Global ocean resistome revealed: exploring Antibiotic Resistance Genes (ARGs) abundance and distribution on TARA oceans samples through machine learning tools

Rafael R. C. Cuadrat[1], Maria Sorokina[2], Bruno G. Andrade[3], Alberto M. R. Dávila[4]

1 - Molecular Epidemiology Department, German Institute of Human Nutrition - DIfE
2 - Friedrich-Schiller University, Lessingstrasse 8, 07743 Jena, Germany, ORCID: 0000-0001-9359-7149
3 - Embrapa Southeast Livestock - EMBRAPA
4 - Computational and Systems Biology Laboratory, Oswaldo Cruz Institute, FIOCRUZ. Av Brasil 4365, Rio de Janeiro, RJ, Brasil. 21040-900.

Abstract

The rise of antibiotic resistance (AR) in clinical settings is one of the biggest modern global public health concerns, therefore, the understanding of its mechanisms, evolution and global distribution is a priority due to its impact on the treatment course and patient survivability. Besides all efforts in the elucidation of AR mechanisms in clinical strains, little is known about its prevalence and evolution in environmental uncultivable microorganisms. In this study, 293 metagenomic and 10 metatranscriptomic samples from the TARA oceans project were used to detect and quantify environmental Antibiotic Resistance Genes (ARGs) using modern machine learning tools. After extensive manual curation, we show their global distribution, abundance, taxonomy and phylogeny, their potential to be horizontally transferred by plasmids or viruses and their correlation with environmental and geographical parameters. After manual curation, we identified a total of 99,205 environmental ORFs as potential ARGs. These ORFs belong to 560 ARG families that confer resistance to 26 antibiotic classes. A total of 149 ORFs were classified as viral sequences. In addition, 24,567 ORFs were found in contigs classified as plasmidial sequences, suggesting the importance of mobile genetic elements in the dynamics of ARGs transmission. From the 13,163 identified ARGs passing all the criteria for quantification analysis, 4,224 were expressed in at least one of the 10 metatranscriptomic samples (FPKM >5). Moreover, 4,804 contigs with more than 2 ARGs were found, including 2 plasmids with 5 different ARGs, highlighting the potential presence of multi-resistant microorganisms in natural environment and/or non-impacted by human presence oceans, together with the possibility of Horizontal Gene Transfer (HGT) between clinical and natural environments. The abundance of ARGs in 293 samples showed different patterns of distribution, with some classes being significatively more abundant in Coastal Biomes. Finally, we identified ARGs conferring resistance to some of the most relevant clinical antibiotics, revealing the presence of 15 ARGs from the recently discovered MCR-1 family with high abundance on Polar Biomes. A total of 5 MCR-1 ORFs are present in the genus *Psychrobacter*, an opportunistic bacteria that can cause fatal infections in humans.

Our results are available on Zenodo in MySQL database dump format and all the code used for the analyses, including jupyter notebooks can be accessed on github (https://github.com/rcuadrat/ocean_resistome).

Keywords:

Antibiotic resistance, ARG, marine metagenomics, MCR-1, TARA oceans

Introduction

The world is facing a dramatic emergence of cases of multi-resistant bacterial infections, and leading organizations as the Center for Disease Control and Prevention (CDC) and World Health Organization (WHO) classify the problem as “urgent, serious and concerning” [1]. Antibiotic resistant bacteria causes 700.000 deaths per year, being an economic burden to the entire world and in particular for developing countries. Furthermore, if the emergence of multi-resistant bacteria continues at the same rate projections then they will cause 10 million deaths per year by 2050, which is more than deaths caused by cancer [2,3].

Antibiotic resistance (AR) is a natural phenomenon, and some of the most common resistance mechanisms like beta-lactamase are estimated to have emerged more than 1 billion years ago [4,5]. This estimation is based on the prediction of ancestral enzymes, in the Precambrian era, through Bayesian statistics based on modern ARGs [4] (1.3-3 Gyr). The collection of antibiotic resistance genes (ARGs) in a given environment, also called resistome, is a natural feature of microbial communities. The resistome is part of both inter- and intra-community communication and defense repertoires of organisms sharing the same biological niche [6,7].

Over the years, evidence of the role of natural environments as a reservoir of resistance genes has been shown in different environments, such as hospital wastewater [11], water supply [12], remote pristine Antarctic soils [13] or impacted Arctic tundra wetlands [14], and it is believed that anthropogenic activity (e.g. over-usage of antibiotics and their subsequent release in waste waters and in the environment) could lead to the rise of ARGs in the clinical environment. Therefore, the investigation of the natural context of ARGs such as the geographic location, dynamics and in particular their presence on horizontally transferable mobile genetic elements (MGEs), such as plasmids, transposons and phages, is crucial to assess their potential to emerge and spread [8–10]. Thanks to the development of DNA sequencing technology, the metagenomics and bioinformatics fields, it is now possible to study the presence and prevalence of ARGs in different environments, however, most of the studies investigated one or few classes of ARGs, being limited to specific environments and to the technology that were still being developed.

Recently the number of metagenomic projects stored in public databases have been growing, but the lack of related metadata makes it difficult to conduct high-throughput gene screenings and correlations with environmental factors. Fortunately,

the TARA oceans project [15], measured several environmental conditions, such as temperature, salinity, geographical location, pH, etc, across the globe and stored them as structured metadata, which, together with the deep sequencing of metagenomic samples, allows the use of machine learning and deep learning approaches to search for patterns of species and genes with environmental parameters and extract relevant biological information.

In this study, we applied DeepARG [16], a deep learning approach for ARG identification, trained to search for over 30 classes of ARG based on the similarity distribution of sequences in the 3 ARG databases [17] on the assembled TARA oceans data [18] for 12 regions. The ARG identification was then followed by abundance quantification for each individual sample for 293 samples and association analyses between the quantification of ARGs and environmental parameters using Ordinary Least Squares (OLS) regression. We also explored the presence of ARGs in mobile genetic elements (MGEs) in TARA samples in order to verify the potential of a given environment to act as a reservoir of ARGs.

Material and Methods

Metagenomic data

A total of 12 co-assembled metagenomes, with contigs larger than 1 kilobase from the regions explored by the Tara Oceans expedition were obtained from the dataset published in 2017 by Delmont et al. [19]. Raw reads of 378 runs from 243 samples from the Tara Oceans project were downloaded from the EBI ENA database (https://www.ebi.ac.uk/ena) from studies PRJEB1787, PRJEB6606 and PRJEB4419. The sequence identifier and metadata were obtained from Companion Tables Ocean Microbiome (EMBL) [20]. All these samples were collected in seawater on different depth and filtered to retain different fractions (cell sizes). Also, 10 metatranscriptome samples from the TARA Oceans expedition were obtained from the EBI repository (https://www.ebi.ac.uk/ena/data/view/PRJEB6608) and used in order to evaluate the expression of the previously detected ARGs.

Obtaining and analysing putative ARGs from Tara Ocean Project contigs

Open reading frame (ORF) prediction and extraction was performed on the 12 co-assembled metagenomes using the software MetaGeneMark v3.26 [21] with default parameters.

The screening for ARGs with DeepARG [16] was performed on the extracted ORFs using the default gene models and sequences with probability equal or superior to 0.8 of being an ARG. The contigs containing at least one putative ARG were analysed with the tool PlasFlow 1.1 [23] in order to check if those contigs are plasmids or chromosomal sequences. We also investigated the number and

distribution of contigs with 2 or more putative ARGs, in order to check for multiple resistance presence and/or whole ARG operons from environmental samples.
The ORFs of putative ARGs were extracted with a Python script for following analyses. All ORFs classified as ARG were submitted to Kaiju v1.6.2 [22] for taxonomic classification, with the option “run mode” set as “greedy”.

