PNAS RESEARCH ARTICLE | EVOLUTION OPEN ACCESS

# Plasmid mutation rates scale with copy number

Paula Ramiro-Martínez[a,b], Ignacio de Quinto[a], Laura Jaraba-Soto[a], Val F. Lanza[a,c], Cristina Herencias-Rodríguez[a,c], Adrián González Casanova[d], Rafael Peña-Miller[e,1], and Jerónimo Rodríguez-Beltrán[a,c,1]

Affiliations are included on p. 9.

Edited by Eduardo P. Rocha, Institut Pasteur, Paris, France; received September 16, 2025; accepted December 19, 2025 by Editorial Board Member Eugene V. Koonin

**Plasmids are extrachromosomal DNA molecules that spread by horizontal transfer and shape bacterial evolution. Plasmids are typically present at multiple copies per bacterial cell, and these extra copies increase the supply of plasmid mutations, potentially accelerating their evolution. However, the segregation of plasmid copies to daughter cells is random, introducing an additional layer of genetic drift, termed segregational drift, that might delay plasmid evolution. The interplay between plasmid mutational supply and segregational drift determines the evolutionary rate of plasmid-encoded genes, yet the relative contribution of these opposite forces in plasmid evolution remains unclear. Here, we develop a population genetics framework to predict the rate of plasmid mutations in bacterial populations and validate these predictions using computational, experimental, and bioinformatic approaches. Our findings show that plasmid mutation rates scale logarithmically with copy number and that the supply of new mutations consistently surpasses the impact of segregational drift across all copy numbers. These results underscore plasmids as powerful drivers of bacterial evolvability, where they can potentiate the evolution of critical traits such as antibiotic resistance.**

plasmids | polyploidy | mutation rates | segregational drift | bacterial evolution

## Significance

The classic Luria–Delbrück experiment showed how mutations arise in single-copy bacterial chromosomes. Yet many genetic systems, including plasmids and the genomes of mitochondria, plastids, and cancer cells, exist in multiple copies. In such systems, polyploidy generates opposing forces: Extra copies increase the supply of new mutations, while random segregation slows their fixation. Therefore, it remains unclear whether polyploidy promotes or constrains evolution. Using plasmids as a case study, we combine theory, modeling, experiments, and a bioinformatic extension of the Luria–Delbrück framework to show that mutation rates increase logarithmically with copy number. Our results settle a long-standing debate, demonstrating that plasmids are hotspots of bacterial evolution and revealing the rules by which polyploidy sets the tempo of evolution.

Since the origins of cellular life, all species depend on genomic flexibility to adapt to the ever-changing environment. Life is contingent on permanent change. Bacteria readily acquire genomic changes through mutation and horizontal gene transfer (HGT). Mutation is inherent to DNA replication: No polymerase can copy DNA without mistakes. And it is from those mistakes that life evolved on Earth. However, excessive mutations cause organisms to lose their identity as gene function collapses and their fitness declines (1). It is thus not surprising that life balances evolvability with genome stability and has evolved mechanisms to keep mutations in check (2). But what if bacteria could benefit from specialized genetic platforms in which specific genes could evolve faster? Theoretically, this would be achieved by locally increasing the supply of mutations, either by directly increasing the rate at which mutations are produced (i.e., decreasing replication fidelity) or by increasing the target size of mutations (i.e., the number of DNA molecules that can acquire mutations) (3).

Plasmids—autonomously replicating DNA molecules that coexist with bacterial chromosomes—are ideal candidates to fulfill that role (4). Plasmids are key agents of HGT and play a fundamental role in bacterial ecology and evolution, where they contribute to disseminating relevant traits such as antibiotic resistance (5). Several factors highlight plasmids as potential drivers of bacterial evolution. Plasmid genes might evolve faster than chromosomal genes because they are not essential to their host (with notable exceptions, e.g., refs. 6, 7) and are less connected to cellular metabolic and physiological networks (8–10). This relieves plasmid-encoded genes of the functionality constraints typically associated with the evolution of chromosomal genes. Second, plasmids might be more prone to recombination than chromosomes (11–13). Recombination provides bacterial plasmids with extraordinary modularity, shuffling genes, and combining beneficial alleles in the same backbone (14, 15). Third, plasmid-encoded genes experience an increase in gene dosage that typically translates into higher expression. This magnifies the selective impact of mutations with slight beneficial effects (16–18). And fourth, many plasmids carry other mobile genetic elements, such as integrons and insertion sequences, which drive genetic rearrangements that can accelerate the evolution of both plasmids and chromosomes (19, 20).

Critically, most plasmids are stably present at more than one copy per cell (median= 3.02, IQR= 10.29 copies per chromosome; data from ref. 21), providing an effective polyploid genetic platform (4). This raises complex evolutionary implications that are far from understood. On one hand, classical evolutionary theory predicts that the rate of new mutations

Author contributions: V.F.L., A.G.C., R.P.-M., and J.R.-B. designed research; P.R.-M., I.d.Q., L.J.-S., C.H.-R., A.G.C., R.P.-M., and J.R.-B. performed research; V.F.L., C.H.-R., and J.R.-B. contributed new reagents/analytic tools; P.R.-M., I.d.Q., A.G.C., R.P.-M., and J.R.-B. analyzed data; and P.R.-M., A.G.C., R.P.-M., and J.R.-B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. E.P.R. is a guest editor invited by the Editorial Board.



[1]To whom correspondence may be addressed. Email: rafael.penamiller@gmail.com or jeronimo.rodriguez.beltran@gmail.com.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2526088123/-/DCSupplemental.

Published January 22, 2026.

should increase with gene copy number because each additional copy has an independent chance of acquiring a mutation. Therefore, due to their stable polyploid nature, any given gene should acquire mutations faster when encoded on a plasmid. Supporting this view, experimental evolution showed that de novo mutations conferring antibiotic resistance occur more frequently when the target gene is encoded on a plasmid (18, 22). Moreover, numerical simulations and experimental approaches demonstrate that mutation rates per gene increase with plasmid copy number (PCN) as long as the mutations are dominant (23, 24). This indicates that plasmids can enhance the rate of beneficial mutations, thereby increasing the evolvability of the genes they carry (17, 18, 25).

On the other hand, novel plasmid mutations experience complex allele dynamics (Fig. 1*A*) (4, 26–28). New plasmid mutations only appear in a single plasmid molecule. This mutated plasmid copy will coexist intracellularly under heteroplasmy (also known as heterozygosity) with other nonmutated plasmid molecules until they eventually segregate into different cell lineages (26, 29). Since plasmid segregation is random, daughter cells inherit an allelic set that may differ from the one in the mother cell (for example, only wild-type copies). The establishment probability of plasmid mutations is thus reduced by an additional layer of genetic drift termed segregational drift (27, 30). Recent studies suggest that segregational drift hinders the establishment of plasmid-encoded beneficial alleles in growing populations, as they are more prone to being lost than chromosomal ones (31–34). However, these studies examine the establishment of mutations at single loci and the dynamics of multiple, sequentially arising mutations remain underexplored.

Therefore, the evolutionary rate of plasmid-encoded genes depends on the interplay between two opposing forces: increased mutational supply and segregational drift (Fig. 1*B*). The relative contribution of these two evolutionary forces remains controversial, with studies supporting (17, 18, 25) or opposing (24, 27, 30, 31–33) the role of plasmids as potential catalysts of bacterial evolution. To settle this controversy, here we used a combined theoretical, computational, experimental, and bioinformatic approach to estimate how plasmid mutation rates scale with PCN.

## Results

**A Plasmid-Centered Population Genetics Model Predicts Logarithmic Increases in Mutation Rates As PCN Increases.** To investigate the mutational dynamics of plasmid-encoded genes, we first took a theoretical approach rooted in classical population genetics (35, 36). Specifically, we developed a plasmid-centered population genetics model that shifts the focus from bacterial hosts to plasmids as independent evolutionary entities. In this approach, the key parameter N, representing population size in traditional population genetics, corresponds to the PCN. Thus, each bacterial reproduction event along the ancestral lineage is treated as a separate plasmid generation with a fixed population size set equal to the PCN. Notably, the fixed population size assumption central to classical population genetics is a natural feature of plasmid biology, making plasmids a compelling system for studying evolutionary dynamics through population genetics frameworks (31, 37).

