A wastewater plasmid-host network is dominated by specialist plasmids

Alice Risely[1], Benno I. Simmons[2], Angus Buckling[2,3], Dirk Sanders[2,3] *

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

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

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

*Corresponding author:

Dirk Sanders

Email: d.sanders@exeter.ac.uk

AR Orcid ID: 0000-0002-0731-2934

BIS Orcid ID: 0000-0002-2751-9430

DS Orcid ID: 0000-0003-2383-8693

Abstract. Plasmids are ubiquitous and important vectors for horizontal gene transfer; however, little is known about the structure of interaction networks between plasmids and their hosts in natural environments. Here we analyse a natural host-plasmid network extracted from wastewater samples. We found that plasmids were highly specific to their bacterial hosts, yet a small number of super generalists connected the entire network, allowing inter-class horizontal gene transfer and indirect interactions across broad taxonomic scales. Beta and Gamma proteobacteria exhibited more generalist interactions with plasmids, and this larger gene pool may explain the greater number of antimicrobial resistance genes associated with these classes.

Main

Plasmids - ubiquitous symbionts of bacteria - play a key role in the spread of antimicrobial resistance (AMR) and other genes, both within and between bacterial taxa[1, 2, 3, 4, 5]. While plasmids have been shown to vary in their host range [6], we know very little about the structure of natural plasmid-host interaction networks. The structure of such ecological networks[7, 8] gives insights into their function [9], how robust they are against perturbations[10, 11] as well as the likely spread of AMR and virulence genes [4].

Proximity-ligation methods such as Hi-C have been used to detect associations between DNA molecules originating in the same cell within microbial communities [12, 13]. These methods allow the construction of natural plasmid-host interaction networks. Here we analyse a dataset from Stalder *et al.*[12] that used Hi-C to link plasmids to their hosts in samples from wastewater. Because a relatively small number of plasmids have been fully characterised, and because plasmids are challenging to identify from metagenomic data, we apply a machine-learning method that distinguishes between chromosomal and plasmid sequences across assembled contigs[14]. Indirect interactions within ecological networks (mediated by at least a third species) are as important as direct interactions for ecological and evolutionary dynamics, and we employ a recently developed framework to quantify indirect interactions within bipartite ecological networks. Motifs[15] are subnetworks representing different patterns of interactions between small numbers of nodes. By calculating how often nodes occur in different positions within motifs, we gain information about functional similarities, i.e., the

number and structure of direct and indirect interactions. We further discuss the potential of network structures in promoting the spread of AMR.

The bipartite host-plasmid network included 191 bacterial hosts and 249 contigs identified as plasmids. Plasmids were more specialised than their hosts (Figure S1), with most plasmids linked to only one host species. As expected for a more specialist network, nestedness, measured as NODF [16], was relatively low (Table S1). Consistent with previous studies[6], the number of plasmids associated with a given host species was highest found among Beta and Gamma proteobacteria (Figure 1a, Table S1), with *Neisseria dentiae* and *Acidovorax soli* (Beta proteobacteria) having the largest number of plasmid associations in weighted and unweighted networks, respectively (Figure 1b,c). This means that connectance as a measure of the proportion of realized network links was also highest in the proteobacteria (Table S1), showing that plasmid and hosts were more widely shared. We found that Bacteroidea and Clostridia had more network compartments than the other classes (see Table1, Figure S2), again an indication of more specialist interactions. Crucially, Beta and Gamma proteobacteria had more generalised interactions with plasmids than the other bacteria classes (Figure S1) and had the highest prevalence of plasmid-associated antibiotic resistance genes (Figure 1a). This suggests that network properties may indeed facilitate the spread of AMR. While Beta and Gama proteobacteria were associated with the highest number of plasmids, plasmid hubs were found across bacterial classes, with *Anaerobium acetethylicum* (Clostridia) and *Parabacteroides chartae* (Bacteroidia) also being associated with particularly high numbers of plasmids. Host class subsections of the network were linked by low frequency plasmid associations (Figure 1c), with proteobacteria and a few generalist plasmids assuming a central role in the overall network (Figure 1d). These generalist plasmids may contribute to horizontal gene transfer (HGT) between different classes of bacteria.