In order to check for mis-annotations and inconsistencies, a manual curation on each of the ARG was conducted in sequences from both deepARGdb [16] and obtained from environmental samples. Next, blastp queries [24] were performed against the non-redundant protein database, with default parameters and the results were manually inspected. Conserved domains (CDDs) and annotations in the source databases (ARDB, CARD and UNIPROT) were manually inspected.

ARGs quantification on metagenome and metatranscriptome samples

Environmental ARGs identified in the assembled database after the manual curation were used as reference for mapping all the raw reads from the 378 metagenomic and 10 metatranscriptomic samples, using BBMAP v 37.90. The coverage and abundance of each ARG class for every sample was obtained from BBMAP (FPKM). Next, data was normalized by Reads Per Kilo Genome equivalents (RPKG), being the Average Genome Size (AGS) and Genome Equivalents (GE) estimated by the software MicrobeCensus v 1.0.7 [25].

Phylogenetic analysis of environmental ARGs

The clinically relevant ARG family MCR-1 was used for phylogenetic analyses with the ARGs found in public databases, such as NCBI and deepARGdb. A multiple alignment and phylogenetic trees were generated using the standard pipeline found in phylogeny.fr [26]. Sequences were aligned with Muscle [27], then conserved blocks extracted with gblocks [28], phylogenetic trees generated with phyML [29], using “WAG” as substitution model and the statistical test alrt.

Statistical analyses

RPKG values for all ORFs of each ARG family were summed for each sample. The features were grouped by environmental features, such as the sample depth, biogeographic biomes, ocean and sea regions and the concatenation of size fraction lower and upper thresholds (here defined as “Fraction”). Box plots were generated for ARG classes grouped by each of the previously cited parameters. Pairwise Tukey HSD tests and PCA analyses were also performed on the same. The multivariate linear regression using OLS models was performed considering the following formula:

ARG_{RPKG} ~ Marine provinces + Environmental Feature + Ocean sea regions + Fraction + Biogeographic biomes + Latitude + Longitude + NO2 + PO4 + NO2NO3 + SI + miTAGSILVATaxo Richness + miTAG_SILVA_Phylo_Diversity + miTAG_SILVA_Chao + miTAG_SILVA_ace + miTAG_SILVA_Shannon + OG_Shannon + OG_Richness + OG_Evenness + FC_heterotrophs_cells_mL + FC_autotrophs_cells_mL + FC_bacteria_cells_mL + FC_picoeukaryotes_cells_mL

Where ARG_{RPKG} (the dependent variable) is the sum of RPKM of all ARGs in a given class and all the dependent variables are selected environmental parameters. Anova test was conducted on the coefficients obtained from the OLS regression. Statistical analysis were performed using Python 2.7 scripts with data science and machine learning libraries, such as pandas, scipy, numpy, seaborn and sklearn.

Database design and implementation

A manually curated MySQL database was created with the environment ARGs described in this study plus the original ARGs from deepARGdb. We provide the SQL file and in the future a web interface to query, visualize and analyse similar datasets will be made available. The full MySQL dump is available at https://doi.org/10.5281/zenodo.3404245.

Results

Environmental ARGs prediction and manual curation

The identification of ORFs, performed with the software MetaGeneMark on the 12 co-assembled metagenomes, returned a total of 41,249,791 ORFs from 15,600,278 assembled contigs. The identified ORFs were used in a screening for ARGs with the software DeepARG, of which 116,425 (0.28% of ORFs) were classified as putative antibiotic resistance genes (ARGs), belonging to 594 gene families and 28 ARG classes. The number of contigs, ORFs and putative ARGs for each region (metagenomic co-assembly) can be found on Table 1 and the distribution of putative ARG classes within each co-assembled metagenome can be found in Supplementary Table 1.

Table 1: Number of contigs, ORFs and putative ARGs for each region (metagenomic co-assembly)

Sample	**Contigs**	**ORFs**	**#ARG**

TARA_ANE_RAW	1382239	3686619	11283
TARA_ANW_RAW	1267057	3308183	9994
TARA_ASE_RAW	766472	1842937	4955
TARA_ASW_RAW	989154	2502625	6902
TARA_ION_RAW	1608737	4257000	9830
TARA_IOS_RAW	1475271	3736229	9909
TARA_MED_RAW	1146485	3344341	10160
TARA_PON_RAW	1663221	4368436	11814
TARA_PSE_RAW	2722083	7155344	20587
TARA_PSW_RAW	1124813	3080817	9224
TARA_RED_RAW	1039053	2937951	8924
TARA_SOC_RAW	415693	1029309	2843
Total	15,600,278	41,249,791	116,425

Afterwards, we conducted a manual curation on all 594 genes, in order to investigate different scenarios: (i) miss-annotated genes or with low experimental evidence as ARGs on the databases; (ii) housekeeping genes that confers resistance when specific mutations arise; (iii) housekeeping genes conferring resistance when overexpressed; (iv): regulatory sequences, responsible for ARGs activation or overexpression of housekeeping genes (leading to resistance phenotype); (v): sequences with both similarity to ARGs and non-ARGS, for example, from the same super family and/or sharing domains.

For the scenario (i), the genes were completely removed from both our database and downstream analysis. Genes on the scenario (ii), (iii) and (iv) were kept in the database for further studies but not used in the quantification and statistical analyses. The first due to technical limitations for differentiate the wild type from

mutant with our approach and the later two because we do not have expression data from all the samples analysed. For the scenario (v), when it was possible, we conducted phylogenetic analysis in order to try to separate ARGs from non-ARGs homologous, and when impossible, those genes were kept in the database as "waiting for validation" ARGs, (and excluded from quantification and statistical analyses).
Regarding the scenario (i), we identified 34 ARG families as missannotated in the source database or with low quality annotation. For instance, the *msr*B gene encodes an ABC-F subfamily protein, leads to resistance to the antibiotics erythromycin and streptogramin B, but the fasta sequence on the database is actually a Methionine Sulfoxide Reductases B, also called *msr*B but not a resistance gene. In addition, another miss annotated ARG family is *pat*A, an ABC transporter of *Streptococcus pneumoniae*, conferring resistance to fluoroquinolones, however, the sequence in the CARD database was a putrescine aminotransferase (*pat*A).

After removing ORFs identified for the scenario (i), a total of 99,205 ORFs remained for non-quantitative analyses (taxonomical classification, presence of multi-ARGs in the same contig, and potential HGT ARGs - by checking if they are in plasmids or phage).

Regarding the scenario (ii): we identified 10 families which genes are housekeeping and mutations could lead to resistance; scenario (iii): we found 9 ARGs in which the overexpression leads to resistance; scenario (iv): we found 41 regulatory sequences, and for scenario (v): we identified 187 families that cannot be distinguished by similarity alone from genes not involved with the resistance phenotype. After the removal of these genes, a total of 13,163 ORFs (from 116,425) belonging to 313 (from 594) families were retained for quantification and statistical analysis.

The supplementary table 1 shows the results of our manual curation.

Figure 1 shows the 20 more abundant ARG families (in number of ORFs found in all 12 metagenomes) with their antibiotic classes (only for the ARGs marked for quantification analyses).

