.26	0.87 ± 2.99	0.04
05-08-08	106.11	14.32 ± 1.64	15.19				
02-09-08	73.46	12.74 ± 2.65	9.36				

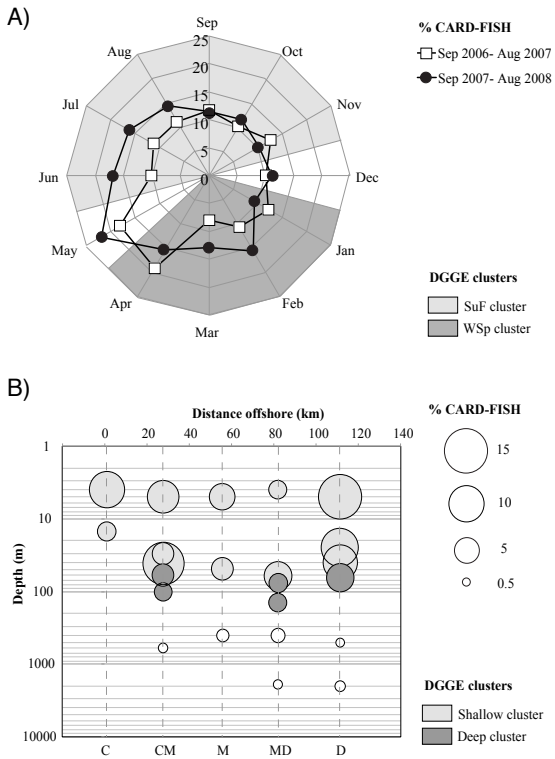


Fig 1. Percentage of Bacteroidetes in CARD-FISH counts: A) over two consecutive years in the temporal study, represented as circles and squares and B) in the spatial study as circle areas for each sample point. The groups defined by the DGGE clustering analyses (SuF/WSp or Shallow/Deep) are also represented as different grey tones. Since transitional months (May and Dec) fell into different clusters in the 2 years analyzed, then are colored in white. Samples under 200 m (not included in the DGGEs) are also in white.

Table 1. Spearman rank correlations between environmental factors and Bacteroidetes percentage contribution to the bacterial community. Statistically significant correlations, P < 0.01 (**) or P < 0.05 (*), are highlighted in bold. N= 25 for the temporal series and N=17 for the spatial transect.

		Temporal		Spatial	
		rs value	P value	rs value	P value
Physical	Depth	n.a.	n.a.	-0.388	0.124
	Temperature	0.053	0.801	**0.560****	**0.019**
	Fluorescence	n.a	n.a	-0.047	0.857
	Day length	**0.444****	**0.026**	n.a.	n.a.
	Oxygen	n.a.	n.a.	0.161	0.538
	Salinity	-0.326	0.112	**-0.609****	**0.009**
	PO_4	0.185	0.477	-0.166	0,524
	NH_4	-0.499	0.011	-0.053	0.839
	NO_2	-0.017	0.936	-0.095	0.716
	NO_3	0.252	0.225	-0.451	0.069
	SiO_4	-0.083	0.693	**-0,589***	**0,013**
Biological	Chl *a*	0.044	0.833	0.087	0.740
	Chl *a* < 3 µm	-0.006	0.977	0.064	0.808
	Prochlorococcus	**-0.525****	**0.007**	0.195	0.453
	Synechococcus	-0.055	0.793	0.216	0.405
	Photosynthetic picoeukaryotes	-0.105	0.619	0.294	0.251
	Bacteria	**-0.442***	**0.027**	0.137	0.599

Table 2 Summary of multiple regression models (forward procedure) for the temporal series and the spatial transect. The independent contributions of each variable to the prediction of Bacteroidetes percentages after controlling for all other independent variables are shown (**ß**, or standardized regression coefficients). P-values are the significance of each predictor variable within the model. R^2 corresponds to corrected variance explanation.

Dependent variable	Predictor variables	Cumulative R^2	ß	P
Bacteroidetes percentage (Temporal series)	Day Lenght	0.218	0.627	0.000
	Bacteria	0.467	-0.514	0.002
	NH4	0.575	-0.342	0.021
Bacteroidetes percentage (Spatial transects)	Salinity	0.436	-0.944	0.000
	Synechococcus	0.653	-0.538	0.002
	Picoeukaryotes	0.786	0.363	0.008

Community composition by DGGE analyses.

The temporal dynamics were analyzed from September 2006 to September 2008, using DGGE fingerprinting with Bacteroidetes-specific primers. Samples from November 2006, June and July 2008 did not yield enough PCR product to be included in the dataset. Image analyses of the 22 remaining samples resulted in 292 bands identified in 31 unique positions (Fig. S3). The number of bands per sample was between 6 and 17. In general, 50% of the bands accounted for more than 80% of the total intensity per lane. No band appeared consistently in all samples throughout the year. In the clustering analyses, samples were classified in two main groups at 35% similarity between them: “Winter-Spring” (WSp) and “Summer-Fall” (SuF) (Fig. 2). We observed more bands in the WSp group (mean = 15.6, n = 10) than in the SuF group (mean = 11.3, n = 12) ($p<0.001$, ANOVA $F(1, 20) = 16.68$). The band pattern shifted significantly in the transition from winter to spring (Fig. S3). However, from the early summer to fall, the fingerprint pattern was relatively stable. Within the WSp group, two subgroups could be distinguished at 47% similarity: early and late within this period. Likewise, within the SuF group, samples taken at the end of the period clustered apart from earlier samplings although these groups were not significantly different (Fig. 2). ANOSIM tests significantly supported the existence of 3 groups in this seasonal series (T1, T2 and T3 in Fig. 2) (Global $R=0.824$, one-way ANOSIM $p < 0.001$).

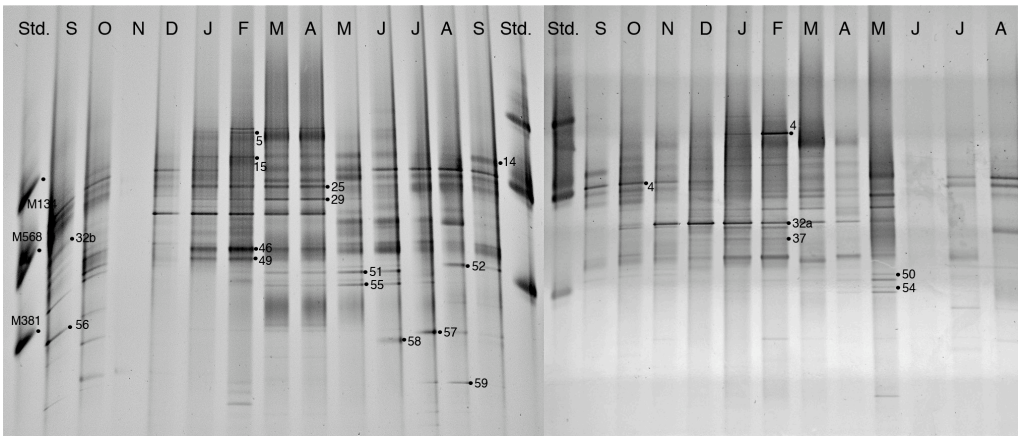


Fig. S3 DGGE analysis of 16S rDNA amplicons using Bacteroidetes-specific primers. Monthly samples over two consecutive years at the coastal sampling point (BBMO). Numbered dots mark the bands that were further excised and sequenced (Table S3). Std standard containing DNA from three marine Bacteroidetes isolates: Dokdonia sp. MED134 (DQ481462), ***Polaribacter*** sp. MED 568 (EU253599), and Maribacter sp. MED381 (EU253579).

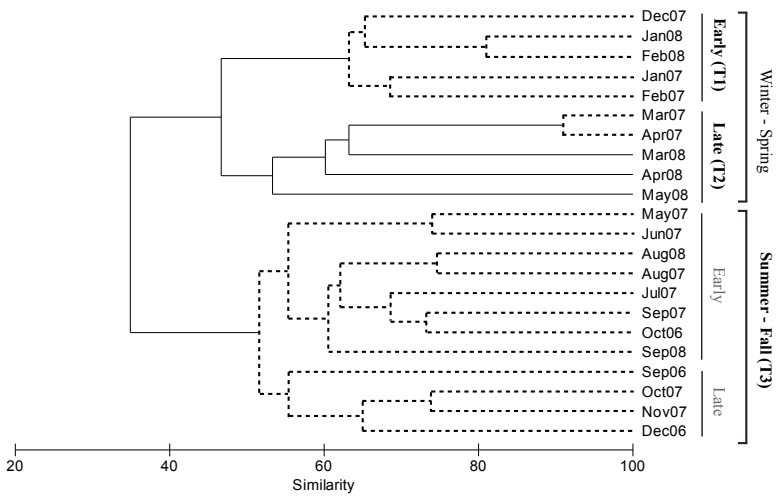


Fig 2. Cluster analyses of Bacteroidetes DGGE banding patterns using UPGMA. (W) Winter, (Sp) Spring, (S) Summer, (F) Fall. T1, T2 and T3 represent significantly different clusters in the seasonal series. The Distance matrix was calculated using the Bray-Curtis similarity algorithm. (Branches marked with dotted lines, indicate that the SIMPROF analysis could not find statistical evidence for any sub-structure within these clusters).

In order to test whether the dominant bacterioplankton populations mirrored this seasonal clustering observed among Bacteroidetes, a 1-year time series (September 2007-September 2008) was analyzed by DGGE using universal bacterial primers (Fig. S5A) and it was compared with the Bacteroidetes-specific DGGE from the same dates (Fig. 2).

The analysis of the universal bacteria gel yielded a total of 420 detectable bands in 61 different positions among the 13 samples analyzed. The number of bands per sample was rather stable varying between 30 and 35. Minor differences were observed between the bacterial (Fig. S5B) and the Bacteroidetes-specific dendrograms (Fig. 2), e.g. in the Dec07 and Jan08 samples, which branched off from other winter samples. Samples Jun08 and Jul08 were not present in the Bacteroidetes dendogram and we cannot compare their position within the clustering.

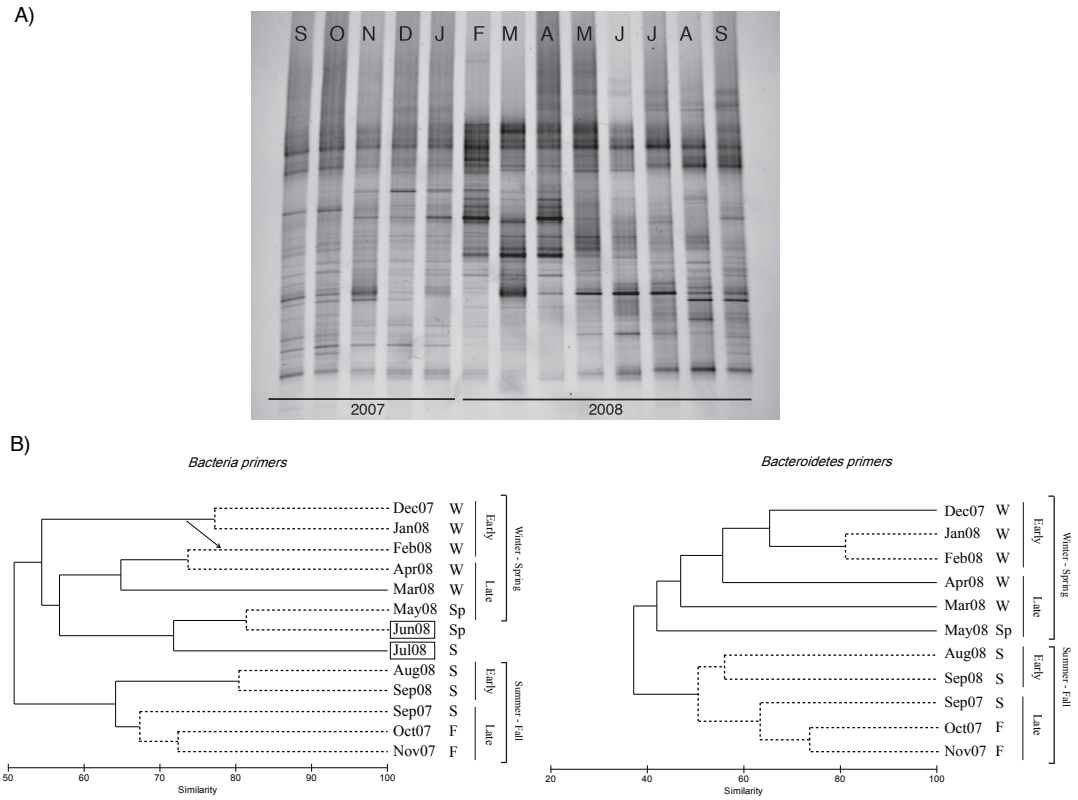


Figure S5. A) DGGE analysis of 16S rDNA amplicons using Universal Bacterial primers. The lanes contain monthly samples over a 1-year time series at the coastal sampling point (BBMO). B) Dendrograms generated from the analysis of DGGE band fingerprints using Bacterial primers (left) and Bacteroidetes-specific primers (right). The distance matrix was calculated using the Bray-Curtis similarity algorithm. The statistical significance of the clustering was calculated using the "similarity profile" permutation tests with Primer 6 software. Samples connected by dashed lines are not significantly different. The arrow marks a difference between both dendrograms in the clustering of samples Dec07 and Jan08. Months that could not be included in the Bacteroidetes study are framed in boxes in the Bacteria dendrogram.

Spatial Transect

The spatial sampling (September 2007) included stations along a transect from coastal waters to open ocean stations above a maximal water column of 2,500 m (Fig. S1). At the time of sampling, surface water temperature was 23-24 ℃ from coast to offshore while it dropped under 14 ℃ below 100 m (Fig. S2B). The thermocline was located between 30 and 40 m depth at stations CM, M and D, whereas station MD presented a superficial temperature drop (from 22 to 19 ℃ in the first 10 m) and a second temperature drop at 60 m depth (from 18 to 14 ℃). This was caused by a small cyclonic gyre (details not shown) that upwelled some colder, deep water. Maximum values of Chl *a* were found between 60 and 80 m (Fig. S2B). A Chl *a* peak, reaching up to 1.15 µg l^{-1}, was present in the DCM-layer at 44 m in station CM (mostly associated to >3 µm particles, i.e. nano- and microphytoplankton). Synechoccocus presented maximal abundance inshore and at upper water layers (above 27 m), whereas *Prochlorococcus* peaked offshore and at deeper layers (50 m). Phototrophic picoeukaryotes values reached $7.87 \cdot 10^3$ cells ml^{-1} at 44 meters in station CM (details not shown).

Spatial patterns of Bacteroidetes abundance and their environmental predictors.

In the spatial transects, the maximal abundance of Bacteroidetes was usually observed in surface waters (5 meters) (average 9.6% of DAPI counts or $5.2 \pm 2.3 \cdot 10^4$ cells ml^{-1}), with the exception of station MD (at the upwelling core), where they represented only 2.6% of the DAPI counts (Table S1). In general, the contribution of Bacteroidetes decreased with depth (Fig. 1B), although high relative values were detected at 44 m in station CM, reaching up to 13.7% ($1.35 \cdot 10^5$ cells ml^{-1}) matching the aforementioned Chl *a* peak at this station (1.15 µg Chl *a* l^{-1}) (Fig. S2).

Table S2 Summary of the multiple regression models (forward procedures) for the temporal series and the spatial transect. The independent contributions of each variable to the prediction of Bacteroidetes abundances after controlling for all other independent variables are presented (**β**, or standardized regression coefficients). P-values are the significance of each predictor variable within the model.

Dependent variable	Predictor variables	Cumulative R^2	β	P
Bacteroidetes abundance (Temporal series)	Day Length	0.404	0.567	0.000
	Bacteria	0.528	0.395	0.007
	NH4	0.628	-0.327	0.018
Bcateroidetes abundance (Spatial transects)	Peuk	0.572	0.736	0.000
	Salinity	0.732	-0.410	0.007

Bivariate correlations between Bacteroidetes percentage and the environmental and biological variables for the spatial transect (in samples above 140 meters), indicated a positive correlation with temperature and a negative correlation with salinity and SiO_4 (Table 1). The best model derived from multiple regression analyses for the spatial transect explained 78.6% of Bacteroidetes percentage with 3 variables: salinity (-) and *Synechococcus* abundance (-) and picoeukaryotes concentration (+) (Table 2). We tested for collinearity between salinity and *Synechococcus* because they were correlated ($r = -0.762$, $p < 0.001$). These two predictor variables explained a different proportion of Bacteroidetes percentage (Table 2) and since the tolerance value was relatively high (0.708) we rejected multicollinearity. Again, model residuals were normally distributed. Multiple regression analyses using Bacteroidetes abundance gave almost identical results (Table S2).

Community composition by DGGE analyses.

DGGE fingerprinting was carried out in the spatial transect samples using the Bacteroidetes-specific primers. Due to the weak amplification of samples below 400 m depth, downstream analyses could not be done although we detected Bacteroidetes by CARD-FISH down to 2,000 m depth (Table S1). The gel displayed 211 detectable bands in 47 positions (Fig. S4). The number of bands per sample was rather variable, ranging between 7 and 20. No band appeared consistently in all samples, and only 10 bands were present in at least half of the samples.

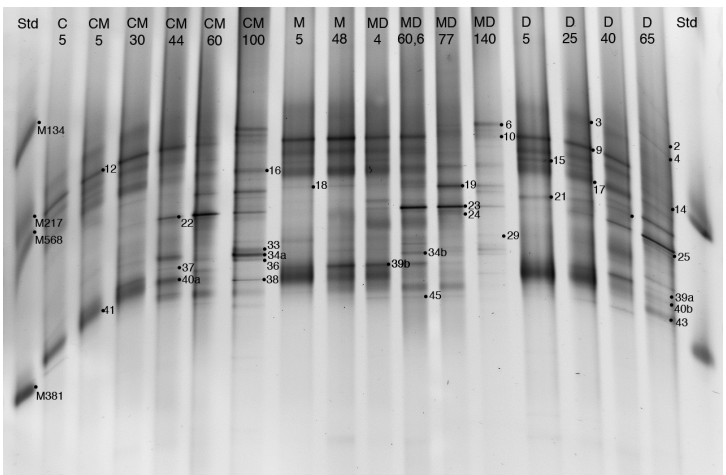


Figure S4 DGGE analysis of 16S rDNA amplicons using Bacteroidetes-specific primers. The lanes correspond to the different horizontal stations and vertical profiles sampled during the Modivus cruise. Numbered dots mark bands which were further excised and sequenced (Table S4). Std: standards containing DNA from four marine Bacteroidetes isolates: ***Dokdonia*** sp. MED134 (DQ481462), ***Leeuwenhoekiella blandensis*** MED217 (DQ294291), ***Polaribacter*** sp. MED568 (EU253599), and ***Maribacter*** sp. MED381 (EU253579).

Changes in Bacteroidetes composition were mostly associated with depth rather than with our horizontal gradient (i.e., distance from shore). Two main groups were detected with less than 20% similarity between them (Fig. 3). The "Deep" group showed higher number of bands (mean = 16.8, n = 5) than the "Shallow" group (mean = 11.5, n = 10), (ANOVA p=0.03, F (1,13) = 13.94). Sample CM44 presented a rather unique profile, clustering loosely to "Shallow" samples (40% similarity). Within these two groups, samples above 30 meters (i.e., above the thermocline as defined before) clustered apart from those collected below the thermocline. In the "Deep" group, samples collected in the DCM (between 60 to 80 meters) clustered separately from those sampled below this region (CM100, MD140) (Fig. 3). These 4 groups (S1 to S4) were significantly different (Global R = 0.930, one-way ANOSIM p < 0.001).

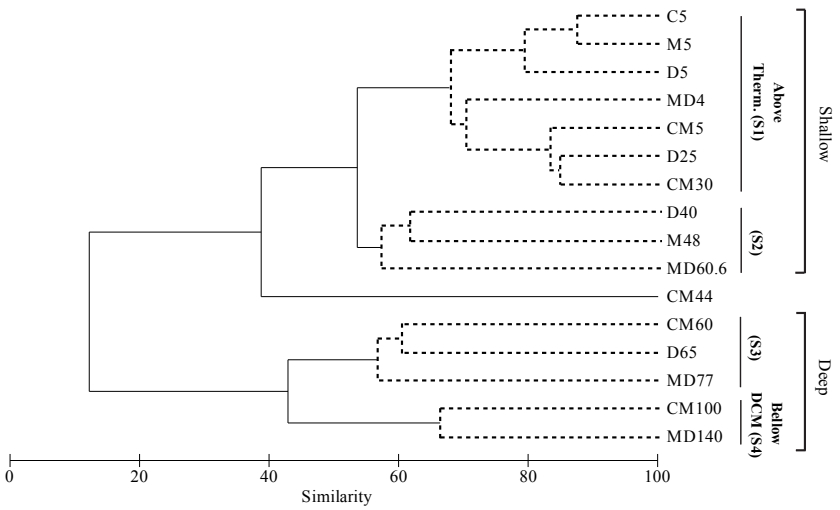


Figure 3. Cluster analyses of Bacteroidetes DGGE banding patterns using UPGMA. Therm.: thermocline; DCM: Deep chlorophyll maximum. S1, S2, S3 and S4 represent significantly different clusters in the spatial samples. The distance matrix was calculated using the Bray-Curtis similarity algorithm. (Branches marked with dotted lines, indicate that the SIMPROF analysis could not find statistical evidence for any sub-structure within these clusters).

Table S3. GenBank accession numbers and taxonomic affiliation of sequences obtained from DGGE bands, according to the Silva classifier system (Ludwig et al., 2004), and their closest match in the GenBank database (the % similarity, accession number, and origin of the sequences are listed). Sample names are coded after the sampling location (BBMO), date, depth (S=surface), filter fraction (F= small-fraction) and band ID#.

			Closest neighbor		
Acc. #	**DGGE band name**	**Taxonomic affiliation**	**%**	**Acc. #**	**Origin**
JX865538	BBMO_Feb08_SF_04	*Polaribacter*	100	EU394574	northern Bay of Biscay, pelagic, particle-associated
JX865539	BBMO_Feb07_SF_05	*Tenacibaculum*	100	EU394574	northern Bay of Biscay, pelagic, particle-associated
JX865540	BBMO_Sep07_SF_14	*Tenacibaculum*	98	AB557541	seawater
JX865541	BBMO_Feb07_SF_15	NS2b marine group	99	GU940787	South China Sea
JX865542	BBMO_Feb08_SF_19	NS4 marine group	99	JF692410	coastal urban watershed; Brazil: Jacarepagua
JX865543	BBMO_Aug08_SF_22	NS4 marine group	99	JF692410	coastal urban watershed; Brazil: Jacarepagua
JX865544	BBMO_Apr07_SF_25	NS4 marine group	100	FJ826359	filtered surface sea water at the contrasting non-bloom station in the Yellow Sea
JX865545	BBMO_Apr07_SF_29	NS4 marine group	100	HQ242382	Line P Station P4 depth of 10m
JX865546	BBMO_Feb08_SF_32a*	NS9 marine group	99	HQ242263	Line P Station P4 depth of 10m
JX865547	BBMO_Feb08_SF_32b*	NS4 marine group	99	GU940946	South China Sea
JX865548	BBMO_Aug08_SF_35	*Owenweeksia*	99	DQ473560	NW Mediterranean seawater microcosm
JX865549	BBMO_Feb07_SF_37	*Fluviicola*	99	JN233066	coastal North Carolina surface water,Beaufort Inlet
JX865550	BBMO_Feb07_SF_46	*Fluviicola*	100	JN233066	coastal North Carolina surface water,Beaufort Inlet
JX865551	BBMO_Feb07_SF_49	*Fluviicola*	99	JN233066	coastal North Carolina surface water,Beaufort Inlet
JX865552	BBMO_May08_SF_50	Saprospiraceae	99	FJ406506	mesocosm; Spain: Bay of Blanes
JX865553	BBMO_May07_SF_51	Saprospiraceae	99	FJ406506	mesocosm; Spain: Bay of Blanes
JX865554	BBMO_Aug07_SF_52	NS9 marine group	100	HM747225	Sanggou Bay (Yellow Sea) marine water assemblages between 0.22 and 3 microns"
JX865555	BBMO_May08_SF_54	Saprospiraceae	99	FJ406506	mesocosm; Spain: Bay of Blanes
JX865556	BBMO_May07_SF_55	Saprospiraceae	99	FJ406506	mesocosm; Spain: Bay of Blanes
JX865557	BBMO_Sep06_SF_56	Flavobacteriaceae	99	FJ745043	surface water at the UGA Marine Institute"; USA: Georgia, Sapelo Island
JX865558	BBMO_Jul07_SF_57	NS9 marine group	99	JN233667	coastal North Carolina surface water,Beaufort Inlet
JX865559	BBMO_Jun07_SF_58	*Owenweeksia*	99	FJ745055	surface water at the UGA Marine Institute"; USA: Georgia, Sapelo Island
JX865560	BBMO_Aug07_SF_59	Verrucomicrobia	99	GU940910	South China Sea

(*) Suffixes "a" and "b" denote sequences obtained from bands located at the same position but on different lanes and containing different sequences, thus corresponding to different phylotypes.

Table S4. GenBank accession numbers, and taxonomic affiliation of sequences obtained from DGGE bands according to the Silva classifier system, and their closest match in the GenBank database (the % similarity, accession number, origin, and depth are listed). Band names are coded after the sampling location (MOD), date, sampling station (C, CM, MD, D), depth, filter fraction (F= small-fraction) and band ID#.

			Closest neighbour		
Acc. #	**DGGE band name**	**Taxonomic affiliation**	**%**	**Acc. #**	**Origin; Depth**
JX865561	MOD_Sep07_D65F_02	NS2b marine group	99	HE647151	Pacific Ocean; Ahe atoll; n.s
JX865562	MOD_Sep07_D25F_03	NS5 marine group	99	FN433302	Atlantic Ocean; n.s.
JX865563	MOD_Sep07_D65F_04	NS2b marine group	99	GU940787	South China Sea; n.s.
JX865564	MOD_Sep07_MD140F_06	NS4 marine group	98	FN433303	Atlantic Ocean; n.s
JX865565	MOD_Sep07_D25F_09	NS4 marine group	99	FN435495	Spain;Mallorca Island; surface coastal seawater
JX865566	MOD_Sep07_MD140F_10	NS5 marine group	98	EU804964	250 miles from Panama City; n.s.
JX865567	MOD_Sep07_C5F_12	NS4 marine group	99	GQ916085	USA; west coast of Florida, Eastern Gulf of Mexico; n.s.
JX865568	MOD_Sep07_D65F_14	NS4 marine group	98	FN433303	Atlantic Ocean; n.s.
JX865569	MOD_Sep07_D5F_15	NS4 marine group	99	DQ656199	East China Sea; surface
JX865570	MOD_Sep07_D65F_16	NS4 marine group	99	HQ163249	Canada; British Columbia, Vancouver Island, Saanich Inlet; 100 m
JX865571	MOD_Sep07_D25F_17	NS4 marine group	99	DQ656199	East China Sea; surface
JX865572	MOD_Sep07_M5F_18	NS4 marine group	99	GU940946	South China Sea; n.s.
JX865573	MOD_Sep07_MD140_19	NS4 marine group	99	FN433461	Atlantic Ocean; n.s.
JX865574	MOD_Sep07_D5F_21	NS9 marine group	96	AB355756	western subarctic Pacific; 5 m
JX865575	MOD_Sep07_MD4F_22	NS4 marine group	99	GU061592	South China Sea; 5 m
JX865576	MOD_Sep07_CM60F_23	NS9 marine group	99	HQ242263	North East Pacific; 10 m
JX865577	MOD_Sep07_MD77F_24	NS9 marine group	99	HQ242263	North East Pacific; 10 m
JX865578	MOD_Sep07_D65F_25	NS9 marine group	100	HQ242263	North East Pacific; 10 m
JX865579	MOD_Sep07_MD140F_29	NS9 marine group	98	GQ915983	USA; west coast of Florida, Eastern Gulf of Mexico; n.s.
JX865580	MOD_Sep07_CM100F_33	*Fluviicola*	99	HQ673135	Northeast subarctic Pacific Ocean; 1000 m
JX865581	MOD_Sep07_CM44F_34a*	*Fluviicola*	100	DQ009095	San Pedro Ocean; 5 m
JX865582	MOD_Sep07_D65F_34b*	NS9 marine group	99	EF572198	Coco's Island site 23 (Costa Rica); n.s.
JX865583	MOD_Sep07_CM100F_36	*Fluviicola*	100	FJ202110	host="Montastraea faveolata; n.s.
JX865584	MOD_Sep07_CM44F_37	NS4 marine group	100	GU940762	South China Sea; n.s.
JX865585	MOD_Sep07_CM100F_38	WCHB1-69	98	HM799100	Puerto Rico Trench; 6000 m
JX865586	MOD_Sep07_MD4F_39a*	NS4 marine group	100	HM117334	Elbo Bay in the NW Meditrranean Sea; surface
JX865587	MOD_Sep07_D65F_39b*	*Owenweeksia*	99	DQ656206	East China Sea; surface
JX865588	MOD_Sep07_CM44F_40a*	NS4 marine group	98	HM117334	Elbo Bay in the NW Meditrranean Sea; surface
JX865589	MOD_Sep07_D65F_40b*	*Owenweeksia*	99	GQ348808	Saanich Inlet; 10 m
JX865590	MOD_Sep07_CM5F_41	NS4 marine group	99	HQ242031	North East Pacific; 10 m
JX865591	MOD_Sep07_D65F_43	*Owenweeksia*	99	GQ348808	Saanich Inlet; 10 m
JX865592	MOD_Sep07_MD60F_45	NS9 marine group	100	GQ915956	USA: west coast of Florida, Eastern Gulf of Mexico; n.s.

(n.s) Not specified

Sequencing and relative abundance of marine Bacteroidetes phylotypes

In both Bacteroidetes-specific gels (spatial transects and temporal series), the 55 predominant bands were excised and successfully sequenced (300 to 450 pb). 23 sequences were obtained from bands of the temporal analysis gels and 32 sequences from the spatial ones. Several bands occupying the same position in gels were excised from different lanes and sequenced to confirm the matching analysis. Most of the DGGE signal could be sequenced, since these bands accounted for roughly 80% of the average intensity per lane. All the obtained sequences from both the temporal and the spatial studies were related to the phylum Bacteroidetes, except one identified as phylum Verrucomicrobia (BBMO-59) (Table S3 and S4). In the phylogenetic analyses, 4 sequences belonged to class Sphingobacteria and the remaining 51 to class Flavobacteria. The Sphingobacteria sequences were related to the family Saprospiraceae and they were recovered exclusively in the temporal series study during spring and summer. These were present at the lowermost end of the gels (Fig. S3).

The Flavobacteria sequences grouped into two families: Flavobacteriaceae and Cryomorphaceae. Fourteen of them grouped with clusters that included cultured representatives (*Fluviicola*, *Owenweeksia*, *Polaribacter* and *Tenacibaculum*) (see Fig. 6) while the remaining were assigned to uncultured isolates, mainly to North Sea clusters (as characterized by Alonso *et al.,* [2007]).

In order to explore the contribution of different Bacteroidetes phylotypes throughout the temporal and spatial scales, we measured the relative intensity of the bands in the DGGE analyses. During the 2-year time series, this relative intensity of bands corresponding to specific phylotypes changed gradually over the seasons. A representation of these patterns (limited to the sequenced bands) is shown in Fig. 4. Some phylotypes (e.g. BBMO -4, -5, -15, -25, -29, -32, -37, -49, -54) were mostly present in “WSp” months whereas others were dominant in the “SuF” group (e.g. BBMO -14, -19, -22, -35, -52, -56). These season-specific phylotypes were recurrent over the 2 years analyzed (Fig. 4). A remarkable pattern shift between these communities was noticed between April and May 2007 together with an increase in the intensity of other 3 phylotypes (BBMO -50, -51, -55) only during the transition (Fig. 4). A similar pattern shift was detected between October and November 2007. We expected the same shifts to occur from SuF to WSp in 2006 and WSp to SuF in 2008, but unfortunately these samples did not amplify. Only one phylotype (BBMO -46) was abundant throughout both years.

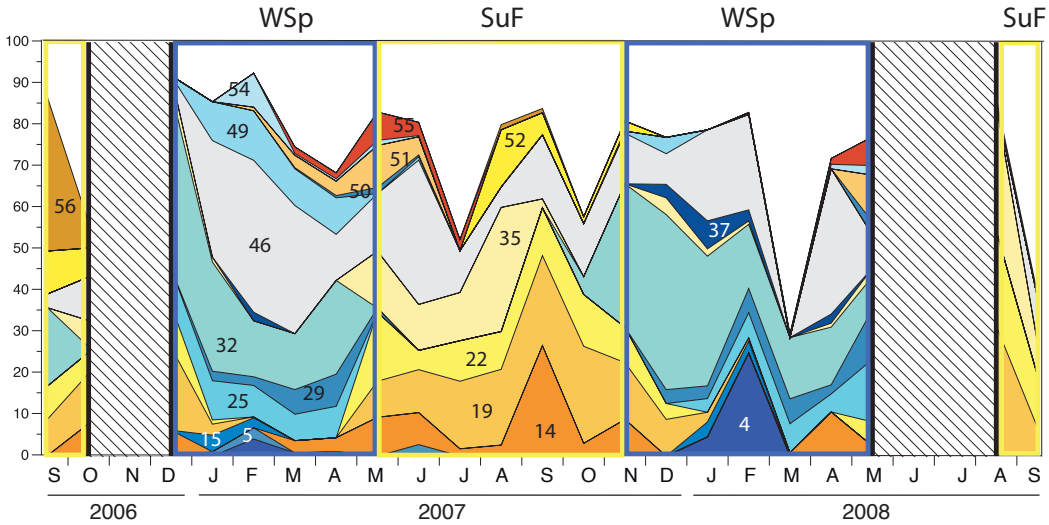


Figure 4. Relative intensity of the sequenced DGGE bands during the 2-year time series. Coloured blocks split the samples in WSp (blue) or SuF (yellow) groups according to the clustering obtained in the DGGE profile analyses. Then, the relative intensity of the bands is represented using yellow (WSp) or blue (SuF) colour palettes. (Nov 06, Jun 08 and Jul 08 were not included in the analyses). The numbers indicate the #ID of the band (Fig. S3 and Table S3).

Along the spatial gradient, a clear shift in the Bacteroidetes communities was observed from surface to depth (Fig. 5). Samples from moderate depths (i.e. 40-70 m.) typically presented transitional patterns in which both "Shallow" and "Deep" bands could be detected (Fig. 5).

According to the sequence data analysis, the Bacteroidetes sequences presented an average pairwise percent similarity of 90.3% (Table S5). After assigning the sequences to taxonomic groups, mean similarity within each phylogenetic cluster ranged from 93.9% to 99.5% whereas similarity between clusters only averaged 88.8% (Table S5). In the phylogenetic analysis (Fig. 6), the main phylogenetic clusters included sequences from mixed seasons and depths: for example, the predominant sequences in groups "WSp" group (BBMO-29) or "SuF" (BBMO-19, -22, & -25) branched within the NS4 cluster (similarity of the NS4 cluster= 97.5%) (Fig. 6). Similarly, "Shallow" samples (MOD-9, -12, -17, -18, -39a & -41) and "Deep" samples (MOD-6, -10, -16 & -19) branched in the same NS4 cluster.

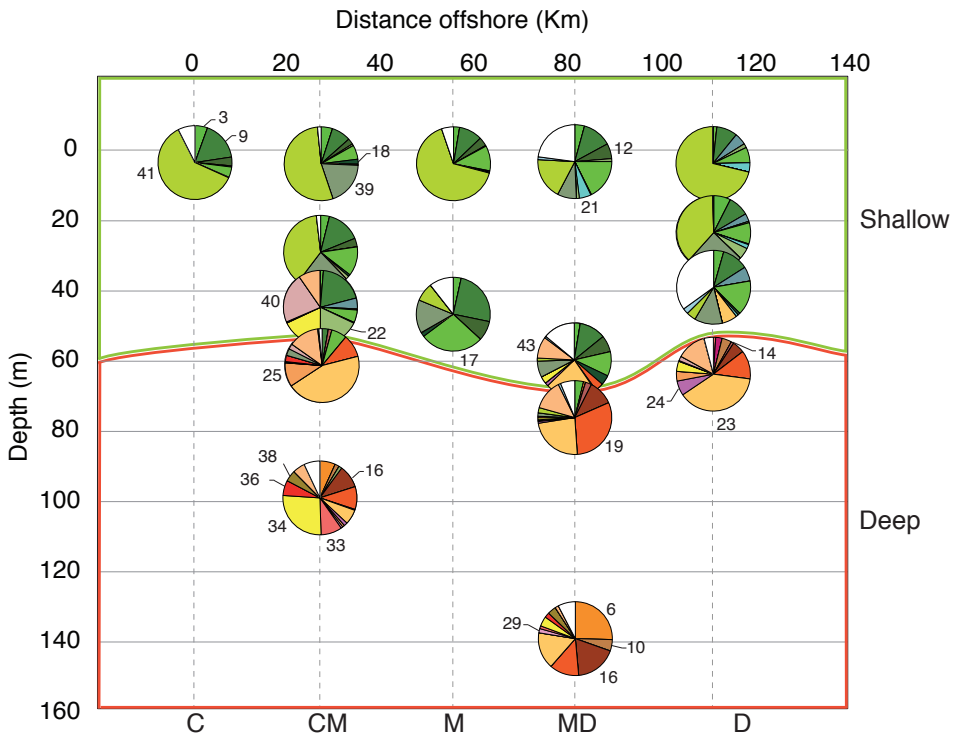


Figure 5. Relative intensity of the DGGE bands sequenced from the horizontal and vertical profiles in the spatial study (cruise Modivus). Colour blocks split the samples in Shallow (green) or Deep (red) groups according to the clustering obtained in the DGGE profile analyses. The relative intensity of the main bands is represented using green (Shallow cluster) or red (Deep cluster) colour palettes. The numbers indicate the #ID of the bands (Fig.S4 and Table S4).

Table S5. Percentage similarity of partial 16S rDNA sequences from the Bacteroidetes clusters identified in the DGGE study in the Northwestern Mediterranean Sea. Average similarities of pairwise comparisons within (in boxes) and between groups were estimated from an alignment of 54 sequences using MEGA4 software. The overall similarity (average pairwise percentage similarity of all the sequences retrieved in the study) was 90.97%.

