Article type : Original Article

Delineation of ecologically distinct units of marine Bacteroidetes in the Northwestern Mediterranean Sea

Running title: Delineation of ESTUs of marine Bacteroidetes

Cristina Díez-Vives^{1,2}, Shaun Nielsen³, Pablo Sánchez¹, Oswaldo Palenzuela⁴, Isabel Ferrera^{1,5}, Marta Sebastián^{1,6}, Carlos Pedrós-Alió^{1,7}, Josep M. Gasol^{1,8} and Silvia G. Acinas^{1*}

¹ Department of Marine Biology and Oceanography, Institut de Ciències del Mar (CSIC), 08003 Barcelona, Catalunya, Spain.

² Life Sciences Department, Natural History Museum, SW7 5BD London, UK.

³ School of Biotechnology and Biomolecular Sciences, Centre for Marine Bio-Innovation, The University of New South Wales, Sydney, NSW, Australia

⁴ Department of Biology, Culture and Pathology of Marine Species, Instituto de Acuicultura Torre de la Sal, CSIC. 12595 Ribera de Cabanes, Castellón, Spain

⁵ Instituto Español de Oceanografía, Centro Oceanográfico de Málaga, 29640 Fuengirola, Málaga, Spain

⁶ Instituto de Oceanografía y Cambio Global, IOCAG, Universidad de las Palmas de Gran Canaria, ULPGC, Parque Científico Tecnológico Marino de Taliarte, s/n. 35214, Gran Canaria, Spain.

⁷ Departamento de Biología de Sistemas. Centro Nacional de Biotecnología, CNB-CSIC, Madrid, Spain.

⁸ Centre for Marine Ecosystems Research, School of Science. Edith Cowan University, Joondalup, WA, Australia

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.15068

This article is protected by copyright. All rights reserved.

*Corresponding author: E-mail sacinas@icm.csic.es, Tel (+34) 93 230 8565; Fax (+34) 93 230 95 55.

Abstract

Bacteroidetes is one of the dominant phyla of ocean bacterioplankton, yet its diversity and population structure is poorly understood. To advance in the delineation of ecologically meaningful units within this group, we constructed near full-length 16S rRNA gene clone libraries from contrasting marine environments in the NW Mediterranean. Based on phylogeny and the associated ecological variables (depth and season), 24 different Bacteroidetes clades were delineated. By considering their relative abundance (from iTag amplicon sequencing studies), we described the distribution patterns of each of these clades, delimiting them as ecologically significant taxonomic units (ESTUs). Spatially, there was almost no overlap among ESTUs at different depths. In deep waters there was predominance of *Owenweeksia*, *Leeuwenhoekiella*, *Muricauda*-related genera, and some depth-associated ESTUs within the NS5 and NS2b marine clades. Seasonally, multi-annual dynamics of recurring ESTUs were present with dominance of some ESTUs within the NS4, NS5 and NS2b marine clades along most of the year, but with variable relative frequencies between months. A drastic change towards the predominance of *Formosa*-related ESTUs and one ESTU from the NS5 marine clade was typically present after the spring bloom. Even though there are no isolates available for these ESTUs to determine their physiology, correlation analyses identified the environmental preference of some of them. Overall, our results suggest that there is a high degree of niche specialisation within these closely related clades. This work constitutes a step forward in disentangling the ecology of marine Bacteroidetes, which are essential players in organic matter processing in the oceans.

Keywords: Bacteroidetes, ecological units, community structure, 16S rRNA gene, NW Mediterranean.

Introduction

Bacteria and archaea are extremely diverse with global estimations of about 10^{12} species (Locey & Lennon, 2016), yet we know little about microbial speciation processes in natural marine populations and to what extent coexisting genotypic diversity reflects ecological meaningful units (Cohan 2002, 2006; Polz *et al.* 2006; Cohan & Perry, 2007). One bacterial speciation model (the stable ecotype model) predicts that phylogenetic clusters likely correspond to ecological distinct units that arose by neutral diversification after niche-specific selection (Wiedenbeck & Cohan 2011). Such clusters have been observed in many bacterial taxa, and associated with environmental factors such as light, depth, seasonal mixing, or water stratification (Gordon & Giovannoni 1996; Moore *et al.* 1998; Morris *et al.* 2004; Carlson *et al.* 2009; Preheim *et al.* 2011). However, linking phylogenetic clusters to distinct ecotypes is still challenging, because the phylogenetic clusters can result from neutral evolution alone (Fraser *et al.* 2007), and the relationship between function and phylogeny can be distorted by recombination events or horizontal gene transfer (Doolittle & Papke 2006; Retchless & Lawrence 2007).

The phylum Bacteroidetes is known to be one of the most abundant bacterial groups in aquatic ecosystems, contributing approximately from 4 to 22% of total bacteria in different marine ecosystems (Cottrell and Kirchman, 2000; Alonso-Sáez *et al.*, 2007; Acinas *et al.* 2015) and having a pivotal role in the processing of polymeric organic material (Kirchman, 2002). Despite their importance, relatively little is known about the fine-scale diversity and structure of this phylum. Substantial progress in describing the diversity of class Flavobacteriia has been made by studies defining novel phylogenetic lineages among

uncultured species in the North Sea (i.e. marine clades NS; Alonso *et al.* 2007; Gómez-Pereira *et al.* 2010), but this was done using general bacterial primers and through clone library construction. A few additional studies used Bacteroidetes-specific primers in an attempt to obtain detailed insights into their diversity in certain habitats (Kirchman *et al.* 2003; Chen *et al.* 2008; Mühling *et al.* 2008; Díez-Vives *et al.* 2013). However, these studies were either limited by the low resolution of the fingerprinting approach used (i.e. denaturing gradient gel electrophoresis) or by the rather low number of clones analysed.

In contrast, bacterial groups such as the Cyanobacteria, have received more attention, and segmentation of clades into ecologically significant groups (ecotypes) was reported by pioneering studies (Moore *et al.* 1998; West & Scanlan 1999). Since then, an increasing number of studies have focused on ecotype abundance and dynamics, and on refining the delineation of these ecotypes with newer techniques. For instance, ITS-based phylogenies partitioned previously observed divisions of cultured isolates of *Prochlorococcus* and *Synechococcus* into several genetically distinct new clades associated with particular depths, phenotypes, and geography (Rocap *et al.* 2002; Ferris *et al.* 2003). More recently, Farrant and coworkers used metagenomic reads of a single functional gene marker (i.e. *petB*) to obtain spatial distribution patterns which divided well-defined populations of *Prochlorococcus* and *Synechococcus* into new Ecologically Significant Taxonomic Units (ESTUs; Farrant *et al.* 2016).

