Gain and loss of plasmid-borne antibiotic resistance genes are associated with their chromosomal counterparts and the broader plasmid gene repertoire in Enterobacteriaceae

Yang Liu[1, *], Yue Liu[1,2, *]

[1] Department of Microbiology, Zhejiang University School of Medicine, Zhejiang, China

[2] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases of the First Affiliated Hospital, 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 and loss 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, pARGs were co-transferred with a broad range of plasmid functional genes, including defense, metabolic, and mobilome modules throughout evolutionary history. Moreover, bacterial clades harboring chromosomal ARGs (cARGs) showed significantly higher acquisition of plasmid-borne resistance than did their sister clades lacking corresponding cARGs. Our results demonstrate that pARGs in Enterobacteriaceae are co-transferred with the broader plasmid gene repertoire on an evolutionary timescale, whereas copy-number dynamics are distinct from plasmids. Moreover, cARGs can serve as markers of lineages predisposed to accumulate plasmid-mediated resistance.

KEYWORDS antibiotic resistance genes; Enterobacterales; gene gain and loss, plasmid; mobile genetic elements.

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 and loss processes as other plasmid genes but exhibit markedly higher and species-dependent copy number changes. Moreover, bacterial lineages carrying chromosomal ARGs are significantly more likely to acquire additional plasmid-mediated resistance. This association suggests that plasmid and chromosomal ARGs could act synergistically to accelerate resistance accumulation within specific lineages. Identifying such evolutionarily high-risk 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 (Davies and Davies 2010). In bacterial genome evolution, gene gain through horizontal gene transfer (HGT) and gene loss (via selection) are the primary forces shaping genomic repertoires (Koonin and Wolf 2008, Koonin and Wolf 2012). Mobile genetic elements (MGEs), particularly conjugative plasmids, serve as principal vehicles of HGT and are central to the spread of antibiotic resistance genes (ARGs) (Bennett 2008, Partridge, Kwong et al. 2018, Redondo-Salvo, Fernández-López et al. 2020).

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 (Coluzzi and Rocha 2025). Their plasticity is enhanced by high densities of transposable elements and recombination systems, which accelerate gene repertoire turnover (Amos, Ploumakis et al. 2018, Che, Yang et al. 2021). Moreover, evolutionary innovations such as replication-control mutations can increase plasmid copy number, thereby amplifying resistance levels and enhancing transfer efficiency (Dimitriu, Matthews et al. 2021). Thus, plasmids are highly dynamic replicons characterized by frequent horizontal transfer, rapid turnover, and extensive genetic plasticity (San Millan and MacLean 2017, San Millan 2018).

Given their evolutionary plasticity, understanding plasmid-mediated ARG turnover is particularly critical in clinically important bacterial families such as Enterobacteriaceae (San Millan 2018). The Enterobacteriaceae family harbors extensive plasmid diversity and constitutes one of the largest reservoirs of clinically relevant ARGs (Carattoli 2013, Mathers, Peirano et al. 2015, Rozwandowicz, Brouwer et al. 2018). Members of this family, including *Escherichia*, *Klebsiella*, *Enterobacter*, and *Salmonella*, are major contributors to multidrug-resistant infections in hospitals and communities worldwide (van Duin and Paterson 2016). Although numerous clinical studies have documented outbreaks of antibiotic resistance within the Enterobacteriaceae family (Grundmann, Livermore et al. 2010, Mathers, Peirano et al. 2015, van Duin and Paterson 2016), most of these reports have focused on the dissemination of particular resistance plasmids or the epidemiology of individual ARGs (Carattoli 2013, Mathers, Peirano et al. 2015, Rozwandowicz, Brouwer et al. 2018). 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) identify non-ARG gene modules that consistently co-transfer with ARGs, and (iii) 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 and loss 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 (Csuros 2010) to quantify gene turnover rates across 6,895 genomes of which 6,130 contained at least one identified plasmid sequence (Supplementary Table S1). The approach estimates four key evolutionary processes: gain (appearance of new gene families), loss (disappearance of entire gene families), expansion (increase in gene copy number within existing families), and reduction (decrease in gene copy number within existing families). Since we focused exclusively on plasmid-borne genes, expansion and reduction can be interpreted as copy number increases or decreases of plasmid gene families within individual genomes. We

compared overall rates between plasmid borne ARGs (pARGs) and all other plasmid genes (non-ARGs) across 98 species. Our analysis revealed that pARGs exhibit substantially different dynamics for gene gain and loss versus copy number variation (Figure 1, Supplementary Table S2). Both gain and loss rates showed no significant differences between pARGs and non-ARGs ($n = 98$, gain: adjusted $p = 0.173$, loss: adjusted $p = 0.080$; Figure 1A-B), with data points forming relatively compact distributions across species. This suggests that the fundamental processes governing the acquisition and elimination of entire gene families operate similarly for resistance and non-resistance genes. In contrast, pARGs exhibited significantly elevated rates for both expansion ($n = 97$, adjusted $p = 3.21\times10^{-5}$, median $\log_2$(ARG/ALL) = 1.09; Figure 1C) and reduction (adjusted $p = 4.17\times10^{-4}$, median $\log_2$(ARG/ALL) = 0.39; Figure 1D) events. The high scatter in these data points indicated substantial species-specific variation in the extent to which expansion and reduction dynamics differ between pARGs and other plasmid genes, suggesting that genes copy number fluctuations on the genomes are strongly influenced by which bacterial species carries the plasmid.

The analysis of copy-number changes revealed that pARGs showed balanced dynamics between expansion and reduction rates (median $\log_2$ ratio = -1.38, adjusted $p = 0.47$, Supplementary Figure S1). In contrast, the overall gain/loss ratios displayed gain bias for both pARGs and non-ARGs (ARGs: median $\log_2$(gain/loss) = 3.87, adjusted $p = 4.27\times10^{-8}$; non-ARGs: median = 4.69, adjusted $p = 6.35\times10^{-15}$; Figure 1E-F).

In summary, across Enterobacteriaceae, gain and loss rates show no pARG-specific signal, indicating that acquisition and elimination of pARG families follow common plasmid evolutionary pathways. In contrast, pARGs exhibit significantly

elevated expansion and reduction rates with pronounced species-level variation, revealing that copy-number dynamics are shaped by lineage-specific factors.

Turnover dynamics of antibiotic resistance genes are primarily species dependent

To test whether the species-level patterns in Figure 1 are consistent across resistance mechanisms or vary among drug classes, which are defined by their mechanisms of action, we assigned orthologous gene families to CARD-defined drug classes and estimated gain, loss, expansion, and reduction rates for each class across species. Using the CARD database (Alcock, Huynh et al. 2023), we identified and analyzed the turnover dynamics of specific drug classes, which are groups of antibiotics with similar chemical structures and mechanisms of action. Our analysis revealed that the evolutionary rates for all four event types were strongly species-dependent (Kruskal–Wallis test; gain: $H = 735.9$, $p = 7.6\times10^{-99}$, $\eta^2 = 0.84$; loss: $H = 660.9$, $p = 9.1\times10^{-85}$, $\eta^2 = 0.74$; expansion: $H = 511.5$, $p = 5.9\times10^{-58}$, $\eta^2 = 0.56$; reduction: $H = 583.2$, $p = 7.5\times10^{-71}$, $\eta^2 = 0.66$), with substantial variation across species for gain, loss, expansion, and reduction (Supplementary Figure S2A–B; Figure 2A–B). In contrast, drug class effects were weak: only marginal signals were detected for gain ($p = 0.021$, $\eta^2 = 0.02$) and expansion ($p = 0.019$, $\eta^2 = 0.02$), while loss ($p = 0.97$, $\eta^2 \approx 0$) and reduction ($p = 0.35$, $\eta^2 < 0.01$) showed no significant drug class dependence. Full results are provided in Supplementary Table S3.

Furthermore, species that exhibited high expansion rates typically also showed elevated reduction rates for the same drug classes (n=96, Spearman's rank correlation

with 5,000 permutations: $\rho=0.57$, permutation $p\approx0.0002$). By contrast, species with high gain rates did not necessarily display the corresponding high loss rates ($n=98$, $\rho=-0.15$, $p\approx0.145$), indicating that plasmid-borne gene copy number increases and decreases within existing gene families are more tightly coupled than complete gene family acquisition and loss.

In summary, antibiotic resistance gene evolution shows distinct signatures across event types: gain and loss are strongly species dependent with negligible drug class influence, expansion is likewise dominated by species with only minor drug class effects, and reduction is also primarily species dependent. Moreover, expansion and reduction are coupled, whereas gain and loss are largely independent.

Plasmid-borne antibiotic resistance genes co-transfer with diverse plasmid functional modules and often entire plasmids

To identify genes that are potentially co-transferred with antibiotic resistance genes on plasmids, we quantified the relationship between per-branch gain and loss rates of pARGs with those of individual COG-annotated gene families within each species. For each species tree, branch-level rates were pooled to compute overall correlations, and significance was evaluated using Spearman correlation followed by phylogenetic Brownian motion simulations to control for phylogenetic effects.

Focusing on 85 species with plasmid borne pARGs exhibiting non-zero gain or loss rates, we found correlated evolutionary dynamics between pARGs and genes from diverse functional categories within plasmids (Figure 3, Supplementary Table S4). Although our study focuses on plasmid-encoded genes, many correlated COG

179 categories were not mobilomes (COG category X), but rather belonged to categories
180 such as defense system (V), various metabolic functions, and cellular processes, often
181 with high numbers of unique COG families involved (Figure 3A; line thickness reflects
182 this number). Considering both the number of unique COG families involved (Figure
183 3A) and the number of species showing significant correlations (Figure 3B), gain and
184 loss correlations occurred at similar frequencies, indicating that co-dynamics were not
185 restricted to a single event type (full results in Supplementary Table S4). At the species
186 level, genes that tended to co-gain with pARGs also tended to co-loss with them over
187 evolutionary timescales (Figure 3B).

188 Additionally, 29 species showed significant correlations when all COG categories
189 were analyzed together, 18 of which involved both gain and loss correlations
190 (Supplementary Table S4), suggesting that in these species, pARGs may be
191 transferred together with entire plasmids rather than as isolated genes. A comparison
192 of plasmid length between the 29 species with significant correlations across all COG
193 categories and the 56 species without such correlations revealed a statistically
194 significant difference (Mann–Whitney U test, $p < 0.001$); however, the effect size was
195 negligible (Cliff's delta = –0.034, Cohen's d = –0.081), indicating that plasmid size does
196 not account for the observed correlations.

197 These findings provide quantitative evidence that co-transfer patterns extend
198 beyond a few functional categories and encompass a broad spectrum of plasmid-
199 encoded genes, with different COG categories contributing to varying degrees. The
200 consistent coupling of high gain rates with high loss rates across gene categories
201 indicates that plasmid-borne genes experience bidirectional turnover, with frequent
202 acquisition accompanied by frequent loss. Importantly, in nearly one-third of the

species analyzed, pARGs co-transferred together with the broader plasmid gene repertoire, indicating that entire plasmids often represent the fundamental units of resistance dissemination.

