Supporting Information for

Determining hosts of antibiotic resistance genes: A review of methodological advances

*Eric W. Rice[1, §], Phillip Wang[2, §], Adam L. Smith[2], and Lauren B. Stadler[1, *]*

[1]Department of Civil and Environmental Engineering, Rice University, 6100 Main Street, MS-519, Houston, TX 77005, USA

[2]Astani Department of Civil and Environmental Engineering, University of Southern California, 3620 South Vermont Avenue, Los Angeles, CA 90089, USA

[§]These authors contributed equally to this work.

[*]Corresponding author – lauren.stadler@rice.edu

Phone: +1 713-348-3307

Fax: +1 713-348-5268

Studies included in ARG-host data summary

We searched the literature for studies that reported ARG-host relationships. Search terms that were used to identify papers with published ARG-host information included: antibiotic resistance, antibiotic resistance genes (ARGs), pathogen, epicPCR, wastewater, class, host, mobile, resistome, metagenomic, correlation, soil, swine, cross, metal, co-occurrence, and Hi-C. Additional studies were identified within the PLOS|ONE Metagenomics Journal via their search and article similarity suggestions. We also identified papers based on those referenced in other manuscripts reporting ARG-host relationships. In order to be included in our data summary, the data from the publication had to meet the following criteria: (1) used a culture-independent method that was not donor-plasmid dependent; (2) provided bacterial host assignment of ARGs; (3) information could be gleaned from legible figures or tables. Data was manually compiled into a spreadsheet from articles that met the aforementioned criteria to generate a table of ARGs and their reported host assignments. Taxonomy assignments were confirmed using the NCBI Database.[1] Duplicate ARG host assignment entries within a single study were reduced to a single entry. ARGs were classified using the CARD Database.[2] If they were not found in CARD, classification was performed via review of scientific literature and the UniProt Database. Categorical assignments of ARGs and ARG Class are summarized in Table S1. ARGs were only included in the analysis if the host was classified at the family-level or greater resolution (*e.g.* if the host was unclassified at the family or phylum level, it was not included in the analysis). The total number of unique ARGs included in the analyses are summarized in Table S2. A complete list of the studies included in the data summary is provided in Table S3. The complete table of ARGs and host assignments included in the analysis can be made available upon email request.

Table S1. Categorical ARG Class assignments of ARGs across all included studies (n = 21).