We model plasmid inheritance using a discrete-generation approach related to the Wright–Fisher model (38), where plasmids act as individuals that replicate and segregate randomly across generations. More formally, our model falls within the broader Cannings class of population genetics models, which allows for variable offspring distributions while maintaining a fixed population size. During each cellular reproduction event, a random subset of plasmids is inherited, and the remaining copies are replenished by replication, preserving a constant copy number. This setup captures the random segregational dynamics of plasmids and allows us to trace the ancestry of plasmids across time (Fig. 2*A*). Assuming that a low-probability neutral mutation can occur at each plasmid replication event, the count of distinct mutations in a sample correlates with the length of the genealogy tree. Consequently, the model suggests that nonlinear effects influence the genealogy of plasmid samples, and at certain scales, these genealogies align with the Kingman Coalescent universality class (Fig. 2*B*, see *Materials and Methods*). Under these conditions, the probability of two plasmids sharing a common ancestor decreases with PCN, and the total branch length of the genealogy grows

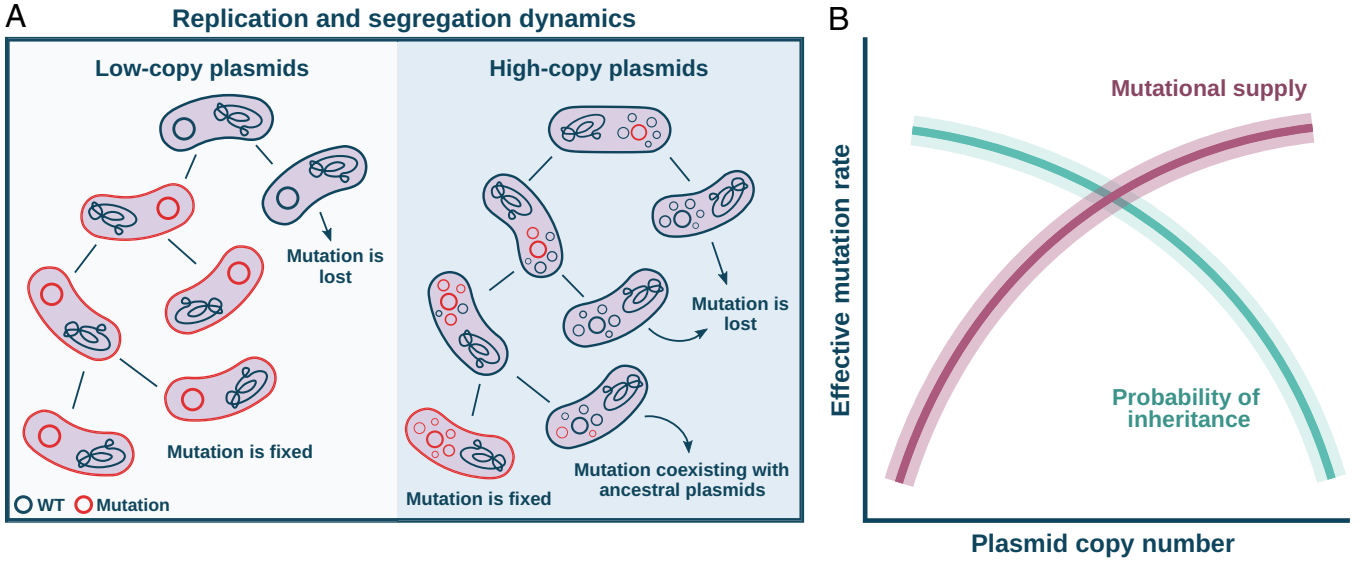


**Fig. 1.** Plasmid-encoded genes face opposing evolutionary forces. (*A*) Plasmid segregation dynamics. The *Left* panel illustrates the segregation of plasmids with a single copy, where a mutant plasmid (red) is inherited during cell division. The *Right* panel shows the dynamics of multicopy plasmids, where segregation occurs randomly upon division and mutant plasmids are either inherited or lost during cell division due to segregational drift. (*B*) The illustration conceptually shows the trade-off between the probability of a new mutation arising (red line) and the probability of that mutation being inherited (green line) in subsequent generations. The mutational supply scales with PCN. In contrast, the chance that any given mutation is retained across generations decreases with PCN, as mutant plasmid copies are more likely to be lost due to segregational drift. This figure is illustrative and not based on real data.

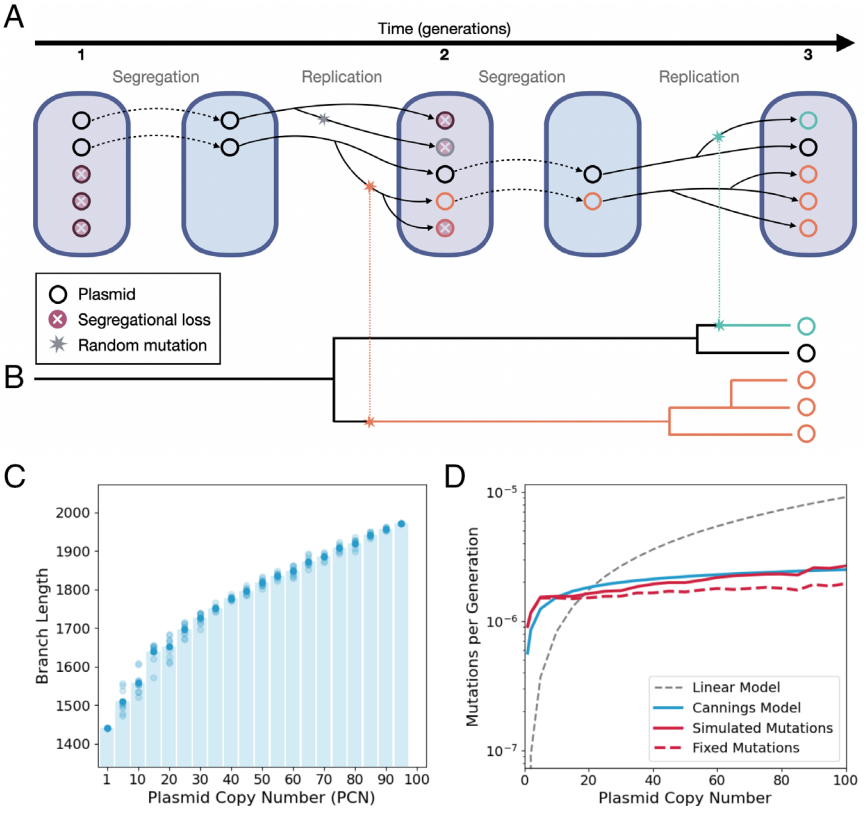


**Fig. 2.** Population genetics model for plasmid mutation accumulation. (*A*) Schematic diagram of the population genetics model. Each cell harbors a population of plasmids, with plasmid segregation during cell division modeled as a binomial process ($P$ = 1/2). Reproduction events are represented within a Cannings framework, where PCN remains constant across generations. (*B*) Example lineage tree of plasmids derived from the model. Genealogies demonstrate the coalescence dynamics driven by segregational drift, with branch lengths measured in generations, corresponding to the time to the most recent common ancestor. (*C*) Total branch length of the lineage tree as a function of PCN. Branch length is defined as the cumulative distance from the root to the tips of the tree. Simulation results show that while total branch length increases with PCN, the rate of increase slows at higher PCN values. (*D*) Number of mutations per generation per plasmid as a function of PCN. The gray dotted line represents a linear extrapolation based on the chromosomal mutation rate ($\mu = 9.2 \times 10^{-8}$; see Fig. 3), while the blue line depicts the Cannings model, where the constants $c_1$ and $c_2$ were adjusted to match the observed relationship between plasmid copy number and mutation accumulation. The red lines correspond to computer simulations: the solid red line shows the average total mutation rate across 1,000 simulations per PCN, and the red dashed line shows the average number of fixed mutations, defined as mutations present in all plasmids within a cell.

logarithmically with copy number (Fig. 2*C*). As a result, the effective mutation rate ($\mu_{eff}$), defined as the average number of distinct mutations per generation per plasmid, is expected to scale with the logarithm of PCN: $\mu_{eff} = \mu\left(c_1 + c_2 ln(PCN)\right)$, where $\mu$ denotes the underlying mutation rate (in mutations per generation and base pair) and $c_1$ and $c_2$ are constants. This relationship suggests that increasing PCN raises the mutational supply, yet the rate of increase diminishes at higher PCNs (Fig. 2*D*).

**Simulating Multilevel Evolution in Multicopy Plasmids.** To evaluate the predictions of the population genetics model, we implemented a computational framework based on object-oriented programming principles (*Materials and Methods*). The model simulates bacterial populations as collections of individual cells, each containing plasmids modeled as independent entities (*SI Appendix*, Fig. S1). This structure allows for the simulation of multilevel evolutionary processes, capturing key mechanisms—including random plasmid replication, segregation, and mutation—across multiple generations under controlled in silico conditions.

In the model, each cell in the population contains a defined number of plasmid copies that replicate during cell growth until the target PCN is reached. Plasmid copies are randomly distributed to daughter cells during cell division, and a single bacterial cell is randomly selected at each generation to seed the next population, creating a bottleneck that drives genetic drift. Mutations occur with a nonzero probability in each plasmid copy during replication. Each mutation is recorded as an attribute of the corresponding plasmid object and includes details such as the generation of occurrence, allowing precise tracking of mutations over time within the simulation framework (*SI Appendix*, Figs. S2–S5).

We performed 1,000 independent simulations for a range of PCN, each elapsing 1,400 bacterial generations. At the end of the simulation, plasmids were sampled from the final bacterial population to reconstruct their evolutionary relationships. We then built a lineage tree by tracing the ancestry of each plasmid based on its generational history. Plasmids that share a common lineage (i.e., descend from the same ancestral plasmid in earlier generations) are grouped together, with branches representing points where lineages diverged due to random segregation events during cell division (*SI Appendix*, Fig. S6).

