nature communications

Article

https://doi.org/10.1038/s41467-026-77447-7

Plasmid–Host Networks Reveal Persistent Mobile Elements Across Wastewater Treatment

Received: 12 November 2025

Accepted: 21 August 2026

Published online: 07 September 2026

Check for updates

Siyi Zhou[1], Sarah E. Philo[2], Michael A. Saldana[2], Jeseth Delgado Vela[3], Adam L. Smith[2] & Lauren B. Stadler[1]

Plasmids drive horizontal gene transfer and antimicrobial resistance spread in water resource recovery facilities, yet their in situ dynamics in complex microbial communities remain poorly characterized. We apply Hi-C metagenomics to resolve plasmid–host associations across influent, activated sludge, and effluent at three facilities. We identify 944 plasmid clusters with facility- and stage-specific distributions. Plasmid host range narrows from influent to effluent, indicating that treatment selectively constrains plasmid–host associations. Persistent plasmid clusters with detectable mobility markers show a trend toward broader, less modular host networks than those with no detected markers. Hi-C-supported antibiotic resistance gene associations highlight *Burkholderiaceae* and *Rhodocyclaceae* as prominent hosts of beta-lactam, tetracycline and sulfonamide resistance genes across facilities. These findings indicate that wastewater treatment narrows plasmid host range while maintaining dominant plasmid-carrying taxa and adaptive mobile elements, providing a framework for understanding plasmid persistence across treatment systems.

Plasmids are extrachromosomal genetic elements that facilitate bacterial adaptation and evolution by mediating horizontal gene transfer (HGT)[1–3]. These mobile genetic elements (MGEs) contribute to the spread of antimicrobial resistance (AMR), virulence factors, and metabolic capabilities, significantly shaping the function and dynamics of microbial communities[4–6]. While plasmid biology has been extensively studied under controlled laboratory conditions, the mechanisms governing plasmid persistence, such as conjugative transfer, fitness impacts, selective pressures, and interspecies competition, remain to be assessed in complex environmental microbial communities[7–9]. Distinguishing between these mechanisms in representative environmental systems is critical for advancing our understanding of plasmid-host ecology.

Water resource recovery facilities (WRRFs) serve as critical interfaces between the built and natural environments, facilitating the exchange and dissemination of genetic material, including AMR genes[9,10]. Biological treatment systems like activated sludge provide ideal conditions for HGT and the spread of adaptive traits due to high microbial densities, abundant biofilms, and active microbial metabolism[11,12]. Broad-host-range plasmids are key drivers of this process, transferring between diverse microbial hosts and mobilizing otherwise non-mobilizable plasmids, thereby accelerating the spread of AMR[6,13,14]. Although disinfection processes inactivate bacteria in wastewater effluents, they may not fully eliminate plasmid-carrying bacteria or extracellular DNA, allowing AMR genes to persist[15,16]. Residual plasmids in effluent may contribute to the reintroduction of resistance genes into natural ecosystems, where they could be acquired by environmental or opportunistic pathogens[10]. Understanding plasmid persistence throughout treatment processes is crucial for assessing AMR risks and developing strategies to mitigate

[1]Civil and Environmental Engineering, Rice University, Houston, TX, USA. [2]Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, CA, USA. [3]Department of Civil and Environmental Engineering, Duke University, Durham, NC, USA.
e-mail: lauren.stadler@rice.edu

their dissemination from WRRFs to environmental microbial communities.

Despite advances in sequencing technologies, tracking plasmid dissemination and host range is methodologically challenging. Laboratory-based approaches, such as synthetic plasmid systems and culture-based conjugation assays, often rely on well-characterized, culturable hosts and controlled conditions, which may not fully capture the complexity of natural microbial ecosystems[17,18]. While emerging metagenomic approaches, including high-throughput single-cell sequencing, can be used to reveal host-MGE associations, they are limited by incomplete genome recovery, amplification biases, and insufficient sequencing depth, hindering the reconstruction of complete plasmid sequences and transmission networks[19,20].

Chromosome conformation capture sequencing, like Hi-C, preserves spatial interactions between plasmids and their hosts within intact microbial communities. This provides key ecological context for understanding plasmid dissemination in microbial communities[21–23]. Unlike single-cell sequencing, which isolates individual cells and may miss community interactions, Hi-C maintains in situ associations, enabling a more comprehensive reconstruction of plasmid-host networks. This capability is particularly valuable in complex environments such as WRRFs, where plasmids circulate among diverse taxa. Hi-C studies have inferred plasmid-host associations using either contig-level or bin-level data[22,23]. Contig-level analyses are highly sensitive, detecting links even for fragmented and low-abundance plasmids, but the loss of broader genomic context can inflate apparent host ranges. Bin-level analyses retain genomic context and improve host-range estimates, but they are more vulnerable to Hi-C noise. Artefactual links from contig ends or IS elements, as well as incomplete binning, can result in host misassignments[24]. Here, to track plasmids from influent to effluent and examine their genetic context, we adapted a recent Hi-C-based plasmid clustering framework[21], and implemented a graph-based approach to group plasmid contigs into putative plasmid units with stringent linkage thresholds, aiming to provide a conservative, context-preserving basis for plasmid-host associations.

In this study, we applied Hi-C metagenomics and performed plasmid clustering to systematically track plasmid-host linkages across multiple unit processes within WRRFs in three U.S. states. Our objectives were to: (1) characterize the distribution of plasmids across treatment stages, (2) resolve plasmid-host associations with improved resolution of plasmid genetic context (e.g., encoded functions, plasmid mobility), and (3) identify bacterial taxa that may serve as plasmid reservoirs across treatment stages. This study enhances our understanding of plasmid persistence and transmission in WRRFs by offering insights into their microbial ecology.

Results

Plasmid-associated communities exhibit greater variability across facilities than total microbial communities

We began by comparing microbial and plasmid-associated communities across three facilities (WRRFs in State A, B, and C), with samples collected from influent (Inf), activated sludge (AS), and effluent (Eff). The total and plasmid-associated communities were assessed at the family level, with total community profiles derived from all MAGs and plasmid-associated profiles restricted to MAGs linked to plasmid clusters via Hi-C contacts. Within each facility and treatment stage, the three weekly samples were aggregated and re-normalized to generate a single representative community profile. Community-level variation was assessed using non-metric multidimensional scaling (NMDS) based on Bray–Curtis dissimilarities (Fig. 1a). Both total and plasmid-associated communities revealed facility-specific clustering of the microbial communities, with samples from the same facility appearing more similar than those from different facilities. Furthermore, the plasmid-associated communities showed more distant separation across facilities as compared to total communities. PERMANOVA analysis confirmed a significant effect of facility identity on community composition ($p = 0.015$ for total community; $p = 0.004$ for plasmid-associated community). These results were robust to weekly disaggregation: per-sample PERMANOVA confirmed a significant facility effect for plasmid-associated communities ($p = 0.001$, $R^2 = 0.21$), while the facility effect for total communities was weaker at weekly resolution ($p = 0.109$), likely reflecting greater week-to-week variability in overall community composition. Notably, facility identity explained a larger proportion of variation in plasmid-associated communities ($R^2 = 0.508$) than in total communities ($R^2 = 0.364$), indicating that plasmid-associated communities were more variable across facilities than the total communities. This finding suggests that plasmid-host associations may be more sensitive to differences in influent sources or treatment technologies across facilities than the overall microbial community structure.

We next assessed alpha diversity using the Shannon index, calculated at the family level from MAG taxonomy (Fig. 1b). Across all three facilities and stages, total communities exhibited consistently higher diversity as compared to plasmid-associated communities. This reflects that plasmid-associated taxa represent only a subset of the overall microbial community. Community-level alpha diversity did not differ significantly across facilities or stages ($p > 0.05$). However, as detailed in subsequent analyses, the host range of individual plasmid clusters (measured by the number of MAGs and bacterial classes per cluster) varied across stages. Additionally, Hi-C crosslinking may not capture all plasmid-bearing taxa, which could further contribute to lower observed diversity in the plasmid-associated fraction.

To better understand these differences in community composition between the total and plasmid-associated communities, we examined the taxonomic profiles at the family level for both total and plasmid-associated communities (Fig. 1c). The dominant taxa differed between the two groups: total communities were dominated by *Comamonadaceae* (18.2% on average), followed by *Pseudomonadaceae* (8.8%), *Aeromonadaceae* (4.9%) and *Burkholderiaceae* (4.1%); in contrast, plasmid-associated communities displayed a narrower taxonomic spectrum, dominated by *Burkholderiaceae* (34.3%), *UBA10799* (18.0%), *Neisseriaceae* (15.3%), *Nitrospiraceae* (14.2%), and *Rhodocyclaceae* (12.6%). These findings highlight that the families that were most abundant in the overall community may not necessarily dominate plasmid-host associations. Instead, plasmid-host interactions appear concentrated within a distinct subset of bacterial families.

