

1    **Supporting Information**

2

3    **Evidence for Long-term Anthropogenic Pollution: Hadal Trench as Depository**  
4    **and Indicator for Dissemination of Antibiotic Resistance Genes**

5

6    **Author:** Huiying Yang<sup>1</sup>, Rulong Liu<sup>2</sup>, Huafeng Liu<sup>1</sup>, Chen Wang<sup>1</sup>, Xiaole Yin<sup>3</sup>, Ming  
7    Zhang<sup>1</sup>, Jiasong Fang<sup>2</sup>, Tong Zhang<sup>3</sup>, Liping Ma<sup>1\*</sup>

8    **Author affiliation:**

9    <sup>1</sup>School of Ecological and Environmental Sciences, East China Normal University,  
10    Shanghai, China.

11    <sup>2</sup>Shanghai Engineering Research Center of Hadal Science and Technology, College of  
12    Marine Sciences, Shanghai Ocean University, Shanghai, China

13    <sup>3</sup>Environmental Microbiome Engineering and Biotechnology Laboratory, The University  
14    of Hong Kong, Hong Kong SAR, China.

15    **\* Corresponding author:**

16    Dr. Liping Ma

17    Email: lpma@des.ecnu.edu.cn

18    Tel: +86-21-54836059

19

20    **This PDF file includes:**

21    Texts S1 to S3

22    Tables S1 to S10

23    Figures S1 to S9

## Table of Contents

<b>Text S1</b> Identification of ARG-like sequences.....	S4
<b>Text S2</b> Co-occurrence of ARGs and microbial communities.....	S5
<b>Text S3</b> Source prediction of ARGs and method validation.....	S5
<b>Table S1</b> Sample information of collected hadal sediments .....	S6
<b>Table S2</b> Abundance of ARG types in MT sediments (copy/cell).....	S7
<b>Table S3</b> Abundance of ARG subtypes in MT sediments (copy/cell).....	S8
<b>Table S4</b> Public metagenomic datasets of sediment samples downloaded from NCBI database .....	S14
<b>Table S5</b> Information of collected 584 environmental samples for source prediction.	
<b>Table S6</b> SourceTracker prediction proportion variation (indicated by mean, SD and RSD) between runs in leave-one-out cross-validation by 584 samples.....	S16
<b>Table S7</b> The correlation coefficient of ARGs subtypes with genera (occurrence in $\geq 50\%$ samples and Spearman's $r > 0.7$ ).....	S17
<b>Table S8</b> Taxonomic comparison of ARG-carrying contigs.....	S19
<b>Table S9</b> Information of the retrieved ARG-carrying genomes and ARGs they carried.....	S20
<b>Table S10</b> Identification of ARG biomarkers .....	S21
<b>Figure S1</b> Experimental pipeline of sample collection, pretreatment and DNA extraction.....	S23
<b>Figure S2</b> Rarefaction curves of ARG subtypes.....	S24
<b>Figure S3</b> The distribution of ARG abundance and diversity of sediment metagenomic datasets (river and fish pond, estuary and offshore, and deep sea) downloaded from NCBI database.....	S25

<b>Figure S4</b> Geographic distribution of 584 data sets divided into four groups (AF, HF, WA and NT) covering 6 continents .....	S26
<b>Figure S5</b> Network analysis of ARGs and bacterial genera with sample occurrence $\geq 50\%$ and Spearman's correlation coefficient $r \geq 0.7$ ( $P$ -value $\leq 0.05$ ).....	S27
<b>Figure S6</b> Shared resistome and bacterial hosts of ARG-carrying contigs .....	S28
<b>Figure S7</b> Abundance shifts of recovered ARG-carrying genomes across depths.....	S29
<b>Figure S8</b> The shared resistome of six retrieved ARG-carrying genomes.....	S30
<b>Figure S9</b> Identification of anthropogenic ARGs.....	S31

## **Text S1 Identification of ARG-like sequences**

24 All the metagenomic sequencing data after quality control were searched for ARGs  
25 against the structured ARG database SARG v2.0 using UBLAST strategy with an E-value  
26  $\leq 10^{-7}$ <sup>1</sup>. A read was identified as an ARG-like sequence if its best alignment showed a  
27 similarity of  $\geq 80\%$  and alignment length of  $\geq 75\%$  comparing with the referenced ARG  
28 sequences in SARG database. The identified ARG-like sequences were subsequently  
29 classified according to the catalogue of 24 ARG types (e.g., tetracycline resistance genes)  
30 and 1,208 ARG subtypes (e.g., *tetA*, *tetB*, *tetC*)<sup>1</sup>. To evaluate the distribution of ARG in  
31 MT sediments, the abundance of ARGs was normalized and expressed as "copy of ARG  
32 per cell" (copy/cell) with the following equation<sup>2, 3</sup>:

$$33 \text{Abundance} = \sum_i^n \frac{N_{i(\text{ARG-like sequence})} \times L_{\text{reads}} / L_{i(\text{ARGs reference sequence})}}{N_{16S \text{ sequence}} \times L_{\text{reads}} / L_{16S \text{ sequence}}} \times N_{16S \text{ copy number}}$$

34 where  $N_{i(\text{ARG-like sequence})}$  is the number of the ARG-like reads annotated as one specific  
35 ARG reference sequence,  $L_{i(\text{ARGs reference sequence})}$  is the sequence length (bp) of the  
36 corresponding ARG reference sequence,  $N_{16S \text{ sequence}}$  is the number of the 16S rRNA gene  
37 sequence identified for the metagenomic sequencing data by comparison to Greengenes  
38 database<sup>4</sup>,  $L_{16S \text{ sequence}}$  is the average length of 16S rRNA genes (1,432 bp) in Greengenes  
39 database,  $n$  is the number of mapped ARG reference sequences belonging to that ARG  
40 type or subtype,  $L_{\text{reads}}$  is the sequence length (bp) of the metagenomic reads, and  $N_{16S \text{ copy}}$   
41  $\text{number}$  is the average copy number of 16S rRNA genes per cell. The average copy number  
42 within cells was calculated using CopyRighter<sup>5</sup> based on the ".biom" file generated by  
43 QIIME.  
44

45 **Text S2 Co-occurrence of ARGs and microbial communities**

46 Network analysis using the matrix of ARGs (with ARGs-OAP<sup>1</sup>) and taxonomic data (with  
47 MetaPhlAn2<sup>6</sup>) was performed with R “psychc” and “reshape2” packages, and then  
48 visualized using the interactive Gephi platform (version 0.9.2)<sup>7</sup>. ARGs or bacterial genera  
49 which occurred in at least 50% of samples were selected, and a correlation between two  
50 nodes was regarded as statistically significant for Spearman’s correlation coefficient  $r \geq$   
51 0.7 and  $P$  value  $\leq 0.05$  (fdr adjusted). As a result, the obtained network plot consisted of  
52 38 nodes (9 microbial taxa and 29 ARG subtypes) and 61 edges, and the modularity index  
53 was 0.606.

54

55 **Text S3 Source prediction of ARGs and method validation**

56 A Bayesian-based machine-learning classification SourceTracker v1.0<sup>8</sup> was employed to  
57 predict the probable sources of ARGs in hadal sediment based on the matrix of ARG  
58 subtypes and abundances. The model validation and ARG prediction was conducted  
59 following previous studies<sup>8, 9</sup>. A total of 584 environmental samples (**Table S5**) were  
60 collected covering 12 niches and 6 continents for training and building “source” models,  
61 mainly divided into four ecotypes, including animal feces (AF), human feces (HF),  
62 wastewater treatment plants (WA), and natural environments (NT). The established model  
63 was validated by using Leave-one-out strategies<sup>8</sup> (**Table S6**).