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

Our previous analyses revealed that pARGs exhibit similar gain and loss rates compared to other plasmid genes. We further investigated whether additional factors might be associated with the evolutionary dynamics of pARGs, specifically examining whether the presence of antibiotic resistance genes in bacterial chromosomes is associated with altered evolutionary rates of pARGs. We employed a comparative phylogenetic approach using sister clade analysis (Figure 4A), identifying 436 pairs of phylogenetic clades where one clade (lineage) contained ≥ 80% of genomes harboring a specific ARG family in their chromosome (ARG-containing clade) while <20% of genomes in the sister clade (lineage) contained that homologous ARG family. For each clade pair, we calculated gain and loss rates for plasmid genes by summing evolutionary events across all branches within each clade and normalizing by total branch length. In total, 78 species and 79 RGI-defined ARG families were included in the analysis (Supplementary Table S5 and S6). Linear mixed-effects models (LMM) were used as statistical test for the analysis, including fixed effects for clade type, gene class, and their interaction, with random effects for species, ARG family, and sister-clade comparisons to capture lineage-specific variation (see Methods).

Our analysis revealed that bacterial clades with cARGs exhibit significantly

226 different plasmid gene gain rates compared with their sister clades (Figure 4B-C).
227 ARG-containing clades showed approximately 10-fold higher gain rates for pARGs
228 (estimate = 2.30 on the log scale, $p = 1.9\times10^{-7}$; Figure 4B, Supplementary Table S7)
229 and modestly but significantly lower loss rates ($p = 0.034$). In contrast, other plasmid
230 genes showed significantly lower gain rates in ARG-containing clades compared with
231 sister clades (about threefold lower; estimate = −1.12, $p = 0.0069$), while loss rates
232 showed no significant differences between clade types ($p = 0.30$). The interaction term
233 indicated that the correlation between cARG presence and evolutionary rates differs
234 significantly between gene types for gain events ($p = 1.7 \times 10^{-8}$) but not for loss events
235 ($p = 0.418$, Figure 4C, Supplementary Table S7).

236 In addition to the overall association, we performed per-family LMM analyses for
237 those with at least three available clades in Enterobacteriaceae (Supplementary Table
238 S8), the overall gene gain rate remained higher in cARG clades (Supplementary Figure
239 S3; $p < 0.05$). For each ARG family, we found significantly higher gene-gain rates on
240 the ARG clade relative to its sister clade for ANT(3''), chloramphenicol
241 acetyltransferase (CAT), fluoroquinolone-resistant *gyrB*, sulfonamide-resistant *sul*, and
242 trimethoprim-resistant dihydrofolate reductase (*dfr*) (Supplementary Figure S4). We
243 further summarized the species distribution and lineage-level genome counts for these
244 five ARG families (Figure 5), suggesting that in these species, ARGs tend to be more
245 frequently acquired on plasmids over evolutionary timescales. In contrast, several
246 families showed lower gene-loss rates in the ARG-containing clade, including APH(3'),
247 OXA-10-like β-lactamase, and SHV β-lactamase (Supplementary Figure S4). These
248 results suggest that elevated gene gain rates are not consistent across all ARG families.

249 These results indicate that bacterial lineages harboring cARGs exhibit enhanced

acquisition of plasmid-borne resistance, with preferential gain of pARGs but not other plasmid genes in cARG-containing clades.

DISCUSSION

Across 6,895 Enterobacteriaceae genomes, we show that plasmid-borne antibiotic resistance genes (pARGs) and other plasmid genes exhibit similar species-level gain and loss rates (Figure 1A–B). This parallels our co-transfer analyses, in which pARG gain and loss dynamics were significantly correlated with those of diverse plasmid functions, including defense, metabolic, and mobilome modules (Figure 3A-B). Moreover, we observed 29 species with significant correlations between pARGs and overall plasmid gene dynamics (Supplementary Table S4). Together, these findings suggest that pARGs are integrated into broader plasmid evolutionary processes, with plasmids acting as units of turnover and selection rather than ARGs following distinct evolutionary rules. This observation supports the view that plasmid genomes evolve as integrated genetic modules rather than collections of independently evolving genes (Puigbò, Lobkovsky et al. 2014), and suggesting monitoring and controlling resistance may require targeting plasmid populations as whole evolutionary units.

Importantly, our observed gain bias in pARGs (Figure 1E-F) contrasts with the loss bias dynamics that characterize bacterial genomes and other mobile genetic elements more generally (Puigbò, Lobkovsky et al. 2014, Puigbo, Makarova et al. 2017). One possible explanation is that lineage-specific features of Enterobacteriaceae favor acquisition, consistent with recent evidence that mobile element dynamics are strongly species dependent (Liu, Botelho et al. 2025). Moreover, plasmid turnover may not conform to the general loss-biased evolutionary rule governing bacterial genome

274 dynamics.

275 In contrast to gene family gain and loss, pARGs exhibited significantly elevated
276 expansion and reduction rates compared to other plasmid genes (Figure 1C-D). These
277 copy-number dynamics were highly variable across species, and statistical tests
278 showed that species identity, rather than resistance class, explained most of the
279 observed variance (Figure 2A-B, Supplementary Table S3). Weak drug class signals
280 were detectable for expansion and gain, but effect sizes were negligible
281 (Supplementary Table S3). Elevated copy-number turnover may reflect the activity of
282 gene-level mobile elements such as IS26 translocatable units (He, Hickman et al. 2015,
283 Harmer and Hall 2016) and integron cassette rearrangements (Partridge, Tsafnat et al.
284 2009, Souque, Escudero et al. 2021), as well as selective amplification events that
285 duplicate or individual genes without affecting entire plasmids (Nicoloff, Hjort et al.
286 2024). Besides, recent study revealed that 77% of studied ARG-carrying plasmids
287 contain insertion sequences, facilitating inter-plasmid resistance gene transfer(Coluzzi
288 and Rocha 2025). And the replication-control mutations (e.g., *copA*) may also elevate
289 plasmid copy number and increase ARG transfer rates (Dimitriu, Matthews et al. 2021)

290 Beyond species-specific variation in copy-number turnover, which likely reflects
291 differing balances between antibiotic-driven selection and fitness costs(Vogwill and
292 MacLean 2015, Pal and Andersson 2024), we observed that bacterial lineages
293 harboring cARGs acquired pARGs at ~10-fold higher rates than their sister clades
294 (Figure 4B, Supplementary Table S7). This pattern mirrors clinical observations of high-
295 risk clones that accumulate resistance genes across both chromosomal and plasmid
296 compartments (Mathers, Peirano et al. 2015, Rozwandowicz, Brouwer et al. 2018).
297 Rather than implying causation, the association suggests that ARG accumulation is a

298 lineage-level property: clades predisposed to high plasmid turnover are more likely to
299 fix ARGs in their chromosomes, functioning as 'super-accumulators' of resistance
300 (Woodford, Turton et al. 2011, Wang and Dagan 2024, Zongo, Cabanel et al. 2024).
301 From a clinical perspective, cARGs may thus serve as genomic markers for identifying
302 bacterial lineages with elevated capacity for plasmid-mediated resistance acquisition,
303 underscoring the need for enhanced surveillance and infection control targeting these
304 high-risk lineages.

305 We note that the cARG families showing elevated acquisition of pARGs in our
306 analyses (e.g., ANT(3"), CAT, *sul*, and *dfr*) are consistently highlighted across lineages
307 (Figure 5). These families correspond to antibiotic classes that remain widely applied
308 or frequently detected in clinical and agricultural settings (Manyi-Loh, Mamphweli et al.
309 2018, Bacanli and Basaran 2019, Dominguez, Miranda et al. 2019). For example,
310 *Providencia rettgeri* and *Citrobacter spp*. have been reported in agricultural and
311 zoonotic contexts to carry resistance to aminoglycosides and sulfonamides (Liu, Zhang
312 et al. 2021, Ahmed, Islam et al. 2023, Wang, Li et al. 2023, Valiatti, Santos et al. 2024).
313 Such lineages are likely exposed to overlapping selective pressures from multiple
314 antibiotic classes, which may accelerate pARG turnover, promote their integration into
315 chromosomes, and in turn facilitate the acquisition of additional pARGs. Thus, the
316 combined action of chromosomal and plasmid-borne ARGs may drive the rapid
317 accumulation of multidrug resistance within these taxa.

318 **Limitations**

319 This study focused exclusively on plasmid-associated genes, excluding other
320 mobile elements such as Integrative and conjugative elements (ICEs) and integrative

321 mobilizable elements (IMEs), which also mediate chromosomal integration and
322 excision of ARGs (Johnson and Grossman 2015, Cury, Touchon et al. 2017, Botelho
323 2023). Bacteriophages further can contribute to ARG dissemination through
324 transduction and phage–plasmid hybrids (Touchon, Moura de Sousa et al. 2017).
325 Consequently, our observations represent only a subset of the overall resistance
326 mobilome. Moreover, we lacked ecological information about the habitats and antibiotic
327 exposure histories of the analyzed organisms. Expanding future analyses to include
328 additional mobile genetic element and ecological datasets would provide a more
329 comprehensive understanding of ARG dissemination in natural environments.

330 **Conclusion**

331 Our findings highlight a duality in plasmid-mediated resistance evolution in
332 Enterobacteriaceae. At the level of whole gene families, pARGs follow the same
333 turnover dynamics as do other plasmid genes, indicating that resistance dissemination
334 largely reflects plasmid-level processes rather than ARG-specific rules. At the level of
335 copy-number changes, however, pARGs show elevated expansion and reduction with
336 strong lineage dependence and negligible drug-class effects. Such species-driven
337 volatility suggests recurrent expansion–reduction cycles as a flexible resistance
338 strategy. Moreover, chromosomal resistance content predicts a lineage's propensity
339 for further plasmid-mediated resistance acquisition, positioning cARGs as potential
340 markers of high-risk lineages.

341

342 **METHODS**

343 **Genome Collection and Gene Annotation**

We parsed the Genome Taxonomy Database (GTDB; https://gtdb.ecogenomic.org) release 220 (Parks, Chuvochina et al. 2020) to identify genomes belonging to Enterobacteriaceae. For species with more than 300 genomes, we randomly selected 300 representatives with random seed 19940421. We applied MIMAG criteria (Bowers, Kyrpides et al. 2017) with completeness >99%, contamination <1%, mean contig length >5,000 bp, and contig count <500 to filter high-quality genomes. To reduce computational cost, species with more than 100 genomes were subsampled to retain at most 100 genomes (random seed 19940421). Only species with more than 10 genomes were considered for further analysis. The genomes that passed these filters were downloaded from the NCBI FTP site (https://ftp.ncbi.nlm.nih.gov).

