RESEARCH

Global ocean resistome revealed: Exploring antibiotic resistance gene abundance and distribution in TARA Oceans samples

Rafael R. C. Cuadrat [ID][1], Maria Sorokina [ID][2], Bruno G. Andrade[3], Tobias Goris[4] and Alberto M. R. Dávila [ID][5,6,*]

[1]Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke - DIfE, Arthur-Scheunert-Allee 114–116, 14558 Nuthetal, Germany; [2]Institute for Inorganic and Analytical Chemistry, Friedrich-Schiller University, Lessingstrasse 8, 07743 Jena, Germany; [3]Animal Biotechnology Laboratory, Embrapa Southeast Livestock, EMBRAPA, Rodovia Washington Luiz, Km 234 s/n°, 13560-970 São Carlos, SP, Brazil; [4]Department of Molecular Toxicology, Research Group Intestinal Microbiology, German Institute of Human Nutrition Potsdam-Rehbruecke - DIfE, Arthur-Scheunert-Allee 114–116, 14558 Nuthetal, Germany; [5]Computational and Systems Biology Laboratory, Oswaldo Cruz Institute, FIOCRUZ, Av Brasil 4365, 21040-900 Rio de Janeiro, RJ, Brazil and [6]Graduate Program in Biodiversity and Health, Oswaldo Cruz Institute, FIOCRUZ, Av. Brasil 4365, 21040-900 Rio de Janeiro, RJ, Brazil

*Correspondence address. Alberto M. R. Dávila, Computational and Systems Biology Laboratory, Oswaldo Cruz Institute, FIOCRUZ, Av Brasil 4365, 21040-900 Rio de Janeiro, RJ, Brazil. E-mail: alberto.davila@fiocruz.br [ID] http://orcid.org/0000-0002-6918-7673

Abstract

Background: The rise of antibiotic resistance (AR) in clinical settings is of great concern. Therefore, the understanding of AR mechanisms, evolution, and global distribution is a priority for patient survival. Despite all efforts in the elucidation of AR mechanisms in clinical strains, little is known about its prevalence and evolution in environmental microorganisms. We used 293 metagenomic samples from the TARA Oceans project to detect and quantify environmental antibiotic resistance genes (ARGs) using machine learning tools. **Results:** After manual curation of ARGs, their abundance and distribution in the global ocean are presented. Additionally, the potential of horizontal ARG transfer by plasmids and their correlation with environmental and geographical parameters is shown. A total of 99,205 environmental open reading frames (ORFs) were classified as 1 of 560 different ARGs conferring resistance to 26 antibiotic classes. We found 24,567 ORFs in putative plasmid sequences, suggesting the importance of mobile genetic elements in the dynamics of environmental ARG transmission. Moreover, 4,804 contigs with >=2 putative ARGs were found, including 2 plasmid-like contigs with 5 different ARGs, highlighting the potential presence of multi-resistant microorganisms in the natural ocean environment. Finally, we identified ARGs conferring resistance to some of the most relevant clinical antibiotics, revealing the presence of 15 ARGs similar to mobilized colistin resistance genes (*mcr*) with high abundance on polar biomes. Of these, 5 are assigned to *Psychrobacter*, a genus including opportunistic human pathogens. **Conclusions:** This study uncovers the diversity and abundance of ARGs in the global ocean metagenome. Our results are available on Zenodo in MySQL database dump format,

and all the code used for the analyses, including a Jupyter notebook js avaliable on Github. We also developed a dashboard web application (http://www.resistomedb.com) for data visualization.

Keywords: β-lactamase; machine learning; marine metagenomics; colistin; tetracycline; multidrug resistance

Introduction

Antibiotic-resistant bacteria are a global public health issue and an economic burden to the entire world, especially in developing countries. Projections have shown that, if the emergence of multi-resistant bacteria continues at the same rate, they will cause 10 million deaths per year, which would outnumber cancer-related deaths [1, 2]. Despite its impact on human health, antibiotic resistance (AR) is a natural phenomenon and one of the most common bacterial defense mechanisms. For example, the resistance to β-lactam antibiotics, conferred by β-lactamase activity, is estimated to have emerged >1 billion years ago [3, 4]. Some authors argue that β-lactamase genes are part of inter- and intra-community communication and used in the defense repertoires of organisms sharing the same biological niche [5, 6].

The collection of antibiotic resistance genes (ARGs) in a given environment or organism is known as the resistome, and such genes have been detected in different natural environments, such as oceans [7], lakes [8], rivers [9], remote pristine Antarctic soils [10], and impacted Arctic tundra wetlands [11]. Studies also showed that anthropogenic activity (e.g., overuse of antibiotics and their subsequent release via wastewater into the environment) could lead to the spread of clinically relevant ARGs across natural environments [12, 13]. Therefore, the investigation of the natural context of ARGs, their geographic distribution, dynamics, and, in particular, their presence on horizontally transferable mobile genetic elements, such as plasmids, transposons, and phages, is crucial to assess their potential to emerge and spread [14–16]. Owing to modern advances in DNA sequencing and bioinformatics, it is now possible to study the presence and prevalence of ARGs in different environments. However, most of the published studies targeted only 1 or a few classes of ARGs and were limited to specific environments and geographic locations. The oceans cover ∼70% of Earth's surface, harbouring a significant diversity of planktonic microorganisms, forming a complex ecological network that is still understudied [17, 18]. To tackle this problem, the number of ocean metagenomic projects stored in public databases has been growing. Again, the lack of related metadata have made it challenging to conduct high-throughput gene screenings and correlations with environmental factors. Fortunately, the TARA Oceans project [19] measured several marine environmental conditions across the globe and stored them as structured metadata. This rich and unique dataset, together with the metagenome sequences [19], will allow the use of machine and deep learning approaches to search for gene and species distributions and their correlation to environmental parameters. In this study, we applied deepARG [20], a deep learning approach for ARG identification, to screen co-assembled TARA Oceans contigs [21]. After the manual curation of ARGs, we classified the results of the deepARG screening taxonomically. Furthermore, ARG abundance was quantified, and ordinary least squares (OLS) regression with association analyses between the quantification of ARGs and environmental parameters was used. We also explored the presence of ARGs located on putative plasmids to investigate the potential of these oceanic environments to act as a reservoir of potentially mobile ARGs.

Methods

Metagenomic data

A total of 12 co-assembled metagenomes from different oceanic regions explored by the TARA Oceans expedition, with contigs larger than 1 kb, were obtained from the dataset published in 2017 by Delmont et al. [22]. Raw reads of 243 samples (378 sequencing runs; accession numbers PRJEB1787, PRJEB6606, and PRJEB4419) were obtained from the European Bioinformatics Institute (EBI) European Nucleotide Archive (ENA) database [23].

Sample identifiers and metadata were obtained from the TARA Oceans companion website tables [24]. Samples were collected at different sites and depths and successively filtered using a single or a combination of membranes with pore sizes of 0.1, 0.2, 0.45, 0.8, 1.6, and 3 μm to retain different size fractions (i.e., viruses, giant viruses, and prokaryotes) [24]. We created a variable called "fraction," where the upper and lower filtration membrane size were used together to define groups. However, owing to methodological limitations (described in the Results and Methods sections), viruses and giant viruses (giruses) enriched samples were excluded from quantitative analysis.