Overall	Similarity (%)						
Bacteroidetes Clusters	**NS2b**	**NS4**	**NS9**	***Fluviic.***	***Owen.***	***Polar.***	***Sapros.***
NS2b	99.45						
NS4	94.65	97.04					
NS9	89.52	89.64	93.87				
Fluviicola	88.73	88.95	89.65	97.47			
Owenweeksia	90.15	90.24	92.36	90.25	96.40		
Polaribacter	91.71	92.15	88.29	88.29	87.87	97.80	
Saprospira	85.12	86.19	86.81	86.81	85.04	83.97	99.45

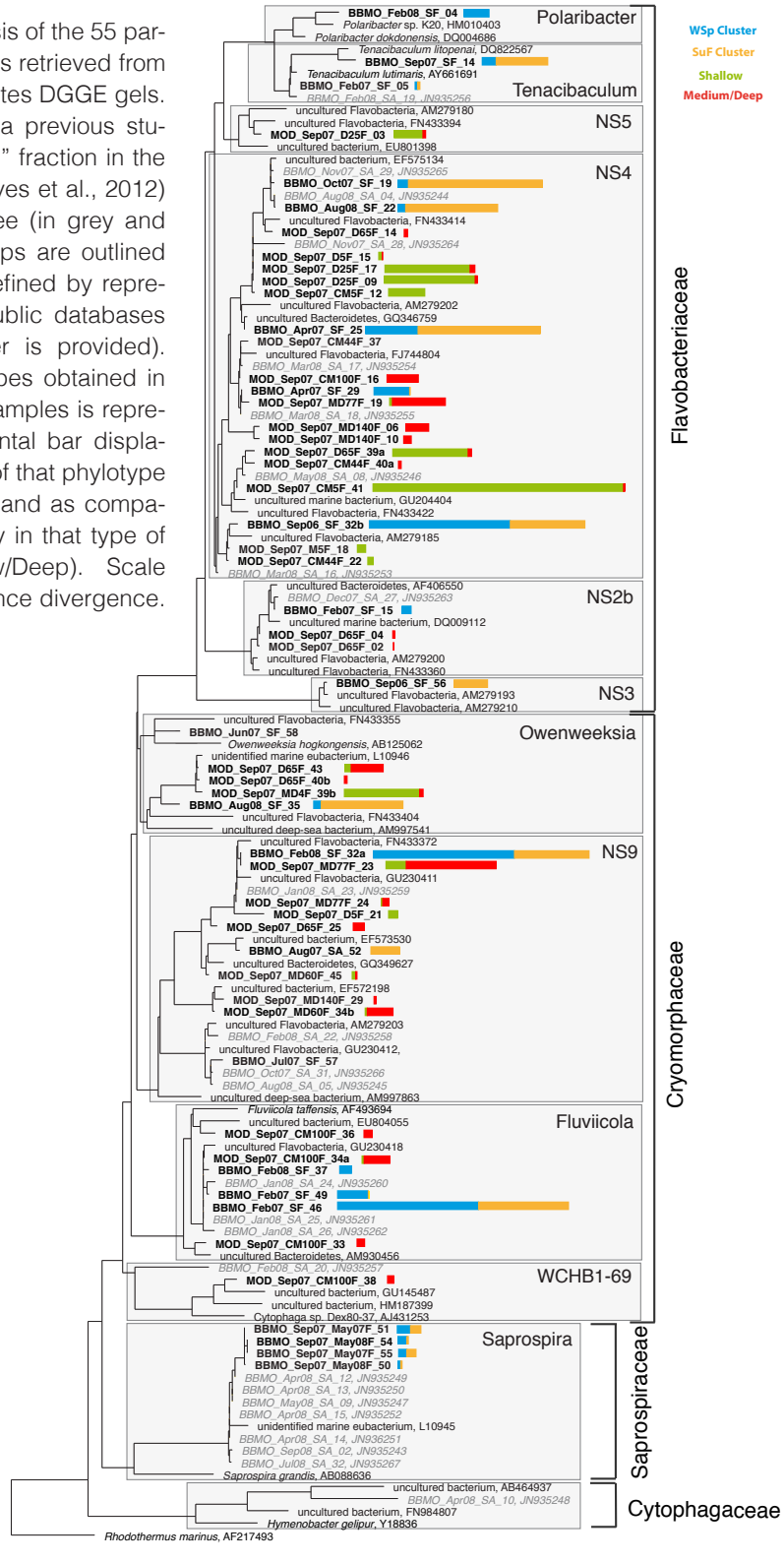


Figure 6. Phylogenetic analysis of the 55 partial 16S rRNA gene sequences retrieved from the group-specific Bacteroidetes DGGE gels. Twenty-five sequences from a previous study of the "particle-associated" fraction in the same sampling area (Díez-Vives et al., 2012) were also included in the tree (in grey and italics type). Taxonomic groups are outlined in shaded blocks and are defined by representative sequences from public databases (Genbank accession number is provided). The presence of the phylotypes obtained in this study in each group of samples is represented by a coloured horizontal bar displaying the relative contribution of that phylotype (i.e. intensity of the DGGE band as compared to the total band intensity in that type of analysis: WSp/SuF, Shallow/Deep). Scale bar: 10% of estimated sequence divergence.

Discussion

We used an oligonucleotide group-specific probe and primers to describe the abundance and distribution of Bacteroidetes, a key bacterial group in marine environments, across a spatial gradient and through a temporal series. The current study was focused on "free-living" communities (fraction 0.22-3 µm). In previous studies we had analyzed the particulate fraction (3-20 µm) of some of the same samples included here. We obtained similar DGGE patterns and phylotypes upon sequencing of the bands to those obtained here with the free-living fraction [Díez-Vives *et al.,* 2012] (Fig. 6). This agrees with studies that have found the same phylotypes in different fractions at least for the most abundant Bacteroidetes [Abell & Bowman, 2005a; Riemann & Winding 2001; Worm *et al.,* 2001] (but see [Fandino *et al.,* 2001; Rink *et al.,* 2007]).

Temporal changes of Bacteroidetes assemblages

We found that the contribution of Bacteroidetes abundance to total bacterioplankton varied between 8 and 22%, a range typically reported in coastal environments. In addition, the percentage contribution values peaked in spring and early summer, similarly to reports from other marine habitats [Pinhassi *et al.,* 2000; Eilers *et al.,* 2001; Mary *et al.,* 2001; Kan *et al.,* 2006; Alderkamps *et al.,* 2003]. However, these changes in the abundance of Bacteroidetes were not synchronized with changes in their community structure. Thus, we observed that Bacteroidetes abundances usually increased from March to April but a similar community structure remained until May (Fig. 1A). A similar pattern was observed after the abundance drop by the end of the summer (from August to September), maintaining the summer community structure until November when it shifted to the winter pattern.

A substantial temporal variability was found among the 31 Bacteroidetes phylotypes characterized here (see Fig. 4). The detailed analysis of the Bacteroidetes community patterns supported the clustering into two main groups (Winter-Spring, "WSp" and Summer-Fall, "SuF") following a seasonal scheme. Interestingly, the fingerprinting of total bacterioplankton (with universal bacterial primers) also resulted in a similar clustering (Fig. S5B), suggesting that the Bacteroidetes seasonal changes paralleled those of the whole bacterial assemblage. Each group of samples (WSp and SuF) was characterized by markedly different communities (i.e. specific phylotypes in each group) that were recurrent for two consecutive years (Fig. 4). However, these distinct

phylotypes were not distantly phylogenetically grouped as they grouped together in the same phylogenetic clusters (Fig. 6), at least based on our partial sequences and the limited sequence data obtained in the current study (e.g. NS4 cluster included phylotypes from "WSp" and "SuF" samples). Additionally, it has been shown that organisms with very similar 16S rDNA can be very different in their physiological traits [Béjà *et al.,* 2002; Jaspers *et al.,* 2004]. The substitution of related, yet ecologically distinct populations over time has been reported previously in bacterioplankton analyses [Casamayor *et al.,* 2002; Schauer *et al.,* 2003]. The present study, focused on the most abundant Bacteroidetes, supports the concept that certain microbial populations become predominant over short periods of time but they can be substituted rapidly by related genotypes probably better adapted to the new changing conditions, thus keeping the taxonomic structure at a genus/family/group level rather stable.

Spatial changes of Bacteroidetes assemblages

Previous studies of microbial distribution along space commonly reported small changes in bacterioplankton community structure over short superficial transects [Riemann *et al.,* 1999; Riemann & Middelboe, 2002; Bano & Hollibaugh 2002; Ghiglione *et al.,* 2005]. However, some studies have suggested that Bacteroidetes diversity at the coast is higher than at the open sea [Eilers *et al.,* 2001; Kirchman *et al.,* 2003; Alonso *et al.,* 2007]. In all surface water samples studied here (above 30 m depth), spanning a >100-km coast-offshore transect, the contribution of Bacteroidetes to total bacterial abundance was uniform and community structure was similar, clustering together in the dendrogram (Fig. 3). An exception was noticed at station MD, in which lower abundance but higher diversity was detected coinciding with an upwelling event bringing up cold water from the deep in the slope region. A similar situation was described by Suzuki *et al.* [2001b] at Monterey Bay, where lower percentages of Bacteroidetes were observed in the core of an upwelling plume.

Drastic changes in bacterioplankton abundance and community composition do usually occur with depth, even across small distances [Lee & Fuhrman 1991; Acinas *et al.,* 1999; Moeseneder *et al.,* 2001; Ghiglione *et al.,* 2005, Pommier *et al.,* 2010]. Bacteroidetes are usually more abundant in the surface and DCM layers [Wells and Deming 2003; Schattenhofer *et al.,* 2009; Gómez-Pereira *et al.,* 2010] and the values decrease with depth, usually at a much steeper rate than the bulk bacterial community (Table S1). Our results agree with these previously reported patterns, showing average

surface contributions of around 15%, which decreased to 2.5% at 150 m. However, Bacteroidetes richness increased with depth even though their abundance decreased (Fig. 1B). Additionally, it was previously described that global bacterioplankton richness increased with depth in the same vertical profiles, suggesting that deeper, likely more stable, ecosystems present a higher bacterial diversity [Pommier *et al.,* 2010].

The analysis of the Bacteroidetes community patterns in the spatial transects supported the clustering of two main groups ("Shallow" and "Deep") (Fig. 3) with different phylotypes dominating each one. Again, these phylotypes were not phylogenetically different (Fig. 6). Some data from Bacteroidetes-specific DGGEs and gene libraries hypothesized a putative depth-specific distribution of the group [Blümel *et al.,* 2007]. We think that the current evidence is not enough to clearly resolve the existence of stratified, depth-specific Bacteroidetes clades. More detailed studies implying high sequencing coverage will be necessary to solve this issue.

Influence of environmental factors

Day length and total Bacteria abundance explained a very large variability of Bacteroidetes contribution to community structure in the temporal series. Day length has been considered an important variable affecting bacterial diversity in coastal temperate waters [Gilbert *et al.,* 2012]. In the particular case of Bacteroidetes, this correlation with day length could be attributed to the higher resistance of these organisms to solar radiation compared to other bacterioplankton groups [Alonso-Sáez *et al.,* 2006; Ruiz-González *et al.,* 2012b], or to the presence of proteorhodopsines in a relatively large fraction of marine Bacteroidetes for which genomic data are available [Fernández-Gómez *et al.,* 2013]. Alternatively, the observed correlation could be attributed to additional, unidentified factors covarying with daylength.

In the spatial transect, salinity was the best predictor of the Bacteroidetes contribution to community structure. Bacteroidetes abundance has been negatively correlated with salinity in other studies, since they became less abundant in deep waters [Wells & Deming, 2003]. Moreover, they are relatively scarce in recently upwelled waters but their abundance increases as the water ages and phytoplankton blooms and decays [Suzuki *et al.,* 2001b].

Although other authors have found associations between Bacteroidetes abundance and algal blooms [Cottrell & Kirchman, 2000b; Kirchman, 2002; Abell

& Bowman, 2005b], we did not find a relationship between Chl *a* concentration and Bacteroidetes percentages. The response of Bacteroidetes to the blooms can occur coinciding with the Chl *a* peak or during the detrital phase of the bloom [Alderkamp *et al.,* 2006]. This may depend on the composition of the blooms [Fandino *et al.,* 2001] but it has also been shown that different Bacteroidetes genera may display different responses [Teeling *et al.,* 2012]. Our samples showed maximal abundances one or two months after the Chl *a* peaks, which could indicate a delayed response to the bloom. However, shifts in abundance and composition during blooms may also occur in the order of a few days [Fandino *et al.,* 2005]. In that sense, our monthly samplings would certainly have overlooked this direct effect and short-term fluctuations.

Interestingly, we observed a positive correlation with temperature in the spatial transects ($r = 0.560$, $p < 0.05$, $n = 17$), even when the analysis included only surface (above 30 m) water samples ($r = 0.881$, $p < 0.01$, $n = 8$). This contrasts with previous reports in which Bacteroidetes abundance correlated negatively with surface temperature [Gómez-Pereira *et al.,* 2010; Abell & Bowman 2005b]. Focusing on the taxonomic data obtained in this study, we identified genotypes from genera *Fluviicola* and *Owenweeksia*, which have mesophilic type strains [O'Sullivan *et al.,* 2004; Lau *et al.,* 2005]. The later have been found to be abundantly represented in a pyrosequencing approach with the same samples used in this study [Crespo *et al.,* 2013], and are considered typical Mediterranean species. Moreover, a *Polaribacter* genotype related to an isolate from North Western Mediterranean Sea (*Polaribacter* sp. strain MED152) [González *et al.,* 2008] was also retrieved. Although *Polaribacter* are usually psychrophilic, several isolates have been found to grow optimally at 25-28 ℃ [Yoon *et al.,* 2006] and sequences from the Mediterranean might belong to a subgroup of warm-water phylotypes, as it was suggested for North Atlantic waters [Gómez-Pereira *et al.,* 2010]. The presence of these mesophilic phylotypes may partly explain the positive correlation with temperature. Bacteroidetes is one of the most diverse groups in phylogenetic studies of natural environments [Kirchman *et al.,* 2005]. We found an overall similarity of 90.9 % among all our Flavobacteria and Sphingobacteria sequences, indicating highly divergent phylotypes. Some values lower than this have also been reported, e.g. Bacteroidetes sequences were 84.9 % similar whereas Alpha-proteobacteria ones were 99.2 % similarity, on average, in the Delaware River [Kirchman *et al.,* 2005].

Conclusions

The study of Bacteroidetes dynamics at temporal scale during two consecutive years displayed specific phylotypes, detected for Winter-Spring and Summer-Fall, following a recurrent seasonality. Similarly, the Bacteroidetes community structure changed with depth, but interestingly, Bacteroidetes exhibited its highest diversity in deepest waters where the abundance was lowest. Furthermore, we pointed out different physical and biological factors responsible for the changes in the Bacteroidetes contributions to total abundance of bacterioplankton. These factors were markedly different over a temporal series than across a surface-deep gradient. This can be simply due to different environmental conditions varying differently across each gradient- e.g. day length does not affect communities in a surface-depth gradients. As currently defined [Ludwig *et al.,* 2011], the phylum Bacteroidetes seems highly diverse genetically, and narrower phylogenetic levels within this group would be necessary to identify meaningful units (subgroups, clades or ecotypes) for conducting further detailed ecological studies.

Acknowledgments

We thank Prof. C. Pedrós-Alió and Dr. O. Palenzuela for valuable suggestions and critical reading of the manuscript. B. Díez, V. Balagué, C. Cardelús and I. Forn for DGGE training, and sampling and DNA preparation of the BBMO time series, M. Coll for sampling during the the Modivus cruise, and T. Lefort for the flow cytometry data. CDV was a recipient of a JAE-predoc fellowship from CSIC. SGA was supported by a Ramon y Cajal contract from the Spanish Ministry of Science and Innovation and by MicroB3 (FP7-OCEAN-2011). This research was supported by grants MICRODIVERSITY (CGL2008-00762/BOS), PANGENOMICS (CGL2011-26848/BOS), GEMMA (CTM2007-63753-C02-01/MAR), SUMMER (CTM2008-03309/MAR) and STORM (CTM2009-09352/MAR) from the Spanish Ministry of Science and Innovation (MICINN).

Chapter 4

Disentangling microdiversity and structure within marine Bacteroidetes through study of the 16S rRNA and the Internal Transcribed Spacer region

Abstract

The sequence of the 16S rRNA has been used as both, an evolutionary and an identification marker in the study of bacterioplankton. However, it is well known that some organisms with very similar, or identical, 16SrRNA sequences might have different ecological strategies. The level of resolution of bacterial identities can be improved by using the full rRNA operon that includes the internal transcriber space (ITS), tRNAs and the 23S rRNA. We tried to improve our knowledge of marine Bacteroidetes by performing extensive clone libraries including a large fragment (ca. 2,000 bp) of the ribosomal operon. We chose contrasting environments (i.e. winter *vs.* summer and surface *vs.* deep communities) to cover as much diversity of Bacteroidetes communities as possible. A total of 646 high quality sequences were obtained from the four libraries. Since the sequence variability of the ITS fragment was much higher than that of the 16S rRNA, both regions were split and analyzed separately. Specific analysis evaluating the richness, diversity, phylogenetic composition, and other descriptors of the communities were compared among the different libraries (i.e., environments). Fine-scale phylogenetic trees of the 16S rRNA revealed some closely related habitat-specific clusters, mainly associated with deep waters. The topology of 16S rRNA-inferred phylogenies was compared with ITS-based trees. The ITS region was highly variable showing large differences in size and variable sequences for near identical 16S rRNA sequences (>99% similar). This enhanced resolution was evidenced in the trees, which displayed new subclusters associated to seasons. The ITS region was an informative region allowing differentiation of genotypes showing identical 16S rRNA, extremely helpful to infer relationships among closely related taxa. In this work, the identification of different habitat-specific Bacteroidetes clades is reported for the first time using extensive and high quality sequences. The quantitative study of these clades would test whether they can be treated as ecologically distinct populations.

Disentangling Microdiversity and Structure within Marine Bacteroidetes through study of the 16S rRNA and the Internal Transcribed Spacer genes. **Cristina Díez-Vives, Pablo Sánchez, Josep M. Gasol and Silvia G. Acinas.** Manuscript.

Introduction

The phylum Bacteroidetes is known to be one of the most abundant bacterial groups in aquatic ecosystems (up to 40% of total bacterioplankton of total counts), having a pivotal role in the processing of organic material [Kirchman, 2002]. However, relatively little is know about the diversity and structure of uncultured members of this phylum, mainly because they have generally been targeted together with other bacterial groups by doing environmental clone libraries using universal primers [Cottrell *et al.,* 2000; O´Sullivan *et al.,* 2004; Alonso *et al.,* 2007; Gómez-Pereira *et al.,* 2012]. Some studies reflect that Bacteroidetes are usually underrepresented in clone libraries of the 16S rRNA gene [Kirchman, 2002; Kirchman *et al.,* 2003; Alonso-Sáez *et al.,* 2007]. Only a few studies have used Bacteroidetes-specific primers in DGGE-based analyses [Abell *et al.*, 2004; Blümel *et al.,* 2007; Xihan *et al.,* 2008; Díez-Vives *et al.,* in press] or in the generation of more specific gene libraries [Kirchman *et al.,* 2003], attempting to obtain insight into their diversity in certain habitats. However, these studies analyzed a rather limited number of clones (i.e. 107 Bacteroidetes-like clones in the latter study).

The existence of ecotypes (niche-specific phylogenetic clades) in uncultured microbes is nowadays a central theme in microbial ecology. For instance, 16S rRNA clades with depth-specific distributions have been reported in several bacterial phyla, such as the *Prochlorococcus* High-Light (HL) and Low-Light (LLA) adapted ecotypes [Moore *et al.,* 1998; Rocap, *et al.,* 2002], green non-sulfur bacteria [Giovannoni *et al.,* 1996], the ubiquitous SAR11 cluster [Field *et al.,* 1997], or the Deltaproteobacteria [Wright *et al.,* 1997]. SAR11 has also showed distinct annual distribution patterns for three of its ecotypes [Carlson *et al.,* 2009], which were correlated with seasonal mixing and stratification of the water column. Similarly, new ecotypes of SAR11 identified recently by NGS technologies have shown clear spatiotemporal distribution patterns [Vergin *et al.,* 2013]. This notwithstanding, the Phylum Bacteroidetes has scarcely been the focus of fine-scale phylogenetic studies of the 16S rRNA gene dealing with habitat-specific clusters with a few exceptions [Blümel *et al.,* 2005]. Studies of marine Bacteroidetes are usually concentrated on high phylogenetic levels such as class Flavobacteria, which according to current classifications includes at least 100 highly divergent genera (Silva database release 115, August 2013; Quast *et al.,* 2013]. Such broad scale studies may overlook the structure of different Bacteroidetes populations and hypothetical associations of particular clades with environmental factors. Comprehensive work on identifying ecotypes within Bacteroidetes from aquatic systems is still lacking and therefore, a high

phylogenetic resolution analyses is required for the identification of habitat-specific clades from sequence-based trees of the 16S rRNA gene.

Such phylogenetic analyses can greatly benefit from the study of the internal transcribed spacer (ITS) of the ribosomal operon. Although the ITS plays a role in rRNA processing and may encode for up to 2 tRNA genes, most of its variable regions are exempt of the requirement for structural integrity that constrain mutations in the ribosomal operon molecules. Consequently, the ITS displays highly variable length and sequence heterogeneity, and therefore is adequate for such fine-scale phylogenetic studies [Schloter *et al.,* 2000].

ITS variation has been extensively used as a phylogenetic marker in clinical identification of bacterial strains [García-Martínez *et al.,* 1996, Graham *et al.,* 1997, Liguori *et al.,* 2011]. Among environmental bacteria, ITS-based phylogenies partitioned previously observed divisions of cultured isolates *Prochlorococcus* and *Synechococcus* into several new genetically distinct clades not resolved by the analysis of the 16S rRNA gene. These new ecotypes were associated with particular depths, phenotypes and geography [Rocap *et al.,* 2002, Ferris *et al.,* 2003]. In the case of mainly uncultured environmental groups, the ITS region gave a better and refined picture of SAR11 distribution in the water column, allowing to identify depth-specific clusters [García-Martínez & Rodríguez-Valera 2000; Brown *et al.,* 2005; Ngugi *et al.,* 2012]. It remained unclear whether the ITS would provide a deeper phylogenetic signal for Bacteroidetes populations from different environmental conditions.

In order to characterize the unknown diversity, structure and distribution of Bacteroidetes, we constructed 4 large Bacteroidetes-specific 16S–23S rRNA environmental clone libraries using samples stratified according to temporal (seasonal) and spatial (depth) variables in the NW Mediterranean Sea. We conducted phylogenetic analyses of the 16S and ITS sequences, demarcating phylogenetically specific clusters and their association with contrasting environmental variables, which could represent distinct ecotypes within the phylum Bacteroidetes. The ultimate aim is to achieve the identification of putative environmental Bacteroidetes ecotypes associated to ecological niches in time or space.

Material and Methods

Sample Collection and DNA extraction

Bacterioplankton was collected from seawater samples taken in two different campaigns. Firstly, a 2-year seasonal study was conducted in the Northwestern Mediterranean Sea, at the Blanes Bay Microbial Observatory (BBMO, www.icm.csic.es/bio/projects/icmicrobis/bbmo) (41°40N´, 2°48E), from September 2006 to September 2008. This is a coastal oligotrophic site, just 1 km offshore the city of Blanes. Surface (0.5 m depth) waters were sampled monthly, filtered through a 200 µm mesh net and transported to the lab under dim light (within 1.5 h) in 25 l polycarbonate carboys. Secondly, two additional samples were taken in the same area, one in a coastal-surface (5 meters depth) station and the other in a deeper station (2000 meters) at 111 km offshore from the coastal station, both sampled from on board BO 'García del Cid' during the "Modivus" cruise (September 20th to the 23rd, 2007). Water was prefiltered through a 200 µm-mesh and processed immediately on board.

To collect bacterioplankton biomass, 8-10 l of seawater were filtered using a peristaltic pump sequentially through GF/A or polycarbonate 3 µm prefilters (Whatman) and 0.22 µm Sterivex filters (Durapore, Millipore). The filters were saturated with 1.8 ml lysis buffer (50 mM Tris- HCl, 40 mM EDTA, 0.75 M sucrose, pH 8.0) or flash-frozen in liquid $N_{2,}$ and kept at -80 °C until used. Microbial biomass was digested with lysozyme, proteinase K and sodium dodecyl sulfate, and the nucleic acids were extracted with phenol and concentrated in a Centricon-100 (Millipore), as described in Massana *et al.* [1997b].

Generation of clone libraries and sequencing

Samples were selected to construct four different clone libraries: two libraries at the spatial scale with one representative of the coastal surface station and another in off-shore at 2000 m-depth sampled in September 2007. These two clone libraries are refereed as "Surface" and "Deep" libraries (or SFL and DPL respectively). Additionally, two libraries were done at temporal scale, representing the opposite seasonal conditions in the BBMO area. For the later, the selection of the samples was based on our previous results of a seasonal study based on DGGE and CARD-FISH, which showed clearly different Bacteroidetes communities during the year (Winter-Spring *vs.* Summer-Fall) [Diez-Vives *et al.,* in press, chapter 3]. Thus, the two temporal libraries were generated

with pooled samples representing the Winter-Spring community ("Winter" library, WTL), and the Summer-Fall community ("Summer" library, SML). Eight samples were pooled for the WTL and twelve samples for the SML from the two consecutive years (Fig. 1).

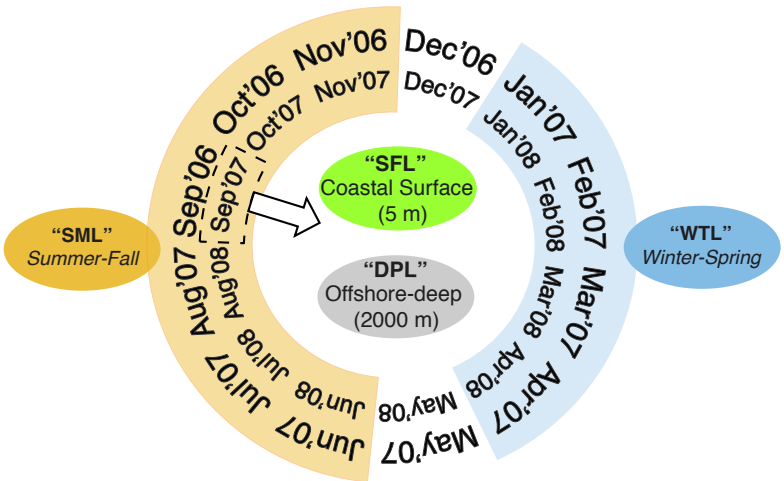


Figure 1. Design of the gene libraries. Orange: "Summer" samples selected from monthly samplings taken during two consecutive years at a surface coastal point (Blanes Bay Microbial Observatory) and used to build SML. Blue: "Winter" samples from the same sampling chosen to build WTL. Samples taken in December and March were not included in any seasonal library because they represented transitional months harboring a mixture of summer and winter communities in previous studies (see chapter 2). Samples taken in September 2007 during the spatial study (Modivus cruise) were chosen to build SFL and DPL.

The DNA samples were amplified for subsequent cloning using two Bacteroidetes-specific primers: A forward primer close to the 5' end of the 16S rRNA (CF319aF) [Manz *et al.,* 1996] and a reverse primer in the 23S ribosomal subunit (CF434R) (5' CACTATCGGTCTCTCAGG 3') that was specifically designed for making the Bacteroidetes clone libraries [Acinas *et al.,* unpublished]. This reverse primer was designed in ARB using the "Probe Design" tool and a specific Bacteroidetes dataset retrieved from Silva SSU Ref 108 dataset, that contained 618,442 sequences [Pruesse *et al.,* 2007]. An updated revision of the primer pair coverage was done using the "TestPrime" online tool in Silva website (http://www.arb-silva.de/search/testprobe/) using the latest Silva release SSU Ref NR 115 (August 2013) with 1,426,414 sequences [Quast *et al.,* 2013]. The amplification product of these primers included almost the complete 16S rRNA, the internal transcribed spacer, and 434 bases of the 23S rDNA. PCR reactions (50 µL volume) contained 200 µM

each deoxynucleoside triphosphates, 0.2 µM each primer, 2 mM $MgCl_2$, 1x PCR buffer, and 1 U Taq DNA polymerase (Invitrogen). PCR cycles consisted of an initial denaturation step (94 °C, 5 min); 27 cycles of DNA denaturation (94 °C, 30 s), primer annealing (55 °C, 30 s), and Taq extension (72 °C, 1 min); and a final extension step (72 °C, 10 min). A reconditioning PCR was made for each sample as described in Acinas *et al.* [2005]. Briefly, PCR products were analyzed and quantified by agarose gel electrophoresis using size standards (Low DNA Mass Ladder, GIBCO BRL) and purified using micro spin columns (Qiagen). For each template, triplicate reactions were pooled and concentrated using a vacuum pump. Twenty microliters of the purified and concentrated pool were used for a new PCR with fresh components, with only 3 cycles of amplification in a final reaction volume of 100 µl.

The clone libraries were constructed using the Topo TA cloning system (Invitrogen) following the manufacturer's instructions. Clones containing inserts of the expected size were selected from each library and Sanger-sequenced with three primers: CF319aF [Manz *et al.,* 1996], 968F [Chen *et al.,* 2006] and 1492F [Lane, 1991], and the sequences were assembled using software Geneious Pro (http://genious.com). Chimeric sequences were evaluated using software Mallard [Ashelford *et al.,* 2006], and removed before further downstream analyses. The small subunit ribosomal gene, ITS region and large subunit ribosomal gene were identified based on alignments with reference Flavobacteria sequences. These fragments were split and treated as separate datasets.

Clustering of OTUs and 16S rRNA-based phylogenetic analysis

The small subunit dataset was aligned in the SINA web alignment tool (http://www.ar-silva.de/aligner/) [Pruesse *et al.,* 2012] and imported into the ARB software package (http://www.arb-silva.de) [Ludwig *et al.,* 2004]. The alignment was curated by eye following secondary structure criteria. The resulting dataset was used as input for software Mothur [Schloss *et al.,* 2009] to generate a pairwise uncorrected p-distances matrix, which was used to cluster sequences in OTUs (Operational Taxonomic Units) at different sequence identity levels, choosing the options "furthest neighbor" and "precision=100".

A phylogeny for 16S rRNA sequences was calculated using MEGA 5 [Tamura *et al.,* 2007], by maximum-likelihood inference using partial deletion (cutoff of 20%) with

a General Time Reversible model [Tavaré, 1986] with invariant sites and a Gamma-distributed site rate variation. The bootstrap support was calculated using 100 replicate resamplings of the dataset. The tree was rooted using a close outgroup sequence (Rhodothermus marinus, accession number: AF217493). The resulting consensus tree was edited with the iTol software [Letunic & Bork 2011].

The sequences were used as queries against the GenBank database, National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) using BLAST [Altschul *et al.,* 1990] to determine their closest hits and putative classification. Sequences were classified further with the SINA classifier [Pruesse *et al.,* 2012], using both Silva and RDP (Ribosomal Database Project) taxonomies. Complete rrn operons of Bacteroidetes were extracted from genomes deposited in public databases such as rrnDB (http://rrndb.mmg.msu.edu) and GenBank (http://www.ncbi.nlm.nih.gov/), and used as references for intragenomic variability studies.

Clone library analyses

Quantitative analyses. Cluster similarity curves (sequences represented in decreasing % of similarity) and rarefaction curves (number of OTUs resulting as a function of sample size) of each clone library were analyzed and coverage values were calculated to estimate how efficiently the libraries described the complexity of a theoretical community of infinite size, i.e., the original community. The coverage of each library is given as C = 1 - (n1/N), where n1 is the number of clones which occurred only once in the library and N is the total number of clones examined. We calculated this value at different sequence similarity cutoffs (99% and 97%), for each library separately and for the 4 libraries combined. We also calculated the Chao estimator of the species richness [Chao *et al.,* 2005] and the Shannon diversity index [Hill, 1973], which accounts not only for species richness but also for the evenness of the community. To determine differences among libraries we used the program S-Libshuff [Schloss, 2004], which calculates the integral of the Cramér-von Mises statistic. We compared dissimilarity of communities from different environments by calculating the Jaccard index (qualitative β diversity), using the Mothur package [Schloss *et al.,* 2009].

Sequence-based analyses (Community similarity). We tested for environment-specific genetic bacterial diversity by employing an analysis of molecular variance (AMOVA) [Excoffier at a., 1992; Anderson *et al.,* 2001; Martin *et al.,* 2002]. This method is widely used in population genetics, but it can be applied to community ecology as well [Schloss *et al.,* 2008]. It is a nonparametric analogue of ANOVA, based on a distance matrix of sequences retrieved from different communities. The test was performed with 1,000 randomizations using the Mothur package.

Phylogenetic-based analyses (Bacterial phylogenetic structure). To estimate the phylogenetic diversity across libraries we used the Faith's phylogenetic diversity index [Faith, 1992] and phylogenetic species variability (PSV) [Helmus *et al.,* 2007] using the package Picante [Kembel *et al.,* 2010] as implemented in R [R development Core Team, 2013]. Faith's PD calculates the diversity in a community by summing the lengths of all those branches that are members of the corresponding minimum spanning path connecting species of the same community. PSV reflects how a hypothetical unselected or neutral trait shared by all species in a community is affected by phylogenetic relatedness, being closer to 1 when all taxa are poorly related (i.e., star phylogeny) and closer to 0 when taxa are closely related. To correct for unequal sample sizes, 1,000 randomized subsamples were run for each library and the phylogenetic signal of community composition was tested against the null models taking into account either the community or the species pool level. We also calculated PD rarefaction curves to correct for sampling effort differences using the R function phylocurve.perm with 100 randomizations [Nipperess & Matsen, 2013].

The β diversity among samples can be tested incorporating the degree of divergence between sequences, using phylogenetic β diversity measures such as the Unifrac distance U [Lozupone & Knight, 2005]. This measure allows to test whether or not two or more communities have the same phylogenetic structure. UniFrac measures the difference between two collections of sequences (e.g. environments) as the amount of evolutionary history that is unique to either of the two. We performed the UniFrac analyses with 1,000 randomizations, using Mothur. Two different datasets were analyzed: first including all the 511 unique sequences and secondly using only one representative sequence for each >99% similarity OTU (115 OTUs). We used both the unweighted and weighted versions of UniFrac as qualitative and quantitative approaches for both dataset

(511 seq. and 115 seq.). Qualitative measures are informative when communities differ primarily by the different taxa present and quantitative measures (e.g. number of times that each sequence was observed) reveal community differences that are due to changes in relative taxon abundances [Lozupone *et al.,* 2007].

Principal component analysis and LTT

In order to calculate the evolution of the lineages through time (LTT), inferred phylogenetic trees were transformed to ultrametric with the package ape [Paradis *et al.,* 2004] in R using Grafen's computation of branch lengths, scaling the height of each node such as the root height is 1 and branch lengths were computed as the difference between height of lower node and height of upper node. Then log-transformed LTT plots were generated standardizing the number of lineages of each library to 100. Departure from the pure-birth model was tested with the g statistic [Pybus & Harvey, 2000], a value that describes the distribution of the nodes in a phylogeny with respect to time, such as positive values of g indicate late cladogenesis, or an acceleration of the diversification rate, and vice-versa.

Internal transcribed spacer

The presence of tRNA genes within the ITS fragments was predicted using the ARAGORN program [Laslett & Canback, 2004]. ITS length heterogeneity and its sequence divergence hampered quality alignments of all the sequences. Only subgroups of related sequences were aligned in MAFFT using the E-INS-i option [Katoh & Standley, 2013], which is recommended for sequences with multiple conserved domains and long gaps. These alignments were processed in Mothur to generate p-distances matrixes as above.

Due to the heterogeneity of the ITS, phylogenetic trees based on the ITS region were built only for some of the largest, closely related clusters. The trees were inferred in MEGA 5, using all the positions of the alignments. The best evolution model for each sequence dataset was calculated by the software. Bootstrap values were calculated for 100 resamplings of the data. Comparisons of 16S and ITS-based phylogenies were used as additional criteria to identify chimeric sequences. Trees were edited with the iTol software [Letunic & Bork, 2011].

Results

Diversity within Bacteroidetes environmental clone libraries

Four libraries were built from samples taken at temporal and spatial scales (DPL, SFL, WTL and SML), using specific primers for the ribosomal RNA operon. An updated evaluation of the Bacteroidetes-specific primer set used is shown in Table 1. The libraries were sequenced with a similar effort (around 180 sequences each) using three overlapping primers to get almost the complete fragment comprising the 16S-ITS-23S regions. A total of 701 sequences were obtained, of which 55 were flagged as putative chimeras, leaving a final dataset of 646 sequences. Chimeric sequences were found particularly abundant in the WTL (34 out of 55) and, as a consequence, this curated library contained a lower number of sequences (Table 2). Since we sequenced roughly 1,875 bp per clone, we obtained over $1.2 \cdot 10^6$ bp of Bacteroidetes ribosomal operon data.

CF319aF (in the 16s rRNA gene)	**Eligible**	**Coverage**	**Specificity**
Phylum Bacteroidetes	79437	35,2	99,8
Class Bacteroidia	56154	24,5	97,7
Class Cytophagia	4514	7,6	96
Class Sphingobacteria	6402	37,7	96,3
Class Flavobacteria	10452	93,2	97,3
> Family Blattabacteriaceae	338	55	96
> Family Cryomorphaceae	1165	96,1	96,2
> Family Flavobacteriaceae	8071	94	97
CF434R (in the 23S rRNA gene)	**Eligible**	**Coverage**	**Specificity**
Phylum Bacteroidetes	351	33,3	100
Class Bacteroidia	173	2,3	99,4
Class Cytophagia	25	36	99,4
Class Sphingobacteria	124	70,2	99,8
Class Flavobacteria	28	60,7	99,5
> Family Blattabacteriaceae	9	0	0
> Family Cryomorphaceae	7	0	0
> Family Flavobacteriaceae	101	85,1	99,8

Table 1. In silico coverage of primers performed with the Silva's "TestPrime" online tool. Total number of eligible sequences (number of sequences with data available at the position of the primer), coverage of the primers (matches/elegible sequences*100) and specificity (100-(outgroup matches/elegible sequences)*100) are provided for the whole phylum and segmented at the class level. Within Class Flavobacteria, detailed information at the family level is also given.

Three levels of operational taxonomic units (OTU) definition were used according to 16S rRNA sequence similarity: 100% as unique sequences, i.e., removing redundancies (which resulted in 511 OTUs); >99% as "species" delimitation (resulted in 115 OTUs); and >97% as a rough "classical species definition" (which resulted in 73 OTUs). Cluster

similarity curves showed a drastic increase in the number of distinct sequences over the 99% similarity threshold. The SFL presented the largest number of sequences (82%) sharing >99% similarity, whilst the remaining libraries showed a percentage of between 62 to 70% of sequences with >99% similarity (Fig. 2A-B).

Rarefaction curves were calculated for >97, 99 and 100% similarity cut-offs. All libraries approached asymptotes at the 99% and 97% OTU definition (Fig. 2C-D). As expected, SFL reached saturation with relatively few sequenced clones. Actually the surface library was the best represented with a coverage of 94% at 99% similarity, while the other libraries reached similar coverage values only at lower similarity level (97%), and specifically the WTL was the less represented (80% at 99% similarity) (Table 2). Overall, the high coverage of the libraries (from 80 to 94% for a >99% similarity level cutoff) indicates that the data presented here provide a relatively reliable inventory of the bacterial rRNA gene sequences present in BBMO.

