

ORIGINAL ARTICLE

Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes

Bing Li¹, Ying Yang¹, Liping Ma¹, Feng Ju¹, Feng Guo¹, James M Tiedje² and Tong Zhang¹
¹Environmental Biotechnology Laboratory, The University of Hong Kong, Hong Kong and ²Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI, USA

A metagenomic approach and network analysis was used to investigate the wide-spectrum profiles of antibiotic resistance genes (ARGs) and their co-occurrence patterns in 50 samples from 10 typical environments. In total, 260 ARG subtypes belonging to 18 ARG types were detected with an abundance range of 5.4×10^{-6} – 2.2×10^{-1} copy of ARG per copy of 16S-rRNA gene. The trend of the total ARG abundances in environments matched well with the levels of anthropogenic impacts on these environments. From the less impacted environments to the seriously impacted environments, the total ARG abundances increased up to three orders of magnitude, that is, from 3.2×10^{-3} to 3.1×10^0 copy of ARG per copy of 16S-rRNA gene. The abundant ARGs were associated with aminoglycoside, bacitracin, β -lactam, chloramphenicol, macrolide-lincosamide-streptogramin, quinolone, sulphonamide and tetracycline, in agreement with the antibiotics extensively used in human medicine or veterinary medicine/promoters. The widespread occurrences and abundance variation trend of vancomycin resistance genes in different environments might imply the spread of vancomycin resistance genes because of the selective pressure resulting from vancomycin use. The simultaneous enrichment of 12 ARG types in adult chicken faeces suggests the coselection of multiple ARGs in this production system. Non-metric multidimensional scaling analysis revealed that samples belonging to the same environment generally possessed similar ARG compositions. Based on the co-occurrence pattern revealed by network analysis, *tetM* and aminoglycoside resistance protein, the hubs of the ARG network, are proposed to be indicators to quantitatively estimate the abundance of 23 other co-occurring ARG subtypes by power functions.

The ISME Journal (2015) 9, 2490–2502; doi:10.1038/ismej.2015.59; published online 28 April 2015

Introduction

Antibiotic resistance was identified as a major health security challenge of the twenty-first century in the 2013 G8 Science Ministers Statement (<https://www.gov.uk/government/publications/g8-science-ministers-statement-london-12-june-2013>). It is not just a regional or national phenomenon but a global problem, indicated by the famously typical cases of the rapid dissemination of *Klebsiella pneumoniae* carbapenemase-positive bacteria and New Delhi metallo- β -lactamase-positive bacteria in Asia, Europe and North America (McKenna, 2013). Soil, sediment, surface water, sewage, sludge and animal waste have been considered important reservoirs for antibiotic resistance genes (ARGs) because abundant ARGs have been frequently detected in these environments (LaPara *et al.*, 2011; Zhang and Zhang, 2011;

Koczura *et al.*, 2012; Burch *et al.*, 2013; Zhu *et al.*, 2013; Czekalski *et al.*, 2014). Strikingly, recent studies revealed that the exchange of ARGs between bacteria from farm animals/soils and clinical pathogens occurred via horizontal gene transfer (Smillie *et al.*, 2011; Forsberg *et al.*, 2012). This phenomenon emphasizes the clinical importance of the environmental bacteria, that is, the transmission of ARGs from natural environments to clinic establishes the natural resistome as a potential direct source of pathogenic resistance genes (Forsberg *et al.*, 2012). Consequently, the in-depth investigation of the diversity and abundance of ARGs in various environments is central to establishing the overall picture that is essential for management decision frameworks for controlling antibiotic resistance.

Currently, the global monitoring efforts including the European Antimicrobial Resistance Surveillance Network (EARS-Net) (<http://www.ecdc.europa.eu/en/activities/surveillance/EARS-Net>) and the US National Antimicrobial Resistance Monitoring System for Enteric Bacteria (<http://www.cdc.gov/narms/>) mainly focus on the antibiotic consumption and

Correspondence: T Zhang, Environmental Biotechnology Laboratory, The University of Hong Kong, Pokfulam Road, Hong Kong 999077, Hong Kong.

E-mail: zhangt@hku.hk

Received 19 September 2014; revised 6 December 2014; accepted 19 March 2015; published online 28 April 2015

antibiotic resistance isolates in clinical and public health laboratories, whereas ARGs, an emerging environmental pollutant, are not included in these surveillance systems (Grundmann *et al.*, 2011). One of the dominant reasons is the lack of a rapid, universal and accurate analysis method for the broad-spectrum detection and quantification of ARGs in environmental samples. Various molecular technologies, such as PCR, quantitative PCR (qPCR) and DNA microarray approaches have been commonly used to determine the occurrence/fate of environmental ARGs and valuable insights have been gained. However, there are several limitations of amplification-based methods (PCR and qPCR), including low-throughput, limited availability of primers, amplification bias, false-negative results due to inhibition in PCR and false-positive results due to nonspecific amplification. High-capacity quantitative PCR arrays have been applied to detect ARGs in manure, compost and soil (Looft *et al.*, 2012; Zhu *et al.*, 2013) to overcome the capacity limitations; however, it cannot overcome the other drawbacks mentioned above. DNA microarray is a genomic analysis method that can simultaneously detect a large number of ARGs in a single assay (Zhang *et al.*, 2009); the possibility for cross-talk of different probes coupled with low-sensitivity restricts its applications for comprehensive surveys of ARGs in complicated environmental samples (Yang *et al.*, 2013).

High-throughput sequencing (HTS)-based metagenomic analysis is a powerful tool that could overcome the drawbacks of the above methods (Schmieder and Edwards, 2012) if the sequencing depth and analysis tools are suitable. A novel 'environmental ARG diagnostic approach', that is, a metagenomic analysis method using the Structured Non-redundant Clean Antibiotic Resistance Genes Database (SNC-ARDB) and customized scripts, has been developed in our previous study to facilitate the detection and quantification of a broad-spectrum profile of ARGs (Yang *et al.*, 2013). Combined with network analysis tools, which have been widely used to explore the interactions/associations among entities, such as the species in a food web (Krause *et al.*, 2003), proteins in metabolic pathways (Guimerà and Amaral, 2005), coexisting patterns among microbial taxa in soils (Barberán *et al.*, 2012), activated sludge (Ju *et al.*, 2014) and human gastrointestinal tract (Zhang *et al.*, 2014), we could also assess the co-occurrence patterns among ARGs in complex environmental samples across spatial gradients. Previous study indicated that ARG composition correlated with microbial phylogenetic and taxonomic structure both across and within soil types, that is, bacterial community composition was the primary determinant of soil ARG contents (Forsberg *et al.*, 2014); therefore, the co-occurrence patterns between ARGs and microbial taxa in multiple environments could help exploring the association of the bacteria and ARGs.

The objectives of this study were (1) to conduct a more comprehensive profiling of ARG diversity and abundance in 50 environmental samples including water, soils, sediments, sludge, biofilm and faeces using the HTS-based metagenomic analysis; (2) to evaluate the similarity/difference of ARG compositions among different environmental samples using non-metric multidimensional scaling (NMDS) analysis; and (3) to identify several specific ARG subtypes as the indicators for ARG contamination based on the ARG co-occurrence patterns obtained using network analysis.

Materials and methods

Sampling and data sources

Basic information on the 50 samples in this study is summarized in Supplementary Table S1 (also see Supplementary Information), covering various typical environments, including 13 water samples (sewage, swine wastewater sample, treated wastewater, river water and drinking water), three soils, three sediments, one wastewater biofilm, 18 sludge samples (activated sludge and anaerobic digestion sludge) and 12 faecal samples (human, chicken and pig). Among these 50 samples, three soil and two human faeces samples were downloaded from MG-RAST. Eighteen data sets, including eight AS samples (Yang *et al.*, 2013), three sediments, one ADS sample, one biofilm sample (Ma *et al.*, 2014), three river water samples and two tap water samples (Chao *et al.*, 2013), have been used in our previous studies of ARGs in a single environment compartment. The detailed sample collection procedures are described in Supplementary Information S1.