With the goal of investigating the ecology of naturally occurring marine Bacteroidetes, we aimed to identify phylogenetic clusters potentially representing ecologically distinct populations. We constructed a reference tree of the near full-length 16S rRNA gene sequences of samples obtained from contrasting environments to identify habitat-specific

clades. The temporal and spatial distribution of the defined clades was assessed by comparing them with two large-scale Illumina amplicon 16S rRNA gene tags (16S iTags) datasets from the North-western Mediterranean Sea. Representative sequences of Operational Taxonomic Units (OTUs) defined at 99% identity were included by parsimony into our pre-existent phylogenetic tree. Clustering based on the temporal and spatial distribution patterns of the Bacteroidetes clades allowed defining ESTUs (i.e. subgroups within clades that co-occur in the field), which could be further associated with specific environmental parameters.

Material and Methods

Sample Collection and DNA extraction

Bacterioplankton was collected from seawater samples taken with two different sampling strategies. Firstly, surface water samples were taken at the Blanes Bay Microbial Observatory (BBMO, 41°40'N, 2°48'E; Gasol *et al.* 2016) during a two-year seasonal study (from September 2006 to September 2008). Secondly, we compared this coastal location at the surface with a deep sample (2,000 m depth) obtained from an offshore location 111 km away. Both samples were taken aboard the RV 'García del Cid' during the 'Modivus' cruise (September 20th to 23rd, 2007). About 8-10 L of seawater were *in situ* prefiltered through a 200 µm-mesh, and sequentially filtered through a 20 µm-mesh, GF/A or polycarbonate 3 µm prefilters (Whatman), and 0.22 µm Sterivex filters (Durapore, Millipore) using a peristaltic pump. The 0.22 µm filters, were saturated with 1.8 ml lysis buffer (50 mM Tris- HCl, 40 mM EDTA, 0.75 M sucrose, pH 8.0), flash-frozen in liquid N₂, and kept at -80 °C until extraction. Microbial cells were lysed with lysozyme, proteinase K and sodium dodecyl sulfate, and the nucleic acids extracted with phenol and concentrated in an Amicon Ultracentrifugal Unit (Millipore), as described in Massana *et al.* (1997).

Bacteroidetes-specific DNA amplification

In a previous study using denaturing gradient gel electrophoresis (DGGE) fingerprinting, we demonstrated that the Bacteroidetes community structure at the BBMO displayed a marked seasonality with two distinct assemblages: one in Summer-Fall, and one in Winter-Spring (Díez-Vives *et al.* 2013). We built a Bacteroidetes-specific clone library for each of these seasonal communities, using pooled monthly samples from the corresponding periods taken along the 2-year study: eight samples in the Winter-Spring library ‘WS’ and twelve in the Summer-Fall library ‘SF’ (see Table S1 for details). For the comparison between depths, the libraries were constructed using DNA from two individual samples (surface and 2,000 m depth), referred to as the Surface community ‘S’ and the Deep community ‘D’. Note that these libraries were obtained from samples taken at the end of summer.

DNA amplification was carried out with a forward primer on the 16S rRNA gene and a reverse primer on the 23S rRNA. The primers were specific to the Bacteroidetes phylum (Table S2), CF319aF (Manz *et al.* 1996) and CF434R (5’ CACTATCGGTCTCTCAGG 3’). The latter is a primer designed on the 23S rRNA gene to overcome some cross amplification issues of the forward primer (Díez-Vives *et al.* 2012). Conventional PCR was modified with a reconditioning PCR to minimize the occurrence of heteroduplex amplicons as described in Acinas *et al.* (2005). Details on the primer election and PCR conditions can be found in the Methods S1 section of the supplementary material.

Generation of clone libraries and sequencing

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). Visual inspection of the

electropherograms was conducted, and high quality sequences were assembled using the software Geneious Pro (<http://www.geneious.com>). Chimeric sequences were identified using the software Mallard (Ashelford *et al.* 2006), and removed before further downstream analyses. The end of the small subunit rRNA gene was identified based on alignments with some representative Bacteroidetes sequences from the SILVA123 database. Trimmed sequences had a total length of 1062 - 1074 bp. These gene sequences were deposited in GenBank under accession numbers KP887163 - KP887808.

Classification and alignment of 16S rRNA sequences

The ribosomal small subunit sequence dataset was used to query against the GenBank database, National Centre for Biotechnology Information (Benson *et al.* 2013) using BLAST (Altschul *et al.* 1990) to determine their closest relatives. Sequences were further classified with the SINA classifier (Pruesse *et al.* 2012) using both SILVA (Quast *et al.* 2013) and RDP (Ribosomal Database Project) taxonomies, with a cut off of 0.95 for the classification. Alignment of the small subunit dataset was performed in the SINA web alignment tool (Pruesse *et al.* 2012) and imported into the ARB software v5.1 (Ludwig *et al.* 2004). The alignment was manually refined according to secondary structure criteria.

Clone library descriptors

Operational Taxonomic Units (OTUs) were defined at 99% sequence similarity cut-off, for the near full 16S rRNA gene, using the furthest neighbour algorithm (complete clustering) within the software package Mothur (Schloss *et al.* 2009). Cluster similarity curves (unique sequence clusters represented as a function of sequence similarity), accumulation curves (number of OTUs as a function of number of sequences sampled), Good's coverage (Good 1953), Chao-1 estimator of species richness (Chao *et al.* 2009) and the Shannon diversity

index (Hill 1973) were generated for each of the four libraries. The Jaccard index (Jaccard 1901) was calculated to detect similarities among libraries.

Identification of ‘clades and habitats’ using AdaptML

A reference phylogenetic tree was built with the near full-length 16S rRNA sequences. The evolutionary model AdaptML (Hunt *et al.* 2008) was used to identify putative clades (named ‘populations’ in that software) based on the phylogenetic information and ecological data associated with each branch (i.e. season and depth). AdaptML identifies ecological populations based on associated environmental categories without a predetermined genetic similarity cut off and without knowledge of existing species, and offers a better approach than simplistic OTU clustering. It is a maximum likelihood method that employs a hidden Markov model to learn ‘projected habitats’ (distribution patterns among environmental categories) and ecologically cohesive ‘populations’ (groups of related strains sharing the same projected habitat). To define our environmental categories, we combined the temporal (Summer-Fall ‘SF’ or Winter-Spring ‘WS’) and spatial (Surface ‘S’ or Deep ‘D’) categories. In this case, S and D were both sampled in September (Table S1) and hence were considered SF sequences, while WS and SF were sampled at 5 m and thus were tagged as S sequences. The phylogenetic tree of the small subunit rRNA gene was generated using PhyML v.3.0 with 100 bootstraps (Guindon *et al.* 2010). All parameters for AdaptML were kept at default values except *rand*=20 and the model threshold value that was set at 0.05. Model predictions and ecological data for the sequences were visualized using the interactive Tree of Life web application (iTol; Letunic & Bork 2011).