Gene	ARG Class	Gene	ARG Class	Gene	ARG Class	Gene	ARG Class	Gene	ARG Class	Gene	ARG Class	Gene	ARG Class
aac(3')	aminoglycoside	apha1	aminoglycoside	cat3	phenicol	emrb	multidrug	mdto	multidrug	pmra	multidrug	teth	tetracycline
aac(3)-vi	aminoglycoside	apha3	aminoglycoside	cata	phenicol	emrb-qaca	multidrug	mdtp	multidrug	pmrb	peptide	tetl	tetracycline
aac(6')	aminoglycoside	arna	peptide	cata11	phenicol	emrd	multidrug	mef(a/e)	mls	pmrc	peptide	tetm	tetracycline
aac(6')-ie	aminoglycoside	arnb	peptide	cata13	phenicol	emre	multidrug	mefa	mls	pmre	peptide	teto	tetracycline
aac(6')-ii	aminoglycoside	arnc	peptide	cata2	phenicol	emrk	multidrug	mefb	mls	qach	multidrug	tetpa	tetracycline
aaca4	aminoglycoside	arnt	peptide	catb	phenicol	erea	mls	mel	mls	qacΔ1	quaternary ammonium	tetpb	tetracycline
aad(9)-ib	aminoglycoside	arr-2	ansamycin	catb3	phenicol	erma	mls	mexb	multidrug	qnrb6	quinolone	tetq	tetracycline
aada	aminoglycoside	baca	peptide	catb5	phenicol	ermb	mls	mexc	multidrug	qnrs	quinolone	tetr	tetracycline
aada2	aminoglycoside	bcea	polysaccharide	cbla	b-lactam	ermf	mls	mexd	multidrug	qnrvc	quinolone	tets	tetracycline
aada5	aminoglycoside	bcr	sulfonamide	ccra	b-lactam	ermg	mls	mexe	multidrug	rbpa	ansamycin	tetw	tetracycline
aadb	aminoglycoside	bcra	peptide	ceob	multidrug	ermq	mls	mexf	multidrug	rosa	peptide	tetx	tetracycline
aade	aminoglycoside	bla1	b-lactam	cepa	b-lactam	ermt	mls	mexi	multidrug	rosb	peptide	tolc	multidrug
abes	multidrug	blaa	b-lactam	cfx	b-lactam	erythromycin ribosome methylase	mls	mexk	multidrug	sdey	multidrug	vana	peptide
acra	multidrug	blac	b-lactam	cfxa	b-lactam	flor	phenicol	mexw	multidrug	shv	b-lactam	vanb	peptide
acrb	multidrug	b-lactam (class a - cfx, cbla, cepa)	b-lactam	cfxa2	b-lactam	fosa	fosfomycin	mexx	multidrug	shv-14	b-lactam	vanc	peptide
acrd	aminoglycoside	b-lactam (class c)	b-lactam	cfxa3	b-lactam	fosx	fosfomycin	mexy	multidrug	shv-24	b-lactam	vand	peptide
acre	multidrug	blamox-6	b-lactam	chloramphenicol exporter	phenicol	fox	b-lactam	mox	b-lactam	shv-46	b-lactam	vang	peptide
acrf	multidrug	blanps-1	b-lactam	cmeb	multidrug	hydrophobe amphiphile efflux family protein 1	multidrug	mpha	mls	shv-61	b-lactam	vanhb	peptide
acrs	multidrug	blaoxa	b-lactam	cml	phenicol	kdpe	aminoglycoside	mphb	mls	sm resistance protein a/b	aminoglycoside	vanr	peptide
act	b-lactam	blaoxa-10	b-lactam	cmla	phenicol	kluy-3	b-lactam	mphe	mls	smeb	multidrug	vanrb	peptide
adek	multidrug	blaoxa-2	b-lactam	cmla1	phenicol	ksga	aminoglycoside	mphg	mls	smee	multidrug	vanrd	peptide
aer	b-lactam	blaoxa-347	b-lactam	cmy	b-lactam	lmrb	mls	msrd	mls	stra	aminoglycoside	vanrg	peptide
ampc	b-lactam	blaoxa-58	b-lactam	cmy-21	b-lactam	lnub	mls	multidrug transporter	multidrug	strab	aminoglycoside	vans	peptide
amph	b-lactam	blaoxa-9	b-lactam	cpxr	multidrug	lnuc	mls	myra	mls	strb	aminoglycoside	vansb	peptide
amrb	multidrug	blashv-2	b-lactam	dfra1	pyrimidine inhibitor	maca	mls	nimd	nitroimidazole	sul1	sulfonamide	vantg	peptide
ant(3")	aminoglycoside	blatem	b-lactam	dfra14	pyrimidine inhibitor	macb	mls	nime	nitroimidazole	sul2	sulfonamide	vanug	peptide
ant(3")-ia	aminoglycoside	blatem-1	b-lactam	dfra17	pyrimidine inhibitor	mcr-5	peptide	norb	multidrug	sul3	sulfonamide	vanw	peptide
ant(6)	aminoglycoside	blaveb-1	b-lactam	dfra2	pyrimidine inhibitor	mdfa	multidrug	norm	multidrug	tet32	tetracycline	vanwb	peptide
ant(6)-ia	aminoglycoside	blaveb-3	b-lactam	dfra20	pyrimidine inhibitor	mdta	multidrug	nps-1	b-lactam	tet34	tetracycline	vanwg	peptide
ant(9)	aminoglycoside	blaveb-9	b-lactam	dfra22	pyrimidine inhibitor	mdtb	multidrug	omp36	multidrug	tet36	tetracycline	vanxb	peptide
aph(3")-ib	aminoglycoside	blavim-10	b-lactam	dfra26	pyrimidine inhibitor	mdtc	multidrug	ompf	multidrug	tet37	tetracycline	vanxd	peptide
aph(3')	aminoglycoside	blavim-14	b-lactam	dfra7	pyrimidine inhibitor	mdtd	multidrug	ompr	multidrug	tet39	tetracycline	vanxyg	peptide
aph(3")	aminoglycoside	blavim-15	b-lactam	dfrf	pyrimidine inhibitor	mdte	multidrug	oprj	multidrug	tet40	tetracycline	vany	peptide
aph(3')-ia	aminoglycoside	blavim-16	b-lactam	dhfr	pyrimidine inhibitor	mdtf	multidrug	oprm	multidrug	tet44	tetracycline	vanyb	peptide
aph(3")-ib	aminoglycoside	blavim-3	b-lactam	dhfr"	pyrimidine inhibitor	mdtg	multidrug	pac	aminonucleoside	teta	tetracycline	vanz	peptide
aph(3")-id	aminoglycoside	blavim-4	b-lactam	dhfr-iic	pyrimidine inhibitor	mdth	multidrug	pbp1b	b-lactam	tetb	tetracycline	vatb	mls
aph(3')-iiia	aminoglycoside	ble	peptide	drra	anthracycline	mdtk	multidrug	pbp2	b-lactam	tetc	tetracycline	ykkd	multidrug
aph(3')-iiib	aminoglycoside	bpef	multidrug	efpa	multidrug	mdtl	multidrug	pbp2x	b-lactam	tetd	tetracycline		
aph(6)	aminoglycoside	carb	b-lactam	emea	multidrug	mdtm	multidrug	pbp4b	b-lactam	tete	tetracycline		
aph(6)-id	aminoglycoside	cat	phenicol	emra	multidrug	mdtn	multidrug	pena	b-lactam	tetg	tetracycline		