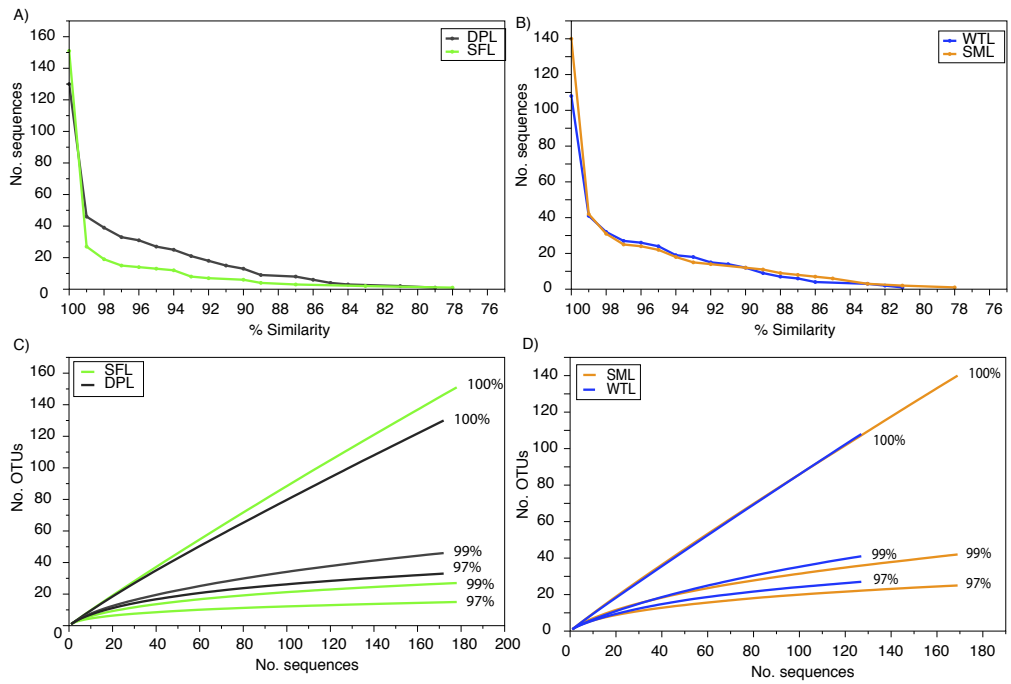


Figure 2. A-B) Cluster similarity curves from the four clone libraries of marine Bacteroidetes predicted by Mothur. The graphs show the number of different sequences present at decreasing percentages of 16S rRNA sequence similarity, for the spatial (A) and the temporal (B) libraries. **C-D)** Rarefaction curves of the four clone libraries at three different similarity cut-offs (100, 99 and 97%). At the plateau, the library is considered to be 'large enough' to provide an unbiased estimate of OTU richness (Kemp & Aller 2004).

The Chao-1 estimator of richness indicated that, at >99% similarity, SML contained the largest number of predicted OTUs (78.1), followed by WTL (71). SFL presented the smallest number of predicted OTUs (40.7), and DPL showed an intermediate richness value (65.5) (Table 2). However, when sequences were clustered at >97% similarity, DPL showed the highest richness (51.2), and the other libraries dropped to values around 30-38. The Shannon diversity (at 99% similarity) index was generally high (from 2.38 to 3.11), with the maximum value corresponding to the DPL (Table 2). Diversity in all the samples considered as a whole increased only slightly (3.54) because, despite increasing the number of different OTUs (115 at >99% similarity), the abundance distribution of these OTUs was not even, causing almost no increase in the global diversity estimate. The value of a diversity index is maximal when all types are equally abundant.

Table 2. Clone library descriptors at different similarity cut-offs, and estimates of richness and diversity.

Clone Library	No. of seq (Unique seq.)	Average seq. similarity	No. of OTUs (99%; 97%)	Coverage (99%; 97%)	Chao 1 (99%; 97%)	Shannon (99%; 97%)
WTL	127 (108)	91.4%	41; 27	0.80; 0.91	71; 38	2.65; 2.26
SML	169 (140)	90.2%	42; 25	0.86; 0.94	78.14; 32.5	2.83; 2.27
SFL	178 (151)	92.1%	27; 15	0.94; 0.97	40.75; 30	2.38; 1.76
DPL	172 (130)	88.3%	46; 33	0.87; 0.92	65.46; 51.2	3.11; 2.81
ALL	646 (511)	90%	115; 73	0.92; 0.96	186.55; 95.08	3.54; 2.98

Results of the S-Libshuff analyses (using 646 sequences) confirmed that the composition of the SFL was not different to that of SML ($p>0.05$) (Table 3). All the remaining pairwise comparisons were significantly different ($p<0.0001$), thus indicating that different communities were represented in the libraries. In order to analyze further the extent of the similarity between these communities we used the Jaccard index (Table 4). As expected, SFL and SML were the most similar communities sharing 35.6% of their OTUs. The second most similar communities were SML and WTL (19.4%). DPL was the most dissimilar community.

Ecological diversity comparisons using these standard statistics are influenced by the cutoffs chosen for OTU delimitation, and most importantly, by considering all OTUs equivalently. However, some OTUs may be highly divergent and phylogenetically unique whereas others may be part of a closely related cluster and thus phylogenetically redundant. To address this issue, we also computed statistics that use genetic data for

comparing the community diversity among the environments. Thus, AMOVA (analysis of the molecular variance) tests showed that SFL and SML contain different bacterial genetic diversity (F_s= 569.9, p-value < 0.001). Pairwise comparisons of the libraries from the other environments also confirmed different genetic diversity within their populations (Table 4).

Table 3. Results of the S-Libshuff analysis testing the differences between each pair of libraries using all the sequences (646).

Comparison	dCXYScore	Significance
DPL > SML	0.01954862	<0.0001
SML > DPL	0.00548060	<0.0001
DPL > SFL	0.02690846	<0.0001
SFL > DPL	0.00232126	<0.0001
DPL > WTL	0.02320122	<0.0001
WTL > DPL	0.00428187	<0.0001
SML > SFL	0.00308606	<0.0001
SFL > SML	0.00003119	0.3766*
SML > WTL	0.00161312	<0.0001
WTL > SML	0.00142426	<0.0001

(*) Significantly different libraries after multiple-testing with Bonferoni correction (p<0.0041).

Table 4. Comparative pairwise analyses of the gene libraries: Jaccard index, AMOVA and Unifrac test.

Test		DPL-SFL	SML-WTL	DPL-SML	DPL-WTL	SFL-SML	SFL-WTL
Jaccard (100%)	J (%)	10.29	19.44	9.64	7.23	35.85	14.75
Jaccard (>99%)	J (%)	1.08	1.22	0.75	0.42	4.30	1.17
AMOVA (100 %)	Fs p-value	26.02 <0.001*	481.832 <0.001*	123.57 <0.001*	123.92 <0.001*	569.90 <0.001*	110.76 <0.001*
Unifrac (100 %)	Unweighted	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
	Weighted	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
Unifrac (>99 %)	Unweighted	**<0.001**	0.03401	**<0.001**	**<0.001**	0.009009	0.03001
	Weighted	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*

(*) Significantly different libraries after multiple-testing with Bonferoni correction (p<0.0083).

Phylogenetic diversity and Bacteroidetes community structure at the temporal and spatial scales.

Community structure across environments was assessed as phylogenetic diversity (PD) and phylogenetic species variation (PSV), incorporating phylogenetic information from the maximum likelihood inferred trees of OTUs clustered at 99% similarity (Table 5). Overall, SFL showed lower values of both metrics, indicating a closer relatedness of taxa within this environment. The PSV analysis revealed significant nonrandom community structure, due to the presence of phylogenetically related species in the same environments more frequently than expected by chance (Table 5). Overall, rarefaction curves increased the PD values, as adding new sequences increased branch length of the inferred phylogenetic trees. The SFL curve approached saturation, indicating a smaller phylogenetic diversity for an equivalent sampling effort (Fig. 3). DPL, WTL and SML, although decreased their slope, didn't reach a plateau for PD accumulation, indicating that phylogenetic diversity may not be fully recovered in these libraries (Fig. 3).

Table 5. Species richness (SR), phylogenetic diversity (PD ± std.) and phylogenetic species variability (PSV) of the 4 Bacteroidetes clone libraries used in this study.

Library	SR	PD	PSV(*)
DPL	47	1.75±0.132	0.78
SML	44	1.60±0.132	0.59
SFL	28	0.85±0.118	0.52
WTL	42	1.47±0.128	0.58

(*) The standard deviation for PSV values was < 0.0001 for all libraries.

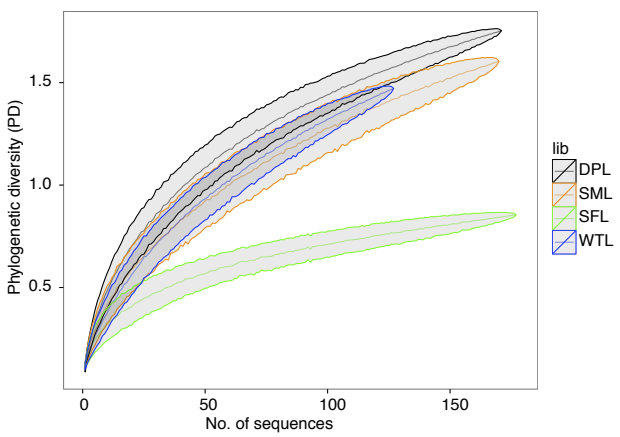


Figure 3. Phylogenetic diversity rarefaction curves for the four gene libraries. Larger phylogenetic diversity is observed in DPL for an equivalent sampling effort.

Lineage-through-time plots show the evolution of diversity of a reconstructed phylogeny (Fig. 4). The deviation of a constant diversification rate or pure-birth model can be assessed with a g test, in which positive values indicate a higher degree of branching towards the tips of the tree, or higher cladogenesis towards the present, and vice-versa. Similar concave shape LTT plots were found for all four libraries, analyzed with all 511 non-redundant sequences. Initially, all the communities except for the DPL one showed fast lineage accumulation rates, and together with recent diversification processes yielded highly significant positive gamma values for all the libraries (between 7.3 and 11.7). The maximum diversification of lineages occurs at the right-side of the plot, indicating an increase in the rate of cladogenesis towards the present. DPL has the lower gamma value (7.3), and unless the rest of the libraries, does not show the slightly convex pattern at the left end of the plot and presents a steeper slope towards relative time 0. This may suggest that DPL has experienced a more recent diversification process.

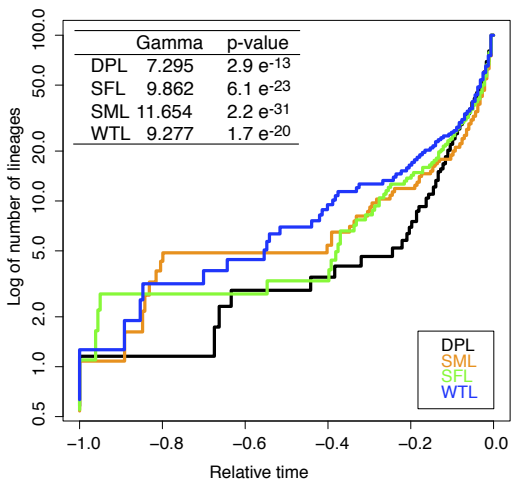


Figure 4. Dynamics of the cladogenesis events versus relative time using transformed number of lineages through time (Lineage Through Time plot) for the inferred phylogeny using the 16S rRNA dataset. Time 0: the present. Statistics for each plot are provided.

In order to compare if lineages from two communities underwent equal rates of diversification, UniFrac analyses were carried out with different phylogenetic trees (Table 4). The analysis of the 4 libraries representing different environments with all the unique sequences (511 OTUs) showed significant differences in microbial community structure for all of them. This result indicates that the sequences clustered significantly

by environment. When the tree was replicated keeping only representative sequences at 99% similarity (115 OTUs), only DPL was different from the other libraries with the unweighted option. Surface samples did not contain significantly different bacterial communities at 99% similarity (considering only presence/absence of OTUs). However, when accounting for the number of sequences classified in each OTU by using the UniFrac weighted option, the bacterial communities of surface environment showed dissimilarities in SFL compared with either SML or WTL, and when comparing WTL and SML (Table 4).

Taxonomic diversity of taxa within the phylum Bacteroidetes

We successfully amplified and sequenced 646 16S rRNA. At a 99% similarity cutoff, these grouped in 115 OTUs, of which OTU 1 was highly abundant (178 sequences), whereas OTUs 62 to 115 were represented by a single sequence. Overall, the sequences obtained in this study presented 87-100% similarity to their closest BLAST hits. Classification of the sequences by SINA and RDP (Fig. 5) sorted them as 7 class Cytophagia, 6 class Bacteroidia, 28 class Sphingobacteria, and 602 class Flavobacteria (of these, 59 sequences belonging to Family Cryomorpha and 543 sequences to Family Flavobacteriaceae). Some sequences were not classified further than to family level by any classifier tool, and they are referenced as “Uncultured”.

The phylogenetic position of class Bacteroidia, represented only by 6 sequences, was somewhat unstable. It generally rooted within the class Flavobacteria, but with no bootstrap support in the consensus tree (Fig. 5).

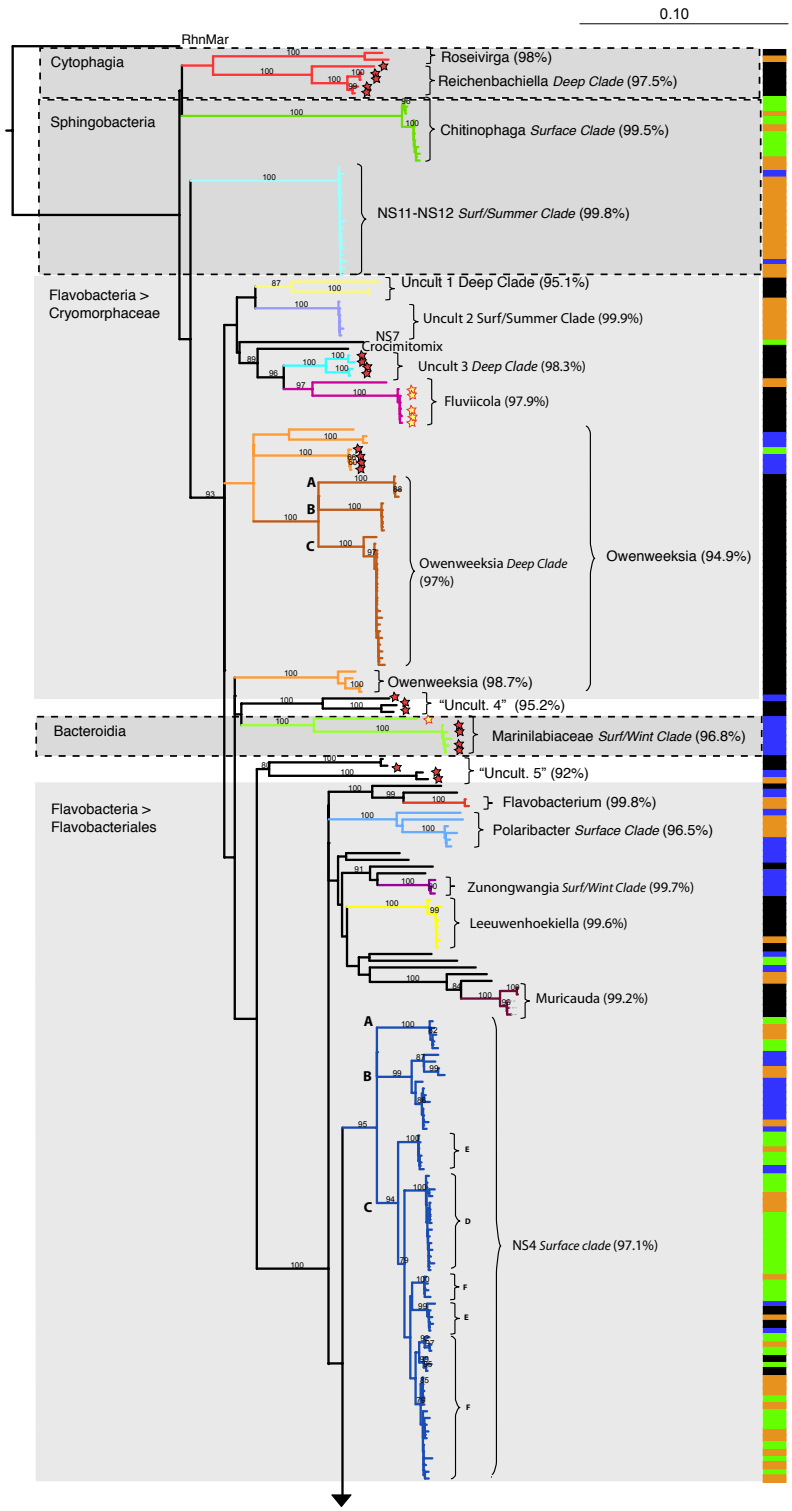
Within class Flavobacteria, the majority of the sequences belonged to uncultured clusters NS2b, NS4, NS5, and NS7 (as defined by Alonso *et al.,* 2007), but several genotypes matched clusters with cultured organisms such as the genera *Fluviicola*, *Owenweeksia, Winogradskiella, Formosa, Leeuwenhoekiella, Muricauda, Zunongwangia*, *Polaribacter*, etc. (Fig. 5). *Owenweeksia* did not appear as a monophyletic clade, and the classifier tools failed to classify consistently sequences in this genus (e.g. some sequences were classified as class Flavobacteriales > Family Cryomorphaceae by RDP and as class Sphingobacteriales > WCHB1-69 by Silva). This notwithstanding, they were firmly resolved within Genus *Owenweeksia* in our phylogenetic analyses. The overall mean similarity for all our sequences was 90.7%. The mean similarity within clusters was 97.6% and among different clusters 86.4%.

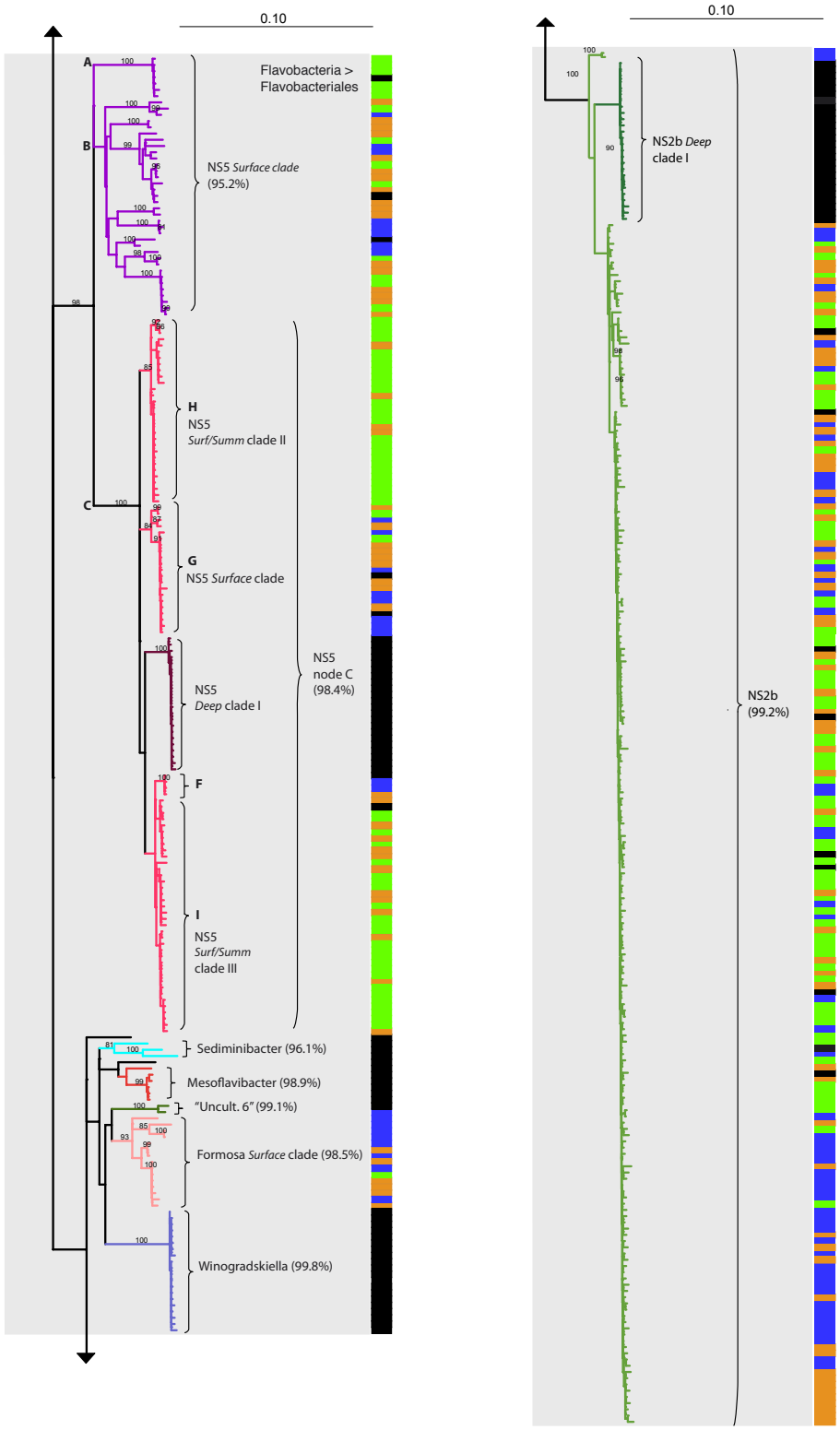
Habitat-specific clusters as defined by 16S rRNA sequences

Interestingly, some of the clades contained sequences from specific environments with closest BLAST hits in sequences retrieved from similar environments (i.e. deep or surface). Sequences belonging to class Cytophagia were almost exclusively retrieved from DPL, and many of their closest hits were retrieved from deep-sea or sediment samples. Similarly, sequences belonging to class Sphingobacteria were present only in libraries from surface samples, mainly in SML (only 2 sequences in WTL), and their closest hits were also from surface samples. All the sequences belonging to class Bacteroidia (7 seq.) were retrieved from WTL. Within the class Flavobacteria, Family Cryomorphaceae was mainly present in DPL, with few exceptions (i.e. "Uncult 2" from SML, and some *Owenweeksia* sequences from WTL; Fig. 5) and, in general, the closests hits were also retrieved from the same habitats. For instance, *Owenweeksia*-deep clade I, had closest hits exclusively from deep waters (1,000 m to 6,000 m). An exception was *Fluviicola* that was present mainly in the DPL, although its closest hits were usually from surface samples.

The diverse Family Flavobacteriaceae (within class Flavobacteria) was represented in all libraries. Nevertheless, some specific subgroups were retrieved mainly from DPL (i.e., *Leeuwenhoekiella*, *Muricauda*, Sediminibacter, *Mesoflavibacter*, and *Winogradskiella*) although their closest BLAST hits were found among sequences from surface waters. We only found some deep-specific clusters whose closest hits were also from deep waters in a few nodes within the NS5 clade (deep clade I) and the NS2b clade (deep clade II) (from 90 m for NS5 sequences, and between 500 and 2,000 m for the NS2b). In addition, some season-specific clusters were identified within NS4 clade (node B from WTL, nodes A & C from SML), within NS5 clade (nodes H and I from SML) and within *Formosa* clade (node A from WTL). No apparent season specific cluster was found among sequences corresponding to NS2b clade.

Figure 5. Cladogram of Bacteroidetes 16S rRNA gene sequences present in the clone libraries obtained in this study. The regions are outlined in 5 grey boxes according to the taxonomic affiliation of the main clusters. The main subclades are colored randomly, and their taxonomy and mean intra-group sequence percentage similarity is provided. The origin of each sequence relative to the different gene libraries (Fig. 1) is represented in a vertical column (right), colored according to the following code: Blue= WTL; Yellow= SML Green= SFL; Red= DPL. Sequences are also coded with stars according to their tRNA composition in the ITS region analysis as: Red star=No tRNA; Gold star= 1 tRNAs; No star=2 tRNAs (next two pages)


0.10
RhnMar
Cytophagia
Roseivirga (98%)
Reichenbachiella Deep Clade (97.5%)
Sphingobacteria
Chitinophaga Surface Clade (99.5%)
NS11-NS12 Surf/Summer Clade (99.8%)
Flavobacteria > Cryomorphaceae
Uncult 1 Deep Clade (95.1%)
Uncult 2 Surf/Summer Clade (99.9%)
NS7
Crocimitomix
Uncult 3 Deep Clade (98.3%)
Fluviicola (97.9%)
A
B
C
Owenweeksia Deep Clade (97%)
Owenweeksia (94.9%)
Owenweeksia (98.7%)
"Uncult. 4" (95.2%)
Bacteroidia
Marinilabiaceae Surf/Wint Clade (96.8%)
"Uncult. 5" (92%)
Flavobacteria > Flavobacteriales
Flavobacterium (99.8%)
Polaribacter Surface Clade (96.5%)
Zunongwangia Surf/Wint Clade (99.7%)
Leeuwenhoekiella (99.6%)
Muricauda (99.2%)
A
B
C
D
E
F
NS4 Surface clade (97.1%)


0.10
A
B
C
Flavobacteria >
Flavobacteriales
NS5 Surface clade
(95.2%)
H
NS5
Surf/Summ clade II
G
NS5 Surface clade
NS5
Deep clade I
NS5
node C
(98.4%)
F
I
NS5
Surf/Summ
clade III
Sediminibacter (96.1%)
Mesoflavibacter (98.9%)
"Uncult. 6" (99.1%)
Formosa Surface clade (98.5%)
Winogradskiella (99.8%)
0.10
NS2b Deep
clade I
NS2b
(99.2%)

Exploring microdiversity using the ITS sequences

ITS length was substantially variable among Bacteroidetes, ranging from 93 bp in Crocinitomix to 739 bp in *Muricauda*. The majority (88.5%) ranged between 400-600 bp, but differences among libraries were observed (Fig. 6). Sequences from SFL and SML presented a narrower range of ITS length (Mean=461.6 and 482.7; SD=41.4 and 58.1 respectively), compared to DPL (Mean=454.2; SD=107.4) or WTL (Mean=459.3; SD=95.7) (Fig. 6).

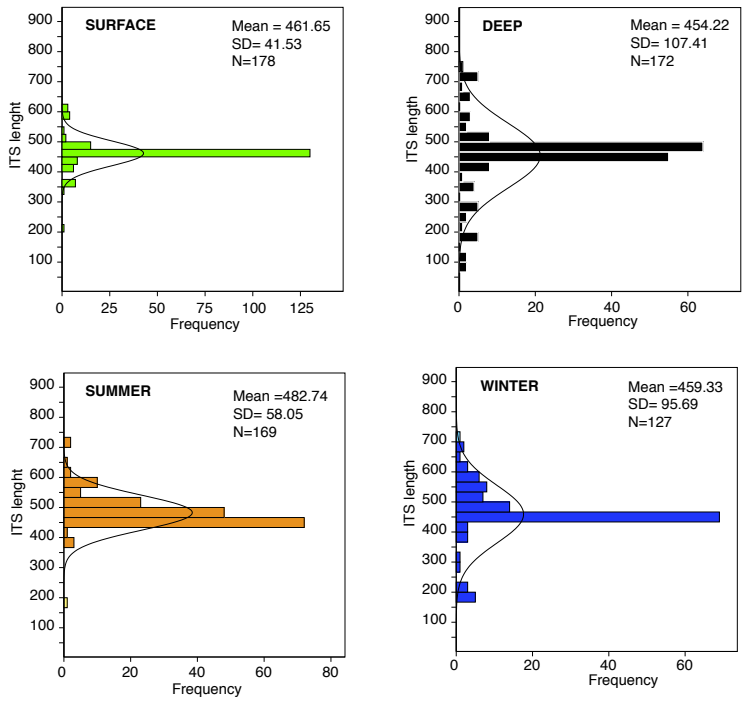


Figure 6. Frequency histograms of ITS length (bps) for each gene library. Each graph represents the number of sequences in each 33 bp length range bin.

Attending to the tRNA composition, the ITS regions could be discriminated in five classes: tRNA-Ile+Ala (613 sequences), tRNA-Ile (5 sequences), tRNA-Ala (1 sequences), tRNA-Val+Ala (1 sequence) or ITS with no tRNA genes (24 sequences). It was observed that, for ITS lengths >342 bp the gene presented both tRNAs, whereas ITS regions <342 bp contained no tRNA with few exceptions. ITS length did not appear to be related with 16S-based phylogeny, and unrelated groups presented similar ITS sizes (Table 6).

Lack of tRNAs was restricted to certain clusters: *Reichenbachiella*, *Fluviicola*, *Owenweeksia* (2), Marinilabiaceae, and Uncult. 2, 3 and 4 (Fig. 5). Interestingly, within these clusters other sequences presented one or both tRNAs (e.g. *Fluviicola*, or Marinilabiaceae). Sequences without any tRNA still conserved the Box A element. The most prevalent motif for this element was GTTC(T/A)TTGACATATTG, with few variations (Table 7) affecting all but 4 invariable sites. The lack of one, or both, tRNA was also more prevalent in sequences from WTL (11) and DPL (17) than in SFL or SML (only one on each). Among the class Flavobacteria no lack of tRNA was detected.

Differences in ITS length within phylogenetic groups ranged from 0 bp for some groups to 275 bp among Oweenweksia or 481 bp among *Fluviicola*. Major differences were associated to ITS presenting different tRNA composition, but substantial differences in ITS length were also found among groups that included both tRNAs, such as *Zunongwagia*, *Leeuwenhoekiella* and NS5 node A-B (260, 277 and 199 bp of difference in size respectively) (Table 6).

Table 6. Analysis of rRNA sequences and ITS sequences present in the clone libraries, grouped according to taxonomic level (as determined by the Silva Classifier tool and displayed in Fig. 5. Average intra-group similarities (AIS) are calculated using all positions of the 16S rRNA dataset.

	No. seq.	AIS (*) at 16S rRNA (%)	ITS length range	Maximum difference (bp)	tRNA composition
Family Flavobacteriaceae					
NS2b	222	99.23	436 - 470	34	Ile+Ala
NS5	157	96.44	417 - 664	199	Ile+Ala
NS4	69	97.07	401 - 496	95	Ile+Ala
Winogradskiella	20	99.8	452 - 480	28	Ile+Ala
Formosa	15	98.54	513 - 600	87	Ile+Ala
Leweenhoekiella	8	99.65	451 - 728	277	Ile+Ala
Polaribacter	6	96.58	526 - 619	93	Ile+Ala
Mesoflavibacter	6	98.94	420 - 433	13	Ile+Ala
Muricauda	5	99.21	539 - 575	36	Ile+Ala
Zunongwangia	4	99.67	454 - 714	260	Ile+Ala
Sedimibacter	3	96.06	494 - 518	24	Ile+Ala
Flavobacterium	3	96.40	499 - 552	53	Ile+Ala
Croceitalea	2	96.40	732 - 811	79	Ile+Ala
Uncult 6	2	99.1	456 - 480	24	Ile+Ala
Phsychroserpens	1	-	496	0	Ile+Ala
Salegentibacter	1	-	633	0	Ile+Ala
Salinimicrobium	1	-	739	0	Ile+Ala
Ulvibacter	1	-	616	0	Ile+Ala
Marixanthomonas	1	-	603	0	Ile+Ala
uncultured	1	-	538	0	Ile+Ala
Croceibacter	1	-	770	0	Ile+Ala
Myroides	1	-	503	0	Ile+Ala

Table 6. Continue

	No. seq.	AIS (*) at 16S rRNA (%)	ITS length range	Maximum difference (bp)	tRNA composition
Family Cryomorphaceae					
BBMO5	29	97.01	445 - 501	56	Ile+Ala
Owenweeksia	17	94.9	226 - 501	275	Ile + Ala / none
Fluviicola	7	97.87	282 - 574	481	Ile + Ala / Ile
Uncult. 2	6	99.88	469	0	Ile+Ala
Owenweek. missclassified	4	98.7	349-356	7	Ile+Ala
Uncult. 3	4	98.33	93 - 114	21	none
Uncult. 1	3	95.1	527-532	5	Ile+Ala
Class Sphingobacteria					
NS12	17	99.79	524	0	Ile+Ala
Chitinophaga	10	99.45	373 (+1)	1	Ile+Ala
Class Cytophagia					
Roseovirga	**2**	98	560 - 669	29	Ile+Ala
Reichenbachiella	**5**	97.5	178 - 207	109	none
Class Bacteroidia					
Marinilabiaceae	**7**	96.83	711 - 836	125	Ile + Ala / none
Unclassified					
Uncult. 4	3	95.21	236 - 310	74	none
Uncult. 5	4	91.98	181 - 453	272	Ile + Ala / none

To accurately compare the divergence of ITS sequences within closely related organisms, we focused on the classification resulting from the similarity of 16S rRNA sequences at 99% cutoff (646 sequences arranged as 115 OTUs, from which 61 OTUs included more than two sequences) (Table 8 and Fig. 7). Most of them presented few differences in their ITS region length. Thus, 50 OTUs (representing 332 sequences) had differences < 10 bp. Five OTUs (representing 209 sequences) differed between 10 and 40 bp, and nine OTUs (35 sequences) had differences >40 bp (Table 8). The largest difference was found in OTU 45, belonging to the "Uncult 5" clade, which presented a difference of 328 bp in ITS length between two almost identical 16S rRNA (>99% similarity). Other large differences were observed in OTUs 41, 51 and 59 (all belonging to NS5), OTU 15 (*Leeuwenhoekiella*), OTU 20 (Bacteroidia) and OTU 23 (*Fluviicola*).

Considering the ITS sequence similarity, 35 of the OTUs (defined by >99% similarity in the 16S rRNA gene) presented >99% pairwise similarity in the ITS region as well. Eight OTUs were 97 to 99% similar, and 8 OTUs were between 93% and 97% similar in this region. Only 10 OTUs (representing 35 sequences) presented large ITS differences (below 91.5% similarity). In these cases, the divergence was mostly due to the organization of the intergenic spacer region in blocks or modules alternating conserved and variable regions, so the similarity increased significantly by excluding gaps and regions of dubious alignment (e.g. from 34.1 to 86.6% for the "Uncult 5" clade or OUT 45) (Table 8).

Table 7. Frequency of nucleotides present at the 16 positions of the "Box A" element of the ITS. The upper line shows the majority-rule consensus at each position and the bottom line shows the prevalence of that nucleotide in the 646 sequences studied.

Nucleotide	G	T	T	C	T/A	T	T	G	A	C	A	T	A	T	T	G
Frequency	100	92	100	100	49.7/50	98	99.4	98.6	100	95.8	97.2	99.7	79.9	89.9	99.4	98.3

Table 8. Analysis of the ITS region of sequences grouped according to the 16S rRNA phylogeny (OTUs defined at >99% similarity cut-off). Average intra-group similarity values (AIS) were calculated both using all positions (All Gaps), and excluding gaps from the analysis (No Gaps). OTUs with low ITS sequence similarity (i.e. <90%) that raised high similarity values when excluding gaps are presented in bold italics.

OTU	Classification	No Seq.	ITS length range	Maximum difference (bp)	tRNA type	AIS at ITS region (%) All Gaps	AIS at ITS region (%) No Gaps
1	NS2b	176	436 - 458	39*	Ile + Ala	95.39	96.93
2	NS5a	37	472 - 475	3	Ile + Ala	98.48	98.60
3	NS5a	27	465 - 468	3	Ile + Ala	99.24	99.36
4	NS2b	26	468	0	Ile + Ala	99.84	99.84
5	NS5a	22	465 - 467	2	Ile + Ala	99.33	99.46
6	Winog	19	521 - 522	1	Ile + Ala	99.52	99.54
7	BBMO5	19	498 - 499	1	Ile + Ala	99.65	99.71
8	NS4	16	472 - 475	3	Ile + Ala	99.32	99.37
9	NS12	17	524	0	Ile + Ala	99.52	99.55
10	NS5a	17	473 - 474	1	Ile + Ala	99.58	99.69
11	NS4	15	463 - 488	25*	Ile + Ala	95.84	99.03
12	NS2b	10	438 - 455	17*	Ile + Ala	97.29	98.30
13	*Formosa*	8	598 - 600	2	Ile + Ala	99.15	99.37
14	NS5a	8	585	0	Ile + Ala	98.71	98.71
15	*Leuweenhoekiella*	7	644 - 728	83**	Ile + Ala	93.68	96.70
16	NS5b	7	417 - 421	4	Ile + Ala	99.28	99.68
17	NS4	7	455 - 457	2	Ile + Ala	95.52	97.13
18	NS5b	7	510 - 515	5	Ile + Ala	91.10	93.03
19	Chitinophaga	7	373	0	Ile + Ala	99.92	99.92
20	Bacteroidales	5	172 - 278	107**	Ile + Ala / none	***78.55***	***91.31***
21	NS4	6	456	0	Ile + Ala	99.85	99.85
22	Uncult. 2	6	469	0	Ile + Ala	99.79	99.79
23	*Fluviicola*	6	285 - 441	156**	Ile + Ala / Ile	***83.47***	***97.04***
24	NS4	5	460 - 461	1	Ile + Ala	99.57	99.65
25	BBMO5	5	453	0	Ile + Ala	99.91	99.91
26	NS2b	5	436	0	Ile + Ala	99.45	99.45
27	BBMO1	5	420	0	Ile + Ala	99.48	99.48
28	*Zunongwangia*	3	681 - 725	44**	Ile + Ala	93.75	97.75
29	BBMO5	4	445	0	Ile + Ala	99.78	99.78

Table 8. Continue

OTU	Classification	No Seq.	ITS length range	Maximum difference (bp)	tRNA type	AIS at ITS region (%) All Gaps	AIS at ITS region (%) No Gaps
30	*Owenweeksia*	4	202 - 206	4	none	95.62	96.95
31	*Polaribacter*	4	534 - 542	8	Ile + Ala	***88.42***	***89.74***
32	NS5a	4	472 -744	2	Ile + Ala	98.21	98.41
33	NS5a	4	472 - 746	4	Ile + Ala	96.83	97.91
34	NS4	4	469 - 495	26*	Ile + Ala	95.70	98.44
35	NS4	4	475	0	Ile + Ala	99.79	99.79
36	*Muricauda*	3	575	0	Ile + Ala	99.77	99.77
37	NS4	3	471 - 472	1	Ile + Ala	99.72	99.86
38	NS4	3	471 - 474	3	Ile + Ala	98.74	99.43
39	NS5a	3	466 - 471	5	Ile + Ala	97.03	98.13
40	*Formosa*	3	517	0	Ile + Ala	99.74	99.74
41	**NS5b**	3	429 - 563	134**	Ile + Ala	***77.50***	***91.30***
42	Chitinophaga	3	374	0	Ile + Ala	99.82	99.82
43	Uncult. 3	2	114	0	none	100	100
44	Uncult. 3	2	93	0	none	98.93	98.93
45	Uncult. 5	2	181 - 453	328**	Ile + Ala / none	***34.07***	***86.59***
46	*Winogradskiella*	2	456 - 480	24*	Ile + Ala	***75.10***	***85.39***
47	NS2b	2	450 - 451	1	Ile + Ala	99.56	99.78
48	NS4	2	451 - 552	1	Ile + Ala	99.34	99.56
49	*Roseovirga*	2	179	0	none	100	100
50	*Owenweeksia*	2	395	0	Ile + Ala	100	100
51	NS5b	2	508 - 594	86**	Ile + Ala	***75.08***	***87.80***
52	*Flavobacterium*	2	499	0	Ile + Ala	99.80	99.80
53	*Owenweeksia*	2	532	0	Ile + Ala	99.81	99.81
54	*Roseovirga*	2	178	0	none	100	100
55	*Formosa*	2	542 - 543	1	Ile + Ala	99.45	99.63
56	NS5b	2	553 - 618	65**	Ile + Ala	***81.16***	***91.64***
57	NS5b	2	605	0	Ile + Ala	97.36	97.68
58	NS5b	2	444	0	Ile + Ala	99.77	99.77
59	NS5b	2	518 - 664	146**	Ile + Ala	***64.77***	***88.96***
60	*Muricauda*	2	539	0	Ile + Ala	100	100
61	*Owenweeksia*	2	350	0	Ile + Ala	100	100

Length differences among 10 and 40 pb (*) and >40 pb (**).

