Utilizing co-abundances of antimicrobial resistance genes to identify potential co-selection in the resistome

Authors:

Hannah-Marie Martiny[1]

Patrick Munk[1]

Christian Brinch[1]

Frank M. Aarestrup[1]

M. Luz Calle[2]

Thomas N. Petersen[1]

Affiliations:

[1]Research Group for Genomic Epidemiology, Technical University of Denmark, Kongens Lyngby, Denmark

[2]Biosciences Department, Faculty of Sciences and Technology, University of Vic - Central University of Catalonia, Vic, Spain

Abstract

The rapid spread of antimicrobial resistance (AMR) is a threat to global health, and the nature of co-occurring antimicrobial resistance genes (ARGs) may cause collateral AMR effects once antimicrobial agents are used. Therefore, it is essential to identify which pairs of ARGs co-occur. Given the wealth of NGS data available in public repositories, we have investigated the correlation between ARG abundances in a collection of 214,095 metagenomic datasets. Using more than $6.76 \cdot 10^8$ read fragments aligned to ARGs to infer pairwise correlation coefficients, we found that more ARGs correlated with each other in human and animal sampling origins than in soil and water environments. Furthermore, we showed that the correlations serve as risk profiles of resistance co-occurring to critically important antimicrobials. Using these profiles, we found several key ARGs indirectly but

strongly selecting for ARGs of critical importance, such as tetracycline ARGs correlating with most forms of resistances. In conclusion, this study highlights the important ARG players indirectly involved in shaping the resistomes of various environments that can serve as monitoring targets in AMR surveillance programs.

Introduction

Antimicrobial resistance (AMR) is one of the biggest threats to human and animal health[1,2], and it is widely acknowledged that the misuse of antimicrobials has accelerated the dissemination and prevalence of antimicrobial resistance genes (ARGs)[3]. Most attempts to reduce the burden of AMR have focused on reducing the use of single classes of antimicrobial agents considered of critical importance[4]. Despite considerable efforts in various settings, such as livestock, these regulations have not significantly reduced the spread of ARGs[5,6]. It is known that even after banning the use of specific antimicrobials, the ARGs conferring resistance will still be prevalent in the environment[7]. Furthermore, studies have shown that ARGs can be indirectly selected if they co-occur with an ARG conferring resistance to the antimicrobial, causing the selective pressure[6,8–10]. We have also recently observed that changes in the selective pressure of even a single antimicrobial agent influence several ARGs in pig metagenomes[10]. It is, however, not known whether this is due to co-selection due to the genetic linkage of ARGs, the presence of ARGs in the same bacterial clones or because different bacterial species with different ARGs are co-selected in a microbial network.

Most studies associating AMU and AMR have, however, focused on one antimicrobial agent at that time and how the use impacts the development of resistance to that agent and not to other unrelated antimicrobials. This needs to be understood better to stop these collateral damage effects[11,12]. Co-occurrence of microbes has been evaluated in soil[13,14] and marine environments[15,16], whereas ARG co-occurrences have been studied in sewage sludge[17], freshwater[18], marine[19], swine[7], and cattle[20]. However, most of these studies have only been evaluated on a smaller scale and not always using the same methods. There are currently a large collection of next-generation sequencing datasets from metagenomic samples available in public repositories, providing an excellent resource to quantify the prevalence of ARGs by analyzing the read abundances [21–24].

59

60 In this study, we have studied the co-abundance of sequencing reads aligned to ARGs to
61 assess how resistance to one specific antimicrobial agent is linked to the abundance of
62 another class of antimicrobial. With a collection of 214,095 metagenomic datasets[25], we
63 examined the correlation between pairwise ARG read abundances with a compositional
64 approach[26,27] using SparCC (Sparse Correlations for Compositional data)[28]. Our results
65 demonstrated that many ARG pairs interact but are highly specific to the environment. We
66 believe that these interactions provide a new foundation to understand how ARGs are being
67 co-selected independently of the microbial context, and the findings can be used to design
68 targeted interventions to limit the spread of AMR.

69

70 Methods

71 Data collection and pre-processing

72 We have previously described in detail the process of downloading and analyzing 214,095
73 metagenomic samples[23,25], but in brief: We downloaded raw sequencing reads corresponding
74 to 442 Tbp from metagenomic samples deposited in the European Nucleotide Archive[29] that
75 were uploaded between 2010-01-01 and 2020-01-01 and had at least 100,000 reads and were
76 shotgun sequenced. The raw sequencing reads were quality-checked with FASTQC v.0.11.15
77 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed with BBduk2
78 36.49[30]. The trimmed reads were then globally aligned using KMA[31] 1.2.21 against two
79 reference sequence databases: ResFinder[32] (downloaded 25-01-2020) and Silva[33] (version 38,
80 downloaded 16-01-2020). ResFinder is a database of 3,085 acquired ARGs, whereas Silva is a
81 16S/18S rRNA database of 2,225,272 sequences.

82 Homology-reduced ResFinder

83 Since there is a possibility that KMA might assign reads to closely related homologs differently
84 across samples, we decided to homology reduce the 3,085 ARGs of ResFinder using
85 USEARCH[34] 11.0.667 with 90% nucleotide identity, producing a total of 716 ARG groups. For
86 each of these groups, the read counts for the genes in that group were first adjusted by the
87 length of the gene and then aggregated together. These new group counts are used in the

correlation analysis described in the next section. For the rest of the manuscript, we refer to these 716 ARG clusters as simply ARGs.

Calculating relative abundances

For a category label, we calculated the relative abundance of fragment counts assigned to different genes or classes as:

$$\text{Relative abundance}(x) = \frac{\kappa}{\sum n_i} n_i$$

where x is the label, n_i is the count of read fragments assigned to gene i, and $\kappa = 100$ is a scaling constant.

Inferring pairwise correlations with SparCC

The SparCC algorithm[28] was used to obtain correlations using pairs of log-ratio transformed ARG read counts to infer linear Pearson correlations. SparCC obtains linear Pearson correlations and p-values through an iterative approach that adjusts for spurious correlations and lowers the false discovery rate. The ARG-ARG correlations were inferred as the average over 50 iterations, and one-sided pseudo p-values were obtained through a bootstrapping procedure of 100 rounds. In each bootstrapping round, the input count matrix was shuffled and correlations were averaged over 10 iterations to infer one-sided p-values to test whether the correlation for the observed data was statistically significant. Correlations values ≥ 0.6 with p-values ≤ 0.01 were selected for further analysis. We implemented SparCC to run on GPUs on the Danish National Supercomputer for Life Sciences (https://www.computerome.dk).

We ran SparCC on the entire dataset of the 214K metagenomic samples and subsets of samples grouped by sampling host and environment, where at least 800 samples existed. Due to inconsistent labeling of the sampling sources, we made new source groups, as shown in Table 1. We only consider genes for SparCC analysis that are present in at least 10 samples with a minimum read fragment count of 50. In total, 11 different correlation matrices were made, i.e., one for each of the source groups listed in Table 1. The correlation networks were visualized in R 4.1.0[35] with packages igraph[36], qgraph[37], and ggraph[38] using the Fruchterman-Reingold layout algorithm[39].

Source group	Sampling source label(s)	Number of samples	Number of samples with ARGs
Air	Air metagenome	914	870
Chicken	*Gallus gallus* (1,219), chicken gut metagenome (4)	1,223	1,215
Cow	*Bos taurus* (872), cow dung metagenome (14)	886	824
Dog	*Canis lupus familiaris*	3,439	3,182
Freshwater	Freshwater metagenome	4,494	585
Human	*Homo sapiens*	95,003	57,239
Marine	Marine metagenome	30,002	5,444
Mouse	*Mus musculus* (1,462), mouse metagenome (50), mouse gut metagenome (2,435)	3,947	3,909
Pig	*Sus scrofa* (673), *Sus scrofa domesticus* (355), pig metagenome (2,129), pig gut metagenome (72)	3,229	3,461
Soil	Soil metagenome	6,533	2,822
All		214,095	119,206

Table 1: Grouped labels for hosts and environments. The parenthesis after each sampling source label denotes the number of samples assigned to that label if the group consisted of multiple labels.

Network comparisons

The topology of the different networks is described using different metrics: the number of nodes (N) and edges (E), the global clustering coefficient, network density, edge density, and the number of components. The global clustering coefficient, or the graph transitivity, measures the density of node triplets in the network[40]. The network density is calculated as $2E/N(N-1)$ as given in Parente et al. (2018)[41]. Edge density is the number of edges over the number of possible edges[42]. The average correlation between ARGs of two antimicrobial classes was calculated using Fisher's z-transformation on the correlation values, averaging the z-scores and converting it back to a correlation score with the inverse Fisher transformation[43].

Data and code availability

The matrix of raw ARG read counts is available at https://zenodo.org/record/6919377. The code used to run the analysis and create figures and the output files of SparCC are available at https://github.com/hmmartiny/global_resistome_correlations. Classifications of antimicrobial importance were retrieved from the 6th revision of critically important antimicrobials for human medicine from https://www.who.int/publications/i/item/9789241515528, accessed 2022-10-10.