**Table S1** Sample information of collected hadal sediments.

Sample ID	Sampling location	Sampling depth	Dataset size (GB)	Length
MT1		0-2 cm	32	90
MT2		2-3 cm	20	95
MT3		3-4 cm	21	95
MT4		4-5 cm	19	95
MT5	Mariana Trench (MT)	5-6 cm	24	95
MT6		6-7 cm	21	95
MT7		7-8 cm	21	95
MT8		8-9 cm	18	95
MT9		9-10 cm	24	95

**Table S2** Abundance of ARG types in MT sediments (copy/cell).

ARG type	MT1	MT2	MT3	MT4	MT5	MT6	MT7	MT8	MT9
Aminoglycoside	4.75E-05	3.47E-04	3.18E-04	4.51E-05	4.76E-05	4.69E-04	2.52E-04	4.74E-05	3.99E-05
Bacitracin	3.16E-04	7.46E-04	8.98E-04	5.24E-04	4.04E-04	6.33E-04	7.69E-04	6.13E-04	6.41E-04
Beta-lactam	1.99E-03	8.26E-03	1.67E-03	1.10E-03	1.17E-03	1.80E-03	1.78E-03	7.64E-04	1.72E-03
Bleomycin	0	0	0	4.90E-05	0	0	0	0	0
Chloramphenicol	1.05E-04	4.04E-04	0	8.99E-05	9.04E-05	5.09E-05	2.87E-05	0	2.57E-05
Fosfomycin	3.42E-05	0	2.33E-04	3.45E-04	1.58E-04	9.27E-05	9.07E-05	1.04E-04	8.09E-05
Fosmidomycin	1.13E-05	1.50E-04	1.37E-04	3.33E-05	9.01E-05	9.41E-05	7.43E-05	1.78E-04	6.40E-05
Kasugamycin	1.67E-05	1.94E-04	0	0	1.92E-05	0	2.22E-05	0	0
MLS <sup>a</sup>	1.04E-03	6.80E-03	7.52E-05	6.54E-05	2.63E-05	0	6.79E-05	2.14E-05	2.50E-05
Multidrug	4.41E-04	2.24E-03	8.69E-04	7.29E-04	7.31E-04	6.90E-04	6.29E-04	7.38E-04	6.76E-04
Polymyxin	1.38E-05	7.99E-05	1.32E-04	1.09E-04	9.55E-05	8.41E-05	1.83E-05	1.05E-05	6.52E-05
Rifamycin	0	1.59E-05	1.31E-05	6.76E-05	7.92E-05	7.11E-05	3.80E-05	9.46E-05	2.25E-05
Sulfonamide	0	0	0	8.29E-05	1.47E-04	0	0	0	0
Tetracycline	8.02E-04	5.34E-03	5.17E-04	4.82E-04	3.42E-04	3.34E-04	4.21E-04	3.35E-04	4.47E-04
Trimethoprim	0	0	0	0	0	4.06E-05	0	4.09E-05	5.71E-05
Vancomycin	4.22E-05	1.31E-04	1.37E-04	1.60E-04	1.06E-04	1.07E-04	1.48E-04	6.35E-05	9.63E-05

<sup>a</sup>macrolide-lincosamide-streptogramin

**Table S3** Abundance of ARG subtypes in MT sediments (copy/cell).

ARG type	ARG subtype	MT1	MT2	MT3	MT4	MT5	MT6	MT7	MT8	MT9
Aminoglycoside	<i>aac(2')</i> -I	0	0	0	0	0	3.41E-05	6.42E-05	0	0
	<i>aac(3)</i> -I	0	0	0	0	0	1.85E-04	0	0	0
	<i>aac(3)</i> -VII	0	0	0	0	0	0	0	0	1.87E-05
	<i>aadA</i>	0	2.87E-05	1.14E-04	0	0	1.27E-04	4.59E-05	4.74E-05	0
	<i>ant(2")</i> -I	0	0	8.76E-05	0	4.76E-05	2.48E-05	1.19E-04	0	0
	<i>aph(3')</i> -I	0	0	4.76E-05	0	0	0	0	0	2.13E-05
	<i>aph(3")</i> -I	0	0	0	0	0	7.56E-05	0	0	0
	<i>aph(3")</i> -IIb	0	0	2.32E-05	0	0	2.31E-05	2.26E-05	0	0
	<i>aph(3")</i> -III	0	1.14E-04	0	0	0	0	0	0	0
	<i>aph(6)</i> -I	0	0	4.61E-05	4.51E-05	0	0	0	0	0
bifunctional aminoglycoside										
N-acetyltransferase and phosphotransferase		4.75E-05	2.03E-04	0	0	0	0	0	0	0
Beta-lactam	<i>bacA</i>	3.16E-04	7.46E-04	8.98E-04	5.24E-04	4.04E-04	6.33E-04	7.69E-04	6.13E-04	6.41E-04
	<i>cfxA2</i>	1.42E-04	2.00E-03	0	0	0	0	0	0	0
	<i>cfxA3</i>	0	7.04E-05	0	0	0	0	0	0	0
	class A beta-lactamase	0	0	4.22E-05	0	0	0	0	0	0
	class C beta-lactamase	0	3.81E-05	3.03E-05	0	2.79E-05	0	0	0	0
	LRA-3	1.51E-05	0	0	0	0	0	0	0	0
	metallo-beta-lactamase	0	0	4.28E-05	4.53E-05	0	0	0	0	0
OXA-22		2.99E-04	4.40E-04	1.36E-04	3.59E-04	3.45E-04	3.60E-04	1.98E-04	3.53E-04	1.57E-04

OXA-60	3.02E-04	5.01E-04	3.21E-04	1.94E-04	1.36E-04	1.36E-04	4.00E-04	2.55E-04	3.57E-04
OXA-85	0	5.94E-05	0	0	0	0	0	0	0
PBP-1A	1.92E-04	1.13E-03	0	0	0	0	0	0	0
PBP-1B	1.33E-04	6.70E-04	0	0	0	0	0	0	0
PBP-2X	3.09E-04	1.31E-03	0	0	0	0	1.61E-05	0	0
<i>penA</i>	8.60E-05	8.81E-04	0	0	0	0	0	0	0
TEM-1	3.07E-04	2.89E-04	2.19E-04	1.27E-04	2.06E-04	3.62E-04	4.92E-04	8.26E-05	3.51E-04
TEM-102	0	0	6.51E-05	0	1.84E-05	4.31E-05	0	0	0
TEM-106	0	0	0	0	0	0	0	0	1.88E-05
TEM-117	8.95E-05	2.37E-04	1.96E-04	2.58E-05	1.03E-04	1.94E-04	3.09E-04	0	1.48E-04
TEM-118	1.74E-05	8.66E-05	2.38E-05	2.51E-05	2.01E-05	4.72E-05	6.93E-05	0	6.18E-05
TEM-123	0	2.64E-05	0	0	0	0	0	0	0
TEM-138	0	0	4.34E-05	0	3.67E-05	0	0	0	3.76E-05
TEM-157	0	1.05E-04	4.34E-05	2.30E-05	1.84E-05	4.31E-05	6.32E-05	0	3.76E-05
TEM-171	0	2.64E-05	2.17E-05	0	0	2.16E-05	0	0	0
TEM-177	0	0	0	0	0	0	2.11E-05	0	0
TEM-178	1.60E-05	0	1.09E-04	4.61E-05	1.84E-05	1.30E-04	2.12E-05	0	3.78E-05
TEM-187	1.60E-05	5.29E-05	6.53E-05	2.30E-05	1.84E-05	4.33E-05	2.12E-05	0	3.78E-05
TEM-188	0	2.64E-05	2.17E-05	0	0	0	0	0	0
TEM-195	1.59E-05	2.64E-05	2.17E-05	2.30E-05	3.67E-05	0	4.22E-05	4.84E-05	1.88E-05
TEM-205	1.59E-05	1.05E-04	4.34E-05	6.89E-05	9.18E-05	4.31E-05	0	0	1.13E-04
TEM-209	0	5.27E-05	4.34E-05	2.30E-05	1.84E-05	4.31E-05	4.22E-05	0	0
TEM-216	0	0	0	0	0	0	2.11E-05	0	3.76E-05
TEM-29	0	2.64E-05	0	0	0	2.16E-05	0	0	0
TEM-6	0	0	2.17E-05	4.59E-05	1.84E-05	8.62E-05	4.22E-05	0	1.88E-05
TEM-63	0	0	2.17E-05	0	0	0	0	0	0

		TEM-75	0	5.78E-05	2.38E-05	0	2.01E-05	4.72E-05	2.31E-05	0	8.24E-05
		TEM-89	1.64E-05	2.71E-05	2.23E-05	2.36E-05	1.89E-05	6.65E-05	0	0	5.81E-05
		TEM-91	1.59E-05	2.64E-05	8.68E-05	4.59E-05	1.84E-05	1.08E-04	0	2.42E-05	1.50E-04
Bleomycin	bleomycin resistance protein		0	0	0	4.90E-05	0	0	0	0	0
Chloramphenicol	<i>cat</i>		0	0	0	0	0	0	0	0	2.57E-05
	<i>catB</i>		0	0	0	0	0	0	2.87E-05	0	0
	<i>catD</i>		2.15E-05	7.11E-05	0	0	0	0	0	0	0
	<i>catS</i>		8.32E-05	2.30E-04	0	0	0	0	0	0	0
	chloramphenicol exporter		0	0	0	0	1.74E-05	2.04E-05	0	0	0
	<i>cmlA</i>		0	1.03E-04	0	8.99E-05	7.30E-05	3.05E-05	0	0	0
Fosfomycin	<i>fosX</i>		3.42E-05	0	2.33E-04	3.45E-04	1.58E-04	9.27E-05	9.07E-05	1.04E-04	8.09E-05
	<i>rosA</i>		1.13E-05	1.50E-04	1.26E-04	3.33E-05	8.04E-05	9.41E-05	7.43E-05	1.78E-04	5.41E-05
	<i>rosB</i>		0	0	1.15E-05	0	9.71E-06	0	0	0	9.94E-06
MLS	<i>ksgA</i>		1.67E-05	1.94E-04	0	0	1.92E-05	0	2.22E-05	0	0
	<i>ermB</i>		6.48E-04	2.18E-03	0	0	0	0	4.92E-05	0	0
	<i>ermF</i>		2.57E-04	3.77E-03	0	0	0	0	0	0	0
	<i>ermX</i>		0	2.29E-04	0	0	0	0	0	0	0
	<i>lsa</i>		9.25E-06	0	0	0	0	0	0	0	0
	<i>macA</i>		0	2.12E-04	0	0	0	0	0	0	0
	<i>macB</i>		7.02E-06	4.66E-05	1.92E-05	1.01E-05	1.63E-05	0	1.87E-05	2.14E-05	2.50E-05
	<i>mefA</i>		1.24E-04	3.53E-04	0	0	0	0	0	0	0
	<i>mphC</i>		0	0	0	2.20E-05	0	0	0	0	0
	<i>msrA</i>		0	1.58E-05	5.60E-05	3.33E-05	0	0	0	0	0
Multidrug	<i>acrB</i>		4.33E-06	7.19E-06	0	1.90E-05	1.00E-05	1.18E-05	1.74E-05	6.60E-06	1.54E-05

