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 acquired 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 argued that the correlations could serve as risk profiles of resistance co-occurring to critically important antimicrobials. Using these profiles, we found evidence of

28 several ARGs conferring resistance for critically important antimicrobials (CIA) being co-
29 abundant, such as tetracycline ARGs correlating with most other forms of resistance. In
30 conclusion, this study highlights the important ARG players indirectly involved in shaping the
31 resistomes of various environments that can serve as monitoring targets in AMR
32 surveillance programs.

33 Introduction

34 Antimicrobial resistance (AMR) has become a global health threat, with ramifications not
35 only for modern medicine but also for agriculture and environmental health[1–3]. It is widely
36 acknowledged that the misuse of antimicrobials has accelerated the dissemination and
37 prevalence of antimicrobial resistance genes (ARGs)[4]. Most attempts to reduce the burden
38 of AMR have focused on reducing the use of single classes of antimicrobial agents
39 considered of critical importance[5]. Despite considerable efforts in various settings, such as
40 livestock, these regulations have not significantly reduced the spread of ARGs[6,7]. It has been
41 shown that after banning the use of specific antimicrobials, resistance to that antibiotic will
42 still be prevalent in the environment [8]. Indirect co-occurrence of ARGs has also been shown
43 to happen under selective antibiotic pressure[7,9–11]. We have also recently observed that
44 changes in the selective pressure of even a single antimicrobial agent influence several ARGs
45 in pig metagenomes[11]. However, there is also evidence that adding mobile ARGs to a
46 microbiome increased its stability, whereas if the ARGs were encoded in the chromosome,
47 the stability decreased[12]. The origin of the interplay between the microbiome and
48 resistome, however, remains largely unresolved, as there are many ecological, functional,
49 and evolutionary properties unaccounted for. It is unclear whether the co-selection of ARGs
50 happens due to genetic linkage or inter-species interactions.
51
52 Studies focused on studying the link between antimicrobial usages (AMUs) and the
53 prevalence of AMR have typically only focused on the effects of using one antimicrobial
54 agent and the impact on developing resistance to that agent, ignoring the overall effect on
55 the abundance of ARGs in the environment. This effect on the resistome must be
56 understood better to prevent these collateral damage effects from happening[13,14]. Co-
57 occurrence of microbes has been evaluated in soil[15,16] and marine environments[17,18],
58 whereas ARG co-occurrences have been studied in sewage sludge[19], freshwater[20], marine[21],

59 swine[8], and cattle[22]. However, most of these studies have only been evaluated on a smaller
60 scale and not always using the same methods. There is currently a large collection of next-
61 generation sequencing datasets from metagenomic samples available in public repositories,
62 providing an excellent resource to quantify the prevalence of ARGs by analyzing the read
63 abundances [23–26].
64
65 In this study, we have analyzed the co-abundance of sequencing reads aligned to acquired
66 ARGs to assess how resistance to one specific antimicrobial agent is linked to the abundance
67 of another class of antimicrobial. With a collection of 214,095 metagenomic datasets from
68 public repositories[27], we examined the correlation between pairwise ARG read abundances
69 with a compositional approach[28,29] using SparCC (Sparse Correlations for Compositional
70 data)[30]. Our results demonstrated that many ARG pairs interact but are highly specific to the
71 environment. We believe that these interactions provide another way to study how ARGs
72 are being co-selected independently of the microbial context, and the findings can be used
73 to design targeted interventions to limit the spread of AMR.
74

75 Methods

76 Data collection and pre-processing

77 We have previously described in detail the process of downloading and analyzing 214,095
78 metagenomic samples[25,27], but in brief: We downloaded raw sequencing reads corresponding
79 to 442 Tbp from metagenomic samples deposited in the European Nucleotide Archive[31] that
80 were uploaded between 2010-01-01 and 2020-01-01 and had at least 100,000 reads and were
81 shotgun sequenced. The raw sequencing reads were quality-checked with FASTQC v.0.11.15
82 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed with BBduk2
83 36.49[32]. The trimmed reads were then globally aligned using KMA[33] 1.2.21 against two
84 reference sequence databases: ResFinder[34] (downloaded 25-01-2020) and Silva[35] (version 38,
85 downloaded 16-01-2020). ResFinder is a database of 3,085 acquired ARGs, whereas Silva is a
86 16S/18S rRNA database of 2,225,272 sequences.
87

Homology-based clustering of ResFinder sequences

With the possibility of detecting up to $C(3085,2) = \frac{3085!}{(2!(3085-2)!)} = 4{,}757{,}070$ pairwise interactions between the 3,085 ARGs of ResFinder, we decided to cluster similar reference sequences together to reduce the number of possible combinations. Using USEARCH[36] 11.0.667 and 90% nucleotide identity, we did a homology-based clustering that grouped the ARGs into a total of 716 groups. For each of these groups, the read counts for the genes in that group were first adjusted by the 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[30] 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[37] with packages igraph[38], qgraph[39], and ggraph[40] using the Fruchterman-Reingold layout algorithm[41].

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[42]. The network density is calculated as $2E/N(N-1)$ as given in Parente et al. (2018)[43]. Edge density is the number of edges over the number of possible edges[44]. 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[45].

Data and code availability

The matrix of raw ARG read counts is available at https://doi.org/10.5281/zenodo.6519843 . 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.

146 # Results

147 This study investigated the correlation of pairwise ARG abundances across a set of 214,095
148 metagenomic sequencing datasets. This collection represents a highly diverse set of publicly
149 available metagenomic samples collected between 2000 and 2020 from locations all over
150 the world, where the sequencing datasets consist of short-read reads produced on Illumina
151 platforms (Figure S1). We observed that not all samples contained sequence read
152 fragments aligned to ARGs, so we only included those that did in the correlation analyses (n
153 = 119,206; Table 1). Looking at the various sampling sources, read fragments aligned to
154 ARGs conferring resistance to tetracycline and beta-lactams were very common in host-
155 associated sources, whereas phenicol resistance was more frequently observed in
156 environmental sources (freshwater, marine, and soil) (Table 2). *catA1* was the most
157 dominant gene in the environmental samples, especially in marine samples, whereas
158 various *tet* genes had high relative abundances in livestock or human samples: *tet(W)* in
159 chicken and pig samples; *tet(Q)* in cow, human, and pig metagenomes. *blaTEM-52B*
160 accounted for more than 30% of the read fragments assigned to ARGs in air metagenomes
161 (Figure S2).
162

163 ## Balancing the sparsity and network complexity

164 Initially, we inferred correlations with SparCC on all samples that had at least one read
165 fragment aligned to an ARG. We observed that most ARGs were found to correlate with
166 each other in all environments, even in cases where there was only one read fragment per
167 reference (Table S1). Across all samples, we observed template hits ranging from 1 to 252
168 with a median of 20 and fragment counts spanning 1 to $4.5 \cdot 10^7$, with a median of 880
169 fragment counts. To balance the sparsity and network complexity, we introduced a filtering
170 step. This step determined whether an ARG should be included or excluded based on a
171 minimum number of fragment counts and a minimum number of samples supporting hits to
172 that ARG template. Among the 716 ARGs considered, we noticed that without any filters,
173 SparCC indicated correlations even with very low read fragment counts. Therefore, we
174 required that for an ARG to be included, it had to have a minimum fragment count of 50
175 across at least 10 samples (Table S1).
176

177 ## Analysis of large-scale metagenomic correlation networks

178 Based on our filter settings, we constructed a global network using the correlation
179 coefficients for the entire collection of metagenomes, with each node representing an ARG
180 and each edge representing a pairwise ARG connection (correlation ≥ 0.6, p-value < 0.01,
181 Figure S3). The global network, nicknamed 'all', contained 225 ARGs connected through
182 2,344 correlation edges (Figure 1a). As this all-network was hard to interpret due to the
183 many highly interconnected ARGs, we also inferred pairwise ARG correlations in specific
184 sampling groups (Table 1). The genes that were part of these networks were found to
185 correlate with varying degrees of strength (Figure 1a-b). For example, the human network
186 contained many correlation coefficients, but most were less than 0.8 (Figure 1c). Another
187 example is the marine network, where only a few ARGs were found to correlate, but with
188 values above 0.9 (Figure 1b-c). Despite the networks reflecting the composition of the
189 various environments, we still observed overlaps between which ARGs were found to
190 correlate. One example is that all the correlations inferred from the pig metagenomes also
191 existed in the human metagenomic network (Figure 1d).
192
193 There was a limited number of correlations for ARGs encoding resistance to
194 fluoroquinolones, steroid antibiotics (fusidic acid), colistin, and rifampicin. On the other
195 hand, beta-lactam, tetracycline, and aminoglycoside ARGs had many correlations with each
196 other and with other classes (Figure S4, Figure S6). Despite resistance to some antimicrobial
197 classes being the most abundant, the ARGs did not always correlate to many others. For
198 example, tetracycline ARGs were the most abundant in dog metagenomes, but no
199 correlations were inferred for these ARGs. Similarly, in the human samples where
200 aminoglycoside and beta-lactam ARGs were less abundant than tetracycline ARGs,
201 aminoglycoside and beta-lactam resistance genes had a higher number of correlation
202 coefficients were reported (Table 2, Figure S4).
203
204 On the level of ARG abundance, we observed that just because an ARG was highly abundant
205 in a sampling group, it did not automatically mean that it correlated to many other ARGs.
206 The highly abundant *catA1* gene in marine (97.4% of all ARG reads), freshwater (55.1%), and
207 soil samples (84.7%) (Figure S2) did only correlate with one or two other genes in the water
208 environmental networks, and none in the soil network. On the other hand, *catA1* did seem

