



# Exploring Microdiversity in Novel *Kordia* sp. (Bacteroidetes) with Proteorhodopsin from the Tropical Indian Ocean via Single Amplified Genomes

Marta Royo-Llonch<sup>1</sup>, Isabel Ferrera<sup>1</sup>, Francisco M. Cornejo-Castillo<sup>1</sup>, Pablo Sánchez<sup>1</sup>, Guillem Salazar<sup>1</sup>, Ramunas Stepanauskas<sup>2</sup>, José M. González<sup>3</sup>, Michael E. Sieracki<sup>4</sup>, Sabrina Speich<sup>5</sup>, Lars Stemann<sup>6</sup>, Carlos Pedrós-Alió<sup>7</sup> and Silvia G. Acinas<sup>1,8\*</sup>

## OPEN ACCESS

### Edited by:

Roland Hatzepichler,  
Montana State University,  
United States

### Reviewed by:

Jayashree Ray,  
Lawrence Berkeley National  
Laboratory, United States  
Christian Rinke,  
The University of Queensland,  
Australia  
Albert Leopold Mueller,  
Stanford University, United States

### \*Correspondence:

Silvia G. Acinas  
sacinas@icm.csic.es

### Specialty section:

This article was submitted to  
Microbial Physiology and Metabolism,  
a section of the journal  
Frontiers in Microbiology

**Received:** 10 February 2017

**Accepted:** 29 June 2017

**Published:** 25 July 2017

### Citation:

Royo-Llonch M, Ferrera I,  
Cornejo-Castillo FM, Sánchez P,  
Salazar G, Stepanauskas R,  
González JM, Sieracki ME, Speich S,  
Stemann L, Pedrós-Alió C and  
Acinas SG (2017) Exploring  
Microdiversity in Novel *Kordia* sp.  
(Bacteroidetes) with Proteorhodopsin  
from the Tropical Indian Ocean via  
Single Amplified Genomes.  
Front. Microbiol. 8:1317.  
doi: 10.3389/fmicb.2017.01317

<sup>1</sup> Department of Marine Biology and Oceanography, Institut de Ciències del Mar, Consejo Superior de Investigaciones Científicas, Barcelona, Spain, <sup>2</sup> Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States, <sup>3</sup> Department of Microbiology, University of La Laguna, La Laguna, Spain, <sup>4</sup> National Science Foundation, Arlington, VA, United States, <sup>5</sup> École Normale Supérieure, Département de Géosciences, Laboratoire de Météorologie Dynamique, UMR 8539 ENS-CNRS- École Polytechnique, Paris, France, <sup>6</sup> Sorbonne Universités, UPMC Université Paris 06, CNRS, Laboratoire d'Océanographie de Villefranche (LOV) UMR7093, Observatoire Océanologique, Villefranche-sur-Mer, France, <sup>7</sup> Systems Biology Program, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, Madrid, Spain, <sup>8</sup> Departament de Genètica i de Microbiologia, Facultat de Biociències, Universitat Autònoma de Barcelona, Bellaterra, Spain

Marine Bacteroidetes constitute a very abundant bacterioplankton group in the oceans that plays a key role in recycling particulate organic matter and includes several photoheterotrophic members containing proteorhodopsin. Relatively few marine Bacteroidetes species have been described and, moreover, they correspond to cultured isolates, which in most cases do not represent the actual abundant or ecologically relevant microorganisms in the natural environment. In this study, we explored the microdiversity of 98 Single Amplified Genomes (SAGs) retrieved from the surface waters of the underexplored North Indian Ocean, whose most closely related isolate is *Kordia algicida* OT-1. Using Multi Locus Sequencing Analysis (MLSA) we found no microdiversity in the tested conserved phylogenetic markers (16S rRNA and 23S rRNA genes), the fast-evolving Internal Transcribed Spacer and the functional markers proteorhodopsin and the beta-subunit of RNA polymerase. Furthermore, we carried out a Fragment Recruitment Analysis (FRA) with marine metagenomes to learn about the distribution and dynamics of this microorganism in different locations, depths and size fractions. This analysis indicated that this taxon belongs to the rare biosphere, showing its highest abundance after upwelling-induced phytoplankton blooms and sinking to the deep ocean with large organic matter particles. This uncultured *Kordia* lineage likely represents a novel *Kordia* species (*Kordia* sp. CFSAG39SUR) that contains the proteorhodopsin gene and has a widespread spatial and vertical distribution. The combination of SAGs and MLSA makes a valuable approach to infer putative ecological roles of uncultured abundant microorganisms.

**Keywords:** MLSA, SAGs, Bacteroidetes, *Kordia*, proteorhodopsin

## INTRODUCTION

The phylum Bacteroidetes is the third most abundant group of bacteria in the oceans (Kirchman, 2002) but has been poorly studied at the species level compared to the two other main marine microbial phyla, i.e., Proteobacteria and Cyanobacteria. Bacteroidetes is a cosmopolitan phylum that typically constitutes between 4 and 22% of marine bacterioplankton cells (Glöckner et al., 1999; Cottrell and Kirchman, 2000; Alonso-Sáez et al., 2007; Ruiz-González et al., 2012; Lefort and Gasol, 2013; Acinas et al., 2014). The relative abundance of Bacteroidetes can reach up to 53% and it often correlates with phytoplankton blooms (Abell and Bowman, 2005; Fandino et al., 2005). Bacteroidetes are proficient at the degradation of particulate organic matter (POM) (Cottrell and Kirchman, 2000; Gómez-Pereira et al., 2012; Fernández-Gómez et al., 2013; Swan et al., 2013) and present a generally high growth rate when substrate is available (e.g., following phytoplankton blooms) (Kirchman, 2002; Ferrera et al., 2011; Buchan et al., 2014). Some representatives contain the gene coding for the light-dependent proton-pump proteorhodopsin, that has been shown to stimulate bacterial growth or work as a source of additional energy in the presence of light in some strains (Gómez-Consarnau et al., 2007; Stepanauskas and Sieracki, 2007; González et al., 2008; Woyke et al., 2009; Martínez-García et al., 2012). Although Bacteroidetes can be found in the free-living plankton size fraction, they seem to be predominantly particle-attached (DeLong et al., 1993; Schattenhofer et al., 2009; Crespo et al., 2013; Díez-Vives et al., 2014; Salazar et al., 2015).

Population genetics studies of marine Bacteroidetes are scarce, and have been generally conducted using very small populations of isolates or comparing several distant species (Fernández-Gómez et al., 2013; Swan et al., 2013), giving a general overview of common genotypic traits and differences above the species level. Single cell sequencing enables the retrieval of discrete microbial genomes from natural samples. A direct link between phylogenetic markers and functional information from single cells allows a finer resolution in microdiversity studies (Stepanauskas and Sieracki, 2007). When combined with multi locus sequencing analysis (MLSA), this approach can easily provide a robust phylogenetic reconstruction of the selected population. MLSA has been widely used for population genetics analyses since it is a feasible approach for delineating ecologically relevant units within species (Gevers et al., 2005). It is based on the sequencing of several marker genes and so far, its use in marine bacterial taxa has been limited to a few cultured groups such as *Alteromonas macleodii* (Ivars-Martínez et al., 2008), members of *Vibrio* (Thompson et al., 2005) including *Vibrio cholerae* (Boucher et al., 2011), *Tenacibaculum maritimum* (Habib et al., 2014), or *Prochlorococcus* (Kashtan et al., 2014). Since it is well known that many ecologically relevant bacterial taxa elude isolation in culture (Rappé and Giovannoni, 2003), the combination of single cell genomics of environmental samples and MLSA can help in the exploration of intra-specific genetic variability as well as the different evolutionary processes involved in microbial speciation (such as homologous recombination and divergent selection).

The genus *Kordia* (Bacteroidetes, Flavobacteriaceae) contains representatives with isolation sources suggesting different lifestyles and habitats: (i) marine surface seawater for *Kordia antarctica* (Baek et al., 2013), *K. aquimaris* (Hameed et al., 2013), and *K. algicida* (Sohn et al., 2004), (ii) surface freshwater for *K. zhangzhouensis* (Lai et al., 2015), (iii) the interphase between the ocean and a freshwater spring for *K. jejudonensis* (Park et al., 2014), (iv) the surface of the green alga *Ulva* sp. for *K. ulvae* (Qi et al., 2016) or (v) the digestive tract of a marine polychaete for *K. periserrulae* (Choi et al., 2011).

Here we report the first MLSA analysis of the genus *Kordia*, performed with 98 Single Amplified Genomes (SAGs) from the genus *Kordia* that were retrieved from a seawater sample from the North Indian Ocean, a location subjected to seasonal monsoon winds and coastal upwelling events as well as open ocean oligotrophy. As a result, the microbial community structure varies and large phytoplankton blooms are common, especially those formed by diatoms (Landry et al., 1998). One characteristic feature of the North Indian Ocean is the marked Oxygen Minimum Zone (OMZ) layered in the subsurface of the upwelling regions (Roullier et al., 2014). Even though there have been studies of the bacterial community composition of the Indian Ocean seawater (Sunagawa et al., 2015; Wang et al., 2016) and deep-sea sediments (Wu et al., 2011; Khandeparker et al., 2014), the diversity and genomics of the Bacteroidetes populations remain poorly characterized in this area. This is then a good opportunity to analyze whether micro-diversity exists in a particular Bacteroidetes taxon dominating after the phytoplankton bloom event. We used the well-conserved ribosomal RNA genes, the fast-evolving internal transcribed spacer (ITS) and the beta subunit DNA-directed RNA polymerase *rpoB*. We also screened these genomes for a gene that can provide ecological information of the targeted population: proteorhodopsin (PR), which is present in ecologically relevant Bacteroidetes genomes (González et al., 2008; Pinhassi et al., 2016). Fragment Recruitment Analysis (FRA) of metagenomic reads suggested the preferred habitat and an estimated abundance of the novel *Kordia* sp. CFSAG39SUR in different oceans, depths and size fractions.

## MATERIALS AND METHODS

### Sampling, Generation and Selection of Single Amplified Genomes

Surface (SUR) seawater samples (5 m deep, not pre-filtered, Sample ID: TARA\_G000000266) were collected for single cell analysis from the North Indian Ocean during the circumnavigation expedition *Tara* Oceans, station TARA\_039<sup>1</sup> (Figure 1). The environmental setting of the station was inferred from its physical report and can be found at Pangaea website in the following link: [http://store.pangaea.de/Projects/TARA-OCEANS/Station\\_Reports/TARA\\_039\\_oceanographic\\_context\\_report.pdf](http://store.pangaea.de/Projects/TARA-OCEANS/Station_Reports/TARA_039_oceanographic_context_report.pdf).