Finally, we quantified the total length of the lineage tree, defined as the cumulative distance from the root to the tips across all plasmid lineages. Consistent with the theoretical predictions, as PCN increases and we sample more plasmids per cell, the total branch length also increases, reflecting more coalescing events and longer genealogical paths. However, as shown in Fig. 2*C*, this increase slows at higher PCN levels, consistent with a logarithmic increase in tree length as the copy number rises, even though the total tree length continues to grow (Fig. 2*D* and *SI Appendix*, Fig. S7).

**A Mutation Accumulation Experiment Validates Computational and Theoretical Predictions.** To experimentally test our theoretical and computational predictions, we used a mutation accumulation (MA) approach. MA experiments typically consist of serially passaging multiple replicate clonal lines by repeated streaking on rich agar plates to generate new colonies initiated by a single cell. This severe bottleneck minimizes natural selection, allowing genomes to accumulate nearly any mutation regardless of its selective effect, except those that are lethal or highly deleterious (39). After a number of passages, mutations across entire genomes are identified by sequencing, allowing for the direct estimation of mutation rates by simply counting the number of genetic changes accumulated over a defined number of generations (39, 40).

To reduce the number of passages required to accumulate and detect sufficient mutations in the MA experiment, we used a hypermutator derivative (Δ*mutS;* see *Materials and Methods*) of

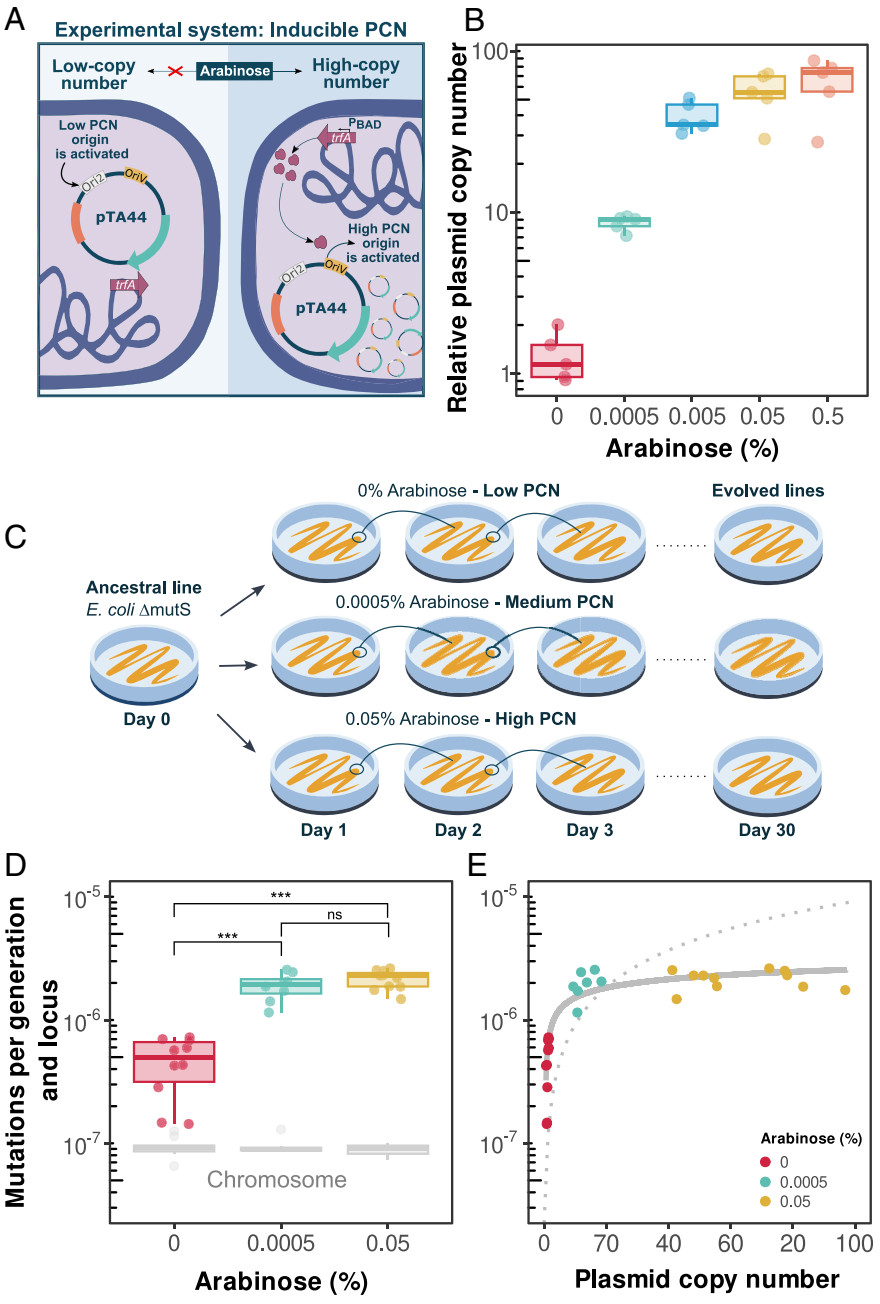


**Fig. 3.** An MA approach validates theoretical predictions. (*A*) Plasmid copy number induction in the pTA44 plasmid. In the presence of arabinose, expression of TrfA protein from the $P_{BAD}$ promoter is turned on, initiating replication from the high-copy replication origin (oriV). Without arabinose replication is driven by the low-copy replication origin (ori2). (*B*) PCN of the pTA44 plasmid at different arabinose concentrations. The copy number of five biological replicates was calculated by qPCR. The average PCN of 9 technical replicates was 1.31 copies for 0% of arabinose, 9.32 for 0.0005%, 40.07 for 0.005%, 57.70 for 0.05%, and 67.65 for 0.5%. (*C*) Experimental setup. We conducted a MA experiment at different arabinose concentrations. Each day, one random colony of each line was transferred to a new petri dish with the appropriate arabinose concentration. This cycle was repeated for 30 d. (*D*) Boxplots show plasmid mutation rates (mutations per generation and locus) at different arabinose concentrations. Each point represents an independent biological replicate. Boxes indicate the median and interquartile range; whiskers extend to 1.5× IQR. Chromosomal mutation rates (gray) are shown for reference. Statistical significance was assessed using pairwise Wilcoxon rank-sum exact tests (*$P < 10^{-4}$; ns, not significant). (*E*) Plasmid mutation rates scale logarithmically with copy number (Datasets S1 and S2). Each dot represents an independent experimental line, colored by the arabinose concentration used during evolution (see legend). For the analyses, we used PCN measurements obtained through qPCR. The dashed line indicates the expected linear increase in the absence of segregational drift, based on the chromosomal mutation rate. The solid line shows the fitted Cannings model, with constants $c_1$ and $c_2$ adjusted to match the experimental data (Dataset S3).

the *Escherichia coli* EPI300 strain carrying a plasmid with a tunable PCN. The plasmid, dubbed pTA44, can initiate replication from two distinct origins: *ori2* from the F plasmid and *oriV* from the RK2 plasmid (41). Replication from *ori2* is constitutive and leads to approximately one copy per cell. Replication from *oriV* depends on the presence of the replication initiation protein TrfA, whose expression is chromosomal in the EPI300 strain and controlled by the arabinose-inducible $P_{BAD}$ promoter (Fig. 3*A* and *SI Appendix*, Fig. S8). The PCN of the pTA44 plasmid shows a wide dynamic range spanning from ~1 copy per cell in the absence of arabinose to nearly 70 copies at high arabinose concentrations (Fig. 3*B*). Since bacterial growth was relatively robust to changes in PCN (*SI Appendix*, Fig. S9), we selected three arabinose concentrations (0, 0.0005, and 0.05%) for subsequent experiments. These yielded, on average, 1, 10, and 60 plasmid copies per cell, corresponding to low, medium, and high PCN regimes, respectively (Fig. 3*B*).

We serially propagated 12 independent lineages under each PCN regimen (low, medium, and high), for a total of 36 lineages. Each day, we streaked a randomly selected colony onto fresh LB agar plates supplemented with the appropriate arabinose concentration (Fig. 3*C*). During each passage, populations expanded from a single cell to ~$10^7$ cells, which correspond to ~24 bacterial generations (*SI Appendix*, Fig. S10). Repeating this process for 28 d yielded ~670 generations per lineage. After discarding eight lineages due to cross-contamination or other issues (see *Materials and Methods* and below), we retained 28 lineages, representing ~6,700 generations of evolution per PCN regimen and nearly ~20,000 generations across all regimens (Dataset S1). The number of generations per passage was consistent across lineages over the duration of the experiment, suggesting selection played a minimal role under our experimental conditions (*SI Appendix*, Fig. S10, Spearman's rank correlation test between number of generations and days, $P = 0.81$, $r_s = 0.014$). We also verified that the plasmid

was stable within individual colonies (*SI Appendix*, Fig. S11) and that PCN remained stable across arabinose concentrations. The only exception were two lineages in the medium-copy regimen (p5.1 and p5.7), which were omitted for further analysis (*SI Appendix*, Figs. S12 and S13).