209 to be correlated with 15 other genes in the pig correlation network despite not being highly
210 abundant in that group of samples (Figure S5a). *mef(A)*_1 accounted for 15.9% of the reads
211 aligned to ARGs in cow samples and 6.63% in pigs (Figure S2) and was also strongly
212 correlating with other genes, which mainly conferred resistance to aminoglycosides,
213 (fluoro)quinolones and tetracyclines (Figure S5b). *tet(L)_4* only accounted for 4.01% of the
214 read fragments aligned to ARGs in metagenomes of chicken origins (Figure S2) but was
215 shown to correlate in its abundance with the abundance of 8 other ARGs, e.g., with a
216 correlation of 0.77 with *lnu(A)_1 (*Figure S5c*).*
217

218 The hidden signals between ARGs profile the potential risk of co-selection in different 219 environmental contexts

220 Using the correlations between ARGs in the various environments (Figure 1), we calculated
221 the average correlation between ARGs of different antimicrobial classes (Figure S6). These
222 average correlations can then serve as profiles to assess the risk of indirectly selecting ARGs
223 that confer resistance to different ARGs through co- and cross-resistance. These risk profiles
224 can then be used to judge the strength of interactions upon using one antimicrobial in each
225 setting. Upon constructing these profiles, we observed that the strength and the number of
226 correlations highly depend on the antimicrobial classes and the environmental context
227 (Figure S6). We hypothesize that if an important (IA) or highly important antimicrobial (HIA)
228 is used, first, the resistance to the antimicrobial class will likely flourish, and, secondly,
229 through co- and cross-resistance, so will ARGs conferring resistance to other classes,
230 including those that are CIA.
231

232 Figure 2 shows two risk profiles for ARG correlations for glycopeptide and macrolide, two
233 highly critically important antimicrobials. Correlations between ARGs of glycopeptide
234 resistance to other resistance classes were much rarer than those connected with macrolide
235 ARGs (Figure S6) and those correlations that were observed were relatively low
236 (correlation<0.8, Figure 2). Different vancomycin resistance cassettes were responsible in
237 different environments, namely *VanHAX*, *VanC2*, and *VanX_bc* in human samples and
238 *VanHDX* and *VanC1XY* in pig samples (Figure S7a). *VanHAX* only has one correlation,
239 whereas the remaining five correlate with many different ARGs. On the contrary, ARGs

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

256 Discussion

257 Considering the complexity of microbiomes, studying how microbial composition shapes the
258 distribution of ARGs is a challenging task but one that could shed light on how ARGs
259 indirectly select one another. However, with the high-throughput sequencing technologies
260 and many metagenomic datasets available in public repositories, it is now much more
261 feasible to extract the patterns of how ARGs co-occur without knowing their microbial
262 origin. Using our recently published collection of 214K metagenomic datasets[27], we have
263 inferred correlations of pairwise ARG abundances to profile which types of resistance
264 influence the shape of resistome and which genes are the key players. To the best of our
265 knowledge, this is the first study to relate ARG abundances on such a large and broad scale.
266
267 Our correlation networks revealed that not all ARGs are connected, as the pairwise ARG
268 interactions were largely shaped by the composition of the environmental resistome. We
269 observed cases where a highly abundant ARG had no correlations and the opposite where a
270 sparse ARG had many correlations, highlighted by our effort to apply appropriate filters to

271 ensure robust correlations (Table S1). However, by filtering on the individual abundance of
272 ARGs in each resistome, there is a risk of removing important links between ARGs, especially
273 in environments with very low abundant ARGs. We continued with settings that ensured
274 that sparse ARGs in complex environments were removed but that ARGs in overall sparse
275 environments could still be included, such as for the air resistome. More sensitive filtering
276 could be applied to further improve the robustness of the inferred ARGs for each resistome.
277
278 The network showing correlations for all metagenomic samples was the most complex,
279 which we speculate is due to both the wide variety of sampling sources and the
280 overrepresentation of human metagenomes (Table 1). We found the differences in the
281 animal and environmental networks much more interesting, as they reflect the dynamics of
282 ARG abundances in more localized settings (Figure 1). On the level of which antimicrobial
283 class an ARG confers resistance to, we could also observe that some resistance classes had
284 more and stronger correlations than others (Figure S6). We found several cases of strongly
285 correlated ARG pairs that have also been reported in other studies. These cases were often
286 due to the ARGs being present in the same microbial genome. For example, the gene
287 cassette *vanHAX* has been found together with *msr(C)* and *aac(6')* in genomes of human
288 isolates[46] and *mef(A)* linked with *tet(O)*[47] and *mdf(A)* with *blaTEM, aph(6), sul2, and, tet(A)*[48]
289 (Figure S7).
290
291 As highlighted in Figure 2 and Figure 3, we argue that the correlations can serve as a way to
292 profile the collateral effects occurring in a resistome when the abundance of different ARG
293 mechanisms changes. Antimicrobials have been classified differently to reflect their
294 importance to human medicine, of which glycopeptides and macrolides are CIA with the
295 highest priorities. Glycopeptide and macrolide resistance have previously been linked
296 genetically[7,9] in pigs, where we also can report the presence of correlations between ARGs
297 of glycopeptide and macrolide resistances not only in pigs but also in human and mouse
298 environments (Figure 2). Pleuromutilin and tetracycline antimicrobials are classified as being
299 less critically important[49], but SparCC reported many correlations of ARGs for these two
300 classes in various networks (Figure 3), which suggests that there are still risks associated
301 with the use of these two. Tetracycline resistance has been found to occur together with
302 resistance to macrolides[50–52], aminoglycosides[50,53], folate pathway antagonists[54,55],

303 lincosamide[11], and beta-lactams[53], to name a few studies. This high connectivity of
304 tetracycline ARGs seems in line with our results, as this specific group of ARGs was
305 connected to almost all classes of antimicrobials in most of our networks (Figure 3, Figure
306 S8b). The variety of connections to antimicrobials of less importance to human health
307 should be more in focus, as our results show that there are risks of critical antimicrobial
308 resistances emerging from the enrichment of less essential resistance genes.
309
310 Following this line of thought of focusing on ARGs that give resistance to less critically
311 important antimicrobials, a recent study by Tarek and Garner (2022)[56] proposed to create a
312 monitoring framework based on isolated components, or clusters, in correlation networks.
313 They constructed a correlation network encompassing ARG abundances in samples from
314 wastewater treatment plants and argued that representative ARG members from each
315 cluster in their network should be monitoring targets. Our networks suggest that limiting to
316 only a few ARGs would fail to capture the complete picture in some environments, such as
317 human microbiomes (Figure 1). Instead, we propose to limit the set of monitoring targets to
318 only what happens to ARG abundances during exposure to one type of antimicrobials
319 (Figures Figure 2,Figure 3, Figure S7Figure S8). If a monitoring system is implemented, it
320 would need to be updated regularly to show the changes in antimicrobial usage and ARG co-
321 occurrences since the correlations we have inferred in this study only reflect the current and
322 past abundances.
323
324 In order to use correlation networks for surveillance of AMR, more work is needed to
325 confirm that the observed interactions do indeed exist in nature[57]. We have defined
326 interactions as being indirect since SparCC does not indicate which way the interaction
327 occurred. One way to investigate this could be to investigate physical linkages or by using
328 other analytical methods. For example, to determine whether *vanHAX* influences *msr(C)* or
329 the other way around (Figure S7a), the SPIEC-EASI[58] method could be used to infer such
330 directional dependences. A directional correlation could be included in a risk profile.
331
332 Our choice of only studying the co-abundances of ARGs does not capture the causality and
333 directionality of the interactions. We have not disentangled whether the interactions are
334 due to co-selection or cross-selection, and neither investigating genetic linkages, co-

exposure, and the presence of other genetic elements such as mobile genetic elements. Understanding some of these factors could be achieved by expanding our analysis with bacterial read counts and counts of mobile genetic elements. We have speculated that the resulting correlations could reflect the community's response to external stressors, such as different usages of antimicrobials. While there has been work on linking the ARG prevalence with AMUs in various settings[7,9,11], incorporating AMU values, other non-antimicrobial stressors[59] and microbial signatures[12] in our analysis would likely help us better understand the microbiomes' response to selective pressures.

By utilizing the wealth of information on ARG abundances available in a collection of 214K metagenomic datasets, we have studied the co-abundance of ARGs to discover how these interactions shape the prevalence of resistances in different environments. The inferred correlation networks provide insights into how two resistance types indirectly and species-independent select for each other in different habitats. Our results further highlight that there are instances of genes of one type of resistance often co-occurring with many other types of resistance and that the environmental context plays an important role, revealing them as important targets in surveillance programs to limit their impact on global health.