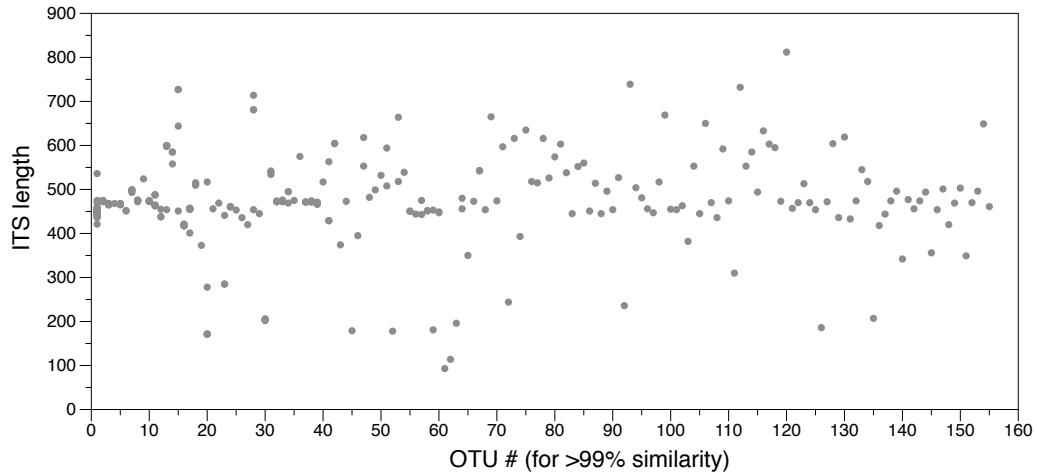


Figure 7. Relationship between 16SrRNA OTU clustering and ITS length. X axis represents the 155 OTUs grouped by >99% similarity at the 16S rRNA and the circles at the Y axis represent the variable ITS lengths present in the different sequences included in the OTU.

Habitat-specific clusters by ITS region

Finally, in order to test the usefulness of the ITS region as a microdiversity marker for phylogenetic analyses, trees derived from alignments of the ITS region were constructed and compared with the 16S rRNA gene phylogeny. We limited this analysis to the largest, best-supported clades inferred by the 16S rRNA gene phylogeny that were not identified as deep-specific clusters (i.e. NS2b, NS4, NS5 and *Formosa*). The ITS trees presented roughly the same topology than the 16S rRNA tree, but the larger heterogeneity of the ITS region allowed to identify new subclades, sometimes with good bootstrap support. Several of these seemed to sort the samples from WTL and SML/SFL.

Specifically, the ITS phylogeny of NS2b (Fig. 8) displayed two main divisions: node A, with sequences from DPL (clade I, ITS length = 468 bp); and node B, with sequences from surface samples libraries. Within node B, further subclades were resolved, two of them mainly constituted by sequences from the WTL (node D or clade II, ITS length = 440 -445 bp; and node H or clade V, ITS length = 447-451 bp), and two additional subgroups by clones from the SFL or SML (node E or clade III, ITS length = 454 bp; node F or clade IV, ITS length = 453 bp). These clades had not been discovered by the lower resolution allowed by the 16S rRNA phylogenetic inference analyses.

In the NS5 group ITS-based phylogeny, the topology differed from the 16S rRNA analyses, particularly affecting the resolution of NS5 (node A-B), which was poorly supported in the 16S rRNA-based analysis (Fig. 5). However, clade NS5 (node C) in the 16S-based tree appeared quite similar to the ITS analyses (Fig. 9). This node was divided in several distinct subclades, in congruence with the 16S rRNA analyses, although the branching order was slightly different. One division, branching off in this tree, contained the DPL sequences (node G or clade I, ITS length = 466-677 bp) while two other divisions contained sequences from the SFL and SML, providing further support to the surface/summer-specific clusters detected by the 16S rRNA-derived tree (node H or Clade II, ITS length = 466-470 bp; and node I or Clade III, ITS length = 473-474p bp). The remaining two subdivisions within Node E (node F and G) grouped sequences present in all the libraries. Interestingly, these two nodes clustered apart from each other in the 16S rRNA-based phylogeny, but appeared as sister groups in the ITS analysis.

The ITS analysis of the NS4 group maintained similar terminal branching as displayed with the 16S rRNA-based tree, but presented slightly different branching order (Fig. 10). Most of the sequences of this group were from surface libraries (only 4 sequences were retrieved from the DPL). Node B grouped mainly sequences form the WTL (NS4 surface/winter-clade II, ITS length = 401-481 bp), and the other four habitat-specific clades contained sequences from the SML or SFL (clade I, ITS length = 495-496 bp; clade III, ITS length = 463-499 bp; clade IV, ITS length = 456 bp; clade V, ITS length = 471-475 bp). These groups were similar to those identified using the 16S rRNA gene, and the ITS did not reveal any further consistent subdivision. Similarly, the *Formosa* group was only constituted by sequences from surface libraries, and the branching were congruent with those identified using the 16S rRNA gene, only resolving one winter specific-clade (clade I, 542-544 bp) (Fig. 11).

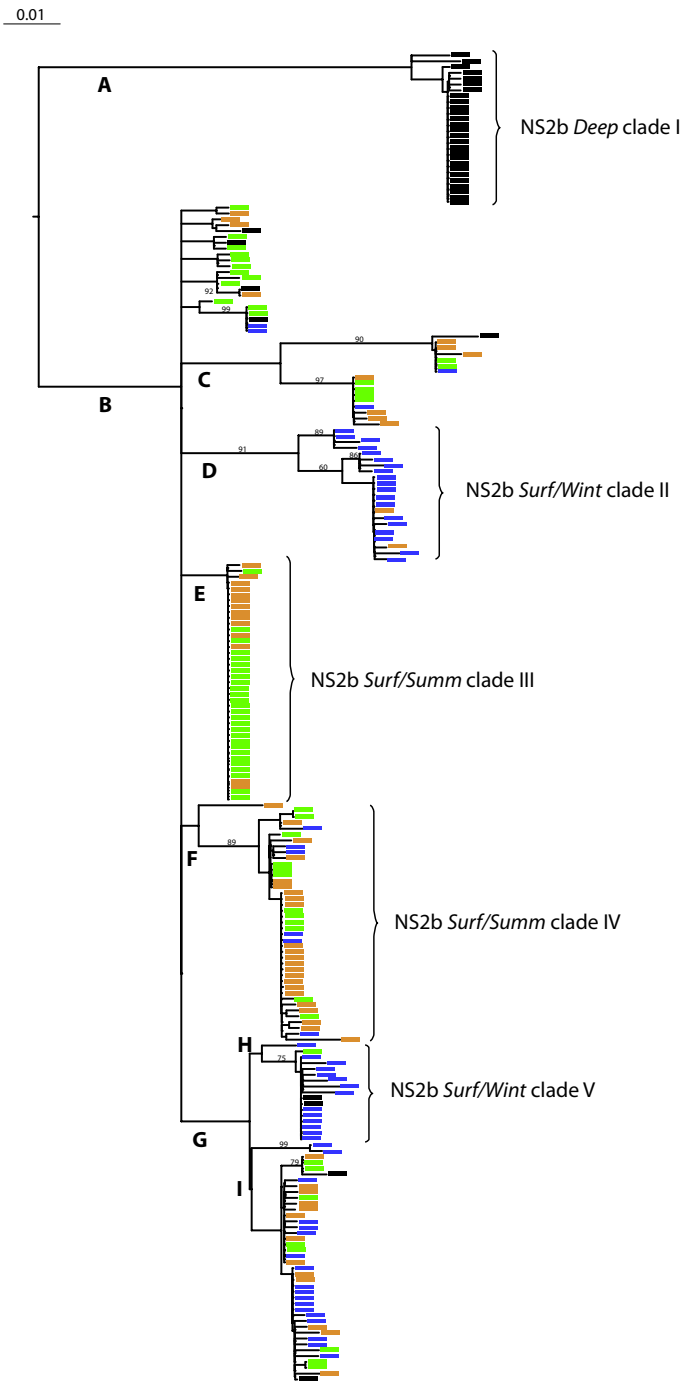


Figure 8. ITS-based cladogram of the NS2b marine group inferred using sequences from the clone libraries obtained in this study. The main nodes are marked with capital letters (A-I). The origin of each sequence relative to the different Gene Libraries is represented as a horizontal bar colored according to the following code: Blue= WTL; Yellow= SML Green= SFL; Red= DPL. Habitat-specific subclades are named and marked with braces.

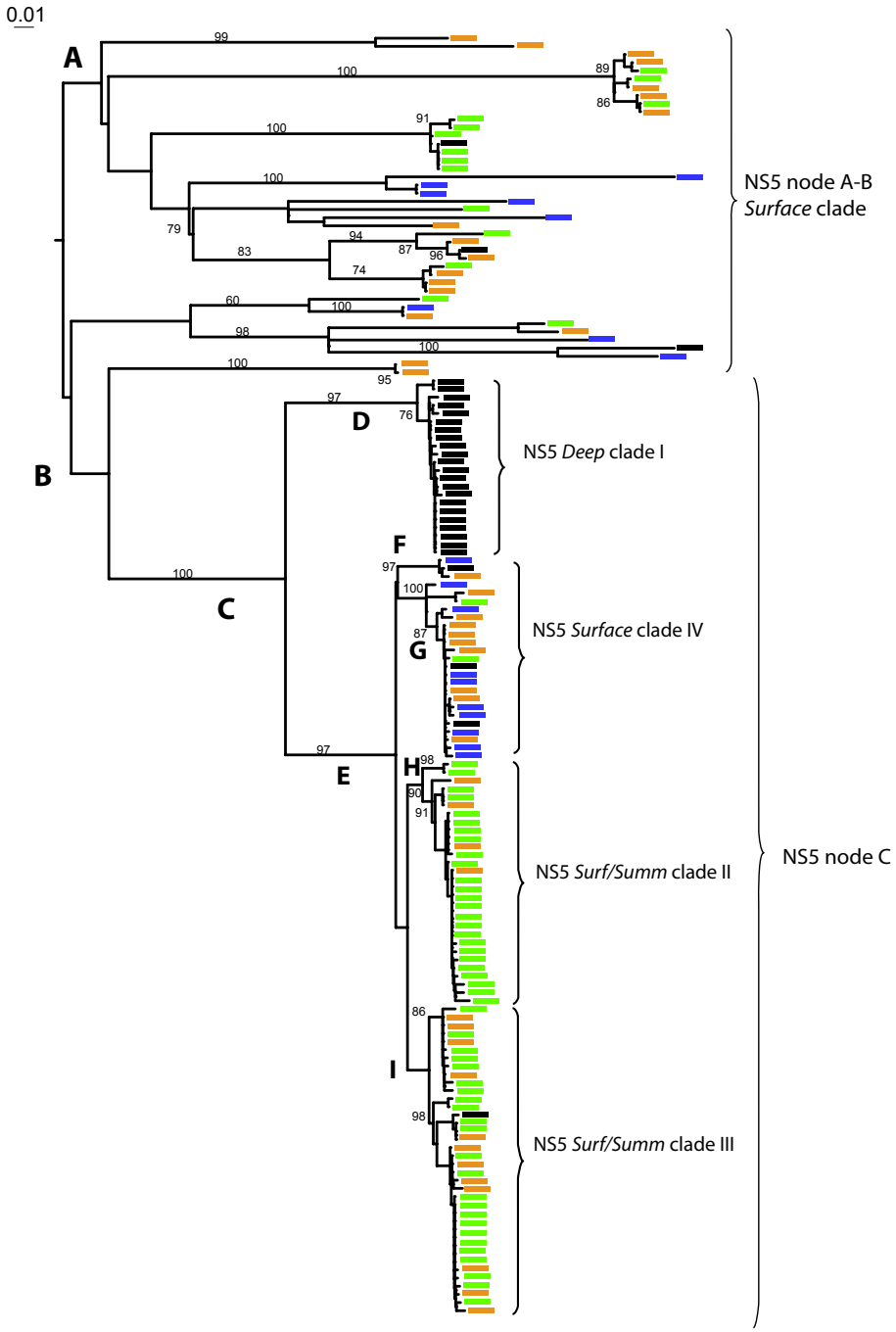


Figure 9. ITS-based cladrogram as in Fig. 8 for of the NS5 marine group.

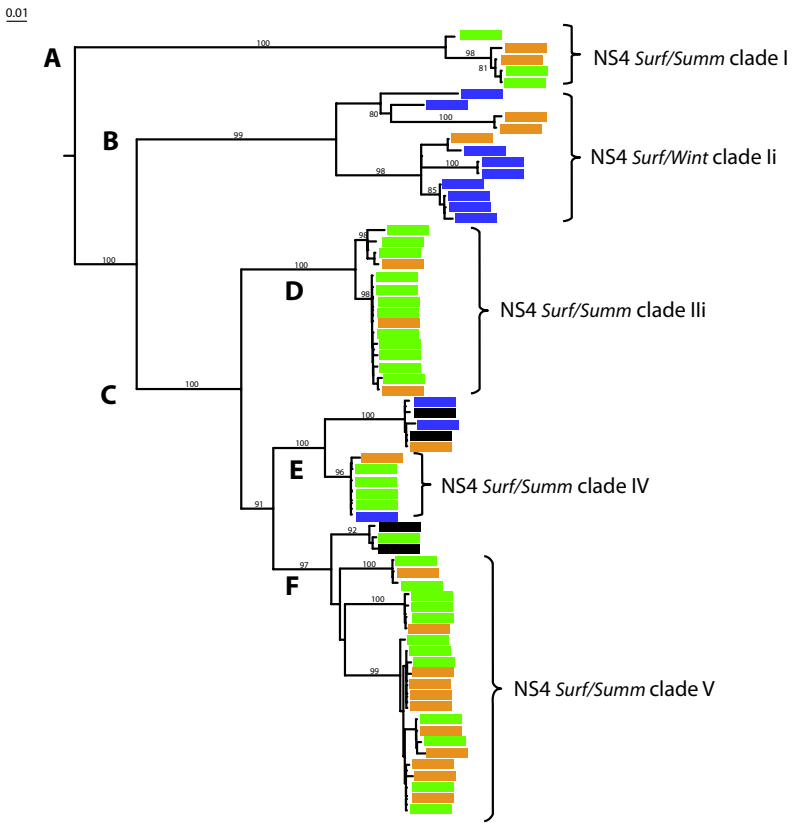


Figure 10. ITS-based cladrogram as in Fig. 8 for of the NS4 marine group.

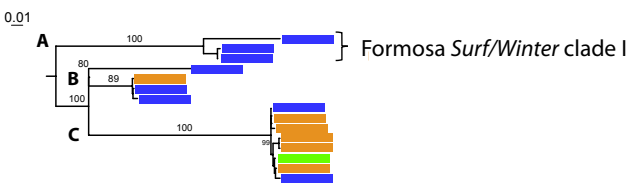


Figure 11. ITS-based cladrogram as in Fig. 8 for of the Formosa clade.

Discussion

A classic question in microbial ecology deals with what are the processes that structure and govern the species composition of communities. Since most species in microbial communities are uncultured, accurate predictions on how they interact with the changing biotic and abiotic factors are quite difficult. Responding to this challenge, taxonomic phylogenies have been incorporated into the study of microbial communities and ecology. In this context, Cohan [2001] presented several universal sequence-based approaches for defining ecotypes and incorporating them into bacterial taxonomy. A bacterial "ecotype" is a set of strains that use the same or similar ecological resources, and it is expected to be identifiable as a monophyletic group in a DNA-based phylogeny [Cohan, 2002]. The correspondence between sequence clusters and ecologically distinct populations has been widely reported [Field *et al.,* 1997; Morris *et al.,* 2004; Wright *et al.,* 1997; Vergin *et al.,* 2013].

This approach has not been applied comprehensively among marine Bacteroidetes, and therefore a detailed analysis of how the group is organized in different clusters, which might represent niche-specific groups, does not exist. In this study, we have analyzed 16S rRNA data to test for the existence of a phylogenetic signal in community composition that would indicate the presence of specific ecotypes. Our dataset represents a significant fraction of the taxa present in the natural assemblages in our study area, as shown by the saturated gene libraries accumulation curves (for a threshold below 99%). Moreover, the number of OTUs in the libraries dropped to 1/3rd when sequences were grouped at 99% sequence similarity, indicating that the library was composed of several bacterial clusters with high sequence similarity. These clusters of microdiversity are commonly identified in bacterioplankton phylogenies, and they have been proposed as ecologically meaningful "units of differentiation" [Acinas *et al.,* 2004a].

Although there is little consensus about a unified view of species definition for prokaryotes, several options have been proposed. Defining species by 97% 16S rRNA identity is a common procedure because it fitted well with the accepted 70% DNA–DNA hybridization threshold [Stackebrandt, *et al.,* 1994]. However, comparative genomics has revealed that species defined at that level are known to encompass large diversity in genome content, physiology and ecology [Keswani & Whitman, 2001; Thompson *et al.,* 2005; Staley *et al.,* 2006; Ward, 2006] and are expected to underestimate total diversity. We used the 99% similarity cutoff as the threshold for "species" delimitation. This value

also correlates well with new studies using whole genome analysis. All pairs of genomes from the same "species" showing ca. 94% ANI (average nucleotide identity), correspond to the traditional 70% DNA-DNA similarity and at least 98.5% of 16S rRNA gene identity [Konstantinidis & Tiedje, 2005; Goris *et al.* 2007]. Moreover, a recent study using a clustering analyses with identity cutoff between 98.7 and 99.1% based on 40 universal protein coding genes (pMGs) encountered that the best cutoff to accurately delineate prokaryotic species was well correlated with the range 98.5-99% in 16S rRNA sequence similarity [Mende *et al.,* 2013].

The overall conclusion extracted from the analysis of 16S rRNA sequences and the consequent phylogenetic inferences indicated that the SFL community composition was not different from that of the SML. Nevertheless, the genetic diversity contained in both libraries was different, being SFL less diverse. This could suggest that the SFL would be, to some extent, a subsample of the SML. Furthermore, our phylogenetic analyses indicated that (for >99% similarity of the 16S rRNA) sequences from libraries built from surface samples (SFL, SML and WTL) presented similar community structure in terms of the presence/absence of taxa. Only by weighted Unifrac analyses, i.e., when relative abundance data are considered, it was confirmed that communities were different over the year in our libraries (i.e., SML *vs.* WTL). This observations would support the idea that seasonal changes in community composition correspond to shift in the relative abundance of persistent taxa, rather than taxa extinction and recolonization in the ecosystem through time [Campell *et al.,* 2011; Caporaso *et al.,* 2011].

Considering the spatial distribution, the DPL presented the most different phylogenetic structure compared to any other environment, both in presence/absence of taxa and in their abundance. Moreover, even though the DPL was generated from a single sample, it contained much more diversity than the SFL (from one single sample as well), but also more diversity than the SML and the WTL, which were pools of several samples. Some studies have already identified higher diversity in deep environments compared to the surface ones [García-Martínez 2000, Pommier *et al.,* 2010], and this appears also to be true when analyzing only Bacteroidetes communities.

Lineage-through-time plots show accelerating rates of diversification in all four communities towards the present, indicating a high speciation rate of marine Bacteroidetes, as it has been shown in other marine prokaryotes [Barberán *et al.,*2011] . This effect is slightly more evident in the DPL library, suggesting that this community has experienced a more recent diversification process.

A detailed analysis of the 16S rRNA phylogenies studied here showed a number of clusters particularly associated with depth. This suggests niche partitioning of ecologically distinct populations that may have diverged during evolution due to "speciation" under depth-related factors. Depth-related clades have been previously suggested for Bacteroidetes using the 16S rRNA gene [Blümel *et al.,* 2005]. Season-specific clusters were also identified within some groups, and they were corroborated by the ITS-based phylogenetic analysis. Interestingly, in some instances (e.g., within NS2b clade), 16S rRNA sequences did not resolve any season-specific clade, but the enhanced resolution of the ITS region allowed to clearly identify four new season-clades.

Studies relaying on PCR amplification of the rRNA operon may run into the problem of coexisting non-identical rrn operons [Crosby *et al.,* 2003]. Multiple non-identical rRNA operons may impact studies that depend on the comparison of homologous sequences. Bacterial genomes contain 1 to 15 operons, but ca. 40% have either one or two [Acinas *et al.,* 2004b]. Analyses of genomes of cultured Bacteroidetes demonstrate that class Bacteroida contain 5 operons on average, while classes Flavobacteria and Sphingobacteria present 2.4 and 2.7 average copies respectively (rrnDB) (http://ribosome.mmg.msu.edu/rrndb/index.php) [Klappenbach *et al.,* 2001; Lee at al., 2009]. Although microheterogeneity between paralogous 16S rRNA genes is relatively common, the degree of sequence divergence in bacteria is relatively low, with the vast majority showing <1% nucleotide differences [Acinas *et al.,* 1994b]. It can be assumed that intra–genomic 16S rRNA gene heterogeneity has generally a relatively small impact on species delineation and inference of evolutionary histories [e.g. Engene & Gerwick, 2011]. However, ITS region heterogeneity among operons is larger. It has been extensively reported that cultured bacteria may present high ITS divergence in the same strain [e.g. Milyutina *et al.,* 2004; Liguori *et al.,* 2011; Osorio *et al.,* 2005; Man *et al.,* 2010]. Among Bacteroidetes, one study has used the ITS region for molecular identification of medically important strains, reporting different ITS types for the same cultured species [Kuwahara *et al.,* 2001].

In our study, the analyses of the ITS region revealed that length variation in marine Bacteroidetes was larger than what had been reported for Bacteroidetes cultured strains [Kuwahara *et al.,* 2001, Steward *et al.,* 2007]. Furthermore, Bacteroidetes usually present two tRNAs (Ile + Ala), although we found one ITS with tRNA-Val. This gene is relatively rare within the ITS region and seemed to be confined to bacteria of the genus Vibrio (Gammaproteobacteria) and to the related bacterium *Photobacterium profundum*, and

to date it had not been found in any Bacteroidetes [Steward *et al.,* 2007]. No variation was found regarding the use of different isoacceptors: all tRNA-Ile and tRNA-Ala used the GAU and UGC anticodons, respectively, which are the most frequent for these amino acids [Rocha, 2004]. Interestingly, several sequences did not contain tRNAs, something that has been reported to be common among Firmicutes and Actinobacteria [Boyer *et al.,* 2001], but not within Bacteroidetes [Steward *et al.,* 2007]. However, upon close inspection of available genomes of Bacteroidetes we also found some instances of tRNAs absence in *Fluviicola* taffensis NC_015321 and Chitinophaga pinenesis NC_013132 (details not shown). These genomes contain several ribosomal operons, some of them with two tRNAs and others lacking one or both. The Box A element was found in all the sequences, even in those lacking tRNAs. This element is highly conserved among all bacteria and plays an important role in antitermination of rRNA transcription [Pfeiffer *et al.,* 1997]. The sequence was similar to the previously identified Box A of Bacteroides species [Kuwahara *et al.,* 2001].

Previous studies have reported no length or sequence variation at the ITS locus within identical 16S rRNA sequences in marine bacteria [Brown *et al.,* 2005]. However, we found large ITS divergence (14.2 % differences excluding gaps) among sequences >99.7% similarity in the 16S rRNA (i.e. in *Fluviicola*, *Zunongwangia*, *Leeuwenhoekiella*, NS2b clade and NS5 clade). This could reflect real variability in the ITS region among closely related taxa, but it could also be due to the existence of intragenomic variability, i.e., among rrn operons. Hence, we carried out more detailed comparisons of these heterogeneous ITS sequences with publicly available related genomes, both from culture isolates and from single amplified genomes (SAGs) of uncultured Bacteroidetes (details in Supplemental Material). Genomes from cultured isolates displayed large variability among operons presenting 2 or 3 operons differing in length and sequence. In our study, the ITS sequences related to these cultured taxa (i.e. *F. taffensis* and *L. blandensis*) were similar to the ITS of both operons. This suggests that, at least in some cases, our different ITS sequence of closely related 16S rRNA sequences, may correspond to different operons of the same organism rather than to different discrete organisms.

Although ITS variability among operons for some cultured strains can be high, it had been suggested that marine bacteria in oligotrophic environments contain only one or few operon copies and, if multiple operons were detected, they were identical [Brown *et al.,* 2005b]. This was based on observations of slow-growing bacteria (as most bacteria from oligotrophic environments tend to be) having fewer copies than bacteria

responding rapidly to added resources [Klappenbach *et al.,* 2000]. We support this idea from the observation of the SAGs from uncultured groups (e.g., NS5 and NS2b clades). These genomes presented only one or two operons, in the latter case with identical sequences. However, more genomes from uncultured bacteria are needed to assure that each unique ITS sequence corresponds to a discrete organism [García-Martínez & Rodríguez-Valera, 2000; Brown *et al.,* 2005]. An error in this assumption could result in misinterpretation of some taxa in the ITS-based phylogenies, because a single phylotype can be misclassified into different phylotypes, leading also to an overestimation of microbial diversity.

The spatial and temporal distribution of different populations is related to several factors (light regimen, nutrients, temperature, etc) that are known to vary locally and seasonally. Clades delineated with 16S rRNA or ITS sequences from cultured organisms, e.g., cyanobacterial ecotypes, have in the past been related to a particular phenotype such as chlorophyll content, motility, chromatic adaptation or lack of phycourobilin [Moore & Chisholm, 1999; Rocap *et al.,* 2002]. However, uncultured isolates cannot be associated to phenotypic characteristics. Most natural marine Bacteroidetes communities retrieved in this study represented uncultured clusters for which we lack physiological information, and thus, the environmental preferences of the habitat-specific clades described here are difficult to determine from current knowledge. A plausible hypothesis for Bacteroidetes is that depth-specific clusters may be related to dissolved organic matter (DOM) quantity and quality, which typically changes through depth gradients [Hedges, 2002; Benner, 2002], including the relative amount of high molecular weight compounds. Furthermore, the use of complex macromolecules varies among cultured genera of the class Flavobacteria [Bernardet & Nakagawa, 2006]. Genomic data also point to specialization of distinct flavobacterial clades for degradation of polysaccharides and proteins [Bauer *et al.,* 2006, Woyke *et al.,* 2009; Fernández-Gómez *et al.,* 2013]. It is also known that some Bacteroidetes contain the proteorhosopsin and can express a mixotrophic life style [González *et al.,* 2008] and therefore such lineages are expected to be absent from the deep oceans. Besides depth-specific clusters, temporal-clades would likely specialize in response to particularly changing environmental conditions. The design of specific probes or primers for the quantitative analysis of the habitat-specific clusters identified here could contribute to elucidate the distributions and dynamics of the different Bacteroidetes clades.

Concluding remarks. Using the 16S rRNA and ITS sequences, we have successfully outlined the extent of diversity and the structure of marine Bacteroidetes and uncovered some habitat-related groups, which likely represent ecological units. Many of these clades cannot be related to previously isolated taxa and therefore the phenotypes associated to such ecotypes can not be described. However, identifying sequence similarity clusters will allow a more informed selection of clades to design probes and primers that can be used in further quantitative studies. Doing so will allow to examine their distribution and relative abundance under different environmental conditions, ultimately providing a better understanding of the forces affecting the evolution and population dynamics of this abundant group of heterotrophic bacteria.

Acknowledgments

We thank Montse Coll, Vanessa Balagué and the Blanes Bay sampling team for collecting the samples used in these analyses and Dr. Fred Cohan for hosting CDV at his laboratory in Wesleyan University. Critical reading of the manuscript by Dr. O. Palenzuela is also appreciated. This work was supported by grant MICRODIVERSITY (CGL2008-00762/BOS) and MicroOcean PANGENOMICS (CGL2011-26848/BOS) to SGA and STORM (CTM2009-09352/MAR) and HOTMIX (CTM2011-30010/MAR) to JMG from the Spanish Ministry of Science and Innovation (MICINN). SGA was supported by a RyC contract Spanish Government of Science and Innovation (MICINN) and FP7-OCEAN-2011 (MicroB3) from the EU. Financial support was provided by Ph.D. JAE-Predoc fellowships from CSIC to CDV. PS was supported by project Malaspinomics (CTM2011-15461) from the Spanish Ministry of Economy & Competitivity.

SUPPLEMENTAL MATERIAL.

Comparison with public genomes

We carried out detailed comparisons of some of the heterogenous ITS sequences with those from publicly available related genomes, both culture isolated and single amplified genomes (SAGs) from uncultured Bacteroidetes.

The genome of *Fluviicola taffensis* NC_015321 (96% similarity to our *Fluviicola* 16S rRNA sequences) presents two operons, one of them with tRNA-Ile (289 pb) and the other with both tRNAs (438 nt). These ITS were rather similar to the two ITS versions found in our *Fluviicola* genotypes, also differing in length and in tRNA genes encoded (285 pb and 441 pb respectively).

Similarly, the genome from *Zunongwangia profunda* available in GenBank (NC_014041) (99.6% identity to our sequences at the 16S locus), presented three operons with the same ITS size (730 bp) which were quite similar (97.7% pairwise ID) to the "long" ITS type retrieved by us (726 pb). Our "short" type (681 bp) was also rather similar, but presented a 43 pb deletion between the tRNAs loci.

The genome of *Leeuwenhoekiella blandensis* (NC_043695) includes two operons with similar ITS length (699 & 707 bps), but rather different sequences (88% similarity). Our *Leewenhoekiella* sequences also presented two ITS types (644 & 727 bps) with 81.3% similarity among them, and showing a possible event of homologous recombination (see below).

Regarding the genotypes classified in uncultured clades, we only found two available single amplified genomes (SAGs) that were closely related to our sequences in the NS5 clade and in the large NS2b clade. The SAG MS024-2A (97.7% to the closest 16S rRNA sequence of our NS5 clade) had two operons with identical ITS types (632 pb), which were only 77.3% similar to our ITS sequence (605 bp). Finally, SAG SCGC AAA536-G18 (100% to the closest 16S rRNA sequence of our NS2b clade) had only one operon with an ITS sequence 100% to several of our sequences.

Homologous recombination in the ITS

The sequence comparisons of *Leeuwenhoekiella blandensis* (NC_043695) genome with our genotypes also showed interesting patterns. This genome includes two operons with similar ITS length (699 & 707 bps), but rather different sequences (88% similarity). Our *Leewenhoekiella* sequences also presented two ITS types (644 & 727 bps) with 81,3% similarity among them. Interestingly the first half of our "727 bps type" was almost identical (99.1%) to one L. blandensis operon (the "699 pbs type"), and then it switched to be more similar to the other operon type (i.e. "707 pbs type"), with a 99.8% similarity. A recombination point was noted in the position 322 (between both tRNAs), making our "727 bps type" strikingly close to a hybrid sequence of both L. blandesis operons, in a similar fashion to the mosaic structure described in Photobacterium damselae [Osorio *et al.,* 2005]. This sequence was retrieved from 6 clones belonging to two different libraries (DPL and SML), so the probability of it being a chimeric sequence was disregarded. This pattern has been attributed to homologous recombination between paralogous rrn operons. This mechanism rearranges tRNA genes and sequences blocks within the ITS region [Gürtler & Stanisich 1996; Antón *et al.,* 1998; Lan & Reeves 1998; Liao 2000; Wenner *et al.,* 2002; Osorio *et al.,* 2005], contributing to the homogenization of parts or all of the rrn operon across all copies within a genome , a phenomenon known as concerted evolution [Liao *et al.,* 2000]. However, for many taxa (including Bacteroidetes), the extent to which recombination either homogenizes ITS regions or facilitates the emergence of new combinations of sequence blocks within the ITS is unclear. The available L. blandensis genome was generated from an isolate originally retrieved from our same sampling station (BBMO) [Gómez-Consarnau *et al.,* 2007], so probably our sequences represent the same species/ecotype presenting different combinations of sequence blocks in the ITS region. This could also be the case with other uncultured species clusters, but it is not possible to discern if the ITS variability corresponds to different operons or to different strains.

Chapter 5

Primer design and optimization of the qPCR workflow to explore Bacteroidetes ecotypes

Abstract

PCR-based approaches are fast, effective methods enabling the selective quantification of specific bacterial taxa within environmental samples. We designed several clade-specific primer sets targeting putative Bacteroidetes ecotypes identified among extensive analysis of Bacteroidetes sequences associated to environmental variables. These primers were thoroughly tested as the best possible compromise between specificity and performance for the intended purpose, and they were extensively evaluated *in silico*. An optimization of the qPCR technique workflow and a demonstration of its usefulness were carried out using N22b group, which covered 248 sequences and represent a cosmopolitan Bacteroidetes ecotype. The different primers we have designed offer a toolbox for quantitative studies characterizing Bacteroidetes populations in different environments. Such studies would help to explore and test links of the groups with specific habitats and/or environmental variables.

Primer design and optimization of the qPCR workflow to explore Bacteroidetes ecotypes.
Cristina Díez-Vives, Josep M. Gasol and Silvia G. Acinas. Manuscript.

Introduction

Bacteria belonging to the phylum Bacteroidetes are recognized as important members of the bacterial community in many aquatic environments [Glöcker *et al.,* 1999; Cottrell & Kirchman, 2000a; Alonso *et al.,* 2007]. Earlier studies focusing on the contribution of Bacteroidetes to community structure have studied the group at a broad 'phylum' or 'class' levels. Hence, results of cell abundances and distribution of members of this phylum have been obtained by fluorescence in situ hybridization (FISH) using Bacteroidetes-specific probes, such as CF319a [Manz *et al.,* 1996], CFB560 [O'Sullivan *et al.,* 2002] or CF968 [Acinas *et al.,* submitted]. Some of these results revealed that they contribute to as much as 18-39% of the bacterioplankton cells in coastal surface seawater samples in California [Cottrell &Kirchman, 2000a] and up to 55% of total picoplankton in early summer in coastal North Sea surface waters [Eilers *et al.,* 2001]. In the NW Mediterranean Sea they constitute the second most important group throughout the year with an average 11% of DAPI counts [Alonso-Sáez *et al.,* 2007], with spring peaks up to 22% [Díez-Vives *et al.,* 2012].

However, evidence for distribution patterns of distinct Bacteroidetes subclades is scarce. Examples include FISH-based quantifications of some, few clades. For instance, Cytophaga marinoflava-latercula was shown to contribute up to 6% of the total picoplankton population during spring and summer off Helgoland Island, in the German North Sea [Eilers *et al.,* 2001]. Also, Clade DE2, originally retrieved from the Delaware Estuary, was found to account up to 10% of picoplankton in this estuary as well as in the Arctic's Chukchi Sea [Kirchman *et al.,* 2003; Malmstrom *et al.,* 2007]. Members of the genera *Polaribacter* have also been shown to constitute prominent fractions in polar oceans [Brinkmeyer *et al.,* 2003; Malmstrom *et al.,* 2007].

Beyond the quantification of certain clusters in specific areas of study, only two studies associated some specific clusters to environmental conditions, suggesting the existence of several ecotypes. One of such works was performed in a spatial study in the Easter Mediterranean Sea, using DGGE and clone libraries [Blümel *et al.,* 2005]. These authors identified 3 sub-clusters of Bacteroidetes associated with specific water masses, two of them related to intermediate water layers and the other to deep waters. Another study consisted in a geographic survey along a latitudinal transect in the North Atlantic Ocean, from cold and rich water masses to oligotrophic warm waters [Gómez-Pereira *et al.,* 2011). In this study, the abundance of 11 Bacteroidetes clusters was quantified

using specific FISH probes, and the authors suggested that different provinces harbor distinct Flavobacterial assemblages. Moreover, some of the clades were correlated with environmental variables, suggesting that different groups occupy distinct niches and have different life strategies, probably representing different ecotypes.

Ecotypes are phylogenetically-determined subgroups of bacteria that differ in physiological requirements that determine their niche specificity [Cohan, 2006]. The discovery of ecotypes within *Prochlorococcus* [Moore *et al.,* 1998] was supported by genome sequences and physiological characteristics that could explain their differing vertical distributions, in particular adaptations to different levels of light and nutrients [Rocap *et al.,* 2003; Moore & Chisholm, 1999]. However, the specific traits selected during the divergence of ecotypes are in most cases unknown, at least for most uncultured bacterial groups. As an example, subgroups of the SAR11 clade were reported to present a strong depth-specific distribution [Field *et al.,* 1997], also suggesting the existence of ecotypes in this group. However, this possibility had to be confirmed through quantitative studies on the distribution of these subgroups in depth gradients. In recent years, environmental data combined with quantitative abundance data are accumulating evidence on dynamics of different SAR11 ecotypes, which are revealing details of their ecology [Carlson *et al.,* 2009, Brown *et al.,* 2012, Vergin *et al.,* 2013].

Quantification of specific genotypes has relied mainly on FISH probes targeting distinct ecotypes [Field *et al.,* 1997; Carlson *et al.,* 2009]. However, low abundance of cells and the limited throughput of this technique hamper its application in numerous aquatic samples. Since the pioneering applications of qPCR in microbial ecology [Becker *et al.,* 2002; Suzuki *et al.,* 2005; Takai & Horikoshi, 2000], this technique is now widely used in microbial ecology to estimate gene abundances in environmental samples to explore dynamics of different taxa at high-throughput scale. This method detects amplified rDNAs in early cycles of the PCR, and quantification is performed during the exponential phase of the reactions. The method also allows the estimation of PCR amplification efficiency and thus an estimate of potential biases that could be associated to inhibitory residues in the samples, primer mismatches, master mix or Taq polymerase performance. The specificity is determined by the design of the primers or labeled probes, that allows quantification of broad phylogenetic groups, but also of individual species or strains at a higher resolution. Quantitative measurements by qPCR have already been reported for Proclorococcus ecotypes [Zinser *et al.,* 2006; Malmstrom *et al.,* 2010] and Marine group

I Crenarchaeota [Hu *et al.,* 2011]. For Bacteroidetes, whole phylum or class-specific qPCR primers have been described [Suzuki *et al.,* 2001b; Fandino *et al.,* 2005; Eilers *et al.,* 2007; de Gregorios *et al.,* 2011], but cluster-specific dynamics studies have not been attempted, to the best of our knowledge. Exploring quantitatively the distribution and abundances of different genotype clusters will provide a better understanding of the population dynamics of this abundant group and uncover potential ecotypes adapted to different environmental conditions.

The objective of this study was to design a methodological framework to quantify the relevant Bacteroidetes subclades that we had previously identified during our studies of this group (Chapter 4). These clusters potentially represent distinct ecotypes likely correlated with environmental factors. We designed primers, suitable for qPCR techniques, for the quantification of these specific clades of Bacteroidetes. We tested the primers in qPCR with plasmidic DNA containing the target genotypes, cloned from environmental samples collected in Mediterranean Sea waters. Our preliminary results indicate that these assays can accurately estimate relative proportions of these genotypes in complex mixtures. Our results also reveal some of the inherent limitations of the approach when dealing with closely related genotypes.

