Supplementary Material

Class 1 Integrons and Carbapenemase Genes Distinguish Hospital Effluent Across Urban Wastewater Sources

---

Contents: Supplementary Methods, Supplementary Results, Supplementary Figures S1-S5, Supplementary Tables S1-S17, Supplementary References.

Supplementary Methods

Sample collection and biomass concentration. Wastewater samples were kept on ice for transport, arriving at the laboratory within six hours of collection and held at 4°C until processing. Biomass recovered by vacuum-filtration (0.22 µm membrane filters, Millipore, USA) was weighed gravimetrically before being transferred into sterile cryovials for storage at −80°C.

DNA extraction and quality assurance. Genomic DNA was extracted from filters using the PureLink™ Microbiome DNA Purification Kit (Invitrogen, Thermo Fisher Scientific, USA), combining bead-beating with enzymatic and chemical lysis before silica-membrane spin-column purification, eluting into low-salt buffer for storage at −20°C. NanoDrop spectrophotometry confirmed yield and purity against target ratios of A260/A280 1.8–2.0 and A260/A230 2.0–2.2. Extract integrity was additionally checked on 1% agarose gels (TAE buffer, ethidium bromide); only extracts with high-molecular-weight bands above 10 kb and minimal smearing proceeded to library preparation.

Bioinformatics pipeline and host decontamination. Raw read quality was assessed with FastQC v0.11.9 (Andrews, 2010), then adapters trimmed using Trimmomatic v0.39 (Bolger et al., 2014) under the settings ILLUMINACLIP:adapters:2:30:10, SLIDINGWINDOW:4:20, and MINLEN:50, retaining only reads with Phred scores above Q20. Human-derived reads were removed with KneadData v0.10.0, running Bowtie2 v2.4.5 (Langmead & Salzberg, 2012) in very-sensitive mode against the GRCh38/hg38 reference; unmapped reads carried forward into assembly.

Metagenomic assembly and gene catalog construction. Contigs were assembled from host-depleted reads in MEGAHIT v1.2.9 (Li et al., 2015) under its --meta-sensitive preset with default k-mer sizes (21, 29, 39, 59, 79, 99, 119, 141), and assembly quality checked with QUAST v5.2.0 (Gurevich et al., 2013). Gene calling used Prodigal v2.6.3 (Hyatt et al., 2010) in metagenomic mode (-p meta); predicted genes were clustered at 95% identity with CD-HIT v4.8.1 (Li & Godzik, 2006) to build a non-redundant gene catalogue, and Salmon v1.9.0 (Patro et al., 2017) quantified gene-level abundance as TPM.

Taxonomic profiling parameters. Kraken2 v2.1.2 (Wood et al., 2019; confidence threshold 0.05) classified reads against a custom RefSeq database covering bacteria, archaea, viruses, and the human genome; Bracken v2.6 (Lu et al., 2017; 150 bp read length, k-mer threshold 10) refined these to species-level estimates. Assignments were cross-checked using an LCA approach in DIAMOND v2.0.15 (Buchfink et al., 2015; Kahlke & Ralph, 2019).

Resistome profiling parameters. ARG identification used DIAMOND v2.0.15 (Alcock et al., 2023) alignment against CARD v3.2.7 with thresholds of ≥80% sequence identity, ≥70% query coverage, and E-value ≤ 1×10^{-5}.

Data visualisation software. Hierarchically clustered heatmaps were generated using BioinCloud's visualisation module together with the pheatmap v1.0.12 package; bar

charts and composition plots used ggplot2 v3.4.0 (R v4.2.0); circos plots used circlize v0.4.15 (Gu et al., 2014); Matplotlib v3.6.0 and seaborn v0.12.0 (Waskom, 2021) in Python v3.9 handled further figure refinement. Complete software versions are listed in Table S2.

Supplementary Results

Kingdom- and phylum-level composition. Bacteria accounted for 95.87–99.62% of classified reads across all six samples, the dominant kingdom throughout. Archaea formed the next-largest fraction (0.34% in SLW2 up to 3.90% in CW2), while viral sequences, pooled across seven viral phyla, reached only 0.01–0.57% per sample, and fungal reads were rarer still (0.009–0.073%). Hospital wastewater carried the lowest bacterial share of any sector (mean 97.35%), against a high of 98.94% in slaughterhouse samples. These kingdom-level proportions held steady between T1 and T2 within each sector. Within Bacteria, *Pseudomonadota* was the dominant phylum in all six samples, reaching its highest levels in community wastewater (CW1 : 73.45%, CW2 : 75.92%) and its lowest in slaughterhouse samples (SLW1 : 36.00%, SLW2 : 53.02%). *Campylobacterota* showed the largest single-phylum deviation among non-dominant taxa, reaching 17.11% in HW1 . *Bacteroidota* and *Bacillota* together defined the slaughterhouse profile: *Bacteroidota* peaked at 39.44% in SLW2 , and *Bacillota* at 24.90% in SLW1 . *Actinomycetota* ranged 3.27–9.33% without any sector-specific pattern; together these four phyla made up 71.58–93.78% of classified reads per sample. *Thermodesulfobacteriota* rose to 6.57% in HW2 , pulling the hospital sector mean to 3.52% against 0.58% in community samples. *Fusobacteriota* reached appreciable abundance only in SLW1 (2.35%), staying at trace levels in every other sample. Neither Kruskal-Wallis nor ANCOM found significant inter-sector differences after Bonferroni

correction, consistent with the limited statistical power given two timepoints per sector rather than independent spatial replicates.

Genus-level diversity and clinically relevant taxa. Genus-level profiling picked out clear sector-specific signatures. *Thauera* dominated hospital wastewater (HW mean : 24.34%), reaching 32.44% in HW2 . *Aliarcobacter* made up 15.71% of HW1 reads, the single highest abundance recorded for any clinically relevant genus in this study. *Acinetobacter* was significantly more abundant in slaughterhouse than hospital wastewater (SLW mean : 8.08% vs HW mean : 0.64%; Dunn adjusted p = 0.049). *Enterococcus* reached 4.08% in SLW1 , *Escherichia* 2.33%; *Flavobacterium* peaked at 18.95% in SLW2 and *Acetoanaerobium* at 13.80% in SLW1 . *Streptococcus* stayed within 0.10–2.66% across all sectors, and *Salmonella* within 0.01–0.43% per sample. *Vibrio* made up 71.06% of CW2 's classified reads against just 0.03% in CW1 , an episodic bloom confined to a single timepoint. *Campylobacter* itself reached only 0.01–0.02% despite *Campylobacterota* comprising 17.11% of HW1 reads; *Aliarcobacter* was the actual driver behind that phylum-level signal. *Bifidobacterium* was significantly more abundant in hospital than slaughterhouse wastewater (HW mean : 1.65% vs SLW mean : 0.42%; Dunn adjusted p = 0.049). Thirty bacterial genera reached significance after Bonferroni correction, though none met the ANCOM rejection threshold; the top 19 named genera together accounted for 51.99–78.64% of classified reads per sample.

Alpha and beta diversity. Shannon-index alpha diversity ranked sectors as slaughterhouse highest (SLW mean : 4.56; range 4.52–4.60), hospital intermediate (HW mean : 4.10; range 3.93–4.26), and community lowest (CW mean : 4.03; range 3.50–4.57); Simpson evenness followed the identical order (SLW : 0.968, HW : 0.923, CW : 0.915). The community sector's wide range traces to a single timepoint: Shannon diversity collapsed to 3.50 in CW2 , coinciding with the *Vibrio* bloom there. Bray-Curtis PCoA captured 64.5% of total variance (Axis 1: 42.1%; Axis 2: 22.4%), yet PERMANOVA found no significant sector-based structuring ($R^2 = 0.374$, $F = 0.897$, $p = 0.667$). Temporal distances within sectors varied widely: HW1 –HW2 = 0.501, SLW1 –SLW2 = 0.798, CW1 –CW2 = 0.913; in both community and slaughterhouse

sectors, that within-sector distance met or exceeded the 0.723 inter-sector mean, meaning temporal turnover rivalled cross-sector divergence in scale. NMDS gave the same ordination profile (stress = 0.000, a perfect 2D fit at n = 6).