<sup>1</sup>oceans.taraexpeditions.org

Replicated 1 mL aliquots of seawater were cryopreserved with 6% glycine betaine (Sigma) and stored at  $-80^{\circ}\text{C}$ . Samples used in this study were collected from TARA\_039 (SUR; 5 m deep) located in the North Indian Ocean (18.59–66.62) on March 18th, 2010 (Supplementary Table 1). SAGs generation and preliminary 16S rRNA screening with primers 27F and 907R (Supplementary Table 2) were performed at the Bigelow Laboratory Single Cell Genomics<sup>2</sup>. More information can be found in Swan et al. (2011). Partial 16S rRNA gene sequences (640–850 bp) were compared to previously deposited sequences using the RDP Naive Bayesian rRNA Classifier Version 2.4 tool, and 98 SAGs with >99% similarity to the reference strain *K. algicida* OT-1 (AY195836) were selected for further analyses. These represented 84% of successfully amplified SAGs (total of 117 SAGs). They were named CFSAG39SUR referring to their taxonomic assignment (Cytophaga-Flavobacteria), retrieval method (SAG) and sampling station (TARA\_039, SUR). Other SAGs used in this study (AAA285-F05 and AAA242-P21) were collected at 3,000 m at 4,800 m, respectively, in the North Pacific Sub-tropical Gyre. They belong to the Hawaii Ocean Time-Series (HOT, **Figure 1A**) and were generated following the same procedures.

## Amplification of Phylogenetic Markers 16S rRNA, ITS and 23S rRNA Genes

Amplification of the 16S rRNA gene, ITS and 23S rRNA gene was done by a single Polymerase Chain Reaction (PCR) amplification with the 358F and CF434R primers (Supplementary Table 2). Amplification was performed in a Biorad Thermocycler by an initial denaturation at  $94^{\circ}\text{C}$  for 5 min, followed by 40 cycles of  $94^{\circ}\text{C}$  for 1 min,  $55^{\circ}\text{C}$  for 1 min and  $72^{\circ}\text{C}$  for 2 min, and a final extension at  $72^{\circ}\text{C}$  for 10 min. Each amplification reaction contained: 3 to 20 ng of template DNA, dNTPs (200  $\mu\text{M}$  each),  $\text{MgCl}_2$  (2 mM), primers (0.5  $\mu\text{M}$  each), Taq DNA polymerase (1.25 U), the PCR buffer supplied by the manufacturer (Invitrogen, Paisley, United Kingdom) and MilliQ water up to the final volume of 25.

## Amplification of Functional Genes: *rpoB* and Proteorhodopsin

A set of newly designed primers was used for *rpoB* amplification (Supplementary Table 2). Each PCR reaction followed an initial denaturation at  $94^{\circ}\text{C}$  for 5 min, 40 cycles of  $94^{\circ}\text{C}$  for 1 min,  $40^{\circ}\text{C}$  for 45 s and  $72^{\circ}\text{C}$  for 1 min, and a final extension at  $72^{\circ}\text{C}$  for 5 min. Each amplification reaction contained: 1.5 to 10 ng of template DNA, dNTPs (0.2 mM each),  $\text{MgCl}_2$  (1.5 mM), primers (0.25  $\mu\text{M}$  each), KAPA2G Robust HotStart DNA Polymerase (1 U), KAPA2G Buffer A (KAPA BIOSYSTEMS, Wilmington, MA, United States) and MilliQ water up to the final volume of 25  $\mu\text{l}$ . For proteorhodopsin (PR) amplification, the primers used and the PCR protocol were those described by Yoshizawa et al. (2012) (Supplementary Table 2) using the KAPA2G Robust HotStart DNA Polymerase and KAPA2G Buffer A from KAPA BIOSYSTEMS. Each PCR reaction followed an initial

denaturation at  $94^{\circ}\text{C}$  for 5 min, 35 cycles of  $94^{\circ}\text{C}$  for 1 min,  $44^{\circ}\text{C}$  for 45 s and  $72^{\circ}\text{C}$  for 1 min, and a final extension at  $72^{\circ}\text{C}$  for 5 min. Each amplification reaction contained: 3 to 20 ng of template DNA, dNTPs (0.2 mM each),  $\text{MgCl}_2$  (3 mM), primers (0.5  $\mu\text{M}$  each), Taq DNA polymerase (1.25 U), the PCR buffer supplied by the manufacturer (Invitrogen, Paisley, United Kingdom) and MilliQ water up to 25  $\mu\text{l}$ .

## Sequencing and Phylogenetic Analysis

Polymerase Chain Reaction (PCR) products were purified and sequenced by Genoscreen (Lille, France) with OneShot Sanger sequencing and aligned against NCBI's nBLAST and xBLAST databases for identification. Reference sequences used in this study were retrieved from Integrated Microbial Genomes and Metagenomes (IMG) 4 Data Management or NCBI database. Using Geneious Pro 4.8.5 software all sequences were aligned (using default Geneious progressive pairwise aligner) trimmed and manually checked. The software was also used to obtain the complete 16S rRNA gene assembling the amplicon sequences from MDA screening and the phylogenetic markers amplification. Alignments were processed into phylogenetic trees with Mega 5.2.2 Software choosing Maximum Likelihood statistical methods along with 1000 bootstrap replications. The best-fit substitution models for Maximum Likelihood phylogenies of the studied markers were: GTR with Gamma distribution (16S rRNA gene), Kimura-2 with Gamma distribution (23S rRNA gene) and WAG with Gamma distribution (PR and *rpoB*). Representatives from the Proteobacteria and Cyanobacteria phyla were used as outgroups in all phylogenies. Phylogeny reconstruction of functional genes used the closest sequences to the SAGs' genes obtained from the Ocean Microbiome Reference Gene Catalog (Sunagawa et al., 2015).

## Fragment Recruitment Analysis (FRA)

Nucleotide-Nucleotide BLAST 2.2.28+ was used to recruit metagenomic reads from several metagenomic samples similar to all the available sequenced *Kordia* genomes. Considering the aim of this part of the study, which was to determine the presence and distribution of the novel *Kordia* sp. CFSAG39SUR in different oceanic regions, depths and size fractions, a database containing all 4 *Kordia* genomes was generated. This ensured a competitive recruitment between the genomes for each metagenomic sample analyzed. The quality and genome completeness of the four reference genomes in the database was measured with software checkM (Parks et al., 2015) and fetchMG (Sunagawa et al., 2013) (Supplementary Table 3). The available genomic sequences of *Kordia* AAA285-F05 did not code for any of the marker genes or COGs used by the software, resulting in a genome completeness estimation of 0% despite being 283.58 Kb long. The other three reference genomes (*K. algicida*, *K. jejudonensis*, and *K. zhangzhouensis*) were estimated to be complete and free of contamination. Recruitment regarded only one read per target gene ( $-\text{max\_target\_seqs } 1$ ), an identity percentage higher than 70% ( $-\text{perc\_identity } 70$ ), as well as an *e*-value lower than 0.000001 ( $-\text{e-value } 0.000001$ ). All other nBLAST settings were set on default. In order to avoid random alignments, a

<sup>2</sup>scgc.bigelow.org

filtering process was applied using R software (R Core Team, 2013) (i) excluding alignments with coverage lower than 90% of the sequence length, (ii) removing duplicated reads within each genome and (iii) masking reads belonging to the ribosomal operon. This operon does not follow the species delimiting recruitment patterns as the rest of the genome does and its presence would overestimate read recruitment (Caro-Quintero and Konstantinidis, 2012). All contigs from each genome were concatenated to a single sequence and the file containing the four sequences (one for each genome) was used to generate the database. This step was done using NCBI's *makeblastdb* application. As query we selected those metagenomic samples which contained the most metagenomic reads 16S rRNA and metagenomic Illumina tags (miTags) (Logares et al., 2013) annotated as *Kordia* (data not shown) using SILVA database Release 115 (Quast et al., 2013). These were stations TARA\_039 (0.2–20  $\mu\text{m}$ ) and TARA\_085 (0.2–3  $\mu\text{m}$ ) from *Tara* Oceans (Supplementary Table 1 and **Figure 1A**) (Sunagawa et al., 2015). In addition, we used 4 metagenomes from the deep waters of Malaspina 2010 circumnavigation expedition (Acinas et al., in preparation), two of them from the Brazil Basin in the Atlantic Ocean and two of them from Circumpolar Deep Waters (Supplementary Table 1 and **Figure 1A**). Due to the flexibility of Bacteroidetes' lifestyle, the metagenomic selection was done from a limited pool of deep ocean metagenomes covering the free-living fraction of prokaryotes (0.2–0.8  $\mu\text{m}$ ) and the fraction of prokaryotes attached to particles or aggregates (0.8–20  $\mu\text{m}$ ) of a same seawater sample. The coverage of different deep ocean biomes and the highest abundance of *Kordia* related metagenomic read recruitments was also taken into account for the metagenomes' selection for the study. All queries consisted on merged, paired-end reads with different sequencing depths. Merging was done using Mothur v.1.33.3 (Schloss et al., 2009). Recruited data normalization was done: (i) by reference genome size (resulting in recruited reads/genomic bp), then (ii) upscaling this ratio to 1 Mb (as done in Swan et al., 2011), and finally (iii) by metagenomic sequencing depth. As sequencing depths varied between metagenomes, normalization was performed to the smallest metagenomic sequencing depth of the study.

## Accession Numbers

Accession numbers for PCR products of *Kordia* CFSAG39SUR: 16S rRNA gene (MF187452), ITS (MF187454), 23S rRNA gene (MF187453), *rpoB* (MF187456), and proteorhodopsin (MF187455). Accession numbers for partial 16S rRNA gene sequences are: MF187458 for AAA285-F05 and MF187457 for AAA242-P21.

The genomes used as reference for the FRA are the following: Bacteroidetes bacterium SCGC AAA285-F05 (JGI GoldStamp Id Go0060979), *K. algicida* strain OT-1 (NCBI RefSeq NZ\_ABIB000000000.1), *K. jejudonensis* strain SSK3-3 (NCBI RefSeq NZ\_LBMG000000000.1), *K. zhangzouensis* strain MCCC 1A00726 (NCBI RefSeq NZ\_LBMH000000000.1). The *Tara* Oceans metagenomes used as query are the following (INSDC run accession numbers): TARA\_039\_DCM\_0.22-1.6 (ERR599145), TARA\_039\_MES\_0.22-1.6 (ERR599037| ERR599172), TARA\_085\_

SRF\_0.22-3 (ERR599090| ERR599176), TARA\_085\_DCM\_0.22-3 (ERR599104| ERR599121), TARA\_085\_MES\_0.22-3 (ERR599008| ERR599125). The Expedición Malaspina metagenomes are the following (JGI Genome Portal, genome.jgi.doe.gov/): MP1493 (/DeeseametaMP1493\_FD), MP1494 (/DeeseametaMP1494\_FD), MP0326 (/DeeseametaMP0326\_FD), MP0327 (/DeeseametaMP0327\_FD) with the previous agreement of the PI of the JGI CSP 602 grant.