140 # Results

141 This study investigated the correlation of pairwise ARG abundances across a highly diverse
142 set of 214,095 metagenomic samples. The samples included in the 214K metagenomic
143 collection come from a variety of backgrounds, such as the sampling period spanning
144 between 2000 and 2020 (Figure S1a), primarily sequenced on Illumina platforms (Figure
145 S1b), and hailing from all over the world (Figure S1c). We observed that not all samples
146 contained sequence read fragments aligned to ARGs, so we only included those that did in
147 the correlation analyses (n = 119,206; Table 1). Looking at the various sampling sources,
148 read fragments aligned to ARGs conferring resistance to tetracycline and beta-lactams were
149 very common in host-associated sources, whereas phenicol resistance was more frequently
150 observed in environmental sources (freshwater, marine, and soil) (Table 2). *catA1* was the
151 most dominant gene in the environmental samples, especially in marine samples, whereas
152 various *tet* genes had high relative abundances in livestock or human samples: *tet(W)* in
153 chicken and pig samples; *tet(Q)* in cow, human, and pig metagenomes. *blaTEM-52B*
154 accounted for more than 30% of the read fragments assigned to ARGs in air metagenomes
155 (Figure S2).
156

157 ## Balancing the sparsity and network complexity

158 We selected ARGs to infer abundance-based correlations by considering how sparse the
159 input count matrix was and how that sparsity affected the results of SparCC. Out of the 716
160 ARGs that could be included, we observed that without filtering the raw counts, SparCC
161 found that most of the ARGs correlated with each other, even if the count of read fragments
162 was low. Based on this observation, we decided to require that for an ARG to be included, it
163 had to have a minimum count across a specific number of samples for a group. Applying
164 these two filters, the number of correlations inferred for each sampling group decreased as
165 compared to the no-filter results (Table S1). We decided to pick filter settings to balance the
166 amount of sparsity allowed in the sample groups and the number of correlation coefficients,
167 which ended up being a minimum fragment count of 50 across at least 10 samples.
168

169 ## Analysis of large-scale metagenomic correlation networks

170 Based on our filter settings, we constructed a global network using the correlation
171 coefficients for the entire collection of metagenomes, with each node representing an ARG
172 and each edge representing a pairwise ARG connection (correlation ≥ 0.6, p-value < 0.01,
173 Figure S3). The global network, nicknamed 'all', contained 225 ARGs connected through
174 2,344 correlation edges (Figure 1a). As this all-network was hard to interpret due to the many
175 highly interconnected ARGs, we also inferred pairwise ARG correlations in specific sampling
176 groups (Table 1). The genes that were part of these networks were found to correlate with
177 varying degrees of strength (Figure 1a-b). For example, the human network contained many
178 correlation coefficients, but most were less than 0.8 (Figure 1c). Another example is the
179 marine network, where only a few ARGs were found to correlate, but with values above 0.9
180 (Figure 1b-c). Despite the networks reflecting the composition of the various environments,
181 we still observed overlaps between which ARGs were found to correlate. One example is
182 that all the correlations inferred from the pig metagenomes also existed in the human
183 metagenomic network (Figure 1d).
184
185 There was a limited number of correlations for ARGs encoding resistance to
186 fluoroquinolones, steroid antibiotics (fusidic acid), colistin, and rifampicin. On the other
187 hand, beta-lactam, tetracycline, and aminoglycoside ARGs had many correlations with each
188 other and with other classes (Figure S4, Figure S6). Despite resistance to some antimicrobial
189 classes being the most abundant, the ARGs did not always correlate to many others. For
190 example, tetracycline ARGs were the most abundant in dog metagenomes, but no
191 correlations were inferred for these ARGs. Similarly, in the human samples where
192 aminoglycoside and beta-lactam ARGs were less abundant than tetracycline ARGs,
193 aminoglycoside and beta-lactam resistance genes had a higher number of correlations
194 coefficients were reported (Table 2, Figure S4).
195
196 On the level of ARG abundance, we observed that just because an ARG was highly abundant
197 in a sampling group, it did not automatically mean that it correlated to many other ARGs.
198 The highly abundant *catA1* gene in marine (97.4% of all ARG reads), freshwater (55.1%), and
199 soil samples (84.7%) (Figure S2) did only correlate with one or two other genes in the water
200 environmental networks, and none in the soil network. On the other hand, *catA1* did seem

201 to be correlated with 15 other genes in the pig correlation network despite not being highly
202 abundant in that group of samples (Figure S5a). *mef(A)*_1 accounted for 15.9% of the reads
203 aligned to ARGs in cow samples and 6.63% in pigs (Figure S2) and was also strongly
204 correlating with other genes, which mainly conferred resistance to aminoglycosides,
205 (fluoro)quinolones and tetracyclines (Figure S5b). *tet(L)_4* only accounted for 4.01% of the
206 read fragments aligned to ARGs in metagenomes of chicken origins (Figure S2) but was
207 shown to correlate in its abundance with the abundance of 8 other ARGs, e.g., with a
208 correlation of 0.77 with *lnu(A)_1 (*Figure S5c*).*
209

210 The hidden signals between ARGs profile the potential risk of co-selection in different
211 environmental contexts

212 Using the correlations between ARGs in the various environments (Figure 1), we calculated
213 the average correlation between ARGs of different antimicrobial classes (Figure S6). These
214 average correlations can then serve as profiles to assess the risk of indirectly selecting ARGs
215 that gives resistance to different ARGs through co- and cross-resistance. These risk profiles
216 can then be used to judge the strength of interactions upon using one antimicrobial in each
217 setting. Upon constructing these profiles, we observed that the strength and the number of
218 correlations highly depend on the antimicrobial classes and the environmental
219 contextFigure S6. We hypothesize that if an important (IA) or highly important antimicrobial
220 (HIA) is used, first, the resistance to the antimicrobial class will likely flourish, and, secondly,
221 through co- and cross-resistance, so will ARGs conferring resistance to other classes,
222 including those that are critically important antimicrobials (CIA).
223

224 Figure 2 shows two risk profiles for ARG correlations for two highly critical important
225 antimicrobials glycopeptide and macrolide. Correlations between ARGs of glycopeptide
226 resistance to other resistance classes were much rarer than those connected with macrolide
227 ARGs (Figure S6) and those correlations that were observed were relatively low
228 (correlation<0.8, Figure 2). Different vancomycin resistance cassettes were responsible in
229 different environments, namely *VanHAX*, *VanC2*, and *VanX_bc* in human samples and
230 *VanHDX* and *VanC1XY* in pig samples (Figure S7a). *VanHAX* only has one correlation,
231 whereas the remaining five correlate with many different ARGs.

232 On the contrary, ARGs conferring resistance to macrolide were much more interactive with
233 other classes of resistance genes in all networks and had strong correlations with specific
234 classes (correlation > 0.9, Figure 2, Figure S6). The macrolide ARGs co-abundant with other
235 ARGs were many but separated into distinct network clusters (Figure S7b). For example,
236 *mef(A)* and *msr(D)* were usually found together in different environments, both in small and
237 large clusters.
238
239 While Figure 2 highlights how CIA ARGs interact, it is just as important to investigate what
240 ARGs of less critically important antimicrobials correlate with. As seen in Figure 3, ARGs for
241 pleuromutilin resistance (IA) and for tetracycline resistance (HIA) were found to interact
242 with many other classes, including those that are critically important. For example,
243 pleuromutilin ARGs are few but well connected (Figure 3, Figure S6), as seen with the
244 connections with *lsa(E)* and *cfr(C)* (Figure S8a). Tetracycline ARGs correlated with the
245 abundance of multiple ARGs, such as those conferring resistance to lincosamides,
246 macrolides, and phenicols (Figure 3). While there were many ARGs for tetracycline
247 resistance, they often correlated to the same ARGs (Figure S8b).

248 Discussion

249 Considering the complexity of microbiomes, studying how microbial composition shapes the
250 distribution of ARGs is a challenging task but one that could shed light on how ARGs
251 indirectly select one another. However, with the high-throughput sequencing technologies
252 and many metagenomic datasets available in public repositories, it is now much more
253 feasible to extract the patterns of how ARGs co-occur without knowing their microbial
254 origin. Using our recently published collection of 214K metagenomic datasets[25], we have
255 inferred correlations of pairwise ARG abundances to profile which types of resistance
256 influence the shape of resistome and which genes are the key players. To the best of our
257 knowledge, this is the first study to relate ARG abundances on such a large and broad scale.
258
259 Our correlation networks revealed that not all ARGs are connected, as we observed that the
260 interactions were largely shaped by the composition of the environmental resistome.
261 Nonetheless, we did not always observe strong correlations if an ARG was highly abundant.
262 The all-network with correlation for all metagenomic samples was the most complex, which

263 we speculate is due to both the wide variety of sampling sources and the
264 overrepresentation of human metagenomes (Table 1). We found the differences in the
265 animal and environmental networks much more interesting, as they reflect the dynamics of
266 ARG abundances under exposure to different antimicrobials (Figure 1). On the level of which
267 antimicrobial class an ARG confers resistance to, we could also observe that some resistance
268 classes had more and stronger correlations (Figure S6). There are several cases of our
269 observed co-abundant ARGs that have been reported in the literature, and some of the
270 strongest correlation pairs have been detected due to the ARGs sitting together in the same
271 genome. For example, the gene cassette *vanHAX* has been found together with *msr(C)* and
272 *aac(6')* in genomes of human isolates[52] and *mef(A)* linked with *tet(O)*[53] and *mdf(A)* with
273 *blaTEM, aph(6), sul2, and, tet(A)*[54] (Figure S7).
274
275 As highlighted in Figure 2 and Figure 3, we argue that the correlations can serve as a way to
276 profile the risk of co-selection of ARGs occurring in a setting if exposed to an antimicrobial.
277 Some of these relationships between ARGs of different antimicrobial classes have been
278 observed in other studies.
279
280 Antimicrobials have been classified differently to reflect their importance to human
281 medicine, of which glycopeptides and macrolides are CIA with the highest priorities.
282 Glycopeptide and macrolide resistance has previously been linked genetically[6,8] in pigs,
283 where we also can report the presence of correlations between ARGs of glycopeptide and
284 macrolide resistances not only in pigs but also in human and mouse environments (Figure
285 2). Pleuromutilin and tetracycline antimicrobials have less critical importance of
286 antimicrobials [44], but many correlations for ARGs of pleuromutilin and tetracycline exist in
287 various networks (Figure 3) suggest that there are still risks associated with the use of these
288 two. Tetracycline resistance has been found to occur together with resistance to
289 macrolides[45–47], aminoglycosides[45,48], folate pathway antagonists[49,50], lincosamide[10], and
290 beta-lactams[48], to name a few studies. This high connectivity of tetracycline ARGs seems in
291 line with our results, as this specific group of ARGs was connected to almost all classes of
292 antimicrobials in most of our networks (Figure 3, Figure S8b). The variety of connections to
293 antimicrobials of less importance to human health should be more in focus, as our results