Broader host range and higher plasmid prevalence drive plasmid–host connectivity in the State A WRRF

To characterize plasmid–host association patterns, we quantified: (1) plasmid host range, defined as either the number of unique metagenome-assembled genomes (MAGs; individual genome bins) or the number of distinct bacterial classes (taxonomic groups) associated with each plasmid cluster, and (2) plasmid burden, defined as the number of unique plasmid clusters associated with each MAG. Plasmids in the State A WRRF associated with the widest range of bacterial classes, spanning on average 5.5 ± 0.20 classes per cluster, significantly more than in the State B (3.4 ± 0.20) or State C (2.5 ± 0.26) WRRFs (Fig. 2a; both $p < 0.0001$; two-sided Wilcoxon rank-sum test with Benjamini–Hochberg adjustment). By MAG count, clusters from the State A and State B WRRFs reached comparably broad host sets (28.2 ± 1.68 and 28.7 ± 2.31 MAGs per cluster, respectively), both far exceeding State C (13.6 ± 1.92 MAGs; both $p < 0.0001$) (Fig. 2b). Although State A and State B plasmids associated with similar numbers of host MAGs, those in State A were distributed across a wider range of bacterial classes. Plasmid burden per MAG showed a similar trend (Fig. 2c; Supplementary Fig. 1): MAGs in the State A WRRF carried on average 18.6 ± 0.98 plasmid clusters (max = 57), compared to 10.5 ± 0.62 in State B (max = 46) and 5.1 ± 0.42 in State C (max = 30). These differences were also significant ($p < 0.0001$), indicating a higher

https://doi.org/10.1038/s41467-026-77447-7

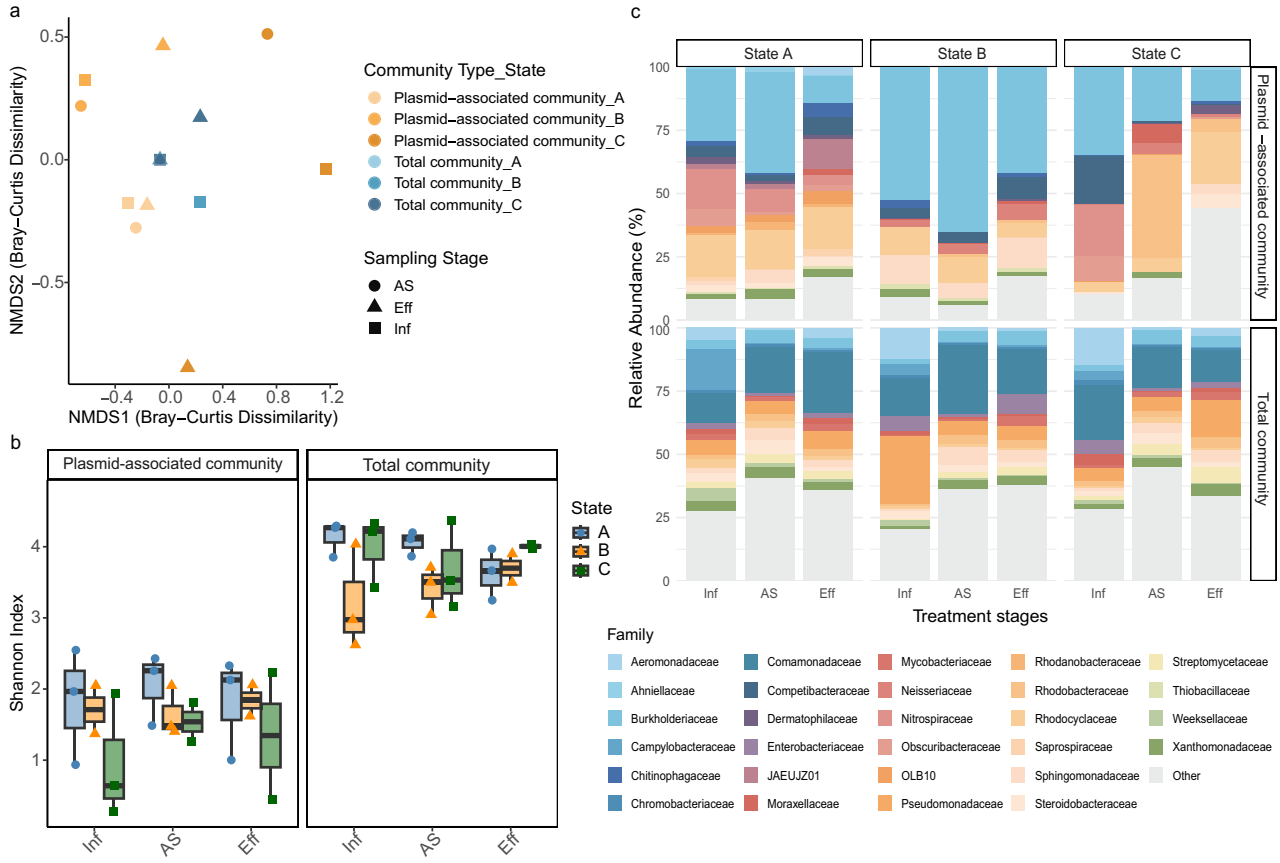


Fig. 1 | Microbial community composition, plasmid-associated taxa, and diversity patterns across wastewater treatment facilities. a Non-metric multi-dimensional scaling (NMDS) plots based on Bray–Curtis dissimilarities for the plasmid-associated taxa and total communities. Inf influent, AS activated sludge, Eff effluent. PERMANOVA revealed significant differences across states (plasmid-associated: $p = 0.004$, $R^2 = 0.508$; total: $p = 0.015$, $R^2 = 0.364$). **b** Box plots show the Shannon index for each state: the center line indicates the median, the box bounds mark the 25th and 75th percentiles (interquartile range), and whiskers extend to the minimum and maximum values within 1.5 × the interquartile range; individual weekly samples are overlaid as points. Differences across facilities were tested with a two-sided Kruskal–Wallis test, performed separately for each community type, using all available samples from each facility. For the plasmid-associated community, $n = 9/7/7$ for States A/B/C; $H = 4.47$, df = 2, $p = 0.107$. For the total community, $n = 9/8/8$ for States A/B/C; $H = 5.43$, df = 2, $p = 0.066$. ns, not significant. **c** Relative abundance of the top 20 bacterial families in (top) plasmid-associated communities (MAGs linked to plasmid clusters via Hi-C contacts) and (bottom) total communities (all MAGs).

plasmid load in the microbial community of the State A WRRF. The co-occurrence of broader host associations and elevated plasmid burden may indicate a community more conducive to plasmid exchange and maintenance. State A WRRF lacks primary clarification and uses pure oxygen aeration, which supports higher biomass density and sustained aerobic activity, conditions known to favor conjugation and plasmid maintenance[25,26]. In contrast, State C WRRF operates a five-stage Bardenpho process with alternating redox conditions that may impose energetic constraints on plasmid stability[27,28].

To further investigate WRRF-specific differences in plasmid-host interaction structure, we constructed bipartite networks linking plasmid clusters to bacterial host classes (Fig. 3a) and quantified network properties (Supplementary Data 1). The number of host-linked plasmid clusters varied among facilities (State A WRRF: 299; State B WRRF: 120; State C WRRF: 62), which may contribute to differences in apparent network complexity. Despite this, these networks revealed marked contrasts in connectivity across states. The State A WRRF showed the most interconnected network, with the highest connectance (0.50) and mean node degree (9.18), and the lowest modularity ($Q = 0.16$), indicating extensive cross-lineage plasmid–host connectivity. The State B WRRF displayed reduced connectivity (connectance = 0.21, mean node degree = 3.90), with plasmid-host associations more concentrated among *Gammaproteobacteria, Bacteroidia* and *Alphaproteobacteria*, though modularity remained low ($Q = 0.17$). The State C WRRF showed the most compartmentalized network structure, with the highest modularity ($Q = 0.35$), indicating that plasmid clusters were largely confined to specific host classes, reflecting a narrower host-association breadth.

To explore the observed differences, we analyzed the persistence of plasmid-host associations across treatment stages within each state (Fig. 3b). The State A WRRF retained a diverse and stable set of host classes across all stages, with many associations maintained from influent to effluent. In contrast, the State B and State C WRRFs showed more stage-specific patterns, with fewer host classes shared between stages, indicating lower persistence of plasmid-host associations. We also quantified the occurrence of broadly distributed plasmids (e.g., ≥3 bacterial classes, ≥5 MAGs) across treatment stages (Fig. 3c). Cross-class transfer has been documented experimentally for promiscuous conjugative plasmids, including IncP-1, IncU, IncW and IncQ groups[29,30]; we also detected IncQ-type plasmid clusters associated with multiple classes in our dataset. The State A WRRF consistently carried more broadly distributed plasmids, whereas most plasmids in the State C WRRF were confined to a single host class and a small number of MAGs. These findings suggest that the greater network connectivity observed in the State A WRRF is driven by broader plasmid host ranges and higher plasmid prevalence.

https://doi.org/10.1038/s41467-026-77447-7

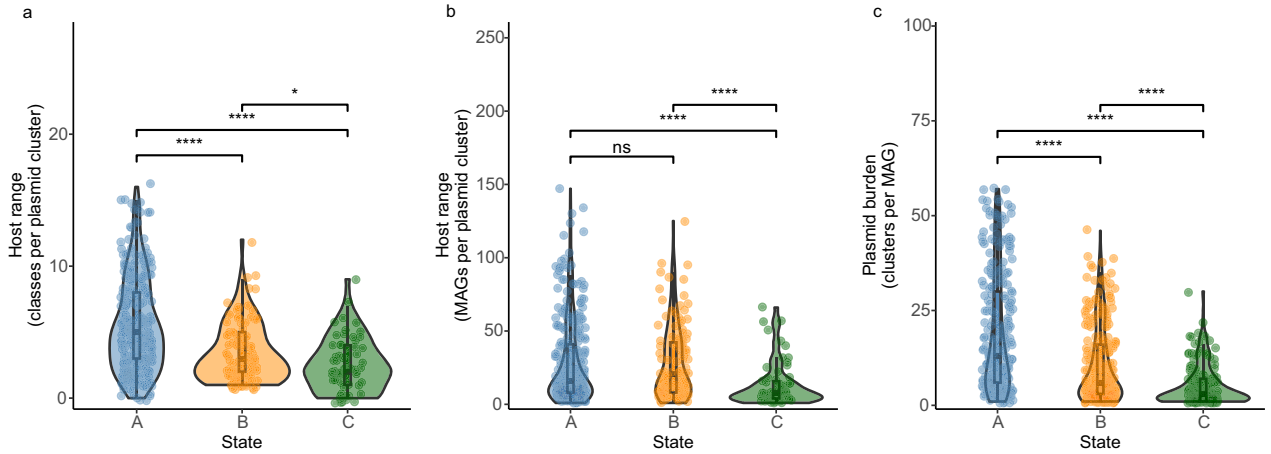


Fig. 2 | Plasmid-host association metrics across wastewater treatment samples. **a**, **b** Violin plots showing the distribution of plasmid host range, defined as the number of bacterial classes (**a**) and the number of MAGs (**b**) associated with each plasmid cluster in States A, B, and C ($n = 299$, 120, and 62 clusters, respectively; biological replicates). Adjusted P values: **a** A vs B, $p = 1.4 \times 10^{-8}$; A vs C, $p = 3.2 \times 10^{-10}$; B vs C, $p = 0.011$; **b** A vs B, $p = 0.41$; A vs C, $p = 1.2 \times 10^{-5}$; B vs C, $p = 1.1 \times 10^{-5}$. **c** Plasmid burden, measured as the number of plasmid clusters per MAG ($n = 259$, 263, and 156 MAGs for States A, B, and C, respectively; biological replicates; A vs B, $p = 6.0 \times 10^{-11}$; A vs C, $p = 8.3 \times 10^{-27}$; B vs C, $p = 2.0 \times 10^{-9}$). In all panels the violin shows the kernel density distribution, and the overlaid box plot indicates the median (center line), the 25th and 75th percentiles (box bounds), and whiskers extending to the minimum and maximum values within 1.5 × the interquartile range; all individual data points are overlaid. For all panels, pairwise differences between states were evaluated using two-sided unpaired Wilcoxon rank-sum tests with Benjamini–Hochberg adjustment. Significance is denoted **** ($p < 0.0001$), * ($p < 0.05$), and ns (not significant).