Shared, unique, and differentially abundant taxa. We detected 6,272 bacterial species across the three sectors combined; 2,865 of these (45.7%) formed a core microbiome common to all three wastewater sources. Sector-specific richness was greatest in community wastewater (894 unique species), then slaughterhouse (632) and hospital (331). Looking at pairwise overlaps, community and slaughterhouse shared 1,069 species exclusively, 3.5 times the community–hospital overlap (302) and 6.0 times the hospital–slaughterhouse overlap (179). Total species counts followed a different order: hospital wastewater held the smallest pool (3,677), below slaughterhouse (4,745) and community (5,130). The community–slaughterhouse pairwise overlap was the largest of the three exclusive-sharing categories, making up 17.0% of all species detected. LEfSe picked out 324 genus-level biomarkers across the three sectors at LDA > 2.0: 117 enriched in hospital, 106 in community, and 101 in slaughterhouse wastewater. Hospital wastewater's top discriminators were *Thauera* (LDA = 5.45) and *Aliarcobacter* (4.96), followed by *Diaphorobacter* (4.77), *Acidovorax* (4.59), and *Desulfovibrio* (4.38). *Flavobacterium* (LDA = 5.00), *Acinetobacter* (4.94), *Acetoanaerobium* (4.89), and *Cloacibacterium* (4.83) characterised slaughterhouse wastewater, with *Enterococcus* (4.35), *Escherichia* (4.16), and *Fusobacterium* (4.11) as further significant biomarkers there. Community wastewater's strongest single discriminator across all sectors was *Vibrio* (LDA = 5.57), followed by *Pseudoxanthomonas* (4.71), *Pseudomonas* (4.70), and *Mycolicibacterium* (4.30), with *Malaciobacter* (4.22) and *Stenotrophomonas* (4.03) rounding out that profile. *Bifidobacterium* reached LDA = 4.28 as a hospital biomarker, in line with its Dunn-significant enrichment in HW over SLW .

MGE richness, abundance, and Level 1 type composition. Across the six Peshawar wastewater samples we identified 37 distinct MGE types spanning transposases, insertion sequences, integrons, conjugative transposons, and plasmid replicons. Total MGE abundance ranged from 2.075 CPC in SLW1 up to 8.338 CPC in HW1 , and hospital

wastewater carried the heaviest sector-mean burden overall (HW mean : 7.775 CPC), 1.85-fold above community (4.202 CPC) and 3.34-fold above slaughterhouse (2.329 CPC). Community wastewater told a different story between its two timepoints: CW1 (6.237 CPC) nearly tripled CW2 (2.168 CPC), an inverse pattern to the ARG spike seen at CW2 , suggesting the *Vibrio*-dominated community carried a proportionally lighter mobilome load. Hospital samples, by contrast, barely moved between timepoints (HW1 : 8.338 CPC; HW2 : 7.213 CPC; 1.16-fold variation). Transposases were the largest single contributor to the mobilome at every site (72.7% of total MGE abundance, 17.972 CPC across six samples), ahead of unclassified elements (10.2%), integrases (7.7%), and *qacE·1* (5.7%).

Family-level and site-specific MGE patterns. Transposases were the leading MGE class in every sector, with *tnpA* variants making up most of that signal and hospital wastewater carrying the heaviest load (HW mean : 4.843 CPC, versus 2.623 in community and 1.520 in slaughterhouse); the largest hospital-associated jump among individual insertion sequences belonged to IS26 (HW mean : 0.074 CPC, 9.6-fold over slaughterhouse). IS91-family insertion sequences came second (9.6%; 2.753 CPC total), spread broadly across sectors. Integrases, chiefly *intI1*, ranked third (7.7%; 2.201 CPC) and were most concentrated in hospital wastewater (HW mean : 0.726 CPC, 5.9-fold above slaughterhouse), as were the transposition genes *tniA* and *tniB*, which mobilise class 1 integrons (*tniA* HW mean : 0.465 CPC; *tniB*: 0.284 CPC, against 0.084 and 0.052 CPC in slaughterhouse). *qacE·1* followed the same hospital-skewed pattern (5.7%; HW mean 0.507 CPC versus SLW mean 0.114 CPC). ISCR elements made up a further 1.8%, with Tn916 and plasmid replicons (IncQ1, IncQ2, IncP(6), ColKP3) each under 1%. Within community wastewater, CW1 outpaced CW2 across every MGE class (6.237 versus 2.168 CPC total), an inversion of the ARG pattern at that timepoint. Testing 1,236 individual MGE features by Kruskal-Wallis found no significant inter-sector differences after correction (minimum raw $p = 0.091$), mirroring the ARG result.

Temporal patterns in MGE profiles. Bray-Curtis similarity applied to MGE Level 2 profiles across the 30-day sampling window revealed sharply different temporal stability

by sector. Hospital wastewater was the most stable (HW1 vs HW2 : BC = 0.911), slaughterhouse close behind (SLW1 vs SLW2 : 0.856), while community wastewater was far less stable (CW1 vs CW2 : BC = 0.469), its own within-sector, between-timepoint dissimilarity actually exceeded the between-sector dissimilarity of the hospital-slaughterhouse pair (0.547). The shift within CW between timepoints traced mainly to IS91-family insertion sequences (a 6.95-fold drop, CW1 : 0.988 CPC to CW2 : 0.142 CPC) and *qacE·1* (5.93-fold), both class 1 integron-associated elements, alongside the *intI1* integrase itself (3.84-fold drop), pointing to a real decline in integron-bearing populations at the second community timepoint. Total MGE abundance told the same story: hospital wastewater shifted only 1.16-fold between timepoints (HW1 : 8.338 CPC; HW2 : 7.213 CPC), while community wastewater dropped 2.88-fold (CW1 : 6.237 CPC; CW2 : 2.168 CPC). No PERMANOVA was run on the MGE compositional data; Bray-Curtis values are reported here purely as descriptive indices of temporal stability, given insufficient replication for formal hypothesis testing.

Supplementary Figures

Figure S1. Bacterial alpha diversity across the six Peshawar wastewater samples. (A) Shannon index per sample, coloured by sector; (B) Simpson evenness (1 − D) sector means. Values as reported in Section 3.1.4.

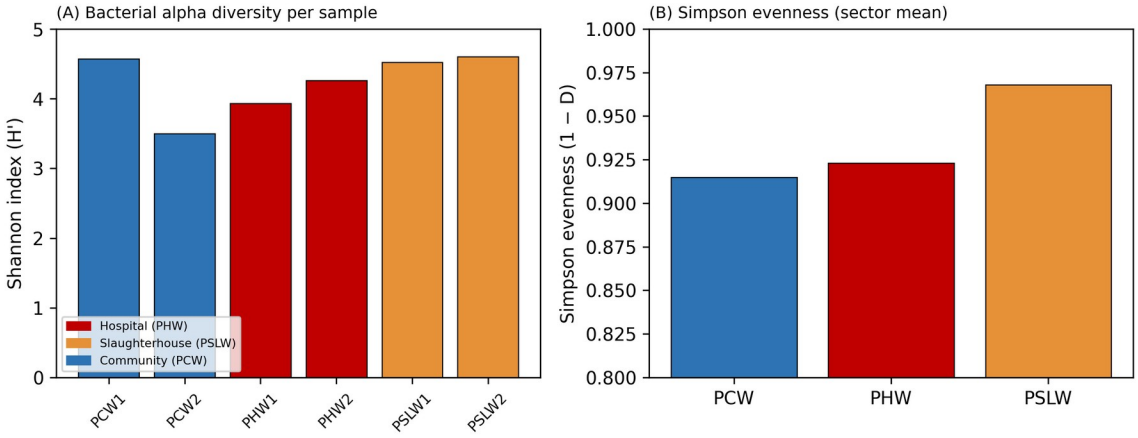


Figure S2. LEfSe biomarker analysis of antibiotic resistance genes. Linear discriminant analysis (LDA) bar chart of ARG-level biomarkers (LDA score > 2.0); gene names labelled at each bar; bars colour-coded by enriched sector.

