The diversity of antibiotic resistance and virulence genes are correlated in human gut and environmental microbiomes

Pedro Escudeiro,[a]* Joël Pothier,[b] Francisco Dionisio,[a] Teresa Nogueira[a]#

[a] cE3c – Centro de Ecologia, Evolução e Alterações Ambientais, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

[b] Atelier de Bioinformatique, ISYEB, UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, Paris, France

Running Head: Antibiotic resistance and virulence in microbiomes

#Address correspondence to Teresa Nogueira, teresainogueira@gmail.com.

*Present address: Pedro Escudeiro, Microbiology & Biotechnology Group, Biosystems & Integrative Sciences Institute (BioISI), Faculdade de Ciências da Universidade de Lisboa, Portugal.

P.E. and J.P. contributed equally to this work.

Abstract

Human beings have used large amounts of antibiotics, not only in medical contexts but also, for example, as growth factors in agriculture and livestock, resulting in the contamination of the environment. Even when pathogenic bacteria are the targets of antibiotics, hundreds of non-pathogenic bacterial species are affected as well. Therefore, both pathogenic and non-pathogenic bacteria have gradually become resistant to antibiotics. We tested whether there is still co-occurrence of resistance and virulence determinants. We performed a comparative study of environmental and human gut metagenomes issuing from different individuals and from distinct human populations across the world. We found a great diversity of antibiotic resistance determinants (ARd) and virulence factors (VFd) in metagenomes. Importantly there is a correlation between ARd and VFd, even after correcting for protein family richness. In the human gut there are less ARd and VFd than in more diversified environments, and yet correlations between the ARd and VFd are stronger. They can vary from very high in Malawi, where antibiotic consumption is unattended, to non-existent in the uncontacted Amerindians. We conclude that there is co-occurrence of resistance and virulence determinants, suggesting a possible co-selective mechanism. For example, by selecting for resistant bacteria, we may be selecting for more virulent strains as a side effect of antimicrobial therapy.

Importance

Every year thousands of tons of antibiotics are used, not only in human and animal health but also as growth promoters in livestock. Consequently, during the last 75 years, antibiotic resistant bacterial strains have been selected in human and environmental microbial communities. This implies that, even when pathogenic bacteria are the targets of antibiotics, hundreds of non-pathogenic bacterial species are also affected. Here, we performed a comparative study of environmental and human gut microbial communities issuing from different individuals and from distinct human populations across the world. We found that

48 antibiotic resistance and pathogenicity are correlated and anticipate that, by selecting for
49 resistant bacteria, we may be selecting for more virulent strains, as a side effect of
50 antimicrobial therapy.

Introduction

Antibiotics are present in microbial communities, not only as a result of the natural lifestyle of microorganisms but also due to the usage of these drugs in agriculture, food industry, livestock, or in human health (1). Therefore, antibiotics can affect bacterial communities as a whole, comprising both pathogenic and non-pathogenic bacteria. Take, for example, the human microbiome, defined as the set of microorganisms that colonize humans. This microbiome is composed of about 3.8e+13 bacterial cells (2) spanning thousands of taxa and colonizing our body's surfaces and bio fluids, including tissues such as skin, mucosa and most importantly, the gastrointestinal tract. Thus, even when virulent bacteria are the targets of antibiotics, the administration of these drugs may also affect many non-pathogenic mutualistic or commensal bacterial species present in individuals undergoing treatment (3).

The resistome (the collection of all antibiotic resistance (AR) genes, which exist in both pathogenic and non-pathogenic bacteria (4)) is frequently encoded on mobile genetic elements. Similarly, the virulome, the set of genes encoding virulence can also be encoded on the mobilome (5–7). Therefore, many bacterial virulence factors (VF) are easily spread in bacterial populations by horizontal gene transfer, converting mutualistic or commensal bacteria into potential opportunistic pathogens. Several examples of highly virulent and multi-resistant disseminated clones have already been reported throughout the literature (for a review see (8)).

A given bacterial population may constitute a life insurance to other bacterial populations present in the same microbiome through at least two different mechanisms. First, if antibiotic resistance genes are coded by plasmids or other mobile genetic elements, they may transfer to pathogenic cells and save them from the negative effects of antibiotics. Indeed, these genetic elements can spread into the bacterial community by horizontal gene transfer, even crossing species (9–11). Curiously, some bacterial strains from different species have been shown to be extremely good donors of certain plasmids. As such, they are able to amplify the number of plasmids in a bacterial community and spread those plasmids to other bacterial

78 cells (12), a phenomenon probably explained by interactions between different plasmids (13,
79 14).

80 Secondly, even bacterial cells not coding for antibiotic resistance determinants nor receiving
81 mobile elements may be protected by cells coding for certain drug resistance genes. Indeed,
82 gene products that inactivate antibiotics by degrading or modifying antibiotic molecules are
83 also decreasing their concentration in the local environment (for a review about mechanisms
84 of possible indirect resistance see (15)). This mechanism of indirect resistance has been
85 shown to occur in different systems and to be pervasive (16–20).

86 In this paper we ask whether the abundance of antibiotic resistance protein families correlate
87 with that of virulence, among bacterial populations. For this, we have chosen to study
88 metagenomes. There are three main reasons to use metagenomes to address this question.
89 First, it is known that, for millions of years, bacteria had to cope with the presence of many
90 other species, mostly competing with them (21), but also cooperating, both phenomena
91 relying on the full set of genes of the metagenome (21). Secondly, horizontal gene transfer
92 promotes the genetic relationship between species thus enforcing cooperation (6) avoiding
93 the emergence of cheaters within the microbiome. Third, the study of metagenomes gives us
94 access to the repertoire of genes involved in adaptation to the environment, given that many
95 of these traits are often encoded in the mobilome and thus can be shared by different,
96 eventually unrelated bacteria. In this context, mining for genes coding for antibiotic
97 resistance and virulence traits in metagenomes is a reliable way to access the selective
98 pressures the population is subject to, as well as the co-occurrence of genetic traits of the
99 whole microbiome.

100 The present study aims at understanding the relationship between antibiotic resistance
101 genes and those coding for virulence. Here we show that there is indeed a linkage between
102 the dissemination of virulence factors and genes coding for antibiotic resistance, within
103 natural microbiomes, and that this relationship could be influenced by the behaviours of
104 human populations, spanning from very different geographical locations across the world.

105

Results

Antibiotic resistance (AR) gene families in the metagenomes

We started by evaluating the quantity of different AR proteins present in the metagenomes under study. We wanted to know how does the diversity of antibiotic resistance gene families' homologues (that is, ARd, the number of different protein families) vary with the metagenome protein family richness (that is the total number of proteins families in a metagenome). To answer this question, we used a data set of environmental metagenomes issuing from diverse ecosystems and biomes, such as oceans, coral atolls, deep oceans, Antarctic aquatic environments and snow, soils, hyper saline sediments, sludge's, microbial fuel cell biofilms, and animal microbial populations (22). In this dataset, there is a broad variation in AR gene diversity (ARd) (Fig 1A). However, there is a correlation between ARd and the protein family richness of the various metagenomes (overall correlation of 0.754). In Fig 1A, regression lines for human gut microbiome (in blue, slope 0.0037) and for other biomes (in black, slope 0.0052) are shown. Nevertheless, the p-value for the difference of these slopes is 1.4e-02, which is significant but not highly significant, does not allow, in our opinion, to draw solid conclusions from this difference.

Another dataset composed by 110 human gut metagenomes, sampled from healthy individuals with ages ranging from 0.05 to 53 years of age, spanning different regions of the world such as: US, Malawi, and the Venezuelan Amazon was also studied (23).

Virulence factor (VF) gene families in the metagenomes

Bacterial virulence factors are proteins that enable pathogenic bacteria to parasitize a host, including gene products involved in adhesion and invasion, secretion, toxins, and iron acquisition systems (24).

The environmental microbiomes of our dataset reveal a great diversity of virulence factors protein families' (VFd) densities (Fig 1B). Since virulence may also be associated with the colonization of different types of biomes (some virulent traits are involved in adaptation to

new environments, by allowing bacteria to adhere and colonize substrates, to access to resources such as iron, among other functions) besides the context of infection, one can expect different types of these genes in environmental microbiomes. Moreover, Fig 1B shows that in human gut microbiomes, VFd versus protein family richness of the metagenomes achieves an overall correlation coefficient of 0.669. Regression lines for human gut microbiome (in blue, slope 0.0035) and for other biomes (black, slope 0.0106) are shown. Here the p-value for difference of these slopes is 1.7e-08, indicating that there could be a difference, but the points are very spread over all types of biome, and we prefer not to draw conclusion about these data.