Generation of high-throughput sequencing (HTS) data

To add distribution and ecological context to these clades, we included the relative abundance of OTUs (defined at 99% cut-off) from two different 16S rRNA gene Illumina-Tag (16S iTags) sequencing datasets from the North-western Mediterranean Sea. These were generated from water samples collected either during the trans-Mediterranean cruise HOTMIX aboard the R/V *Sarmiento de Gamboa* in spring 2014 (April 27th to May 29th, 2014; Table S3) or from surface waters in a time-series station sampled monthly at the BBMO from September 2006 to December 2011 (Table S4), which overlaps with the 2 years of samples used in the Bacteroidetes clone library construction (Table S1). Details of the DNA extraction, amplification and sequencing can be found in the Methods S2 of the supplementary material. Briefly, the V4-V5 region of the 16S rRNA gene was amplified for the HOTMIX samples, and the V3-V4 region was amplified for the BBMO series. Sequences were clustered into OTUs at 99% cut-off, and were taxonomically annotated using BLASTN against the SILVA123 database as reference.

The HOTMIX dataset covers the whole Mediterranean Sea, yet for this study we considered the samples from the NW region only (65 samples from 9 locations), including 14,421 OTUs defined at 99% identity. The Blanes Bay dataset included 33,005 OTUs defined at 99% identity (64 samples from 5 years monthly sampling). Our clone library sequences were used to recruit OTUs with >99% similarity by BLASTN v.2.6.0 using default parameters. The candidate OTUs were then included in our 16S rRNA gene tree by Parsimony using ARB. Only sequences that were located within the limits of our predefined clades were kept. Bacteroidetes sequences from both HTS datasets have been submitted to the Genbank Sequence Read Archive under Bioproject ID PRJNA495818.

Defining Ecologically Significant Units (ESTUs)

ESTUs were determined by clustering recruited OTUs from both amplicon 16S rRNA gene iTags datasets within each clade, based on relative abundances using model-based clustering methods implemented in the MCLUST package (Scrucca *et al.* 2016) within R (R Core Team 2015). Heatmaps were used to visualize the relative abundances of OTUs clustered into each ESTU, and to validate the associations, i.e. OTUs with similar distribution patterns were clustered together. The relative abundance of each OTU belonging to a cluster was summed up to represent the ESTU abundance.

The relative contribution of each ESTUs to the Bacteroidetes population was then calculated for each sample. Stacked barplots were used to visualize ESTUs. These ESTU profiles were compared using the Bray-Curtis similarity coefficient and clustered into groups by similarity using SIMPROF with an alpha level of 5% and 999 permutations (Somerfield & Clarke, 2013). The resulting groups were termed 'Environmental Groups' (EG), as they responded similarly to the environmental factors studied. PCoA ordination was used to visualize sample clustering, with environmental parameters overlaid on the ordination using distance based linear modelling. A generalised additive model (GAM) was further used to fit either temperature or depth contour lines to ordinations. Ordinations, overlaying of environmental parameters, and covariates were conducted using the R package 'vegan' (Oksanen *et al.* 2018).

The environmental parameters to which these data were compared included nutrients, chlorophyll *a*, salinity, organic carbon concentration, and temperature in the Blanes Bay Microbial Observatory and the methods used to obtain these variables are detailed in

Romera-Castillo *et al.* (2013) and Gasol *et al.* (2016). The environmental variables collected during the HOTMIX cruise are described in Martínez-Pérez *et al.* (2017a, 2017b).

Results

Diversity of Bacteroidetes using environmental clone libraries

The four libraries were sequenced with a similar effort (around 180 sequences each) using three overlapping primers to obtain almost complete 16S rRNA gene sequences. A total of 701 sequences were obtained, of which 55 were flagged as putative chimeras, resulting in a final dataset of 646 sequences. Chimeric sequences were found particularly abundant in the WS (34 out of a total of 55) and, as a consequence, this curated library contained a lower number of sequences (Table S5).

Cluster similarity curves of the 16S rRNA gene showed a drastic decrease in the number of distinct OTU clusters from 100 to 99% sequence similarity (Fig. S1A) suggesting significant sequence diversity within only 1% divergence (~ 10 nt). Likewise, accumulation curves approached asymptotes by 99% OTU definition in all libraries (Fig. S1B). The S library had the lowest number of OTUs at 99% cluster similarity (27; Table S5) compared with the other three libraries, which had relatively similar number of OTUs (41 – 46, Table S5). Good's coverage ranged from 80 to 94% among the libraries (Table S5) and indicated that, at 99% similarity clustering, the libraries provided a relatively reliable inventory of the Bacteroidetes ribosomal operons present in BBMO given the method used. The Chao-1 estimator of richness revealed that the SF library contained the largest number of predicted OTUs (78.1), followed by the WS library (71), the D library (65.5) and S library (40.7). However, the D library was the most diverse (Shannon = 3.11) given the evenness in abundance of the OTUs within this sample (Table S5). The S and SF libraries were the most similar ones, sharing

36% of OTUs, followed by the SF and WS libraries with 19% similarity shared between them. The D library had the least similarity (7.2, 9.6 and 10.3 %) with the other libraries (Table S6).

Taxonomic diversity and habitat-specific clades within the phylum Bacteroidetes

The 16S rRNA gene sequences obtained in this study presented 87-100% sequence similarity to their closest BLAST hit (including uncultured/environmental sequences, Table S7). Classification of the sequences by SINA and RDP sorted them into the Flavobacteriia (602 sequences), Sphingobacteriia (28 sequences), Cytophagia (7 sequences) and Bacteroidia (6 sequences) classes. Within the class Flavobacteriia, 59 sequences belonged to the family Cryomorphaceae and 543 sequences to the family Flavobacteriaceae (Table S7). Within the latter, the majority of sequences belonged to uncultured marine clusters NS2b, NS4, NS5, and NS7 (as defined by Alonso *et al.* 2007), but several phylotypes belonged to clusters that contained cultured organisms of the genera *Fluviicola*, *Owenweeksia*, *Winogradskiella*, *Formosa*, *Leeuwenhoekella*, *Muricauda*, *Zunongwangia*, or *Polaribacter* (Table S7). Some sequences were not classified further than to family level by any of the classifier tools used, and they are here referred to as ‘Unclassified’. Interestingly, genera with cultured organisms were overrepresented in deep environments (55% of the clone sequences), and represented a smaller fraction among the surface environments (12% of the clone sequences).

Using the phylogenetic tree based on all 16S rRNA gene sequences and the associated environmental categories, AdaptML identified 3 non-redundant ‘habitats’ (H_A, H_B, H_C; Fig. 1). Habitat H_A was dominated by sequences from surface environments during summer; habitat H_B indicated a preference for deep waters, and habitat H_C included winter specific communities in surface waters. It should be noted that we did not sample the combination of

winter and deep. The algorithm identified 24 clades, most of which presented a remarkable preference for one of the specific habitats (e.g. clades C4 and C12 associated to H_C included only Winter-Spring sequences in surface, C5-C8 from H_B presented only Summer-Fall sequences in deep waters). Only a few clades were more broadly distributed across different environments (e.g. C17, C21 and C22 associated to H_C consisted of a mix of SF and WS sequences mainly in shallow water; Fig. 1B).