DNA extraction and HTS

DNA extraction and concentration determination are described in Supplementary Information S2. Beijing Genomics Institute (BGI) provided shotgun library construction and Illumina HTS on HiSeq2000 for 45 DNA samples (6 µg of DNA for each sample). The base-calling pipeline (Version Illumina Pipeline-0.3) was used to process the raw fluorescence images and call sequences (Qin *et al.*, 2010). The entire data set is 187 Gb (giga base pairs), which is the largest sequence data set reported so far on the study of ARGs in environmental samples.

Bioinformatics analysis

Data filtration was performed to guarantee the quality of the downstream analysis (Supplementary Information S3). Subsequently, all the metagenomic sequencing data were searched for ARGs against the SNC-ARDB using BLASTX with E-value $\leq 1 \times 10^{-5}$ (Yang *et al.*, 2013). A sequence was annotated as an ARG-like fragment if its best hit in the SNC-ARDB had $\geq 90\%$ sequence identity and the alignment length was ≥ 25 amino acids (Kristiansson *et al.*, 2011).

The high accuracy for positive hits (>99.5%) of the metagenomic analysis approach using these similarity/alignment length cutoffs has been validated in our previous study (Yang *et al.*, 2013). To save sorting time and avoid human errors in ARG-like sequence classification, we developed a package of customized scripts that could automatically sort the ARG-like sequences obtained from the BLAST results into different types and subtypes of ARGs. Additionally, the number of ARG-like sequences in each subtype can be counted automatically. There are 25 'ARG types', 618 'ARG subtypes' and 2998 non-redundant reference sequences in the SNC-ARDB. For example, 'aminoglycoside resistance genes' is one example of an 'ARG type', whereas 'aadA' is one of the aminoglycoside resistance subtypes and there are three reference sequences belonging to *aadA*. It should be pointed out that the methodology used in the present study would only detect ARGs that have been annotated in SNC-ARDB. Some novel types of ARGs present in the samples might be missed since the analysis is based on a similarity search. Additionally, SNC-ARDB contains a number of efflux proteins (e.g., *acrA* and *acrB* and so on). Although they are present in both antibiotic-susceptible and antibiotic-resistant bacteria and cannot be good markers of resistance phenotype, they were usually related to efflux of antibiotic and thus were classified as ARGs (genotype) in previous references (Mikolosko *et al.*, 2006; Szczepanowski *et al.*, 2009; Nesme *et al.*, 2014) and another commonly used database, The Comprehensive Antibiotic Resistance Database (CARD, <http://arpcard.mcmaster.ca/?q=CARD/ontology/36298>). Therefore, efflux pump-related ARGs were still included in the SNC-ARDB to evaluate the antibiotic resistance potential.

MetaPhlAn was applied to conduct taxonomic classification and quantify the corresponding relative abundance (in terms of the number of cells rather than fraction of reads) by mapping metagenomic reads against a catalogue of clade-specific marker sequences currently spanning the bacterial and archaeal phylogenies (Segata *et al.*, 2012).

Statistical analysis and network analysis

In our previous study (Yang *et al.*, 2013), the portion of types or subtypes of ARG-like sequences in the 'total metagenome sequences' were defined as the 'abundance' ('p.p.m.', one read in one million reads). However, this calculation did not consider the impact of the sequence length of the ARG reference sequences on the final abundance results and may have biases, especially when comparing the ARG-like sequence abundance among different ARG subtypes or types that have different gene lengths. In the SNC-ARDB, the ARG reference sequences range widely from 186 to 4728 bp (Supplementary Table S2). To avoid the bias, normalization by the ARG reference sequence length was conducted in

this study. Additionally, normalization by the 16S-rRNA gene sequence length was also conducted, and ARG 'abundance' was expressed as 'copy of ARG per copy of 16S-rRNA gene' (thereafter called 'ratio'), which is the same as that of qPCR results reported in most previous literature. Therefore, the abundance obtained using the metagenomic analysis approach could be directly compared with those acquired via qPCR in other studies. The 'abundance' of the ARG type or subtype was calculated using the following equation:

$$\text{Abundance} = \sum_1^n \frac{N_{\text{ARG-like sequence}} \times L_{\text{reads}} / L_{\text{ARG reference sequence}}}{N_{16\text{S sequence}} \times L_{\text{reads}} / L_{16\text{S sequence}}} \quad (1)$$

where $N_{\text{ARG-like sequence}}$ is the number of the ARG-like sequence annotated as one specific ARG reference sequence; $L_{\text{ARG reference sequence}}$ is the sequence length of the corresponding specific ARG reference sequence; $N_{16\text{S sequence}}$ is the number of the 16S sequence identified from the metagenomic data; $L_{16\text{S sequence}}$ is the average length of the 16S sequence in the Greengenes database, which was used as the reference database for the 16S sequence identification via the local BLAST approach (Albertsen *et al.*, 2013), that is, 1432 bp was used in Equation (1); n is the number of the mapped ARG reference sequence belonging to the ARG type or subtype; L_{reads} is the sequence length of the Illumina reads (100 nt) or 454 pyrosequencing reads (200 nt) that was used in the present study.

NMDS was performed using the abundance correlation matrix of the ARG subtypes. Additionally, the Mann-Whitney test was implemented to compare whether the medians of the ARG abundances among various environments are significantly different (Hu *et al.*, 2013). All statistical analyses were performed by PAleontological STatistics software (version 2.15).

To visualize the correlations in the network interface, we constructed a correlation matrix by calculating all possible pairwise Spearman's rank correlations between the 84 ARG subtypes that occurred in at least 16 samples out of all environmental samples in the present study (41 samples excluding biological duplicates) (Steele *et al.*, 2011). This preliminary filtering step removed those poorly represented ARG subtypes that occurred in a limited number of samples and thus reduced the artificial association bias. A correlation between two items was considered statistically robust if the Spearman's correlation coefficient (ρ) was >0.8 and the P -value was <0.01 (Junker and Schreiber, 2008). To reduce the chances of obtaining false-positive results, the P -values were adjusted with a multiple testing correction using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). The robust pairwise correlations of the ARG subtypes formed their co-occurrence networks. Network analyses were performed in R environment using VEGAN (Oksanen *et al.*, 2007), igraph (Csárdi and Nepusz,

2006) and Hmisc (Harrell and Frank, 2008) packages. Network visualization was conducted on the interactive platform of Gephi (Bastian *et al.*, 2009).

Results and discussion

Broad-spectrum profile of ARG abundances in various environments

The average relative standard deviation of ARG abundance at total ARG, ARG type and subtype levels were 8.6%, 15.6%, and 21.4%, respectively which indicate that the HTS-based metagenomic approach was reproducible for ARG quantification (Supplementary Figure S1). Additionally, the sequencing depth of these 50 data sets was sufficient to characterize the ARG profiles at the subtype level (Supplementary Figure S2). In total, 18 of all 25 ARG types included in the SNC-ARDB were detected in at least one of the 50 samples (Figure 1a). The abundances of different ARG types in the samples varied greatly, from a 5.4×10^{-6} ratio (acridine resistance genes in 'Pigfarm.STP.Influent') to a 1.0×10^0 ratio (multidrug resistance genes in 'Faeces-Chicken-80d-b'). In general, the resistance genes for aminoglycoside, bacitracin, β -lactam, chloramphenicol, macrolide-lincosamide-streptogramin (MLS), multidrug, quinolone, sulphonamide and tetracycline were more abundant and commonly distributed than the other ARG types in these samples. As expected, these abundant ARGs were usually associated with the antibiotics used extensively as human medicine or veterinary medicine including growth promotion.