The AR / VF correlations

The main purpose of the present work is to evaluate the relationship, if any, between ARd and VFd in microbiomes. Therefore, we excluded from our analysis all the gene products that are both homologues to AR and VF determinants. Thus, we avoid the introduction of a bias due to both correlations of AR and VF diversities (ARd and VFd) with metagenomes richness. In the analysis, we systematically analysed the residuals of the regression of ARd on metagenome richness and the residuals of the regression of VFd on metagenome richness. In Fig 3 one can see that the residuals of ARd and those of VFd are positively correlated. Although there is a large variation of ARd (Fig 1A), and VFd (Fig 1B) in environmental metagenomes, there is a strong correlation between ARd and VFd (Fig 3): the overall partial correlation between ARd and VFd given metagenome richness is 0.84 (p-value 0.8e-35~i.e. 0.0). Indeed, the partial correlation between ARd and VFd given the metagenome richness is the same as the correlation between residuals of ARd and residuals of VFd (the partial correlation coefficient for environmental biomes is 0.86 while the one for the human gut is 0.90). When we draw the regression line for the human gut microbiomes from this dataset (human gut), we can see that the ratio of (residuals ARd)/(residuals VFd) is higher with slope of 0.63 than for the case of environmental metagenomes with slope of 0.44. The difference in slopes is significant, but not very significant, with p-value of 3e-02. Despite the fact that the values for ARd and VFd are lower in gut metagenomes (see Fig 1), one can witness a higher ratio ARd/VFd (larger slope) of metagenomes pertaining to the human faeces biome relatively

161 to the slope calculated for all environmental metagenomes (Fig 3). This suggests that there
162 could be a greater accumulation of ARd over VFd for this particular biome than in
163 environmental ones.

164 Fig 4 shows residuals of AR diversity versus the residuals of VF diversity in the human gut
165 metagenomes for each population (country) and for the three populations together. The
166 partial Pearson correlation coefficient between ARd and VFd given metagenome richness is
167 lower in the human gut metagenomes dataset (r=0.68, p-value 3e-22~0.0, Fig 4) than in the
168 environmental samples (r=0.86 as seen above, Fig 3).

169 We can further distinguish different trends upon geographical localization of the human
170 populations under study. The largest contribution to this graph comes from the North
171 American samples, which account for 66/110 (60%) of the individuals (Fig 4). The
172 Amerindians (Venezuela) metagenomes account for 21/110 metagenomes (19%) and the
173 Malawians for 23/110 (21%). These two non-western human populations have very
174 contrasting lifestyles and exposure to antibiotics: the uncontacted Amerindians, harbouring a
175 resistome similar to that of pre-antibiotic era; and those of the people from Malawi with a
176 high prevalence of infectious diseases and widespread use of unprescribed antibiotics (25–
177 27).

178 The Pearson partial correlation coefficients for metagenomes in the three populations are
179 positive and in the same range (North americans: 0.66, Malawians: 0.81, Amerindians: 0.70).
180 Regressions of the residuals of ARd on the residuals of VFd (Fig 4) shows a smaller slope for
181 Amerindians (0.20) than for USA (0.31) and Malawians (0.33), but the differences are not
182 significant: p-values spray from 8e-02 for the difference in Malawians-Amerindians slopes to
183 7.2e-01 for the difference in Malawians-North americans slopes.

184 **The co-occurrence of AR and VF belonging to the cell envelope**

185 Our results suggest that co-occurrence of AR and VF might be taking place amidst bacterial
186 communities. We wondered, however, which were the genetic traits that were more prone to
187 this effect. We have then computed the partial correlations, given the metagenome richness,

188 between every AR and VF protein families in the metagenomes under study to generate the
189 correlogram presented in Fig 5. The vast majority of associations falls into the functional
190 category of multi-drug efflux pumps (AR's) associated with either secretion systems, as well
191 with iron uptake and adhesion mechanisms (VF's), respectively (Fig 5). Amongst the most
192 representative associations between AR and VF traits are those belonging to the cell
193 envelope and the general secretion mechanisms. On the other hand, the AR protein families
194 for beta-lactamases, and general beta-lactam resistance mechanisms, bore no statistically
195 significant correlations whatsoever.

196

Discussion

Antibiotic resistance among environmental and human gut metagenomes

As expected, we found different homologues for antibiotic resistance (AR) gene products belonging to different protein families (AR diversity or ARd) among environmental metagenomes (Fig 1A). We have shown that, in the environmental samples, ARd varies a lot from metagenome to metagenome. This may result from the differential microbial community composition of the metagenomes, whose genetic diversity can be grouped according to the adaptation to the environment in question (22); but also from the fact that the selective pressure for the maintenance of antibiotic resistance genes in environmental microbiome varies widely from environment to environment (28–30). Human gut metagenomes, on the contrary, have a less diversified repertoire of AR determinants (lower ARd) than that of their environmental counterparts.

We found a very strong correlation between the diversity of AR determinants and the metagenome's protein family richness independently of the geographic origin of the human populations. These similar densities of ARd can indicate that, in human gut microbiomes, the number of different AR genes is not influenced by the human lifestyle, such as diet, medical care, access to antibiotics or other cultural habits (31), and/or that the adaptation to the intestinal tract shapes microbial AR diversity as well. This is a surprising results given that antibiotic use can vary from country to country (CDDEP), and, as consequence, individuals from different human populations are under different antibiotic exposition. Forslund and co-workers have demonstrated that there are robust differences in the antibiotic resistance arsenal between countries and that these differences follow the veterinary and human antibiotic use (27).

Virulence among environmental and human gut metagenomes

In what concerns virulence, we can assert that there is a wide diversity and density differences of VFs in environmental metagenomes, which poses as evidence of the plasticity

223 portrayed by environmental bacteria to adapt to different hosts and niches (Fig 1B). On the
224 other hand, human gut microbiomes harbour a less diverse VF repertoire, especially in the US
225 samples, which seem to indicate an evolution towards adaptation to the human gut, or lower
226 contact with pathogens, eventually due to vaccinations and sanitation.

227 **Association of AR / VF**

228 According to Yatsunenko and colleagues, once the human gut metagenome is established at
229 the age of 3 years old, it does not diverge much from individual to individual, in what
230 concerns both phylogenetic diversity and functional richness (23).

231 The most relevant results shown here are that, when corrected for metagenome's protein
232 family richness, ARd and VFd are strongly correlated both in the environmental samples
233 (Figs 3) and in the human gut samples (Fig 4), with special emphasis on the human gut.

234 The North American intestinal samples (Fig 4) show a wide variety of associations between
235 ARd and VFd, always presenting a statistically significant correlation between these genetic
236 traits. This result, in itself, reinforces our hypothesis that antibiotic resistance and virulence
237 are in fact co-associated, suggesting that the mobilization of mobile genetic elements as gene
238 cassettes, driving to multirresistance, might be taking place. The US population, like other
239 industrialized countries, is culturally exposed to antibiotics from health care facilities such as
240 hospitals, antibiotic therapy, and the use of antibiotics in agriculture and livestock.

241 Malawians and amerindians from Venezuela share, phylogenetically, more similar human gut
242 microbiomes than the North Americans (23), which belong mainly to a different enterotype
243 (23, 32). Yet, whereas there is no ARd/VFd correlation among the Venezuelan individuals,
244 there is a very strong correlation among the Malawian ones. This result highlights the
245 potential effect of the exposure to antibacterial drugs on promoting the dissemination of
246 antibiotic resistance by horizontal gene transfer, and the potential co-selection of virulence
247 traits within bacterial communities. The uncontacted Amerindian microbiome represents a
248 "frozen" relic of a pre-antibiotic era of the human resistome, while the Malawian gut

249 microbiome is much more exposed, both to antibiotics, and to colonization by pathogens (25–
250 27).

