

1 **Antimicrobial resistance genes predict plasmid generalism and network
2 structure in wastewater**

3 Alice Risely¹, Thibault Stalder^{4,5}, Benno I. Simmons², Eva M. Top^{4,5}, Angus Buckling^{2,3}, Dirk
4 Sanders^{2,3 *}

5

6 ¹ Institute of Evolutionary Ecology and Conservation Genomics, University of Ulm, Ulm, Germany

7 ² Centre for Ecology & Conservation, College of Life and Environmental Sciences, University of
8 Exeter, Penryn, Cornwall TR10 9FE, United Kingdom

9 ³ Environment and Sustainability Institute, University of Exeter, Penryn, Cornwall TR10 9FE, United
10 Kingdom

11 ⁴Department of Biological Sciences, University of Idaho, Moscow, ID, USA

12 ⁵Institute for Interdisciplinary Data Sciences, University of Idaho, Moscow, ID, USA
13

14 *Corresponding author:

15 Dirk Sanders

16 Email: d.sanders@exeter.ac.uk

17

18 AR Orcid ID: 0000-0002-0731-2934

19 BIS Orcid ID: 0000-0002-2751-9430

20 DS Orcid ID: 0000-0003-2383-8693

21

22

23 **Keywords:** microbial communities, antibiotic resistance, mutualistic and antagonistic interactions,
24 network structure, Hi-C, ecological networks

25

26

27 **Abstract.** Plasmids are mobile genetic elements that can act as mutualists or parasites to their
28 bacterial hosts depending on their accessory genes and environment. Ecological network theory
29 predicts that mutualists, such as plasmids with antimicrobial resistance (AMR) genes in the presence
30 of antimicrobials, should act as generalists, while plasmids without beneficial genes are expected to
31 be more specialised. Therefore, whether the relationship between plasmid and host is mutualistic or
32 antagonistic is likely to have a strong impact on the formation of interaction network structures and the
33 spread of AMR genes across microbial networks. Here we re-analyse Hi-C metagenome data from
34 wastewater samples and identify plasmid signatures with machine learning to generate a natural host-
35 plasmid network. We found that AMR-carrying plasmids indeed interacted with more hosts than non-
36 AMR plasmids (on average 14 versus 3, respectively). The AMR plasmid-host subnetwork showed a
37 much higher connectedness and nestedness than the subnetwork associated with non-AMR
38 plasmids. The overall network was clustered around Proteobacteria and AMR-carrying plasmids
39 giving them a crucial role in network connectivity. Therefore, by forming mutualistic networks with their
40 hosts, beneficial AMR plasmids lead to more connected network structures that ultimately share a
41 larger gene pool of AMR genes across the network.

42

43 **Introduction**

44 Plasmids play a key role in the spread of antimicrobial resistance (AMR) and other genes
45 (e.g., metal resistance, biodegradation, virulence), both within and between bacterial taxa (Bennett
46 2008; Martínez 2008; Dang *et al.* 2017; San Millan 2018; Acman *et al.* 2020). Understanding the
47 ecological mechanisms that underpin plasmid transmission within bacterial communities is important
48 for combating the spread of AMR and associated bacterial epidemics (Dimitriu *et al.* 2021). However,
49 our knowledge about plasmid-host interactions is mostly gained from laboratory research on a limited
50 number of bacteria and plasmids. Therefore, there remains considerable uncertainty surrounding the
51 role of plasmids within larger communities and the resulting plasmid-host networks in nature. This
52 limits our understanding of the ecological and evolutionary processes driving plasmid transmission
53 across natural microbial communities.

54 Ecological interaction network analysis can provide crucial insights into community structure
55 (Kaiser-Bunbury *et al.* 2017), the speed of disease transmission (González-Salazar & Stephens
56 2012), the dynamics of coevolution (Guimarães *et al.* 2017) and community stability (Thébault &
57 Fontaine 2010; Veron *et al.* 2018). Recently, the ecological network approach has been applied to
58 microbiological systems to elucidate the ecological mechanisms that underpin microbial dynamics
59 (Flores *et al.* 2013; Weitz *et al.* 2013; Coyte *et al.* 2015; Wang *et al.* 2016). The structure of ecological
60 interaction networks is often related to the prevalent type of interaction between species, with
61 antagonistic and mutualistic interactions associated with different network structures (Thébault &
62 Fontaine 2010; Montesinos-Navarro *et al.* 2017). Theory predicts that the structure for each network
63 type leads to increased stability: mutualistic networks tend to be nested with generalists linking the
64 whole network while antagonistic networks often have a modular structure dominated by specialist
65 interactions (Thébault & Fontaine 2010). Observation studies support theoretical predictions that
66 ecological interaction networks dominated by antagonists tend to have fewer generalist interactions
67 than mutualistic networks (Fontaine *et al.* 2009; Thébault & Fontaine 2010; Montesinos-Navarro *et al.*
68 2017; Newbury *et al.* in press). Network structure is also influenced by the coevolutionary history of
69 interacting species, because many interactions are evolutionary constrained (Segar *et al.* 2020). For
70 example, due to coevolutionary arms races we can expect interactions in antagonistic networks to be
71 characterised by stronger phylogenetic signal than in mutualist networks (Rohr & Bascompte 2014).

72 These theories can potentially be applied to bacteria-plasmid networks, because plasmids
73 have highly variable host ranges (Suzuki *et al.* 2010; Klümper *et al.* 2015) and can be parasitic and
74 mutualistic to their bacterial hosts (Lili *et al.* 2007; Harrison & Brockhurst 2012). In many cases,
75 plasmids simply impose a fitness cost on their host, and survive in host communities through the
76 evolution of a lower fitness cost, high transmission rates, high fidelity of partitioning, and sophisticated
77 killing systems ensuring their stable presence within bacterial lines (Jensen & Gerdes 1995).
78 However, plasmids often carry accessory genes, such as AMR genes, that promote the survival of the
79 host bacteria under certain environmental contexts (Lili *et al.* 2007; Harrison & Brockhurst 2012).
80 Plasmids that carry context-dependent beneficial genes may therefore be expected to interact with
81 more bacterial taxa than those that don't (Fig. 1A-C). This will be driven by both ecological dynamics
82 (selection for hosts associated with a mutualistic plasmid) and coevolutionary dynamics between
83 hosts and plasmids leading to the evolution of generalism.