As shown in Figure 1b, the 50 samples from 10 typical environments, including river water, drinking water, STP influent, STP effluent, activated sludge and biofilm, anaerobic digestion sludge, soil, sediment, human faeces and faeces and wastewater from livestock farms, could be clustered into four groups according to their total ARG abundance levels. The Mann–Whitney test indicated that total ARG abundances were significantly different (P -value: <0.005 ; Supplementary Table S3) among these four groups and followed the order of Group I $<$ Group II $<$ Group III $<$ Group IV with successive increments of ~ 0.5 – 1 orders of magnitude. The ARG abundances in faeces and wastewater from livestock farms in Group IV (5.4×10^{-1} – 3.1×10^0 ratio) were 1–3 orders of magnitude higher than those of samples in Group I, such as sediment (4.0×10^{-3} – 3.0×10^{-2} ratio), soil (1.6×10^{-2} – 1.8×10^{-2} ratio), river water (1.7×10^{-2} – 3.1×10^{-2} ratio) and drinking water (1.2×10^{-2} – 4.7×10^{-2} ratio). These data further verified the conclusion that livestock farms were hotspots for ARGs (Zhu *et al.*, 2013). The ARG abundances in Group II (STP AS and BF, STP ADS and STP effluents) and Group III (STP influents and human faeces) were within the ratio range of 2.7×10^{-2} – 2.2×10^{-1} and 2.4×10^{-1} – 4.4×10^{-1} , respectively. With regard to a specific ARG type,

tetracycline resistance genes also exhibited a similar abundance trend with total ARG abundance among these four groups (P -value: <0.01). It should be noted that these four groups were representative of the typical environments affected by the anthropogenic activities and antibiotic selection pressures at different levels, from the slightly impacted Group I to the most seriously impacted Group IV. The ARG abundance variations closely matched the levels of anthropogenic impact and, presumably, antibiotic selection pressures in the various environments. We should keep in mind that one possible reason resulting in the above grouping pattern might be because of database bias of SNC-ARDB. After all, the metagenomic analysis method using SNC-ARDB only obtain the 'broad-spectrum' profile rather than the 'full-spectrum' profile of ARGs. Some novel types of ARGs present in the sample might be missed as the analysis is based on a similarity search. Additionally, some types of ARGs might be underestimated owing to the integrity of the databases because not all known ARGs were included in SNC-ARDB.

Supplementary Figure S3 illustrates the ARG-type compositions in 10 different environmental types. For soils, river water and human faeces, the top 2 or 3 dominant ARG types belonging to bacitracin, multidrug, tetracycline, β -lactam or MLS resistance genes could account for over 80% of the total ARG abundances. However, for sediments, drinking water, environments related to STP (influent, effluents, AS and BF, ADS) and livestock farms (faeces and wastewater), the compositions of the ARG types mainly included the resistance genes of tetracycline, aminoglycoside, MLS, multidrug, β -lactam, chloramphenicol, sulphonamide, bacitracin and quinolone.

Valuable insights have been gained by PCR-based approaches to investigate a few groups of ARGs (Supplementary Table S4). However, these previous studies only covered a limited number of well-studied ARG types among the resistance genes of tetracycline, sulphonamide, β -lactam, vancomycin, chloramphenicol and aminoglycoside. Tetracycline resistance genes and sulphonamide resistance genes were the predominant target ARG types in 90% of previous studies (Supplementary Table S4). With regard to the ARG subtypes, <40 subtypes were covered in the ARGs detection list in previous studies (Supplementary Table S4). In this aspect, the current PCR-based approaches provided a mere snapshot of the ARG profiles in environmental samples. The metagenomic analysis method could be used to investigate ARGs across a broader spectrum without PCR bias and capture a more complete picture of the ARG profiles because 25 ARG types, consisting of 618 ARG subtypes, could be analysed simultaneously. Among these 618 ARG subtypes, 260 ARG subtypes were detected in our samples (Supplementary Figure S4), 1–2 orders of magnitude more ARG subtypes than those detected

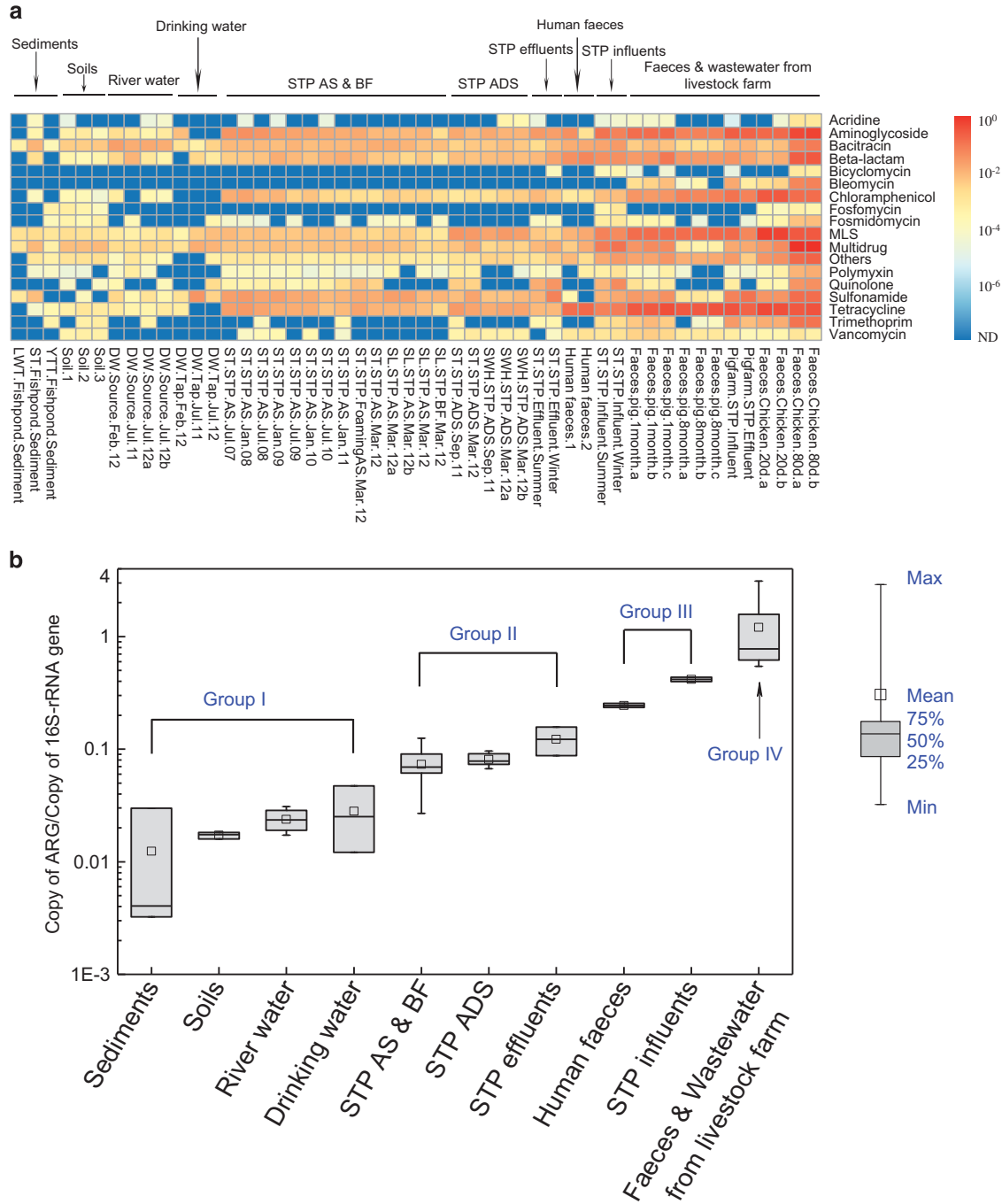


Figure 1 (a) Broad-spectrum quantitative profile of the ARG types (copy of ARG per copy of 16S-rRNA gene) in 50 environmental samples. (b) Comparison of total ARG abundance in different environments.

in similar environments reported previously (Supplementary Table S4). Our results indicate that, besides tetracycline resistance genes and sulphonamide resistance genes, the resistance genes of aminoglycoside, MLS, multidrug, β -lactam, chloramphenicol, bacitracin and quinolone were also very abundant in the different sampled environments. For example, bacitracin resistance genes in river water, multidrug/MLS/quinolone resistance genes in STP influents, MLS/chloramphenicol in

livestock faeces, and so on would likely be missed if only the small subset of the ARGs were assayed.