Material and Methods

In silico design and accuracy tests of the QPCR primers

Primers for qPCR were designed to be specific for several 16S rDNA genotype clusters defined in our previous study on the diversity and distribution of Bacteroidetes (chapter 4). These clusters were delineated with a dataset comprising 646 sequences from four 16S rDNA environmental clone libraries (SFL, DPL, WTL and SML). To increase the number of target and non-target sequences, we downloaded the 111NR Silva database (http://www.arb-silva.de), July 2012) [Quast *et al.,* 2013] including 286,858 sequences, of which 25,584 corresponded to phylum Bacteroidetes. The 646 sequences were inserted into the Silva guide phylogeny tree of this dataset using the “quick add parsimony” tool included in the ARB software. In order to define the target and non-target hits among the database, sequences that clustered together with our predefined clusters in monophyletic groups, were included as positive hits, and the remaining were defined as out-groups. The alignments of these clusters were used to test existing primers and to design new ones. We used the “Probe Design” tool in ARB for the design of new

primers and the "Probe Check" tool for the subsequent evaluation of mismatches with non-target sequences. The chances of non-desired amplifications were examined with allowance of 1, 2, and 3 mismatches in the binding loci (details not shown). For instance, if one primer presented a perfect binding to non-target sequences with 0 mismatches, complementary primers could not match the same out-groups even allowing up to 2 mismatches in the binding criteria. The sharing of non-target sequences was only allowed when both primers presented at least one (ideally more) mismatches with the sequences of the non-binding targets. Whenever possible, the mismatches were at the 3' end of the binding regions, since mismatches at the 3' termini are extended by DNA polymerases with lower efficiency that perfectly matched probes [Petruska *et al.,* 1988].

Every possible primer pair was further evaluated to meet the requirements of the qPCR technique and to simulate its performance, quality and compatibility (self-complementarity, stability of dimers, etc.), using AlleleID software (Premier Biosoft, Inc.). The main criteria for desired primer characteristics were: (i) amplicon length shorter than 300 bps, (ii) melting temperature (Tm) values of both primers differing no more than 2℃, and (iii) avoidance of cross-homologies and regions of stable secondary structure in the template.

In addition, we chose a universal bacterial primer set, previously tested in qPCR [Fierer *et al.,* 2005], and a Bacteroidetes-specific qPCR set [de Gregoris *et al.,* 2011]. Both primer pairs were checked with ARB's "Probe Match" tool to test their coverage and specificity and to corroborate their perfect match with our sequences.

Preparation of plasmid standards

Environmental 16S rDNA genes from assorted samples taken from the NW Mediterranean Sea were amplified with the Bacteroidetes-specific primers CF319aF [Manz *et al.,* 1996] and 464R (5' CACTATCGGTCTCTCAGG 3') [Acinas *et al.,* unpublished], and cloned using the Topo TA cloning kit (Invitrogen) (chapter 4). Several of these clones were chosen and used as templates for the qPCR standard curves in this study. The plasmids were purified using a commercial silica-based kit (Plasmid DNA Purification kit, Qiagen) and they were linearized by restriction digestion of the supercoiled plasmid with the endonuclease Not I (Sigma). This enzyme's restriction target motif is present only at one site in the vector, and none within the insert in any clone. The restriction was verified by analyzing a subsample of the reaction for each clone in an agarose gel. DNA concentration and purity

of the plasmid extracts was assessed with a NanoDrop Spectrophotometer (ND-1000, Thermo) and with a Qubit Fluorometer (Invitrogen). The number of rDNA molecules in the extracts was calculated using the following formula:

$$\text{molecules } \mu l^{-1} = [a/(6000 \times 660)] \times 6.023 \times 10^{23}$$

where a is the plasmid DNA concentration (g μl^{-1}), 6,000 is the length of the plasmid (3,931 bp of Invitrogen TA vector plus roughly 2,000 bp of the 16S-23S rDNA insert), 660 is the average molecular weight of one DNA base pair and 6.023×10^{23} is the molar constant (Avogadro constant). Standard curves were generated using duplicate series of 10-fold dilutions of the linearized plasmids, containing from 10 to 10^8 target sequences per reaction.

Performance of the PCR and qPCR assays

In order to test the correct performance of the primer sets, all the group-specific combinations were checked first by standard end-point PCR followed by agarose-gel analysis, and subsequently further adapted for qPCR. Although optimized assay parameters between different primer sets varied (i.e., primer concentration, and annealing temperature (T_a), the following conditions were fixed: in a final volume of 15 µl, the reaction mixtures contained 2.5 µl of DNA template, the selected concentration of primers (see below), and 7.5 µl of 2x SYBR Green Supermix (Bio-Rad), which contains Fast Stard Taq DNA polymerase, SYBER Green I, ROX (reference dye), reaction buffer and nucleotides. The primers were diluted in nuclease-free, PCR-grade water (Gibco) while dilutions of plasmids and samples were prepared in Tris-EDTA buffer (Sigma).

The reaction mixtures were assembled in white, nuclease and Pyrogen-free microtubes with clear caps, grouped in strips (AB-gene), and inside a UV-sterilized chamber. Reactions were run in a 7300 Real-Time PCR System (Applied Biosystems) programmed with an enzyme activation step (95°C, 10 min), followed by 40 cycles consisting on DNA denaturing (95°C, 15 s) primer annealing (20 s at the specific temperature), and DNA synthesis plus data collection (72°C, 30 s). The annealing temperatures (see Tables 1-9) were experimentally optimized to maximize the group specificity of the amplification. Each run included duplicate reactions per each DNA sample, an appropriate set of standards, and two non-template samples (blanks) as negative controls of the reaction. Data were analyzed using the 7300 System SDS software (Applied Biosystems).

Optimization of the qPCR efficiency.

To optimize the conditions for each primer pair, we focused on the qPCR efficiency parameter, targeting values between 90–110% (−3.6 ≥ slope ≥ −3.1). This slope is calculated from standard curves plotting the quantity of template and the C_t (threshold cycle). The C_t is the intersection between an amplification curve and a threshold line. It is a relative estimate of the concentration of target in the PCR reaction, i.e., a higher C_t means that less copies were present at the beginning of the reaction. In a "perfect" standard amplification curve with 100% efficiency, the C_t value of a 10-fold dilution should be 3.3 cycles apart (meaning that the amount of product doubles in each cycle). Another critical parameter to evaluate PCR efficiency is the regression coefficient of the standard curve, R^2. An R^2 value >0.99 provides good confidence in the quality of this standard curve.

The proper concentration of template to carry out efficiency tests is the one resulting in C_ts between 20 and 30. A qPCR with standard conditions (400 nM each primer and the theoretical optimal T_m for each primer pair) was run with 10-fold dilutions of the different plasmids (from 10 to 10^6 rDNA molecules per reaction). The optimum concentration of primers was determined by testing a matrix of variable concentrations of forward and reverse primers. The goal of these optimizations was to seek the primer concentration yielding minimal C_t values, and consequently the highest amplification efficiencies. Primer concentrations tested were 100, 200, 300, 400 and 500 nM each. Different annealing temperatures (T_a) were also evaluated in order to improve the efficiency.

Primer specificity and melting curve analysis

Several positive clones (target) for each primer pair were tested for differences in the amplification efficiency of all the target sequences. Several out-group (non-target) clones, with differences in the number and location of mismatches at the binding loci, were also tested for each primer pair in order to verify specificity. While standard curves with 10-fold dilutions were prepared for target sequences, non-target clones were tested using a fixed value of 10^5 copies per reaction. Moreover, mixtures of two target clones, or target plus non-target sequences, were run in the same reaction to test whether or not the presence of different template sequences could cause inhibition of the PCRs.

The specificity tests were also aided by the analysis of melting curves, which provide accurate identification of discrete amplified products and reveal the presence of primer dimers and other amplification artifacts. SYBR Green I binds to all double-stranded DNA, including specific and unspecific PCR products, but these present different melting temperatures, which depends on their base composition and length. The dye is released upon melting (denaturation) of the double-stranded DNA, providing accurate Tm data for every single amplified product. Melting peaks are calculated by taking the differential (the first negative derivative (–dF/dT) of the melt curve. These peaks are analogous to the bands on an electrophoresis gel and allow the qualitative monitoring of amplification products present at the end of a run. A dissociation curve from 55°C to 94°C was generated after the last qPCR cycle in all samples.

Preliminary tests on environmental samples.

Tests for quantitative analyses of environmental marine samples were run in series of 10 fold dilutions (1:10, 1:100 and 1:1,000) using the specific primer pairs for each Bacteroidetes cluster, and using cloned plasmids with the target sequences as standard curves. Those dilutions resulting in low PCR efficiency or deviating from the expected values were excluded from further analyses. This could be due to the presence of reaction inhibitors or interfering targets, or simply reflect that the number of copies was too low to be accurately measured at the highest dilution. To test if environmental samples presented amplification inhibitors, we also spiked dilutions of environmental samples with known numbers of plasmids with target sequences. These spiked samples were tested to determine if the results resembled the perfect addition of both abundances or a fraction of the expected value.

After the qPCR run, the number of rDNA molecules present in each reaction tube was interpolated from the standard curves. These values were converted to true concentration (molecules ml^{-1} of seawater) in three steps: (i) considering the dilution factor of the tested template to find out the number of molecules per µl of DNA extract; (ii) multiplying this number by the volume of DNA extract (200– 300 µl) to obtain the total number of molecules in the sample; and (iii) dividing the later number by the volume of seawater filtered (6–10 l).

Results

In silico coverage and specificity of the clade-specific primers sets

The first step of the workflow included the design and *in silico* evaluation of the primers. We used the latest database available for Bacteroidetes (Silva database NR 111 released on July 27, 2012) in the software ARB, in order to maximize the theoretical specificity and coverage of the newly designed primers. Primers were designed to cover the habitat-specific clusters previously delineated in chapter 4, which included the additional sequences from many environments worldwide, but also to higher-level clades containing these clusters. This "bottom to top" hierarchical targeting of clades and sub-clades was designed to allow quantification of the relative contribution of smaller habitat-specific clades to the larger taxonomic groups. For some wide groups, no primers were found to cover the whole desired group without matching also a significant numbers of non-target hits. This was caused by the limited number of target regions being conserved in the group of interest but with enough variability in the non-target groups, i.e., the Signature Sequences. In these cases, primer design was restricted to the smaller clusters including our clone sequences, but not to higher levels. Tables 1 to 14 show relevant information of the fourteen newly designed primers, including size of the taxonomic target group as defined by the Silva release, number of sequences in the target cluster (T), number of those sequences appearing in our Mediterranean libraries (Mediterranean Target, MT), coverage of the primers, specificity, etc.

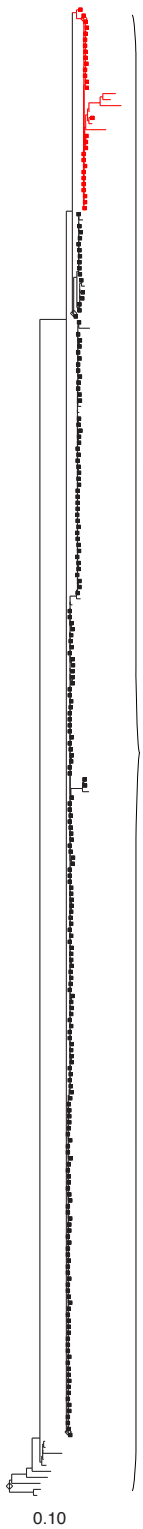
A single pair of primers, NS2b_F & R, covered almost the complete NS2 marine group as defined in Silva's guide tree (including 248 sequences), with high specificity (Table 1). Moreover, we designed two additional primers sets (NS2b_deep_F & R) covering the deep-specific subcluster (NS2b Deep Clade, Table 2) and the remaining surface sequences (NS2b_surface_F & R, Table 3), respectively. The *Mesoflavibacter* clade was also completely covered with newly designed Meso_F & Meso_R primers (Table 9).

Cluster NS4 was less homogeneous. It was divided in 3 major nodes (chapter 4) and we only found feasible to design primers covering node C (pair NS4C_F & R), which contained the largest number of our genotypes (75.7%) (Table 4). A similar situation was observed for group NS5, for which specific primers were restricted to a major subgroup (node C, Table 5) containing 72.3% of our sequences (see tree chapter 4). Within this node C, a subcluster of deep-specific sequences was also chosen as target of additional

specific primers (NS5C_deep_F & R, Table 6), and the remaining superficial sequences were covered with other primer set (NS5C_surface_F & R, Table 7). Sequences within the nodes A and B of NS5 were very diverse, and specific primers could be designed only for small lineages containing a few sequences each (data not shown). The branch of *Owenweeksia*-Deep Clade I also presented 3 subdivisions (chapter 4), which could not be covered specifically with a single primer set. Specific pairs were designed to satisfactorily cover nodes B + C (Table 8) and node A (Table 9) separately. Among the remaining clusters of interest (i.e. *Winogradskiella*, *Leewenhoekiella*, Chitinophagaceae and NS11-NS12), our sequences were phylogenetically resolved as small monophyletic clades within these groups. Hence, the new primer sets matched only these small clades that included our sequences, as well as some additional sequences from other environments that clustered nearby (tables 10 to 13).

For these newly designed primer sets, the in-silico evaluation using primer ARB's match tool with relaxed specificity criteria (i.e., allowing certain mismatches in the binding region) revealed their tendency to amplify non-target sequences in these circumstances (up to 2 mismatches in both primers). The results of this evaluation are also shown for each primer individually, in tables 1 to 14. In addition, details on the number and relevant taxonomical affiliation of shared non-target sequences among other Bacteroidetes or other phylums of Bacteria are also provided.

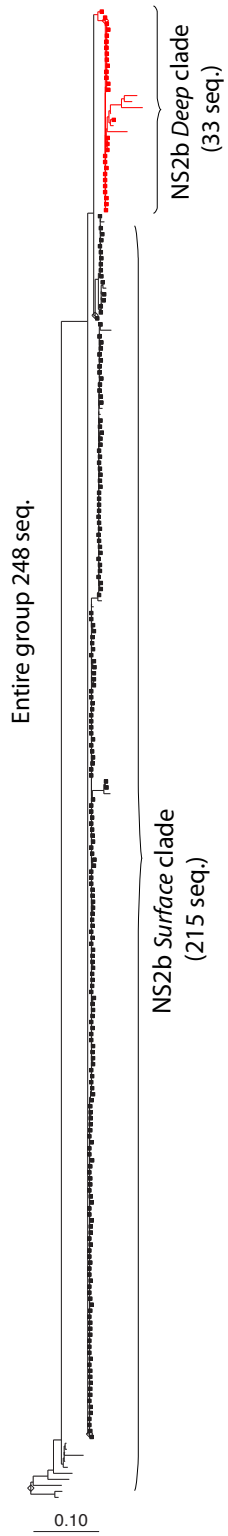
Tables 1-14. Design and in-silico evaluation of Bacteroidetes clade-specific qPCR primers. For each set, its target clade is marked in braces and positioned in a phylogenetic tree based on Silva release. Clades defined as "deep" are cloroured in red. The specificity of the primers is evaluated allowing up to 2 mismatches in the binding locus for each primer, and the number of Target (T) and Non-Target (NT) hits under these conditions is listed independently for the forward and reverse primer. Categories are stablished in the T and NT sequences in order to specify the number of our Mediterranean clones included in each set (Mediterranean-Target, MT; and Mediterranean Non-Target, MNT). NT hits among Bacteroidetes and among other Bacteria are also listed separately. Detailed information on NT hits shared by both primers when allowing different combinations of mismatches is also provided. n.e. = not evaluated.



NS2b marine group (Entire group)

GROUP INFO		No. Seq.			
Silva Clade size		248			
Target group (T) size		248			
Mediterranean Target group (MT) size		224			
PRIMER INFO		**E.coli position**	**Tm**	**Primer sequence 5' -> 3'**	
NS2b_F		1136-1157	63.7	TAATGTCGGGAACTCTAGCAAG	
NS2b _R		1266-1286	62.6	ATAGATTTGCTCATGCTCGC	
PAIR EVALUATION					
Forward (mismatches)	**Total Hits**	**T (MT) NS2b**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	321	232 (220)	86 (9)	3	MNT among *Mesoflavibacter*
1 mm	682	245 (224)	437 (18)	19	MNT among *Mesoflavibacter* and *Zunongwangia*
2 mm	1,712	246 (224)	1200 (36)	266	MNT among *Mesoflavibacter*, *Zunongwangia*, *Formosa* and NS4
Reverse (mismatches)	**Total Hits**	**T (MT) NS2b**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	208	205 (198)	2 (0)	1	none
1 mm	220	215(202)	3 (0)	2	none
2 mm	348	238 (224)	98 (10)	12	MNT among NS4
Shared outgroups					
Number of mismatches					
Forward	**Reverse**	**NT (MNT) Bacteroidetes**		**Relevant shared Bacteroidetes**	
1 mm	1 mm	2 (0)		none	
1 mm	2 mm	16 (1)		none	
2 mm	1 mm	2 (0)		none	
2 mm	2 mm	44 (8)		MNT among NS4	
PRODUCT INFO					
Product lenght		151			
Optimal anneling T°		60.8			

Table 1. Specific primer pair for NS2b marine group (entire group).



NS2b marine group (deep clade)

GROUP INFO		No. Seq.
Silva Clade size		248
Target group (T) size		33
Mediterranean Target group (MT) size		27

PRIMER INFO	E.coli position	Tm	Primer sequence 5' -> 3'
NS2b_ deep_F	1031-1048	59.4	CGGACACTAAACAAGGTG
NS2b _deep_R	1198-1211	60.7	CATAAGGGCCGTGATGAT

PAIR EVALUATION

Forward (mismatches)	Total Hits	T (MT)	NT (MNT) Bacteroidetes	NT Bacteria	Relevant non target Bacteroidetes
0 mm	33	33 (27)	0	0	none
1 mm	246	33 (27)	212 (195)	0	MNT among NS2b surface
2 mm	826	33 (27)	612 (200)	177	MNT among NS2b surface
Reverse (mismatches)	**Total Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	272	33 (27)	123 (27)	115	MNT among NS5b deep
1 mm	7,702	33 (27)	6,617 (525)	1,051	MNT amnong NS5b deep and NS2b surface
2 mm	91,681	33 (27)	18,603 (564)	73,044	n.e.

Shared outgroups

Number of mismatches Forward	Reverse	NT (MNT) Bacteroidetes	Relevant shared Bacteroidetes
1 mm	1 mm	198 (191)	MNT among NS2b surface
1 mm	2 mm	207 (195)	MNT among NS2b surface
2 mm	1 mm	340 (193)	MNT among NS2b surface
2 mm	2 mm	458 (198)	MNT among NS2b surface

PRODUCT INFO	
Product lenght	184
Optimal anneling T°	60.9

Table 2. Specific primer pair for NS2b marine group (deep clade)

NS2b marine group (surface clades)

GROUP INFO		No. Seq.
Silva Clade size		248
Target group (T) size		215
Mediterranean Target group (MT) size		197

PRIMER INFO	E.coli position	Tm	Primer sequence 5' -> 3'
NS2b_surf_F	822-839	59.1	GTCACTAGCTGTTTGGAC
NS2b _surf_R	1031-1048	59.4	CACCTTGTTAAGTGTCCG

PAIR EVALUATION

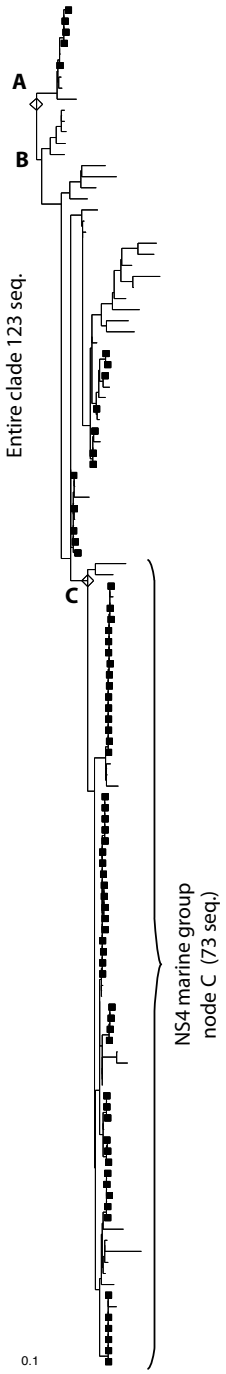
Forward (mismatches)	Total Hits	T (MT) NS2 surf.	NT (MNT) Bacteroidetes	NT Bacteria	Relevant non target Bacteroidetes
0 mm	253	199 (190)	54 (26)	0	
1 mm	391	211 (197)	179 (47)	2	MNT among NS2b deep
2 mm	648	215 (197)	431 (70)	3	MNT among NS2b deep
Reverse (mismatches)	**Total Hits**	**T (MT) NS2 surf.**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	202	201 (196)	1 (0)	0	none
1 mm	369	211 (197)	157 (30)	2	Estan en el NS2b deep
2 mm	1053	215 (197)	786 (20)	23	Estan en el NS2b deep

Shared outgroups

Number of mismatches Forward	Reverse	NT (MNT) Bacteroidetes	Relevant shared Bacteroidetes
1 mm	1 mm	5 (0)	none
1 mm	2 mm	18 (0)	none
2 mm	1 mm	15 (1)	No en NS2b deep
2 mm	2 mm	67 (1)	No en NS2b deep

PRODUCT INFO	
Product lenght	223
Optimal anneling T°	59.8

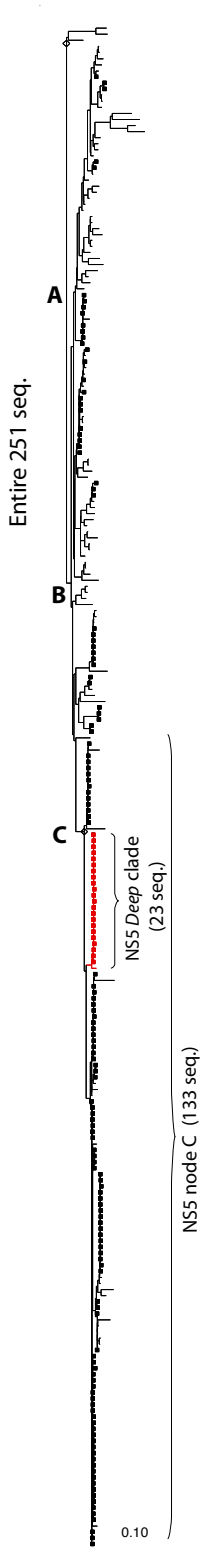
Table 3. Specific primer pair for NS2b marine group (surface clades)



NS4 marine group (node C)

GROUP INFO		No. Seq.		
Silva Clade size		123		
Target group (T) size		73		
Mediterranean Target group (MT) size		53		
PRIMER INFO	**E.coli position**	**Tm**	**Primer sequence 5' -> 3'**	
NS4C_F	668-685	57.7	TGGTTAGAATGTGTGGT	
NS4C_R	822-841	56.3	GTAGACATACAGCTAGTATC	
PAIR EVALUATION				
Forward (mismatches)	**Total Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	116	72 (53)	44 (13)	0	NT among NS4-clade A-B
1 mm	529	73 (53)	454 (216)	2	NT among NS4-clade A-B and NS2b
2 mm	1,408	73 (53)	1,319 (381)	9	NT among our NS4-clade A-B and NS2b
Reverse (mismatches)	**Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	64	64 (45)	0	0	none
1 mm	73	73 (53)	0	0	none
2 mm	73	73 (53)	0	0	none
Shared outgroups					
Number of mismatches					
Forward	**Reverse**	**NT (MNT) Bacteroidetes**		**Relevant shared Bacteroidetes**	
1 mm	1 mm	0 (0)		none	
1 mm	2 mm	0 (0)		none	
2 mm	1 mm	0 (0)		none	
2 mm	2 mm	0 (0)		none	
PRODUCT INFO					
Product lenght		164			
Optimal anneling T°		57.7			

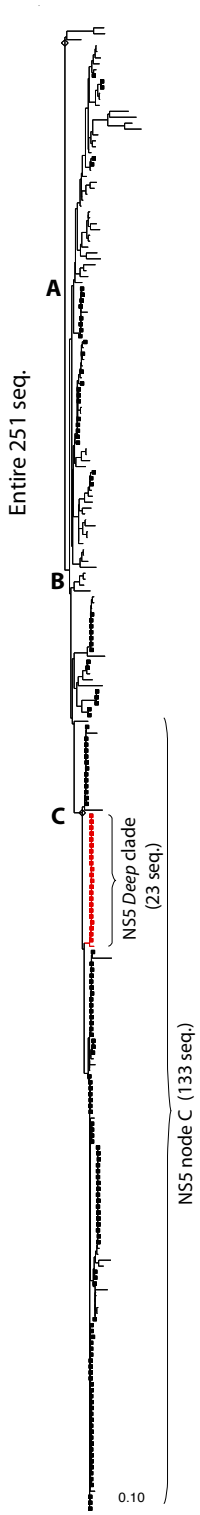
Table 4. Specific primer pairs for NS4 marine group (node C)



NS5 Node C

GROUP INFO		No. Seq.			
Silva Clade size		251			
Target group (T) size		133			
Mediterranean Target group (MT) size		115			
PRIMER INFO	**E.coli position**	**Tm**	**Primer sequence 5' -> 3'**		
NS5C_F	**972-991**	63.3	CGAGGAACCTTACCAGGAC		
NS5C_R	**1135-1157**	64.4	TTCGCTAGAGTTCCCAACTTTAC		
PAIR EVALUATION					
Forward (mismatches)	**Total Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	188	129 (114)	33 (6)	25	NT among NS5 node A-B
1 mm	6,625	130 (115)	3,693 (456)	2,8	NT among NS5 node A-B
2 mm	72,328	132 (115)	13,309 (518)	58,884	n.e.
Reverse (mismatches)	**Total Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	136	131 (114)	3 (0)	2	none
1 mm	140	132 (115)	6 (0)	2	NT among NS4
2 mm	336	132 (115)	146 (5)	58	NT among Chitinphaga and Bacteroidia; MNT among NS4
Shared outgroups					
Number of mismatches					
Forward	**Reverse**	**NT (MNT) Bacteroidetes**		**Relevant shared Bacteroidetes**	
1 mm	1 mm	4 (0)		none	
1 mm	2 mm	45 (5)		MNT among NS4	
2 mm	1 mm	4 (0)		none	
2 mm	2 mm	142 (5)		MNT among NS4	
PRODUCT INFO					
Product lenght		183			
Optimal anneling T°		60.1			

Table 5. Specific primer pair for NS5 marine group (clade C)



NS5 Node C (deep clade)

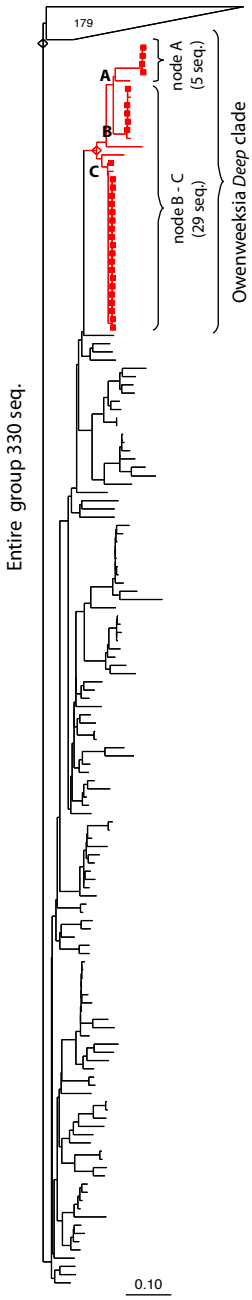
GROUP INFO		No. seq.			
Silva Clade size		251			
Target group (T) size		23			
Mediterranean Target group (MT) size		22			
PRIMER INFO		**E.coli position**	**Tm**	**Primer sequence 5' -> 3'**	
NS5C_deep_F		**1253- 1274**	68.8	AAGCAGCCACTTTGCGAAAAGG	
NS5C_deep_R		**1434-1463**	67.7	AGTTCCTTTCGGTCACCGACT	
PAIR EVALUATION					
Forward (mismatches)	**Total Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	23	22 (21)	1 (0)	0	none
1 mm	26	23 (22)	3 (0)	0	NT are NOT among NS5 node C surface
2 mm	138	23 (22)	113 (85)	2	MNT among NS5 node C surface
Reverse (mismatches)	**Total Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	22	22 (22)	0	0	none
1 mm	149	23 (22)	118 (52)	8	MNT among NS5 node C surface
2 mm	1,317		1,233 (137)	61	MNT among NS5 node C surface
Shared outgroups Number of mismatches					
Forward	**Reverse**	**NT (MNT) Bacteroidetes**		**Relevant shared Bacteroidetes**	
1 mm	1 mm	0		none	
1 mm	2 mm	0		none	
2 mm	1 mm	21 (19)		MNT among NS5 node C surface	
2 mm	2 mm	65 (62)		MNT among NS5 node C surface	
PRODUCT INFO					
Product lenght		202			
Optimal anneling T°		63.8			

Table 6. Specific primer pair for NS5 Node C (Deep clade)

NS5 Node C (surface clades)

GROUP INFO		No. Seq.			
Silva Clade size		251			
Target group (T) size		110			
Mediterranean Target group (MT) size		93			
PRIMER INFO		**E.coli position**	**Tm**	**Primer sequence 5' -> 3'**	
NS5C_surface_F		1140-1161	67	TTGGGAACTCTAGCGAAACTGC	
NS5C_surface_R		1246-1267	65.7	GCATAGTGGCTGCTTTCTGTAC	
PAIR EVALUATION					
Forward (mismatches)	**Total Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	138	109 (93)	27 (22)	2	MNT among NS5 node C deep
1 mm	222	109 (93)	38 (27)	75	MNT among NS5 node C deep
2 mm	6,753	109 (93)	1,292 (109)	5,352	MNT among NS5 node C deep and node A-B
Reverse (mismatches)	**Total Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	110	106 (92)	2 (0)	2	NT are NOT among NS5 node C deep
1 mm	134	109 (93)	23 (2)	2	MNT among NS5 node C deep
2 mm	452	109 (93)	320 (85)	23	MNT among NS5 node C deep
Shared outgroups Number of mismatches					
Forward	**Reverse**	**NT (MNT) Bacteroidetes**		**Relevant shared Bacteroidetes**	
1 mm	1 mm	5 (2)		MNT among NS5 node C deep	
1 mm	2 mm	25 (21)		MNT among NS5 node C deep	
2 mm	1 mm	5 (2)		MNT among NS5 node C deep	
2 mm	2 mm	128 (67)		MNT among NS5 node C deep	
PRODUCT INFO					
Product lenght		128			
Optimal anneling T°		62			

Table 7. Specific primer pair for NS5 Node C (surface clades)



Owenweeksia Clade A

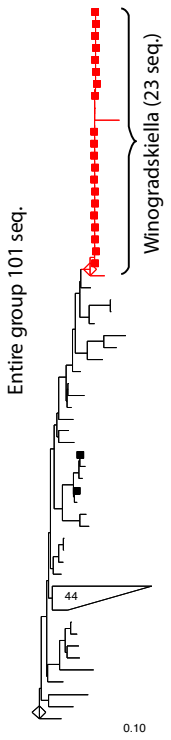
GROUP INFO		No. Seq.			
Silva Clade size		330			
Target group (T) size		5			
Mediterranean Target group (MT) size		4			
PRIMER INFO	**E.coli position**	**Tm**	**Primer sequence 5' -> 3'**		
OweenA_F	579-596	60.5	GTGTAGGCGGGAATGTAA		
OweenA_R	838-858	60.8	GGACGCCGACAGTATATC		
PAIR EVALUATION					
Forward (mismatches)	**Total Hits**	**T (MT) clade A**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	71	4 (4)	0	67	none
1 mm	382	5 (4)	10 (0)	367	NT are NOT in the clades B-C
2 mm	2322	5 (4)	209 (0)	2106	NT are NOT in the clades B-C
Reverse (mismatches)	**Total Hits**	**T (MT) clade A**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	4	4 (4)	0	0	none
1 mm	43	4 (4)	39 (2)	0	NT among Roseovirga
2 mm	474	5 (4)	459 (46)	10	NT amonge clades B-C
Shared outgroups					
Number of mismatches					
Forward	**Reverse**	**NT (MNT) Bacteroidetes**		**Relevant shared Bacteroidetes**	
1 mm	1 mm	0		none	
1 mm	2 mm	0		none	
2 mm	1 mm	0		none	
2 mm	2 mm	0		none	
PRODUCT INFO					
Product lenght	277				
Optimal anneling T°	61.1				

Table 9. Specific primer clade for Owenweeksia clade A

Owenweeksia Clade B-C

GROUP INFO		No. Seq.			
Silva Clade size		330			
Target group (T) size		29			
Mediterranean Target group (MT) size		25			
PRIMER INFO	**E.coli position**	**Tm**	**Primer sequence 5' -> 3'**		
OweenBC_F	999-1018	63.6	AGATGACAGGCTGGGAAAC		
OweenBC_R	1134-1154	63.9	CTTAGAGTCCCCGGCATTATC		
PAIR EVALUATION					
Forward (mismatches)	**Total Hits**	**T (MT) clade BC**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	25	25 (23)	0	0	none
1 mm	29	28 (25)	1 (0)	0	none
2 mm	36	29 (25)	5 (1)	2	MNT are clasified as *Owenweeksia* in Silva.
Reverse (mismatches)	**Total Hits**	**T (MT) clade BC**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	65	28 (24)	13 (0)	24	none
1 mm	426	29 (25)	188 (1)	209	MNT are clasified as *Owenweeksia* in Silva.
2 mm	2,043	29 (25)	757 (10)	1257	MNT are clasified as *Owenweeksia* in Silva.
Shared outgroups					
Number of mismatches					
Forward	**Reverse**	**NT (MNT) Bacteroidetes**		**Relevant shared Bacteroidetes**	
1 mm	1 mm	0		none	
1 mm	2 mm	0		none	
2 mm	1 mm	2 (0)		none	
2 mm	2 mm	2 (0)		none	
PRODUCT INFO					
Product lenght	152				
Optimal anneling T°	62.7				

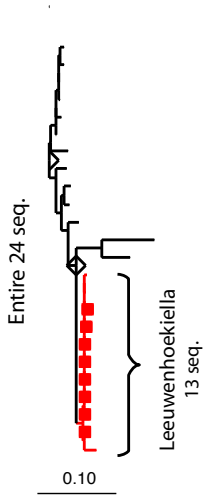
Table 8. Specific primer pair for Owenweeksia clade B-C



Winogradskiella

GROUP INFO		**No. Seq.**		
Silva Clade size		101		
Target group (T) size		22		
Mediterranean Target group (MT) size		20		
PRIMER INFO	**E.coli position**	**Tm**	**Primer sequence 5' -> 3'**	
Wino_Med_F	1019-1046	61.4	GCTATTTCTTCGGACTATCTACA	
Wino_Med_R	1270-1293	61.7	TAGTTTGTAGATTCGCTCCTTC	
PAIR EVALUATION				
Forward hits	**Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	38	22 (20)	16 (1)	0	NT among other *Winogradskiella*
1 mm	55	22 (20)	32 (8)	1	MNT among NS5
2 mm	155	22 (20)	132 (9)	1	MNT among NS5
Reverse hits	**Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	23	22 (20)	1 (0)	0	NT among other *Winogradskiella*
1 mm	30	22 (20)	8 (0)	0	NT among other *Winogradskiella*
2 mm	86	22 (20)	64 (1)	0	NT among tenacibacullum, Salegentibacter and Lutibacter
Shared outgroups					
Number of mismatches					
Forward	**Reverse**	**NT (MNT) Bacteroidetes**		**Relevant shared Bacteroidetes**	
1 mm	1 mm	3 (0)		none	
1 mm	2 mm	4 (0)		none	
2 mm	1 mm	4 (0)		none	
2 mm	2 mm	10 (0)		none	
PRODUCT INFO					
Product lenght		270			
Optimal anneling T°		61.9			

Table 10. Specific primer pair for Winogradskiella



Leeuwenhoekiella

GROUP INFO		No. Seq.		
Silva Clade size		24		
Target group (T) size		13		
Mediterranean Target group (MT) size		8		
PRIMER INFO	**E.coli position**	**Tm**	**Primer sequence 5' -> 3'**	
Leewen_F	464-484	58.4	CGTGTAGCACGATGAC	
Leewen_R	**603-621**	61.7	AAGCTGCAGACTTTCACC	
PAIR EVALUATION				
Forward hits	**Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**
0 mm	13	13 (8)	0	0
1 mm	17	13 (8)	3 (0)	1
2 mm	141	13 (8)	116 (0)	12
Reverse hits	**Hits**	**T (MT)**	**NT (MNT) Bacteroidetes**	**NT Bacteria**
0 mm	63	13 (8)	50 (4)	0
1 mm	701	13 (8)	638 (7)	50
2 mm	3229	13 (8)	2,837 (68)	379
Shared outgroups				
Number of mismatches				
Forward	**Reverse**	**NT (MNT) Bacteroidetes**	**Relevant shared Bacteroidetes**	
1 mm	1 mm	0	none	
1 mm	2 mm	0	none	
2 mm	1 mm	12 (0)	none	
2 mm	2 mm	21 (0)	none	
PRODUCT INFO				
Product lenght	153			
Optimal anneling T°	60			

Relevant non target Bacteroidetes
none
none
none
Relevant non target Bacteroidetes
NT among other *Leeuwenhoekiella*; MNT among *Zunongwangia*
NT among other *Leeuwenhoekiella*; MNT among *Zunongwangia*
n.e.

Table 11. Specific primer pair for Leeuwenhoekiella.

Chitinophagaceae

GROUP INFO		No. Seq.		
Silva Clade size		44		
Target group (T) size		12		
Mediterranean Target group (MT) size		10		
PRIMER INFO	**E.coli position**	**Tm**	**Primer sequence 5' -> 3'**	
Chitin_F	851-869	64.9	GAGGGTCCAAGCGAAAGC	
Chitin_R	**978-995**	60.7	TAGAGCCCAGGTAAGGTT	
PAIR EVALUATION				
Forward (mismatches)	**Total Hits**	**T (MT) Chitinop.**	**NT (MNT) Bacteroidetes**	**NT Bacteria**
0 mm	11	11 (10)	0	0
1 mm	68	12 (10)	56 (0)	0
2 mm	792	12 (10)	763 (3)	17
Reverse (Mismatches)	**Total Hits**	**T (MT) Chitinop.**	**NT (MNT) Bacteroidetes**	**NT Bacteria**
0 mm	49	12 (10)	35 (0)	0
1 mm	3,728	12 (10)	730 (5)	2986
2 mm	26,294	12 (10)	9,486(46)	16,795
Shared outgroups				
Number of mismatches				
Forward	**Reverse**	**NT (MNT) Bacteroidetes**	**Relevant shared Bacteroidetes**	
1 mm	1 mm	6 (0)	none	
1 mm	2 mm	36 (0)	none	
2 mm	1 mm	57 (0)	none	
2 mm	2 mm	309 (2)	MNT among Roseovirga	
PRODUCT INFO				
Product lenght	144			
Optimal anneling T°	60			

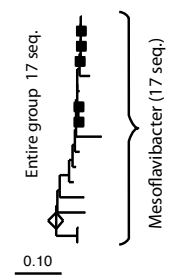
Relevant non target Bacteroidetes
none
none
NT among other Chitinoophagas; MNT among Roseovirga
Relevant non target Bacteroidetes
none
NT among other Chitinoophagas; MNT among *Reichenbachiella*
NT among other Chitinoophagas; MNT among *Reichenbachiella* and Oweenweksia

Table 12. Specific primer pair for Chitinophagaceae

NS11-12 marine group

GROUP INFO		No. Seq.			
Silva Clade size		125			
Target group (T) size		26			
Mediterranean Target group (MT) size		16			
PRIMER INFO		**E.coli position**	**Tm**	**Primer sequence 5' -> 3'**	
NS11-12_F		1002-1019	68.2	TGACCGTCGCCGAAAGGT	
NS11-12_R		1149-1167	60	CGTAGGCAGTGTCTCTAG	
PAIR EVALUATION					
Forward hits	**Total Hits**	**T (MT) NS11-12**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	24	23 (15)	1 (0)	0	none
1 mm	31	25 (16)	7 (0)	0	none
2 mm	124	25 (16)	67 (0)	33	NT among other NS11-12
Reverse hits	**Hits**	**T (MT) NS11.12**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	24	23 (16)	0	0	none
1 mm	100	24 (16)	53 (0)	24	NT among other NS11-12
2 mm	1276	25 (16)	982 (16)	269	NT among other NS11-12, Chitinophaga and *Fluviicola*
Shared outgroups					
Number of mismatches					
Forward	**Reverse**	**NT (MNT) Bacteroidetes**		**Relevant shared Bacteroidetes**	
1 mm	1 mm	0		none	
1 mm	2 mm	0		none	
2 mm	1 mm	9 (0)		none	
2 mm	2 mm	29 (0)		none	
PRODUCT INFO					
Product lenght		163			
Optimal anneling T°		61.3			

Table 13. Specific primers for NS211-12 marine group



Mesoflavibacter

GROUP INFO		No. Seq.			
Silva Clade size		16			
Target group **(T)** size		16			
Mediterranean Target group **(MT)** size		5			
PRIMER INFO		***E.coli* position**	**Tm**	**Primer sequence 5' -> 3'**	
Meso_F		828-848	63.2	AGCTGTTCGGGTTTCGAT	
Meso_R		984-1000	61.1	ACCACATTTAAGCCCTGG	
PAIR EVALUATION					
Forward (mismatches)	**Total Hits**	**T (MT) Mesoflavib.**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	8	8 (5)	0	0	none
1 mm	17	9 (5)	6 (0)	2	none
2 mm	116	10 (5)	74 (19)	32	MNT among Winogradkiella
Reverse (mismatches)	**Total Hits**	**T (MT) Mesoflavib.**	**NT (MNT) Bacteroidetes**	**NT Bacteria**	**Relevant non target Bacteroidetes**
0 mm	68	12 (5)	55 (0)	1	none
1 mm	677	12 (5)	653 (30)	12	MNT among NS4 and *Leeuwenhoekiella*
2 mm	1,940	16 (5)	1,869 (109)	55	n.e.
Shared outgroups					
Number of mismatches					
Forward	**Reverse**	**NT (MNT) Bacteroidetes**		**Relevant shared Bacteroidetes**	
1 mm	1 mm	2 (0)		none	
1 mm	2 mm	4 (0)		none	
2 mm	1 mm	5 (0)		none	
2 mm	2 mm	57 (18)		MNT among *Winogradskiella*	
PRODUCT INFO					
Product lenght		171			
Optimal anneling T°		60.7			

Table 14. Specific primer pair for Mesoflavibacter.