84 Over 10,000 plasmids have been described and referenced to date (likely only a small
85 fraction of real numbers), the majority of which are associated with the bacterial phylum
86 Proteobacteria, especially Alpha- and Gammaproteobacteria (Redondo-Salvo *et al.* 2020). Both
87 specialist and generalist plasmids have been identified, yet an analysis of a subset of described
88 plasmids suggests that up to 60% of plasmids are associated with multiple bacterial host species, and
89 that transmission is limited by host phylogeny with only a small number of super-generalist plasmids
90 (Suzuki *et al.* 2010; Redondo-Salvo *et al.* 2020). However, the link between a plasmid's range of
91 hosts in a natural microbial community and its likely effect on bacterial host fitness in that community
92 remains unknown. Note that in this study we distinguish between the concepts of host range and
93 generalism. Host range, as it has been defined by plasmid biologists, is the ability of a plasmid to
94 replicate in phylogenetically diverse hosts (i.e., narrow host range versus broad host range). This view
95 of plasmid host range does not necessarily reflect the actual spread of the plasmid in a bacterial
96 community (e.g., a broad-host-range plasmid could be found in only one phylogenetically distinct
97 species of a community and thus be considered a specialist while a narrow-host range plasmid could
98 be found in many different closely related species and be a generalist plasmid). Hereafter we refer to
99 generalist plasmids as plasmids that are found in many species within a community, and specialist
100 plasmids as those that are found in very few host species.

101 To understand the functional role of plasmids in ecological communities, ranging from soil, to
102 wastewater, to the gut microbiome, we need to identify interactions between plasmids and their hosts
103 in their natural environment. Recently, proximity-ligation methods such as Hi-C allow us to overcome
104 previous sampling limitations and have been used to detect associations between DNA molecules
105 originating in the same cell within microbial communities (Stalder *et al.* 2019; Kent *et al.* 2020; Yaffe &
106 Relman 2020). Moreover, machine learning techniques that can identify plasmids based on their
107 genomic signatures, such as GC content and (tetra)nucleotide composition-specific sequences, with
108 96% accuracy, allowing for the detection of plasmid genomes that have not officially been described
109 (Krawczyk *et al.* 2018; Pellow *et al.* 2020). Together, these novel methods allow us to quantify host
110 range distributions of plasmids in natural bacterial communities.

111 Here we apply an ecological network analysis to test the effect of mutualistic interactions on
112 the structure of a natural plasmid-bacteria network, using plasmid-bacteria interaction data from
113 Stalder *et al.* (2019) that used Hi-C to link mobile genetic elements to their bacterial hosts in a

114 wastewater sample. Here we advance/expand on that study by using machine learning methods to
115 identify clusters of plasmid contigs that original from the same cell (hereafter termed ‘putative
116 plasmids’), and test whether AMR presence on these putative plasmids is associated with altered
117 interaction distributions with 374 bacterial metagenome-assembled genomes (MAGs).

118 While plasmids can carry a range of host-beneficial genes, we make the assumption that
119 putative plasmids that carry AMR genes will be on average more beneficial than those that don’t,
120 because the concentrations of antibiotics in wastewater are likely to confer an advantage to at least
121 some AMR genes (Karkman *et al.* 2018). In this context, we predict that plasmids that carry AMR
122 should interact with more bacterial hosts (i.e., have higher ‘prevalence’; Fig. 1D, see Fig. 1E for
123 glossary of terms) than plasmids without AMR, and thereby lead to more connected and generalist
124 networks. Because evolutionary arms races in antagonistic networks tend to generate strong
125 phylogenetic signal (Rohr & Bascompte 2014), we would also expect that AMR-carrying plasmids to
126 demonstrate weaker phylogenetic signal than those without AMR genes (Fig. 1D). We further discuss
127 the potential of network structures in promoting the spread of AMR.

128

129 **Results**

130 The inferred wastewater bacteria-plasmid network was made up of 374 bacterial MAGs
131 (metagenome-assembled genomes) and 289 putative plasmids. Most MAGs identified belonged to
132 either the class Betaproteobacteria (Phylum Proteobacteria), Gammaproteobacteria (Phylum
133 Proteobacteria), Clostridia (Phylum Firmicutes), Bacteroidia (Phylum Bacteroidetes) or Actinobacteria
134 (Phylum Actinobacteria) (Fig. 2A). Bacterial MAGs associated with an average of 3.5 putative
135 plasmids (median = 1, min. = 0, max. = 47), whilst putative plasmids associated with an average of
136 4.5 MAGs (median = 2, min. = 1, max. = 80). Note that this does not equate to one bacterial cell
137 having 47 plasmids; rather, 47 putative plasmids were found to be associated with that MAG across
138 its entire population within the wastewater community. While MAG relate to a metagenomic bin that
139 can be asserted to be a close representation to an actual individual genome, here the quality of most
140 of the MAG identified (according to genome size and completeness scores computed by CheckM)
141 does not allow us to do this assumption and rather consider the MAGs not as one bacterial cell’s
142 genome, but the genome (or fragment of a genome) of closely related strains within a species.

143

144 MAGs that associated with a high number of plasmids were distributed across the phylogenetic tree,
145 although MAGs belonging to Betaproteobacteria, and Gammaproteobacteria tended to associate with
146 a higher number of plasmids (Fig. 2A). Putative plasmids that were associated with AMR genes
147 tended to be more widely distributed across MAGs than those that were not associated with AMR
148 genes (mean AMR = 14, mean no AMR = 3; Wilcoxon W = 1379, $p < 0.0001$; Fig. 2B).

149 The full network clustered strongly by bacterial taxonomy, with MAGs belonging to Proteobacteria,
150 Firmicutes, and Bacteroidetes largely clustering separately (Fig. 2C). The large majority of AMR
151 plasmids clustered together with Proteobacteria, with classes Betaproteobacteria and
152 Gammaproteobacteria clustering together and sharing most of the AMR plasmids (Fig. 2D).

153

154 We next investigated the role of AMR genes on network structure by comparing sub-networks based
155 on whether putative plasmids associated with AMR or not (Fig. 3). When AMR putative plasmids were
156 excluded (Fig. 3A), networks were more modular (higher number of not linked sub networks) and less
157 nested (see Figure 1E) than networks based on solely AMR putative plasmids (Fig. 3B). Putative
158 plasmids carrying AMR genes further connected Proteobacteria to other phyla linking large parts of
159 the whole network.

160

161 We next visualized the distribution of the 15 most prevalent putative plasmids (super generalists)
162 across the bacterial phylogenetic tree (Fig. 4). The most prevalent putative plasmids were largely
163 shared amongst members of the same phyla, although some were occasionally shared more widely.
164 Putative plasmids that were associated with Proteobacteria were often shared across both Beta- and
165 Gamma-proteobacteria (Fig. 4). Whilst most putative plasmids remain undescribed, plasmid 3 (Fig. 4)
166 was identified as a broad range plasmid belonging to the IncP-β group.

167

168 Lastly, we estimated phylogenetic signal in these super generalists and tested whether AMR
169 presence was associated with phylogenetic signal strength. Phylogenetic signal quantifies the
170 relationship between interaction strength and MAG phylogenetic distance, with high phylogenetic
171 signal indicating that MAGs that are closely related are very likely to have a similar number of
172 interactions with a particular plasmid. Importantly, two main methods are available to quantify
173 phylogenetic signal: the first simply measures autocorrelation between trait differences and

174 phylogenetic distance (measured by Abouheif's C_{mean}); the second applies more complex evolutionary
175 models to test whether distributions match what would be expected if traits coevolved measured by
176 Pagel's λ (Münkemüller *et al.* 2012). We found that most putative plasmids demonstrated significant
177 phylogenetic signal when measured by both methods (Fig. 5). Contrary to our predictions, putative
178 plasmids carrying AMR had significantly stronger phylogenetic signal than those without when
179 measured by autocorrelation (Fig. 5A) and by an evolutionary model (Fig. 5B), although estimates
180 may be biased by AMR putative plasmids being generally more prevalent than those without AMR
181 (Fig. 5).