294 show that there are risks of critical antimicrobial resistances emerging from the enrichment
295 of less essential resistance genes.
296
297 Following this line of thought of focusing on ARGs that gives resistance to less critically
298 important antimicrobials, a recent study by Tarek and Garner (2022)[51] proposed to create a
299 monitoring framework based on isolated components, or clusters, in correlation networks.
300 They constructed a correlation network encompassing ARG abundances in samples from
301 wastewater treatment plants and argued that representative ARG members from each
302 cluster in their network should be monitoring targets. Our networks suggest that limiting to
303 only a few ARGs would fail to capture the complete picture in some environments, such as
304 human microbiomes (Figure 1). However, as our risk profiles show, limiting the set of
305 monitoring targets by only focusing on what happens to ARG abundances during exposure
306 to one type of antimicrobials would be a way to implement a monitoring system (Figures
307 Figure 2Figure 3, Figure S7-Figure S8). If a monitoring system is implemented, it would need
308 to be updated regularly to show the changes in antimicrobial usage and ARGs co-
309 occurrences since the correlations we have inferred in this study only reflects the current
310 and past usage of antimicrobials.
311
312 In order to use correlation networks for surveillance of AMR, more work is needed to
313 confirm that the observed interactions do indeed exist in nature[55]. We have defined
314 interactions as being indirect since we need to determine the degree that co-abundance is
315 explained by physical linkage on a genome or plasmid or sharing of a cell. To investigate the
316 direction of correlation, for example, whether *vanHAX* influences *msr(C)* or the other way
317 around (Figure S7a), the SPIEC-EASI[56] method could be used to infer such directional
318 dependences. A directional correlation could be included in a risk profile.
319
320 We have only included ARGs of different resistance classes that are co-abundant in our
321 analysis, but the abundance of bacterial reads could also have been analyzed. By including
322 the bacterial counts, it would likely be possible to understand how the dynamics of
323 microbiomes and resistomes are linked. Similarly, redoing the alignment procedure to
324 include the count of reads aligned to other drivers of resistance, e.g., mobile genetic
325 elements, could also lead to the discovery of new patterns. There has been much work on

326 linking the prevalence of ARGs with antimicrobial usage (AMU) in various settings[6,8,10], and
327 incorporating AMU could confirm our speculation that many of the co-abundances are
328 driven by selective pressures of various antimicrobials. Pooling the various data types
329 together in this kind of analysis would be a complicated task that nevertheless would shed
330 light on how the abundance and selective pressures contribute to the shape of resistomes.
331
332 By utilizing the wealth of information on ARG abundances available in a collection of 214K
333 metagenomic datasets, we have studied the co-abundance of ARGs to discover how these
334 interactions shape the prevalence of resistances in different environments. The inferred
335 correlation networks provide insights into how two resistance types indirectly and species-
336 independent select for each other in different habitats. Our results further highlight that
337 there are instances of genes of one type of resistance often co-occurring with many other
338 types of resistance and that the environmental context plays an important role, revealing
339 them as important targets in surveillance programs to limit their impact on global health.
340

341 ## References

342 1. O'Neill, J. & The Review on Antimicrobial Resistance (Chaired by Jim O'Neill). Tackling
343 Drug-Resistant Infections Globally: Final Report and Recommendations. *Rev.*
344 *Antimicrob. Resist.* 1–80 (2016).
345 2. Murray, C. J. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a
346 systematic analysis. *Lancet* **399**, 629–655 (2022).
347 3. Holmes, A. H. *et al.* Understanding the mechanisms and drivers of antimicrobial
348 resistance. *Lancet* **387**, 176–187 (2016).
349 4. Collignon, P., Powers, J. H., Chiller, T. M., Aidara-Kane, A. & Aarestrup, F. M. World
350 health organization ranking of antimicrobials according to their importance in human
351 medicine: a critical step for developing risk management strategies for the use of
352 antimicrobials in food production animals. *Clin. Infect. Dis.* **49**, 132–141 (2009).
353 5. Jensen, H. H. & Hayes, D. J. Impact of Denmark's ban on antimicrobials for growth
354 promotion. *Curr. Opin. Microbiol.* **19**, 30–36 (2014).
355 6. Aarestrup, F. M. *et al.* Effect of abolishment of the use of antimicrobial agents for
356 growth promotion on occurrence of antimicrobial resistance in fecal enterococci from

357 food animals in Denmark. *Antimicrob. Agents Chemother.* **45**, 2054–2059 (2001).
358 7. Sun, J. *et al.* Development of aminoglycoside and β-lactamase resistance among
359 intestinal microbiota of swine treated with lincomycin, chlortetracycline, and
360 amoxicillin. *Front. Microbiol.* **5**, 580 (2014).
361 8. Aarestrup, F. M. Characterization of glycopeptide-resistant Enterococcus faecium
362 (GRE) from broilers and pigs in Denmark: Genetic evidence that persistence of GRE in
363 pig herds is associated with coselection by resistance to macrolides. *J. Clin. Microbiol.*
364 **38**, 2774–2777 (2000).
365 9. Looft, T. *et al.* In-feed antibiotic effects on the swine intestinal microbiome. *Proc.*
366 *Natl. Acad. Sci. U. S. A.* **109**, 1691–1696 (2012).
367 10. Andersen, V. D. *et al.* Predicting effects of changed antimicrobial usage on the
368 abundance of antimicrobial resistance genes in finisher' gut microbiomes. *Prev. Vet.*
369 *Med.* **174**, 104853 (2020).
370 11. Peterson, L. R. Squeezing the antibiotic balloon: The impact of antimicrobial classes
371 on emerging resistance. *Clin. Microbiol. Infect. Suppl.* **11**, 4–16 (2005).
372 12. Paterson, D. L. 'Collateral damage' from cephalosporin or quinolone antibiotic
373 therapy. *Clin. Infect. Dis.* **38**, 341–345 (2004).
374 13. Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to
375 explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351
376 (2012).
377 14. Yuan, M. M. *et al.* Climate warming enhances microbial network complexity and
378 stability. *Nat. Clim. Chang.* **11**, 343–348 (2021).
379 15. Zhang, Z. *et al.* The large-scale spatial patterns of ecological networks between
380 phytoplankton and zooplankton in coastal marine ecosystems. *Sci. Total Environ.* **827**,
381 154285 (2022).
382 16. Li, L. G., Xia, Y. & Zhang, T. Co-occurrence of antibiotic and metal resistance genes
383 revealed in complete genome collection. *ISME J.* **11**, 651–662 (2017).
384 17. Ju, F. *et al.* Antibiotic resistance genes and human bacterial pathogens: Co-
385 occurrence, removal, and enrichment in municipal sewage sludge digesters. *Water*
386 *Res.* **91**, 1–10 (2016).
387 18. Pan, X., Lin, L., Zhang, W., Dong, L. & Yang, Y. Metagenome sequencing to unveil the
388 resistome in a deep subtropical lake on the Yunnan-Guizhou Plateau, China. *Environ.*

389 *Pollut.* **263**, 114470 (2020).
390 19. Wu, J. *et al.* Prevalence and distribution of antibiotic resistance in marine fish farming
391 areas in Hainan, China. *Sci. Total Environ.* **653**, 605–611 (2019).
392 20. Liu, J. *et al.* The fecal resistome of dairy cattle is associated with diet during nursing.
393 *Nat. Commun.* **10**, (2019).
394 21. Munk, P. *et al.* Abundance and diversity of the faecal resistome in slaughter pigs and
395 broilers in nine European countries. *Nat. Microbiol.* **3**, 898–908 (2018).
396 22. Hendriksen, R. S. *et al.* Global monitoring of antimicrobial resistance based on
397 metagenomics analyses of urban sewage. *Nat. Commun.* **10**, (2019).
398 23. Martiny, H.-M. *et al.* Global distribution of mcr gene variants in 214,095 metagenomic
399 samples. *mSystems* (2022). doi:10.1128/msystems.00105-22
400 24. Karkman, A., Do, T. T., Walsh, F. & Virta, M. P. J. Antibiotic-Resistance Genes in Waste
401 Water. *Trends Microbiol.* **26**, 220–228 (2018).
402 25. Martiny, H.-M., Id, P. M., Id, C. B., Aarestrup, F. M. & Petersenid, T. N. A curated data
403 resource of 214K metagenomes for characterization of the global antimicrobial
404 resistome. *PLOS Biol.* **20**, e3001792 (2022).
405 26. Aitchison, J. The Statistical Analysis of Compositional Data. *J. R. Stat. Soc. Ser. B* **44**,
406 139–160 (1982).
407 27. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome
408 datasets are compositional: And this is not optional. *Front. Microbiol.* **8**, 1–6 (2017).
409 28. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey Data.
410 *PLoS Comput. Biol.* **8**, 1–11 (2012).
411 29. Leinonen, R. *et al.* The European nucleotide archive. *Nucleic Acids Res.* **39**, 44–47
412 (2011).
413 30. Bushnell, B. BBMap. (2014).
414 31. Clausen, P. T. L. C., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw
415 reads against redundant databases with KMA. *BMC Bioinformatics* **19**, 1–8 (2018).
416 32. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J.*
417 *Antimicrob. Chemother.* **67**, 2640–2644 (2012).
418 33. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data
419 processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).
420 34. Edgar, R. C. & Bateman, A. Search and clustering orders of magnitude faster than

