Supplemental information

Plasmids as Persistent Genetic Reservoirs of Bacterial Defense Systems in Wastewater Treatment

Haotian Zheng[1], Leighton Payne[1], Wanli He[1], Mario Rodríguez Mestre[1], Lili Yang[1], Arnaud Dechesne[2], Rafael Pinilla-Redondo[1], Joseph Nesme[1], Søren J. Sørensen[1,3,*]

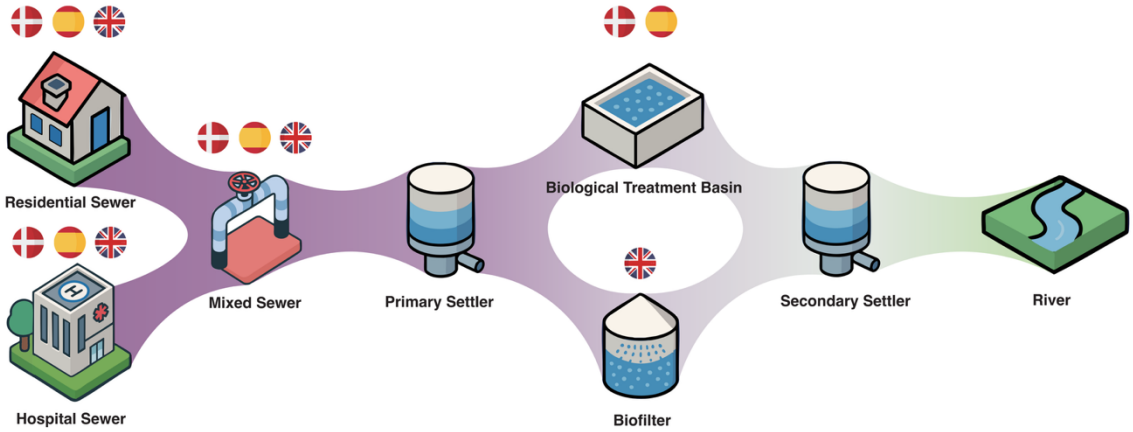

Figure S1. Overview of the UWS Treatment Flow and Sampling Strategy. This schematic illustrates the key treatment steps and corresponding sampling points across three representative urban wastewater systems (UWSs). Wastewater originated from hospital sewage (HS) and residential sewage (RS) first entering the mixed sewer (MS) pipeline, followed by passage through the primary settler. In Denmark and Spain, the flow subsequently continued into a biological treatment basin, whereas in the United Kingdom, a biofilter was used instead. Effluent from all systems then passed through a secondary settler before being discharged into the downstream river (RU). The infrastructure layout and sampling framework were adapted from (Yu et al. 2024), which provided a detailed description of the UWS designs. National flags indicate the sampling sites corresponding to each country's UWS.

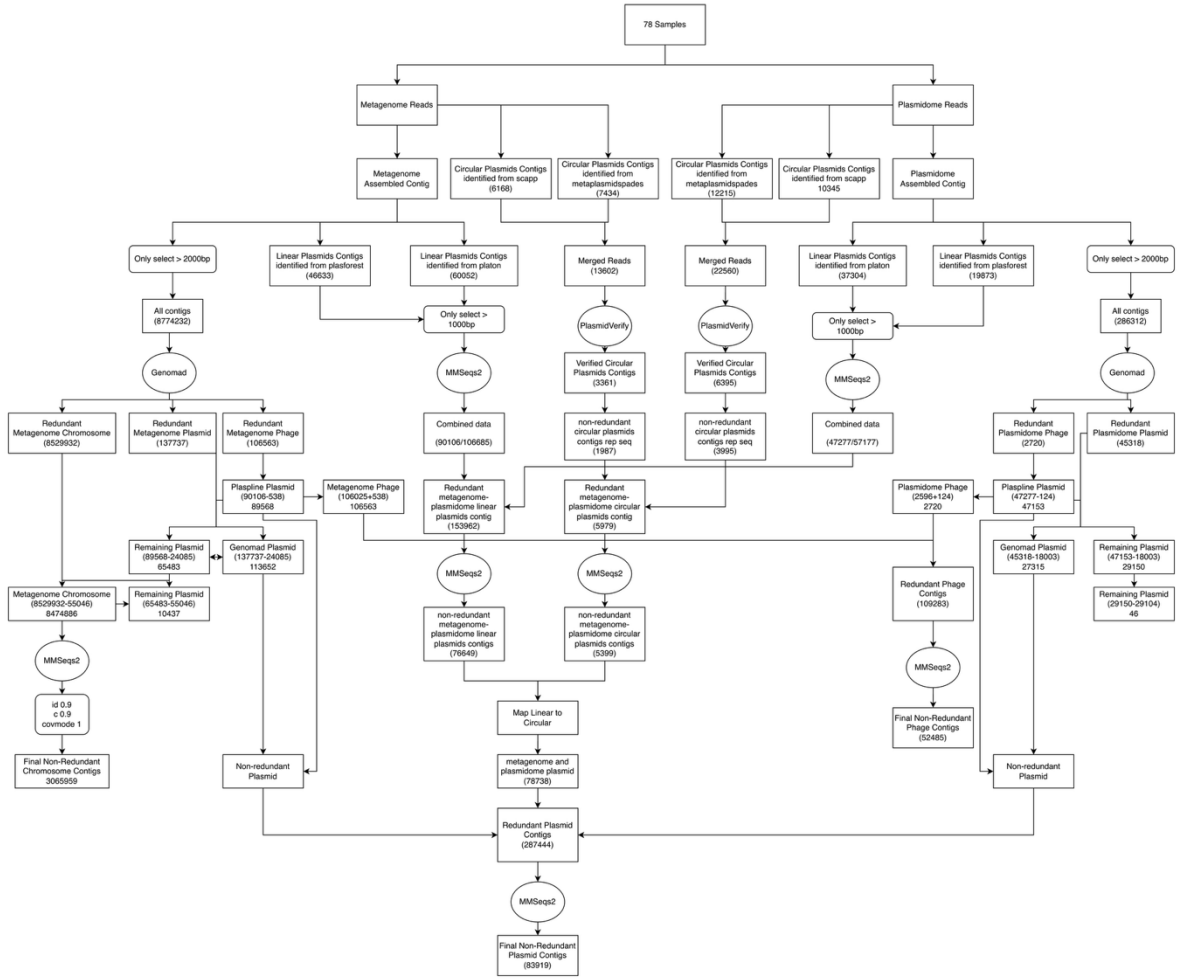

Figure S2. Workflow for the Identification and Classification of Chromosome, Plasmids and Phages Contigs from Metagenomic and Plasmidome Datasets. This workflow outlines the processing of 78 samples through parallel metagenome and plasmidome assembly pipelines. Plasmid contigs, both linear and circular, are identified using Platon, PlasForest, SCAPP, and MetaPlasmidSPAdes. Contig classification into chromosomes, plasmids and phages is performed using GeNomad and Plaspline. Redundant sequences are removed using MMSeqs2 with a 90% identity and 90% coverage threshold. Circular plasmids are further validated with PlasmidVerify. The final output consists of 3,065,959 chromosomal contigs, 83,919 plasmid contigs, and 52,485 phage contigs. Numbers in parentheses indicate the number of contigs retained at each processing step.