182

183 **DISCUSSION**

184 The wastewater network analysis demonstrates that while the natural plasmid-host
185 community is dominated by specialist putative plasmids, those carrying AMR genes tend to be more
186 generalist and markedly increase the connectivity of the network. As predicted, the network structure
187 for the AMR plasmid-host subnetwork differed substantially from the non-AMR plasmid network. The
188 AMR plasmid – host network showed a high degree of generalism and nestedness, with an overall
189 high level of connectedness. Extrapolations from ecological network theory (Thebault & Fontaine
190 2010) and experimental data (Heß *et al.* 2021; Newbury *et al.* in press) suggest that this pattern can
191 be explained by the two different types of plasmids in this network: (1) Costly, more specialised
192 plasmids and (2) beneficial, more generalist plasmids. It is reasonable to think that AMR genes can
193 be directly beneficial in the wastewater environment because they can confer a selective advantage
194 even in the presence of low concentrations of antibiotics and other biocides (Murray *et al.* 2018). Even
195 if not directly beneficial in the current wastewater environment experienced by these organisms, AMR
196 genes will almost certainly have provided a benefit in the environments they originate from, such as
197 hospitals and a community of people consuming various antibiotics. This may have led to interactions
198 that evolved to be more generalised (Guimarães Jr *et al.* 2011; Nuismer *et al.* 2013).).

199

200 It is not known precisely why mutualistic networks (e.g., seed dispersal, pollination, symbiosis) often
201 show greater generalism than antagonistic networks (e.g., herbivory and parasitism), with mechanism
202 largely inferred from theory alone. It may be purely short-term ecological consequences, with the
203 greater fitness of mutualistic partners leading to the subsequent spread of mutualists to new species.

204 More generalism may also lead to greater stability in communities dominated by mutualistic
205 interaction (i.e., species are less likely to go extinct, resulting in changes in network properties), while
206 generalism decreases stability of antagonistic communities (Thébault & Fontaine 2010). Theory also
207 suggests that coevolution may drive this pattern under the assumption that trait matching (e.g.,
208 attack-defence traits) determines the strength of antagonistic interactions while trait differences (e.g.,
209 barriers for transmission) determine mutualistic interactions (Nuismer *et al.* 2013; de Andreazzi *et al.*
210 2020). While it is impossible to deduce mechanism from our correlational study, recent experiments
211 and theory work using simplified bacteria-plasmid networks demonstrates that short-term growth rate
212 advantages conferred by a beneficial plasmid can result in greater plasmid ecological generalism
213 (Newbury *et al.* in press). Specifically, if a plasmid increases the frequency of its host, the plasmid
214 then has greater opportunities to be transmitted to other host taxa. While the distribution of hosts of a
215 plasmid can be influenced by factors affecting its ability to transfer into a new host, after entry it is
216 primarily the plasmid-encoded replication system and its interaction with host factors that determines
217 the ability of a plasmid to survive in that host (del Solar *et al.* 1996; Toukdarian 2004). This suggests
218 that increased ecological generalism of beneficial plasmids could in turn promote greater evolutionary
219 generalism as a consequence of mutualistic coevolution for plasmid maintenance occurring between
220 plasmids and the multiple hosts they interact with (Harrison & Brockhurst 2012).

221
222 We predicted that plasmids with AMR would show lower phylogenetic signal than those without,
223 because mutualists tend to evolve weaker phylogenetic signal than antagonists (Rohr & Bascompte
224 2014). Yet, we found the opposite: plasmids with AMR genes showed higher phylogenetic signal
225 when measured by both autocorrelation and evolutionary models. This finding may either reflect that
226 AMR plasmids can still be parasitic in some contexts, and therefore require hosts to carefully control
227 plasmid entry and maintenance in a way that might evolutionarily constrain interactions. Another
228 possibility is that AMR plasmids are indeed generally beneficial, yet if most plasmids are parasitic this
229 would still promote the evolution of specific interactions between bacterial and mutualists that act to
230 constrain interactions within phylogenies (Thrall *et al.* 2007). In general, whilst most putative plasmids
231 were highly specialist, super generalist putative plasmids were still largely shared within phyla, and
232 only rarely interacted with bacterial hosts outside of the dominant host phylum. This pattern was
233 irrespective of whether they carried AMR genes or not. This indicates strong barriers to plasmid

234 transmission between phyla (Redondo-Salvo *et al.* 2020), yet not between different classes within
235 phyla. For example, Gamma- and Proteo-bacteria appeared to freely share putative plasmids and did
236 not form separate clusters within networks.

237

238 While greater generalism associated with AMR plasmids obviously has important implications for the
239 spread of the specific AMR genes encoded by the plasmids, it is also likely to affect the spread of
240 additional AMR genes, even those not currently under selection. First, a new AMR gene that gets
241 incorporated into a generalist plasmid will have more chance to spread. Second, generalist plasmids,
242 are more likely to acquire additional AMR genes (e.g., by transposition), given the greater diversity of
243 hosts they interact with. Generalist plasmids and in particular generalist plasmids with AMR were
244 mostly associated with Proteobacteria, although plasmid hubs were found across bacterial classes.
245 Generalist AMR plasmids assumed a central role in the overall network by linking Proteobacteria to
246 other classes, although these interactions were relatively rare, and these generalist plasmids may
247 contribute to the spread of AMR gene transfer in general between different classes of bacteria. This
248 might reflect greater selection for AMR in Proteobacteria because many common human pathogens
249 are found in the Proteobacteria.

250

251 Our approach advances on the analysis from Stalder *et al.* (2019) by utilizing novel machine learning
252 methods to identify undescribed plasmid signatures. Whilst this method considerably increases our
253 understanding of how undescribed plasmids contribute to interaction network structure, we assume
254 that connected clusters of sequences represent one plasmid. Yet, it is possible that these clusters in
255 fact represent multiple co-occurring plasmids, or, conversely, that some sequences treated as
256 separate plasmids are in fact part of the same plasmid. Nevertheless, our sensitivity analyses suggest
257 that our results and interpretations are robust to changes to methodology. An additional limitation to
258 our approach is that some shared genes or mobile elements between different plasmids could have
259 amplified the connections of generalist putative plasmids to more hosts. We strived to remove any
260 such genomic elements, such as transposons, AMR, metal resistance, biocide resistance and
261 virulence, yet it is possible that at least some super generalist putative plasmids may be a product of
262 other plasmid accessory genes commonly shared among different plasmids of this bacterial
263 community. Future advances in Hi-C technology paired with long-read sequencing methods will

264 further our ability to distinguish and describe plasmids in natural communities using high throughput
265 sequencing technology. Lastly, this is a correlational study and AMR presence and generalism may
266 also be driven by host taxa. Indeed, proteobacteria are a very ecologically diverse phylum (Woese
267 1987), so they may be more likely to be associated with promiscuous plasmids carrying genes
268 beneficial in a range of hostile environments. Therefore, to fully understand the ecological and
269 evolutionary dynamics under varying plasmid-host interaction type we need experimental approaches
270 that measure fitness consequences and link those to changes in observed network structures.