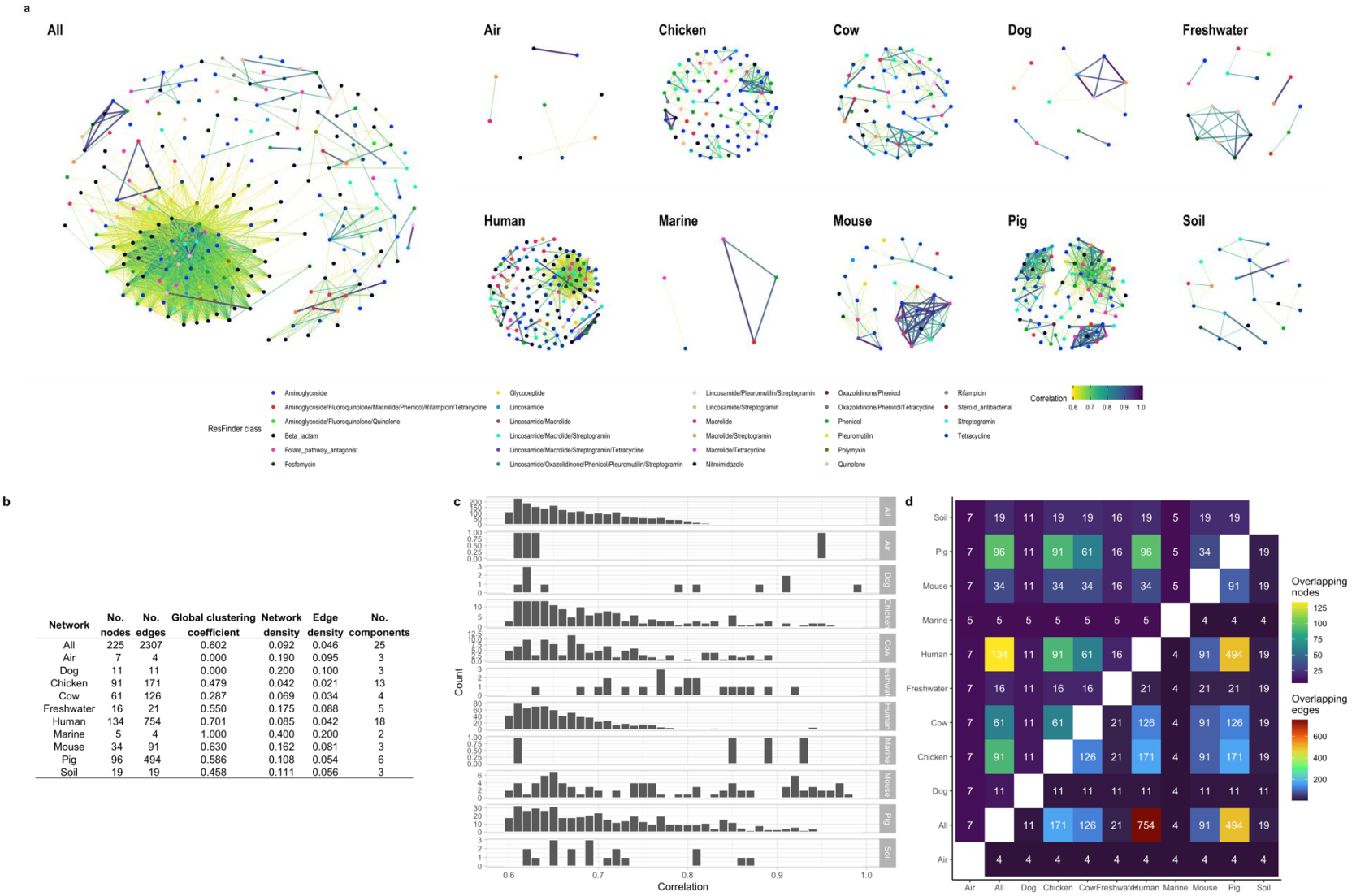
421 BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
422 35. Team, R. C. R: A Language and Environment for Statistical Computing. (2021).
423 36. Csardi, G. & Nepusz, T. The igraph software package for complex network research.
424 *InterJournal* **Complex Sy**, 1695 (2006).
425 37. Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D. & Borsboom, D.
426 qgraph: Network Visualizations of Relationships in Psychometric Data. *J. Stat. Softw.*
427 **48**, 1–18 (2012).
428 38. Pedersen, T. L. ggraph: An Implementation of Grammar of Graphics for Graphs and
429 Networks. (2021).
430 39. Fructherman, T. M. . & Reingold, E. M. Graph Drawing by Force-directed Placement.
431 *Softw. - Pract. Exp.* **21**, 1129–1164 (1991).
432 40. Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of
433 complex weighted networks. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 3747 (2004).
434 41. Parente, E., Zotta, T., Faust, K., De Filippis, F. & Ercolini, D. Structure of association
435 networks in food bacterial communities. *Food Microbiol.* **73**, 49–60 (2018).
436 42. Wasserman, S. & Faust, K. *Social network analysis: Methods and applications.*
437 (Cambridge University Press, 1994).
438 43. Silver, N. C. & Dunlap, W. P. Averaging Correlation Coefficients: Should Fisher's z
439 Transformation Be Used? *J. Appl. Psychol.* **72**, 146–148 (1987).
440 44. Group, W. H. O. A., Surveillance, I. & Resistance, A. *WHO | WHO list of Critically*
441 *Important Antimicrobials (CIA)*.
442 45. Arredondo, A., Blanc, V., Mor, C., Nart, J. & León, R. Tetracycline and multidrug
443 resistance in the oral microbiota: differences between healthy subjects and patients
444 with periodontitis in Spain. *J. Oral Microbiol.* **13**, (2021).
445 46. Nielsen, H. U. K. *et al.* Tetracycline and macrolide co-resistance in Streptococcus
446 pyogenes: Co-selection as a reason for increase in macrolide-resistant S. pyogenes?
447 *Microb. Drug Resist.* **10**, 231–238 (2004).
448 47. Ardic, N., Ozyurt, M., Sareyyupoglu, B. & Haznedaroglu, T. Investigation of
449 erythromycin and tetracycline resistance genes in methicillin-resistant staphylococci.
450 *Int. J. Antimicrob. Agents* **26**, 213–218 (2005).
451 48. Stepanauskas, R. *et al.* Coselection for microbial resistance to metals and antibiotics
452 in freshwater microcosms. *Environ. Microbiol.* **8**, 1510–1514 (2006).

453 49. Gao, P., Munir, M. & Xagoraraki, I. Correlation of tetracycline and sulfonamide
454 antibiotics with corresponding resistance genes and resistant bacteria in a
455 conventional municipal wastewater treatment plant. *Sci. Total Environ.* **421**–**422**,
456 173–183 (2012).
457 50. Zhang, M. Q., Yuan, L., Li, Z. H., Zhang, H. C. & Sheng, G. P. Tetracycline exposure
458 shifted microbial communities and enriched antibiotic resistance genes in the aerobic
459 granular sludge. *Environ. Int.* **130**, 104902 (2019).
460 51. Tarek, M. H. & Garner, E. A proposed framework for the identification of indicator
461 genes for monitoring antibiotic resistance in wastewater: Insights from metagenomic
462 sequencing. *Sci. Total Environ.* 158698 (2022). doi:10.1016/J.SCITOTENV.2022.158698
463 52. Freitas, A. R. *et al.* High-Resolution Genotyping Unveils Identical Ampicillin-Resistant
464 Enterococcus faecium Strains in Different Sources and Countries: A One Health
465 Approach. *Microorganisms* **10**, (2022).
466 53. Giovanetti, E., Brenciani, A., Lupidi, R., Roberts, M. C. & Varaldo, P. E. Presence of the
467 tet(O) gene in erythromycin- and tetracycline-resistant strains of Streptococcus
468 pyogenes and linkage with either the mef(A) or the erm(A) gene. *Antimicrob. Agents*
469 *Chemother.* **47**, 2844–2849 (2003).
470 54. Trongjit, S. & Chuanchuen, R. Whole genome sequencing and characteristics of
471 Escherichia coli with coexistence of ESBL and mcr genes from pigs. *PLoS One* **16**, 1–17
472 (2021).
473 55. Blanchet, F. G., Cazelles, K. & Gravel, D. Co-occurrence is not evidence of ecological
474 interactions. *Ecol. Lett.* **23**, 1050–1063 (2020).
475 56. Kurtz, Z. D. *et al.* Sparse and Compositionally Robust Inference of Microbial Ecological
476 Networks. *PLOS Comput. Biol.* **11**, e1004226 (2015).
477