Per-cluster plasmid host range narrows across treatment stages

We next investigated the distribution of plasmid clusters across treatment stages by quantifying their host range, defined as the number of associated MAGs and bacterial classes per plasmid cluster. Host range was calculated within each WRRF, and per-WRRF counts were pooled for stage-level comparisons. Plasmid host range differed significantly between stages: plasmids from influent and AS were associated with more MAGs and classes than those from effluent ($p < 0.05$, Wilcoxon test; Fig. 4a, b). Facility-specific analyses (Supplementary Fig. 2a, b) showed a consistent overall pattern, though the degree of host range reduction from influent to effluent varied, likely reflecting differences in treatment configuration and/or influent composition. This suggests that plasmids with broader host ranges are more common in earlier stages of treatment, while those persisting in the effluent tend to be more host-restricted. Across all states, the number of MAGs per plasmid cluster was strongly correlated with the number of host classes per cluster ($r_s \geq 0.79$, $p < 0.001$; Fig. 4c), indicating that plasmids associating with more genomes also tend to span greater phylogenetic breadth.

Mobility status and host-association patterns of persistent plasmid clusters

To characterize the mobility potential of plasmid clusters, we annotated them using MOB-suite to identify key mobility markers, including mating pair formation genes (MPF), relaxase/mobilization genes (MOB), and origins of transfer (oriT). Each cluster was categorized as conjugative (MPF present), mobilizable/conjugative (relaxase/MOB or oriT without MPF), or no marker detected. Of 481 clusters across three WRRFs, 181 (37.6%) were conjugative, 57 (11.9%) mobilizable/conjugative, and 243 (50.5%) contained no detectable mobility markers. This distribution did not differ significantly across facilities ($\chi^2 = 7.60$, $p = 0.11$) or stages (Fisher's exact tests, all $p > 0.24$). MPF types G and T predominated (33 and 27 clusters), with *virB4* as the most prevalent conjugation gene across all facilities (Supplementary Data 2; Supplementary Fig. 3).

To assess whether mobility potential influences host association breadth, we first identified persistent plasmid clusters, defined as clusters sharing ≥95% nucleotide identity across two or more treatment stages within the same facility and sampling week. We then constructed weighted bipartite networks connecting persistent clusters to host families via normalized Hi-C linkage, grouped by mobility status: Marker Positive (conjugative or mobilizable/conjugative evidence detected) versus No Marker (no mobility markers detected). Because short-read metagenomic assemblies are inherently fragmented, the No Marker designation indicates that markers were not detected, rather than that non-mobilizability was confirmed. To compare network structure between the two groups, we calculated connectance, modularity, and the number of communities for each subnetwork (Supplementary Fig. 4; Supplementary Data 1). The Marker Positive subnetwork (30 persistent plasmid clusters) shared a greater proportion of possible host links (connectance: 0.18 vs. 0.15), was less compartmentalized (modularity: 0.30 vs. 0.52), and formed fewer distinct host–plasmid groups (6 vs. 8 communities) than the No Marker subnetwork (35 persistent plasmid clusters). Marker Positive groups associated with a median of 16.5 host families versus 10 for No Marker groups, although this per-cluster comparison was not statistically significant (Wilcoxon $p = 0.10$). The network-level metrics point in the same direction, which may suggest that mobility-marker-bearing plasmid clusters tend to maintain broader, less compartmentalized host associations.

Among the 65 persistent plasmid clusters identified, 13 were detected across all three stages (Inf-AS-Eff) at 98.7–100% identity (Supplementary Data 3). We examined the breadth and composition of Hi-C-linked hosts for each persistent plasmid cluster at each stage, resolved to genus level where available. Several conjugative persistent plasmid clusters exhibited broad, stage-variable host profiles. For example, A2-L5 (State A, MPF type T) was primarily associated with *Phenylobacterium* sp. in influent, *Novosphingobium* sp. in activated sludge and *OLB10* in effluent based on normalized Hi-C contacts, at 99.1–100% identity. Taxonomic profiling of the total microbial community showed that the genera associated with persistent plasmid clusters persisted or increased in relative abundance across treatment stages (Supplementary Fig. 5), indicating that the observed changes in host association profiles cannot be attributed to shifts in host relative abundance alone. While this is consistent with expanded host associations, we acknowledge that metagenomic

https://doi.org/10.1038/s41467-026-77447-7

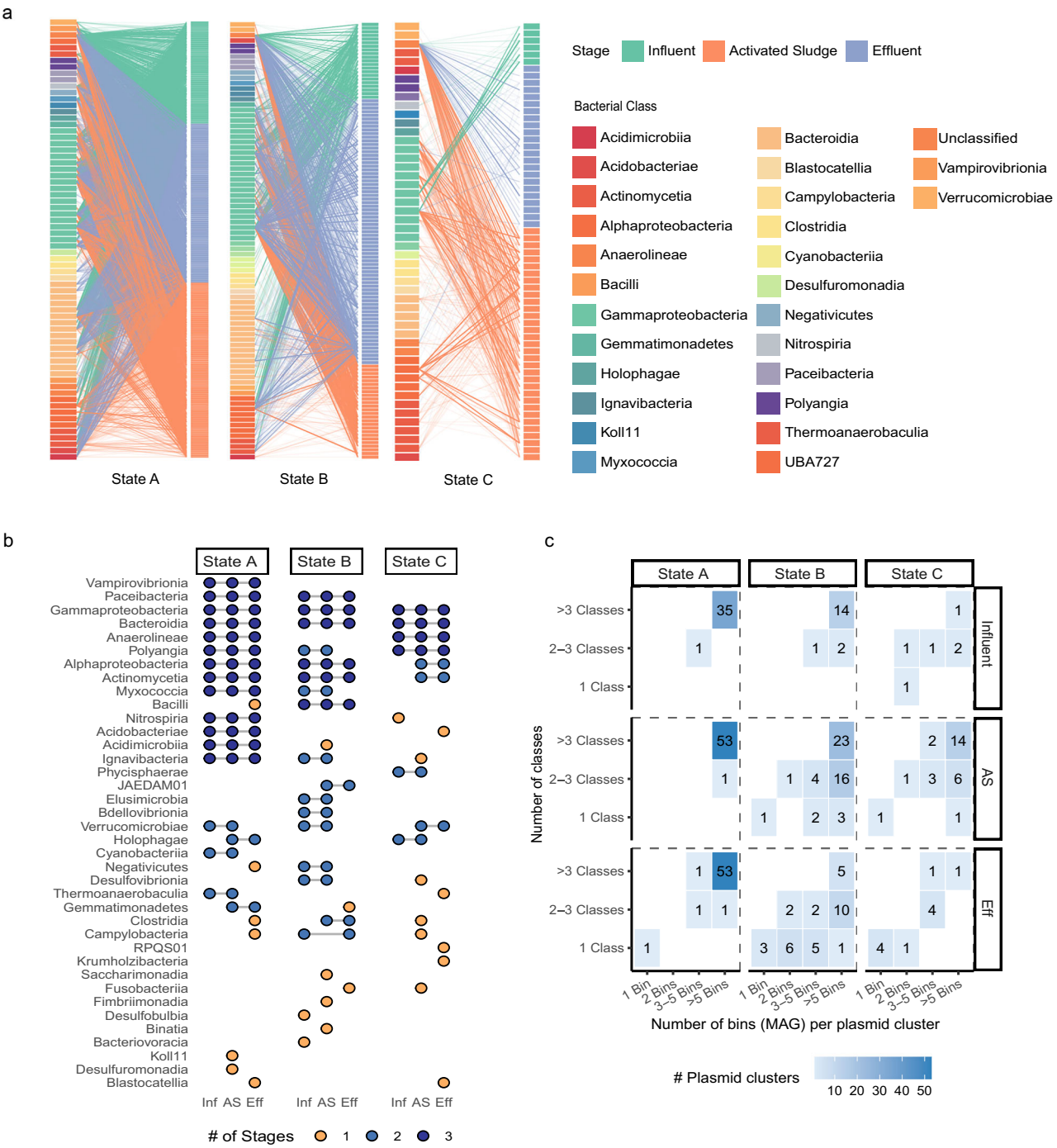


Fig. 3 | WRRF-specific patterns of plasmid-host association structure, host distribution across stages, and plasmid dissemination metrics. a Bipartite networks linking plasmid clusters (right columns, colored by treatment stages) to bacterial hosts (left columns, colored by taxonomic class) in the State A, State B, and State C WRRFs. Edges represent plasmid–host associations inferred from Hi-C linkages. **b** Presence–absence matrix of plasmid-carrying bacterial classes across treatment stages. Dot color indicates persistence across stages: dark blue = detected in all three stages, light blue = detected in two stages, and yellow = detected in only one stage. **c** Heatmaps showing the number of plasmid clusters binned by host burden (i.e., the number of associated MAGs) and host range (i.e., the number of associated bacterial classes and bins) across treatment stages in each WRRF.

co-occurrence does not constitute direct evidence of transfer. In contrast, No Marker persistent plasmid clusters such as B1-L2 (*Ottowia* sp., all stages, 99.5–100% identity) exhibited narrow, stage-stable host profiles. Their limited host breadth and lack of detected mobility markers are consistent with vertical co-inheritance, i.e., persistence of a plasmid element within a stable host population across treatment stages.