Bacterial genes were predicted using pyrodigal-gv v0.3.2 integrated in geNomad. Plasmids and their encoded genes were identified using geNomad v1.11.0 with database v1.9 and default parameters (Camargo, Roux et al. 2024). Antibiotic resistance genes were identified with Resistance Gene Identifier (RGI) v6.0.5 using protein sequences with "include_nudge" option and default parameters, CARD version 4.0.1 (Alcock, Huynh et al. 2023). Pan-genomes of each species were constructed using the orthology-based clustering method OrthoFinder v2.5.5 (Emms and Kelly 2019) following established guidelines (Manzano-Morales, Liu et al. 2023).

For COG functional annotation, we constructed HMM profiles from the COG 2024 database accessed on 2025-02-20 (Galperin, Vera Alvarez et al. 2025). Protein sequences from each COG family were grouped and aligned using MAFFT v7.526 (Katoh and Standley 2013) with --auto for algorithm selection. Multiple sequence alignments were trimmed using ClipKIT v2.6.1 (Steenwyk, Buida et al. 2020) with smart-gap mode to remove poorly aligned regions while preserving phylogenetic signal.

368 HMM profiles were built for each COG family using HMMER v3.4 (http://hmmer.org)
369 hmmbuild with default parameters. Predicted proteins from each genome were
370 searched against the COG HMM database using hmmsearch with an E-value cutoff of
371 1e-3.

372 After applying these criteria, 6,895 complete or nearly complete genomes
373 belonging to 159 Enterobacteriaceae species (sensu GTDB) were included in the
374 analysis (Supplementary Table S1), of which 98 contained antibiotic resistant genes
375 on plasmids (Supplementary Data).

376

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

378 From the plasmid gene annotations obtained through geNomad and the orthologous
379 gene families and species trees generated by OrthoFinder for each species, we
380 extracted plasmid orthologous gene families. An orthologous gene family was
381 considered plasmid-associated if the proportion of plasmid genes within that
382 orthologous gene family exceeded 70% of all genes in the family. Species trees were
383 pruned to retain only genomes contributing to the filtered orthologous gene families.

384 Gene gain and loss dynamics were quantified using the phylogenetic birth-death
385 model implemented in Count v10.04 (Csuros 2010). This probabilistic framework
386 models the evolution of gene family sizes along phylogenetic trees as a continuous-
387 time Markov process. The model accounts for three fundamental processes: gene gain
388 (κ), gene loss (μ), and gene duplication (λ). A gene family of size n changes according
389 to the following rates: the family decreases at rate $n\mu$ due to individual gene losses,
390 increases through duplication at rate $n\lambda$, and gains new genes at rate κ through

horizontal transfer or other acquisition mechanisms.

For each species dataset, we performed iterative maximum likelihood optimization to estimate birth-death model parameters (Puigbò, Lobkovsky et al. 2014). 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 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

Antibiotic resistance gene families were identified using a guilt-by-association approach. If any gene within a plasmid-associated orthologous gene family matched CARD reference sequence through RGI analysis, the entire orthologous gene family was classified as ARG gene families. Branch lengths were extracted from the original species trees and combined with corresponding Count posterior probabilities. For rate

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

419 $$rate_i = \frac{\sum(events_{\{i,j\}})}{\sum(branch\_length_j)}$$

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

422

423 **Comparative Analysis of Evolutionary Rates**

424 For comparative analysis between ARG and non-ARG plasmid gene families, data
425 were aggregated by species, event type, and ARG status. Median rates were
426 calculated within each group, and the data were restructured into a paired format
427 where each species served as a matched pair with corresponding median rates for
428 ARG and non-ARG families. Paired Wilcoxon signed-rank tests compared ARG versus
429 non-ARG families within species to control for species-specific evolutionary factors.
430 Additionally, gain/loss and expansion/reduction ratios were calculated per species to
431 assess the relative balance of these complementary processes. Multiple testing
432 correction was performed using the Benjamini-Hochberg procedure. Effect sizes were
433 quantified as median log2-fold changes, and statistical significance was assessed at p
434 < 0.05 and q < 0.05 (FDR-adjusted p).

435 To examine rate variation across antibiotic resistance classes, drug class
436 annotations were extracted from RGI output. Each ARG-containing orthologous gene

437 family was annotated to its corresponding drug classes. If an orthologous gene family
438 matched multiple antibiotic classes, we count it in each matching class. Evolutionary
439 rates were then visualized across different species and drug class within
440 Enterobacteriaceae. To assess concordance between event types across species, we
441 compared median per-species rates using Spearman's rank correlation with
442 permutation tests (5,000 iterations, random seed = 19940421) to obtain empirical p-
443 values.

444

445 **Identification of Genes Co-transferred with Antibiotic Resistance Genes**

446 To identify genes potentially co-transferred with ARGs through horizontal gene transfer,
447 we assessed whether ARG and non-ARG plasmid genes exhibited similar evolutionary
448 patterns across phylogenetic branches within each species.

449 For data preparation, branch-specific evolutionary events from Count analysis
450 were merged with ARG annotations from RGI and COG functional annotations. In
451 particular, gene families were classified as ARG-containing based on RGI matches,
452 and all plasmid gene families were annotated with their consensus COG assignments
453 (requiring ≥75% consistency within each orthologous gene family). Only COG families
454 with at least 5 genes per species were retained for analysis. Event rates were
455 calculated by summing gain and loss events for each gene family at each phylogenetic
456 branch (both tips and internal nodes). ARG event rates were aggregated across all
457 resistance genes, while rates of non-ARG plasmid genes were calculated both globally
458 and per individual COG family.

459 We then compared the branch-specific event rates between ARGs and individual

460 COG-annotated gene families using phylogenetic Brownian motion (BM) simulations
461 to test for correlations while accounting for phylogenetic structure. The analysis used
462 Spearman correlation on both tip and internal node data, with BM simulations providing
463 null distributions for significance testing. Initial screening was performed with a
464 minimum of 5 data points per comparison and 100 simulations. Gene families showing
465 significant correlations (adjusted $p < 0.05$ after Benjamini-Hochberg correction) were
466 subjected to progressively stringent re-analyses: first with a minimum of 8 data points
467 and 200 simulations, then with 10 data points and 500 simulations. Only gene families
468 maintaining significance through all three rounds were considered as potential co-
469 transfer candidates. Species trees were pre-processed using midpoint rooting, with
470 zero-length branches adjusted by adding small random values (1e-13 to 5e-13) to
471 avoid artifacts. Phylogenetic Brownian-motion traits were simulated using fastBM, and
472 ancestral states were reconstructed using fastAnc, both in the phytools package v2.4.4
473 (Revell 2024) for R v4.3.0.

474 The identified co-transfer candidate genes were then categorized according to
475 COG functional categories. Individual COG categories were mapped to four major
476 process groups: information storage and processing, cellular processes and signaling,
477 metabolism, and poorly characterized functions.

478

479 **Comparison of Plasmid Gene Dynamics Between ARG-containing and Sister**
480 **Clades**

481 To assess whether the presence of antibiotic resistance genes in bacterial
482 chromosomes influences the evolutionary dynamics of plasmid-borne genes, we

483 identified pairs of phylogenetic clades that differed in the chromosomal presence of an
484 ARG family. For each combination of species tree and ARG family, we scanned for
485 internal nodes where ≥80% of descendant genomes harbored the specific ARG family
486 in their chromosome (ARG-containing clade), while <20% of genomes in the
487 corresponding sister clade contained that ARG family. Only clade pairs with at least
488 two genomes in each clade were retained for analysis.

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

492 $$Rate_{OG,\,clade} = \frac{\sum_{b \in clade} Events_{OG,\,b}}{\sum_{b \in clade} Branch\ Length_b}$$

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

498 Event rates were analyzed using linear mixed-effects models on the log scale, with
499 gain and loss modeled separately; the response was $\log(rate + \varepsilon)$, where $\varepsilon =$
500 $0.1 \times \min(\text{non-zero rate})$ within the corresponding event type. Fixed effects were clade
501 type (ARG-containing vs sister clade), gene class (plasmid ARGs vs other plasmid
502 genes), and their interaction; random effects were species, resistance gene family, and
503 the clade pair linking each ARG-containing clade with its sister (implemented as a
504 random intercept on Clade ID). Models were fitted using restricted maximum likelihood
505 and restricted to paired observations with both clade types available within a given

gene class. Post hoc contrasts from estimated marginal means assessed (i) ARG-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 (Bates, Mächler et al. 2015), lmerTest v3.1.3 (Kuznetsova, Brockhoff et al. 2017), and emmeans v 1.11.2.8.

ACKNOWLEDGEMENTS

This work was funded by the ZJU100 Young Professor Start-up Fund from Zhejiang University of China (to Yue Liu).

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

During the preparation of this work, the author(s) used ChatGPT 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.

DECLARATION OF COMPETING INTERESTS

The authors declare no competing financial interests.

ETHICS APPROVAL

There are no ethical concerns in this study.

DATA AVAILABILITY

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

AUTHOR CONTRIBUTIONS

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

References

Ahmed, T., M. S. Islam, N. Haider, L. Elton, B. Hasan, M. Nuruzzaman, M. T. Rahman, S. M. L. Kabir and M. S. R. Khan (2023). "Phenotypic and Genotypic Characteristics of Antimicrobial Resistance in Citrobacter freundii Isolated from Domestic Ducks (Anas platyrhynchos domesticus) in Bangladesh." Antibiotics (Basel) **12**(4).

Alcock, B. P., W. Huynh, R. Chalil, K. W. Smith, A. R. Raphenya, M. A. Wlodarski, A. Edalatmand, A. Petkau, S. A. Syed, K. K. Tsang, S. J. C. Baker, M. Dave, M. C. McCarthy, K. M. Mukiri, J. A. Nasir, B. Golbon, H. Imtiaz, X. J. Jiang, K. Kaur, M. Kwong, Z. C. Liang, K. C. Niu, P. Shan, J. Y. J. Yang, K. L. Gray, G. R. Hoad, B. F. Jia, T. Bhando, L. A. Carfrae, M. A. Farha, S. French, R. Gordzevich, K. Rachwalski, M. M. Tu, E. Bordeleau, D. Dooley, E. Griffiths, H. L. Zubyk, E. D. Brown, F. Maguire, R. G. Beiko, W. W. L. Hsiao, F. S. L. Brinkman, G. Van Domselaar and A. G. McArthur (2023). "CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database." Nucleic Acids Research **51**(D1): D690-D699.

Amos, G. C. A., S. Ploumakis, L. Zhang, P. M. Hawkey, W. H. Gaze and E. M. H.

558 Wellington (2018). "The widespread dissemination of integrons throughout bacterial
559 communities in a riverine system." ISME J **12**(3): 681-691.

560 Bacanli, M. and N. Basaran (2019). "Importance of antibiotic residues in animal food."
561 Food Chem Toxicol **125**: 462-466.