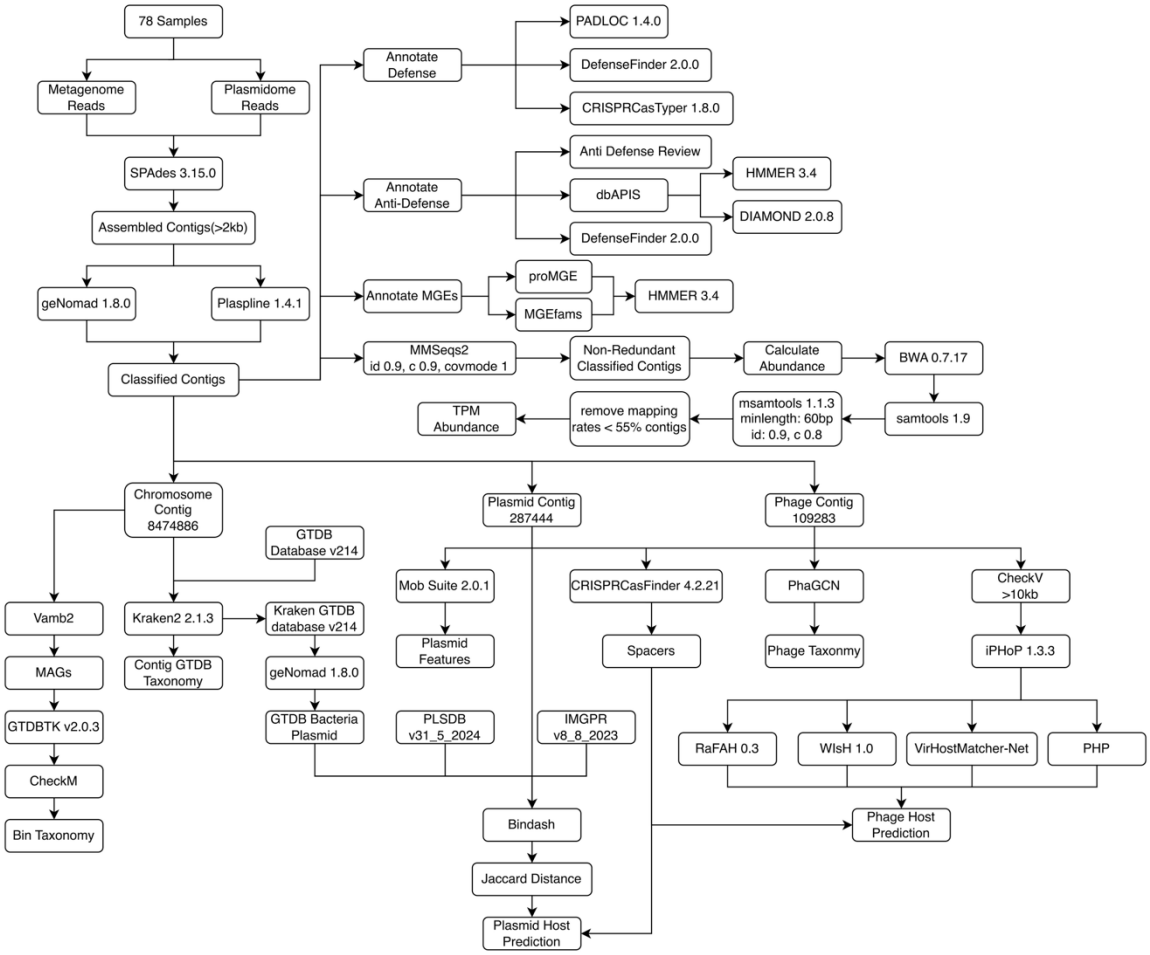

Figure S3. Gene annotation and downstream analysis pipeline for classified genomic contigs. Functional annotation of classified contigs from 78 samples using multiple tools including PADLOC, DefenseFinder, and CRISPRCasTyper for defense systems; dbAPIS database for anti-defense annotation; proMGE database and MGEfams for MGEs annotation; and various specialized tools for plasmid-host prediction, phage taxonomy, and spacer identification. Abundance calculations are performed using BWA mapping and TPM normalization. Quality filtering removes contigs with mapping rates <55%, and redundancy is removed using MMSeqs2 (90% identity, 90% coverage). Final outputs include annotated defense/anti-defense systems, taxonomically classified contigs, and abundance profiles.

