

Supplementary Information for

Duplicated antibiotic resistance genes reveal ongoing selection and horizontal gene transfer in bacteria

Rohan Maddamsetti^{1,2}, Yi Yao^{1,2}, Teng Wang^{1,2}, Junheng Gao³, Vincent T. Huang^{1,2}, Grayson S. Hamrick^{1,2,4}, Hye-In Son^{1,2}, Lingchong You^{1,2,4,5*}

¹Center for Quantitative Biodesign, Duke University, Durham, NC, USA

²Department of Biomedical Engineering, Duke University, Durham, NC, USA.

³Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, USA.

⁴Center for Biomolecular and Tissue Engineering, Duke University, Durham, NC, USA

⁵Department of Molecular Genetics and Microbiology, Duke University School of Medicine, Durham, NC, USA

* Email for correspondence: you@duke.edu*

Supplementary Table 1. Proportion of isolates containing duplicated antibiotic resistance genes (ARGs).

Annotation	Isolates	Isolates with Duplicated ARGs	Proportion of isolates with Duplicated ARGs	95% confidence interval
Humans	7777	1104	0.1420	[0.1342, 0.1497]
Plants & Animals	3123	61	0.0195	[0.0147, 0.0244]
Earth	2495	38	0.0152	[0.0104, 0.0200]
Livestock	2025	272	0.1343	[0.1195, 0.1492]
Water	1224	31	0.0253	[0.0165, 0.0341]
Human-impacted	1191	88	0.0739	[0.0590, 0.0887]
Food	1103	46	0.0417	[0.0299, 0.0535]

Supplementary Table 2. Proportion of isolates containing single-copy antibiotic resistance genes (ARGs).

Annotation	Isolates	Isolates with single-copy ARGs	Proportion of isolates with single-copy ARGs	95% confidence interval
Humans	7777	7742	0.9955	[0.9940, 0.9970]
Plants & Animals	3123	2983	0.9552	[0.9479, 0.9624]
Earth	2495	2489	0.9976	[0.9957, 0.9995]
Livestock	2025	1949	0.9625	[0.9542, 0.9707]
Water	1224	1218	0.9951	[0.9912, 0.9990]
Human-impacted	1191	1187	0.9966	[0.9934, 0.9999]
Food	1103	1103	1.000	[1.0000, 1.0000]

Supplementary Table 3. Proportion of isolates containing duplicated genes.

Annotation	Isolates	Isolates with duplicated genes	Proportion of isolates with duplicated genes	95% confidence interval
Humans	7777	7249	0.9321	[0.9265, 0.9377]
Plants & Animals	3123	2659	0.8514	[0.8389, 0.8639]
Earth	2495	2246	0.9002	[0.8884, 0.9120]
Livestock	2025	1795	0.8864	[0.8726, 0.9002]
Water	1224	1100	0.8987	[0.8818, 0.9156]
Human-impacted	1191	1129	0.9479	[0.9353, 0.9606]
Food	1103	1001	0.9075	[0.8904, 0.9246]

Supplementary Table 4. Duplicated ARGs on chromosomes and plasmids.

Annotation	Chromosomal	Plasmid	Chromosomal	Plasmid
	duplicated	duplicated	duplicated	duplicated
	genes	genes	ARGs	ARGs
Humans	406,232	47,554	2,392	2,825
Livestock	96,593	10,728	553	814
Human-impacted	49,951	12,330	136	284
Plants & Animals	147,270	30,317	103	136
Food	37,409	6,727	68	101
Water	37,441	4,370	42	83
Earth	78,806	12,200	66	46

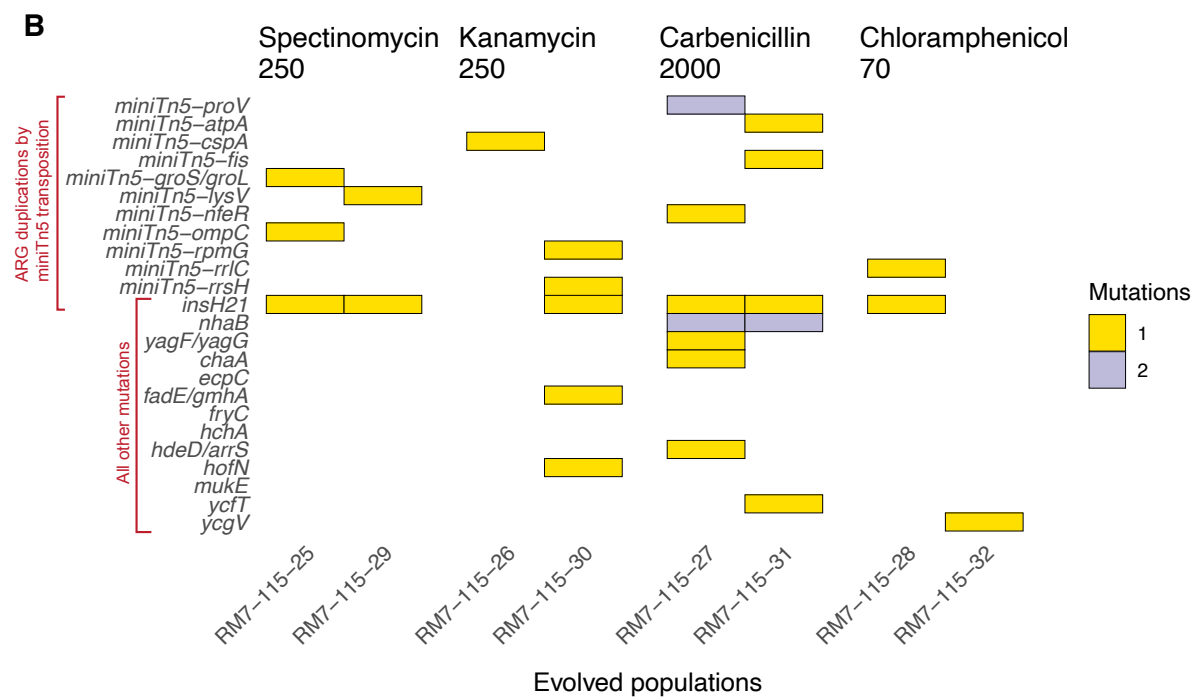
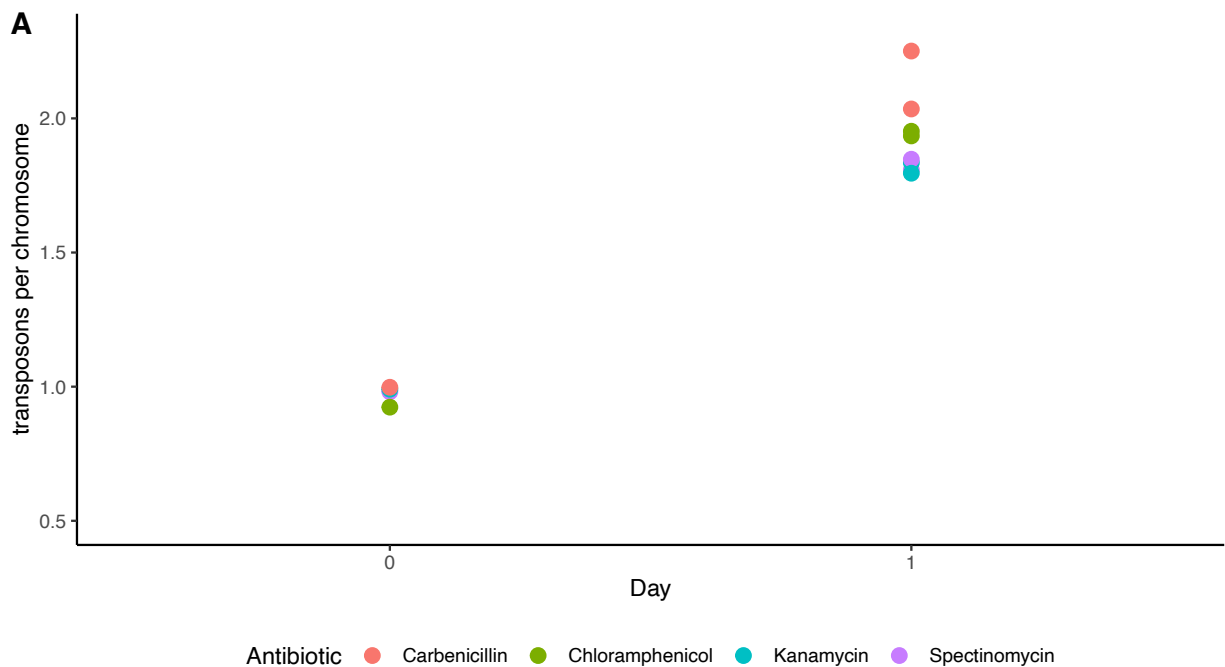
Supplementary Table 5. Single-copy ARGs on chromosomes and plasmids.