Widespread occurrence of vancomycin resistance genes
As the last line of defence against Gram-positive bacteria such as *Streptococcus pneumoniae* and *Enterococcus*, for which some strains are resistant to most other antibiotics, vancomycin has been prudently prescribed during the past several decades

(Jovetic *et al.*, 2010; McKenna, 2013). As shown in Figure 1, the abundance of vancomycin resistance genes in the environments was significantly lower than those of other resistance genes related to the widely used antibiotics (aminoglycoside, bacitracin, β -lactam, chloramphenicol, MLS, quinolone, sulphonamide and tetracycline). In total, 19 unique subtypes of vancomycin resistance genes were detected in the seven environments but not river water, drinking water and sediments (Supplementary Figure S4). Among these seven environments, the human faeces, STP influent and faeces and wastewater from livestock farms possessed the most abundant vancomycin resistance genes. Previous studies also indicated that a number of vancomycin resistance genes were detected in swine manure samples (Zhu *et al.*, 2013) and human faeces samples (Hu *et al.*, 2013). *vanRG* even ranked in the top 10 most abundant ARGs in the human faeces samples of Chinese, Spanish and Danish populations. We wish to emphasize that although vancomycin resistance genes have also been found in permafrost sediments, which were never affected by the anthropogenic activities before (D'Costa *et al.*, 2011), and the average abundance of vancomycin resistance genes in the samples of this study was relatively low (1.6×10^{-5} – 1.8×10^{-4} ratio), the findings of their widespread occurrences and abundance variation trend in the present study, that is, human faeces, STP influent and faeces and wastewater from livestock farms possessed the most abundant vancomycin resistance genes, whereas no detection in river water, drinking water or sediments still deserved more attention because they might imply the spread of vancomycin resistance genes because of the selective pressure resulting from vancomycin use.

The enrichment of ARG abundances in adult chicken faeces and the decrease of ARG abundances in adult pig faeces

ARG abundance increased significantly in the swine microbiome after 14 days of feeding antibiotics (Looft *et al.*, 2012), indicating that the selective pressure of non-therapeutic levels antibiotics greatly enhanced the abundance of ARGs. This conclusion was also supported by our ARG data from faeces of commercially grown chicks (20 days old) and adult chickens (80 days old). The abundance of ARGs increased from a ratio of 1.62–1.53 for chick faeces to 2.74–3.11 for adult chicken faeces. The ARGs with enrichment over 10-fold consisted of resistance genes for acridine, bacitracin, β -lactam, bleomycin, fosmidomycin, multidrug, polymyxin and quinolone (Supplementary Figure S5A). Apart from these eight ARG types, the resistance genes of aminoglycoside, sulphonamide, 'others' and trimethoprim also increased beyond from their already high background abundance ratios of 1.6×10^{-1} , 2.9×10^{-2} , 3.2×10^{-2} and 1.1×10^{-2} to even higher ratios of 4.7×10^{-1} , 1.2×10^{-1} , 1.8×10^{-1} and 1.0×10^{-1} ,

respectively. Although we did not analyse the types and concentrations of antibiotics existing in the chicken feed and faeces, judging by the practical situation on the livestock farm, there should not be so many types of antibiotics added into the chicken feed/water. The significant enrichment of 12 ARG types raises an interesting question as to why so many ARGs have increased. Similar results were also found in Looft's study (2012), that is, aminoglycoside O-phosphotransferases conferring resistance to aminoglycosides were markedly enriched in pig faeces even though aminoglycosides were not used. This finding suggests an indirect mechanism of selection (coselection) of multiple ARGs, perhaps by co-occurrence on mobile elements conferring resistance to the antibiotics fed (Looft *et al.*, 2012).

In contrast to the trend of ARG profiles in chick and adult chicken faeces, there was no enrichment of the ARGs in adult pig (8 months old) faeces compared with the piglet (1-month-old) faeces. Instead, the abundances of some ARG types, such as resistance genes of bleomycin, fosmidomycin, polymyxin, multidrug, sulphonamide and trimethoprim, decreased greatly and their abundances in adult pig faeces were less than one-tenth of the amounts in young pig faeces (Supplementary Figure S5B). In addition to these six ARG types, aminoglycoside resistance genes also decreased from their high background abundance ratio of 1.4×10^{-1} – 6.0×10^{-2} . This might be due to the practice of feeding more antibiotics to young pigs more susceptible to disease, and reducing or eliminating such feed amendments to older pigs (Zhou *et al.*, 2013; Jensen and Hayes, 2014).

Representative ARG subtypes in different environments

At the subtype level, 260 ARG subtypes were detected in these 50 samples at the abundance ratio range of 5.4×10^{-6} – 2.2×10^{-1} (Supplementary Figure S4). Among the 260 detected subtypes, the profile of 129 major ARG subtypes ($> 1.0 \times 10^{-3}$ ratio in at least one sample) was shown in Figure 2. The top 10% (referring to the ratio of ARGs subtype numbers) most abundant ARG subtypes in each environment, which were considered the representative ARGs, were summarized in Supplementary Table S5. Most of the representative ARGs in the environments of this study have not been previously reported, indicating there are new findings from the present study using the novel approach.

STP influents and STP effluents

In STP influents, the 20 representative ARG subtypes with the abundance ratio range of 5.7×10^{-3} – 3.2×10^{-2} belonged to nine types and accounted for 62.7% of the total ARG abundance in influents. Among these subtypes, *tetM*, *tetW*, *ermB* and *sul1* have also been frequently detected in STP influents of US (Gao *et al.*, 2012), China (Chen and Zhang, 2013)