Using the distribution of Bacteroidetes clades to define ESTUs

In order to infer Ecologically Significant Units (ESTUs), we analysed the distribution patterns of the Bacteroidetes clades previously delineated, using counts of OTUs (at the 99% identity) obtained from two amplicons 16S rRNA gene iTags datasets from the North-western Mediterranean Sea (see Material and Methods). In the temporal dataset (BBMO), 318 OTUs were recovered with >99% similarity to our clones from all Bacteroidetes OTUs (1,380 in the dataset), which corresponded to 42% of the total Bacteroidetes sequences. From these, 183 OTUs (20.6% of sequences) could be placed within the limits of our predefined clades in the 16S rRNA phylogenetic tree (Table S8). This was a stringent step that left half of the OTUs outside of the branch limits, but it was done to ensure that the OTUs were confidently related to the previous Bacteroidetes clades. As expected, clone sequences related to deep environments did not match many OTUs from the temporal scale study (C5, C6, C7, C8, C11, C18 and C20 included 0 to 3 OTUs, Table S8). Abundance and distribution could not be assessed for clades that did not include any OTU from the BBMO high-throughput sequencing dataset (i.e. six clades). Clades displaying more than one coherent OTU cluster with distinct distribution patterns, likely occupying different ecological niches, were split into independent ESTUs (Fig. S2A, Table S8). For instance, clades C16, C21 and C22 were divided into 4 ESTUs each (named A to D), C17 into 3 ESTUs, and C2, C3, C9 and C24 into

2 ESTUs each, based on clustering of their abundances per sample. Clades displaying a homogenous distribution pattern for all OTUs were not split but letter A was still added for consistency (Table S8).

In the spatial dataset (HOTMIX), 219 OTUs were recovered with > 99% similarity to our clones, from the 1,978 OTUs classified as Bacteroidetes in the dataset. The sequences included in those OTUs represented 32.6% of all Bacteroidetes sequences. From these, 147 OTUs (18.2% of Bacteroidetes sequences) were confidently placed within our predefined clades in the phylogenetic tree. In this case, only two clades did not include any OTUs (C4 assigned to WS community, and C5 to SF community), likely because we also included bathypelagic samples. Similarly, some clades had homogenous distributions of OTUs and were maintained as one ESTU. Others were split into 2 ESTUs each (i.e. C16 and C24), or 3 ESTUs (i.e. C7, C15, C17 and C23; Fig. S2B and Table S8). Thus, although our previous phylogenetic analysis divided Bacteroidetes clusters in several clades, which were based on sampling origin (i.e. seasonal and spatial), the inclusion of relative abundance values for these clades indicates that these ESTUs provide a finer estimate of Bacteroidetes ecologically meaningful diversity. Interestingly, when recruiting OTUs from the two 16S rRNA gene iTags using our clone sequences, we observed that these matched OTUs covered a wide range of relative abundances, from abundant to rare ones (see Fig. S3).

Temporal distribution patterns of Bacteroidetes ESTUs

The relative abundances of the identified ESTUs (with respect to the Bacteroidetes community) were examined for each monthly sample. Differential distribution of the ESTUs revealed five Environmental temporal Groups clustered by similarity (EtG; Fig. 2A). Environmental groups including single samples were not considered. The largest group

(EtG4) included 25 samples mostly from winter-spring samples (Fig. 2A-B). The distribution patterns of the bacterial assemblages within EtG4 seemed to be mainly driven by temperature, the concentration of Chl *a*, NO₃⁻ and Si, as indicated by the principal coordinate analyses (PCoA) (Fig. 2C). EtG5 formed by summer-fall samples, included assemblages mostly correlated to temperature and DOC concentration, whereas EtG1, EtG2 and EtG3 included samples from the transition period from late spring to early summer, particularly from April, May and June of several years (Fig. 2B).

Looking at the individual distribution of ESTUs (Fig. 2A), in the transition from EtG5 to EtG4 many ESTUs varied their abundances. The largest change occurred for ESTU 14A (belonging to the NS4 marine group), whose abundance increased notably. This ESTU 14A was positively correlated with Chl *a* concentration, salinity, NO₂⁻ and NO₃⁻ concentrations, and negatively with DOC concentration and temperature (Fig. 3A). However, ESTU 15A (also within the NS4 marine group) was already present in the previous months and did not show relevant abundance changes. ESTU 14A remained abundant in the next months (EtG 3) but ESTU 15A decreased to undetectable limits in certain months. Within the NS5 marine group, ESTUs 16A-D abundances decreased notably, but ESTUs 17A abundance increased. While ESTUs 16A-D were positively correlated with temperature and DOC concentration, and negatively with the concentration of Chl *a*, NO₂⁻, and NO₃⁻, the ESTU 17 A and C showed an opposite pattern (Fig S4A). Finally, within the NS2 marine group we observed an increase in ESTUs 22D and 24A-B, but a decrease in ESTU 22C, which became undetectable after winter. Clade 24 seemed to be a branch of the NS2 marine group dominant in mesopelagic waters (see below), and probably reached the surface at detectable levels during the winter mix.

Another drastic change in the Bacteroidetes communities occurred from EtG 4 to EtG 3 (roughly from March to April). In this case, the largest change involved Clade 21 (assigned to *Formosa*), which bloomed in March and contributed up to 47% of the Bacteroidetes community in EtG3, with ESTUs 21B and D showing higher abundances than 21A and C. The seven ESTUs forming the NS2 marine group, that were in general abundant over EtG4 and EtG5, decreased radically in numbers in EtG3. Clade 16 (of the NS5 marine group), after a decrease in the EtG4, showed a new increase in abundance for ESTU 16C. The other ESTUs belonging to this NS5 marine group, increased later on in EtG 2 when *Formosa*-related ESTUs dropped in abundance.

Spatial distribution patterns of Bacteroidetes ESTUs

Similarly, the spatial samples could be grouped in 7 Environmental spatial Groups (excluding groups with single samples) based on the distribution of the identified ESTUs (clustered by similarity, Fig. 4A). Environmental spatial Groups EsG1 and EsG2 included deep-water samples (Fig. 4A-B) that were correlated with high salinities and nutrient concentrations (Fig. 4C). EsG3 included mesopelagic samples, and EsG from 4 to 7 represented surface and DCM samples, linked to higher pH, O₂, and TOC. Samples from shallow waters were overall more similar to each other than any of the deep-water samples, as indicated by the short distance among them in the PCoA plot (Fig. 4C).

Distribution of individual ESTUs was characterized by the dominance of *Owenweeksia*-related ESTUs in deep-water samples, mostly ESTU 7C, followed by ESTUs 7A and 7B. Other ESTUs that were abundant and specific of deep waters were ESTUs 8A (unclassified *Cryomorphaceae*), ESTU 13A (*Leeuwenhoekiella*-related), ESTU 18A (a deep water branch of the NS5 marine group) and ESTU 19A (*Muricauda*-related). In mesopelagic samples,

ESTUs 7A-C was still abundant but ESTUs 24A-B became predominant (from a deep water branch of NS2b marine group). This ESTU 24B was positively correlated with depth and other factors such as salinity and nutrients, and negatively correlated with oxygen, pH, temperature, and TOC (Fig. 3B). For shallower waters, the assemblages were dominated by different ESTUs related to NS5 marine clade (i.e. ESTUs 16A-B and 17A-C), NS4 marine group (ESTUs 15A-C and 14A), followed by ESTUs 23A-C (NS2 marine group), and ESTU 21A (*Formosa*-related). All these ESTUs presented opposite correlations than those of the deep-water ESTUs (Fig. 3B).