We sequenced the genomes of all evolved lineages and the ancestor and identified all mutations (single-nucleotide polymorphisms—SNPs—and small indels). A key feature of plasmid segregation and replication dynamics is that multiple plasmid alleles can coexist within the same cell, leading to plasmid-mediated heteroplasmy or heterozygosity (26, 29, 31). Therefore, plasmid-encoded alleles can be polymorphic even in otherwise clonal lineages. To account for this possibility, we kept mutations with an allelic frequency of at least 5%, as done previously (31). Following this approach, we identified 284 plasmid and 7,620 chromosomal mutations (Dataset S1).

We calculated the spontaneous mutation rate for both plasmids and chromosomes by dividing the number of mutations by the length of each replicon and the number of generations elapsed for each line. As expected, chromosomal mutation rates remained unaffected by the presence of arabinose (Kruskal–Wallis rank-sum test, $P = 0.79$; Fig. 3*D*), and were consistent with previous estimations (42). In contrast, plasmid mutation rates were significantly higher in the medium and high PCN regimes than in the low PCN (Pairwise comparisons using the Wilcoxon rank-sum exact test, $P < 10^{-4}$ between low and medium, and low and high copy number regimes; Fig. 3*D*). There were no significant differences between the medium and high PCN regimes, despite a ~5-fold difference in PCN (Wilcoxon rank-sum test, $P = 0.27$). Strikingly, at the low PCN regime, where plasmids and chromosomes were present at comparable copy numbers (~1 copy per cell), plasmid mutation rates were approximately five times higher than those of the chromosome.

We then analyzed the mutation rates of each evolved line individually. Plasmid mutation rates increased with PCN and were positively correlated, either calculated by qPCR (Fig. 3*E*, Spearman's rank correlation, $r_s = 0.77$, $P < 10^{-6}$) or with the sequencing coverage (*SI Appendix*, Fig. S14, Spearman's rank correlation, $r_s = 0.79$, $P < 10^{-6}$). However, the rate of increase diminished progressively as PCN rose. We observed a growing divergence between the observed mutation rate and the expected rate if mutations accumulated linearly with PCN, using the chromosomal mutation rate as a baseline (Fig. 3*E*). This aligns with our theoretical and computational predictions, as a logarithmic regression best fitted the experimental data ($R^2 = 0.79$; Dataset S2).

**Estimation of Plasmid Mutation Rates From Whole-Genome Sequencing Data.** We hypothesized that if mutation rates scale with PCN beyond experimental conditions, a compatible signature should be detectable in natural plasmids. To test this idea, we analyzed the extensive collection of whole-genome sequences in the NCBI database. In a typical sequencing experiment, a culture is inoculated from a small number of cells, incubated overnight, and used for DNA extraction and sequencing. Sequencing reads are then assembled into a consensus genome, representing the dominant genotype. However, mutations can arise during the growth phase in both chromosomal and plasmid DNA. While most of these mutations may be too infrequent to be detected, some might reach higher frequencies through drift or selection and become detectable in the reads. We reasoned that, by aligning sequencing reads to their corresponding consensus assemblies, we could capture within-sample mutations that most likely arose during culture growth, and the number of such mutations would be roughly proportional to the underlying mutation rate (Fig. 4*A*).

We analyzed a previously published dataset comprising 723 *E. coli* genomes and 1,598 associated plasmids, for which complete assemblies, sequencing reads, and PCN estimates were available (21). These plasmids belonged to 155 replicon types, 57 plasmid taxonomic units (PTUs), and 136 sequence-based clusters of highly similar plasmids (90% identity and k = 3 neighbors). We identified putative mutations by aligning sequencing reads to their corresponding consensus assemblies. To reduce potential false positives from sequencing errors, we implemented a strict filtering criterion: We retained only genomes with high overall coverage (≥30×) and considered only variants with allele frequencies of ≥5% (31). Additionally, we excluded common sequencing-related artifacts by discarding SNPs unsupported by reads from both strands and variants located within homopolymeric regions (43) (*Materials and Methods*). This process resulted in the identification of 13,196 plasmidic and 192,322 chromosomal putative mutations.

To estimate plasmid mutation rates, we treated each plasmid within a sequence-based cluster as an independent replicate and estimated the number of mutations per plasmid cluster using a maximum likelihood estimation (MLE) algorithm (44). We then normalized mutation counts by median plasmid cluster length to obtain the mutation density, defined here as the number of within-cluster, high-confidence mutations per base pair (Fig. 4*B*). This approach mirrors the classical Luria–Delbrück fluctuation test, where parallel cultures are propagated, and the distribution of mutant counts across cultures is used to infer mutation rate (39, 45, 46).

We first focused on host chromosomal mutation densities and found that as expected, they remained stable across PCN categories (Fig. 4*D*, gray boxplots, Kruskal–Wallis test $P = 0.57$). Then, we classified plasmids into low, medium, and high copy number groups (Fig. 4*C*) and compared their mutation density. High-copy number plasmids exhibited significantly higher mutation densities than both medium-copy plasmids and low-copy plasmids (Fig. 4*D*; Kruskal–Wallis test $P < 10^{-11}$ followed by the Wilcoxon rank-sum exact test, $P < 0.0019$, effect size > 0.28 for all pairwise comparisons; Dataset S4). Moreover, PCN and mutation density (Fig. 4*E*) were strongly and positively correlated (Spearman's rank correlation $r_s = 0.82$, $P < 10^{-16}$), even after normalizing by chromosomal mutation density (*SI Appendix*, Fig. S15, Spearman's rank correlation $r_s = 0.80$, $P < 10^{-16}$). This relationship was well fitted by both a logarithmic model ($R^2 = 0.63$, $P < 10^{-16}$, Dataset S2) and the theoretical expectation derived from the Cannings model (residual sum-of-squares of $1.34 \times 10^{-7}$ and one iteration with tolerance = $7.92 \times 10^{-9}$). Alternative methods for estimating plasmid and chromosomal mutation densities yielded qualitatively similar results (*SI Appendix*, Figs. S16–S18). Of note, and in agreement with the MA experiment, we observed a fivefold increase in mutation density of low-copy plasmids compared with chromosomal rates. Altogether, despite large variation in experimental conditions and other uncontrolled factors across datasets, our bioinformatic approach yielded a mutational signature that is highly compatible with our theoretical, computational, and experimental observations.

## Discussion

The mechanisms of evolution do not operate in isolation; instead, highly intertwined and often opposing forces shape evolutionary trajectories. This is particularly true for plasmids. On the one hand, plasmid polyploidy causes an increase in the availability of mutational targets relative to the chromosome, potentially accelerating the evolution of plasmid-encoded genes (4, 18, 20, 21, 47). On the other hand, because plasmid alleles must first reach fixation