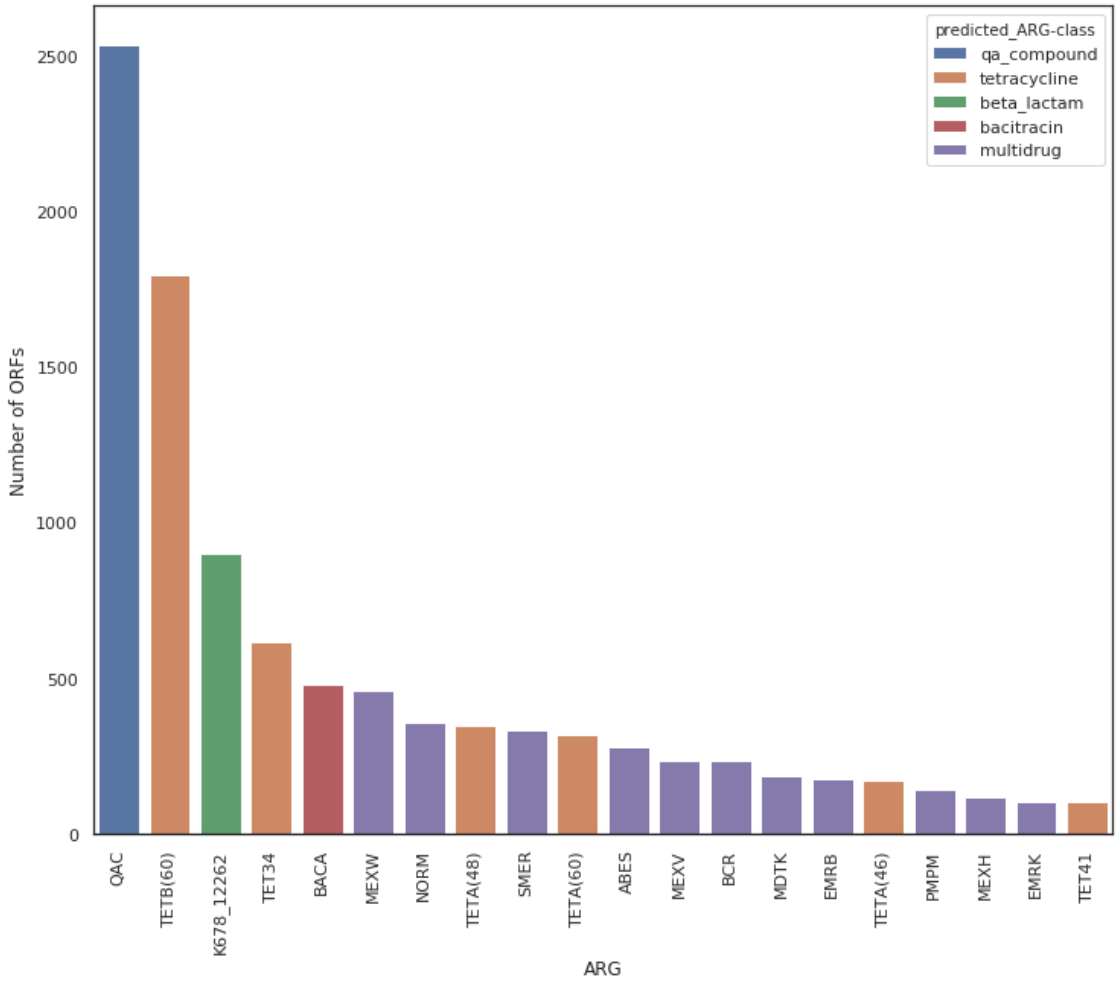

Figure 1: The 20 most abundant ARGs (in number of ORFs detected on the study) and their respective antibiotic classes.

Multiple ARGs in chromosomes and plasmids and its taxonomic classification

The presence of 2 or more ARGs in a single contig was also checked in order to find possible multi-resistant organisms. For this analysis, we only removed the ARGs from scenario (i), because the presence of putative ARGs in the same contig and/or plasmid can give us additional functional evidence, for example, regulatory sequences (scenario iv) in the same contig of multiple sequences from scenario v (not easily distinguished from non-ARGs homologs), can be an AR operon.

We identified 4,804 contigs with 2 or more putative ARGs, of which 2 contigs harboured a maximum of 11 putative ARGs, and 2 contigs classified as plasmids with 5 putative ARGs, suggesting the presence of multi-resistant microorganisms in these environments (Figure 2). In fact, a total of 24,567 putative ARGs (24.76% of the 99,205 ARGs in scenario ii,iii,iv and v) were classified as plasmids. Figure 2

shows the distribution of contigs and their respective number of ARGs in plasmids and chromosomes.

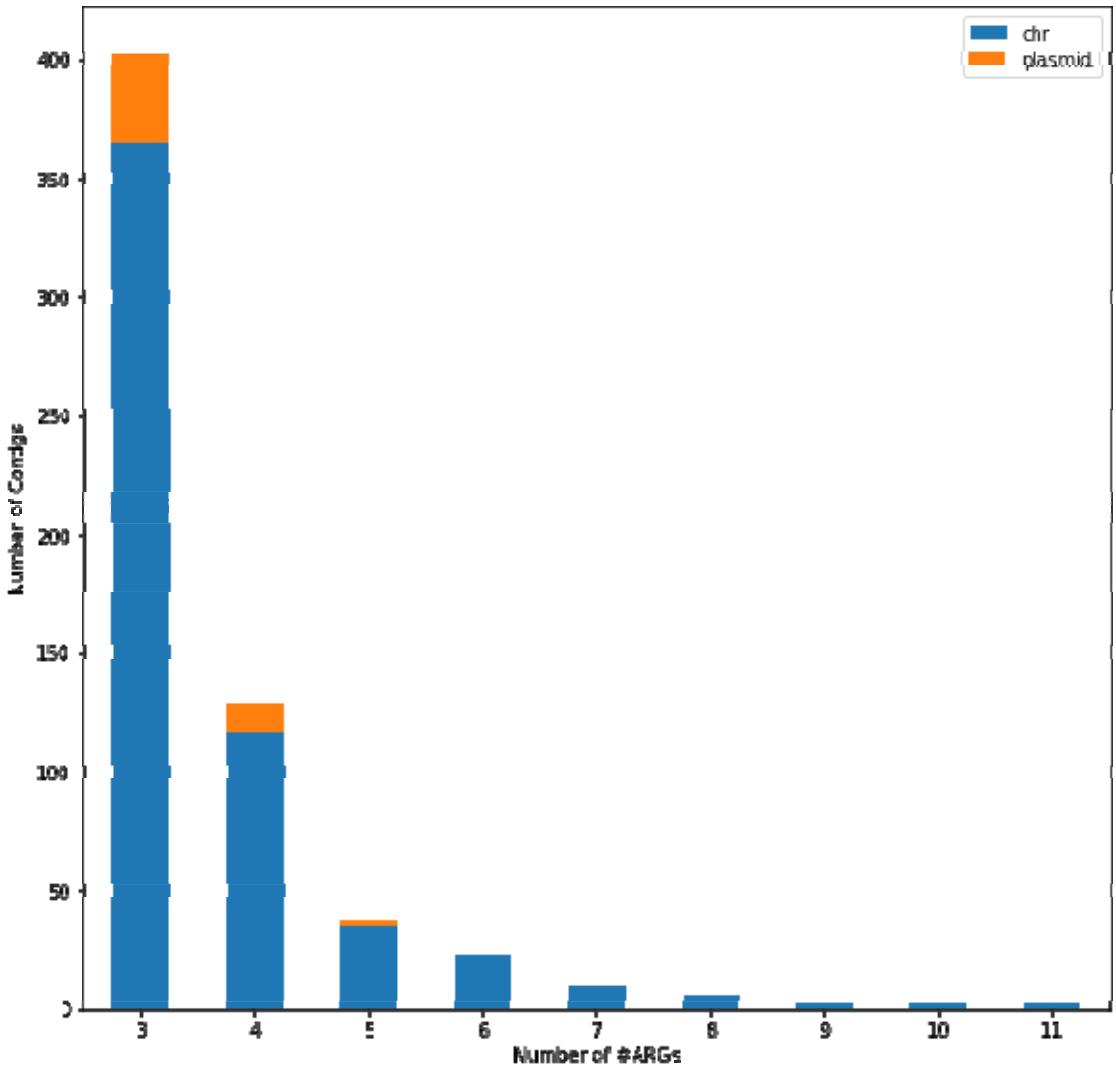

Figure 2: Distribution of contigs with multiple ARGs (more than 2 per contig) in chromosome (blue bar) and plasmids (orange bar).