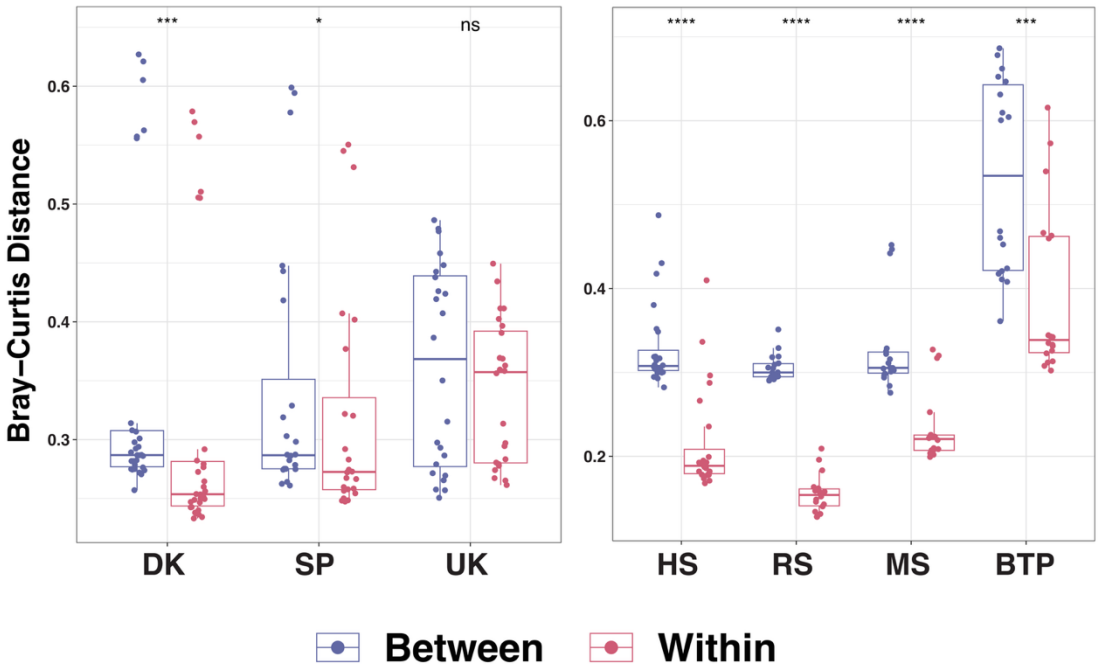

Figure S4. Differences in Bray–Curtis distances of defense system composition across groups. Boxplots show within-group (red) and between-group (blue) Bray–Curtis dissimilarities of defense system gene abundance profiles, calculated across 78 wastewater samples. (Left) Comparisons by country (DK: Denmark, SP: Spain, UK: United Kingdom); (Right) comparisons by compartment (HS: Hospital Sewer, RS: Residential Sewer, MS: Mixed Sewer, BTP: Biological Treatment Process). Significance was assessed by two-sided Wilcoxon rank-sum tests. *P* values were adjusted using the Bonferroni method. Asterisks indicate significance levels ($P \leq 0.05$, $P \leq 0.01$, $P \leq 0.001$, $P \leq 0.0001$); "ns" denotes non-significant differences.

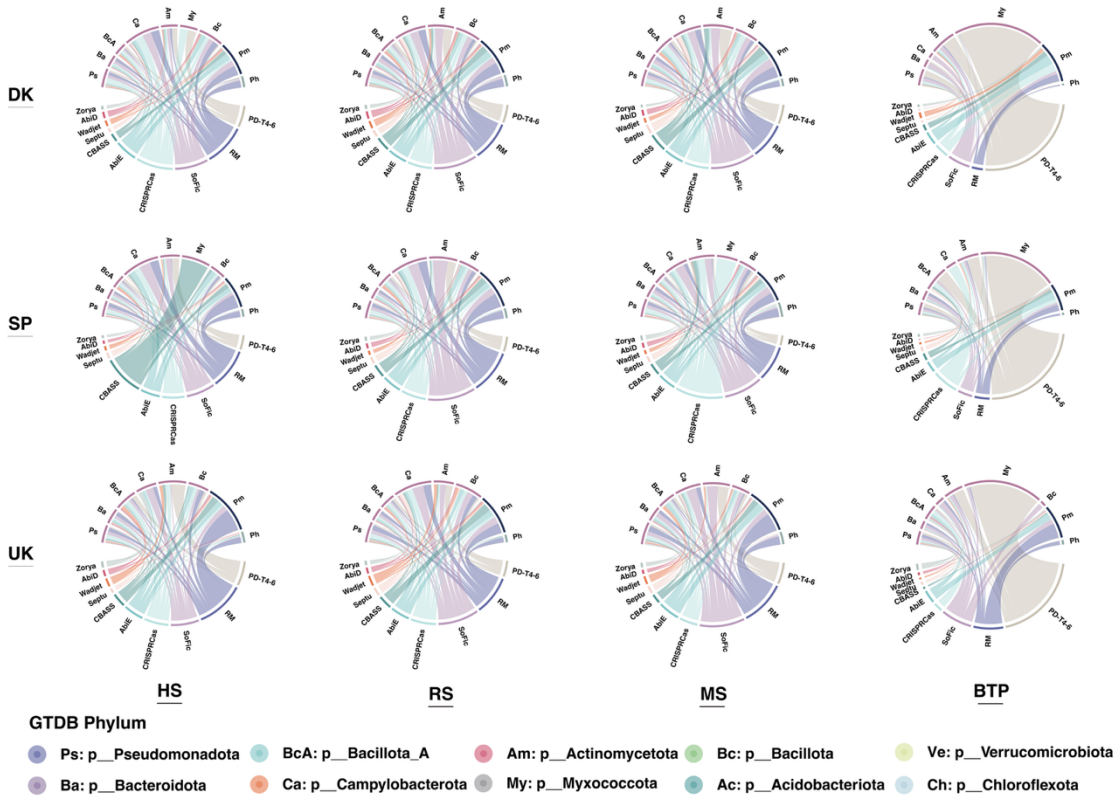

Figure S5. Phylum-defense system associations across wastewater compartments in Denmark, Spain and United Kingdom. Chord diagrams showing compartment-specific associations between bacterial phyla and defense systems based on gene abundance in hospital sewer (HS), residential sewer (RS), mixed sewer (MS), and biological treatment process (BTP) in DK (Denmark, top row), SP (Spain, middle row) and UK (United Kingdom, bottom row). Each chord connects a bacterial phylum (colored according to legend) with a defense system, with chord thickness representing the strength of association. All three countries show reduced complexity of phylum-defense connections in BTP compared to sewage compartments (HS, RS, MS), while maintaining distinct country-specific patterns.