<i>adeJ</i>	0	0	0	1.24E-05	0	0	0	6.54E-06	0
<i>amrB</i>	0	0	0	0	0	0	0	6.61E-06	0
<i>bcr</i>	0	0	0	0	1.32E-05	0	0	0	0
<i>bpeF</i>	0	0	5.85E-06	1.86E-05	0	1.74E-05	4.55E-05	1.96E-05	1.01E-05
<i>ceoB</i>	2.19E-05	5.84E-05	6.60E-05	2.50E-05	4.04E-05	4.12E-05	4.68E-05	4.70E-05	4.17E-05
<i>emrA</i>	1.25E-05	1.38E-04	1.59E-05	3.37E-05	0	0	1.55E-05	0	1.47E-05
<i>emrB</i>	2.72E-05	3.42E-04	0	0	2.05E-05	0	0	0	0
<i>emrB-qacA</i> family major facilitator transporter	0	0	2.39E-05	1.26E-05	0	0	1.16E-05	1.33E-05	2.16E-05
<i>emrD</i>	0	0	1.62E-05	1.67E-05	0	3.18E-05	0	0	0
<i>emrE</i>	0	0	0	0	0	0	0	0	7.09E-05
major facilitator superfamily transporter	5.22E-05	2.88E-05	4.90E-05	0	3.01E-05	1.18E-05	2.60E-05	5.63E-05	3.49E-05
<i>mdfA</i>	1.10E-05	0	0	0	0	0	0	0	1.22E-05
<i>mdtB</i>	0	5.07E-05	5.96E-05	3.15E-05	4.54E-05	2.37E-05	8.10E-05	3.33E-05	4.65E-05
<i>mdtC</i>	0	0	3.02E-05	0	5.11E-06	0	5.87E-06	1.35E-05	5.25E-06
<i>mdtF</i>	0	0	1.20E-05	0	0	5.94E-06	0	0	0
<i>mdtG</i>	3.42E-05	9.45E-05	0	0	0	0	0	0	0
<i>mdtK</i>	4.96E-05	3.61E-04	0	0	1.15E-05	0	0	0	0
<i>mdtN</i>	0	0	0	0	1.53E-05	0	0	0	0
<i>mepA</i>	0	0	0	0	0	0	0	0	1.19E-05
<i>mexA</i>	0	0	1.85E-05	0	0	0	0	0	0
<i>mexB</i>	4.40E-06	1.45E-05	0	1.34E-05	0	0	1.31E-05	0	0
<i>mexC</i>	0	0	1.61E-05	0	0	0	0	0	0
<i>mexD</i>	0	7.23E-06	0	0	0	0	0	6.63E-06	0
<i>mexE</i>	0	0	1.52E-05	6.60E-05	3.93E-05	1.52E-05	1.51E-05	3.40E-05	4.02E-05

	<i>mexF</i>	0	4.27E-05	7.02E-05	3.10E-05	5.94E-05	6.40E-05	1.71E-05	5.22E-05	4.57E-05
	<i>mexT</i>	1.88E-05	2.29E-05	7.70E-05	0	0	2.03E-05	0	4.21E-05	0
	<i>mexW</i>	0	7.48E-06	6.10E-06	0	5.18E-06	0	1.78E-05	0	0
	<i>mtrE</i>	0	4.84E-05	0	0	0	0	0	0	0
	multidrug ABC transporter	1.48E-05	1.37E-04	9.44E-05	1.75E-04	1.65E-04	1.60E-04	1.62E-04	7.86E-05	1.39E-04
	multidrug transporter	1.69E-04	7.53E-04	1.93E-04	1.50E-04	1.91E-04	1.97E-04	1.22E-04	1.71E-04	1.15E-04
	<i>ompR</i>	0	0	2.46E-05	0	2.08E-05	4.89E-05	0	8.24E-05	2.13E-05
	<i>opcM</i>	0	0	0	0	1.02E-05	0	0	1.34E-05	0
	<i>opmD</i>	9.34E-06	0	0	0	0	0	0	0	0
	<i>oprC</i>	0	1.91E-05	7.88E-06	0	3.33E-06	7.82E-06	3.83E-06	0	6.83E-06
	<i>oprM</i>	0	4.80E-05	0	1.39E-05	0	0	0	0	0
	<i>oprN</i>	0	0	0	0	1.11E-05	0	1.28E-05	0	0
	<i>smeB</i>	0	1.44E-05	5.92E-06	1.88E-05	5.01E-06	1.76E-05	0	6.59E-06	5.13E-06
	<i>smeC</i>	0	3.20E-05	0	0	0	0	0	0	0
	<i>smeD</i>	1.16E-05	0	3.16E-05	5.01E-05	1.33E-05	1.57E-05	1.53E-05	3.51E-05	1.37E-05
	<i>smeE</i>	0	1.45E-05	2.98E-05	1.26E-05	5.05E-06	0	0	1.33E-05	5.17E-06
	<i>smeF</i>	0	0	0	2.79E-05	1.13E-05	0	0	0	0
Polymyxin	<i>arnA</i>	1.38E-05	7.99E-05	1.32E-04	1.09E-04	9.55E-05	8.41E-05	1.83E-05	1.05E-05	6.52E-05
Rifamycin	<i>arr</i>	0	0	0	4.03E-05	3.50E-05	4.53E-05	0	5.09E-05	0
	monooxygenase	0	1.59E-05	1.31E-05	2.73E-05	4.41E-05	2.58E-05	3.80E-05	4.37E-05	2.25E-05
Sulfonamide	<i>sul1</i>	0	0	0	8.29E-05	1.47E-04	0	0	0	0
	<i>ter32</i>	0	3.54E-05	0	0	0	0	0	0	0
	<i>ter35</i>	2.47E-05	3.27E-04	4.54E-04	4.09E-04	3.42E-04	3.34E-04	3.92E-04	3.19E-04	4.23E-04
Tetracycline	<i>ter37</i>	4.21E-05	9.77E-04	0	0	0	0	0	0	0
	<i>ter39</i>	2.30E-05	0	0	3.32E-05	0	0	0	0	0
	<i>tetA</i>	0	1.13E-04	0	0	0	0	0	0	0

<i>tetC</i>		2.32E-05	3.84E-04	0	0	0	0	0	0
<i>tetD</i>		0	1.09E-04	0	0	0	0	0	0
<i>tetL</i>		0	0	0	0	0	0	0	2.45E-05
<i>tetM</i>		3.27E-04	1.20E-03	0	0	0	0	0	0
<i>tetO</i>		0	2.36E-05	0	0	0	0	0	0
<i>tetQ</i>		2.25E-04	1.30E-03	0	0	0	0	0	0
tetracycline	resistance protein	1.08E-04	8.09E-04	0	3.92E-05	0	0	0	0
<i>tetS</i>		7.06E-06	1.17E-05	0	0	0	0	0	0
<i>tetV</i>		0	1.76E-05	4.38E-05	0	0	0	2.91E-05	1.65E-05
<i>tetW</i>		2.14E-05	3.54E-05	1.94E-05	0	0	0	0	0
<i>dfrA16</i>		0	0	0	0	0	4.06E-05	0	0
<i>dfrA20</i>		0	0	0	0	0	0	4.09E-05	3.18E-05
<i>dfrA5</i>		0	0	0	0	0	0	0	2.53E-05
Vancomycin	<i>vanB</i>	0	0	0	0	0	0	1.76E-05	0
	<i>vanR</i>	1.97E-05	1.31E-04	1.07E-04	1.41E-04	9.09E-05	1.07E-04	1.31E-04	2.99E-05
	<i>vanS</i>	0	0	0	1.90E-05	1.52E-05	0	0	0
	<i>vanX</i>	2.25E-05	0	3.01E-05	0	0	0	3.36E-05	2.66E-05