The taxonomic classification of putative ARGs using Kaiju [22], allowed us to classify 97,397 (98.17%) sequences to some taxonomic level. The largest class is alphaproteobacteria (35,638 sequences), as expected for marine environments. A total of 149 ARGs were classified as virus (0.15% of total ARGs in scenario ii,iii,iv and v). The most abundant is Chrysochromulina ericina virus (CeV) (28 ARGs).

However, all the 149 ARGs are in the scenario iv, and further investigations should be done in order to curate these sequences.

Impact of filter size on average genome size (AGS) on Tara Oceans samples

For all 378 sample runs, the software MicrobeCensus v 1.0.7 [25] was used to estimate the average genome size of the sample. The method is used to avoid bias of genome sequence coverage between samples populated with different size of genomes, what is expected due to filtration methods in the TARA study.
The authors of the software proposed the method RPKG for gene abundance normalization in microbiome studies, described in [25]. Since the TARA samples had different lower and upper thresholds of filtration size, we grouped samples by both thresholds to investigate the differences in the AGS. This tool use a set of gene markers to infer the AVS, and as the markers are mainly present in cellular genomes, the fractions enriched for virus and giant virus (< 0.22 and 0.1-0.22) showed very biased and aberrant results for AVS (very high, up to 395.4 megabases) and genome equivalents (consequently very low, minimal), being then removed from the quantitative analysis. The remaining 293 samples were mapped to the ARGs (all versus all) and the RPKG calculated for each of the environmental ARGs.
Also, 10 RNA-seq samples (metatranscriptome) were used for evidence of gene expression, and a total of 4,224 ARGs (from a total of 13,163) were expressed in at least one sample. From those, 841 are plasmidial ARGs.

Statistical tests

In order to explore the abundance of ARG classes across different marine regions, fractions and layers, we ran pairwise tukey hsd tests on the ARG classes (The sum of RPKG for each class). For example, we wanted to investigate if the Coastal biomes had significatively more ARGs from any class than other more pristine biomes, such as Coastal. For quinolone and bacitracin classes, the Coastal biome shown significant more abundant than Westerlies biome (adjusted p-values 0.0169 and 0.0076). Also for quinolone, Coastal biome shown also borderline significantly more abundance than Trades Biome (adjusted p-value 0.0438).
For fosmidomycin, the Coastal biome shown significantly more abundance of ARGs than Trades Biome (adjusted p-value 0.0014) (Figure 3).

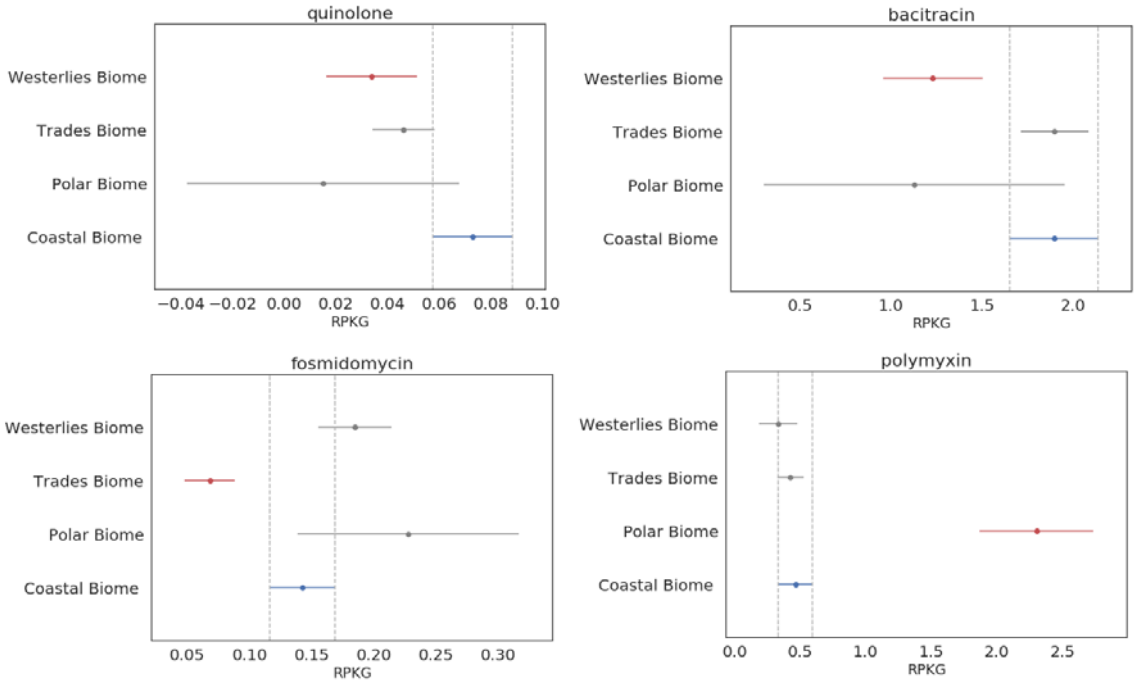

Figure 3: Tukey HSD comparing the RPKG of ARG classes for 4 biomes of Tara Oceans study. a- RPKG for Quinolone ARGs; b- RPKG for Bacitracin ARGs c- RPKG for Fosmidomycin ARGs d- RPKG for Polymyxin ARGs

However, surprisingly, the Polar Biome shown significantly higher RPKG values for ARGs from the class Polymyxin than any other biome.

When comparing provinces, we found significant difference of bleomycin class in 2 Indian provinces compared to most of the other provinces, even if the RPKG values are very low for all the samples.

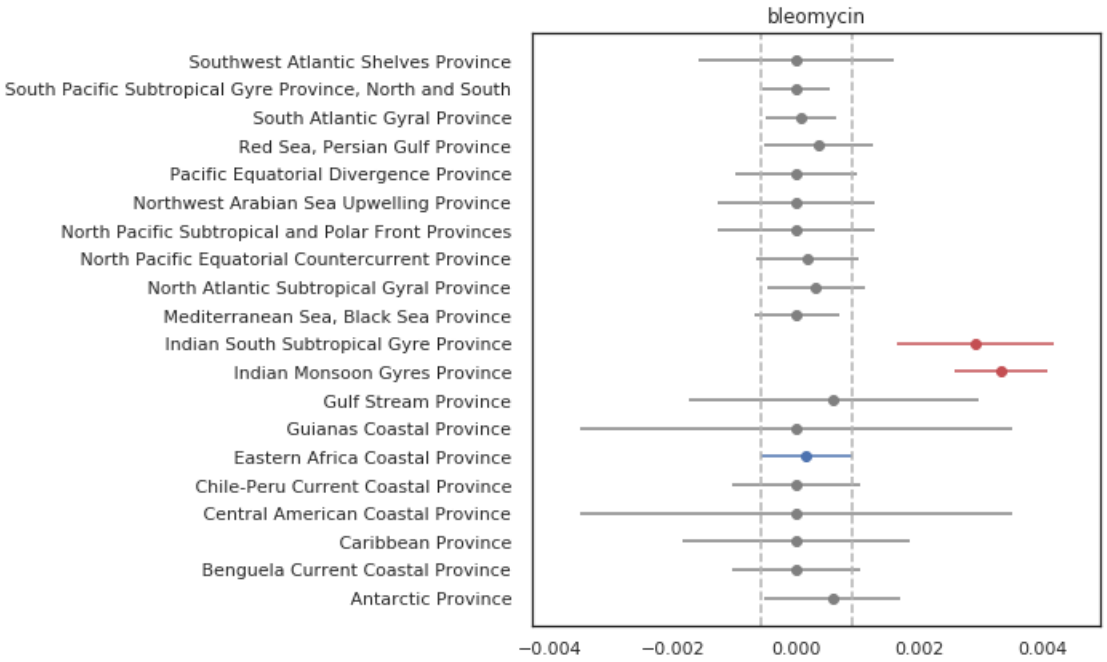

Figure 4: Tukey HSD comparing the RPKG of ARGs (bleomycin class) for Ocean Provinces.