Table S2. Number of ARGs in each ARG class for all included studies (n=21).

ARG class	Unique ARGs
Tetracycline	154
Multidrug	137
Aminoglycoside	109
β-Lactam	107
Peptide	87
MLS (Macrolide, lincosamide, streptogramin)	75
Quaternary Ammonium	43
Sulfonamide	34
Phenicol	25
Pyrimidine Inhibitor	17
Ansamycin	4
Polysaccharide	4
Quinolone	3
Fosfomycin	2
Nitroimidazole	2
Anthracycline	1
Aminonucleoside	1

Table S3. Studies included in the data summary that reported ARG hosts (n=21).

Study	Method Used
Stalder, et al. ISME 2019[3]	Hi-C
Hultman, et al. FEMS 2018[4]	EpicPCR
Liu, et al. Env Int 2019[5]	Metagenomics
Ma, et al. ES&T 2016[6]	Metagenomics
Che, et al. Microbiome 2019[7]	Metagenomics
Yadav, et al. STOTEN 2019[8]	Metagenomics
Xiong, et al. Microbiome 2018[9]	Metagenomics
Forslund et al. Genome Res 2013[10]	Metagenomics
Forsberg et al. Nature 2014[11]	Metagenomics
Goethem et al. Microbiome 2018[12]	Metagenomics
Li et al. ISME 2017[13]	Metagenomics
Ma et al. Microbiome 2017[14]	Metagenomics
Jia et al. Water Res 2017[15]	Metagenomics
Su, et al. Microbiome 2017[16]	Correlation
Li et al. ISME 2015[17]	Correlation
Johnson et al. ASM 2016[18]	Correlation
Zhu et al. Nat Microbio. 2017[19]	Correlation
Feng et al. Environ Microbiol 2018[20]	Correlation
Tian et al. Water Res 2016[21]	Correlation
Luo et al. ES&T 2017[22]	Correlation
Zhao et al. Water Res 2018[23]	Correlation

Analysis of ARG hosts

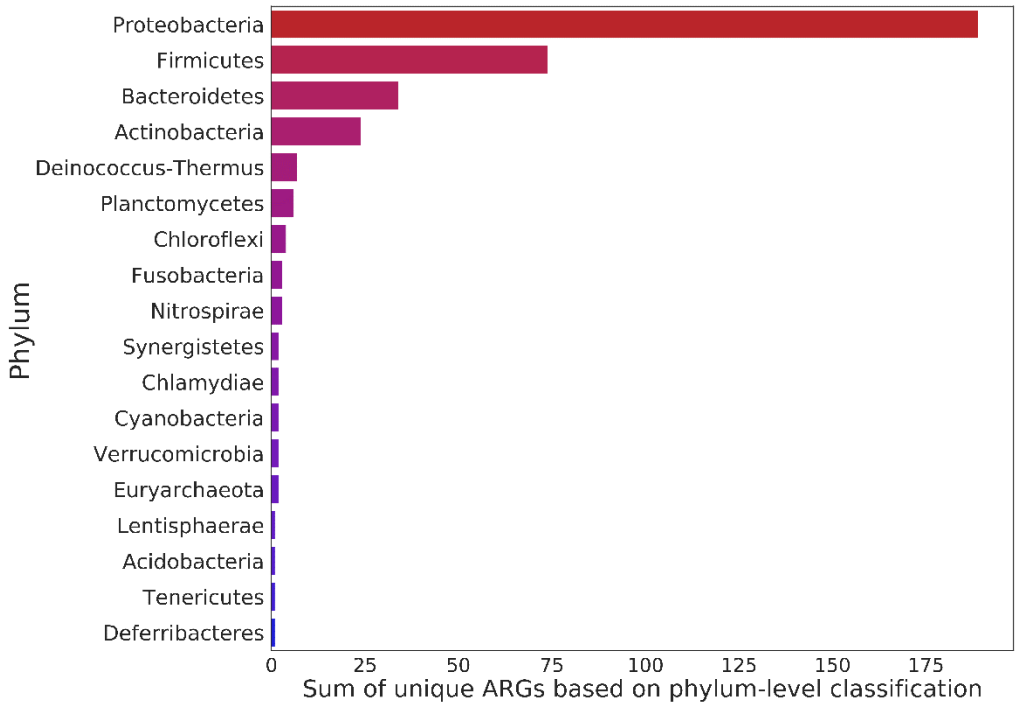

Figure S1. The sum of unique ARGs based on phylum-level classification across all studies (n=21)