Environmental ARG prediction

Open reading frame (ORF) prediction was performed on the 12 co-assembled metagenomes using MetaGeneMark v3.26 [25] with default parameters (sequences larger than 60 nucleotides). The screening for ARGs was performed with DeepARG [20] on the predicted ORFs using gene models. The deepARG tool was developed, taking into account a dissimilarity matrix using all ARG categories of 3 curated and merged databases (Antibiotic Resistance Genes Database [ARDB], Comprehensive Antibiotic Resistance Database [CARD], and UniProt) [20]. This approach is an alternative to the "best hits" of sequence searches against existing databases, which produces a high rate of false-negative results [20]. An ORF was classified as ARG if the estimated probability was ≥0.8. Contigs containing ≥1 putative ARG were analysed with PlasFlow 1.1 [26] using a probability threshold of 0.7 to check for a potential plasmidial location of ARGs. We also investigated the number and distribution of contigs with 2 or more putative ARGs to check for multiple resistance and/or whole ARG operons from environmental samples. Putative ARGs (and their respective contig) were submitted to Kaiju v1.6.2 [27] for taxonomic classification, with the option "run mode" set as "greedy." Later, we conducted a manual curation of each ARG to check for misannotations and inconsistencies. BLASTp searches [28] were performed against the non-redundant (NR) protein database, with default parameters. Results with an e-value lower than e^{-5} were considered. Conserved domains (CDDs) and annotations in the source databases (ARDB [29], CARD [30], and UniProt [31]) were manually inspected. These results were used to classify misannotated/misclassified ARGs into different categories: (i) misannotated genes or gene families with low support for ARG prediction, i.e., all source database sequences exhibiting non-ARGs as top 5 BLASTp (against NR database) hits with an e-value cut-off of e^{-5}. Included are especially cases with an unambiguously erroneous original annotation (examples are de-

scribed in the Results). All of these misannotated ARGs were removed from our database and the downstream analyses; (ii) housekeeping genes that confer resistance only when specifically mutated; (iii) housekeeping genes conferring resistance when overexpressed; (iv) regulatory sequences responsible for ARG activation or overexpression of housekeeping genes leading to a resistance phenotype. The ARG family descriptions of the source databases (mainly those of the CARD database) were used (in addition to literature information) to classify ARGs into this scenario; (v) sequences with both similarities to ARGs and non-ARGs, belonging to the same superfamily and/or sharing domains. BLASTp and CDD analysis were used to classify ARGs into this scenario in cases where the TARA sequences show non-ARGs and no specific CDD domain for that ARG among the top 10 BLASTp hits.

ARG quantification and statistical tests on metagenomic samples

Environmental ARGs identified were used as a reference for raw read mapping by BBMAP v37.90 (default parameters) [32] after manual curation. The coverage, in terms of read count per gene, and the abundance, in terms of fragments per kilobase per million mapped reads (FPKM), of each ARG was then calculated for each sample by BBMAP. The average genome size (AGS) and genome equivalents (GEs) were estimated by the software MicrobeCensus v1.0.7 (default parameters) [33] to calculate reads per kilobase per genome equivalents (RPKG) as described [33]. The RPKG of an ARG in a metagenome was calculated by (i) counting the number of reads mapped to the ARG, (ii) dividing (i) by the length of the ARG in kilobase pairs, and (iii) dividing the result of (ii) by the number of sequenced genome equivalents:

$$RPKG = \frac{Mapped\,reads/Gene\,Length\,(kb)}{Genome\,equivalents},$$

where

$$GE = \frac{Library\,size\,(bp)}{AGS\,(bp)}$$

and library size is the total number of sequenced base pairs.

RPKG values for all ORFs classified as the same ARG were summed for each sample. Environmental features, such as sample depth, biogeographic biomes, ocean and sea regions, and fractions, were used for sample grouping and statistical tests. Pairwise Tukey HSD and multivariate linear regression using OLS models were conducted in Python 3.6 using the library "statsmodels." The OLS was performed considering the following formula:

$$ARG_{RPKG} \sim fraction + Latitude + Longitude + depth$$
$$+temp\_c + NO_2NO_3 + PO_4 + SI + Mean\_Oxygen$$
$$+Mean\_Salinity + OG\_Shannon,$$

where ARG_{RPKG} (the dependent variable) is the sum of RPKM of all ARGs in a given class, and all the dependent variables are the selected environmental features. A 2-way ANOVA analysis was conducted on the coefficients obtained from the OLS regression to infer the significance of a feature. A Python Jupyter note-

book with the code and the results for all the exploratory and statistical analyses is provided on GitHub [34].

Phylogenetic analysis of environmental ARGs

Phylogenetic analyses were performed on environmental nucleotide sequences identified as clinically relevant ARGs, such as mobilized colistin resistance (MCR)-related sequences, for which reference sequences were retrieved from public databases (e.g., NCBI and deepARGdb). Multiple protein sequence alignments and phylogenetic trees were generated using the standard pipeline of Phylogeny.fr [35]. In short, sequences were aligned using MUSCLE (default parameters) [36], conserved blocks extracted with gblocks (default parameters) [37], and phylogenetic trees generated with phyML [38], using Whelan and Goldman (WAG) matrix substitution model and approximate likelihood-ratio test (ALRT) statistical test.

Database design and implementation

A manually curated MySQL database was created with the environmental ARGs described and all the subsequent analysis results. Data downloaded and processed as described above were parsed with Java 8 and stored in the database with Hibernate. The database model is also managed by Hibernate in Java. The code is available on GitHub [39]. The resulting database contains 5 main data tables: "orf", "arg", "sample", "organism", and "xref", containing cross-references between the different data sources. The additional 5 connection tables map in an SQL engine-free way the correspondences between the items from different tables. We provide the SQL dump and the database schema at Zenodo [40].

Dash web application for data exploration and visualization

We developed a Python dashboard web application where the user can explore the results through interactive graphics (plotted with the plotly library). The application includes a geographical scatterplot, where it is possible to visualize the abundance of each ARG (or antibiotic class) selected by the user across all the samples in a world map; a boxplot, where environmental features can be chosen to group the samples and compare their abundances; a barplot with taxonomic classification of the selected ARG (different taxonomic levels for the visualization can be chosen); and a scatterplot with marginal distribution plots and trend line (OLS), where the X-axis represents the selected ARG, and the Y-axis, the environmental variables selected by the user (e.g., oxygen concentration, salinity, temperature, depth). In addition, a table containing information for each ORF is displayed. The additional information includes ORF ID, contig ID, antibiotic class, deepARG probability value, plasmid classification by PlasFlow, taxonomic classification by Kaiju (on the deepest level), the abundance of additional ARG ORFs in the same contig, and the total of ARG ORFs in the contig. A link to download the multi-fasta file of the selected ARG is also provided. The application can be accessed at [41]. The code and data for the dash app can be accessed at [42].

Pipeline and code availability

The code of the complete pipeline (Fig. 1) is in Bash and Python and is available at the project repository on GitHub [43].
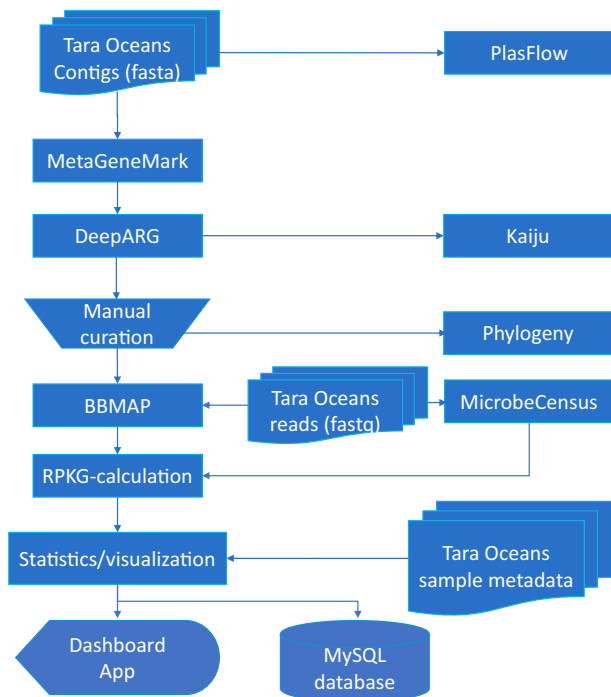
Figure 1: Flow chart used for ARG classification. The single steps and data used in the pipeline applied for the analyses presented in this work.

Results and Discussion

Environmental ARG prediction and manual curation

A total of 41,249,791 ORFs were predicted from 15,600,278 assembled contigs by MetaGeneMark. These ORFs were used as input for ARG screening with the deepARG software [20], resulting in the classification of 116,425 TARA ORFs (0.28%) as putative ARGs, related to 594 clinically relevant ARGs that confer resistance to 28 antibiotic classes (classes defined in the deepARGdb). The number of contigs, ORFs, and putative ARGs from each oceanic region is available in Supplementary Table 1. It was necessary to conduct an extensive manual curation on the results owing to misannotations and misclassifications of ARGs in the databases used by deepARG. This curated dataset represents an important resource for further studies, including evolutionary and comparative studies.

