Gain and loss of plasmid-borne antibiotic resistance genes are associated with chromosomal resistance presence in Enterobacteriaceae

Yang Liu[1, *], Yue Liu[1, *]

[1] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases of the First Affiliated Hospital and Department of Microbiology, Zhejiang University School of Medicine, Zhejiang, China

[*]Correspondence: yang.liu29@zju.edu.cn (Yang L.); yuel@zju.edu.cn (Yue L.)

ABSTRACT

Plasmids are central vehicles for the dissemination of antibiotic resistance genes (ARGs). They are among the most mobile and evolvable genetic elements, with broad host ranges and high rates of gene turnover, making them especially effective in spreading antibiotic resistance across bacterial lineages. Using the phylogeny-aware gene gain and loss model applied to 6,895 Enterobacteriaceae genomes, we quantified four evolutionary processes—gene gain, loss, expansion, and reduction—for plasmid-borne genes. We found that, overall, plasmid-borne ARGs (pARGs) exhibit similar gain rates compared with other plasmid genes, but significantly higher expansion and reduction rates. All four processes were strongly species dependent, with only a minor influence of antibiotic class. Further, bacterial clades harboring chromosomal ARGs (cARGs) showed significantly higher acquisition and lower loss of plasmid-borne resistance than did their sister clades lacking corresponding cARGs. Moreover, we found that the IncQ2 backbone was associated with *qnrS2* and exclusively identified in *Leclercia adecarboxylata*, while Col(VCM04) plasmids carrying *mprF* were predominantly (71.4%) distributed within the *Citrobacter* genus. In summary, plasmid-mediated resistance is primarily species-dependent, and cARGs effectively mark lineages with a high capacity for plasmid-borne resistance acquisition.

KEYWORDS antibiotic resistance genes; Enterobacterales; gene gain and loss, plasmid; horizontal gene transfer.

Importance

Plasmids play a central role in the spread of antibiotic resistance genes (ARGs), and

the long-term evolutionary behavior of plasmid-borne ARGs (pARGs) could provide insights into the emergence of novel multidrug resistance. We studied nearly 7,000 Enterobacteriaceae genomes and show that pARGs evolve through the same gain processes as other plasmid genes but exhibit markedly higher and species-dependent copy number changes. Crucially, the strong association between chromosomal and plasmid ARGs reflect a lineage-level pattern of resistance retention, likely shaped by historical selective pressures or specific genomic backgrounds. Identifying such evolutionarily lineages may provide a basis for predicting and monitoring the emergence of multidrug resistance.

INTRODUCTION

Antibiotic resistance is a major, ongoing challenge in both medicine and evolutionary biology (1). In bacterial genome evolution, gene gain through horizontal gene transfer (HGT) and gene loss (via selection) are the primary forces shaping genomic repertoires (2, 3). Mobile genetic elements (MGEs), particularly conjugative plasmids, serve as principal vehicles of HGT and are central to the spread of antibiotic resistance genes (ARGs) (4-6).

Comparative genomic analyses show that plasmid taxonomic units carrying ARGs exhibit significantly higher mobility, broader host ranges, and elevated rates of homologous recombination and gene turnover, indicating that resistance dissemination is facilitated by intrinsically plastic and mobile plasmid lineages (7, 8). Their plasticity is enhanced by high densities of transposable elements and recombination systems, which accelerate gene repertoire turnover (9, 10). Moreover, evolutionary innovations such as replication-control mutations can increase plasmid copy number, thereby

amplifying resistance levels and enhancing transfer efficiency (11). Thus, plasmids are highly dynamic replicons characterized by frequent horizontal transfer, rapid turnover, and extensive genetic plasticity (12, 13).

Given their evolutionary plasticity, understanding plasmid-mediated ARG turnover is particularly critical in clinically important bacterial families such as Enterobacteriaceae (12). The Enterobacteriaceae family harbors extensive plasmid diversity and constitutes one of the largest reservoirs of clinically relevant ARGs (14-16). Members of this family, including *Escherichia*, *Klebsiella*, *Enterobacter*, and *Salmonella*, are major contributors to multidrug-resistant infections in hospitals and communities worldwide (17). Although numerous clinical studies have documented outbreaks of antibiotic resistance within the Enterobacteriaceae family (16-18), most of these reports have focused on the dissemination of particular resistance plasmids or the epidemiology of individual ARGs (19, 20). However, key questions remain as to how plasmid-borne ARGs in Enterobacteriaceae are maintained and amplified across lineages, and whether their evolutionary dynamics differ from those of other plasmid-encoded functions.

Here, we systematically investigate ARG turnover dynamics in Enterobacteriaceae plasmids by applying a phylogeny-aware birth-death model to 6,895 genomes. Specifically, we (i) quantify the balance of ARG gain, loss, expansion, and reduction across diverse species, (ii) assess the association between chromosomal and plasmid-borne resistance repertoires.

RESULTS

Plasmid-borne antibiotic resistance genes show higher expansion and reduction rates but similar gain dynamics compared with other plasmid genes

To investigate the evolutionary dynamics driving the current distribution of antibiotic resistance genes (ARGs) in Enterobacteriaceae plasmids, we applied a phylogenetic birth-death model using Count (21) to quantify gene turnover rates across 6,895 genomes of which 5,335 contained at least one identified plasmid sequence (Supplementary Table S1). The approach estimates four key evolutionary processes: (i) gene family turnover, comprising gain (acquisition of new gene families) and loss (complete disappearance of an existing family) (Supplementary Figure S1A, C); and (ii) copy-number variation, comprising expansion (increase in gene copies within an existing family) and reduction (decrease in gene copies while maintaining at least one copy) (Supplementary Figure S1B-C). Since we focused exclusively on plasmid-associated gene families (defined by a $\geq 70\%$ plasmid-occurrence threshold), expansion can be biologically interpreted as intra- or inter-plasmid gene duplication (e.g., mediated by transposons), the acquisition of additional plasmid carrying the same gene family, or even chromosomal integration of a plasmid-borne gene copy, while reduction represents the subsequent loss of these redundant copies (Supplementary Figure S1B).

We compared overall rates between plasmid borne ARGs (pARGs) and all other plasmid genes (non-ARGs) across 80 species. Our analysis revealed that pARGs exhibit substantially different dynamics for gene gain and loss versus expansion and reduction (Figure 1, Supplementary Table S2). Gain rates showed no significant differences between pARGs and non-ARGs (n = 80, gain: adjusted p = 0.257; Figure

104 1A), with data points forming relatively compact distributions across species. This
105 suggests that the fundamental processes governing the acquisition of entire gene
106 families operate similarly for resistance and non-resistance genes. In contrast, pARGs
107 exhibited significantly elevated rates for loss (n = 80, adjusted p = 1.23×10^{-2}, median
108 $\log_2$(ARG/ALL) = 0.15; Figure 1B), expansion (adjusted p = 3.34×10^{-6}, median
109 $\log_2$(ARG/ALL) = 1.28; Figure 1C) and reduction (adjusted p = 4.03×10^{-5}, median
110 $\log_2$(ARG/ALL) = 0.83; Figure 1D) events. The high scatter in these data points
111 indicates substantial species-specific variation in the extent to which loss, expansion
112 and reduction dynamics differ between pARGs and other plasmid genes. Regarding
113 the ratios of gain-to-loss and expansion-to-reduction, no significant differences were
114 observed for pARGs ($p > 0.05$; Supplementary Figure S2).

115 In summary, across Enterobacteriaceae, gain rates show no pARG-specific signal,
116 indicating that acquisition of pARG families follows common plasmid evolutionary
117 pathways. In contrast, pARGs exhibit significantly elevated expansion and reduction
118 rates with pronounced species-level variation, revealing that these copy-number
119 dynamics are primarily driven by species identity. While gene loss rates showed a
120 statistically significant difference, the effect size was small (median log2 ratio = 0.15).
121

122 **Turnover dynamics of antibiotic resistance genes are primarily species**
123 **dependent**

124 To test whether the species-level patterns in Figure 1 are consistent across resistance
125 mechanisms or vary among drug classes, which are defined by their mechanisms of
126 action, we assigned orthologous gene families to CARD-defined drug classes and

127 estimated gain, loss, expansion, and reduction rates for each class across species.
128 Using the CARD database (22), we identified and analyzed the turnover dynamics of
129 specific drug classes, which are groups of antibiotics with similar chemical structures
130 and mechanisms of action.

131 Our analysis revealed that the evolutionary rates for all four event types were
132 strongly species-dependent (Kruskal–Wallis test; gain: $H = 536.9$, $p = 2.9\times10^{-69}$, $\eta^2 =$
133 0.76; loss: $H = 506.3$, $p = 1.4\times10^{-63}$, $\eta^2 = 0.71$; expansion: $H = 312.5$, $p = 1.6\times10^{-29}$, η^2
134 $= 0.39$; reduction: $H = 436.9$, $p = 2.4\times10^{-51}$, $\eta^2 = 0.60$), with substantial variation across
135 species for gain, loss, expansion, and reduction (Supplementary Figure S2A–B; Figure
136 2A–B). In contrast, drug class effects were weak: no significant signals were detected
137 for gain ($p = 0.74$) and loss ($p = 0.58$), while expansion ($p = 5.2\times10^{-9}$, $\eta^2=0.09$) and
138 reduction ($p = 1.1\times10^{-3}$, $\eta^2=0.04$) showed significant but weak drug class effects, which
139 were an order of magnitude smaller than those of species. To account for potential
140 biases from multi-counting genes across drug classes, we used an Ordinary Least
141 Squares (OLS) model for variance partitioning. This analysis confirmed species identity
142 as the primary driver of pARG turnover, explaining 66.8%, 63.1%, 42.1%, and 51.5%
143 of the total variance for gain, loss, expansion, and reduction events, respectively. In
144 contrast, drug class identity accounted for minimal variance across all events: 2.1% for
145 gain, 1.7% for loss, 6.9% for expansion, and 4.1% for reduction. Full results are
146 provided in Supplementary Table S3.

147 Furthermore, species that exhibited high expansion rates typically also showed
148 elevated reduction rates for the same drug classes (n=80, Spearman's rank correlation
149 with 5,000 permutations: $\rho=0.792$, permutation $p\approx0.0002$). By contrast, species with
150 high gain rates did not necessarily display the corresponding high loss rates (n=80,

151 ρ=-0.196, p≈0.079), indicating that plasmid-borne gene expansion and reduction
152 increases and decreases within existing gene families are more tightly coupled than
153 complete gene family gain and loss.

