Abundance and diversity of resistomes differ between healthy human oral cavities and gut

Victoria R. Carr[1], Elizabeth Witherden[1], Sunjae Lee[1], Saeed Shoaie[1,2], Peter Mullany[3], Gordon B. Proctor[1], David Gomez-Cabrero[1,4,5], David L. Moyes[1]

[1] Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral and Craniofacial Sciences, King's College London, UK

[2] Science for Life Laboratory, KTH—Royal Institute of Technology, SE-171 21 Stockholm, Sweden

[3] Microbial Diseases, Eastman Dental Institute, University College London, UK

[4] Translational Bioinformatics Unit, NavarraBiomed, Departamento de Salud-Universidad Pública de Navarra, Pamplona, 31008, Navarra, Spain

[5] Unit for Computational Medicine, Karolinska Institutet (L8:05). Karolinska University Hospital, SE-171 76 Stockholm, Sweden.

Correspondence and requests for materials: victoria.carr@kcl.ac.uk, david.moyes@kcl.ac.uk

Keywords

resistome; antimicrobial resistance; microbiome; metagenomics; oral; gut

Abstract

The global threat of the "antimicrobial resistance apocalypse" that has arisen in recent years has driven the use of high-throughput sequencing techniques to monitor the profile of resistance genes in microbial populations. The human oral cavity contains a poorly explored reservoir of these genes, and little is known about their abundance and diversity, or how their profile compares with antimicrobial resistance genes in the gut. Here we analyse the resistome profiles of 790 oral cavities worldwide and compare these profiles with paired stool samples from shotgun metagenomic data. We find country-specific differences in the prevalence of antimicrobial resistance gene classes and mechanisms in oral and stool samples. Countries with a higher prevalence of resistance to antibiotic

classes relative to their use, contain genes resistant to those classes that co-localise with bacteriophages, suggesting the occurrence of horizontal gene transfer of these genes. Between individuals, the oral cavity contains a significantly higher abundance, but lower diversity of antimicrobial resistance genes compared to the gut, which is likely influenced by differences in microbial hosts and mobile genetic element associations. This is the first study to date that characterises the oral cavity resistome worldwide, identifying its distinctive signatures compared to the gut, and its role in the maintenance of antimicrobial resistance.

Introduction

Recent years have highlighted antimicrobial resistance (AMR) as one of the biggest threats to global health, food production and economic development[1]. Given this rapidly developing global crisis, it is imperative that the current gaps in our understanding of the distribution, spread and associations of all AMR factors is filled. AMR is most often conferred through the expression of antimicrobial resistance genes (ARGs) that reduce a microbe's susceptibility to the effects of an antimicrobial compound. As such, monitoring the abundance profiles of these ARGs, or "resistome" profile, has huge potential to increase our understanding of the spread and persistence of AMR within a population. Next-generation sequencing technologies are beginning to be used as tools for screening canonical and novel ARGs for potential surveillance of antimicrobial resistance worldwide. Shotgun metagenomic data mapped against dedicated ARG databases have begun to provide a wealth of insight into the resistome in the human[2–7] and animal gut[8,9] and the environment[10–13]. However, no such study on a large scale has, to date, attempted to characterise the resistome in the human oral cavity. Antimicrobial resistant microbes from the oral cavity have considerable potential to lead to antimicrobial resistant infections at other body sites. For example, ß-lactam, clindamycin, and erythromycin resistant strains of streptococci resident in the oral cavity can lead to infections at other body sites such as infective endocarditis[14].

There are unique ecological pressures on the oral microbial community, such as mechanical force, nutritional availability, pH levels, oxidative stress and redox potential. However, the oral microbial community has been shown to be relatively stable under these ever-changing conditions[15] and even after short-term antibiotic exposure[16]. In the oral cavity, horizontal gene transfer (HGT) has been documented as an important mechanism for the transfer and acquisition of ARGs within and between oral bacterial species[17,18]. The erythromycin resistance *mefA* and *mefE* genes have been found on the *MEGA* mobile genetic element associated with Tn*916*-like conjugative transposons (also called integrative conjugative elements ICE), and this has been implicated in conjugative transfer between viridans group streptococci (VGS) and other streptococci[19].

Metagenomic studies of the oral cavity indicate that this site potentially contains a diverse range of ARGs, including those encoding resistance to tetracycline, amoxycillin and gentamicin in saliva and plaque samples[20,21]. Oral ARGs appear to be natural feature of the human oral cavity, as the oral resistome of isolated Amerindian communities and ancient humans, with no known exposure to antibiotics, were found to contain aminoglycoside, ß-lactam, bacitracin, bacteriocin, macrolide, phenicol and tetracycline ARGs[22,23].

Here, we derived the oral and the gut resistomes from 790 and 386 shotgun metagenomes, respectively, from healthy individuals from China[24], Fiji[25], the Philippines[26], Western Europe[27–29] and the US[30]. We found country-specific differences in the proportion of saliva, dental and stool samples containing ARGs, ARG classes and mechanisms. Metagenomes with a higher prevalence of antibiotic classes relative to their use comprise of ARGs, resistant to those classes, that co-localise with bacteriophages, suggesting the occurrence of horizontal gene transfer. We made 410 comparisons of oral resistomes with paired gut resistomes derived from stool shotgun metagenomes

3

from the same individuals. As expected, there was a greater similarity in interpersonal resistome profiles between the same body sites, than intrapersonal profiles across different body sites. Further, the oral resistome contains a higher abundance but a lower diversity of ARGs, than the gut resistome. Finally, co-occurrence analyses between ARG and species abundance profiles, for saliva and stool from China and saliva from Philippines, that contain a higher incidence of these ARGs across individuals, show contrasting co-associations between saliva and stool, and predict the species from which an ARG is derived. It is imperative, however, to note that the detection and quantities of these ARGs are indicative of potential rather than conferred AMR[31]. Overall, these results demonstrate the requirement for wider AMR surveillance studies at different body sites, including the oral cavity, to understand the composition of the resistome across different human microbial habitats.

Results

Country-specific differences in ARG profiles from oral and stool metagenomes

To establish the incidence of ARGs in oral as well as stool metagenomes collected from various locations, metagenomes were mapped and quantified against the Comprehensive Antibiotic Resistance Database (CARD)[32]. Saliva samples were only available from China, Fiji, the Philippines and Western Europe. The percentages of saliva samples that contain at least one ARG for each class and mechanism from these cohorts were evaluated. To account for varying read depth across samples, the samples were subsampled to the same number of sequences. The percentages of saliva samples from China, Fiji, the Philippines and Western Europe contain ARGs from 17, 14, 27 and 17 classes, respectively (Fig. 1a). More ARG classes are found in Philippines saliva samples, but most of this variability originates from one individual: a farmer from Zambal who has aminocoumarin, carbapenem, diaminopyrimidine, fosfomycin, nitrofuran, nitroimidazole, rifamycin, and triclosan ARGs[26]. A high percentage (above 50 %) of saliva samples from all cohorts

4

contain cephamycin, fluoroquinolone, lincosamide, macrolide, pleuromutilin, streptogramin and tetracycline ARGs. Aminoglycoside and glycylcycline classes are found in mostly Chinese saliva samples, whilst cephamycin is found less in Chinese saliva samples compared to other cohorts. The peptide ARG class is only found in Chinese and Philippines saliva. Antibiotic efflux, inactivation, target alteration and target protection mechanisms are present in all samples across all cohorts (Fig. 1b), whilst the antibiotic target replacement mechanism is found in China, Philippines and Western Europe but not Fiji. The reduced permeability to antibiotic mechanism is only found in one individual from the Philippines (Fig. 1b). Dental samples were only available from China and the US. The percentages of the China and US dental samples containing at least one ARG class and mechanism were compared, and found to consist of 16 and 18 ARG classes, respectively (Fig. 1c). A greater percentage of China than US dental samples contain sulfonamide/sulfone ARGs. Similarly to saliva, a high percentage (above 50 %) of dental samples contain cephamycin, fluoroquinolone, glycylcycline, lincosamide, macrolide, pleuromutilin, streptogramin and tetracycline ARGs for all cohorts. Antibiotic efflux, inactivation, target alteration, target protection and antibiotic target replacement mechanisms are present in all samples across both cohorts (Fig. 1d). Stool samples were available from all locations, apart from the Philippines. Stool samples from China, Fiji, the US and Western Europe were found to contain 31, 31, 19 and 30 ARG classes, respectively (Fig. 1e). A high percentage of stool samples in most cohorts contain cephalosporins, cephamycins, diaminopyrimidines, macrolides, lincosamides, streptogramins and tetracyclines. Fiji samples have the lowest percentages of these ARG classes. In addition to the US containing the fewest number of ARG classes, the US also contains a low proportion of fluoroquinolones and phenicols that are found in a high proportion of stool samples from China, Fiji and Western Europe. Antibiotic efflux, inactivation, target alteration, target protection and antibiotic target replacement mechanisms are present in all samples across all cohorts, but the reduced permeability to antibiotic mechanism is only found in China, Fiji and Western Europe, and not the US (Fig. 1f).