Optimization of qPCR and Plasmid tests

A first round of tests for the evaluation of the primers' performance was carried out by standard end-point PCRs on a variety of target and non-target sequences cloned into plasmids. Some of the pairs resulted in non-specific amplifications, hence they were modified to increase their annealing temperature, and/or to incorporate more mismatches with the non-target groups whenever possible. The primer sets presented (Tables 1 to 14) are the best-performing final versions, and data on previous or substandard-performing designs is not shown. We considered the primers suitable for qPCR only when absence of PCR products from non-target DNA templates in PCR was observed.

Optimization of the primer concentration and annealing temperature for the qPCR protocol was conducted as a trial test with universal bacterial primers (Eub338 and Eub518), and with the NS2b group-specific set, using 10-fold dilutions of the target sequences (10 to 10^8) copies per reaction. For universal bacterial primers, efficiency between 90-110% was achieved using 300 nM of each primer and 62℃ as annealing temperature whereas for the NS2b primer pair (Table 1), optimal efficiencies were achieved using 500 nM of each primer and an annealing temperature of 61℃. In all cases, the performance of the primers resulted in regression coefficients of the standard curves >0.99.

We assessed the performance and specificity of the primer pair for the NS2b group in the presence of non-target Bacteroidetes rDNA. We used a target clone (clone ID # 950) for the standard curve and 14 non-target clones, containing different number and combinations of mismatches with both primers. The non-target clones were amplified with low efficiency, to a maximum of 2.3% of the molecules added to the reactions. Often, they were not amplified at all (Table 15), showing that the primer set was highly specific for NS2b sequences. The out-group clone that amplified the most was Clone ID# 702, which corresponded to a *Formosa* group sequence presenting 2 mismatches in each primer-binding loci.

Table 15. In vitro specificity tests with non-target clones. The standard curve was prepared using a target clone (ID# 950) in 10-fold dilutions containing 10 to 10^7 gene copies/Rx. Copy numbers deduced from the Ct values obtained in reactions with 105 non-target copies are given. The percentage of copies amplified is also provided. Tm represents the melting temperature of the amplicon (melting curve peak).

Clone number	Phylogenetic group	Forward	Reverse	Copies / Rx added	Mean of amplification	% of copies amplified	Tm
950	NS2b (Standard)	0	0	10000000			86.4
950	NS2b (Standard)	0	0	1000000			86.4
950	NS2b (Standard)	0	0	100000			86.6
950	NS2b (Standard)	0	0	10000			86.6
950	NS2b (Standard)	0	0	1000			86.4
950	NS2b (Standard)	0	0	100			86.4
950	NS2b (Standard)	0	0	10			86.4
825	NS4 clade B	3	1	100000	0.33	0.0003	86.9
849	Uncult. 3	3	2	100000	316.33	0.3163	87.1
707	*Marixanthomonas*	1	2	100000	1131.78	1.1318	87.8
702	*Formosa*	2	2	100000	2348.04	2.3480	87.8
58	NS4 clade C	2	2	100000	n.a.	0	69.6
186	NS4 clade C	2	2	100000	n.a.	0	69.4
811	NS5	4	5	100000	n.a.	0	70
272	*Mesoflavibacter*	0	4	100000	0.18	0.0002	77.3
233	*Leewenhoekiella*	0	3	100000	4.44	0.0044	88.3
246	Wino.	3	4	100000	55.16	0.0552	86.6
631	*Owenweeksia*	5	5	100000	147.91	0.1479	86.6
564	NS5	5	0	100000	0.66	0.0007	86.6
752	NS4 clade B	3	4	100000	n.a.	0	69
78	NS4 clade C	2	2	100000	n.a.	0	70.2

* n.a. No amplification

Spiking experiments and melting curves

Experiments to address the performance of primers in complex templates were carried out by mixing target/non-target plasmids as templates for the qPCR reactions. Mixing two different targets for NS2b group (clones IDs# 950 and 703) at $5{\cdot}10^4$ copies per reaction (copies/Rx) resulted in the correct quantification of both (Fig. 1a). To test the effect of the presence of a non-target group (with low but positive cross-amplification), on the quantification of the target sequences, we mixed clone ID# 950 (target) with 702 (non-target). Increasing amounts of each sequence (50 to $5{\cdot}10^5$ copies/Rx) were

tested against a fixed amount ($5 \cdot 10^5$ copies/Rx) of the other one (Fig. 1b-c). When target sequences were present, increasing amounts of the non-target sequence did not distort the correct quantification, even in the presence of equimolar amounts of target and non-target template (Fig. 1b). When non-target template was present, cross-reacting sequences were dominant (fixed at $5 \cdot 10^5$ copies/Rx) and the target sequence was absent or in low numbers (from 0 to 500 copies/Rx), and the amplification was around 1,430 ± 300 copies/Rx (0.28 % of the actual copies present in the mixture). But when the target sequence number was above this background amplification (i.e., when target sequence represented approximately >1% of the total copies present in the reaction) the reaction quantified only the target sequence, being the addition of cross-reacting non-targets negligible (Fig. 1c).

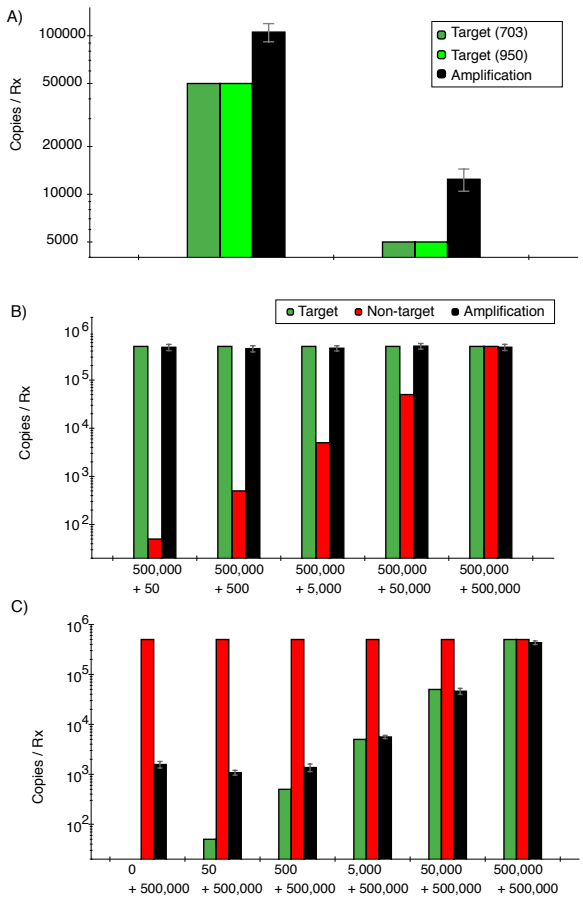


Figure 1. Performance of qPCR with complex templates. A) Mixing of two target clones. B and C) Mixing of target and non-target clones in different combinations of amounts. Color bars indicates the copies of plasmid added to each reaction (green for target groups and red for non-target groups). Black bars indicate the copy number calculated after the qPCR of the mixture.

In addition, non-target sequences could also be identified by different melting curves. Target sequences presented an 81.3℃ T_m peak, *vs.* an 82.5℃ peak of the non-target sequence (Fig. 2). The analysis of the melting curve showed that even in the presence of $5 \cdot 10^5$ non-target copies/Rx, as low as 500 target copies/Rx could be detected as a small peak in the corresponding target product position. When 5000 or more target copies/Rx were present in the mix, the unspecific peak (82.5℃) disappeared, and only the target product was apparent (Fig. 2).

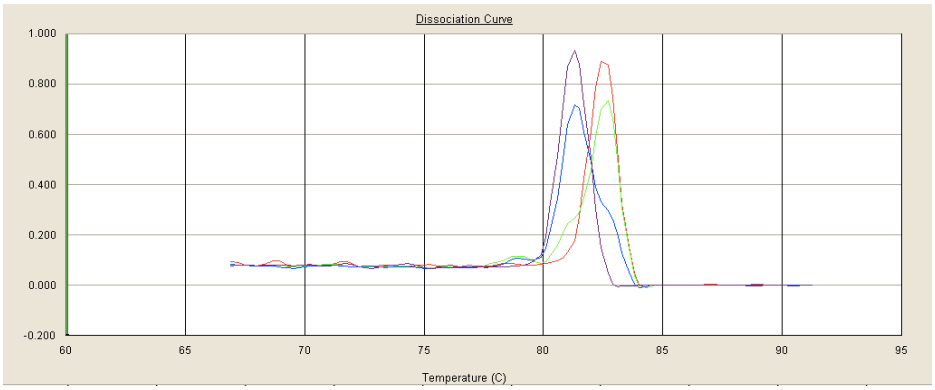


Figure 2. Analysis of melting curves. Example of the run represented in Fig 1c, in which non-target sequences were always present at 500,000 copies/Rx. Red line: Only non-target sequences (Tm=82.5). Green line: non-target and 50 target copies. Blue: non-target and 500 target copies. Purple: non-target and >5,000 target copies (Tm=81.3).

Quantification of NS2b group in environmental samples

A preliminary test to address the usability of NS2b-Entire group primers for group-specific Bacteroidetes quantification on actual environmental samples was carried out in four DNA samples extracted from filtered BBMO seawater corresponding to different months and years were selected from the panel of samples used in previous studies (see chapters 3 & 4). The DNA extractions contained variable amounts of DNA (57 - 344 ng μl^{-1}), and they were used in 10-fold dilutions (1:10, 1:100 and 1:1,000) in replicate reactions. A standard curve was run with a cloned target sequence (clone ID# 950). Quantifications of these environmental samples generally yielded positive results on samples up to 1:1,000 dilutions. However, the PCR efficiency dropped when considering the entire range and, generally, the linear range resulting in best efficiencies was limited to 2 (10-fold) dilution points (Fig. 3). Sample June 2008, presented a very low numbers of target bacteria in the sample, which could compromise the accuracy of the quantification (Fig. 3). With these samples, the technique allowed quantification of roughly 33,000 to 92,000 NS2b rDNA

gene copies per reaction (i.e., corresponding to 370 – 2000 copies/ml) in three of the samples, or approximately 1000 copies per reaction in a summer sample (i.e., 8 copies/ml in June 2008), thus revealing important quantitative differences among some samples (Fig. 3).

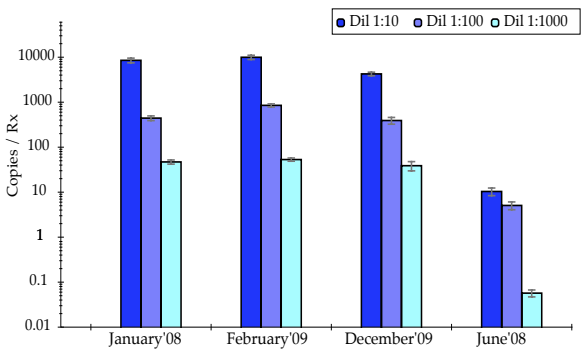


Figure 3. Quantification of NS2b group in assorted environmental samples. The bars indicate the target gene copy number calculated after the qPCR of the samples tested as 3 10-fold dilutions of the extracted DNA.

The possible inhibition of the PCR reactions in environmental samples was also assessed by spiking tests with one of the quantified environmental sample (February 2009). Dilutions 1:10 and 1:100 were mixed with known numbers of plasmid copies containing the target sequence (clone ID# 950). These mixtures were quantified showing that 90 to 114% of the copies from the mixed template, including plasmids and natural genomes, were properly amplified (Table 16). Analysis of the melting curves showed a single peak, demonstrating the selective amplification of the target group (details not shown).

Table 16. Inhibition tests on environmental samples spiked with target sequence (clone ID# 950). Known amounts of both templates were mixed in the qPCR reaction, allowing to calculate the % of the actual target copies present in the mixture that were effectively amplified.

Sample	Copies/rx added*		Quantification Results		
	Env. Sample	Clone	Theoretical result	Actual quantification (Mean)	% of copies quantified
Env. (1:10) + clone	4984	500000	504984,00	468789,00	92,83
Env. (1:10)+ clone	4984	5000	9984,00	8876,69	88,90
Env. (1:10) + clone	4984	50	5034,00	4558,12	90,54
Env. (1:100) + clone	425	500000	500425,00	500129,00	99,94
Env. (1:100) + clone	425	5000	5425,00	4801,91	88,51
Env. (1:100) + clone	425	50	475,00	542,04	114,11

* Environmental sample from BBMO, February 2009.

Discussion

Quantification in spatial or temporal gradients of the different lineages of Bacteroidetes identified in previous phylogenetic analyses (see chapter 4), would allow not only to define and further corroborate the existence of putative ecotypes in this group but also to reveal their distribution in specific niches. Such knowledge would grant a better understanding of the ecological framework and scale in which Bacteroidetes play a significant role.

The most widely used method for enumerating bacterial cells in environmental marine samples is FISH. This technique offers direct detection of the bacteria retained in filters, after they are labeled with group-specific DNA probes. Since the labeling depends on the specificity and the efficiency of probe binding, these factors largely affect the accuracy of data extracted from this technique. In addition, it requires a direct count, which is usually conducted manually under the fluorescence microscope, and therefore it is a very labor-intensive technique and there are a limited number of probes that can be tested in a single sample. Thus, whereas accuracy is considered one of the main strengths of the technique and it is absent of PCR biases, low sensitivity and, particularly, very low throughput can be considered its main weaknesses. The possibility of carrying out enumeration by flow cytometry exists [e.g. Sekar *et al.,* 2004] but has not gain wide use.

The qPCR technique has increasingly gained relevance for the selective quantification gene copies in complex samples due to its sensitivity, flexibility, low cost, and extremely high throughput. These characteristics make it a valuable tool for characterizing aquatic microbial communities [Fierer *et al.,* 2005; Blackwood *et al.,* 2005; Smith & Osborn, 2008]. However, as for FISH, the accuracy of the technique largely depends on the design of appropriate primers or probes, as well as the optimization of the experimental conditions. A number of reasons can affect the accuracy of the technique resulting in estimated abundances of the target microbial groups that may not always parallel true percentages of these groups in complex, marine samples. Among them, DNA extraction bias may alter the estimated abundances of certain groups [von Witntsingerode 1997; Boström *et al.,* 2004]. Heterogeneity in ribosomal operon numbers [Klappenbach 2001; Tourova, 2003; Crosby *et al.,* 2003] may also affect relative estimates of group abundances and, furthermore, some operons can be preferentially amplified [Davis *et al.,* 2003]. Finally, due to the compromises faced in the primer design steps, carried out with a limited subset of the real microbial genomic diversity, the qPCR assays

do not necessarily amplify rDNA genes belonging to all members of each targeted group. This is more problematic for primers matching large phylogenetic groups in contrast to more restricted clade-specific primers [Wagner *et al.,* 2007], and can be lessened, to some extent, by accurate knowledge of the genetic diversity existing in the group to be studied.

In this work, the design of different Bacteroidetes clade-specific qPCR primers was produced after a major effort in the generation of genetic data representing the communities of Bacteroidetes, their phylogenetic relationships, as well as their association to environmental variables (chapter 4). However, due to the limited intragroup variability found in the rDNA genes, some difficulties were encountered in the specificity and coverage of the primers targeting wide groups. In some instances, it was not possible to design primers for the whole taxonomic cluster as delineated in Silva's database, and we limited the design to relevant sub-clusters that included the maximum number of the sequences we had retrieved from the Mediterranean. Thus, although a primer pair specific for the NS2b marine clade was designed covering the entire group (248 sequences), other primers had to be designed to target much smaller clusters within the entire group (e.g. *Winogradskiella* primers matched 23 sequences out of the 101 sequences in the whole group).

In any case, as discussed above, it is important to remark that any designed probe or primer will be influenced by the diversity of the sequences available at the time of the design. Even though the current work was performed after the generation of extensive libraries representing the diversity of Bacteroidetes present in the studied area (see chapter 4), periodic revisiting of *in silico* evaluation and reassignment of the primer coverage and specificity is continuously needed [Díez-Vives *et al.,* 2012].

The *in silico* results showed a high potential of the primers in terms of specificity and coverage, thus setting a framework for the quantitative studies of the main Bacteroidetes subclades present in the area of study. In order to evaluate the usability of the approach and the practical limitations of the technique, we optimized and tested in-depth the performance of a primer pair specific for the NS2b marine clade. This set was chosen because this Bacteroidetes clade included the largest number of sequences in our libraries (chapter 4), and, most likely, it represents a diverse and abundant genotype in the marine environment of the area of study. Moreover, this clade also contains many sequences from worldwide marine waters, representing a cosmopolitan group. Quantification results with this primer pair showed an accurate and specific amplification

of clones belonging to the group. No apparent biases for different target templates were noticed as demonstrated by similar slopes of standard curves with different target clones and by a precise quantification of samples in which different target clones had been mixed.

Extensive testing of this primer set demonstrated almost negligible amplification in a few non-target clones, which could also be traced in the analysis of melting curves. Furthermore, mixing of target and non-target clones revealed that this false amplification was present only when target sequences were scarce, and it disappeared with a ratio of only 1% target/non-target sequences present in the reactions. Finally, the presence of PCR-inhibitory substances is a well-known issue that can affect the accuracy of quantifications by qPCR in environmental samples, or even completely hamper it [Green & Field, 2012]. This often results in the need to use dilutions of the templates, which can by itself compromise quantification of the most rare templates in complex matrixes. In our hands, similar efficiencies were achieved with either plasmidic templates or environmental samples diluted >1:10, and we observed a reliably quantification of mixtures with both templates. In a limited-scale test with environmental samples, the qPCR results were encouraging allowing accurate and linear quantification of NS2b genotypes in at least two 10-fold dilutions, showing differences among samples which were, in general, well above the technique thresholds.

We did not attempt to compare these preliminary results with other quantifications approaches such as FISH, due to the large differences in methodology and the lack of Bacteroidetes subgroup-specific FISH probes for the NS2b marine clade. Comparison of qPCR with non-amplificative techniques, or even with results of the same technique run with different workflows and/or in other laboratories can be extremely problematic [Smith *et al.,* 2006].

It must be stressed that the qPCR technique provides information on the amount of 16S rDNA gene copies and that, due to the variability in rDNA operon copy number in different bacteria, these inferred numbers cannot be directly transformed into numbers of cells [Klappenbach, 2001; Smith & Osborn, 2009]. Moreover, the qPCR protocol may be affected by several variables such as the initial extraction of nucleic acids, the preparation and amplification of the standard curve template [Love at al., 2006], and also the presence of PCR inhibitors [Stults *et al.,* 2001]. Therefore it has been recommended that direct comparison of absolute gene numbers should be used only for targets determined within a single qPCR assay using the same standard curve [Smith *et al.,* 2006].

We consider that the comparison of gene copy numbers for a given sample and laboratory workflow (i.e., implying the same inhibitors and extraction method), using a panel of different primer sets should be reliable if the individual reactions are similarly optimized and they reach similar efficiencies. Thus, our goal was to develop a framework to quantitatively analyze samples in a hierarchical way, first by enumerating the whole bacterioplankton and Bacteroidetes communities, and subsequently smaller levels (e.g. NS2b marine clade followed by NS2b_deep Clade I). These results would determine the percentage contribution of each group to the phylum and to the whole bacterioplankton. Similar approaches have been already applied for the quantification of Bacteroidetes relative to the entire bacterioplankton [Fandino *et al.,* 2005; Abell & Bowman, 2005b, Eilers *et al.,* 2007]. The effort done here in designing specific primers and the first attempts to test their performance in qPCR was designed to provide the tools to be applied in the analysis of our temporal series and spatial transects in the BBMO but can likely be applied to other environmental samples.

Although recently developed high-throughput sequencing (HTS) technologies [Pareek *et al.,* 2011; Logares *et al.,* 2012; Edwards *et al.,* 2006] are increasingly gaining relevance in the field of microbiology due to the huge amount of genetic data that they provide with little effort, the ability of the PCR to specifically target particular taxonomic groups at almost any phylogenetic levels still makes this technique an invaluable method in the molecular microbial ecologist's toolbox.

Acknowledgments

We thank to Dra. Cecilia Alonso for hosting CDV at her laboratory in Montevideo and for her valuable help with primer design in ARB. This work was supported by grant MICRODIVERSITY (CGL2008-00762/BOS) and MicroOcean PANGENOMICS (CGL2011-26848/BOS) to SGA and STORM (CTM2009-09352/MAR). Financial support was provided by Ph.D. JAE-Predoc fellowships from CSIC to CDV. Also, SGA was supported by a RyC contract Spanish Government of Science and Innovation (MICINN) and FP7-OCEAN-2011 (MicroB3) from the EU.

Synthesis of Results and General Discussion

The main goal of this thesis was the study of the diversity, structure and distribution of marine Bacteroidetes with the use of various molecular-based approaches. Different levels of analysis have been considered: (i) determination of community composition and structure using DGGE fingerprinting techniques; (ii) fluorescent in situ hybridization (FISH); (iii) fine-scale genetic diversity studies including 16S and ITS ribosomal gene phylogenies, searching for niche-specific sub-clades, likely to be ecotypes; and (iv) design of clade-specific primers and qPCR tests for an accurate quantification of these ecotypes. In the following discussion, the results of the different analyses will be evaluated and combined with other studies to gain insight into the diversity and distribution of Bacteroidetes communities.

1. Warming up… improving the tools

Before starting a detailed study about a specific group of bacteria, a review and evaluation of the tools available for this purpose should be carried out. We conducted such revision in parallel for different molecular approaches as detailed in chapters 1 and 2. First, we performed a literature survey of all the DNA-based Bacteroidetes primers and probes published to date (table 1, chapter 2). The 8 most commonly used probes were analyzed thoroughly *in silico*. Two of these probes were experimentally tested and compared using CARD-FISH, and 6 of them were further evaluated for their usefulness in PCR-DGGE studies. The *in silico* analysis revealed that coverage and specificity of the oligonucleotides did not change dramatically over time when using the same database, despite the exponential growth of available sequences in recent years. However, notable discrepancies were noticed when comparing different databases, mainly caused by the different taxonomical schemes assumed in each of them (table 2SM, chapter 2). Experimental tests were also conducted with cultured isolates and environmental samples, in order to further validate and verify the predictions of the *in silico* studies.

1.1. What are the best probes for FISH?

I will discuss here the results of the tests of coverage and specificity of two of the main Bacteroidetes probes used in this work and elsewhere: CF319a and CF968R. Probe CF319a [Manz *et al.,* 1996] was reviewed because it is the most common FISH probe used

for quantitative analysis in the literature for over a decade. This notwithstanding, several studies have pointed out its low coverage for the group [Weller *et al.,* 2000; O'Sullivan *et al.,* 2002; Chen *et al.,* 2006; Amann & Fuchs, 2008]. As seen in our results (Table 1, chapter 2), this primer presented high coverage within class Flavobacteria (92.7%), but low (34%) coverage of classes Sphingobacteria and Cytophagia in the Silva updated database (104 NR; [Pruesse *et al.,* 2007]). On the other hand, primer CF968 [Chen *et al.,* 2006] was chosen and optimized as a FISH probe because it displayed the highest coverage for the whole phylum (96.8%) and for class Flavobacteria (96.2%) (Table 1, chapter 2). Moreover, it presented good coverage of classes Sphingobacteria and Cytophagia (90 and 84%, respectively).

We performed different hybridization tests of these probes with cultured isolates and with environmental samples. In 70 % of the samples taken along a spatial transect, probe CF319a systematically resulted in higher number of labeled cells than probe CF968 (Fig. 4, chapter 1). In the temporal study, counts with probe CF319a probe were also higher than counts with CF968 in the period from August to January (Fig. 6, chapter 1). This discrepancy was assigned to the probes targeting different phylotypes, some of them more abundant, or to non-specific hybridization of the probes.

To further explore the reasons behind these differences, we analyzed the 16S rRNA sequences retrieved from 27 metagenomes from the Mediterranean Sea, Red Sea and North Indian Ocean obtained in the TARA expedition [Karsenti *et al.,* 2011], looking for the target sequences of both tested probes. The RDP taxonomic affiliation of such 16S fragment sequences (named miTAGs; [Logares *et al.,* 2013]) extracted from the metagenomes revealed that probe CF319a matched 22 different genera, 9 of them exclusively, whereas probe CF968 targeted a total of 25 genera, 12 exclusively. Thus, only 38% of the genera were matched by both probes, which also showed different coverage of these genera (Fig. 7, chapter 1).

These findings confirmed differences in the targeting of CF319a and CF968 probes but, in contrast with the FISH results, probe CF319a covered a lower genetic diversity (expressed as number of matched genera) than probe CF968. This would suggest that the higher FISH counts obtained with CF319a might be due to non-specific hybridization of the probe, among other possibilities. Oligonucleotide CF319a matched many outgroups (1,902 seq.) in the *in silico* analysis, most of them within Phylum Crenarchaeota (Table 1, chapter 2). The low specificity of this primer was reported previously by Kirchman *et al.* [2003]. Oligonucleotide CF968 also presented many outgroup hits (2,863 seq.)

among different phyla, mainly Spirochaeta (Table 1, chapter 2). Both oligonucleotides also presented matches among orders Rhodobacterales (in the Alphaproteobacteria division) and order Alteromonadales (within the Gammaproteobacteria division). Upon closer inspection of the sequences at the target region, the oligonucleotides displayed two or three mismatches in the binding region with these cross-reacting organisms (Fig. 1SM, chapter 2). However, mismatches of the CF968 were located close to the 3' end of the primers, which prevented amplification [Petruska *et al.,* 1988] and this would thus explain the higher specificity of this probe.

In vitro results of PCR tests with cultured isolates showed that primer CF319a amplified most of the Alpha and Gammaproteobacteria isolates, whilst CF968 did not amplify any (Table 3SM, chapter 2). To further corroborate this striking result we also tested both oligonucleotides as FISH probes (Fig. 3, chapter 2) with four isolates belonging to the cross-reacting groups (i.e. Alphaproteobacteria and Gammaproteobacteria). All cultures were labeled with CF319a and no positive hybridizations were observed for CF968.

These two groups of cross-reacting bacteria are ubiquitous in marine environments; The Roseobacter clade (within family Rhodobacteraceae) may account for up to 20% of total bacterioplankton in coastal areas [Buchan *et al.,* 2005] and Alteromonadales can also be locally abundant in marine communities [Acinas *et al.,* 1999; Alonso-Sáez *et al.,* 2007]. In fact, they were found to represent relevant components of bacterioplankton in our coastal area of study (Fig. 2SM, chapter 1). Not surprisingly, the highest discrepancy between counts conducted with both probes in the temporal series (6%) was coincident with a peak of of *Alteromonas* and *Roseobacter*, supporting the idea that the presence and relative abundance of these phyla is a major cause of the differences observed between the probes in enumerating Bacteroidetes.

1.2. Best primers for DGGE analyses.

Primer CF968 seemed to be the best candidate for Bacteroidetes analyses. However attempts to use it in the DGGE technique failed to resolve clear banding patterns (Fig. 2SM & 3SM, chapter 2). For this reason, we designed a new primer, CF418, that presented good coverage of classes Flavobacteria and Sphingobacteria (53.8 and 35.4% respectively) and very low number of outgroup hits (only 210 sequences) (Table 1, chapter 2). This primer passed all the tests with cultured isolates (Table 3SM, chapter 2)

and yielded clear and diverse banding patterns when used in DGGE (Fig. 4SM, chapter 2). We also tested primers CFB555 and CFB560 as candidates for DGGE analysis obtaining good banding patterns and specificity. However, these primers contained degeneracies, which might introduce unexpected biases, and primer CFB560 needed a nested PCR protocol increasing the potential of chimera formation and amplification errors, and also increasing cost and labor time.

Our comparative results of characteristics and performance in PCR-DGGE of these primers (Table 2, chapter 2) indicated that the new CF418 was the best choice for DGGE-based fingerprinting analyses and we corroborated it with a one-year temporal series study in which primers CF319a and CF418 were compared. Although both primers provided diverse and sharp band patterns reflecting different communities from winter to summer, as expected from the previous results, the sequencing of the bands obtained with primer CF319a confirmed that 10 of 12 bands belonged to non-Bacteroidetes genotypes. These results discourage the use of primer CF319a for Bacteroidetes-specific dynamic analysis despite its apparent good DGGE resolution. The new primer CF418 matched 16 different Bacteroidetes taxa confirming its high specificity and capability to retrieve a high diversity of Bacteroiretes sequences (Table 4SM, chapter 2).

2. Application to samples collected in spatial transects and temporal series

We used the improved molecular tools in temporal and spatial studies to quantify abundances of Bacteroidetes (using probe CF968 in FISH) and to analyze their community composition (using primer CF418 in DGGEs). The goal of these studies was to find correlations among abundances and environmental parameters that might explain their distributions, and to identify the presence of different phylotypes in the community associated to different seasons or locations.

The spatial samples of the studies presented in chapters 3 and 4 were taken along an inshore-offshore spatial transect in the NW Mediterranean Sea during the Modivus cruise. Vertical profiles were also sampled, down to 2,000 meters in the open ocean (Fig. 1SM, chapter 3). In addition, temporal analyses were carried out in a coastal surface site where a permanent observatory has been implemented (the BBMO, [Gasol *et al.,* 2012]). This variety of samples represents a range of temperatures and extensive gradients in Chl *a* concentrations, salinity, and other physical and biological factors. Thus, it was an interesting testing ground for hypotheses on the effect of environmental factors in the abundance and composition of the different Bacteroidetes subclades.

2.1 Temporal dynamics of Bacteroidetes

In the temporal analysis, combining the results of FISH quantifications from chapter 2 and 3 (Fig. 1), we observed higher abundances of Bacteroidetes in the spring and early summer (April-June) over the 3 years studied, as reported from other marine habitats [Pinhassi & Hagström, 2000; Eilers *et al.,* 2001; Mary *et al.,* 2006; Kan *et al.,* 2006; Alderkamp *et al.,* 2006].

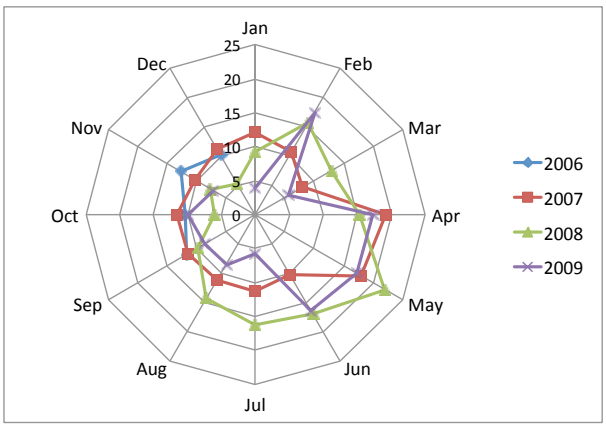


Figure 1. Percentage of Bacteroidetes over total DAPI counts in CARD-FISH analyses with probe CF968 over three consecutive years in the temporal studies at the BBMO.

Correlation of these abundance values with environmental parameters was restricted to two years in chapter 3 (Sept 2006 to Sept 2008), which were also studied at a community composition level. The contribution of Bacteroidetes was variable over time accounting from 8 to 22% of DAPI counts, a range typically reported in coastal environments [Cottrell & Kirchman, 2000a; Alonso-Sáez *et al.,* 2007; Alonso *et al.,* 2007]. A model explaining 57.5% of this variability included only 3 variables: day length (+), total bacterial abundance (-) and NH_4 (-). Correlation with Chl *a* concentration was not detected, although other authors had reported associations between Bacteroidetes abundance and algal blooms [Simon *et al.,* 1999; Kirchman, 2002; O'Sullivan *et al.,* 2004; Pinhassi *et al.,* 2004; Abell & Bowman, 2005b].

This is a controversial topic, because different observations, some difficult to conciliate, have been reported. The response of Bacteroidetes to the blooms can occur coinciding with the Chl *a* peak or during the detrital phase of the bloom [Alderkamp *et al.,* 2006]. This may depend on the composition of the blooms [Fandino *et al.,* 2001] but it has also been shown that different Bacteroidetes genera may display different responses

[Teeling *et al.,* 2012]. Moreover, some Bacteroidetes phylotypes seem to dominate also in conditions when blooms do not occur [Riemann *et al.,* 2000; Pinhassi *et al.,* 2004; West *et al.,* 2008]. A more detailed study considering different clades within the Bacteroidetes could reveal which ones are responding primarily to the phytoplankton blooms. Another possibility for the lack of correlation is that shifts in abundance and composition during phytoplankton blooms can occur in the order of a few days [Fandino *et al.,* 2005]. Hence, detecting the changes of this fast response of the Bacteroidetes would require a more intensive sampling than the frequency used in our study.

Changes in the abundance of Bacteroidetes were not synchronized with changes in their community structure, which usually shifts one or two months later than the major changes in abundance. Bacteroidetes communities were grouped in two clusters with markedly different communities (Winter-Spring and Summer-Fall). In the phylogenetic analyses of sequences obtained from DGGE bands, the distinct phylotypes belonging to these two different seasonal clusters grouped together in the same phylogenetic clades (Fig. 6, chapter 3).

2.2 Spatial distribution of Bacteroidetes

Comparing samples from the Mediterranean Sea, Red Sea, North Indian Ocean, Arctic and Antarctica, the highest abundance of Bacteroidetes were found in Antarctic waters (up to 46% of DAPI counts). This high abundance in cold and nutrient rich waters is well known from previous studies [Simon *et al.,* 1999; Brinkmeyer *et al.,* 2003; Kirchman *et al.,* 2003; Abell & Bowman, 2005b; Gómez-Pereira *et al.,* 2010, Schattenhofer *et al.,* 2013]. Moreover, these particular peaks in the samples analyzed were associated with a *Thalassiosira* (diatom) bloom [Ruiz-González *et al.,* 2012c]. High abundance of marine Bacteroidetes in cold waters associated to algal blooms has recurrently been reported in Polar regions, with values up to 72%, during a Phaeocystis bloom measured with FISH [Simon *et al.,* 1999; Ruiz-González *et al.,* 2012c] and between 17-30% estimated with qPCR [Abell & Bowman, 2005b].

Focusing in the Mediterranean Sea, a study in a transect from the coast to offshore and down to 2,000 meters is reported in chapter 3. Surface abundance was uniform over the 100 km coast-offshore transect (Fig. 1B, chapter 3). While some studies have suggested that Bacteroidetes diversity at the coast is higher than in the open sea [Eilers *et al.,* 2001; Kirchman *et al.,* 2003; Alonso *et al.,* 2007], other results have shown that certain clades of Bacteroidetes are equally abundant in coastal waters and in the

open sea [Gómez-Pereira, 2010]. Drastic changes in Bacteroidetes abundance occurred with depth (from contributions of about 15% in surface and DCM waters down to 2.5% at 150 meters). Such depth-related changes are similar to those observed for the bacterioplankton as a whole [Lee & Fuhrman, 1991; Acinas *et al.,* 1999; Moeseneder *et al.,* 2001; Ghiglione *et al.,* 2005; Pommier *et al.,* 2010]. However, the diversity of Bacteroidetes increased with depth despite their abundance decrease. This pattern has also been reported for total bacterioplankton, and it was assigned to the idea that more stable ecosystems would present a higher bacterial diversity [Pommier *et al.,* 2010]. Salinity was the best predictor for Bacteroidetes abundance distributions in the spatial study as previously found elsewhere [Suzuki *et al.,* 2001b; Wells & Deming, 2003]. Moreover a positive correlation with temperature was also observed contrasting with previous reports [Abell & Bowman, 2005b; Gómez-Pereira *et al.,* 2010]. Bacteroidetes are usually more abundant in cold waters, but sequences from the Mediterranean might belong to a subgroup of warm-water phylotypes responding positively to higher temperatures, as it was suggested for warm-water phylotypes of genus *Polaribacter* in the North Atlantic waters [Gómez-Pereira *et al.,* 2010]. The highest Bacteroidetes counts were also found in surface waters at the most temperate province (Northem Atlantic Drift) of an Atlantic Ocean transect [Schattenhofer *et al.,* 2009]. The community structure was rather similar over the superficial transect, but it changed clearly from shallow to deep waters (Fig. 3, chapter 3). However, the phylotypes dominating at different depths were not phylogenetically different at the resolution level used in chapter 3 (Fig. 6, chapter 3).

In summary, although Bacteroidetes seem to prefer cold and nutrient-rich waters, we observed that they are present in high abundance in surface warm waters as well (around 15% of total counts). Bacteroidetes follow seasonal dynamics and spatial patterns, as does the bulk bacterioplankton, and different environmental factors seem to influence their spatial and temporal dynamics. Other factors not considered in this study might also influence their population sizes. For instance, picoplankton abundance can be tightly controlled by protistan grazing and viral lysis [Pernthaler, 2005], mechanisms that are considered “top–down” controls. Moreover, protists might have differential prey preferences [Massana *et al.,* 2009], therefore differentially affecting the population sizes of different picoplankton clades [Pernthaler *et al.,* 2004; Massana *et al.,* 2009; Ferrera *et al.,* 2011]. Knowledge about species-specific grazing on different Bacteroidetes clades is not available, and studies addressing this topic should also be considered in order to understand their impact over the Bacteroidetes populations and their distribution.