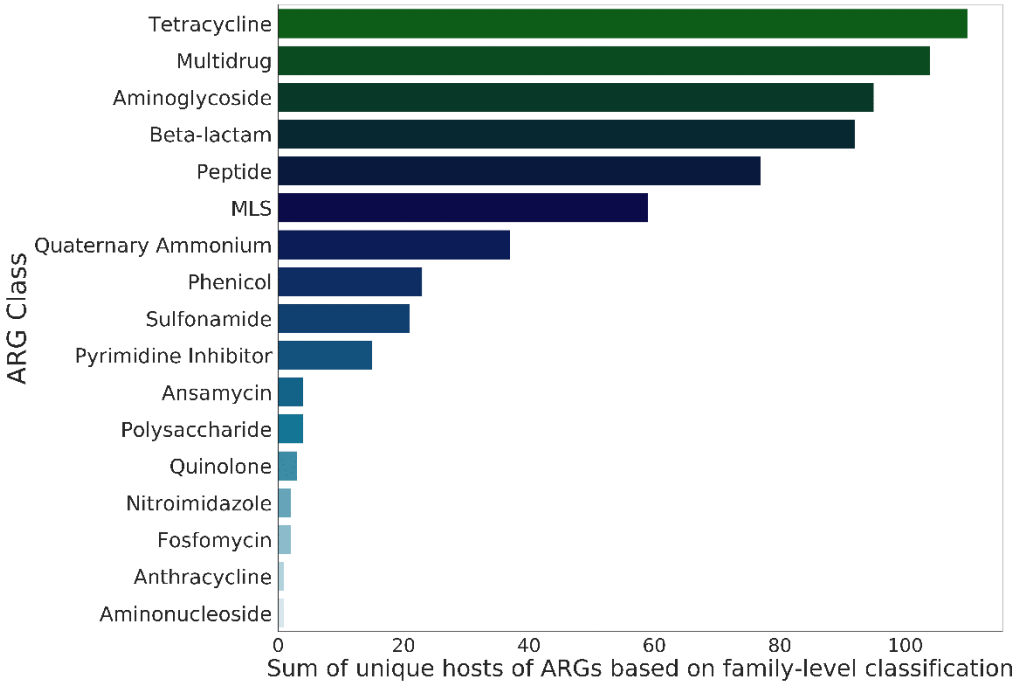

Figure S2. The sum of unique taxa hosting ARGs based on family-level classification, across all studies (n=21)