562 Bates, D., M. Mächler, B. M. Bolker and S. C. Walker (2015). "Fitting Linear Mixed-
563 Effects Models Using lme4." Journal of Statistical Software **67**(1): 1-48.

564 Bennett, P. M. (2008). "Plasmid encoded antibiotic resistance: acquisition and transfer
565 of antibiotic resistance genes in bacteria." Br J Pharmacol **153 Suppl 1**(Suppl 1): S347-
566 357.

567 Botelho, J. (2023). "Defense systems are pervasive across chromosomally integrated
568 mobile genetic elements and are inversely correlated to virulence and antimicrobial
569 resistance." Nucleic Acids Res **51**(9): 4385-4397.

570 Bowers, R. M., N. C. Kyrpides, R. Stepanauskas, M. Harmon-Smith, D. Doud, T. B. K.
571 Reddy, F. Schulz, J. Jarett, A. R. Rivers, E. A. Eloe-Fadrosh, S. G. Tringe, N. N. Ivanova,
572 A. Copeland, A. Clum, E. D. Becraft, R. R. Malmstrom, B. Birren, M. Podar, P. Bork, G.
573 M. Weinstock, G. M. Garrity, J. A. Dodsworth, S. Yooseph, G. Sutton, F. O. Glöckner,
574 J. A. Gilbert, W. C. Nelson, S. J. Hallam, S. P. Jungbluth, T. J. G. Ettema, S. Tighe, K.
575 T. Konstantinidis, W. T. Liu, B. J. Baker, T. Rattei, J. A. Eisen, B. Hedlund, K. D.
576 McMahon, N. Fierer, R. Knight, R. Finn, G. Cochrane, I. Karsch-Mizrachi, G. W. Tyson,
577 C. Rinke, A. Lapidus, F. Meyer, P. Yilmaz, D. H. Parks, A. M. Eren, L. Schriml, J. F.
578 Banfield, P. Hugenholtz, T. Woyke and G. S. Consortium (2017). "Minimum information
579 about a single amplified genome (MISAG) and a metagenome-assembled genome
580 (MIMAG) of bacteria and archaea." Nature Biotechnology **35**(8): 725-731.

581 Camargo, A. P., S. Roux, F. Schulz, M. Babinski, Y. Xu, B. Hu, P. S. G. Chain, S.
582 Nayfach and N. C. Kyrpides (2024). "Identification of mobile genetic elements with
583 geNomad." Nat Biotechnol **42**(8): 1303-1312.

584 Carattoli, A. (2013). "Plasmids and the spread of resistance." Int J Med Microbiol
585 **303**(6-7): 298-304.

586 Che, Y., Y. Yang, X. Xu, K. Brinda, M. F. Polz, W. P. Hanage and T. Zhang (2021).
587 "Conjugative plasmids interact with insertion sequences to shape the horizontal
588 transfer of antimicrobial resistance genes." Proc Natl Acad Sci U S A **118**(6).

589 Coluzzi, C. and E. P. C. Rocha (2025). "The Spread of Antibiotic Resistance Is Driven
590 by Plasmids Among the Fastest Evolving and of Broadest Host Range." Molecular
591 Biology and Evolution **42**(3).

592 Csuros, M. (2010). "Count: evolutionary analysis of phylogenetic profiles with
593 parsimony and likelihood." Bioinformatics **26**(15): 1910-1912.

594 Cury, J., M. Touchon and E. P. C. Rocha (2017). "Integrative and conjugative elements
595 and their hosts: composition, distribution and organization." Nucleic Acids Res **45**(15):
596 8943-8956.

597 Davies, J. and D. Davies (2010). "Origins and evolution of antibiotic resistance."
598 Microbiol Mol Biol Rev **74**(3): 417-433.

599 Dimitriu, T., A. C. Matthews and A. Buckling (2021). "Increased copy number couples
600 the evolution of plasmid horizontal transmission and plasmid-encoded antibiotic
601 resistance." Proc Natl Acad Sci U S A **118**(31).

602 Dominguez, M., C. D. Miranda, O. Fuentes, M. de la Fuente, F. A. Godoy, H. Bello-
603 Toledo and G. Gonzalez-Rocha (2019). "Occurrence of Transferable Integrons and sul
604 and dfr Genes Among Sulfonamide-and/or Trimethoprim-Resistant Bacteria Isolated
605 From Chilean Salmonid Farms." Front Microbiol **10**: 748.

606 Emms, D. M. and S. Kelly (2019). "OrthoFinder: phylogenetic orthology inference for
607 comparative genomics." Genome Biol **20**(1): 238.

608 Galperin, M. Y., R. Vera Alvarez, S. Karamycheva, K. S. Makarova, Y. I. Wolf, D.
609 Landsman and E. V. Koonin (2025). "COG database update 2024." Nucleic Acids Res
610 **53**(D1): D356-D363.

611 Grundmann, H., D. M. Livermore, C. G. Giske, R. Canton, G. M. Rossolini, J. Campos,
612 A. Vatopoulos, M. Gniadkowski, A. Toth, Y. Pfeifer, V. Jarlier, Y. Carmeli and C. W.
613 Group (2010). "Carbapenem-non-susceptible Enterobacteriaceae in Europe:
614 conclusions from a meeting of national experts." Euro Surveill **15**(46).

615 Harmer, C. J. and R. M. Hall (2016). "IS26-Mediated Formation of Transposons
616 Carrying Antibiotic Resistance Genes." mSphere **1**(2).

617 He, S., A. B. Hickman, A. M. Varani, P. Siguier, M. Chandler, J. P. Dekker and F. Dyda
618 (2015). "Insertion Sequence IS26 Reorganizes Plasmids in Clinically Isolated
619 Multidrug-Resistant Bacteria by Replicative Transposition." mBio **6**(3): e00762.

620 Johnson, C. M. and A. D. Grossman (2015). "Integrative and Conjugative Elements
621 (ICEs): What They Do and How They Work." Annu Rev Genet **49**: 577-601.

622 Katoh, K. and D. M. Standley (2013). "MAFFT multiple sequence alignment software
623 version 7: improvements in performance and usability." Mol Biol Evol **30**(4): 772-780.

624 Koonin, E. V. and Y. I. Wolf (2008). "Genomics of bacteria and archaea: the emerging
625 dynamic view of the prokaryotic world." Nucleic Acids Res **36**(21): 6688-6719.

Koonin, E. V. and Y. I. Wolf (2012). "Evolution of microbes and viruses: a paradigm shift in evolutionary biology?" Front Cell Infect Microbiol **2**: 119.

Kuznetsova, A., P. B. Brockhoff and R. H. B. Christensen (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." Journal of Statistical Software **82**(13): 1-26.

Liu, L., L. Zhang, H. Zhou, M. Yuan, D. Hu, Y. Wang, H. Sun, J. Xu and R. Lan (2021). "Antimicrobial Resistance and Molecular Characterization of Citrobacter spp. Causing Extraintestinal Infections." Front Cell Infect Microbiol **11**: 737636.

Liu, Y., J. Botelho and J. Iranzo (2025). "Timescale and genetic linkage explain the variable impact of defense systems on horizontal gene transfer." Genome Res **35**(2): 268-278.

Manyi-Loh, C., S. Mamphweli, E. Meyer and A. Okoh (2018). "Antibiotic Use in Agriculture and Its Consequential Resistance in Environmental Sources: Potential Public Health Implications." Molecules **23**(4).

Manzano-Morales, S., Y. Liu, S. Gonzalez-Bodi, J. Huerta-Cepas and J. Iranzo (2023). "Comparison of gene clustering criteria reveals intrinsic uncertainty in pangenome analyses." Genome Biol **24**(1): 250.

Mathers, A. J., G. Peirano and J. D. Pitout (2015). "The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae." Clin Microbiol Rev **28**(3): 565-591.

Nicoloff, H., K. Hjort, D. I. Andersson and H. Wang (2024). "Three concurrent mechanisms generate gene copy number variation and transient antibiotic heteroresistance." Nat Commun **15**(1): 3981.

Pal, A. and D. I. Andersson (2024). "Bacteria can compensate the fitness costs of amplified resistance genes via a bypass mechanism." Nat Commun **15**(1): 2333.

Parks, D. H., M. Chuvochina, P. A. Chaumeil, C. Rinke, A. J. Mussig and P. Hugenholtz (2020). "A complete domain-to-species taxonomy for Bacteria and Archaea." Nat Biotechnol **38**(9): 1079-1086.

Partridge, S. R., S. M. Kwong, N. Firth and S. O. Jensen (2018). "Mobile Genetic Elements Associated with Antimicrobial Resistance." Clin Microbiol Rev **31**(4).

Partridge, S. R., G. Tsafnat, E. Coiera and J. R. Iredell (2009). "Gene cassettes and cassette arrays in mobile resistance integrons." FEMS Microbiol Rev **33**(4): 757-784.

Puigbò, P., A. E. Lobkovsky, D. M. Kristensen, Y. I. Wolf and E. V. Koonin (2014). "Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes." Bmc Biology **12**.

660 Puigbo, P., K. S. Makarova, D. M. Kristensen, Y. I. Wolf and E. V. Koonin (2017).
661 "Reconstruction of the evolution of microbial defense systems." BMC Evol Biol **17**(1):
662 94.

663 Redondo-Salvo, S., R. Fernández-López, R. Ruiz, L. Vielva, M. de Toro, E. P. C. Rocha,
664 M. P. Garcillán-Barcia and F. de la Cruz (2020). "Pathways for horizontal gene transfer
665 in bacteria revealed by a global map of their plasmids." Nature Communications **11**(1).

666 Revell, L. J. (2024). "phytools 2.0: an updated R ecosystem for phylogenetic
667 comparative methods (and other things)." Peerj **12**.

668 Rozwandowicz, M., M. S. M. Brouwer, J. Fischer, J. A. Wagenaar, B. Gonzalez-Zorn,
669 B. Guerra, D. J. Mevius and J. Hordijk (2018). "Plasmids carrying antimicrobial
670 resistance genes in Enterobacteriaceae." J Antimicrob Chemother **73**(5): 1121-1137.

671 San Millan, A. (2018). "Evolution of Plasmid-Mediated Antibiotic Resistance in the
672 Clinical Context." Trends Microbiol **26**(12): 978-985.

673 San Millan, A. and R. C. MacLean (2017). "Fitness Costs of Plasmids: a Limit to
674 Plasmid Transmission." Microbiol Spectr **5**(5).

675 Souque, C., J. A. Escudero and R. C. MacLean (2021). "Integron activity accelerates
676 the evolution of antibiotic resistance." Elife **10**.

677 Steenwyk, J. L., T. J. Buida, 3rd, Y. Li, X. X. Shen and A. Rokas (2020). "ClipKIT: A
678 multiple sequence alignment trimming software for accurate phylogenomic inference."
679 PLoS Biol **18**(12): e3001007.