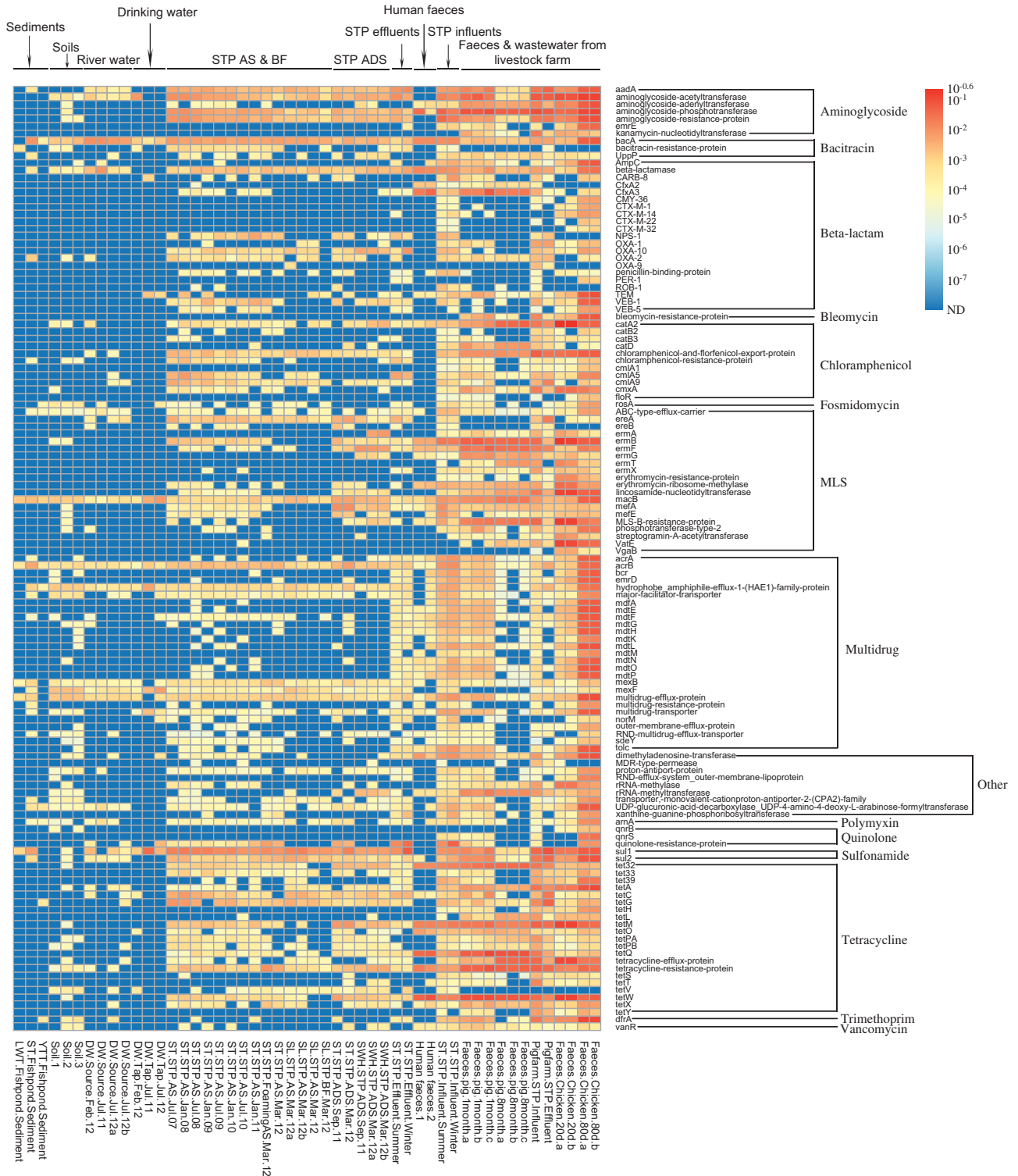


Figure 2 Abundance of the 129 major ($> 1.0 \times 10^{-3}$ ratio in at least one sample) ARG subtypes in the 50 environmental samples.

and Estonia (Nõlvak *et al.*, 2013) using qPCR methods, within the abundance ratio ranging from $\sim 2.0 \times 10^{-5}$ to $\sim 5.0 \times 10^{-2}$. In STP effluents, the 12 representative ARG subtypes with the abundance ratio of 1.9×10^{-3} – 3.2×10^{-2} belonged to nine types and contributed as much as 76.9% of the total ARG abundance in effluents.

The target ARGs detected in STP influents and effluents in previous studies only focused on genes resistant to tetracycline, sulphonamide, β -lactam and vancomycin (Supplementary Table S4). However, considering the results in the present study, many other ARG types not reported previously, such as those associated with resistance to aminoglycoside,

bacitracin, chloramphenicol, MLS, quinolone and multidrug, also occurred in influents and effluents with high abundances. This further highlighted the importance of broad-spectrum scanning of ARGs using the metagenomic approach.

The representative ARG subtypes in river water, drinking water, human faeces and faeces/wastewater from livestock farms are discussed in Supplementary Information S4.

Similarity analysis of ARG profiles in 50 environmental samples

The similarity of ARG compositions in the 50 environmental samples was evaluated using NMDS (Figure 3), which revealed that the grouping pattern was primarily influenced by the types of the environment with a few exceptions. Not surprisingly, samples from the same type of environment generally clustered more closely. For instance, river water, drinking water, STP influents, STP effluents, STP AS and BF, STP ADS, human faeces, soils and faeces and wastewater from livestock farms formed distinct clusters, respectively.

We hypothesized that there would be remarkable similarities of the ARG compositions between STP influent samples and human faeces samples, especially when the STP (like Shatin STP) mainly treats domestic wastewater without significant contributions from animal sources (like slaughterhouses and livestock farms). In other words, the ARG profiles in influents might reflect the average ARG abundance and diversity in the gastrointestinal tracts of urban residents of the STP catchment. The Venn diagram (Supplementary Figure S6) and the grouping pattern demonstrated by the NMDS (Figure 3) clearly confirmed this hypothesis. In total, 68 ARG subtypes belonging to 11 types were shared by the STP influents and human faeces with comparable abundances. The shared ARGs accounted for as much as 67.2% and 100% of the total ARG abundance in STP influents and human faeces, respectively (Supplementary Table S7 and Supplementary Figure S6). It should be noted that the STP influents were collected from Hong Kong, whereas the human faeces samples were collected from residents living in the United States (Turnbaugh *et al.*, 2009). More closely clustered pattern could be expected if the human faeces samples were collected from the residents who lived within the STP catchment area.

Moreover, the ARG profiles in pig faeces (sample codes 39–44 in Figure 3) and chicken faeces (sample codes 47–50 in Figure 3) could be used to represent the corresponding ARG abundance and diversity in pig and chicken gastrointestinal tracts, respectively. Notably, among the three types of animal faeces, the human faeces samples clustered more closely with the pig faeces samples, indicating a higher similarity between the ARG profiles of human gastrointestinal tracts and pig gastrointestinal tracts, consistent with their known digestive similarities.

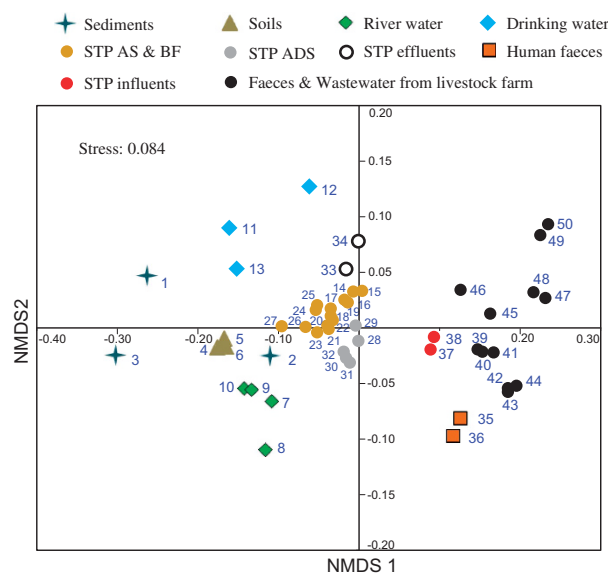


Figure 3 NMDS plot showing the ARG composition differences among the 50 environmental samples. The sample code in the NMDS plot and the corresponding sample names are listed in Supplementary Table S6.

Shared ARGs among human faeces, pig faeces and chicken faeces

Determining the shared ARGs between human and livestock faeces made it possible to compare directly the similarity of ARGs composition in gastrointestinal tracts of humans and farm animals. A total of 99 ARGs (at the reference sequences level) belonging to 10 types were shared by human and young livestock faeces (Figure 4 and Supplementary Table S8). The shared ARGs accounted for $30.5 \pm 0.2\%$, $71.8 \pm 4.4\%$ and $85.6 \pm 1.2\%$ of the total abundance of ARGs detected in chick faeces, piglet faeces and human faeces, respectively (Supplementary Table S9). Among these shared ARGs, the resistance genes of tetracycline, β -lactam, aminoglycoside and bacitracin were dominant in human faeces, whereas the resistance genes of tetracycline, MLS, aminoglycoside and multidrug were abundant in young livestock faeces (Supplementary Figure S7). Figure 4b further exhibited the detailed information on the types of shared ARGs and their abundance comparison among human faeces, piglet faeces and chick faeces. In Figure 4b, the percentage of one specific ARG in each faeces is equal to its corresponding abundance divided by the abundance sum of this ARG in the three types of faeces. Among these shared ARGs, the abundance of bacitracin resistance genes was much higher in human faeces than in young livestock faeces. However, MLS, tetracycline, multidrug, sulphonamide and aminoglycoside resistance genes were more abundant in young livestock faeces.