**Table S4** Public metagenomic datasets of sediment samples downloaded from NCBI database.

Sample ID	Sample type	Geographic Location	Latitude and longitude	Depth (m)	Accession number	# of Bases	References	Diversity	
								Abundance (copy/cell)	(number of ARG subtype)
SD01	Marine fish farm sediment	Tai Po, Hong Kong	22.45N, 114.17E	--	--	2.7G	[10]	2.01E-02	26
SD02	Marine fish farm sediment	Lamma Island, Hong Kong	22.21N, 114.13E	--	--	1G	[10]	3.36E-02	13
SD03	Freshwater fish pond sediment	San Tin, Hong Kong	22.50N, 114.07E	--	SRR1022348	3G	[10]	9.79E-02	53
SD04	River sediment	Kanawha River, USA: West Virginia, South Charleston	38.37N, 81.69W	0	SRR6814546	4.6G	[11]	6.74E-02	83
SD05	Estuary	Pearl River Estuary, site B2	22.44N, 113.72E	6.4	SRR1022378	3G	[12]	3.16E-02	33
SD06	Offshore sediment	Atlantic Ocean, Baltic Sea	55.50N, 15.48E	84.4	SRR12059197	5.8G	[13]	1.77E-02	12
SD07	Offshore sediment	South China sea, site E208	22.49N, 116.50E	35	--	3G	[12]	2.38E-02	23
SD08	Offshore sediment	Hong Kong: Port Island	22.50N, 114.36E	--	SRR8361708	5.4G	[14]	2.74E-02	20
SD09	Offshore sediment	Atlantic Ocean; Mexico: Gulf of Mexico	26.59N, 87.51W	2,800	SRR7716469	12.6G	[15]	4.62E-02	94

SD10	Offshore sediment	Yellow Sea	35.50N, 121.67E	39	SRR6441346	17.7G	[16]	5.12E-02	117
SD11	Offshore sediment	South China sea, site E201	19.69N, 119.33E	2,918	--	3G	[12]	7.17E-02	23
SD12	Deep sea sediment	Mariana Trench, the north slope	11.74N, 142.11E	5,838	DRR090556	2.2G	[17]	1.13E-03	6
SD13	Deep sea sediment	Arctic Ocean: Gakkel Ridge, Loki's castle	73.76N, 8.46E	3,250	SRR8452060	16.4G	[18]	1.26E-03	21
SD14	Deep sea sediment	Pacific Ocean	11.91N, 144.93E	7,942	SRR8709623	7.1G	[19]	1.26E-03	16

**Table S6** SourceTracker prediction proportion variation (indicated by mean, SD and RSD) between runs in leave-one-out cross-validation by 584 samples.

Expected ecotype	Predicted ecotype					
	AF	HF	WA	NT	UN	
AF	Mean (%)	79.73	14.31	1.24	1.9	2.82
	SD	0.07	0.04	0	0	0
	RSD (%)	8.3	27.35	20.99	25.1	15.91
HF	Mean (%)	7.79	89.7	0.74	0.7	1.08
	SD	0.03	0.04	0	0	0
	RSD (%)	32.5	4.06	19.61	21.37	6.93
WA	Mean (%)	0.85	1.64	82.95	12.07	2.5
	SD	0	0	0.02	0.01	0
	RSD (%)	7.38	22.39	2.87	12.37	4.44
NT	Mean (%)	0.29	0.22	7.75	88.83	2.91
	SD	0	0	0.03	0.04	0
	RSD (%)	14.42	6.85	39.13	3.94	8.44

**Table S7** The correlation coefficient of ARGs subtypes with genera (occurrence in  $\geq 50\%$  samples and Spearman's  $r \geq 0.7$ ).

No.	Genus	ARG/Genus	Spearman's $r$
1	<i>Burkholderia</i>	Beta-lactam TEM-178	0.912
2	<i>Thaumarchaeota</i>	<i>Delta</i> proteobacteria	0.883
3	<i>Chloroflexi</i>	Multidrug <i>ompR</i>	0.831
4	<i>Stenotrophomonas</i>	Polymyxin <i>arnA</i>	0.817
5	<i>Chloroflexi</i>	Multidrug <i>bpeF</i>	0.814
6	<i>Burkholderia</i>	Beta-lactam TEM-6	0.797
7	<i>Ralstonia</i>	<i>Burkholderia</i>	0.783
8	<i>Propionibacterium</i>	<i>Stenotrophomonas</i>	0.783
9	<i>Propionibacterium</i>	Multidrug <i>mexE</i>	0.778
10	<i>Burkholderia</i>	<i>Stenotrophomonas</i>	0.767
11	<i>Stenotrophomonas</i>	Tetracycline <i>tetA</i>	0.767
12	<i>Nitrosopumilaceae</i>	Multidrug <i>mexE</i>	0.762
13	<i>Propionibacterium</i>	Beta-lactam TEM-205	0.762
14	<i>Pseudomonas</i>	Multidrug <i>bpeF</i>	0.759
15	<i>Stenotrophomonas</i>	Fosfomycin <i>fosX</i>	0.750

16	<i>Ralstonia</i>	Beta-lactam TEM-178	0.745
17	<i>Deltaproteobacteria</i>	Multidrug <i>ompR</i>	0.729
18	<i>Ralstonia</i>	<i>Pseudomonas</i>	0.729
19	<i>Pseudomonas</i>	Beta-lactam TEM-6	0.724
20	<i>Stenotrophomonas</i>	Multidrug <i>mexE</i>	0.720
21	<i>Propionibacterium</i>	Beta-lactam TEM-91	0.717
22	<i>Ralstonia</i>	Beta-lactam TEM-6	0.712
23	<i>Thaumarchaeota</i>	<i>Propionibacterium</i>	0.700

**Table S8** Taxonomic comparison of ARG-carrying contigs.

Contig ID	NCBI best hit			This study	Match (✓) or not (✗)
	Assigned taxon	Identity (>90%)	Hit length (> 80%)	Assigned taxon	
MT0_k109_515212	Firmicutes	99.50%	99%	Firmicutes	✓
MT2_k113_125021	Bacteroidetes	99.30%	100%	Bacteroidetes	✓
MT2_k113_490872	Firmicutes	93.50%	99%	Firmicutes	✓
MT2_k113_562447	Proteobacteria	97.10%	100%	Proteobacteria	✓
MT2_k113_765422	Bacteroidetes	99.90%	88%	Bacteroidetes	✓
MT2_k113_832425	Firmicutes	99.80%	100%	Firmicutes	✓
MT2_k113_996763	Proteobacteria	95.90%	100%	Proteobacteria	✓
MT2_k113_1281130	Bacteroidetes	99.50%	100%	Bacteroidetes	✓
MT2_k113_1462292	Bacteroidetes	100%	100%	Bacteroidetes	✓
MT3_k99_1290732	Proteobacteria	98.20%	100%	Proteobacteria	✓

**Table S9** Information of the retrieved ARG-carrying genomes and ARGs they carried.

Parameters	bin1	bin2	bin3	bin4	bin5	bin6
Genome size (Mbp)	5.3	2.1	2	1.5	2.3	4.2
Number of scaffold	275	135	77	858	400	424
Number of ORFs	5,253	2,074	1,973	2,124	2,461	4,299
Mean length of scaffold (bp)	19,781	15,980	26,332	1,699	5,696	10,094
Completeness (%)	90.27	93.73	80.92	52.97	93.89	88.47
Contamination (%)	2.89	0	1.23	9.56	3.06	3.10
GC content (%)	65.8	44.1	56.7	40.6	69.3	64.2
ARGs	bacitracin- <i>bacA</i>	+	+			+
	beta-lactam-CfxA2			+		
	beta-lactam-OXA-22	+				+
	beta-lactam-OXA-60	+				+
	beta-lactam-PBP-1A			+		
	beta-lactam-PBP-2X			+		
	MLS- <i>ermF</i>			+		
	multidrug-ABC-transporter		+		+	
	multidrug- <i>acrB</i>	+				
	multidrug- <i>mdtB</i>					+
	multidrug- <i>ompR</i>	+				
	multidrug-transporter	+				+
	tetracycline- <i>tetM</i>			+		
	tetracycline- <i>tetQ</i>			+		

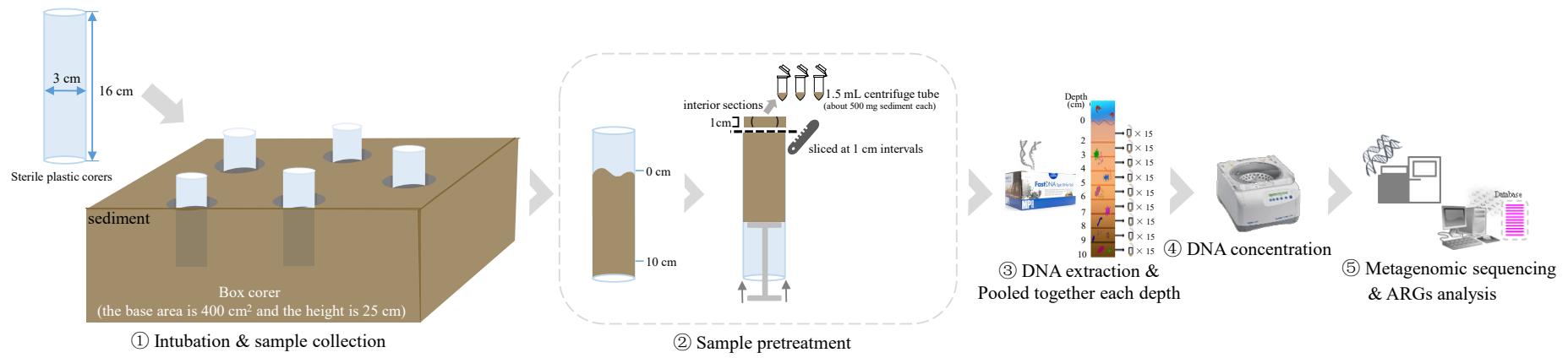
**Table S10** Identification of ARG biomarkers.

