Molecular Biology and Evolution, 2025, **42**, 1–17
https://doi.org/10.1093/molbev/msaf062
Advance access publication 26 March 2025
Discoveries

MBE OXFORD

Evolution of a Plasmid Regulatory Circuit Ameliorates Plasmid Fitness Cost

Clinton A. Elg [1,2,3], **Erin Mack** [2,3], **Michael Rolfsmeier** [3], **Thomas C. McLean** [4], **David Sneddon** [1,2,3], **Olivia Kosterlitz** [2,3,5], **Elizabeth Soderling**[3], **Solana Narum** [3], **Paul A. Rowley** [3], **Jack Sullivan** [1,2,3], **Christopher M. Thomas** [6], **Eva M. Top** [1,2,3,*]

[1]Bioinformatics and Computational Biology Program, University of Idaho, Moscow, ID, USA
[2]Institute for Interdisciplinary Data Sciences, University of Idaho, Moscow, ID, USA
[3]Department of Biological Sciences, University of Idaho, Moscow, ID, USA
[4]Department of Molecular Microbiology, John Innes Centre, Norwich, UK
[5]Biology Department, University of Washington, Seattle, WA, USA
[6]School of Biosciences, University of Birmingham, Edgbaston, Birmingham, UK

***Corresponding author:** E-mail: evatop@uidaho.edu.
Associate editor: Miriam Barlow

Abstract

Plasmids promote adaptation of bacteria by facilitating horizontal transfer of diverse genes, notably those conferring antibiotic resistance. Some plasmids, like those of the incompatibility group IncP-1, are known to replicate and persist in a broad range of bacteria. We investigated a poorly understood exception, the IncP-1β plasmid pBP136 from a clinical *Bordetella pertussis* isolate, which quickly became extinct in laboratory *Escherichia coli* populations. Through experimental evolution, we found that the inactivation of a previously uncharacterized plasmid gene, *upf31*, drastically improved plasmid persistence in *E. coli*. The gene inactivation caused alterations in the plasmid regulatory system, including decreased transcription of the global plasmid regulators (*korA*, *korB*, and *korC*) and numerous genes in their regulons. This is consistent with our findings that Upf31 represses its own transcription. It also caused secondary transcriptional changes in many chromosomal genes. In silico analyses predicted that Upf31 interacts with the plasmid regulator KorB at its C-terminal dimerization domain (CTD). We showed experimentally that adding the CTD of *upf31*/pBP136 to the naturally truncated *upf31* allele of the stable IncP-1β archetype R751 results in plasmid destabilization in *E. coli*. Moreover, mutagenesis showed that *upf31* alleles encoded on nearly half of the sequenced IncP-1β plasmids also possess this destabilization phenotype. While Upf31 might be beneficial in many hosts, we show that in *E. coli* some alleles have harmful effects that can be rapidly alleviated with a single mutation. Thus, broad-host-range plasmid adaptation to new hosts can involve fine-tuning their transcriptional circuitry through evolutionary changes in a single gene.

Keywords: plasmids, compensatory evolution, transcription regulation, multidrug resistance, IncP-1

Introduction

Bacterial evolution and rapid adaptation are frequently shaped by the horizontal transfer of genes, including those transferred by extra-chromosomal DNA elements known as plasmids (Norman et al. 2009; Soucy et al. 2015). One example is the plasmid-encoded spread of antibiotic resistance, which enables bacterial pathogens to resist traditional therapeutic antibiotic treatments, leading to the rise of highly resistant "superbugs" that increasingly threaten human health (San Millan 2018; Antimicrobial Resistance Collaborators 2022). Understanding how plasmids evolve to successfully transfer and maintain themselves in bacterial populations and communities is thus a critical step toward limiting the spread of antibiotic resistance.