The complete results of all the pairwise tests can be found in supplementary table 2.

Estimating the environmental factors influence on ARGs abundance by linear regression (OLS)

Our OLS models shown mostly geographical parameters affecting the variance of ARGs. However, for some classes, the influence of non-geographical parameters (such as nutrient concentration) on the abundance of ARGs could be demonstrated. For example, for quinolone, the model shows PO4 inverted correlated (beta -0.13, p-value 0.0002), NO2NO3 (Nitrite+Nitrate concentration) direct correlated (beta 0.007, p-value 0.002). The R^2 of this model is 0.57 and p-value $9.82E^{-17}$.
The supplementary table 3 shows all the results anova test on the coefficients of OLS for each class (significant results).

Mobilized colistin resistance genes (MCR-1)

Genes of great clinical relevance were found, for example 15 ORFs of MRC-1 (resistance to polymyxin).

The figure 5 shows the distribution of MRC-1 over the marine regions.

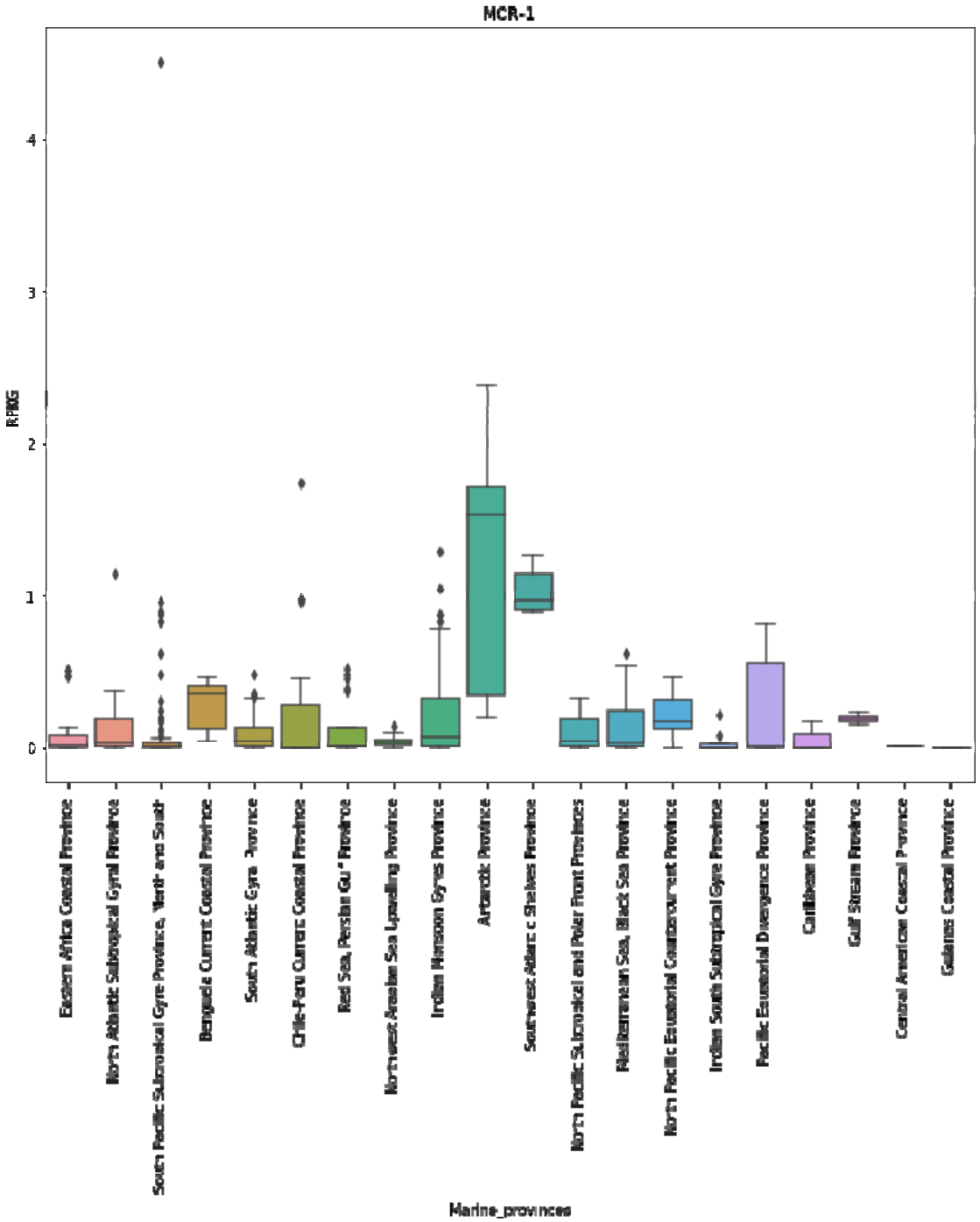

Figure 5: MCR-1 RPKG distribution for each marine province. The most abundant regions were Antarctic Province followed by its adjacent region Atlantic Southwest Shelves Province.

Phylogenetic analysis of MCR-1

The MCR-1 genes are similar to chromosomal polymyxin resistance eptA genes, then in order to classify the environmental ARGs we conducted a phylogenetic analysis using EPTA sequences as outgroups, and also clinical MCR-1 genes. Our results show that the environmental sequences felt on the same big clade with clinical MCR-1 sequences (support value 0.954), being eptA genes external group as expected. The 5 sequences classified by kaiju as belonging to Psychrobacter genus were the closest to the clinical clade, with support value of 0.986 (figure 6).