Discussion

Our results indicate that Bacteroidetes clades can be divided into ecologically distinct units based on their phylogeny and also in their relative abundance in different environments. The ecological partition of Bacteroidetes clades has been poorly explored in comparison to other bacterial groups (e.g. Cyanobacteria) despite being one of the most abundant groups in the ocean and playing key functions in global carbon biogeochemical cycling. In this work, we performed a comprehensive study to increase the knowledge of marine Bacteroidetes bringing novel insights into the ecology of this group.

To characterise the fine-scale phylogenetic diversity of Bacteroidetes, a high quality reference tree was constructed from clone libraries including near full-length 16S rRNA gene sequences, in order to identify consistently supported clades. Phylogenetic trees that are constructed from short marker gene sequences, such as those from next-generation sequencing technologies typically lack resolution and struggle to confidently detect fine-scale community architectures (Acinas *et al.* 2004a; Singer *et al.* 2016)

Some mandatory methodological considerations in any study relying on the rRNA operon include the definition of the threshold for species discrimination, and the coexistence of non-identical *rrn* operons within a genome (Crosby & Criddle 2003). Comparative genomics of prokaryotic isolates suggested an average nucleotide identity (ANI) threshold range of 94-96% for species demarcation (Richter & Rosselló-Móra, 2009). These values corresponded to 98.2 - 99% of 16S rRNA sequence similarity (Kim *et al.* 2014). Therefore, the cut-off of 99% 16S rRNA sequence similarity used for OTU delineation in this work is appropriate for comparative purposes with other microbial ecology studies. Moreover, two thirds of the sequences grouped above the 99% sequence similarity, indicating a high microdiversity within Bacteroidetes species. Microdiversity clusters are commonly observed in bacterioplankton phylogenies, and it has been suggested that may be meaningful units of differentiation (Acinas *et al.* 2004a). On the other hand, multiple non-identical rRNA operons represent a source of systematic error in community analysis (Crosby & Criddle 2003). Bacterial genomes usually contain from 1 to 15 operons, but ca. 40% have either only one or two (Acinas *et al.* 2004b; Espejo & Plaza 2018). Although microheterogeneity between paralogous 16S rRNA genes is relatively common, the degree of sequence divergence in bacteria is rather low, with the vast majority showing < 1% nucleotide differences (Acinas *et al.* 2004a; Pei *et al.* 2010). 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 in bacteria (e.g. Coenye & Vandamme, 2003; Engene & Gerwick, 2011).

Saturated clone library accumulation curves indicated that all libraries reached their diversity plateau, suggesting sufficient sequencing depth with this approach. The SF (Summer-Fall) library presented the largest richness for Bacteroidetes populations, in contrast with the

common trend of whole bacterial communities, which show diversity peaks in winter (Gilbert *et al.* 2012). However, the deep library contained more diversity than the surface libraries, in agreement with previous observations for marine bacterioplankton communities (Pommier *et al.*, 2010; Sunagawa *et al.* 2015).

Delineation of ecologically distinct populations, based on different gene-based phylogenies, has been successfully applied using the AdaptML algorithm with both bacterial isolates and uncultured groups (Connor *et al.* 2010; Biller *et al.* 2012; Szabo *et al.* 2013; Shepherd *et al.* 2013; Chimetto Tonon *et al.* 2015). In our study, the model identified three non-redundant ‘habitats’ and 24 ecological clades associated with these habitats. The clades were clearly segregated by depth but showed some mixing between seasons. Most marine Bacteroidetes sequences retrieved in this study represent uncultured clusters (76% of clone sequences) for which physiological information is lacking, and thus the environmental preferences of the clades described here are difficult to determine. Nonetheless, the use of Bacteroidetes OTU counts based on Illumina 16S rRNA gene iTags associated with our clades helped to understand some of the group distribution and their relation with environmental parameters.

The Bacteroidetes assemblages from deep waters were substantially different from those of surface samples. In the spatial transect there was almost no overlap among ESTUs at different depths (the water was well stratified in the HOTMIX cruise; Martínez-Pérez *et al.* 2017b). The notable occurrence of depth-specific clades had previously been suggested for Bacteroidetes using the 16S rRNA gene in DGGE studies (Blümel *et al.* 2007; Díez-Vives *et al.* 2013). A plausible hypothesis for such niche partitioning is that depth-specific clusters could be related to different types of organic matter, which typically change through depth gradients (Santinelli *et al.* 2002; Hedges 2002; Benner 2002). Recent findings of a clear

depth stratification of Flavobacteria in the water column were revealed by the relative abundance of TonB-dependent transporters (TBDTs), which play a role in high-molecular weight organic matter uptake (Bergauer *et al.* 2017). In fact, the use of complex macromolecules varies among cultured genera of the class Flavobacteriia (Bernadet & Nakagawa, 2006), and genomic data also point to specialization of distinct flavobacterial clades for degradation of different polysaccharides and proteins (Bauer *et al.* 2006; Gómez-Pereira *et al.*, 2012; Fernández-Gómez *et al.*, 2013; Bennke *et al.* 2016). This specialization is in tune with our findings on the large marine clade NS2b, which despite seeming generalist as a whole, is likely constituted by a complex suite of specialists adapted to narrower niches. This marine clade NS2b was first described from coastal North Sea surface waters (Alonso *et al.* 2007), and ribotypes related to this group appeared in all our libraries. A first phylogenetic approximation distinguished three clades with different depth distributions, and further analyses of abundance distribution within these clades confirmed the presence of ESTUs with different niche preferences, not only related with depth but also with season.

Seasonally, Bacteroidetes presented multi-annual dynamic of recurring clusters, similarly to temporal patterns found in other bacterial phyla (Vergin *et al.* 2013, Restrepo-Ortiz *et al.* 2013). Specifically, five periods with fairly stable ESTU assemblages were identified (EtGs). The two longest stable periods occurred from July to October (EtG 5) and from November to April (EtG 4), with a notable transition from October to November. Likewise, drastic changes occurred from March to April. Groups that increased their abundance along this transition included *Formosa*-related and the NS5 marine group, while ESTUs related to NS2b and NS4 marine groups decreased considerably. Previous studies have shown that flavobacterial phylotypes are one of the dominant groups responding to phytoplankton blooms (Teeling, *et*

al. 2012 and 2016, Buchan *et al.* 2014). This response consists of a succession of particular clades, namely *Ulvibacter* spp., *Tenacibaculum*, *Formosa*, *Polaribacter* spp., NS9, NS4, NS5, NS3a marine clades, and others, related to the progressive consumption of the algal-derived organic matter, and the gene repertoires that allow the uptake and use of specific biopolymers during and after the bloom (Teeling *et al.* 2016, Unfried *et al.* 2018). We also detected a similar increase of *Formosa*-related and NS5 marine clade members after the chlorophyll peak in March. More importantly, we showed that these groups are not homogeneous but contain recognizable ‘Units’ (ESTUs) that respond differently to environmental conditions. Although we do not have information on the functional gene content of these Units, they most likely display distinct gene repertoires driving their adaptation to available substrates. We further observed that other temporal-specific ESTUs seemed to specialize in response to particularly changing environmental conditions (e.g. temperature, inorganic nutrients). It would be interesting to further investigate the temporal variability of Bacteroidetes in deep environments in the light of the observations presented here.