ARG subtype	Abundance (copy/cell)										A <sub>0-2cm</sub> /AVG(A <sub>3-10cm</sub> ) ≥ 5	A <sub>2-3cm</sub> /AVG(A <sub>3-10cm</sub> ) ≥ 5	A <sub>0-2cm</sub> ≥ 10 <sup>-4</sup> (copy/cell)	A <sub>2-3cm</sub> ≥ 10 <sup>-4</sup> (copy/cell)	(P <sub>AF</sub> + P <sub>HF</sub> + P <sub>WA</sub> ) ≥ 90%
	MT1	MT2	MT3	MT4	MT5	MT6	MT7	MT8	MT9						
Aminoglycoside aph(3'')-III	0	1.14E-04	0	0	0	0	0	0	0		✓			✓	
Aminoglycoside bifunctional															
N-acetyltransferase and phosphotransferase	4.75E-05	2.03E-04	0	0	0	0	0	0	0	✓	✓			✓	
Beta-lactam <i>cfxA2</i>	1.42E-04	2.00E-03	0	0	0	0	0	0	0	✓	✓	✓	✓	✓	✓
Beta-lactam LRA-3	1.51E-05	0	0	0	0	0	0	0	0	✓					
Beta-lactam OXA-85	0	5.94E-05	0	0	0	0	0	0	0		✓				
Beta-lactam PBP-1A	1.92E-04	1.13E-03	0	0	0	0	0	0	0	✓	✓	✓	✓	✓	
Beta-lactam PBP-1B	1.33E-04	6.70E-04	0	0	0	0	0	0	0	✓	✓	✓	✓	✓	
Beta-lactam PBP-2X	3.09E-04	1.31E-03	0	0	0	0	1.61E-05	0	0	✓	✓	✓	✓	✓	
Beta-lactam <i>penA</i>	8.60E-05	8.81E-04	0	0	0	0	0	0	0	✓	✓			✓	
Beta-lactam TEM-123	0	2.64E-05	0	0	0	0	0	0	0		✓				
Beta-lactam TEM-188	0	2.64E-05	2.17E-05	0	0	0	0	0	0		✓				
Beta-lactam TEM-29	0	2.64E-05	0	0	0	2.16E-05	0	0	0		✓				
Chloramphenicol <i>catD</i>	2.15E-05	7.11E-05	0	0	0	0	0	0	0	✓	✓				
Chloramphenicol <i>catS</i>	8.32E-05	2.30E-04	0	0	0	0	0	0	0	✓	✓			✓	
Kasugamycin <i>ksgA</i>	1.67E-05	1.94E-04	0	0	1.92E-05	0	2.22E-05	0	0		✓				
MLS <i>ermF</i>	2.57E-04	3.77E-03	0	0	0	0	0	0	0	✓	✓	✓	✓	✓	✓
MLS <i>ermX</i>	0	2.29E-04	0	0	0	0	0	0	0		✓				
MLS <i>lsa</i>	9.25E-06	0	0	0	0	0	0	0	0	✓					
MLS <i>macA</i>	0	2.12E-04	0	0	0	0	0	0	0		✓			✓	

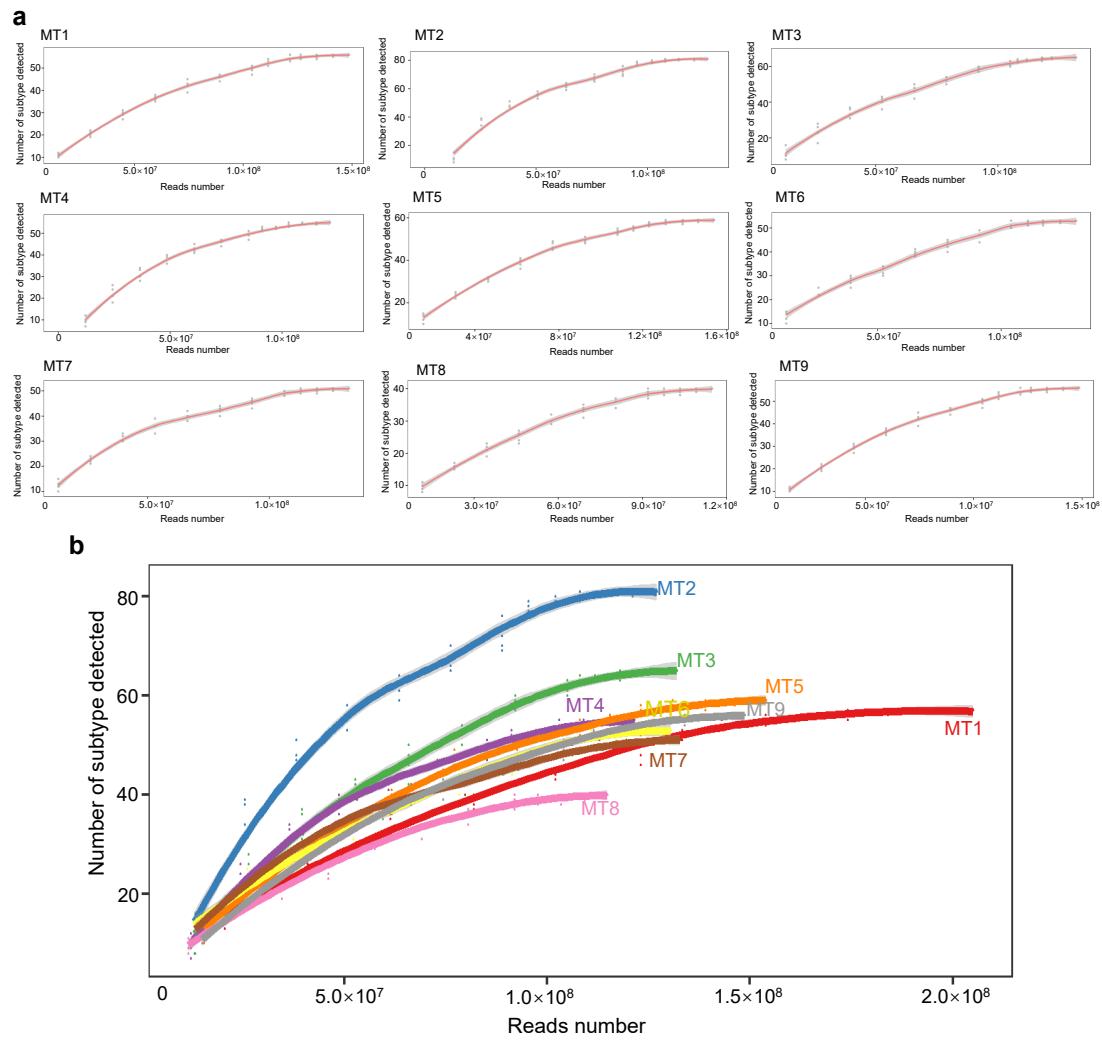
MLS <i>mefA</i>	1.24E-04	3.53E-04	0	0	0	0	0	0	0	✓	✓	✓	✓	✓
Multidrug <i>enrA</i>	1.25E-05	1.38E-04	1.59E-05	3.37E-05	0	0	1.55E-05	0	1.47E-05		✓			
Multidrug <i>enrB</i>	2.72E-05	3.42E-04	0	0	2.05E-05	0	0	0	0	✓	✓			
Multidrug <i>mdfA</i>	1.10E-05	0	0	0	0	0	0	0	1.22E-05	✓				
Multidrug <i>mdtG</i>	3.42E-05	9.45E-05	0	0	0	0	0	0	0	✓	✓			
Multidrug <i>mdtK</i>	4.96E-05	3.61E-04	0	0	1.15E-05	0	0	0	0	✓	✓			
Multidrug <i>mexD</i>	0	7.23E-06	0	0	0	0	0	6.63E-06	0		✓			
Multidrug <i>oprD</i>	9.34E-06	0	0	0	0	0	0	0	0	✓				
Multidrug <i>oprM</i>	0	4.80E-05	0	1.39E-05	0	0	0	0	0		✓			
Multidrug <i>smeC</i>	0	3.20E-05	0	0	0	0	0	0	0		✓			
Tetracycline <i>tet32</i>	0	3.54E-05	0	0	0	0	0	0	0		✓			
Tetracycline <i>tet37</i>	4.21E-05	9.77E-04	0	0	0	0	0	0	0	✓	✓			
Tetracycline <i>tetA</i>	0	1.13E-04	0	0	0	0	0	0	0		✓			
Tetracycline <i>tetC</i>	2.32E-05	3.84E-04	0	0	0	0	0	0	0	✓	✓			
Tetracycline <i>tetD</i>	0	1.09E-04	0	0	0	0	0	0	0		✓			
Tetracycline <i>tetM</i>	3.27E-04	1.20E-03	0	0	0	0	0	0	0	✓	✓	✓		
Tetracycline <i>tetO</i>	0	2.36E-05	0	0	0	0	0	0	0		✓			
Tetracycline <i>tetQ</i>	2.25E-04	1.30E-03	0	0	0	0	0	0	0	✓	✓	✓		
Tetracycline resistance protein	1.08E-04	8.09E-04	0	3.92E-05	0	0	0	0	0	✓	✓	✓		
Tetracycline <i>tetS</i>	7.06E-06	1.17E-05	0	0	0	0	0	0	0	✓	✓			
Tetracycline <i>tetW</i>	2.14E-05	3.54E-05	1.94E-05	0	0	0	0	0	0	✓	✓			