271

272 By conducting ecological network analyses on a wastewater Hi-C metagenome, we have been able to
273 describe a natural plasmid-host network. The patterns we observe are consistent with theory. First,
274 networks are primarily driven by specialism, consistent with a predominantly parasitic impact of
275 plasmids in the absence of carriage of beneficial accessory genes. Second, greater prevalence of
276 AMR genes – which are often transferred by plasmids – in generalist and abundant plasmids lead to a
277 more connected network. Third, this offers a large potential for sharing of a few generalist plasmids
278 across the network, promoting inter-class HGT and indirect network interactions. Further work is
279 clearly required to determine the generality of our findings and the mechanisms underpinning them.
280 This includes other types of networks, such as bacteria-bacteriophage (Flores *et al.* 2013), where
281 interactions while primarily antagonistic can also be mutualistic (Harrison & Brockhurst 2017;
282 Wendling *et al.* 2021). A closer look at the types of plasmids that cause higher network
283 connectedness would also help understand the drivers of AMR spread in various environments.

284

285 **Materials and Methods**

286 Processing data

287 Generating bacterial MAG data

288 Hi-C metagenome data from Stalder *et al.* was assembled into MAGs using an updated algorithm of
289 ProxiMeta™ on April 4th 2021 (Phase Genomics, Inc. 2021). This generated 374 MAGs analyzed in
290 this study. We assigned MAG taxonomy by running MAGs through Phylophlan (Asnicar *et al.* 2020),
291 which calls MASH for taxonomic assignment, and used taxonomy as a proxy for phylogeny (Table
292 S1).

293

294 Contig filtering

295 The raw Hi-C data from one wastewater sample contains over 2.5 million contigs representing DNA
296 fragments from diverse organisms. Because we are interested specifically in links between bacteria
297 and plasmids, we initially filtered out contigs that were 1) identified as transposons, and 2) were not
298 identified as either a MAG, a plasmid, or carrying other AMR, virulence, metal and biocide genes.
299 Transposons and IS elements were identified by performing a homology search with BLASTp on
300 predicted gene from all contigs using an e-value < 0.01 against all known transposase proteins from
301 the databases from IS finder (Siguier *et al.* 2006) available from
302 <https://github.com/thanhleviet/ISfinder-sequences/blob/master/README.md> and from Tn3
303 Transposon Finder (Ross *et al.* 2021) available from
304 <https://tncentral.proteininformationresource.org/TnFinder.html>. Protein-coding genes were predicted
305 from all contigs using prodigal in metagenomic mode using the option ‘-p meta’ available from
306 <https://github.com/hyattpd/Prodigal> (Hyatt *et al.* 2010). Contigs with gene coding for antimicrobial
307 resistance, virulence factors, metal resistance or resistance to biocides were identified using AMR
308 finder plus (Feldgarden *et al.* 2019), using ‘-n’ and ‘—plus’ parameters.

309

310 Identifying putative plasmids

311 All remaining contigs were run through PlassClass (Pellow *et al.* 2020) using default parameters to
312 distinguish between contigs that were of chromosomal or plasmid origin. This method performs better
313 on short reads < 1000 bp than other machine learning algorithms such as Plasflow (Krawczyk *et al.*
314 2018). To be conservative, only contigs where 95% of sequence signatures were classified as
315 plasmid signatures. Only one of these was identified as a well-characterised broad-range plasmid
316 (IncP-β group). Contigs with gene coding for antimicrobial resistance, virulence factors, metal
317 resistance or resistance to biocides were not treated as belonging to plasmids because, like
318 transposons, they can be shared across multiple plasmids and therefore hinder the identification of
319 unique plasmid signatures. Because the contigs identified as plasmids were mostly constituted by
320 short contigs (the median length was 379 bp) and plasmids are likely to be > 1000 bp long, we
321 reasoned that for a contig to belong to a plasmid it should be found to be consistently connected to at
322 least one other. In order to account for this, we retained only plasmid contigs that were linked to other
323 plasmid contigs at least five times (n = 841 contigs). We then performed a cluster analysis on these

324 plasmid contigs using the Walktrap clustering algorithm using the *igraph::walktrap.community* function
325 and with a step length of 4 (Csardi & Nepusz 2006). This clustering step identified 331 plasmid
326 clusters which we treated as putative plasmids.

327

328 We conducted several quality checks to assess the reliability of these cluster of contigs we called
329 putative plasmids. We first checked the total length of plasmid contigs. The average total length
330 (4,200 bp) and the general distribution (median = 1,300 bp, min = 500 bp, max. = 78,700 bp) were
331 below typical plasmid length found in natural communities suggesting those clusters of plasmids
332 contigs were part of plasmids but we did not assemble complete plasmids (Dunivin *et al.* 2019). Of
333 these, 39 clusters were found not to associate with any MAGs and were excluded. In addition, we
334 used BLASTn (Megablast against the non-redundant nucleotide database) to manually check the
335 gene content of 132 contigs (out of 841) that were part of putative plasmid clusters that were
336 subsequently found to associate with over 10% of MAGs. Thirty-three contigs were removed at this
337 stage for having genes that could plausibly be associated with transposons or bacterial
338 chromosomes. Lastly, associations characterised by only one Hi-C link were considered unreliable
339 and removed. After this quality filtering, 289 putative plasmid clusters made up of 729 contigs were
340 retained for analysis (Fig. S1a). The remaining putative plasmids we classified as associating with an
341 AMR gene if at least one contig within the cluster was connected to an AMR contig at least twice (Fig.
342 S1b). Hi-C link counts were then normalised Hi-C by both MAG abundance and by putative plasmid
343 size.

344

345 Analysis

346 An adjacency matrix was generated from the processed Hi-C association data, and data was handled
347 using the packages phyloseq and igraph. Networks were visualized using the ggnetwork (Briatte
348 2020) using graphopt layout. Network statistics for the five major host classes were generated with
349 the *bipartite::networklevel* function (Dormann *et al.* 2009). Phylogenetic trees and their attributes were
350 visualized with the ggtree package (Yu *et al.* 2017). Phylogenetic signal was estimated using the
351 phylosignal::phylosignal function (Keck *et al.* 2016), applying two different measures: Abouheif's
352 C_{mean}, which calculates autocorrelation between phylogenetic distance and trait distributions, and
353 Pagel's λ , which uses a Brownian motion (BM) model of trait evolution. We chose these two metrics

354 because they perform the best out of a number of metrics available, and are also insensitive to branch
355 length (Münkemüller *et al.* 2012), therefore are appropriate to use on our phylogenetic tree based on
356 taxonomy. To test whether plasmids with and without AMR genes differ in their phylogenetic signal,
357 be performed a t-test.