154 In summary, antibiotic resistance gene evolution dynamics are primarily driven by
155 species identity, with only minor drug class effects. Moreover, expansion and reduction
156 rates are coupled, whereas gain and loss are largely independent.

157

158 **Chromosomal antibiotic resistance genes (cARGs) are associated with elevated**
159 **pARG acquisition rates**

160 Our previous analyses revealed that pARGs exhibit similar gain rates compared to
161 other plasmid genes. We further investigated whether additional factors might be
162 associated with the evolutionary dynamics of pARGs, specifically examining whether
163 the presence of antibiotic resistance genes in bacterial chromosomes is associated
164 with altered evolutionary rates of pARGs. We employed a comparative phylogenetic
165 approach using sister clade analysis (Figure 3A), identifying 464 pairs of phylogenetic
166 clades where one clade (lineage) contained ≥80% of genomes harboring a specific
167 ARG family in their chromosome (ARG-containing clade) while <20% of genomes in
168 the sister clade (lineage) contained that homologous ARG family. For each clade pair,
169 we calculated gain and loss rates for plasmid genes by summing evolutionary events
170 across all branches within each clade and normalizing by total branch length. In total,
171 80 species and 84 RGI-defined ARG families were included in the analysis
172 (Supplementary Tables S4 and S5). Linear mixed-effects models (LMM) were used as
173 a statistical test for the analysis, including fixed effects for clade type, gene class, and

174 their interaction, with random effects for species, ARG family, and sister-clade
175 comparisons to capture lineage-specific variation (see Methods).

176 Our analysis revealed that bacterial clades with cARGs exhibit significantly
177 different plasmid gene gain rates compared with their sister clades (Figure 3B-C).
178 ARG-containing clades showed approximately 2.47-fold higher gain rates for pARGs
179 (estimate = 0.91 on the log scale, $p = 1.9\times10^{-7}$; Figure 3B, Supplementary Table S6)
180 and significantly lower loss rates (fold-change = 0.138, p = 0.034). In contrast, other
181 plasmid genes showed significantly lower gain rates in ARG-containing clades
182 compared with sister clades (about 3.7-fold lower; estimate = -1.3, $p = 2.97\times10^{-7}$), while
183 their loss rates also showed a decrease between clade types (estimate = -0.94, p =
184 1.44×10^{-7}). The interaction term indicated that the correlation between cARG presence
185 and evolutionary rates differs significantly between gene types for both gain (estimate
186 = -2.2, $p = 2.2 \times 10^{-9}$) and loss events (estimate = -1.0, $p = 5.6 \times 10^{-5}$) (Figure 3C,
187 Supplementary Table S6).

188 To test the robustness of our results across different clade definitions, we
189 performed an analysis across a gradient of sister-clade thresholds (70/30, 80/20, 85/15,
190 and 90/10 for cARG/sister clades). Although the magnitude of the effect was sensitive
191 to the specific cutoff, the positive association remained robust across all thresholds
192 (Supplementary Figure S4, Supplementary Table S5). As the threshold for defining
193 ARG-containing clades became more stringent, representing lineages with higher
194 levels of chromosomal resistance prevalence, the magnitude of the pARG gain bias
195 became more pronounced. Specifically, the fold-change in pARG gain rates increased
196 progressively from 1.65 at the 70/30 threshold (p = 0.015) to 2.47 at the 80/20 threshold
197 ($p = 6.7 \times 10^{-4}$), 3.21 at the 85/15 threshold ($p = 1.5 \times 10^{-4}$), and 4.91 at the 90/10

threshold ($p = 1.19 \times 10^{-6}$) (Supplementary Figure S4, Supplementary Table S6). Parallel to the gain dynamics, the suppression of pARG loss rates in cARG clades remained remarkably stable and highly significant across all tested categories. The loss rate fold-changes ranged from 0.183 to 0.125 with all *p* values below 0.001 (Supplementary Figure S4, Supplementary Table S6), indicating that these lineages consistently maintain acquired resistance genes more effectively than their sister clades.

In addition to the overall association, we performed per-family LMM analyses for those with at least three available clades in Enterobacteriaceae (Supplementary Table S7). The overall gene gain rate remained higher in cARG clades (Supplementary Figure S5; $p < 0.05$). Specific ARG families, such as sulfonamide-resistant *sul* and trimethoprim-resistant dihydrofolate reductase (*dfr*), display higher gene gain but lower gene loss rates (Supplementary Figure S6).

Notably, individual tests for each family revealed that a larger proportion of cARG families exhibited significantly reduced gene loss rates than elevated gain rates (see Supplementary Figure S6). Along with the reduced loss of other plasmid genes within these lineages (Supplementary Figure S3), this stability may represent an evolutionary tendency to maintain plasmids in cARG-containing clades. Consistent with this, a pairwise comparison of 436 clade pairs revealed that cARG clades maintain a significantly higher plasmid burden than their sister counterparts, with a median of 2.64 unique plasmid clusters compared to 1.66 ($p = 2.42 \times 10^{-21}$) (Supplementary Figure S7A).

To further study whether the accelerated pARG acquisition in cARG clades be attributed to lower plasmid backbone incompatibility, we then calculated the frequency

222 of these pARG-associated replicon types within their respective sister clades. We
223 observed a 0.00% median sharing ratio of backbone replicon genes between sister
224 clades (Supplementary Figure S7B), suggesting that most of these pARG-carrying
225 plasmids were absent in the most recent common ancestor and were acquired
226 specifically in the cARG clades. Together, these evidences help strengthen the
227 interpretation that these cARG-containing lineages are genuinely more prone to
228 plasmid-mediated resistance acquisition.

229 For most species, cARG-clade associated genomes exhibit a wide and diverse
230 spectrum of antibiotic resistance (Figure 4A). Many species (e.g., *Enterobacter*
231 *cloacae_M*, *Enterobacter chengduensis*, *Erwinia persicina*, *Klebsiella terrigena*, and
232 *Kosakonia cowanii*) exhibit widespread resistance to clinically and agriculturally used
233 antibiotics (e.g., penicillins, cephalosporins, and tetracyclines). In contrast, only a few
234 species accumulate antibiotic-resistant strains for a specific antibiotic; the most notable
235 examples are *Serratia nevei* and *Klebsiella michiganensis*, which exhibit high levels of
236 aminoglycoside antibiotic resistance, and *Klebsiella quasipneumoniae* for glycopeptide
237 antibiotics. The observed resistance profiles across these lineages can be compared
238 to the representative consumption levels of corresponding antibiotic classes in clinical
239 and agricultural settings (Figure 4B and 4C).

240

241 **Co-occurrence of plasmid-borne antibiotic resistance genes with plasmid**
242 **replicon types**

243 To identify the associations between the presence of ARGs and specific plasmid types,
244 we reconstructed plasmid units from draft assemblies and performed replicon typing

245 for both draft and complete assemblies. For plasmid groups defined by a particular
246 replicon type, we calculated all-against-all Mash distances and constructed plasmid
247 similarity networks. Notably, we employed a Mash distance cutoff of 0.025, which is
248 equivalent to the secondary clustering criterion used by MOB-recon tool. Consequently,
249 each panel in Figure 5 visualizes the fine-grained genetic similarity within specific
250 plasmid types.

251 By mapping the ARGs to plasmids, we observed that specific plasmid types, such
252 as ColRNAI_rep_cluster_1987, Col(VCM04), and IncQ2, exhibited strong associations
253 with sulfonamide, peptide, and glycopeptide antibiotics, respectively (Figure 5A-C and
254 Table 1). Specifically, the ARGs on ColRNAI_rep_cluster_1987 plasmids were
255 disinfectant resistance genes *qacG*, while those on Col(VCM04) were primarily *mprF*.
256 The IncQ2 plasmids were mainly associated with the *qnrS2* (Supplementary Table S8).
257 In contrast, such distinct gene-replicon associations were rarely observed in plasmids
258 harboring multidrug resistance, which often exhibited more complex and
259 heterogeneous ARG compositions (Figure 5D-F, Table 1).

260 In terms of host distribution, the *qacG* genes associated with
261 ColRNAI_rep_cluster_1987 were identified across 16 species, with 34 of the 71 total
262 plasmids (47.9%) found in *Serratia nevei* (Supplementary Table S8). The *mprF* genes
263 associated with Col(VCM04) spanned 14 species, with 40 of the 56 total plasmids
264 (71.4%) originating from the *Citrobacter* genus. The *qnrS2* genes associated with
265 IncQ2 were exclusively identified in *Leclercia adecarboxylata*.

266

267 **DISCUSSION**

268 Across 6,895 Enterobacteriaceae genomes, we show that plasmid-borne antibiotic
269 resistance genes (pARGs) and other plasmid genes exhibit similar species-level gain
270 rates (Figure 1A). However, the expansion and reduction rates of ARGs, which depend
271 on species, are much more diverse than those of other plasmids. Elevated expansion
272 and reduction may reflect the activity of gene-level mobile elements such as IS26
273 translocatable units (23, 24) and integron cassette rearrangements (25, 26), as well as
274 selective amplification events that duplicate individual genes without affecting entire
275 plasmids (27). Besides, recent study revealed that 77% of studied ARG-carrying
276 plasmids contain insertion sequences, facilitating inter-plasmid resistance gene
277 transfer (7). And the replication-control mutations (e.g., *copA*) may also elevate
278 plasmid copy number and increase ARG transfer rates (11)

279 Beyond species-specific variation in expansion and reduction, which likely reflects
280 differing balances between antibiotic-driven selection and fitness costs (28, 29), we
281 observed that bacterial lineages harboring cARGs acquired pARGs at 2.47-fold higher
282 and 3.7-fold lower rates than their sister clades (Figure 6B, Supplementary Table S6).
283 Rather than implying direct causation, the lack of temporal resolution in static genomic
284 data precludes a definitive conclusion regarding the directionality of this evolutionary
285 process. We consider two distinct evolutionary scenarios to explain this association. In
286 one scenario, chromosomal integration acts as a prerequisite for further resistance
287 accumulation. By fixing critical resistance genes into the core genome, lineages ensure
288 vertical inheritance and minimize the metabolic burden associated with maintaining
289 plasmids. This stabilized genetic background may facilitate the subsequent uptake of
290 a broader plasmid repertoire. Alternatively, chromosomal integration may be the
291 evolutionary consequence of intense plasmid flux under sustained antibiotic selection.