Theory and empirical data suggests that networks dominated by antagonists have fewer generalist interactions than mutualistic networks[10]. Plasmids can act as parasites (antagonists) or mutualists depending on the accessory genes they carry and the environmental context[17]. The host specificity of most plasmids (Figure 1e, Figure S1) suggests that the majority of interactions are antagonistic. It is possible that the prevalence of AMR genes, which can confer a selective advantage at the low

concentrations of antibiotics and other biocides experienced by wastewater communities [18], may contribute to differences in mutualistic interactions between classes. Our results show similarities to bacteria-bacteriophage networks, in which bacteriophages are specialised on host bacteria at broader taxonomic scale, with varying structural patterns at different scales [19]. Similar to the plasmids our data, bacteria-bacteriophages are antagonists in general but can also be mutualistic in some contexts[20].

We next determined the structure of indirect interactions, and functional similarity of host and plasmids in general, within the network. We calculated network motif node positions for every bacterial host and plasmid and classified them via k-means clustering into groups with similar motif signatures (Figure 2a, b, ellipses denote clusters). Indirect interactions between bacteria are mediated to a large extent by the degree of generalism of plasmids, and we found extensive variation. Motif signatures for bacterial hosts were not strongly clustered by taxonomy (Figure 2a), although we found some clustering for Beta and Gamma proteobacteria, which typically showed relatively large numbers of indirect interactions (Figure 2a). The presence of indirect interaction throughout the network suggests the operation of apparent competition[21] or apparent mutualism[22], both of which can affect coexistence, operating between taxa occupying very different ecological niches. We also find considerable variation in plasmid motifs (Figure 2b), which is largely mediated by the extent multiple plasmids can infect a given bacteria.

By conducting ecological network analyses on a wastewater Hi-C metagenome, we have been able to describe a natural plasmid-host network. The patterns we observe are consistent with theory. First, networks are primarily driven by specialism, fitting with the parasitic impact of plasmids when they are not beneficial to their hosts through ecological beneficial functions, such as access to nutrients or protection in the presence of high concentrations of antimicrobials or other biocides. Second, greater prevalence of AMR genes – which are often transferred by plasmids – in bacterial classes where bacterial taxa interact with the greatest number of plasmids. Third, there was a large potential for sharing of a few generalist plasmids across the network, promoting inter-class HGT and apparent competition and mutualism.

Materials and Methods

Processing data

Metagenome assembled genomes (MAGs) from the Stalder et al. [12] wastewater sample dataset (WW without *E. coli* addition) with over 50% completeness (n = 191) were included in this study. These 191 taxa were estimated to make up approximately 16% of the total abundance of the microbial community. As common for environmental samples, the most abundant bacteria were not well-characterised and had extremely low completeness. MAGs were run through Plasflow [14] to distinguish between contigs that were of chromosomal or plasmid origin with 97% certainty. This identified 2770 contigs as belonging to plasmids. Only one of these was identified as a well-characterised broad-range plasmid (IncP(beta)). Once chromosomal and plasmid contigs were identified, we filtered for Hi-C associations that were between these specific chromosomal and plasmid contigs. Associations between two chromosomal contigs and two plasmid contigs were removed. To ensure associations were robust, we only included plasmid contigs that were represented by at least 50 associations across the whole dataset (n = 249). Varying this threshold did not alter results or interpretations of this study but allowed us to present the networks graphically. We assigned MAG taxonomy and generated a MAG phylogenetic tree by running MAGS through Phylophlan[23], which calls MASH for taxonomic assignment. To predict antibiotic resistance genes hosted by plasmids, we ran plasmid contigs through the Resistance Gene Identified software hosted by the Comprehensive Antibiotic Resistance Database (CARD [24]) applying the 'loose' setting. The RGI tool predicts resistome genes based on homology and SNP models.