358

359 The Rmarkdown report is available at <https://github.com/Riselya/Plasmid-networks>. Sequencing data
360 are available in FASTQ format at SRA accession PRJNA506462. Processed data and scripts for
361 linking contigs to genome clusters using Hi-C data are available at <https://osf.io/ezb8j/>.

362

363

364 **Acknowledgments**

365 We thank Mike Brockhurst for the discussion about the manuscript and Suzanne Kay for help with the
366 data analysis. This research is funded by NERC, grant number NE/S000771/1. BIS is supported by a
367 Royal Commission for the Exhibition of 1851 Research Fellowship. TS is in part supported by the
368 program Understanding Antimicrobial Resistance grant no 2018-67017-27630 / project accession no
369 1015171 from the USDA National Institute of Food and Agriculture.

370

371 **Author Contributions**

372 DS, AB and AR conceived and designed the study. AR, TS, and BIS analyzed the data. DS and AR
373 wrote the first manuscript draft and all authors contributed.

374

375 **Competing Interest Statement**

376 The authors have no competing interests.

377

378 **References**

- 379 1.
- 380 Acman, M., van Dorp, L., Santini, J.M. & Balloux, F. (2020). Large-scale network analysis captures
381 biological features of bacterial plasmids. *Nature Communications*, 11, 2452.
- 382 2.
- 383 Asnicar, F., Thomas, A.M., Beghini, F., Mengoni, C., Manara, S., Manghi, P. *et al.* (2020). Precise
384 phylogenetic analysis of microbial isolates and genomes from metagenomes using
385 PhyloPhiAn 3.0. *Nature Communications*, 11, 2500.
- 386 3.
- 387 Bennett, P.M. (2008). Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic
388 resistance genes in bacteria. *British Journal of Pharmacology*, 153, S347-S357.

389 4.

390 Briatte, F. (2020). *ggnetwork: geometries to plot networks with 'ggplot2'*. R package version 0.5. 8.

391 5.

392 Coyte, K.Z., Schluter, J. & Foster, K.R. (2015). The ecology of the microbiome: Networks,

393 competition, and stability. *Science*, 350, 663-666.

394 6.

395 Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research.

396 *InterJournal, complex systems*, 1695, 1-9.

397 7.

398 Dang, B., Mao, D., Xu, Y. & Luo, Y. (2017). Conjugative multi-resistant plasmids in Haihe River and

399 their impacts on the abundance and spatial distribution of antibiotic resistance genes. *Water*

400 *Research*, 111, 81-91.

401 8.

402 de Andreazzi, C.S., Astegiano, J. & Guimarães Jr, P.R. (2020). Coevolution by different functional

403 mechanisms modulates the structure and dynamics of antagonistic and mutualistic networks.

404 *Oikos*, 129, 224-237.

405 9.

406 del Solar, G., Alonso, J.C., Espinosa, M. & Díaz-Orejas, R. (1996). Broad-host-range plasmid

407 replication: an open question. *Molecular microbiology*, 21, 661-666.

408 10.

409 Dimitriu, T., Matthews, A.C. & Buckling, A. (2021). Increased copy number couples the evolution of

410 plasmid horizontal transmission and plasmid-encoded antibiotic resistance. *Proceedings of*

411 *the National Academy of Sciences*, 118, e2107818118.

412 11.

413 Dormann, C.F., Fründ, J., Blüthgen, N. & Gruber, B. (2009). Indices, graphs and null models:

414 analyzing bipartite ecological networks. *The Open Ecology Journal*, 2.

415 12.

416 Dunivin, T.K., Choi, J., Howe, A., Shade, A. & Sharpton, T.J. (2019). RefSoil+: a Reference Database

417 for Genes and Traits of Soil Plasmids. *mSystems*, 4, e00349-00318.

418 13.

419 Feldgarden, M., Brover, V., Haft, D.H., Prasad, A.B., Slotta, D.J., Tolstoy, I. et al. (2019). Validating

420 the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance

421 Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrobial Agents and*

422 *Chemotherapy*, 63, e00483-00419.

423 14.

424 Flores, C.O., Valverde, S. & Weitz, J.S. (2013). Multi-scale structure and geographic drivers of cross-

425 infection within marine bacteria and phages. *The ISME Journal*, 7, 520-532.

426 15.

427 Fontaine, C., Thébault, E. & Dajoz, I. (2009). Are insect pollinators more generalist than insect

428 herbivores? *Proceedings of the Royal Society B: Biological Sciences*, 276, 3027-3033.

429 16.

430 González-Salazar, C. & Stephens, C.R. (2012). Constructing Ecological Networks: A Tool to Infer

431 Risk of Transmission and Dispersal of Leishmaniasis. *Zoonoses and Public Health*, 59, 179-

432 193.

433 17.

434 Guimarães Jr, P.R., Jordano, P. & Thompson, J.N. (2011). Evolution and coevolution in mutualistic

435 networks. *Ecology Letters*, 14, 877-885.

436 18.

437 Guimarães, P.R., Pires, M.M., Jordano, P., Bascompte, J. & Thompson, J.N. (2017). Indirect effects

438 drive coevolution in mutualistic networks. *Nature*, 550, 511-514.

439 19.

440 Harrison, E. & Brockhurst, M.A. (2012). Plasmid-mediated horizontal gene transfer is a coevolutionary

441 process. *Trends in Microbiology*, 20, 262-267.

442 20.

443 Harrison, E. & Brockhurst, M.A. (2017). Ecological and Evolutionary Benefits of Temperate Phage:

444 What Does or Doesn't Kill You Makes You Stronger. *BioEssays*, 39, 1700112.

445 21.

446 Heß, S., Kneis, D., Virta, M. & Hiltunen, T. (2021). The spread of the plasmid RP4 in a synthetic

447 bacterial community is dependent on the particular donor strain. *FEMS Microbiology Ecology*,

448 97.

449 22.

450 Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W. & Hauser, L.J. (2010). Prodigal: 451 prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 452 11, 119.

453 23.

454 Jensen, R.B. & Gerdes, K. (1995). Programmed cell death in bacteria: proteic plasmid stabilization 455 systems. *Molecular Microbiology*, 17, 205-210.

456 24.

457 Kaiser-Bunbury, C.N., Mougal, J., Whittington, A.E., Valentin, T., Gabriel, R., Olesen, J.M. et al. 458 (2017). Ecosystem restoration strengthens pollination network resilience and function. *Nature*, 459 542, 223-227.

460 25.

461 Karkman, A., Do, T.T., Walsh, F. & Virta, M.P. (2018). Antibiotic-resistance genes in waste water. 462 *Trends in microbiology*, 26, 220-228.

463 26.

464 Keck, F., Rimet, F., Bouchez, A. & Franc, A. (2016). phylosignal: an R package to measure, test, and 465 explore the phylogenetic signal. *Ecology and Evolution*, 6, 2774-2780.