Resistance class	All	Air	Dog	Chicken	Cow	Freshwater	Human	Marine	Mouse	Pig	Soil
Aminoglycoside (144)	9.12	17.64	8.42	11.51	5.24	16.70	8.51	0.08	10.12	14.08	1.58
Aminoglycoside/Fluoroquinolone/Macrolide/Phenicol/Rifampicin/Tetracycline (1)	1.08	0.63	0.52	0.12	0.07	0.34	1.08	0.20	1.20	0.17	0.02
Aminoglycoside/Fluoroquinolone/Quinolone (1)	0.19	<0.01	<0.01	0.09	-	0.46	0.15	<0.01	0.01	<0.01	<0.01
Beta_lactam (227)	20.55	41.60	20.07	0.79	1.64	8.44	18.50	1.85	3.02	5.59	1.89
Folate_pathway_antagonist (37)	3.20	2.92	0.97	2.19	0.41	5.45	3.13	0.36	4.33	1.03	0.82
Fosfomycin (28)	1.07	4.36	0.01	0.20	<0.01	1.29	1.49	<0.01	0.22	0.03	0.19
Glycopeptide (27)	0.29	<0.01	-	0.01	<0.01	0.01	0.58	<0.01	0.72	0.02	0.01
Lincosamide (9)	3.33	1.88	29.00	3.42	8.29	0.36	2.02	<0.01	4.01	11.61	0.71
Lincosamide/Macrolide (1)	<0.01	<0.01	-	<0.01	<0.01	<0.01	<0.01	<0.01	-	-	-
Lincosamide/Macrolide/Streptogramin (44)	12.33	6.12	1.35	14.51	2.26	0.32	17.71	0.03	8.19	9.52	5.95
Lincosamide/Macrolide/Streptogramin/Tetracycline (1)	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.07	<0.01	<0.01
Lincosamide/Oxazolidinone/Phenicol/Pleuromutilin/Streptogramin (3)	0.49	0.41	0.01	1.36	0.50	<0.01	0.45	<0.01	1.64	1.25	0.04
Lincosamide/Pleuromutilin/Streptogramin (7)	0.27	0.48	0.18	0.09	0.02	0.02	0.34	<0.01	<0.01	0.21	0.05
Lincosamide/Streptogramin (2)	0.28	0.12	<0.01	0.02	<0.01	0.01	0.38	<0.01	0.06	<0.01	<0.01
Macrolide (24)	3.04	1.86	5.78	0.36	16.71	1.90	2.53	<0.01	4.22	6.83	0.25

Macrolide/Streptogramin (4)	1.10	1.69	1.01	0.12	0.39	1.18	1.43	<0.01	1.83	0.48	0.11
Macrolide/Tetracycline (1)	<0.01	<0.01	-	<0.01	<0.01	<0.01	<0.01	<0.01	-	<0.01	<0.01
Nitroimidazole (9)	0.32	<0.01	0.04	0.01	0.05	<0.01	0.56	<0.01	0.02	0.14	<0.01
Oxazolidinone/Phenicol (1)	0.02	0.03	-	0.10	<0.01	<0.01	<0.01	<0.01	-	0.09	<0.01
Oxazolidinone/Phenicol/ Tetracycline (1)	0.07	0.02	<0.01	1.54	<0.01	<0.01	<0.01	<0.01	0.03	0.04	<0.01
Phenicol (33)	9.21	6.19	0.36	6.10	4.33	55.89	3.33	97.37	0.14	0.81	85.04
Pleuromutilin (1)	0.06	0.02	-	<0.01	<0.01	0.39	0.06	<0.01	0.01	0.07	<0.01
Polymyxin (10)	1.19	0.30	<0.01	0.14	<0.01	1.85	1.26	<0.01	0.18	0.14	0.12
Quinolone (17)	0.11	<0.01	-	0.01	-	0.18	0.06	<0.01	-	<0.01	<0.01
Rifampicin (6)	0.06	0.09	-	<0.01	<0.01	<0.01	0.03	<0.01	<0.01	<0.01	<0.01
Steroid_antibacterial (2)	0.15	<0.01	-	3.30	<0.01	<0.01	<0.01	<0.01	<0.01	0.11	0.02
Streptogramin (11)	32.48	13.61	32.27	53.99	60.05	5.20	36.39	0.06	59.97	47.76	3.18
Tetracycline (64)	<0.01	-	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

478 *Table 2: Relative abundance of read fragments aligned to each resistance class (row) for each sampling source (column). The relative abundance is the number of read fragments aligned to the group out of all*
479 *fragments aligned to ARGs. For each resistance class, the parenthesis shows the number of ARGs belonging to that category.*



Network	No. nodes	No. edges	Global clustering coefficient	Network density	Edge density	No. components
All	225	2307	0.602	0.092	0.046	25
Air	7	4	0.000	0.190	0.095	3
Dog	11	11	0.000	0.200	0.100	3
Chicken	91	171	0.479	0.042	0.021	13
Cow	61	126	0.287	0.069	0.034	4
Freshwater	16	21	0.550	0.175	0.088	5
Human	134	754	0.701	0.085	0.042	18
Marine	5	4	1.000	0.400	0.200	2
Mouse	34	91	0.630	0.162	0.081	3
Pig	96	494	0.586	0.108	0.054	6
Soil	19	19	0.458	0.111	0.056	3

*Figure 1: The resistome networks consisting of correlations (edges) between pairs of ARGs (nodes). **a**. Each correlation network is visualized, where each ARG node is colored by resistance class, and edges are colored by the correlation coefficient value. **b.** Metrics of the correlation networks reveal how interconnected and complex the networks are. No. stands for number of. **c**. Distribution of included correlation coefficients in each network. Figure S3 shows the distribution of all inferred correlations and their p-values. **d**. A heatmap showing how similar the content of the two networks is. In the upper half of the heatmap, the number of overlapping ARG nodes is shown. An overlap means that a specific ARG has a correlation in both networks, ignoring what it is co-abundant with. In the lower half of the heatmap, the number of overlapping correlation edges is shown. An overlapping edge is defined as whether the correlation coefficient is reported in both networks, regardless of the value of the coefficient.*

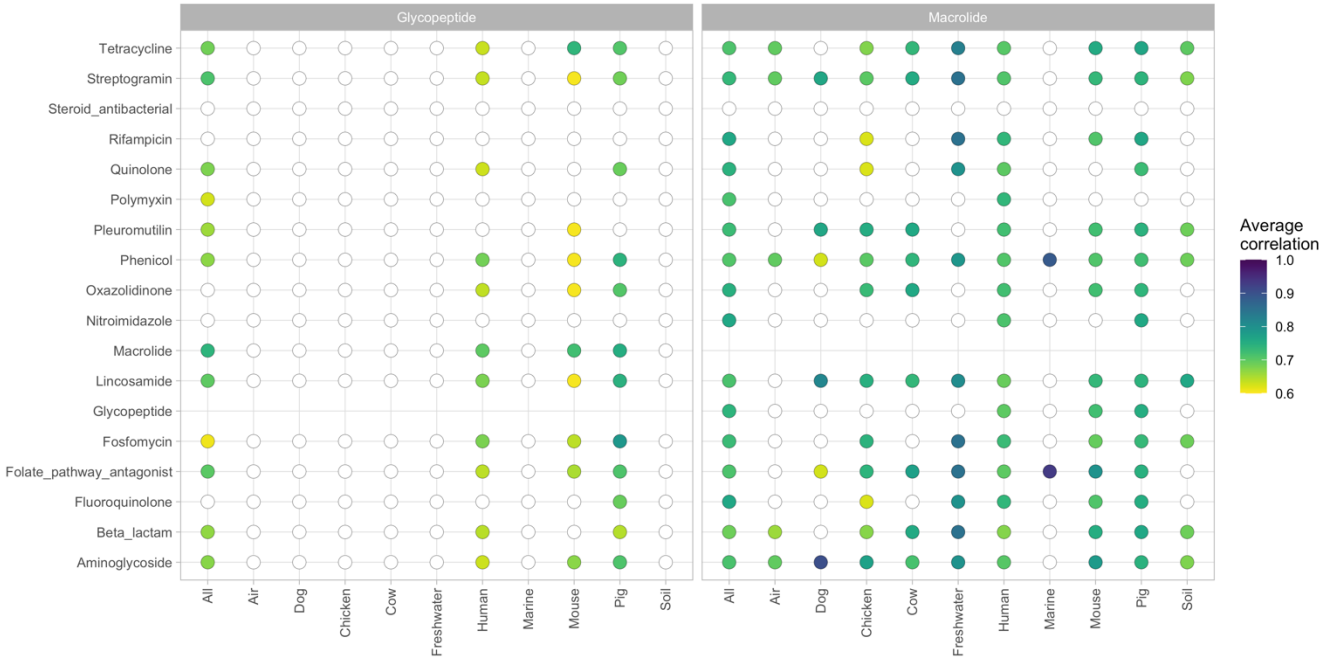


Figure 2: ***Correlation profiles for ARGs conferring resistance to the critically important antimicrobial classes glycopeptides (left) and macrolides (right).*** *Each column shows the average correlation from, e.g., macrolide ARGs to ARGs for other antimicrobial classes. The circle is colored by the average correlation, where a white circle indicates no statistically significant correlations of ARGs observed between the two antimicrobial classes.*

Figure 3: ***Correlation profiles for ARGs conferring resistance to important pleuromutilins (left) and highly important tetracyclines (right).*** *Each column shows the average correlation from, e.g., tetracycline ARGs to ARGs for other antimicrobial classes. The circle is colored by the average correlation, where a white circle indicates no statistically significant correlations of ARGs observed between the two antimicrobial classes.*