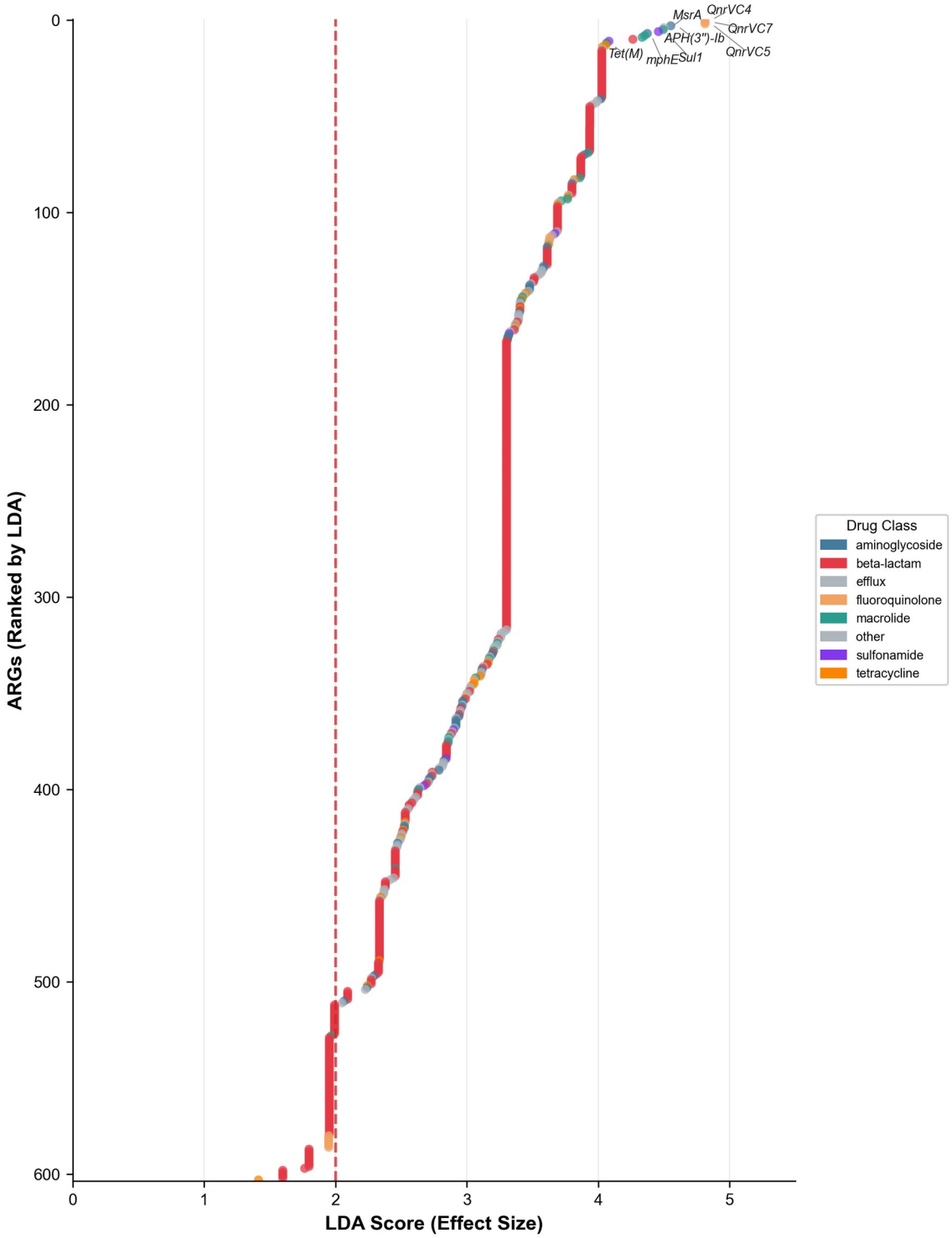


Figure S3. Sector-specific MGE family abundance patterns. (A) Kruskal-Wallis horizontal column plots for the top MGE Level 2 families by mean abundance difference across sectors; none reached significance after correction. (B) ANOVA horizontal column plots for the same comparison; error bars represent standard error; none reached significance after correction.

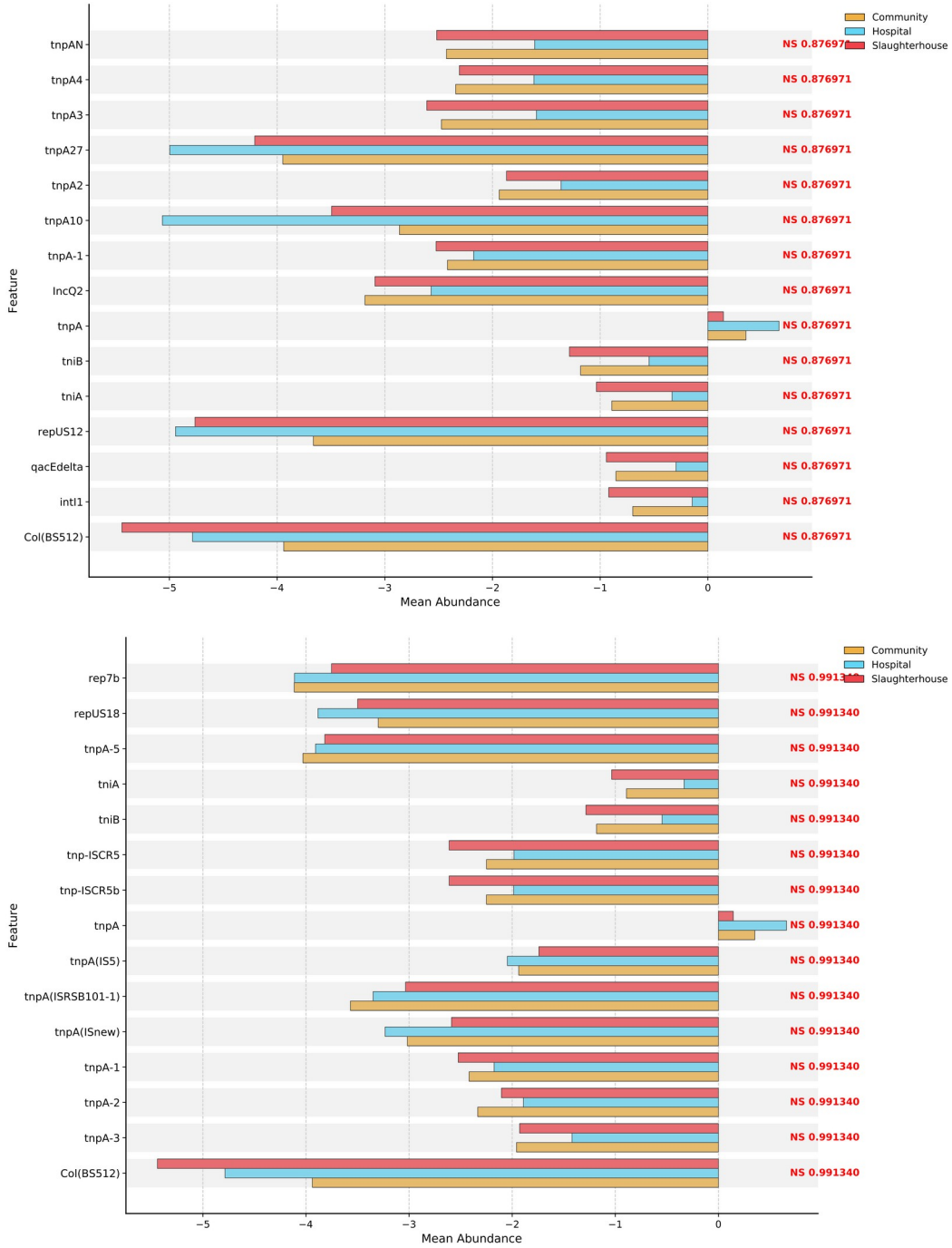


Figure S4. Genus-level compositional profiles across the six Peshawar wastewater samples. (A) Clustered genus-level heatmap (Z-score normalised) with hierarchical dendrogram showing the top bacterial genera. (B) LEfSe linear discriminant analysis (LDA) bar chart of genus-level biomarkers (LDA score > 2.0); bars colour-coded by enriched sector (HW = hospital; SLW = slaughterhouse; CW = community).

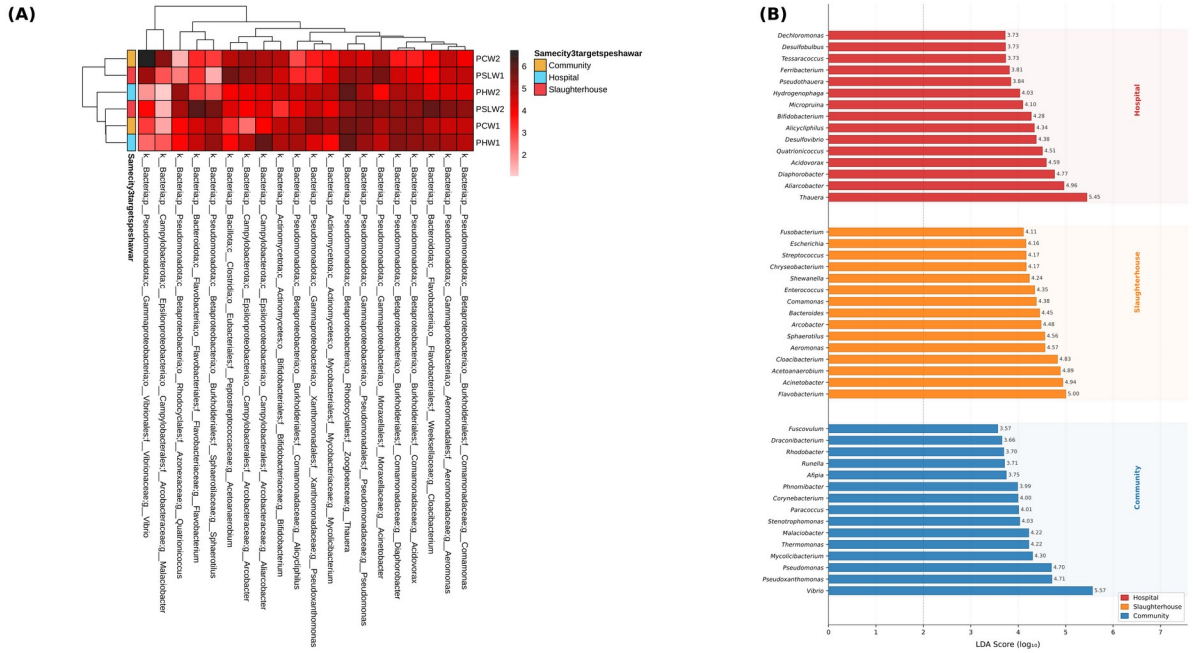


Figure S5. MGE Level 2 family-level composition and temporal dynamics across the 30-day sampling interval. (A) Clustered heatmap of MGE family abundances (TPM) across all six samples, with hierarchical dendrogram; colour scale reflects log-transformed abundance. (B) Principal coordinates analysis (PCoA) of Bray-Curtis dissimilarities in MGE profiles; samples labelled by sector and timepoint (T1, T2). (C) Slope plot showing changes in the top five MGE families between timepoints T1 and T2 per sector; connecting lines indicate direction and magnitude of temporal change.