A total of 34 ARGs were identified as misannotated or with low-quality annotation in the source database, leaving 560 ARGs for further analyses. A prominent example of a misannotated ARG is the *msr*B gene: while the *msr*B classified as ARG encodes an ABC-F subfamily protein leading to erythromycin and streptogramin B resistance, the corresponding fasta sequence in the CARD database [30] belongs to the *msr*B gene encoding methionine sulfoxide reductases B, not conferring AR. Another misannotated ARG is the *pat*A gene, an ABC transporter of *Streptococcus pneumoniae*, conferring resistance to fluoroquinolones, whose sequence is a putrescine aminotransferase (*pat*A) in the CARD database. A total of 99,205 ORFs identified as putative ARGs in categories (ii), (iii), (iv), and (v) (see Methods parts) were kept in the MySQL database for further studies, while they were not used in the quantification and statistical analyses. Category (ii) includes the identification of 10 families of housekeeping genes and the corresponding mutations that could infer resistance. Category (iii) included 9 ARGs whose overexpression can

Table 1: Distribution of multiple ARGs in chromosome and plasmids (classified by PlasFlow)

No. of ARGs	In chromosome	In plasmid
2	3,503	689
3	365	37
4	116	13
5	35	2
6	22	0
7	10	0
8	6	0
9	2	0
10	2	0
11	2	0

lead to resistance. For category (iv), we identified 41 regulatory sequences that have been identified as responsible for ARG expression or overexpression of housekeeping genes, causing the resistance phenotype. Category (v) included 187 putative ARGs that cannot be distinguished from non-ARGs by similarity alone (mostly due to commonly shared domains, e.g., ATPases). After the removal of those genes, a total of 13,163 ORFs (from the initial 116,425) classified as 313 ARGs were retained for quantification and further analysis (Supplementary Table 2).

The most frequent ARG (in number of ORFs) identified in the co-assembly dataset was Qac (multidrug efflux pumps named after their conferring resistance to quaternary ammonium compounds) with >2,500 overall occurrences, followed by TETB(60) (Fig. 2). The latter is an ABC transporter that confers resistance to tetracycline and tigecycline identified in a human saliva metagenomic library [44]. The ORFs conferring resistance to tetracycline combined are the most widespread, with several Tet and TetA classes accounting for ~4,000 occurrences. Also, the most frequent ARG that confers resistance to β-lactams was identified as K678_12262, with ~1,000 occurrences.

Environmental ARGs in chromosomes and plasmids

We found a total of 24,567 putative ARGs (24.76% of the ORFs considered for the downstream analysis) present in contigs classified as plasmids by PlasFlow, which indicates the potential of horizontal genetic transfer (HGT). The occurrence of HGT of ARGs has already been described in clinical environments [45], wastewater treatment plants (activated sludge) [14, 46], and in fertilized soil [47], but little is known about ARG HGT in aquatic environments, especially in open ocean regions. As discussed in the section on *mcr* genes, it should be noted here that PlasFlow analyses bear a small chance (~4%) of false-positive results as described [26], which especially could be the case with chromosomally integrated plasmids or very short contig sequence sizes.

Multiple resistance presence in environmental contigs

The presence of 2 or more ARGs in a single contig was analysed to identify possible multi-resistant organisms. For this analysis, we only removed the ARGs from category (i) (misannotated sequences) because the presence of putative ARGs in the same contig and/or plasmid can give us additional functional evidence. We identified 4,063 contigs with multiple putative ARGs in contigs classified as chromosomes (up to 11 ARGs in the same contig), and 741 in contigs classified as a plasmid (up to 5 ARGs in the same contig), suggesting the presence of multi-resistant microorganisms in these environments (Table 1). We cannot ex-
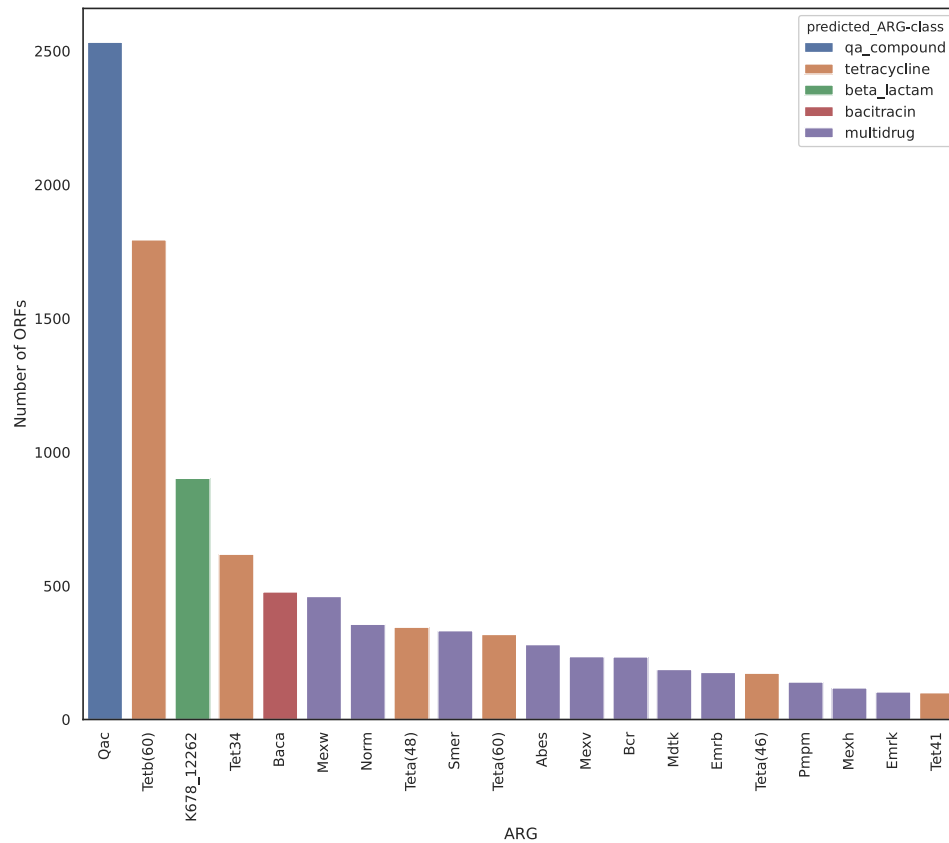
Figure 2: The 20 most frequent ARGs after manual curation (in number of ORFs on co-assembled contigs). The corresponding antibiotic resistance classes is depicted in the upper right.

clude the possibility of multiple ARGs in both ends of plasmidial contigs being, in fact, artefacts, such as pieces of the same ARG in a circular contig. From the 4,192 contigs with 2 ARGs, 74 showed the same annotation for both ARGs (33 classified as plasmid). In Fig. S1, we show the distribution of the ARGs in the 2 putative plasmids containing 5 ARGs each.

Taxonomic classification of environmental ARGs

We classified 97,244 ARGs (98.02%) up to ≥1 taxonomic level using Kaiju [27]. Alphaproteobacteria (37,360 sequences) was identified as the largest taxonomic unit, followed by Gammaproteobacteria (19,355 sequences). A total of 124 ARGs were classified as of viral origin. The most frequent taxonomic viral groups identified were Prymnesiovirus (21 ARGs) and *Chrysochromulina ericina* virus (19 ARGs). However, all 124 viral ARGs were classified into category (v), and further investigations should be performed to confirm these findings. The presence of ARGs in phages and their potential HGT has been described in a Mediterranean river [48], pig faecal samples [15], fresh-cut vegetables, and agricultural soil [16].

In the contig containing 11 ARGs (TARA_ANW-k99_1343221), 9 were classified as HGW-Alphaproteobacteria-3 or HGW-Alphaproteobacteria-12, and as generic Alphaproteobacteria. The 2 residual ARGs were classified as belonging to *Parvibaculum lavamentivorans*, an alphaproteobacterial species first isolated from activated sludge in Germany [49]. A previous study showed the presence of ARGs in a strain of *Parvibaculum* from marine samples by functional metagenomics [7], which might indicate a broader ARG distribution among this clade. All ARGs from the

other contig containing 11 ARGs (TARA_ANE-k99_4428305) were classified as *Micavibrio sp.*, an obligate predatory bacterium exhibiting "vampire-like" behavior on gram-negative pathogens [50]. First isolated from wastewater samples, this genus has been considered as a potential new therapeutic approach against multi-resistant bacteria [51], including *mcr-1* positive strains [51], because no species from the genus *Micavibrio* was found to be pathogenic for humans [50]. However, if *Micavibrio* species are confirmed to contain 1 or multiple ARGs, this would raise concerns about any clinical therapeutic approaches with these bacteria. One of the putative plasmids containing 5 ARGs (contig TARA_PSE-k99_4996023, Supplementary Fig. S1) showed a taxonomic agreement between the classification of all its ARGs, which were assigned to the species *Tistrella mobilis*. Strains of this species were isolated from Thailand wastewater [52] and the Red Sea [53]. The other contig containing 5 ARGs of plasmidial origin was classified as *Halomonas desiderata*, a denitrifying bacterium first isolated from a municipal sewage treatment plant [54]. Two of the putative 5 ARGs in this contig were classified as DfrE and DfrA3, which confer resistance to trimethoprim. Previous work showed that another bacterial species of the same genus (*Halomonas marisflavi* type strain) is resistant to trimethoprim in vitro [55]. However, in the same study, *H. desiderata* did not show resistance to any of the antibiotics tested.