Supplementary materials

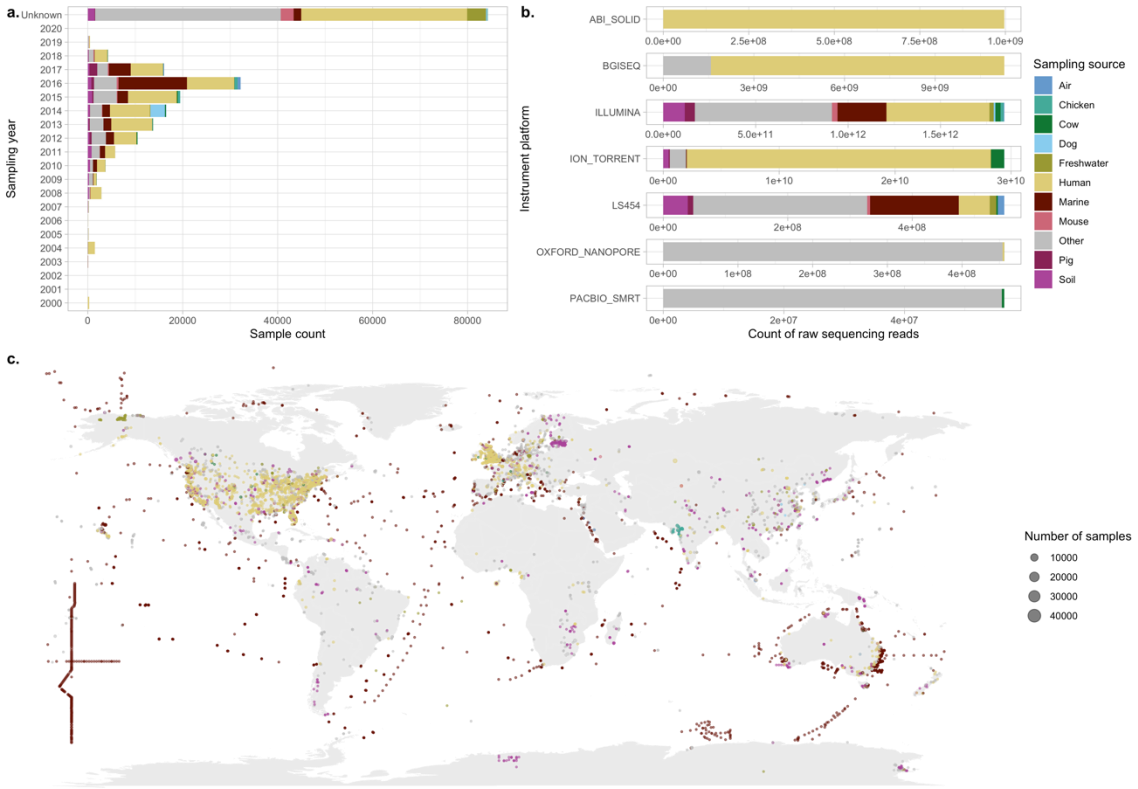


Figure S1: ***Metagenomic origins colored by the sampling group****. a. Overview of sampling year for metagenomic samples colored by their sampling origin. 100 samples were taken before 2000, and 84,238 did not have a valid sampling date. b. Overview of the amount of sequencing reads available for each sampling origin. c. Sampling locations of metagenomic samples used in the correlation analysis were split by their sampling source. Number of samples with no coordinates available; All: 83,361; Air: 16; Dog: 3,159; Chicken: 570; Cow: 262; Freshwater: 61; Human: 40,003; Marine: 363; Mouse: 2,842; Pig: 320; Soil: 477. The 'Other' label refers to those that are not in one of the source-specific networks but are included in the 'All' network.*

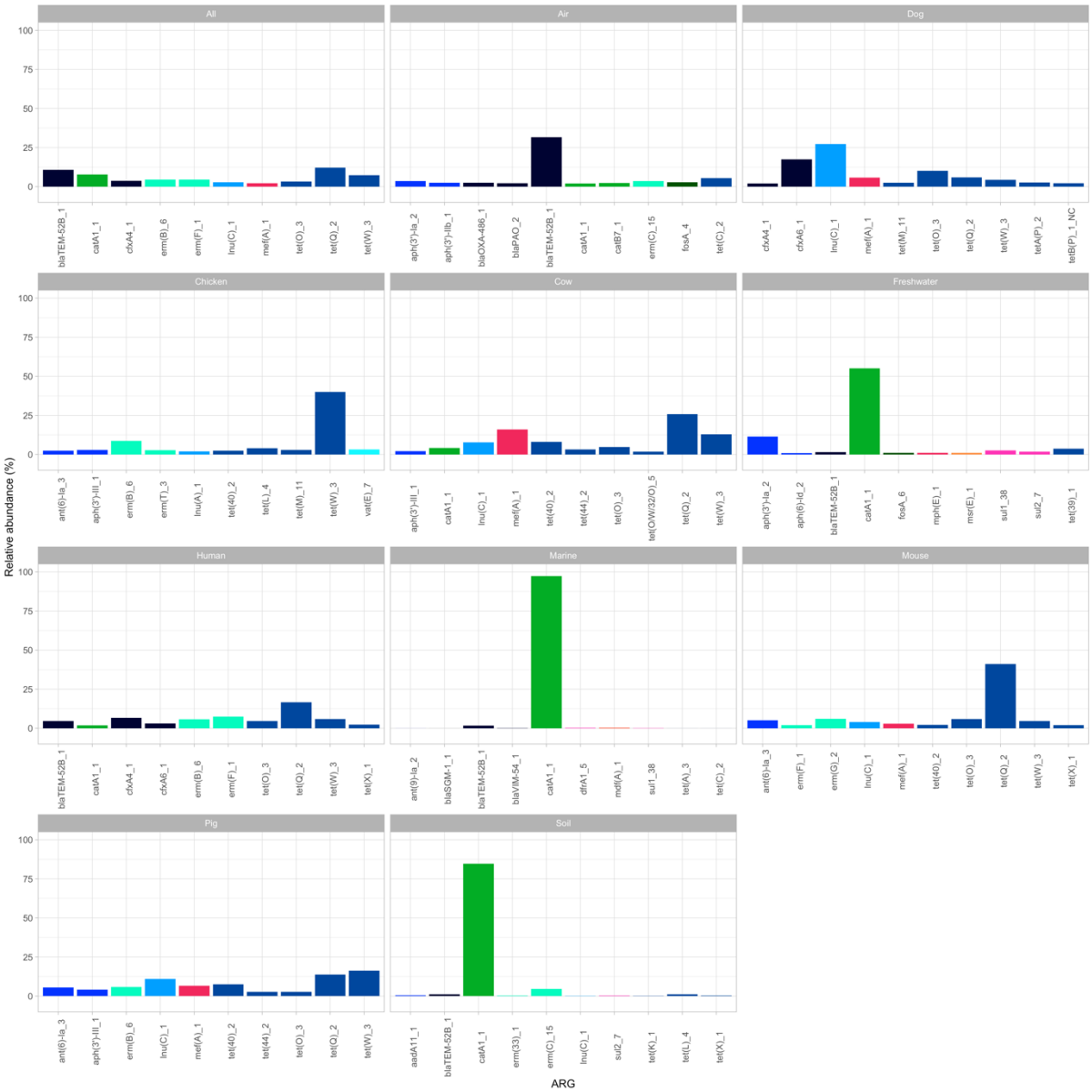


Figure S2: ***Top 10 most abundant ARGs per sampling group.*** *A barplot showing the ARGs having the highest 10 relative abundances per sampling source colored by resistance class.*

	No filters			minF	minS		minF	minS		minF	minS		minF	minS	
				25	10		50	10		50	20		50	50	
Source	**ARG cols**	**#nodes**	**#edges**	**ARG cols**	**#nodes**	**#edges**	**ARG cols**	**#nodes**	**#edges**	**ARG cols**	**#nodes**	**#edges**	**ARG cols**	**#nodes**	**#edges**
All	716	443	34,815	459	275	4,786	425	257	2,970	313	165	320	313	165	320
Air	436	565	66,612	34	14	10	19	9	6	0	-	-	0	-	-
Chicken	391	634	92,563	145	123	289	123	191	226	69	56	87	69	56	87
Cow	445	614	92,881	94	75	242	85	68	198	37	25	57	37	25	57
Dog	158	72	85	45	25	35	45	25	35	45	25	35	45	25	35
Freshwater	579	589	43,054	56	40	74	36	24	33	1	-	-	1	-	-
Human	692	474	53,408	307	179	1,445	278	158	902	201	91	120	201	91	120
Marine	668	13	8	24	7	5	10	7	5	4	-	-	4	-	-
Mouse	326	549	80,254	55	45	122	41	35	107	25	21	70	25	21	70
Pig	408	634	87,754	136	114	864	118	99	606	61	52	175	61	52	175
Soil	614	510	22,415	72	53	85	47	28	37	3	-	-	3	-	-

*Table S1: **Filtering the counts impacts the number of input ARG columns, as well as the resulting number of correlations.** minF is the minimum number of fragments available for an ARG in one sample, and minS is the minimum number of samples to fulfill the minF requirement for the column to be included in the inference procedure*