Hi-C-based assignment of ARG contigs to plasmid clusters

Because insertion sequence (IS)/transposase-encoding contigs were excluded prior to clustering, ARG-bearing contigs were not incorporated into plasmid clusters. To identify associations between ARG contigs and plasmid clusters, we used normalized Hi-C contacts to associate plasmid-derived ARG-bearing contigs with plasmid clusters as well as host MAGs. Across all three facilities,

https://doi.org/10.1038/s41467-026-77447-7

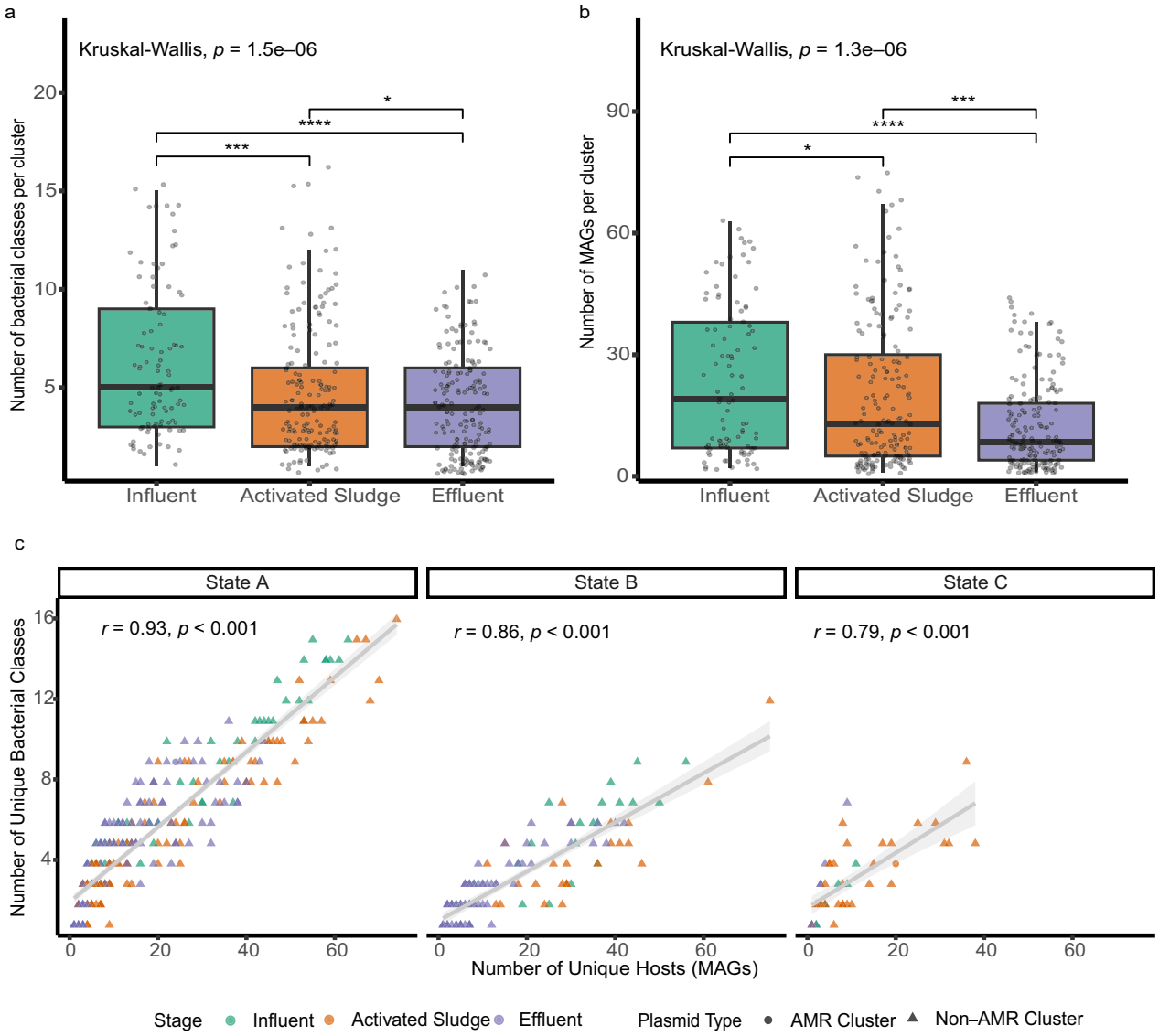


Fig. 4 | Plasmid host range narrows across wastewater treatment stages. a, b Plasmid host range: bacterial classes (**a**) and MAGs (**b**) per plasmid cluster across stages (Influent, Activated Sludge, Effluent), states pooled. Each point is one plasmid cluster with ≥1 classified host ($n = 97$, 179, and 186 clusters for Influent, Activated Sludge, and Effluent; biological replicates). Box: center line, median; bounds, 25th–75th percentiles; whiskers, minimum and maximum values within 1.5 × IQR; all points overlaid. Stages were compared by a two-sided Kruskal–Wallis test with two-sided Wilcoxon rank-sum pairwise tests (unadjusted). **a** $H = 26.8$, df = 2, $p = 1.5 \times 10^{-6}$; Inf vs AS, $p = 5.1 \times 10^{-4}$; Inf vs Eff, $p = 4.7 \times 10^{-7}$; AS vs Eff, $p = 0.024$. **b** $H = 27.1$, df = 2, $p = 1.3 \times 10^{-6}$; Inf vs AS, $p = 0.049$; Inf vs Eff, $p = 6.2 \times 10^{-7}$; AS vs Eff, $p = 5.2 \times 10^{-4}$. **c** Host range in MAGs (x) versus bacterial classes (y) per cluster, faceted by state ($n = 292$, 120, and 50 for States A, B, and C; biological replicates), colored by stage and shaped by plasmid type (AMR/non-AMR). Line, ordinary least-squares fit (center); band, 95% confidence interval. Per-state two-sided Spearman's rank correlation (r_s), computed independently per state (no adjustment for multiple comparisons): State A, $r_s = 0.93$, 95% CI [0.91, 0.94], $p = 3.1 \times 10^{-126}$; State B, $r_s = 0.86$, 95% CI [0.80, 0.90], $p = 1.4 \times 10^{-35}$; State C, $r_s = 0.79$, 95% CI [0.65, 0.88], $p = 1.0 \times 10^{-11}$. Colors denote treatment stage (Influent, green; Activated Sludge, orange; Effluent, purple). In **c**, point shape denotes plasmid type (circles, AMR clusters; triangles, non-AMR clusters). Significance in a and b: **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns, not significant.

90 plasmid clusters had Hi-C links to at least one ARG contig. All of the plasmid clusters with links to ARG contigs were also linked to one or more host MAGs. To assess confidence in associations between ARG contigs, plasmid clusters, and MAGs, we evaluated whether each ARG contig also had Hi-C contacts with the host MAG carrying its associated plasmid cluster (Fig. 5; Supplementary Fig. 9). Of the 90 plasmid clusters linked to ARG contigs, 43 (47.8%) showed Hi-C support across all three pairwise links (ARG–cluster, ARG–MAG, and MAG–cluster), hereafter referred to as concordant associations. For the remaining clusters (47 of 90; 52.2%), ARG–cluster links were detected, but the ARG contig did not show direct Hi-C contact with any of the host MAGs associated with that cluster. This may reflect insufficient Hi-C contacts from small or low-coverage ARG contigs, incomplete MAG recovery, or spurious ARG–cluster links. Both concordant and non-concordant associations are reported in Fig. 5 and Supplementary Data 4.

ARG profiles showed facility-specific patterns (Fig. 5d). In the State A WRRF, beta-lactam (*bla*), sulfonamide (*sul1*), and tetracycline (*tet*(C), *tet*(A)) resistance genes were detected across multiple host families, including *Rhodocyclaceae, OLB10* and *Competibacteraceae*. In the State B WRRF, beta-lactam, streptomycin, and tetracycline resistance genes were the most frequently detected, with *Burkholderiaceae* carrying a diverse set of ARG classes, including macrolide, quinolone, and sulfonamide markers.