251 Amerindians have no known access to pharmaceutical drugs, as they usually make use of the
252 traditional indigenous medicine. Malawi is one of the poorest countries in Africa, where most
253 people live with less than one dollar a day, many people cope with AIDS and bacterial
254 infections, and where many children suffer from severe malnutrition (25). Nutrition has been
255 reported to have a big impact on both the human gut microbiome composition and resistome
256 (31, 33, 34). In Malawi, UNICEF has been implementing a program of Ready-to-Use-
257 Therapeutic-Food (RUTF) to reduce mortality among children. RUTF often contains antibiotics
258 such as co-trimoxazole. It has been questioned, however, whether the success of this therapy
259 is due to re-nutrition or to the combination with antibiotics (35, 36) that could have an effect
260 on the microbial gut composition. It has been reported that there is widespread resistance to
261 almost all of antibiotics that are empirically used in Malawi due to the lack of routine
262 microbiological culture and sensitivity testing (36), but also due to self-medication. Human
263 gut samples from the Malawi people of the dataset used in the present study are described
264 as having a high overall resistance potential with an over representation of cephalosporin
265 and tetracycline resistance genes which may suggest extensive use of old, broad-spectrum
266 antibiotics, a known problem in many developing countries (27).

267 The co-representation of AR and VF targeting the cell envelope

268 Between all the possible statistical AR and VF protein families' correlations, the strongest
269 ones are among proteins belonging to the bacterial cell envelope (Fig 5). This result is not
270 surprising, as many proteins that belong to the secretory system or to the secretome itself,
271 and thus target the bacterial envelope, are frequently encoded on mobile genetic elements
272 (6). As such, this co-occurrence may also be regarded as a direct consequence of the
273 dynamics of these genetic elements coding for both types of determinants.

274 In human gut, the strongest correlations involve efflux pumps that can extrude antibiotics
275 non-specifically, and adhesion and iron scavenging mechanisms of pathogenicity. One
276 possible explanation for the fact that the most frequent AR and VF associations in human gut

277 involve efflux pumps could be that they allow a fast and efficient response to new man-made
278 antibiotic molecules, while in environmental genomes, specific resistance mechanisms
279 targeted to specific antibiotic molecules may have had more time to evolve to specific
280 strategies.

281 Another possible explanation is that extrusion can also be implicated in pathogenic
282 mechanisms of human associated pathogens as well as bacteria-host interactions. For
283 example, biofilm formation in a *Staphylococcus aureus* methicillin resistant (MRSA) strain has
284 been known to be essentially reliant on the activity of fibronectin binding proteins (37), and
285 multi-drug efflux pumps have direct implications with the formation and maintenance of such
286 biofilms (38). Furthermore quinolone resistant *S. aureus* strains up regulate the production of
287 fibronectin binding proteins when subjected to sub-lethal dosages of ciprofloxacin (39). It has
288 also been acknowledged that physiological levels of some cations present within the host
289 promote the up regulation of genes encoding putative efflux transporters, oxidoreductases,
290 and mechanisms of iron uptake either in *Acinetobacter baumannii* (40) as in *Burkholderia*
291 *cenocepacia* (41), which could explain the co-association of iron acquisition systems with
292 those of multi-drug efflux pumps.

293 We show here that, even after correcting for protein family richness, there is a correlation
294 between the diversity of antibiotic resistance and virulence traits in human gut microbiomes
295 and in environmental microbiomes. Interestingly, this correlation varies from a human
296 population to another. Human antibiotic exposure due, either to therapy or to the
297 environment and food, can have an effect on selecting for potentially pathogenic bacteria in
298 the human gut microbiomes. It can also drive and shape changes on the gene pool of
299 microbiomes. This means that, by selecting for resistant bacteria, we might be also selecting
300 for more virulent strains, as a side effect of antimicrobial therapy.

301 # Materials and methods

302 ## Metagenomic datasets

303 Our human gut query cohort included 110 previously studied, and publicly available
304 metagenomes pertaining to individuals from different regions of Venezuela, Malawi, and the
305 US, as well as a broad age span (0.05 to 53 years) (23) (dataset of the project mgp98 "Human
306 gut microbiome viewed across age and geography (WGS)" at the MG-RAST metagenomic
307 analysis server). All the metagenomes files were generated by the same team and project,
308 and by using the same bioinformatics pipeline. Our environmental cohort comprised 64
309 previously selected, and publicly available environmental metagenomes, belonging to 12
310 different biomes (dataset collection of the metagenomes referred in the "Metagenomic
311 mining for microbiologists" study) (22). Although Delmont's team report using a dataset
312 comprised of 77 metagenomes, there are only 70 MG-RAST accession numbers present in the
313 article's appendix, of which only 64 are publicly available. Both project's metagenomes were
314 downloaded from MG-RAST on the 29th of January 2016, under FASTA format, making use of
315 successive calls to its Application Programming Interface (API) (42) using the respective MG-
316 RAST's accession numbers available in the aforementioned bibliography. Each FASTA file
317 comprised clustered protein-coding sequences, retrieved from MG-RAST's file-formatting
318 pipeline (550.cluster.aa90.faa files). The protein sequences enclosed in these files are
319 clustered at 90% identity, containing non-redundant translated sequences. These protein-
320 coding formatted FASTA files contain the translation of one representative from each cluster.
321 Thus, the numbers of protein sequences for a given metagenome used here represent its
322 richness.

323 ## BLAST, VFDB and Resfams

324 For every metagenome present in our query, a BLASTP (43) search was performed against
325 the VFDB database for bacterial virulence factors families (Virulence Factors of Pathogenic
326 Bacteria VF.pfasta dataset of the Institute of Pathogen Biology, China) (24) and the Resfams
327 AR Proteins database of bacterial antibiotic resistance protein families (Resfams AR Proteins

328 - v1.2 database of the Dantas Lab at the Washington University School of Medicine, St Louis)
329 (44). An alternative approach made use of the Antibiotic Resistance Genes Database (ARDB)
330 (reference antibiotic resistance genes dataset resisGenes.pfasta of the Center for
331 Bioinformatics and Computational Biology University of Maryland) (45), but several
332 hindrances concerning its sub-classification by functional antibiotic resistance protein
333 families made us discard the possibility of using such database to address of AR and VF
334 protein family co-occurrence in metagenomes.

335 The BLAST+ executables package was downloaded on the 17th of November 2015 (ncbi-
336 blast-2.2.31+ version). The VFDB was downloaded on the 11th of November 2013 (31
337 classified FASTA files of bacterial virulence factor sub-families), and the Resfams database
338 was downloaded on the 29th of January 2016, (123 classified FASTA files of bacterial
339 antibiotic resistance protein subfamilies. Every protein enclosed in each of the addressed
340 metagenomes, was used as a query in order to search for similarities to either AR or VF
341 protein-coding traits. Hence, we aimed at retrieving the best hit (best scored alignment) that
342 enabled us to assign an AR or VF function to each of the aforementioned metagenomic
343 proteins. Every BLASTP search was performed with a very stringent e-value cut-off of 1e-15
344 (10 orders of magnitude lower than the conservative commonly used E-value of 1e-5). Next,
345 we filtered the resulting output files, as to only retrieve the alignment hits with >60%
346 coverage of the query size, and whose query and subject length ratio is between 0.75 and 1.5.
347 When using a combination of this very stringent e-value of 1e-15 together with coverage, and
348 length ratio filters, we expect to only retrieve true homologues, avoiding false positives.
349 Furthermore, from all the generated alignments between our metagenomic query cohort and
350 the preceding databases, we computed the representative counts for the different gene
351 families that coded for either AR or VF traits. Thus, the amounts of different classes (gene
352 families) that are present in a given metagenome represent its diversity in terms of AR or VF
353 traits.

354 We have also removed hits for proteins that aligned with both antibiotic resistance and
355 virulence factor proteins (25.9% of hits against the VFDB, and 29,9% of the hits against the
356 Resfams database). The abovementioned filters and algorithms were implemented making

use of Unix scripting languages (GNU Awk version 4.0.1, and Z-Shell version 5.0.2), under a Linux environment.

Statistical Analysis

To test for relationships in metagenomes between the presence of both antibiotic resistance and virulence factors traits, we proceeded as follows. Our expectation is that the diversity number for homologues of antibiotic resistance protein families (henceforth denoted as ARd) and the diversity number for homologues of virulence factors protein families (denoted as VFd) in each metagenome (clustered at 90% identity) increases with the protein family richness of the latter (being that the protein family richness here means the total number of cluster representative proteins). Given the potential diversity of these protein families, one does not expect ARd and VFd to level off with a metagenome's protein family richness. Therefore, we assume a linear relationship between ARd and metagenome's protein family richness, with a fixed 0 intercept. The same for VFd. Our results will show that these assumptions are reasonable. Thus, to avoid spurious correlation, we corrected the diversity of ARd and VFd in a metagenome by taking into account its protein family richness: we have computed the partial correlations between ARd and VFd given the metagenome richness, and equivalently, when plotting ARd versus VFd, we have instead taken residuals of regression of ARd on metagenome richness and residuals of regression of VFd on metagenome richness. Recall that the partial correlation measures the degree of association between two random variables, with the effect of a set of controlling random variables removed. For example, the partial correlation between A and B given a third variable C correlated with A and B, is the correlation between the residuals e_a and e_b resulting from the linear regression of A with C and of B with C. If r_{AB}, r_{AC} and r_{BC} are the correlation coefficients between A and B, A and C, and B and C, the partial correlation between A and B given C, $r_{AB.C}$, can be computed as $r_{AB.C} = \frac{r_{AB} - r_{AC} r_{BC}}{\sqrt{1-r^2{}_{AC}}\sqrt{1-r^2{}_{BC}}}$.