5

In saliva, dental and stool samples, most of these highly prevalent ARG classes are represented by the presence of only a few ARGs. In particular, only one ARG, *APH(3')-Ia*, encapsulates the high prevalence of the aminoglycoside ARG class in saliva from China (Supplementary Fig. 1a), and only one ARG, *RlmA(II)*, represents the high proportion of the lincosamide and macrolide ARG classes in saliva and dental samples from all cohorts (Supplementary Fig. 1a, b). Again, the ARGs, tetQ, tetO and tetW comprise the majority of the tetracycline ARG class across all cohorts in stool samples (Supplementary Fig. 1c).

To evaluate whether the prevalent ARGs in the oral cavity are related to antibiotic use, the percentage of samples containing ARG classes were compared to the "Defined Daily Doses Per 1000 individuals" for the equivalent antimicrobial drug class for each region. Percentages of saliva, dental and stool samples containing ARG classes do not follow antibiotic use linearly for each location. In particular, there is a high percentage of samples containing aminoglycoside ARG classes in China, macrolide and tetracycline ARG classes in all sites from all locations, cephalosporin ARG classes in stool samples from all locations, and fluoroquinolone ARG classes in stool samples from China and in oral sites from all locations (Fig. 1g). Analyses of the predicted synteny between phages and ARGs in assembled metagenomes indicated that ARGs that are found in most samples or in particular body sites, and confer resistance to these classes, may have been previously acquired rapidly via HGT. For example, all samples containing the aforementioned *APH(3')-Ia* ARG in saliva samples from China are associated with *Enterobacteria* phage DE3 along with the tetracycline ARG *tet(C)* (Supplementary Fig. 2). Most dental and saliva samples across all locations contain macrolide ARGs, *mefA* and *mel*, that are associated with *Streptococcus* phage phiNJ3/phiSC070807 and often with *ermB*, which is likely to form the *MEGA* (macrolide efflux genetic assembly) element previous found in *Streptococcus* species[33–35]. In another instance,

6

aminoglycoside ARGs, in particular *APH(3')-IIIa*, *aad(6)*, and *AAC(6')-Ie-APH(2'')-Ia*, found in most stool samples from Western Europe and China (Supplementary Fig. 1c), are syntenic with *Escherichia* phage 1720a-02, *Staphylococcus* phage SPbeta-like, and *Streptococcus* phage phiSC070807/phi-SsUD.1/phiJH1301-2 (Supplementary Fig. 2).

Another possibility is that antibiotic use may be underestimated in China and the Philippines especially, as antibiotics are more readily available across-counter in these countries as well as via prescriptions and are found as containments in water and food supplies[36–38]. Antibiotics are also used more widely in husbandry and the fishing industry in China, the Philippines and the US, whereas the European Union banned the use of antimicrobials as animal growth promoters in 2006[39,40]. Although antimicrobial resistant microbes and ARGs can transfer from animals and the environment to humans, the impact of antimicrobial use in agriculture and livestock farming on AMR incidence and prevalence in humans is poorly understood[41].

Comparisons between oral and gut resistomes reveal differences in ARG composition, abundance and diversity

To determine whether there are differences in ARG composition between oral and gut samples, the separation of the ARG incidence profiles, i.e. resistotypes, for each sample were evaluated using principal coordinates analysis. Resistotypes were extracted using hierarchical clustering and silhouette analysis[42]. Five resistotypes in total were identified, where most samples are found in four out of the five resistotypes, and oral resistotypes clustered separately from stool resistotypes (Fig. 2a). Saliva from China and the Philippines have similar resistome profiles to those from US dorsum of tongue, buccal mucosa and dental samples in Resistotype R1, whereas resistome profiles from Fiji and Western Europe saliva cluster separately in Resistotype R3 (Fig. 2b). Similarly, there is a

7

177 strong separation of Resistotype R4 represented by mainly China and some Fiji stool samples, and
178 Resistotype R5 represented by mainly US and the remaining Fiji stool samples.
179
180 To further investigate the differences between oral and gut resistome profiles, the abundance and
181 diversity between oral and gut samples were evaluated and compared. The total abundance,
182 measured as the total Reads Per Kilobase Million (RPKM), of all ARGs in gut samples is lower
183 than in oral (buccal mucosa, dental dorsum of tongue and saliva) samples across all pairwise
184 comparisons in all cohorts (Fig. 3a). The overall abundance is similar between paired US oral sites,
185 with buccal mucosa and dental having a slightly higher abundance than the tongue dorsum samples
186 (Supplementary Fig. 3). Oral samples contain a higher abundance of fluoroquinolone, macrolide,
187 multidrug macrolide/lincosamide, and multidrug macrolide/streptogramin ARGs than stool samples
188 (Fig. 3b, Supplementary Fig. 4). These classes are mostly dominated by one of two ARGs across all
189 cohorts, such as *patB* and *pmrA* ARGs in the fluoroquinolone class (Supplementary Fig. 5a, b, c, d).
190 Stool contains a higher abundance of tetracyclines and "Other" ARGs less commonly found in all
191 samples across all cohorts and more cephamycin ARGs in stool than saliva, particular from Fiji
192 (Fig. 3b, Supplementary Fig. 4). These "Other" ARGs are most found in aminocoumarins,
193 aminoglycosides, cephalosporins, diaminopyrimidines, penams, penems and peptides across all
194 cohorts (Supplementary 5a, b, c, d). Unlike the higher abundance of a few ARGs in each class from
195 oral samples, stool samples contain a diverse mixture of ARGs for each class. Between oral samples
196 from the US, there are more fluoroquinolone, macrolide and lincosamide ARGs, but fewer
197 cephamycin ARGs in buccal mucosa, and more tetracycline ARGs in the tongue dorsum.
198
199 The availability of the longitudinal buccal mucosa, dental, tongue dorsum and stool samples from
200 the US over two years, from individuals who did not take antimicrobial agents, afforded us the
201 ability to investigate the stability of resistomes over time. Hierarchical clustering revealed that the

8

same individuals and sample types cluster together, verifying resistomes remain stable over a long period in both the oral cavity and gut with no antimicrobial use (Supplementary Fig. 6).

The abundance of ARGs were compared between sample types using differential analysis with DESeq2[43]. Stool samples contain more diverse ARGs with a significantly higher abundance than oral samples across all cohorts (Supplementary Fig. 7). Specifically, these include *CfxA* beta-lactamase family ARGs, the cephalosporin *CblA-1* ARG (found in *Bacteroides uniformis*, a common commensal in the gut microbiota[44]), the diaminopyrimidine *dfrF* ARG, MLS (macrolide-lincosamide-streptogramin) Erm 235 ribosomal RNA methyltransferase family ARGs, multidrug ARGs (all conferring resistance via multidrug efflux pumps), and tetracyclines *tet(40)*, *tet32*, *tetQ*, *tetO* and *tetW* ARGs. Oral samples have a higher abundance of ARGs than in stool samples, but fewer are significant compared to stool samples. In particular, macrolide ARGs, *macA* and *macB*, fluoroquinolone ARGs, *patB* and *pmrA*, and the macrolide/lincosamide ARG *RlmA(II)*, are found to be significantly more abundant in oral than stool samples across all cohorts. There are also differentially abundant ARGs between different sites in the oral cavity. For instance, between the US dorsum of tongue and dental samples, and between the US dorsum of tongue and buccal mucosa, all significantly abundant ARGs are found in the dorsum of tongue. Similarly, between dental and buccal mucosa, all significantly abundant ARGs are found in dental samples. Most of these ARGs are resistant to cephamycin, fluoroquinolone, MLS and tetracycline antibiotics. From China, saliva contains more significantly abundant ARGs than dental, in particular, resistance to aminoglycoside, cephalosporin, fluoroquinolone (specifically *pmrA* and *patB*), lincosamides, macrolides (specifically *macB* and *mefA* ARGs), MLS (in particular to the *Erm 235* ribosomal RNA methyltransferase family) and tetracycline antibiotics. Overall, stool samples contain ARGs conferring resistance to a greater diversity of ARG classes compared to oral samples, especially

9

multidrug ARGs that express efflux pumps. In comparison, oral samples contain ARGs, where each ARG confers resistance to no more than three antibiotic classes.