Annotation	Chromosomal	Plasmid	Chromosomal	Plasmid
	single-copy	single-copy	single-copy	single-copy
	genes	genes	ARGs	ARGs
Humans	26,809,779	571,349	78,328	14,708
Livestock	6,158,097	144,296	20,121	3,808
Plants & Animals	10,630,513	613,492	25,991	1,328
Human-impacted	4,695,471	182,323	13,353	1,308
Food	3,436,339	75,345	11,536	1,016
Earth	11,193,086	306,348	30,732	626
Water	4,344,815	97,867	9,076	521

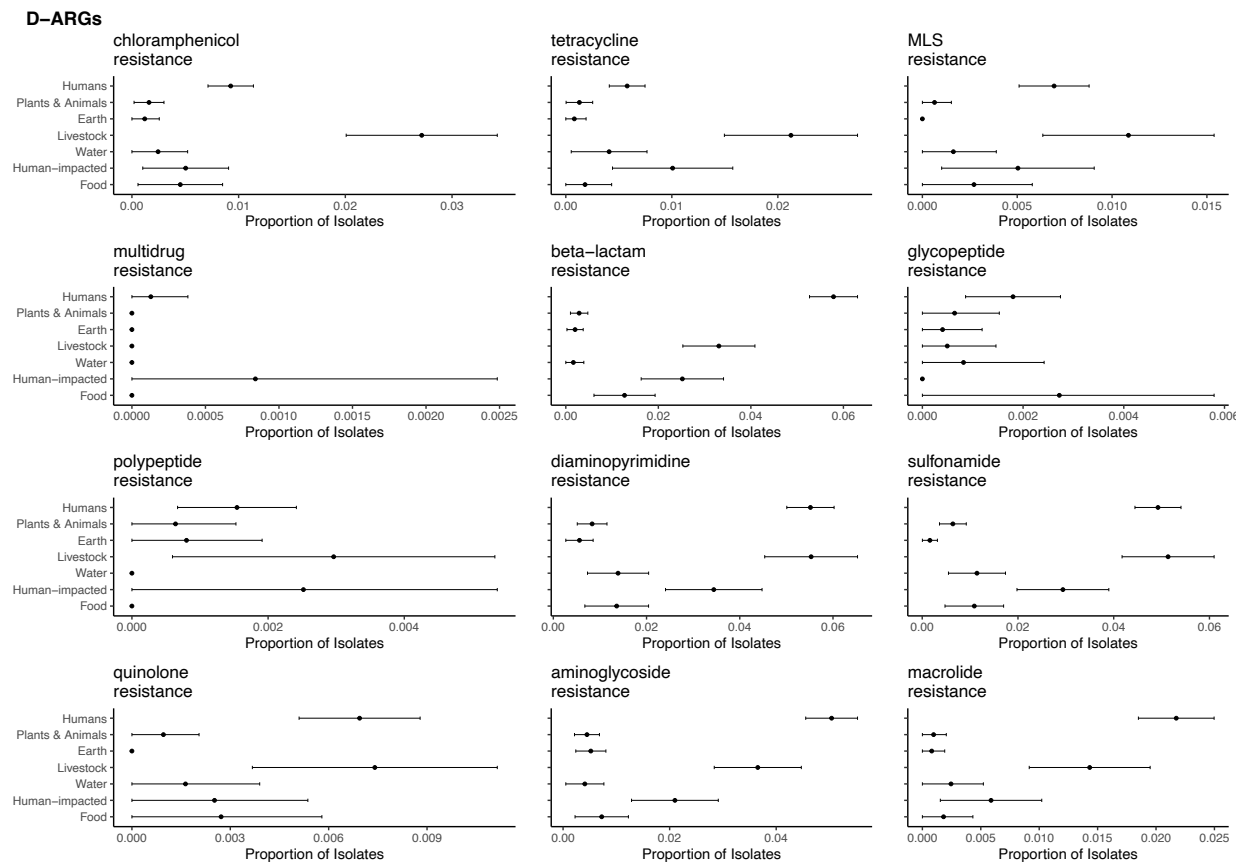
Supplementary Figure 1. Laboratory evolution with *E. coli* K-12 MG1655 demonstrate that antibiotic selection is sufficient to drive the rapid evolution of antibiotic resistance

through the duplication of antibiotic resistance genes. Eight populations (two replicates per treatment) were evolved in LB supplemented with one of four antibiotics (2000 µg/mL carbenicillin, 70 µg/mL chloramphenicol, 250 µg/mL kanamycin, or 250 µg/mL spectinomycin). Each panel shows a result generated by Illumina sequencing of ancestral clones and whole-population samples isolated after one day of antibiotic selection.

- A. One day of antibiotic selection was sufficient to drive an increase in ARG copy number in all treatments. ARG copy number in the ancestral clones is shown at the pre-treatment Day 0 timepoint for comparison.
- B. Genome sequencing reveals targets of positive selection after 1 day of growth under antibiotic treatment. All 8 populations show duplicated antibiotic resistance genes, driven by chromosomal transpositions of the Tn5 mini-transposons (miniTn5). Transposons in the RM7-115-32 population show an increase in copy number in panel A, but are not shown in panel B, because these mobilized into multi-copy 23S ribosomal RNA genes, and their precise chromosomal location is ambiguous.



Supplementary Figure 2. Proportions of isolates with duplicated ARGs, across 12 chemical classes of antibiotics.



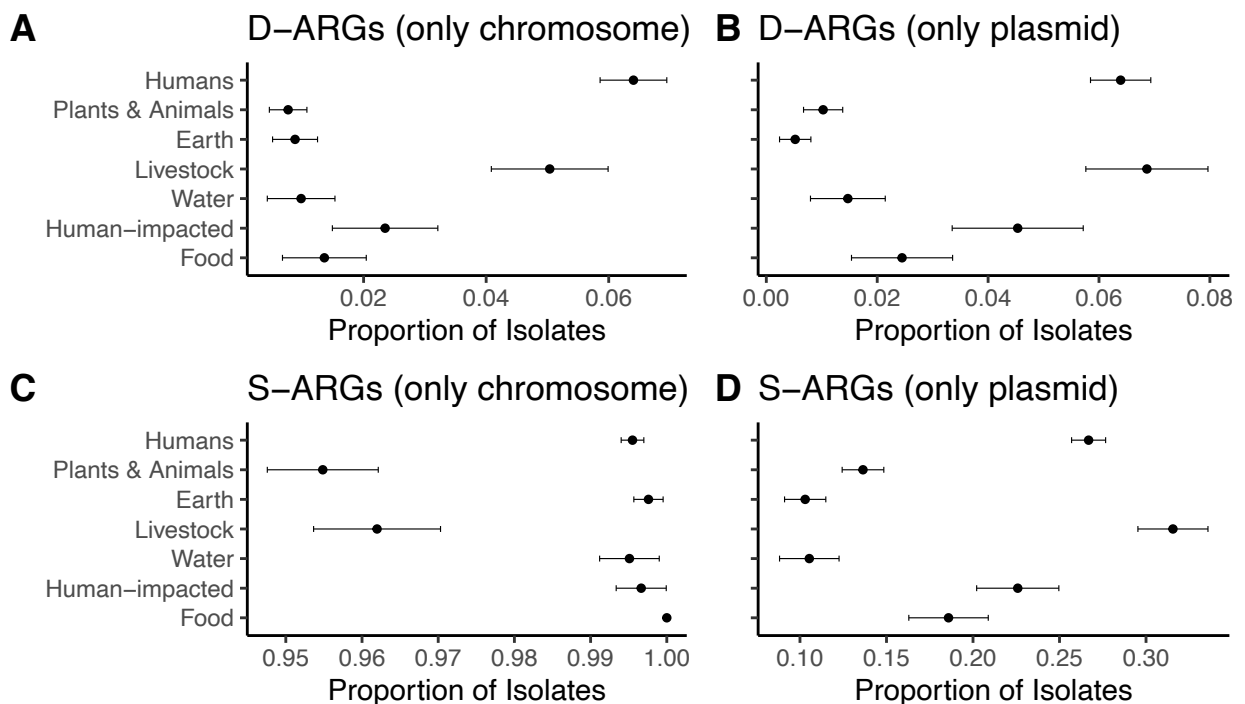
Supplementary Figure 3. D-ARGs are enriched on the chromosomes and plasmids of

bacteria isolated from humans and livestock. Error bars are 95% binomial proportion