292 High rates of pARG acquisition increase the intracellular density of resistance genes,
293 elevating the statistical probability of stochastic transposition to the chromosome via
294 mobile elements such as IS26 or integrons (24, 30-33). In this scenario, cARGs
295 represent the consequence of a highly dynamic resistance mobilome. In any case,
296 from a long-term evolutionary perspective, cARGs can serve as genomic markers for
297 bacterial lineages prone to accumulate plasmid-mediated ARGs.

298 cARG-containing clades have a higher plasmid burden, with a median of 2.64
299 plasmids, compared to 1.66 in sister clades (Supplementary Table S7). Since the
300 median sharing ratio of pARG-associated replicons with sister clades is 0%, these
301 plasmids were likely acquired independently after the lineages diverged, rather than
302 inherited from a common ancestor. This independent acquisition, combined with
303 significantly lower rates of plasmid gene loss (Figure 3B), suggests that these clades
304 are particularly effective at obtaining and retaining the environmental mobilome. The
305 presence of more plasmids increases the physical proximity of different pARGs within
306 the same cell. This co-localization creates the conditions necessary for recombination
307 between different plasmid backbones or the assembly of new resistance islands (34).
308 Consequently, these clades may be important sites for the exchange of ARGs. The
309 newly formed elements are subsequently stabilized by reduced loss rates, ensuring
310 persistent multidrug resistance and contributing to the dissemination of these traits to
311 the broader bacterial population.

312 Finally, we manually selected few Count-inferred events (posterior
313 probability >0.90) and verified whether the events matched previous reported
314 outbreaks (Supplementary Table S9). We identified a clade of 24 genomes from France
315 that acquired the *blaCTX-M* gene family at node N13 of the *Enterobacter*

316 *quasihormaechei* species tree. These genomes were assigned to ST873 (using the
317 ecloacae scheme), matching a previously documented *blaCTX-M-15* positive ST873
318 outbreak in French hospitals (35). Similarly, a gain event of the *mcr*
319 phosphoethanolamine transferase family was inferred at node N22 of the *Escherichia*
320 *fergusonii* species tree, involving 10 Chinese isolates. These genomes were assigned
321 to the novel ST16672 type (using the ecoli_achtman_4 scheme), for which no known
322 clinical outbreaks have been reported to date. However, by looking at their isolation
323 locations, this gain event aligns with the landmark emergence of plasmid-mediated
324 mcr-1 in Chinese animal and human populations (36). Furthermore, we identified
325 multiple *sul* gain events across diverse species that align with regional antimicrobial
326 resistance (AMR) trends. For instance, a gain at *Salmonella enterica* node N9 is
327 consistent with the high *sul* prevalence (34.37%) reported in southern Brazil (37).
328 Similarly, a gain at *Salmonella arizonae* node N34 matches documented MDR
329 outbreaks in U.S. turkey operations (38), while a gain at *Edwardsiella anguillarum* node
330 N5 reflects the high detection frequency of *sul* genes in fish pathogens along the Fujian
331 coast of China (39).

332 **Limitations**

333 This study focused exclusively on plasmid-associated genes, excluding other
334 mobile elements such as Integrative and conjugative elements (ICEs) and integrative
335 mobilizable elements (IMEs), which also mediate chromosomal integration and
336 excision of ARGs (40-42). Bacteriophages further can contribute to ARG dissemination
337 through transduction and phage–plasmid hybrids (43). Consequently, our observations
338 represent only a subset of the overall resistance mobilome. Moreover, we lacked

339 ecological information about antibiotic exposure histories of the analyzed organisms.
340 Expanding future analyses to include additional mobile genetic elements and
341 ecological datasets would provide a more comprehensive understanding of ARG
342 dissemination in natural environments.

343 **Conclusion**

344 Our findings highlight a duality in plasmid-mediated resistance evolution in
345 Enterobacteriaceae. At the level of whole gene families, pARGs follow the same
346 turnover dynamics as do other plasmid genes, indicating that resistance dissemination
347 largely reflects plasmid-level processes rather than ARG-specific rules. At the level of
348 copy-number changes, however, pARGs show elevated expansion and reduction with
349 strong lineage dependence and minor drug-class effects. Such species-driven
350 plasticity suggests recurrent expansion-reduction cycles as a flexible resistance
351 strategy. Moreover, chromosomal resistance content serves as a genomic marker of a
352 lineage's potential for plasmid-mediated resistance accumulation on evolutionary
353 timescales.

354

355 **METHODS**

356 **Genome Collection and Gene Annotation**

357 We parsed the Genome Taxonomy Database (GTDB; https://gtdb.ecogenomic.org)
358 release 220 (44) to identify genomes belonging to Enterobacteriaceae. For species
359 with more than 300 genomes, we randomly selected 300 representatives with random
360 seed 19940421. We applied MIMAG criteria (45) with completeness >99%,
361 contamination <1%, mean contig length >5,000 bp, and contig count <500 to filter high-

362 quality genomes. To reduce computational cost, species with more than 100 genomes
363 were subsampled to retain at most 100 genomes (random seed 19940421). Only
364 species with more than 10 genomes were considered for further analysis. The
365 genomes that passed these filters were downloaded from the NCBI FTP site
366 (https://ftp.ncbi.nlm.nih.gov). For sequence typing, the MLST software
367 (https://github.com/tseemann/mlst) was used to against PubMLST database (46).

368 Bacterial genes were predicted using pyrodigal-gv v0.3.2 integrated in geNomad.
369 Plasmids and their encoded genes were identified using geNomad v1.11.0 with
370 database v1.9 and default parameters (47), as well as the Platon v1.7 accuracy model
371 (48). we defined a gene as plasmid-derived only if it is independently identified by both
372 geNomad and Platon. Antibiotic resistance genes were identified with Resistance
373 Gene Identifier (RGI) v6.0.5 using protein sequences with "include_nudge" option and
374 default parameters, CARD version 4.0.1 (22). Pan-genomes of each species were
375 constructed using the orthology-based clustering method OrthoFinder v2.5.5 (49)
376 following established guidelines (50).

377 After applying these criteria, 6,895 complete or nearly complete genomes
378 belonging to 159 Enterobacteriaceae species (sensu GTDB) were included in the
379 analysis (Supplementary Table S1), of which 80 contained antibiotic resistant genes
380 on plasmids (Supplementary Data).

381

382 **Applying Phylogenetic Birth-Death Model for Gene Gain and Loss Inference**

383 From the orthologous gene families and species trees generated by OrthoFinder, we
384 extracted plasmid orthologous gene families. An orthologous gene family was

385 considered plasmid-associated if the proportion of plasmid genes within that
386 orthologous gene family exceeded 70% of all genes in the family. Species trees were
387 pruned to retain only genomes contributing to the filtered orthologous gene families.

388 Gene gain and loss dynamics were quantified using the phylogenetic birth-death
389 model implemented in Count v10.04 (21). This probabilistic framework models the
390 evolution of gene family sizes along phylogenetic trees as a continuous-time Markov
391 process. The model accounts for three fundamental processes: gene gain (κ), gene
392 loss (μ), and gene duplication (λ). A gene family of size n changes according to the
393 following rates: the family decreases at rate $n\mu$ due to individual gene losses, increases
394 through duplication at rate $n\lambda$, and gains new genes at rate κ through horizontal transfer
395 or other acquisition mechanisms.

396 Note that the biological processes underlying gene gain, loss, expansion, and
397 reduction can be interpreted as follows: 1) Gain (0 to 1): A species acquires a new
398 plasmid carrying a previously absent resistance gene family, or a new gene family
399 integrates into an existing replicon via horizontal gene transfer. 2) Loss (n to 0 for $n >$
400 1): The total loss of a plasmid harboring all copies of a gene family, or the complete
401 deletion of a gene family from all replicons within the cell. 3) Expansion (n to m for $m >$
402 1): An increase in intra-replicon copy number via tandem duplication, which is often
403 mediated by insertion sequences; the acquisition of additional homologous copies from
404 external sources into existing plasmids; an increase in plasmid copy number via
405 replication-control mutations; or the translocation and subsequent duplication of pARG
406 onto the chromosome or other co-resident plasmids. 4) Reduction (n to m for $n > m \geq$
407 1): Loss of some duplicated copies of a gene family while retaining at least one copy,
408 or loss of one plasmid in a cell harboring multiple replicons with the same gene family.

These scenarios illustrate the most probable biological drivers captured by the model. However, they are not an exhaustive representation of all complex scenarios that may occur in Enterobacteriaceae genomes.

For each species dataset, we performed iterative maximum likelihood optimization to estimate birth-death model parameters using whole bacterial genomes (51). We implemented a systematic search strategy starting with initial parameter values of $\kappa=\mu=\lambda=1$ and incrementally exploring the parameter space up to $\kappa, \mu, \lambda \leq 4$. This resulted in a total of 10 optimization rounds. The iterative approach followed the pattern: if all parameters are equal, increase loss rate (μ); if duplication and gain rates are equal, increase duplication rate (λ); otherwise increase gain rate (κ). Each round uses the previous round's results as starting values. The optimization utilized maximum likelihood estimation with maximum paralogs of 200, optimization tolerance of 0.1, and minimum lineages of 2.

After parameter optimization, we calculated posterior probabilities for plasmid-borne ancestral gene family states and evolutionary events using the Posteriors module of Count. This reconstructs the evolutionary history of each gene family, providing probabilistic estimates for gain, loss, expansion, and reduction events along each branch. Count analysis was performed separately for each species dataset.

Gene Gain and Loss Data Integration and Rate Calculation

An orthologous gene family was defined as an ARG family if at least 50% of its member genes were identified as ARGs by RGI. Branch lengths were extracted from the original species trees and combined with corresponding Count posterior probabilities. For rate

calculations, we summed gain, loss, expansion, and reduction events across all branches within each orthologous gene family and divided by the corresponding total branch length. Only branches with branch lengths > 1e-12 were included to avoid computational artifacts from extremely short branches. Evolutionary rates were calculated per species per orthologous gene family as:

$$rate_i = \frac{\sum\left(events_{\{i,j\}}\right)}{\sum\left(branch_length_j\right)}$$

where $events_{\{i,j\}}$ is the number of events for orthologous gene family i on branch j, and $branch_length_{\{j\}}$ is the length of branch j in substitutions per site.