ARG abundance and statistical tests on metagenomic samples

In previous sections, we aimed to find and characterize ARGs in metagenomic contigs obtained from co-assembled samples

(by oceanic regions). In this section, we aim to quantify ARGs in individual samples, to understand their geographical distribution and the environmental features driving their abundance. The AGS of samples of fractions enriched for virus and girus showed biased and aberrant results for AGS (up to 395.4 Mb). These results are because AGS values (calculated by Microbe-Census [33]) are inversely proportional to the number of reads mapping to housekeeping gene markers, and such genes have low abundance in virus-enriched samples. On the basis of this information, we kept only the 293 non–virus enriched sample runs for downstream quantitative analyses.

For example, comparing biogeographical biomes, the quinolone and bacitracin ARG classes were significantly more abundant in the coastal biome than in the westerlies biome (adjusted Tukey HSD P-values 0.0476 and 0.0027, respectively). Furthermore, fosmidomycin ARGs were significantly (adjusted Tukey HSD P-value 0.0011) more abundant in the coastal biome than in the trades biome (Fig. 3, Supplementary Table 3). Quinolone ARGs were previously reported as highly abundant in Chinese coastal areas [56]. These results might indicate that quinolone, bacitracin, and fosmidomycin ARGs are under anthropogenic pressure in coastal environments, and future studies should be carried out to investigate this assumption in greater detail.

The pristine polar biome showed significantly higher RPKG values for polymyxin ARGs than any other biome. The antibiotics polymyxin B and E (also known as colistin) are the last resorts against gram-negative bacterial infections when modern antibiotics are ineffective, especially in cases of multiple drug-resistant *Pseudomonas aeruginosa* or carbapenemase-producing Enterobacteriaceae [57, 58]. We discuss mobilized colistin resistance genes (*mcr*) in greater detail below.

When comparing the abundances of ARG classes in marine provinces, we found a significant difference (P-value < 0.05) of bleomycin class in 2 Indian provinces when compared to most of the other provinces (Fig. 4). Bleomycin resistance genes were previously reported to be in association with New Delhi metallo-β-lactamase (*ndm*-1) genes [59, 60]. In this study, *ndm*-like genes (classified by deepARG as *ndm*-17 variant) were also found in greater abundance in Indian South Subtropical Gyre province. The first variant of *ndm* was identified in a *Klebsiella pneumoniae* strain isolated from a Swedish patient who travelled to New Delhi, India [61]. Shortly after, it was spread globally in a few years and was also detected in other Enterobacteriaceae, which was a reason to classify NDMs as a potential worldwide public health problem [62].

In addition to the geographical location, we investigated the influence of other environmental parameters on the abundance of ARG classes. In our OLS models, the variables with significant P-values (<0.05 ANOVA test) for the largest number of antibiotic classes were "fraction" (14 classes), "sampling depth," and "Shannon-Wiener index" (11 classes each). The fraction is a categorical variable, and the smallest size fraction (0.22–0.45 μm) was used as a reference for computing the coefficients in the model. This fraction is enriched for free-living, non-aggregating bacteria, which are smaller than other size fractions. For most classes (11 of 14), ≥1 category of fraction showed positive coefficients. For 3 of them, all fractions showed significantly more ARGs than the smallest fraction (tetracycline, aminoglycoside, and fosmidomycin). This result may indicate that free-living bacteria, in general, have a lower abundance of ARGs than particle-associated bacteria. These results corroborate a previous study, in which the antagonistic activity among pelagic marine bacteria (i.e., production of antibiotics) was

more common in particle-associated bacteria than free-living bacteria [63].

For sampling depth, 5 of 11 classes were negatively correlated, indicating an increased abundance of ARGs in the deep water. For the Shannon-Wiener index, the only negative correlation was tetracycline, indicating an increased abundance of ARGs in samples with lower species richness.

The regression model for tetracycline presented the highest adjusted R^2 (0.666) of all classes, with fraction, temperature, and sampling depth the most significant variables. For polymyxin, the adjusted R^2 was the second highest (0.559), with temperature, Shannon index, and sampling depth the most significant variables.

In general, among the nutrients, nitrite+nitrate concentration (NO_2NO_3) was significant for the largest number of classes (7 classes), followed by inorganic phosphate $(PO_4)^{3-}$ (6 classes). Silicon (Si) was only significant for the classes fosmidomycin and tetracycline.

The role of inorganic nutrient concentration in ARG abundance is poorly understood and sometimes controversial. Some studies suggest that a high concentration of nutrients is negatively associated with ARGs because competitive interactions in nutrient-rich environments are less important [64]. However, the abundances of ARGs are increased in wastewater treatment plants [65] and agricultural soil receiving dairy manure [66], both environments rich in nutrients. Further studies should be conducted to better understand the role played by different nutrients in the abundance of ARGs of different classes in both pristine oligotrophic and impacted environments. Supplementary Table 4 reports all significant results of an ANOVA test on the coefficients of OLS for each class, and Supplementary Table 5 shows all the OLS results. A Q-Q plot of the OLS residuals is shown in Supplementary Fig. 2.