A: abundance of ARGs; P: percentage of source; AVG( $A_{3-10\text{cm}}$ ): average abundance of ARGs in 3-10 cm.

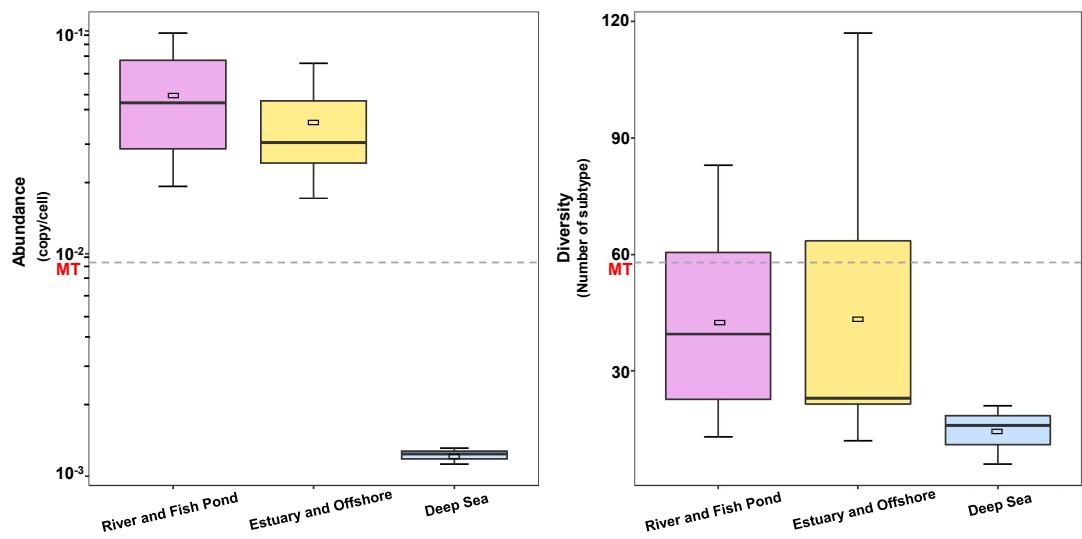
The highlighted ARGs (*cfxA2*, *ermF* and *mefA*) were identified as biomarkers considering their 1) significant increase in the surface layer, 2) relatively high abundance, and 3) high anthropogenic origins.



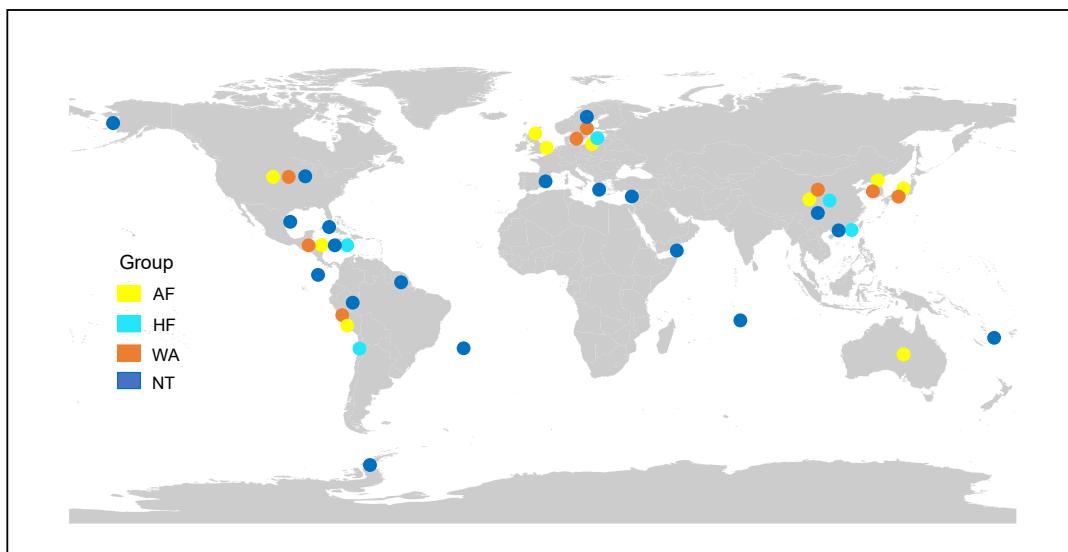
**Figure S1** Experimental pipeline of sample collection, pretreatment and DNA extraction.



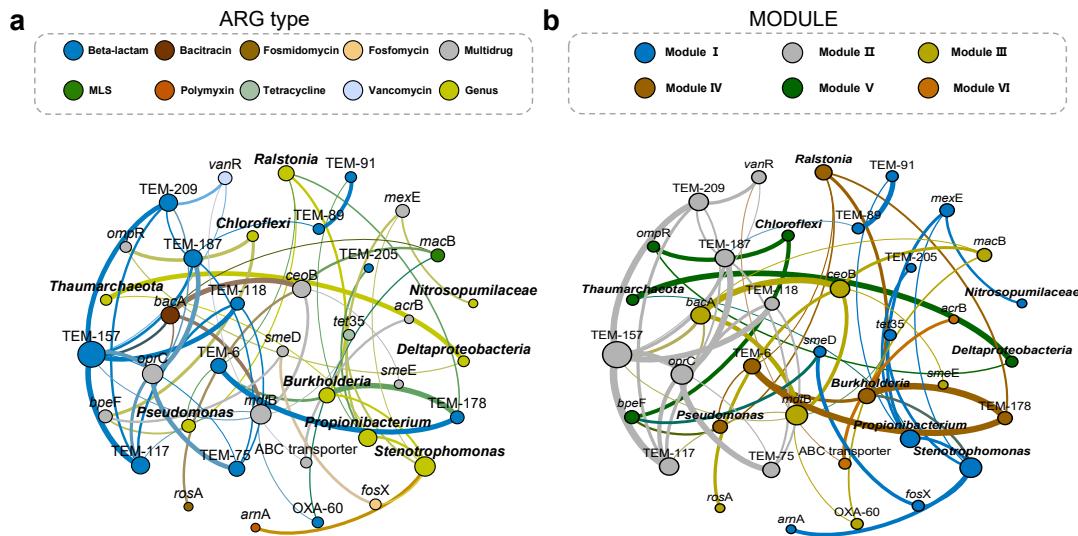
**Figure S2 Rarefaction curves of ARG subtypes. (a)** The rarefaction curves of each sample. **(b)** The rarefaction curve of all samples.



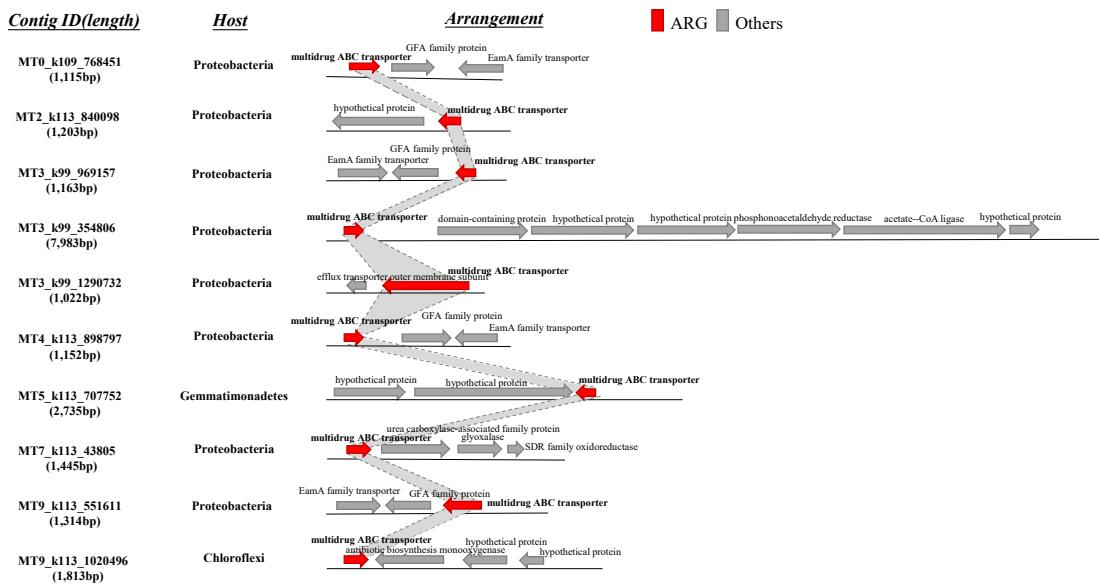
**Figure S3** The distribution of ARG abundance and diversity of sediment metagenomic datasets (river and fish pond, estuary and offshore, and deep sea) downloaded from NCBI database. Boxplots show the maximum, minimum, 1st quartile, 3rd quartile, median and mean abundances and diversities of ARGs in metagenomic datasets.