Comparative Analysis of Evolutionary Rates

For comparative analysis between ARG and non-ARG plasmid gene families, data were aggregated by species, event type, and ARG status. Median rates were calculated within each group, and the data were restructured into a paired format where each species served as a matched pair with corresponding median rates for ARG and non-ARG families. Paired Wilcoxon signed-rank tests compared ARG versus non-ARG families within species to control for species-specific evolutionary factors. Additionally, gain/loss and expansion/reduction ratios were calculated per species to assess the relative balance of these complementary processes. Multiple testing correction was performed using the Benjamini-Hochberg procedure. Effect sizes were quantified as median log2-fold changes, and statistical significance was assessed at $p < 0.05$ and $q < 0.05$ (FDR-adjusted p).

To examine rate variation across antibiotic resistance classes, drug class annotations were extracted from RGI output. For ARG families associated with multiple

455 drug classes, evolutionary rates were assigned to each corresponding class.
456 Evolutionary rates were then visualized across different species and drug class within
457 Enterobacteriaceae. The concordance between event types across species was
458 assessed by compared median per-species rates using Spearman's rank correlation
459 with permutation tests (5,000 iterations, random seed = 19940421) to obtain empirical
460 p-values.

461 To evaluate the relative contributions of species background and drug class to
462 pARG turnover, variance partitioning was performed using Ordinary Least Squares
463 (OLS) models. Evolutionary rates (y) were log-transformed as $\ln(y + 10^{-12})$ to satisfy
464 normality assumptions. We compared the explanatory power of two categorical models,
465 Species-driven model, $ln(y + \epsilon) \sim \beta_0 + \beta_i \cdot Species_i + \epsilon$, and a drug-class-driven
466 model, $ln(y + \epsilon) \sim \beta_0 + \beta_i \cdot Drug_class_j + \epsilon$, where Species and drug class represent
467 categorical predictors for bacterial lineages and antibiotic classes, respectively. The
468 coefficient of determination (R^2) was used to quantify the proportion of total variance
469 explained by each factor.

470

471 **Comparison of Plasmid Gene Dynamics Between ARG-containing and Sister**
472 **Clades**

473 To assess whether the presence of antibiotic resistance genes in bacterial
474 chromosomes influences the evolutionary dynamics of plasmid-borne genes, we
475 identified pairs of phylogenetic clades that differed in the chromosomal presence of an
476 ARG family. For each combination of species tree and ARG family, we scanned for
477 internal nodes where ≥80% of descendant genomes harbored the specific ARG family

478 in their chromosome (cARG-containing clade), while <20% of genomes in the
479 corresponding sister clade contained that ARG family. Only clade pairs with at least
480 two genomes in each clade were retained for analysis. A representative case involving
481 Citrobacter portucalensis is illustrated in Supplementary Figure S8.

482 For each identified clade pair, we calculated the gain and loss rates of plasmid
483 genes, distinguishing between plasmid-borne ARGs and other plasmid genes. For
484 each orthologous gene family, rates were computed as:

485 $$Rate_{OG,\,clade} = \frac{\sum_{b \in clade} Events_{OG,\,b}}{\sum_{b \in clade} Branch\ Length_b}$$

486 where events were summed across all branches within each clade, normalized by the
487 total branch length of that entire clade. Plasmid ARGs that matched the chromosomal
488 ARG family used to define the clade were excluded from the analysis. The median rate
489 across all orthologous gene families within each gene class (ARGs or non-ARGs) was
490 then used as the representative value for each clade.

491 Event rates were analyzed using linear mixed-effects models on the log scale, with
492 gain and loss modeled separately; the response was $\log(rate + \varepsilon)$, where $\varepsilon =$
493 $0.1 \times \min(\text{non-zero rate})$ within the corresponding event type. Fixed effects were clade
494 type (ARG-containing vs sister clade), gene class (plasmid ARGs vs other plasmid
495 genes), and their interaction; random effects were species, resistance gene family, and
496 the clade pair linking each ARG-containing clade with its sister (implemented as a
497 random intercept on Clade ID). Models were fitted using restricted maximum likelihood
498 and restricted to paired observations with both clade types available within a given
499 gene class. Post hoc contrasts from estimated marginal means assessed (i) ARG-
500 containing vs sister differences within each gene class and (ii) the clade-type × gene-

class interaction. In addition to the overall model, we performed family-specific analyses. For each resistance gene family with ≥ 4 available paired clades, we re-fit the model within that family using the same fixed-effects structure and random intercepts for species and clade pair, but omitting the family random effect (not identifiable within a single family). Estimates and 95% CIs were back-transformed to the original rate scale, with two-sided tests and Benjamini–Hochberg adjustment for multiple contrasts. Analyses were performed in R v4.4.0 using lme4 v1.1.35.5 (52), lmerTest v3.1.3 (53), and emmeans v 1.11.2.8.

To evaluate the robustness of our sister-clade analysis, we performed a sensitivity analysis by testing a gradient of clade-defining thresholds for cARG frequency: 70%/30%, 80%/20% (the primary threshold used in this study), 85%/15%, and 90%/10%. These alternative thresholds were applied to verify whether the observed associations between cARG presence and pARG gain/loss rates were sensitive to the stringency of clade selection. The fold-change in pARG gain rates and the decrease in pARG loss rates were assessed across all thresholds using the linear mixed-effects model structure described above.

The usage levels of ARG-associated antibiotics in clinical (human, Spain, 2023) and agricultural (veterinary, Spain, 2022) data were obtained from the One Health Trust's ResistanceMap (https://resistancemap.onehealthtrust.org/) and the European Medicines Agency's (EMA) ESVAC report (https://www.ema.europa.eu/), respectively, and were standardized into the CARD drug class.

Plasmid Unit Reconstruction and Similarity Network Analysis

To accurately determine plasmid numbers and mitigate overestimation, mob_recon (MOB-suite v3.1.9) (54) was employed to reconstruct plasmids across the entire dataset. In this section, each reconstructed plasmid was treated as an individual genomic unit. Only plasmid groups (primary clusters) comprising at least ten independent units and a consistent replicon type were included in the analysis. Genomic similarity within each group was assessed by calculating Mash distances, excluding edges between plasmid pairs with distances exceeding 0.025. The resulting similarity networks were visualized using Cytoscape v3.10.2 (55).

Resistance genes conferring resistance to monobactams, carbapenems, cephalosporins, or penicillins were grouped into a single beta-lactam category for the drug-class analysis. Similarly, the MLS group was used to encompass macrolides, lincosamides, pleuromutilins, and streptogramins (A and B).

ACKNOWLEDGEMENTS

Yue Liu is supported by the State Key Laboratory for Diagnosis and Treatment of Infectious Diseases (No. 202506) and the ZJU100 Young Professor Start-up Fund from Zhejiang University of China. Yang Liu is supported by the China Postdoctoral Science Foundation (No. 2025M783825).

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the author(s) used ChatGPT and Gemini to polish the language. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

548

549 **DECLARATION OF COMPETING INTERESTS**

550 The authors declare no competing financial interests.

551

552 **ETHICS APPROVAL**

553 There are no ethical concerns in this study.

554

555 **DATA AVAILABILITY**

556 Supplementary material is available for this article. Also, the supplement data, and core
557 script are available on Figshare under the DOI: 10.6084/m9.figshare.30186307.

558

559 **AUTHOR CONTRIBUTIONS**

560 **Yang Liu:** conceptualization, bioinformatics analysis, writing – original draft, writing –
561 review & editing. **Yue Liu:** writing – review & editing, funding acquisition.

562

563 **References**

564 1. Davies J, Davies D. 2010. Origins and evolution of antibiotic resistance.
565 Microbiol Mol Biol Rev 74:417-33.
566 2. Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging
567 dynamic view of the prokaryotic world. Nucleic Acids Res 36:6688-719.
568 3. Koonin EV, Wolf YI. 2012. Evolution of microbes and viruses: a paradigm shift
569 in evolutionary biology? Front Cell Infect Microbiol 2:119.
570 4. Bennett PM. 2008. Plasmid encoded antibiotic resistance: acquisition and
571 transfer of antibiotic resistance genes in bacteria. Br J Pharmacol 153 Suppl
572 1:S347-57.
573 5. Partridge SR, Kwong SM, Firth N, Jensen SO. 2018. Mobile Genetic Elements

Associated with Antimicrobial Resistance. Clin Microbiol Rev 31.

6. Redondo-Salvo S, Fernández-López R, Ruiz R, Vielva L, de Toro M, Rocha EPC, Garcillán-Barcia MP, de la Cruz F. 2020. Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. Nature Communications 11.

7. Coluzzi C, Rocha EPC. 2025. The Spread of Antibiotic Resistance Is Driven by Plasmids Among the Fastest Evolving and of Broadest Host Range. Molecular Biology and Evolution 42.

8. Benz F, Hall AR. 2023. Host-specific plasmid evolution explains the variable spread of clinical antibiotic-resistance plasmids. Proc Natl Acad Sci U S A 120:e2212147120.

9. Amos GCA, Ploumakis S, Zhang L, Hawkey PM, Gaze WH, Wellington EMH. 2018. The widespread dissemination of integrons throughout bacterial communities in a riverine system. ISME J 12:681-691.

10. Che Y, Yang Y, Xu X, Brinda K, Polz MF, Hanage WP, Zhang T. 2021. Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. Proc Natl Acad Sci U S A 118.

11. Dimitriu T, Matthews AC, Buckling A. 2021. Increased copy number couples the evolution of plasmid horizontal transmission and plasmid-encoded antibiotic resistance. Proc Natl Acad Sci U S A 118.

12. San Millan A. 2018. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. Trends Microbiol 26:978-985.

13. San Millan A, MacLean RC. 2017. Fitness Costs of Plasmids: a Limit to Plasmid Transmission. Microbiol Spectr 5.

14. Carattoli A. 2013. Plasmids and the spread of resistance. Int J Med Microbiol 303:298-304.

15. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius DJ, Hordijk J. 2018. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. J Antimicrob Chemother 73:1121-1137.

16. Mathers AJ, Peirano G, Pitout JD. 2015. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae. Clin Microbiol Rev 28:565-91.

17. van Duin D, Paterson DL. 2016. Multidrug-Resistant Bacteria in the Community:

608 Trends and Lessons Learned. Infect Dis Clin North Am 30:377-390.

609 18. Grundmann H, Livermore DM, Giske CG, Canton R, Rossolini GM, Campos J,
610 Vatopoulos A, Gniadkowski M, Toth A, Pfeifer Y, Jarlier V, Carmeli Y, Group CW.
611 2010. Carbapenem-non-susceptible Enterobacteriaceae in Europe:
612 conclusions from a meeting of national experts. Euro Surveill 15.