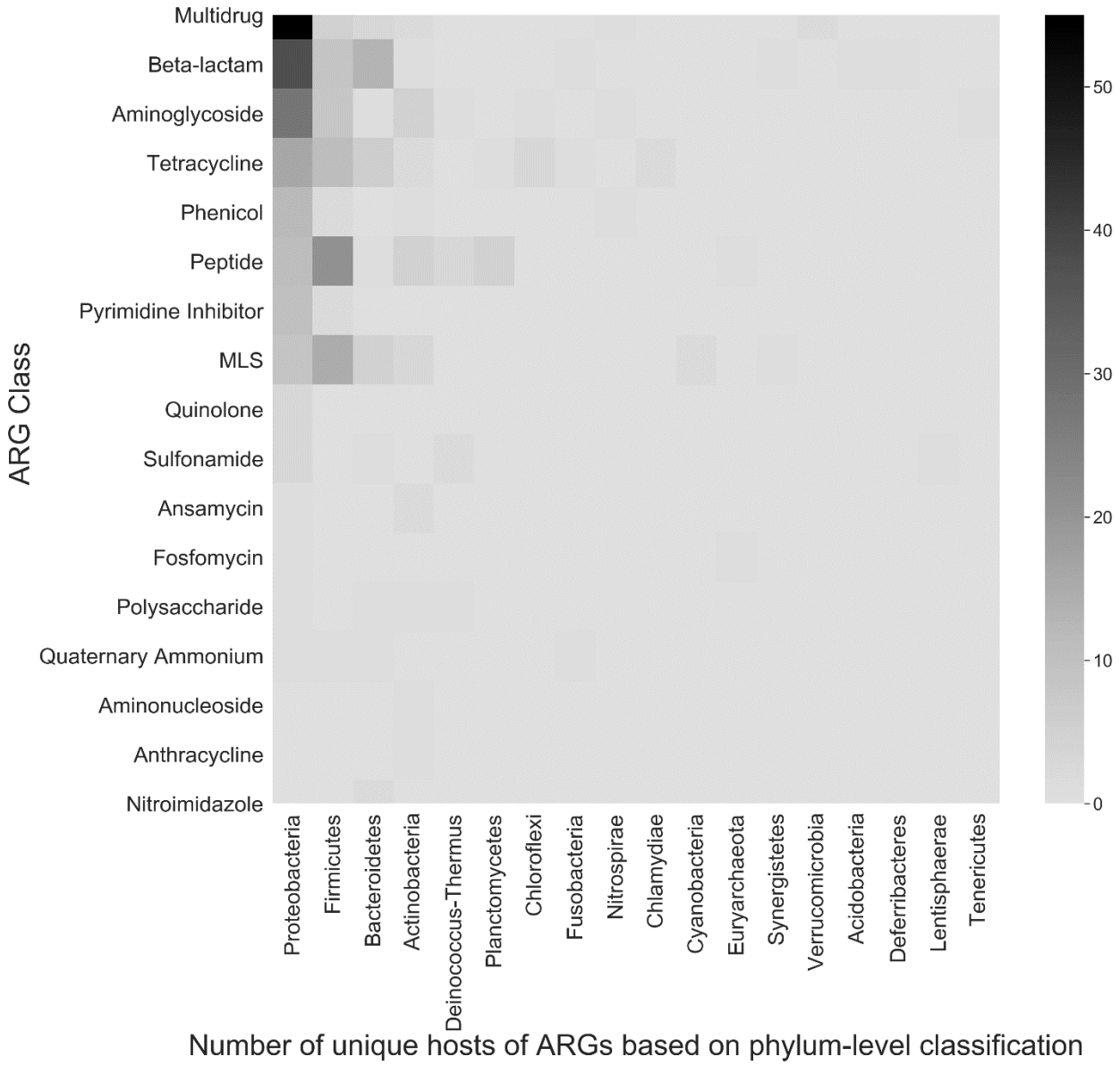

Figure S3. The number of unique ARGs (grouped by ARG class) hosted by unique taxa based on phylum-level classifications across all studies (n=21)

Wastewater ARG hosts analysis

We compared ARG hosts in wastewater influent, effluent, and activated sludge samples reported in all wastewater environment studies (n = 7 studies; Table S4 and Figure S4, S5). The dominant hosts of ARGs at the family level were *Comamonadaceae* followed by *Aeromonadaceae* and *Rhodocyclaceae* (Figure S4). Others have found *Comamonadaceae* to be a dominant host of ARGs in lettuces when using correlation network analysis.[1] In addition, we compared ARG hosts reported by Che et al. (2019)[7] and Hultman et al. (2018)[4] as they were the only two studies that assessed ARG hosts in both influent and effluent samples from the same WWTPs (Figure S6). The findings show that the effluent community of ARG hosts are a subset of the influent ARG host community based on family-level classifications. In addition, the ARG hosts in activated sludge were also identified in both influent and effluent samples (Figure S6). We also compared the reported ARG hosts *across studies* in wastewater influent, effluent, and activated sludge to assess the variability in reported ARG hosts (Figure S7). This variability across studies is likely due to many factors such as location, wastewater characteristics, process configuration, and methodology used to determine ARG hosts. We see that influent, effluent, and activated sludge ARG hosts were in general very study-specific (the majority of unique hosts were unique to each individual study; Figure S7). These results highlight the need for more ARG-host information to assess whether there are dominant ARG hosts shared across many WWTP environments, and whether treatment process or wastewater characteristics significantly associate or select for certain ARG hosts.

Table S4. Studies that reported ARG-host information in wastewater environments. N represents the number of wastewater treatment plants included in the individual study.