Acknowledgements

We thank Eduardo P. C. Rocha for critical reading of the manuscript and constant interest

and suggestions since the beginning of this work, Guillaume Sapriel for suggestions to improve the methods, Pedro David, Eric Duchaud, Guillaume Achaz, and all the members of our teams for the fruitful discussions, Octávio Paulo, Francisco Pina-Martins and Sophie Brouillet for the technical help, and granting access to the servers.

Conceived and designed the experiments: JP TN. Performed the experiments: PE. Analysed the data: JP FD TN. Wrote the paper: JP FD TN.

References

1. Castanon JIR. 2007. History of the use of antibiotic as growth promoters in European poultry feeds. Poult Sci 86:2466–2471.

2. Sender R, Fuchs S, Milo R. 2016. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol 14:e1002533.

3. Francino MP. 2015. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. Front Microbiol 6:1543.

4. Wright GD. 2007. The antibiotic resistome: the nexus of chemical and genetic diversity. Nat Rev Microbiol 5:175–186.

5. Svara F, Rankin DJ. 2011. The evolution of plasmid-carried antibiotic resistance. BMC Evol Biol 11:130.

6. Nogueira T, Rankin DJ, Touchon M, Taddei F, Brown SP, Rocha EPC. 2009. Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. Curr Biol 19:1683–1691.

7. Nogueira T, Touchon M, Rocha EPC. 2012. Rapid evolution of the sequences and gene repertoires of secreted proteins in bacteria. PLoS One 7:e49403.

408 8. Beceiro A, Tomas M, Bou G. 2013. Antimicrobial Resistance and Virulence: a Successful
409 or Deleterious Association in the Bacterial World? Clin Microbiol Rev 26:185–230.

410 9. Sommer MOA, Church GM, Dantas G. 2010. The human microbiome harbors a diverse
411 reservoir of antibiotic resistance genes. Virulence 1:299–303.

412 10. Andersson DI, Hughes D. 2014. Microbiological effects of sublethal levels of antibiotics.
413 Nat Rev Microbiol 12:465–478.

414 11. Vega NM, Gore J. 2014. Collective antibiotic resistance: mechanisms and implications.
415 Curr Opin Microbiol 21:28–34.

416 12. Dionisio F, Matic I, Radman M, Rodrigues OR, Taddei F. 2002. Plasmids spread very fast
417 in heterogeneous bacterial communities. Genetics 162:1525–1532.

418 13. Gama JA, Zilhão R, Dionisio F. 2017. Multiple plasmid interference – Pledging allegiance
419 to my enemy's enemy. Plasmid 93:17–23.

420 14. Gama JA, Zilhão R, Dionisio F. 2017. Conjugation efficiency depends on intra and
421 intercellular interactions between distinct plasmids: Plasmids promote the immigration
422 of other plasmids but repress co-colonizing plasmids. Plasmid 93:6–16.

423 15. Wright G. 2005. Bacterial resistance to antibiotics: Enzymatic degradation and
424 modification. Adv Drug Deliv Rev 57:1451–1470.

425 16. Nicoloff H, Andersson DI. 2015. Indirect resistance to several classes of antibiotics in
426 cocultures with resistant bacteria expressing antibiotic-modifying or -degrading
427 enzymes. J Antimicrob Chemother 71:100–110.

428 17. Sorg RA, Lin L, van Doorn GS, Sorg M, Olson J, Nizet V, Veening J-W. 2016. Collective
429 Resistance in Microbial Communities by Intracellular Antibiotic Deactivation. PLoS Biol
430 14:e2000631.

431 18. Domingues IL, Gama JA, Carvalho LM, Dionisio F. 2017. Social behaviour involving drug
432 resistance: the role of initial density, initial frequency and population structure in
433 shaping the effect of antibiotic resistance as a public good. Heredity 119:295–301.

434 19. Ponomarova O, Patil KR. 2015. Metabolic interactions in microbial communities:
435 untangling the Gordian knot. Curr Opin Microbiol 27:37–44.

436 20. Rosenthal AZ, Matson EG, Eldar A, Leadbetter JR. 2011. RNA-seq reveals cooperative
437 metabolic interactions between two termite-gut spirochete species in co-culture. ISME J
438 5:1133–1142.

439 21. Foster KR, Bell T. 2012. Competition, Not Cooperation, Dominates Interactions among
440 Culturable Microbial Species. Curr Biol 22:1845–1850.

441 22. Delmont TO, Malandain C, Prestat E, Larose C, Monier J-M, Simonet P, Vogel TM. 2011.
442 Metagenomic mining for microbiologists. ISME J 5:1837–1843.

443 23. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris
444 M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J,
445 Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI.
446 2012. Human gut microbiome viewed across age and geography. Nature 486:222–227.

447 24. Chen L, Xiong Z, Sun L, Yang J, Jin Q. 2012. VFDB 2012 update: toward the genetic
448 diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res
449 40:D641–5.

450 25. WHO. Global Health Observatory (GHO) data. WHO.

451 26. UNICEF. Committing to Child Survival: A Promise Renewed – Progress Report 2015.
452 UNICEF.

453 27. Forslund K, Sunagawa S, Kultima JR, Mende DR, Arumugam M, Typas A, Bork P. 2013.
454 Country-specific antibiotic use practices impact the human gut resistome. Genome Res
455 23:1163–1169.

456 28. Martinez JL. 2008. Antibiotics and Antibiotic Resistance Genes in Natural Environments.
457 Science 321:365–367.

458 29. Martínez JL, Baquero F. 2002. Interactions among strategies associated with bacterial
459 infection: pathogenicity, epidemicity, and antibiotic resistance. Clin Microbiol Rev

460 15:647–679.

461 30. Fitzpatrick D, Walsh F. 2016. Antibiotic resistance genes across a wide variety of
462 metagenomes. FEMS Microbiol Ecol 92.

463 31. Ghosh TS, Gupta SS, Nair GB, Mande SS. 2013. In silico analysis of antibiotic resistance
464 genes in the gut microflora of individuals from diverse geographies and age-groups. PLoS
465 One 8:e83823.

466 32. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap
467 J, Bruls T, Batto J-M, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen
468 T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh
469 C, Bjørn Nielsen H, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S,
470 Torrents D, Ugarte E, Zoetendal EG, JunWang, Guarner F, Pedersen O, de Vos WM,
471 Brunak S, Doré J, MetaHIT Consortium, Weissenbach J, Dusko Ehrlich S, Bork P. 2011.
472 Enterotypes of the human gut microbiome. Nature 474:666–666.

473 33. Ghosh TS, Gupta SS, Bhattacharya T, Yadav D, Barik A, Chowdhury A, Das B, Mande SS,
474 Nair GB. 2014. Gut microbiomes of Indian children of varying nutritional status. PLoS One
475 9:e95547.

476 34. Pehrsson EC, Tsukayama P, Patel S, Mejía-Bautista M, Sosa-Soto G, Navarrete KM,
477 Calderon M, Cabrera L, Hoyos-Arango W, Bertoli MT, Berg DE, Gilman RH, Dantas G.
478 2016. Interconnected microbiomes and resistomes in low-income human habitats. Nature
479 533:212–216.

480 35. Alcoba G, Kerac M, Breysse S, Salpeteur C, Galetto-Lacour A, Briend A, Gervaix A. 2013.
481 Do children with uncomplicated severe acute malnutrition need antibiotics? A
482 systematic review and meta-analysis. PLoS One 8:e53184.

483 36. Makoka MH, Miller WC, Hoffman IF, Cholera R, Gilligan PH, Kamwendo D, Malunga G,
484 Joaki G, Martinson F, Hosseinipour MC. 2012. Bacterial infections in Lilongwe, Malawi:
485 aetiology and antibiotic resistance. BMC Infect Dis 12:67.