https://doi.org/10.1038/s41467-026-77447-7

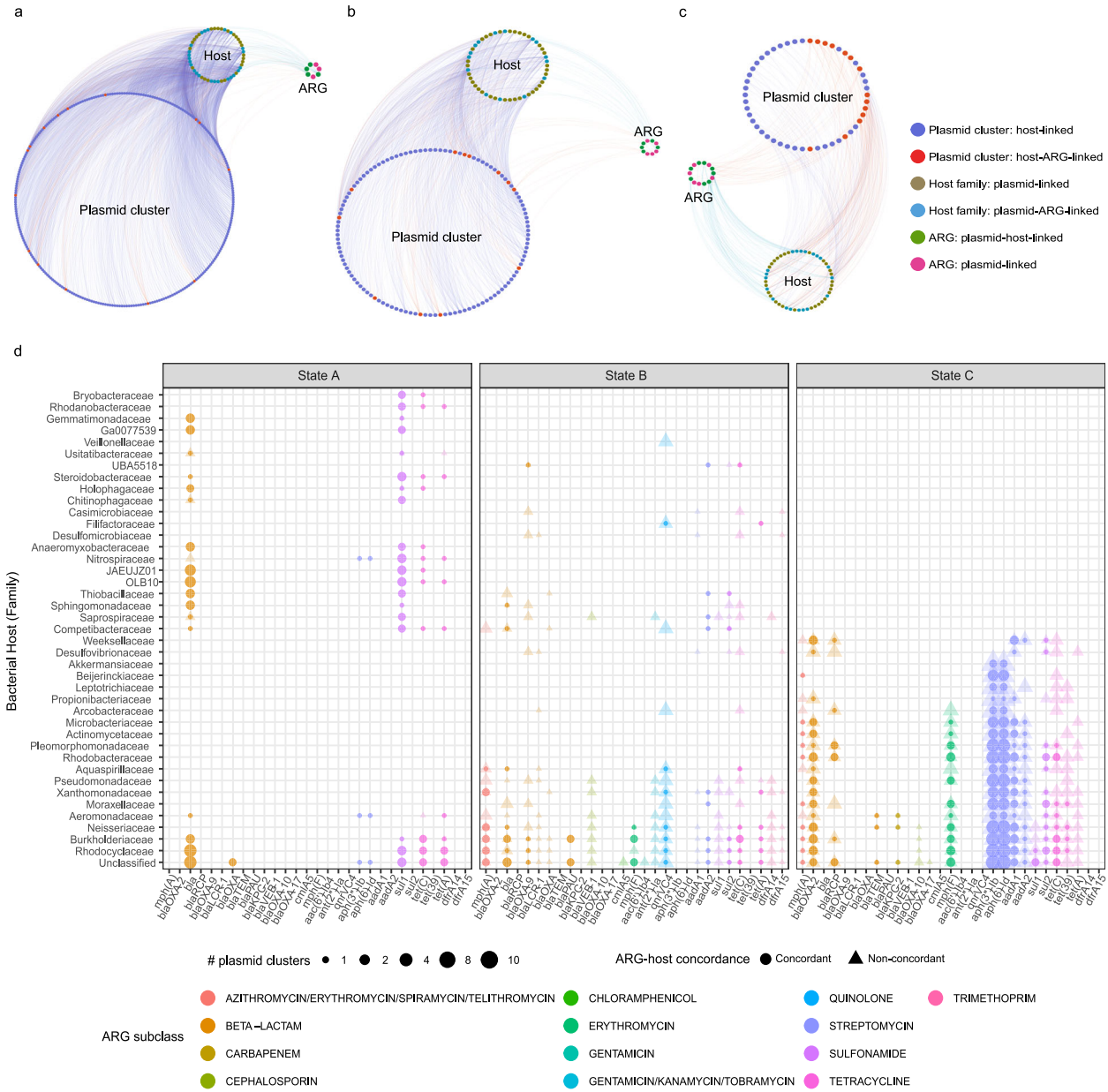


Fig. 5 | ARG–plasmid cluster–host family associations identified across three wastewater treatment systems. a–c Hi-C proximity linkages between ARGs, plasmid clusters, and bacterial host families in State A (**a**), State B (**b**), and State C (**c**). Nodes represent plasmid clusters, ARGs, and bacterial host families. Edges denote Hi-C proximity-based linkages: plasmid cluster–host family links, plasmid cluster–ARG links, and ARG–host family links. **d** ARG–host family associations grouped by state and ARG subclass. The top 20 host families per facility are shown; each point represents a unique ARG–host family pair; point size indicates the number of plasmid clusters, color denotes ARG subclass, and shape distinguishes concordant (all three pairwise Hi-C links detected; circle) from non-concordant (triangle) associations.

In the State C WRRF, streptomycin resistance genes *aph(3″)-Ib* and *aph(6)-Id* were the most prominent, each linked to 19 concordant host families across multiple plasmid clusters. Despite this facility-level variation, *tet*(C)–*Burkholderiaceae* associations were observed in all three states, suggesting a shared plasmid-associated tetracycline resistance signal across treatment configurations.

Discussion

We evaluated plasmid-host associations across wastewater treatment stages in three U.S. states to understand how plasmids persist in complex microbial communities. We achieved this by adopting a recently developed methodology to group plasmid contigs into plasmid clusters. To improve upon contig-based limitations in host assignment and plasmid resolution, we applied a graph-based clustering approach that integrates Hi-C contact data with sequence similarity, grouping contigs into coherent plasmid clusters. To reduce false-positive host associations, contigs encoding IS elements and transposases were removed prior to clustering, and only plasmid-derived contigs with ≥15 Hi-C contacts to another plasmid contig were retained[21]. The resulting plasmid clusters represent Hi-C-supported

groups of plasmid-derived contigs and may not correspond to complete plasmid genomes. They nonetheless provide a conservative basis for tracking plasmid distribution and mobility potential across treatment stages. We further identified persistent plasmid clusters, defined as plasmid clusters sharing ≥95% nucleotide identity across two or more treatment stages within the same facility and sampling week, and grouped the plasmid clusters based on the presence of mobility markers to evaluate host association patterns during wastewater treatment.

Across all WRRFs and treatment stages, plasmid-associated taxa constituted only a narrow subset of the total microbial community, indicating strong host–plasmid selectivity. Certain bacterial families (e.g., *Burkholderiaceae* and *Rhodocyclaceae*, and *Nitrospiraceae* in State A) were more represented among plasmid-linked hosts than expected from their relative abundance in the total community, suggesting higher permissiveness to plasmid acquisition and maintenance in these groups (Fig. 1c). These families harbor traits that likely support plasmid persistence. *Burkholderiaceae* are metabolically versatile and stress-tolerant, capable of producing antifungal compounds and withstanding oxidative stress, facilitating survival in dynamic environments[31,32]. Members of *Nitrospiraceae* exhibit redox adaptability and flexible energy metabolisms[33], while *Rhodocyclaceae*, such as *Azonexus*, are denitrifiers with broad substrate utilization, commonly found in wastewater systems[34]. These traits may facilitate their role as stable plasmid hosts in dynamic treatment environments. Plasmid persistence also depends on host fitness, conjugation rates, and interspecies competition[35]. Together, our findings suggest these families act as potential plasmid reservoirs in wastewater microbiomes. Future work should identify and characterize plasmid-host pairs that consistently promote plasmid maintenance to elucidate the ecological and genetic factors stabilizing plasmids in complex communities.

Despite standardized sampling across treatment stages, plasmid prevalence and host range varied substantially across facilities. These differences likely reflect variation in operational design, influent characteristics, and environmental conditions. Because we analyzed one WRRF per state and did not collect operational metadata, the following links between facility configuration and plasmid ecology are interpreted as hypotheses requiring multi-facility testing rather than established mechanisms. The WRRF in State A, which lacks primary clarification and uses a high-capacity activated sludge system with pure oxygen aeration, exhibited the broadest plasmid host range. This configuration may support high microbial biomass density and sustained aerobic activity, conditions favorable for long-term plasmid maintenance[25,26]. Prior studies have shown that aeration and high dissolved oxygen concentrations can enhance plasmid host diversity by shaping microbial communities more permissive to HGT[36,37]. In contrast, the WRRF in State B also employs pure oxygen aeration but includes primary clarification. Primary clarification may result in the removal of influent bacteria, resulting in a community of plasmid-linked hosts belonging to fewer taxonomic classes, which may explain why State B shows relatively high MAG-level associations but less expansion in class-level host range (Fig. 4c)[38].

The WRRF in State C, which operates a five-stage Bardenpho process, exhibited the lowest overall plasmid host range and plasmid burden across stages. The alternating aerobic, anoxic, and anaerobic conditions inherent to this configuration likely impose energetic and regulatory constraints that limit plasmid stability. Previous work suggests that redox fluctuations can impair plasmid retention and horizontal transfer[27,28]. Additionally, frequent redox shifts, elevated biomass turnover, and rapid changes in microbial community composition may impose selective pressure against metabolically costly plasmids, narrowing host range and reducing plasmid permissiveness[39,40]. The relatively small catchment population served by the WRRF in State C may have also contributed to the lower observed diversity of incoming bacterial taxa and plasmids. These findings underscore the potential influence of redox regime and influent source diversity as key drivers of plasmid ecology in WRRFs, shaping both plasmid persistence and host range.

We observed a significant reduction in plasmid host range from influent to effluent, which may reflect selective pressures during wastewater treatment[41,42]. Some persistent plasmid clusters showed changes in host family associations across treatment stages, particularly those carrying conjugative gene markers. Taxonomic profiling of the total microbial community showed that the genera associated with persistent plasmid clusters (e.g., *Novosphingobium, Phenylobacterium*) remained relatively stable in abundance across stages (Supplementary Fig. 5), indicating that changes in plasmid–host associations are not driven solely by shifts in host population abundance. Rather, near-identical plasmid clusters (≥99% identity) co-occurred with different host genera across stages, together with the presence of conjugation machinery. These observations suggest that plasmid-host compatibility and mobilization potential may contribute to persistence alongside environmental selection.

Persistent plasmid clusters spanning influent to effluent often encoded mobilization genes. In fact, nearly half (49.5%) of host-linked plasmid clusters encoded at least one mobilization or conjugation marker (e.g., *MOBQ, MOBP, MOBH*), type IV secretion system components, and adaptive traits such as oxidative-stress resistance, membrane repair, and nutrient acquisition functions (Supplementary Fig. 3; Supplementary Data 2 and 8). These features may contribute to plasmid maintenance under treatment-imposed stressors like redox fluctuations and nutrient limitation. Previous studies similarly found that WRRFs act as ecological filters, favoring mobile and functionally adaptive plasmids while attenuating others[43,44]. These findings underscore that both host dynamics and plasmid-intrinsic features contribute to the long-term stability and maintenance of plasmids in complex treatment environments.

Grouping plasmid–host networks by mobility status revealed structural differences consistent with distinct persistence modes. The network of plasmid clusters that contained conjugation and/or mobilization markers exhibited lower modularity and higher connectance than the network of clusters without detected mobilization markers, which may suggest a tendency toward broader host associations among mobility-marker-bearing clusters. At the individual-cluster level, conjugative persistent clusters such as A2-L5 were linked to taxonomically diverse genera across stages at near-identical sequence identity, while some clusters lacking mobilization markers, such as B1-L2, had associations with a single genus consistent with vertical co-inheritance. We note that host association variability was observed in both mobility groups, and that Hi-C captures physical co-localization within cells rather than confirmed transfer events. Assembly fragmentation inherent to short-read metagenomics may also limit recovery of complete plasmid sequences across stages. Definitive evidence of individual conjugation events would require validation, such as conjugation assays or plasmid isolation from representative hosts.

Our study has several limitations. First, Hi-C captures physical proximity, not definitive biological interaction, and our conclusions are accordingly limited to plasmid–host co-localization and association patterns. Because crosslinking is performed on intact cells prior to lysis, each read pair originates from DNA within the same cell rather than from neighboring cells in dense flocs[23,45]. Misassignment can still occur due to chimeric ligations, and we applied conservative thresholds to mitigate this[21,23,24]. Second, plasmid recovery and clustering are limited by read depth and contig resolution, which can underrepresent low-abundance or cryptic plasmids. We reduced these effects by normalizing Hi-C contacts to MAG-mapped reads and plasmid cluster lengths, but detection and assembly biases remain. A central trade-off in this analysis is the

removal of IS element– and transposase-encoding contigs prior to clustering, necessary to prevent multi-copy mobile elements from generating spurious Hi-C links. Because ARGs frequently co-localize with IS elements, this filtering excluded many ARG-bearing contigs. To address this, we used Hi-C contacts to link plasmid-derived ARG contigs to plasmid clusters and host MAGs, and evaluated concordant Hi-C support across all three pairwise links (ARG–cluster, ARG–MAG, and MAG–cluster) to increase confidence in these associations. The resulting ARG–host family associations, including *Rhodocyclaceae* and *Burkholderiaceae* as carriers of sulfonamide, beta-lactam, and tetracycline resistance genes, are consistent with previous reports from wastewater systems[46–48]. Third, without meta-transcriptomics or proteomics, we cannot determine the activity or expression of plasmid-borne genes, limiting functional inference[49]. Lastly, our sampling was limited to three states over three timepoints, which constrains statistical power and the generalizability of observed patterns. Broader spatiotemporal sampling, coupled with measurements of environmental and operational parameters, will be needed to rigorously test mechanistic drivers of plasmid persistence and host-association dynamics.