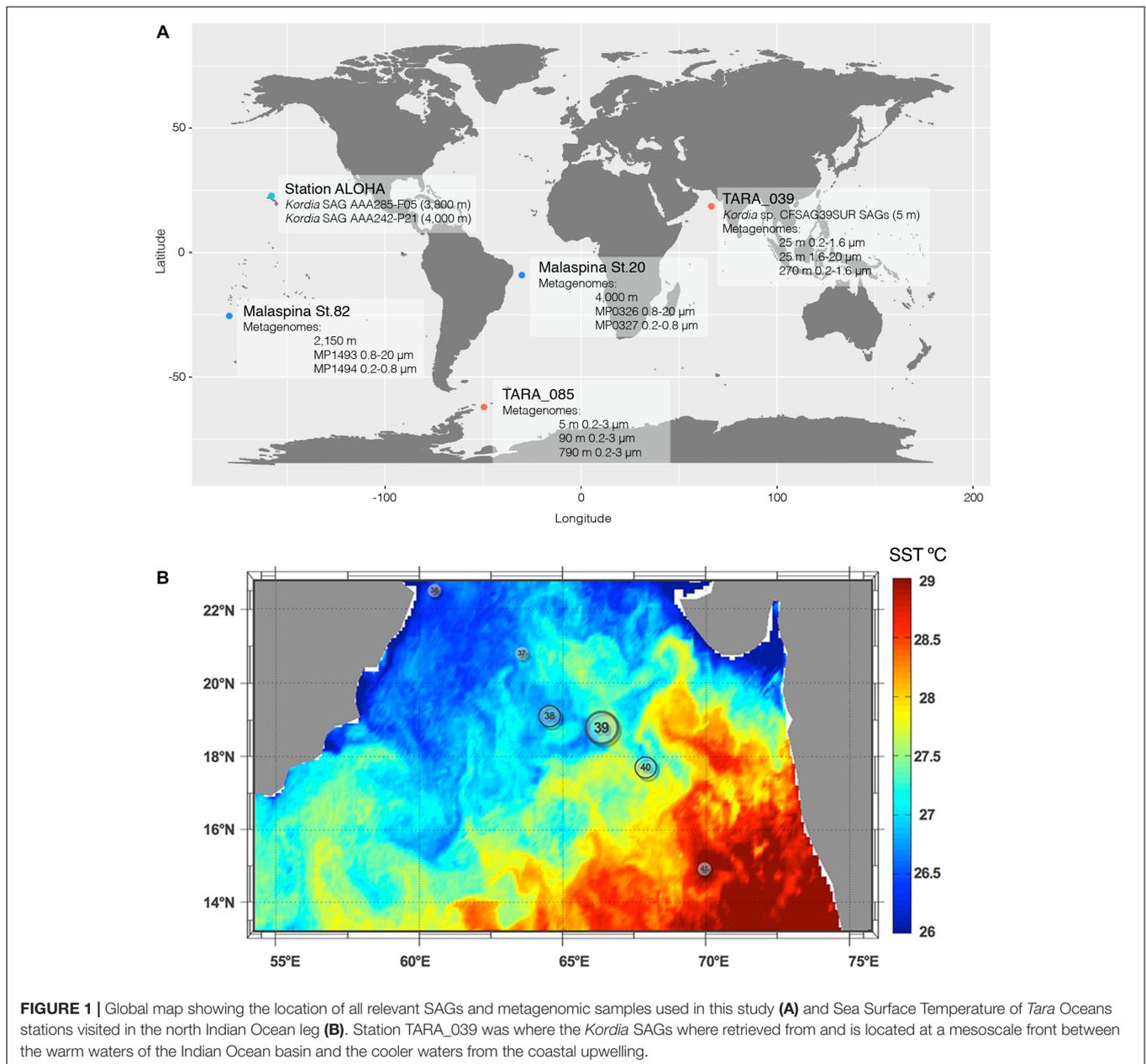
## RESULTS

### Environmental Setting and Bacterial Diversity of Station TARA\_039

On February 2010, 1 month prior to sampling, a moderate bloom of phytoplankton occurred in the Oman Gulf due to a coastal upwelling under strong wind conditions. It propagated to the central basin of the Arabian Sea reaching only up to station TARA\_039 (see Figure 3 in Roullier et al., 2014), which was the last mesotrophic station occupied by the *Tara* Oceans expedition before entering the warm oligotrophic waters of the central basin. Previous stations occupied during this leg in the North Indian Ocean (TARA\_037 and TARA\_038) were in the cooler waters of the upwelling, whereas station TARA\_039 was at a mesoscale frontal region with the oligotrophic zone (**Figure 1B**). Particles produced during the phytoplankton bloom decreased in size as the cruise reached the central basin (see Figure 10 in Roullier et al., 2014). Diatom contribution to total phytoplankton decreased from 20 to 5% from stations close to the Gulf of Oman down to station TARA\_40, outside the upwelling waters. Sunagawa et al. (2015) found that the bacterial assemblage at station TARA\_039 was dominated by Proteobacteria in both surface and DCM (Deep Chlorophyll Maximum) layers, followed by Cyanobacteria and other, less abundant groups including Euryarchaeota, Actinobacteria, Deferribacteres, and Bacteroidetes (Sunagawa et al., 2015). In the mesopelagic (MES) zone there was a decrease in the abundance of Proteobacteria, a small increase of Deferribacteres and Actinobacteria and a remarkable increase in Thaumarchaeota and unclassified Bacteria. There was an increase in Bacteroidetes abundance toward the end of the *Tara* Oceans' sampling leg in the North Indian Ocean (Sunagawa et al., 2015).

### Phylogenetic Reconstruction of a Natural *Kordia* Population

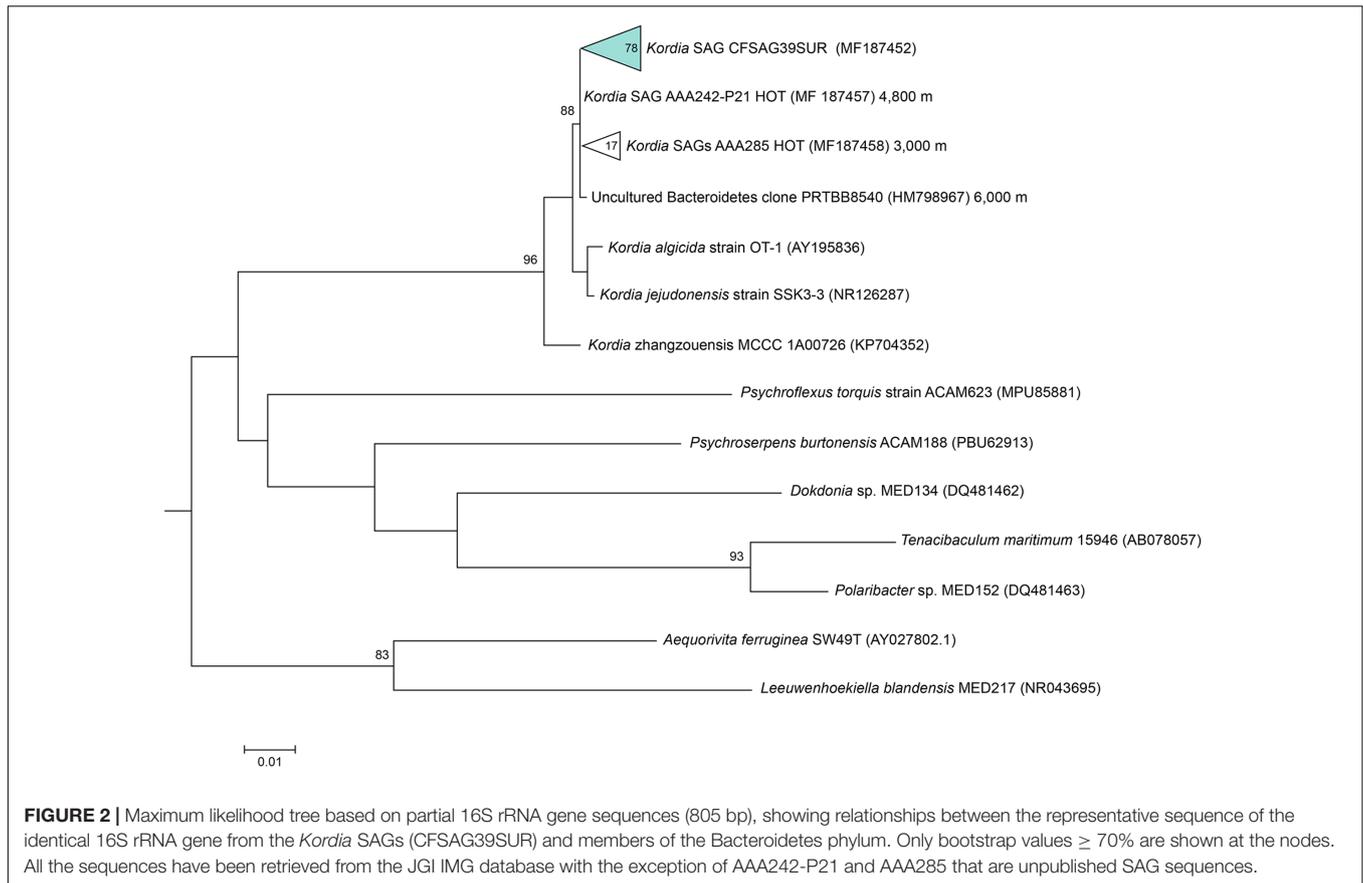
Amplification of the three phylogenetic markers of the ribosomal operon was successful in 78 out of 98 *Kordia* SAGs, and their alignment indicated a 100% identity among them. The resulting sequences contained full 16S rRNA gene (1,495 bp), ITS-1 (536 bp) and partial 23S rRNA gene (415 bp). The alignment of the resulting full 16S rRNA gene sequence against NCBI database assigned the SAGs to the genus *Kordia*, family Flavobacteriaceae. Its best hit with a nucleotide identity of 99% and sequence coverage of 97% was an uncultured Bacteroidetes sampled in the Puerto Rico Trench (clone PRTBB8540, acc. HM798967) at 6,000 m depth. The closest sequenced cultured relatives were



*K. algicida* strain OT-1 with a 96.8% of nucleotide identity and 99% coverage (complete sequence, acc. NR027568) and *K. jejudonensis* strain SSK3-3 with the same identity percentage but 96% coverage (partial sequence, NR\_126287). A 100% nucleotide identity was found with partial 16S rRNA gene sequences from 17 SAGs collected at 3,000 m (AAA285) and one SAG from 4,800 m depth in the North Pacific Sub-tropical Gyre (AAA242-P21), which belong to the Hawaii Ocean Time-series (HOT) (Stepanuskas, unpublished results). The phylogenetic reconstruction based on the representative sequences of the 16S rRNA genes from the 78 SAGs (**Figure 2**) strongly indicated that they belonged to the genus *Kordia* (phylum Bacteroidetes, family Flavobacteriaceae). High bootstrap values supported the cluster

formed by the 78 CFSAG39 SAGs with the 17 AAA285-SAGs, the one AAA242-P21 SAG and the uncultured Bacteroidetes clone PRTBB8540. This cluster's sequences were all retrieved from deep waters except for the SAGs reported here. The resulting sequence from the 78 identical partial 23S rRNA genes, when aligned against the NCBI database, showed a lower identity against *K. algicida* OT-1 (90.5%). Nevertheless, it was still its closest culture hit and the phylogenetic tree supported the assignment of the SAGs to the genus *Kordia* ITS with high bootstrap values (Supplementary Figure 1).