In conclusion, using fine-scale phylogenies we have successfully outlined the extent of diversity and structure of marine Bacteroidetes. The examination of the specific distribution of phylogenetic clades supported their spatial and temporal niche partitioning, and hence their identification as ecologically significant taxonomic units (ESTUs). A number of ESTUs, from both surface and deep water samples, were taxonomically associated to genera including bacterial isolates, however the most abundant ESTUs could not be related to previously isolated taxa preventing their phenotypic description. Despite that, we could identify some physico-chemical variables that could be acting as drivers, shaping the distribution and evolution of the various subgroups of this abundant group of heterotrophic bacteria.

Acknowledgements

We thank M. Coll, V. Balagué, M. Zamanillo and the Blanes Bay sampling team for collecting the samples used in these analyses; E. Garcés for the BBMO sequencing analyses; J. Arístegui, M. Álvarez, X.A. Alvarez-Salgado, V. Vieitez and M.J. Pazó for the HOTMIX data, and Dr. Frederick Cohan for hosting CDV in his laboratory at Wesleyan University. This work was supported by grants MAGGY (CTM2017-87736-R) to SGA and REMEI (CTM2015-70340-R) to JMG from the Spanish Ministry of Science and Innovation (MICINN). MS was supported by a Viera y Clavijo contract funded by the ACIISI and the ULPGC. 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.

References

- Acinas SG, Klepac-Ceraj V, Hunt D *et al.* (2004a) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature*, **430**, 551–554.
- Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF (2004b) Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *Journal Of Bacteriology*, **186**, 2629–2635.
- Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz M (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology*, **71**, 8966–8969.
- Acinas SG, Ferrera I, Sarmiento H, Díez-Vives C, Forn I, Ruiz-González C, Cornejo-Castillo FM, Salazar G, Gasol JM (2015) Validation of a new catalysed reporter deposition–fluorescence *in situ* hybridization probe for the accurate quantification of marine Bacteroidetes populations. *Environmental Microbiology*, **17**, 3557–3569.
- Alonso C, Warnecke F, Amann R, Pernthaler J (2007) High local and global diversity of Flavobacteria in marine plankton. *Environmental Microbiology*, **9**, 1253–1266.
- Alonso-Sáez L, Gasol JM (2007) Seasonal variations in the contributions of different bacterial groups to the uptake of low-molecular-weight compounds in northwestern Mediterranean coastal waters. *Applied and Environmental Microbiology*, **73**, 3528–3535.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool (BLAST). *Journal of Molecular Biology*, **215**, 403–410.

- 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. *Applied and Environmental Microbiology*, **72**, 5734–5741.
- 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. *Environmental Microbiology*, **8**, 2201–2213.
- Benner R (2002) Chemical composition and reactivity. In Carlson CA & DA Hansell (eds) Biogeochemistry of marine dissolved organic matter. Elsevier, pp. 59-90.
- Bennke CM, Krüger K, Kappelmann L, *et al.* (2016) Polysaccharide utilisation loci of Bacteroidetes from two contrasting open ocean sites in the North Atlantic. *Environmental Microbiology*, **18**, 4456–4470.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. *Nucleic Acids Research*, **41**, D36–42.
- Bergauer K, Fernández-Guerra A, Garcia JAL, Sprenger RR, Stepanauskas R, Pachiadaki MG *et al.* (2018) Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, E400–E408.
- Bernardet JF, Nakagawa Y (2006) An introduction to the family *Flavobacteriaceae*. In: Dworkin M, Flakow S, Rosenberg E, Schleifer KH, Stackebrandt E (ed.) The Prokaryotes. 3rd edn., vol 7. Springer, New York, pp. 455-480.
- Biller SJ, Mosier AC, Wells GF, Francis CA (2012) Global Biodiversity of Aquatic Ammonia-Oxidizing Archaea is Partitioned by Habitat. *Frontiers in microbiology*, **3**, 252.
- Blümel M, Süling J, Imhoff JF (2007) Depth-specific distribution of *Bacteroidetes* in the oligotrophic eastern Mediterranean Sea. *Aquatic Microbial Ecology*, **46**, 209–224.
- Buchan A, LeClerc GR, Gulvik CA, González JM (2014) Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nature Reviews Microbiology*, **12**, 686–698.
- Carlson CA, Morris R, Parsons R *et al.* (2009) Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *The ISME journal*, **3**, 283–295.
- Chao A, Colwell RK, Lin C-W, Gotelli NJ (2009) Sufficient sampling for asymptotic minimum species richness estimators. *Ecology*, **90**, 1125–1133.
- Chen X, Zeng Y, Jiao N (2006) Development and evaluation of specific 16S rDNA primers for marine *Cytophaga-Flavobacteria* cluster. *Molecular Ecology Notes*, **6**, 1278–1281.
- Chen X, Zeng Y, Jiao N (2008) Characterization of *Cytophaga-Flavobacteria* community structure in the Bering Sea by cluster-specific 16S rRNA gene amplification analysis. *Journal of Microbiology and Biotechnology*, **18**, 194–198.
- Chimetto Tonon LA, Silva BS de O, Moreira APB *et al.* (2015) Diversity and ecological structure of vibrios in benthic and pelagic habitats along a latitudinal gradient in the Southwest Atlantic Ocean. *PeerJ*, **3**, e741.
- Coenye T, Vandamme P (2003) Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiology Letters*, **228**, 45–49.
- Cohan FM (2002) What are bacterial species? *Annual Review of Microbiology*, **56**, 457–487.
- Cohan FM (2006) Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, **361**, 1985–1996.
- Cohan FM, Perry E (2007) A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol*, **17**, R373–86.