Seventy-seven ARGs (at the reference sequences level) belonging to 10 types were shared by humans and adult livestock (Supplementary Figure S8 and Supplementary Table S10). The abundance sum of these 77 shared ARGs contributed to $32.4 \pm 0.8\%$,

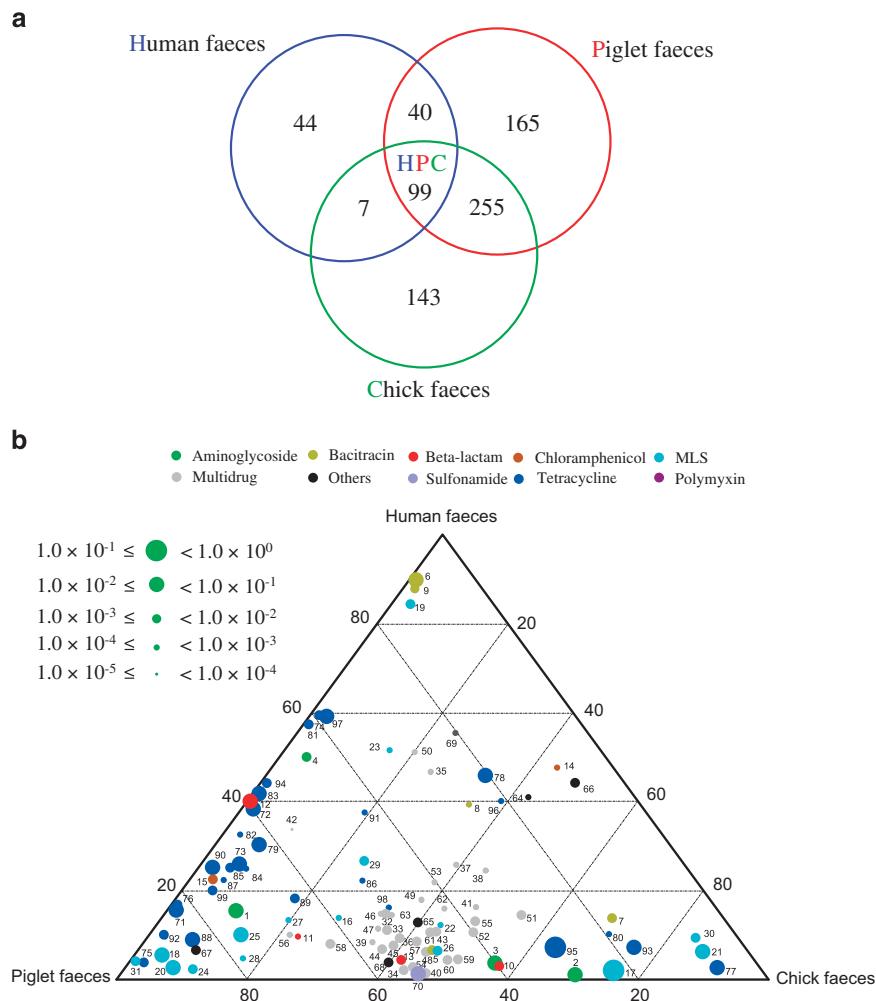


Figure 4 (a) Venn diagram showing the number of shared and unique ARGs (at the reference sequences level) among human gut, piglet gut and chick gut. (b) Ternary plot showing the abundance comparison of the 99 shared ARGs in human gut, piglet gut and chick gut. The sum of the abundance for one specific ARG in these three types of gut was set as 100%. In the ternary plot, the percentage (%) of one specific ARG in each gut is equal to its corresponding abundance divided by the abundance sum of this ARG in the three types of gut. The symbol size indicated the abundance of the ARGs.

75.2 ± 0.8% and 88.9 ± 0.6% of the total ARGs occurring in adult chicken faeces, adult pig faeces and human faeces, respectively (Supplementary Table S11). Among these shared ARGs, the resistance genes of tetracycline, β -lactam and aminoglycoside were dominant in human faeces, whereas the resistance genes of tetracycline, MLS and aminoglycoside were abundant in adult pig faeces (Supplementary Figure S9). Unlike the human faeces and adult pig faeces, multidrug resistance genes, instead of tetracycline resistance genes, were the most abundant shared ARGs in adult chicken faeces.

Co-occurrence patterns among ARG subtypes

The co-occurrence patterns among ARG subtypes were explored using network inference based on strong ($\rho > 0.8$) and significant (P -value < 0.01) correlations (Junker and Schreiber, 2008). Figure 5 consists of 46 nodes (ARG subtypes) and 98 edges. Certain topological properties widely used in

network analysis were calculated to describe the complex pattern of the interrelationships among the ARG subtypes (Supplementary Information S5). The modularity index of 0.472 suggested that the network had a modular structure (Newman, 2006). Based on the modularity class, the entire network could be parsed into eight major modules (i.e., clusters of nodes that interact more among themselves than with other nodes, compared with a random association), with 31 of 46 total vertices occupied by the two largest modules: Modules I and II. The most densely connected node in each module was defined as the 'hub' in the following statements. The co-occurring ARG subtypes of the module hubs were summarized in Supplementary Table S12.

The 'tetM' was the hub of Module I, whereas the 'aminoglycoside resistance protein' was the hub for Module II (Figure 5). One possible explanation for the hubs and related co-occurring ARGs in each module is that they might be harboured in some specific microbial taxa that are shared by different

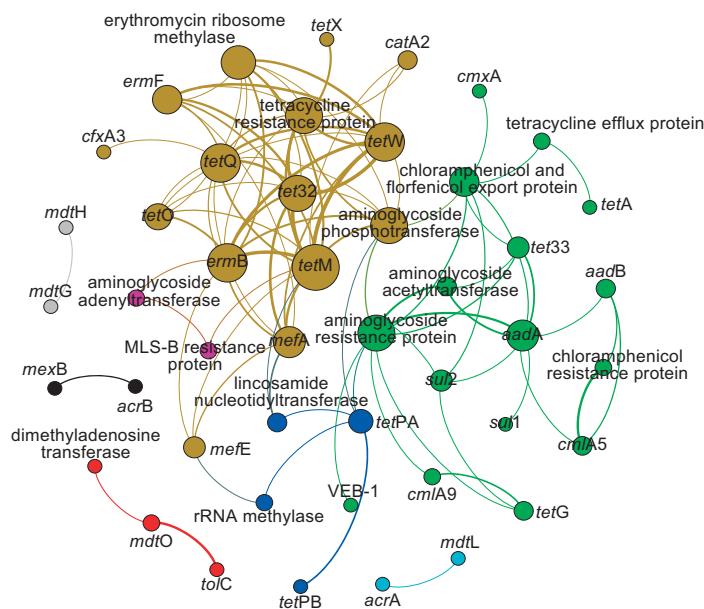


Figure 5 The network analysis revealing the co-occurrence patterns among ARG subtypes. The nodes were coloured according to modularity class. A connection represents a strong (Spearman’s correlation coefficient $\rho > 0.8$) and significant (P -value < 0.01) correlation. The size of each node is proportional to the number of connections, that is, the degree.