466 27.

467 Kent, A.G., Vill, A.C., Shi, Q., Satlin, M.J. & Brito, I.L. (2020). Widespread transfer of mobile antibiotic 468 resistance genes within individual gut microbiomes revealed through bacterial Hi-C. *Nature 469 Communications*, 11, 4379.

470 28.

471 Klümper, U., Riber, L., Dechesne, A., Sannazzaro, A., Hansen, L.H., Sørensen, S.J. et al. (2015). 472 Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial 473 community. *The ISME Journal*, 9, 934-945.

474 29.

475 Krawczyk, P.S., Lipinski, L. & Dziembowski, A. (2018). PlasFlow: predicting plasmid sequences in 476 metagenomic data using genome signatures. *Nucleic Acids Research*, 46, e35-e35.

477 30.

478 Lili, L.N., Britton, N.F. & Feil, E.J. (2007). The Persistence of Parasitic Plasmids. *Genetics*, 177, 399- 479 405.

480 31.

481 Martínez, J.L. (2008). Antibiotics and Antibiotic Resistance Genes in Natural Environments. *Science*, 482 321, 365-367.

483 32.

484 Montesinos-Navarro, A., Hiraldo, F., Tella, J.L. & Blanco, G. (2017). Network structure embracing 485 mutualism–antagonism continuums increases community robustness. *Nature Ecology & 486 Evolution*, 1, 1661-1669.

487 33.

488 Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K. et al. (2012). How to 489 measure and test phylogenetic signal. *Methods in Ecology and Evolution*, 3, 743-756.

490 34.

491 Murray, A.K., Zhang, L., Yin, X., Zhang, T., Buckling, A., Snape, J. et al. (2018). Novel Insights into 492 Selection for Antibiotic Resistance in Complex Microbial Communities. *mBio*, 9, e00969- 493 00918.

494 35.

495 Newbury, A., Dawson, B., Klümper, U., Hesse, E., Castledine, M., Fontaine, F. et al. (in press). 496 Fitness effects of plasmids shape the structure of bacteria-plasmid interaction networks. 497 *Proceedings of the National Academy of Sciences*.

498 36.

499 Nuismer, S.L., Jordano, P. & Bascompte, J. (2013). Coevolution and the architecture of mutualistic 500 networks *Evolution*, 67, 338-354.

501 37.

502 Pellow, D., Mizrahi, I. & Shamir, R. (2020). PlasClass improves plasmid sequence classification. 503 *PLoS computational biology*, 16, e1007781.

504 38.

505 Redondo-Salvo, S., Fernández-López, R., Ruiz, R., Vielva, L., de Toro, M., Rocha, E.P. et al. (2020). 506 Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. 507 *Nature communications*, 11, 1-13.

508 39.

509 Rohr, R.P. & Bascompte, J. (2014). Components of Phylogenetic Signal in Antagonistic and
510 Mutualistic Networks. *The American Naturalist*, 184, 556-564.

511 40.

512 Ross, K., Varani, A.M., Snesrud, E., Huang, H., Alvarenga, D.O., Zhang, J. *et al.* (2021). TnCentral: a
513 Prokaryotic Transposable Element Database and Web Portal for Transposon Analysis. *mBio*,
514 12, e02060-02021.

515 41.

516 San Millan, A. (2018). Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context.
517 *Trends in Microbiology*, 26, 978-985.

518 42.

519 Segar, S.T., Fayle, T.M., Srivastava, D.S., Lewinsohn, T.M., Lewis, O.T., Novotny, V. *et al.* (2020).
520 The role of evolution in shaping ecological networks. *Trends in Ecology & Evolution*, 35, 454-
521 466.

522 43.

523 Siguier, P., Pérochon, J., Lestrade, L., Mahillon, J. & Chandler, M. (2006). ISfinder: the reference
524 centre for bacterial insertion sequences. *Nucleic acids research*, 34, D32-D36.

525 44.

526 Stalder, T., Press, M.O., Sullivan, S., Liachko, I. & Top, E.M. (2019). Linking the resistome and
527 plasmidome to the microbiome. *The ISME Journal*, 13, 2437-2446.

528 45.

529 Suzuki, H., Yano, H., Brown, C.J. & Top, E.M. (2010). Predicting Plasmid Promiscuity Based on
530 Genomic Signature. *Journal of Bacteriology*, 192, 6045-6055.

531 46.

532 Thébault, E. & Fontaine, C. (2010). Stability of Ecological Communities and the Architecture of
533 Mutualistic and Trophic Networks. *Science*, 329, 853-856.

534 47.

535 Thrall, P.H., Hochberg, M.E., Burdon, J.J. & Bever, J.D. (2007). Coevolution of symbiotic mutualists
536 and parasites in a community context. *Trends in Ecology & Evolution*, 22, 120-126.

537 48.

538 Toukdarian, A. (2004). Plasmid strategies for broad-host-range replication in Gram-negative bacteria.
539 *Plasmid biology*, 257-270.

540 49.

541 Veron, S., Fontaine, C., Dubos, N., Clergeau, P. & Pavoine, S. (2018). Predicting the impacts of co-
542 extinctions on phylogenetic diversity in mutualistic networks. *Biological Conservation*, 219,
543 161-171.

544 50.

545 Wang, J., Gao, Y. & Zhao, F. (2016). Phage–bacteria interaction network in human oral microbiome.
546 *Environmental Microbiology*, 18, 2143-2158.

547 51.

548 Weitz, J.S., Poisot, T., Meyer, J.R., Flores, C.O., Valverde, S., Sullivan, M.B. *et al.* (2013). Phage–
549 bacteria infection networks. *Trends in Microbiology*, 21, 82-91.

550 52.

551 Wendling, C.C., Refardt, D. & Hall, A.R. (2021). Fitness benefits to bacteria of carrying prophages
552 and prophage-encoded antibiotic-resistance genes peak in different environments. *Evolution*,
553 75, 515-528.

554 53.

555 Woese, C.R. (1987). Bacterial evolution. *Microbiological reviews*, 51, 221-271.

556 54.

557 Yaffe, E. & Relman, D.A. (2020). Tracking microbial evolution in the human gut using Hi-C reveals
558 extensive horizontal gene transfer, persistence and adaptation. *Nature Microbiology*, 5, 343-
559 353.

560 55.