680 Touchon, M., J. A. Moura de Sousa and E. P. Rocha (2017). "Embracing the enemy:
681 the diversification of microbial gene repertoires by phage-mediated horizontal gene
682 transfer." Curr Opin Microbiol **38**: 66-73.

683 Valiatti, T. B., F. F. Santos, F. O. Bessa-Neto, R. Veiga, S. Simionatto, G. H. de Almeida
684 Souza, M. S. M. Vaz, A. C. C. Pignatari, R. Cayo, A. C. Gales and N. Guarani (2024).
685 "Emergence of multidrug-resistant Providencia rettgeri clone in food-producing
686 animals: A public health threat." One Health **19**: 100887.

687 van Duin, D. and D. L. Paterson (2016). "Multidrug-Resistant Bacteria in the
688 Community: Trends and Lessons Learned." Infect Dis Clin North Am **30**(2): 377-390.

689 Vogwill, T. and R. C. MacLean (2015). "The genetic basis of the fitness costs of
690 antimicrobial resistance: a meta-analysis approach." Evol Appl **8**(3): 284-295.

691 Wang, P., C. Li, Z. Yin, X. Jiang, X. Li, X. Mu, N. Wu, F. Chen and D. Zhou (2023).
692 "Genomic epidemiology and heterogeneity of Providencia and their bla(NDM-1)-
693 carrying plasmids." Emerg Microbes Infect **12**(2): 2275596.

694 Wang, Y. and T. Dagan (2024). "The evolution of antibiotic resistance islands occurs
695 within the framework of plasmid lineages." Nat Commun **15**(1): 4555.
696 Woodford, N., J. F. Turton and D. M. Livermore (2011). "Multiresistant Gram-negative
697 bacteria: the role of high-risk clones in the dissemination of antibiotic resistance."
698 FEMS Microbiol Rev **35**(5): 736-755.
699 Zongo, P. D., N. Cabanel, G. Royer, F. Depardieu, A. Hartmann, T. Naas, P. Glaser and
700 I. Rosinski-Chupin (2024). "An antiplasmid system drives antibiotic resistance gene
701 integration in carbapenemase-producing Escherichia coli lineages." Nat Commun
702 **15**(1): 4093.
703

704 **Figure legends**

705 **Figure 1. Evolutionary dynamics of plasmid-borne antibiotic resistance genes**
706 **compared to other plasmid genes in Enterobacteriaceae.**

707 (A) Gain rates (appearance of new gene families per substitution per site) for ARGs
708 versus other plasmid genes. (B) Loss rates (disappearance of gene families per
709 substitution per site) for ARGs versus other plasmid genes. (C) Expansion rates
710 (addition of gene copies within existing families per substitution per site) for ARGs
711 versus other plasmid genes. (D) Reduction rates (loss of gene copies within existing
712 families per substitution per site) for ARGs versus other plasmid genes. (E) Log_2 ratio
713 of gain to loss rates for ARGs. (F) Log_2 ratio of gain to loss rates for other plasmid
714 genes. Each point represents one Enterobacteriaceae species. Box plots show median
715 and quartiles. Gene families were defined as orthologous groups (OGs).

716

717 **Figure 2. Expansion and reduction rate patterns across antibiotic resistance**
718 **classes and Enterobacteriaceae species.**

719 Heatmaps showing evolutionary rates for different antibiotic resistance classes across

720 bacterial species. (A) Expansion rates. (B) Reduction rates. Rows represent
721 Enterobacteriaceae species, and columns represent CARD-defined antibiotic
722 resistance classes. Color intensity indicates log10 scaled rate magnitude, with darker
723 colors representing higher rates. White cells indicate absence of the ARG class in the
724 species. Color scales are truncated for visualization clarity: gain rates (20th-80th
725 percentile), loss/expansion/reduction rates (5th-95th percentile); all statistical analyses
726 were performed on complete untruncated datasets.

727

728 **Figure 3. Correlation between ARGs and other plasmid functional genes on**
729 **evolutionary timescales.**

730 Sankey diagram showing the size of significant COG gene families that correlated with
731 ARGs across phylogenetic tree. Flow width represents the number of unique COG
732 families within each functional category that show significant correlation with antibiotic
733 resistance genes. (B) Bar plot showing the number of species have at least a significant
734 correlation for each COG functional category. X-axis represents the count of species
735 showing significant case between ARGs and genes from each COG category. The
736 triangle represents the median branches number involved in the statistical test for each
737 category. The border color of the bar indicates whether both gain and loss are
738 significant or only one of the tests is significant. Abbreviations of functional categories,
739 A: RNA processing and modification; B: Chromatin structure and dynamics; C: energy
740 production and conversion; D: cell cycle control, cell division, chromosome partitioning;
741 E: amino acid transport and metabolism; F: nucleotide transport and metabolism; G:
742 carbohydrate transport and metabolism; H: coenzyme transport and metabolism; I:
743 lipid transport and metabolism; J: translation, ribosomal structure and biogenesis; K:

transcription; L: replication, recombination and repair; M: cell wall/membrane/envelope biogenesis; N: cell motility; O: posttranslational modification, protein turnover, chaperones; P: inorganic ion transport and metabolism; Q: secondary metabolites biosynthesis, transport and catabolism; R: general function prediction only; S: function unknown; T: signal transduction mechanisms; U: intracellular trafficking, secretion, and vesicular transport; V: defense mechanisms; W: extracellular structures; X: mobilome: prophages, transposons; Y: nuclear structure; Z: cytoskeleton.

Figure 4. Association between presence of cARGs in bacterial lineages and plasmid borne ARG acquisition.

(A) Schematic showing sister clade comparison approach. Phylogenetic clades containing high frequencies (≥80%) of specific cARG families are compared to sister clades with low frequencies (<20%). Colored circles represent different bacterial clades. (B) Fold-change comparison of evolutionary rates between ARG-containing clades and sister clades. Y-axis shows the fold-change ratio (ARG clade / Sister clade) for gain (purple) and loss (orange) rates of plasmid ARGs versus other plasmid genes. Dashed line at y=1 indicates no difference between clades. (C) Interaction effect showing the differential impact of cARG presence on gain versus loss dynamics. Y-axis represents the interaction effect (fold-change, log scale) between clade type and evolutionary process. Error bars show 95% confidence intervals. Statistical significance indicators: ***$p < 0.001$, *$p < 0.05$, ns = not significant.

767 **Figure 5. Species distribution and lineage-level genome counts for ARG families**

768 **showing significantly higher plasmid gene-gain rates.**

769 The y-axis indicates the antibiotic classes conferring resistance by the corresponding

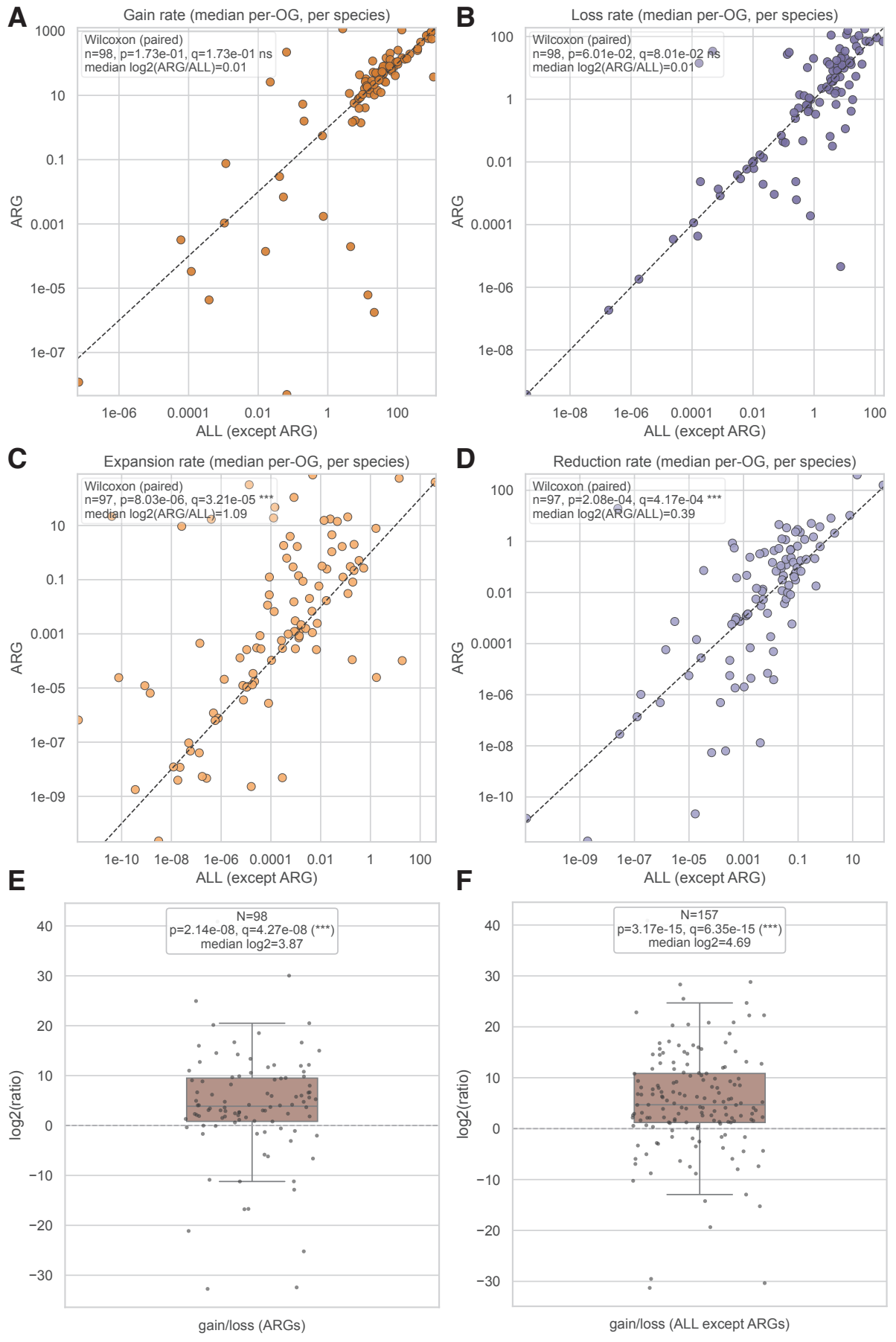
770 gene families: aminoglycosides (ANT(3")), phenicols (CAT), fluoroquinolones (gyrB),