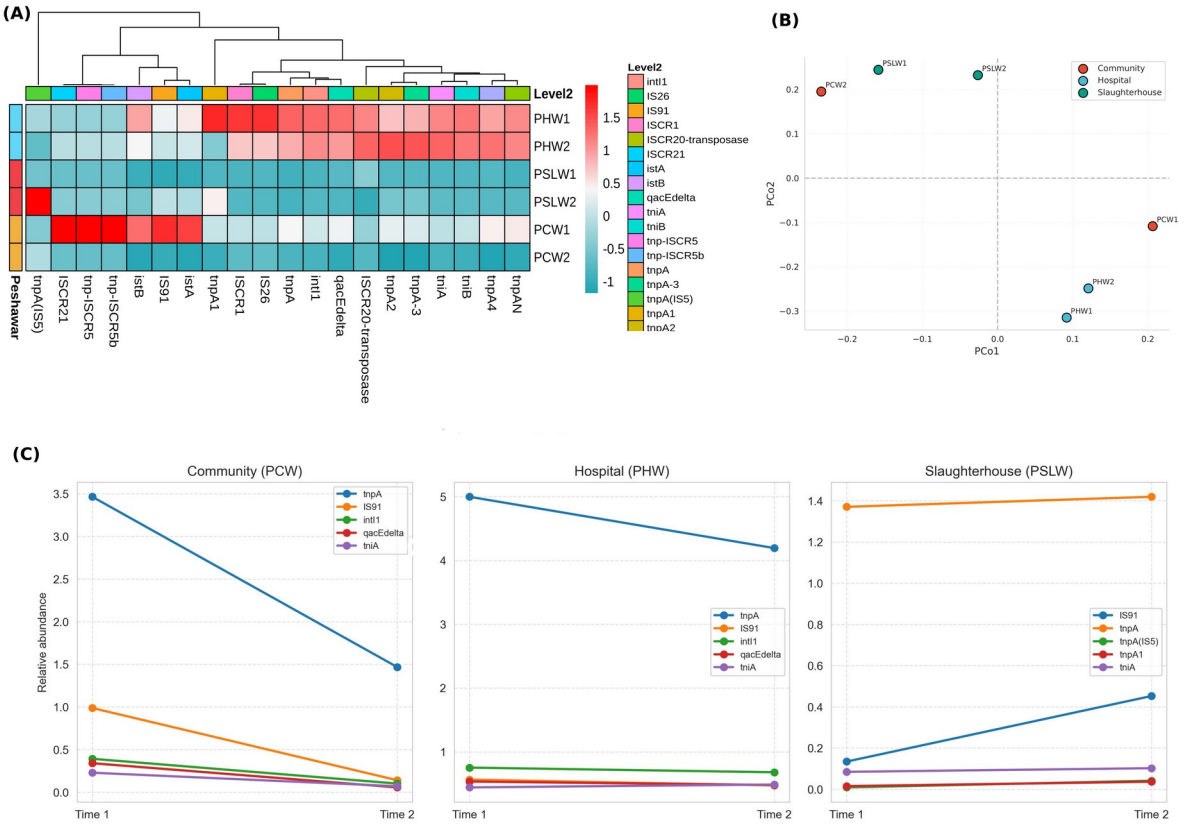


Figure S6. Circos plot of the top 30 antibiotic resistance genes (ARGs) by abundance across the six Peshawar wastewater samples. Ribbon thickness is proportional to TPM and ribbons are colour-coded by sector. Previously presented as Fig. 2A.

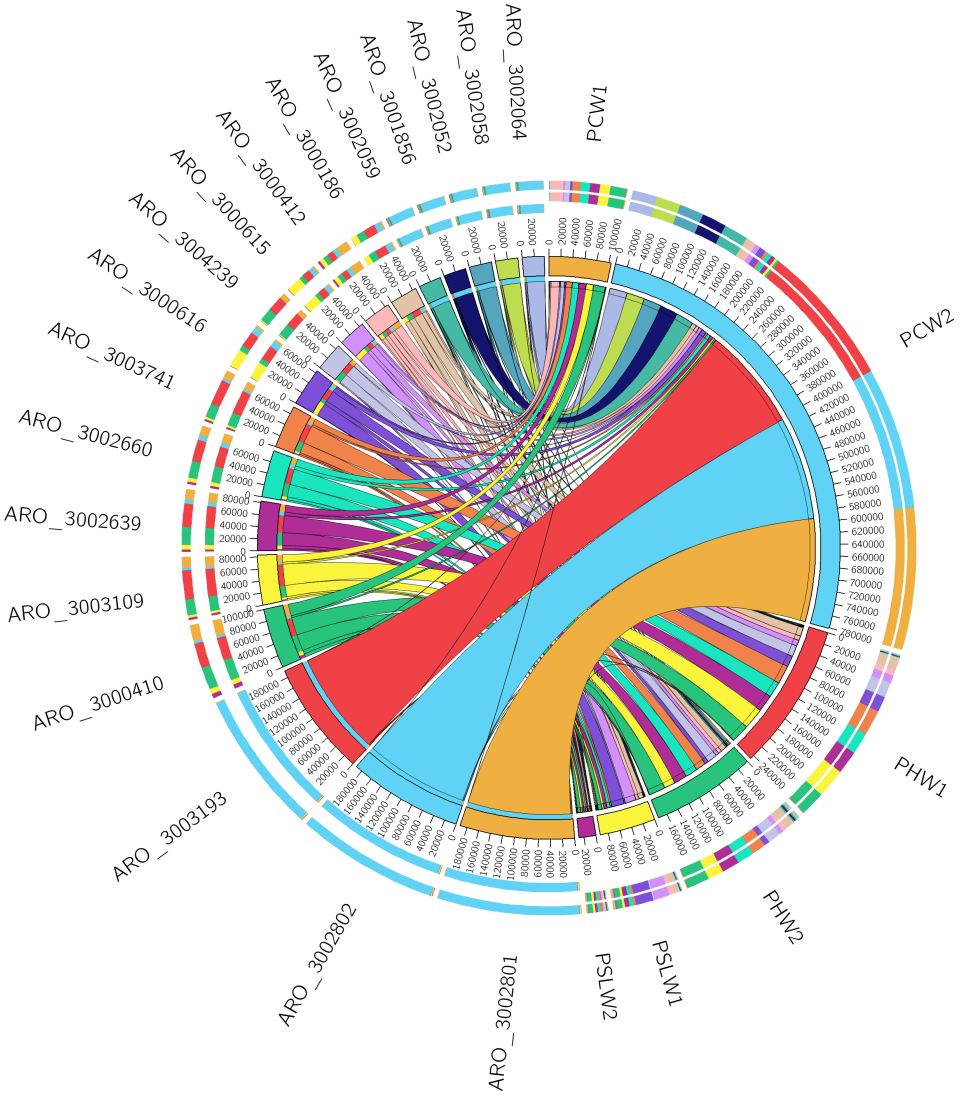
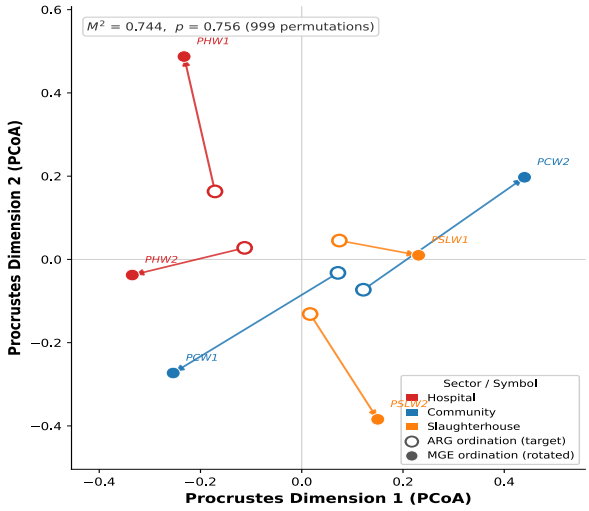


Figure S7. Procrustes superimposition of the ARG and MGE community ordinations ($m^2 = 0.744$; $p = 0.756$; 999 permutations). Open circles mark each sample's position in the ARG ordination (target), filled circles its position in the rotated MGE ordination, and arrows connect the two positions for each sample, colour-coded by sector (hospital, community, slaughterhouse). Previously presented as Fig. 4C.



Supplementary Tables

Table S1. Sequencing statistics for the six Peshawar wastewater samples (Illumina NovaSeq 6000, 2 × 150 bp). Raw and quality-filtered read counts, bases, GC content, and Q20/Q30 scores.