Study	Influent	Effluent	Activated Sludge	Method	Country
Che, Microbiome 2019[7]	n = 3	n = 3	n = 3	Metagenomics	Hong Kong
Hultman, FEMS 2018[4]	n = 2	n = 2		EpicPCR	Finland
Liu, Env Int 2019[5]			n = 3	Metagenomics	Taiwan
Ma, ES&T 2016[6]	n = 1			Metagenomics	Hong Kong
Stalder, ISME 2019[3]	n = 1			Hi-C	USA
Su, Microbiome 2017[16]	n = 32			Correlation	China
Yadav, STOTEN 2019[8]			n = 2	Metagenomics	India

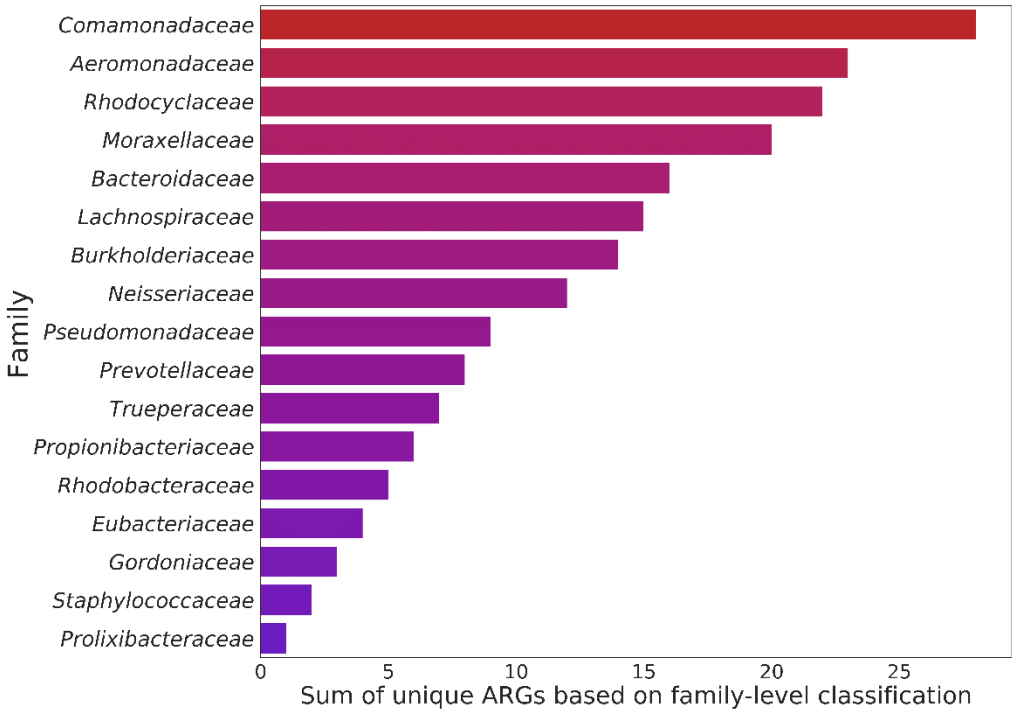

Figure S4. The sum of unique ARGs based on family-level classification across wastewater studies (n=7). Only the top 17 families are shown.

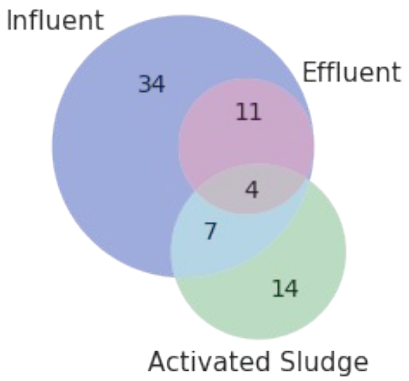

Figure S5. Unique ARG hosts (classified at the family-level) present in wastewater influent, activated sludge, and effluent samples (n =7 studies total).

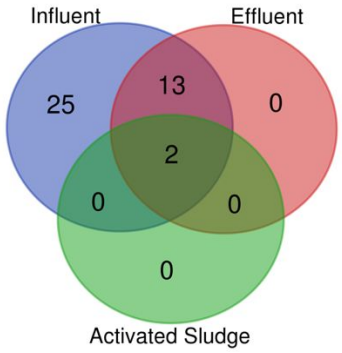

Figure S6. Unique ARG hosts (classified at the family-level) present in wastewater influent, activated sludge, and effluent samples reported by Che *et al.* (2019) and Hultman *et al.* (2018).[4,7]