486 37. McCourt J, O'Halloran DP, McCarthy H, O'Gara JP, Geoghegan JA. 2014. Fibronectin-

487 binding proteins are required for biofilm formation by community-associated methicillin-
488 resistant Staphylococcus aureus strain LAC. FEMS Microbiol Lett 353:157–164.

489 38. Soto SM. 2013. Role of efflux pumps in the antibiotic resistance of bacteria embedded in a
490 biofilm. Virulence 4:223–229.

491 39. Bisognano C, Kelley WL, Estoppey T, Francois P, Schrenzel J, Li D, Lew DP, Hooper DC,
492 Cheung AL, Vaudaux P. 2004. A recA-LexA-dependent pathway mediates ciprofloxacin-
493 induced fibronectin binding in Staphylococcus aureus. J Biol Chem 279:9064–9071.

494 40. Hood MI, Jacobs AC, Sayood K, Dunman PM, Skaar EP. 2010. Acinetobacter baumannii
495 increases tolerance to antibiotics in response to monovalent cations. Antimicrob Agents
496 Chemother 54:1029–1041.

497 41. Drevinek P, Holden MTG, Ge Z, Jones AM, Ketchell I, Gill RT, Mahenthiralingam E. 2008.
498 Gene expression changes linked to antimicrobial resistance, oxidative stress, iron
499 depletion and retained motility are observed when Burkholderia cenocepaciagrows in
500 cystic fibrosis sputum. BMC Infect Dis 8.

501 42. Wilke A, Bischof J, Harrison T, Brettin T, D'Souza M, Gerlach W, Matthews H, Paczian T,
502 Wilkening J, Glass EM, Desai N, Meyer F. 2015. A RESTful API for accessing microbial
503 community data for MG-RAST. PLoS Comput Biol 11:e1004008.

504 43. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997.
505 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.
506 Nucleic Acids Res 25:3389–3402.

507 44. Gibson MK, Forsberg KJ, Dantas G. 2015. Improved annotation of antibiotic resistance
508 determinants reveals microbial resistomes cluster by ecology. ISME J 9:207–216.

509 45. Liu B, Pop M. 2009. ARDB--Antibiotic Resistance Genes Database. Nucleic Acids Res
510 37:D443–D447.

511 ## Figure captions

512 **Figure 1. Distribution of the diversity number of antibiotic resistance and virulence**
513 **factor protein families by environmental metagenomes' protein family richness**

514 The vertical axes represent the total diversity count of AR protein families' homologues (ARd)
515 (1. A) and VF protein families' homologues (VFd) (1.B), present in metagenomes. The
516 horizontal axes represent the protein family richness of the metagenome, that is, the number
517 of cluster representative sequences - see Methods. Each dot represents one of the 64
518 metagenomes. In 1.A (resp.1B), the black lines represent the regressions of the ARd (resp.
519 VFd) of all the metagenomes on the metagenome richness. The points are scattered showing
520 that the diversity of AR and even more the diversity of VF gene families can vary greatly from
521 metagenome to metagenome. The light green lines represent the linear regressions of the
522 ARd (resp. VFd) of the human faeces metagenomes subset on metagenome richness.

523 **Figure 2. Distribution of the diversity number of antibiotic resistance and virulence**
524 **factor protein families in human gut metagenomes versus metagenome protein**
525 **family richness**

526 The vertical axes represent the total diversity count of AR protein families' homologues (ARd)
527 (2. A) and VF protein families' homologues (VFd) (2.B), present in the 110 human gut
528 metagenomes (see Material and Method). As in Figure 1, the horizontal axes represent the
529 protein family richness of the metagenomes. Each dot represents one of the 110 human gut
530 metagenomes. In 2.A (resp.2B), the black lines represent the regressions of the ARd (resp.
531 VFd) of all the metagenomes on the metagenome richness. The green lines represent the
532 linear regression of the ARd (resp. VFd) of the USA metagenomes subset on metagenome
533 richness, the red lines are regression lines for the Malawian metagenomes, and the blue lines
534 are those for the Venezuelan metagenomes.

535 **Figure 3. Distribution of AR by VF diversity in environmental metagenomes.**

536 Scatter plot ARd residuals versus VFd residuals of each environmental and human faeces 61
537 metagenomes (see Material and Method). Residuals are taken from the regressions shown in
538 Figure 1.A and 1.B. Each dot represents a metagenome. The green line represents the
539 regression line of ARd residuals environmental metagenomes on VFd residuals (the grey
540 shading is the 95% confidence interval) and its slope is 0.44. The red line represents the
541 regression line of ARd residuals of human faeces metagenomes on VFd residuals and the
542 slope is 0.63. The difference in slopes is significant, but not very significant, with p-value of
543 3e-02.

544 **Figure 4. Distribution of AR by VF diversity in Human gut metagenomes.**

545 Scatter plot ARd residuals versus VFd residuals of each human gut 110 metagenomes.
546 Residuals are taken from the regressions shown in Figure 2.A and 2.B. Each dot represents a
547 metagenome. In 2.A (resp.2B), the black lines represent the regression of the ARd residuals
548 on VFd residuals for all the metagenomes. The green line represents the linear regression of

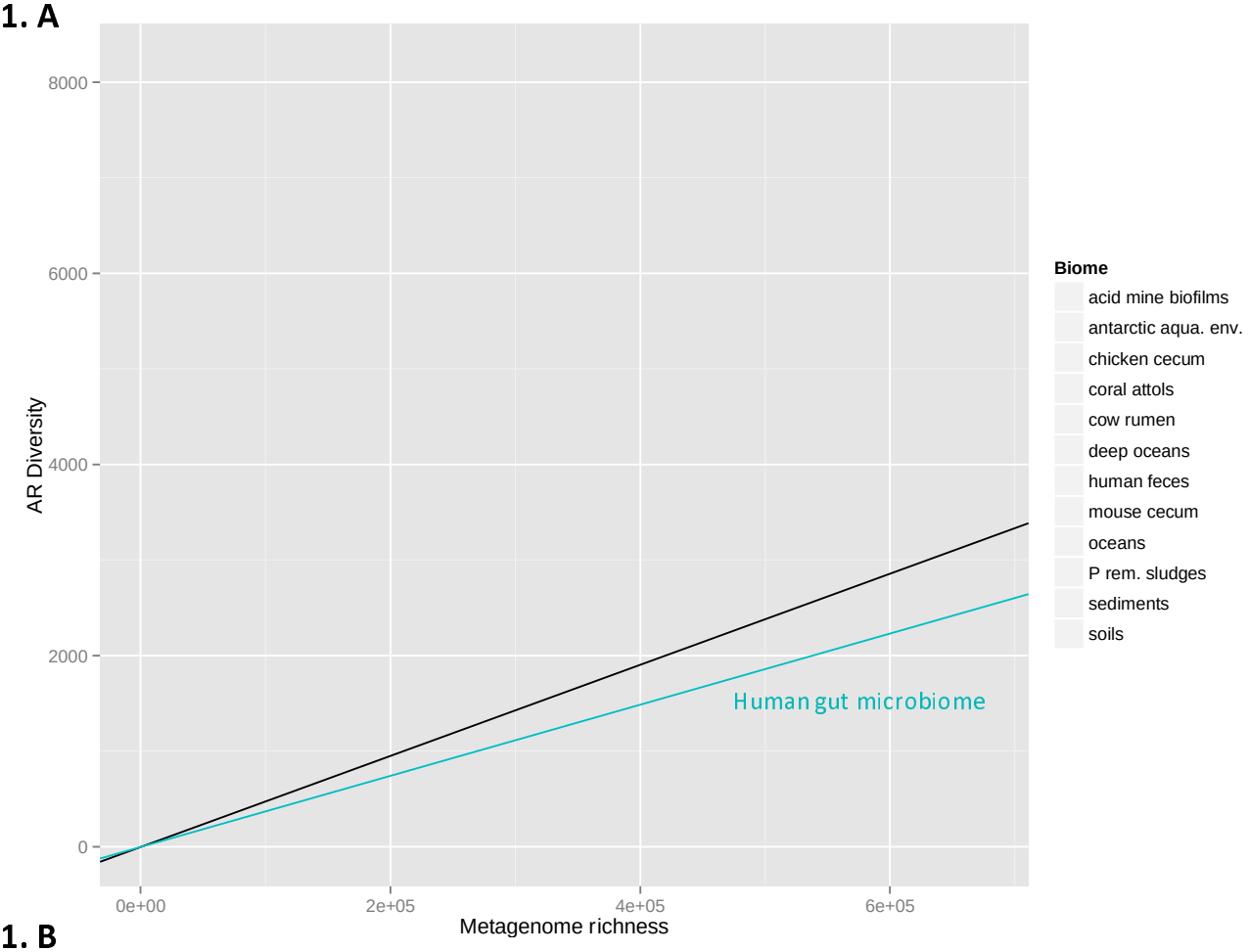
549 the ARd residuals on VFd residuals for the USA metagenomes subset, the red line is the
550 regression line for the Malawian metagenomes, and the blue line is the one for the
551 Venezuelan metagenomes.