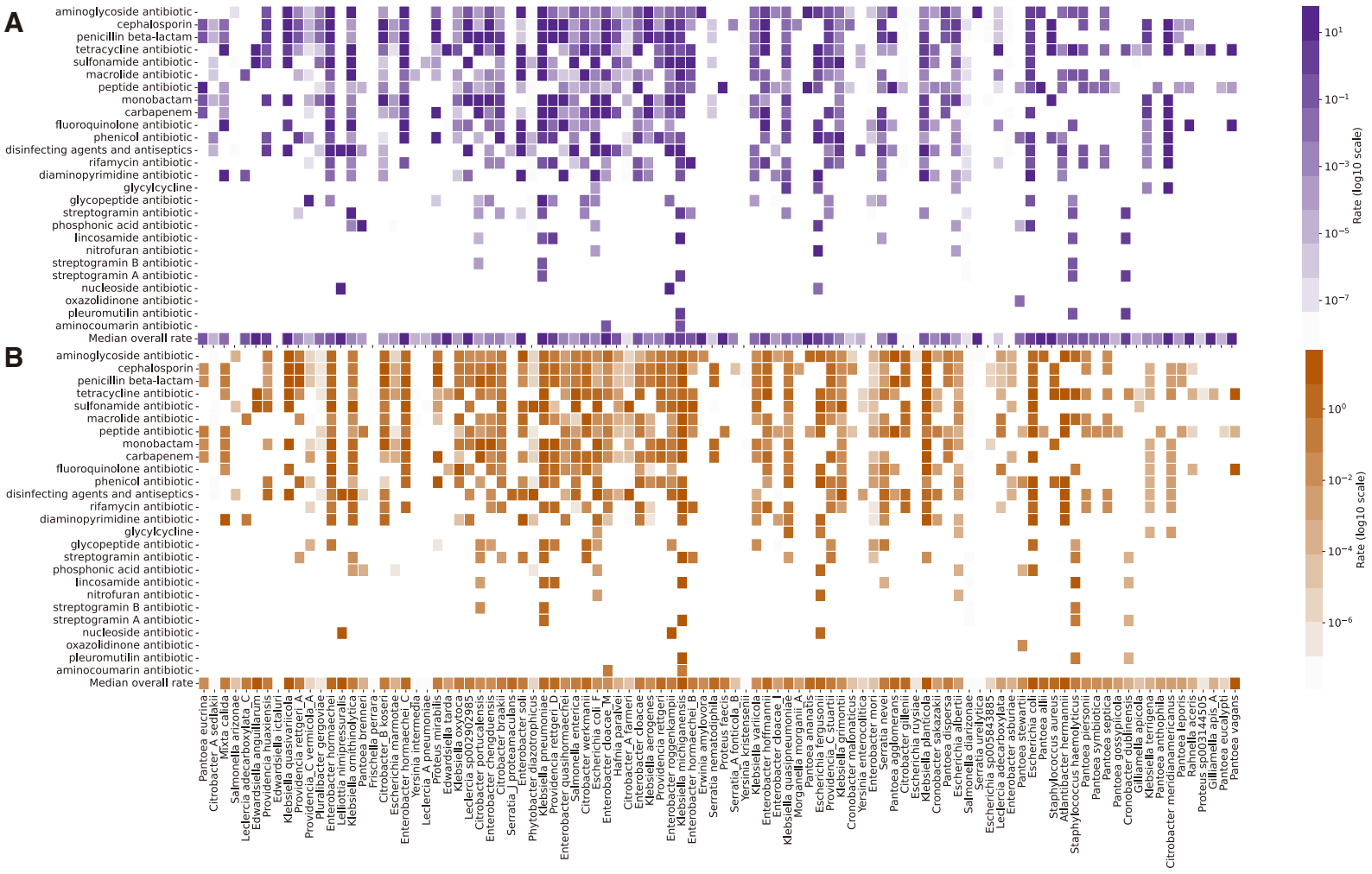
771 sulfonamides (sul), and diaminopyrimidines (dfr). The x-axis represents species as

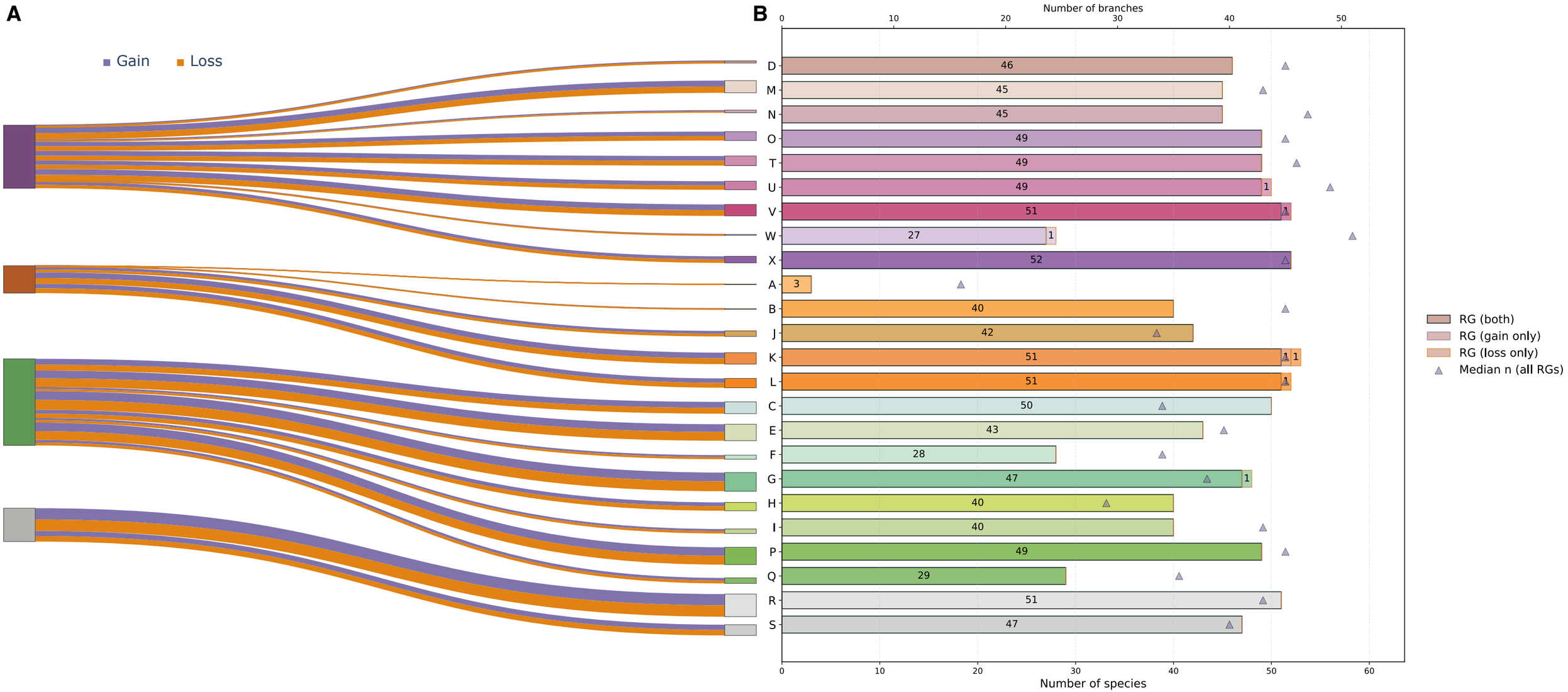
772 defined by the Genome Taxonomy Database (GTDB).