- Connor N, Sikorski J, Rooney AP *et al.* (2010) Ecology of Speciation in the Genus *Bacillus*. *Applied and Environmental Microbiology*, **76**, 1349–1358.
- Cottrell M, Kirchman D (2000) Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. *Applied and Environmental Microbiology*, **66**, 5116–5122.
- Crosby LD, Criddle CS (2003) Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. *BioTechniques*, **34**, 790–4, 796, 798 passim.
- Diez-Vives C, Gasol JM, Acinas SG (2012) Evaluation of marine Bacteroidetes-specific primers for microbial diversity and dynamics studies. *Microbial Ecology*, **64**, 1047–1055.
- Diez-Vives C, Gasol JM, Acinas SG (2013) Spatial and temporal variability among marine *Bacteroidetes* populations in the NW Mediterranean Sea. *Systematic Applied Microbiology*, **37**, 68–78.
- Doolittle WF, Papke RT (2006) Genomics and the bacterial species problem. *Genome Biology*, **7**, 116.
- Engene N, Gerwick WH (2011) Intra-genomic 16S rRNA gene heterogeneity in cyanobacterial genomes. *Fottea*, **1**, 1–8.
- Espejo RT, Plaza N (2018) Multiple Ribosomal RNA Operons in Bacteria; Their Concerted Evolution and Potential Consequences on the Rate of Evolution of Their 16S rRNA. *Frontiers in Microbiology*, **9**, 2629–6.
- Farrant GK, Doré H, Cornejo-Castillo FM *et al.* (2016) Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci U S A*, **113**, E3365–74.
- Fernández-Gómez B, Richter M, Schüller M *et al.* (2013) Ecology of marine Bacteroidetes: a comparative genomics approach. *The ISME journal*, 1–12.
- Ferris MJ., Kühl M, Wieland A, Ward DM (2003) Cyanobacterial Ecotypes in Different Optical Microenvironments of a 68°C Hot Spring Mat Community Revealed by 16S-23S rRNA Internal Transcribed Spacer Region Variation. *Applied and Environmental Microbiology*, **69**, 2893–2898.
- Fraser C, Hanage WP, Spratt BG (2007) Recombination and the Nature of Bacterial Speciation. *Science (New York, NY)*, **315**, 476–480.
- Gasol JM, Cardelús C, Morán XAG *et al.* (2016) Seasonal patterns in phytoplankton photosynthetic parameters and primary production at a coastal NW Mediterranean site. *Scientia Marina*, **80**, 63–77.
- Gilbert JA, Steele JA, Caporaso JG, *et al.* (2012) Defining seasonal marine microbial community dynamics. *The ISME Journal*, **6**, 298–308.
- Gómez-Pereira PR, Fuchs BM, Alonso C *et al.* (2010) Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. *The ISME journal*, **4**, 472–487.
- Gómez-Pereira PR, Schüller M, Fuchs BM, *et al.* (2012) Genomic content of uncultured Bacteroidetes from contrasting oceanic provinces in the North Atlantic Ocean. *Environmental Microbiology*, **14**, 52–66.
- Good IJ (1953) The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika*, **40**, 237–264.
- Gordon DA, Giovannoni SJ (1996) Detection of stratified microbial populations related to *Chlorobium* and *Fibrobacter* species in the Atlantic and Pacific oceans. *Applied and Environmental Microbiology*, **62**, 1171–1177.
- Guindon S, Dufayard J-F, Lefort V *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

- Hedges JI (2002) Why dissolved organic matter? In Carlson CA & DA Hansell (eds) Biogeochemistry of marine dissolved organic matter. Elsevier, pp. 1-33.
- Hill MO (1973) Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, **54**, 427–432.
- Hunt DE, David LA, Gevers D *et al.* (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science (New York, NY)*, **320**, 1081–1085.
- Jaccard P (1901) Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, **37**, 547–579.
- Kim M, Oh HS, Park SC, Chun J (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, **64**, 346–351.
- Kirchman DL (2002) The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS microbiology ecology*, **39**, 91–100.
- Kirchman DL, Yu L, Cottrell MT (2003) Diversity and abundance of uncultured *Cytophaga*-like bacteria in the Delaware Estuary. *Applied and Environmental Microbiology*, **69**, 6587–6596.
- Lane DJ (1991). 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds). *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley & Sons: New York, NY, USA. pp 115–147.
- Letunic I, Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic acids research*, **39**, W475–W478.
- Locey KJ and Lennon JC (2016) Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 5970–5975.
- Ludwig W, Strunk O, Westram R *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Research*, **32**, 1363–1371.
- 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**, 1097–1106.
- Martínez-Pérez AM, Nieto-Cid M, Osterholz H, *et al.* (2017a). Linking optical and molecular signatures of dissolved organic matter in the Mediterranean Sea. *Scientific Reports*, **7**, 487–11.
- Martínez-Pérez AM, Osterholz H, Nieto-Cid M, *et al.* (2017b) Molecular composition of dissolved organic matter in the Mediterranean Sea. *Limnology and Oceanography*, **62**, 2699–2712.
- Massana R, Murray AE, Preston CM, Delong EF (1997) Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Applied and Environmental Microbiology*, **63**, 50–56.
- Moore L, Rocap G, Chisholm S (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature*, **393**, 464–467.
- Morris RM, Rappé MS, Urbach E, Cannon SA, Giovannoni SJ (2004) Prevalence of the *Chloroflexi*-related SAR202 bacterioplankton cluster throughout the mesopelagic zone and deep ocean. *Applied and Environmental Microbiology*, **70**, 2836–2842.
- 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 journal*, **2**, 379–392.
- Oksanen J, Blanchet FG, Kindt R *et al.* (2018) Vegan: Community Ecology Package. R package version 2.0-2, 1–293.

- Pei AY, Oberdorf WE, Nossa CW *et al.* (2010) Diversity of 16S rRNA Genes within Individual Prokaryotic Genomes. *Applied and Environmental Microbiology*, **76**, 3886–3897.
- Polz M, Hunt D, Preheim S, Weinreich D (2006) Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, **361**, 2009–2021.
- Pommier T, Neal P, Gasol JM *et al.* (2010) Spatial patterns of bacterial richness and evenness in the NW Mediterranean Sea explored by pyrosequencing of the 16S rRNA. *Aquatic Microbial Ecology*, **61**, 221–233.
- Preheim SP, Timberlake S, Polz MF (2011) Merging taxonomy with ecological population prediction in a case study of *Vibrionaceae*. *Applied and Environmental Microbiology*, **77**, 7195–7206.
- Pruesse E, Peplies J, Glöckner FO (2012) SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes, **28**, 1823–1829.
- Quast C, Pruesse E, Yilmaz P *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, **41**, D590–6.
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- 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. *Environmental Microbiology*, **16**, 689–700.
- Retchless AC, Lawrence JG (2007) Temporal fragmentation of speciation in bacteria. *Science (New York, NY)*, **317**, 1093–1096.
- Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA*, **106**, 19126–19131.
- 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. *Applied and Environmental Microbiology*, **68**, 1180–1191.
- Romera-Castillo C, Álvarez-Salgado XA, Galí M, Gasol JM, Marrasé C (2013) Combined effect of light exposure and microbial activity on distinct dissolved organic matter pools. A seasonal field study in an oligotrophic coastal system (Blanes Bay, NW Mediterranean). *Marine Chemistry*, **148**, 44–51.
- Santinelli C, Gasparini GP, Nannicini L, Seritti A (2002) Vertical distribution of dissolved organic carbon (DOC) in the western Mediterranean Sea in relation to the hydrological characteristics. *Deep-Sea Research Part I: Oceanographic Research Papers*, **49**, 2203–2219.
- Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.
- Scrucca L, Fop M, Murphy TB, Raftery AE (2016) mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R Journal*, **8**, 289–317.
- Shepherd MA, Fleming VM, Connor TR *et al.* (2013) Historical zoonoses and other changes in host tropism of *Staphylococcus aureus*, identified by phylogenetic analysis of a population dataset, *PLoS One*, **8**, e62369.
- Singer E, Bushnell B, Coleman-Derr D *et al.* (2016) High-resolution phylogenetic microbial community profiling. *The ISME Journal*, **10**, 2020–2032.