References

1. O'Neill, J. & The Review on Antimicrobial Resistance (Chaired by Jim O'Neill). Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. *Rev. Antimicrob. Resist.* 1–80 (2016).
2. Murray, C. J. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629–655 (2022).
3. Zhang, Z. *et al.* Assessment of global health risk of antibiotic resistance genes. *Nat. Commun.* **13**, (2022).
4. Holmes, A. H. *et al.* Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* **387**, 176–187 (2016).
5. Collignon, P., Powers, J. H., Chiller, T. M., Aidara-Kane, A. & Aarestrup, F. M. World health organization ranking of antimicrobials according to their importance in human medicine: a critical step for developing risk management strategies for the use of

366 antimicrobials in food production animals. *Clin. Infect. Dis.* **49**, 132–141 (2009).
367 6. Jensen, H. H. & Hayes, D. J. Impact of Denmark's ban on antimicrobials for growth
368 promotion. *Curr. Opin. Microbiol.* **19**, 30–36 (2014).
369 7. Aarestrup, F. M. *et al.* Effect of abolishment of the use of antimicrobial agents for
370 growth promotion on occurrence of antimicrobial resistance in fecal enterococci from
371 food animals in Denmark. *Antimicrob. Agents Chemother.* **45**, 2054–2059 (2001).
372 8. Sun, J. *et al.* Development of aminoglycoside and β-lactamase resistance among
373 intestinal microbiota of swine treated with lincomycin, chlortetracycline, and
374 amoxicillin. *Front. Microbiol.* **5**, 580 (2014).
375 9. Aarestrup, F. M. Characterization of glycopeptide-resistant Enterococcus faecium
376 (GRE) from broilers and pigs in Denmark: Genetic evidence that persistence of GRE in
377 pig herds is associated with coselection by resistance to macrolides. *J. Clin. Microbiol.*
378 **38**, 2774–2777 (2000).
379 10. Looft, T. *et al.* In-feed antibiotic effects on the swine intestinal microbiome. *Proc.*
380 *Natl. Acad. Sci. U. S. A.* **109**, 1691–1696 (2012).
381 11. Andersen, V. D. *et al.* Predicting effects of changed antimicrobial usage on the
382 abundance of antimicrobial resistance genes in finisher' gut microbiomes. *Prev. Vet.*
383 *Med.* **174**, 104853 (2020).
384 12. Coyte, K. Z. *et al.* Horizontal gene transfer and ecological interactions jointly control
385 microbiome stability. *PLoS Biol.* **20**, 1–18 (2022).
386 13. Peterson, L. R. Squeezing the antibiotic balloon: The impact of antimicrobial classes
387 on emerging resistance. *Clin. Microbiol. Infect. Suppl.* **11**, 4–16 (2005).
388 14. Paterson, D. L. 'Collateral damage' from cephalosporin or quinolone antibiotic
389 therapy. *Clin. Infect. Dis.* **38**, 341–345 (2004).
390 15. Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to
391 explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351
392 (2012).
393 16. Yuan, M. M. *et al.* Climate warming enhances microbial network complexity and
394 stability. *Nat. Clim. Chang.* **11**, 343–348 (2021).
395 17. Zhang, Z. *et al.* The large-scale spatial patterns of ecological networks between
396 phytoplankton and zooplankton in coastal marine ecosystems. *Sci. Total Environ.* **827**,
397 154285 (2022).

398 18. Li, L. G., Xia, Y. & Zhang, T. Co-occurrence of antibiotic and metal resistance genes
399 revealed in complete genome collection. *ISME J.* **11**, 651–662 (2017).
400 19. Ju, F. *et al.* Antibiotic resistance genes and human bacterial pathogens: Co-
401 occurrence, removal, and enrichment in municipal sewage sludge digesters. *Water*
402 *Res.* **91**, 1–10 (2016).
403 20. Pan, X., Lin, L., Zhang, W., Dong, L. & Yang, Y. Metagenome sequencing to unveil the
404 resistome in a deep subtropical lake on the Yunnan-Guizhou Plateau, China. *Environ.*
405 *Pollut.* **263**, 114470 (2020).
406 21. Wu, J. *et al.* Prevalence and distribution of antibiotic resistance in marine fish farming
407 areas in Hainan, China. *Sci. Total Environ.* **653**, 605–611 (2019).
408 22. Liu, J. *et al.* The fecal resistome of dairy cattle is associated with diet during nursing.
409 *Nat. Commun.* **10**, (2019).
410 23. Munk, P. *et al.* Abundance and diversity of the faecal resistome in slaughter pigs and
411 broilers in nine European countries. *Nat. Microbiol.* **3**, 898–908 (2018).
412 24. Hendriksen, R. S. *et al.* Global monitoring of antimicrobial resistance based on
413 metagenomics analyses of urban sewage. *Nat. Commun.* **10**, (2019).
414 25. Martiny, H.-M. *et al.* Global distribution of mcr gene variants in 214,095 metagenomic
415 samples. *mSystems* (2022). doi:10.1128/msystems.00105-22
416 26. Karkman, A., Do, T. T., Walsh, F. & Virta, M. P. J. Antibiotic-Resistance Genes in Waste
417 Water. *Trends Microbiol.* **26**, 220–228 (2018).
418 27. Martiny, H.-M., Munk, P., Brinch, C., Aarestrup, F. M. & Petersen, T. N. A curated data
419 resource of 214K metagenomes for characterization of the global antimicrobial
420 resistome. *PLOS Biol.* **20**, e3001792 (2022).
421 28. Aitchison, J. The Statistical Analysis of Compositional Data. *J. R. Stat. Soc. Ser. B* **44**,
422 139–160 (1982).
423 29. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome
424 datasets are compositional: And this is not optional. *Front. Microbiol.* **8**, 1–6 (2017).
425 30. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey Data.
426 *PLoS Comput. Biol.* **8**, 1–11 (2012).
427 31. Leinonen, R. *et al.* The European nucleotide archive. *Nucleic Acids Res.* **39**, 44–47
428 (2011).
429 32. Bushnell, B. BBMap. (2014).

430 33. Clausen, P. T. L. C., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw
431 reads against redundant databases with KMA. *BMC Bioinformatics* **19**, 1–8 (2018).
432 34. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J.*
433 *Antimicrob. Chemother.* **67**, 2640–2644 (2012).
434 35. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data
435 processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).
436 36. Edgar, R. C. & Bateman, A. Search and clustering orders of magnitude faster than
437 BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
438 37. Team, R. C. R: A Language and Environment for Statistical Computing. (2021).
439 38. Csardi, G. & Nepusz, T. The igraph software package for complex network research.
440 *InterJournal* **Complex Sy**, 1695 (2006).
441 39. Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D. & Borsboom, D.
442 qgraph: Network Visualizations of Relationships in Psychometric Data. *J. Stat. Softw.*
443 **48**, 1–18 (2012).
444 40. Pedersen, T. L. ggraph: An Implementation of Grammar of Graphics for Graphs and
445 Networks. (2021).
446 41. Fructherman, T. M. J. & Reingold, E. M. Graph Drawing by Force-directed Placement.
447 *Softw. - Pract. Exp.* **21**, 1129–1164 (1991).
448 42. Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of
449 complex weighted networks. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 3747 (2004).
450 43. Parente, E., Zotta, T., Faust, K., De Filippis, F. & Ercolini, D. Structure of association
451 networks in food bacterial communities. *Food Microbiol.* **73**, 49–60 (2018).
452 44. Wasserman, S. & Faust, K. *Social network analysis: Methods and applications.*
453 (Cambridge University Press, 1994).
454 45. Silver, N. C. & Dunlap, W. P. Averaging Correlation Coefficients: Should Fisher's z
455 Transformation Be Used? *J. Appl. Psychol.* **72**, 146–148 (1987).
456 46. Freitas, A. R. *et al.* High-Resolution Genotyping Unveils Identical Ampicillin-Resistant
457 Enterococcus faecium Strains in Different Sources and Countries: A One Health
458 Approach. *Microorganisms* **10**, (2022).
459 47. Giovanetti, E., Brenciani, A., Lupidi, R., Roberts, M. C. & Varaldo, P. E. Presence of the
460 tet(O) gene in erythromycin- and tetracycline-resistant strains of Streptococcus
461 pyogenes and linkage with either the mef(A) or the erm(A) gene. *Antimicrob. Agents*

462 *Chemother.* **47**, 2844–2849 (2003).
463 48. Trongjit, S. & Chuanchuen, R. Whole genome sequencing and characteristics of
464 Escherichia coli with coexistence of ESBL and mcr genes from pigs. *PLoS One* **16**, 1–17
465 (2021).
466 49. WHO. *Critically Important Antimicrobials for Human Medicine, 6th Revision*. (2018).
467 50. Arredondo, A., Blanc, V., Mor, C., Nart, J. & León, R. Tetracycline and multidrug
468 resistance in the oral microbiota: differences between healthy subjects and patients
469 with periodontitis in Spain. *J. Oral Microbiol.* **13**, (2021).
470 51. Nielsen, H. U. K. *et al.* Tetracycline and macrolide co-resistance in Streptococcus
471 pyogenes: Co-selection as a reason for increase in macrolide-resistant S. pyogenes?
472 *Microb. Drug Resist.* **10**, 231–238 (2004).
473 52. Ardic, N., Ozyurt, M., Sareyyupoglu, B. & Haznedaroglu, T. Investigation of
474 erythromycin and tetracycline resistance genes in methicillin-resistant staphylococci.
475 *Int. J. Antimicrob. Agents* **26**, 213–218 (2005).
476 53. Stepanauskas, R. *et al.* Coselection for microbial resistance to metals and antibiotics
477 in freshwater microcosms. *Environ. Microbiol.* **8**, 1510–1514 (2006).
478 54. Gao, P., Munir, M. & Xagoraraki, I. Correlation of tetracycline and sulfonamide
479 antibiotics with corresponding resistance genes and resistant bacteria in a
480 conventional municipal wastewater treatment plant. *Sci. Total Environ.* **421**–**422**,
481 173–183 (2012).
482 55. Zhang, M. Q., Yuan, L., Li, Z. H., Zhang, H. C. & Sheng, G. P. Tetracycline exposure
483 shifted microbial communities and enriched antibiotic resistance genes in the aerobic
484 granular sludge. *Environ. Int.* **130**, 104902 (2019).
485 56. Tarek, M. H. & Garner, E. A proposed framework for the identification of indicator
486 genes for monitoring antibiotic resistance in wastewater: Insights from metagenomic
487 sequencing. *Sci. Total Environ.* 158698 (2022). doi:10.1016/J.SCITOTENV.2022.158698
488 57. Blanchet, F. G., Cazelles, K. & Gravel, D. Co-occurrence is not evidence of ecological
489 interactions. *Ecol. Lett.* **23**, 1050–1063 (2020).
490 58. Kurtz, Z. D. *et al.* Sparse and Compositionally Robust Inference of Microbial Ecological
491 Networks. *PLOS Comput. Biol.* **11**, e1004226 (2015).
492 59. Nagata, N. *et al.* Population-level Metagenomics Uncovers Distinct Effects of Multiple
493 Medications on the Human Gut Microbiome. *Gastroenterology* **163**, 1038–1052

494 (2022).