Analysis

An adjacency matrix was generated from the processed Hi-C association data. Bipartite networks were visualised using *BipartiteD3* package in R[25]. Network statistics for the five major host classes were generated with the *networklevel* function from *bipartite* package [26]. Phylogenenetic trees and their attributes were visualised with the *ggtree* package [27].

To understand the importance of indirect interactions within the network – either plasmid-mediated for hosts or host-mediated for plasmids – we used bipartite network motifs. These motifs describe all possible ways of arranging links between two- to six-nodes, subject to the constraint that all nodes

have at least one link, resulting in 44 different possible motifs. Nodes can occupy different, topologically distinct positions within motifs[15]. There are 148 unique node position across all two- to six-node motifs. We calculated the number of times each host and plasmid occurred in each of these motif positions using the R package bmotif [28]. This gives a high-dimensional 'signature' of a node's indirect interactions in a network.

We separately clustered hosts and plasmids by their motif node position signatures with k-means cluster analyses and allowed for a six-cluster solution after visualising the extent to which each additional cluster added to the total sum of squares. To visualize clusters, we ordinated the node position matrix using Non-Metric Multi-Dimensional Scaling (NMDS) with two axes and using Jaccard distance (stress = 0.13 and 0.05 for plasmid and host ordinations, respectively). We further calculated the number of indirect associations with other plasmids/hosts within each motif and coloured the arrows accordingly ranging from 0 to a maximum of 4. We then added arrows to the NMDS plot that represent node positions particularly influential in the ordination (for $r > 0.3$).

Acknowledgments

We thank Mike Brockhurst for the discussion about the manuscript and Suzanne Kay for help with the data analysis. This research is funded by NERC, grant number NE/S000771/1. BIS is supported by a Royal Commission for the Exhibition of 1851 Research Fellowship.

Author Contributions

DS, AB and AR conceived and designed the study. AR and BIS analysed the data. DS wrote the first manuscript draft and all authors contributed.

Competing Interest Statement

The authors have no competing interests.

References

1. Bennett, P.M. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *British Journal of Pharmacology* **153**, S347-S357 (2008).

2. Dang, B., Mao, D., Xu, Y. & Luo, Y. Conjugative multi-resistant plasmids in Haihe River and their impacts on the abundance and spatial distribution of antibiotic resistance genes. *Water Research* **111**, 81-91 (2017).

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

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

5. Acman, M., van Dorp, L., Santini, J.M. & Balloux, F. Large-scale network analysis captures biological features of bacterial plasmids. *Nature Communications* **11**, 2452 (2020).

6. Klümper, U. *et al.* Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. *The ISME Journal* **9**, 934-945 (2015).

7. Montoya, J.M., Pimm, S.L. & Solé, R.V. Ecological networks and their fragility. *Nature* **442**, 259-264 (2006).

8. Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nature Reviews Microbiology* **10**, 538-550 (2012).

9. Kaiser-Bunbury, C.N. *et al.* Ecosystem restoration strengthens pollination network resilience and function. *Nature* **542**, 223-227 (2017).

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

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

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

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

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

15. Simmons, B.I. *et al.* Motifs in bipartite ecological networks: uncovering indirect interactions. *Oikos* **128**, 154-170 (2019).

16. Almeida-Neto, M., Guimarães, P., Guimarães Jr, P.R., Loyola, R.D. & Ulrich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**, 1227-1239 (2008).

17. Harrison, E. & Brockhurst, M.A. Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in Microbiology* **20**, 262-267 (2012).

18. Murray, A.K. *et al.* Novel Insights into Selection for Antibiotic Resistance in Complex Microbial Communities. *mBio* **9**, e00969-00918 (2018).

19. Flores, C.O., Valverde, S. & Weitz, J.S. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *The ISME Journal* **7**, 520-532 (2013).

20. Harrison, E. & Brockhurst, M.A. Ecological and Evolutionary Benefits of Temperate Phage: What Does or Doesn't Kill You Makes You Stronger. *BioEssays* **39**, 1700112 (2017).

21. Holt, R.D. & Bonsall, M.B. Apparent Competition. *Annual Review of Ecology, Evolution, and Systematics* **48**, 447-471 (2017).