Overall, this study provides insights into plasmid–host associations in wastewater microbial communities and highlights that certain bacterial groups may play key roles in plasmid maintenance and host associations. Although wastewater microbial communities are highly diverse, plasmid-associated hosts are constrained to a limited number of bacterial families. Plasmid host range consistently narrowed from influent to effluent across all three WRRFs, which may reflect treatment-related selective effects. A subset of plasmid clusters persisted across stages, and mobility-stratified network analysis revealed that plasmids carrying mobilization markers were associated with broader, less modular host associations than plasmids lacking detected mobility markers. Analysis of ARG-plasmid-host associations further showed that plasmid clusters were associated with known resistance determinants across multiple drug classes, linking plasmid persistence to ARG ecology in engineered water systems.

Methods

Sample Collection and Concentration

Wastewater samples were collected weekly from three U.S. WRRFs in States A, B, and C, three times over a month, as described previously[50]. Primary influent was obtained as 24-hour composite samples (States A and B) or via grab sampling (State C). AS samples were collected via grab sampling from all sites, and effluent was grab-sampled after chlorination in States A and B, and combined UV/chlorination in State C. Effluent from State A WRRF underwent tertiary filtration to remove fine solids prior to chlorination. Samples from WRRFs in States A and C were shipped on ice to State B and stored at 4 °C until processing. State A WRRF does not employ primary clarification; instead, influent undergoes preliminary screening and flows directly into the activated sludge process. State C WRRF operated a five-stage Bardenpho system with filtration and UV disinfection, while WRRFs in States A and B used high-purity oxygen activated sludge systems. All WRRFs are discharged to local surface waters. Influent and AS samples were aliquoted into 50 mL and 2.0 mL tubes, respectively, centrifuged at 12,000 × g for 10 min at 4 °C, and pellets were stored at 4 °C. Effluent (500 mL) was centrifuged at 3,500 × g for 30 min, pooled, recentrifuged, and pellets stored at 4 °C to preserve cell integrity for Hi-C crosslinking, which requires intact cells. All samples were processed in duplicate: one replicate for Hi-C sequencing and the other for metagenomic DNA extraction.

Hi-C sample preparation

One replicate per sample (by site, type, and date) was processed using the ProxiMeta™ Hi-C Kit (Phase Genomics, Seattle, WA, USA) following protocol v4.5.4, with modified crosslinking. Briefly, samples were incubated with 10 mL of chilled 1× PBS containing 1% formaldehyde for 10 min at room temperature, then quenched with 527 µL of 2.5 M glycine for 5 min. After a 10 min incubation on ice with agitation, samples were centrifuged at 3500 × *g* for 10 min at 4 °C, washed twice with cold 1× PBS, and resuspended in 1.0 mL PBS. A final spin at 800 × *g* for 5 min at 4 °C was performed before storing the pellets at −80 °C. Remaining steps followed the manufacturer's instructions.

DNA extraction and sequencing

Genomic DNA was extracted from pellet replicates using the Maxwell® RSC 48 system and Blood DNA Kit (Promega, Madison, WI, USA), with modified lysis to address solids in wastewater. Pellets were combined with 0.5 g zirconia beads (BioSpec) and 300 µL lysis buffer, bead-beaten at 2400 RPM for 2 minutes, and centrifuged at 15,000 × *g* for 5 min. The supernatant (200 µL) was collected. This process was repeated twice with fresh lysis buffer, yielding 600 µL total lysate. Proteinase K (60 µL) was added, followed by incubation at 56 °C for 30 minutes. DNA extracts were submitted to the Oklahoma Medical Research Foundation Clinical Genomics Core (Oklahoma City, OK) for library preparation using the Watchmaker DNA Library Prep Kit (Watchmaker Genomics, Boulder, CO) and sequenced on an Illumina NovaSeq X Plus (Illumina, San Diego, CA) using 150 bp paired-end reads to a target depth of 100 million reads per sample. Per-sample sequencing yields and read counts are reported in Supplementary Data 5–6 and Supplementary Fig. 6a–b. Indexed Hi-C libraries were pooled and sequenced separately at the Duke Sequencing and Genomic Technologies Core (Durham, NC) under the same platform and read depth.

Shotgun metagenomic and Hi-C data analysis

Shotgun reads were quality-filtered and adapter-trimmed using BBDuk (BBTools v39.01, Joint Genome Institute) with the following parameters: k = 23, ktrim=r, mink=12, hdist=1, minlength=50, –tpe, –tbo for adapter removal, and qtrim=rl, trimq=10, chastityfilter=True for low-quality base trimming. Trimmed reads were assembled using MEGAHIT (v1.2.9) with default parameters[51]. Shotgun and Hi-C reads were mapped to the assemblies using Bowtie2 (v2.5.3) and processed with Samtools (v1.18), while Hi-C-specific alignments were performed using BWA-MEM (v0.7.17) with default parameters[52–54]. Genome binning was conducted with MetaCC (v1.2.0), which uses Hi-C contact frequencies to cluster contigs[55]. Contact matrices were derived from Hi-C BAM files using the custom script bam2links.py, which parses Hi-C BAM alignments and counts read pairs linking distinct contigs to build a contig-by-contig contact matrix, as described by Press et al.[56]. MAGs were annotated using Prokka (v1.14.6), taxonomically classified with GTDB-Tk (v2.2.6), and assessed for quality using CheckM2 (v1.0.1)[57–59]. In total, binning across all samples recovered 993 metagenome-assembled genomes (MAGs) for State A, 1,475 MAGs for State B, and 882 MAGs for State C. Of these, 79 (2.4%) were high quality (completeness ≥90%, contamination ≤5%), 348 (10.4%) were medium quality (completeness ≥50%, contamination ≤10%), and the remainder were partial genomes (Supplementary Data 7). MAG quality (completeness and contamination) was comparable across facilities (Supplementary Fig. 6). While MAGs can represent near-complete genomes, most MAGs in this study were partial. Accordingly, we interpret MAGs more broadly as genomes or genome fragments from closely related strains within a species. Plasmid contigs were identified and characterized using geNomad (v1.8.1) end-to-end command[60].

Plasmid clusters identification and functional annotation

Putative plasmid clusters were identified using a Hi-C-based approach adapted from Risely et al.[21]. First, protein-coding genes were predicted from assembled contigs using Prodigal (v2.6.3) in metagenomic mode

(-p meta)[61]. To minimize false-positive associations driven by mobile elements, contigs encoding transposases or insertion sequences (IS elements) were excluded. These were identified via BLASTp searches (e-value < 0.01) against curated transposase databases from ISfinder and the Tn3 Transposon Finder[62,63]. Contigs predicted to be plasmid-derived using geNomad were retained only if they exhibited strong linkage (≥15 Hi-C contacts) to at least one other plasmid contig, reflecting the assumption that genuine plasmid-derived contigs tend to co-occur stably in Hi-C networks. While this filtering may exclude small or non-transferable plasmids, it helps enrich for conjugative or mobilizable elements that are more ecologically relevant to HGT, as noted by Risely et al.[21]. Graph-based clustering of the retained contigs was then performed using the Walktrap algorithm (igraph::walktrap.community, step length = 10)[64], which groups contigs that share frequent Hi-C contacts into clusters, yielding a total of 944 plasmid clusters (Supplementary Fig. 7). Each plasmid cluster represents a group of plasmid-derived contigs that share Hi-C contacts; clusters may represent complete or partial plasmid sequences given the fragmented nature of short-read assemblies. Plasmid-cluster lengths ranged from 2,051 to 140,087 bp and were largely comparable across states (Supplementary Fig. 8). Lengths fall within reported ranges for natural plasmids[65,66]. Hi-C contact data were then integrated to identify host associations, revealing 481 plasmid clusters linked to at least one metagenome-assembled genome (MAG). These host-linked clusters were used in subsequent analyses. Hi-C contact counts were normalized for differences in host abundance and plasmid size, following the general approach of Risely et al. and Stalder et al.[21,23]. Specifically, for each MAG-plasmid pair, the raw Hi-C contact value was scaled by the average shotgun-derived MAG read count across all MAGs in the dataset relative to the read count for that MAG, and by the average plasmid cluster length across all plasmid clusters relative to the length of the individual plasmid cluster. This correction reduces detection biases introduced by variation in microbial abundance and plasmid size, enabling more comparable contact frequencies across samples.

Functional annotation targeted plasmid mobility and virulence. MOB-suite (v3.1.9; https://github.com/phac-nml/mob-suite) was used to classify plasmid clusters based on mobility-associated genes, including relaxases, oriT sites and MPF systems[67]. Virulence factors were identified using ABRicate (v1.0.1; https://github.com/tseemann/abricate) with the Virulence Factors Database (VFDB) (≥75% identity, ≥70% coverage)[68].

Mobility annotation and network stratification

Each plasmid cluster was assigned a mobility category by integrating MOB-suite marker detection with conjugation-gene identification via BLASTp against curated databases. Clusters were categorized hierarchically: conjugative (MPF evidence present), mobilizable/conjugative (relaxase/MOB or oriT detected without MPF), or no marker detected. For network analyses, these were collapsed into a binary classification: Marker Positive (conjugative or mobilizable/conjugative) versus No Marker. Weighted bipartite networks connecting plasmids to host families were constructed separately for each mobility class, and network metrics (connectance, modularity, community number) were computed using igraph (v2.0.3)[64].