495

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

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

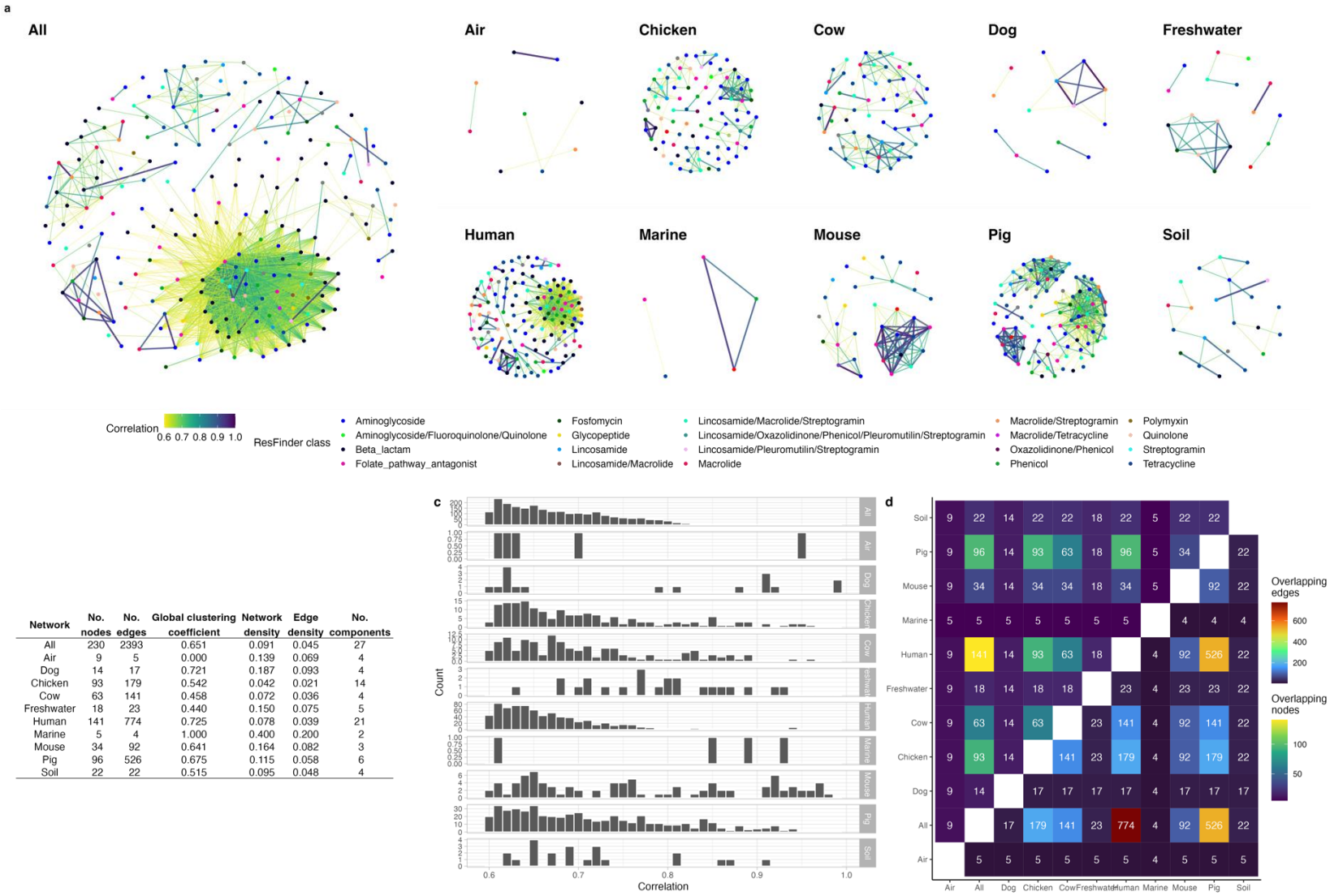
498

Resistance class	Air	All	Chicken	Cow	Dog	Freshwater	Human	Marine	Mouse	Pig	Soil
Tetracycline (64)	13.61	32.48	53.99	60.05	32.27	5.2	36.39	0.06	59.97	47.76	3.18
Streptogramin (11)	<0.01	0.15	3.3	<0.01	-	<0.01	<0.01	<0.01	<0.01	0.11	0.02
Steroid_antibacterial (2)	0.09	0.06	<0.01	<0.01	-	<0.01	0.03	<0.01	<0.01	<0.01	<0.01
Quinolone (17)	0.3	1.19	0.14	<0.01	<0.01	1.85	1.26	<0.01	0.18	0.14	0.12
Polymyxin (10)	0.02	0.06	<0.01	<0.01	-	0.39	0.06	<0.01	0.01	0.07	<0.01
Pleuromutilin (1)	-	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Phenicol (33)	6.19	9.21	6.1	4.33	0.36	55.89	3.33	97.37	0.14	0.81	85.04
Oxazolidinone/Phenicol/Tetracycline (1)	0.02	0.07	1.54	<0.01	<0.01	<0.01	<0.01	<0.01	0.03	0.04	<0.01
Oxazolidinone/Phenicol (1)	0.03	0.02	0.1	<0.01	-	<0.01	<0.01	<0.01	-	0.09	<0.01
Missing (26)	0.31	6.65	0.17	1.13	19.4	1.34	10.77	<0.01	0.9	2.83	0.2
Macrolide/Tetracycline (1)	<0.01	<0.01	<0.01	<0.01	-	<0.01	<0.01	<0.01	-	<0.01	<0.01
Macrolide/Streptogramin (4)	1.69	1.1	0.12	0.39	1.01	1.18	1.43	<0.01	1.83	0.48	0.11
Macrolide (24)	1.86	3.04	0.36	16.71	5.78	1.9	2.53	<0.01	4.22	6.83	0.25
Lincosamide/Streptogramin (2)	0.12	0.28	0.02	<0.01	<0.01	0.01	0.38	<0.01	0.06	<0.01	<0.01
Lincosamide/Pleuromutilin/Streptogramin (7)	0.48	0.27	0.09	0.02	0.18	0.02	0.34	<0.01	<0.01	0.21	0.05
Lincosamide/Oxazolidinone/Phenicol/Pleuromutilin/Streptogramin (3)	0.41	0.49	1.36	0.5	0.01	<0.01	0.45	<0.01	1.64	1.25	0.04
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/Macrolide/Streptogramin (44)	6.12	12.33	14.51	2.26	1.35	0.32	17.71	0.03	8.19	9.52	5.95
Lincosamide/Macrolide (1)	<0.01	<0.01	<0.01	<0.01	-	<0.01	<0.01	<0.01	-	-	-
Lincosamide (9)	1.88	3.33	3.42	8.29	29	0.36	2.02	<0.01	4.01	11.61	0.71
Glycopeptide (27)	<0.01	0.29	0.01	<0.01	-	0.01	0.58	<0.01	0.72	0.02	0.01
Fosfomycin (28)	4.36	1.07	0.2	<0.01	0.01	1.29	1.49	<0.01	0.22	0.03	0.19
Folate_pathway_antagonist (37)	2.92	3.2	2.19	0.41	0.97	5.45	3.13	0.36	4.33	1.03	0.82
Beta_lactam (216)	41.3	14.32	0.64	0.57	0.7	7.28	8.34	1.84	2.14	2.89	1.69
Aminoglycoside/Fluoroquinolone/Quinolone (1)	<0.01	0.19	0.09	-	<0.01	0.46	0.15	<0.01	0.01	<0.01	<0.01
Aminoglycoside/Fluoroquinolone/Macrolide/Phenicol/Rifampicin/Tetracycline (1)	0.63	1.08	0.12	0.07	0.52	0.34	1.08	0.2	1.2	0.17	0.02
Aminoglycoside (144)	17.64	9.12	11.51	5.24	8.42	16.7	8.51	0.08	10.12	14.08	1.58

Source

Relative abundance: 75, 50, 25

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 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	230	2393	0.651	0.091	0.045	27
Air	9	5	0.000	0.139	0.069	4
Dog	14	17	0.721	0.187	0.093	4
Chicken	93	179	0.542	0.042	0.021	14
Cow	63	141	0.458	0.072	0.036	4
Freshwater	18	23	0.440	0.150	0.075	5
Human	141	774	0.725	0.078	0.039	21
Marine	5	4	1.000	0.400	0.200	2
Mouse	34	92	0.641	0.164	0.082	3
Pig	96	526	0.675	0.115	0.058	6
Soil	22	22	0.515	0.095	0.048	4

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.