Mobilized colistin resistance genes (mcr) and other polymyxin resistance genes

Most mechanisms that confer resistance to colistin act against modifications of the lipid A moiety of lipopolysaccharide, with the addition of L-ara4N and/or phosphoethanolamine (PEA) to lipid A as the main mechanisms [67]. We found evidence for the occurrence of putative mobilized colistin resistance genes related to the recently discovered *mcr*-1 [68], which relies on the PEA addition to lipid A. The Mcr-1 enzyme was described as 41% and 40% identical to the PEA transferases LptA and EptC, respectively, and sequence comparisons suggest that the active-site residues are conserved. However, until the discovery of the plasmid-borne *mcr*-1 in *Escherichia coli* from pig [68], colistin resistance has always been linked to chromosomally encoded genes with low or no possibility of horizontal transfer. Further studies showed a high prevalence of the *mcr*-1 gene (e.g., 20% in animal-specific bacterial strains and 1% in human-specific bacterial strains in China), and the plasmid has been detected in several countries covering Europe, Asia, South America, North America, and Africa [69–76]. Further *mcr* variants have been described as *mcr*-1 to 9 as of December 2019 [77, 78]. In the present data, we detected 15 proteins classified as Mcr-1 by deepARG, most abundant in the Atlantic Southwest Shelves Province, followed by its adjacent region, Antarctic Province (Fig. 5). However, the version of deepARG that we used did not classify these sequences into the more recently described Mcr-2 to 9. Therefore, we performed a phylogenetic analysis (Fig. 6), which included sequences of different Mcrs (Mcr-1 to 5) and LptA (encoded by the gene *eptA*, used here as outgroup). The results suggested that 5
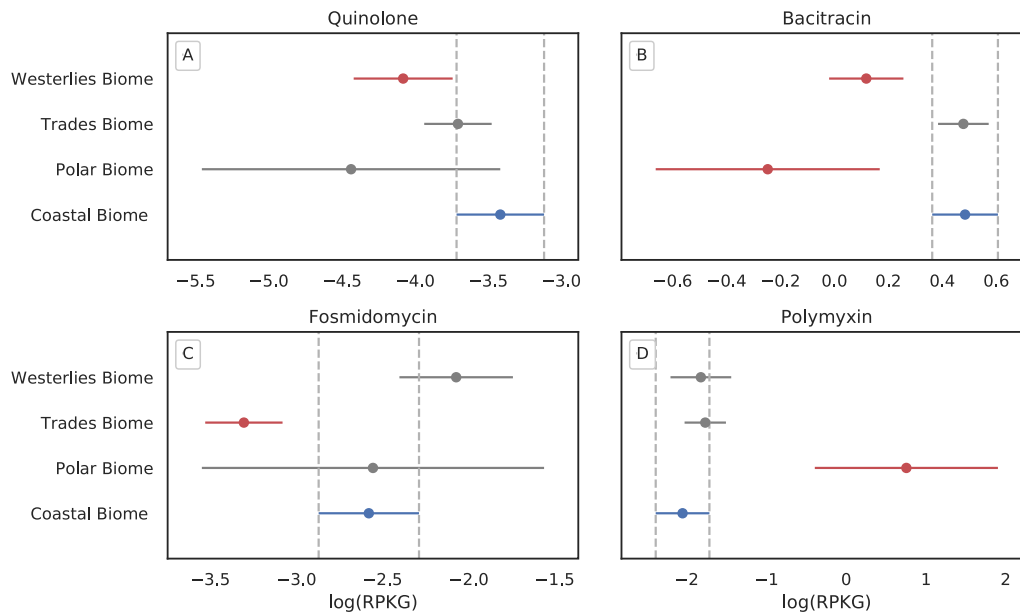
Figure 3: Significantly different mean abundances of ARG classes from oceanic biomes. Tukey HSD comparing the log-transformed RPKG of ARG classes for 4 biomes of TARA Oceans study. Shown are the means and 95% confidence intervals of RPKG for (A) quinolone ARGs, (B) bacitracin ARGs, (C) fosmidomycin ARGs, and (D) polymyxin ARGs. Blue indicates the reference for the test (coastal biome, chosen on the basis of its ecological relevance), and red, biome significantly different from the reference (P-value < 0.05).
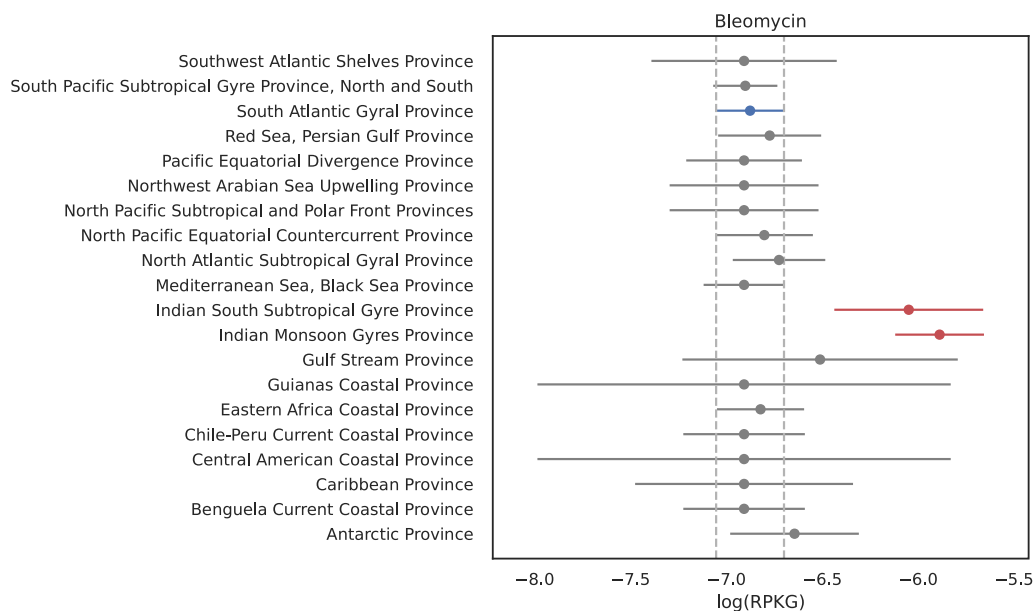
Figure 4: Bleomycin ARG abundance in marine provinces. Tukey HSD comparing the mean RPKG of ARGs from the class bleomycin. Blue indicates the reference for the test, and red, biome significantly different from the reference (P-value < 0.05). Error bars indicate 95% confidence intervals. The reference was chosen randomly.

ORFs (from the genus *Psychrobacter*, family Moraxellaceae [79]) are close to the Mcr-1/2 clade with a support value of 1 (Fig. 6). Members of the genus *Psychrobacter* were isolated from a wide range of habitats, including food, clinical samples, skin, gills, and intestines of fish, seawater, and Antarctic sea ice [80–84]. Importantly, ≥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 [81]. Coincidently, the regions with greater RPKG mean values for Mcr-1 abundance in our study were Southwest Atlantic and Antarctic Province. Our results support that *Psychrobacter* might be an ecological reservoir for the transfer of PEA transferases to other pathogens, and further studies should be conducted to better elucidate the dynamics and evolution of ARGs in this genus. Also, some species of this genus were reported to cause opportunistic infections in humans, including ≥1 case reported to be associated with marine environment exposure [85]. In this context, it is therefore essential to increase monitoring by, e.g., including screenings specific for *mcr*-related genes in these genera.
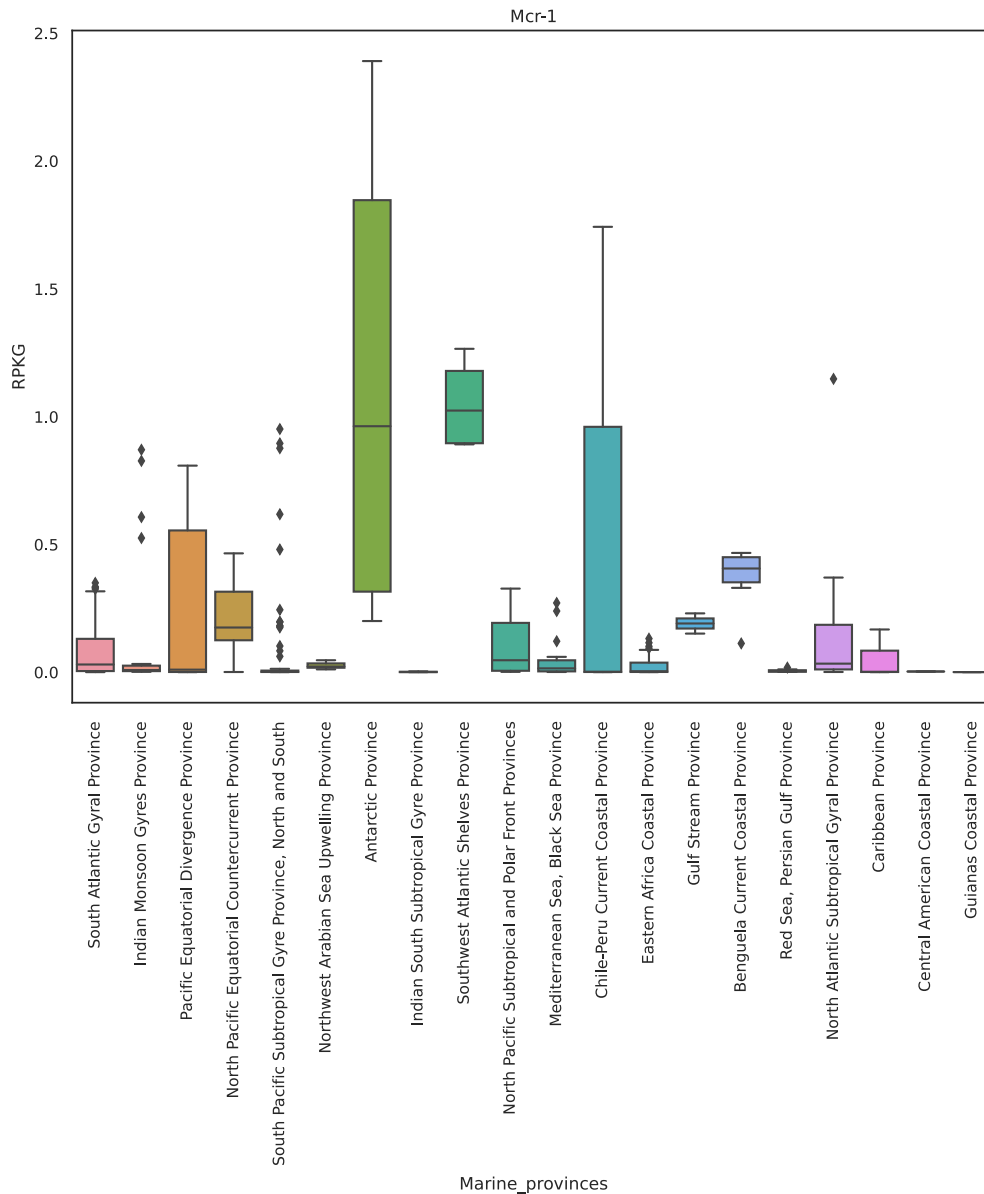
Figure 5: Mcr-1 distribution in TARA Oceans marine provinces. The boxplot shows the sum of RPKG values for all Mcr-1 ORFs.

The residual Mcr sequences, mostly belonging to the *Thioglobus* genus, were phylogenetically farther away from Mcr-1/2 and might constitute new, distinct Mcr variants (Fig. 6). Important to note is that the phylogenetically close relationship to Mcr sequences does not prove the function as a colistin-resistant gene, which awaits further experiments to confirm this role.