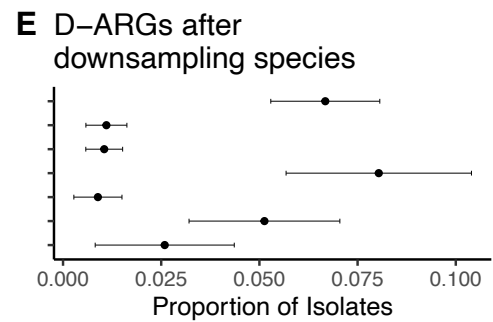
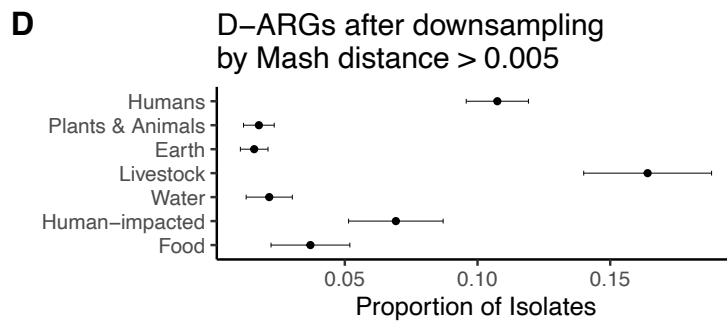
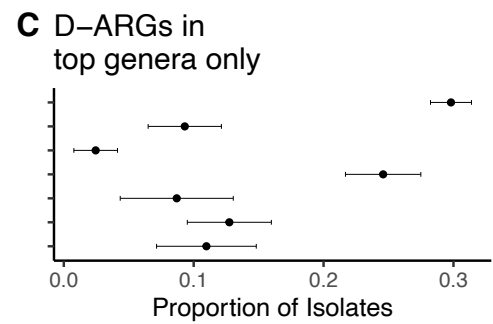
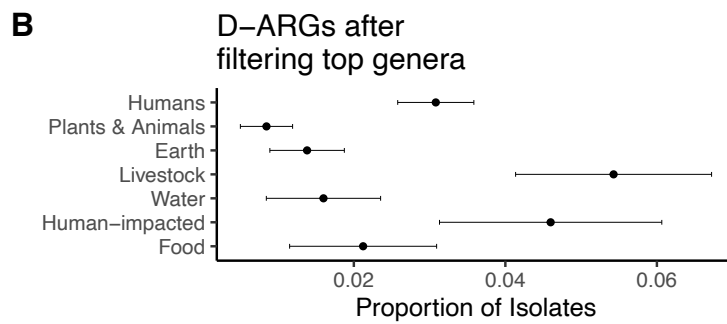
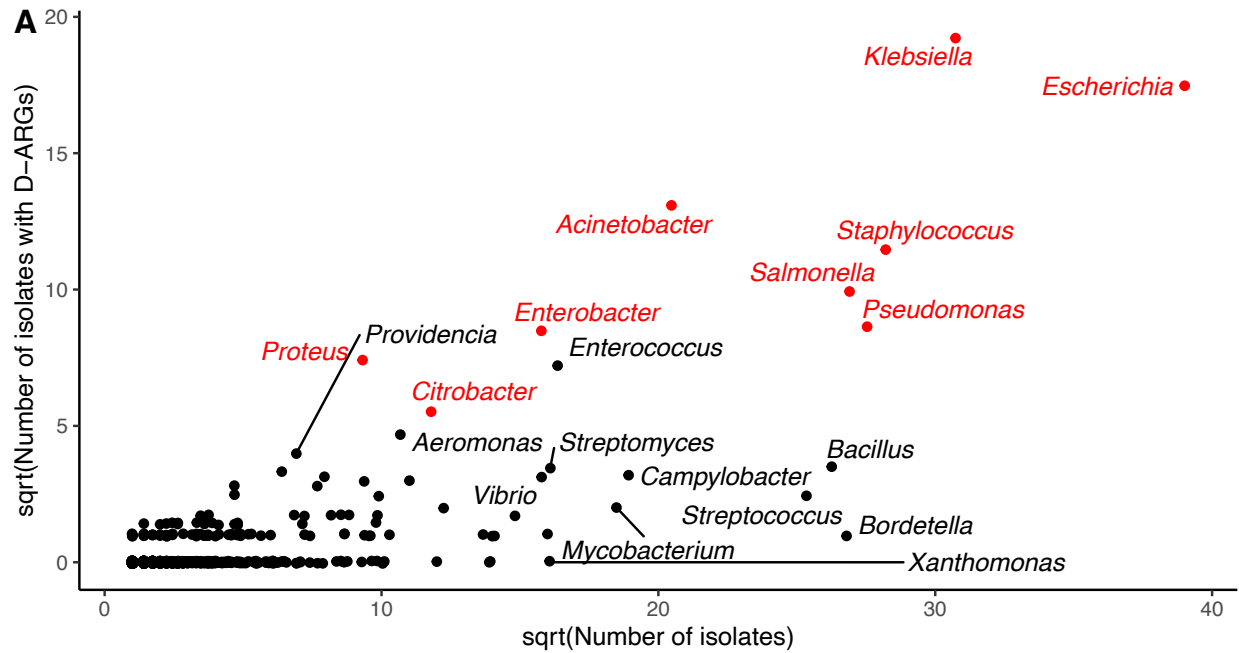
confidence intervals, calculated using the formula $p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$, where p is the proportion, n

is the sample size, and $Z_{\alpha/2} = 1.96$.

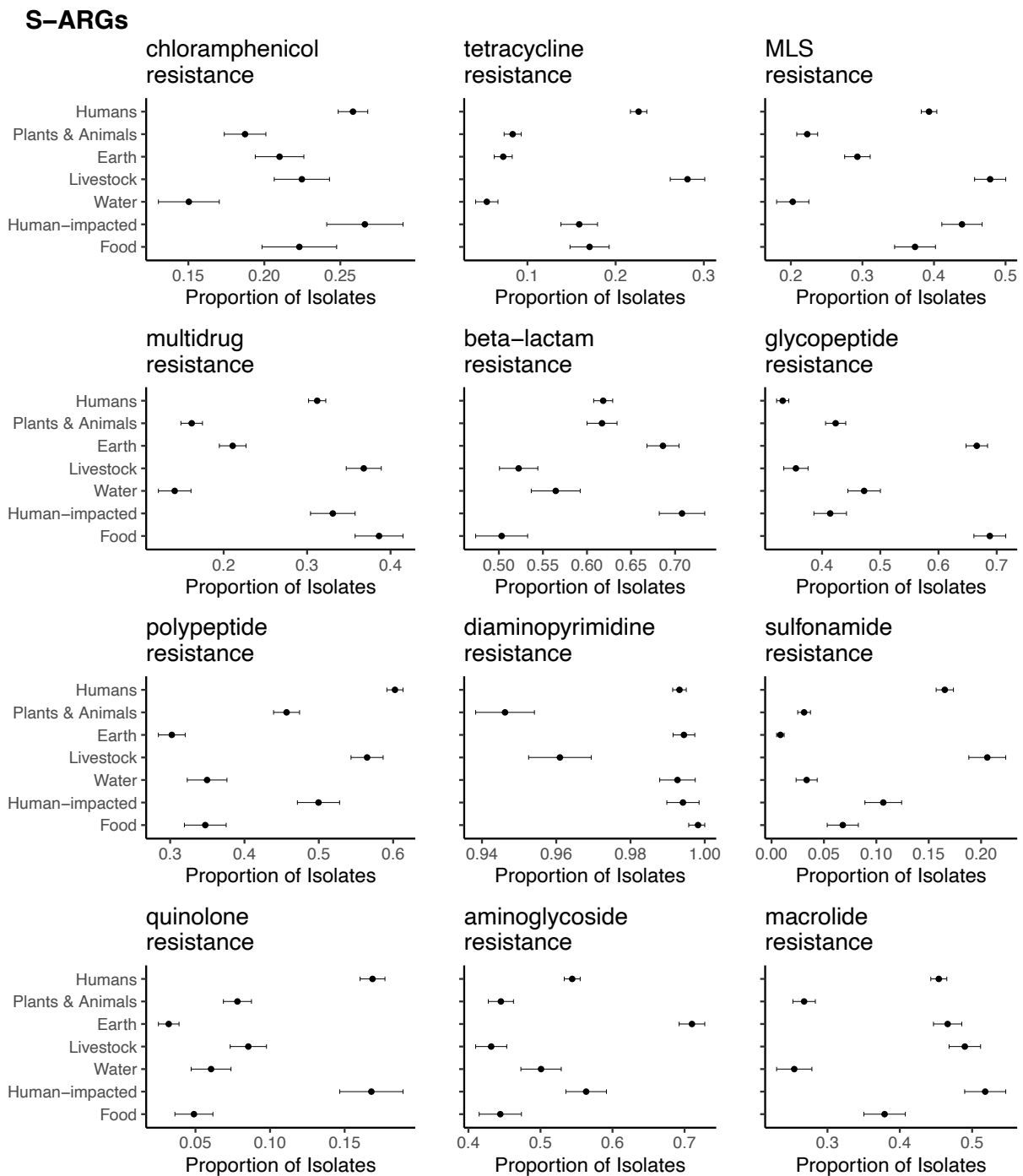
- A. Bacterial isolates with D-ARGs on chromosomes are specifically associated with humans and livestock.
- B. Bacterial isolates with D-ARGs on plasmids are specifically associated with humans and livestock.
- C. Most isolates contain at least one single-copy ARG on their chromosome.
- D. Bacterial isolates with S-ARGs on plasmids are specifically associated with humans, livestock, and human-impacted environments.