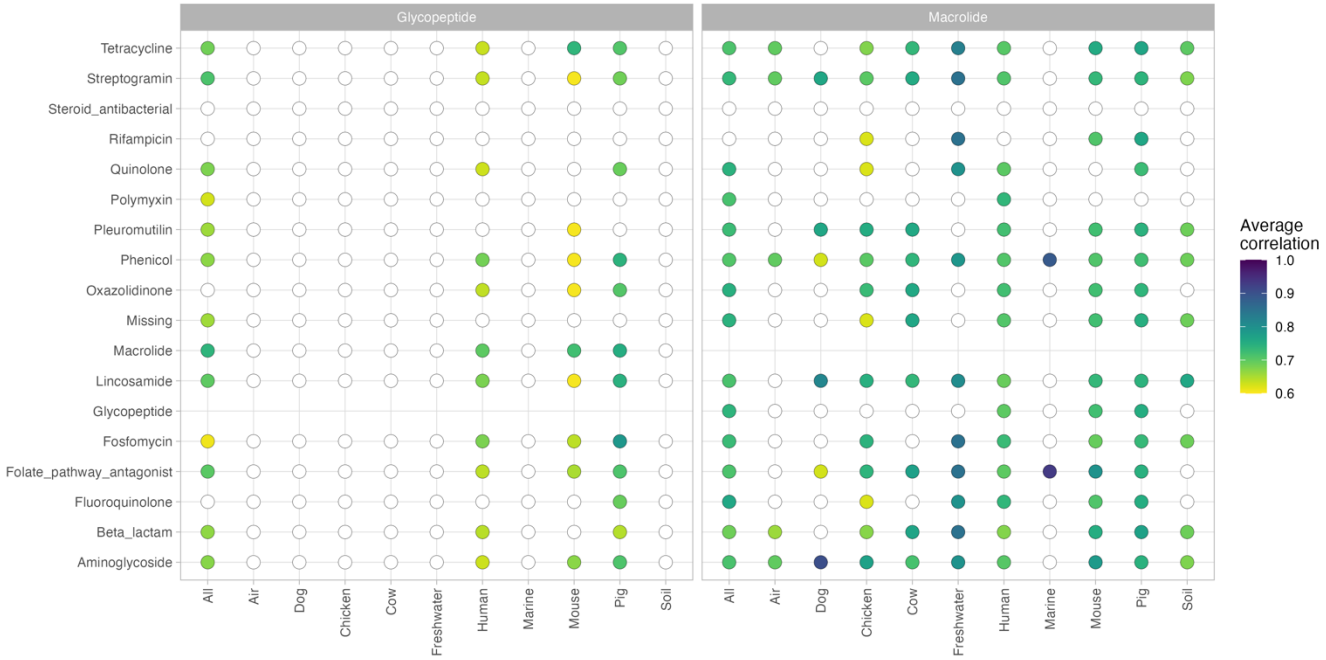


Figure 2: ***Correlation profiles for ARGs conferring resistance to the critically important antimicrobial classes glycopeptides (left) and macrolides (right).*** *Each column shows the average correlation from, e.g., macrolide ARGs to ARGs for other antimicrobial classes. The circle is colored by the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

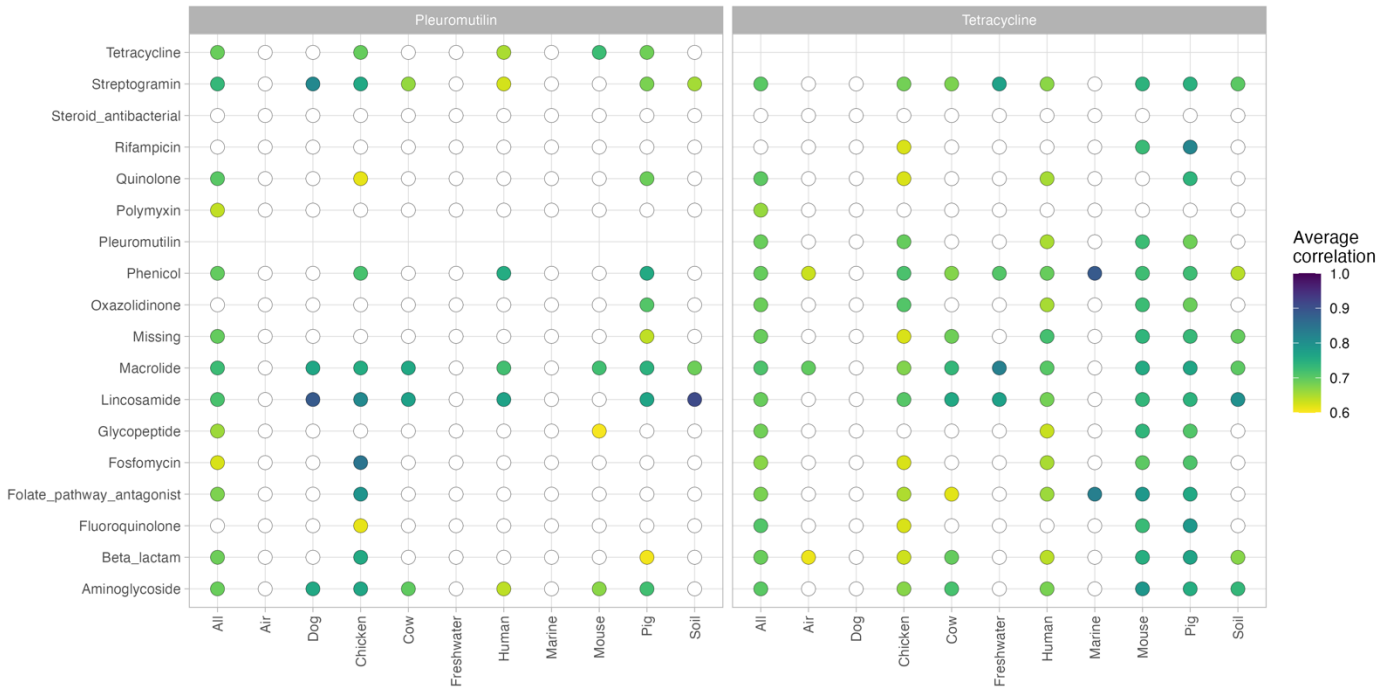


Figure 3: ***Correlation profiles for ARGs conferring resistance to important pleuromutilins (left) and highly important tetracyclines (right).*** *Each column shows the average correlation from, e.g., tetracycline ARGs to ARGs for other antimicrobial classes. The circle is colored by the average correlation,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

Supplementary materials

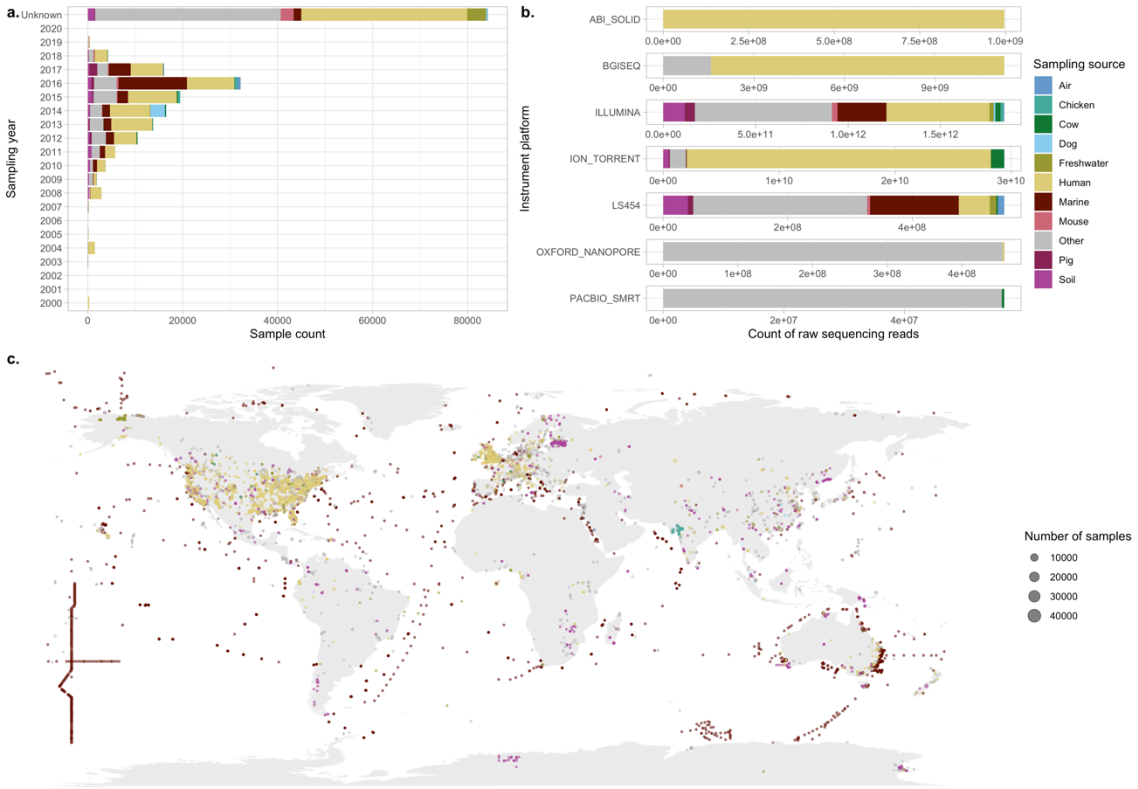


Figure S1: ***Metagenomic origins colored by the sampling group****. a. Overview of sampling year for metagenomic samples colored by their sampling origin. 100 samples were taken before 2000, and 84,238 did not have a valid sampling date. b. Overview of the amount of sequencing reads available for each 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.*