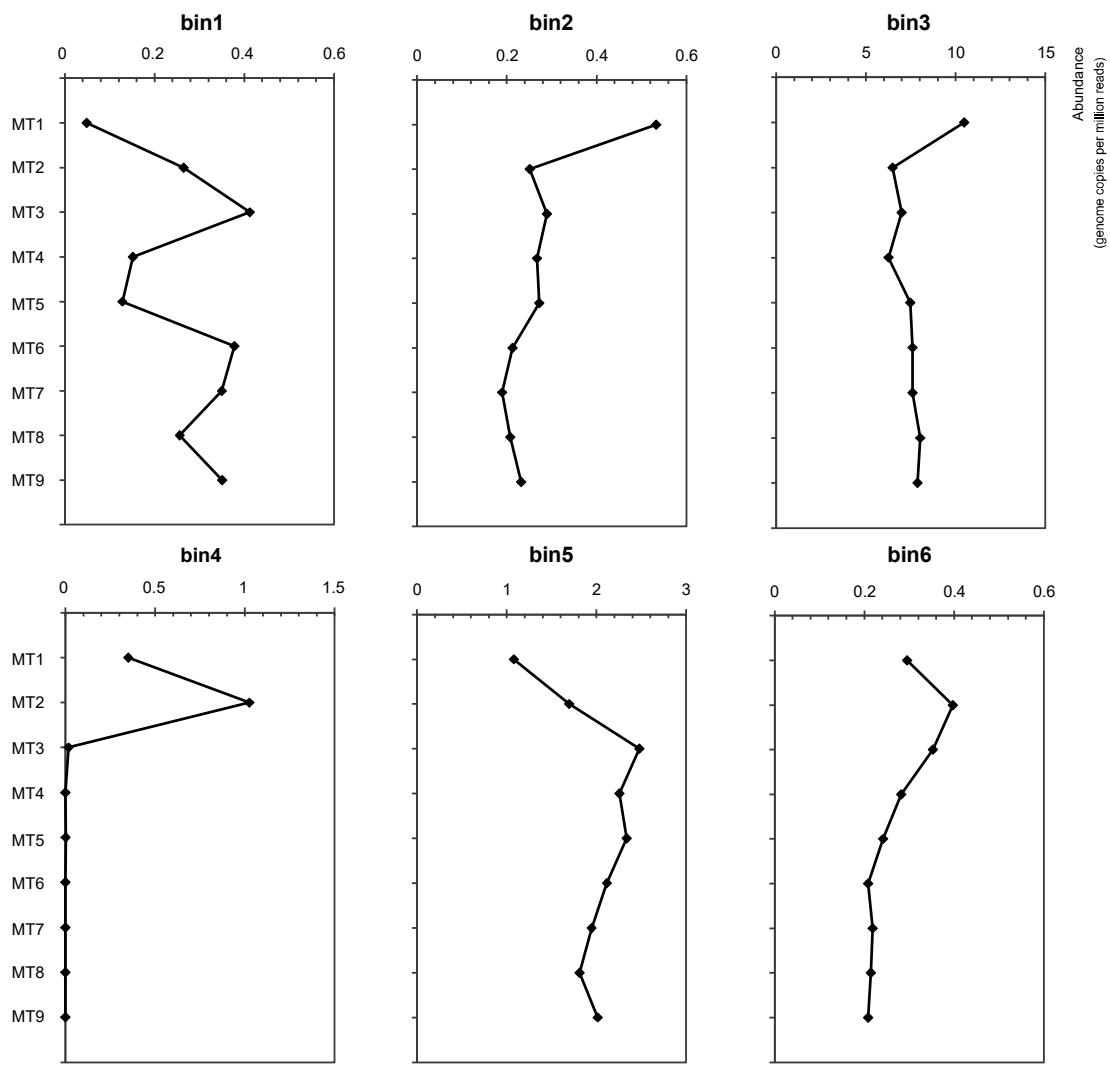
**Figure S4** Geographic distribution of 584 data sets divided into four groups (AF, HF, WA and NT) covering 6 continents.



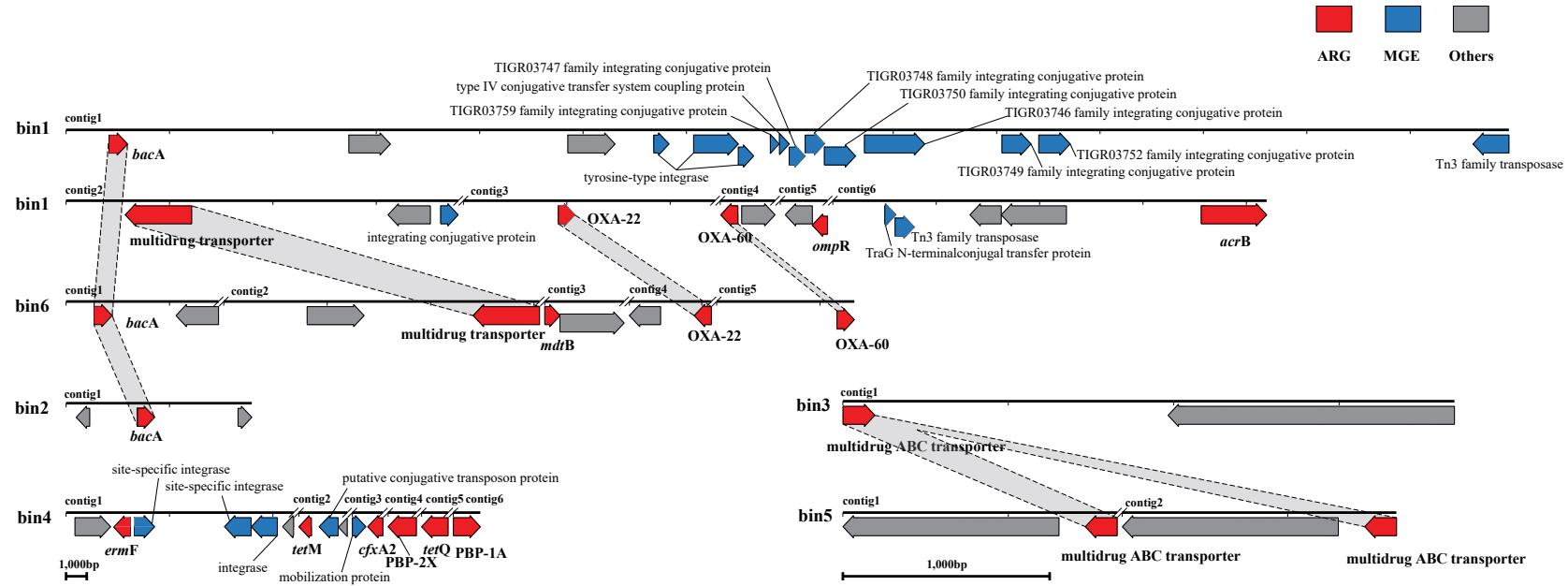
**Figure S5** Network analysis of ARGs and bacterial genera with sample occurrence  $\geq 50\%$  and Spearman's correlation coefficient  $r \geq 0.7$  ( $P$ -value  $\leq 0.05$ ). **(a)** Colored by ARG types and genus; **(b)** Colored by module (modularity = 0.606, 6 modules).