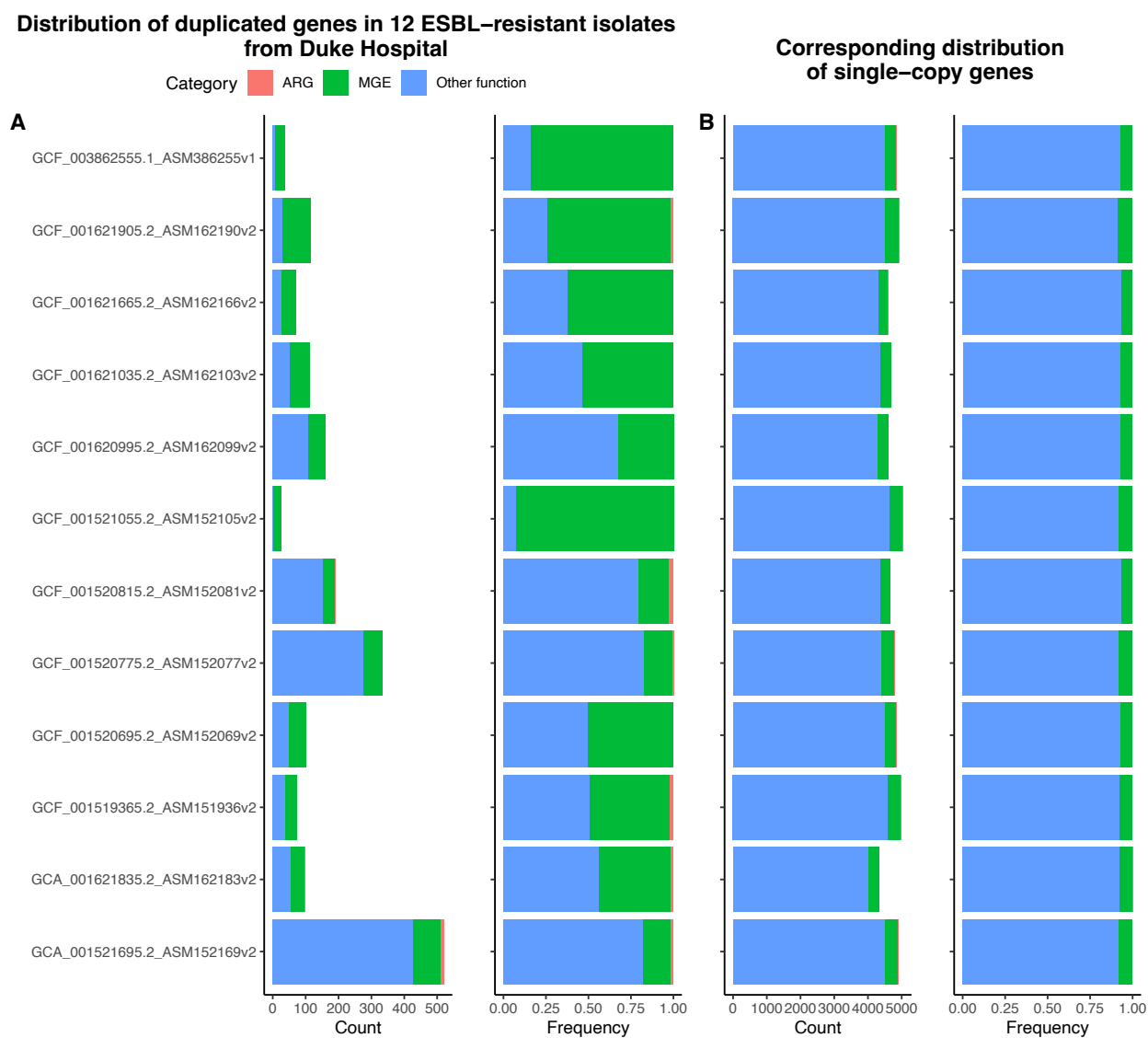
Supplementary Figure 4. Bacteria isolated from humans and livestock are more likely to have duplicated antibiotic resistance genes (D-ARGs) compared to bacteria isolated from other environments, even after taxonomic down-sampling. Error bars are 95% binomial proportion confidence intervals, calculated using the formula $p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$, where p is the proportion, n is the sample size, and $Z_{\alpha/2} = 1.96$. A) The number of D-ARGs per bacterial genus positively correlates with the number of isolates per genus. The most common genera in the data are shown in red. B) Removing the genera colored in red in panel A does not affect the enrichment of D-ARGs in livestock (5,638 isolates were removed, leaving 13,300 isolates). C) The same trend is seen, when just examining the genera colored in red in panel A. D) The enrichment of D-ARGs in humans and livestock is maintained, after downsampling the isolates by Mash distance > 0.005 . E) The enrichment of D-ARGs in humans and livestock is maintained, after downsampling the data to one genome per species.



Supplementary Figure 5. Proportions of isolates with single-copy ARGs, across 12 chemical classes of antibiotics.



Supplementary Figure 6. Distribution of duplicated ARGs in 12 clinical extended-spectrum beta-lactam (ESBL) resistant *E. coli* isolates from Duke Hospital. Long-read genome assemblies were downloaded from NCBI BioProject PRJNA290784. A) The number of duplicated genes in the 12 isolates is highly variable. 6 out of the 12 isolates harbor duplicated antibiotic resistance genes. B) All 12 isolates have similar numbers and fractions of single-copy genes encoding mobile genetic element functions and antibiotic resistance.



Supplementary Figure 7. Distribution of duplicated ARGs in 46 clinical antibiotic resistant isolates sequenced as part of the BARNARDS study. Long-read genome assemblies were downloaded from NCBI BioProject PRJNA767644. A) The number of duplicated genes in the 46 isolates is highly variable. 23 out of the 46 isolates harbor duplicated antibiotic resistance genes. B) All 46 isolates have similar numbers and fractions of single-copy genes encoding mobile genetic element functions and antibiotic resistance.