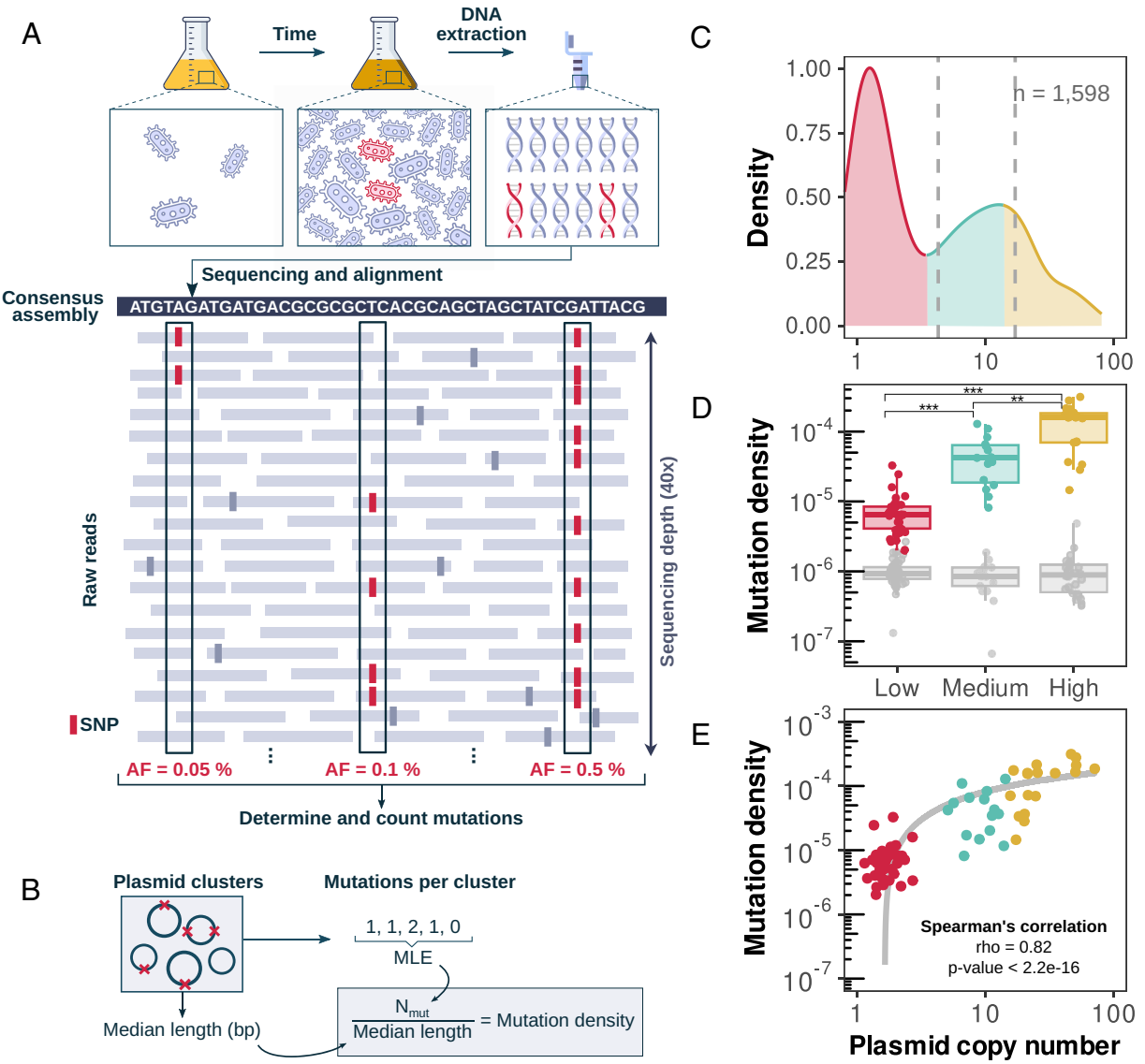


**Fig. 4.** Mutation density determination using whole-genome sequencing data. (*A*) This panel illustrates the method used to identify low-frequency mutations in whole-genome sequencing assays. First, cultures were grown to extract the genomic DNA. This growth time is sufficient to accumulate mutations within the culture population. Once the DNA was extracted and sequenced, we used the available reads in the NCBI database to search for these mutations. Reads are mapped against their respective assemblies to detect mutations within the "population" of reads, with a minimum allele frequency (AF) of 5% and a maximum of 50%. (*B*) Schematic representation of how mutation density was estimated using sequencing data. For each plasmid cluster, the total number of detected mutations was used to determine the number of mutations per cluster ($N_{mut}$) using a maximum likelihood estimation (MLE) algorithm. Mutation density was then calculated by normalizing the number of mutations by the within-cluster median plasmid length. (*C*) Distribution of PCN in *E. coli* plasmids (n = 1,598). Colors indicate thresholds used to classify plasmids as low, medium, or high PCN, based on the antimode (4.3 copies) and the second mode of the PCN distribution (17.06 copies). (*D*) Plasmid mutation density (mutations per bp) by plasmid cluster (y-axis) for each condition of PCN: high (n = 27), medium (n = 17), and low (n = 40) (Kruskal–Wallis test ($P < 10^{-11}$) followed by the Wilcoxon rank-sum test, low vs. medium, $P < 10^{-9}$, effect size = 0.79; low vs. high, $P < 10^{-13}$, effect size = 0.91; medium vs. high, $P = 0.0018$, effect size = 0.47, Dataset S4). The chromosomal mutation densities are shown in gray. (*E*) Correlation between mutation density (mutations per bp) and PCN grouping by plasmid clusters (Spearman's correlation, $r_s = 0.82$ and $P < 10^{-16}$). The gray line represents the Cannings model, where the constants $c_1$ and $c_2$ were adjusted to match the observed relationship between PCN and mutations per bp (Dataset S3). Only plasmids with fewer than 100 copies are shown.

within individual cells before reaching fixation at the population level, plasmid-mediated adaptation can be significantly delayed (24, 27, 30, 31–33). These opposing forces have been studied separately, leading to controversial results that either support or refute the role of plasmids as catalysts of bacterial evolution.

We addressed this controversy by integrating theoretical, computational, experimental, and bioinformatic approaches. Using a Wright–Fisher-based model and agent-based simulations, we showed that plasmid evolution can be described as a Cannings process and predicted a logarithmic relationship between PCN and mutation accumulation (Fig. 2). To empirically test these predictions, we performed an MA experiment using a plasmid with tunable PCN. Mutation rates markedly increased from ~1 to 10 copies but seem to plateau beyond this point, even at 60 copies. These results were best explained by a logarithmic regression, which captured the decelerating relationship between PCN and mutation rates (Fig. 3). We then analyzed extensive sequencing data as independent fluctuation assays and showed that plasmids with higher PCNs tend to accumulate more detectable mutations per base pair (Fig. 4). Together, these results indicate that plasmid evolvability scales logarithmically with PCN beyond laboratory conditions, with mutation rates plateauing around the

natural boundary separating high- and low-copy plasmids [i.e., ~6 copies per cell (21)].

But what is the biological meaning of this logarithmic relationship? Population genetics offers an intuitive explanation: During severe bottlenecks and over many bacterial generations, most plasmid mutations are lost due to segregational drift, reducing their contribution to final plasmid diversity. However, over shorter time scales, the increase in mutational supply surpasses segregational drift, leading to a rise in the mutation rate associated with plasmids. Consequently, under our study conditions, most of the plasmid mutations we detected likely arose shortly before sampling. This has broader implications for evolutionary analyses, as the rapid and recent accumulation of mutations in HCP can potentially complicate efforts to infer accurate evolutionary trajectories from endpoint data.

A noteworthy observation is that plasmid mutation rates were about fivefold higher than chromosomal rates at comparable copy number in both our MA experiment and our bioinformatic analyses (Figs. 3 and 4). Several non–mutually exclusive hypotheses could explain this pattern. i) Plasmids might evolve faster because they are less essential to the host, yet, under our conditions, the fraction of essential loci is similar on the plasmid and chromosome (~10%) and selection is minimized in the MA experiment making this unlikely to be the main explanation. ii) Systematic underestimation of PCN would inflate plasmid mutation rates. However, PCN was assessed by independent methods (qPCR and WGS coverage) and datasets (MA experiment and NCBI) that closely agree, arguing against a pervasive bias. iii) Conflicts between the two replication origins in our model plasmid could elevate mutation rates, but the same fivefold excess appears in natural plasmids, so this cannot fully account for the result. iv) Plasmid replication may be intrinsically more error-prone or less efficiently repaired. However, plasmids rely on host polymerases and repair machineries for their replication, so this would involve an unknown mechanism. Further work will be needed to test the generality of this pattern and to dissect the underlying mechanisms.

A potential limitation of our theoretical, computational, and experimental (but not bioinformatic) analyses is that the conditions we tested depart from natural settings. We used a single-cell bottleneck regime that maximizes genetic drift and minimizes the effective population size of plasmid lineages. Theory predicts that under less stringent bottlenecks, the effective mutation rate should be higher due to reduced effect of drift, suggesting that our estimates likely represent a lower bound. Future work applying similar approaches across a range of bottleneck sizes will help estimate effective mutation rates for plasmids in diverse ecological settings. A second limitation is the absence of horizontal transfer in our analysis. Conjugation provides additional opportunities for plasmid replication (that involve transient single-stranded DNA) that could further increase plasmid mutation rates, although this effect should be partially independent of PCN. Conjugation also enables plasmids to circulate among hosts, promoting distributed evolution within bacterial communities and allowing species to access and exploit genetic variation generated by their neighbors (48).

Beyond plasmids, our findings likely extend to any genetic element present in multiple copies and subject to random segregation during cell division, such as plastid and mitochondrial DNA (mtDNA). For example, human mitochondria contain thousands of mtDNA copies and undergo significant bottlenecks during maternal transmission (49). Vertebrate mtDNA exhibits a mutation rate approximately 20 times higher than nuclear DNA (50), a pattern that aligns with our results and underscores the robustness of the relationship between copy number and mutation rate across different genetic systems.

Collectively, our results have important implications for microbial biotechnology. Increasing PCN could accelerate plasmid evolution in directed-evolution platforms, whereas low-copy plasmids could help suppress unwanted evolution in bioproduction. Moreover, our findings, along with other studies (13, 18, 26), highlight that plasmids are key drivers of bacterial evolution, extending beyond their role as gene delivery platforms. By functioning as autonomous genetic elements with increased mutation rates, plasmids can reduce the burden they cause to their host and potentiate the evolution of critical bacterial traits such as antibiotic resistance, virulence, and metabolism, underscoring their role as the evolutionary powerhouses of the bacterial cell.

## Materials and Methods

**Population Genetics Model of Plasmid Evolution.** We described the dynamics of plasmid evolution using a Cannings model, a classical population genetics model that captures stochastic reproduction while maintaining a fixed population size. In our context, each generation corresponds to a bacterial reproduction event, and the population size at each generation is defined by the PCN, such that the key parameter N is equal to the number of plasmids per cell. During each reproduction event, plasmids have a 50% chance of being inherited by the daughter cell or remaining in the mother cell, independently of other plasmids. This results in a binomial distribution of plasmid counts in each new cell, with parameters $P = 1/2$ and $N = PCN$. A random binomial subset of plasmids, denoted as $B \sim Binomial(N, 1/2)$, moves to the daughter cell, while the remaining $N - B$ plasmids are replicated independently, each copying one of the $B$ ancestors.