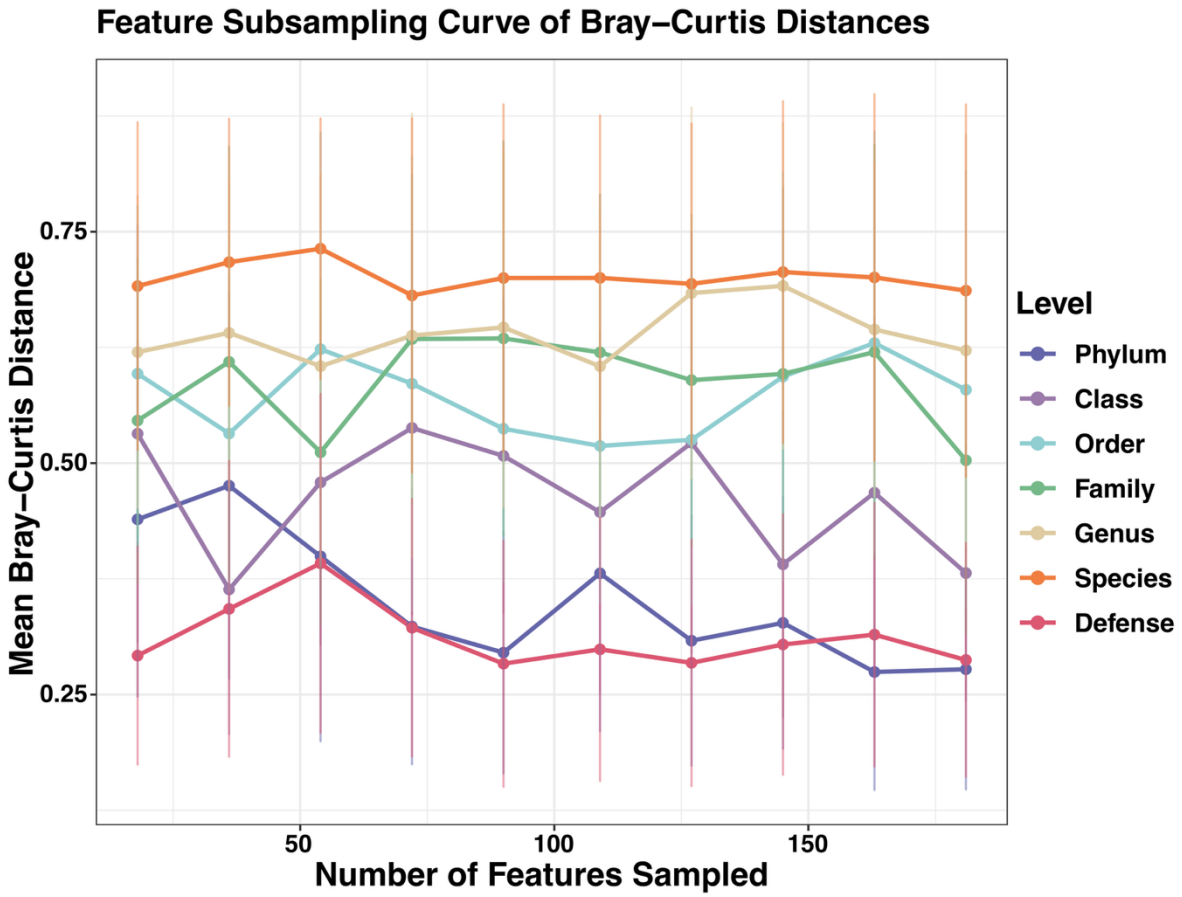

Figure S6. Feature subsampling curve of Bray-Curtis distances. Rarefaction analysis showing mean Bray-Curtis distances across different numbers of features sampled for taxonomic profiles at different levels (phylum to species) and defense systems. Each line represents a different taxonomic level or defense systems, with error bars indicating standard deviation. The analysis confirms that defense systems (pink line) maintain consistently lower Bray-Curtis dissimilarity values compared to all taxonomic levels, regardless of the number of features sampled. This rarefaction approach validates that the observed conservation of defense system composition across samples (as shown in Figure 2g) is not an artifact of differential feature abundance but represents a genuine biological pattern of defense system stability in urban wastewater environments.

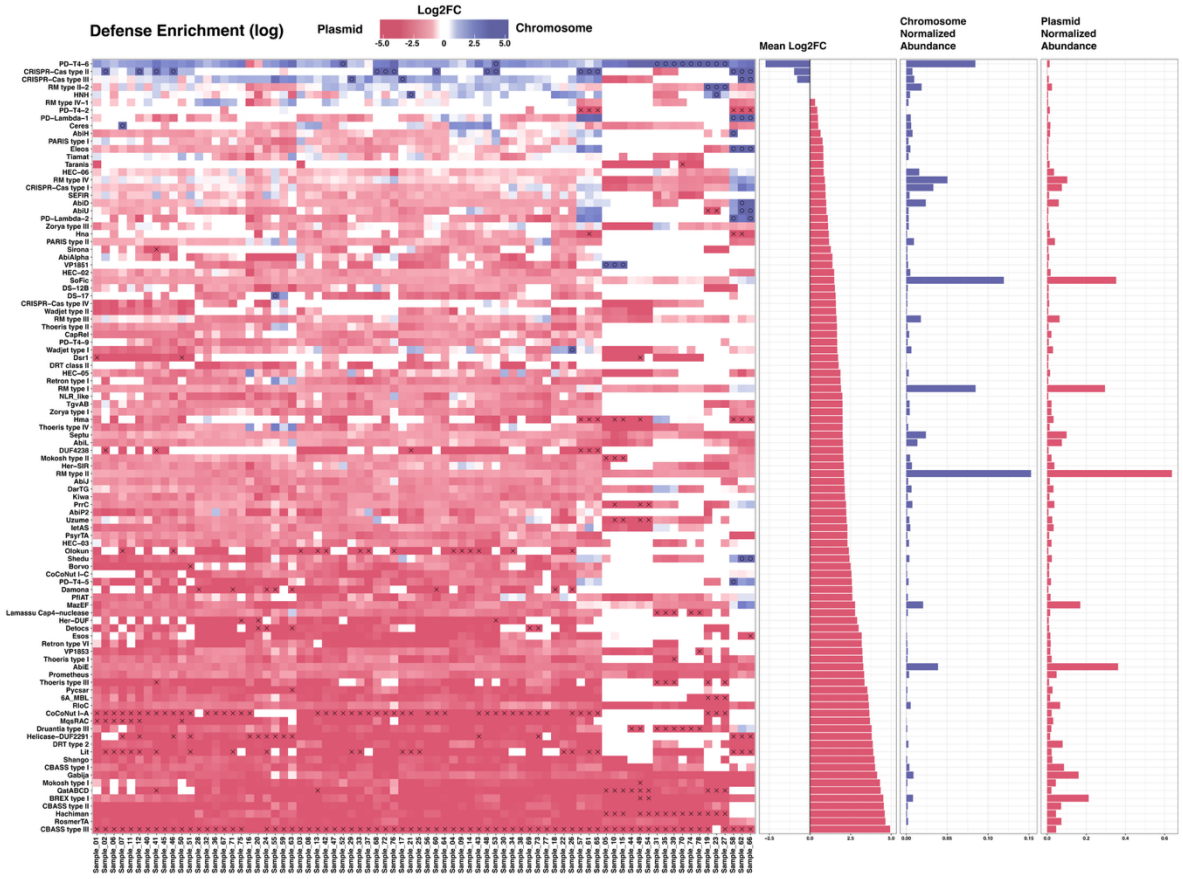

Figure S7. Extended analysis of defense system enrichment patterns between plasmids and chromosomes. Heatmap showing the abundance and genomic location (plasmid vs. chromosome) of the top 100 defense systems across all 78 samples. Samples are grouped by compartment and country; defense systems are ordered by increasing tendency for plasmid association. The left panel shows log-transformed relative abundance across samples, with intensity reflecting defense system abundance. The right panels display: (1) Defense enrichment (log scale) showing overall abundance, (2) Normalized abundance on plasmids (blue), (3) Normalized abundance on chromosomes (red), and (4) Log2 fold change (Log2FC) indicating plasmid vs. chromosome preference. Most defense systems show enrichment on plasmids (positive Log2FC values), with only a few systems like PD-T4-6 showing strong chromosomal preference (negative Log2FC values). This comprehensive analysis expands upon Figure 3a by including all defense system categories and demonstrating the consistent pattern of plasmid enrichment across the broader defense system repertoire.