Distribution of duplicated genes in 46 genomes from the BARNARDS study

Corresponding distribution of single-copy genes

Category ARG MGE Other function

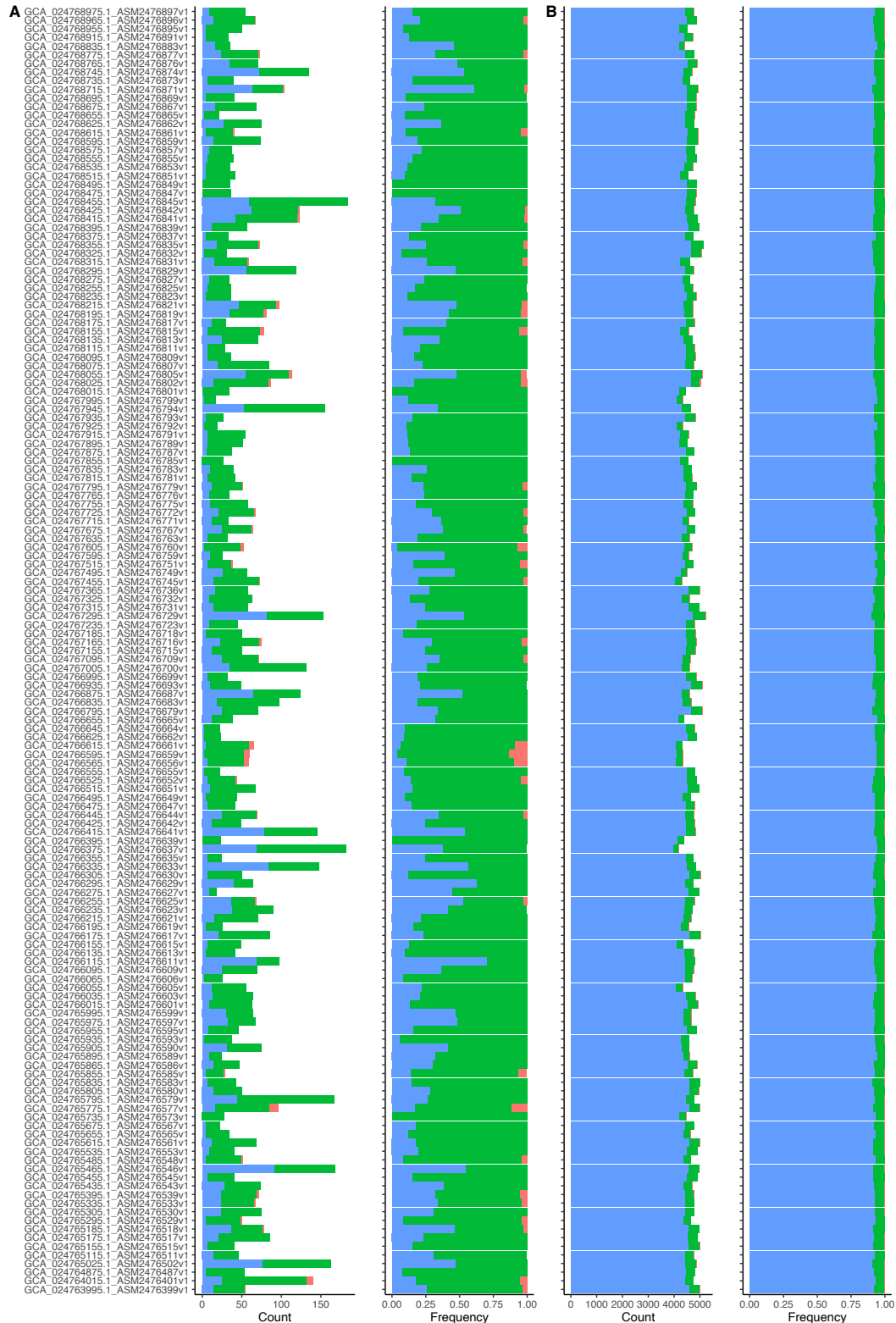


Supplementary Figure 8. Distribution of duplicated ARGs in 149 clinical antibiotic resistant isolates sequenced by Mahmud et al. (2022). Hybrid genome assemblies were downloaded from NCBI BioProject PRJNA824420. A) The number of duplicated genes in the 149 isolates is highly variable. 36 out of the 149 isolates harbor duplicated antibiotic resistance genes. B) All 149 isolates have similar numbers and fractions of single-copy genes encoding mobile genetic element functions and antibiotic resistance.

Distribution of duplicated genes in 149 genomes
from Barnes-Jewish Hospital (Mahmud et al. 2022)

Corresponding distribution
of single-copy genes

Category ■ ARG ■ MGE ■ Other function

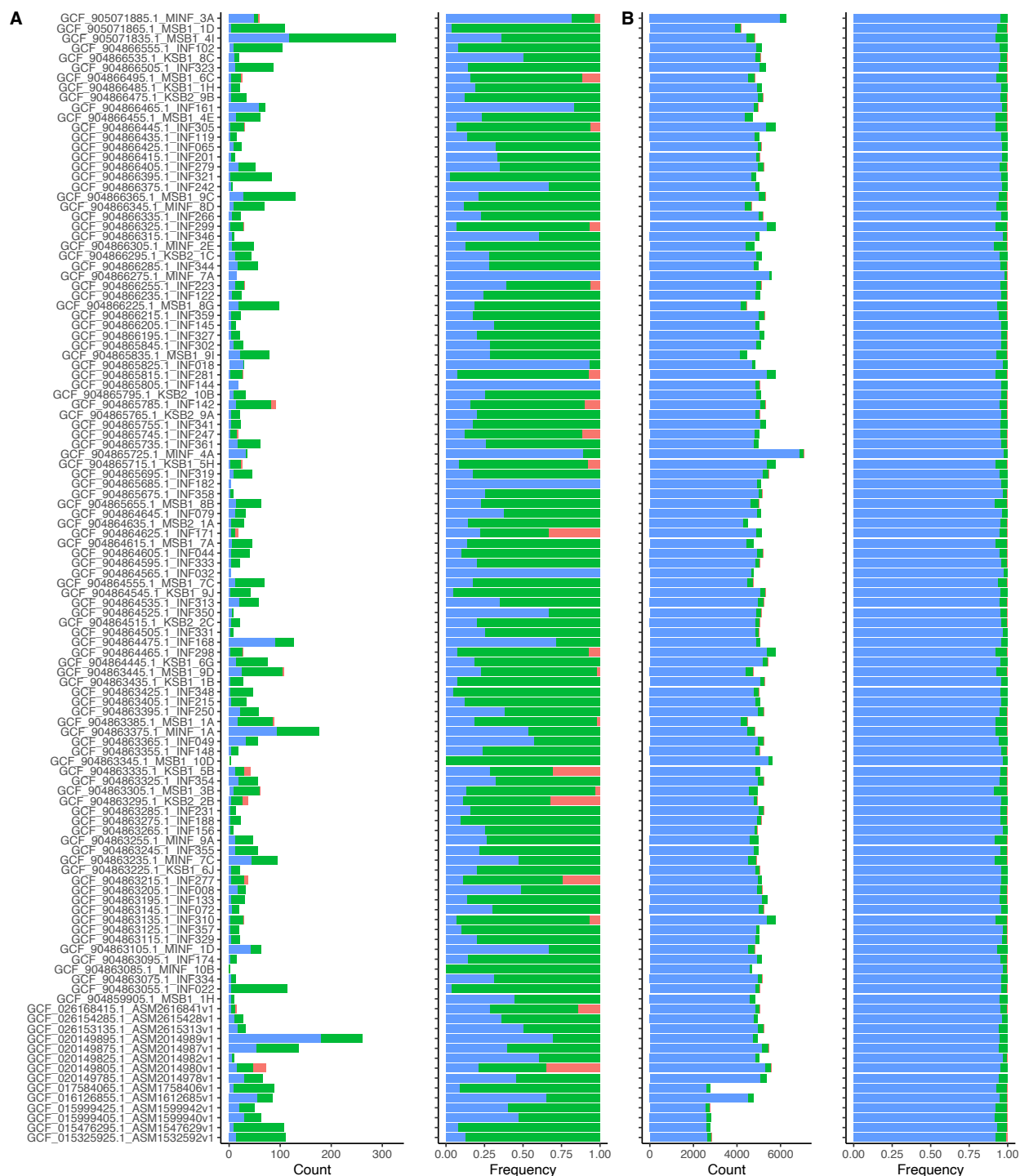


Supplementary Figure 9. Distribution of duplicated ARGs in 114 clinical antibiotic-resistant isolates from an Australian ICU (Hawkey et al. 2022). Complete genome assemblies were downloaded from NCBI BioProject PRJNA646837. A) The number of duplicated genes in the 114 isolates is highly variable. 20 out of the 114 isolates harbor duplicated antibiotic resistance genes. B) All 114 isolates have similar fractions of single-copy genes encoding mobile genetic element functions and antibiotic resistance.