Only 2 *mcr* sequences were classified as present on plasmids via PlasFlow, which can be explained by the small size of many *mcr*-containing contigs (with 8 of them smaller than 3 kb). Additionally, a false-negative result from PlasFlow could be the result of a re-integration of plasmidial sequences into the chromosome—or that these *mcr* genes may constitute an ancestor of the plasmidial *E. coli mcr* sequences, as suggested for *mcr*-1 encoded by *Moraxella* species [79]. The 2 ARGs classified as located on a plasmid are detected in contigs with a size of 2 and 38 kb. The former, classified as belonging to a *Thioglobus* species, is challenging to validate as a plasmidial sequence owing to its small size. The latter is classified as a sequence of a

Poseidonibacter species, a marine group of bacteria recently reclassified from the *Arcobacter* genus, the latter containing several pathogenic species [86]. A toxin-antitoxin system is encoded 2 ORFs upstream of the *mcr* gene, which might be an indication for a plasmidial location. However, no further genes that are usually located on *Arcobacter* spp. plasmids [87] were found on this contig, hampering its correct classification as a plasmidial *mcr*. That said, various mobile element genes located on this contig (Fig. 7) strengthen the assumption that this contig is related to a mobile genetic region. An unusual synteny of *mcr*, *pap2*, and a downstream encoded *dagK* was observed (Fig. 7), of which *dagK* only appears in *mcr-3* genetic environments [88]. Related genes (amino acid sequence identity of ~70%) with a conserved gene synteny are found in several *Arcobacter* species (Fig. 7). A few *Arcobacter* species with a similar *mcr* gene were susceptible to colistin treatment [89], arguing against a colistin resistance conferred by this gene product. Further research is necessary to confirm or refute colistin resistance in marine *Poseidonibacter*.
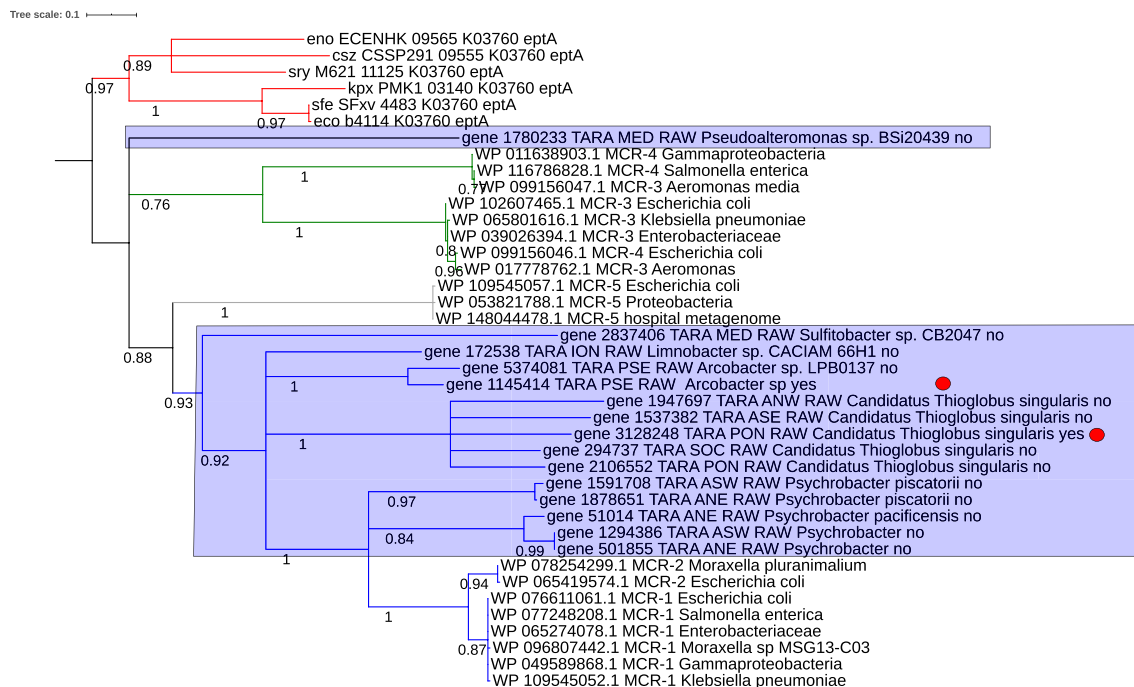
Figure 6: Phylogenetic tree of MCR sequences. The phylogenetic tree was inferred using the standard pipeline from phylogeny.fr (phyML with the "WAG" model and statistical test approximate likelihood-ratio [ALRT] for support values). Sequences for the outgroup *eptA* and clinical Mcr-1 to Mcr-5 were obtained from NCBI and used in addition to the sequences obtained from our results from TARA Oceans co-assemblies. The names of the TARA Oceans sequences displayed in the tree are defined with the ID of sequence, co-assembly ID, taxon name from Kaiju, and yes/no for plasmid classification from PlasFlow. The blue rectangles mark TARA sequences. The blue clade depicts the MCR-1/2 clade, the grey clade Mcr-5, the green clade MCR-3/4, and the red clade *eptA*. The red circles indicate sequences located on contigs classified as plasmids by PlasFlow. Numbers indicate ALRT support values.
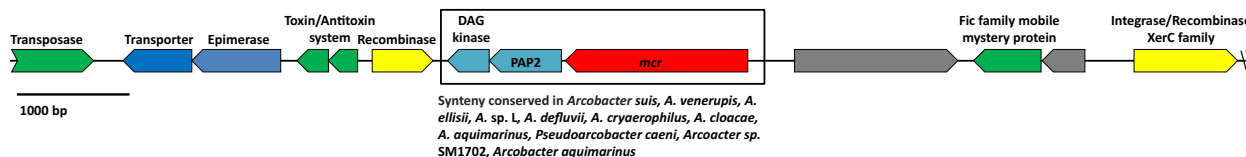

Figure 7: Genomic context of the *mcr* gene of contig TARA_PSE_k99_4834589. This contig was classified to be plasmidial by PlasFlow. Depicted are the first 13 ORFs from 28 of the whole contig, showing *mcr* (red) and surrounding genes and including the mobile element–related genes (green). DAG: diacylglycerol; PAP2: phosphatase PAP2 family protein; *mcr*: mobilized colistin resistance protein. Blue indicates other/metabolic genes; yellow, DNA-related genes; light blue, Mcr-accessory genes; and grey, hypothetical protein. Annotations from MetaGeneMark were manually refined using the conserved domains database and BLASTp against the SwissProt database. The taxonomy of *Arcobacter* species is stated as of December 2019 in the GenaBnk taxonomy database.

The presence of *mcr*-related genes in both Antarctic and adjacent regions can also raise concerns about gene flow due to ice melting, a problem already discussed previously for other ARGs [90].

Conclusions

This study uncovers the diversity and abundance of ARGs in the global ocean metagenome, conferring putative resistance to 26 classes of antibiotics. The extensive analysis leads to a detailed taxonomic classification and distribution of ARG abundance in different biomes. The present study also exposes the importance of monitoring coastal water for anthropogenic impact because the inflow of antibiotic-resistant strains by, e.g., wastewater might provide input of ARGs by HGT for environmental strains. This study could also affect investigations dealing with the evolutionary history of ARGs, with the herein-presented genes as ancestors of common ARGs in clinically relevant strains. Last but not least, the combination of multiple

modern machine learning tools and other open source data science libraries such as Dash and Plotly produced a valuable resource for the scientific community working on further studies on ARGs in different environments.

Availability of Source Code and Requirements

- Project name: ResistomeDB
- Project home page: https://resistomedb.com
- Operating system(s): Platform independent
- Programming language: Python
- Other requirements: None
- License: MIT
- RRID:SCR_018305

Availability of Supporting Data and Materials

Snapshots of code and other supporting data are available in the *GigaScience* repository, GigaDB [91].

Additional Files

Figure S1: ARG distribution in the 2 plasmids showing 5 ARGs each. The sizes of genes and distances are not scaled. PBP2B: methicillin-resistant PBP2; MTRA: transcriptional activator of the MtrCDE multidrug efflux pump; DFRE: dihydrofolate reductase; DFRA3: integron-encoded dihydrofolate reductase; BCR: bicyclomycin resistance protein; VANXA: variant of VANX D, D-dipeptidase; MEXH: membrane fusion protein of the efflux complex MexGHI-OpmD; VANSO: variant of VANS, required for high-level transcription of other van glycopeptide resistance genes; VANRI: regulatory transcriptional activator in the VanSR regulator within the VanI glycopeptide resistance gene cluster.