The 78 SAGs' ITSs were identical and coded for Ala-tRNA and Ile-tRNA. These tRNAs showed only one polymorphism each when compared to those of *K. algicida*. The nucleotidic change



was located at the next-to-last base, not affecting the structure of the functional molecule. The box A conserved region present at the end of the ITS was identical to that of *K. algicida* although the former was located slightly upstream (17 bp) (Supplementary Figure 2).

Amplification of the beta subunit of RNA polymerase (*rpoB*) was successful in 18 SAGs, obtaining  $\sim 1,400$  bp amplicons with 100% of both amino acid and nucleotide identity among them. Their closest hit in NCBI database was the *rpoB* gene of *K. algicida* OT-1 with a 98% amino acid identity. High bootstrap values supported the location of the SAGs' *rpoB* sequence in the phylogenetic tree, within the Flavobacteriaceae cluster and closest to *K. algicida* OT-1 (Supplementary Figure 3).

An amplification summary can be found in Table 1.

### Novel Proteorhodopsin Gene

The PR gene was amplified in 34 out of the 98 *Kordia* SAGs. The resulting 167 amino acid sequences turned out to be 100% identical to one another. The sequence was closest to the Flavobacteriaceae *Flagellimonas* sp. DIK-ALG-169's proteorhodopsin (89% identity, 100% coverage) (AHN13811) and *Kordia* sp. PC-4 (Yoshizawa et al., 2012) (88% identity, 91% coverage) in the NCBI database. When comparing sequences in the Ocean Microbiome Reference Gene Catalog (OMRGC), the closest hit shared 84.2% amino acid similarity (unclassified Flavobacteria, NOG136807 ID.v1.022398661). Phylogenetic

reconstruction placed the SAGs' proteorhodopsin in a cluster of Flavobacteriaceae PR genes (Figure 3). We tried to amplify the proteorhodopsin gene from the 17 SAGs retrieved from 3,000 m at the North Pacific Gyre (AAA285-F05) but the results were negative.

The SAGs' proteorhodopsin alignment with other closely related PR sequences showed that this proteorhodopsin had the same structural features as its relatives. It had the predicted transmembrane helices, key amino acids for functionality and the location of the spectral tuning amino acid, which in this case was methionine (Figure 4), indicating absorbance of light spectrum between 518 and 535 nm (green light).

### Distribution of Novel *Kordia* sp. in Different Oceanic Regions

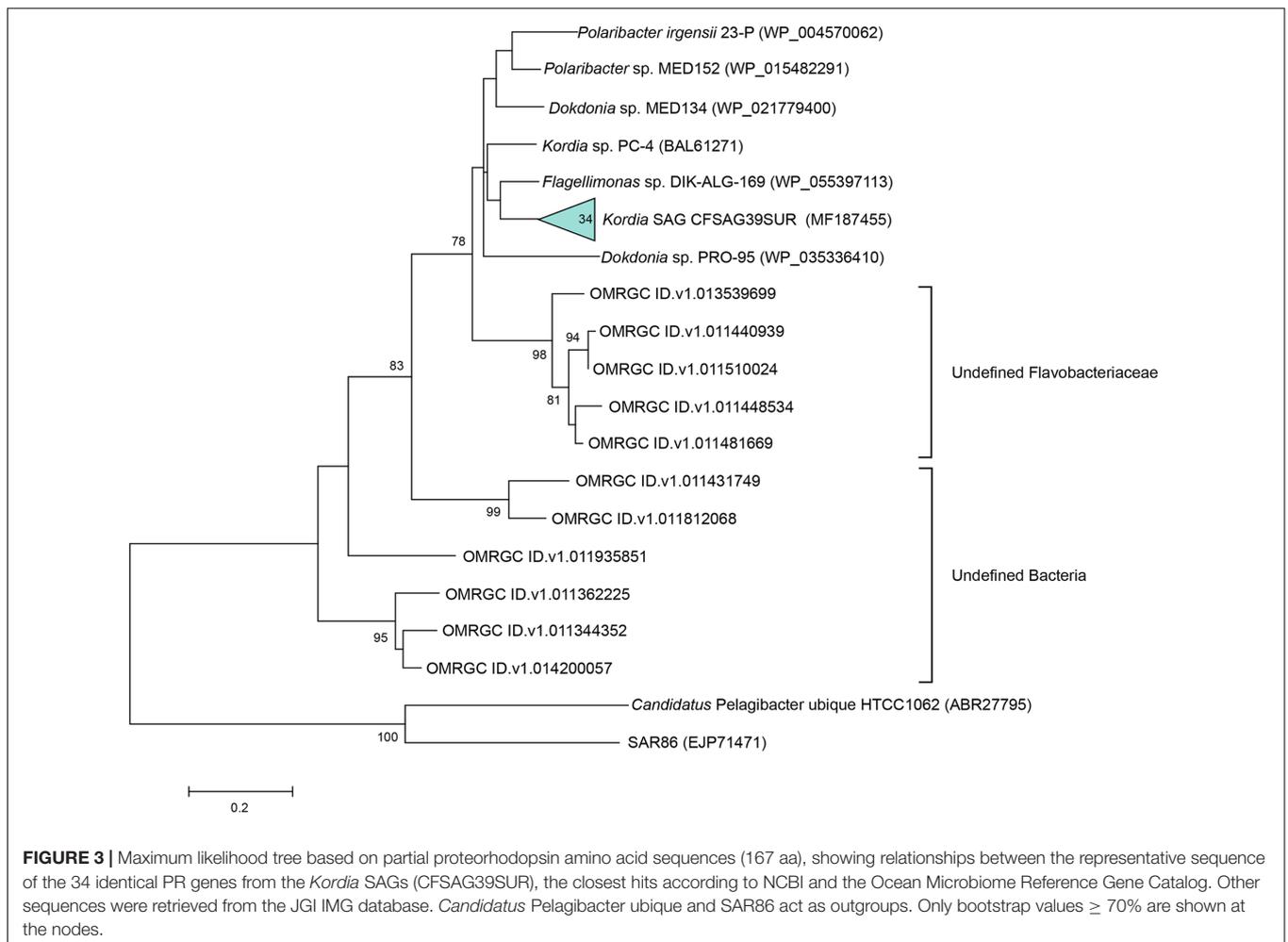
Metagenomic reads from different stations were recruited through a competitive FRA with the available sequenced genomes of the genus *Kordia*: the deep-sea *Kordia* SAG AAA285-F05 (283.58 Kbp), the seawater surface isolated in culture *K. algicida* OT-1 (5.01 Mbp), the freshwater *K. zhangzhouensis* MCCC 1A00726 (4.03 Mbp) and *K. jejudonensis* SSK3-3 (5.3 Mbp), isolated from a region where spring freshwater and seawater meet. We counted reads with 70–100% identity to these reference genomes.

For *Kordia* AAA285-F05, the number of recruited reads increased with depth in all stations tested. The percentage

**TABLE 1** | Summary on genetic information of *Kordia* Single Amplified Genomes (SAGs) and closest SAGs, isolates and clones. Identity % against *Kordia* sp. CFSAG39SUR's corresponding genes (16S, ITS and 23S) and proteins (RpoB and PR).

<i>Kordia</i> sequence information	Origin	Depth (m)	16S		ITS		23S		<i>rpoB</i>		PR	
			# seq.	id %	# seq	id %	# seq	id %	# seq	id %	# seq	id %
CFSAG39SUR (98 SAGs)	North Indian Ocean	5	78	100	78	100	78	100	18	100	34	100
AAA242-P21 (1 SAG)	NPSG ALOHA	4000	1	100	-	-	-	-	-	-	-	-
AAA285 (17 SAGs)	NPSG ALOHA	3800	1	100	-	-	-	-	-	-	-	-
<i>Kordia algicida</i> OT-1 (1 isolate)	Masan Bay, S.Korea	0	3	96.8	3	76.1	3	90.5	1	89.3	0	-
<i>Kordia</i> sp. PC-4 (1 isolate)	Sagami Bay, Japan	100	1	98.4	-	-	-	-	-	-	1	89.2
Unc. Bact. PRTBB8540 (1 clone)	Puerto Rico Trench	6000	1	99.9	-	-	-	-	-	-	-	-

For each gene and protein, “# seq” refers to the number of sequences available for comparison.