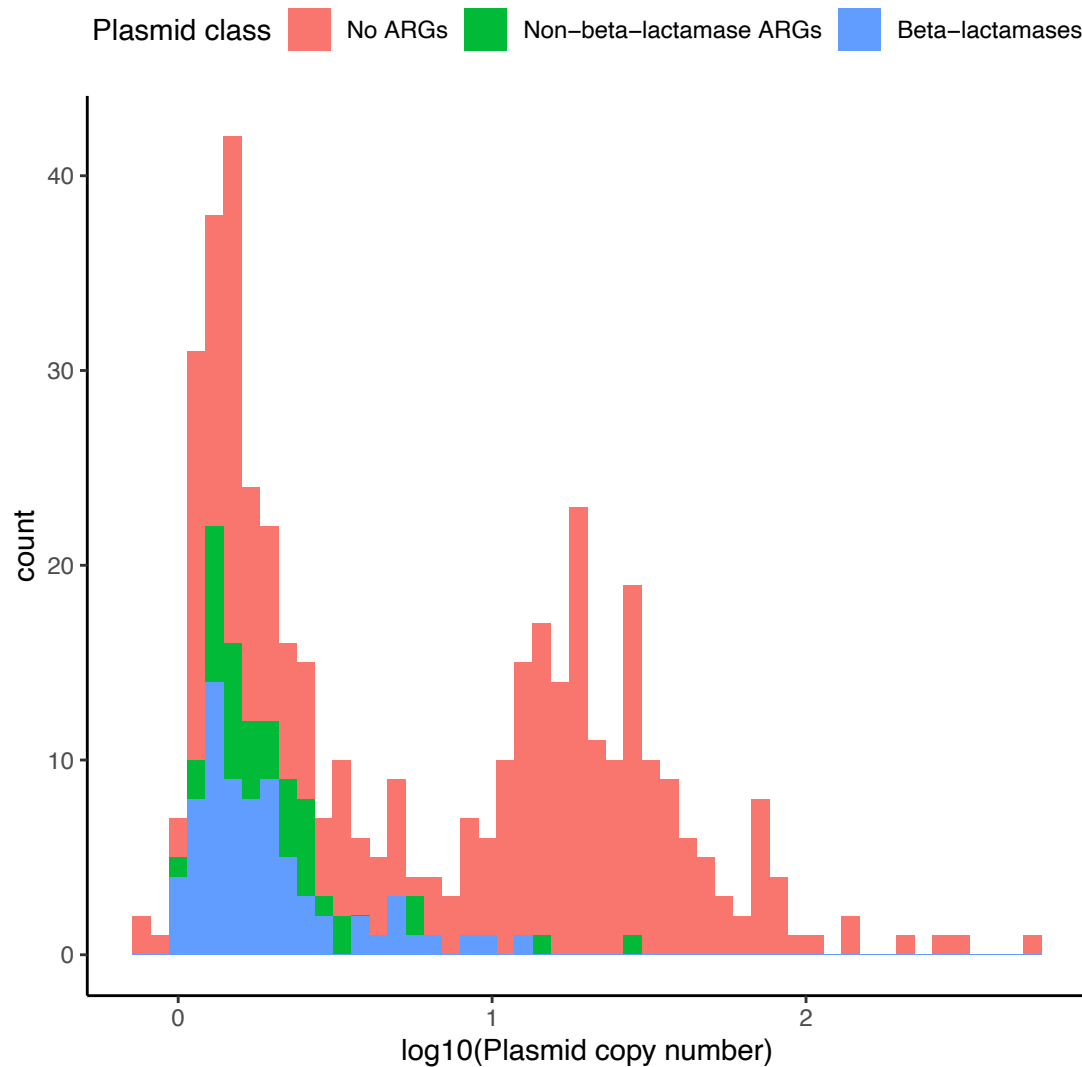
Distribution of duplicated genes in 114 genomes from an Australian ICU (Hawkey et al. 2022)

Corresponding distribution of single-copy genes

Category ARG MGE Other function

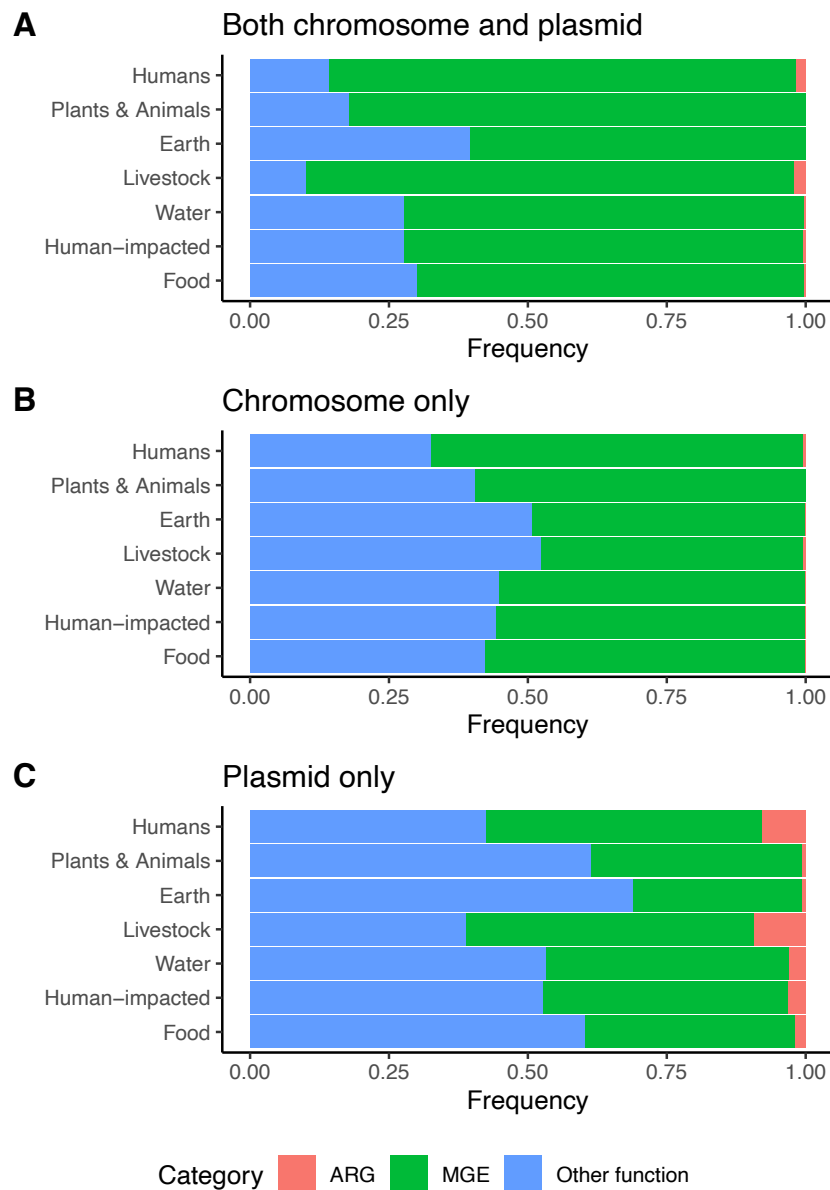


Supplementary Figure 10. Plasmid copy number relative to chromosome among 114 clinical antibiotic-resistant isolates from an Australian ICU. Complete genome assemblies were downloaded from NCBI BioProject PRJNA646837. Plasmid copy number relative to chromosome was measured by dividing the average Illumina short-read sequencing coverage per plasmid by the average Illumina short-read sequencing coverage per chromosome in each genome assembly.

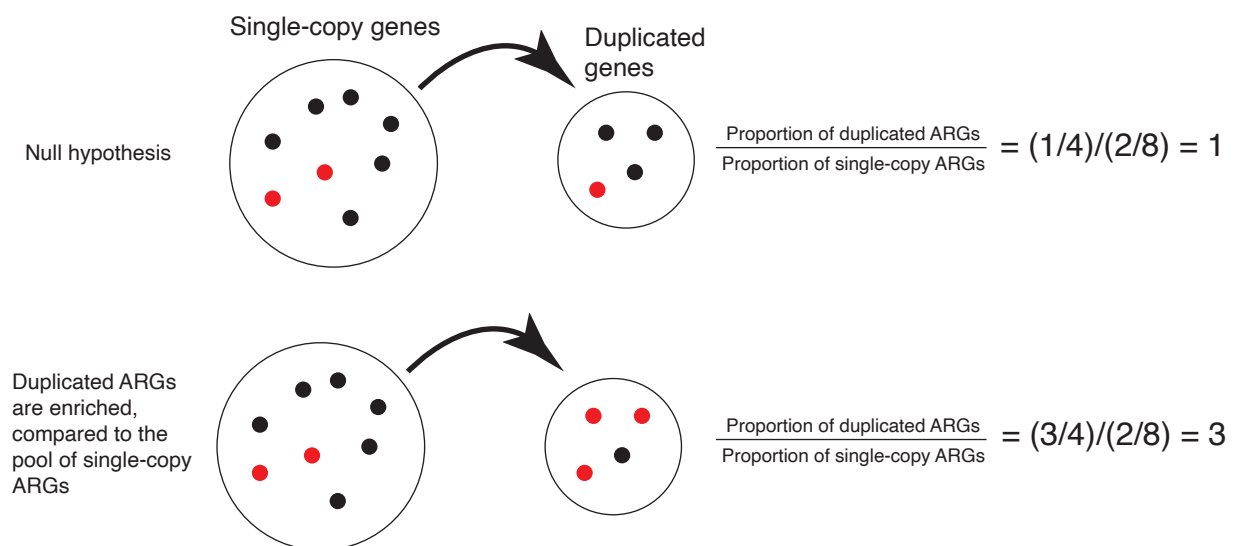


Supplementary Figure 11. Genomic distribution of duplicated genes (D-genes). D-genes encoded solely on plasmids are more likely to encode functions other than those associated with mobile genetic elements, in comparison to both D-genes encoded solely on the chromosome, and D-genes encoded on plasmids and the chromosome.