environments. On the other hand, the hubs could act as the ARG indicators to indicate the quantity of the correspondingly co-occurring ARGs. In other words, these two hubs could be used as representatives of 23 unique ARGs subtype (Supplementary Table S12). Supplementary Figure S10 illustrates the correlations between the hub abundance and the co-occurring ARG abundance. Further investigation indicated that the abundance of ‘hubs’ and the correspondingly co-occurring ARGs followed a power function with R^2 values ranging from 0.86 to 0.92 (Supplementary Table S13 and Supplementary Figure S11). Validation has been conducted to guarantee the applicability and accuracy of the model by comparing the model predicted value and the detected ARG abundance of co-occurring ARGs (Supplementary Figure S6 and Supplementary Table S14). The validation results imply that the abundance of co-occurring ARGs could be accurately estimated using the hub’s abundance via this power function. In addition, the high abundance of these two ARGs in various environments facilitates their applicability as ARG indicators (Figure 2). To indicate the abundance of the other 23 ARGs in samples, a simple and quick qPCR assay of ‘*tetM*’ and ‘aminoglycoside resistance protein’ as the target genes could be used. This will substantially save on the detection time and labour for monitoring these ARGs in multiple environmental samples. We suggest that this approach be more widely tried and tested.

Co-occurrence between ARG subtypes and microbial taxa

Comparison with microbial diversity of the 50 samples was summarized in Supplementary Table S15. As shown in Supplementary Figure S12, there was a significant Spearman’s rank correlation (Spearman’s $\rho = 0.58 \sim 0.61$, P -value $< 1.0E - 4$) between the microbial diversity and the ARGs diversity. The co-occurrence patterns between ARG subtypes and microbial taxa were also investigated using network analysis approach (Figure 6). Some topological properties of network analysis were summarized in section Supplementary Information S7. The detailed co-occurrence between ARGs subtype and microbial taxa were summarized in Supplementary Table S16. In the present study, it was hypothesized that the non-random co-occurrence patterns between ARGs and microbial taxa could indicate the possible host information of ARGs if the ARGs and the co-existed microbial taxa possessed the significantly similar abundance trends among the different environments (Spearman’s $\rho > 0.8$, P -value < 0.01). In other words, one of the reasonable explanations of the corresponding similar abundance trends was because of some specific microbial taxa carrying some specific ARGs, which has been verified by Forsberg’s study (2014).

As shown in Supplementary Table S16, five bacterial genera and one archaea genus were speculated as the possible ARGs host based on the co-occurrence results. For instance, *Blautia* was the

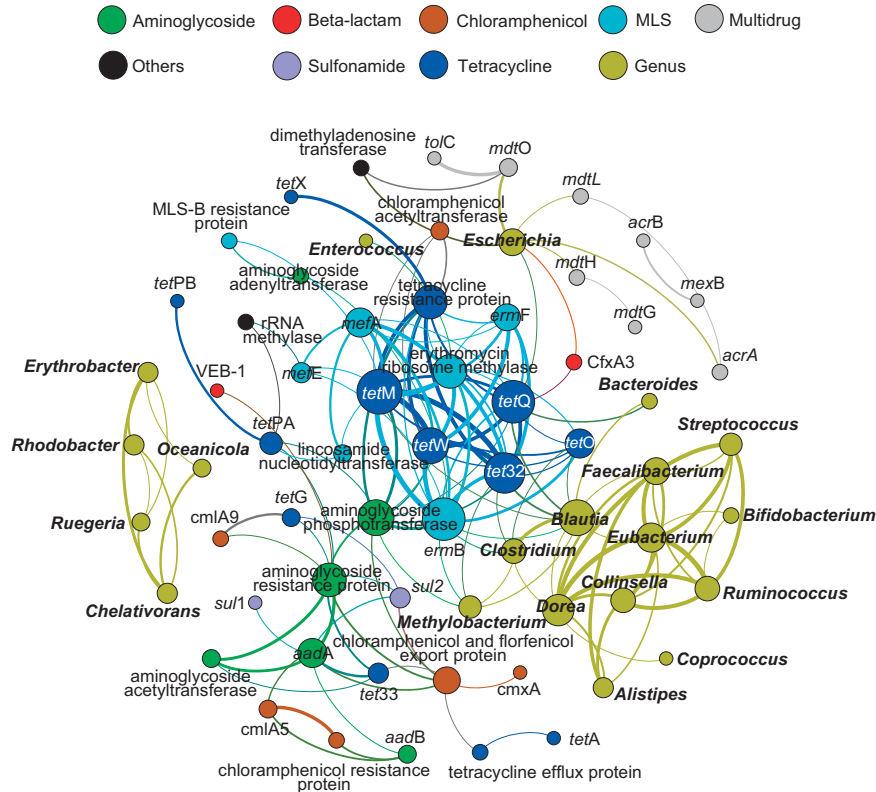


Figure 6 The Network analysis revealing the co-occurrence patterns between ARG subtypes and microbial taxa. The nodes were coloured according to ARG types and genus. A connection represents a strong (Spearman's correlation coefficient $\rho > 0.8$) and significant (P -value < 0.01) correlation. The size of each node is proportional to the number of connections, that is, degree.

host of tetracycline resistance genes (*tet32*, *tetM*, *tetQ* and *tetO*) and MLS resistance gene (*ermB*). Similar to *Blautia*, *Clostridium* also took along tetracycline resistance genes (*tet32* and *tetO*) and MLS resistance gene (*ermB*). *Enterococcus* was found to be the host of ARG subtypes of erythromycin ribosome methylase, whereas *Bacteroides* mainly carried *tetQ*. Compared with the above genera, *Escherichia* took along more diverse ARGs, including resistance genes of β -lactam (*cfxA3*), tetracycline (*tetQ*), multidrug (*acrA*, *mdtH*, *mdtL* and *mdtO*) and others (dimethyladenosine transferase). *Methanobrevibacter*, a typical methanogenic archaea genus (Daquiado *et al.*, 2014), mainly carries *tet32*, *ermB* and aminoglycoside phosphotransferase. A few of the ARG hosts have been verified in previous studies, which showed very consistent results (Supplementary Table S16). For instance, *tetQ* was commonly carried by *Escherichia*, and *Bacteroides* (Zhang *et al.*, 2009; Shoemaker *et al.*, 2001; Forslund *et al.*, 2013). Forslund *et al.* (2013) reported that both *Blautia* and *Clostridium* mainly harboured *tet32*, *tetO* and *ermB*, whereas *Escherichia* took along *acrA*, *mdtH*, *mdtL* and *mdtO*.

The high consistency between our results and the previous studies for the cases mentioned above indicated that the network analysis is a reasonable and powerful tool to provide us new insights into the ARGs and their possible hosts in complex

environmental examples. Such co-occurrence relationships revealed by network analysis need to be further validated using other approaches.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank the Hong Kong General Research Fund (HKU7201/11E) for financially supporting this study. Dr Bing Li and Dr Feng Guo thank The University of Hong Kong for the postdoctoral fellowship. Miss Ying Yang, Miss Liping Ma and Mr Feng Ju thank The University of Hong Kong for the postgraduate studentship.

References

- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**: 533–538.
- Barberán A, Bates ST, Casamayor EO, Fierer N (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J* **6**: 343–351.