3. A Fine-scale phylogenetic tree of the marine Bacteroidetes.

In the phylogenetic study carried out in chapter 3 with sequences from DGGE bands, most Bacteroidetes clades (covering from 93.9 to 99.5 % sequence identity; Table S5, chapter 3) encompassed phylotypes from contrasting environments (i.e. surface and deep, or winter and summer), suggesting that the Bacteroidetes community was rather similar among these habitats. However, sequences from DGGE bands may not provide enough resolution to detect habitat-specific clusters of several closely related sequences because only the most abundant taxa can be retrieved and the sequences are usually short (<500 bp). For that reason, we extended the resolution of the study by building 4 clone libraries, representing the diversity present in heterogeneous spatial and temporal samples (chapter 4). We constructed two libraries from opposite sites of our spatial transects, i.e. a "Surface" library (SFL, using a sample from a coastal surface point) and a "Deep" library (DPL, using a sample taken at 2,000 meters depth). In the temporal study (chapter 3), we focused in the community structure as resolved by the DGGE analysis of the monthly time-series. Hence, we built a "Summer" library (SML, including samples from 9 months from the Summer-Fall season) and a "Winter" library (WTL, including 14 samples from the Winter-Spring season). All four gene libraries contained the nearly complete 16S, the ITS and partial 23S rRNA genes.

The goal of this chapter was to look for niche partitioning among sequence types (or putative ecotypes) using fine-scale phylogenetic trees. This would allow us to test whether genetic variation in ribosomal gene clusters is correlated with long-term evolutionary divergence. Ecotypes are closely related groups of bacteria that differ in physiological details that determine their niche specificity [Cohan, 2006; Cohan & Perry, 2007]. The ecotype concept applied relies in the interpretation of phylogenetic structure in environmental data (i.e. ecotypes as genetic variants that can be distinguished phylogenetically and by their differing spatiotemporal distributions in the environment).

3.1. 16S rRNA gene phylogenies.

Our data from these Mediterranean Bacteroidetes clone libraries revealed that the vast majority of sequences were phylogenetically affiliated with class Flavobacteria (as previously observed within the 27 genomes analyzed in chapter 1), thus providing additional evidence that Flavobacteria is the dominant Bacteroidetes class in marine bacterioplankton [e.g. Kirchman *et al.,* 2003; Alonso *et al.,* 2007; Chen *et al.,* 2008]. Other

sequences were affiliated with the classes Sphingobacteria, Cytophagia and Bacteroida. Most sequences (76 %) clustered with uncultured groups, mainly defined after samples from the North Atlantic Ocean and the North Sea (incorporated in the Silva taxonomy as "NS marine clades") (Fig. 2).

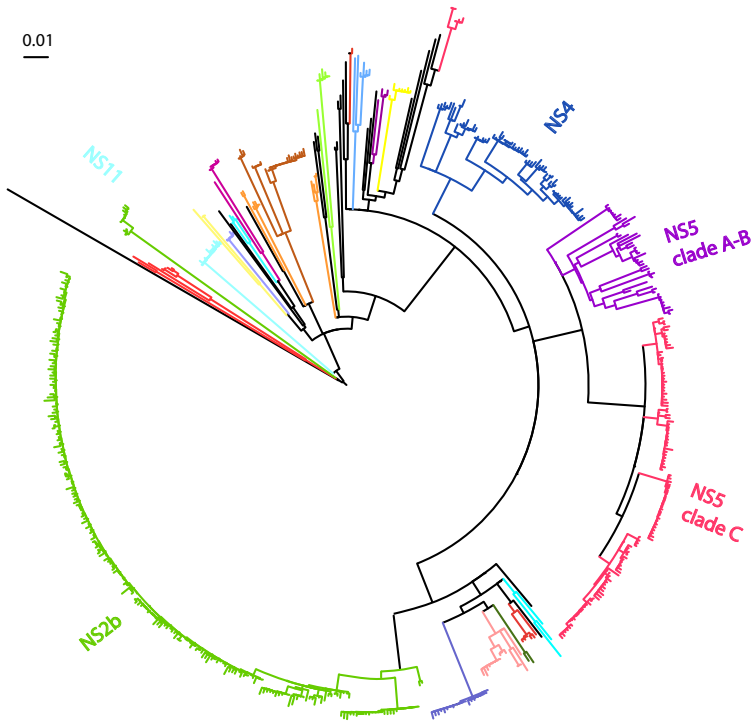


Figure 2. Maximum Likelihood phylogenetic tree inferred with 646 16S rRNA sequences from the environmental libraries. The main groups belonging to the North Sea "NS marine clades" are named.

Sequences from the DGGE bands (chapter 2) and clone libraries (chapter 3) were associated to similar cultured and uncultured groups, though both techniques were performed with different primer sets (CF418F/907RM and CF319a/CF434R respectively). Shared cultured groups were *Fluviicola*, *Owenweeksia* and *Polaribacter*. Among the uncultured groups, both studies retrieved phylotypes affiliated with NS2b, NS5 and NS4 clades [Alonso *et al.,* 2007]. NS9 and Saprospiraceae were common among the DGGE bands but absent in the clone libraries. The forward primer used for the construction of gene libraries (CF319a) presented perfect match with these groups, but the limited amount of data available for 23S rRNA genes did not allow to thoroughly examine the matching of the reverse primer (CF434R), which could present a bias against these groups.

Analyses of the high-resolution phylogenetic tree built with nearly full-length 646 16S rRNA gene sequences indicated that more than 60% of the sequences fall into discrete clusters of closely related organism, as it had been shown for bacterial communities as a whole [Palys *et al.,* 1997; Casamayor *et al.,* 2002; Acinas *et al.,* 2004a]. Some of these clusters were associated to specific environments, e.g., deep-specific clusters were clearly separated from surface-specific clades, sometimes an entire class was present in one of these environments (i.e. *Winogradskiella* exclusively from DPL). However, seasonal-specific clades (winter *vs.* summer) were restricted to smaller clusters and sometimes they were intermixed in the cladograms, not being resolved as monophyletic groups with solid bootstrap support (Fig. 5, chapter 4).

Analyses of the phylogenetic structure of these communities (i.e. using the Unifrac tool) confirmed that the deep environment presented a distinct microbial community structure compared to any other environment (both in presence/absence and in relative abundances of OTUs of the phylogenetic lineages). Differences among surface environments (SFL, WTL and SML) were only detected when considering the total number of sequences in each season, but not by the presence of specific phylotypes. When using the 16S rRNA gene, it seems that final associations of the microdiverse clusters in terms of taxa present is more influenced by depth-related factors than seasonal factors.

Few studies have observed habitat-specific Bacteroidetes clades. One such example was reported in the eastern Mediterranean Sea, where some depth-specific Bacteroidetes clades were encountered [Blümel *et al.,* 2007]. Moreover, two ecotypes have also been suggested for *Polaribacter* sequences: one "temperate" cluster with sequences associated to *P. dockdonensis* strains, originally isolated from a coastal temperate marine environment and presenting optimal growth temperatures between 25 - 28°C. A second *Polaribacter* ecotype includes genotypes associated with strains *P. irgensii*, *P. filamentus* and *P. franzmannii*, which have been isolated from waters in Antarctic and Arctic seas [Gosink *et al.,* 1998] and constitute "polar" clades [Gómez-Pereira, 2010].

3.2 ITS region variability.

In some cases, the variability of the 16S rRNA sequences has not been sufficient for establishing species identity and thus, was not appropriate for studies at the intraspecific level [Fox *et al.,* 1992]. As a result, researchers have increasingly turned to the more

variable ITS region. High diversity has been revealed studying the ITS region of specific bacterial strains [Milyutina *et al.,* 2004; Jaspers & Overmann, 2004; Osorio *et al.,* 2005; Acinas *et al.,* 2009].

This work (chapter 4) constitutes the first extensive analysis of the ITS region for Bacteroidetes from environmental samples. Interesting results were the high variability in ITS length detected among our sequences (ranging from 93 to 739 bps), which was larger than previously reported for some cultured strains of Bacteroidetes [Kuwahara *et al.,* 2001; Steward & Cavanaugh, 2007]. However, most of the sequences ranged between 400-600 bps. The largest differences in the ITS length was found among sequences from deep waters, thus corroborating the high diversity present in this environment. The tRNA composition of the ITS region was dominated by tRNA - Ile+Ala, but ITS regions with only one tRNA, or none at all, were also found in all groups except in class Sphingobacteria and in family Flavobacteriaceae (Table 6, chapter 4). To study the variability of the ITS region compared to the 16S rRNA gene, 61 OTUs defined at >99% 16S rRNA similarity were analyzed (table 8, chapter 4). Twenty-three of them showed < 99s% similarity in the ITS region, indicating higher variability at this locus. Of these, nine OTUs presented very low sequence similarity (from 34% to 93.7%), with length differences >40 bps. The low similarity was mostly due to the presence of indels (insertion or the deletion of bases in the DNA), hence increasing when gaps were excluded from the analysis (e.g. from 34% to 86.6% in OTU# 45; Table 8, chapter 4).

These instances with almost identical 16S rRNA sequence but different ITS region, could respond to two different scenarios: (i) they could reflect real variability in the ITS region among closely related taxa, but (ii) they could also be due to the variability between paralogous ribosomal genes within the same genome (i.e., among rrn operons). Comparing our ITS sequences with available cultured and uncultured Bacteroidetes genomes (SAGs), we observed that our sequences were similar to the ITS region of different operons within a genome. This suggests that at least in some cases, differences in ITS sequences of closely related 16S rRNA sequences may correspond to different operons of the same organism rather than to different discrete organisms. However, more genomes are needed to extract patterns of correspondence of unique ITS sequence to discrete organisms. In any case, only a short fraction of our sequences from environmental samples affiliated primarily with cultured isolates, not affecting the ITS-based phylogenetic analyses of the main groups (i.e. NS2b, NS4, NS5 and *Formosa*, chapter 4).

An additional consideration in the interpretation of these results is that sequences sharing >99% similarity in the 16S rRNA gene may also represent genetically and physiologically distinct populations adapted to different niches. Ecologically relevant phenotype differentiation has been observed in organisms with identical or nearly identical 16S rRNA gene sequences [Ward *et al.,* 1998; Jasper & Overman, 2004]. Similarly, significant genome diversification (and hence potentially phenotypic diversification) has been widely reported for closely related ribosomal seuqences [Keswani & Whitman, 2001; Bejà *et al.,* 2002, Tyson *et al.,* 2004; Jasper & Overman, 2004; Thompson *et al.,* 2005].

3.3 ITS-based phylogenies.

The 16S rRNA phylogenies revealed several depth-specific clades (i.e. Deep or Surface), but seasonal-specific clades (i.e. Winter or Summer in surface water) were less obvious. For instance, most of the sequences of the NS2b marine clade belonged to surface samples with high similarity in the 16S rRNA (>99.2%). Hence, the low phylogenetic signal did not allow resolving seasonal-subclades with reliable bootstrap support. In other groups (e.g. *Cyanobacteria* and SAR11), the ITS region analysis has revealed specific ecotypes not identified before using the 16S rRNA [Rocap *et al.,* 2002; Ferris *et al.,* 2003; Acinas *et al.,* 2009; Ngugi *et al.,* 2012]. The ITS has an evolutionary rate 10 times faster than that of the 16S rRNA gene [Leblond-Bourget *et al.,* 1996], and therefore can give insights into diversity at a higher resolution [Woebken *et al.,* 2008].

We extended the 16S rRNA phylogenetic analyses by constructing ITS-based phylogenetic trees for the main groups present in our study area (i.e., clusters with large number of sequences that may represent abundant groups in the environment). Large variations in sequence and length of the ITS region prevented us from building a general phylogenetic tree including all the phylotypes, and it was only possible to align closely related phylotypes. The ITS-based phylogeny showed rather similar topology to the 16S rRNA-based trees, thus confirming the previous habitat-specific clusters described in those, as it has previously been observed for other taxa [García-Martínez & Rodríguez-Valera, 2000]. Moreover, the enhanced resolution of this region partitioned several new seasonal-specific clusters among the closely related NS2b marine clade. The use of the ITS revealed large variability among these sequences, delineating 4 specific-clades of summer or winter environments (Fig. 9, chapter 4).

4. Quantification of habitat-specific clades

Enumeration of the whole bacterioplankton or of specific subgroups has been generally done with oligonucleotide probes [e.g. Field *et al.,* 1997; Carlson *et al.,* 2009]. A suit of oligonucleotides probes targeting Bacteroidetes in particular habitats were designed by Weller at 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]. Some probes have been designed to match cultured genera e.g., Leweenhoekiella, Gramella, C. marinofava-latercula, *Polaribacter*, *Dockdonia* [Brinkmeyer *et al.,* 2003; Malstrom *et al.,* 2007; Gómez-Pereira *et al.,* 2010] and some uncultured clades e.g., DE clade, AGG58, VISION clades [Kirchman *et al.,* 2003; O´Sullivan *et al.,* 2004; Malmstrom *et al.,* 2007; Gómez-Pereira *et al.,* 2010].

Using the sequences of the 16S rRNA and ITS regions, we have successfully delineated marine Bacteroidetes into some habitat-related clades (chapter 4), which likely represent ecological units. These clades cannot be related to distinct phenotypes to explain the different niches these lineages might occupy. However, examining their distribution and relative abundances under different environmental conditions, may provide understanding of niche partitioning among sequences types, evidencing a "speciation" related to specific factors. The objective of the last chapter of this thesis (chapter 5) was to design a methodological workflow to quantify these relevant Bacteroidetes subclades, which represent putative ecotypes.

4.1. The qPCR approach

One of the limitations of fluorescence in situ hybridization (FISH) is its low sensitivity. 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 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].

Since the pioneering applications of qPCR in microbial ecology, this technique is now widely used to determine the gene abundance present within environmental samples [Becker *et al.,* 2000; Takai & Horikoshi, 2000; Suzuki *et al.,* 2005; Restrepo-Ortiz *et al.,* 2013, Restrepo-Ortiz & Casamayor, 2013]. The specificity of this method is

determined by the design of the primers or labeled probes that allow quantification of broad phylogenetic groups, but also of individual species or strains.

4.2. Clade-specific primers for Bacteroidetes

The 646 sequences from the four Bacteroidetes gene libraries (chapter 4) were inserted in the Silva guide phylogeny tree accompanying the 111NR dataset (Fig. 3) using the ARB software. Fourteen primer sets were designed to specifically amplify major Bacteroidetes clades and some of their subclades (tables 1-9, chapter 5). Primer sets were designed to cover some habitat specific-clades, which included additional sequences from many environments worldwide, but also to cover the higher-level clades containing these clusters. This hierarchical targeting of clades and sub-clades was designed to allow quantification of the relative contribution of smaller habitat-specific clades to the larger taxonomic groups.

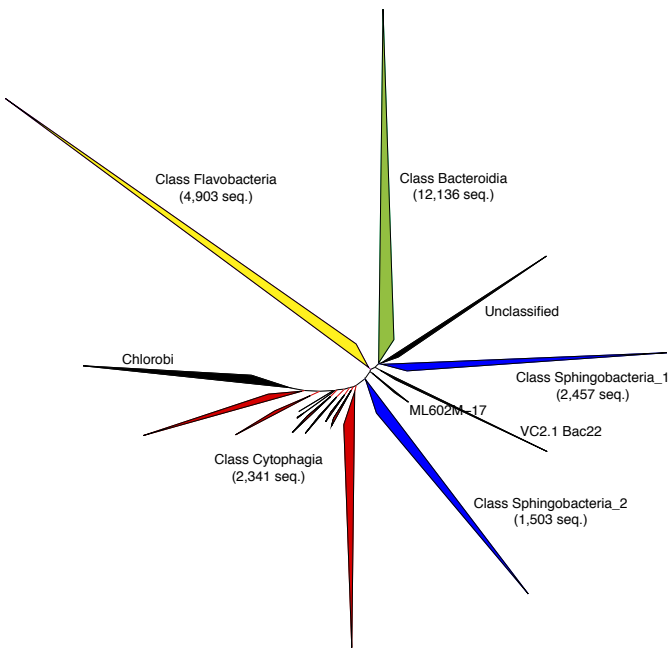


Figure 3. Schematic phylogenetic tree of the 16S rRNA gene sequences of Bacteroidetes (using Chlorobi as outgroup) in the 111NR Silva database [Pruesse et al., 2007]. The color of the clades indicate the four different clases. The total numbers of sequences available in Silva for each class are given. The Mediterranean sequences from our libraries (646 seq. from chapter 4) are not included in these numbers.

For these newly designed primer sets, the *in silico* evaluation of non-target matches was made allowing 1, 2 and 3 mismatches in the binding loci (Tables 1 - 9, chapter 5). Stringent criteria for allowance of matching on non-target sequences, shared by both primers, were followed. For instance, the sharing of non-target sequences was only allowed when both primers presented at least one (ideally more) mismatch with non-targets sequences. Moreover, every possible primer pair was further evaluated to meet the requirements of the qPCR technique and to test their performance, quality and compatibility (Tables 1-9).

We further assessed the *in vitro* performance and specificity of the primer pair designed for the entire NS2b group, as one of the most abundant and diverse subclades. A first round of tests was carried out to evaluate the optimal concentration of primers and the best annealing temperature (T_m) in the qPCR procedure. Later, the coverage and specificity of the primer set was analyzed with a series of amplifications with different target clones, non-target clones and mixes of both types of templates. The results showed a similar efficiency for different target clones, and a specific amplification of the group with negligible amplification of non-target sequences, which disappeared when target sequences were present in the reaction with a ratio of only 1% target/non-target. Furthermore, a preliminary test to address the usability of NS2b primers for quantification on actual environmental samples was carried out and successfully identified important quantitative differences among some of the samples. Efficiencies of the reactions were comparable with cloned sequences or environmental samples as templates. Quantification of the natural samples was achieved even with high dilutions (1:1000) and reliable results were obtained for as few as 5 copies per reaction, demonstrating the high sensibility of this technique.

Summarizing, the primers designed in this study and thoroughly evaluated *in silico*, and the preliminary results of the developing and testing of their application protocols are encouraging for the use of quantitative PCR in the enumeration of some major clade-specific Bacteroidetes constituting the populations present in different niches at least in the studied area and probably elsewhere in the sea. The study of the different distributions of these clades is necessary to understand the phylogenetic levels at which Bacteroidetes should be studied in future research, and to examine their distribution and relative abundances under different environmental conditions. Such studies will ultimately provide a better understanding of the forces affecting the evolution and population dynamics of Bacteroidetes.

5. Perspectives of genomic approaches for the study of marine microbiology

New methodologies, including the so-called high-throughput sequencing (HTS) [Pareek *et al.,* 2011; Logares *et al.,* 2012] are lately replacing Sanger sequencing techniques [Sanger *et al.,* 1977] as the most common tool to explore the diversity and structure of microbial communities. Advantages of these methodologies are their ability to produce vast amounts of data and their speed and relatively affordable cost, which explain why they have become popular in recent years. 454/Roche pyrosequencing technology and the Illumina system are the most common sequencing platforms and they both can be applied in microbial ecology as (i) amplicon TAGs (sequencing of libraries of a PCR-amplified gene of interest), and (ii) metagenomic (sequencing of all genes from all the members of the sampled community).

The most common approach for investigating environmental prokaryotic diversity has been the amplicon TAGs (i.e., the sequencing of 16S rRNA amplicons). However, this approach is affected by PCR-related artifacts (such as primer mismatches, heteroduplex related to long cycles, DNA polymerase errors, differential amplification of complex templates, chimeric amplicons, etc), which can distort the estimations of richness and evenness in microbial communities [Acinas *et al.,* 2005; Hong *et al.,* 2009; Engelbrektson *et al.,* 2010; Ahn *et al.,* 2012; Pinto & Raskin 2012].

In the course of other studies not included in this work, this methodology has been used with some of the samples used here. For instance, some samples enumerated by FISH with probe CF968 in the spatial study (chapter 3) were also enumerated by 454 pyrosequencing in the framework of the ICoMM (international Census of Marine Microbes, http://icomm.mbl,edu) program. Comparison of both approaches shows some tendencies: FISH tends to enumerate more Bacteroidetes than 454 pyrosequencing, but a general trend showing their relatively higher contribution in surface waters rather than in deep waters is captured by both methodologies.

A powerful alternative to amplicon TAGs for investigating the diversity and structure of prokaryotic communities is to identify rRNA fragments from (e.g. Illumina) metagenomic data. By overcoming PCR biases, this technique provides what is believed to be more realistic estimates of community richness and evenness than amplicon TAGs [Logares *et al.,* 2013]. However, the 16S rRNA fragments detected by this approach also undergo amplification steps associated with the sequencing platform (a problem shared with amplicon TAGs), which may also suffer base-composition biases [Aird *et al.,*

2011; Nakamura *et al.,* 2011]. However, a number of protocols and base call algorithms have been developed to minimize such biases and improve the error rate of Illumina sequencing [Harismendy *et al.,* 2009; Aird *et al.,* 2011].

A study such as the one we have performed with almost full sequencing of the 16S rRNA and the full ITS would not be possible nowadays with HTS. The current maximum length of 454 pyrosequencing reads is 500 - 700 bp, but errors accumulate after the first 300 bp [Claesson *et al.,* 2010]. Although Illumina platforms can generate 2–3 billion of shorter (150 bp-200 bp) reads, the identification of ITS regions from HTS data can only proceed on sequence reads including both 16S-ITS fragments, but the length of the tags would rarely allow this. The assembly of these tags into longer contigs can be difficult and prone to errors, given the large variability of the ITS (even within a single 16S genotype) and to the limited ITS databases available. Such short reads (< 150 pb) can be used for rough taxonomic profiling of communities [Hao & Chen, 2012], but are not sufficient for high-resolution taxonomies such as the one conducted in this study using clone libraries with the near entire 16S rRNA gene and the more variable ITS region. As the cost of HTS techniques and the assembly algorithms improve, these approaches will gain more relevance in the near future.

Quantitative PCR methods can provide accurate estimation of bacterial groups at any taxonomic level (e.g. ecotypes) to confirm their association to different niches. Thus, it can complement HTS efforts providing a higher level of resolution, the opportunity to target specific groups, and offering a quantitative layer to the data. Furthermore, the high sensitivity offers exciting opportunities for new developing fields. As an example, this technique allows tracing whether or not taxa become extinct from their habitat, or if a “rare biosphere” of latent taxa is present [Sogin *et al.,* 2006], a possibility known as the seed-bank hypothesis [Lennon & Jones, 2011]. Using HST this hypothesis can only be tested through rather deep sequencing of many samples.

CONCLUSIONS

1. Probe CF319a is not appropriate for the enumeration of Bacteroidetes in fluorescence in situ hybridization (FISH) protocols as we have observed unspecific hybridization with other taxa. CF968 is an excellent alternative for quantifications by FISH.

2. Designing new primers or probes should always include thorough *in silico* and *in vitro* testing. In addition, periodical revisions of the probe coverage are needed to account for the rapidly growing sequence databases.

3. Bacteroidetes populations showed a clear seasonal fluctuation in the NW Mediterranean Sea, with higher abundances in spring and early summer. Day length was the environmental variable that best explained relative abundances.

Changes in Bacteroidetes composition were detected one to two months later than changes in their abundance and two significantly different community structures were observed with seasonal changes during the year (Winter-Spring and Summer-Fall).

4. Bacteroidetes are known to be associated to cold waters, but we also found large abundances in temperate waters, which were furthermore positively correlated with temperature. This suggests the existence of specific warm-water phylotypes in the Mediterranean.

5. An analysis along a spatial gradient showed that Bacteroidetes abundance was rather similar throughout a horizontal surface transect from the coast to offshore waters. Interestingly Bacteroidetes presented higher diversity in deeper waters where their abundance was lower, rather than at surface. Community composition was clearly different in shallow and in deep waters.

6. Different surface communities (i.e. summer and winter communities) presented different phylotypes, but their overall phylogenetic structure was not significantly different. Differences in the seasonal communities were found to affect only the relative abundances of the different phylotypes.

7. Deep water communities were not only genetically different from surface communities, but also their phylogenetic structure differed regarding both presence/absence and abundance. This allowed us to identify depth-specific clusters of closely related phylotypes in 16S rRNA-inferred trees (i.e. deep and surface clusters).

8. We report, for the first time, the large variability within the Bacteroidetes ITS region, in terms of length ranges and on tRNA composition. Furthermore, large sequence divergence was found in the ITS region for almost identical (>99% similarity) 16S rRNA sequences.

9. The ITS region was very useful for the identification of seasonally-defined clusters previously not resolved by the 16S rRNA gene. Several winter or summer-specific clusters could be identified among almost identical 16S rRNA sequences (>99% similarity).

10. In some cases, different ITS sequences associated to similar 16S rRNA genes could be related to different *rrn* operons of available cultured and uncultured (i.e. single-amplified) genomes, highlighting the presence of differing operons within the same individual and some of the associated implications in phylogenetic inference.

11. A battery of primers for quantitavive PCR analyses was designed covering the main Bacteroidetes subclades associated to variable ecological contexts. *In silico* features of the primers and preliminary *in vitro* results, are encouraging for the use of quantitative PCR in the enumeration of clade-specific Bacteroidetes.

12. We have designed a methodological framework to study the abundances and dynamics of specific Bacteroidetes subclades in a hierarchical fashion. Such studies will facilitate the study of ecologically different populations of Bacteroidetes (putative ecotypes).

General References

Abell GCJ, Bowman JP (2005a) Colonization and community dynamics of class *Flavobacteria* on diatom detritus in experimental mesocosms based on Southern Ocean seawater. *FEMS Microbiol Ecol* **53**: 379-391

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

Acinas SG, Rodríguez-Valera F, Pedrós-Alió C (1997) Spatial and temporal variation in marine bacterioplankton diversity as shown by RFLP fingerprinting of PCR amplified 16S rDNA. *FEMS Microbiol ecol* **24**: 27-40

Acinas SG, Antón J, Rodríguez-Valera F (1999) Diversity of free-living and attached bacteria in offshore western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. *Appl Environ Microbiol* **65:** 514–522

Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL, Polz MF (2004a) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**:551–554

Acinas SG, Marcelino L, Klepac-Ceraj V, Polz M (2004b) Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. *J Bacteriol* **168**: 2629-2635

Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **71**: 8966–8969

Acinas SG, Haverkamp T, Huisman J, Stal LS (2009) Phenotypic and genetic diversification of *Pseudanabaena* spp (*Cyanobacteria*). *ISME J* **3**: 31-46

Ahn JH, Kim BY, Song J, Weon HY (2012) Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *J Microbiol* **50**: 1071–1074

Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, *et al.* (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* **12**: R18

Alderkamp AC, Sintes E, Herndl GJ (2006) Abundance and activity of major groups of prokaryotic plankton in the coastal North Sea during spring and summer. *Aquat Microb Ecol* **45**: 237-246

Allewalt JP, Bateson MM, Revsbech NP, Slack K, Ward DM (2006) Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the octopus spring microbial mat community of Yellowstone National Park. *Appl Environ Microbiol* **72**: 544-50

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-Sáez L, Gasol JM, Lefort T, Höfer J, Sommaruga R (2006) Impact of sunlight radiation on bacterial activity and differential sensitivity of natural bacterioplankton groups in NW Mediterranean coastal waters. *Appl Environ Microbiol* **72:** 5806–5813

Alonso-Sáez L, Balagué V, Sà EL, Sánchez O, González JM, Pinhassi J, Massana R, Pernthaler J, Pedrós-Alió C, Gasol JM (2007) Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. *FEMS Microbiol Ecol* **60**: 98-112

Alonso-Sáez L, Vázquez-Domínguez E, Cardelús C, Pinhassi J, Sala MM, Lekumberri I, Balagué V, Vila-Costa M, Unrein F, Massana R, Simó R, Gasol JM (2008) Factors controlling the year-round variability in carbon flux through bacteria in a coastal marine system. *Ecosystems* **11**: 397-409

Alonso-Sáez L, Sánchez O, Gasol JM (2012) Bacterial uptake of low molecular weight organics in the subtropical Atlantic: Are major phylogenetic groups functionally different? *Limnol Oceanogr* **57:** 798–808

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool (BLAST). *J Molec Biol* **215**: 403-410

Amann RI, Zarda B, Stahl DA, Schleifer KH (1992) Identification of individual prokaryotic cells by using enzyme-labeled, rRNA-targeted oligonucleotide probes. *App Environ Microb* **58**: 3007-3011

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

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26**: 32–46

Antoine D, Morel A, André JM (1995) Algal pigment distribution and primary production in the Eastern Mediterranean as derived from Coastal Zone Color Scanner observations, *J Geophys Res* **100**: 16,193-16,209

Antón AI, Martínez-Murcia AJ, Rodríguez-Valera F (1998) Sequence diversity in the 16S-23S intergenic spacer region (ISR) of the rRNA operons in representatives of the *Escherichia coli* ECOR collection. *J Mol Evol* **47**: 62–72

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2006) New Screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl Environ Microbiol* **72**: 5734-5741

Azam F, Fenchel T, Field JG, Gray JS, Meyerreil LA, Thingstad F (1983) The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser* **10**: 257–263

Azam F, Malfatti F (2007) Microbial structuring of marine ecosystems. *Nat Rev Micro* **5**: 782–791

Baltar F, Arístegui J, Gasol JM, Hernández-León S, Herndl GJ (2007) Strong coast-

ocean and surface-depth gradients in prokaryotic assemblage structure and activity in a coastal transition zone region. *Aquat Microb Ecol* **50**: 63-74

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

Barberán A, Fernández-Guerra A, Auguet J-C, Galand PE, Casamayor EO (2011) Phylogenetic ecology of widespread uncultured clades of the Kingdom Euryarchaeota. *Mol Ecol* **20**: 1988-1996

Bauer M, Kube M, Teeling H*, et al.* (2006) Whole genome analysis of the marine Bacteroidetes '*Gramella forsetii*' reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8**: 2201-2213

Becker S, Böger P, Oehlmann R, Ernst A (2000) PCR bias in ecological analysis: a case study for quantitative Taq nuclease assays in analyses of microbial communities. *Appl Environ Microbiol* **66**: 4945–4953

Becker S, Fahrbach M, Böger P, Ernst A (2002) Quantitative Tracing, by *Taq* Nuclease Assays, of a *Synechococcus* Ecotype in a Highly Diversified Natural Population. *Appl Environ Microbiol* **68**: 4486

Behrens S, Fuchs BM, Mueller F, Amann R (2003) Is the *in situ* accessibility of the 16S rRNA of *Escherichia coli* for Cy3-labeled oligonucleotide probes predicted by a three-dimensional structure model of the 30S ribosomal subunit*? Appl Environ Microbiol* **69:** 4935–4941

Béjà O, Koonin EV, Aravind L, Taylor LT, Seitz H, *et al.* (2002) Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl Environ Microbiol* **68**: 335–345

Benner R (2002) Chemical composition and reactivity. In Carlson CA & DA Hansell (eds) *Biogeochemistry of marine dissolved organic matter*. Pp. 59-90

Bergey DH, Harrison FC, Breed RS, Hammer BW, Huntoon FM (1923) Bergey's Manual of Determinative Bacteriology, 1st edn. Baltimore: The Williams and Wilkins Co.

Bernardet JF, Nakagawa Y (2006) An introduction to the family Flavobacteriaceae. In *The Prokaryotes*. 7th edn, Ch. 6.3, pp. 455–480. Edited by E. Stackebrandt, A. Balows, H. G. Trüper, M. Dworkin, W. Harder & K.-H. Schleifer. Berlin: Springer.

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

Blackwood CB, Oaks A, Buyer JS (2005). Phylum- and class-specific PCR primers for general microbial community analysis. *Appl Environ Microbiol* **71**: 6193–6198

Blümel M, Süling J, Imhoff JF (2007) Depth-specific distribution of Bacteroidetes in the oligotrophic eastern Mediterranean Sea. *Aquat Microb Ecol* **46**: 209-224

Boström KH, Simu K, Hagström A, Riemann L (2004) Optimization of DNA extraction for quantitative marine bacterioplankton community analysis. *Limnol Oceanogr Meth* **2**: 365-373

Bowman JP (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* Dworkin M, Flakow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds) *The Prokaryotes, third ed., vol 7.* Springer Science Business Media, Inc., New York, USA

Boyer SL, Flechtner VR, Johansen JR (2001) Is the 16S–23S rRNA internal transcribed spacer region a good tool for use in molecular systematics and population genetics? A case study in cyanobacteria. *Mol Biol Evol* **18**: 1057–1069

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 Microbiol* **69**: 6610–6619

Brown MV, Fuhrman JA (2005) Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat Microb Ecol* **41**: 15–23

Brown MV, Schwalbach MS, Hewson I, Fuhrman JA (2005) Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environ Microbiol* **7**: 1466–1479

Bru D, Martin-Laurent F, Philippot L (2008) Quantification of the Detrimental Effect of a Single Primer-Template Mismatch by Real-Time PCR Using the 16S rRNA Gene as an Example. *Appl Environ Microbiol* **74**: 1660-1663

Buchan A, González JM, Moran MA (2005) Overview of the marine *Roseobacter* lineage. *Appl Environ Microbiol* **71:** 5665–5677

Campbell BJ, Yu L, Heidelberg JF, Kirchman DL (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci* USA **108**: 12776–12781

Caporaso JG, Paszkiewicz K, Field D, Knight R, Gilbert JA (2012) The western english channel contains a persistent microbial seed bank. *ISME J* **6**: 1089–1093

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

Casamayor EO, Pedrós-Alió C, Muyzer G, Amann RI (2002) Microheterogeneity in 16S ribosomal DNA-defined bacterial populations from a stratified planktonic environment is related to temporal changes and to ecological adaptations. *App Environ Microb* **68**: 1706-1714

Castle D, Kirchman DL (2004) Composition of estuarine bacterial communities assessed by denaturing gradient gel electrophoresis and fluorescence in situ hybridization. *Limnol Oceanogr Meth* **2**: 303-314

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 (2006) Development and evaluation of specific 16S rDNA primers for marine Cytophaga-Flavobacteria cluster. *Molec Ecol Notes* **6**: 1278-1281

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

Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, O'Toole PW (2010) Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* **38**: e200

Cohan FM (2001) Bacterial species and speciation. *Syst Biol* **50**: 513–524

Cohan FM (2002) What are bacterial species? *Ann Rev Microbiol* **56**: 457-87

Cohan FM (2006). Toward a conceptual and operational union of bacterial systematics, ecology, and evolution. *Proc Roy Soc Lond Series B* **361**: 1985–1996

Cohan FM, Perry EB (2007) A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol* **17**: R373–R386

Cohan FM, Koeppel AF (2008) The origins of ecological diversity in prokaryotes. *Curr Biol* **18**: R1024–R1034

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**: D141-D145

Cottrell MT, Kirchman D (2000a) Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. *App Environ Microb* **66**: 5116-5122

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

Cottrell MT, Waidner LA, Yu L, Kirchman DL (2005) Bacterial diversity of metagenomic and PCR libraries from the Delaware River. *Environ Microbiol* **7**: 1883-1895

Covert JS, Moran MA (2001) Molecular characterization of estuarine bacterial communities that use high- and low- molecular weight fractions of dissolved organic carbon. *Aquat Microb Ecol* **25**: 127–139

Crespo BG, Pommier T, Fernández-Gómez B, Pedrós-Alió C (2013) Particle attached and free living bacterial community composition in the NW Mediterranean Sea studied by pyrosequencing of the 16S rRNA. *Microbiology Open* **2:** 541–552

Crosby LD, Criddle CS (2003) Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. *Biotechniques* **34**: 790–802

Crump BC, Armbrust EV, Baross JA (1999) Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia river, its estuary, and the adjacent coastal ocean. *App Environ Microb* **65**: 3192-3204

Danovaro R, Dinet A, Duineveld G, Tselepides A (1999) Benthic response to particulate fluxes in different trophic environments: A comparison between the Gulf of Lions–Catalan Sea (western-Mediterranean) and the Cretan Sea (eastern-Mediterranean). *Prog Oceanogr* **44**: 287–312

Dar SA, Kuenen JG, Muyzer G (2005) Nested PCR-denaturing gradient gel electrophoresis approach to determine the diversity of sulfate-reducing bacteria in complex microbial communities. *Appl Environ Microbiol* **71**: 2325-2330

Davis RE, Jomantiene R, Kalvelyte A, Dally EL (2003). Differential amplification of sequence heterogeneous ribosomal RNA genes and classification of the 'Fragaria multicipita' phytoplasma. *Microbiol Res* **158:** 229–236

De Gregoris TB, Aldred N, Clare AS, Burgess JG (2011) Improvement of phylum- and class-specific primers for real-time PCR quantification of bacterial taxa. *J Microbiol Meth* **86**: 351-356

Del Giorgio PA, Prairie YT, Bird DF (1997) Coupling between rates of bacterial production and the abundance of metabolically active bacteria in lakes, enumerated using CTC reduction and flow cytometry. *Microb Ecol* **34**:144–154

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

DeLong E, Pace N (2001) Environmental diversity of Bacteria and Archaea. *Syst Biol* **50**: 470–78

Didelot X, Maiden MC (2010) Impact of recombination on bacterial evolution. *Trends Microbiol* **18**: 315-22

Dieffenbach CW, Lowe TM, Dveksler GS (1993) General concepts for PCR primer design. *PCR Meth Appl* 3: S30-7

Díez-Vives C, Gasol JM, Acinas SG (2012) Evaluation of Marine Bacteroidetes-Specific Primers for Microbial Diversity and Dynamics Studies. *Microb Ecol* **64**: 1047-1055

DiMeo CA, Wilbur AE, Holben WE, Feldman RA, Vrijenhoek RC, Cary SC (2000) Genetic variation among endosymbionts of widely distributed vestimentiferan tubeworms. *Appl Environ Microbiol* **66**: 651–658

Doolittle WF (1999) Phylogenetic Classification and the Universal Tree. *Science* **284**: 2124-2128

Dorador C, Meneses D, Urtuvia V, Demergasso C, Vila I, Witzel KP, Imhoff JF (2009) Diversity of Bacteroidetesin high-altitude saline evaporitic basins in northern Chile. *J Geophys Res* **114**: G00D05

Doval MD, Álvarez-Salgado XA, Gasol JM, Lorenzo LM, Mirón I, Figueiras FG, Pedrós-Alió C (2001) Dissolved and suspended organic carbon in the Atlantic sector of the Southern Ocean. Stock dynamics in upper ocean waters. *Mar Ecol-Prog Ser* **223**: 27-38

Ducklow H (1983) Production and fate of bacteria in the oceans. *BioScience* **33**: 494-499

Ducklow HW, Carlson CA (1992) Oceanic bacterial production. *Adv in Microb Ecol* **12**: 113–181

Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**: e1002195.

Edwards RA, Rodríguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander EC, Rohwer F (2006). "Using pyrosequencing to shed light on deep mine microbial ecology". *BMC Genomics* **7**: 57

Eilers H, Pernthaler J, Glockner F, Amann R (2000) Culturability and In situ abundance of pelagic bacteria from the North Sea. *App Environ Microb* **66**: 3044-3051.