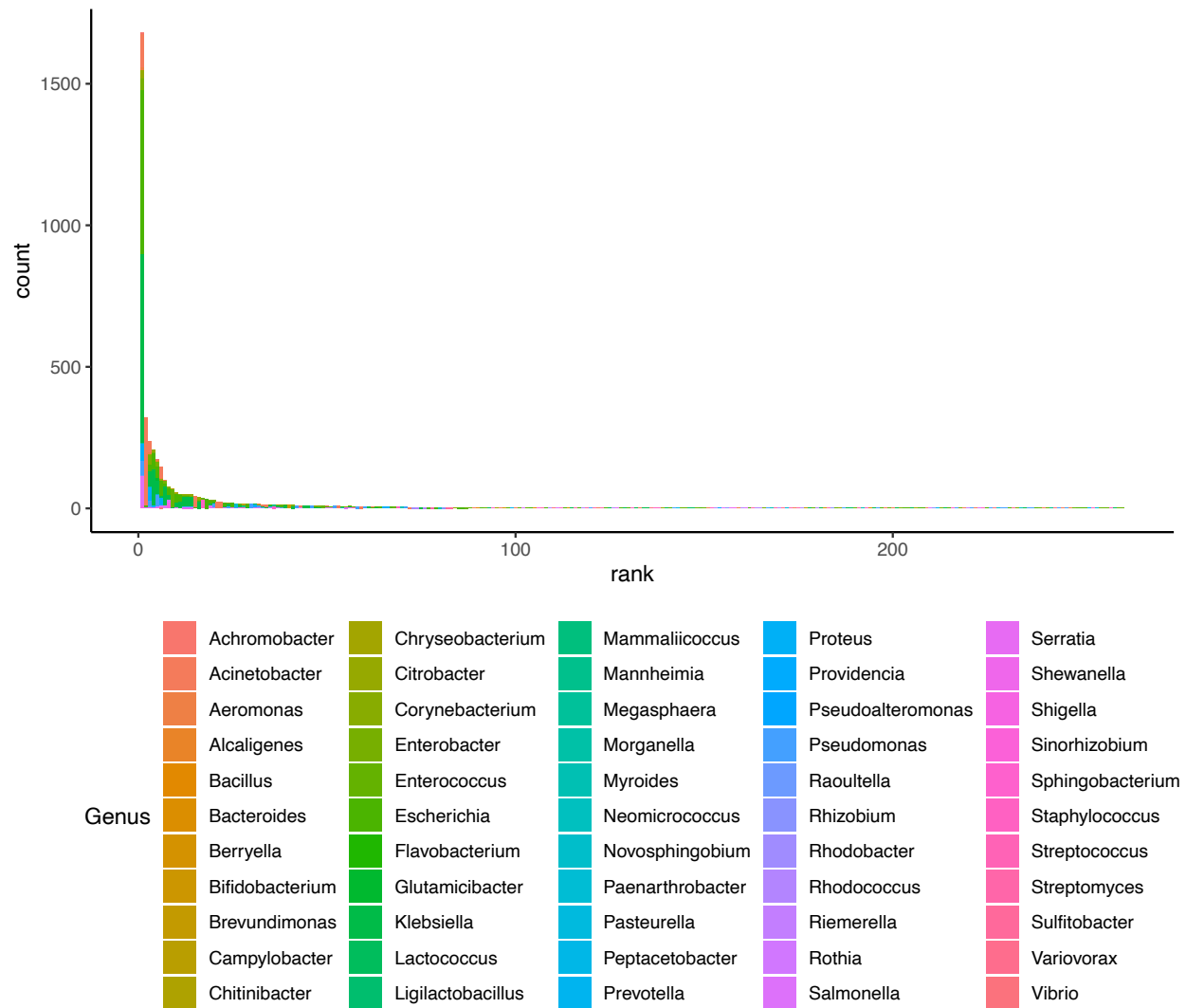
Genomic distribution of D-genes



Supplementary Figure 12. Non-random evolutionary forces such as natural selection and horizontal gene transfer drive the ecological distributions of duplicated ARGs and MGE-associated genes away from equilibrium. Suppose no evolutionary forces such as selection, horizontal gene transfer, or associations with mobile genetic elements affect the probability that a gene undergoes gene duplication. In this scenario, the probability that a gene of a given functional class is duplicated should be proportional to the number of single-copy genes of that functional class in the ecological category. Deviations from this null expectation indicates that the frequency of duplicated genes is being driven away from this equilibrium by evolutionary forces, such as selection and HGT.



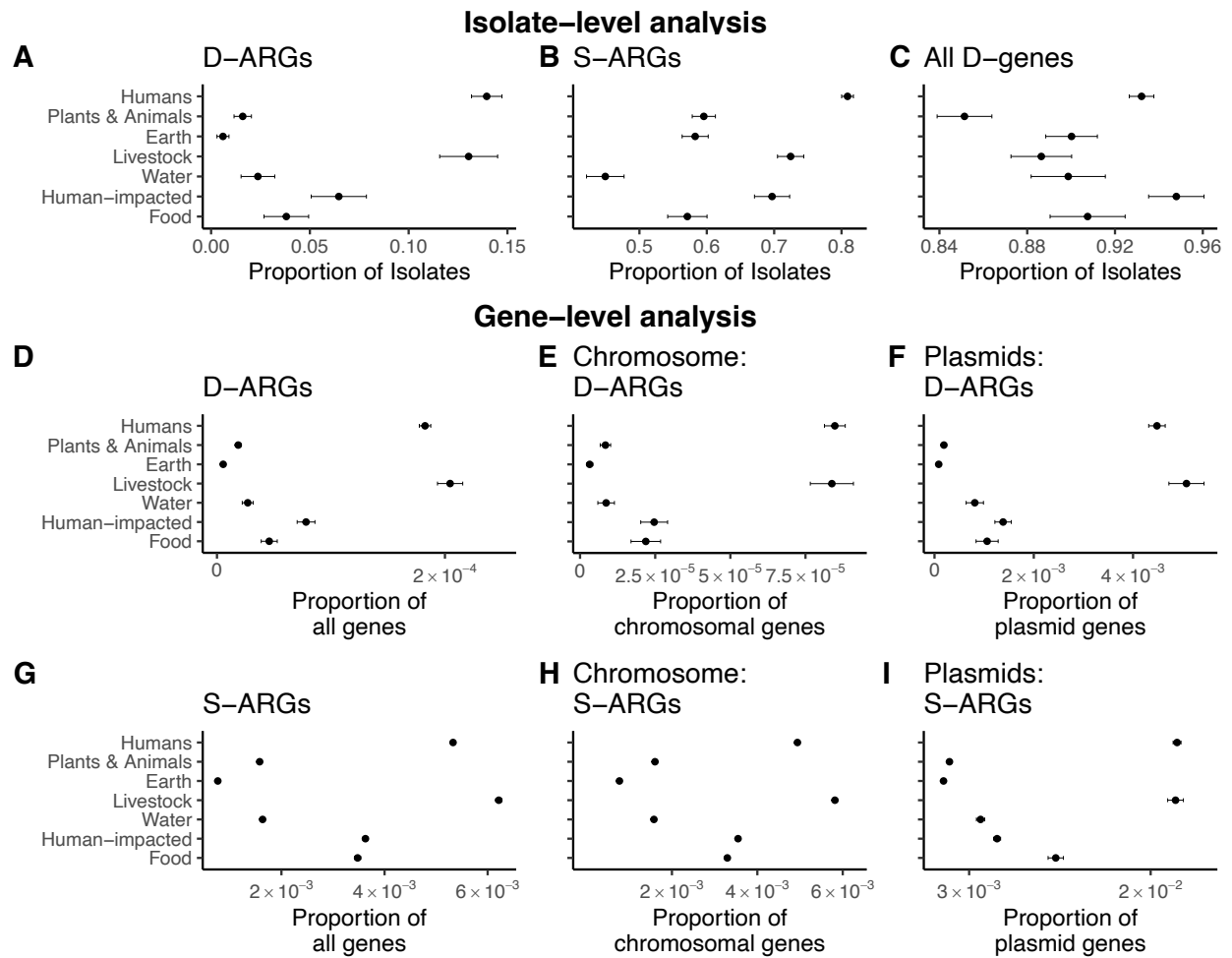
Supplementary Figure 13. Full distribution of transposases associated with ARGs in regions of consecutive duplicated genes.