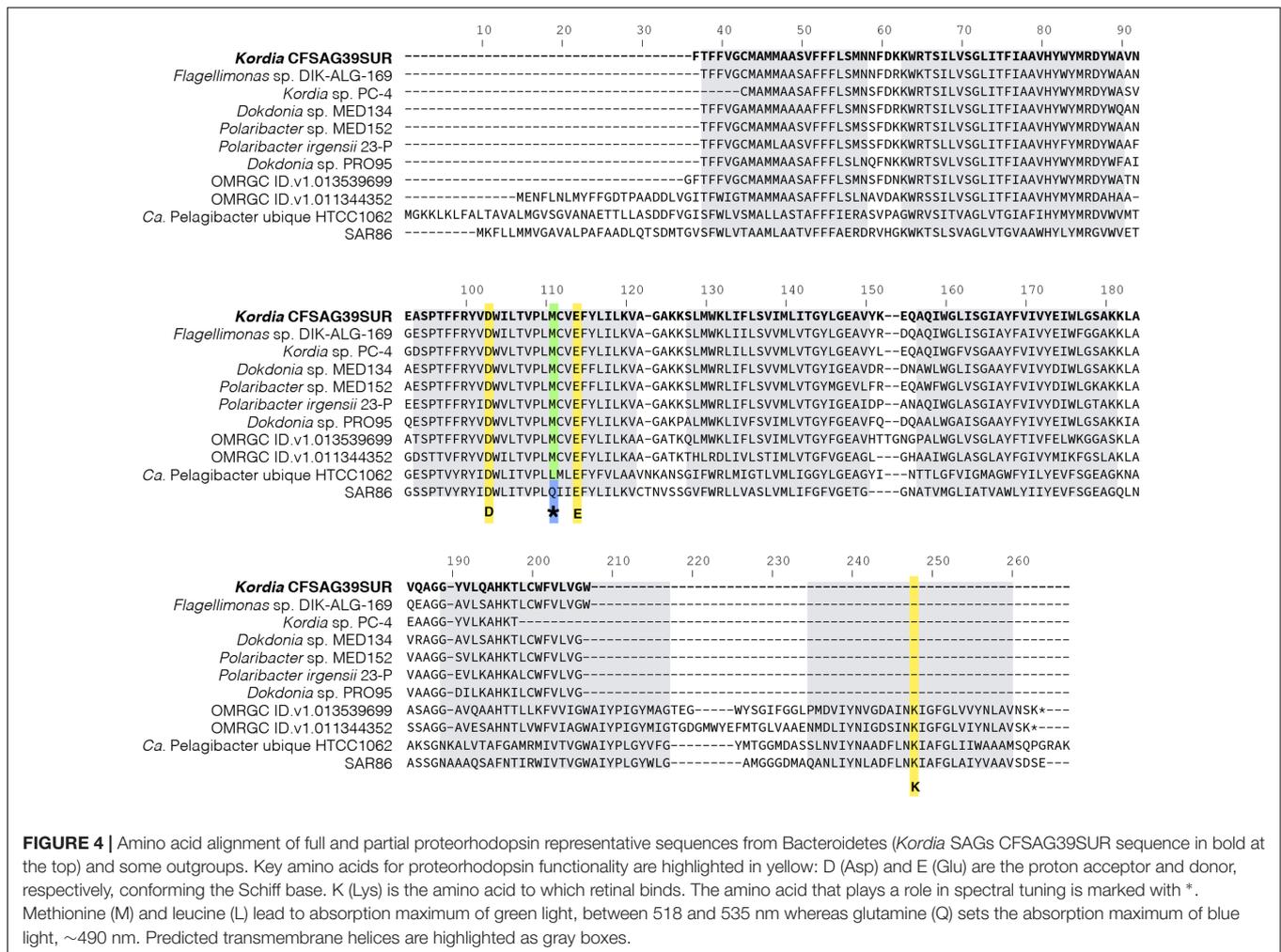
**FIGURE 3** | Maximum likelihood tree based on partial proteorhodopsin amino acid sequences (167 aa), showing relationships between the representative sequence of the 34 identical PR genes from the *Kordia* SAGs (CFSAG39SUR), the closest hits according to NCBI and the Ocean Microbiome Reference Gene Catalog. Other sequences were retrieved from the JGI IMG database. *Candidatus Pelagibacter ubique* and SAR86 act as outgroups. Only bootstrap values  $\geq 70\%$  are shown at the nodes.

of reads recruited per genomic Mbp was several orders of magnitude larger in the free-living than in the particle associated size fractions (Figure 5; numeric data in Supplementary Table 4).

Highest relative abundances of reads per Mbp with 95–100% identity for *Kordia* AAA285-F05's occurred in the free-living fraction of bathypelagic waters from the Brazil Basin (MP0327, 0.51%) and the Circumpolar Deep Waters of the Pacific Ocean (MP0326, 0.46%). Recruitment percentages for this SAG in these

two locations decreased down to 0.04 and 0.01%, respectively. In TARA\_039 highest recruitment % per Mbp was 0.0007% at 270 m, while in the Southern Ocean (TARA\_085), recruitment increased with depth reaching 0.06% at 790 m. Recruitment was virtually zero for the isolates genomes at this recruitment identity percentage.

Recruitment reads with 70–94.9% identity might be indicative of populations related to, but different from, the reference genome present at the tested metagenomic sample. The



**FIGURE 4 |** Amino acid alignment of full and partial proteorhodopsin representative sequences from Bacteroidetes (*Kordia* SAGs CFSAG39SUR sequence in bold at the top) and some outgroups. Key amino acids for proteorhodopsin functionality are highlighted in yellow: D (Asp) and E (Glu) are the proton acceptor and donor, respectively, conforming the Schiff base. K (Lys) is the amino acid to which retinal binds. The amino acid that plays a role in spectral tuning is marked with \*. Methionine (M) and leucine (L) lead to absorption maximum of green light, between 518 and 535 nm whereas glutamine (Q) sets the absorption maximum of blue light, ~490 nm. Predicted transmembrane helices are highlighted as gray boxes.

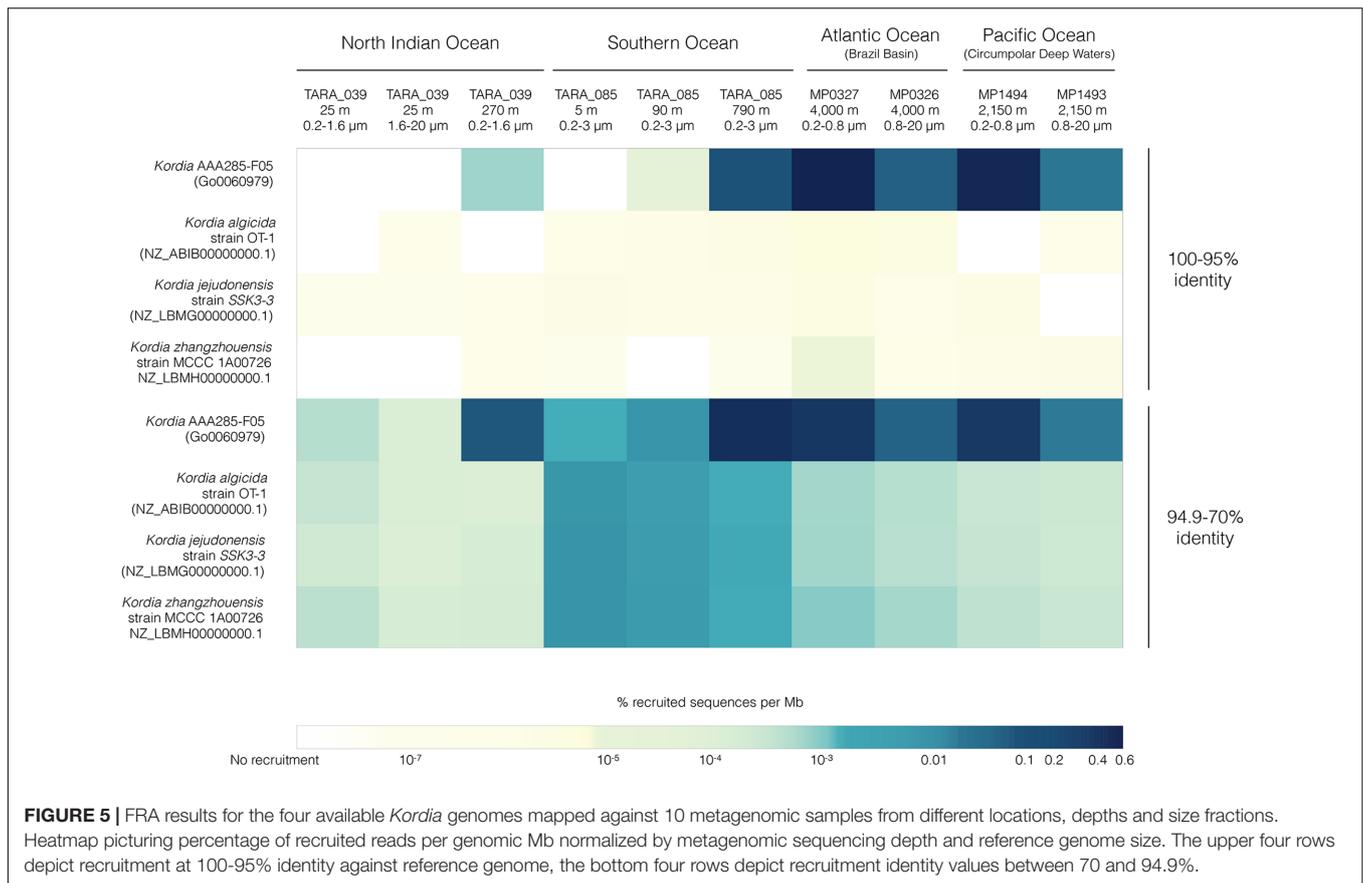
recruitment trend observed in the higher identity percentage for *Kordia* SAG AAA285-F05 was consistent at the lower percentage as well. Recruitment % per genomic Mbp increased at all depths of both TARA\_039 and TARA\_085, now reaching values of 0.05% at 270 m in TARA\_039 and the maximal of 0.38% at 790 m in TARA\_085. Similar abundances were obtained for the Brazil Basin (MP0327, 4,001 m) and the Circumpolar Deep Waters (MP1494, 2,150 m) in their free-living fraction (0.3 and 0.27%, respectively). Recruitment decreased in the larger size fraction of the two Malaspina 2010 circumnavigation expedition metagenomes down to 0.04 and 0.01%, respectively. For the other three *Kordia* genomes competing for recruitment in the same metagenomic samples, recruitment values decreased significantly to values ranging from 0.008% per Mb for *K. zhangzhouensis* in TARA\_085 surface waters to values down to 10<sup>-7</sup>% per Mb. The recruitment for these three genomes followed a similar pattern, with highest values in TARA\_085 surface decreasing with depth. In their case, bathypelagic samples and North Indian ocean's DCM (25 m) recruited similar values (~0.0002%). The lowest values were found at TARA\_039 mesopelagic waters (0.00001%).

FRA plots (Figure 6) showed a similar pattern of *Kordia* AAA285-F05 recruitment for the meso- and bathypelagic metagenomes (TARA\_085\_MES, MP0327, MP1494) of the free-living prokaryotic fraction. There was a very good coverage of the whole reference sequence, with the exception of some fragments where there was a visible decrease in the recruitment at all identity percentages. The highest read densities were found above 95% identity. In the TARA\_039 OMZ recruitment plot, an overall decrease in read density was observed, identities ranged from 85 to 99%. The two genomic regions with very low recruitment were also apparent in this plot.

The three reference *Kordia* genomes (Supplementary Figure 4) showed read clouds at the lower identity percentage in all cases. There was a good coverage of the length of the genome but not as intense as that observed in Figure 6 for reference genome *Kordia* AAA285-F05.