Sample	Raw reads	Raw bases (Gb)	%GC	Raw Q20 (%)	Raw Q30 (%)	Clean reads	Clean (%)	Clean Q20 (%)	Clean Q30 (%)
HW1	21,189,547	6.36	53	98.93	96.59	20,479,223	96.65	99.44	97.79
HW2	22,555,936	6.77	55	98.84	96.42	21,771,904	96.52	99.43	97.77
SLW1	23,539,594	7.06	46	99.08	97.04	22,881,984	97.21	99.51	98.07
SLW2	29,218,612	8.77	47	98.99	96.84	28,299,888	96.86	99.50	98.03
CW1	24,357,273	7.31	59	98.78	96.26	23,419,941	96.15	99.41	97.70
CW2	25,000,696	7.50	46	99.07	97.04	24,327,018	97.31	99.52	98.08

Table S2. Bioinformatics software, databases, and versions used in this study.

Tool / database	Version	Purpose
FastQC	v0.11.9	Raw read quality control
Trimmomatic	v0.39	Adapter and quality trimming
KneadData	v0.10.0	Host read removal
Bowtie2	v2.4.5	Host mapping (GRCh38/hg38)
MEGAHIT	v1.2.9	Metagenome assembly
QUAST	v5.2.0	Assembly quality assessment
Prodigal	v2.6.3	ORF prediction
CD-HIT	v4.8.1	Gene clustering (95% identity)
Salmon	v1.9.0	Gene abundance quantification (TPM)
Kraken2	v2.1.2	Taxonomic classification
Bracken	v2.6	Species abundance estimation
DIAMOND	v2.0.15	Protein alignment (CARD)
CARD	v3.2.7	Antibiotic resistance gene database
vegan	v2.6-4	Diversity and ordination
pheatmap	v1.0.12	Heatmaps
ggplot2	v3.4.0	Plotting
circlize	v0.4.15	Circos plots
matplotlib	v3.6.0	Plotting
seaborn	v0.12.0	Plotting
R	v4.2.0	Statistical computing
Python	v3.9	Analysis
Wekemo BioinCloud	—	Integrated analysis platform

Table S3. Top 20 antibiotic resistance genes (ARGs) by total abundance. Gene name, drug class, resistance mechanism, and per-site abundance (copies per cell, CPC) for the 20 most abundant ARGs across all six Peshawar wastewater samples.

Rank	Gene	ARO ID	Drug Class	Resistance Mechanism	CW1	CW2	HW1	HW2	SLW 1	SLW 2	Total	Note
1	QnrVC4	ARO:3002801	Fluoroquinolone	Target protection	0.0004	0.1888	0.0003	0.0002	0.0015	0.0007	0.1920	*
2	QnrVC5	ARO:3002802	Fluoroquinolone	Target protection	0.0004	0.1888	0.0003	0.0002	0.0015	0.0007	0.1920	*
3	QnrVC7	ARO:3003193	Fluoroquinolone	Target protection	0.0004	0.1888	0.0003	0.0002	0.0015	0.0007	0.1920	*
4	Sul1	ARO:3000410	Sulfonamide	Target replacement	0.0215	0.0034	0.0328	0.0299	0.0071	0.0058	0.1004	
5	MsrE	ARO:3003109	Macrolide	Antibiotic efflux	0.0140	0.0032	0.0388	0.0208	0.0033	0.0032	0.0834	
6	APH(3")-Ib	ARO:3002639	Aminoglycoside	Antibiotic inactivation	0.0128	0.0059	0.0309	0.0237	0.0062	0.0030	0.0826	
7	APH(6)-Id	ARO:3002660	Aminoglycoside	Antibiotic inactivation	0.0126	0.0064	0.0288	0.0225	0.0052	0.0033	0.0787	
8	mphE	ARO:3003741	Macrolide	Antibiotic inactivation	0.0109	0.0032	0.0340	0.0187	0.0030	0.0027	0.0724	
9	Mel	ARO:3000616	Macrolide	Antibiotic efflux	0.0031	0.0090	0.0203	0.0071	0.0238	0.0007	0.0641	
10	NPS-1	ARO:3004239	Beta-lactam	Antibiotic inactivation	0.0068	0.0004	0.0204	0.0176	0.0017	0.0005	0.0473	
11	MsrA	ARO:3000615	Macrolide	Antibiotic efflux	0.0024	0.0062	0.0106	0.0058	0.0204	0.0009	0.0464	
12	Sul2	ARO:3000412	Sulfonamide	Target replacement	0.0155	0.0063	0.0086	0.0070	0.0060	0.0028	0.0463	
13	Tet(M)	ARO:3000186	Tetracycline	Target protection	0.0024	0.0081	0.0158	0.0108	0.0086	0.0001	0.0458	
14	CFE-1	ARO:3001856	Beta-lactam	Antibiotic inactivation	0.0002	0.0326	0.0009	0.0021	0.0007	0.0007	0.0372	
15	CMY-41	ARO:3002052	Beta-lactam	Antibiotic inactivation	0.0002	0.0326	0.0009	0.0021	0.0007	0.0007	0.0372	
16	CMY-47	ARO:3002058	Beta-lactam	Antibiotic inactivation	0.0002	0.0326	0.0009	0.0021	0.0007	0.0007	0.0372	
17	CMY-48	ARO:3002059	Beta-lactam	Antibiotic inactivation	0.0002	0.0326	0.0009	0.0021	0.0007	0.0007	0.0372	
18	CMY-53	ARO:3002064	Beta-lactam	Antibiotic inactivation	0.0002	0.0326	0.0009	0.0021	0.0007	0.0007	0.0372	
19	CMY-110	ARO:3002071	Beta-lactam	Antibiotic inactivation	0.0002	0.0326	0.0009	0.0021	0.0007	0.0007	0.0372	

20	CMY-67	ARO:3002080	Beta-lactam	Antibiotic inactivation	0.0002	0.0326	0.0009	0.0021	0.0007	0.0007	0.0372	

Table S4. Top 15 AMR gene families by cumulative abundance. Gene family name, total TPM, per-site abundances, and drug class association for the 15 most abundant AMR gene families across all six Peshawar samples.

Rank	**AMR Gene Family**	**Drug Class**	**Per-Sample Abundance (CPC)**						**Sector Mean (CPC)**			**Total CPC**
			CW1	CW2	HW1	HW2	SLW 1	SLW 2	CW	HW	SLW	**(all sites)**
1	CMY beta-lactamasea	Beta-lactam (cephamycin)	0.0054	0.8029	0.0222	0.0521	0.0187	0.0177	0.4042	0.0372	0.0182	0.9191
2	TEM beta-lactamase	Beta-lactam (penam)	0.2798	0.0408	0.1453	0.0583	0.0397	0.2337	0.1603	0.1018	0.1367	0.7975
3	GES beta-lactamase	Beta-lactam (carbapenem)	0.0697	0.0029	0.4496	0.1865	0.0191	0.0232	0.0363	0.3180	0.0211	0.7510
4	quinolone resistance protein (qnr)a	Fluoroquinolone	0.0031	0.6031	0.0040	0.0011	0.0080	0.0067	0.3031	0.0026	0.0074	0.6259
5	OXA beta-lactamase	Beta-lactam (carbapenem/penam)	0.0799	0.0462	0.1476	0.1028	0.0352	0.0540	0.0630	0.1252	0.0446	0.4657
6	VEB beta-lactamase	Beta-lactam (cephalosporin)	0.0264	0.0166	0.1314	0.0540	0.0254	0.0196	0.0215	0.0927	0.0225	0.2735
7	ANT(3'') aminoglycoside nucleotidyltransferase	Aminoglycoside	0.0462	0.0058	0.0419	0.0454	0.0124	0.0100	0.0260	0.0436	0.0112	0.1617
8	MFS antibiotic efflux pump	Multi-drug efflux	0.0294	0.0194	0.0429	0.0356	0.0166	0.0162	0.0244	0.0392	0.0164	0.1601
9	sulfonamide resistant sul	Sulfonamide	0.0397	0.0098	0.0435	0.0401	0.0137	0.0088	0.0248	0.0418	0.0113	0.1556
10	msr-type ABC-F protein	Macrolide/ streptogramin	0.0171	0.0123	0.0591	0.0280	0.0271	0.0040	0.0147	0.0435	0.0155	0.1475

11	RND antibiotic efflux pump	Multi-drug efflux	0.0332	0.0181	0.0172	0.0192	0.0172	0.0203	0.0256	0.0182	0.0187	0.1251
12	tetracycline-resistant ribosomal protection protein	Tetracycline	0.0057	0.0148	0.0335	0.0222	0.0268	0.0005	0.0102	0.0279	0.0136	0.1035
13	macrolide phosphotransferase (MPH)	Macrolide	0.0178	0.0060	0.0435	0.0204	0.0054	0.0072	0.0119	0.0319	0.0063	0.1003
14	NDM beta-lactamaseb	Beta-lactam (carbapenem)	0.0000	0.0002	0.0049	0.0037	0.0000	0.0003	0.0001	0.0043	0.0001	0.0091
15	vanH / vanR (glycopeptide resistance)b	Glycopeptide (vancomycin)	5.1×10^{-5}	0.0000	1.8×10^{-4}	9.4×10^{-4}	0.0000	0.0000	2.5×10^{-5}	5.6×10^{-4}	0.0000	0.0013

Table S5. MGE richness and abundance across Peshawar wastewater samples. Total MGE abundance (CPC), unique MGE types identified, and Level 1 type composition (% of total MGE pool) per sample and sector.