Mutations can occur during plasmid replication, with each mutation event occurring with a low probability. To capture the evolutionary dynamics of plasmid lineages, we treat plasmids as independent entities within a Cannings model with constant population size N (51), where at each generation, a random subset of plasmids is selected as ancestors for the next generation. Each plasmid in generation $g + 1$descends from a uniformly chosen ancestor from the subset selected in generation $g$, and the remaining plasmids are replicated to restore copy number. The probability that two plasmids in a cell trace back to the same plasmid in the preceding cell is E[1/B], which is on the order of 1/N. Likewise, the probability that three plasmids share a common ancestor is $E[1/B^2]$, scaling as $1/N^2$. According to Möhle's criteria (52), this places the genealogical process of plasmid lineages within the universality class of the Kingman Coalescent (53). This result implies that, under neutral conditions, the expected total branch length of the genealogy increases logarithmically with PCN, leading to a sublinear scaling of mutation accumulation with increasing copy number.

**Simulation of Plasmid Mutation Accumulation.** The simulation was implemented in Python using standard scientific computing libraries, including NumPy and Matplotlib, with additional functionality from Biopython for lineage tree construction. The model represents bacterial cells as objects containing plasmids, each tracked independently to capture processes such as mutation accumulation and segregational drift. Plasmid replication was modeled as a stochastic process until reaching the maximum PCN, and segregation during cell division was simulated by randomly partitioning plasmids between daughter cells with a probability of 1/2.

The serial transfer protocol involved propagating populations for 60 simulated days, each comprising 24 generations. At the end of each day, a single bacterial cell was randomly selected to seed the next population. This bottleneck amplified genetic drift and allowed the accumulation of mutations to be observed over multiple transfers. Mutation events were introduced probabilistically during plasmid replication at a rate of $1.8 \times 10^{-8}$ per plasmid per generation, with each mutation logged alongside its generation of origin. For each PCN value, we performed simulations until at least 1,000 replicates with one or more mutations were obtained.

Lineage trees were constructed at the end of simulations by tracing plasmid ancestries based on replication and segregation events. The total length of these trees was calculated as the cumulative distance from root to tips, serving as a measure of genealogical divergence. Mutations were analyzed by tracking their frequency and persistence across generations to quantify key dynamics: emergence (appearance of new mutations), fixation (mutations reaching 100% frequency in the population), and loss (disappearance of mutations due to segregational drift or genetic bottlenecks). The simulation generated output files

containing plasmid composition per cell across generations and reconstructed lineage trees, which were used to study genealogical divergence and mutation dynamics. The implementation is available at https://github.com/ccg-esb/MAp.

**Bacterial Strains, Plasmids, and Growth Conditions.** The strains used in this study were *E. coli* EPI300 and a hypermutator derivative of the same strain. To construct the hypermutator derivative strain, the *mutS* gene was replaced by a kanamycin resistance gene flanked by FRT sites, following the protocol described in ref. 54 with minor modifications. Briefly, the thermosensitive pKOBEG plasmid (55) carrying the lambda red system was introduced into the EPI300 strain. Then, a purified PCR product (primers MutS_F/R) amplified from the Δ*mutS::kan* KEIO mutant (55) was electroporated on arabinose-induced EPI300 cells. This PCR fragment contained the kanamycin resistance (*kan*) gene flanked by FRT sites, flanked by two ~50 bp homology arms. After recombination, the pKOBEG plasmid was cured by incubating at 37°C. The thermosensitive pCP20 plasmid carrying site-specific FLP recombinase was then introduced to remove the kanamycin resistance and subsequently cured by incubating at 42 °C. Excision of the *kan* resistance gene was verified by PCR (primers K1/K2 and mutS_F/R), yielding the EPI300 Δ*mutS* strain. The hypermutator status of this strain was verified by performing fluctuation assays to rifampicin resistance.

EPI300 Δ*mutS* was transformed by heat shock with the pTA44 plasmid (10.3 kb) (41) as a tunable copy number plasmid. pTA44 carries two origins of replication: *ori2* and *oriV* (41). The *ori2* is a single-copy replication origin, while replication from *oriV* is proportional to the amount of TrfA initiation protein. The *trfA* gene, encoded in the host *E. coli* EPI300 Δ*mutS* chromosome, is controlled by the $P_{BAD}$ promoter, which allows fine-tuning plasmid replication initiation by simply adding the inducer L-arabinose (41). Confirmation of the plasmid transformation was verified by PCR (repE_pTA44_F/R).

**Growth Curves.** Single colonies from each bacterial population were inoculated into LB starter cultures and incubated at 37 °C for 16 h with shaking at 225 rpm (three biological replicates). Each culture was then diluted 1:2,000 in LB medium supplemented with kanamycin and the appropriate concentration of arabinose, and 200 μL of the diluted culture was transferred to a 96-well microtiter plate. The plates were incubated at 37 °C with orbital shaking for 24 h, and the optical density (OD) at 600 nm was measured every 10 min using a Synergy HTX plate reader (BioTek). The area under the growth curve (AUC) was calculated using the "auc" function from the "flux" R package. AUC was chosen as the primary growth metric because it incorporates key growth parameters, including maximum growth rate, lag phase duration, and carrying capacity.

**Mutation Accumulation Experiment.** Each MA line was originated from a single colony independently isolated from LB agar plates. Each day, each replicate line was streaked to yield isolated colonies on LB agar plates with kanamycin and the appropriate concentration of arabinose and then incubated at 37 °C for 22 h. This process was repeated for 30 consecutive days. To ensure random selection of the colony to be passaged, the furthest colony on the streak's end was chosen.

Initially, 12 Δ*mutS* lines were established for each condition of PCN. Lineages were named using a systematic naming scheme to indicate their experimental conditions. Each name followed the format p[arabinose].[replicate]. The first number (0, 5, or 50) corresponds to the concentration of arabinose used for induction (0, 0.0005, or 0.05%, respectively). The final number (1 to 12) identifies the specific replicate lineage within that group. Some lines were excluded due to highly similar mutation patterns that likely resulted from cross-contamination during the experiment. Specifically, the excluded lines were p0.5 and p0.7 from the 0% arabinose condition, p5.2, p5.5, and p5.12 from the 0.0005% arabinose condition, and p50.4 from the 0.05% arabinose condition. Additionally, two samples (p5.7 and p5.1) were removed due to a decline in PCN (*SI Appendix*, Fig. S12). We could determine the likely cause for this decline in p5.7, which showed a missense mutation in the *trfA* gene (*SI Appendix*, Fig. S13) Consequently, ten lines were retained for the 0% arabinose condition, seven lines for the 0.0005% condition, and eleven lines for the 0.05% condition.

During the MA experiment, a random colony from each of the 12 lines was resuspended in 100 μL of MilliQ water daily to be stored at −20 °C. Before storage, 20 μL of suspension was transferred to 180 μL of saline buffer, serially diluted, and replated to count viable cells and estimate the number of elapsed generations using the following formula: $\frac{log(finalcellcount) - log(initialcellcount)}{log(2)}$, where final cell counts correspond to the final CFUs/mL and the initial cell count is assumed to be 1 (the seeding cell of the colony). The stored samples were also used to estimate PCN every ten days during the experiment, as described below.

**Relative Copy Number Determination.** To determine relative PCN of the pTA44 plasmid, qPCR assays were performed using an Applied Biosystems 7300 Real-Time PCR System and NZYSupreme qPCR Green Master Mix (NZYTech), following the protocol described in ref. 56. We developed a specific primers for pTA44 (repE_F: TCGGATTGACCTCTGCGGAAGC, repE_R: GCCTTTTTCATCGCCGGCATCC, amplicon size: 113 bp, efficiency: 99.09%, $R^2$ = 0.999) and we used a previously described primer set for the dxs chromosomal gene (dxs_F: CGAGAAACTGGCGATCCTTA, dxs_R: CTTCATCAAGCGGTTTCACA, amplicon size: 113 bp, efficiency: 101.7%, $R^2$ = 0.998) (57). Single colonies were resuspended in MilliQ water, heated at 98 °C for 10 min, and directly used in qPCRs as template DNA (58). Reaction efficiencies were determined from standard curves generated using five 8-fold serial dilutions of the template DNA, each tested in duplicate. qPCR cycling conditions included an initial polymerase activation step at 95 °C for 2 min, followed by 40 cycles of denaturation at 95 °C for 5 s and annealing/extension at 60 °C for 30 s. Plasmid copy number was calculated using the following formula: $PCN = [(1 + E_c)^{CTc}/(1 + E_p)^{CTp}] \times (S_c/S_p)$, where $E_c$ and $E_p$ are the efficiencies of the chromosomal and plasmid qPCR, CTc, and CTp are the threshold cycles of chromosomal and plasmid reactions, and $S_c$ and $S_p$ the size (bp) of the chromosomal and plasmid amplicons (56).

Although qPCR is the most widely used method to estimate PCN, we validated these results using whole-genome sequencing data. Specifically, we calculated the PCN of each sample as the ratio of plasmid contig coverage to chromosomal coverage. Comparison of PCN values obtained by qPCR on the final day of the evolution experiment with those estimated from sequencing data revealed a strong and statistically significant correlation between the two methods (Pearson's product-moment correlation, cor = 0.83, $P = 9.7 \times 10^{-9}$, *SI Appendix*, Fig. S19).