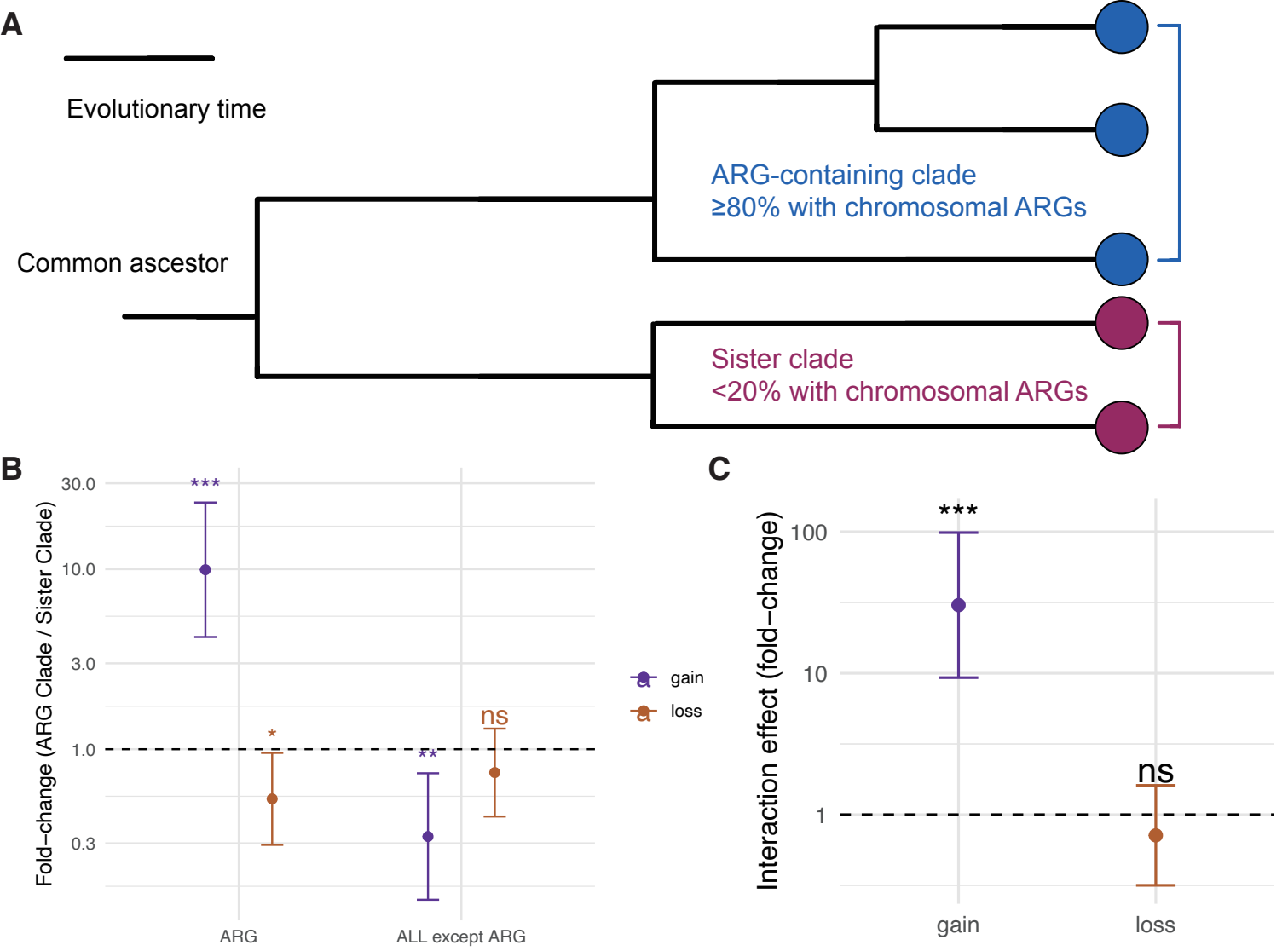
773

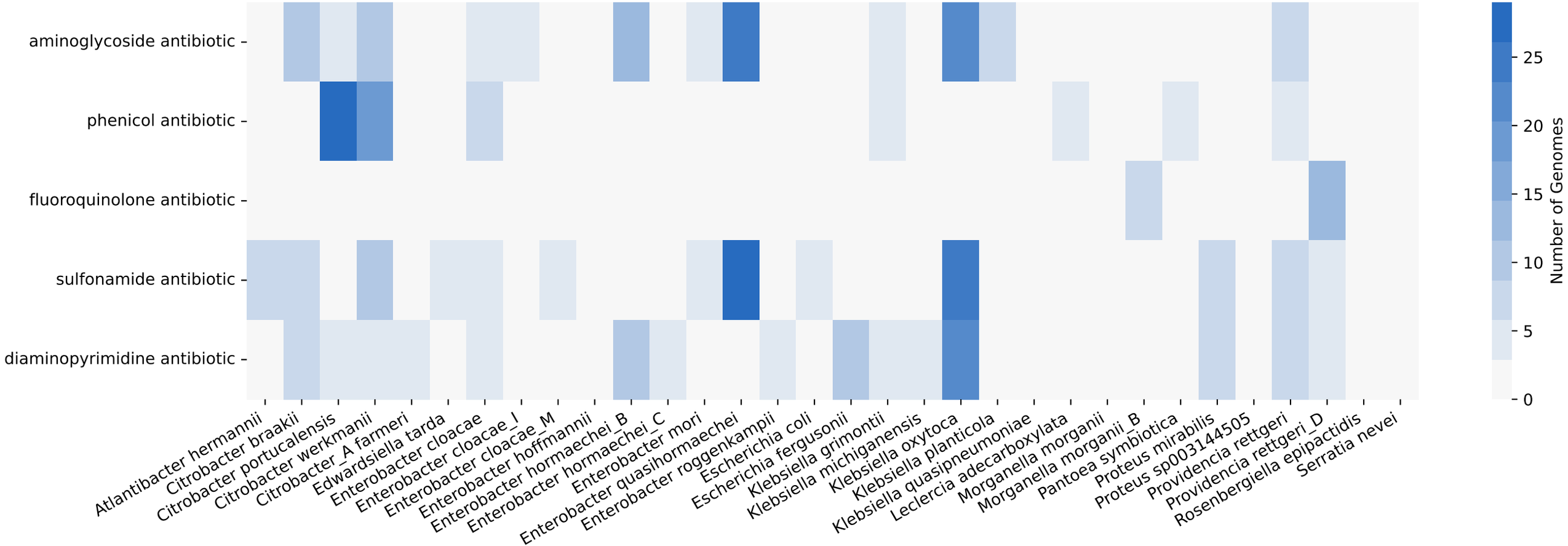
774


A
Gain rate (median per-OG, per species)
Wilcoxon (paired)
n=98, p=1.73e-01, q=1.73e-01 ns
median log2(ARG/ALL)=0.01
ARG
ALL (except ARG)
B
Loss rate (median per-OG, per species)
Wilcoxon (paired)
n=98, p=6.01e-02, q=8.01e-02 ns
median log2(ARG/ALL)=0.01
ARG
ALL (except ARG)
C
Expansion rate (median per-OG, per species)
Wilcoxon (paired)
n=97, p=8.03e-06, q=3.21e-05 ***
median log2(ARG/ALL)=1.09
ARG
ALL (except ARG)
D
Reduction rate (median per-OG, per species)
Wilcoxon (paired)
n=97, p=2.08e-04, q=4.17e-04 ***
median log2(ARG/ALL)=0.39
ARG
ALL (except ARG)
E
N=98
p=2.14e-08, q=4.27e-08 (***)
median log2=3.87
log2(ratio)
gain/loss (ARGs)
F
N=157
p=3.17e-15, q=6.35e-15 (***)
median log2=4.69
log2(ratio)
gain/loss (ALL except ARGs)


A
B
aminoglycoside antibiotic
cephalosporin
penicillin beta-lactam
tetracycline antibiotic
sulfonamide antibiotic
macrolide antibiotic
peptide antibiotic
monobactam
carbapenem
fluoroquinolone antibiotic
phenicol antibiotic
disinfecting agents and antiseptics
rifamycin antibiotic
diaminopyrimidine antibiotic
glycylcycline
glycopeptide antibiotic
streptogramin antibiotic
phosphonic acid antibiotic
lincosamide antibiotic
nitrofuran antibiotic
streptogramin B antibiotic
streptogramin A antibiotic
nucleoside antibiotic
oxazolidinone antibiotic
pleuromutilin antibiotic
aminocoumarin antibiotic
Median overall rate
Rate (log10 scale)
10^{1}
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{0}
10^{-2}
10^{-4}
10^{-6}
Pantoea eucrina
Citrobacter_A sedlakii
Mixta calida
Salmonella arizonae
Leclercia adecarboxylata_C
Edwardsiella anguillarum
Providencia huaxiensis
Edwardsiella ictaluri
Klebsiella quasivariicola
Providencia rettgeri_A
Providencia_C vermicola_A
Pluralibacter gergoviae
Enterobacter hormaechei
Lelliottia nimipressuralis
Klebsiella ornithinolytica
Pantoea brenneri
Frischella perrara
Citrobacter_B koseri
Escherichia marmotae
Enterobacter hormaechei_C
Yersinia intermedia
Leclercia_A pneumoniae
Proteus mirabilis
Edwardsiella tarda
Klebsiella oxytoca
Leclercia sp002902985
Citrobacter portucalensis
Enterobacter chengduensis
Citrobacter braakii
Serratia_J proteamaculans
Enterobacter soli
Phytobacter diazotrophicus
Klebsiella pneumoniae
Providencia rettgeri_D
Enterobacter quasihormaechei
Salmonella enterica
Citrobacter werkmanii
Escherichia coli_F
Enterobacter cloacae_M
Hafnia paralvei
Citrobacter_A farmeri
Enterobacter cloacae
Klebsiella aerogenes
Providencia rettgeri
Enterobacter roggenkampii
Klebsiella michiganensis
Enterobacter hormaechei_B
Erwinia amylovora
Serratia nematodiphila
Proteus faecis
Serratia_A fonticola_B
Yersinia kristensenii
Klebsiella variicola
Enterobacter hoffmannii
Enterobacter cloacae_I
Klebsiella quasipneumoniae
Morganella morganii_A
Pantoea ananatis
Escherichia fergusonii
Providencia_C stuartii
Klebsiella grimontii
Cronobacter malonaticus
Yersinia enterocolitica
Enterobacter mori
Serratia nevei
Pantoea agglomerans
Citrobacter gillenii
Escherichia ruysiae
Klebsiella planticola
Cronobacter sakazakii
Pantoea dispersa
Escherichia albertii
Salmonella diarizonae
Serratia ureilytica
Escherichia sp005843885
Leclercia adecarboxylata
Enterobacter asburiae
Pantoea stewartii
Escherichia coli
Pantoea alii
Staphylococcus aureus
Atlantibacter hermannii
Staphylococcus haemolyticus
Pantoea piersonii
Pantoea symbiotica
Pantoea septica
Pantoea gossypiicola
Cronobacter dublinensis
Gilliamella apicola
Klebsiella terrigena
Pantoea anthophila
Citrobacter meridianamericanus
Pantoea leporis
Rahnella aceris
Proteus sp003144505
Gilliamella apis_A
Pantoea eucalypti
Pantoea vagans


A
Gain
Loss
B
Number of branches
0
10
20
30
40
50
D
M
N
O
T
U
V
W
X
A
B
J
K
L
C
E
F
G
H
I
P
Q
R
S
46
45
45
49
49
49
1
51
1
27
1
52
3
40
42
51
1
1
51
1
50
43
28
47
1
40
40
49
29
51
47
0
10
20
30
40
50
60
Number of species
RG (both)
RG (gain only)
RG (loss only)
Median n (all RGs)


A
Evolutionary time
Common ascestor
ARG-containing clade
≥80% with chromosomal ARGs
Sister clade
<20% with chromosomal ARGs
B
Fold−change (ARG Clade / Sister Clade)
30.0
10.0
3.0
1.0
0.3
***
*
**
ns
ARG
ALL except ARG
gain
loss
C
Interaction effect (fold−change)
100
10
1
***
ns
gain
loss


aminoglycoside antibiotic
phenicol antibiotic
fluoroquinolone antibiotic
sulfonamide antibiotic
diaminopyrimidine antibiotic
Atlantibacter hermannii
Citrobacter braakii
Citrobacter portucalensis
Citrobacter werkmanii
Citrobacter_A farmeri
Edwardsiella tarda
Enterobacter cloacae
Enterobacter cloacae_I
Enterobacter cloacae_M
Enterobacter hoffmannii
Enterobacter hormaechei_B
Enterobacter hormaechei_C
Enterobacter mori
Enterobacter quasihormaechei
Enterobacter roggenkampii
Escherichia coli
Escherichia fergusonii
Klebsiella grimontii
Klebsiella michiganensis
Klebsiella oxytoca
Klebsiella planticola
Klebsiella quasipneumoniae
Leclercia adecarboxylata
Morganella morganii
Morganella morganii_B
Pantoea symbiotica
Proteus mirabilis
Proteus sp003144505
Providencia rettgeri
Providencia rettgeri_D
Rosenbergiella epipactidis
Serratia nevei
Number of Genomes
0
5
10
15
20
25

Gain and loss of plasmid-borne antibiotic resistance genes are associated with their chromosomal counterparts and the broader plasmid gene repertoire in Enterobacteriaceae

Yang Liu[1, *], Yue Liu[1,2, *]

[1] Department of Microbiology, Zhejiang University School of Medicine, Zhejiang, China

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

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

Supplementary Figure Legends

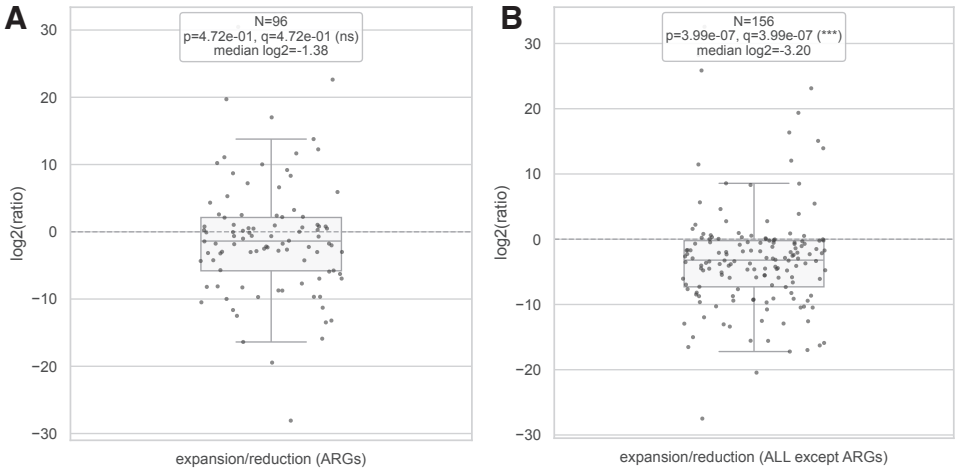


Figure S1. Ratio of gene copy number across species. (E) Log_2 ratio of expansion to reduction rates for ARGs. (F) Log_2 ratio of expansion to reduction rates for other plasmid genes. Each point represents one Enterobacteriaceae species. Box plots show median and quartiles. Gene families were defined as orthologous groups (OGs).