To investigate ARG diversity further, the ARG richness was evaluated between pairwise comparisons of sample types for each cohort. ARG richness is defined as the number of unique ARGs per sample. China, US and Fiji stool samples have a significantly higher ARG richness than paired China dental and saliva, US dental and buccal mucosa, and Fiji saliva samples (Mann-Whitney, paired, two-sided t-test, p-value < 0.05) (Fig. 4). This suggests stool samples contain a more diverse range of ARGs than some oral samples across most cohorts. No significant difference was found between saliva and stool samples from Western Europe. However, the US dorsum of tongue contains a significantly higher ARG richness than US stool. From pairwise comparisons between oral sites (Supplementary Fig. 8), China saliva has a greater ARG richness than dental samples (Mann-Whitney, paired, two-sided t-test, p-value < 0.05). US dorsum of tongue has a higher ARG richness than both dental and buccal mucosa, and dental also has a greater ARG richness than the buccal mucosa.

Oral and gut ARG profiles associate with species and strains

Spearman's correlation analysis between ARG and species/strain abundance was conducted to determine whether it is possible to predict the origin of ARGs. Only significant correlations were found in saliva and stool from China, and saliva from the Philippines. The *CfxA* beta-lactamase family, *tetQ*, *patB* and *pmrA* ARGs are all strongly correlated with specific species or strains found in both countries (Fig. 5a, b). Cephamycin ARGs, CfxA, *CfxA2, CfxA4* and *CfxA5* correlate with *Prevotella sp., Megasphaera micronuciformis* GCF 000165785 and *Selonomonas fluggei* GCF 000160695, and the tetracycline ARG, *tetQ*, co-associates with *Prevotella sp.* in both countries. In addition, fluoroquinolone ARGs, *patB* and *pmrA*, that are highly abundant in oral samples, co-occur

10

with *Gemella haemolysans* and *Streptococcus mitis/oralis/pneumoniae* in both countries (Fig. 5a, b). Also highly abundant in oral samples, macrolide ARG, *macB*, is associated with *Kingella denitrificans GCF 000190695*, *Neisseria sicca* and *N. flavescens* in China (Fig. 5a), and with *Neisseria subflava GCF 000173955*, *N. gonorrhoeae*, *Campylobacter rectus GCF 000174175* and *Aggregatibacter aphrophilus* in the Philippines (Fig. 5b). In contrast to China saliva, *Escherichia coli* in China stool is co-associated with many ARGs that code for multidrug efflux pumps (Supplementary Fig. 9). ARGs *E. coli ampC beta-lactamase*, *E.coli acrA* and *E. coli mdfA* co-occur with *E. coli, Bifidobacterium ileS* with *Bifidobacterium longum* and *Bifidobacterium adolescentis,* and *Klebsiella pneumoniae acrA* with *Klebsiella pnumoniae*, as expected. The multidrug ARGs, *oqxA* and *oqxB,* co-associate with *Klebsiella pneumoniae,* an intestinal commensal which can cause infections such as pneumonia, sepsis, meningitis and urinary tract infections[45]. These observations demonstrate that correlation analysis can potentially be applied as a predictor of ARG origin.

Discussion

This study provides insights into the ARG resistome profiles of different intraoral sites from healthy individuals across different geographical locations, and compares their composition to ARGs in the gut. At a population level, there are country-specific differences in the prevalence of ARGs, ARG classes and resistance mechanisms in saliva, dental plaque and gut. Phageome analysis of assembled metagenomes show ARG classes that are highly represented by these sites relative to the antibiotic class use, may consist of ARGs that are associated with phages, and have been acquired by horizontal gene transfer.

Within individuals, differential abundance analysis between sample types show higher abundances of ARGs in the oral cavity compared to the gut across all locations, especially for macrolide ARGs, *mel and mefA*, fluoroquinolone ARGs, *patB* and *pmrA*, and macrolide/lincosamide ARG *RlmA(II)*. Co-occurrence analysis of saliva samples from China and the Philippines predicts *patB* and *pmrA*

11

originate from *Gemella haemolysans* and *Streptococcus mitis/oralis/pneumoniae*, and phage analysis shows *mel* and *mefA* are located with *Streptococcus* phages phiNJ3 and phiSC070807 in saliva and dental samples.

However, most sites in the oral cavity (saliva, dental and buccal mucosa) have a lower diversity of ARGs than the gut. The sole exception is the dorsum of the tongue from the US. Sites in the oral cavity, including cheek, dental plaque and tongue, host microbial biofilm structures. It is believed the close proximity of microbes within oral biofilms can be a conducive environment for HGT of ARGs[46]. The unique papillary structure on the dorsum of the tongue acts as complex microbiological niche and favours the deposition of oral debris and microbes[47], which may explain a higher diversity of ARGs in the dorsum of the tongue compared to other oral sites. The higher abundance and lower diversity of ARGs in the oral cavity compared to stool samples may be due to several factors. One factor could be the difference in species resident in the gut compared to the oral cavity. As shown in ARG-species/strain correlation analysis of Chinese stool, *Escherichia coli* strains are predicted to contain a variety of ARGs, especially of the multidrug class, whereas species found in saliva were estimated to contain fewer ARGs. Due to stringent constraints of the correlation analysis, however, it was not possible to predict the origins for all ARGs found. Antibiotic use leading to acquisition of ARGs is another potential factor. Pharmacokinetics of orally administered antibiotics suggests that the oral cavity and oesophagus would be only briefly exposed to the antibiotic during swallowing, then the gut would subsequently be exposed, but over a more prolonged period as the antibiotics are held in the stomach lumen and intestine before being absorbed via the intestinal epithelium into the bloodstream. Therefore, microbes in the gut will be exposed to antibiotics for a longer period of time due to their increased bioavailability, which will result in higher antibiotic selection pressures than in the oral cavity[48]. The incidences where the oral

12

cavity is likely to acquire ARGs from selective pressures of antibiotics are from topical antibiotics for periodontal infections or orally-administered antibiotics being absorbed into the bloodstream.

The differences in ARG profiles across body sites has significant implications for the characterisation and interpretation of resistome analyses. Previous shotgun metagenomic studies have focussed almost exclusively on the resistome from the human gut[2–4,6]. While the gut may be a significant reservoir of ARGs, some of these may not be particularly prevalent with very little potential to drive resistant infections at other body sites. Further, the lower abundance of ARGs in the gut may be the result of some of the highly diverse ARG population only being present in species that are unable to transcribe them and are therefore irrelevant. Therefore, to test potential applications of non-cultured based, metagenomic AMR surveillance, future studies need to characterise the resistome of different body sites that have had different pharmacokinetic exposures to antimicrobials, and integrate this information with culture-based susceptibility tests, culturomics[49] and functional metagenomic screens[50].