552 **Figure 5. Correlogram of VR and AR protein families.**

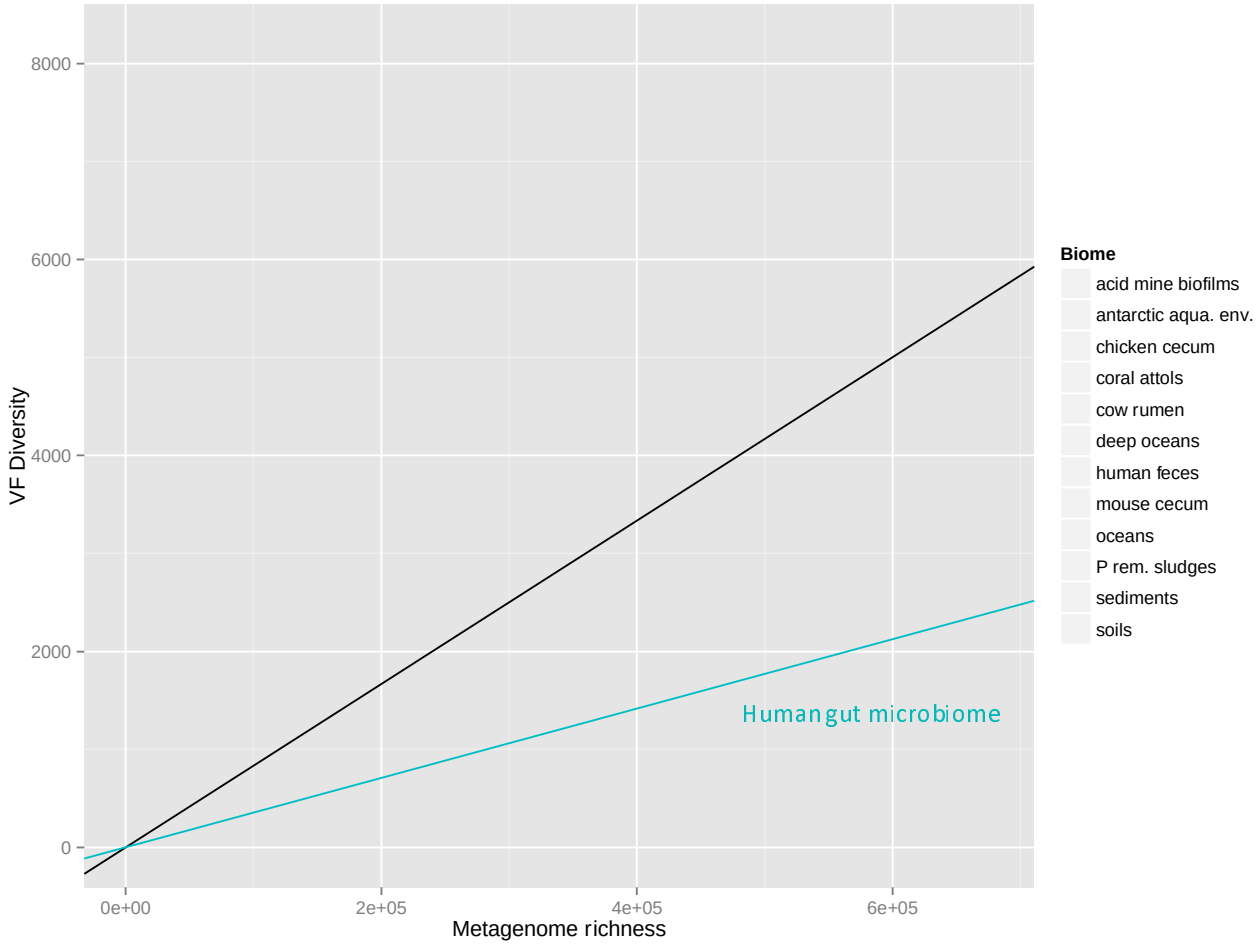
553 Partial correlations for AR and VF protein families given metagenome richness have been
554 computed. The color indicate negative (blue) or positive (red) partial correlation. Intensity of
555 the color indicates the value. The point sizes indicate the significances of the partial
556 correlation (white squares indicate no significant correlation).

557

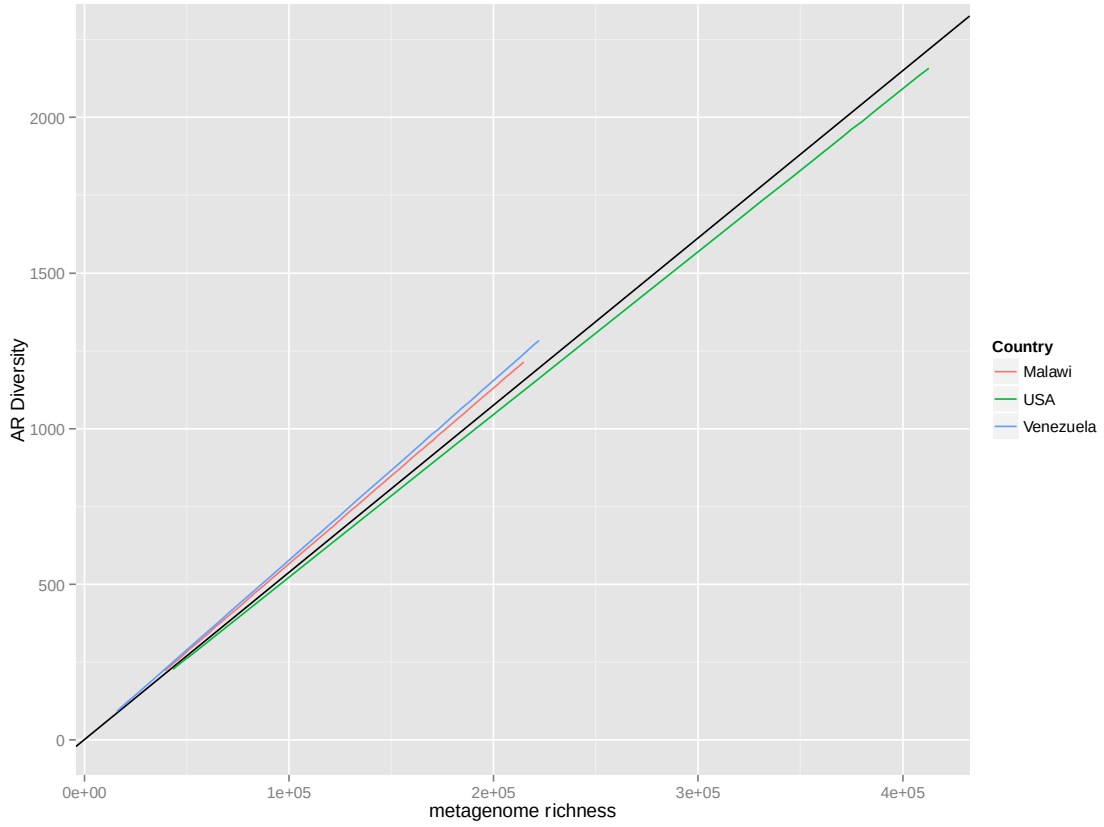
1. A


8000
6000
4000
2000
0
AR Diversity
0e+00
2e+05
4e+05
6e+05
Metagenome richness
Human gut microbiome
Biome
acid mine biofilms
antarctic aqua. env.
chicken cecum
coral attols
cow rumen
deep oceans
human feces
mouse cecum
oceans
P rem. sludges
sediments
soils