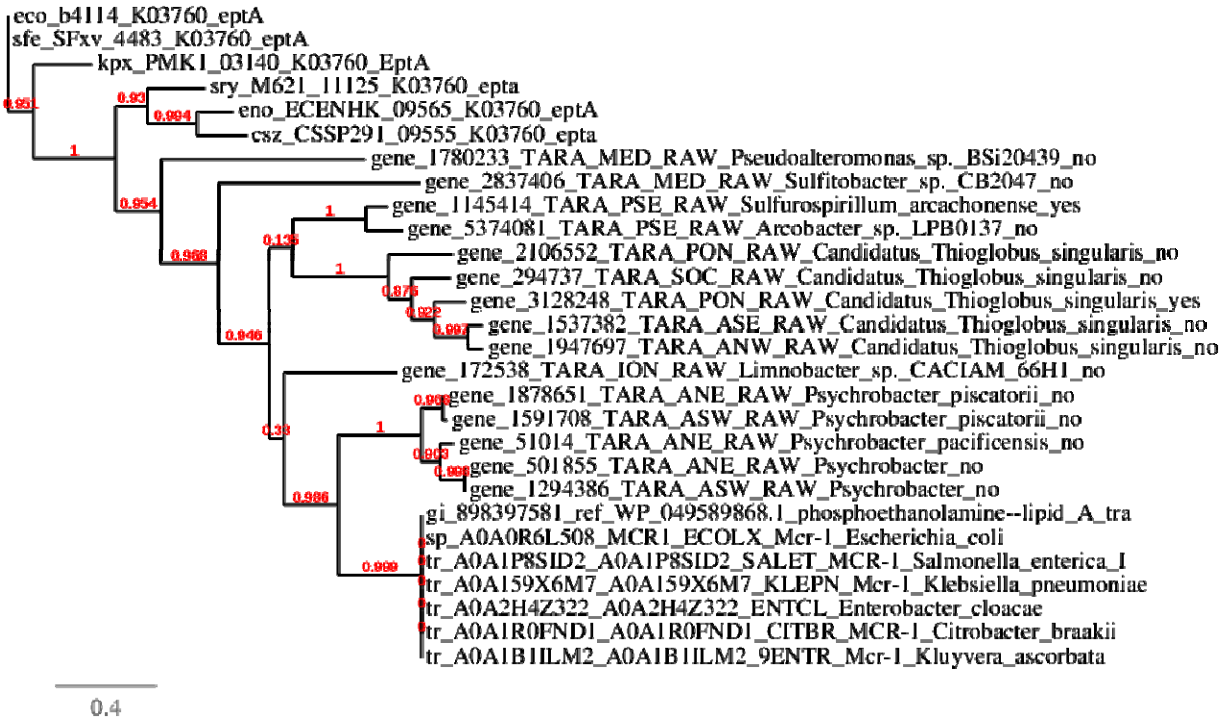

Figure 6: Phylogenetic tree, inferred by standard pipeline from phylogeny.fr (phyML with “WAG” model and statistical test Alrt for support values). Sequences from NCBI for outgroup EPTA and for clinical MCR-1 were used in addition to the samples obtained from our results from Tara co-assemblies.

Discussion

In this study, we explored the distribution, abundance and phylogeny of homologs of antibiotic resistance genes that are conferring resistance to 28 antimicrobial classes, thus revealing the global ocean resistome using public metagenomic databases.
Previous studies screened and analyzed ARGs in genomes and metagenomes of clinical and environmental samples [11,12,13,14], but they focused on one or few ARG classes, being mostly β-lactamases lactamase coding genes. In general, there is a gap in our knowledge about how most of ARGs are distributed in the natural environment, how these environments can act as a reservoir of ARGs and also how

human activities can impact by providing ARGs that evolved under anthropogenic activity pressure, for example, agricultural activity or urban wastewater.

Using the deepARG tool [16] , we found 594 putative ARGs in 28 ARGs classes, in the most comprehensive study so far. From the total of ORFs found in the 12 co-assembled metagenomes from the TARA Oceans expedition [15], 0.28% were classified as an ARG gene. However, due to miss-annotations and miss-classifications, it was necessary to conduct an extensive manual curation on the results. After curation, we completely excluded 31 ARGs (so-called scenario i). For the quantitative analyses, we kept only 313 ARGs (26 classes), the ones passing all the criterias of scenario i-to-v. This represents an important resource for further studies, including evolutionary and comparative studies.

By using contig sequences assembled from multiple samples, we assume the risk of analysing chimeras, where contigs can be artificial assembled between different strains or even species, but we are able to evaluate genetic context of the studied genes, and also the possible presence of multi-resistance, by finding multiple ARGs in the same contig. In fact, it was possible to find several contigs containing 2 or more ARGs, with extreme cases of 2 chromosomal contigs with 11 putative ARGs (figure 2). The presence of multi-resistant bacteria in aquatic samples was not a surprise, and it was already described by several studies in seawater and sediment [30,31] and even in drinking water [32].

In addition, we used PlasFlow [23] to infer if the putative environmental ARGs are in a plasmid or in a chromosomal region. We found 24,567 ARG sequences in plasmids, including 2 plasmidial contigs containing up to 5 ARGs. We also used Kaiju [22] for taxonomic classification of ARGs and we found 149 sequences classified as viruses. The main objective of the latter analyses was to evaluate the potential of horizontal genetic transfer (HGT), as plasmids, bacteriophages, and extracellular DNA are the three primary drivers of HGT. The occurrence of HGT of ARGs was already detected and characterized in clinical environments [33], in wastewater treatment plants (activated sludge) [34,35] and in fertilized soil [36], but still little is known about aquatic environments, especially in open ocean regions. The presence of ARGs in phages and its potential HGT was described in many studies, for example, in a mediterranean river [37], in pig feces samples [38], in fresh-cut vegetables and in agricultural soil [39].

The environmental ARGs sequences found in this study were used as reference for mapping reads from 293 TARA oceans metagenomic samples, in order to quantify each of them around the world. The samples used in the original assembly (93) were included in the samples used in this step, but with the addition of new samples published afterwards.

In order to investigate if environmental features can influence ARG classes abundances, we conducted statistical tests (pairwise tukey HSD and OLS), showing that for 3 ARG classes there are significant higher abundance on Coastal Biomes than some other biome (Figure 3). For quinolone class, the Coastal biomes are the highest compared to all the other 3 biomes. High abundance of ARGs of quinolone

class along other ARG classes was previously reported on China coastal environment [40].

These results could indicate that specifically this class of ARGs are under anthropogenic pressure, and future studies should be carried to investigate it in greater details.

For Polymyxin class, we found a surprising result, showing Polar Biomes as significantly more abundance. Polymyxin B and E (also known as Colistins) are last resort antibiotics used against gram-negative bacteria only when modern antibiotics are ineffective, especially in cases of multiple drug-resistant *Pseudomonas aeruginosa* or carbapenemase-producing Enterobacteriaceae [41,42]. However, some bacteria, such as *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Acinetobacter baumannii*, develop resistance to polymyxins in a process referred to as acquired resistance, whereas other bacteria, such as *Proteus spp*., *Serratia spp*., and *Burkholderia spp*., are naturally resistant to this drug class [43].

Most mechanisms conferring resistance to Colistin are directed against modifications of the lipid A moiety of lipopolysaccharide (LPS), being the addition of l-ara4N and/or phosphoethanolamine (PEtN) to lipid A is one of the main mechanisms [44].

We found ARGs responsible for both resistance mechanisms: for l-ara4N addition we found genes from arnBCADTEF operon (arnA, arnC and arnD), pmrA, pmrB, PBGP/pmrF/yfbF; and for PEtN addition (pmrC/eptA and the recent discovery of mobilized colistin resistance - mrc-1).

The efflux pump gene rosB involved in polymyxin resistance was also found. Until the discovery of the plasmid-borne mrc-1 in *E. coli* from pigs and meat in China [45], colistin resistance has always been linked to a chromosomal-related mechanism with few or no possibility of horizontal transfer. However, further studies showed high prevalence of the mcr-1 gene (for example 20% in animal strains and 1% in human strains in China) and the plasmid has been detected in several countries covering Europe, Asia, South America, North America and Africa [46–53].

In the present study, we found 15 ORFs classified as mcr-1 by deepARG tool, but only 2 were in contigs classified as plasmid by PlasFlow. MCR-1 enzyme was described as 41 % and 40 % identical to the PEA transferases LptA and EptC, respectively, and sequence comparisons suggest the active-site residues are conserved [45]. In our phylogenetic analysis, we included sequences of clinical mcr-1 and LptA (eptA), in order to understand if the environmental sequences are closer to MCR-1 or to the chromosomal PEA transferases. Our results suggested that 5 ORFs are very close to MCR-1 (from genus Psychrobacter, family Moraxellaceae [54]) with support value 0.986 (Figure 6). Those sequences were not classified by PlasFlow as being in a plasmid, what can maybe explained by either a false negative results of PlasFlow, a re-integration of plasmidial sequences in chromosome or they may constitute some of the ancestral of the plasmidial modern *E. coli* MCR-1 sequences, as was already suggested [55].