Persistent plasmid cluster identification

To evaluate plasmid persistence across treatment stages, plasmid clusters from different stages within the same facility and sampling week were compared using BLASTn (≥95% nucleotide identity, e-value ≤ 10^{-10}). For each cluster pair, we computed symmetric coverage (sym_cov), defined as the minimum fraction of either cluster's total assembled sequence covered by the other. Clusters sharing ≥95% identity and sym_cov ≥0.5 across two or more stages were defined as persistent plasmid clusters and classified by the stages they span (Inf-AS, AS-Eff, Inf-Eff, or Inf-AS-Eff). Each persistent cluster was assigned a mobility label based on the strongest mobility category among its members.

Hi-C-based assignment of ARG contigs to plasmid clusters

Because IS/transposase-encoding contigs were removed prior to clustering, ARG-bearing fragments were excluded from plasmid clusters. AMR genes were identified using AMRFinderPlus (v3.12.8) prior to IS/transposon filtering[69]; only ARG contigs classified as plasmid-derived by geNomad were retained. For each ARG contig, Hi-C contact support was quantified to (i) host MAGs and (ii) plasmid clusters using contig-length and mapped-depth-normalized linkage strengths; ARG–MAG linkages were additionally normalized by MAG abundance. Concordance was evaluated by testing whether each ARG contig showed direct Hi-C contact with the host MAG associated with its assigned plasmid cluster (Supplementary Methods).

Statistical analysis

Community structure was assessed via NMDS based on Bray-Curtis dissimilarity. For each facility and stage, the three weekly samples were aggregated to produce a single representative community profile, reducing week-to-week variability. Disaggregated per-week analyses were additionally performed to assess the robustness of facility-level patterns to this aggregation. Facility-level differences in community composition were evaluated using PERMANOVA[70]. Alpha diversity was compared across facilities and treatment stages using Kruskal–Wallis[71]. The number of plasmid clusters associated per MAG and host range per plasmid cluster were compared across facilities using Wilcoxon rank-sum tests with Benjamini-Hochberg adjustment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequencing reads and metagenome-assembled genomes generated in this study have been deposited in the NCBI database under BioProject accession code PRJNA1295499. The processed data generated in this study are available at Zenodo (https://doi.org/10.5281/zenodo.21381023)[72]. Source data are provided with this paper.

Code availability

The code used in this study is available at GitHub (https://github.com/SiyiZhou6/HiC-plasmid-wastewater) and archived at Zenodo (https://doi.org/10.5281/zenodo.21381419)[73].

References

1. Bennett, P. M. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *Br. J. Pharm.* **153**, S347–S357 https://doi.org/10.1038/sj.bjp.0707607 (2008).
2. Smalla, K., Jechalke, S. & Top, E.M. Plasmid Detection, Characterization, and Ecology. *Microbiol Spectr.* American Society for Microbiology **3**, https://doi.org/10.1128/microbiolspec.plas-0038-2014 (2015).
3. Coluzzi, C., Garcillán-Barcia, M. P., de la Cruz, F. & Rocha, E. P. C. Evolution of Plasmid Mobility: Origin and Fate of Conjugative and Nonconjugative Plasmids. *Mol. Biol. Evol.* **39**, msac115 https://doi.org/10.1093/molbev/msac115 (2022).
4. Barlow, M. What antimicrobial resistance has taught us about horizontal gene transfer. *Methods Mol. Biol. Clifton NJ* **532**, 397–411 https://doi.org/10.1007/978-1-60327-853-9_23 (2009).
5. Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C. & de la Cruz F. Mobility of Plasmids. *Microbiol Mol. Biol. Rev. Am. Soc. Microbiol.* **74**, 434–452 https://doi.org/10.1128/MMBR.00020-10 (2010).

6. Che, Y. et al. Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. *Proc. Natl. Acad. Sci. USA* **118**, e2008731118 https://doi.org/10.1073/pnas.2008731118 (2021).
7. Sünderhauf, D., Klümper, U., Gaze, W. H., Westra, E. R. & van Houte, S. Interspecific competition can drive plasmid loss from a focal species in a microbial community. *ISME J. Nat. Publ. Group* **17**, 1765–1773 https://doi.org/10.1038/s41396-023-01487-w (2023).
8. Benz, F. & Hall, A. R. Host-specific plasmid evolution explains the variable spread of clinical antibiotic-resistance plasmids. *Proc. Natl. Acad. Sci. USA* **120**, e2212147120 https://doi.org/10.1073/pnas.2212147120 (2023).
9. Bottery, M. J. Ecological dynamics of plasmid transfer and persistence in microbial communities. *Curr. Opin. Microbiol* **68**, 102152 https://doi.org/10.1016/j.mib.2022.102152 (2022).
10. Li, L.-G. & Zhang, T. Plasmid-mediated antibiotic resistance gene transfer under environmental stresses: Insights from laboratory-based studies. *Sci. Total Environ.* **887**, 163870 https://doi.org/10.1016/j.scitotenv.2023.163870 (2023).
11. van Elsas, J. D. & Bailey, M. J. The ecology of transfer of mobile genetic elements. *FEMS Microbiol Ecol.* **42**, 187–197 https://doi.org/10.1111/j.1574-6941.2002.tb01008.x (2002).
12. Cook, L. C. C. & Dunny, G. M. The influence of biofilms in the biology of plasmids. *Microbiol Spectr.* **2**, 0012 https://doi.org/10.1128/microbiolspec.PLAS-0012-2013 (2014).
13. Schlüter, A. et al. The 64508 bp IncP-1β antibiotic multiresistance plasmid pB10 isolated from a waste-water treatment plant provides evidence for recombination between members of different branches of the IncP-1β group. *Microbiol. Microbiol. Soc.* **149**, 3139–3153 https://doi.org/10.1099/mic.0.26570-0 (2003).
14. Schlüter, A., Szczepanowski, R., Pühler, A. & Top, E. M. Genomics of IncP-1 antibiotic resistance plasmids isolated from wastewater treatment plants provides evidence for a widely accessible drug resistance gene pool. *FEMS Microbiol Rev.* **31**, 449–477 https://doi.org/10.1111/j.1574-6976.2007.00074.x (2007).
15. Mosaka, T. B. M., Unuofin, J. O., Daramola, M. O., Tizaoui, C. & Iwarere S. A. Inactivation of antibiotic-resistant bacteria and antibiotic-resistance genes in wastewater streams: Current challenges and future perspectives. *Front Microbiol [Internet]. Frontiers***13** [cited 2025 Mar 5] https://doi.org/10.3389/fmicb.2022.1100102 (2023).
16. Zhang, S. et al. Chlorine disinfection facilitates natural transformation through ROS-mediated oxidative stress. *ISME J.* **15**, 2969–2985 https://doi.org/10.1038/s41396-021-00980-4 (2021).
17. Lopatkin, A. J. et al. Persistence and reversal of plasmid-mediated antibiotic resistance. *Nat. Commun. Nat. Publ. Group* **8**, 1689 https://doi.org/10.1038/s41467-017-01532-1 (2017).
18. Alonso-del Valle, A. et al. Variability of plasmid fitness effects contributes to plasmid persistence in bacterial communities. *Nat. Commun. Nat. Publ. Group* **12**, 2653 https://doi.org/10.1038/s41467-021-22849-y (2021).
19. Ma, J. et al. Horizontal transfer potential of antibiotic resistance genes in wastewater treatment plants unraveled by microfluidic-based mini-metagenomics. *J. Hazard Mater.* **465**, 133493 https://doi.org/10.1016/j.jhazmat.2024.133493 (2024).
20. Zhang, Y. et al. High-throughput single-cell sequencing of activated sludge microbiome. *Environ. Sci. Ecotechnology* **23**, 100493 https://doi.org/10.1016/j.ese.2024.100493 (2025).
21. Risely, A. et al. Host- plasmid network structure in wastewater is linked to antimicrobial resistance genes. *Nat. Commun. Nat. Publ. Group* **15**, 555 https://doi.org/10.1038/s41467-024-44827-w (2024).
22. McCorison, C. B., Kim, T., Donato, J. J. & LaPara, T. M. Proximity-Ligation Metagenomic Sequence Analysis Reveals That the Antibiotic Resistome Makes Significant Transitions During Municipal Wastewater Treatment. *Environ. Microbiol* **27**, e70036 https://doi.org/10.1111/1462-2920.70036 (2025).
23. Stalder, T., Press, M. O., Sullivan, S., Liachko, I. & Top, E. M. Linking the resistome and plasmidome to the microbiome. *ISME J. Nat. Publ. Group* **13**, 2437–2446 https://doi.org/10.1038/s41396-019-0446-4 (2019).
24. McCallum, G. E. et al. Noise reduction strategies in metagenomic chromosome confirmation capture to link antibiotic resistance genes to microbial hosts. *Micro. Genomics. Microbiol. Soc.* **9**, 001030 https://doi.org/10.1099/mgen.0.001030 (2023).
25. Pei, R. & Gunsch, C. K. Plasmid Conjugation in an Activated Sludge Microbial Community. *Environ. Eng. Sci. Mary Ann. Liebert, Inc., publishers* **26**, 825–831 https://doi.org/10.1089/ees.2008.0236 (2009).
26. Li, L. et al. Plasmids persist in a microbial community by providing fitness benefit to multiple phylotypes. *ISME J.* **14**, 1170–1181 https://doi.org/10.1038/s41396-020-0596-4 (2020).
27. Zhou, S. et al. Metagenomic insights into the distribution, mobility, and hosts of extracellular antibiotic resistance genes in activated sludge under starvation stress. *Water Res* **236**, 119953 https://doi.org/10.1016/j.watres.2023.119953 (2023).
28. Jong, M.-C., Harwood, C. R., Blackburn, A., Snape, J. R. & Graham, D. W. Impact of Redox Conditions on Antibiotic Resistance Conjugative Gene Transfer Frequency and Plasmid Fate in Wastewater Ecosystems. *Environ. Sci. Technol. Am. Chem. Soc.* **54**, 14984–93 https://doi.org/10.1021/acs.est.0c03714 (2020).
29. Sen, D. et al. Broad-Host-Range Plasmids from Agricultural Soils Have IncP-1 Backbones with Diverse Accessory Genes. *Appl Environ. Microbiol. Am. Soc. Microbiol.* **77**, 7975–83 https://doi.org/10.1128/AEM.05439-11 (2011).
30. Brown, C. J. et al. Diverse Broad-Host-Range Plasmids from Freshwater Carry Few Accessory Genes. *Appl Environ. Microbiol* **79**, 7684–7695 https://doi.org/10.1128/AEM.02252-13 (2013).
31. Carrión, V. J. et al. Involvement of Burkholderiaceae and sulfurous volatiles in disease-suppressive soils. *ISME J.* **12**, 2307–2321 https://doi.org/10.1038/s41396-018-0186-x (2018).
32. DeShazer, D. A novel contact-independent T6SS that maintains redox homeostasis via Zn2+ and Mn2+ acquisition is conserved in the Burkholderia pseudomallei complex. *Microbiol Res* **226**, 48–54 https://doi.org/10.1016/j.micres.2019.05.007 (2019).
33. Bayer, B. et al. Metabolic versatility of the nitrite-oxidizing bacterium Nitrospira marina and its proteomic response to oxygen-limited conditions. *ISME J.* **15**, 1025–1039 https://doi.org/10.1038/s41396-020-00828-3 (2021).
34. Petri, R. M. et al. Changes in the Rumen Epimural Bacterial Diversity of Beef Cattle as Affected by Diet and Induced Ruminal Acidosis. *Appl Environ. Microbiol* **79**, 3744–3755 https://doi.org/10.1128/AEM.03983-12 (2013).
35. Benz, F. et al. Plasmid- and strain-specific factors drive variation in ESBL-plasmid spread in vitro and in vivo. *ISME J. Nat. Publ. Group* **15**, 862–878 https://doi.org/10.1038/s41396-020-00819-4 (2021).
36. Qiu, T. et al. Impact of aeration rate on the transfer range of antibiotic-resistant plasmids during manure composting. *Environ. Pollut.* **361**, 124851 https://doi.org/10.1016/j.envpol.2024.124851 (2024).
37. Qin, L. et al. Impact of Dissolved Oxygen on the Performance and Microbial Dynamics in Side-Stream Activated Sludge Hydrolysis Process. *Water Multidiscip. Digital Publ. Inst.* **15**, 1977 https://doi.org/10.3390/w15111977 (2023).
38. Riisgaard-Jensen, M., Dottorini, G., Nierychlo, M. & Nielsen, P. H. Primary settling changes the microbial community of influent wastewater to wastewater treatment plants. *Water Res* **244**, 120495 https://doi.org/10.1016/j.watres.2023.120495 (2023).
39. Kempf, I., Le Devendec, L., Lucas, P., Druilhe, C. & Pourcher, A.-M. Impact of mesophilic anaerobic digestion and post-treatment of