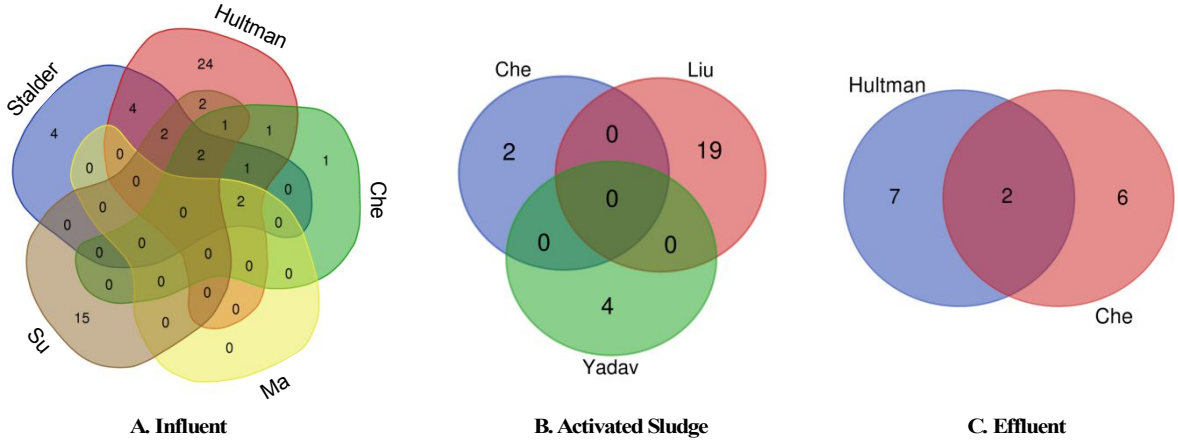

Figure S7. Unique ARG hosts (classified at the family-level) present in wastewater influent (A), activated sludge (B), and effluent (C) samples grouped by study.[3–8,16]

References

(1) NCBI Taxonomy Browser https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi.

(2) The Comprehensive Antibiotic Resistance Database (CARD) https://card.mcmaster.ca/.

(3) Stalder, T.; Press, M. O.; Sullivan, S.; Liachko, I.; Top, E. M. Linking the Resistome and Plasmidome to the Microbiome. *ISME J.* **2019**, *13*, 2437–2446. https://doi.org/10.1038/s41396-019-0446-4.

(4) Hultman, J.; Tamminen, M.; Pärnänen, K.; Cairns, J.; Karkman, A.; Virta, M. Host Range of Antibiotic Resistance Genes in Wastewater Treatment Plant Influent and Effluent. *FEMS Microbiol. Ecol.* **2018**, *94* (4), 1–10. https://doi.org/10.1093/femsec/fiy038.

(5) Liu, Z.; Klümper, U.; Liu, Y.; Yang, Y.; Wei, Q.; Lin, J. G.; Gu, J. D.; Li, M. Metagenomic and Metatranscriptomic Analyses Reveal Activity and Hosts of Antibiotic Resistance Genes in Activated Sludge. *Environ. Int.* **2019**, *129*, 208–220. https://doi.org/10.1016/j.envint.2019.05.036.

(6) Ma, L.; Xia, Y.; Li, B.; Yang, Y.; Li, L.-G.; Tiedje, J. M.; Zhang, T. Metagenomic Assembly Reveals Hosts of Antibiotic Resistance Genes and the Shared Resistome in Pig, Chicken, and Human Feces. *Environ. Sci. Technol.* **2016**, *50* (1), 420–427. https://doi.org/10.1021/acs.est.5b03522.

(7) Che, Y.; Xia, Y.; Liu, L.; Li, A.-D.; Yang, Y.; Zhang, T. Mobile Antibiotic Resistome in Wastewater Treatment Plants Revealed by Nanopore Metagenomic Sequencing. *Microbiome* **2019**, *7* (44), 1–13. https://doi.org/10.1186/s40168-019-0663-0.

(8) Yadav, S.; Kapley, A. Exploration of Activated Sludge Resistome Using Metagenomics. *Sci. Total Environ.* **2019**, *692*, 1155–1164. https://doi.org/10.1016/j.scitotenv.2019.07.267.

(9) Xiong, W.; Wang, Y.; Sun, Y.; Ma, L.; Zeng, Q.; Jiang, X.; Li, A.; Zeng, Z.; Zhang, T. Antibiotic-Mediated Changes in the Fecal Microbiome of Broiler Chickens Define the Incidence of Antibiotic Resistance Genes. *Microbiome* **2018**, *6*, 1–11. https://doi.org/10.1186/s40168-018-0419-2.

(10) Forslund, K.; Sunagawa, S.; Kultima, J. R.; Mende, D. R.; Arumugam, M.; Typas, A.; Bork, P. Country-Specific Antibiotic Use Practices Impact the Human Gut Resistome. *Genome Res.* **2013**, *23* (7), 1163–1169. https://doi.org/10.1101/gr.155465.113.

(11) Forsberg, K. J.; Patel, S.; Gibson, M. K.; Lauber, C. L.; Knight, R.; Fierer, N.; Dantas, G. Bacterial Phylogeny Structures Soil Resistomes across Habitats. *Nature* **2014**, *509*, 612–616. https://doi.org/10.1038/nature13377.

