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

Cristina Díez-Vives, Shaun Nielsen, Pablo Sánchez, Oswaldo Palenzuela, Isabel Ferrera, Marta Sebastián, Carlos Pedrós-Alió, Josep M. Gasol and Silvia G. Acinas

Table of Contents:

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods S1</td>
<td>2</td>
</tr>
<tr>
<td>Methods S2</td>
<td>3</td>
</tr>
<tr>
<td>Figure S1</td>
<td>5</td>
</tr>
<tr>
<td>Figure S2</td>
<td>6</td>
</tr>
<tr>
<td>Figure S3</td>
<td>7</td>
</tr>
<tr>
<td>Tables S1-S8</td>
<td>Excel file</td>
</tr>
</tbody>
</table>
Methods S1

**Bacteroidetes-specific DNA amplification**

DNA was amplified for subsequent cloning using two Bacteroidetes-specific primers. The forward primer, close to the 5’ end of the 16S rRNA gene (CF319aF; Manz *et al.* 1996), has good coverage for class Flavobacteriia (Table S2), which is the dominant class in marine environments (Kirchman *et al.*, 2003; Alonso *et al.*, 2007; Chen *et al.*, 2008), but it has been reported to have some cross amplification with other phyla (Díez-Vives *et al.* 2012). To overcome this problem we designed a new phylum-specific reverse primer in the 23S ribosomal subunit (CF434R; 5’ CACTATCGGTCTCTCAGG 3’), with 60.7% coverage of the class Flavobacteriia, but high specificity for Bacteroidetes sequences (> 99.4 % for any group, Table S2). This reverse primer was designed in ARB (Ludwig *et al.* 2004) using the *Probe Design* tool and a specific Bacteroidetes dataset retrieved from Silva LSU Ref 108 database that contained 23,600 sequences (Pruesse *et al.* 2012). *In silico* evaluation of the primer pair (CF319a and CF434R) performance was done using the *TestPrime* online tool in Silva website (<http://www.arb-silva.de/search/testprobe/> using the Silva SSU Ref NR 115 database with 1,426,414 sequences (Quast *et al.* 2013) (Table S2).

PCR reactions (50 µL volume) contained 200 µM each deoxynucleoside triphosphates, 0.2 µM of each primer, 2 mM MgCl₂, 1x PCR buffer, and 1 U Taq DNA polymerase (Invitrogen). PCR cycles consisted of an initial denaturation step (94 ºC, 5 min); 27 cycles of DNA denaturation (94 ºC, 30 s), primer annealing (55 ºC, 30 s), and Taq extension (72 ºC, 1 min); and a final extension step (72 ºC, 10 min). A reconditioning PCR was made for each sample to minimize the occurrence of heteroduplex amplicons as described in Acinas *et al.* (2005). Briefly, PCR products were analysed 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 concentrator. 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 amplification product of these primers included the almost complete 16S rRNA gene, the internal transcribed spacer, and 434 bases of the 23S rRNA gene (ca. 2,000 bp).
Methods S2

Generation of high-throughput sequencing (HTS) data

Both spatial samples from cruise HOTMIX and temporal ones from monthly sampling at the BBMO were filtered and DNA extracted following the same procedures explained in the main manuscript except that total nucleic acids from the HOTMIX samples were extracted using the PowerWater Sterivex™ DNA isolation Kit (MO BIO Laboratories, Carlsbad, CA) following the manufacturer instructions. DNA was quantified using a Qubit fluorometer assay (Life Technologies, Paisley, UK). The V4-V5 region of the 16S rRNA gene was amplified with the primers 515F-Y (5’-GTGYCAGCMGCGCGGTAA-3’) and 926R (5’-CCGYCAATTYMTTTRAGTTT-3’) (Parada et al. 2016) in the HOTMIX samples and the V3-V4 region was amplified for the BBMO series using primers 341F (5’-CCTACGGGNGGCWGCAG-3’) (Herlemann et al. 2011) and 806RB (5’-GGACTACNVGGGTWTCTAAT-3’) (Apprill et al. 2015) and sequenced in an Illumina MiSeq platform using 2x250bp paired-end approach at the Research and Testing Laboratory facility (Lubbock, Texas, USA; rtlgenomics.com). The reason for the different primers used in both studies can be traced to the timing of sequencing: the Parada et al. (2016) primers were introduced after sequencing the BBMO samples.

Computing analyses were carried out at the Marine Bioinformatics Service of the Institut de Ciències del Mar (ICM-CSIC) in Barcelona. Illumina sequences were processed using an in-house pipeline (Logares 2017). The sequences were clustered into operational taxonomic units (OTUs) at 99% cut-off using the UPARSE algorithm implemented in USEARCH (Edgar 2013). Singletons (i.e. OTUs occurring once in just one sample) and chimerical OTUs were removed. The remaining OTUs were taxonomically annotated using BLAST against the SILVA123 database as reference. OTUs assigned to chloroplasts and mitochondria were removed for subsequent analyses. Data treatment was carried out with R (version 3.3.2) and RStudio software (version 1.0.44). We used the `rrarefy` function (vegan package, Oksanen et al., 2015) to randomly subsample the OTU table down to the minimum reads per sample to avoid statistical artefacts due to an uneven sequencing effort among samples.
References in Methods S2


Figure S1

(A) Cluster similarity curves of the spatial and the temporal 16S rRNA gene clone libraries of marine Bacteroidetes in the Mediterranean Sea. The figure shows the number of unique clusters as a function of decreasing sequence similarity within a cluster. (B) Accumulation curves of the spatial and the temporal clone libraries at 99% similarity OTU definition. D = Deep community, SF = Summer-Fall community, S = Surface community, WS = Winter-Spring community.
Figure S2

**Figure S2.** Determination of ESTUs based on the temporal (A) or spatial (B) distribution patterns of OTUs within each clade. For each sample (months for A and stations for B), the number of reads assigned to a given OTU is normalized by the total number of reads assigned to the clade at that month. OTUs are hierarchically clustered (Bray-Curtis distance) according to the distribution pattern. Only Bacteroidetes clades that were split into different ESTUs are shown (see also Table S6).
Figure S3

**A)**

![Graph A](image)

- **all Bacteroidetes OTUs (1380)**
- **Bacteroidetes OTUs in clades (183)**

**B)**

![Graph B](image)

- **all Bacteroidetes OTUs (1978)**
- **Bacteroidetes OTUs in clades (147)**

**Figure S3.** The ranked mean relative abundances of all Bacteroidetes OTUs (black) and Bacteroidetes OTUs within our phylogenetic clades (coloured, see legend) from temporal samples (A, n=64) and spatial samples (B, n=65).