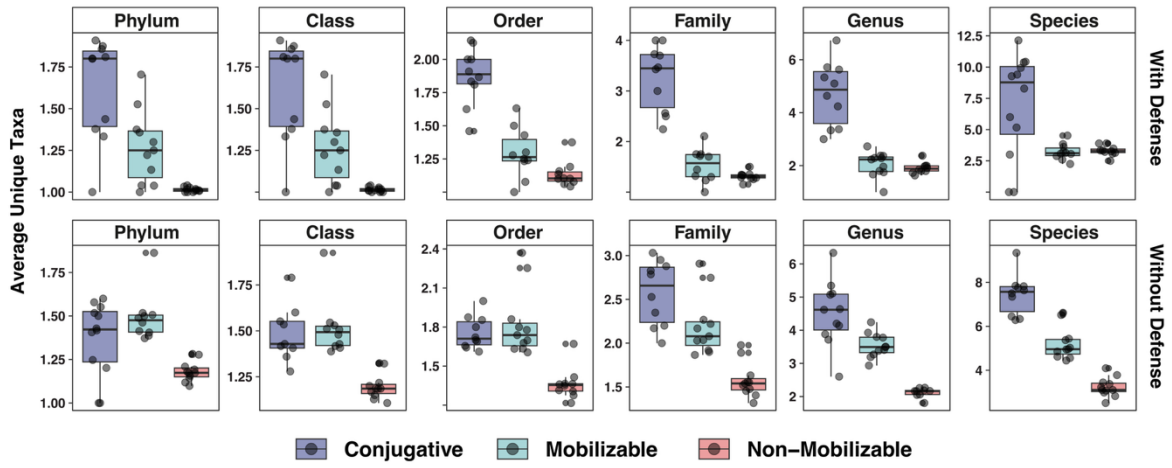

Figure S8. Extended taxonomic breadth analysis of plasmids with different mobility types. Box plots comparing taxonomic breadth across all taxonomic levels (Phylum, Class, Order, Family, Genus, Species) among plasmids with different mobility types, based on the number of unique taxa of plasmids with defense (top row) and plasmids without defense (bottom row). Each dot represents the number of unique predicted host taxa for plasmids of a given mobility type within one country-compartment group. This comprehensive analysis extends Figure 3e by including Class, Order, and Family levels, demonstrating that conjugative plasmids consistently show the highest taxonomic breadth across all taxonomic levels, followed by mobilizable plasmids, while non-mobilizable plasmids show the most restricted host range. The pattern is consistent regardless of defense system presence, with plasmids carrying defense systems showing slightly broader taxonomic breadth at finer taxonomic resolutions.