MGE Type	**Class**	**Per-Sample Abundance (CPC)**						**Sector Mean (CPC)**			**Total CPC**	**% Total**
		CW 1	CW 2	HW 1	HW 2	**PSL W1**	**PSL W2**	CW	HW	**PSL W**		
transposase	Transposase	3.4620	1.4681	4.9992	4.1947	1.3705	1.4194	2.6228	4.8434	1.5199	17.9721	62.8
insertion_element_IS91	IS element	0.9877	0.1421	0.5645	0.4733	0.1335	0.4518	0.5649	0.5189	0.2926	2.7528	9.6
integrase	Integrase	0.3984	0.1059	0.7641	0.6878	0.1187	0.1260	0.2521	0.7259	0.1223	2.2008	7.7
qacEdelta	Integron marker	0.3410	0.0575	0.5344	0.4800	0.1206	0.1079	0.1993	0.5072	0.1143	1.6415	5.7
tniA	Transposition gene	0.2296	0.0715	0.4423	0.4874	0.0838	0.1014	0.1506	0.4648	0.0926	1.4160	4.9
tniB	Transposition gene	0.1468	0.0294	0.2837	0.2845	0.0472	0.0569	0.0881	0.2841	0.0520	0.8485	3.0
ISCR	IS-CR element	0.1197	0.0211	0.1560	0.1372	0.0307	0.0418	0.0704	0.1466	0.0362	0.5064	1.8
IS26	IS	0.0339	0.0115	0.0905	0.0586	0.0072	0.0082	0.0227	0.0745	0.0077	0.2099	0.7

	element											
Tn916	Conjugative Tn	0.0061	0.0379	0.0797	0.0180	0.0347	0.0002	0.0220	0.0489	0.0174	0.1766	0.6
tnp-ISCR	Transposase (ISCR)	0.1135	0.0011	0.0161	0.0268	0.0020	0.0118	0.0573	0.0214	0.0069	0.1713	0.6
Total (all 37 MGE types)	—	6.2371	2.1675	8.3375	7.2129	2.0750	2.5824	4.2023	7.7752	2.3287	28.6124	100.0

Table S6. Genus-level dual ARG–MGE carriage across the Peshawar wastewater network for the 15 strongest of 44 dual-carrying genera (ρ ≥ 0.43; full 44-genus matrix in Table S16). MGE abundance (TPM), ARG abundance (TPM), Spearman co-occurrence coefficient (ρ), exact permutation p-value (all 6! = 720 sample orderings), Benjamini-Hochberg FDR q-value across the 43-genus comparison family, and dominant ARG carriage sector. No genus reaches FDR q < 0.05; associations are ranked by effect size and exact p-value as hypothesis-generating leads, not confirmed statistical findings.

Genus	**Total TPM**		**ARG Sector Mean** (TPM)			**MGE Sector Mean** (TPM)			**Coupling**		**Features**		**Dom. ARG Sector**	FDR q
	ARG	**MGE**	**PHW**	CW	**PSLW**	**PHW**	CW	**PSLW**	**ρ**	Exact p	**ARG**	**MGE**		(43 genera)
Acinetobacter	122.40	2112.18	6.340	13.860	40.998	14.10	36.06	615.87	1.000	0.003	37	43	Slaughterhouse	0.060
Escherichia	274.90	1185.53	34.710	41.340	61.393	62.27	102.65	429.68	0.829	0.058	206	41	Slaughterhouse	0.193
Pseudomonas	201.29	1989.39	44.950	32.130	23.560	424.89	185.56	104.32	0.943	0.017	40	42	Hospital	0.179
Streptococcus	110.43	752.45	14.060	18.130	23.030	33.65	57.42	284.98	0.771	0.103	13	19	Slaughterhouse	0.233
Klebsiella	10.21	419.67	4.070	0.530	0.505	204.34	17.89	5.46	0.429	0.419	25	33	Hospital	0.601
Thauera	10.63	327.19	3.065	1.427	0.824	126.85	34.23	2.21	1.000	0.003	2	11	Hospital	0.060
Enterococcus	9.56	248.20	2.850	0.260	1.560	110.37	8.16	5.58	0.829	0.0	5	9	Hospital	0.193

										58			
Staphylococcus	21.47	178.13	4.180	4.110	2.443	58.74	26.63	21.78	0.829	0.058	5	17	Hospital	0.193
Aeromonas	27.89	117.74	3.280	1.970	8.695	13.20	11.73	37.14	0.771	0.103	49	27	Slaughterhouse	0.233
Acidovorax	27.21	87.42	5.220	3.870	4.519	30.45	5.81	7.90	0.829	0.058	9	13	Hospital	0.193
Comamonas	13.46	57.85	0.990	1.710	4.024	7.93	4.42	17.01	0.771	0.103	15	13	Slaughterhouse	0.233
Shewanella	12.90	66.12	0.160	0.700	5.595	5.41	2.74	24.28	0.464	0.372	18	10	Slaughterhouse	0.552
Stenotrophomonas	11.44	37.98	1.090	3.270	1.356	5.90	9.38	2.43	0.829	0.058	8	11	Community	0.193
Corynebacterium	16.24	13.53	0.680	6.580	0.853	1.62	3.58	1.63	0.829	0.058	5	6	Community	0.193
Aliarcobacter	14.77	14.18	3.540	0.900	2.942	4.07	1.58	3.28	0.131	1.000	1	5	Hospital	1.000

Table S7. Phylum-level relative abundance across hospital (HW), slaughterhouse (SLW), and community (CW) wastewater sectors in Peshawar. Mean abundance per sector for the top 10 phyla; Dunn's test adjusted p-values shown for inter-sector pairwise comparisons after Bonferroni correction.

Phylum	CW mean (%)	HW mean (%)	SLW mean (%)	**KW p-value**	**Dunn's test (Bonferroni-adjusted)**
Pseudomonadota	74.69	60.43	44.51	0.102	CW vs SLW: p=0.049
Bacteroidota	4.32	5.51	25.51	0.156	ns (all ≥ 0.089)
Bacillota	3.83	4.30	12.88	0.867	ns
Campylobacterota	3.14	9.18	4.68	0.565	ns
Actinomycetota	5.75	6.30	2.77	0.651	ns
Thermodesulfobacteriota	0.58	3.52	2.73	0.155	ns (all ≥ 0.088)
Euryarchaeota†	2.21	2.33	0.73	0.368	ns
Fusobacteriota	0.07	0.01	1.17	0.651	ns
Mycoplasmatota	0.16	0.04	0.28	0.651	ns
Spirochaetota	0.07	0.10	0.08	0.867	ns

Table S8. Genus-level abundance and clinical relevance across wastewater sectors. Per-sector mean relative abundance, Dunn's test adjusted p-values, and clinical relevance notation for key genera detected across all six Peshawar samples.

Genus	CW mean (%)	HW mean (%)	SLW mean (%)	**Dunn's test (Bonferroni-adjusted)**	**Notes**
Thauera	6.75	24.34	4.35	ns	
Aliarcobacter	0.66	8.18	1.66	ns	Formerly classified as Arcobacter; drives HW1 Campylobacterota signal
Acinetobacter	1.66	0.64	8.08	SLW vs HW: p = 0.049*	Significantly enriched in SLW vs HW (Dunn p = 0.049)
Flavobacterium	0.35	0.50	9.50	ns	
Acetoanaerobium	0.42	0.26	7.08	ns	
Diaphorobacter	2.87	5.13	2.08	ns	
Pseudomonas	4.41	3.71	4.57	ns	
Acidovorax	2.99	3.43	2.21	ns	
Aeromonas	0.76	1.57	3.44	ns	
Arcobacter	0.86	1.00	2.78	ns	
Enterococcus	0.15	0.41	2.04	ns	
Escherichia	0.44	0.46	1.34	ns	
Streptococcus	0.39	0.85	1.34	ns	
Salmonella	0.13	0.22	0.19	ns	
Bifidobacterium	1.03	1.65	0.42	HW vs SLW: p = 0.049*	Significantly higher in HW vs SLW (Dunn p = 0.049)
Vibrio†	35.55	0.004	1.72	ns	CW2 single-timepoint bloom (71.06%); CW1 = 0.03%
Campylobacter	0.013	0.013	0.003	ns	True Campylobacter genus; phylum signal driven by Aliarcobacter

Table S9. Resistance mechanism distribution across Peshawar wastewater sectors. Cumulative TPM and percentage contribution per resistance mechanism category (antibiotic inactivation, efflux, target alteration, target protection, target replacement, reduced permeability) for HW, SLW, and CW sectors.