Methods

Metagenomic sequence data

A total of 1176 publicly available metagenomic samples from six countries, including longitudinal US samples, were analysed. US samples after the first time point where excluded from the majority of the study to ensure each sample was independent, unless specified otherwise. These samples include 1) longitudinal data across two years with various timepoints from the Human Microbiome Project 1 (referred to as US)[30] containing buccal mucosa (n = 165: 32 with one, 35 with two, 19 with three and 1 with six timepoints); dorsum of tongue (n = 188: 22 with one, 43 with two, 24 with three and 2 with four timepoints); dental (n = 191: 23 with one, 43 with two, 20 with three, 1 with

13

four and 3 with six timepoints); stool (n = 156: 13 with one, 33 with two, 21 with three, 2 with four and 1 with six timepoints), 2) healthy control samples from a Chinese rheumatoid arthritis study[24] containing dental (n = 32); saliva (n = 33); stool (n = 72), 3) saliva (n = 137) and paired stool (n = 137) samples from Fiji[25], 4) saliva samples (n = 23) from healthy hunter-gatherers and traditional farmers from the Philippines[26], and 5) saliva (n = 21) and stool (n = 21) samples from Western Europe (5 saliva and 5 stool samples from Germany[27,29], and 16 saliva and 16 stool samples from France[28,29]).

Raw paired-end metagenomic reads from Chinese and Philippines samples were downloaded from the EBI (https://www.ebi.ac.uk/metagenomics/). Paired-end metagenomic samples from US were downloaded from https://portal.hmpdacc.org/. Raw paired-end metagenomic reads from Fiji (project accession PRJNA217052), France and Germany (project accession PRJEB28422) were downloaded from the NCBI. All US, China, Fiji and Philippines samples, and stool samples from France and Germany, were collected and sequenced as described in the following cited studies[24–28,30]. Saliva samples from France and Germany were collected and sequenced as described in the following cited study[29]. Metadata for the samples can be found in Supplementary Table 1.

Processing metagenomic data

The raw reads for all samples were trimmed using AlienTrimmer 0.4.0[51] with parameters *-k 10 -l 45 -m 5 -p 40 -q 20* and Illumina HiSeq and Ion Proton primers. Human contaminant sequences were removed from all samples by discarding reads that mapped against a human reference genome (downloaded from Human Genome Resources at NCBI on 27th February 2017) using Bowtie2 2.2.3[52] with parameters *-q -N 1 -k 1 --fr --end-to-end --phred33 --very-sensitive --no-discordant*. The quality of the raw reads and the filtered reads of each sample was evaluated using the Fastqc 0.11.3 (https://github.com/s-andrews/FastQC).

14

Extracting ARGs

All filtered, non-subsampled samples were mapped against *nucleotide_fasta_protein_homolog_model* from the antimicrobial resistance database CARD 2.0.0[32] using Bowtie2 2.2.5 with parameter *–very-sensitive*-local. Mapped reads were filtered from unmapped reads, sorted and indexed using Samtools 1.5[53]. Statistics for the length, coverage and number of reads mapped for each ARG were extracted using Bedtools 2.25.0[54]. ARGs with a coverage of reads mapped below and equal to 90 % were discarded. Each ARG was annotated with the "Drug Class" and "Resistance Mechanism" using CARD 2.0.3 metadata.

Percentage of samples containing ARGs, ARG classes and ARG mechanisms

To show whether the percentages of samples containing an ARG class were consistent across the same number of reads, metagenomes were first subsampled using seqtk 1.2 (https://github.com/lh3/seqtk) with parameter seed *-s100.* 6.9 million reads were subsampled from 18 saliva samples with the lowest number of reads greater than 6.9 million reads, from each cohort: China, Fiji, the Philippines and Western Europe.18 million reads were subsampled from 18 dental samples with the lowest number of reads greater than 18 million reads from both China and US cohorts. 16.9 million reads were subsampled from 18 stool samples with the lowest number of reads greater than 16.9 million reads from China, Fiji, the US and Western Europe cohorts. These were mapped to CARD 3.0.0 as described in *Extracting ARGs.*

R 3.5.1 was used for all downstream analysis. Percentages of samples containing an ARG, ARG class and mechanism were calculated from these samples. 95 % confidence intervals (CIs) were evaluated from percentages extracted from bootstrapping samples 100 times for each cohort and sample type. The percentages were also visualised against antibiotic use measured as the Defined Daily Dose Per 1000 individuals in 2015 from China, the Philippines, Western Europe (France and

15

Germany) and the US, from ResistanceMap (https://resistancemap.cddep.org/) accessed on 19th February 2019. No antibiotic use data was available for Fiji.

ARG-Phage analysis

The reads of the samples specified in "Percentage of samples containing ARGs, ARG classes and ARG mechanisms" were assembled using SPAdes 3.9.0 with parameters *-k 21,33,55 –only-assembler --meta*[55]. These assemblies were aligned against CARD 3.0.0 using BLASTN 2.7.1 with parameters *-outfmt 6 -evalue 1e-10*. Genes were predicted from the assemblies using Prodigal 2.6.3[56] with parameter *-p meta*, then predicted genes were aligned against the PATRIC virulence protein database[57] (downloaded 25th January 2019) using BLASTP 2.7.1 with parameter *-outfmt 6 -evalue 1e-10*. Contigs were filtered where a resident predicted gene aligned to at least one putative phage protein. All hits were filtered by an e-value <= 1e-50 and an identity >= 80. For query sequence hits that overlapped greater than 20% of the smallest query sequence hit, hits were identified that had the lowest e-value, or the highest identity for hits with the same e-values. Other hits were discarded. ARGs found on contigs that aligned against a putative phage protein were regarded as syntenic with the phage of that phage protein.

Principal coordinates analysis

Principal coordinates analysis was applied to the binary distance between ARG presence or absence profiles for each sample (excluding longitudinal US samples) using the vegan 2.5-2 package (https://cran.r-project.org/web/packages/vegan/index.html). Resistotypes were extracted using hierarchical clustering of the euclidean distance between principal coordinates with eigenvalues above zero. Silhouette analysis was used to determine the optimum number of resistotypes using the cluster 2.0.7.1 package (https://cran.r-project.org/web/packages/cluster/index.html). The number of resistotypes is defined by the number of clusters greater than two with the largest silhouette width.

16

400

401 **Abundance of ARGs**

402 To quantify the abundance of ARGs within each sample, the reads per kilobase of read per million
403 (RPKM) was calculated for every sample. The relative abundance of ARGs for each country and
404 sample type was calculated by dividing the RPKM by the sum of RPKM for each country and
405 sample type. The relative abundance of ARGs for each sample and sample type was calculated by
406 dividing the RPKM by the sum of the RPKM for each sample. Differential abundance of ARGs
407 between paired sample types from each country where calculated using the DESeq2 1.20.0
408 package[43] as recommended by Jonsson et al.[58].

409

410 **ARG diversity**

411 To ensure the ARG richness could be compared statistically across different sample types from the
412 same individuals[59], the metagenomes (excluding longitudinal US samples) were subsampled using
413 seqtk 1.2 with seed *-s100* . Paired samples from the same individuals in each of the following
414 groups containing two sample types were subsampled to a number rounded down by two significant
415 figures from the lowest number of reads in the group, apart from Fiji, where 1 million reads were
416 subsampled, due to a low number of reads for a few samples. Fiji samples originally containing
417 fewer than 1 million reads were excluded from the analysis.
418 *China dental vs. saliva:* 3.5 million reads were sampled from China dental (n = 30) and paired
419 saliva (n = 30) samples.
420 *China stool vs. saliva:* 3.5 million reads were sampled from China stool (n = 31) and paired saliva
421 (n = 31) samples.
422 *China stool vs. dental:* 14 million reads were sampled from China stool (n = 30) and paired dental
423 (n = 30) samples.

17

US buccal mucosa vs. dental: 690,000 reads were sampled from US buccal mucosa (n = 86) and paired dental (n = 86) samples.

US buccal mucosa vs. dorsum of tongue: 690,000 reads were sampled from US buccal mucosa (n = 86) and paired dorsum of tongue (n = 86) samples.

US buccal mucosa vs. stool: 700,000 reads were sampled from and US buccal mucosa (n = 64) and paired stool (n = 64) samples.

US dental vs. dorsum of tongue: 4.2 million reads were sampled from US dental (n = 89) and paired dorsum of tongue (n = 89) samples.

US dental vs. stool: 4.2 million reads were sampled from US dental (n = 68) and paired stool (n = 68) samples.

US dorsum of tongue vs. stool: 14 million reads were sampled from US dorsum of tongue (n = 69) and paired stool (n = 69) samples.

*Fiji saliva vs. stool:*1 million reads were sampled from Fiji saliva (n = 127) and paired stool (n = 127) samples.