## DISCUSSION

The aim of this study was to explore the potential microdiversity within a population of 98 SAGs of the genus *Kordia* (*Kordia*



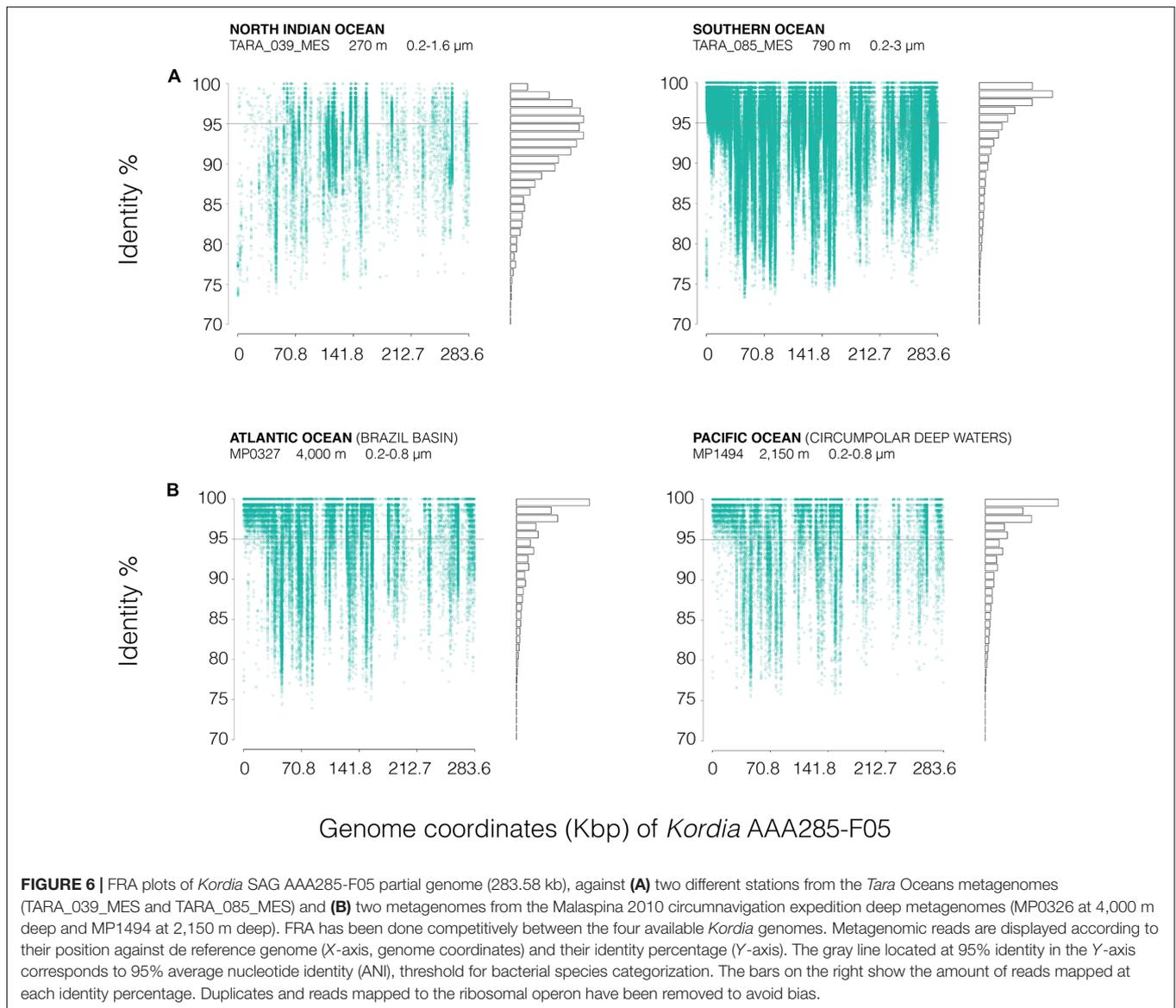
sp. CFSAG39SUR), a Flavobacteriaceae genus established after describing *K. algicida* strain OT-1 (Sohn et al., 2004). Having these SAGs was a unique opportunity to study the extent of microdiversity within a large population of environmental uncultured organisms, since genomic length and composition of marine uncultured bacterial genomes (SAGs) generally differ from those of similar taxa isolated in pure culture (Swan et al., 2013).

Phylogenetic proximity among *Kordia* sp. CFSAG39SUR SAGs was confirmed by the 100% identity of the full 16S rRNA gene, partial 23S rRNA gene, and the ITS-1 region. The latter is located between the 16S and 23S ribosomal genes and has been widely used when more detailed taxonomical resolution was needed, especially in intra-specific diversity studies (Brown and Fuhrman, 2005). Moreover, its variability in length and composition highlights initial genome diversification and evolutionary speciation (Schloter, 2000). The 100% identity at the nucleotide level of functional genes *rpoB* and PR (exposed to a high degree of horizontal gene transfer, Fuhrman et al., 2008) could suggest that the genetic homogeneity may be present throughout the whole genome. However, only the analysis of the whole genomes would confirm it or reject this hypothesis.

Such genetic similarity among SAGs can be attributed to two main factors: (i) to stable environmental conditions providing fewer mutation-prone events, hence low microdiversity in the population (Cohan, 2006), or (ii) recent population expansion,

as nascent populations are usually more genetically homogenous than mature ones, in which products of mutations and recombination may accumulate (Cordero and Polz, 2014). As mentioned above, a month prior to the arrival of the *Tara* schooner to the North Indian Ocean, a coastal upwelling fertilized the surface waters of the Gulf of Oman with new nutrients, resulting in a phytoplankton bloom that extended through the coast of Iran and south-east to station TARA\_039 (Roullier et al., 2014). These blooms have been described to be seasonal, mostly formed by diatoms (Landry et al., 1998). The large particles generated by the bloom were seen to decrease in size through the sampling period (Roullier et al., 2014). Backward Lagrangian particle transport modeling suggested that these particles sunk near station TARA\_039 (Roullier et al., 2014), where a frontal system separated the colder upwelled waters from the warmer waters of the oligotrophic basin. Considering the occurrence of such an event before the sampling at TARA\_039, it is highly possible that the genetic homogeneity found in the 78 *Kordia* sp. CFSAG39SUR SAGs is due to the second possibility mentioned above, that is, that these bacteria were retrieved as a nascent population derived from the phytoplankton bloom.

There is previous knowledge of the relationship between members of the genus *Kordia* and algae. The first isolate described for the genus, *K. algicida* OT-1, was isolated following a red tide of the diatom *Skeletonema costatum* (Sohn et al., 2004).



Additionally, it is the only Bacteroidetes species found to code for R-bodies, a specific system for targeted lysis of eukaryotes (Raymann et al., 2013). Likewise, *K. ulvae* was isolated from the surface of the marine alga *Ulva* sp. (Qi et al., 2016). Abundant algal products or exudates could provide a good environment for the rapid proliferation of our *Kordia* sp. CFSAG39SUR SAGs either attached to large particles or free-living in the water column. This would help to explain the apparent lack of microdiversity found at the time of sampling and that would probably be due to nascent populations from a single genotype after the algal bloom. Interestingly, a population of 18 SAGs identical to this *Kordia* sp. CFSAG39SUR population at the available partial 16S rRNA gene sequence were retrieved from the deep waters of station ALOHA in the North Pacific Subtropical Gyre, where upwelling and seasonal diatom blooms are common events (Villareal et al., 2012).

The *Kordia* sp. CFSAG39SUR SAGs made 84% of the sorted heterotrophic genomes retrieved from the sample, suggesting a high abundance of *Kordia* at the moment of sampling. Even though there was no metagenomic sample available from TARA\_039 surface waters to support this high abundance and being aware of the environmental context of the sampling, we relied on FRA to test its persistence through the water column. Metagenomic read recruitment of *Kordia* spp. was very low in the TARA\_039 DCM metagenome and slightly higher in TARA\_039 MES metagenome. These values pointed to *Kordia* sp. CFSAG39SUR SAGs as members of the less abundant taxa known as the rare biosphere (Pedrós-Alió, 2012). Higher abundances (up to 0.51% of recruited reads per genomic Mbp) were found in other metagenomic samples from *Tara* Oceans and Malaspina 2010 circumnavigation expedition covering different ocean locations, depths (0–4,000 m) and

size fractions (0.2–20  $\mu\text{m}$ ). In many cases read abundances reached values similar to recruitments in the surface waters of different oceans using as reference isolated Rickettsiales, *Planctomyces* or Rhodobacterales, generally considered to be abundant bacteria. Nevertheless, read abundances never reached those of isolated *Prochlorococcus*, *Synechococcus*, or *Pelagibacter*, the most abundant groups in seawater (see Supplementary Figures S10, S11 by Swan et al., 2011). The scarce recruitment in TARA\_039 metagenomes contrasts with the high abundance of SAGs retrieved from the surface waters of the station. We think it is important to highlight the fact that the sampled water processed for single cell genomics was not prefiltered. Together with the slight increase in recruitment at the mesopelagic layer, we cannot reject the possibility that sequences related to *Kordia* sp. CFSAG39SUR SAGs could have been abundant in metagenomes of larger size fractions (20–200  $\mu\text{m}$ ) than those available (0.2–20  $\mu\text{m}$ ), especially since large particles (2–2.5 mm at 520–560 m, see Figure 10 in Roullier et al., 2014) were reported at Station TARA\_039 deeper waters during the *Tara* expedition. It is possible then, that this new *Kordia* species becomes abundant only after upwelling-induced phytoplankton bloom events. To further test the preferred niche for *Kordia*, FRA should be performed on metagenomes of higher planktonic size fractions, either from the same stations or from stations where diatom blooms occur after upwelling episodes. Unfortunately, such metagenomes are not currently available.

This suggests a hypothesis where, after colonization of phytoplankton-derived large particles, *Kordia* sp. CFSAG39SUR would sink with the particles to the deep ocean, becoming part of the bacterial seed bank of the rare biosphere. In fact, FRA of reference genome *Kordia* SAG AAA285-F05 displayed higher read recruitments (i.e., relative read abundance per genomic Mbp) in metagenomes from deeper water samples. Intriguingly, our *Kordia* sp. CFSAG39SUR SAGs coded for a light dependent proteorhodopsin, a metabolic advantageous trait in sunlit water layers. The sequence was similar to those of other Flavobacteria which have been shown to absorb green-light, and have fast photocycles characteristic of proton-pumping rhodopsins rather than sensing rhodopsins (Spudich, 2006). This suggests that *Kordia* sp. CFSAG39SUR has an active metabolism in the upper layers of the ocean. Perhaps their activity decreases or becomes dormant when reaching bottom waters. There, they would accumulate attached to the remaining particles or switching to a free-living lifestyle, explaining the highest recruitments at the free-living fraction of meso- and bathypelagic waters. It must be considered, however, that proteorhodopsins have been found in the deep sea (Tang et al., 2015; Yoshizawa et al., 2015) and that both presence and expression of proteorhodopsin genes were detected in the Arctic Ocean during the Polar night (Nguyen et al., 2015), suggesting that proteorhodopsins may have an unknown function in dark waters.