Resistance Mechanism	**Per-Sample Abundance (CPC)**	**Sector Mean (CPC)**	**% Total**

	CW1	CW2	HW1	HW2	SLW1	SLW2	CW	HW	SLW	**Abundance**
Antibiotic inactivation	0.5934	1.0102	1.1557	0.6592	0.2367	0.3902	0.8018	0.9075	0.3134	72.8
Antibiotic target protection	0.0269	0.6332	0.1020	0.0529	0.0668	0.0113	0.3301	0.0774	0.0390	16.1
Antibiotic efflux	0.0651	0.0543	0.0642	0.0591	0.0359	0.0369	0.0597	0.0617	0.0364	5.7
Antibiotic target replacement	0.0518	0.0183	0.0572	0.0532	0.0169	0.0144	0.0350	0.0552	0.0156	3.8
Antibiotic target alteration	0.0133	0.0111	0.0300	0.0183	0.0150	0.0025	0.0122	0.0242	0.0087	1.6
Reduced permeability	0.0000	9.8×10^{-5}	5.9×10^{-5}	0.0000	1.5×10^{-4}	2.6×10^{-4}	4.9×10^{-5}	3.0×10^{-5}	2.0×10^{-4}	<0.1

Table S10. Kruskal-Wallis and one-way ANOVA statistics for the top 10 MGE families across sectors. No family reached significance after FDR or Bonferroni correction (n = 2 per sector); ANOVA p-values are unadjusted.

MGE family	**KW H**	**KW p**	**ANOVA F**	**ANOVA p**	**Sig. (FDR/Bonf.)**
transposase	4.571	0.102	6.894	0.076	ns
insertion_element_IS91	2.000	0.368	0.309	0.755	ns
integrase	3.429	0.180	13.249	0.032	ns
qacEΔ1	3.429	0.180	6.143	0.087	ns
tniA	3.429	0.180	17.609	0.022	ns
tniB	3.429	0.180	13.484	0.032	ns
ISCR	3.429	0.180	3.755	0.152	ns
IS26	4.571	0.102	9.711	0.049	ns
Tn916	1.143	0.565	0.575	0.615	ns
tnp-ISCR	1.143	0.565	0.629	0.592	ns

Table S11. Key ARG–MGE co-occurrence pairs (Spearman ρ) across the six Peshawar wastewater samples, with exact permutation p-values (all 720 possible sample orderings) and Benjamini-Hochberg FDR-adjusted q-values across the full 10×12 MGE-ARG comparison matrix. Positive coefficients indicate positive co-occurrence; negative coefficients indicate inverse distribution. No pair reaches FDR $q < 0.05$ at this sample size; pairs are ranked and reported by effect size and exact p-value as hypothesis-generating associations.

MGE element	**ARG family**	**Spearman ρ**	**Direction**	Exact p	FDR

					q
integrase	GES beta-lactamase	+1.000	co-occurring	0.003	0.105
ISCR	GES beta-lactamase	+1.000	co-occurring	0.003	0.105
qacEΔ1	VEB beta-lactamase	+1.000	co-occurring	0.003	0.105
tniA	GES beta-lactamase	+0.943	co-occurring	0.017	0.105
tniB	GES beta-lactamase	+0.943	co-occurring	0.017	0.105
IS26	OXA beta-lactamase	+0.943	co-occurring	0.017	0.105
IS26	MFS efflux pump	+0.943	co-occurring	0.017	0.105
tniA	qnr (fluoroquinolone)	−0.943	inverse	0.017	0.105
tniB	qnr (fluoroquinolone)	−0.943	inverse	0.017	0.105
tnp-ISCR	qnr (fluoroquinolone)	−0.943	inverse	0.017	0.105
Tn916	RND efflux pump	−0.886	inverse	0.033	0.160

Table S12. Procrustes analysis results. Per-sample displacement (residual) between ARG and MGE ordination positions (overall $m^2 = 0.744$; $p = 0.756$); T1 versus T2 comparison of residual magnitude per sector.

Sample	Sector	Timepoint	ARG Ordination		MGE Ordination (rotated)		Procrustes Residual
			PCo1	PCo2	PCo1	PCo2	
HW1	Hospital	T1	-0.1713	0.1634	-0.2324	0.4873	0.3297
HW2	Hospital	T2	-0.1128	0.0278	-0.3347	-0.0375	0.2313
CW1	Community	T1	0.0715	-0.0323	-0.2539	-0.2729	0.4047
CW2	Community	T2	0.1217	-0.0729	0.4400	0.1973	0.4175
SLW1	Slaughterhouse	T1	0.0743	0.0453	0.2309	0.0099	0.1605
SLW2	Slaughterhouse	T2	0.0165	-0.1312	0.1501	-0.3842	0.2861
Mean — T1	All sectors	T1	—	—	—	—	0.2983
Mean — T2	All sectors	T2	—	—	—	—	0.3116

Table S13. Shared and unique bacterial species across the three wastewater sectors. Sector totals, unique species counts, pairwise shared counts, and the core microbiome count (species shared by all three sectors).

(A) Sector totals and unique species.

Category	Community (CW)	Hospital (HW)	Slaughterhouse (SLW)	Total / All Sectors
Total species detected	5,130	3,677	4,745	6,272
Sector-unique species	894	331	632	1,857
Sector-unique (%)	14.3%	5.3%	10.1%	29.6%

(B) Pairwise and core shared species.

Shared category	Species count	% of total (6,272)	Fold vs. smallest
Core (all three sectors)	2,865	45.7%	—
Community + Slaughterhouse only	1,069	17.0%	3.5× vs. C+H; 6.0× vs. H+S
Community + Hospital only	302	4.8%	1.7× vs. H+S
Hospital + Slaughterhouse only	179	2.9%	reference

Species counts come from an OTU presence/absence matrix (All_Taxa_OTU.xls) filtered to k__Bacteria, with a species scored present in a sector if it appeared in either of the two timepoints sampled. The core microbiome is defined as species detected across all three sectors: CW = community wastewater (CW1 + CW2); HW = hospital wastewater (HW1 + HW2); SLW = slaughterhouse wastewater (SLW1 + SLW2). Fold differences are calculated relative to the Hospital + Slaughterhouse pairwise overlap (n = 179), the smallest of the pairwise-exclusive categories.

Table S14. Drug class-stratified ARG abundance across Peshawar wastewater sectors. Sector means and network total (copies per cell, CPC), with percentage of total network ARG abundance (grand total = 11.98 CPC across six samples).

Drug class	CW mean	HW mean	SLW mean	**Total CPC**	**% total**
cephalosporin	0.294	0.688	0.239	2.442	20.4
penam	0.302	0.598	0.213	2.228	18.6
carbapenem	0.104	0.465	0.074	1.285	10.7
monobactam	0.185	0.200	0.162	1.095	9.1
cephamycin	0.411	0.059	0.035	1.009	8.4
penem	0.164	0.117	0.143	0.851	7.1
fluoroquinolone	0.322	0.017	0.023	0.723	6.0
macrolide	0.056	0.120	0.046	0.446	3.7
aminoglycoside	0.055	0.123	0.030	0.414	3.5
tetracycline	0.040	0.058	0.033	0.262	2.2
streptogramin	0.022	0.064	0.024	0.221	1.8
sulfonamide	0.026	0.042	0.012	0.161	1.3
phenicol	0.027	0.033	0.020	0.159	1.3
sulfone	0.023	0.040	0.011	0.148	1.2
lincosamide	0.018	0.031	0.018	0.133	1.1
diaminopyrimidine	0.012	0.016	0.010	0.075	0.6
rifamycin	0.008	0.013	0.009	0.060	0.5
streptogramin A	0.005	0.017	0.006	0.056	0.5
streptogramin B	0.005	0.015	0.006	0.053	0.4
disinfectants/ antiseptics	0.008	0.008	0.007	0.046	0.4
peptide	0.007	0.007	0.004	0.036	0.3
aminocoumarin	0.005	0.008	0.003	0.032	0.3
glycylcycline	0.003	0.003	0.004	0.019	0.2
pleuromutilin	0.002	0.003	0.002	0.015	0.1
glycopeptide	0.000	0.001	0.000	0.003	<0.1

CPC = copies per cell; the grand total across all drug classes is 11.98 CPC (six samples combined). The beta-lactam subclasses together, cephalosporin, penam, carbapenem, monobactam, cephamycin, and penem, make up 74.4% of total ARG abundance. HW = hospital; CW = community; SLW = slaughterhouse.

Table S15. WHO Bacterial Priority Pathogens List (BPPL) 2024 pathogens detected in Peshawar wastewater, with associated resistance genes (ARO) and drug classes by sector.

Host Pathogen	Resistance Gene (ARO)	Associated Drug Class	Primary Sector
Acinetobacter baumannii	blaOXA-23, blaOXA-66, aph(3')-VI	Carbapenems, Aminoglycosides	Hospital
Klebsiella pneumoniae	blaNDM-1, blaVIM, blaSHV	Carbapenems, Cephalosporins	Hospital
Escherichia coli	mcr-1, tet(A), blaCTX-M	Colistin, Tetracyclines, ESBLs	Slaughterhouse
Salmonella enterica	qnrVC, sul1, dfrA	Fluoroquinolones, Sulfonamides	Community
Citrobacter spp.	blaOXA-48-like, qnrS	Carbapenems, Fluoroquinolones	Community
Pseudomonas aeruginosa	mexB, triC, aph(3')-IIb	Multidrug Efflux, Aminoglycosides	Hospital / Community

Pathogen selection follows the WHO Bacterial Priority Pathogens List 2024 (World Health Organization, 2024). Primary sector indicates the wastewater compartment of dominant enrichment.