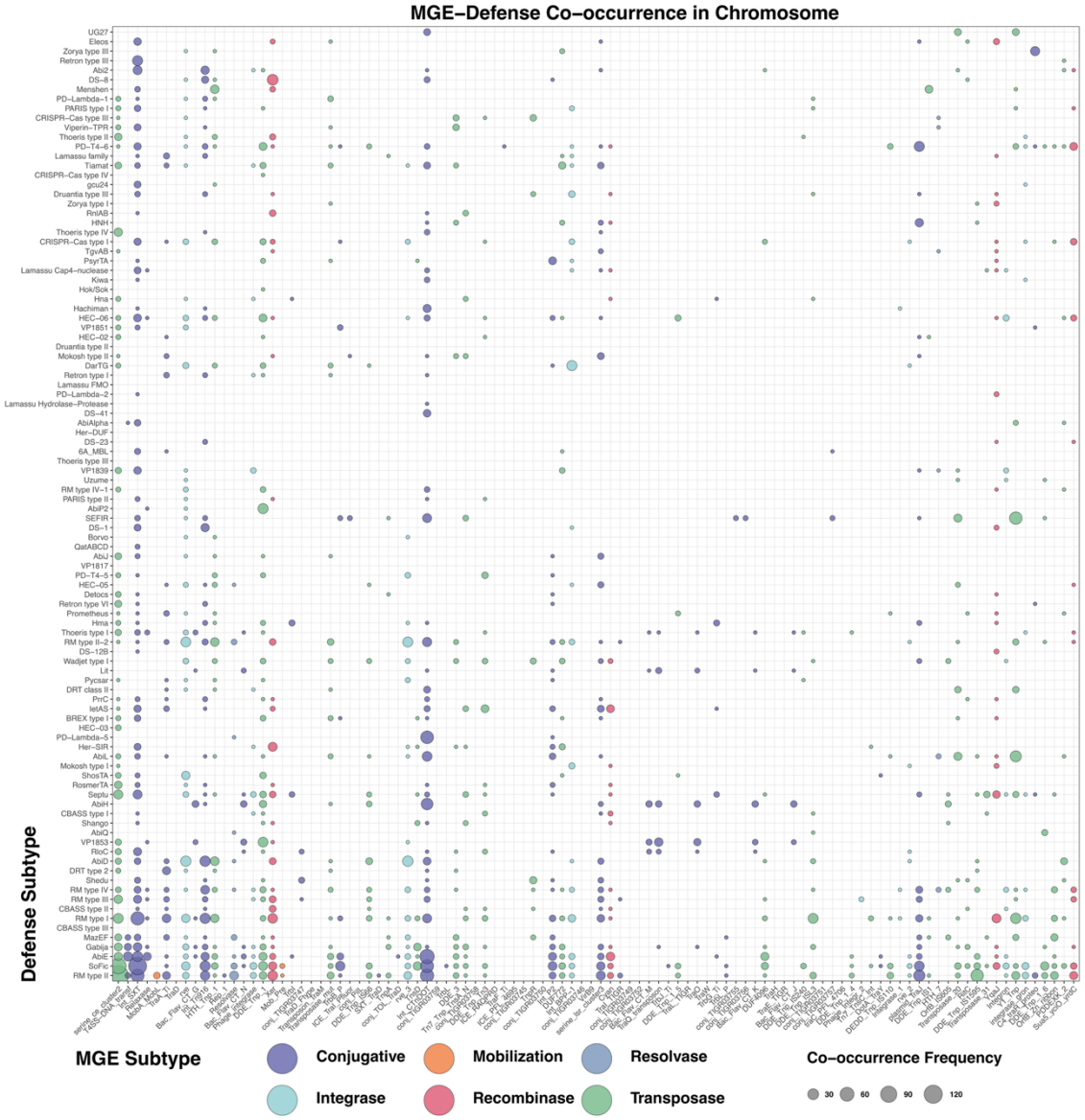

Figure S9. MGE-defense co-occurrence patterns on chromosomes. Dot plot showing co-occurrence frequencies between the top 100 defense system subtypes (y-axis) and top 100 MGE subtypes (x-axis) on chromosomal contigs. Each dot represents a co-occurrence event, with dot size indicating co-occurrence frequency (30-120 events) and color representing MGE functional categories: conjugative (blue), integrase (light blue), mobilization (orange), recombinase (pink), resolvase (light gray), and transposase (green). The analysis reveals distinct co-localization preferences between specific defense systems and MGE elements. This comprehensive view expands the network analysis shown in Figure 4c by providing quantitative co-occurrence data for the full spectrum of defense-MGE interactions on chromosomes, demonstrating the complex mobilization patterns that facilitate horizontal gene transfer of defense systems.

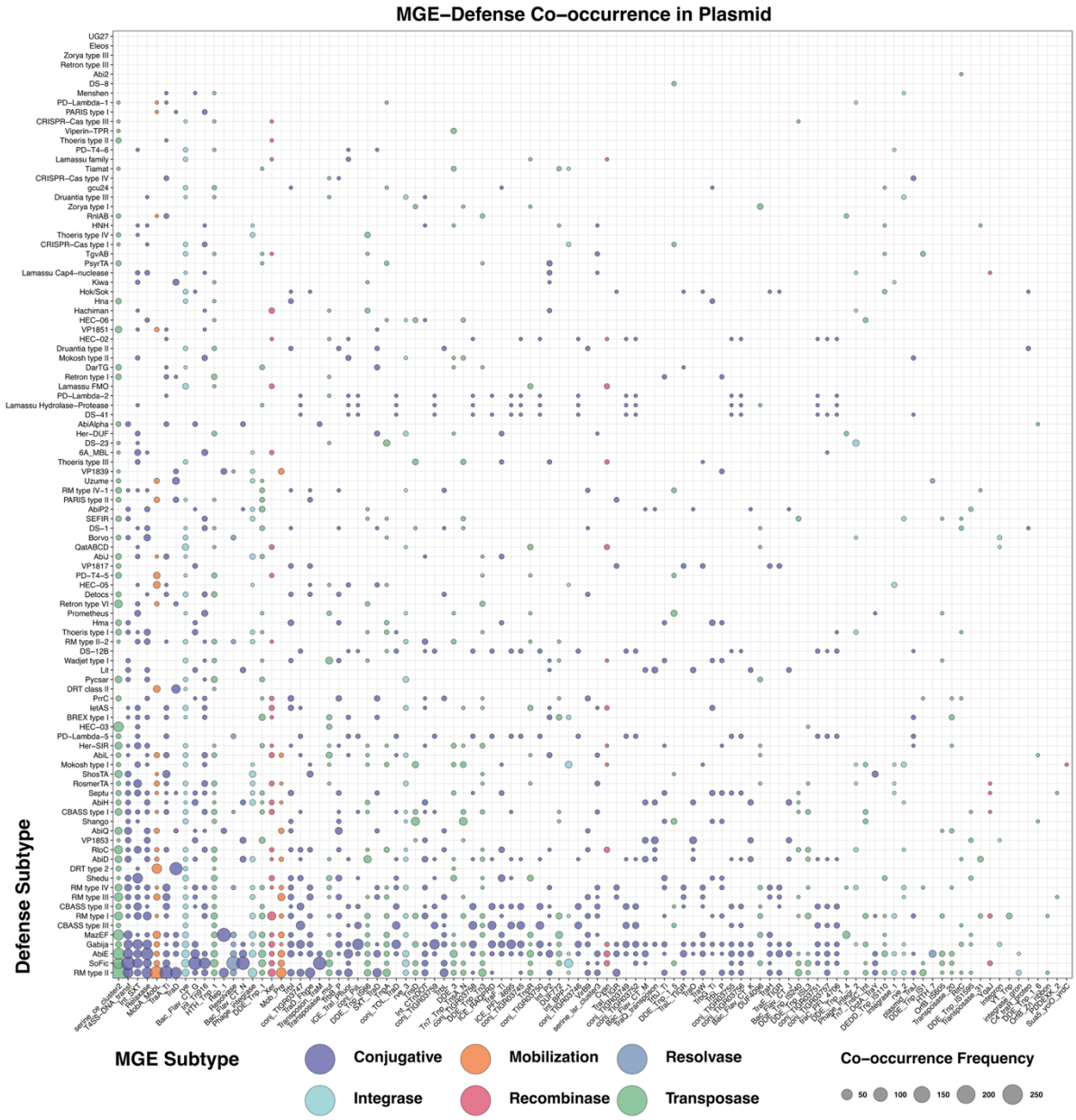

Figure S10. MGE-defense co-occurrence patterns on plasmids. Dot plot showing co-occurrence frequencies between the top 100 defense system subtypes (y-axis) and top 100 MGE subtypes (x-axis) on plasmid contigs. Each dot represents a co-occurrence event, with dot size indicating co-occurrence frequency (50-250 events) and color representing MGE functional categories: conjugative (blue), integrase (light blue), mobilization (orange), recombinase (pink), resolvase (light gray), and transposase (green). The analysis reveals more frequent and intensive co-localization events on plasmids compared to chromosomes (Figure S8), with particularly strong associations between defense systems like SoFic, AbiE, Gabija, and RM systems with various conjugative and transposase elements. This enhanced co-occurrence pattern on plasmids demonstrates their role as active vehicles for coordinated mobilization of defense systems and MGE elements, supporting the findings presented in Figure 4c regarding plasmid-mediated horizontal gene transfer of bacterial immunity.