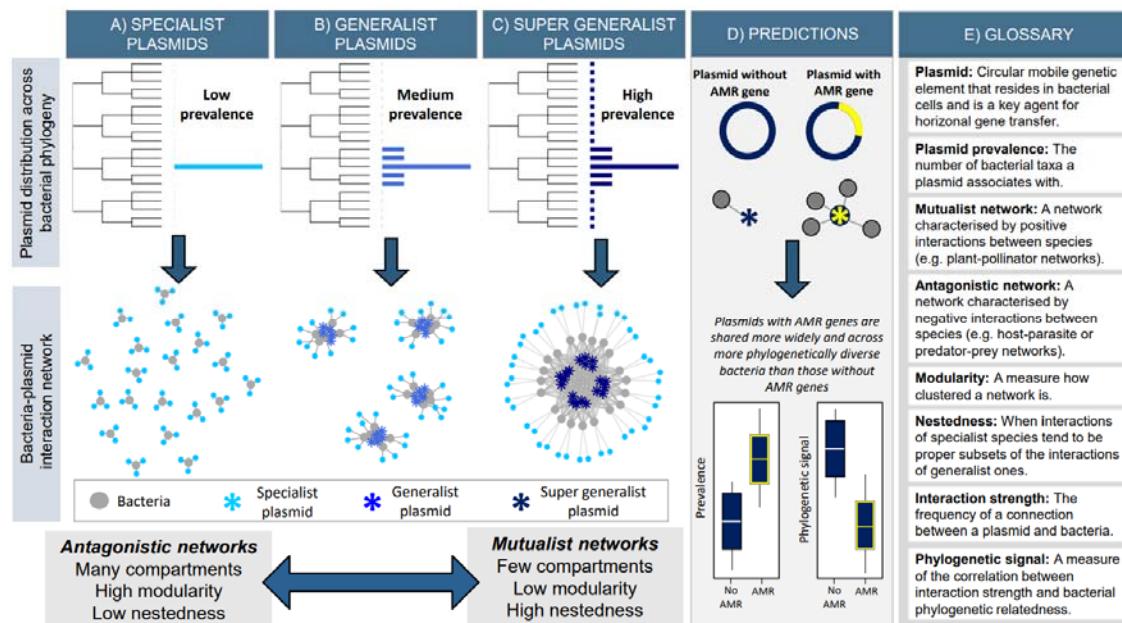
561 Yu, G., Smith, D.K., Zhu, H., Guan, Y. & Lam, T.T.Y. (2017). ggtree: an R package for visualization
562 and annotation of phylogenetic trees with their covariates and other associated data. *Methods
in Ecology and Evolution*, 8, 28-36.

563

564

565 **Figures**

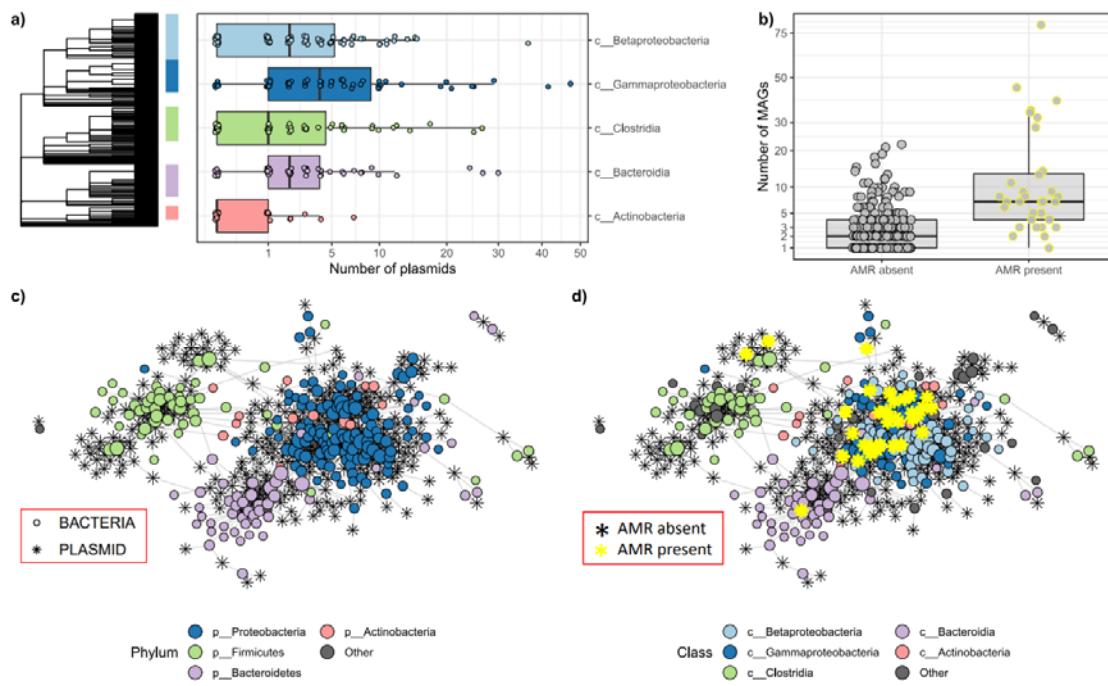
566



568

Figure 1. Predictions for plasmid-host network structure. A-C) Three different hypothetical network structures based on 20 bacterial taxa and 60 plasmids, assuming some level of phylogenetic signal in interaction strength. In each network, 40 out of the 60 plasmids are highly specialist (light blue) and interact with only one bacterial taxa. Network structure changes dramatically if the remaining twenty plasmids are A) also specialist; B) generalist but limited to specific clades (royal blue); or C) super generalist (navy blue) with weak interactions across other clades. Ecological theory suggests that antagonistic networks tend to be made up of high number of specialists that lead to networks with many separate compartments, high in modularity, and low in nestedness. On the contrary, mutualist networks are structured by generalists and form networks that have few compartments, and low in modularity, and high in nestedness. D) Because antimicrobial resistance (AMR) genes are likely to have beneficial effects on bacterial fitness within the context of a wastewater community, we predict that plasmids that carry AMR genes are likely to have higher prevalence and lower phylogenetic signal than plasmids without AMR genes. Because increases in prevalence of at least some plasmids lead to more connected networks (A-C), plasmids with AMR should lead to more connected networks. E) Glossary of terms used throughout the article.