Western Europe saliva vs. stool: 3.2 million reads were sampled from Western Europe saliva (n = 20: 4 from Germany and 16 from France) and paired stool (n = 20: 4 from Germany and 16 from France) samples. These were mapped to the CARD database as described in Methods *Percentage of samples containing ARG classes*.

Once the metagenomes were subsampled, ARGs extracted and filtering by coverage, the ARG diversity per sample was measured as the ARG richness, as recommended previously by Bengtsson-Palme et al.[60]. The ARG richness was calculated as the number of unique ARGs for every sample. The ARG richness between samples in each group were tested for statistical significance with a Mann-Whitney, paired, two-sided t-test.

Correlation analysis

18

MetaPhlAn2 2.6.0[61] was used to extract taxonomic composition from all samples. Spearman's correlation was applied to relative abundances of reads mapped to ARG and MetaPhlAn2 species profiles for paired samples. ARGs and species that were not found in more than half of samples for each country were removed, to alleviate the bias from potential joint-ranking of zero values by Spearman's rank. The rho and p-values were calculated using the stats package in R and the p-values were adjusted with Benjamini-Hochberg where FDR < 5 %. Correlations were found from China saliva and paired stool samples, and Philippines saliva samples. No significant correlations could be found from Fiji, Western Europe or US samples.

Data Availability

ARG data, figures, tables and script to run analysis are available at https://www.dropbox.com/sh/jy9h7ghgen6msgz/AABoMMRXY7BkBKdFY6kIpY3Aa?dl=0

High-resolution figures are available at https://figshare.com/s/a1295c1ece3a89769567

Code Availability

R package for resistome analysis is available at https://github.com/blue-moon22/resistomeAnalysis

References

1. *Antimicrobial resistance: global report on surveillance*. (World Health Organization, 2014).
2. Rose, G. *et al.* Antibiotic resistance potential of the healthy preterm infant gut microbiome. *PeerJ* **5**, e2928 (2017).
3. Feng, J. *et al.* Antibiotic Resistome in a large-scale healthy human gut microbiota deciphered by metagenomic and network analyses. *Environmental Microbiology* n/a-n/a doi:10.1111/1462-2920.14009
4. Forslund, K. *et al.* Country-specific antibiotic use practices impact the human gut resistome. *Genome Res.* **23**, 1163–1169 (2013).

19

5. Rahman, S. F., Olm, M. R., Morowitz, M. J. & Banfield, J. F. Machine Learning Leveraging Genomes from Metagenomes Identifies Influential Antibiotic Resistance Genes in the Infant Gut Microbiome. *mSystems* **3**, e00123-17 (2018).
6. Hu, Y. *et al.* Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nature Communications* **4**, 2151 (2013).
7. Pärnänen, K. *et al.* Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. *Nat Commun* **9**, (2018).
8. Thomas, M. *et al.* Metagenomic characterization of the effect of feed additives on the gut microbiome and antibiotic resistome of feedlot cattle. *Scientific Reports* **7**, 12257 (2017).
9. Munk, P. *et al.* Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. *Nature Microbiology* **3**, 898 (2018).
10. Pehrsson, E. C., Forsberg, K. J., Gibson, M. K., Ahmadi, S. & Dantas, G. Novel resistance functions uncovered using functional metagenomic investigations of resistance reservoirs. *Front. Microbiol.* **4**, (2013).
11. Ma, L. *et al.* Catalogue of antibiotic resistome and host-tracking in drinking water deciphered by a large scale survey. *Microbiome* **5**, 154 (2017).
12. Noyes, N. R. *et al.* Characterization of the resistome in manure, soil and wastewater from dairy and beef production systems. *Scientific Reports* **6**, 24645 (2016).
13. Bengtsson-Palme, J., Boulund, F., Fick, J., Kristiansson, E. & Larsson, D. G. J. Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India. *Front. Microbiol.* **5**, (2014).
14. Knoll, B., Tleyjeh, I. M., Steckelberg, J. M., Wilson, W. R. & Baddour, L. M. Infective Endocarditis Due to Penicillin-Resistant Viridans Group Streptococci. *Clin Infect Dis* **44**, 1585–1592 (2007).
15. Shaw, L. P., Smith, A. M. & Roberts, A. P. The oral microbiome. *Emerging Topics in Life Sciences* **1**, 287–296 (2017).
16. Zaura, E. *et al.* Same Exposure but Two Radically Different Responses to Antibiotics: Resilience of the Salivary Microbiome versus Long-Term Microbial Shifts in Feces. *mBio* **6**, e01693-15 (2015).

20

17. Roberts, A. P. & Kreth, J. The impact of horizontal gene transfer on the adaptive ability of the human oral microbiome. *Front Cell Infect Microbiol* **4**, (2014).
18. Witherden, E. A., Bajanca-Lavado, M. P., Tristram, S. G. & Nunes, A. Role of inter-species recombination of the ftsI gene in the dissemination of altered penicillin-binding-protein-3-mediated resistance in Haemophilus influenzae and Haemophilus haemolyticus. *J. Antimicrob. Chemother.* **69**, 1501–1509 (2014).
19. Chaffanel, F., Charron-Bourgoin, F., Libante, V., Leblond-Bourget, N. & Payot, S. Resistance Genes and Genetic Elements Associated with Antibiotic Resistance in Clinical and Commensal Isolates of Streptococcus salivarius. *Appl. Environ. Microbiol.* **81**, 4155–4163 (2015).
20. Diaz-Torres, M. L. *et al.* Determining the antibiotic resistance potential of the indigenous oral microbiota of humans using a metagenomic approach. *FEMS Microbiol Lett* **258**, 257–262 (2006).
21. Seville, L. A. *et al.* Distribution of tetracycline and erythromycin resistance genes among human oral and fecal metagenomic DNA. *Microb. Drug Resist.* **15**, 159–166 (2009).
22. Clemente, J. C. *et al.* The microbiome of uncontacted Amerindians. *Science Advances* **1**, e1500183 (2015).
23. Warinner, C. *et al.* Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics* **46**, 336–344 (2014).
24. Zhang, X. *et al.* The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nature Medicine* **21**, 895–905 (2015).
25. Brito, I. L. *et al.* Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435–439 (2016).
26. Lassalle, F. *et al.* Oral microbiomes from hunter-gatherers and traditional farmers reveal shifts in commensal balance and pathogen load linked to diet. *Mol. Ecol.* **27**, 182–195 (2018).
27. Voigt, A. Y. *et al.* Temporal and technical variability of human gut metagenomes. *Genome Biology* **16**, 73 (2015).
28. Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Molecular Systems Biology* **10**, 766 (2014).

29. Schmidt, T. S. B. *et al.* Extensive transmission of microbes along the gastrointestinal tract. *eLife* **8**, e42693 (2019).
30. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
31. Piddock, L. J. V. Assess drug-resistance phenotypes, not just genotypes. *Nature Microbiology* **1**, 16120 (2016).
32. McArthur, A. G. *et al.* The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* **57**, 3348–3357 (2013).
33. Ambrose, K. D., Nisbet, R. & Stephens, D. S. Macrolide efflux in Streptococcus pneumoniae is mediated by a dual efflux pump (mel and mef) and is erythromycin inducible. *Antimicrob. Agents Chemother.* **49**, 4203–4209 (2005).
34. Santagati, M., Lupo, A., Scillato, M., Di Martino, A. & Stefani, S. Conjugal mobilization of the mega element carrying mef(E) from Streptococcus salivarius to Streptococcus pneumoniae. *FEMS Microbiol Lett* **290**, 79–84 (2009).
35. Brenciani, A. *et al.* Phim46.1, the main Streptococcus pyogenes element carrying mef(A) and tet(O) genes. *Antimicrob. Agents Chemother.* **54**, 221–229 (2010).
36. Lansang, M. A. *et al.* Purchase of antibiotics without prescription in Manila, The Philippines. Inappropriate choices and doses. *Journal of Clinical Epidemiology* **43**, 61–67 (1990).
37. Chang, J. *et al.* Sale of antibiotics without a prescription at community pharmacies in urban China: a multicentre cross-sectional survey. *J. Antimicrob. Chemother.* **72**, 1235–1242 (2017).
38. Auta, A. *et al.* Global access to antibiotics without prescription in community pharmacies: A systematic review and meta-analysis. *Journal of Infection* (2018). doi:10.1016/j.jinf.2018.07.001
39. Liu, X., Steele, J. C. & Meng, X.-Z. Usage, residue, and human health risk of antibiotics in Chinese aquaculture: A review. *Environ. Pollut.* **223**, 161–169 (2017).
40. Cheng, G. *et al.* Antibiotic alternatives: the substitution of antibiotics in animal husbandry? *Front. Microbiol.* **5**, (2014).
41. Stanton, T. B. A call for antibiotic alternatives research. *Trends in Microbiology* **21**, 111–113 (2013).

42. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53–65 (1987).

43. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

44. Smith, C. J., Bennett, T. K. & Parker, A. C. Molecular and genetic analysis of the Bacteroides uniformis cephalosporinase gene, cblA, encoding the species-specific beta-lactamase. *Antimicrob. Agents Chemother.* **38**, 1711–1715 (1994).

45. Podschun, R. & Ullmann, U. Klebsiella spp. as Nosocomial Pathogens: Epidemiology, Taxonomy, Typing Methods, and Pathogenicity Factors. *Clinical Microbiology Reviews* **11**, 589–603 (1998).

46. Roberts, A. P. & Mullany, P. Oral biofilms: a reservoir of transferable, bacterial, antimicrobial resistance. *Expert Review of Anti-infective Therapy* **8**, 1441–1450 (2010).

47. Roldán, S., Herrera, D. & Sanz, M. Biofilms and the tongue: therapeutical approaches for the control of halitosis. *Clin Oral Invest* **7**, 189–197 (2003).

48. Palleja, A. *et al.* Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nature Microbiology* **3**, 1255 (2018).

49. Greub, G. Culturomics: a new approach to study the human microbiome. *Clinical Microbiology and Infection* **18**, 1157–1159 (2012).

50. Mullany, P. Functional metagenomics for the investigation of antibiotic resistance. *Virulence* **5**, 443–447 (2014).

51. Criscuolo, A. & Brisse, S. AlienTrimmer: A tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics* **102**, 500–506 (2013).

52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

53. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

54. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

23

55. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* **19**, 455–477 (2012).
56. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
57. Wattam, A. R. *et al.* PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* **42**, D581–D591 (2014).
58. Jonsson, V., Österlund, T., Nerman, O. & Kristiansson, E. Statistical evaluation of methods for identification of differentially abundant genes in comparative metagenomics. *BMC Genomics* **17**, 78 (2016).
59. Bengtsson-Palme, J., Larsson, D. G. J. & Kristiansson, E. Using metagenomics to investigate human and environmental resistomes. *J Antimicrob Chemother* **72**, 2690–2703 (2017).
60. Bengtsson-Palme, J. The diversity of uncharacterized antibiotic resistance genes can be predicted from known gene variants—but not always. *Microbiome* **6**, 125 (2018).
61. Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Meth* **12**, 902–903 (2015).

Acknowledgements

The research was supported by the Centre for Host-Microbiome Interactions, King's College London, the Biotechnology and Biological Sciences Research Council (BBSRC) (Grant No. BB/M009513/1) awarded to D.L.M., and Engineering and Physical Sciences Research Council (EPSRC), BBSRC and UK Research and Innovation (UKRI) (Project No. EP/S001301/1) awarded to S.S.. The authors would like to thank Professor William Wade for constructive criticism of the manuscript. S.S. acknowledges Uppmax for providing assistance for computational infrastructure.

Author contributions

24

V.C. and D.L.M. conceived the presented idea. V.C. and D.G-C. conducted the bioinformatics and data analysis. V.C. and D.L.M. wrote the manuscript with support from D.G-C., E.W., G.B.P., P.M., S.L. and S.S.. D.L.M. and D.G-C. oversaw the project.

25

Figures

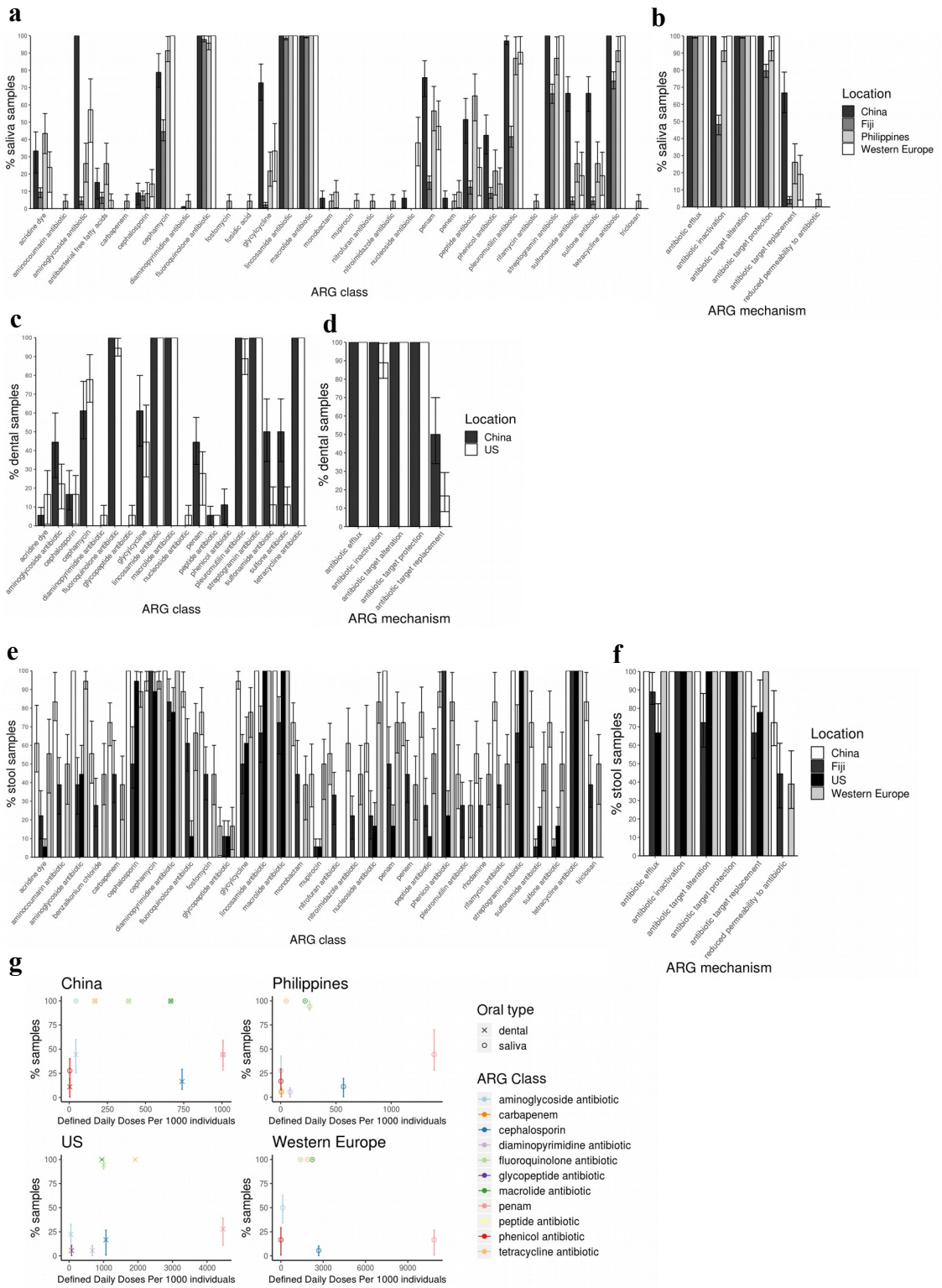
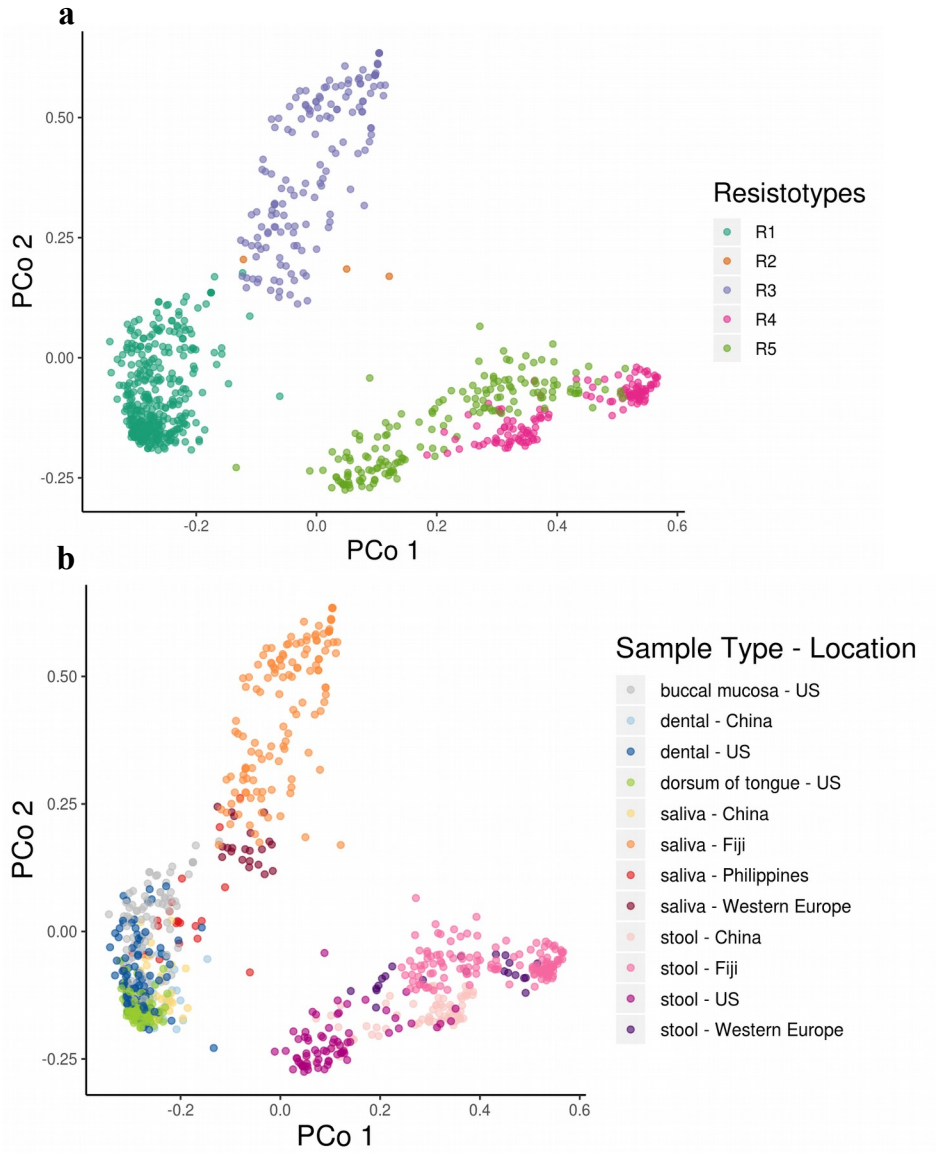