613 19. Gomez-Simmonds A, Annavajhala MK, Wang Z, Macesic N, Hu Y, Giddins MJ,
614 O'Malley A, Toussaint NC, Whittier S, Torres VJ, Uhlemann AC. 2018. Genomic
615 and Geographic Context for the Evolution of High-Risk Carbapenem-Resistant
616 Enterobacter cloacae Complex Clones ST171 and ST78. mBio 9.

617 20. Garcia-Gonzalez N, Beamud B, Fuster B, Giner S, Dominguez MV, Sanchez A,
618 Sevilla J, Coque TM, Gimeno C, Gonzalez-Candelas F, Networked Laboratory
619 for Antimicrobial Resistance Surveillance of Comunitat V. 2023. Tracking the
620 Emergence and Dissemination of a bla(NDM-23) Gene in a Multidrug
621 Resistance Plasmid of Klebsiella pneumoniae. Microbiol Spectr 11:e0258522.

622 21. Csuros M. 2010. Count: evolutionary analysis of phylogenetic profiles with
623 parsimony and likelihood. Bioinformatics 26:1910-2.

624 22. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA,
625 Edalatmand A, Petkau A, Syed SA, Tsang KK, Baker SJC, Dave M, McCarthy
626 MC, Mukiri KM, Nasir JA, Golbon B, Imtiaz H, Jiang XJ, Kaur K, Kwong M,
627 Liang ZC, Niu KC, Shan P, Yang JYJ, Gray KL, Hoad GR, Jia BF, Bhando T,
628 Carfrae LA, Farha MA, French S, Gordzevich R, Rachwalski K, Tu MM,
629 Bordeleau E, Dooley D, Griffiths E, Zubyk HL, Brown ED, Maguire F, Beiko RG,
630 Hsiao WWL, Brinkman FSL, Van Domselaar G, McArthur AG. 2023. CARD
631 2023: expanded curation, support for machine learning, and resistome
632 prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids
633 Research 51:D690-D699.

634 23. He S, Hickman AB, Varani AM, Siguier P, Chandler M, Dekker JP, Dyda F. 2015.
635 Insertion Sequence IS26 Reorganizes Plasmids in Clinically Isolated Multidrug-
636 Resistant Bacteria by Replicative Transposition. mBio 6:e00762.

637 24. Harmer CJ, Hall RM. 2016. IS26-Mediated Formation of Transposons Carrying
638 Antibiotic Resistance Genes. mSphere 1.

639 25. Partridge SR, Tsafnat G, Coiera E, Iredell JR. 2009. Gene cassettes and
640 cassette arrays in mobile resistance integrons. FEMS Microbiol Rev 33:757-84.

641 26. Souque C, Escudero JA, MacLean RC. 2021. Integron activity accelerates the

642 evolution of antibiotic resistance. Elife 10.

643 27. Nicoloff H, Hjort K, Andersson DI, Wang H. 2024. Three concurrent
644 mechanisms generate gene copy number variation and transient antibiotic
645 heteroresistance. Nat Commun 15:3981.

646 28. Vogwill T, MacLean RC. 2015. The genetic basis of the fitness costs of
647 antimicrobial resistance: a meta-analysis approach. Evol Appl 8:284-95.

648 29. Pal A, Andersson DI. 2024. Bacteria can compensate the fitness costs of
649 amplified resistance genes via a bypass mechanism. Nat Commun 15:2333.

650 30. Oliva M, Calia C, Ferrara M, D'Addabbo P, Scrascia M, Mule G, Monno R,
651 Pazzani C. 2020. Antimicrobial resistance gene shuffling and a three-element
652 mobilisation system in the monophasic Salmonella typhimurium strain ST1030.
653 Plasmid 111:102532.

654 31. Domingues S, Harms K, Fricke WF, Johnsen PJ, da Silva GJ, Nielsen KM. 2012.
655 Natural transformation facilitates transfer of transposons, integrons and gene
656 cassettes between bacterial species. PLoS Pathog 8:e1002837.

657 32. Zongo PD, Cabanel N, Royer G, Depardieu F, Hartmann A, Naas T, Glaser P,
658 Rosinski-Chupin I. 2024. An antiplasmid system drives antibiotic resistance
659 gene integration in carbapenemase-producing Escherichia coli lineages. Nat
660 Commun 15:4093.

661 33. Oliva M, Monno R, Addabbo P, Pesole G, Scrascia M, Calia C, Dionisi AM,
662 Chiara M, Horner DS, Manzari C, Pazzani C. 2018. IS26 mediated antimicrobial
663 resistance gene shuffling from the chromosome to a mosaic conjugative FII
664 plasmid. Plasmid 100:22-30.

665 34. Wang Y, Dagan T. 2024. The evolution of antibiotic resistance islands occurs
666 within the framework of plasmid lineages. Nat Commun 15:4555.

667 35. Beyrouthy R, Barets M, Marion E, Dananche C, Dauwalder O, Robin F,
668 Gauthier L, Jousset A, Dortet L, Guerin F, Benet T, Cassier P, Vanhems P,
669 Bonnet R. 2018. Novel Enterobacter Lineage as Leading Cause of Nosocomial
670 Outbreak Involving Carbapenemase-Producing Strains. Emerg Infect Dis
671 24:1505-1515.

672 36. Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, Doi Y, Tian G, Dong B,
673 Huang X, Yu LF, Gu D, Ren H, Chen X, Lv L, He D, Zhou H, Liang Z, Liu JH,
674 Shen J. 2016. Emergence of plasmid-mediated colistin resistance mechanism
675 MCR-1 in animals and human beings in China: a microbiological and molecular

676 biological study. Lancet Infect Dis 16:161-8.

677 37. Vaz CS, Streck AF, Michael GB, Marks FS, Rodrigues DP, Dos Reis EM,
678 Cardoso MR, Canal CW. 2010. Antimicrobial resistance and subtyping of
679 Salmonella enterica subspecies enterica serovar Enteritidis isolated from
680 human outbreaks and poultry in southern Brazil. Poult Sci 89:1530-6.

681 38. Crespo R, Jeffrey JS, Chin RP, Senties-Cue G, Shivaprasad HL. 2004.
682 Phenotypic and genotypic characterization of Salmonella arizonae from an
683 integrated turkey operation. Avian Dis 48:344-50.

684 39. Nogueira T, Botelho A. 2021. Metagenomics and Other Omics Approaches to
685 Bacterial Communities and Antimicrobial Resistance Assessment in
686 Aquacultures. Antibiotics (Basel) 10.

687 40. Cury J, Touchon M, Rocha EPC. 2017. Integrative and conjugative elements
688 and their hosts: composition, distribution and organization. Nucleic Acids Res
689 45:8943-8956.

690 41. Botelho J. 2023. Defense systems are pervasive across chromosomally
691 integrated mobile genetic elements and are inversely correlated to virulence
692 and antimicrobial resistance. Nucleic Acids Res 51:4385-4397.

693 42. Johnson CM, Grossman AD. 2015. Integrative and Conjugative Elements
694 (ICEs): What They Do and How They Work. Annu Rev Genet 49:577-601.

695 43. Touchon M, Moura de Sousa JA, Rocha EP. 2017. Embracing the enemy: the
696 diversification of microbial gene repertoires by phage-mediated horizontal gene
697 transfer. Curr Opin Microbiol 38:66-73.

698 44. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P.
699 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat
700 Biotechnol 38:1079-1086.

701 45. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy
702 TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN,
703 Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P,
704 Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO,
705 Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S,
706 Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon
707 KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW,
708 Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, et al. 2017.
709 Minimum information about a single amplified genome (MISAG) and a

710 metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature
711 Biotechnology 35:725-731.
712 46. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population
713 genomics: BIGSdb software, the PubMLST.org website and their applications.
714 Wellcome Open Res 3:124.
715 47. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, Chain PSG, Nayfach
716 S, Kyrpides NC. 2024. Identification of mobile genetic elements with geNomad.
717 Nat Biotechnol 42:1303-1312.
718 48. Schwengers O, Barth P, Falgenhauer L, Hain T, Chakraborty T, Goesmann A.
719 2020. Platon: identification and characterization of bacterial plasmid contigs in
720 short-read draft assemblies exploiting protein sequence-based replicon
721 distribution scores. Microb Genom 6.
722 49. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for
723 comparative genomics. Genome Biol 20:238.
724 50. Manzano-Morales S, Liu Y, Gonzalez-Bodi S, Huerta-Cepas J, Iranzo J. 2023.
725 Comparison of gene clustering criteria reveals intrinsic uncertainty in
726 pangenome analyses. Genome Biol 24:250.
727 51. Puigbò P, Lobkovsky AE, Kristensen DM, Wolf YI, Koonin EV. 2014. Genomes
728 in turmoil: quantification of genome dynamics in prokaryote supergenomes.
729 Bmc Biology 12.
730 52. Bates D, Mächler M, Bolker BM, Walker SC. 2015. Fitting Linear Mixed-Effects
731 Models Using lme4. Journal of Statistical Software 67:1-48.
732 53. Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest Package: Tests
733 in Linear Mixed Effects Models. Journal of Statistical Software 82:1-26.
734 54. Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering,
735 reconstruction and typing of plasmids from draft assemblies. Microb Genom 4.
736 55. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N,
737 Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for
738 integrated models of biomolecular interaction networks. Genome Res 13:2498-
739 504.
740

741 **Figure legends**

742 **Figure 1. Evolutionary dynamics of plasmid-borne antibiotic resistance genes**

743 **compared to other plasmid genes in Enterobacteriaceae.**

744 (A) Gain rates (appearance of new gene families per substitution per site) for ARGs
745 versus other plasmid genes. (B) Loss rates (disappearance of gene families per
746 substitution per site) for ARGs versus other plasmid genes. (C) Expansion rates
747 (addition of gene copies within existing families per substitution per site) for ARGs
748 versus other plasmid genes. (D) Reduction rates (loss of gene copies within existing
749 families per substitution per site) for ARGs versus other plasmid genes. Gene families
750 were defined as orthologous groups (OGs).

751

752 **Figure 2. Expansion and reduction rate patterns across antibiotic resistance**
753 **classes and Enterobacteriaceae species.**

754 Heatmaps showing evolutionary rates for different antibiotic resistance classes across
755 bacterial species. (A) Expansion rates. (B) Reduction rates. Rows represent
756 Enterobacteriaceae species, and columns represent CARD-defined antibiotic
757 resistance classes. Color intensity indicates log10 scaled rate magnitude, with darker
758 colors representing higher rates. White cells indicate absence of the ARG class in the
759 species.