583



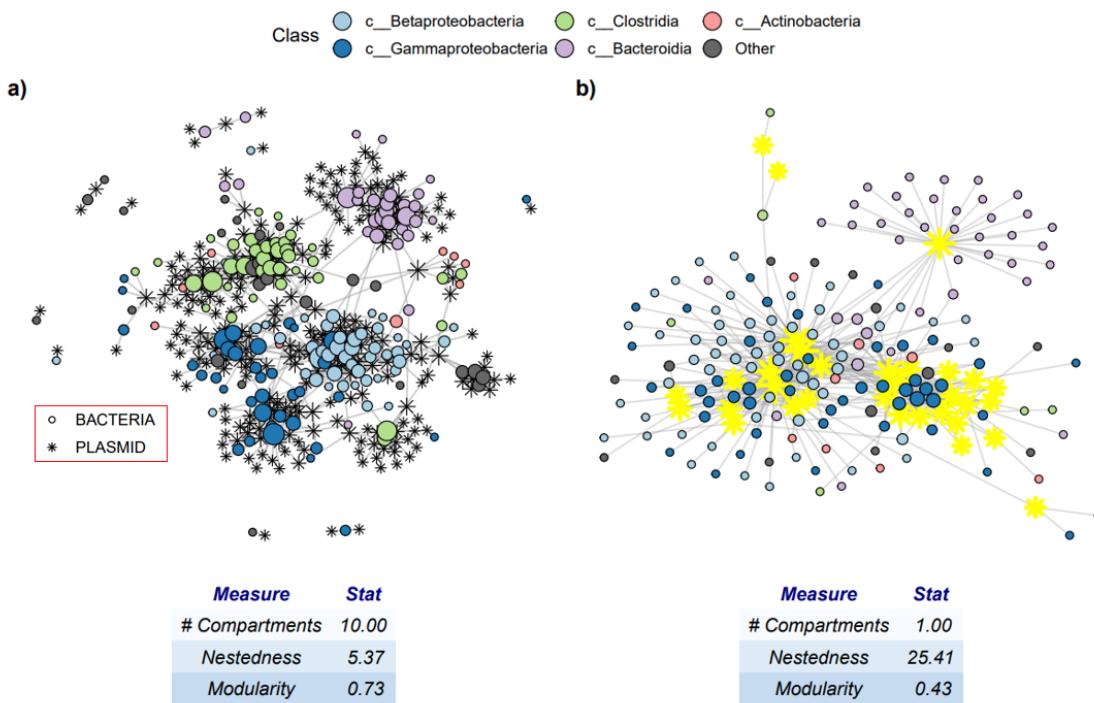
585

586 **Figure 2. MAG-putative plasmid networks based on normalised Hi-C linkage.** A) Distribution of
587 number of plasmids per MAG, split by bacterial class; B) Number of MAG associations per putative
588 plasmid, split by whether plasmids were associated with antimicrobial resistance genes (AMR); C)
589 The full MAG-putative plasmid network made up of 374 MAGs and 289 putative plasmids, with MAGs
590 coloured by phylum and plasmids represented by stars; D) Same network as C) but coloured by class
591 and the position of AMR plasmids highlighted in yellow. Nodes are sized by their network degree.

592

593

594



595

596 **Figure 3. AMR genes and network structure.** Sub-networks and network statistics representing
597 MAG Hi-C associations with A) plasmids without AMR and B) plasmid with AMR. Stars represent
598 plasmids and circles MAGs, with AMR plasmids highlighted in yellow. Nodes are sized by their
599 network degree.

600
601



602

603 **Figure 4. Phylogenetic distributions of the most prevalent plasmids across 374 MAGs.**

604 Phylogenetic weighted distributions of the 15 most prevalent putative plasmids (i.e. plasmids with
605 associations to the highest number of MAGs), ordered by how many MAGs they associate with. Bar
606 length represents interaction strength (i.e. the number of normalised Hi-C links). Putative plasmids
607 with AMR are highlighted yellow.

608

609

610

611

612

613

614

615

616

617

618

619

620

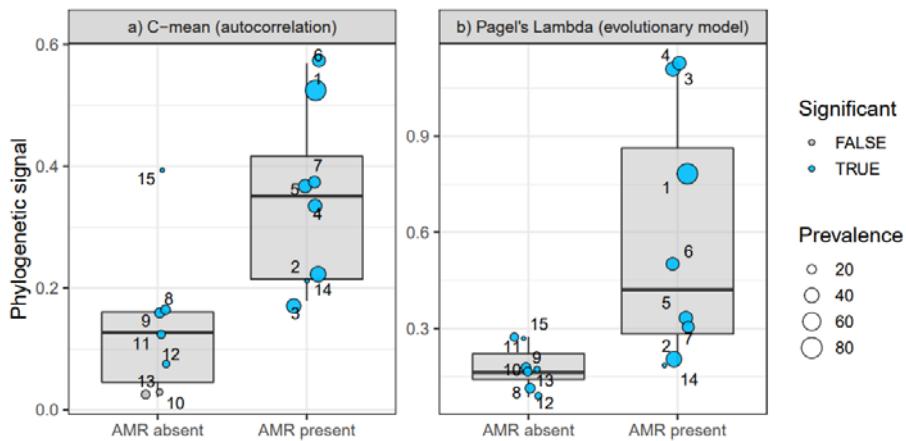
621

622

623

624

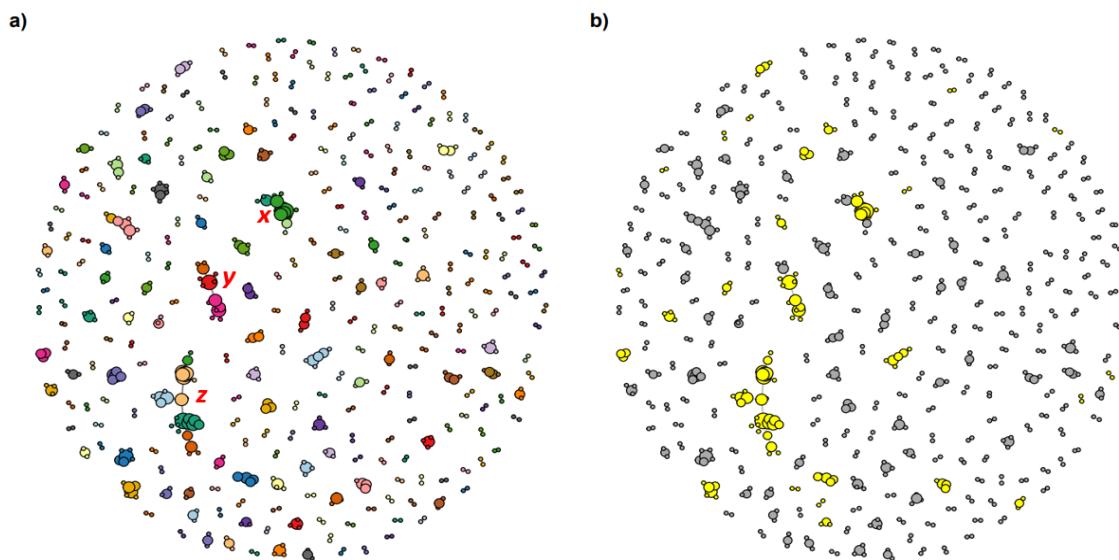
625



626
627
628
629
630
631
632
633
634

Figure 5. Phylogenetic signal and AMR presence. Phylogenetic signal for the 15 most prevalent putative plasmids separated by AMR presence, applying a) Abouheif's C_{mean} , which estimates autocorrelation between trait similarity and phylogenetic distance, and b) Pagel's λ , which applies a model of trait evolution. Points are labelled by the plasmid ID (corresponding to Fig. 4), sized by their prevalence, and coloured by whether estimates for phylogenetic signal are statistically significant.

635 **Supplementary material**



636

637 **Figure S1.** Putative plasmid clusters ($n = 289$ clusters built from 729 contigs) that were retained for
638 analysis. Each node represents a contig identified as having a 95% probability of belonging to a
639 plasmid by PlassClass, coloured by a) its cluster membership based on the walktrap method, and b)
640 whether it is associated with AMR (yellow). Only associations based on at least 5 Hi-C links are
641 included.

642