(12) Van Goethem, M. W.; Pierneef, R.; Bezuidt, O. K. I.; Van De Peer, Y.; Cowan, D. A.; Makhalanyane, T. P. A Reservoir of 'Historical' Antibiotic Resistance Genes in Remote Pristine Antarctic Soils. *Microbiome* **2018**, *6*, 1–12. https://doi.org/10.1186/s40168-018-0424-5.

(13) Li, L.; Xia, Y.; Zhang, T. Co-Occurrence of Antibiotic and Metal Resistance Genes Revealed in Complete Genome Collection. *ISME J.* **2017**, *11*, 651–662. https://doi.org/10.1038/ismej.2016.155.

(14) Ma, L.; Li, B.; Jiang, X.-T.; Wang, Y.-L.; Xia, Y.; Li, A.-D.; Zhang, T. Catalogue of Antibiotic Resistome and Host-Tracking in Drinking Water Deciphered by a Large Scale Survey. *Microbiome* **2017**, *5*, 1–12. https://doi.org/10.1186/s40168-017-0369-0.

(15) Jia, S.; Zhang, X.-X.; Miao, Y.; Zhao, Y.; Ye, L.; Li, B.; Zhang, T. Fate of Antibiotic

Resistance Genes and Their Associations with Bacterial Community in Livestock Breeding Wastewater and Its Receiving River Water. *Water Res.* **2017**, *124*, 259–268. https://doi.org/10.1016/j.watres.2017.07.061.

(16) Su, J.; An, X.; Li, B.; Chen, Q.-L.; Gillings, M. R.; Chen, H.; Zhang, T.; Zhu, Y.-G. Metagenomics of Urban Sewage Identifies an Extensively Shared Antibiotic Resistome in China. *Microbiome* **2017**, *5*, 1–15. https://doi.org/10.1186/s40168-017-0298-y.

(17) Li, B.; Yang, Y.; Ma, L.; Ju, F.; Guo, F.; Tiedje, J. M.; Zhang, T. Metagenomic and Network Analysis Reveal Wide Distribution and Co-Occurrence of Environmental Antibiotic Resistance Genes. *ISME J.* **2015**, *9*, 2490–2502. https://doi.org/10.1038/ismej.2015.59.

(18) Johnson, T. A.; Stedtfeld, R. D.; Wang, Q.; Cole, J. R.; Hashsham, S. A.; Looft, T.; Zhu, Y.; Tiedje, J. M. Clusters of Antibiotic Resistance Genes Enriched Together Stay Together in Swine Agriculture. *MBio* **2016**, *7* (2), e02214-15. https://doi.org/10.1128/mBio.02214-15.

(19) Zhu, Y.-G.; Zhao, Y.; Li, B.; Huang, C.-L.; Zhang, S.-Y.; Yu, S.; Chen, Y.-S.; Zhang, T.; Gillings, M. R.; Su, J.-Q. Continental-Scale Pollution of Estuaries with Antibiotic Resistance Genes. *Nat. Microbiol.* **2017**, *2*, 1–7. https://doi.org/10.1038/nmicrobiol.2016.270.

(20) Feng, J.; Li, B.; Jiang, X.; Yang, Y.; Wells, G. F.; Zhang, T.; Li, X. Antibiotic Resistome in a Large-Scale Healthy Human Gut Microbiota Deciphered by Metagenomic and Network Analyses. *Environ. Microbiol.* **2017**, *20* (1), 355–368. https://doi.org/10.1111/1462-2920.14009.

(21) Tian, Z.; Zhang, Y.; Yu, B.; Yang, M. Changes of Resistome, Mobilome and Potential Hosts of Antibiotic Resistance Genes during the Transformation of Anaerobic Digestion from Mesophilic to Thermophilic. *Water Res.* **2016**, *98*, 261–269. https://doi.org/10.1016/j.watres.2016.04.031.

(22) Luo, G.; Li, B.; Li, L.-G.; Zhang, T.; Angelidaki, I. Antibiotic Resistance Genes and Correlations with Microbial Community and Metal Resistance Genes in Full-Scale Biogas Reactors As Revealed by Metagenomic Analysis. *Environ. Sci. Technol.* **2017**, *51* (7), 4069–4080. https://doi.org/10.1021/acs.est.6b05100.

(23) Zhao, R.; Feng, J.; Yin, X.; Liu, J.; Fu, W.; Berendonk, T. U.; Zhang, T.; Li, X.; Li, B. Antibiotic Resistome in Landfill Leachate from Different Cities of China Deciphered by Metagenomic Analysis. *Water Res.* **2018**, *134*, 126–139. https://doi.org/10.1016/j.watres.2018.01.063.