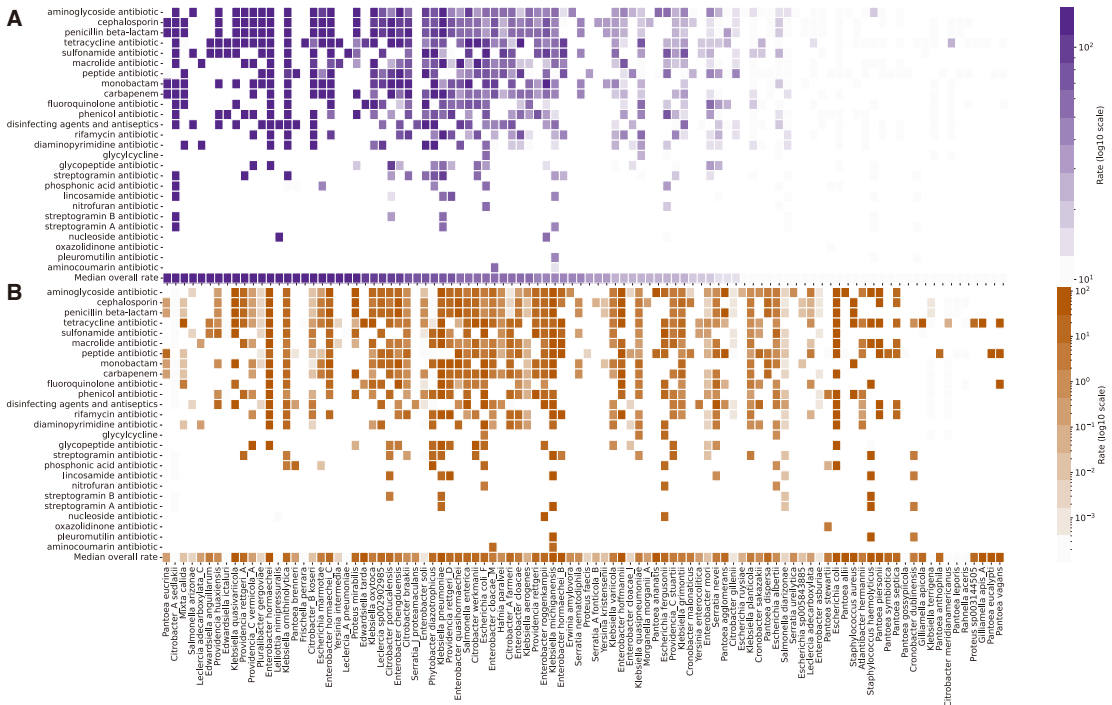


Figure S2. Gain and Loss rate patterns across antibiotic resistance classes and Enterobacteriaceae species. Heatmaps showing evolutionary rates for different

antibiotic resistance classes across bacterial species. (A) Gain rates. (B) Loss rates. Rows represent Enterobacteriaceae species, and columns represent CARD-defined antibiotic resistance classes. Color intensity indicates log10 scaled rate magnitude, with darker colors representing higher rates. White cells indicate absence of the ARG class in the species. Color scales are truncated for visualization clarity: gain rates (20th-80th percentile), loss/expansion/reduction rates (5th-95th percentile); all statistical analyses were performed on complete untruncated datasets.

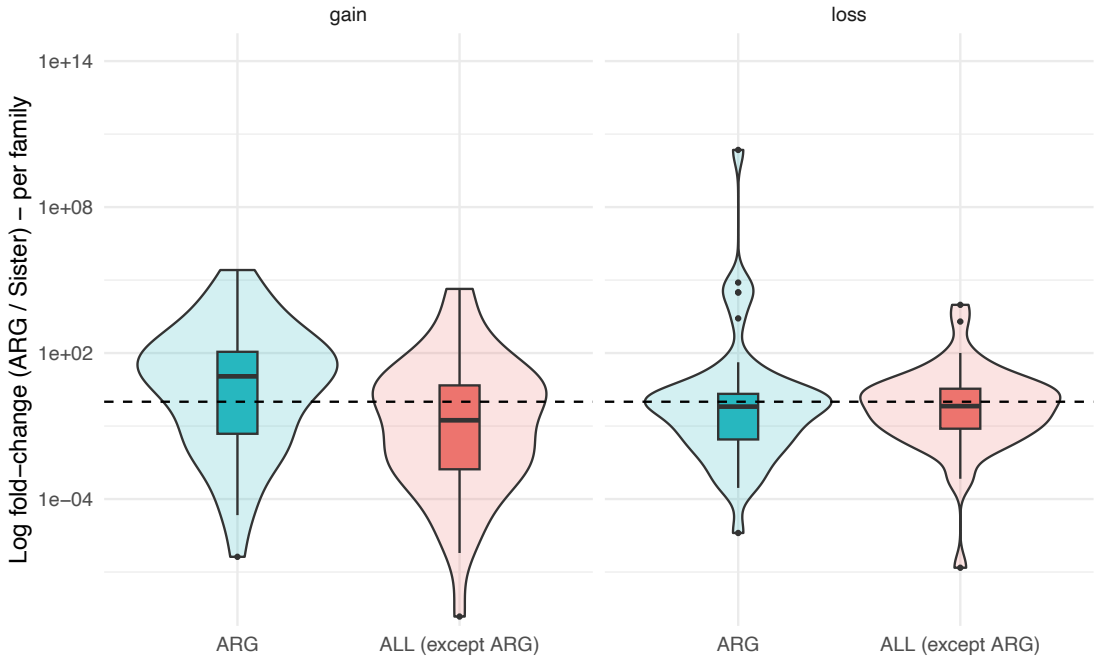


Figure S3. Comparison of evolutionary rates between ARG-containing clades and their sister clades in per-family analyses. Y-axis shows the fold-change ratio (ARG clade / Sister clade) of plasmid ARGs versus other plasmid genes. Dashed line at y=1 indicates no difference between clades. Violin plots show data distribution, and box plots indicate the median and interquartile range.

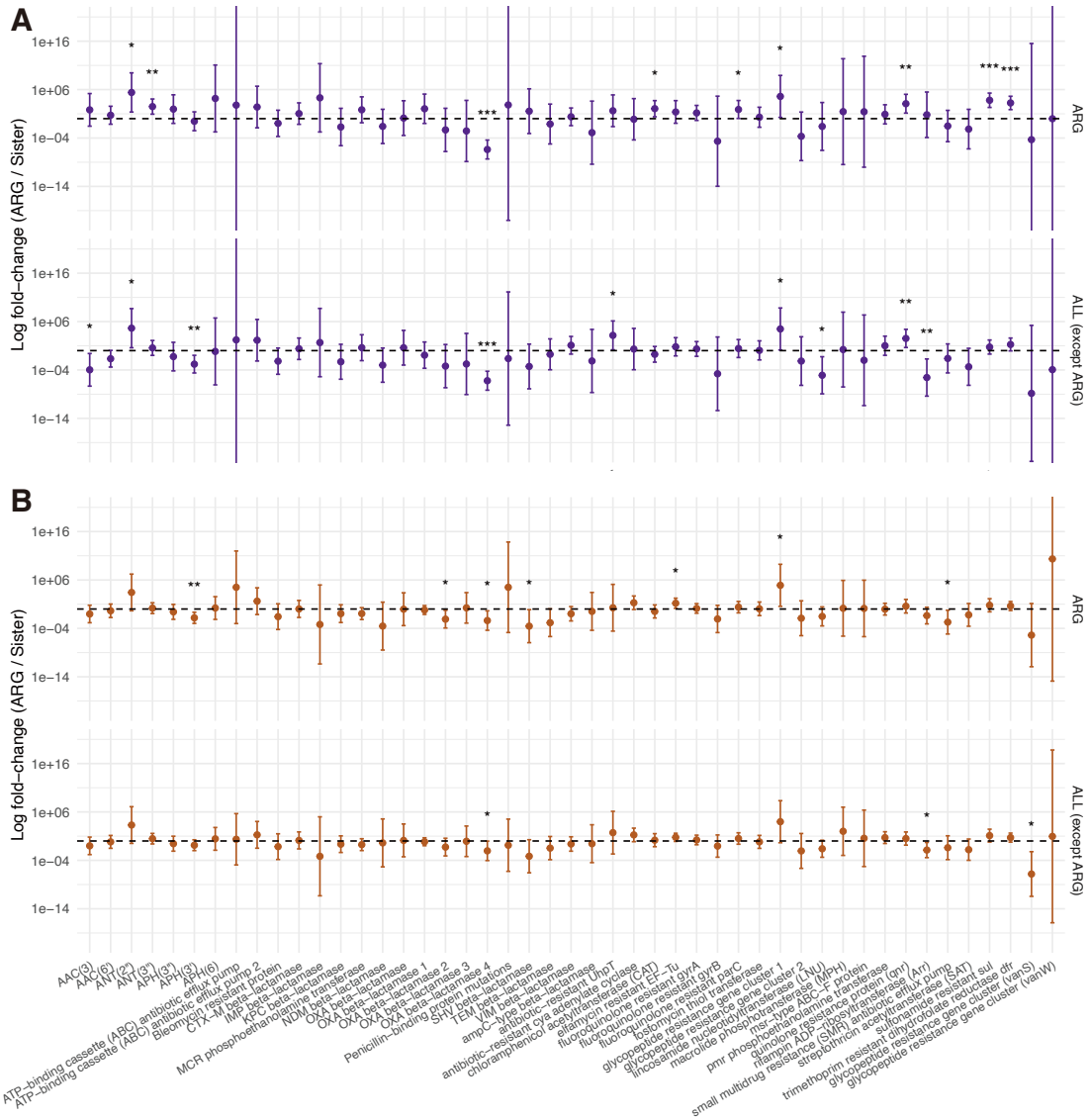


Figure S4. Per-family comparison of evolutionary rates between ARG-containing clades and their sister clades. (A) Gain rates and (B) loss rates are shown for individual antibiotic resistance gene (ARG) families. A dashed line at y = 1 indicates no difference between clades. Asterisks denote statistical significance (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).