Members of the genus Psychrobacter were isolated from a wide range of habitats, including food, clinical samples, skin, gills and intestines of fish, sea water, Antarctic sea ice [56–60] and at least 2 isolates from this genus were already reported to be

resistant to Colistin (Psychrobacter vallis sp. nov. and Psychrobacter aquaticus sp. nov) both isolated from Antarctica [61]. Coincidently, the regions with greater RPKG mean for MCR-1 in our study were Southwest Atlantic and Antarctic Province.

Our results support that Psychrobacter can be an ecological reservoir for mobilization of PEtN transferases to other pathogens and further studies should be conducted to better understand the dynamics and evolution of these genes. The presence of these genes in both Antarctic and adjacent regions can also raise concerns about gene flow due to ice melting, a problem already discussed before for other ARGs [62].

In addition, some species of this genus were also reported to cause opportunistic infections in humans, including at least one case reported to be associated with marine environment exposure [63], being important to increase vigilance (for example including screenings for those genes). In addition, all the 15 ORFs are clustered in the same big clade with the clinical MCR-1 sequences (support value 0.954) (Figure 6).

Conclusion

Our study uncovers the diversity and abundance of ARGs conferring resistance to 26 classes of antibiotics, conducting extensive analysis, including taxonomic classification, abundance in different biomes, potential HGT, multi-resistance, and gene expression. It also exposes the importance of monitoring coastal water for anthropogenic impact as rejection of antibiotics as wastewater and other type of human impact can generate a playground for microorganisms to freely develop antibiotic resistance. This study showed that Antarctic soil and ice are a potential reservoir for ARGs, and one can discuss the future impact of this content of this reservoir on global oceans due to climate change. The code and documentation is publicly available, as all the data generated, and stored in a database. This constitutes a good resource for the scientific community for future studies on antibiotic resistance in different environments.

Acknowledgements

We thank Jorge Boucas and the Bioinformatics Core facility of Max Planck Institute of Biology of Ageing, for the use of the computational resources (HPC cluster) and the fruitful discussions in the initial analysis of this work.

References

1. Ventola CL. The Antibiotic Resistance Crisis. Pharm Ther. 2015;40: 277–283. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4378521/
2. Aslam B, Wang W, Arshad MI, Khurshid M, Muzammil S, Rasool MH, et al.

Antibiotic resistance: a rundown of a global crisis. Infect Drug Resist. 2018;11: 1645–1658. doi:10.2147/IDR.S173867

3. Tagliabue A, Rappuoli R. Changing Priorities in Vaccinology: Antibiotic Resistance Moving to the Top. Front Immunol. 2018;9. doi:10.3389/fimmu.2018.01068
4. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. Hyperstability and Substrate Promiscuity in Laboratory Resurrections of Precambrian β-Lactamases. J Am Chem Soc. 2013;135: 2899–2902. doi:10.1021/ja311630a
5. Hall BG, Barlow M. Evolution of the serine β-lactamases: past, present and future. Drug Resist Updat. 2004;7: 111–123. doi:10.1016/j.drup.2004.02.003
6. Wright GD. The antibiotic resistome: the nexus of chemical and genetic diversity. Nat Rev Microbiol. 2007;5: 175–186. doi:10.1038/nrmicro1614
7. Aminov RI. The role of antibiotics and antibiotic resistance in nature. Environ Microbiol. 2009;11: 2970–2988. doi:10.1111/j.1462-2920.2009.01972.x
8. Singh PK, Bourque G, Craig NL, Dubnau JT, Feschotte C, Flasch DA, et al. Mobile genetic elements and genome evolution 2014. Mob DNA. 2014;5: 26. doi:10.1186/1759-8753-5-26
9. Bennett PM. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. Br J Pharmacol. 2008;153: S347–S357. doi:10.1038/sj.bjp.0707607
10. Zhang Q-Q, Tian G-M, Jin R-C. The occurrence, maintenance, and proliferation of antibiotic resistance genes (ARGs) in the environment: influencing factors, mechanisms, and elimination strategies. Appl Microbiol Biotechnol. 2018;102: 8261–8274. doi:10.1007/s00253-018-9235-7
11. Fróes AM, da Mota FF, Cuadrat RRC, Dávila AMR. Distribution and Classification of Serine β-Lactamases in Brazilian Hospital Sewage and Other Environmental Metagenomes Deposited in Public Databases. Front Microbiol. 2016;7: 1790. doi:10.3389/fmicb.2016.01790
12. Zhang K, Niu Z-G, Lv Z, Zhang Y. Occurrence and distribution of antibiotic resistance genes in water supply reservoirs in Jingjinji area, China. Ecotoxicol Lond Engl. 2017;26: 1284–1292. doi:10.1007/s10646-017-1853-9
13. Van Goethem MW, Pierneef R, Bezuidt OKI, Van De Peer Y, Cowan DA, Makhalanyane TP. A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. Microbiome. 2018;6. doi:10.1186/s40168-018-0424-5
14. Hayward JL, Jackson AJ, Yost CK, Truelstrup Hansen L, Jamieson RC. Fate of antibiotic resistance genes in two Arctic tundra wetlands impacted by municipal wastewater. Sci Total Environ. 2018;642: 1415–1428. doi:10.1016/j.scitotenv.2018.06.083
15. Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, Gorsky G, et al. Open science resources for the discovery and analysis of *Tara* Oceans data. Sci Data. 2015;2: 150023. doi:10.1038/sdata.2015.23
16. Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome. 2018;6: 23. doi:10.1186/s40168-018-0401-z
17. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 2017;45: D566–D573.

doi:10.1093/nar/gkw1004
18. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data. 2018;5: 170203. doi:10.1038/sdata.2017.203
19. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nat Microbiol. 2018;3: 804. doi:10.1038/s41564-018-0176-9
20. Companion Tables Ocean Microbiome EMBL [Internet]. Available: http://ocean-microbiome.embl.de/data/OM.CompanionTables.xlsx
21. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. 2010;38: e132–e132. doi:10.1093/nar/gkq275
22. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7: 11257. doi:10.1038/ncomms11257
23. Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. Nucleic Acids Res. 2018;46: e35–e35. doi:10.1093/nar/gkx1321
24. BLAST [Internet]. Available: https://blast.ncbi.nlm.nih.gov/
25. Nayfach S, Pollard KS. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. Genome Biol. 2015;16: 51. doi:10.1186/s13059-015-0611-7
26. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008;36: W465–W469. doi:10.1093/nar/gkn180
27. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32: 1792–1797. doi:10.1093/nar/gkh340
28. Castresana J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol Biol Evol. 2000;17: 540–552. doi:10.1093/oxfordjournals.molbev.a026334
29. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol. 2010;59: 307–321. doi:10.1093/sysbio/syq010
30. Goodwin KD, McNay M, Cao Y, Ebentier D, Madison M, Griffith JF. A multi-beach study of Staphylococcus aureus, MRSA, and enterococci in seawater and beach sand. Water Res. 2012;46: 4195–4207. doi:10.1016/j.watres.2012.04.001
31. Neela FA, Nonaka L, Suzuki S. The diversity of multi-drug resistance profiles in tetracycline-resistant Vibrio species isolated from coastal sediments and seawater. J Microbiol. 2007;45: 64–68.
32. Armstrong JL, Shigeno DS, Calomiris JJ, Seidler RJ. Antibiotic-resistant bacteria in drinking water. Appl Environ Microbiol. 1981;42: 277–283. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC244002/
33. Lerminiaux NA, Cameron ADS. Horizontal transfer of antibiotic resistance genes in clinical environments. Can J Microbiol. 2019;65: 34–44. doi:10.1139/cjm-2018-0275
34. Zhang T, Zhang X-X, Ye L. Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge.