22. Vandermeer, J. Indirect mutualism: variations on a theme by Stephen Levine. *The American Naturalist* **116**, 441-448 (1980).

23. Asnicar, F. *et al.* Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nature Communications* **11**, 2500 (2020).

24. Alcock, B.P. *et al.* CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research* **48**, D517-D525 (2019).

25. Terry, C. bipartiteD3: Interactive bipartite graphs. R package version 0.1. 0. *Website https://CRAN. R-project. org/package= bipartiteD3* (2018).

26. Dormann, C.F., Fründ, J., Blüthgen, N. & Gruber, B. Indices, graphs and null models: analyzing bipartite ecological networks. *The Open Ecology Journal* **2** (2009).

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

28. Simmons, B.I. *et al.* bmotif: A package for motif analyses of bipartite networks. *Methods in Ecology and Evolution* **10**, 695-701 (2019).

Figures

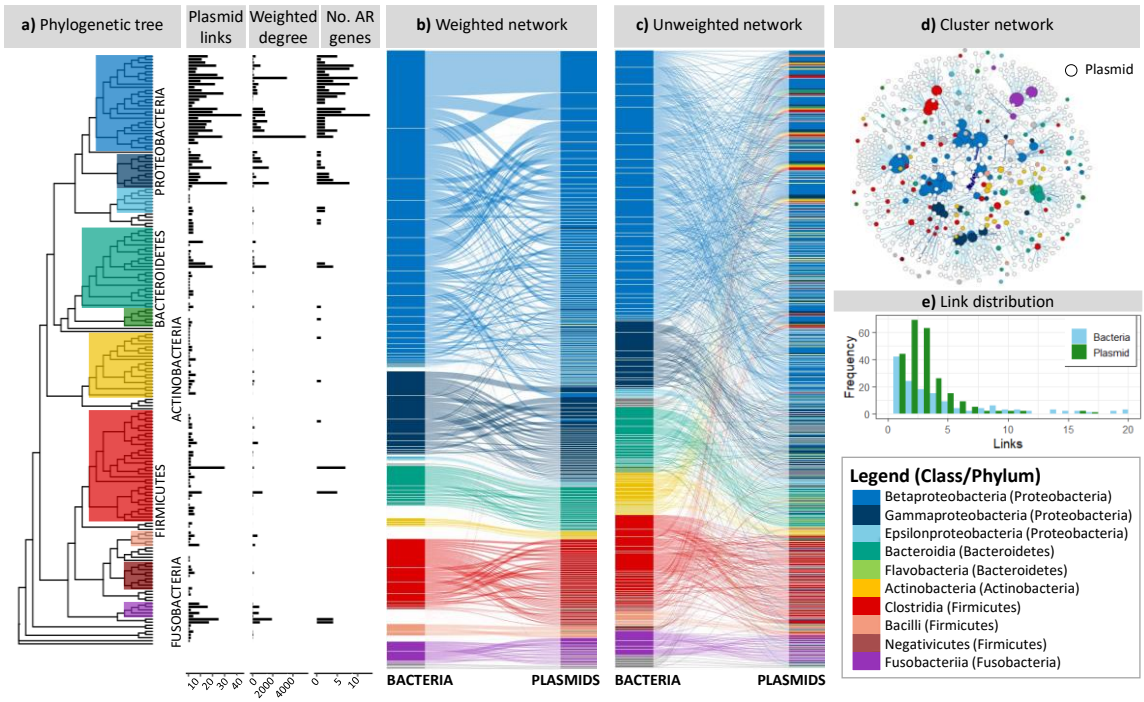


Figure 1. Plasmid-host network. **(**a) Phylogenetic tree for single hosts structured by host classes, with number of plasmid links per host, weighted degree and the number antimicrobial resistance genes for single hosts. (b) Weighted bipartite network with plasmid association according to the frequency of the interaction. Unweighted bipartite (c) and cluster (d) network with plasmid association according to presence- absence of network links. Link distribution (e) for hosts (blue) and plasmids (green).