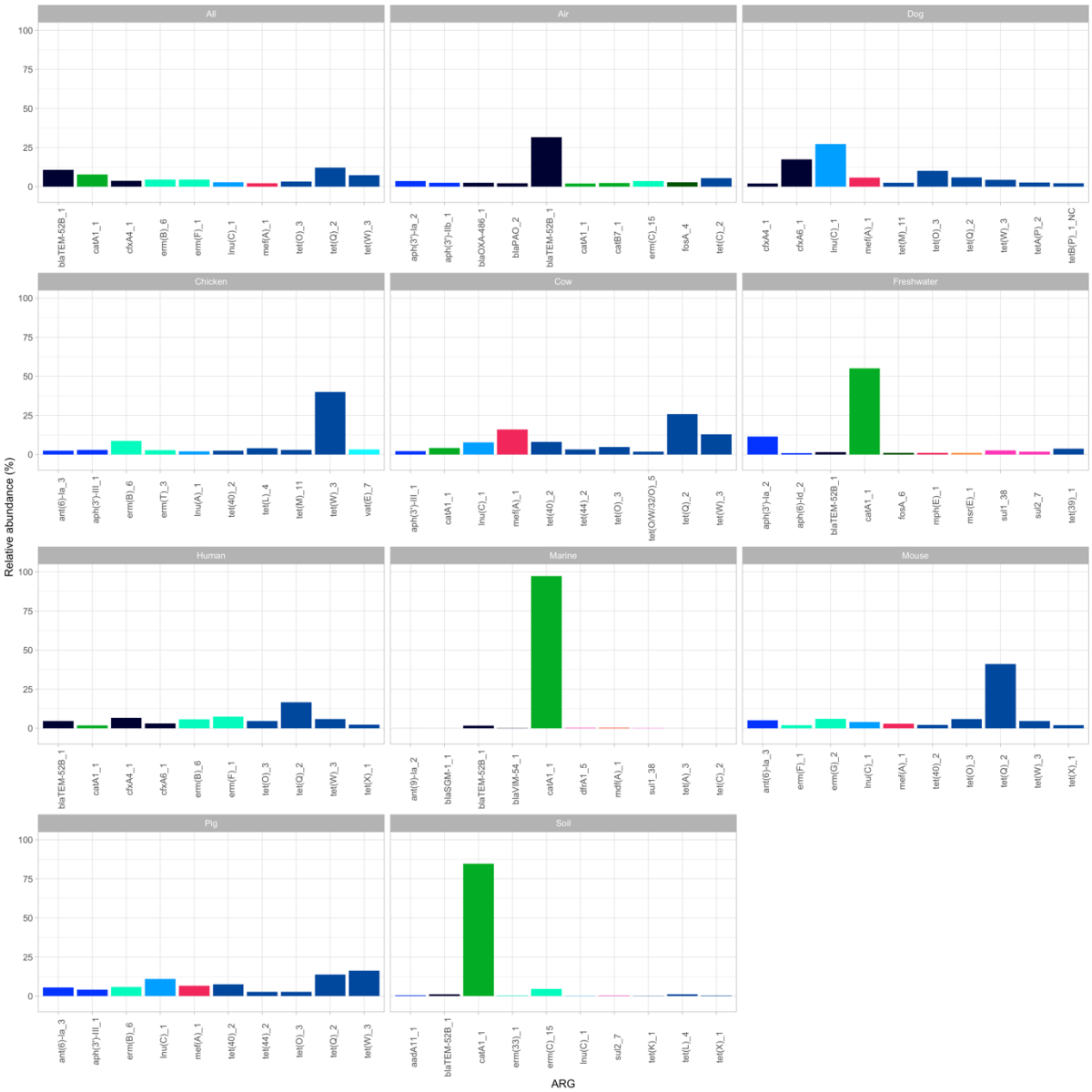
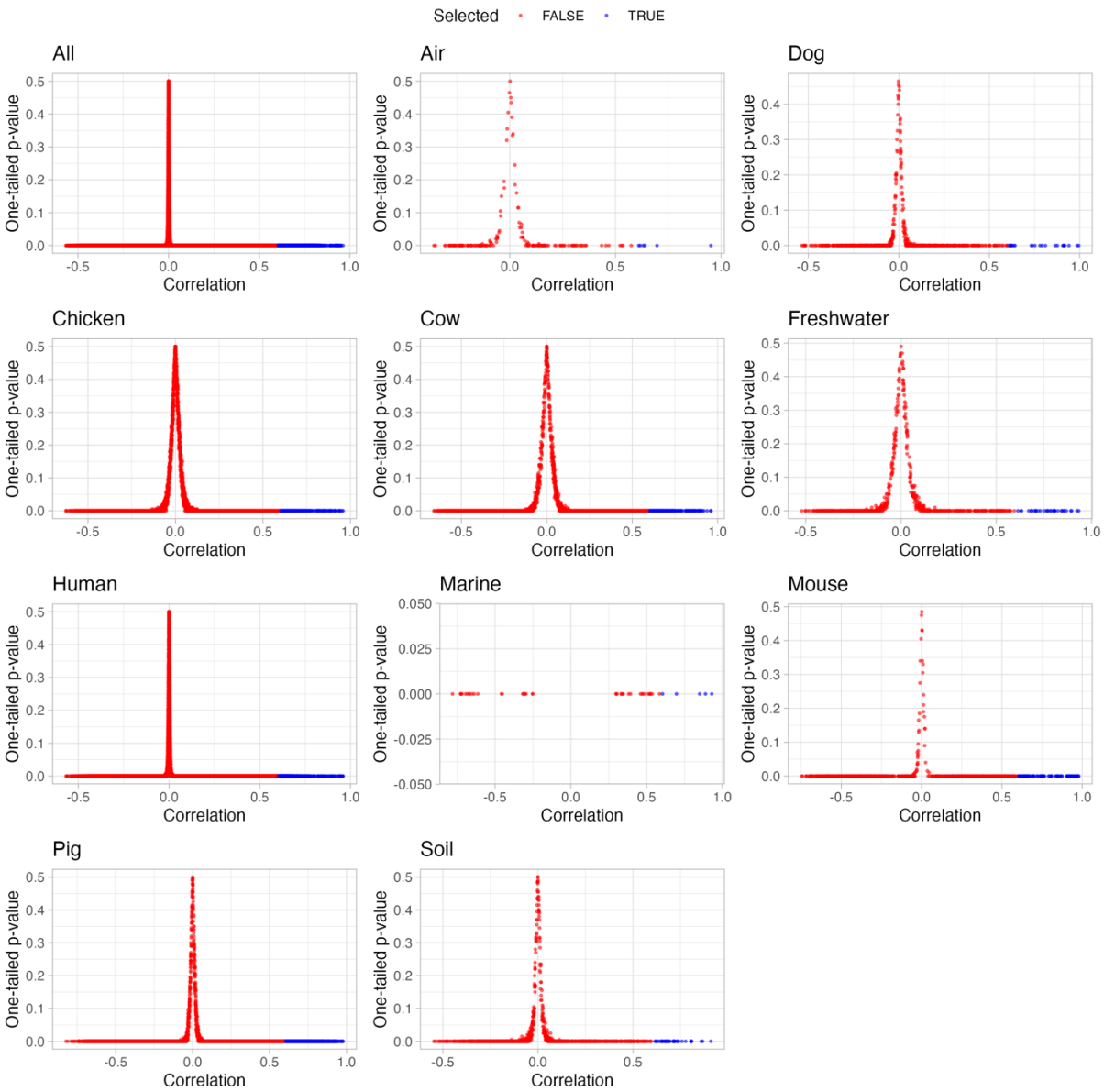


Figure S2: ***Top 10 most abundant ARGs per sampling group.*** *A barplot showing the ARGs having the highest 10 relative abundances per sampling source colored by resistance class.*

	Filter		Network		
Source	minF	minS	I	N	E
All	0	1	716	447	35,573
	25	10	459	241	3,899
	50	10	425	230	2,393
	50	20	313	137	223
	50	50	313	137	223
Air	0	1	436	314	15,079
	25	10	34	14	9
	50	10	19	9	5
Dog	0	1	158	113	1,318
	25	10	45	19	24
	50	10	38	14	17
	50	20	19	7	4
	50	50	19	7	4
Chicken	0	1	389	323	15,658
	25	10	145	114	235
	50	10	123	93	179
	50	20	69	52	71
	50	50	69	52	71
Cow	0	1	443	365	25,632
	25	10	94	72	179
	50	10	85	63	141
	50	20	37	23	32
	50	50	37	23	32
Freshwater	0	1	579	474	19,645
	25	10	56	32	57
	50	10	36	18	23
Human	0	1	692	462	48,821
	25	10	307	158	1210
	50	10	278	141	774
	50	20	201	75	86
	50	50	201	75	86
Marine	0	1	668	15	9
	25	10	24	5	4
	50	10	10	5	4
Mouse	0	1	326	253	10,411
	25	10	55	38	100
	50	10	41	34	92
	50	20	25	21	60
	50	50	25	21	60
Pig	0	1	433	383	21,488
	25	10	136	109	743
	50	10	118	96	526

	Filter		Network		
Source	minF	minS	I	N	E
	50	20	61	49	141
	50	50	61	49	141
Soil	0	1	613	433	9,046
	25	10	72	43	57
	50	10	47	22	22

Table S1: Count filtering of the observed ARG counts (I) greatly influences the number of ARGs (nodes, N) and the correlations (edges, E). For an ARG to be included in the SparCC analysis, it had to pass the two following criteria: it had to have a minimum number of fragments detected (minF) and had to pass the minF criteria for a minimum number of samples (minS).