760

761 **Figure 3. Association between presence of cARGs in bacterial lineages and**
762 **plasmid borne ARG acquisition.**

763 (A) Schematic showing sister clade comparison approach. Phylogenetic clades
764 containing high frequencies (≥80%) of specific cARG families are compared to sister
765 clades with low frequencies (<20%). Colored circles represent different bacterial clades.

766 (B) Fold-change comparison of evolutionary rates between ARG-containing clades and
767 sister clades. Y-axis shows the fold-change ratio (ARG clade / Sister clade) for gain
768 (purple) and loss (orange) rates of plasmid ARGs versus other plasmid genes. Dashed
769 line at y=1 indicates no difference between clades. (C) Interaction effect showing the
770 differential impact of cARG presence on gain versus loss dynamics. Y-axis represents
771 the interaction effect (fold-change, log scale) between clade type and evolutionary
772 process. Error bars show 95% confidence intervals. Statistical significance indicators:
773 ***$p < 0.001$, *$p < 0.05$, ns = not significant.

774

775 **Figure 4. ARG distribution within cARG-associated lineages and antibiotic**
776 **consumption levels.**

777 (A) Heatmap of ARG profiles across cARG-containing clades. The x-axis labels denote
778 the bacterial species and total genome numbers within each analyzed clade. Species
779 names are color-coded according to their primary habitat: Clinical (red), Environmental
780 (blue), Agricultural (green), Mixed (purple), or Unknown (grey). The y-axis lists
781 antibiotic drug classes as defined by the Comprehensive Antibiotic Resistance
782 Database (CARD). Color intensity indicates the number of genomes harboring specific
783 resistance gene classes. (B) Clinical antibiotic consumption. Bar chart representing
784 human antibiotic usage data from Spain (2023). The x-axis measures consumption in
785 Defined Daily Doses (DDD) per 1,000 people per day. (C) Agricultural antibiotic
786 consumption. Bar chart representing veterinary antimicrobial sales data from Spain
787 (2022). The x-axis measures sales in milligrams per population correction unit
788 (mg/PCU). Antibiotic categories are mapped into standardized CARD drug classes for
789 cross-comparison with the heatmap.

Figure 5. Genetic similarity networks of specific plasmid replicon types and their associated antibiotic resistance gene (ARG) profiles.

Each node represents an individual plasmid, and edges connect plasmids with high genetic similarity. Nodes are colored according to the functional class of the carried antibiotic resistance genes (ARGs) as defined by the CARD database.

Figure1

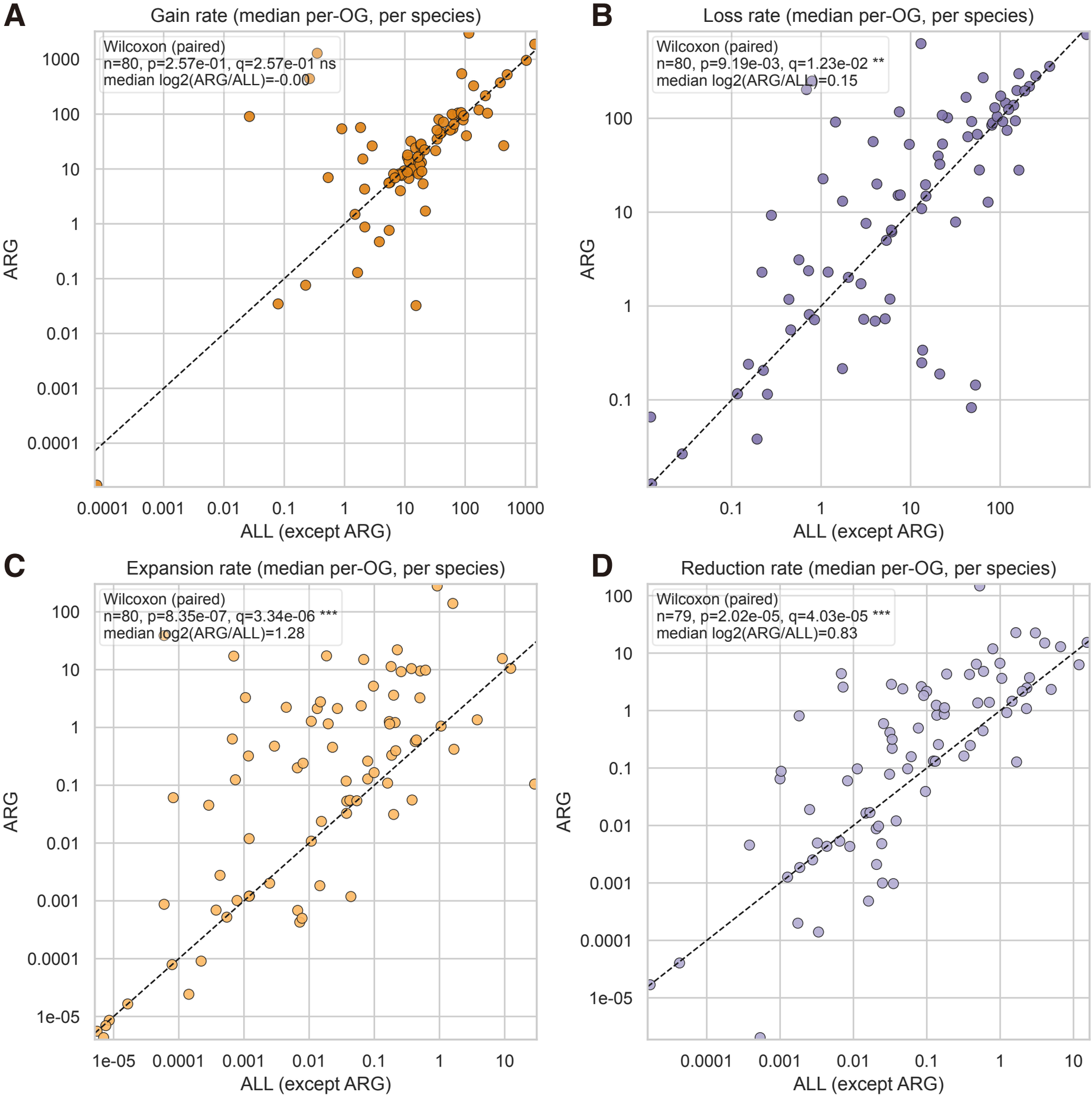

A
Gain rate (median per-OG, per species)
Wilcoxon (paired)
n=80, p=2.57e-01, q=2.57e-01 ns
median log2(ARG/ALL)=-0.00
ARG
ALL (except ARG)
B
Loss rate (median per-OG, per species)
Wilcoxon (paired)
n=80, p=9.19e-03, q=1.23e-02 **
median log2(ARG/ALL)=0.15
ARG
ALL (except ARG)
C
Expansion rate (median per-OG, per species)
Wilcoxon (paired)
n=80, p=8.35e-07, q=3.34e-06 ***
median log2(ARG/ALL)=1.28
ARG
ALL (except ARG)
D
Reduction rate (median per-OG, per species)
Wilcoxon (paired)
n=79, p=2.02e-05, q=4.03e-05 ***
median log2(ARG/ALL)=0.83
ARG
ALL (except ARG)