Fig. 1 Percentage of saliva samples from individuals that contain, **a** an ARG class and **b** an ARG mechanism, from China (n = 18), Fiji (n = 18), the Philippines (n = 18) and Western Europe (n = 18). Percentage of dental samples from individuals that contain, **c** an ARG class and **d** an ARG mechanism, from China (n = 18) and the US (n = 18). Percentage of stool samples from individuals that contain, **e** an ARG class and **f** an ARG mechanism, from China (n = 18), Fiji (n = 18), the US (n = 18) and Western Europe (n = 18). **g** Percentage of the same saliva, dental and stool samples containing an ARG class against the Defined Daily Doses Per 1000 individuals in 2015 from China, the Philippines, Western Europe (France and Germany) and the US. (Fiji antibiotic use data unavailable.) Error bars are 95 % confidence intervals (CIs) that were evaluated from percentages extracted from bootstrapping samples 100 times.

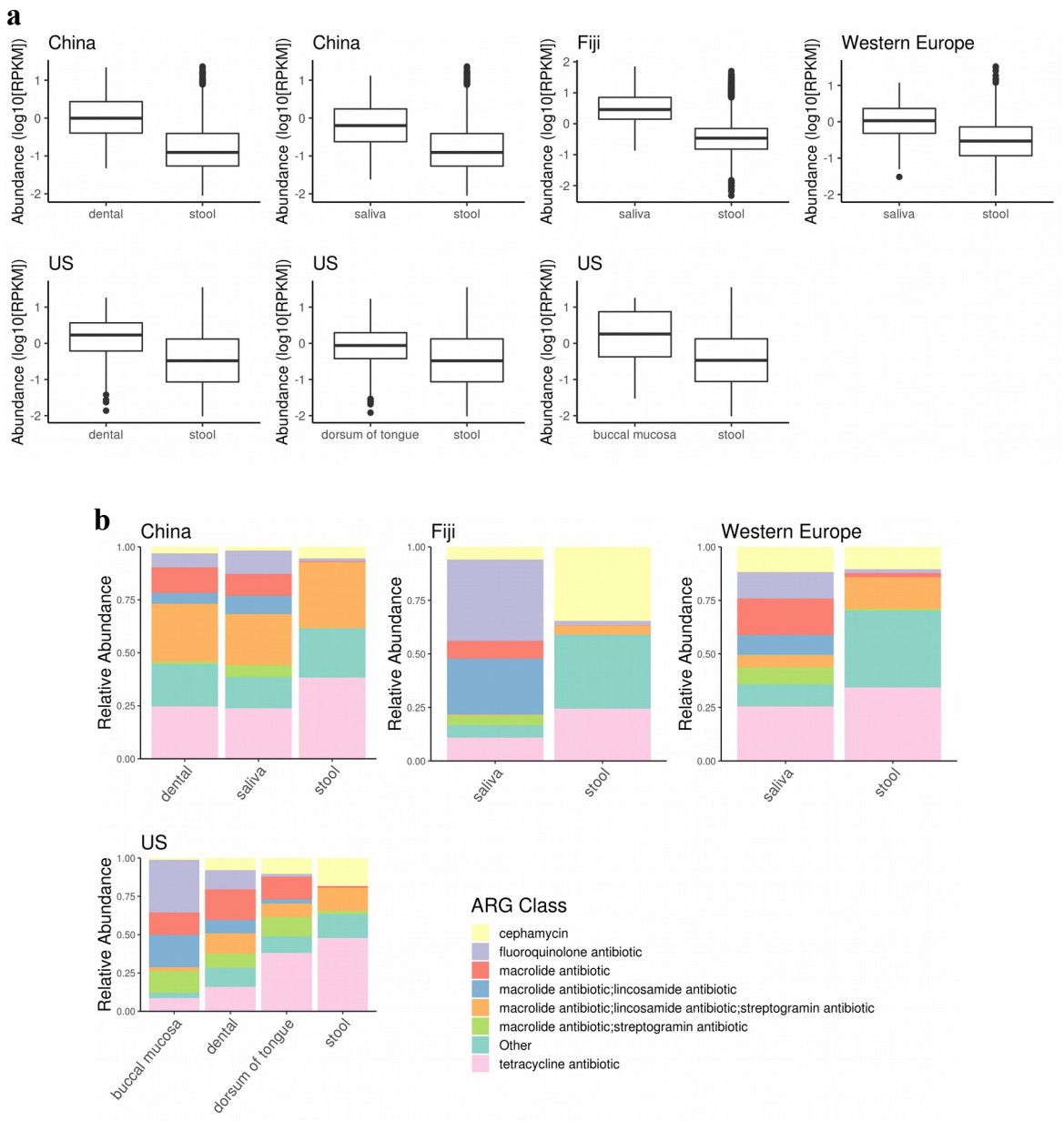
26

511



513 **Fig. 2** Principal Coordinates Analysis of the incidence (presence/absence) of ARGs for all samples where each sample
514 is represented by a point. **a** Samples are labelled as resistotype clusters, evaluated from hierarchical clustering of binary
515 distance between ARG incidence profiles. Number of clusters was selected using silhouette analysis as the number of
516 clusters greater than three with the highest silhouette width. **b** Samples are labelled as sample types and locations: US
517 buccal mucosa (n = 87), China dental (n = 32), US dental (n = 90), US dorsum of tongue (n = 91), China saliva (n = 33),
518 Fiji saliva (n = 137), Philippines saliva (n = 23), Western Europe saliva (n = 21), China stool (n = 72), Fiji stool (n =
519 137), US stool (n = 70), and Western Europe stool (n = 21)