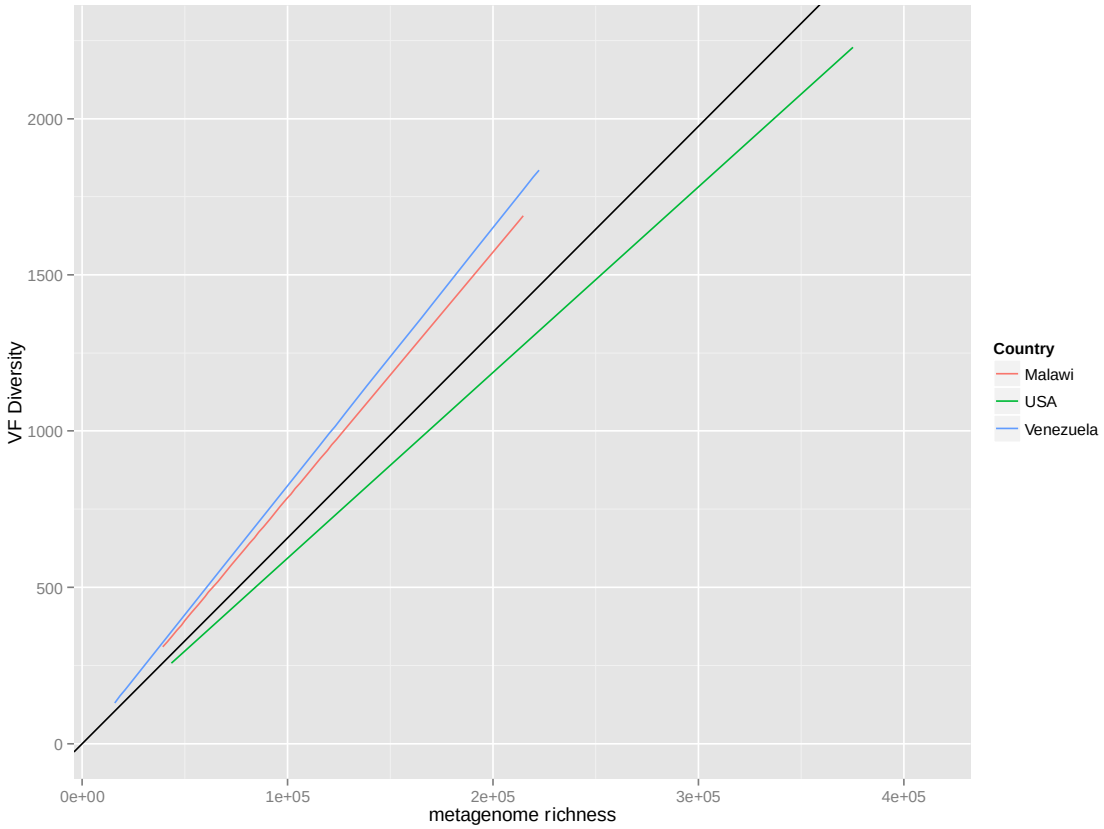
1. B


8000
6000
4000
2000
0
VF Diversity
0e+00
2e+05
4e+05
6e+05
Metagenome richness
Human gut microbiome
Biome
acid mine biofilms
antarctic aqua. env.
chicken cecum
coral attols
cow rumen
deep oceans
human feces
mouse cecum
oceans
P rem. sludges
sediments
soils

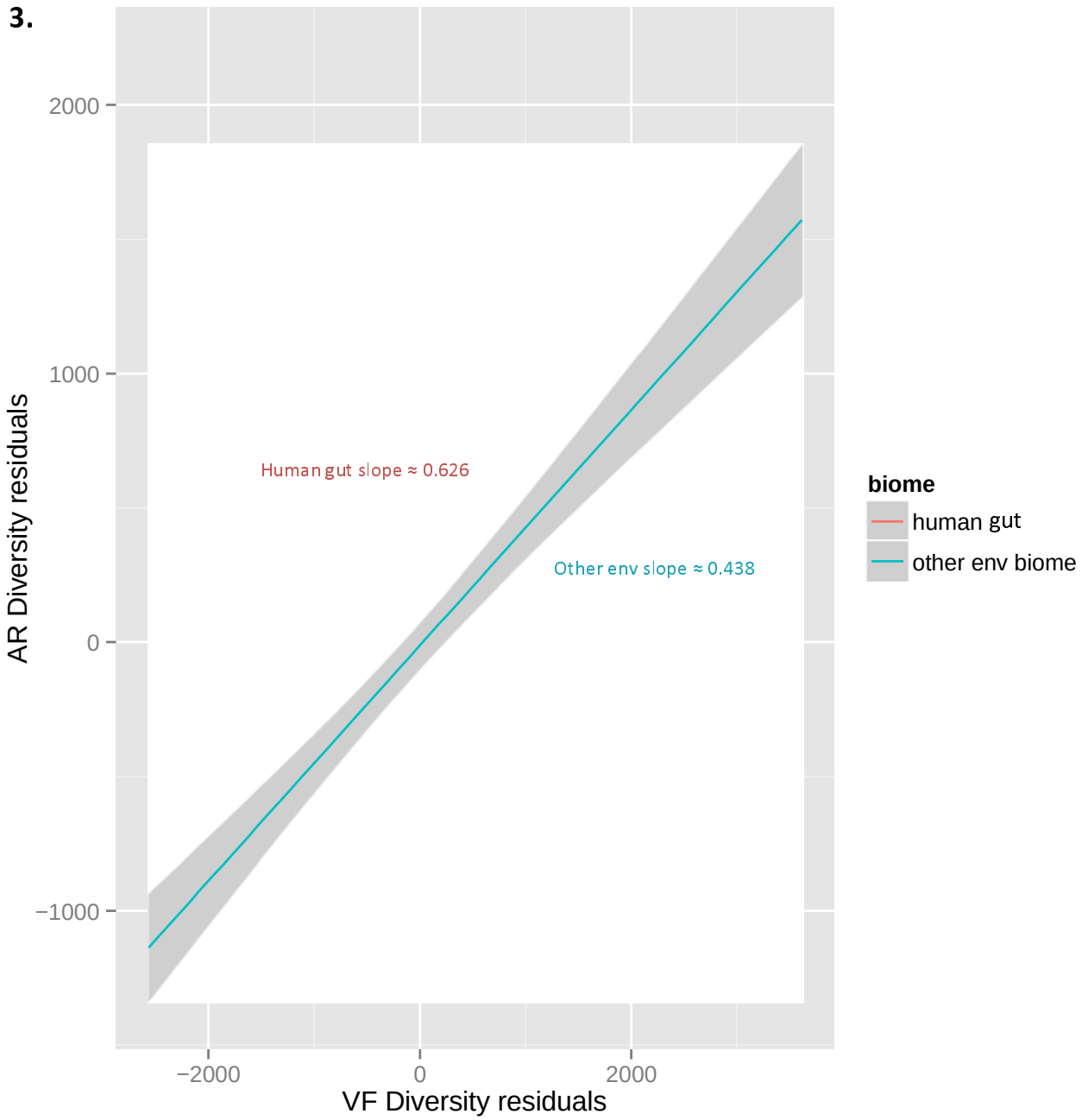
2. A


AR Diversity
2000
1500
1000
500
0
0e+00
1e+05
2e+05
3e+05
4e+05
metagenome richness
Country
Malawi
USA
Venezuela