- Bastian M, Heymann S, Jacomy M (2009). Gephi: an open source software for exploring and manipulating networks. *Proc Third Int ICWSM Conf (ICWSM)* **8**: 361–362.
- Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodological)* **57**: 289–300.
- Burch TR, Sadowsky MJ, LaPara TM (2013). Air-drying beds reduce the quantities of antibiotic resistance genes and class 1 integrons in residual municipal wastewater solids. *Environ Sci Technol* **47**: 9965–9971.
- Chao YQ, Ma LP, Yang Y, Ju F, Zhang XX, Wu WM *et al.* (2013). Metagenomic analysis reveals significant changes of microbial compositions and protective functions during drinking water treatment. *Sci Rep* **3**: 3550.
- Chen H, Zhang MM (2013). Effects of advanced treatment systems on the removal of antibiotic resistance genes in wastewater treatment plants from Hangzhou, China. *Environ Sci Technol* **47**: 8157–8163.
- Csárdi G, Nepusz T. (2006). The igraph software package for complex network research. *Int J Complex Syst* **1695**: 2006.
- Czekalski N, Gascon Diez E, Burgmann H (2014). Wastewater as a point source of antibiotic-resistance genes in the sediment of a freshwater lake. *ISME J* **8**: 1381–1390.
- D’Costa VM, King CE, Kalan L, Morar M, Sung WWL, Schwarz C *et al.* (2011). Antibiotic resistance is ancient. *Nature* **477**: 457–461.
- Daquiado ARP, Cho KM, Kim TY, Kim SC, Chang HH, Lee YB. (2014). Methanogenic archaea diversity in Hanwoo (*Bos taurus coreanae*) rumen fluid, rectal dung, and barn floor manure using a culture-independent method based on *mcrA* gene sequences. *Anaerobe* **27**: 77–81.
- Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MO, Dantas G (2012). The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**: 1107–1111.
- Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, Dantas G (2014). Bacterial phylogeny structures soil resistomes across habitats. *Nature* **509**: 612–616.
- Forslund K, Sunagawa S, Kultima JR, Mende DR, Arumugam M, Typas A *et al.* (2013). Country-specific antibiotic use practices impact the human gut resistome. *Genome Research* **23**: 1163–1169.
- Gao P, Munir M, Xagorarakis I (2012). Correlation of tetracycline and sulfonamide antibiotics with corresponding resistance genes and resistant bacteria in a conventional municipal wastewater treatment plant. *Sci Total Environ* **421–422**: 173–183.
- Grundmann H, Klugman KP, Walsh T, Ramon-Pardo P, Sigauque B, Khan W *et al.* (2011). A framework for global surveillance of antibiotic resistance. *Drug Resist Updates* **14**: 79–87.
- Guimerà R, Amaral LAN (2005). Functional cartography of complex metabolic networks. *Nature* **433**: 895–900.
- Harrell J, Frank E (2008). Hmisc: Harrell miscellaneous. R package version 3.5-2.
- Hu Y, Yang X, Qin J, Lu N, Cheng G, Wu N *et al.* (2013). Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nat Commun* **4**: 2151.
- Jensen HH, Hayes DJ. (2014). Impact of Denmark’s ban on antimicrobials for growth promotion. *Curr Opin Microbiol* **19**: 30–36.
- Jovetic S, Zhu Y, Marccone GL, Marinelli F, Tramper J (2010). β -Lactam and glycopeptide antibiotics: first and last line of defense? *Trends Biotechnol* **28**: 596–604.
- Ju F, Xia Y, Guo F, Wang Z, Zhang T (2014). Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed wastewater treatment plants. *Environ Microbiol* **16**: 2421–2432.
- Junker BH, Schreiber F. (2008). Correlation Networks. In *Analysis of Biological Networks*. Wiley-Interscience.
- Koczura R, Mokracka J, Jabłońska L, Gozdecka E, Kubek M, Kaznowski A (2012). Antimicrobial resistance of integron-harboring *Escherichia coli* isolates from clinical samples, wastewater treatment plant and river water. *Sci Total Environ* **414**: 680–685.
- Krause AE, Frank KA, Mason DM, Ulanowicz RE, Taylor WW (2003). Compartments revealed in food-web structure. *Nature* **426**: 282–285.
- Kristiansson E, Fick J, Janzon A, Grabic R, Rutgersson C, Weijdegård B *et al.* (2011). Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. *PLoS One* **6**: e17038.
- LaPara TM, Burch TR, McNamara PJ, Tan DT, Yan M, Eichmiller JJ (2011). Tertiary-treated municipal wastewater is a significant point source of antibiotic resistance genes into Duluth-Superior Harbor. *Environ Sci Technol* **45**: 9543–9549.
- Looft T, Johnson TA, Allen HK, Bayles DO, Alt DP, Stedtfield RD *et al.* (2012). In-feed antibiotic effects on the swine intestinal microbiome. *Proc Natl Acad Sci USA* **109**: 1691–1696.
- Ma LP, Li B, Zhang T (2014). Abundant rifampin resistance genes and significant correlations of antibiotic resistance genes and plasmids in various environments revealed by metagenomic analysis. *Appl Microbiol Biotechnol* **98**: 5195–5204.
- McKenna M (2013). Antibiotic resistance: the last resort. *Nature* **499**: 394–396.
- Mikolosko J, Bobyk K, Zgurskaya HI, Ghosh P (2006). Conformational flexibility in the multidrug efflux system protein AcrA. *Structure* **14**: 577–587.
- Nesme J, Cécillon S, Delmont TO, Monier JM, Vogel TM, Simonet P (2014). Large-scale metagenomic-based study of antibiotic resistance in the environment. *Curr Biol* **24**: 1096–1100.
- Nölvak H, Truu M, Tiirik K, Oopkaup K, Sildvee T, Kaasik A *et al.* (2013). Dynamics of antibiotic resistance genes and their relationships with system treatment efficiency in a horizontal subsurface flow constructed wetland. *Sci Total Environ* **461–462**: 636–644.
- Newman ME. (2006). Modularity and community structure in networks. *Proc Natl Acad Sci USA* **103**: 8577–8582.
- Oksanen J, Kindt R, Legendre P, O’Hara B, Simpson GL, Solyomos P *et al.* (2007). Vegan: community ecology package. R package version 1.8.
- Qin JJ, Li RQ, Raes J, Arumugam M, Burgdorf KS, Manichanh C *et al.* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**: 59–65.
- Schmieder R, Edwards R (2012). Insights into antibiotic resistance through metagenomic approaches. *Fut Microbiol* **7**: 73–89.

- Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **9**: 811–814.
- Shoemaker N, Vlamakis H, Hayes K, Salyers A. (2001). Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. *Appl Environ Microbiol* **67**: 561–568.
- Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480**: 241–244.
- Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY et al. (2011). Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J* **5**: 1414–1425.
- Szczepanowski R, Linke B, Krahn I, Gartemann KH, Gützkow T et al. (2009). Detection of 140 clinically relevant antibiotic-resistance genes in the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to selected antibiotics. *Microbiology* **155**: 2306–2319.
- Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JL. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* **1**: 6ra14.
- Yang Y, Li B, Ju F, Zhang T (2013). Exploring variation of antibiotic resistance genes in activated sludge over a four-year period through a metagenomic approach. *Environ Sci Technol* **47**: 10197–10205.
- Zhang XX, Zhang T (2011). Occurrence, abundance, and diversity of tetracycline resistance Genes in 15 sewage treatment plants across China and other global locations. *Environ Sci Technol* **45**: 2598–2604.
- Zhang XX, Zhang T, Fang HHP (2009). Antibiotic resistance genes in water environment. *Appl Microbiol Biotechnol* **82**: 397–414.
- Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X et al. (2014). Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *ISME J* **8**: 881–893.
- Zhou LJ, Ying GG, Zhang RQ, Liu S, Lai HJ, Chen ZF et al. (2013). Use patterns, excretion masses and contamination profiles of antibiotics in a typical swine farm, south China. *Environ Sci Proc Imp* **15**: 802–813.
- Zhu YG, Johnson TA, Su JQ, Qiao M, Guo GX, Stedtfeld RD et al. (2013). Diverse and abundant antibiotic resistance genes in Chinese swine farms. *Proc Natl Acad Sci USA* **110**: 3435–3440.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)