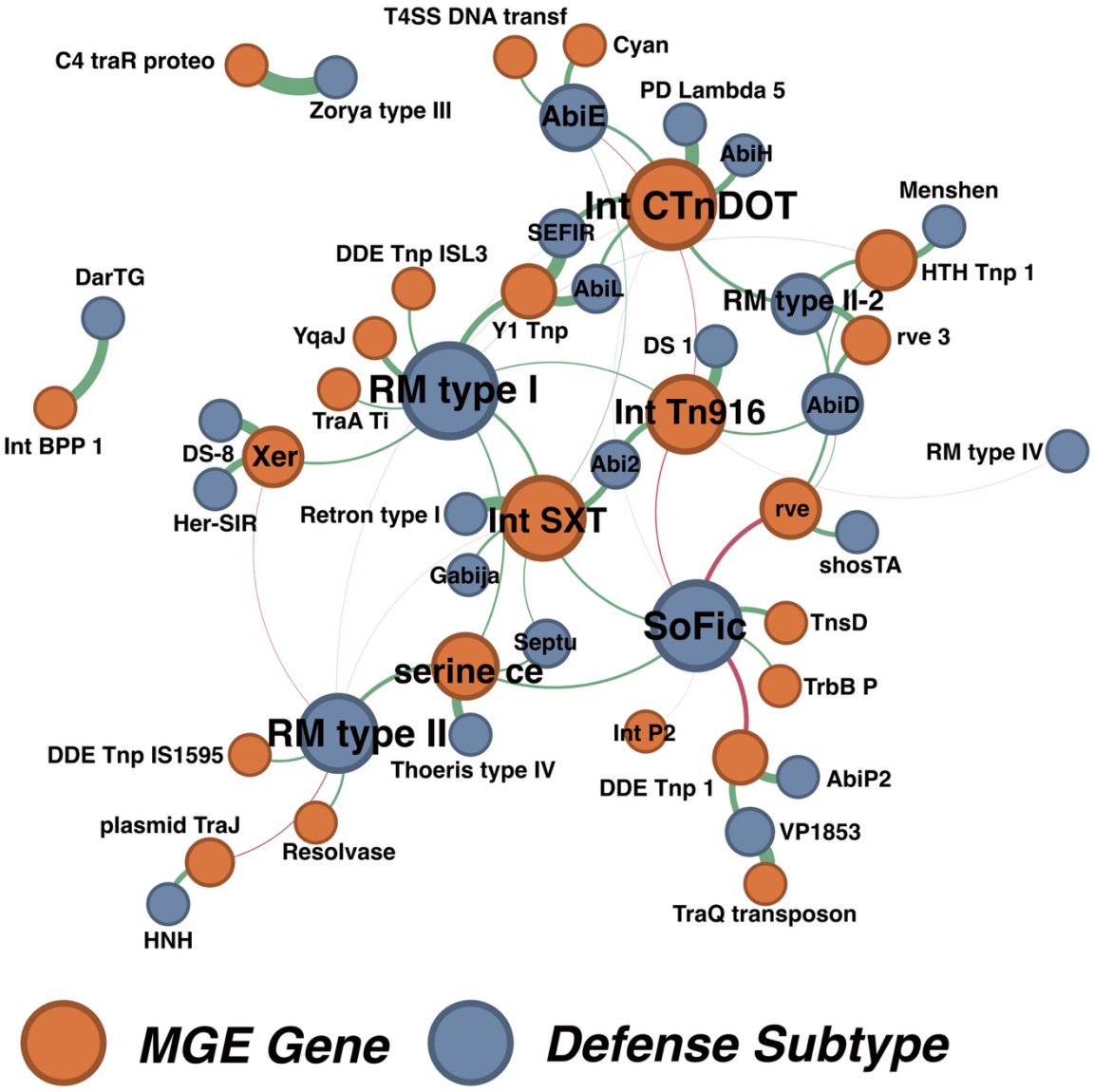

Figure S11. MGE-defense co-localization network on chromosomes. Network visualization showing co-localization patterns between defense systems (blue nodes) and MGE elements (orange nodes) identified on chromosomal contigs. Node size reflects relative abundance, while edge thickness represents the strength of co-localization associations. The network reveals distinct co-localization preferences between specific defense systems and MGE elements on chromosomes, with some defense systems like SoFic, AbiE, and RM systems showing strong associations with various MGE elements including rve, HTH_Tnp_1, and integrase elements. This chromosomal network complements Figure 4c (plasmid network) by demonstrating that while defense-MGE co-localization occurs on both genomic contexts, the patterns and intensities differ between chromosomes and plasmids, reflecting distinct mobilization mechanisms and evolutionary pressures in these two genomic.