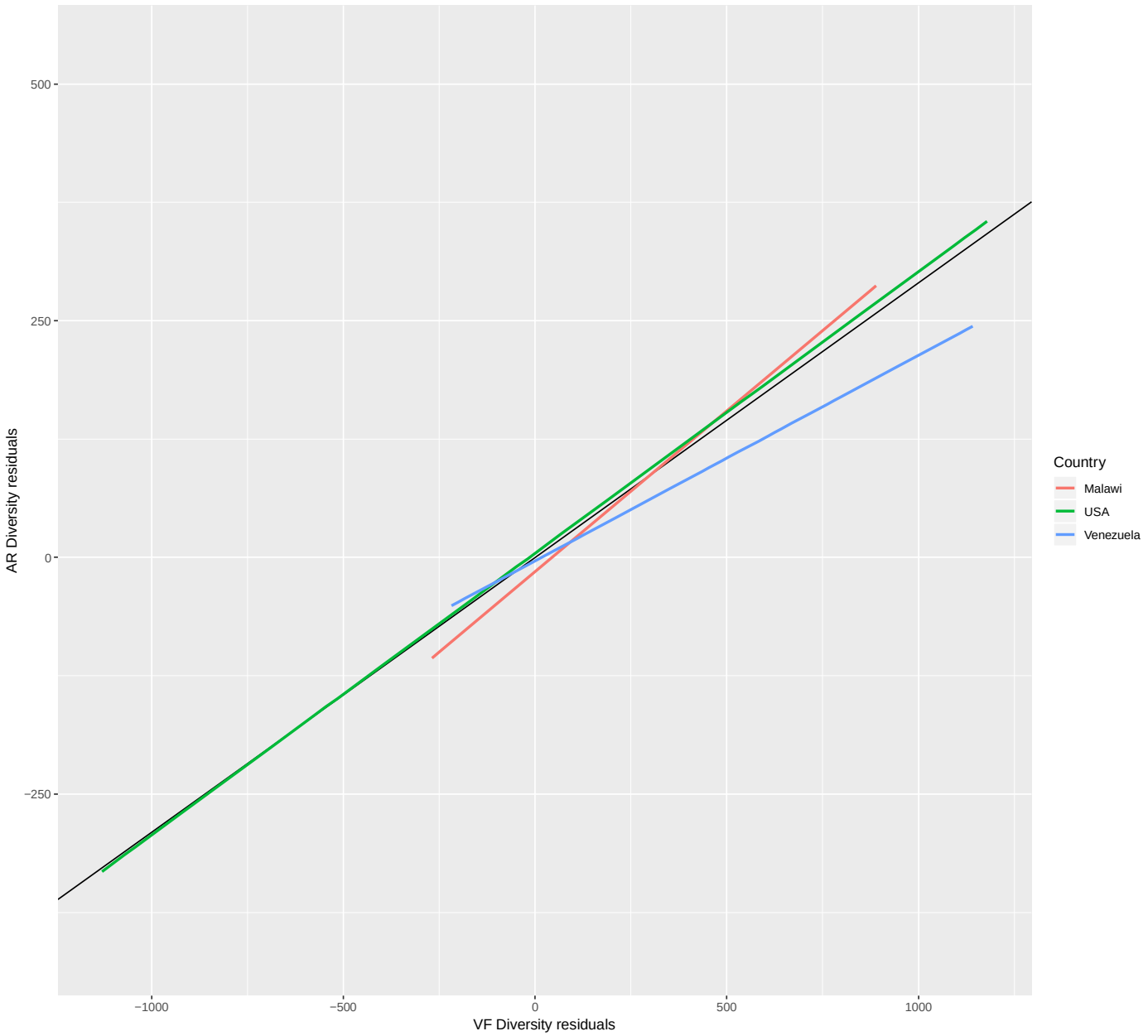
2. B


VF Diversity
2000
1500
1000
500
0
0e+00
1e+05
2e+05
3e+05
4e+05
metagenome richness
Country
Malawi
USA
Venezuela

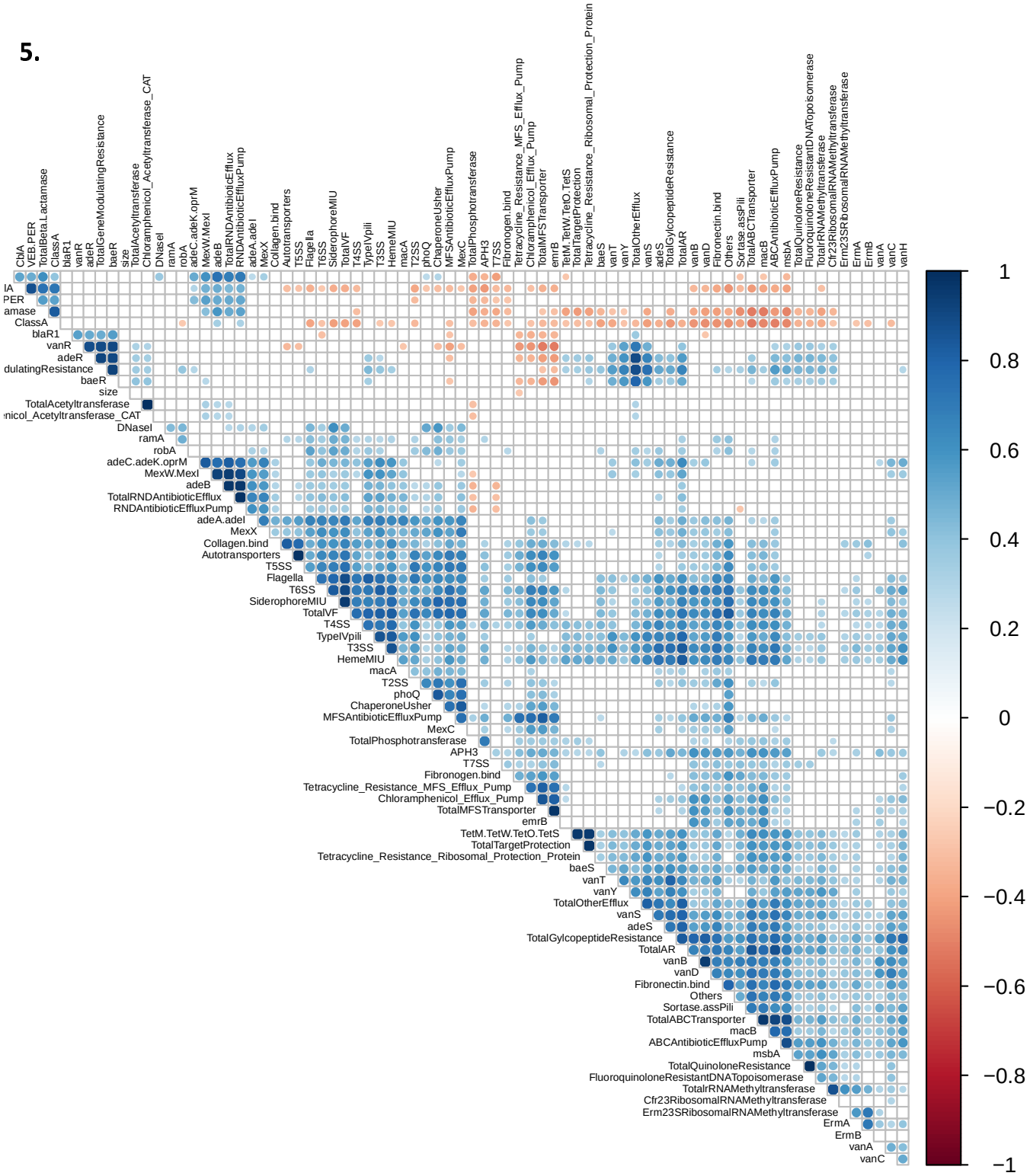
3.


2000
1000
0
−1000
AR Diversity residuals
Human gut slope ≈ 0.626
Other env slope ≈ 0.438
−2000
0
2000
VF Diversity residuals
biome
human gut
other env biome

4.


500
250
0
−250
AR Diversity residuals
−1000
−500
0
500
1000
VF Diversity residuals
Country
Malawi
USA
Venezuela

5.


CblA
VEB.PER
TotalBeta.Lactamase
ClassA
blaR1
vanR
adeR
TotalGeneModulatingResistance
baeR
size
TotalAcetyltransferase
Chloramphenicol_Acetyltransferase_CAT
DNaseI
ramA
robA
adeC.adeK.oprM
MexW.MexI
adeB
TotalRNDAntibioticEfflux
RNDAntibioticEffluxPump
adeA.adeI
MexX
Collagen.bind
Autotransporters
T5SS
Flagella
T6SS
SiderophoreMIU
TotalVF
T4SS
TypeIVpili
T3SS
HemeMIU
macA
T2SS
phoQ
ChaperoneUsher
MFSAntibioticEffluxPump
MexC
TotalPhosphotransferase
APH3
T7SS
Fibronogen.bind
Tetracycline_Resistance_MFS_Efflux_Pump
Chloramphenicol_Efflux_Pump
TotalMFSTransporter
emrB
TetM.TetW.TetO.TetS
TotalTargetProtection
Tetracycline_Resistance_Ribosomal_Protection_Protein
baeS
vanT
vanY
TotalOtherEfflux
vanS
adeS
TotalGylcopeptideResistance
TotalAR
vanB
vanD
Fibronectin.bind
Others
Sortase.assPili
TotalABCTransporter
macB
ABCAntibioticEffluxPump
msbA
TotalQuinoloneResistance
FluoroquinoloneResistantDNATopoisomerase
TotalrRNAMethyltransferase
Cfr23RibosomalRNAMethyltransferase
Erm23SRibosomalRNAMethyltransferase
ErmA
ErmB
vanA
vanC
vanH
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1