Comparative genomics has helped us infer a species threshold based on Average Nucleotide Identity (ANI, Konstantinidis and Tiedje, 2005) where 95% identity corresponds to the previous 70% DNA–DNA hybridization threshold, which was the gold standard for taxonomic species identification (Wayne

et al., 1987). FRA on metagenomes (Caro-Quintero and Konstantinidis, 2012) using a reference genome makes it possible to infer: (i) similar genetic populations (“species”) above >95% ANI and (ii) different sequence-discrete populations within the range of 80–95% ANI. For this study, as the *Kordia* sp. CFSAG39SUR genome sequences are not available, we carried out the analyses with the genome of *Kordia* SAG AAA285-F05 as reference because they shared 100% identity in their 16S rRNA genes. Despite this identity at the 16S rRNA level, there may be significant differences both at the gene content level, as well as in the identity between shared genes. It is common that within a given bacterial species there is a set of genes conserved among all strains for housekeeping functions (core genome), while an array of different functional genes can vary depending on the niche of the specific bacteria (flexible genome) (Medini et al., 2005; Fernández-Gómez et al., 2012). Despite this limitation, the fact that we used *Kordia* AAA285-F05 as our reference genome still may provide an indirect detection of *Kordia* sp. CFSAG39SUR SAGs through the FRA of the fraction of the genome shared between the reference and our SAGs. Since we do not know to what extent *Kordia* sp. CFSAG39SUR resembles *Kordia* SAG AAA285-F05 apart from the 16S rRNA gene in those shared genes, reads recruited between identity values of 95–100% would reveal the presence of *Kordia* belonging to the same species as *Kordia* AAA285-F05. If our *Kordia* sp. CFSAG39SUR on the other hand, was actually a co-occurrent relative of the reference but not the same ecotype (i.e., nucleotidic differences in those shared genes), we would indirectly detect it in those reads recruited below 95% identity (Caro-Quintero and Konstantinidis, 2012).

Being aware of this, we also used all available *Kordia* spp. genomes as reference for the FRAs at the same time. This analysis confirmed that those reads mapping against AAA285-F05’s genome at lower identity ranges belonged, indeed, to a new ecotype from the same novel species (likely our *Kordia* sp. CFSAG39SUR SAGs). The non-existent recruitment of reads at 95–100% for those genomes different from *Kordia* AAA285-F05 backs it up strongly.

## CONCLUSION

This study shows how state-of-the-art single cell genomics can be combined with more traditional techniques such as MLSA to obtain an overview of the genetic composition of the population of study without sequencing whole genomes beforehand. Thus, this can provide both an analysis of population genetics avoiding bias by isolation in culture and an approach to select genomes of interest for further population genomics and biogeography studies. In our case, through MLSA of SAGs we have found that the dominant heterotrophic bacterial taxon in the sample was genetically homogeneous. Moreover, the study of this population in combination with available metagenomic datasets and the oceanographic metadata of the sampling area has helped shedding light on the species distribution, dynamics, and potential ecological niche of a novel *Kordia* species that contains proteorhodopsin.

## AUTHOR CONTRIBUTIONS

SA designed this research. MR-L performed the laboratory and analyses work and wrote the paper. IF, FC-C, GS, and PS helped with the data analyses. SS and LS contributed to oceanographic dataset and comments of the sampling station. MS and RS contributed with the *Tara* Ocean SAGs collection and sorting. SA funded this research with contributions from CP-A and JG. All authors were involved in critical reading for writing the paper with special contribution of CP-A.

## FUNDING

Funding was provided by the Spanish Ministry of Economy and Competitiveness grants GL2011-26848/BOS MicroOcean PANGENOMICS to SA, CTM2013-48292-C3 “EcoBGM” to CP-A and JG as well as grant BIOSENSOMICS through “Ayudas Fundación BBVA a investigadores y creadores culturales” to SA. Additional funding was provided by the Spanish Ministry of Economy and Competitiveness Science and Innovation through the Consolider-Ingenio programme (project Malaspina 2010 Expedition, ref. CSD2008-00077) to Carlos M. Duarte and the

## REFERENCES

- Abell, G. C. J., and Bowman, J. P. (2005). Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbiol. Ecol.* 51, 265–277. doi: 10.1016/j.femsec.2004.09.001
- Acinas, S. G., Ferrera, I., Sarmento, H., Díez-Vives, C., Forn, I., Ruiz-González, C., et al. (2014). Validation of a new catalysed reporter deposition-fluorescence *in situ* hybridization probe for the accurate quantification of marine *Bacteroidetes* populations. *Environ. Microbiol.* 17, 3557–3569. doi: 10.1111/1462-2920.12517
- Alonso-Sáez, L., Balagué, V., Sà, E. L., Sánchez, O., González, J. M., Pinhassi, J., et al. (2007). Seasonality in bacterial diversity in north-west Mediterranean coastal waters: assessment through clone libraries, fingerprinting and FISH. *FEMS Microbiol. Ecol.* 60, 98–112. doi: 10.1111/j.1574-6941.2006.00276.x
- Baek, K., Choi, A., Kang, I., Lee, K., and Cho, J. C. (2013). *Kordia antarctica* sp. nov., isolated from Antarctic seawater. *Int. J. Syst. Evol. Microbiol.* 63, 3617–3622. doi: 10.1099/ijs.0.052738-0
- Boucher, Y., Cordero, O. X., and Takemura, A. (2011). Local mobile gene pools rapidly cross species boundaries to create endemism within global *Vibrio cholerae* populations. *mBio* 2:e00335-10. doi: 10.1128/mBio.00335-10
- Brown, M. V., and Fuhrman, J. A. (2005). Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat. Microb. Ecol.* 41, 15–23.
- Buchan, A., LeCleir, G. R., Gulvik, C. A., and González, J. M. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. *Nat. Rev. Micro* 12, 686–698. doi: 10.1038/nrmicro3326
- Caro-Quintero, A., and Konstantinidis, K. T. (2012). Bacterial species may exist, metagenomics reveal. *Environ. Microbiol.* 14, 347–355. doi: 10.1111/j.1462-2920.2011.02668.x
- Choi, A., Oh, H.-M., Yang, S.-J., and Cho, J.-C. (2011). *Kordia periserrulae* sp. nov., isolated from a marine polychaete *Periserrula leucophryna*, and emended description of the genus *Kordia*. *Int. J. Syst. Evol. Microbiol.* 61, 864–869. doi: 10.1099/ijs.0.022764-0
- Cohan, F. M. (2006). Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1985–1996. doi: 10.1098/rstb.2006.1918
- Cordero, O. X., and Polz, M. F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* 12, 263–273. doi: 10.1038/nrmicro3218

US Department of Energy Joint Genome Institute 2011 Microbes program grant CSP602 grant to SA (Contract No. DE-AC02-05CH11231). Other contributions were the US NSF grants OCE-1232982 and OCE-1335810 to RS. MR-L held a Ph.D. Fellowship FPI (BES-2014-068285) funded by the Spanish Ministry of Economy and Competitiveness.

## ACKNOWLEDGMENTS

We thank our fellow scientists and the crew and chief scientists of the different cruise legs involved in *Tara* Oceans and Malaspina expedition for collecting the samples used in this study. This is *Tara* Oceans contribution paper number 57. The work described here is not to be associated with any policy of or sanctioned by the United States National Science Foundation.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01317/full#supplementary-material>

- Cottrell, M. T., and Kirchman, D. L. (2000). 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.
- Crespo, B. G., Pommier, T., Fernández-Gómez, B., and Pedrós-Alió, C. (2013). Taxonomic composition of the particle-attached and free-living bacterial assemblages in the Northwest Mediterranean Sea analyzed by pyrosequencing of the 16S rRNA. *Microbiologyopen* 2, 541–552. doi: 10.1002/mbo3.92
- DeLong, E. F., Franks, D. G., and Alldredge, A. L. (1993). Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol. Oceanogr.* 38, 924–934. doi: 10.4319/lo.1993.38.5.0924
- Díez-Vives, C., Gasol, J. M., and Acinas, S. G. (2014). Spatial and temporal variability among marine *Bacteroidetes* populations in the NW Mediterranean Sea. *Syst. Appl. Microbiol.* 37, 68–78. doi: 10.1016/j.syapm.2013.08.006
- Fandino, L. B., Riemann, L., Steward, G. F., and Azam, F. (2005). Population dynamics of *Cytophaga-Flavobacteria* during marine phytoplankton blooms analyzed by real-time quantitative PCR. *Aquat. Microb. Ecol.* 40, 251–257.
- Fernández-Gómez, B., Fernández-Guerra, A., Casamayor, E. O., González, J. M., Pedrós-Alió, C., and Acinas, S. G. (2012). Patterns and architecture of genomic islands in marine bacteria. *BMC Genomics* 13:347. doi: 10.1186/1471-2164-13-347
- Fernández-Gómez, B., Richter, M., Schüller, M., Pinhassi, J., Acinas, S. G., González, J. M., et al. (2013). Ecology of marine Bacteroidetes: a comparative genomics approach. *ISME J.* 7, 1026–1037. doi: 10.1038/ismej.2012.169
- Ferrera, I., Gasol, J. M., Sebastian, M., Hojerova, E., and Kobizek, 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. doi: 10.1128/AEM.00208-11
- Fuhrman, J. A., Schwalbach, M. S., and Stingl, U. (2008). Proteorhodopsins: an array of physiological roles? *Nat. Rev. Microbiol.* 6, 488–494. doi: 10.1038/nrmicro1893
- Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., et al. (2005). Opinion: re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* 3, 733–739. doi: 10.1038/nrmicro1236
- Glöckner, F. O., Fuchs, B. M., and Amann, R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence *in situ* hybridization. *Appl. Environ. Microbiol.* 65, 3721–3726.