27



521 **Fig. 3** Resistome abundance for China, Fiji, Western Europe and US. **a** Absolute abundance in log10 of reads per
522 kilobase of read per million (RPKM) of ARGs for paired samples between China stool (n = 30) and dental (n = 30),
523 China stool (n = 31) and saliva (n = 31), Fiji saliva (n = 137) and stool (n = 137), Western Europe saliva (n = 21) and
524 stool (n = 21), US stool (n = 68) and dental (n = 68), US stool (n = 69) and dorsum of tongue (n = 69), and US stool (n
525 = 65) and buccal mucosa (n = 65). **b** Relative abundance of reads labelled by ARG class for each sample type from
526 China saliva (n = 33), dental (n = 32) and stool (n = 72), Fiji saliva (n = 137) and stool (n = 137), Western Europe saliva
527 (n = 21) and stool (n = 21), and US buccal mucosa (n = 87), dorsum of tongue (n = 91), dental (n = 90) and stool (n =
528 70)

28

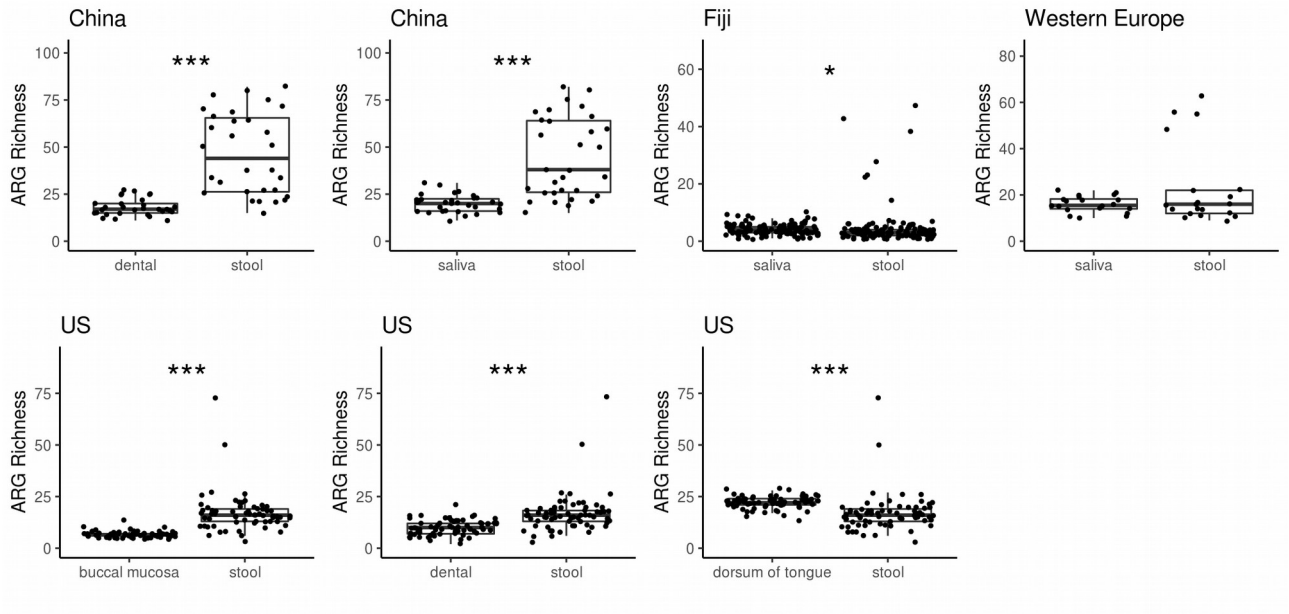
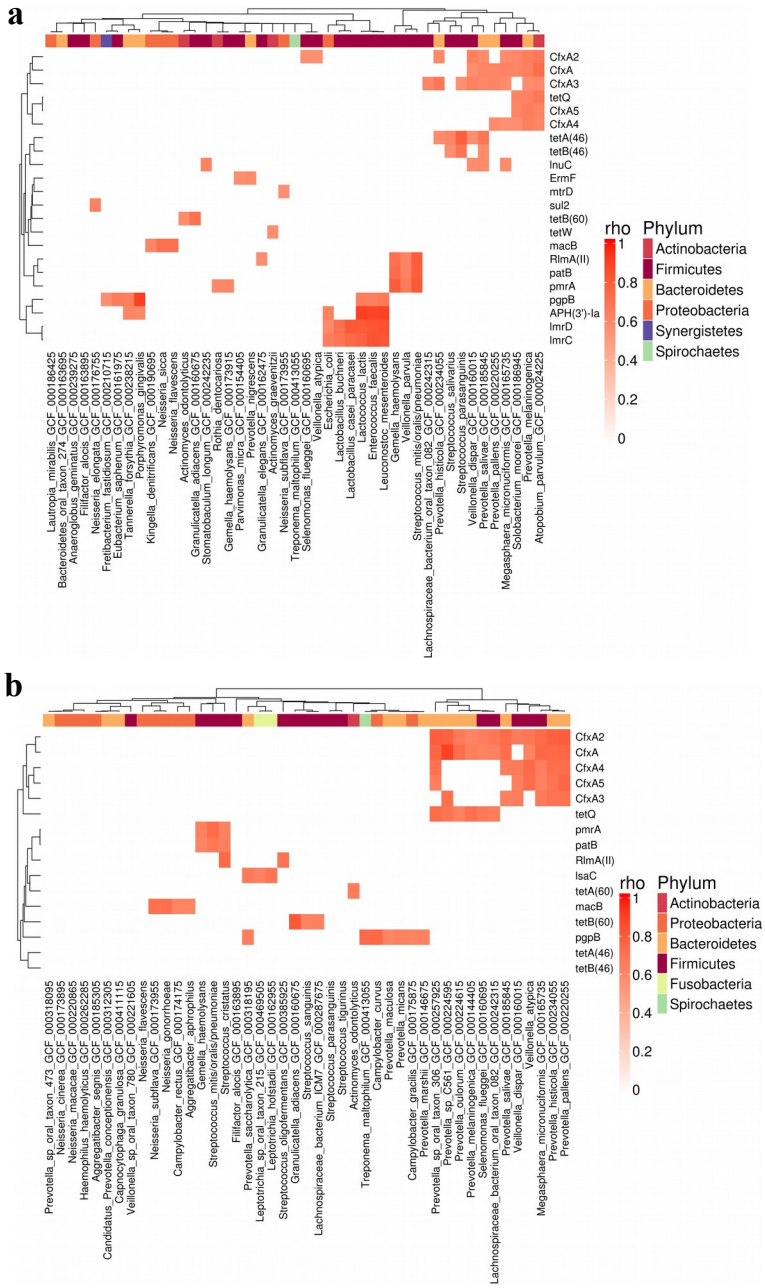


Fig. 4 ARG richness for paired samples between China stool (n = 30) and dental (n = 30), China stool (n = 31) and saliva (n = 31), Fiji saliva (n = 127) and stool (n = 127), Western Europe saliva (n = 21) and stool (n = 21), US buccal mucosa (n = 64) and stool (n = 64), US dental (n = 68) and stool (n = 68), and US dorsum of tongue (n = 69) and stool (n = 69) with Mann-Whitney, paired, two-sided t-test (p-value < 0.05 as *, < 0.01 as **, < 0.005 as ***)

29



534 **Fig. 5** Spearman's correlation of ARG and species/strain abundance from saliva samples represented as heatmaps of *rho*
535 between 0 and 1, only where adjusted p-value < 0.05. Each heatmap represents **a** China (n = 31) and **b** Philippines
536 saliva (n = 23). Rows and columns are clustered by hierarchical clustering of euclidean distance. Columns are coloured
537 by phylum. P-values are adjusted by Bonferroni-Hochberg multiple test correction.

30