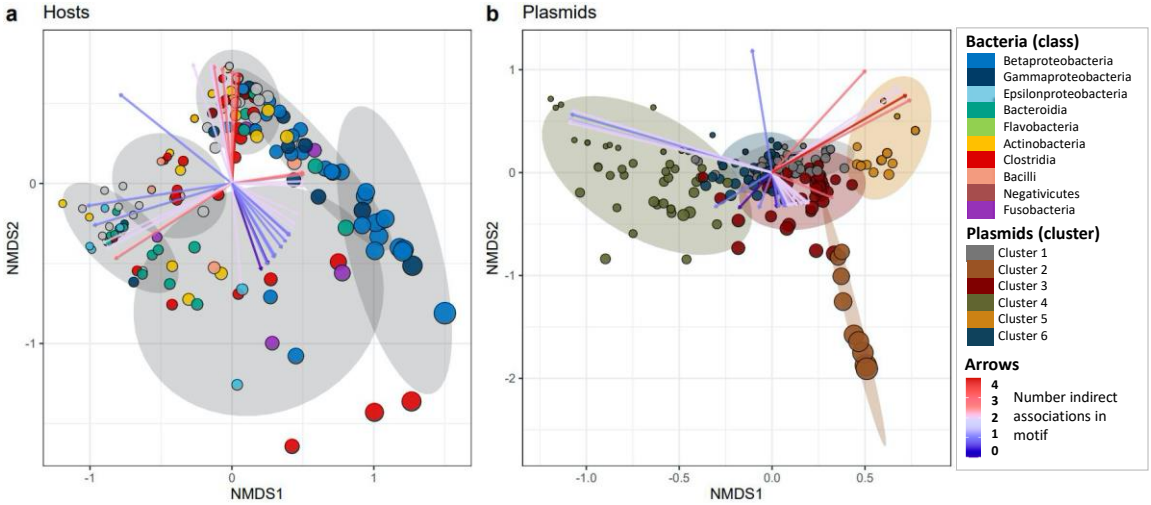


Figure 2. Functional similarity and indirect associations in the network. Nonmetric multidimensional scaling according to the number of indirect associations in network motifs for (a) hosts (b) plasmids. Clusters were formed by motif node position signatures of hosts and plasmids in the networks. The size of each dot reflects the abundance in the network.

Supplementary material

Table S1. Network metrics for bacterial classes.

Class	Number of compartments	NODF (nestedness)	Plasmid Generality	Connectance
Betaproteobacteria	1	18.944	1.707	0.110
Gammaproteobacteria	3	22.889	1.215	0.155
Bacteroidia	7	12.848	1.325	0.080
Clostridia	4	12.608	1.191	0.069
Actinobacteria	5	7.905	1.205	0.074
All classes	**1**	**8.511**	**1.818**	**0.025**