**Whole Genome Sequencing and analysis of the MA experiment.** Thirty-six evolved populations of the hypermutator strain carrying the plasmid pTA44, were sequenced using the Illumina platform at the Wellcome Trust Center for Genomic Sequencing (Oxford, UK). Genomic DNA was extracted from 1 ml of resuspended cells using the Wizard Genomic DNA Purification Kit (Promega) following the manufacturer's instructions. All samples were quantified on a Qubit fluorometer (ThermoFisher) and a Nanodrop 2000c (ThermoFisher). The reads of the sequences generated and analyzed during the current study are available in the EMBL-EBI repository, BioProject ID: PRJEB104260.

Raw Illumina sequencing reads were quality-filtered using TrimGalore v0.6.6 (https://github.com/FelixKrueger/TrimGalore), trimming low-quality bases from both ends with a minimum Phred quality score threshold of 20. Filtered reads were assembled de novo into contigs using SPAdes genome assembler v3.13.1 (https://github.com/ablab/spades) with default parameters. Assembled contigs were annotated using Prokka v1.14.5 (https://github.com/tseemann/prokka) to identify coding sequences and other genomic features. To identify plasmid-derived contigs for each strain, BLASTn v2.9.0 + was used to align all assembled sequences against the reference plasmid sequence.

To detect genetic variants in evolved populations relative to their respective ancestral strains, we used Breseq v0.36.1 (https://github.com/barricklab/breseq) (43) in polymorphism mode, which enables the detection of both fixed and subpopulation (nonfixed) mutations. This approach was critical for capturing mutation rates across multicopy plasmids, where mutations may not be fixed in all copies. SNPs not supported by reads from both strands and variants located within homopolymeric regions were discarded, using the arguments *--polymorphism-coverage-both-strands 2* and *--polymorphism-reject-homopolymer-length 5*. The resulting mutation data were processed and summarized using custom scripts in R v4.1.0, and all statistical analyses were performed in R and RStudio (https://github.com/PaulaRamiro/Mutation_accumulation).

**Mutation Rate Analysis in *Escherichia coli* Genomes.** To detect mutations in naturally occurring plasmids, we first identified and downloaded all available assemblies from *Escherichia coli* annotated as "Complete Genomes" in the NCBI database (n = 4,124) on 05/12/2023. SRA information was extracted using the sra-toolkit v2.11.3 (https://github.com/ncbi/sra-tools) with a custom pipeline (https://github.com/PaulaRamiro/Mutation_accumulation) and used to download available

paired-end reads (corresponding to n =787 assemblies). Then, reads were trimmed to a Phred score of 20 with Trim Galore v0.6.6 (https://github.com/FelixKrueger/TrimGalore) and mapped against their respective assemblies with Breseq v0.38.1 (https://github.com/barricklab/breseq) with the -p flag to analyze nonfixed mutations within the populations. PCN information was extracted from ref. 21.

We classified plasmids using different methods. First, we typed plasmids into different incompatibility groups according to their replication mechanism (59, 60) using MOB-typer from mob_suite v3.1.8 (https://github.com/phac-nml/mob-suite) (60) with the flag *--multi* to type independent plasmids within samples. This method analyzes features within the DNA sequences responsible for plasmid replication, such as the genes that encode replication initiation proteins. This allows us to establish plasmid groups that share similar replication mechanisms, known as replicon types. Second, we used a classification scheme based on similarity across the whole plasmid genetic content with COPLA v1.0 (61). Plasmids that share high homology (>70%) in more than 50% of their sequence are assigned to the same plasmid taxonomic unit (PTU) (61, 62). Third, we classified plasmids into different sequence-based plasmid clusters. Plasmid nucleotide FASTA files were sketched with Mash (63) (k = 21, sketch = 10,000) to obtain an all-against-all pair-wise distance matrix; a k-nearest-neighbor graph was then built by connecting each plasmid to its three closest neighbors when the Mash distance was ≤0.1, and the resulting network was partitioned with the Fast-Greedy community-detection algorithm implemented in igraph (64), assigning every plasmid to a unique cluster.

We treated each plasmid as a replicate within groups categorized by clusters, replicon types, and PTUs as if they were replicates in a classical Luria–Delbrück fluctuation assay. The number of mutations ($N_{mut}$) per plasmid group and 95% CIs were calculated using the maximum likelihood estimator (MLE), applying the *newton.LD* and *confint.LD* functions of the rSalvador package v1.9 for R (44). The $p_0$ method was also used as a control to include the samples without mutations, with the *LD.p0.est* function. $\mu_{eff}$ for both chromosomes and plasmids was then calculated by dividing $\mu$ by the total length of the replicon in base pairs. This led to the estimation of $\mu_{eff}$ for 723 genomes and 1,598 plasmids.

**Materials and Correspondence.** Correspondence and requests for materials should be addressed to Rafael Peña-Miller and Jerónimo Rodríguez-Beltrán.

**Data, Materials, and Software Availability.** Sequencing data, experimental data, and code data have been deposited in Zenodo, EMBL-EBI and Github (https://doi.org/10.5281/zenodo.17132776) (65). Previously published data were used for this work (There's no citation. NCBI data was used).

**ACKNOWLEDGMENTS.** We thank Fernando Baquero, Álvaro San Millán, Javier DelaFuente, Alfonso Santos-López, and João Gama for their helpful suggestions. Work in the evodynamics lab (https://evodynamicslab.com/) is supported by project nos. PI21/01363 and PI23/01945, funded by the Carlos III Health Institute (ISCIII) and cofunded by the European Union; CIBER-Consorcio Centro de Investigación Biomédica en Red-(CIBERINFEC) CB21/13/00084; Convocatoria SEIMC-FUNDACIÓN SORIA MELGUIZO de Investigación 2021; and funded by the European Union (ERC, HorizonGT, 101077809). Views and opinions expressed are, however, those of the author (s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. P.R.-M. is a recipient of a predoctoral PFIS Grant (Grant no. FI22/00265) from the ISCIII, through the Recovery, Transformation and Resilience Plan and Next Generation EU from the European Union. R.P.-M. received support from DGAPA-UNAM (PASPA-2024-III). J.R.-B. acknowledges support by a Miguel Servet contract from the ISCIII (Grant no. CP20/00154), cofounded by the European Social Fund, "Investing in your future." V.F.L. acknowledges support by a Miguel Servet contract from the ISCIII (Grant no. CP22/00164), cofounded by the European Social Fund, "Investing in your future." We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust Grant reference 203141/Z/16/Z) for the generation and initial processing of sequencing data.

Author affiliations: [a]Microbiology Department, Hospital Universitario Ramón y Cajal-Instituto Ramón y Cajal de Investigación Sanitaria, Madrid 28034, Spain; [b]Escuela de Doctorado, Departamento de Microbiología, Universidad Autónoma de Madrid, Madrid 28049, Spain; [c]Centro de Investigación Biomédica en Red de Enfermedades Infecciosas, Instituto de Salud Carlos III, Madrid 28029, Spain; [d]Department of Applied Mathematics, School of Mathematical and Statistical Sciences, and Center for Mechanisms of Evolution (Biodesign Institute), Arizona State University, Tempe, AZ 85281; and [e]Programa de Biología de Sistemas, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca 62210, Mexico