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

Emig CC, Geistdoerfer P (2005) The Mediterranean deep-sea fauna: Historical evolution, bathymetric variations and geographical changes. Carnets de Géologie/Notebooks on Geology, Maintenon, Article 2004/01 (CG2004_A01_CCE-PG)

Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, Hugenholtz P (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J* **4**: 642-647

Engene N, Gerwick WH (2011) Intra-genomic 16S rRNA gene heterogeneity in cyanobacterial genomes. *Fottea* **1**:17-24

Estrada M, Marrasé C, Latasa M, Berdalet E (1993) Variability of deep chlorophyll maximum characteristics in the Northwestern Mediterranean. *Mar Ecol-Prog Ser* **92**: 289-300

Estrada M, Salat J (1989) Phytoplankton assemblages of deep and surface water layers in a Mediterranean frontal zone. *Sci Mar* **53**: 203-214

Estrada M, Varela RA, Salat J, Cruzado A, Arias E (1999) Spatio-temporal variability of the winter phytoplankton distribution across the Catalan and North Balearic fronts (NW Mediterranean). *J Plank Res* **21**: 1-20

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131:** 479-491

Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biol Conservation* **61**: 1-10

Falkowski PG, Barber RT, Smetacek VV (1998) Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**: 200-207

Fandino LB, Riemann L, Steward GF, Long RA, Azam F (2001) Variations in bacterial community structure during a dinoflagellate bloom analyzed by DGGE and 16S rDNA sequencing. *Aqua Microb Ecol* **23**: 119-130

Fandino LB, Riemann L, Steward GF, Azam F (2005) Population dynamics of *Cytophaga-Flavobacteria* during marine phytoplankton blooms analyzed by real-time quantitative PCR. *Aquat Microb Ecol* **40**: 251-257

Fenchel T (1988) Marine plankton food-chains. *Annu Rev Ecol Syst*. **18**: 19–38

Fernández-Gómez B, Fernàndez-Guerra A, Casamayor EO, González JM, Pedrós-Alió C, Acinas SG (2012) Patterns and architecture of genomic islands in marine bacteria. *BMC Genomics* **13:** 347

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

Ferrera I, Gasol JM, Sebastián M, Hojerová E, Koblízek M (2011) Comparison of growth rates of aerobic anoxygenic phototrophic bacteria and other bacterioplankton groups in coastal Mediterranean waters. *Appl Environ Microbiol* **77**: 7451-7458

Ferris MJ, Ward DM (1997) Seasonal distribution of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **63**: 1375–1381

Ferris MJ, Kühl M, Wieland A, Ward DM (2003) Different light-adapted ecotypes in a 68°C *Synechococcus* mat community revealed by analysis of 16S-23S intervening transcribed spacer variation. *Appl Environ Microbiol* **69**: 2893-8

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **21**: 237–240

Field KG, Gordon D, Wright T, Rappe M, Urbach E, Vergin K, Giovannoni SJ (1997) Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl Environ Microbiol* **65**: 63–70

Fierer N, Jackson JA, Vilgalys R, Jackson RB (2005) Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Appl Environ Microbiol* **71**: 4117–4120

Fisher MC, Spratt BG, Staley JT (eds) (2006) Species and speciation in microorganisms. *Philos Trans R Soc Lond B* **361**: 1897–2053

Fogg GE (1983) The ecological significance of extracellular products of phytoplankton photosynthesis. *Bot Mar* **26**: 3–14

Fox GE, Wisotzkey JD, Jurtshuk P (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42**: 166–170

Fuhrman JA, Campbell L (1998) Microbial microdiversity. *Nature* **393**: 410–411

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

Galand PE, Gutiérrez-Provecho C, Massana R, Gasol JM, Casamayor EO (2010) Inter-annual recurrence of archaeal assemblages in the coastal NW Mediterranean Sea (Blanes Bay Microbial Observatory). *Limnol Oceanogr* **55**: 2117–2125

García-Martínez J, Acinas SG, Antón AI, Rodríguez-Valera F (1999) Use of the 16S–23S ribosomal genes spacer region in studies of pro- karyotic diversity. *J Microbiol Meth* **36**: 55–64

García-Martínez J, Rodríguez-Valera F (2000) Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine archaea of group I. *Mol Ecol* **9**: 935–948

García-Martínez J, Acinas SG, Massana R, Rodríguez-Valera F (2002) Prevalence and microdiversity of Alteromonas macleodii-like microorganisms in different oceanic regions. *Environ Microbiol* **4**: 42-50

Gasol JM, Doval MD, Pinhassi J, Calderón-Paz JI, Guixa-Boixareu N, Vaqué D, Pedrós-Alió C (1998) Diel variations in bacterial heterotrophic activity and growth in the northwestern Mediterranean Sea. *Mar Ecol-Prog Ser* **164**: 107-124

Gasol JM, Del Giorgio PA (2000) Using flow cytometry for counting natural planktonic bacteria and understanding the structure of planktonic bacterial communities. *Sci Mar* **64**: 197-224

Gasol JM, Massana R, Simó R, Marrasé C, Acinas SG, Pedrós-Alió P, Pelejero C, Sala MM, Calvo E, Vaqué D, Peters P (2012) Blanes Bay (site 55) *pp. 138-141 in O'Brien, T. D., Li, W. K. W., and Morán, X. A. G. (Eds). 2012. ICES Phytoplankton and Microbial Plankton Status Report 2009/2010. ICES Cooperative Research Report No. 313. 196 pp.*

Ghiglione JF, Larcher M, Lebaron P (2005) Spatial and temporal scales of variation in bacterioplankton community structure in the NW Mediterranean Sea. *Aquat Microb Ecol* **40**: 229-240

Gilbert JA, Dupont CL (2011) Microbial metagenomics: beyond the genome. *Annu Rev Mar Sci* **3**: 347–371

Gilbert JA, Steele JA, Caporaso JG*, et al.* (2012) Defining seasonal marine microbial community dynamics. *ISME J* **6**: 298-308

Giovannoni SJ, Rappe MS, Vergin KL, Adair NL (1996) 16s rRNA genes reveal stratified open ocean bacterioplankton populations related to the green non-sulfur bacteria. *Proc Natl Acad Sci* USA **93**: 7979-7984

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 (8)**: 3721-3726

Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* **3**: 679-87

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

Gómez-Pereira PR, Fuchs BM, Alonso C, Oliver MJ, van Beusekom JE, Amann R (2010) Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. I*SME J* **4**: 472-487

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

Gómez-Pereira PR (2012) Marine Bacteroidetes: distribution patterns and role in the degradation of organic matter. Ph.D. Thesis, Universität Bremen, 317 pp.

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

González JM, Pinhassi J, Fernández-Gómez B, Coll-Lladó M, González-Velázquez M, Puigbò P, Jaenicke S, Gómez-Consarnau L, Fernàndez-Guerra A, Goesmann A, Pedrós-Alió C (2011) Genomics of the proteorhodopsin-containing marine flavobacterium *Dokdonia* sp. strain MED134. *Appl Environ Microbiol* **77**: 8676-8686

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

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. *Internat J Syst Bacteriol* **48**: 223–235

Graham TA, Golsteyn-Thomas EJ, Thomas JE, Gannon VP (1997) Inter- and intraspecies comparison of the 16S-23S rRNA operon intergenic spacer regions of six Listeria spp. *Int J Syst Bacteriol* **47**: 863–869

Green HC, Field KG (2012). Sensitive detection of sample interference in environmental qPCR.*Water Res* **46**: 3251-60

Guixa-Boixereu N, Vaqué D, Gasol JM, Pedrós-Alió C (1999) Distribution of viruses and their potential effect on bacterioplankton in an oligotrophic marine system. *Aquat Microb Ecol* **19**: 205-213

Gurtler V, Stanisich VA (1996) New approaches to typing and identification of bacteria using the 16S-23S rDNA spacer region. *Microbiology* **142**: 3–16

Hao X, Chen T (2012) OTU Analysis Using Metagenomic Shotgun Sequencing Data. *PLoS ONE* **7**: e49785

Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TM, Beeson KY, *et al.* (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* **10**: R32

Hedges JI (2002) Why dissolved organic matter? In *Biogeochemistry of marine dissolved organic matter*. Carlson CA & DA Hansell (eds), Academic Press, San Diego pp. 1-33

Hellebust JA (1965) Excretion of some organic compounds by marine phytoplankton. *Limnol Oceanogr* **10**: 192–206.

Helmus MR, Bland TJ, Williams CK, Ives AR (2007) Phylogenetic measures of biodiversity. *The American Naturalist* **169**: 68–83

Hill MO (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**: 427–432

Hingamp P, Grimsley N, Acinas SG, Clerissi C, Subirana L, Poulain J, *et al.* (2013) Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J* **7:** 1678–1695

Hong S, Bunge J, Leslin C, Jeon S, Epstein SS (2009) Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J* **3**: 1365-1373

Hu A, Jiao N, Zhang R, Yang Z (2011) Niche partitioning of marine group I Crenarchaeota in the euphotic and upper mesopelagic zones of the East China Sea. *Appl Environ Microbiol* **77**: 7469–7478

Huang Y, Gilna P, Li W (2009) Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* **25**: 1338-1340

Hugenholtz PB, Goebel M, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**: 4765–74

Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF (2008). Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081–1085

Hurtado LA, Mateos M, Lutz RA, Vrijenhoek RC (2003) Coupling of bacterial endosymbionts and host mitochondrial genomes in the hydrothermal vent clam Calyptogena magnifica. *Appl Environ Microbiol* **69**: 2058–2064

Jaspers E, Nauhaus K, Cypionka H, Overmann J (2001) Multitude and temporal variability of ecological niches as indicated by the diversity of cultivated bacterioplankton. *FEMS Microbiol Ecol* **36**: 153-164

Jaspers E, Overmann J (2004) Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. *App Environ Microb* **70**: 4831-4839

Kan J, Wang K, Chen F (2006) Temporal variation and detection limit of an estuarine bacterioplankton community analyzed by denaturing gradient gel electrophoresis (DGGE). *Aquat Microb Ecol* **42**: 7-18

Karsenti E, Acinas SG, Bork P, Bowler C, De Vargas C, Raes J, *et al.* (2011) A holistic approach to marine eco-systems biology. *PLoS Biol* **9**: e1001177

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, *et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464

Kenzaka T, Yamaguchi N, Tani K, Nasu M (1998) rRNA-targeted fluorescent in situ hybridization analysis of bacterial community structure in river water. *Microbiology-Sgm* **144**: 2085-2093

Keswani J, Whitman WB (2001) Relationship of 16S rRNA sequence similarity to DNA hybridization in prokaryotes. *Int J Syst Evol Microbiol* **51**: 667–678

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

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

Kirchman DL, Dittel AI, Malmstrom RR, Cottrell MT (2005) Biogeography of major bacterial groups in the Delaware Estuary. *Limnol Oceanogr* **50**: 1697-1706

Kirchman DL, Cottrell MT, Lovejoy C (2010) The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environ Microbiol* **12:** 1132–1143

Kisand V, Wikner J (2003) Limited resolution of 16S rDNA DGGE caused by melting properties and closely related DNA sequences. *J Microbiol Meth* **54**: 183-191

Klappenbach JA, Dunbar JM, Schmidt TM (2000) rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* **66:** 1328–1333

Klappenbach JA, Saxman PR, Cole JR, Schmidt TM (2001) rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res* **29:** 181–184

Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner A, Ward DM, *et al.* (2008) Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc Natl Acad Sci* USA **105**: 2504–2509

Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci* USA **102**: 2567–2572

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.

Krom MD, Kress N, Brenner S, Gordon LI (1991). Phosphorus limitation of primary productivity in the Eastern Mediterranean-Sea. *Limnol Oceanogr* **36**, 424-432.

Kuwahara T, Norimatsu I, Nakayama H, Akimoto S, Kataoka K , Arimochi H, Ohnishi Y (2001) Genetic variation in 16S-23S rDNA internal transcribed spacer regions and the possible use of this genetic variation for molecular diagnosis of Bacteroides species. *Microbiol Immunol* **45**: 191-199

Lane DJ (1991) 16S/23S rRNA sequencing, p. 115-175. In E. Stackebrandt, and M. Goodfellow (ed.), *Nucleic acid techniques in bacterial systematics*. John Wiley & Sons, Chichester, United Kingdom.

Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* **32**: 11–16

Lau KWK, Ng CYM, Ren J*, et al.* (2005) *Owenweeksia hongkongensis* gen. nov., sp. nov., a novel marine bacterium of the phylum 'Bacteroidetes'. *Int J Syst Evol Micr* **55**: 1051-1057

Leblond-Bourget N, Philippe H, Mangin I, Decaris B (1996) 16S rRNA and 16S to 23S internal transcribed spacer sequence analyses reveal inter- and intraspecific Bifidobacterium phylogeny. *Int J Syst Bacteriol* **46**: 102–111

Lee S, Fuhrman JA (1991) Spatial and temporal variation of natural bacterioplankton assemblages studied by total genomic DNA cross-hybridization. *Limnol Oceanogr* **36:** 1277-1287

Lee ZMP, Bussema C, Schmidt TM (2009) rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res* **14**: D489–493

Lefort T, Gasol JM (2013) Global-scale distributions of marine surface bacterioplankton groups along gradients of salinity, temperature, chlorophyll: a meta-analysis of fluorescent *in situ* hybridization studies. *Aquat Microb Ecol* doi: 10.3354/ame01643.

Lennon JT, Jones SE (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol* **9**: 119–130

Letunic I, Bork P (2011) Interactive Tree of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**: W475-W478.

Liguori AP, Warrington SD, Ginther JL, Pearson T, Bowers J, Glass MB, *et al.* (2011). Diversity of 16S-23S rDNA internal transcribed spacer (ITS) reveals phylogenetic relationships in Burkholderia pseudomallei and its near-neighbors. *PlosOne* **6**: e29323

Llobet-Brossa E, Rosselló-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.

Logares R, Haverkamp THA, Kumar S, Lanzén A, Nederbragt AJ, Quince C, Kauserud (2012) Environmental microbiology through the lens of high-throughput DNA sequencing: Synopsis of current platforms and bioinformatics approaches. *J Microbiol Meth* **91**: 106-113

Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmento H, *et al.* (2013) Metagenomic 16S rDNA Illumina Tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ Microbiol* DOI: 10.1111/1462-2920.12250

López-López A, Bartual SG, Stal L, Onyshchenko O, Rodríguez-Valera F (2005) Genetic analysis of housekeeping genes reveals a deep-sea ecotype of *Alteromonas macleodii* in the Mediterranean Sea. *Environ Microbiol* **7**: 649–659

Love JL, Scholes P, Gilpin B, Saville M, Lin S & Samuel L (2006) Evaluation of uncertainty in quantitative real-time PCR. *J Microbiol Meth* **67**: 349–356

Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235

Lozupone CA, Hamady M, Kelley ST, Knight R (2007) Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* **73**: 1576-1585

Ludwig W, Euzéby J, Whitman WB (2010) Road map of the phyla Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes. *In* Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W and Whitman WB (eds) *Bergey's Manual of Systematic Bacteriology, 2nd edn, vol. 4. Springer, New York*, pp. 1-19

Ludwig W, Strunk O, Klugbauer S, Klugbauer N, Weizenegger M, Neumaier J, *et al.* (1998) Bacterial Phylogeny Based on Comparative Sequence Analysis. *Electrophoresis* **19**: 554–568

Ludwig, W, Klenk HP (2001) Overview: a phylogenetic backbone and taxonomic framework for procaryotic systematics. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, pp. 49-65. Edited by D. R. Boone, R. W. Castenholz & G. M. Garrity. New York: Springer.

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

Maidak BL, Cole JR, Lilburn TG, Parker CT, Saxman PR, Stredwick JM, Garrity GM, Li B, Olsen GJ, Pramanik S, Schmidt TM, Tiedje JM (2000) The RDP (Ribosomal Database Project) continues. *Nucleic Acid Research* **28**: 173-174

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.

Malmstrom RR, Coe A, Kettler GC, Martiny AC, Frias-Lopez J, Zinser ER, Chisholm SW (2010) Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *ISME J* **1**: 1-13

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

Man SM, Kaakoush NO, Octavia S, Mitchell H (2010) The internal transcribed spacer region, a new tool for use in species differentiation and delineation of systematic relationships within the*Campylobacter* genus. *Appl Environ Microbiol* **76**: 3071–3081

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. *App Environ Microb* **63**: 186-193

Martin AP (2002) Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Appl Environ Microbiol* **68:**3673–3682

Mary I, Cummings DG, Biegala IC, Burkill PH, Archer SD, Zubkov MV (2006) Seasonal dynamics of bacterioplankton community structure at a coastal station in the western English Channel. *Aquat Microb Ecol* **41**: 119-126

Massana R, Gasol JM, Bjørnsen PK, Blackburn N, Hagstrøm A, Hietanen S, *et al.* (1997a) Measurement of bacterial size via image analysis of epifluorescence preparations: description of an inexpensive system and solutions to some of the most common problems. *Sci Mar* **61:** 397–407

Massana R, Murray AE, Preston CM, Delong EF (1997b) Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *App Environ Microb* **63**: 50-56

Massana R, Unrein F, Rodríguez-Martínez R, Forn I, Lefort T, Pinhassi J, *et al.* (2009) Grazing rates and functional diversity of uncultured heterotrophic flagellates. *ISME J* **3**: 588–596

Matthew CT, Thomas DK, Selinger LB, Inglis GD (2011) Spyder, a new method for *in silico* design and assessment of 16S rRNA gene primers for molecular microbial ecology. *FEMS Microbiol Lett* **320**: 152-159

Mayr E (1982) *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*, Chapter 6. The Belknap Press of Harvard University Press, Cambridge, Massachusetts.

McCarthy MD, Hedges JI, Benner R (1998) Major bacterial contribution to marine dissolved organic nitrogen. *Science* **281**: 231–234

Mende DR, Sunagawa S, Zeller G, Bork P (2013) Accurate and universal delineation of prokaryotic species. *Nature Methods* **10**: 881–884

Milyutina IA, Bobrova VK, Matveeva EV, Schaad NW, Troitsky AV (2004) Intragenomic heterogeneity of the 16S rRNA–23S rRNA internal transcribed spacer among Pseudomonas syringae and Pseudomonas fluorescens strains. *FEMS Microbiol Lett* **239**: 17–23

Moeseneder MM, Winter C, Herndl GJ (2001) Horizontal and vertical complexity of attached and free-living bacteria of the eastern Mediterranean Sea, determined by 16S rDNA and 16S rRNA fingerprints. *Limnol Oceanogr* **46**: 95-107Moore LR, Rocap G, Chisholm SW (1998). Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**: 464–467

Moore LR, Chisholm SW (1999) Photophysiology of the marine cyanobacterium *Prochlorococcus*: ecotypic differences among cultured isolates. *Limnol Oceanogr* **44**: 628–638

Morris RM, Vergin KL, Cho J, Rappé MS, Carlson CA, Giovannoni SJ (2005) Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series study site. *Limnol Oceanogr* **50**: 1687–1696

Mühling M, Woolven-Allen J, Murrell JC, Joint I (2008) Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *ISME J* **2**: 379-392

Muyzer G, de Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**: 695–700

Muyzer G, Brinkhoff TT, Nübel U, Santegoeds C, Schäfer H, Wawer C (1998) Denaturing gradient gel electrophoresis (DGGE) in microbial ecology, p. 3.4.4/1–3.4.4/27. In A. D. L. Akkermans, J. D. van Elsas, F. J. Bruijn (ed.), *Molecular microbial ecology manual*. Kluwer Academic Pub- lishers, Dordrecht, The Netherlands.

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

Nagata T, Kirchman DL (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 (2000) Production mechanisms of dissolved organic matter. In: *Microbial ecology of the oceans*. New York: Wiley-Liss, Inc, pp. 121–152

Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, *et al.* (2011) Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res* **39**: e90

Ngugi DK, Stingl U (2012) Combined analyses of the ITS loci and the corresponding 16S rRNA genes reveal high micro- and macrodiversity of SAR11 populations in the Red Sea. *Plos One* **7**: e50274- e50274

Nipperess DA, Matsen FA (2013) The mean and variance of phylogenetic diversity under rarefaction. *Meth Ecol Evol* ***4***: 566–572

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 Microbiol* **68**: 201-210

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

Oksanen J, Kindt R, Legendre P, O'Hara B, Simpson GL, Solymos P, *et al.* (2009) Vegan: community ecology package. R package version **1.** 15-2. URL http://CRAN.R-project.org/package=vegan.

Osorio CR, Collins MD, Romalde JL, Toranzo AE (2005) Variation in 16S–23S rRNA intergenic spacer regions in *Photobacterium damselae*: a mosaic-like structure. *Appl Environ Microbiol* **71**: 636–645

Palys T, Nakamura LK, Cohan FM (1997) Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int J Sys Bacteriol* **47**: 1145–1156

Paradis E, Claude J, Strimmer K (2004) Ape: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290

Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. *J Appl Genet* **52**: 413–435

Park JW, Crowley DE (2010) Nested PCR bias: a case study of *Pseudomonas spp.* in soil microcosms. *J Environ Monitor* **12**: 985

Pedrós-Alió C, Calderón-Paz JI, Guixa-Boixereu N, Estrada M, Gasol JM (1999) Bacterioplankton and phytoplankton biomass and production during summer stratification in the northwestern Mediterranean Sea. *Deep-Sea Res Pt I* **46**: 985-1019

Pernthaler A, Pernthaler J, Amann R (2002) Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Environ Microbiol* **68**: 3094-3101

Pernthaler J, Pernthaler A, Amann R (2003) Automated enumeration of groups of marine picoplankton after fluorescence in situ hybridization. *Appl Environ Microbiol* **69**: 2631–2637

Pernthaler A, Pernthaler J, Amann R (2004) Sensitive multi-color fluorescence *in situ* hybridization for the identification of environmental microorganisms. In *Molecular Microbial Ecology Manual,* Kowalchuk, G.A., De Bruin, F.J., Head, I.M., Akkermans, A.D.L., Van Elsas, J.D. (eds). Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 711–726

Pernthaler J (2005) Predation on prokaryotes in the water column and its ecological implications. *Nat Rev Micro* **3**: 537-546

Petruska J, Goodman MF, Boosalis MS, Sowers LC, Cheong C, Tinoco I (1988) Comparison between DNA melting thermodynamics and DNA polymerase fidelity. *Proc Natl Acad Sci* USA **85**: 6252–6256

Pinet, PR (2009) Invitation to Oceanography, Fifht ed. *Jones and Bartlett Publishers International*, London, UK (626 pp)

Pinhassi J, Azam F, Hemphälä J, Long RA, Martínez J, Zweifel UL, Hagström A (1999) Coupling between bacterioplankton species composition, population dyna. *Aquat Microb Ecol* **17**: 13-26

Pinhassi J, Hagström A (2000) Seasonal succession in marine bacterioplankton. *Aquat Microb Ecol* **21**: 245-256

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

Pinto AJ, Raskin L (2012) PCR Biases Distort Bacterial and Archaeal Community Structure in Pyrosequencing Datasets. *PLoS ONE* **7**: e43093

Polz MF, Cavanaugh CM (1998) Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**: 3724-3730

Pomeroy LR (1974) The ocean's food web, a changing paradigm. *BioScience* **24**: 499-504

Pomeroy LR, Williams PJI, Azam F, Hobbie JE (2007) The microbial loop. *Oceanography* **20**: 28–33

Pommier T, Canbäck B, Riemann L, Boström KH, Simu K, Lundberg P, *et al.* (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16:** 867–880

Pommier T, Neal P, Gasol JM, Coll M, Acinas SG, Pedrós-Alió C (2010) Spatial patterns of bacterial richness and evenness in the NW Mediterranean Sea explored by pyrosequencing of the 16S rRNA. *Aquat Microb Ecol* **61:** 221-233

Preheim SP, Timberlake S, Martin FP (2011) Merging Taxonomy with Ecological Population Prediction in a Case Study of Vibrionaceae. *Appl Environ Microbiol* **77**: 7195–7206

Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestona MK, Freckleton RP, *et al.* (2007) The role of ecological theory in microbial ecology. *Nat Rev Micro* **5**: 384–392

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acid Research* **35:** 7188-7196

Pruesse E, Peplies J, Glöckner FO (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28:** 1823–1829

Pybus OG, Harvey PH (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc R Soc Lond B* **267**: 2267-2272

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41:** 590–596

R Core Team (2013) R: A language and environment for statistical computing. *R. Foundation for Statistical Computing, Vienna, Austria*. URL http://www.R-project.org/.

Rappe MS, Giovannoni SJ (2003) The uncultured microbial majority. *Annu Rev Microbiol* **57**: 369–394

Ravenschlag K, Sahm K, Amann R (2001) Quantitative molecular analysis of the microbial community in marine Arctic sediments (Svalbard). *Appl Environ Microbiol* **67**: 387–395

Reichenbach H, Weeks OB (Editors) The *Flavobacterium-Cytophaga* Group (1980) Proceedings of the International Symposium on Yellow-Pigmented Gram-Negative Bacteria of the Flavobacterium-Cytophaga Group, Braunschweig-Stöckheim. 217 S., 7 Abb., 107 Tab. Weinheim Deerfield Beach, Florida Basel 1981. Verlag Chemie. DM 88,00 (page 681)

Restrepo-Ortiz CX, Auguet JC, Casamayor EO (2013) Targeting spatiotemporal dynamics of planktonic SAGMGC-1 and segregation of ammonia-oxidizing thaumarchaeota ecotypes by newly designed primers and quantitative polymerase chain reaction. *Environ Microbiol.* **doi**: 10.1111/1462-2920.12191

Restrepo-Ortiz CX, Casamayor EO (2013) Environmental distribution of two widespread uncultured freshwater Euryarchaeota clades unveiled by specific primers and quantitative PCR. *Environ Microbiol Reports* **doi**:10.1111/1758-2229.12088

Riemann L, Steward G, Fandino L, Campbell L, Landry M, Azam F (1999) Bacterial community composition during two consecutive NE Monsoon periods in the Arabian Sea studied by denaturing gradient gel electrophoresis (DGGE) of rRNA genes. *Deep-Sea Res Pt II* **46**: 1791-1811

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

Riemann L, Winding A (2001) Community dynamics of free-living and particle-associated bacterial assemblages during a freshwater phytoplankton bloom. *Microb Ecol* **42**: 274-285

Riemann L, Middelboe M (2002) Stability of bacterial and viral community compositions in Danish coastal waters as depicted by DNA fingerprinting techniques. *Aquat Microb Ecol* **27**: 219-232

Rink B, Seeberger S, Martens T, Duerselen CD, Simon M, Brinkhoff T (2007) Effects of phytoplankton bloom in a coastal ecosystem on the composition of bacterial communities. *Aquat Microb Ecol* **48:** 47-60

Rink B, Martens T, Fischer D, Lemke A, Grossart HP, Simon M, Brinkhoff T (2008) Short-term dynamics of bacterial communities in a tidally affected coastal ecosystem. *FEMS Microbiol Ecol* **66:** 306-319

Robertson BR, Tezuka N, Watanabe MM (2001) Phylogenetic analyses of *Synechococcus* strains (cyanobacteria) using sequences of 16S rDNA and part of the phycocyanin operon reveal multiple evolutionary lines and reflect phycobilin content. *Int J Syst Evol Microbiol* **51:** 861–871

Robinson C, Williams PI (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, Distel DL, Waterbury JB, Chisholm SW (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* **68***: 1180–1191*

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differen- tiation. *Nature* **424:** 1042–1047

Rocha EP (2004) Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res* **14**: 2279–2286

Roux S, Enault F, Bronner G, Debroas D (2011) Comparison of 16S rRNA and protein-coding genes as molecular markers for assessing microbial diversity (Bacteria and Archaea) in ecosystems. *FEMS Microbiol Ecol* **78**: 617-28

Ruiz-González C, Lefort T, Galí M, Montserrat Sala M, Sommaruga R, Simó R, Gasol JM (2011) Seasonal patterns in the sunlight sensitivity of bacterioplankton from Mediterranean surface coastal waters. *FEMS microbiol ecol* **79**: 661-674.

Ruiz-González C, Galí M, Lefort T, Cardelús C, Simó R, Gasol JM (2012a) Annual variability in light modulation of bacterial heterotrophic activity in surface northwestern Mediterranean waters. *Limnol Oceanogr* **57:** 1376–1388

Ruiz-González C, Galí M, Lefort T, Sala M, Sommaruga R, Simó R, Gasol JM (2012b) Seasonal patterns in the sensitivity to sunlight of bacterioplankton from surface Mediterranean coastal waters. *FEMS Microbiol Ecol* **79:** 661–674

Ruiz-González C, Galí M, Simó R, Gasol JM (2012c) Sunlight effects on the DMSP-sulfur and leucine assimilation activities of polar heterotrophic bacterioplankton. *Biogeochemistry* **110:** 57–74

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, *et al.* (2007) The *Sorcerer II* Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5:** e77

Salat J (1996) Review of hydrographic environmental factors that may influence anchovy habitats in northwestern Mediterranean. *Sci Mar* **60**: 21-32

Salyers AA (1984) *Bacteroides* of the human lower intestinal tract. *Annu Rev Microbiol* **38**: 293-313

Sánchez O, Gasol JM, Massana R, Mas J, Pedrós-Alió C (2007) Comparison of different denaturing gradient gel electrophoresis primer sets for the study of marine bacterioplankton communities. *Appl Environ Microbiol* **73:** 5962-5967

Sánchez O, Gasol JM, Balagué V, Massana R, Mas J, Pedrós-Alió C (2009) Influence of primer mismatch and microdiversity on DGGE results: a case study with SAR11. *Aquat Microb Ecol* **54:** 211-216

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463–5467

Santos TMA, Gilbert RO, Bicalho RC (2010) Metagenomic analysis of the uterine bacterial microbiota in healthy and metritic postpartum dairy cows. *Journal of Dairy Science* **94:** 291-302

Sapp J (2005) Microbial Phylogeny and Evolution: Concepts and Controversies. *Jan Sapp (ed.)*. Oxford University Press, New York USA, 326 pp.

Sarmento H, Gasol JM (2012) Use of phytoplankton derived dissolved organic carbon by bacterioplankton large phylogenetic groups. *Environ Microbiol* **14:** 2348–2360

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

Schauer M, Balagué V, Pedrós-Alió C, Massana R (2003) Seasonal changes in the taxonomic composition of bacterioplankton in a coastal oligotrophic system. *Aquat Microb Ecol* **31:** 163-174

Schauer M, Massana R, Pedrós-Alió C (2000) Spatial differences in bacterioplankton composition along the Catalan coast (NW Mediterranean) assessed by molecular fingerprinting. *FEMS Microbiol Ecol* **33:** 51-59

Schloss PD, Larget BR, Handelsman J (2004) Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl Environ Microbiol* **70**: 5485–5492

Schloss PD (2008) Evaluating different approaches that test whether microbial communities have the same structure. *ISME J* **2**: 265–275.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009) Introducing mothur: open- source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microb* **75**: 7537–7541

Schloter M, Lebuhn M, Heulin T, Hartmann A (2000) Ecology and evolution of bacterial microdiversity. *FEMS Microbiol Rev* **24:** 647–660

Sekar R, Fuchs B, Amann R, Pernthaler J (2004) Flow sorting of marine bacterioplankton after fluorescence *in situ* hybridization. *Appl Environ Microbiol* **70:** 6210– 6219

Sheffield VC, Cox DR, Myers RM (1989) Attachment of a 40-bp G+C rich sequence (GC-clamp) to genomic DNA fragments by polymerase chain reaction results in improved detection of single-base changes. *Proc Natl Acad Sci* USA **86**: 232–236

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 Microbiol* **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

Siokou-Frangou I, Christaki U, Mazzocchi MG, Montresor M, Ribera d'Alcala M, Vaqué D, Zingone A (2010) Plankton in the open Mediterranean Sea: A review. *Biogeosciences* **7**: 1543–1586

Smith, C. J., D. B. Nedwell, L. F. Dong, and A. M. Osborn (2006) Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples. Environ. Microbiol. **8:**804-815

Smith CJ, Osborn AM (2008) Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol Ecol* **10**: 6–20

Sogin ML, Morrison HG, Huber JA *et al.* (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *P Natl Acad Sci USA* **103**: 12115–12120

Sokal RR, Rohlf FJ (1995). Biometry, 3rd edn. W.H. Freeman, Company, New York.

Stackebrandt E, Goebel BM (1994) Taxonomic note: a place for DNA-DNA reassociation and 16s rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**: 846-849

Staley JT (2006) The bacterial species dilemma and the genomicphylogenetic species concept. *Philos Trans R Soc Lond B Biol Sci* **361**: 1899–1909

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

Stewart FJ, Cavanaugh CM (2007) Intragenomic variation and evolution of the internal transcribed spacer of the rRNA operon in bacteria. *J Mol Evol* **65**: 44-67

Stults JR, Snoeyenbos-WestcOL, Methe B, Lovley DR, Chandler DP (2001) Application of the 5′ fluorogenic exonuclease assay (TaqMan) for quantitative ribosomal DNA and rRNA analysis in sediments. *Appl Environ Microbiol* **67:** 2781-2789

Suttle CA (2005) Viruses in the sea. *Nature* **437**: 356–361

Suzuki M, Nakagawa Y, Harayama S, Yamamoto S (2001a) 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 *Tenacibaculum mesophilum* sp. nov. and *Tenacibaculum amylolyticum* sp. nov. *Int J Syst Evol Microbiol* **51**: 1639-1652

Suzuki MT, Preston CM, Chavez FP, DeLong EF (2001b) Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in Monterey Bay. *Aquat Microb Ecol* **24:** 117-127

Suzuki N, Yoshida A, Nakano Y (2005) Quantitative analysis of multi-species oral biofilms by TaqMan Real-Time PCR. *Clin Med Res* **3**: 176–185

Szabo G, Preheim SP, Kauffman KM, David LA, Shapiro J, Alm EJ, Polz MF (2012) Reproducibility of Vibrionaceae population structur in coastal bacterioplankton. *ISME J* **7**: 509-519

Takai K, Horikoshi K (2000) Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. *Appl Environ Microbiol* **66**: 5066–5072

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596-1599

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

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

Teira E, Martínez-García S, Lønborg C, Álvarez-Salgado XA (2009) Growth rates of different phylogenetic bacterioplankton groups in a coastal upwelling system. *Environ Microbiol Reports* **1:** 545–554

Thingstad TF, Rassoulzadegan F (1995) Nutrient limitations, microbial food webs, and 'biological C-pumps': suggested interactions in a P-limited Mediterranean. *Mar Ecol Prog Ser* **117**: 299-306

Thingstad TF, Rassoulzadegan F (1999) Conceptual models for the bio-geochemical role of the photic zone microbial food web, with particular reference to the Mediterranean Sea. *Prog Oceanogr* **44**: 271–286

Thingstad TF, Krom MD, Mantoura RFC, Flaten GAF, Groom S, Herut B, *et al.* (2005) Nature of phosphorus limitation in the ultraoligotrophic eastern Mediterranean. *Science* **309**: 1068–1071

Thompson JR, Pacocha S, Pharino C, Klepac-Ceraj V, *et al.* (2005) Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**: 1311–1313

Torsvik V, Daae FL, Sandaa RA, Øvreas L (1998) Novel techniques for analysing microbial diversity in natural and perturbed environments. *J biotechnol* **64**: 53-62

Tourova, TP (2003) Copy number of ribosomal operons in prokaryotes and its effect on phylogenetic analyses. *Microbiology* **72:** 389–402

Troussellier M, Schäfer H, Batailler N*, et al.* (2002) Bacterial activity and genetic richness along an estuarine gradient (Rhone River plume, France). *Aquat Microb Ecol* **28**: 13-24

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, *et al.* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37-43

Urbach E, Scanlan DJ, Distel DL, Waterbury JB, Chisholm SW (1998) Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (*Cyanobacteria*). *J Mol Evol* **46**: 188-201

Urdea MS, Warner BD, Running JA, Stempien M, Clyne J, Horn T (1988) A comparison of non-radioisotopic hybridization assay methods using fluorescent, chemiluminescent and enzyme labeled synthetic oligodeoxyribonucleotide probes. *Nucleic Acids Res* **16:** 4937–4956

Vergin KL, Beszteri B, Monier A, Thrash JC, Temperton B, Treusch AH, *et al.* (2013) High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *ISME J* **7**: 1322–1332

von Wintzingerode F, Göbel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Ecol* **21**: 213–229

Vos M, Didelot X (2009) A comparison of homologous recombination rates in bacteria and archaea. *ISME J* **3:**199-208

Wagner M, Smidt H, Loy A, Zhou J (2007) Unravelling microbial communities with DNA-microarrays: challenges and future directions. *Microb Ecol* **53:** 498–506

Ward DM, Ferris MJ, Nold SC, Bateson MM (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol Mol Biol Rev* **62**: 1353–1370

Ward DM (2006) A macrobiological perspective on microbial species. *Microbe* **1**: 269–278

Ward DM, Bateson MM, Ferris MJ, Kuhl M, Wieland A, Koeppel A, Cohan FM (2006) Cyanobacterial ecotypes in the microbial mat community of mushroom spring (Yellowstone national park, Wyoming) as species-like units linking microbial community composition, structure and function. *Philos Trans R Soc Lond B Biol. Sci.,* **361**: 1997–2008

Weller R, Glöckner FO, Amann R (2000) 16S rRNA-targeted oligonucleotide probes for the in situ detection of members of the phylum Cytophaga-Flavobacterium-Bacteroides. *Syst Appl Microbiol* **23**: 107-114

Wells LE, Deming JW (2003) Abundance of Bacteria, the *Cytophaga-Flavobacterium* cluster and Archaea in cold oligotrophic waters and nepheloid layers of the Northwest Passage, Canadian Archipelago. *Aquat Microb Ecol* **31**: 19-31

West NJ, Scanlan DJ (1999). Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Appl Environ Microbiol* **65**: 2585–2591

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

White TJ, Arnheim N, Erlich HA (1989) The polymerase chain reaction. *Trends Genet* **5**: 185–189

Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: The unseen majority. *Proc Natl Acad Sci* USA **95**: 6578–6583

Woebken D, Lam P, Kuypers MMM, Naqvi SWA, Kartal B, Strous M, *et al.* (2008) A microdiversity study of anammox bacteria reveals a novel CandidatusScalindua phylotype in marine oxygen minimum zones. *Environ Microbiol* **10**: 3106-3119

Woese CR (1987) Bacterial evolution. *Microbiol Rev* **51**: 221- 271

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* **74**: 5088-5090

Worm J, Gustavson K, Garde K, Borch NH, Søndergaard M (2001) Functional similarity of attached and free-living bacteria during freshwater phytoplankton blooms. *Aquat Microb Ecol* **25**: 103-111

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

Wright TD, Vergin KL, Boyd PW, Giovannoni SJ (1997) A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. *Appl Environ Microbiol* **63**: 1441-1448

Yilmaz LS, Noguera DR (2004) Mechanistic approach to the problem of hybridization efficiency in fluorescent *in situ* hybridization. *Appl Environ Microbiol* **70:** 7126–7139

Yilmaz LS, Parkernar S, Noguera DR (2011) mathFISH, a web tool that uses thermodynamics-based mathematical models for *in silico* evalulation of oligonucleotide probes for fluorescence *in situ* hybridization. *Appl Environ Microbiol* **77:** 1118–1122

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

Zinger L, Gobet A, Pommier T (2012) Two decades of describing the unseen majority of aquatic microbial diversity. *Mol Ecol* **21**: 1878–1896

Zinser ER, Coe A, Johnson ZI, Martiny A, Fuller NJ, Scanlan DJ, Chisholm SW (2006) *Prochlorococcus* ecotype abundances in the North Atlantic Ocean revealed by an improved quantitative PCR method. *Appl Environ Microbiol* **72**: 723–732

Twenty years from now you will be more disappointed by the things that you didn't do than by the ones you did do. So throw off the bowlines. Sail away from the safe harbor. Catch the trade winds in your sails. Explore. Dream. Discover.

Mark Twain