- Gómez-Consarnau, L., González, J. M., Coll-Lladó, M., Gourdon, P., Pascher, T., Neutze, R., et al. (2007). Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* 445, 210–213. doi: 10.1038/nature05381
- Gómez-Pereira, P. R., Schüller, M., Fuchs, B. M., 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. doi: 10.1111/j.1462-2920.2011.02555.x
- González, J. M., Fernández-Gómez, B., Fernández-Guerra, A., Gómez-Consarnau, L., Sánchez, O., Coll-Lladó, M., et al. (2008). Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). *Proc. Natl. Acad. Sci. U.S.A.* 105, 8724–8729. doi: 10.1073/pnas.0712027105
- Habib, C., Houel, A., Lunazzi, A., Bernardet, J. F., Olsen, A. B., Nilsen, H., et al. (2014). Multilocus sequence analysis of the marine bacterial genus *Tenacibaculum* suggests parallel evolution of fish pathogenicity and endemic colonization of aquaculture systems. *Appl. Environ. Microbiol.* 80, 5503–5514. doi: 10.1128/AEM.01177-14
- Hameed, A., Shahina, M., Lin, S.-Y., Cho, J.-C., Lai, W.-A., and Young, C.-C. (2013). *Kordia aquimaris* sp. nov., a zeaxanthin-producing member of the family *Flavobacteriaceae* isolated from surface seawater, and emended description of the genus *Kordia*. *Int. J. Syst. Evol. Microbiol.* 63, 4790–4796. doi: 10.1099/ijms.0.056051-0
- Ivars-Martínez, E., D'Auria, G., Rodríguez-Valera, F., Sánchez-Porro, C., Ventosa, A., Joint, I., et al. (2008). Biogeography of the ubiquitous marine bacterium *Alteromonas macleodii* determined by multilocus sequence analysis. *Mol. Ecol.* 17, 4092–4106. doi: 10.1111/j.1365-294X.2008.03883.x
- Kashtan, N., Roggensack, S. E., Rodrigue, S., Thompson, J. W., Biller, S. J., Coe, A., et al. (2014). Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* 344, 416–420. doi: 10.1126/science.1248575
- Khandeparker, R., Meena, R. M., and Deobagkar, D. (2014). Bacterial diversity in deep-sea sediments from Afanasy Nikitin seamount, equatorial Indian Ocean. *Geomicrobiol. J.* 31, 942–949. doi: 10.1080/01490451.2014.918214
- Kirchman, D. L. (2002). The ecology of *Cytophaga*-Flavobacteria in aquatic environments. *FEMS Microbiol. Ecol.* 39, 91–100. doi: 10.1111/j.1574-6941.2002.tb00910.x
- Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102
- Lai, Q., Shao, Z., Liu, Y., Xie, Y., Du, J., and Dong, C. (2015). *Kordia zhangzhouensis* sp. nov., isolated from surface freshwater. *Int. J. Syst. Evol. Microbiol.* 65, 3379–3383. doi: 10.1099/ijsem.0.000424
- Landry, M. R., Brown, S. L., Campbell, L., Constantinou, J., and Liu, H. (1998). Spatial patterns in phytoplankton growth and microzooplankton grazing in the Arabian Sea during monsoon forcing. *Deep. Res. Part II Top. Stud. Oceanogr.* 45, 2353–2368. doi: 10.1016/S0967-0645(98)00074-5
- Lefort, T., and Gasol, J. (2013). Global-scale distributions of marine surface bacterioplankton groups along gradients of salinity, temperature, and chlorophyll: a meta-analysis of fluorescence in situ hybridization studies. *Aquat. Microb. Ecol.* 70, 111–130. doi: 10.3354/ame01643
- Logares, R., Sunagawa, S., Salazar, G., Cornejo-Castillo, F. M., Ferrera, I., Sarmiento, 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.* 16, 2659–2671. doi: 10.1111/1462-2920.12250
- Martínez-García, M., Swan, B. K., Poulton, N. J., Gomez, M. L., Masland, D., Sieracki, M. E., et al. (2012). High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. *ISME J.* 6, 113–123. doi: 10.1038/ismej.2011.84
- Medini, D., Donati, C., Tettelin, H., Massignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi: 10.1016/j.gde.2005.09.006
- Nguyen, D., Maranger, R., Balagué, V., Coll-Lladó, M., Lovejoy, C., and Pedrós-Alió, C. (2015). Winter diversity and expression of proteorhodopsin genes in a polar ocean. *ISME J.* 9, 1835–1845. doi: 10.1038/ismej.2015.1
- Park, S., Jung, Y.-T., and Yoon, J.-H. (2014). *Kordia jejudonensis* sp. nov., isolated from the junction between the ocean and a freshwater spring, and emended description of the genus *Kordia*. *Int. J. Syst. Evol. Microbiol.* 64, 657–662. doi: 10.1099/ijms.0.058776-0
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Pedrós-Alió, C. (2012). The rare bacterial biosphere. *Annu. Rev. Mar. Sci.* 4, 449–466. doi: 10.1146/annurev-marine-120710-100948
- Pinhassi, J., Delong, E. F., Béjà, O., González, J. M., and Pedrós-Alió, C. (2016). Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. *Microbiol. Mol. Biol. Rev.* 80, 929–954. doi: 10.1128/MMBR.00003-16
- Qi, F., Shao, Z., Li, D., Huang, Z., and Lai, Q. (2016). *Kordia ulvae* sp. nov., a bacterium isolated from the surface of green marine algae *Ulva* sp. *Int. J. Syst. Evol. Microbiol.* 66, 2623–2628. doi: 10.1099/ijsem.0.001098
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, 590–596. doi: 10.1093/nar/gks1219
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: <http://www.R-project.org/>
- Rappé, M. S., and Giovannoni, S. J. (2003). The uncultured microbial majority. *Annu. Rev. Microbiol.* 57, 369–394. doi: 10.1146/annurev-micro.57.030502.090759
- Raymann, K., Bobay, L. M., Doak, T. G., Lynch, M., and Gribaldo, S. (2013). A genomic survey of Reb homologs suggests widespread occurrence of R-bodies in proteobacteria. *G3* 3, 505–516. doi: 10.1534/g3.112.005231
- Roullier, F., Berline, L., Guidi, L., Durrieu De Madron, X., Picheral, M., Sciandra, A., et al. (2014). Particle size distribution and estimated carbon flux across the Arabian Sea oxygen minimum zone. *Biogeosciences* 11, 4541–4557. doi: 10.5194/bg-11-4541-2014
- Ruiz-González, C., Lefort, T., Massana, R., Simó, R., and Gasol, J. M. (2012). Diel changes in bulk and single-cell bacterial heterotrophic activity in winter surface waters of the northwestern Mediterranean Sea. *Limnol. Oceanogr.* 57, 29–42. doi: 10.4319/lo.2012.57.1.0029
- Salazar, G., Cornejo-Castillo, F. M., Borrull, E., Díez-Vives, C., Lara, E., Vaqué, D., et al. (2015). Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. *Mol. Ecol.* 24, 5692–5706. doi: 10.1111/mec.13419
- Schattenhofer, M., Fuchs, B. M., Amann, R., Zubkov, M. V., Tarran, G. A., and Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ. Microbiol.* 11, 2078–2093. doi: 10.1111/j.1462-2920.2009.01929.x
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Schlöter, M. (2000). Ecology and evolution of bacterial microdiversity. *FEMS Microbiol. Rev.* 24, 647–660. doi: 10.1016/S0168-6445(00)00051-6
- Sohn, J. H., Lee, J. H., Yi, H., Chun, J., Bae, K. S., Ahn, T. Y., et al. (2004). *Kordia algicida* gen. nov., sp. nov., an algicidal bacterium isolated from red tide. *Int. J. Syst. Evol. Microbiol.* 54, 675–680. doi: 10.1099/ijms.0.02689-0
- Spudich, J. L. (2006). The multitasking microbial sensory rhodopsins. *Trends Microbiol.* 14, 480–487. doi: 10.1016/j.tim.2006.09.005
- Stepanovskas, R., and Sieracki, M. E. (2007). Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9052–9057. doi: 10.1073/pnas.0700496104
- Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science* 348:1261359. doi: 10.1126/science.1261359
- Sunagawa, S., Mende, D. R., Zeller, G., Izquierdo-Carrasco, F., Berger, S. A., Bork, P., et al. (2013). Metagenomic species profiling using universal phylogenetic marker genes. *Nat. Methods* 10, 1196–1199. doi: 10.1038/nmeth.2693
- Swan, B. K., Martínez-García, M., Preston, C. M., Sczyrba, A., Woyke, T., Lamy, D., et al. (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 333, 1296–1300. doi: 10.1126/science.1203690

- Swan, B. K., Tupper, B., Szczyrba, A., Lauro, F. M., Martínez-García, M., González, J. M., et al. (2013). Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11463–11468. doi: 10.1073/pnas.1304246110
- Tang, K., Lin, D., Liu, K., and Jiao, N. (2015). Draft genome sequence of *Parvularcula oceani* JLT2013<sup>T</sup>, a rhodopsin-containing bacterium isolated from deep-sea water of the Southeastern Pacific. *Mar. Genomics* 24, 211–213. doi: 10.1016/j.margen.2015.05.013
- Thompson, F. L., Gevers, D., Thompson, C. C., Dawyndt, P., Naser, S., Hoste, B., et al. (2005). Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. *Appl. Environ. Microbiol.* 71, 5107–5115. doi: 10.1128/AEM.71.9.5107-5115.2005
- Villareal, T. A., Brown, C. G., Brzezinski, M. A., Krause, J. W., and Wilson, C. (2012). Summer diatom blooms in the north Pacific subtropical gyre: 2008–2009. *PLoS ONE* 7:e33109. doi: 10.1371/journal.pone.0033109
- Wang, J., Kan, J., Borecki, L., Zhang, X., Wang, D., and Sun, J. (2016). A snapshot on spatial and vertical distribution of bacterial communities in the eastern Indian Ocean. *Acta Oceanol. Sin.* 35, 85–93. doi: 10.1007/s13131-016-0871-4
- Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., et al. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* 37, 463–464. doi: 10.1099/00207713-37-4-463
- Woyke, T., Xie, G., Copeland, A., González, J. M., Han, C., Kiss, H., et al. (2009). Assembling the marine metagenome, one cell at a time. *PLoS ONE* 4:e5299. doi: 10.1371/journal.pone.0005299
- Wu, H., Guo, Y., Wang, G., Dai, S., and Li, X. (2011). Composition of bacterial communities in deep-sea sediments from the South China Sea, the Andaman Sea and the Indian Ocean. *Afr. J. Microbiol. Res.* 5, 5273–5283. doi: 10.5897/AJMR11.302
- Yoshizawa, S., Kawanabe, A., Ito, H., Kandori, H., and Kogure, K. (2012). Diversity and functional analysis of proteorhodopsin in marine Flavobacteria. *Environ. Microbiol.* 14, 1240–1248. doi: 10.1111/j.1462-2920.2012.02702.x
- Yoshizawa, S., Song, J., Kogure, K., Choi, A., Joung, Y., Cho, J.-C., et al. (2015). *Aurantivirga profunda* gen. nov., sp. nov., isolated from deep-seawater, a novel member of the family Flavobacteriaceae. *Int. J. Syst. Evol. Microbiol.* 65, 4850–4856. doi: 10.1099/ijsem.0.000662

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Royo-Llonch, Ferrera, Cornejo-Castillo, Sánchez, Salazar, Stepanauskas, González, Sieracki, Speich, Stemmann, Pedrós-Alió and Acinas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.