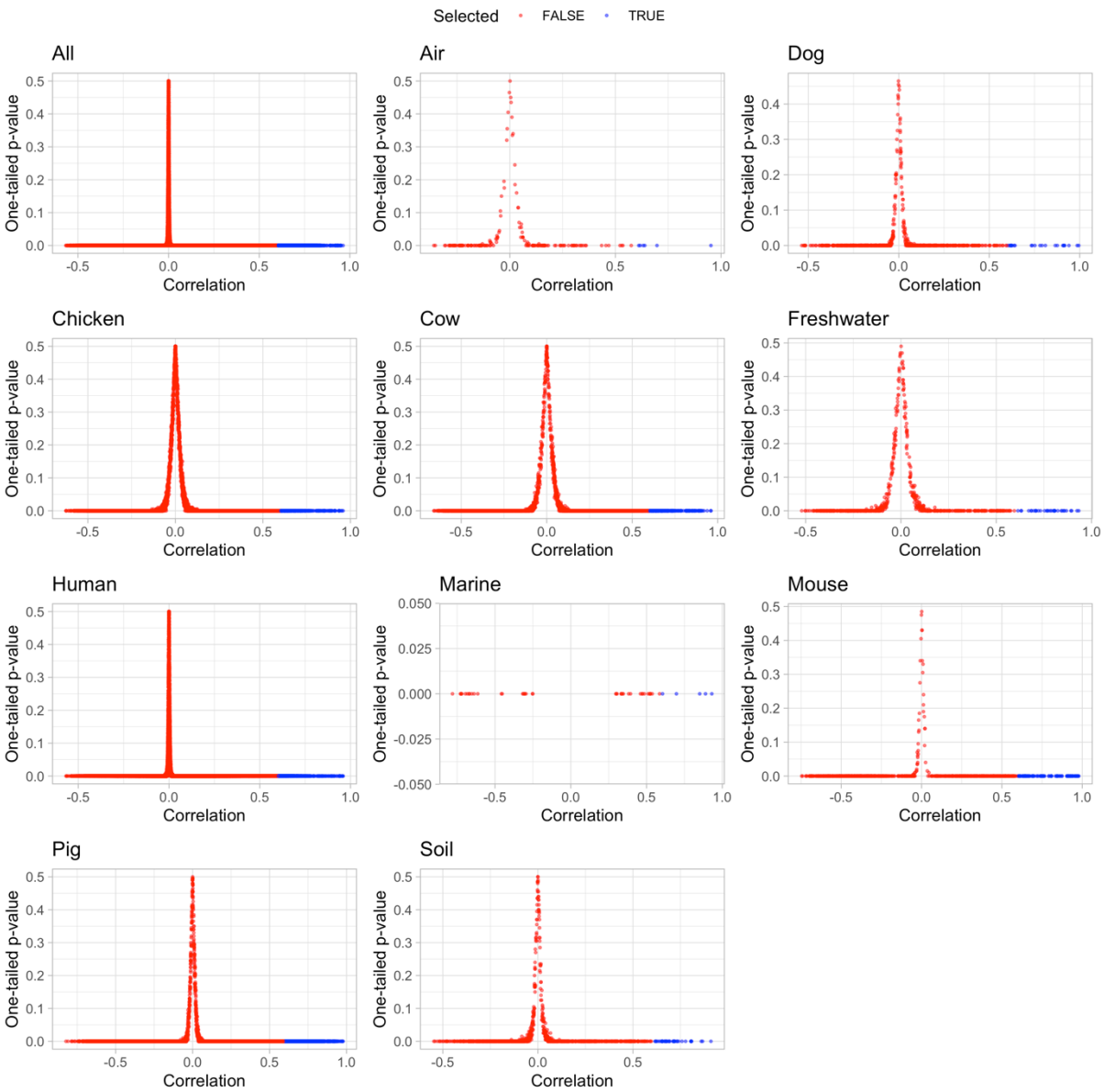


Figure S3: ***Distribution of correlations and p-values produced by SparCC for each data grouping****. Each point is a correlation between two ARGs and is colored by whether the point was selected if the p-value < 0.01 and correlation ≥ 0.6.*

Figure S4: ***The relative abundance, number of ARGs, and the number of correlations for each resistance class in each sampling group.*** *The relative abundance in green shows the percentage of read fragments for each class. In grey is the number of ARGs for each resistance class. Orange coloring indicates the number of correlation coefficients inferred for ARGs in a resistance class.*

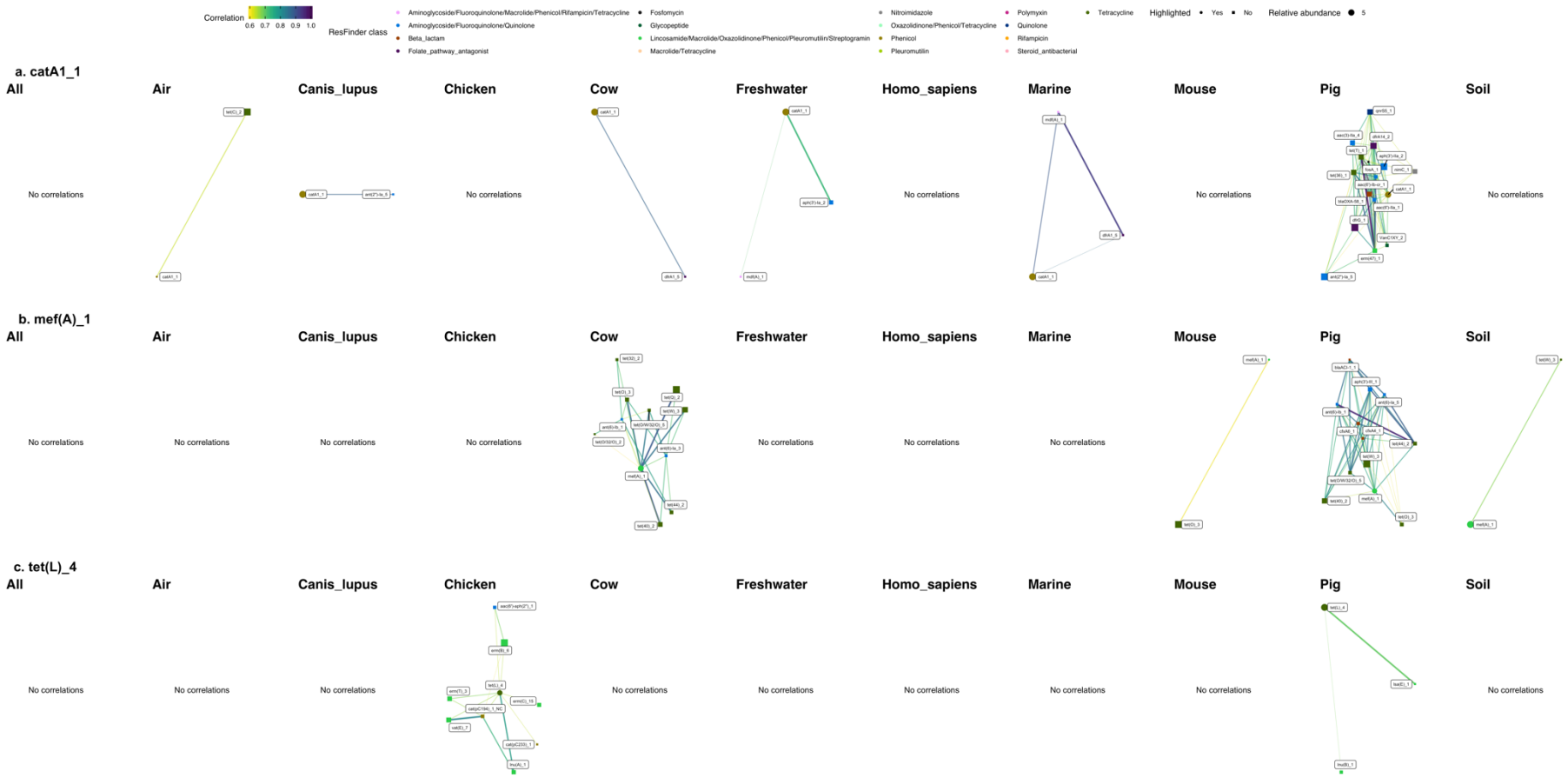


Figure S5: ***Networks illustrating how one ARG interacts differently depending on the sampling environments****. The three genes catA1_1, mef(A)_1, and tet(L)_4 are selected to illustrate that the genes correlate more or less with other genes in their abundance in some habitats.*