Supplementary Figure 14. Bacteria isolated from humans and livestock are much more likely to have duplicated antibiotic resistance genes (D-ARGs) compared to bacteria isolated from other environments; furthermore, D-ARGs are enriched on the chromosomes and plasmids of bacteria isolated from humans and livestock (using CARD to annotate ARGs and mobileOG-db to annotate MGE-associated genes). Error bars are 95% binomial proportion confidence intervals, calculated using the formula $p \pm$

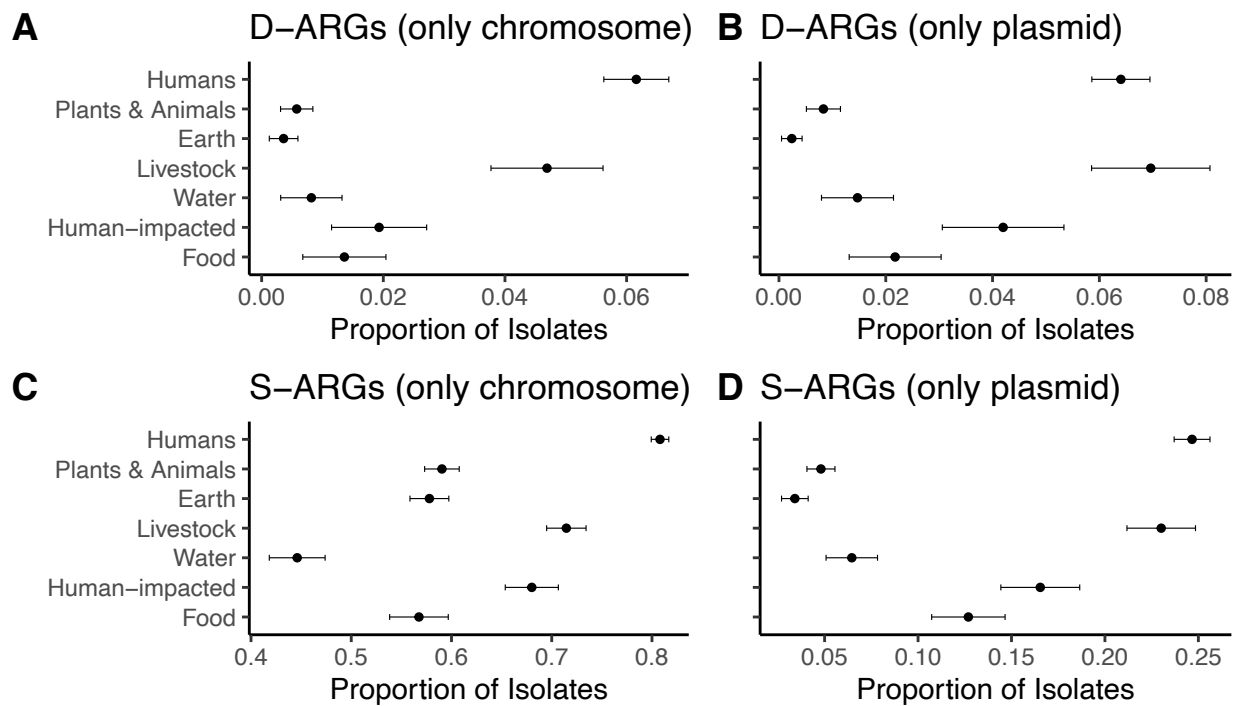
$Z_{\alpha/2} \sqrt{\left(\frac{p(1-p)}{n}\right)}$, where p is the proportion, n is the sample size, and $Z_{\alpha/2} = 1.96$.

- A. D-ARGs are specifically enriched in bacterial isolates from humans and livestock.
- B. The vast majority of isolates contain at least one single-copy antibiotic resistance gene (S-ARG).
- C. The vast majority of isolates contain at least one duplicated gene (D-gene).
- D. D-ARGs represent a higher fraction of genes found in bacteria isolated from humans and livestock compared to bacteria in the other ecological categories.
- E. Chromosomal D-ARGs are enriched in bacteria isolated from humans and livestock.
- F. Plasmid D-ARGs are enriched in bacteria isolated from humans and livestock.
- G. S-ARGs represent a higher fraction of genes found in bacteria isolated from humans and livestock compared to the other ecological categories.
- H. Chromosomal S-ARGs are enriched in humans and livestock.
- I. Plasmid S-ARGs are enriched in bacteria isolated from humans and livestock.



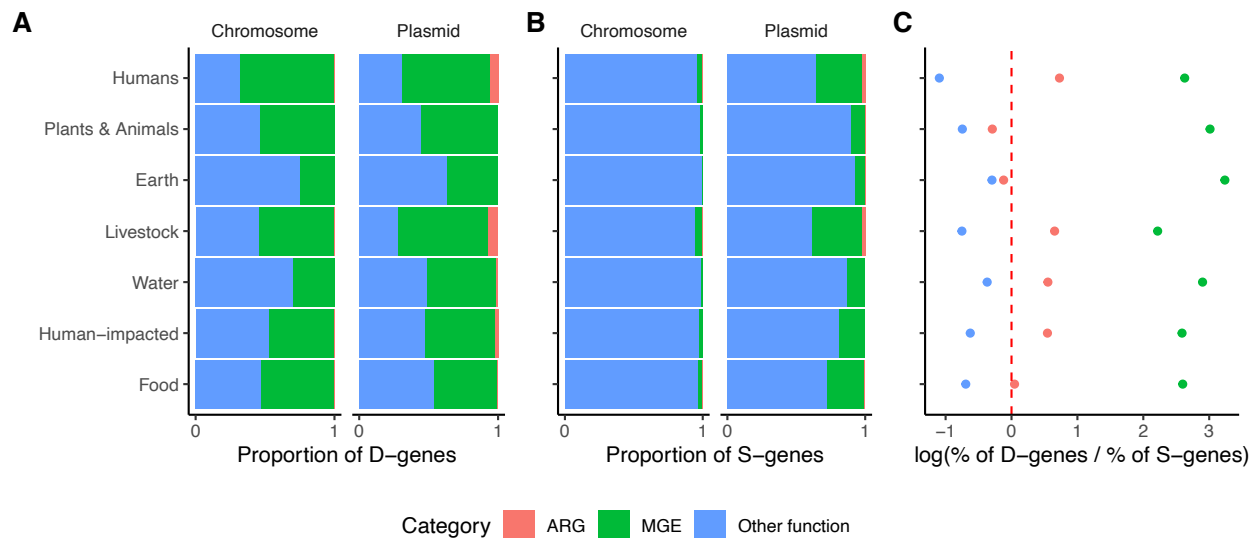
Supplementary Figure 15. D-ARGs are enriched on the chromosomes and plasmids of bacteria isolated from humans and livestock (using CARD to annotate ARGs and mobileOG-db to annotate MGE-associated genes). Error bars are 95% binomial proportion confidence intervals, calculated using the formula $p \pm Z_{\alpha/2} \sqrt{\left(\frac{p(1-p)}{n}\right)}$, where p is the proportion, n is the sample size, and $Z_{\alpha/2} = 1.96$.

- A. Bacterial isolates with D-ARGs on chromosomes are specifically associated with humans and livestock.
- B. Bacterial isolates with D-ARGs on plasmids are specifically associated with humans and livestock.
- C. Most isolates contain at least one single-copy ARG on their chromosome.
- D. Bacterial isolates with S-ARGs on plasmids are specifically associated with humans, livestock, and human-impacted environments.



Supplementary Figure 16. Selection, horizontal gene transfer, and mobile genetic elements shape the ecological distribution of duplicated genes (using CARD to annotate ARGs and mobileOG-db to annotate MGE-associated genes). Proteins associated with mobile genetic elements (MGEs) are shown in green; proteins encoded by antibiotic resistance genes (ARGs) are in red; and all other proteins are shown in blue.

- A. Across all ecological categories, ~50% duplicated genes (D-genes) on chromosomes and plasmids are associated with MGEs.
 - B. MGE-associated proteins account for <10% of single-copy genes (S-genes) on chromosomes, and 5-25% of S-genes on plasmids.
 - C. Duplicated ARGs (D-ARGs) are enriched in humans, livestock, water, and human-impacted environments, and are depleted in isolates from plants and animals, while duplicated genes associated with mobile genetic element functions are enriched in all ecological categories.
- The red dashed line indicates the null hypothesis.



Supplementary Data Files

File Name: Supplementary Data 1

Description: Pluto computational notebook, written in the Julia programming language, allowing for user interaction with the deterministic mathematical model.

File Name: Supplementary Data 2

Description: Mutations found in the evolution experiments.

File Name: Supplementary Data 3

Description: Ecologically annotated complete genomes, annotated by presence and absence of duplicated ARGs.

File Name: Supplementary Data 4

Description: Duplicated ARGs found in the Complete Genomes from NCBI RefSeq.

File Name: Supplementary Data 5

Description: MOBtyper annotations of plasmid found in the Complete Genomes from NCBI RefSeq.

File Name: Source Data

Description: Source Data for main figures.