Figure 2

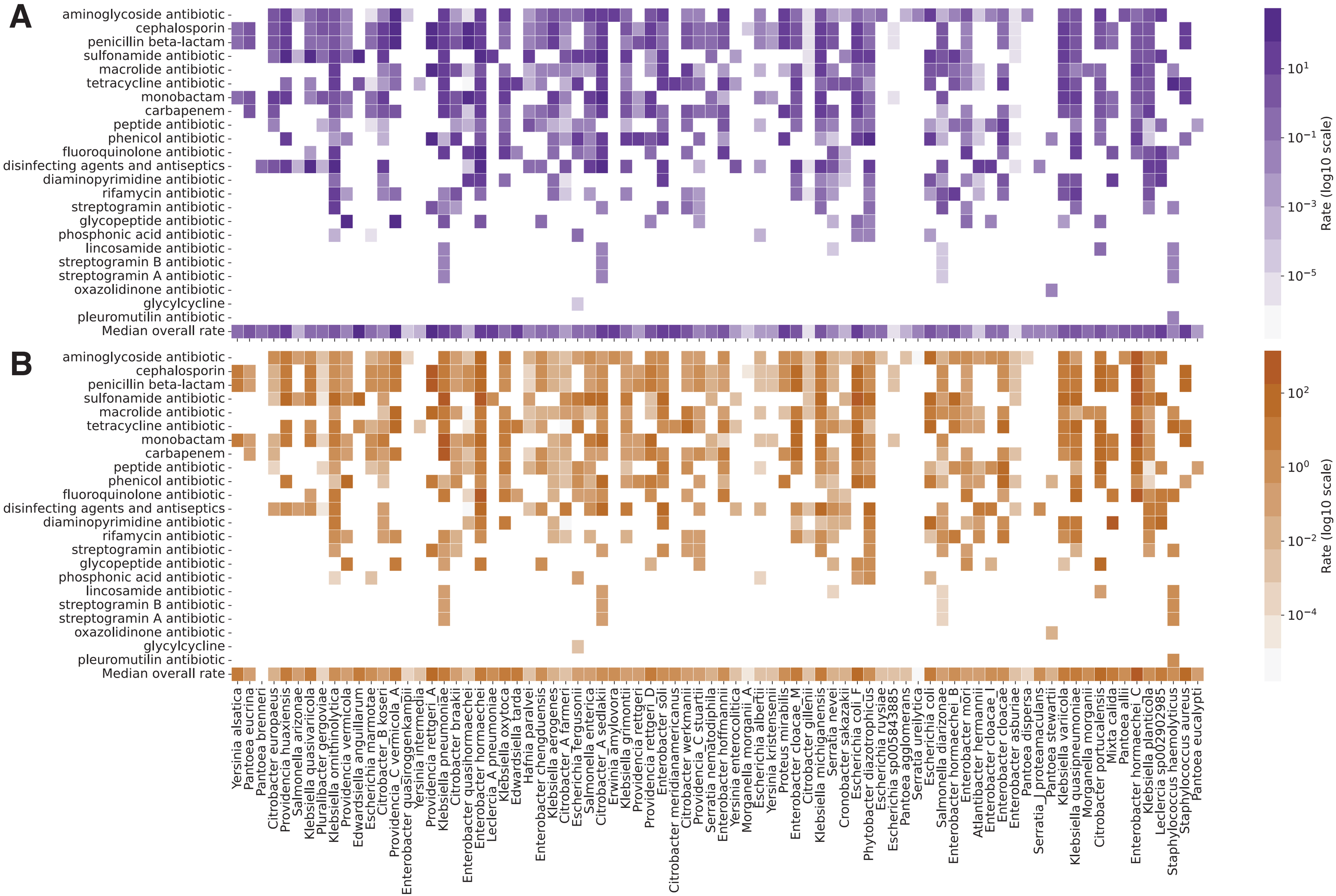

A
aminoglycoside antibiotic
cephalosporin
penicillin beta-lactam
sulfonamide antibiotic
macrolide antibiotic
tetracycline antibiotic
monobactam
carbapenem
peptide antibiotic
phenicol antibiotic
fluoroquinolone antibiotic
disinfecting agents and antiseptics
diaminopyrimidine antibiotic
rifamycin antibiotic
streptogramin antibiotic
glycopeptide antibiotic
phosphonic acid antibiotic
lincosamide antibiotic
streptogramin B antibiotic
streptogramin A antibiotic
oxazolidinone antibiotic
glycylcycline
pleuromutilin antibiotic
Median overall rate
Rate (log10 scale)
10^{1}
10^{-1}
10^{-3}
10^{-5}
B
10^{2}
10^{0}
10^{-2}
10^{-4}
Yersinia alsatica
Pantoea eucrina
Pantoea brenneri
Citrobacter europaeus
Providencia huaxiensis
Salmonella arizonae
Klebsiella quasivariicola
Pluralibacter gergoviae
Klebsiella ornithinolytica
Providencia vermicola
Edwardsiella anguillarum
Escherichia marmotae
Citrobacter_B koseri
Providencia_C vermicola_A
Enterobacter quasiroggenkampii
Yersinia intermedia
Providencia rettgeri_A
Klebsiella pneumoniae
Citrobacter braakii
Enterobacter quasihormaechei
Enterobacter hormaechei
Leclercia_A pneumoniae
Klebsiella oxytoca
Edwardsiella tarda
Hafnia paralvei
Enterobacter chengduensis
Klebsiella aerogenes
Citrobacter_A farmeri
Escherichia fergusonii
Salmonella enterica
Citrobacter_A sedlakii
Erwinia amylovora
Klebsiella grimontii
Providencia rettgeri
Providencia rettgeri_D
Enterobacter soli
Citrobacter meridianamericanus
Citrobacter werkmanii
Providencia_C stuartii
Serratia nematodiphila
Enterobacter hoffmannii
Yersinia enterocolitica
Morganella morganii_A
Escherichia albertii
Yersinia kristensenii
Proteus mirabilis
Enterobacter cloacae_M
Citrobacter gillenii
Klebsiella michiganensis
Serratia nevei
Cronobacter sakazakii
Escherichia coli_F
Phytobacter diazotrophicus
Escherichia ruysiae
Escherichia sp005843885
Pantoea agglomerans
Serratia ureilytica
Escherichia coli
Salmonella diarizonae
Enterobacter hormaechei_B
Enterobacter mori
Atlantibacter hermannii
Enterobacter cloacae_I
Enterobacter cloacae
Enterobacter asburiae
Pantoea dispersa
Serratia_J proteamaculans
Pantoea stewartii
Klebsiella variicola
Klebsiella quasipneumoniae
Morganella morganii
Citrobacter portucalensis
Mixta calida
Pantoea allii
Enterobacter hormaechei_C
Klebsiella planticola
Leclercia sp002902985
Staphylococcus haemolyticus
Staphylococcus aureus
Pantoea eucalypti

Figure 3

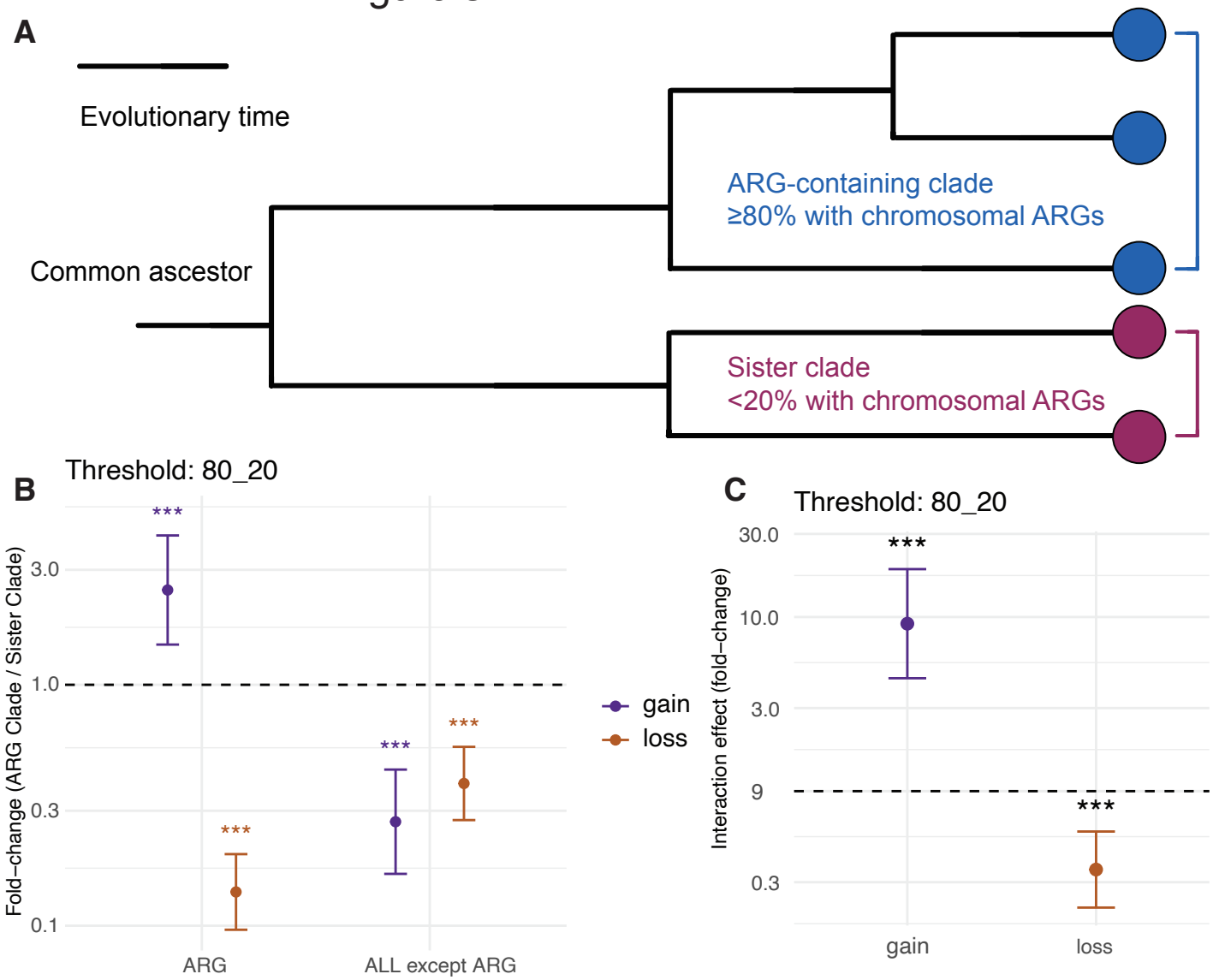

A
Evolutionary time
Common ascestor
ARG-containing clade
≥80% with chromosomal ARGs
Sister clade
<20% with chromosomal ARGs
B
Threshold: 80_20
Fold−change (ARG Clade / Sister Clade)
3.0
1.0
0.3
0.1
***
ARG
ALL except ARG
gain
loss
C
Threshold: 80_20
Interaction effect (fold−change)
30.0
10.0
3.0
9
0.3
gain
loss