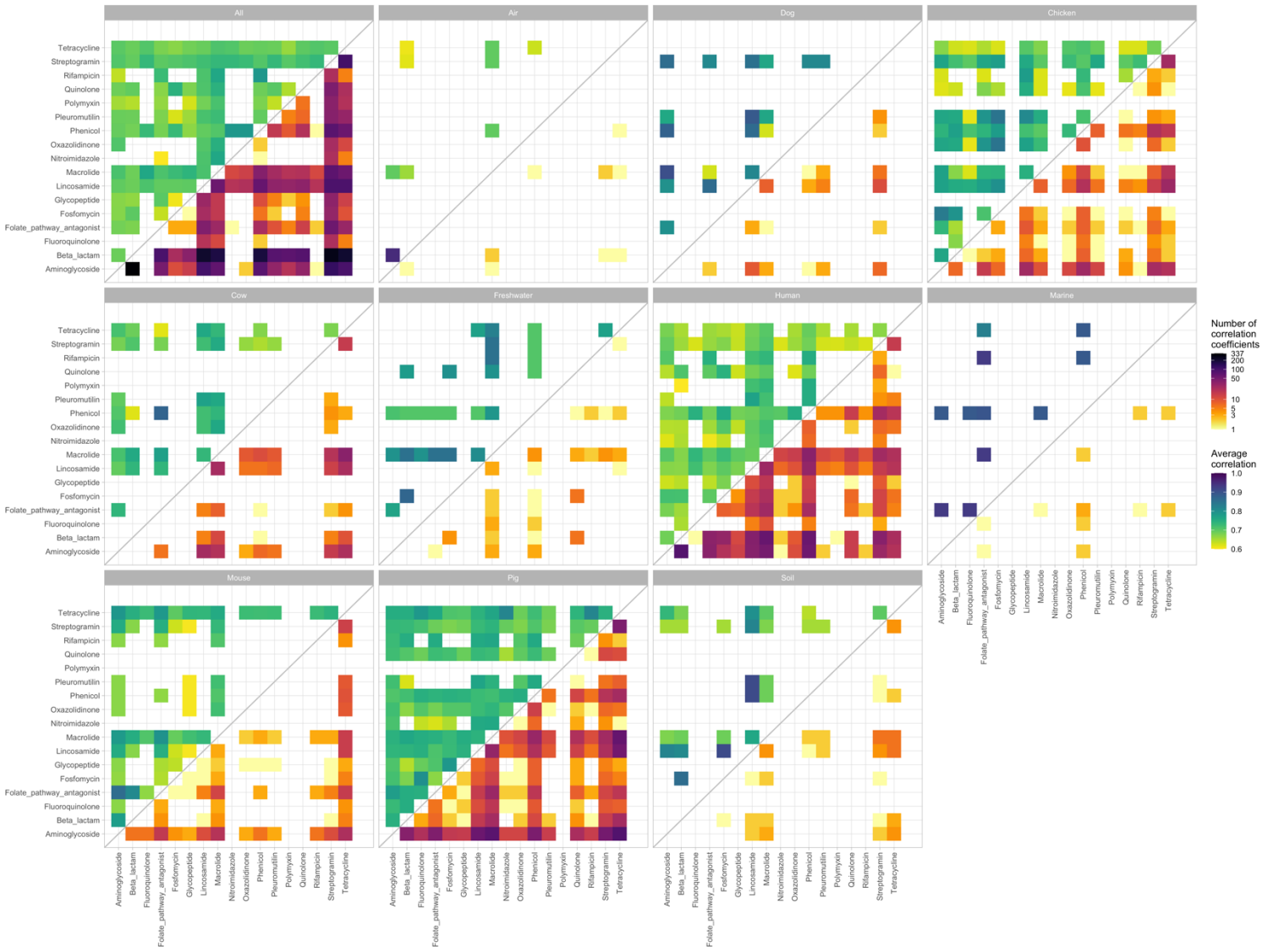
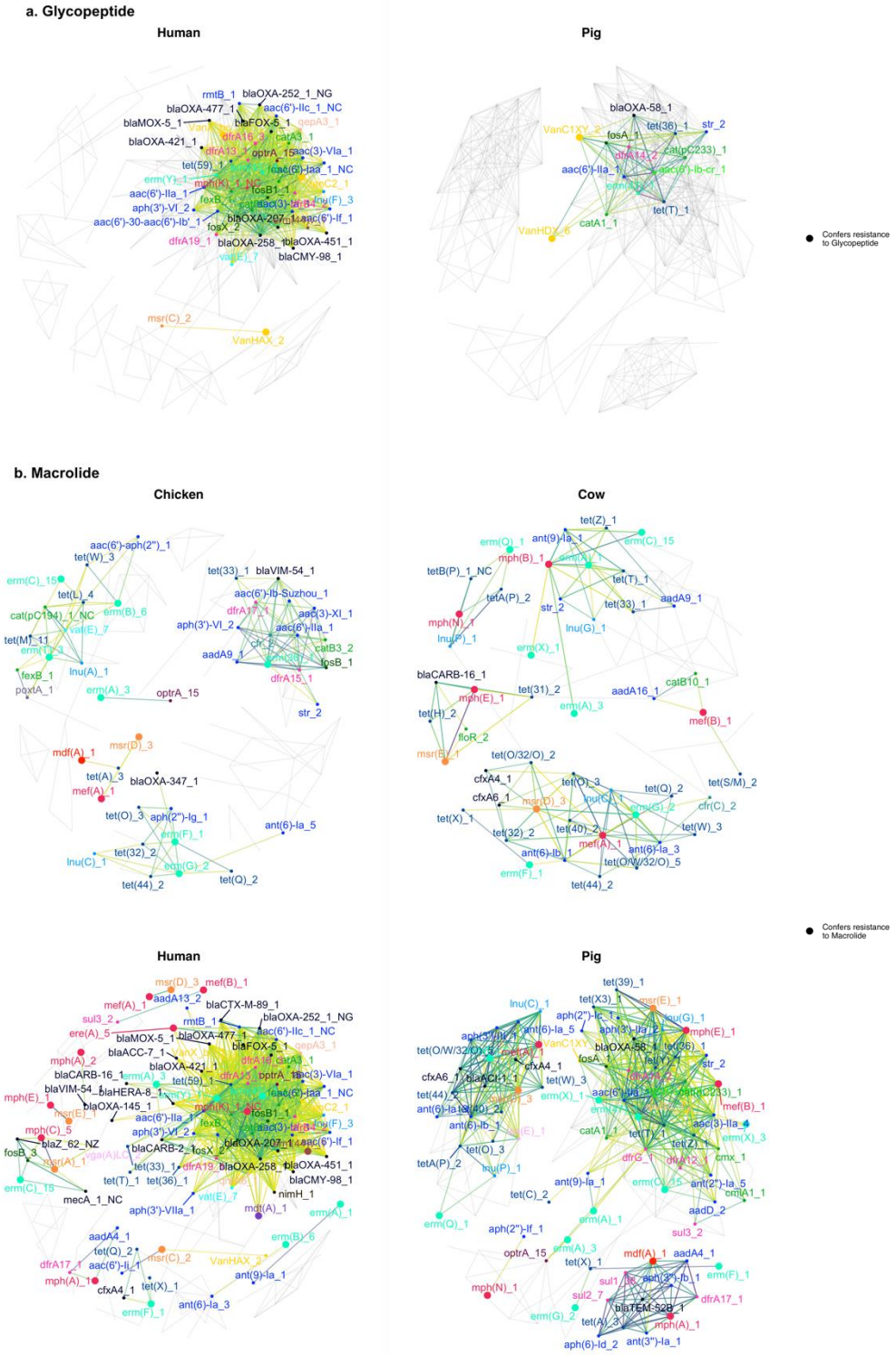


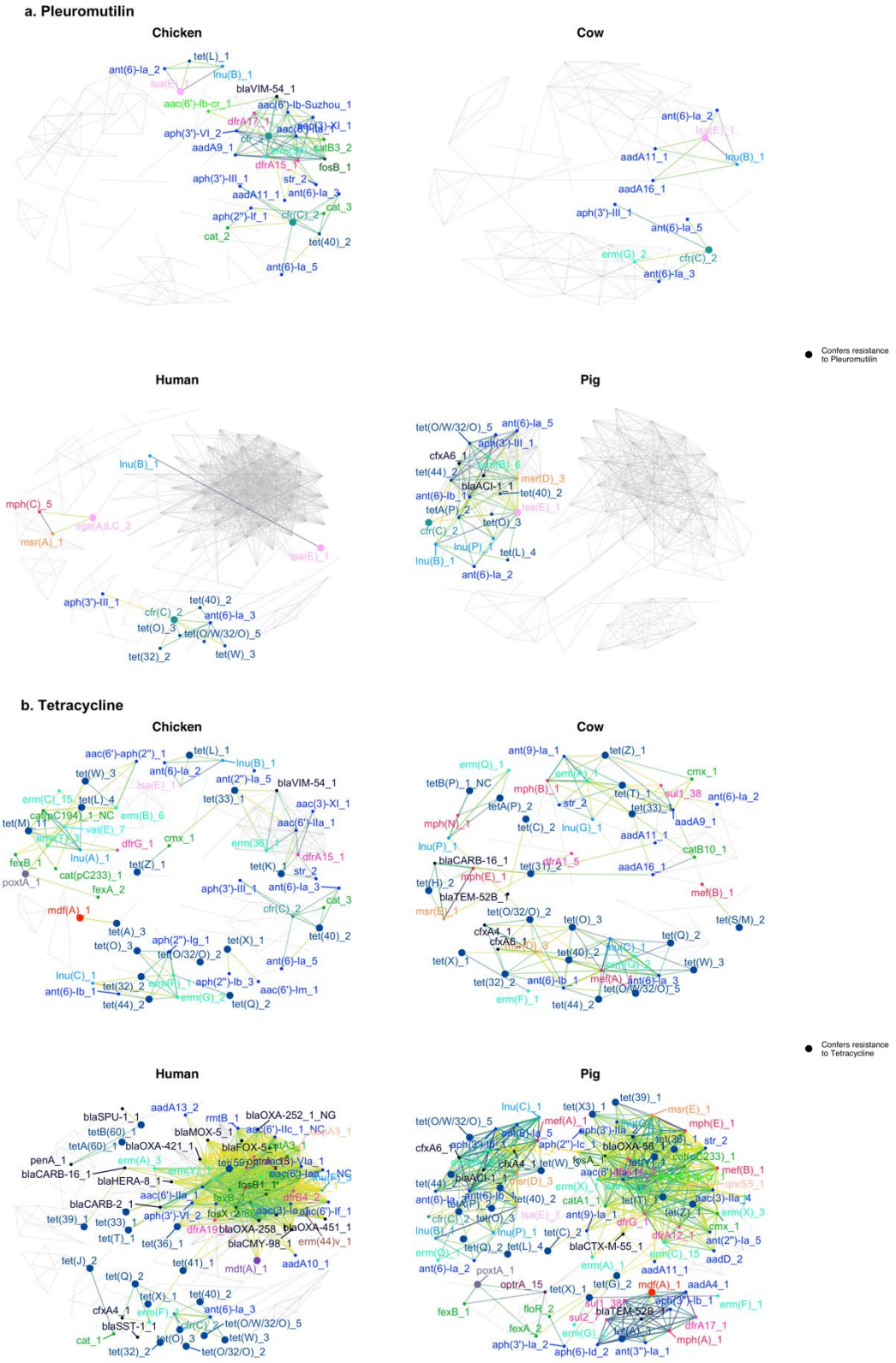
Figure S6: The average correlation between ARGs of different resistance classes in the upper triangle of the heatmaps with the lower half shows the correlation coefficients between the two classes. Note that a correlation coefficient between two ARGs might be present in more than one tile, as some ARGs confer resistance to multiple classes of antimicrobials.



528

529 *Figure S7*: ***Highlights of interactions between ARGs of a. glycopeptide resistance and b. macrolide resistance***. *The ARG node size shows*
530 *whether the corresponding ARG gives resistance to the antimicrobial class in focus. Only correlation edges and ARGs nodes are colored if they*
531 *correlate with the highlighted ARGs; otherwise, they are colored grey. The coloring schemes for nodes and edges are given in Figure 1a.*

532



533

534 *Figure S8:* ***Highlights of interactions between ARGs of a. pleuromutilin resistance and b. tetracycline resistance.*** *The ARG node size shows*
535 *whether the corresponding ARG gives resistance to the antimicrobial class in focus. Only correlation edges and ARGs nodes are colored if they*
536 *correlate with the highlighted ARGs; otherwise, they are colored grey. The coloring schemes for nodes and edges are given in Figure 1a.*

537