1. L. Loewe, W. G. Hill, The population genetics of mutations: Good, bad and indifferent. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 1153–1167 (2010).
2. E. Darmon, D. R. F. Leach, Bacterial genome instability. *Microbiol. Mol. Biol. Rev.* **78**, 1–39 (2014).
3. E. Sano, S. Maisnier-Patin, J. P. Aboubechara, S. Quiñones-Soto, J. R. Roth, Plasmid copy number underlies adaptive mutability in bacteria. *Genetics* **198**, 919–933 (2014).
4. J. Rodríguez-Beltrán, J. DelaFuente, R. León-Sampedro, R. C. MacLean, Á. San Millán, Beyond horizontal gene transfer: The role of plasmids in bacterial evolution. *Nat. Rev. Microbiol.* **19**, 347–359 (2021).
5. J. Wiedenbeck, F. M. Cohan, Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol. Rev.* **35**, 957–976 (2011).
6. M. Anda *et al.*, Bacterial clade with the ribosomal RNA operon on a small plasmid rather than the chromosome. *Proc. Natl. Acad. Sci.* **112**, 14343–14347 (2015).
7. A. Manzano-Marın *et al.*, Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids' di-symbiotic systems. *ISME J.* **141**, 259–273 (2019).
8. V. S. Cooper, S. H. Vohr, S. C. Wrocklage, P. J. Hatcher, Why genes evolve faster on secondary chromosomes in bacteria. *PLoS Comput. Biol.* **6**, e1000732 (2010).
9. O. Cohen, U. Gophna, T. Pupko, The complexity hypothesis revisited: Connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* **28**, 1481–1489 (2011).
10. S. J. Tazzyman, S. Bonhoeffer, Why there are no essential genes on plasmids. *Mol. Biol. Evol.* **32**, 3079–3088 (2015).
11. B. Niaudet, L. Jannière, S. D. Ehrlich, Recombination between repeated DNA sequences occurs more often in plasmids than in the chromosome of *Bacillus subtilis*. *Mol. Gen. Genet. MGG* **197**, 46–54 (1984).
12. J. Rodríguez-Beltrán *et al.*, High recombinant frequency in extraintestinal pathogenic *Escherichia coli* strains. *Mol. Biol. Evol.* **32**, 1708–1716 (2015).
13. X. Zhang, D. E. Deatherage, H. Zheng, S. J. Georgoulis, J. E. Barrick, Evolution of satellite plasmids can prolong the maintenance of newly acquired accessory genes in bacteria. *Nat. Commun.* **10**, 5809 (2019).
14. T. F. Cooper, Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biol.* **5**, e225 (2007).
15. M. Barlow, J. Fatollahi, M. Salverda, Evidence for recombination among the alleles encoding TEM and SHV-lactamases. *J. Antimicrob. Chemother.* **63**, 256–259 (2008).
16. S. Million-Weaver, D. L. Alexander, J. M. Allen, M. Camps, "Quantifying plasmid copy number to investigate plasmid dosage effects associated with directed protein evolution" in *Methods in Molecular Biology (Clifton, N.J.)* (2012), pp. 33–48.
17. A. Couce, A. Rodríguez-Rojas, J. Blázquez, Bypass of genetic constraints during mutator evolution to antibiotic resistance. *Proc. Biol. Sci.* **282**, 20142698 (2015).
18. A. San Millan, J. A. Escudero, D. R. Gifford, D. Mazel, R. C. MacLean, Multicopy plasmids potentiate the evolution of antibiotic resistance in bacteria. *Nat. Ecol. Evol.* **1**, 0010 (2016).
19. C. Souque, J. A. Escudero, R. C. MacLean, Integron activity accelerates the evolution of antibiotic resistance. *eLife* **10**, e62474 (2021).
20. J. Sastre-Dominguez *et al.*, Plasmid-encoded insertion sequences promote rapid adaptation in clinical enterobacteria. *Nat. Ecol. Evol.* **8**, 2097–2112 (2024).
21. P. Ramiro-Martínez, I. de Quinto, V. F. Lanza, J. A. Gama, J. Rodríguez-Beltrán, Universal rules govern plasmid copy number. *Nat. Commun.* **16**, 6022 (2025).
22. J. L. Martinez, F. Baquero, Epidemiology of antibiotic-inactivating enzymes and DNA probes: The problem of quantity. *J. Antimicrob. Chemother.* **26**, 301–303 (1990).
23. J. Rodríguez-Beltrán *et al.*, Genetic dominance governs the evolution and spread of mobile genetic elements in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 15755–15762 (2020).
24. M. Santer, H. Uecker, Evolutionary rescue and drug resistance on multicopy plasmids. *Genetics* **215**, 847–868 (2020).
25. J. Rodríguez-Beltrán *et al.*, Genetic dominance governs the evolution and spread of mobile genetic elements in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 15755–15762 (2020).
26. J. Rodriguez-Beltran *et al.*, Multicopy plasmids allow bacteria to escape from fitness trade-offs during evolutionary innovation. *Nat. Ecol. Evol.* **2**, 873–881 (2018).
27. M. Santer, A. Kupczok, T. Dagan, H. Uecker, Fixation dynamics of beneficial alleles in prokaryotic polyploid chromosomes and plasmids. *Genetics* **222**, iyac121 (2022), 10.1093/genetics/iyac121.
28. F. Rossine, C. Sanchez, D. Eaton, J. Paulsson, M. Baym, Intracellular competition shapes plasmid population dynamics. bioXriv [Preprint] (2025), http://biorxiv.org/lookup/doi/10.1101/2025.02.19.639193 [Accessed 7 May 2025].
29. R. P. Novick, Plasmid incompatibility. *Microbiol. Rev.* **51**, 381–395 (1987).
30. J. Ilhan, *et al.*, Segregational drift and the interplay between plasmid copy number and evolvability. *Mol. Biol. Evol.* **36**, 472–486 (2019).
31. A. Garoña, N. F. Hülter, D. Romero Picazo, T. Dagan, Segregational drift constrains the evolutionary rate of prokaryotic plasmids. *Mol. Biol. Evol.* **38**, 5610–5624 (2021).
32. A. Garoña, M. Santer, N. F. Hülter, H. Uecker, T. Dagan, Segregational drift hinders the evolution of antibiotic resistance on polyploid replicons. *Microbiology* **19**, e1010829 (2023).
33. I. Dewan, H. Uecker, Evolutionary rescue of bacterial populations by heterozygosity on multicopy plasmids. *J. Math. Biol.* **90**, 26 (2025).
34. A. 'Olek' Pisera, C. C. Liu, The role of plasmid copy number and mutation rate in evolutionary outcomes. *Nat. Ecol. Evol.* **9**, 1694–1704 (2025), 10.1038/s41559-025-02792-7.
35. J. Wakeley, *Coalescent Theory: An Introduction* (Roberts & Co, 2009).
36. A. Etheridge, *Some Mathematical Models from Population Genetics: École d'Été de Probabilités de Saint-Flour XXXIX-2009* (Springer, Berlin Heidelberg, 2011).
37. J. C. R. Hernandez-Beltran *et al.*, Segregational instability of multicopy plasmids: A population genetics approach. *Ecol. Evol.* **12**, e9469 (2022).
38. R. Durrett, *Probability Models for DNA Sequence Evolution* (Springer, N.Y., 2008).
39. J. E. Barrick, R. E. Lenski, Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14**, 827–839 (2013).

40. D. L. Halligan, P. D. Keightley, Spontaneous mutation accumulation studies in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* **40**, 151–172 (2009).
41. T. Aakvik *et al.*, A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. *FEMS Microbiol. Lett.* **296**, 149–158 (2009).
42. H. Lee, E. Popodi, H. Tang, P. L. Foster, Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2774–E2783 (2012).
43. D. E. Deatherage, J. E. Barrick, "Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq" in *Engineering and Analyzing Multicellular Systems, Methods in Molecular Biology*, L. Sun, W. Shou, Eds. (Springer, New York, 2014), pp. 165–188.
44. Q. Zheng, rSalvador, An R package for the fluctuation experiment. *G3 Bethesda* **7**, 3849–3856 (2017).
45. S. E. Luria, M. Delbrück, Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491–511 (1943).
46. P. L. Foster, Sorting out mutation rates. *Proc. Natl. Acad. Sci.* **96**, 7617–7618 (1999).
47. J. P. J. Hall, M. A. Brockhurst, C. Dytham, E. Harrison, The evolution of plasmid stability: Are infectious transmission and compensatory evolution competing evolutionary trajectories? *Plasmid* **91**, 90–95 (2017).
48. O. Kosterlitz *et al.*, Evolutionary "crowdsourcing": Alignment of fitness landscapes allows for cross-species adaptation of a horizontally transferred gene. *Mol. Biol. Evol.* **40**, msad237 (2023).
49. E. R. Árnadóttir *et al.*, The rate and nature of mitochondrial DNA mutations in human pedigrees. *Cell* **187**, 3904–3918.e8 (2024).
50. R. Allio, S. Donega, N. Galtier, B. Nabholz, Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol. Biol. Evol.* **34**, 2762–2772 (2017).
51. C. Cannings, The latent roots of certain Markov chains arising in genetics: A new approach. I. Haploid models. *Adv. Appl. Probab.* **6**, 260–290 (1974).
52. M. Möhle, A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Adv. Appl. Probab.* **30**, 493–512 (1998).
53. M. Möhle, S. Sagitov, A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* **29**, 1547–1562 (2001).
54. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6640–6645 (2000).
55. M.-K. Chaveroche, A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans*. *Nucleic Acids Res.* **28**, 97e–997 (2000).
56. A. San Millan *et al.*, Small-plasmid-mediated antibiotic resistance is enhanced by increases in plasmid copy number and bacterial fitness. *Antimicrob. Agents Chemother.* **59**, 3335–3341 (2015).
57. C. Lee, J. Kim, S. G. Shin, S. Hwang, Absolute and relative qPCR quantification of plasmid copy number in *Escherichia coli*. *J. Biotechnol.* **123**, 273–280 (2006).
58. M. Škulj *et al.*, Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes. *Microb. Cell Fact.* **7**, 1–12 (2008).
59. M. Shintani, Z. K. Sanchez, K. Kimbara, Genomics of microbial plasmids: Classification and identification based on replication and transfer systems and host taxonomy. *Front. Microbiol.* **6**, 242 (2015).
60. J. Robertson, J. H. E. Nash, MOB-suite: Software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb. Genomics* **4**, e000206 (2018).
61. S. Redondo-Salvo *et al.*, Copla, a taxonomic classifier of plasmids. *BMC Bioinformatics* **22**, 390 (2021).
62. S. Redondo-Salvo *et al.*, Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat. Commun.* **11**, 3602 (2020).
63. B. D. Ondov *et al.*, Mash: Fast genome and metagenome distance estimation using minhash. *Genome Biol.* **17**, 132 (2016).
64. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal Complex Syst.* **1695**, 1–9 (2006).
65. P. Ramiro-Martínez, Plasmid mutation rates scale with copy number. Zenodo. https://doi.org/10.5281/zenodo.17132776. Deposited 7 January 2026.