digestates on the transfer of conjugative antimicrobial resistance plasmids. *Waste Manag* **152**, 1–5 https://doi.org/10.1016/j.wasman.2022.08.002 (2022).
40. Wang, Y. et al. Redox gradients drive microbial community assembly patterns and molecular ecological networks in the hyporheic zone of effluent-dominated rivers. *Water Res* **248**, 120900 https://doi.org/10.1016/j.watres.2023.120900 (2024).
41. Brown, C. L. et al. Selection and horizontal gene transfer underlie microdiversity-level heterogeneity in resistance gene fate during wastewater treatment. *Nat. Commun. Nat. Publ. Group* **15**, 5412 https://doi.org/10.1038/s41467-024-49742-8 (2024).
42. Jäger, T. et al. Reduction of antibiotic-resistant bacteria during conventional and advanced wastewater treatment, and the disseminated loads released to the environment. *Front Microbiol [Internet]. Frontiers* **9** [cited 2025 Apr 3] https://doi.org/10.3389/fmicb.2018.02599 (2018).
43. An, X.-L. et al. Impact of wastewater treatment on the prevalence of integrons and the genetic diversity of integron gene cassettes. *Appl Environ. Microbiol* **84**, e02766–17 https://doi.org/10.1128/AEM.02766-17 (2018).
44. Zorea, A. et al. Plasmids in the human gut reveal neutral dispersal and recombination that is overpowered by inflammatory diseases. *Nat. Commun. Nat. Publ. Group* **15**, 3147 https://doi.org/10.1038/s41467-024-47272-x (2024).
45. Press, M. O. et al. Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions [Internet]. *bioRxiv* 198713 [cited 2026 Mar 24] https://doi.org/10.1101/198713 (2017).
46. Knecht, C. A. et al. Identification of Antibiotic Resistance Gene Hosts in Treatment Wetlands Using a Single-Cell Based High-Throughput Approach. *Water Multidiscip. Digital Publ. Inst.* **15**, 2432 https://doi.org/10.3390/w15132432 (2023).
47. Tian, Y., Li, J., Meng, J. & Li, J. Deciphering antibiotic resistance gene transfer in activated sludge systems for piggery wastewater: behaviors, hosts and drivers. *Environ. Res* **283**, 122166 https://doi.org/10.1016/j.envres.2025.122166 (2025).
48. Smyth, C., Leigh, R. J., Do, T. T. & Walsh, F. Communities of plasmids as strategies for antimicrobial resistance gene survival in wastewater treatment plant effluent. *Npj Antimicrob. Resist* **3**, 78 https://doi.org/10.1038/s44259-025-00151-x (2025).
49. Bashiardes, S., Zilberman-Schapira, G. & Elinav, E. Use of Metatranscriptomics in Microbiome Research. *Bioinforma. Biol. Insights* **10**, 19–25 https://doi.org/10.4137/BBI.S34610 (2016).
50. Philo, S. E. et al. Assessing resistome host range across water reclamation in three geographically distinct communities using Hi-C sequencing. *Appl Environ. Microbiol. Am. Soc. Microbiol.* **0**, e00701–e00726 https://doi.org/10.1128/aem.00701-26 (2026).
51. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 https://doi.org/10.1093/bioinformatics/btv033 (2015).
52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 https://doi.org/10.1038/nmeth.1923 (2012).
53. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 https://doi.org/10.1093/bioinformatics/btp352 (2009).
54. Li. H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [Internet]. arXiv.org. 2013 [cited 2025 June 16]. https://arxiv.org/abs/1303.3997v2 Accessed (2025).
55. Du, Y. & Sun, F. MetaCC allows scalable and integrative analyses of both long-read and short-read metagenomic Hi-C data. *Nat. Commun. Nat. Publ. Group* **14**, 6231, https://doi.org/10.1038/s41467-023-41209-6 (2023).
56. Press M., Stalder T., Genomics P. Arg-Plasmid-Genome connections from Hi-C. OSF [cited 2025 June 16]; https://osf.io/ezb8j/ Accessed 16 June 2025 (2018).
57. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 https://doi.org/10.1093/bioinformatics/btu153 (2014).
58. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 https://doi.org/10.1093/bioinformatics/btz848 (2020).
59. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043–1055 https://doi.org/10.1101/gr.186072.114 (2015).
60. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol. Nat. Publ. Group* **42**, 1303–1312 https://doi.org/10.1038/s41587-023-01953-y (2024).
61. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma.* **11**, 119 https://doi.org/10.1186/1471-2105-11-119 (2010).
62. Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**, D32–D36 https://doi.org/10.1093/nar/gkj014 (2006).
63. Ross, K. et al. TnCentral: a Prokaryotic Transposable Element Database and Web Portal for Transposon Analysis. mBio. 12:e02060-21 https://doi.org/10.1128/mBio.02060-21
64. The igraph software package for complex network research | BibSonomy [Internet]. [cited 2025 June 16]. https://www.bibsonomy.org/bibtex/bb49a4a77b42229a427fec316e9fe515 (Accessed 16 June 2025).
65. Dewan, I. & Uecker, H. Is the distribution of plasmid lengths bimodal? *Plasmid* **102721**, 129–130. https://doi.org/10.1016/j.plasmid.2024.102721 (2024).
66. Finks, S. S. & Martiny, J. B. H. Plasmid-Encoded Traits Vary across Environments. *mBio.* **14**, e03191-22 https://doi.org/10.1128/mbio.03191-22.
67. Robertson, J. & Nash, J. H. E. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Micro. Genomics* **4**, e000206, https://doi.org/10.1099/mgen.0.000206 (2018).
68. Chen, L. et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* **33**, D325–D328 https://doi.org/10.1093/nar/gki008 (2005).
69. Feldgarden, M. et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep. Nat. Publ. Group* **11**, 12728, https://doi.org/10.1038/s41598-021-91456-0 (2021).
70. Oksanen, J. et al. vegan: Community Ecology Package. R package version 2.8-0, https://vegandevs.github.io/vegan/ (2026).
71. McKight, P. E. & Najab, J. Kruskal-Wallis Test. Corsini Encycl Psychol [Internet]. *John Wiley & Sons*, Ltd; [cited 2025 June 16]. 1–1 https://doi.org/10.1002/9780470479216.corpsy0491 (2010).
72. Zhou, S. et al. Supporting data for "Plasmid–Host Networks Reveal Persistent Mobile Elements Across Wastewater Treatment."*Zenodo* [cited 2026 July 15] https://doi.org/10.5281/zenodo.21381023 (2026).
73. Siyi_Zhou. SiyiZhou6/HiC-plasmid-wastewater: Initial release for publication [Internet]. *Zenodo* [cited 2026 July 15] https://doi.org/10.5281/ZENODO.21381419 (2026).

Acknowledgements

We thank the wastewater utilities for providing access to sampling sites and enabling sample collection.

Author contributions

S.Z. designed the study, collected samples, conducted computational analyses, and drafted the manuscript. S.E.P. and M.A.S. contributed to sample collection, laboratory processing, and manuscript revision. J.D.V. contributed to data interpretation and manuscript editing. A.L.S. provided funding support and contributed to manuscript revision. L.B.S. supervised the project, contributed to the study design, and provided critical feedback. All authors reviewed and approved the final manuscript and agreed to be accountable for all aspects of the work.

Funding

This work was supported by the US–Egypt Science and Technology Joint Fund (NAS Grant G10001728) and under a Cooperative Agreement (W9132T-23-2-0002) with the U.S. Army Corps of Engineers, Engineer Research and Development Center, Construction Engineering Research Laboratory (USACE ERDC-CERL). This work was supported in part by the Engineering Research Centers Program of the National Science Foundation under NSF Cooperative Agreement No. EEC-2133504 and by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award R21AI190938.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-77447-7.

Correspondence and requests for materials should be addressed to Lauren B. Stadler.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026