Figure S2: Q-Q plots for each ARG class. The figure shows the distribution of residuals from the OLS models for each ARG class.

Table S1: Number of contigs, ORFs, and putative ARGs for each oceanic region (metagenomic co-assembly).

Table S2: Manual curation of the ARGs. The table shows if the ARG was assigned for quantification studies and in each category was classified.

Table S3: Pairwise Tukey HSD significant results. The table shows the significant results (adjusted P-value > 0.05) of the paired Pairwise Tukey HSD for each pair of biogeographic biomes.

Table S4: ANOVA results for each ARG class. The table shows the significant results (P-value > 0.05) of the ANOVA for each ARG class.

Table S5: OLS results for each ARG class. The table shows the results of the OLS model for each ARG class, including model parameters and diagnostic

Abbreviations

AGS: average genome size; ALRT: approximate likelihood-ratio test; ANOVA: analysis of variance; AR: antibiotic resistance; ARDB: Antibiotic Resistance Genes Database; ARG: antibiotic resistance gene; BLAST: Basic Local Alignment Search Tool; bp: base pairs; CARD: Comprehensive Antibiotic Resistance Database; CDD: conserved domain; EBI: European Bioinformatics Institute; ENA: European Nucleotide Archive; FPKM: fragments per kilobase per million mapped reads; GE: genome equivalent; HGT: horizontal genetic transfer; HSD: honestly significant difference; kb: kilobase pairs; Mb: megabase pairs; MCR: mobilized colistin resistance; NCBI: National Center for Biotechnology Information; NDM: New Delhi metallo-β-lactamase; OLS: ordinary least squares; ORF: open reading frame; PEA: phosphoethanolamine; RPKG: reads per kilobase per genome equivalents; WAG: Whelan and Goldman.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All authors conceived and designed the analysis; R.R.C.C., M.S., and B.G.A. performed the analysis; R.R.C.C. and M.S. conceived and designed the database; and R.R.C.C. designed the web application. All authors wrote the manuscript and revised it for significant intellectual content.

Acknowledgements

References

1. Aslam B, Wang W, Arshad MI, et al. Antibiotic resistance: a rundown of a global crisis. Infect Drug Resist 2018;**11**:1645–58.
2. Tagliabue A, Rappuoli R. Changing priorities in vaccinology: antibiotic resistance moving to the top. Front Immunol 2018;**9**, doi:10.3389/fimmu.2018.01068.
3. Risso VA, Gavira JA, Mejia-Carmona DF, et al. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. J Am Chem Soc 2013;**135**:2899–902.
4. Hall BG, Barlow M. Evolution of the serine β-lactamases: past, present and future. Drug Resist Updat 2004;**7**:111–23.
5. Wright GD. The antibiotic resistome: the nexus of chemical and genetic diversity. Nat Rev Microbiol 2007;**5**:175–86.
6. Aminov RI. The role of antibiotics and antibiotic resistance in nature. Environ Microbiol 2009;**11**:2970–88.
7. Hatosy SM, Martiny AC. The ocean as a global reservoir of antibiotic resistance genes. Appl Environ Microbiol 2015;**81**:7593–9.
8. Yang Y, Li Z, Song W, et al. Metagenomic insights into the abundance and composition of resistance genes in aquatic environments: influence of stratification and geography. Environ Int 2019;**127**:371–80.
9. McConnell MM, Hansen LT, Neudorf KD, et al. Sources of antibiotic resistance genes in a rural river system. J Environ Qual 2018;**47**:997–1005.
10. Van Goethem MW, Pierneef R, Bezuidt OKI, et al. A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. Microbiome 2018;**6**(1):40.
11. Hayward JL, Jackson AJ, Yost CK, et al. Fate of antibiotic resistance genes in two Arctic tundra wetlands impacted by municipal wastewater. Sci Total Environ 2018;**642**:1415–28.
12. Carney RL, Labbate M, Siboni N, et al. Urban beaches are environmental hotspots for antibiotic resistance following rainfall. Water Res 2019;**167**:115081.
13. Fresia P, Antelo V, Salazar C, et al. Urban metagenomics uncover antibiotic resistance reservoirs in coastal beach and sewage waters. Microbiome 2019;**7**(1):35.
14. 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.
15. Wang M, Liu P, Zhou Q, et al. Estimating the contribution of bacteriophage to the dissemination of antibiotic resistance genes in pig feces. Environ Pollut Barking Essex 2018;**238**:291–8.
16. Larrañaga O, Brown-Jaque M, Quirós P, et al. Phage particles harboring antibiotic resistance genes in fresh-cut vegetables and agricultural soil. Environ Int 2018;**115**:133–41.
17. Ibarbalz FM, Henry N, Brandão MC, et al. Global trends in marine plankton diversity across kingdoms of life. Cell 2019;**179**:1084–97.e21.
18. Lima-Mendez G, Faust K, Henry N, et al. Determinants of community structure in the global plankton interactome. Science 2015;**348**:1262073.
19. Pesant S, Not F, Picheral M, et al. Open science resources for the discovery and analysis of *Tara* Oceans data. Sci Data 2015;**2**:150023.
20. Arango-Argoty G, Garner E, Pruden A, et al. DeepARG: a deep learning approach for predicting antibiotic resis-

20. tance genes from metagenomic data. Microbiome 2018;**6**:23.

21. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data 2018;**5**:170203.

22. Delmont TO, Quince C, Shaiber A, et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nat Microbiol 2018;**3**:804.

23. European Nucleotide Archive. https://www.ebi.ac.uk/ena. Accessed August 2018.

24. Companion Tables Ocean Microbiome EMBL. http://ocean-microbiome.embl.de/data/OM.CompanionTables.xlsx. Accessed August 2018.

25. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res 2010;**38**:e132.

26. Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. Nucleic Acids Res 2018;**46**:e35.

27. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 2016;**7**:11257.

28. BLAST. https://blast.ncbi.nlm.nih.gov/. Accessed December 2018

29. Liu B, Pop M. ARDB–Antibiotic Resistance Genes Database. Nucleic Acids Res 2009;**37**:D443–447.

30. Jia B, Raphenya AR, Alcock B, et al. CARD 2017: expansion and model-centric curation of the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res 2017;**45**:D566–73.

31. UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res 2018;**46**:2699.

32. Bushnell B. BBMap: A fast, accurate, splice-aware aligner. Lawrence Berkeley National Laboratory, Berkeley, CA. 2014. Mar. Report No.: LBNL-7065E. https://www.osti.gov/biblio/1241166.

33. 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.

34. Cuadrat RRC. Resistome Statistical Analysis notebook. 2019. https://github.com/rcuadrat/ocean_resistome/blob/master/exploring.ipynb.

35. Dereeper A, Guignon V, Blanc G, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 2008;**36**:W465–9.

36. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;**32**:1792–7.

37. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 2000;**17**:540–52.

38. Guindon S, Dufayard J-F, Lefort V, et al. New Algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010;**59**:307–21.

39. ResistomeDB. https://github.com/mSorok/ResistomeDB.

40. Cuadrat RRC, Sorokina M, Andrade BG, et al. ResistomeDB. Zenodo 2020, 10.5281/zenodo.3473960.

41. Resistome DB http://resistomedb.com/. Accessed 1st March 2020

42. Resistome dash project repository https://github.com/rcuadrat/resistome_dash. Accessed 1st March 2020

43. Cuadrat RRC. Resistome analysis project repository. 2019. https://github.com/rcuadrat/ocean_resistome.

44. Reynolds LJ, Roberts AP, Anjum MF. Efflux in the oral metagenome: the discovery of a novel tetracycline and tigecycline ABC transporter. Front Microbiol 2016;**7**, doi:10.3389/fmicb.2016.01923.

45. Lerminiaux NA, Cameron ADS. Horizontal transfer of antibiotic resistance genes in clinical environments. Can J Microbiol 2019;**65**:34–44.

46. Qiu Y, Zhang J, Li B, et al. A novel microfluidic system enables visualization and analysis of antibiotic resistance gene transfer to activated sludge bacteria in biofilm. Sci Total Environ 2018;**642**:582–90.

47. Peng S, Dolfing J, Feng Y, et al. Enrichment of the antibiotic resistance gene tet(L) in an alkaline soil fertilized with plant derived organic manure. Front Microbiol 2018;**9**:1140.

48. Calero-Cáceres W, Méndez J, Martín-Díaz J, et al. The occurrence of antibiotic resistance genes in a Mediterranean river and their persistence in the riverbed sediment. Environ Pollut Barking Essex 2017;**223**:384–94.