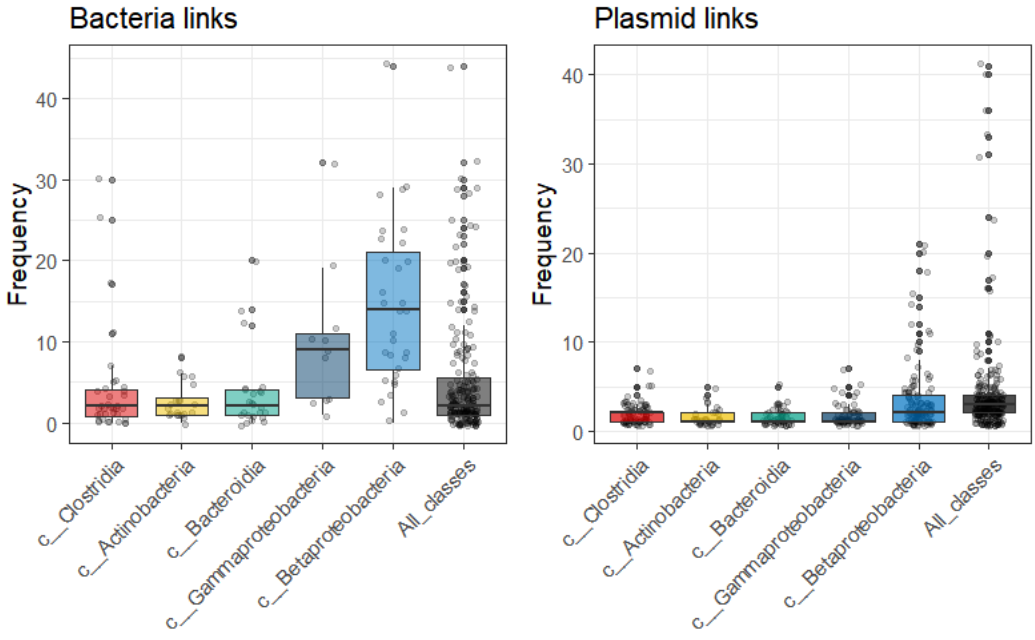
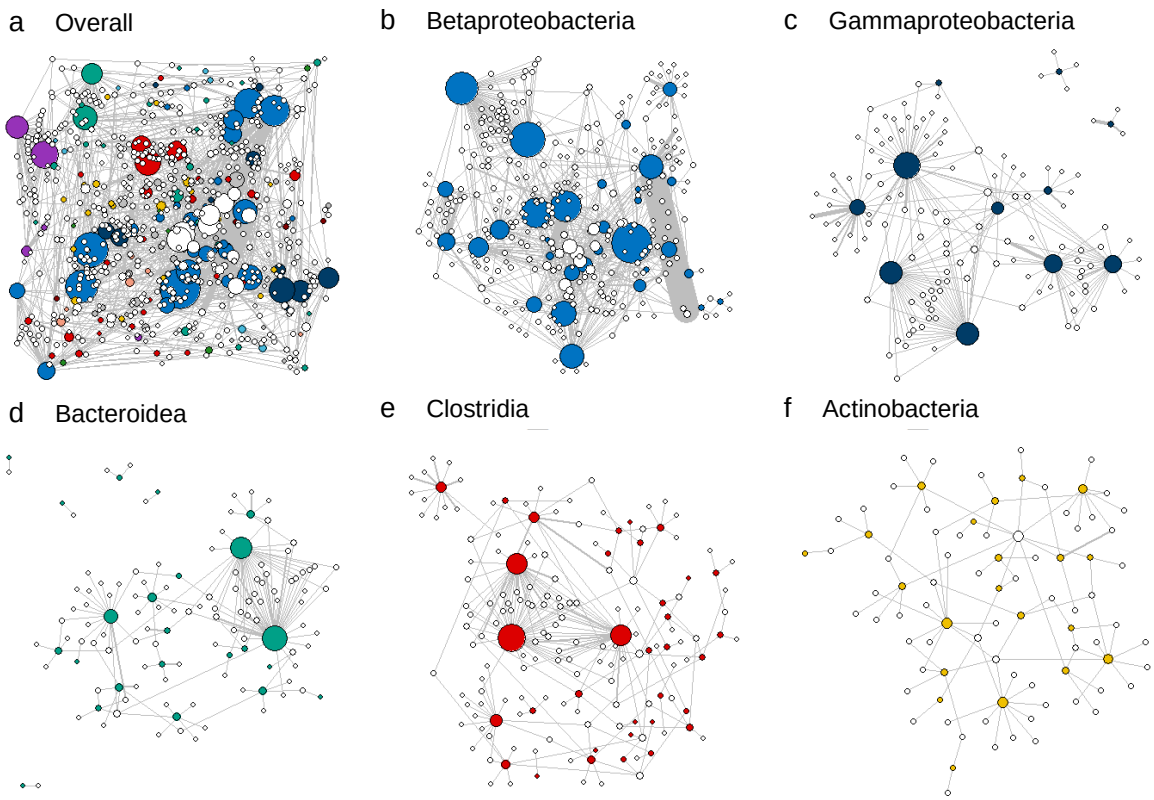


Figure S1. Link distribution for bacteria and plasmids for the overall network and different host classes.



283

284 **Figure S2.** Weighed cluster networks for the overall network and five bacterial classes.

285