PloS One. 2011;6: e26041. doi:10.1371/journal.pone.0026041

35. Qiu Y, Zhang J, Li B, Wen X, Liang P, Huang X. A novel microfluidic system enables visualization and analysis of antibiotic resistance gene transfer to activated sludge bacteria in biofilm. Sci Total Environ. 2018;642: 582–590. doi:10.1016/j.scitotenv.2018.06.012
36. Peng S, Dolfing J, Feng Y, Wang Y, Lin X. Enrichment of the Antibiotic Resistance Gene tet(L) in an Alkaline Soil Fertilized With Plant Derived Organic Manure. Front Microbiol. 2018;9: 1140. doi:10.3389/fmicb.2018.01140
37. Calero-Cáceres W, Méndez J, Martín-Díaz J, Muniesa M. The occurrence of antibiotic resistance genes in a Mediterranean river and their persistence in the riverbed sediment. Environ Pollut Barking Essex 1987. 2017;223: 384–394. doi:10.1016/j.envpol.2017.01.035
38. Wang M, Liu P, Zhou Q, Tao W, Sun Y, Zeng Z. Estimating the contribution of bacteriophage to the dissemination of antibiotic resistance genes in pig feces. Environ Pollut Barking Essex 1987. 2018;238: 291–298. doi:10.1016/j.envpol.2018.03.024
39. Larrañaga O, Brown-Jaque M, Quirós P, Gómez-Gómez C, Blanch AR, Rodríguez-Rubio L, et al. Phage particles harboring antibiotic resistance genes in fresh-cut vegetables and agricultural soil. Environ Int. 2018;115: 133–141. doi:10.1016/j.envint.2018.03.019
40. Lu J, Zhang Y, Wu J, Wang J, Zhang C, Lin Y. Occurrence and spatial distribution of antibiotic resistance genes in the Bohai Sea and Yellow Sea areas, China. Environ Pollut. 2019;252: 450–460. doi:10.1016/j.envpol.2019.05.143
41. Velkov T, Roberts KD, Nation RL, Thompson PE, Li J. Pharmacology of polymyxins: new insights into an 'old' class of antibiotics. Future Microbiol. 2013;8: 711–724. doi:10.2217/fmb.13.39
42. Falagas ME, Kasiakou SK. Toxicity of polymyxins: a systematic review of the evidence from old and recent studies. Crit Care. 2006;10: R27. doi:10.1186/cc3995
43. Olaitan AO, Morand S, Rolain J-M. Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. Front Microbiol. 2014;5. doi:10.3389/fmicb.2014.00643
44. Baron S, Hadjadj L, Rolain J-M, Olaitan AO. Molecular mechanisms of polymyxin resistance: knowns and unknowns. Int J Antimicrob Agents. 2016;48: 583–591. doi:10.1016/j.ijantimicag.2016.06.023
45. Liu Y-Y, Wang Y, Walsh TR, Yi L-X, Zhang R, Spencer J, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis. 2016;16: 161–168. doi:10.1016/S1473-3099(15)00424-7
46. Hasman H, Hammerum AM, Hansen F, Hendriksen RS, Olesen B, Agersø Y, et al. Detection of mcr-1 encoding plasmid-mediated colistin-resistant Escherichia coli isolates from human bloodstream infection and imported chicken meat, Denmark 2015. Eurosurveillance Online Ed. 2015;20: 1–5. doi:10.2807/1560-7917.es.2015.20.49.30085
47. Falgenhauer L, Waezsada S-E, Yao Y, Imirzalioglu C, Käsbohrer A, Roesler U, et al. Colistin resistance gene mcr-1 in extended-spectrum β-lactamase-producing and carbapenemase-producing Gram-negative bacteria in Germany. Lancet Infect Dis. 2016;16: 282–283. doi:10.1016/S1473-3099(16)00009-8
48. Webb HE, Granier SA, Marault M, Millemann Y, Bakker HC den, Nightingale

KK, et al. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis. 2016;16: 144–145. doi:10.1016/S1473-3099(15)00538-1

49. Tse H, Yuen K-Y. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis. 2016;16: 145–146. doi:10.1016/S1473-3099(15)00532-0
50. Zhang R, Huang Y, Chan EW, Zhou H, Chen S. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis. 2016;16: 291–292. doi:10.1016/S1473-3099(16)00062-1
51. Mulvey MR, Mataseje LF, Robertson J, Nash JHE, Boerlin P, Toye B, et al. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis. 2016;16: 289–290. doi:10.1016/S1473-3099(16)00067-0
52. Arcilla MS, Hattem JM van, Matamoros S, Melles DC, Penders J, Jong MD de, et al. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis. 2016;16: 147–149. doi:10.1016/S1473-3099(15)00541-1
53. Malhotra-Kumar S, Xavier BB, Das AJ, Lammens C, Butaye P, Goossens H. Colistin resistance gene mcr-1 harboured on a multidrug resistant plasmid. Lancet Infect Dis. 2016;16: 283–284. doi:10.1016/S1473-3099(16)00012-8
54. Wei W, Srinivas S, Lin J, Tang Z, Wang S, Ullah S, et al. Defining ICR-Mo, an intrinsic colistin resistance determinant from Moraxella osloensis. PLOS Genet. 2018;14: e1007389. doi:10.1371/journal.pgen.1007389
55. Xavier BB, Lammens C, Ruhal R, Kumar-Singh S, Butaye P, Goossens H, et al. Identification of a novel plasmid-mediated colistin-resistance gene, mcr-2, in Escherichia coli, Belgium, June 2016. EuroSurveillance Mon. 2016;21: 30280.
56. Maruyama A, Honda D, Yamamoto H, Kitamura K, Higashihara T. Phylogenetic analysis of psychrophilic bacteria isolated from the Japan Trench, including a description of the deep-sea species Psychrobacter pacificensis sp. nov. Int J Syst Evol Microbiol. 2000;50: 835–846.
57. Bowman JP, Nichols DS, McMeekin TA. Psychrobacter glacincola sp. nov., a halotolerant, psychrophilic bacterium isolated from Antarctic sea ice. Syst Appl Microbiol. 1997;20: 209–215.
58. BOWMAN JP, CAVANAGH J, AUSTIN JJ, SANDERSON K. Novel Psychrobacter Species from Antarctic Ornithogenic Soils. Int J Syst Evol Microbiol. 1996;46: 841–848. doi:10.1099/00207713-46-4-841
59. JUNI E, HEYM GA. Psychrobacter immobilis gen. nov., sp. nov.: Genospecies Composed of Gram-Negative, Aerobic, Oxidase-Positive Coccobacilli. Int J Syst Evol Microbiol. 1986;36: 388–391. doi:10.1099/00207713-36-3-388
60. Yumoto I, Hirota K, Sogabe Y, Nodasaka Y, Yokota Y, Hoshino T. Psychrobacter okhotskensis sp. nov., a lipase-producing facultative psychrophile isolated from the coast of the Okhotsk Sea. Int J Syst Evol Microbiol. 2003;53: 1985–1989. doi:10.1099/ijs.0.02686-0
61. Shivaji S, Reddy GSN, Suresh K, Gupta P, Chintalapati S, Schumann P, et al. Psychrobacter vallis sp. nov. and Psychrobacter aquaticus sp. nov., from Antarctica. Int J Syst Evol Microbiol. 2005;55: 757–762. doi:10.1099/ijs.0.03030-0
62. Edwards A. Coming in from the cold: potential microbial threats from the terrestrial cryosphere. Front Earth Sci. 2015;3. doi:10.3389/feart.2015.00012
63. Bonwitt J, Tran M, Droz A, Gonzalez A, Glover WA. Psychrobacter sanguinis Wound Infection Associated with Marine Environment Exposure, Washington, USA - Volume 24, Number 10—October 2018 - Emerging Infectious Diseases journal - CDC. doi:10.3201/eid2410.171821