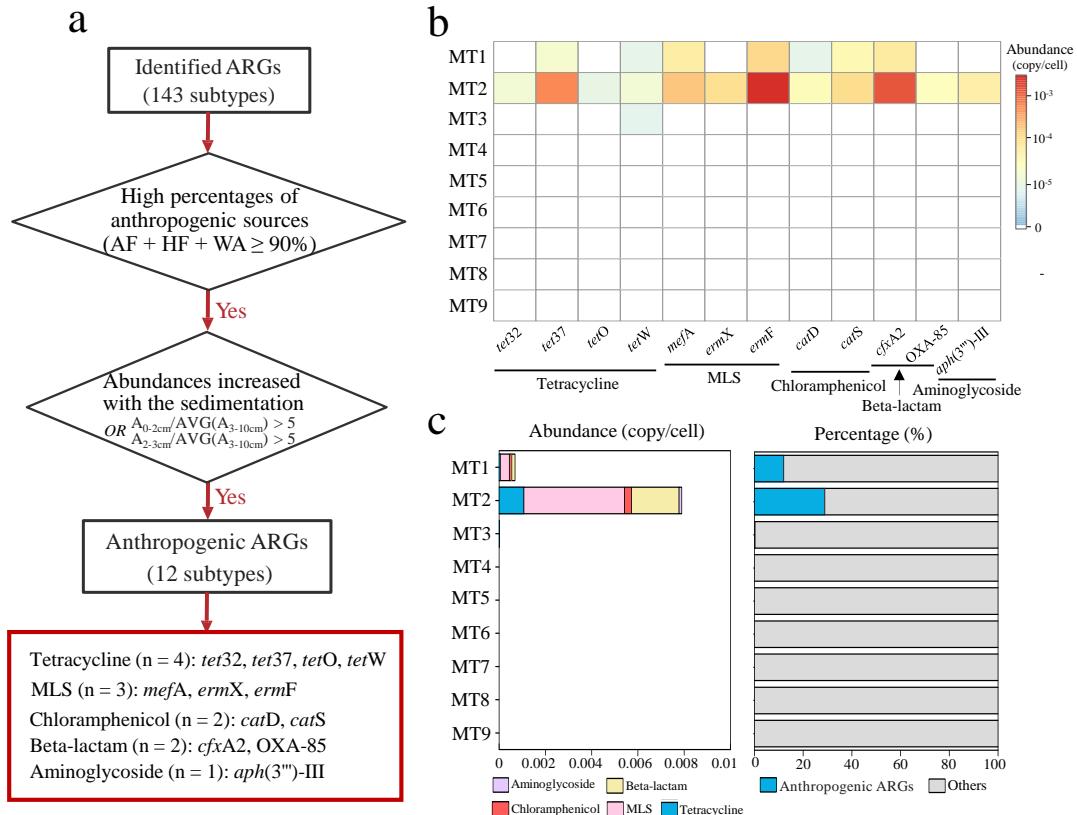
**Figure S6** Shared resistome and bacterial hosts of ARG-carrying contigs.



**Figure S7** Abundance shifts of recovered ARG-carrying genomes across depths.



**Figure S8** The shared resistome of six retrieved ARG-carrying genomes.



**Figure S9 Identification of anthropogenic ARGs.** **(a)** Identification pipeline and criteria of anthropogenic ARGs.  $AVG(A_{3-10cm})$ : average abundance of ARGs in 3-10 cm. **(b)** Abundance heatmap of anthropogenic ARGs in each depth of MT sediments. **(c)** Total abundance and percentage of anthropogenic ARGs. MLS: macrolide-lincosamide-streptogramin.

## References:

1. Yin, X. L.; Jiang, X. T.; Chai, B. L.; Li, L. G.; Yang, Y.; Cole, J. R.; Tiedje, J. M.; Zhang, T., ARGs-OAP v2.0 with an Expanded SARG Database and Hidden Markov Models for Enhancement Characterization and Quantification of Antibiotic Resistance Genes in Environmental Metagenomes. *Bioinformatics* **2018**, *34*, (13), 2263-2270.
2. Yang, Y.; Jiang, X. T.; Chai, B. L.; Ma, L. P.; Li, B.; Zhang, A. N.; Cole, J. R.; Tiedje, J. M.; Zhang, T., ARGs-OAP: Online Analysis Pipeline for Antibiotic Resistance Genes Detection from Metagenomic Data Using an Integrated Structured ARG-Database. *Bioinformatics* **2016**, *32*, (15), 2346-2351.
3. Ma, L. P.; Li, B.; Jiang, X. T.; Wang, Y. L.; Xia, Y.; Li, A. D.; Zhang, T., Catalogue of Antibiotic Resistome and Host-Tracking in Drinking Water Deciphered by a Large Scale Survey. *Microbiome* **2017**, *5*.
4. DeSantis, T. Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E. L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G. L., Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl Environ Microb* **2006**, *72*, (7), 5069-5072.
5. Angly, F. E.; Dennis, P. G.; Skarszewski, A.; Vanwonterghem, I.; Hugenholtz, P.; Tyson, G. W., CopyRighter: A Rapid Tool for Improving the Accuracy of Microbial Community Profiles Through Lineage-Specific Gene Copy Number Correction. *Microbiome* **2014**, *2*.
6. Truong, D. T.; Franzosa, E. A.; Tickle, T. L.; Scholz, M.; Weingart, G.; Pasolli, E.; Tett, A.; Huttenhower, C.; Segata, N., MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling. *Nat Methods* **2015**, *12*, (10), 902-903.
7. Bastian, M.; Heymann, S.; Jacomy, M. J. I., Gephi: An Open Source Software for Exploring and Manipulating Networks. **2009**, *8*, (2009), 361-362.
8. Knights, D.; Kuczynski, J.; Charlson, E. S.; Zaneveld, J.; Mozer, M. C.; Collman, R. G.; Bushman, F. D.; Knight, R.; Kelley, S. T., Bayesian Community-wide Culture-independent Microbial Source Tracking. *Nat Methods* **2011**, *8*, (9), 761-U107.
9. Li, L. G.; Yin, X. L.; Zhang, T., Tracking Antibiotic Resistance Gene Pollution from Different Sources Using Machine-Learning Classification. *Microbiome* **2018**, *6*.
10. Ma, L.; Li, B.; Zhang, T., Abundant Rifampin Resistance Genes and Significant Correlations of Antibiotic Resistance Genes and Plasmids in Various Environments Revealed by Metagenomic Analysis. *Appl Microbiol Biotechnol* **2014**, *98*, (11), 5195-204.
11. Huber, D. H.; Ugwuanyi, I. R.; Malkaram, S. A.; Montenegro-Garcia, N. A.; Noundou, V. L.; Chavarria-Palma, J. E., Metagenome Sequences of Sediment from a Recovering Industrialized Appalachian River in West Virginia. *Microbiol Resour Ann* **2018**, *6*, (18).

12. Chen, B. W.; Yang, Y.; Liang, X. M.; Yu, K.; Zhang, T.; Li, X. D., Metagenomic Profiles of Antibiotic Resistance Genes (ARGs) between Human Impacted Estuary and Deep Ocean Sediments. *Environ Sci Technol* **2013**, *47*, (22), 12753-12760.
13. Marshall, I. P. G.; Karst, S. M.; Nielsen, P. H.; Jorgensen, B. B., Metagenomes from Deep Baltic Sea Sediments Reveal How Past and Present Environmental Conditions Determine Microbial Community Composition. *Mar Genom* **2018**, *37*, 58-68.
14. Chen, J. R.; McIlroy, S. E.; Archana, A.; Baker, D. M.; Panagiotou, G., A Pollution Gradient Contributes to the Taxonomic, Functional, and Resistome Diversity of Microbial Communities in Marine Sediments. *Microbiome* **2019**, *7*.
15. Dong, X. Y.; Greening, C.; Rattray, J. E.; Chakraborty, A.; Chuvochina, M.; Mayumi, D.; Dolfing, J.; Li, C.; Brooks, J. M.; Bernard, B. B.; Groves, R. A.; Lewis, I. A.; Hubert, C. R. J., Metabolic Potential of Uncultured Bacteria and Archaea Associated with Petroleum Seepage in Deep-Sea Sediments. *Nat Commun* **2019**, *10*.
16. Song, D. L.; Zhang, Y. H.; Liu, J.; Zhong, H. H.; Zheng, Y. F.; Zhou, S.; Yu, M.; Todd, J. D.; Zhang, X. H., Metagenomic Insights Into the Cycling of Dimethylsulfoniopropionate and Related Molecules in the Eastern China Marginal Seas. *Front Microbiol* **2020**, *11*.
17. Hirai, M.; Nishi, S.; Tsuda, M.; Sunamura, M.; Takaki, Y.; Nunoura, T., Library Construction from Subnanogram DNA for Pelagic Sea Water and Deep-Sea Sediments. *Microbes Environ* **2017**, *32*, (4), 336-343.
18. Backstrom, D.; Yutin, N.; Jorgensen, S. L.; Dharamshi, J.; Homa, F.; Zaremba-Niedwiedzka, K.; Spang, A.; Wolf, Y. I.; Koonin, E. V.; Ettema, T. J. G., Virus Genomes from Deep Sea Sediments Expand the Ocean Megavirome and Support Independent Origins of Viral Gigantism. *Mbio* **2019**, *10*, (2).
19. Peoples, L. M.; Grammatopoulou, E.; Pombrol, M.; Xu, X. X.; Osuntokun, O.; Blanton, J.; Allen, E. E.; Nunnally, C. C.; Drazen, J. C.; Mayor, D. J.; Bartlett, D. H., Microbial Community Diversity Within Sediments From Two Geographically Separated Hadal Trenches. *Front Microbiol* **2019**, *10*.