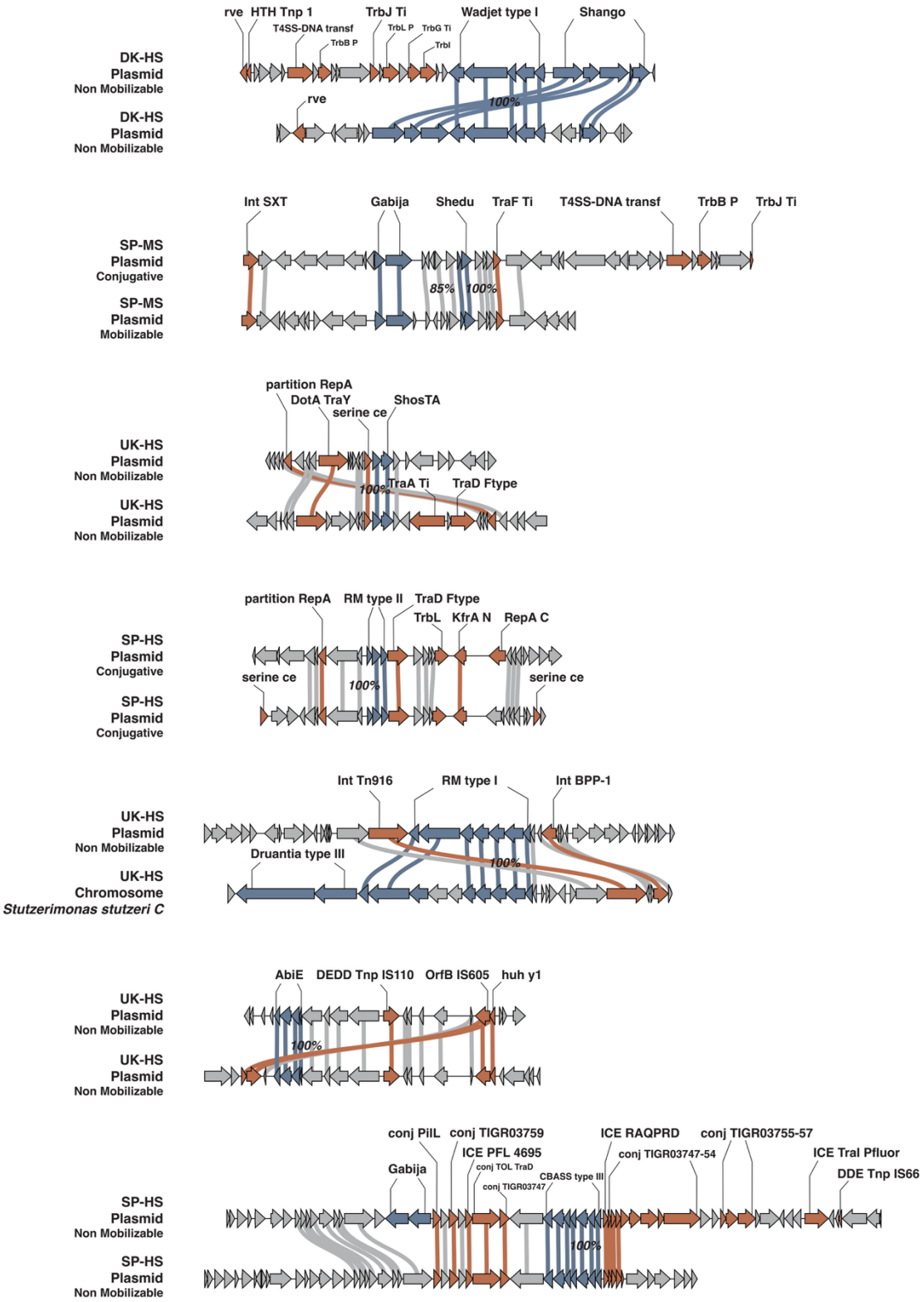

Figure S12. Additional examples of defense system mobilization across genomic contexts. Representative gene map showing co-localization and potential transfer of defense system genes (blue arrows) and MGE-associated genes (orange arrows) across different genomic elements, including plasmids and chromosomes. Each panel represents a distinct transfer scenario observed in the dataset, annotated by sample origin, genomic location (e.g., plasmid mobilizability), and taxonomic assignment. Sequence similarity between homologous regions is indicated by colored blocks

connecting genes (percent identity shown). These cases extend the examples shown in Figure 4d, further illustrating diverse patterns of horizontal gene transfer involving defense systems across environmental compartments and host taxa.

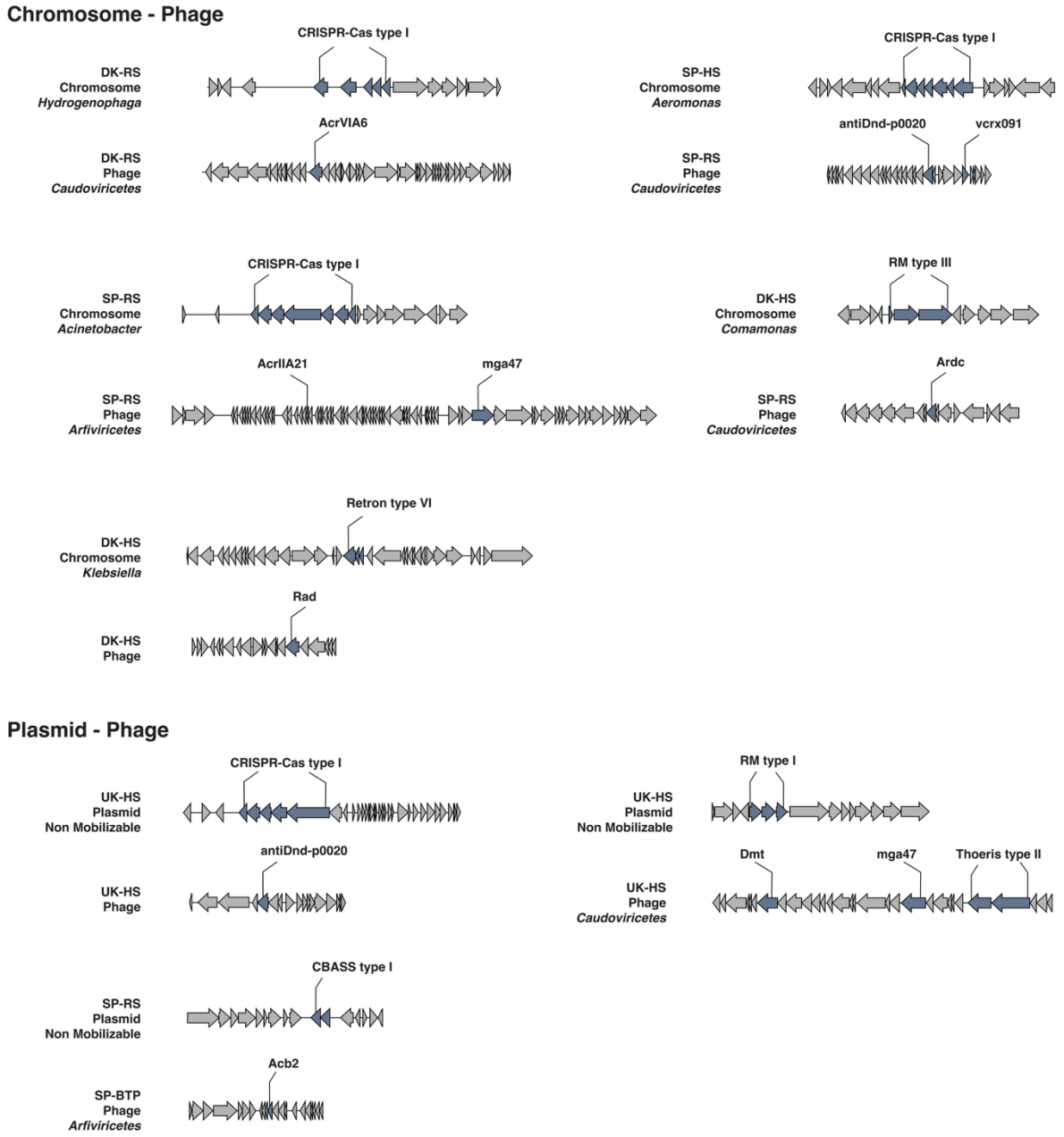

Figure S13. Additional representative examples of matched defense and anti-defense systems across genomic contexts. Gene map diagrams illustrating co-occurrence and pairing of defense systems (dark blue arrows) and their corresponding anti-defense systems (light gray arrows) across phage, plasmid, and chromosomal contigs. Each panel shows a matched pair, identified based on spacer-target network analysis, representing functional couplings such as Anti-CRISPR with CRISPR-Cas systems, Anti-RM with restriction-modification systems, and Anti-CBASS, Anti-Retron, or Anti-Thoeris systems with their respective counterparts. These examples span diverse hosts (e.g., *Hydrogenophaga*, *Klebsiella*, *Acinetobacter*) and genomic locations (e.g., phage-plasmid, chromosome-phage), and demonstrate the structural organization and potential mobilization patterns of immune–counter-immune elements. These cases complement those shown in Figure 5g.