537

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

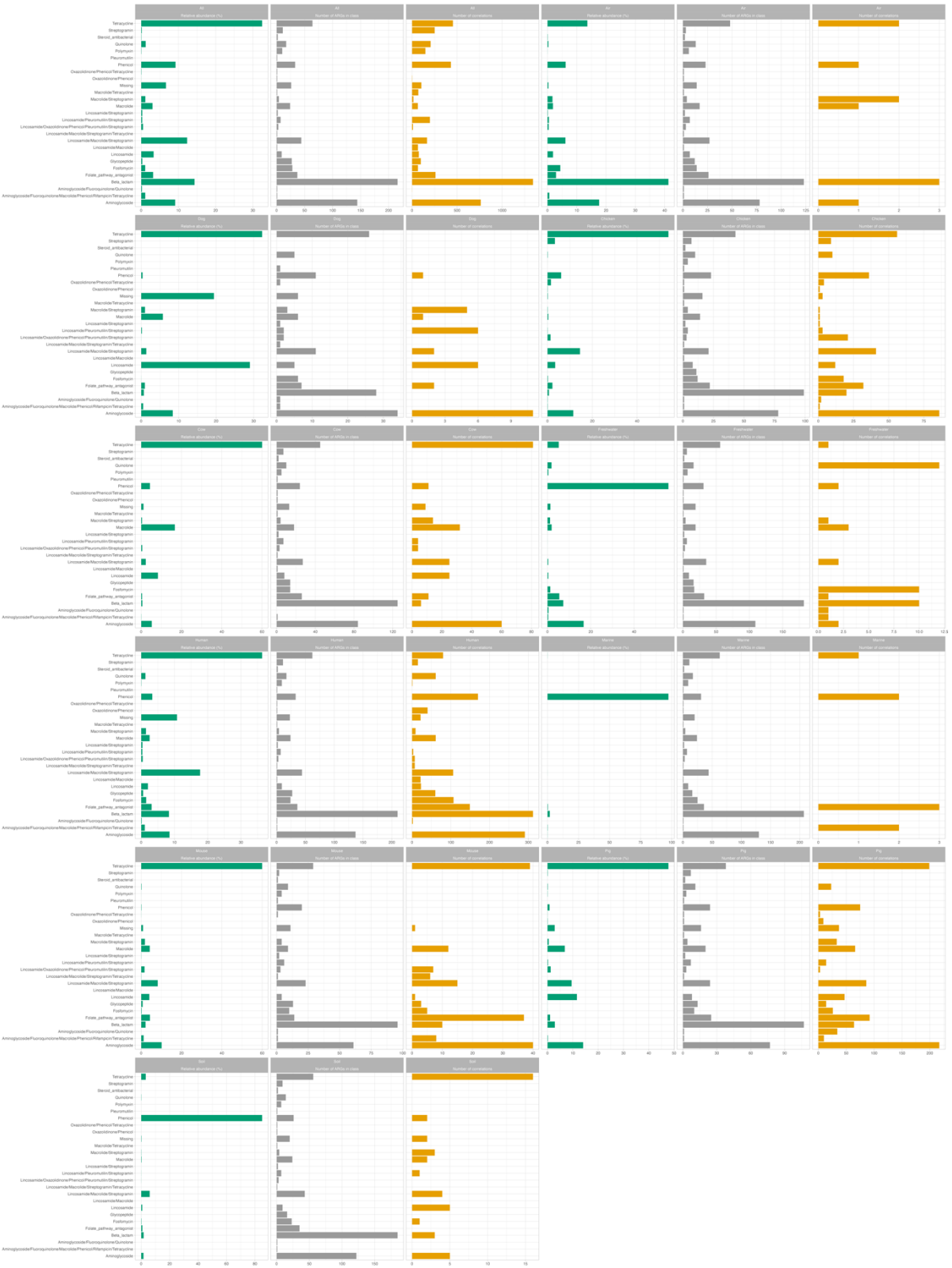


Figure S4: ***The relative abundance, number of ARGs, and the number of correlations for each resistance class in each sampling group.*** *The relative abundance in green shows the percentage of read fragments for each class. In grey is the number of ARGs for each resistance class. Orange coloring indicates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

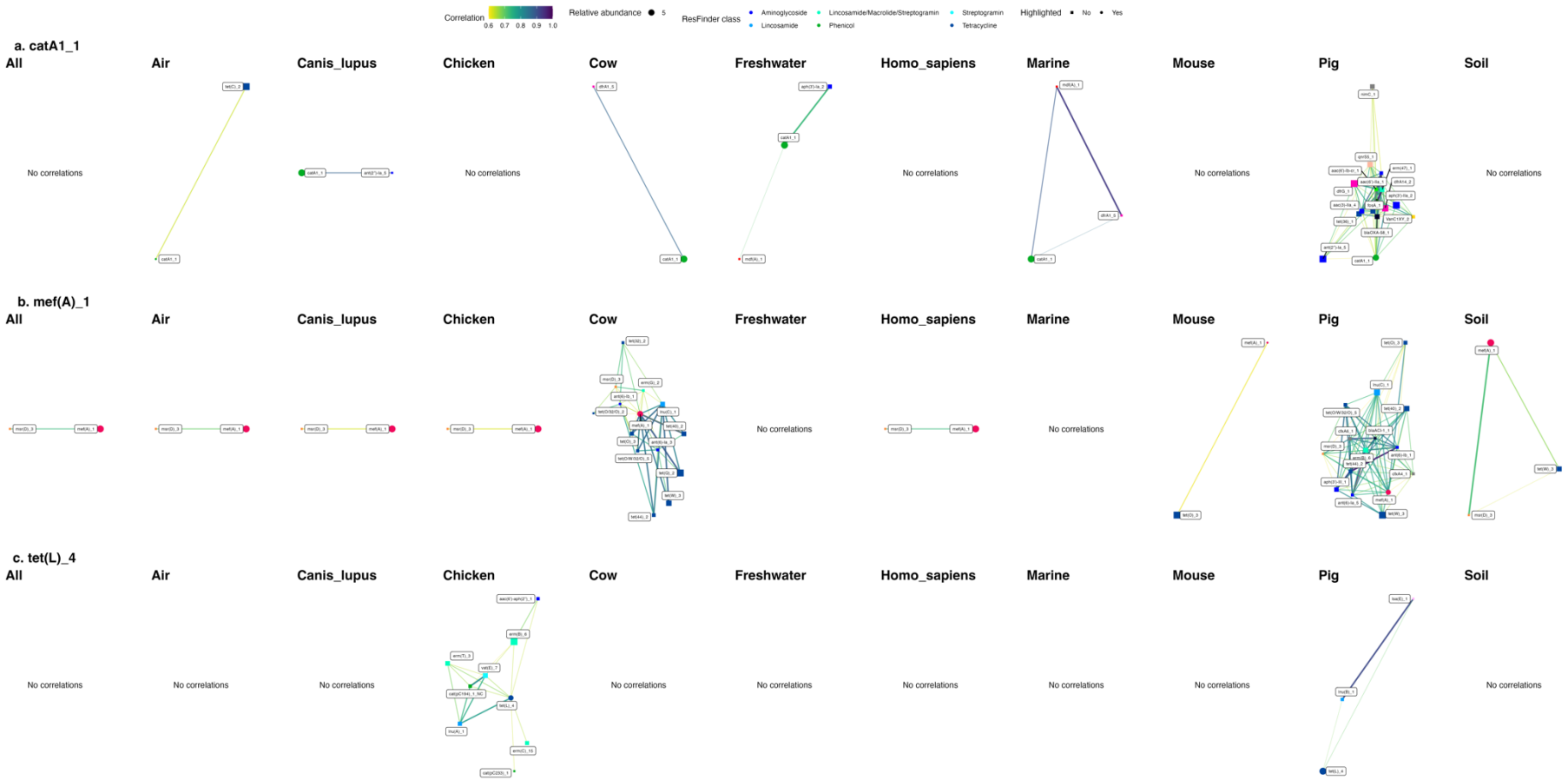


Figure S5: ***Networks illustrating how one ARG interacts differently depending on the sampling environments****. The three genes catA1_1, mef(A)_1, and tet(L)_4 are selected to illustrate that the genes correlate more or less with other genes in their abundance in some habitats.*