- Somerfield PJ, Clarke KR (2013) Inverse analysis in non-parametric multivariate analyses: distinguishing groups of associated species which covary coherently across samples. *Journal of Experimental Marine Biology and Ecology*, **449**, 261–273.
- Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K *et al.* (2015) Structure and function of the global ocean microbiome. *Science*, **348**, 1261359.
- Szabo G, Preheim SP, Kauffman KM *et al.* (2013) Reproducibility of *Vibrionaceae* population structure in coastal bacterioplankton. *The ISME journal*, **7**, 509–519.
- Teeling H, Fuchs BM, Becher D *et al.* (2012) Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. *Science*, **336**, 608–611.
- Teeling H, Fuchs BM, Bennke CM *et al.* (2016) Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *eLife*, **5**, e11888.
- Unfried F, Becker S, Robb CS *et al.* (2018) Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms. *The ISME Journal*, **12**, 2894–2906.
- Vergin KL, Beszteri B, Monier A *et al.* (2013) High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series study site by phylogenetic placement of pyrosequences. *The ISME journal*, **7**, 1322–1332.
- West NJ, Scanlan DJ (1999) Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Applied and Environmental Microbiology*, **65**, 2585–2591.
- Wiedenbeck J, Cohan FM (2011) Origins of bacterial diversity through horizontal gene transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews*, **35**, 957–976.

Data Accessibility

DNA sequences from clone libraries: Genbank accessions KP887163 - KP887808.

High-throughput sequences: Genbank Sequence Read Archive under Bioproject ID PRJNA495818.

Author Contributions

C.D.V. and S.G.A. designed research and J.M.G. sample collection. C.D.V. performed the research and the laboratory work. C.D.V., S.N. and P.S. analysed the data. IF and MS contributed with HTS data. C.D.V. wrote the paper and J.M.G, O.P. C.P.A, M.S, I.F and S.G.A. contributed valuable suggestions and critical reading.

Tables and Figures

Figure 1. Season and depth distributions of Bacteroidetes sequences with habitat predictions mapped onto the phylogeny inferred by maximum likelihood analysis of 16S rRNA gene sequences. **(A)** Phylogenetic tree of all Bacteroidetes sequences; outer capital letters correspond to their taxonomy (see Table S7): FF = family *Flavobacteriaceae* (within class *Flavobacteriia*), FC = family *Cryomorphaceae* (within class *Flavobacteriia*), S = class *Sphingobacteriia*, C = class *Cytophagia*, B = class *Bacteroidia*. Clusters at the genus level are indicated by lowercase letters: a) *Reichenbachiella* and *Roseivirga*; b) *Chitinophaga*; c) NS11-12 marine group and Unclassified 1; d) *Marinilabiaceae* and Unclassified 2; e) *Owenweeksia*; f) Unclassified 3; g) NS7 marine group and Unclassified 4; h) Unclassified 5; i) *Crocimitomix* and Unclassified 6; j) *Fluviicola*; k) *Zunongwangia*, *Saligentibacter*, *Salinimicrobium* and *Croceibacter*; l) *Leeuwenhoekiella*; m) *Marixanthomonas*, *Ulvibacter*, *Kordia* and Unclassified 7; n) *Polaribacter-Tenacibaculum*; o) *Myroides* and *Flavobacterium*; p) NS4 marine group; q) NS5 marine group; t) *Muricauda*; r) *Sedimibacter*; s) *Croceitalea*; t) Unclassified 8; u) *Mesoflavibacter*; u) *Psychroserpens*; v) *Winogradskyella*; w) *Formosa*; x) NS2b marine group. Surrounding colored rings indicate the origin of each sequence: season (inner ring) and depth (outer ring). The codes C1 to C24 correspond to the different ecological clades predicted by AdaptML, and the color shading of these clusters refer to one of the 3 habitats predicted (H_A , H_B and H_C , see Fig 1C) **(B)** Ultrametric tree summarizing habitat-associated units identified by the model and the distribution of each unit among seasons and depths. The clade numbers (C1-C24) match the clusters within the phylogenetic tree. **(C)** Three predicted habitats with distributions of seasons and depths inferred by the model. The habitat legend matches the colored circles used in sections A and B of the figure. To account for effects of uneven sampling, distributions were normalized by the total number of counts from each environmental category.

Figure 2. Temporal distribution of Bacteroidetes ESTUs in the Blanes Bay Microbial Observatory. **(A)** Histogram of the relative abundance of each ESTUs in the monthly samples sorted by similarity. Environmental Group numbers are indicated with vertical bars on the left. Clusters of single samples were not described as environmental groups (grey bars) **(B)** Distribution of samples, color-coded by their Environmental Group as in A, over a 5-year period in the BBMO. **(C)** PCoA analysis of samples according to Bray-Curtis distance between Bacteroidetes assemblages, with fitted statistically significant (adjusted P value < 0.05) physico-chemical parameters. Samples are color-coded as in panel A and clusters are gathered by a background shade of the same color.

Figure 3. Heatmap showing the correlation analysis between ESTUs and various environmental parameters. **(A)** Bacteroidetes ESTUs at a temporal scale during 5 years. **(B)** Bacteroidetes ESTUs at a spatial scale over eight Western Mediterranean stations and with depth. The scale shows the degree of positive correlation (blue) or negative correlation (red) between the two sets of data. Crossing lines indicate correlations with an adjusted P value > 0.05.

Figure 4. Spatial distribution of Bacteroidetes ESTUs across the Western Mediterranean Sea. **(A)** Histogram of the relative abundance of each ESTUs in the spatial samples sorted by similarity. Environmental Groups Numbers are indicated with vertical bars on the left. Clusters of single samples were not described as environmental groups (grey bars) **(B)** Distribution of samples, color-coded by their Environmental Group as in A, for the cruise HOTMIX. Samples from the cruise Modivus are colored in black and are indicated by 'S' for the Surface library sampling point, and 'D' for the Deep library sampling point. **(C)** PCoA analysis of samples according to Bray-Curtis distance between Bacteroidetes assemblages, with fitted statistically significant (adjusted P value < 0.05) physico-chemical parameters. Samples are color-coded as in panel A and clusters are gathered by a background shade of the same color.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

A) Phylogenetic tree of all Bacteroidetes sequences

B) Tree summarizing habitat-associated units

A) Relative abundance of each ESTUs in the monthly samples

B) Sample distribution by month, Chl *a* conc. and EtG

C) PCoA plot of samples with fitted env. parameters

A) Relative abundance of each ESTU in the spatial samples

B) Sample distribution by location, depth and EsG

C) PCoA plot of samples with fitted env. parameters