Figure 4

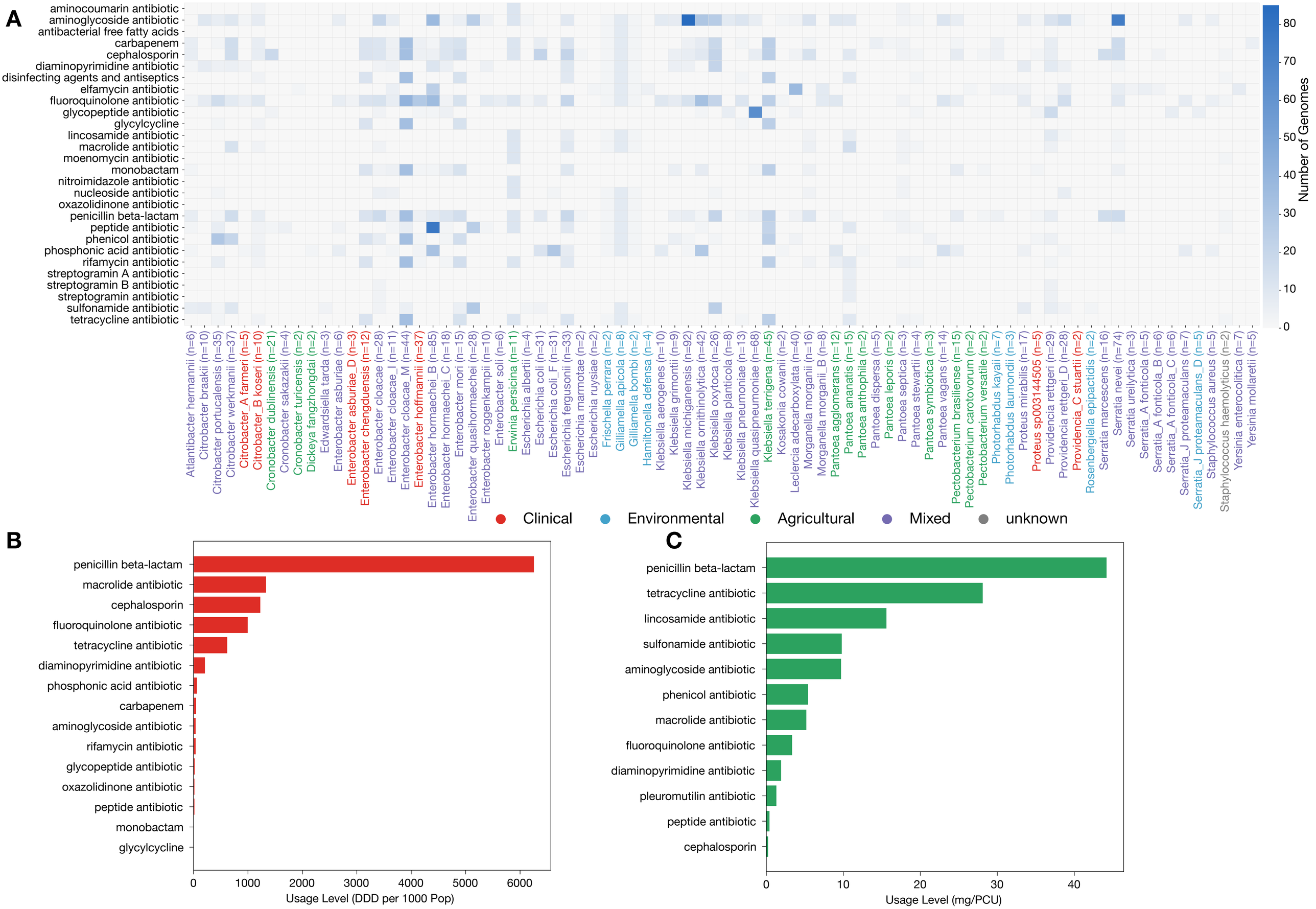
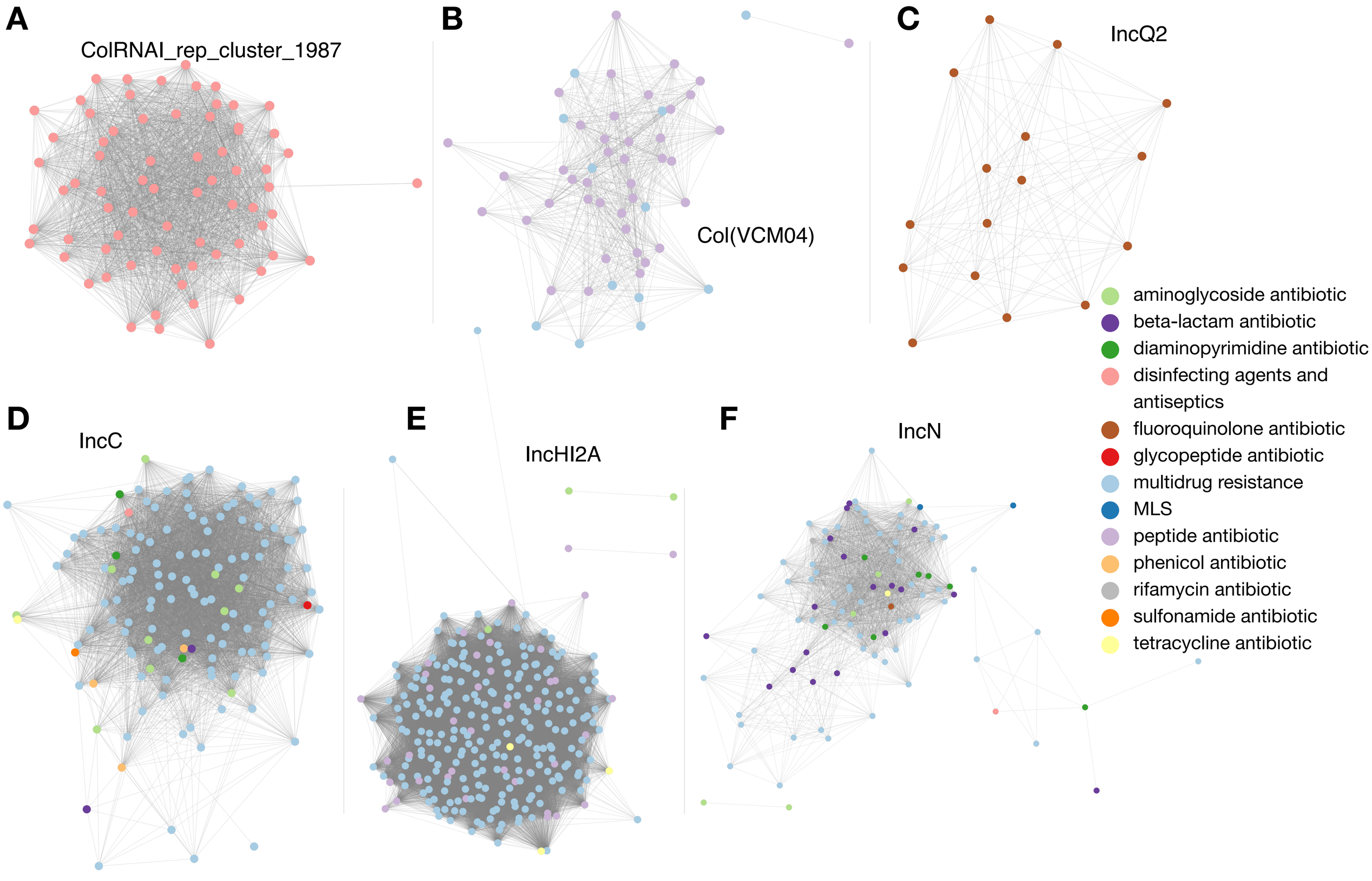

A
aminocoumarin antibiotic
aminoglycoside antibiotic
antibacterial free fatty acids
carbapenem
cephalosporin
diaminopyrimidine antibiotic
disinfecting agents and antiseptics
elfamycin antibiotic
fluoroquinolone antibiotic
glycopeptide antibiotic
glycylcycline
lincosamide antibiotic
macrolide antibiotic
moenomycin antibiotic
monobactam
nitroimidazole antibiotic
nucleoside antibiotic
oxazolidinone antibiotic
penicillin beta-lactam
peptide antibiotic
phenicol antibiotic
phosphonic acid antibiotic
rifamycin antibiotic
streptogramin A antibiotic
streptogramin B antibiotic
streptogramin antibiotic
sulfonamide antibiotic
tetracycline antibiotic
Atlantibacter hermannii (n=6)
Citrobacter braakii (n=10)
Citrobacter portucalensis (n=35)
Citrobacter werkmanii (n=37)
Citrobacter_A farmeri (n=5)
Citrobacter_B koseri (n=10)
Cronobacter dublinensis (n=21)
Cronobacter sakazakii (n=4)
Cronobacter turicensis (n=2)
Dickeya fangzhongdai (n=2)
Edwardsiella tarda (n=3)
Enterobacter asburiae (n=6)
Enterobacter asburiae_D (n=3)
Enterobacter chengduensis (n=12)
Enterobacter cloacae (n=28)
Enterobacter cloacae_I (n=11)
Enterobacter cloacae_M (n=44)
Enterobacter hoffmannii (n=37)
Enterobacter hormaechei_B (n=85)
Enterobacter hormaechei_C (n=18)
Enterobacter mori (n=15)
Enterobacter quasihormaechei (n=28)
Enterobacter roggenkampii (n=10)
Enterobacter soli (n=6)
Erwinia persicina (n=11)
Escherichia albertii (n=4)
Escherichia coli (n=31)
Escherichia coli_F (n=31)
Escherichia fergusonii (n=33)
Escherichia marmotae (n=2)
Escherichia ruysiae (n=2)
Frischella perrara (n=2)
Gilliamella apicola (n=8)
Gilliamella bombi (n=2)
Hamiltonella defensa (n=4)
Klebsiella aerogenes (n=10)
Klebsiella grimontii (n=9)
Klebsiella michiganensis (n=92)
Klebsiella ornithinolytica (n=42)
Klebsiella oxytoca (n=26)
Klebsiella planticola (n=8)
Klebsiella pneumoniae (n=13)
Klebsiella quasipneumoniae (n=68)
Klebsiella terrigena (n=45)
Kosakonia cowanii (n=2)
Leclercia adecarboxylata (n=40)
Morganella morganii (n=16)
Morganella morganii_B (n=8)
Pantoea agglomerans (n=12)
Pantoea ananatis (n=15)
Pantoea anthophila (n=2)
Pantoea dispersa (n=5)
Pantoea leporis (n=2)
Pantoea septica (n=3)
Pantoea stewartii (n=4)
Pantoea symbiotica (n=3)
Pantoea vagans (n=14)
Pectobacterium brasiliense (n=15)
Pectobacterium carotovorum (n=2)
Pectobacterium versatile (n=2)
Photorhabdus kayaii (n=7)
Photorhabdus laumondii (n=3)
Proteus mirabilis (n=17)
Proteus sp003144505 (n=5)
Providencia rettgeri (n=29)
Providencia rettgeri_D (n=28)
Providencia_C stuartii (n=2)
Rosenbergiella epipactidis (n=2)
Serratia marcescens (n=16)
Serratia nevei (n=74)
Serratia ureilytica (n=3)
Serratia_A fonticola (n=5)
Serratia_A fonticola_B (n=6)
Serratia_A fonticola_C (n=6)
Serratia_J proteamaculans (n=7)
Serratia_J proteamaculans_D (n=5)
Staphylococcus aureus (n=5)
Staphylococcus haemolyticus (n=2)
Yersinia enterocolitica (n=7)
Yersinia mollaretii (n=5)
Number of Genomes
0
10
20
30
40
50
60
70
80
Clinical
Environmental
Agricultural
Mixed
unknown
B
penicillin beta-lactam
macrolide antibiotic
cephalosporin
fluoroquinolone antibiotic
tetracycline antibiotic
diaminopyrimidine antibiotic
phosphonic acid antibiotic
carbapenem
aminoglycoside antibiotic
rifamycin antibiotic
glycopeptide antibiotic
oxazolidinone antibiotic
peptide antibiotic
monobactam
glycylcycline
0
1000
2000
3000
4000
5000
6000
Usage Level (DDD per 1000 Pop)
C
penicillin beta-lactam
tetracycline antibiotic
lincosamide antibiotic
sulfonamide antibiotic
aminoglycoside antibiotic
phenicol antibiotic
macrolide antibiotic
fluoroquinolone antibiotic
diaminopyrimidine antibiotic
pleuromutilin antibiotic
peptide antibiotic
cephalosporin
0
10
20
30
40
Usage Level (mg/PCU)

Figure 5


A
ColRNAI_rep_cluster_1987
B
Col(VCM04)
C
IncQ2
D
IncC
E
IncHI2A
F
IncN
aminoglycoside antibiotic
beta-lactam antibiotic
diaminopyrimidine antibiotic
disinfecting agents and antiseptics
fluoroquinolone antibiotic
glycopeptide antibiotic
multidrug resistance
MLS
peptide antibiotic
phenicol antibiotic
rifamycin antibiotic
sulfonamide antibiotic
tetracycline antibiotic