Supplementary References

Alcock, B.P., et al. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction. Nucleic Acids Research, 51(D1), D690–D699.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics.

Bolger, A.M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120.

Buchfink, B., Xie, C., & Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nature Methods, 12(1), 59–60.

Gu, Z., et al. (2014). circlize implements and enhances circular visualization in R. Bioinformatics, 30(19), 2811–2812.

Gurevich, A., et al. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics, 29(8), 1072–1075.

Hyatt, D., et al. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics, 11, 119.

Kahlke, T., & Ralph, P.J. (2019). BASTA – taxonomic classification of sequences and sequence bins using last common ancestor estimations. Methods in Ecology and Evolution, 10(1), 100–103.

Langmead, B., & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357–359.

Li, D., et al. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly. Bioinformatics, 31(10), 1674–1676.

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, 22(13), 1658–1659.

Lu, J., et al. (2017). Bracken: estimating species abundance in metagenomics data. PeerJ Computer Science, 3, e104.

Patro, R., et al. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods, 14(4), 417–419.

Waskom, M.L. (2021). seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021.

Wood, D.E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biology, 20(1), 257.

Table S16a. Full taxonomy–ARG Spearman correlation matrix (15 genera × 12 ARG families = 180 pairs) across the six Peshawar samples, with exact permutation p-values (all 6! = 720 sample orderings) and Benjamini-Hochberg FDR q-values across the full matrix. No pair reaches FDR q < 0.05; the 25 pairs with the largest effect sizes are listed here, ranked by |ρ|. Full 180-pair matrix available from the corresponding authors on request.

Genus	ARG family	ρ	Exact p	FDR q
Aliarcobacter	resistance-nodulation-cell division (RND) antibiotic efflux pump	-1.000	0.003	0.462
Pseudomonas	CMY beta-lactamase	-0.943	0.017	0.462
Thauera	VEB beta-lactamase	0.943	0.017	0.462
Diaphorobacter	GES beta-lactamase	0.943	0.017	0.462
Arcobacter	TEM beta-lactamase	-0.943	0.017	0.462
Thauera	GES beta-lactamase	0.886	0.033	0.462
Thauera	sulfonamide resistant sul	0.886	0.033	0.462
Pseudoxanthomonas	quinolone resistance protein (qnr)	-0.886	0.033	0.462
Diaphorobacter	OXA beta-lactamase	0.886	0.033	0.462
Diaphorobacter	VEB beta-lactamase	0.886	0.033	0.462
Aliarcobacter	tetracycline-resistant ribosomal protection protein	0.886	0.033	0.462
Thauera	quinolone resistance protein (qnr)	-0.886	0.033	0.462
Acidovorax	TEM beta-lactamase	0.886	0.033	0.462
Acetoanaerobium	OXA beta-lactamase	-0.829	0.058	0.477
Sphaerotilus	TEM beta-lactamase	0.829	0.058	0.477
Thauera	msr-type ABC-F protein	0.829	0.058	0.477
Thauera	ANT(3")	0.829	0.058	0.477
Acinetobacter	major facilitator superfamily (MFS) antibiotic efflux pump	-0.829	0.058	0.477
Pseudoxanthomonas	ANT(3")	0.829	0.058	0.477
Arcobacter	resistance-nodulation-cell division (RND) antibiotic efflux pump	-0.829	0.058	0.477
Quatrionicoccus	quinolone resistance	-0.829	0.058	0.477

	protein (qnr)			
Acetoanaerobium	TEM beta-lactamase	-0.829	0.058	0.477
Acetoanaerobium	quinolone resistance protein (qnr)	0.771	0.103	0.544
Diaphorobacter	ANT(3")	0.771	0.103	0.544
Diaphorobacter	quinolone resistance protein (qnr)	-0.771	0.103	0.544

Table S16b. Full MGE–ARG Spearman correlation matrix (10 MGE types × 12 ARG families = 120 pairs) across the six Peshawar samples, with exact permutation p-values and Benjamini-Hochberg FDR q-values across the full matrix. No pair reaches FDR q < 0.05; the 25 pairs with the largest effect sizes are listed here, ranked by |ρ|. Full 120-pair matrix available from the corresponding authors on request.

MGE type	ARG family	ρ	Exact p	FDR q
qacEΔ1	VEB beta-lactamase	1.000	0.003	0.105
integrase	GES beta-lactamase	1.000	0.003	0.105
ISCR	GES beta-lactamase	1.000	0.003	0.105
transposase	major facilitator superfamily (MFS) antibiotic efflux pump	0.943	0.017	0.105
integrase	OXA beta-lactamase	0.943	0.017	0.105
integrase	VEB beta-lactamase	0.943	0.017	0.105
tniA	GES beta-lactamase	0.943	0.017	0.105
transposase	OXA beta-lactamase	0.943	0.017	0.105
qacEΔ1	GES beta-lactamase	0.943	0.017	0.105
qacEΔ1	sulfonamide resistant sul	0.943	0.017	0.105
tnp-ISCR	ANT(3")	0.943	0.017	0.105
ISCR	VEB beta-lactamase	0.943	0.017	0.105
tniB	quinolone resistance protein (qnr)	-0.943	0.017	0.105
ISCR	OXA beta-lactamase	0.943	0.017	0.105
tniB	GES beta-lactamase	0.943	0.017	0.105
tnp-ISCR	quinolone resistance protein (qnr)	-0.943	0.017	0.105
IS26	OXA beta-lactamase	0.943	0.017	0.105
IS26	major facilitator superfamily (MFS) antibiotic efflux pump	0.943	0.017	0.105
tniA	quinolone resistance	-0.943	0.017	0.105

	protein (qnr)			
qacEΔ1	msr-type ABC-F protein	0.886	0.033	0.160
tniA	OXA beta-lactamase	0.886	0.033	0.160
tniB	OXA beta-lactamase	0.886	0.033	0.160
tniB	VEB beta-lactamase	0.886	0.033	0.160
tniA	VEB beta-lactamase	0.886	0.033	0.160
Tn916	resistance-nodulation-cell division (RND) antibiotic efflux pump	-0.886	0.033	0.160

Table S16c. Full genus-level ARG–MGE co-occurrence matrix for all 44 dual-carrying genera (43 with non-zero co-occurrence), with exact permutation p-values and Benjamini-Hochberg FDR q-values across the 43-genus comparison family. No genus reaches FDR q < 0.05 (Acinetobacter and Thauera come closest, q = 0.060). Ranked by |ρ|.

Genus	ρ	Exact p	FDR q
Acinetobacter	1.000	0.003	0.060
Thauera	1.000	0.003	0.060
Phenylobacterium	1.000	0.167	0.299
Pseudomonas	0.943	0.017	0.179
Rheinheimera	0.941	0.017	0.179
Burkholderia	0.868	0.042	0.193
Mycobacterium	0.857	0.033	0.193
Arenimonas	0.845	0.067	0.205
Corynebacterium	0.829	0.058	0.193
Staphylococcus	0.829	0.058	0.193
Acidovorax	0.829	0.058	0.193
Stenotrophomonas	0.829	0.058	0.193
Vibrio	0.829	0.058	0.193
Enterococcus	0.829	0.058	0.193
Escherichia	0.829	0.058	0.193
Marinobacter	-0.823	0.100	0.233
Pseudoxanthomonas	0.812	0.072	0.207
Tessaracoccus	-0.778	0.133	0.279
Aeromonas	0.771	0.103	0.233
Comamonas	0.771	0.103	0.233
Streptococcus	0.771	0.103	0.233
Zoogloea	0.706	0.136	0.279

Cupriavidus	0.696	0.167	0.299
Novosphingobium	0.677	0.150	0.293
Dietzia	-0.647	0.300	0.478
Salmonella	-0.600	0.242	0.413
Microbacterium	0.577	0.250	0.413
Flavobacterium	0.507	0.333	0.512
Shewanella	0.464	0.372	0.552
Klebsiella	0.429	0.419	0.601
Xanthomonas	0.395	0.450	0.624
Citrobacter	0.393	0.667	0.796
Propionibacterium	-0.393	0.667	0.796
Paracoccus	0.371	0.497	0.668
Stutzerimonas	0.273	0.617	0.796
Pseudoalteromonas	0.257	0.633	0.796
Paraburkholderia	-0.216	0.708	0.802
Kocuria	0.203	0.700	0.802
Variovorax	0.152	0.833	0.919
Paenibacillus	0.133	1.000	1.000
Aliarcobacter	0.131	1.000	1.000
Cutibacterium	0.131	1.000	1.000
Rhizobium	0.072	1.000	1.000
Leucobacter	n/a	n/a	n/a

wer (%)