49. Schleheck D, Dong W, Denger K, et al. An α-proteobacterium converts linear alkylbenzenesulfonate surfactants into sulfophenylcarboxylates and linear alkyldiphenyletherdisulfonate surfactants into sulfodiphenylethercarboxylates. Appl Environ Microbiol 2000;**66**:1911–6.

50. Dashiff A, Junka RA, Libera M, et al. Predation of human pathogens by the predatory bacteria *Micavibrio aeruginosavorus* and *Bdellovibrio bacteriovorus*. J Appl Microbiol 2011;**110**:431–44.

51. Dharani S, Kim DH, Shanks RMQ, et al. Susceptibility of colistin-resistant pathogens to predatory bacteria. Res Microbiol 2018;**169**:52–5.

52. Shi B-H, Arunpairojana V, Palakawong S, et al. *Tistrella mobilis* gen nov, sp nov, a novel polyhydroxyalkanoate-producing bacterium belonging to α-*Proteobacteria*. J Gen Appl Microbiol 2002;**48**:335–43.

53. Xu Y, Kersten RD, Nam S-J, et al. Bacterial biosynthesis and maturation of the didemnin anticancer agents. J Am Chem Soc 2012;**134**:8625–32.

54. Berendes F, Gottschalk G, Heine-Dobbernack E, et al. *Halomonas desiderata* sp. nov, a new alkaliphilic, halotolerant and denitrifying bacterium isolated from a municipal sewage works. Syst Appl Microbiol 1996;**19**:158–67.

55. Mata JA, Martínez-Cánovas J, Quesada E, et al. A detailed phenotypic characterisation of the type strains of *Halomonas* species. Syst Appl Microbiol 2002;**25**:360–75.

56. Lu J, Zhang Y, Wu J, et al. Occurrence and spatial distribution of antibiotic resistance genes in the Bohai Sea and Yellow Sea areas, China. Environ Pollut 2019;**252**:450–60.

57. Velkov T, Roberts KD, Nation RL, et al. Pharmacology of polymyxins: new insights into an 'old' class of antibiotics. Future Microbiol 2013;**8**:711–24.

58. Falagas ME, Kasiakou SK. Toxicity of polymyxins: a systematic review of the evidence from old and recent studies. Crit Care 2006;**10**:R27.

59. Zhang L, Calvo-Bado L, Murray AK, et al. Novel clinically relevant antibiotic resistance genes associated with sewage sludge and industrial waste streams revealed by functional metagenomic screening. Environ Int 2019;**132**:105120.

60. Dortet L, Nordmann P, Poirel L. Association of the emerging carbapenemase NDM-1 with a bleomycin resistance protein in Enterobacteriaceae and *Acinetobacter baumannii*. Antimicrob Agents Chemother 2012;**56**:1693–7.

61. Yong D, Toleman MA, Giske CG, et al. Characterization of a new metallo-β-lactamase gene, blaNDM-1, and a novel erythromycin esterase gene carried on a unique genetic struc-

ture in *Klebsiella pneumoniae* Sequence Type 14 from India. Antimicrob Agents Chemother 2009;**53**:5046–54.

62. Kumarasamy KK, Toleman MA, Walsh TR, et al. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect Dis 2010;**10**:597–602.

63. Long RA, Azam F. Antagonistic Interactions among marine pelagic bacteria. Appl Environ Microbiol 2001;**67**:4975–83.

64. Ponce-Soto GY, Aguirre-von-Wobeser E, Eguiarte LE, et al. Enrichment experiment changes microbial interactions in an ultra-oligotrophic environment. Front Microbiol 2015;**6**, doi:10.3389/fmicb.2015.00246.

65. Ju F, Beck K, Yin X, et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. ISME J 2019;**13**:346–60.

66. McKinney CW, Dungan RS, Moore A, et al. Occurrence and abundance of antibiotic resistance genes in agricultural soil receiving dairy manure. FEMS Microbiol Ecol 2018;**94**, doi:10.1093/femsec/fiy010.

67. Baron S, Hadjadj L, Rolain J-M, et al. Molecular mechanisms of polymyxin resistance: knowns and unknowns. Int J Antimicrob Agents 2016;**48**:583–91.

68. Liu Y-Y, Wang Y, Walsh TR, 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**:161–8.

69. Hasman H, Hammerum AM, Hansen F, et al. Detection of mcr-1 encoding plasmid-mediated colistin-resistant *Escherichia coli* isolates from human bloodstream infection and imported chicken meat, Denmark 2015. Euro Surveill 2015;**20**, doi:10.2807/1560-7917.es.2015.20.49.30085.

70. Falgenhauer L, Waezsada S-E, Yao Y, 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**:282–3.

71. Webb HE, Granier SA, Marault M, et al. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis 2016;**16**:144–5.

72. Tse H, Yuen K-Y. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis 2016;**16**:145–6.

73. Zhang R, Huang Y, Chan EW, et al. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis 2016;**16**:291–2.

74. Mulvey MR, Mataseje LF, Robertson J, et al. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis 2016;**16**:289–90.

75. Arcilla MS, van Hattem JM, Matamoros S, et al. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis 2016;**16**:147–9.

76. Malhotra-Kumar S, Xavier BB, Das AJ. Colistin resistance gene mcr-1 harboured on a multidrug resistant plasmid. Lancet Infect Dis 2016;**16**:283–4.

77. Kieffer N, Royer G, Decousser J-W, et al. mcr-9, an inducible gene encoding an acquired phosphoethanolamine transferase in *Escherichia coli*, and Its origin. Antimicrob Agents Chemother 2019;**63**, doi:10.1128/AAC.00965-19.

78. Hadjadj L, Baron SA, Olaitan AO, et al. Co-occurrence of variants of mcr-3 and mcr-8 genes in a *Klebsiella pneumoniae* isolate from Laos. Front Microbiol 2019;**10**, doi:10.3389/fmicb.2019.02720.

79. Wei W, Srinivas S, Lin J, et al. Defining ICR-Mo, an intrinsic colistin resistance determinant from *Moraxella osloensis*. PLoS Genet 2018;**14**:e1007389.

80. Maruyama A, Honda D, Yamamoto H, et al. 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**:835–46.

81. Bowman JP, Nichols DS, McMeekin TA. *Psychrobacter glacincola* sp. nov., a halotolerant, psychrophilic bacterium isolated from Antarctic sea ice. Syst Appl Microbiol 1997;**20**:209–15.

82. Bowman JP, Cavanagh J, Austin JJ, et al. Novel *Psychrobacter* species from Antarctic ornithogenic soils. Int J Syst Evol Microbiol 1996;**46**:841–8.

83. 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**:388–91.

84. Yumoto I, Hirota K, Sogabe Y, et al. *Psychrobacter okhotskensis* sp. nov., a lipase-producing facultative psychrophile isolated from the coast of the Okhotsk Sea. Int J Syst Evol Microbiol 2003;**53**:1985–9.

85. Bonwitt J, Tran M, Droz A, et al. *Psychrobacter sanguinis* wound infection associated with marine environment exposure, Washington, USA. Emerg Infect Dis 2018;**24**, doi:10.3201/eid2410.171821.

86. Pérez-Cataluña A, Salas-Massó N, Diéguez AL, et al. Revisiting the taxonomy of the genus *Arcobacter*: getting order from the chaos. Front Microbiol 2018;**9**, doi:10.3389/fmicb.2018.02077.

87. Douidah L, Zutter LD, Nieuwerburgh FV, et al. Presence and analysis of plasmids in human and animal associated *Arcobacter* species. PLoS One 2014;**9**:e85487.

88. Eichhorn I, Feudi C, Wang Y, et al. Identification of novel variants of the colistin resistance gene mcr-3 in *Aeromonas* spp. from the national resistance monitoring programme GERM-Vet and from diagnostic submissions. J Antimicrob Chemother 2018;**73**:1217–21.

89. Houf K, Devriese LA, Zutter LD, et al. Susceptibility of *Arcobacter butzleri*, *Arcobacter cryaerophilus*, and *Arcobacter skirrowii* to antimicrobial agents used in selective media. J Clin Microbiol 2001;**39**:1654–6.

90. 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.

91. Cuadrat RRC, Sorokina M, Andrade BG, et al. Supporting data for "Global ocean resistome revealed: exploring antibiotic resistance gene abundance and distribution in TARA Oceans samples." GigaScience Database 2020. http://dx.doi.org/10.5524/100739.