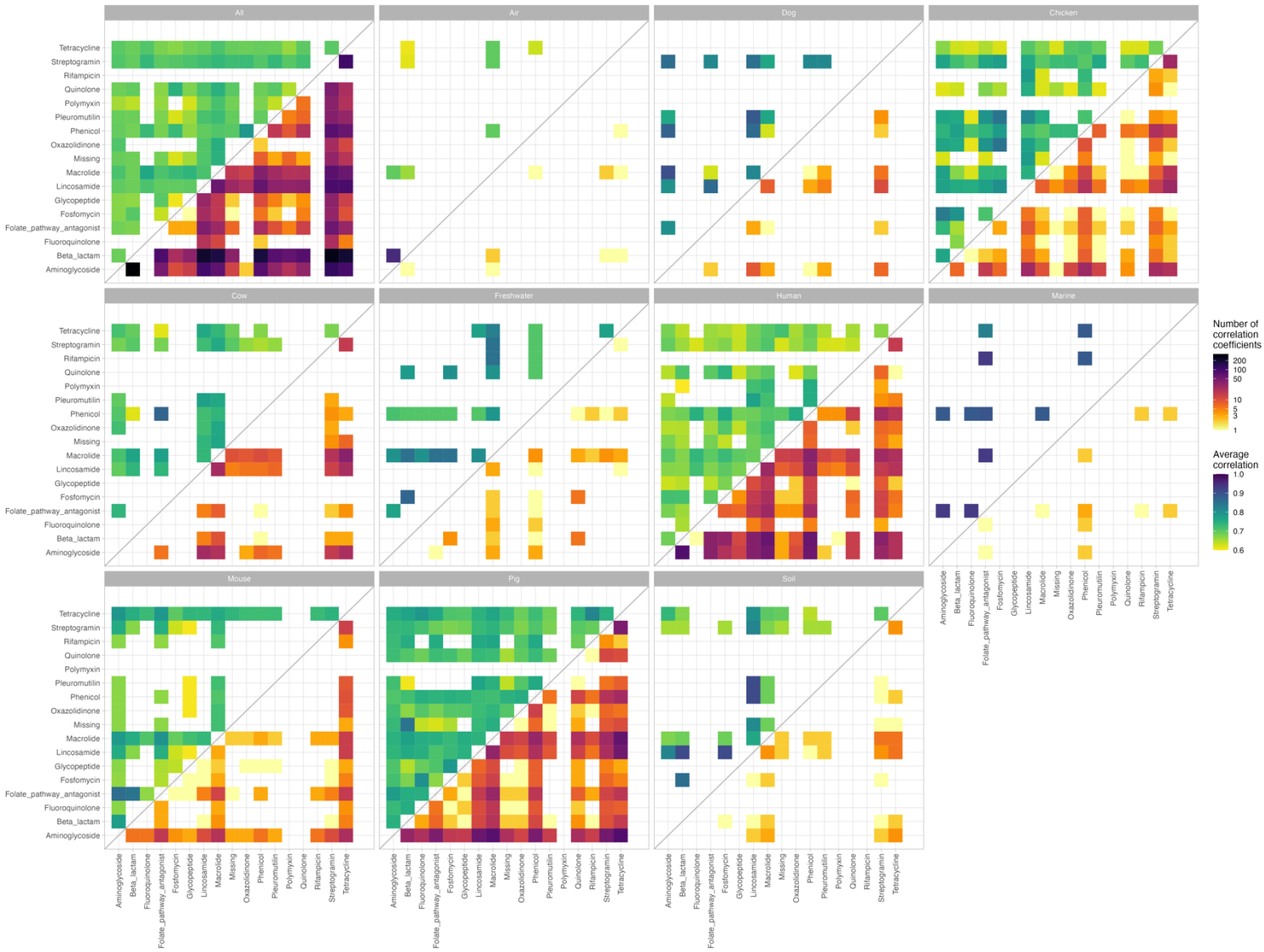
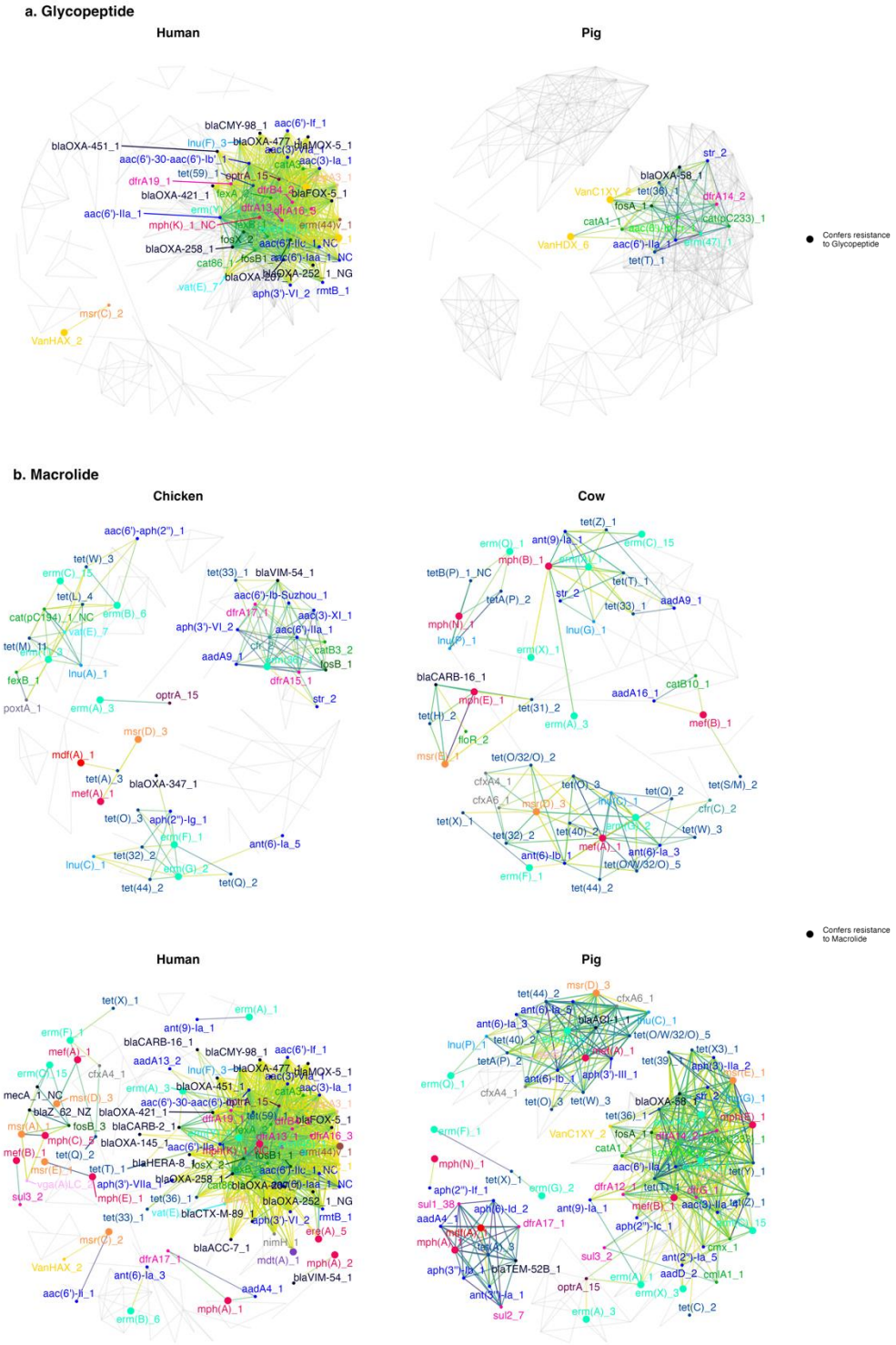


Figure S6: The average correlation between ARGs of different resistance classes in the upper triangle of the heatmaps with the lower half shows the correlation coefficients between the two classes. Note that a correlation coefficient between two ARGs might be present in more than one tile, as some ARGs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



556

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

560

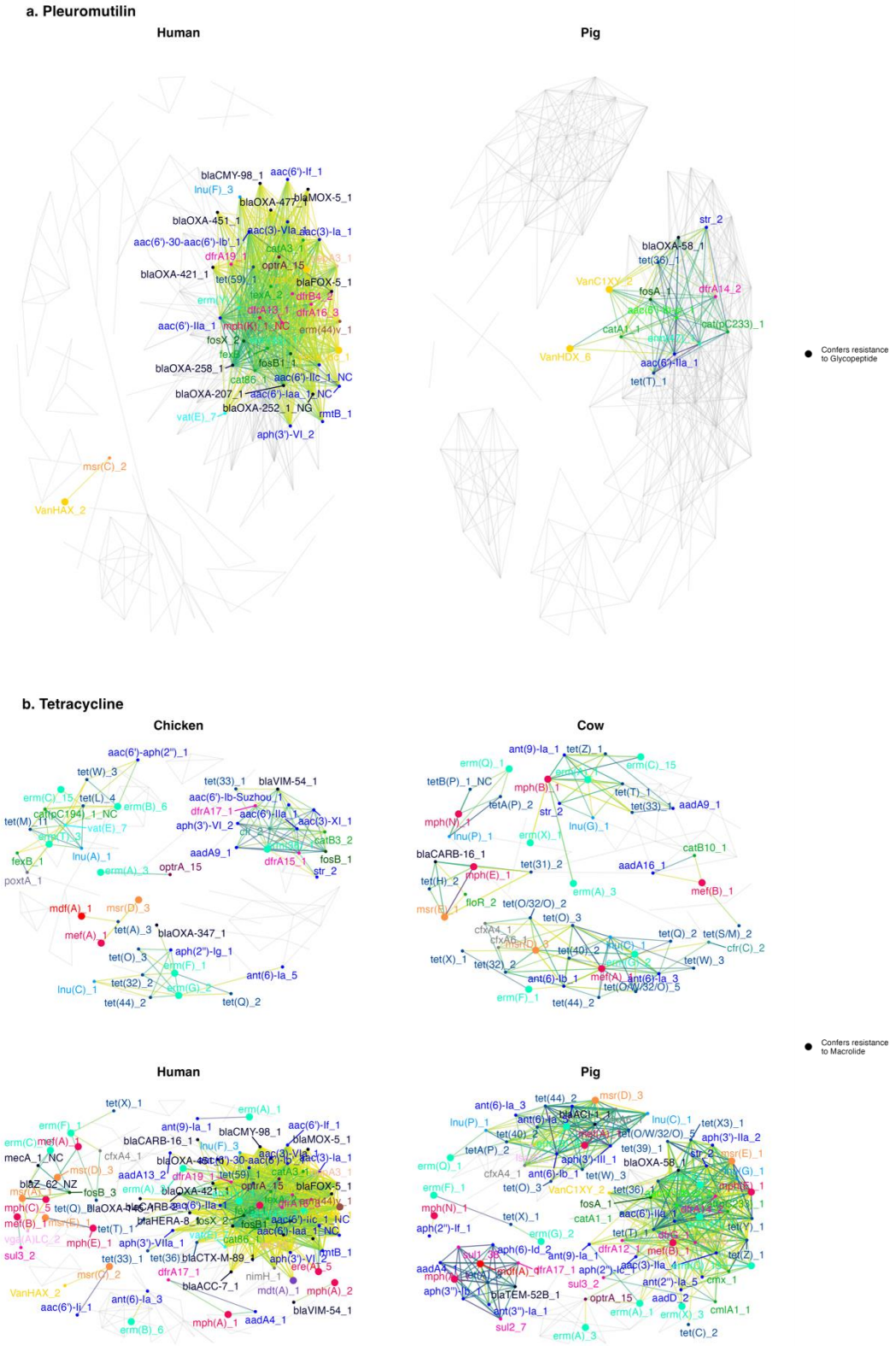


Figure S8: ***Highlights of interactions between ARGs of a. pleuromutilin resistance and b. tetracycline resistance.*** *The ARG node size shows whether the corresponding ARG gives resistance to the antimicrobial class in focus. Only correlation edges and ARG nodes are colored if they correlate with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