A central metric for assessing the degree to which a plasmid remains in a bacterial population over time is termed "plasmid persistence." Plasmid–host pairs often differ in plasmid persistence, and even the same plasmid has been shown to persist differently in closely related species (De Gelder et al. 2007; Kottara et al. 2018). Instances of poor plasmid persistence can often be attributed to decreased fitness of the plasmid-containing bacterium relative to its plasmid-free counterpart. This fitness cost of plasmid carriage imposed on the bacterial host is due to several factors, including metabolic costs and molecular conflicts between plasmid and host machinery (Modi and Adams 1991; San Millan and MacLean 2017). Several studies have shown that compensatory mutations arise during serial batch cultivation of plasmid–host pairs, ameliorating plasmid fitness cost and thereby restoring plasmid persistence. Some of the important early studies showed that such cost amelioration could occur through genetic changes on the plasmid but were unable to explore the underlying molecular mechanisms (Bouma and Lenski 1988; Dahlberg and Chao 2003). With the advent of next-generation sequencing, compensatory mutations have been identified in specific plasmid genes, including those encoding accessory functions (Bottery et al. 2017; Stalder et al. 2017), replication initiation (Sota et al. 2010; Hughes et al. 2012; Yano et al. 2016), and conjugation machinery (De Gelder et al. 2008; Porse et al.

Received: June 13, 2024. **Revised:** February 13, 2025. **Accepted:** February 28, 2025

2016; Jordt et al. 2020; Yang et al. 2023). Additionally, chromosomal evolution can stabilize host–plasmid pairs through amelioration of plasmid cost via mutations targeting global regulators (Harrison et al. 2015; Stalder et al. 2017) and putative helicases (San Millan et al. 2014; Loftie-Eaton et al. 2017). Thus, compensatory mutations that increase plasmid persistence can occur on the plasmid, chromosome, or both (Hall et al. 2021).

The focal plasmid of this work is the self-transmissible plasmid pBP136, which was previously discovered in a *Bordetella pertussis* strain isolated from a lethal infection in an infant (Kamachi et al. 2006). This cryptic 41,268 bp plasmid contains genomic features typical of its incompatibility group IncP-1β, which is 1 of 13 subgroups of the broad host range IncP-1 plasmids (Hayakawa et al. 2022). These features include a replicon with replication initiation gene *trfA*, a *tra* region for DNA processing, *trb* region for conjugation machinery, and a tightly controlled central control operon. This plasmid backbone ensures high plasmid persistence across many Proteobacteria (Schmidhauser and Helinski 1985; Adamczyk and Jagura-Burdzy 2003; Yano et al. 2012, 2013) with a few exceptions (De Gelder et al. 2007). Plasmid pBP136 was previously used as a model plasmid in key studies addressing the evolution of IncP-1 plasmids. The first of these showed that foreign genes tend to insert into specific "accessory regions" of IncP-1 genomes as a result of selection for the integrity of plasmid transferability and maintenance regions (Sota et al. 2007). Next, a pBP136 mini-replicon variant was created which removed the *tra* and *trb* operons related to self-transmission but left the replication, control, and stable inheritance regions intact. Experimental evolution of this vertical inheritance-only mini-replicon showed that plasmid persistence could be improved in various ways. Mutations in the *trfA* gene stabilized the plasmid in *Shewanella oneidensis*, but incurred trade-offs in the form of host range shifts (Sota et al. 2010; Hughes et al. 2012; Yano et al. 2012). In a *Pseudomonas* species, the same mini-replicon improved its persistence and expanded its host range through inter-plasmid transposition of plasmid persistence functions, like toxin–antitoxin (TA) systems, in combination with chromosomal mutations (Loftie-Eaton et al. 2016). A variant of the original full-size plasmid pBP136 which included a kanamycin resistance gene, pBP136Km (42,389 bp), was shown to rapidly adapt to a *S. oneidensis* host via multiple distinct evolutionary pathways (Stalder et al. 2017, 2020). One evolutionary pathway again involved *trfA* mutations, while others involved transposition of a TA system and chromosomal mutations, showing that the conjugative plasmid and its mini-replicon followed similar evolutionary trajectories. While these studies have greatly expanded our understanding of the evolution of IncP-1 plasmids, in some cases, new questions have emerged. In this work, we sought to examine the previously reported but unexplained poor persistence of plasmid pBP136Km in several *Escherichia coli* K-12 strains (Sota and Top 2008).

We report a single genetic change in pBP136Km that occurred within 50 generations of evolution in *E. coli* K-12 in the absence of antibiotic selection and which drastically increased the plasmid's persistence. This improved persistence was explained by a decrease in the plasmid fitness cost due to the partial deletion and presumed inactivation of a plasmid-encoded gene with unknown protein function, *upf31*. This gene was previously proposed to encode a putative adenine methylase that is associated with an intriguing set of clustered repeats specific to the IncP-1β plasmids (Thorsted et al. 1998). A search for *upf31* homologs detected the gene in most IncP-1β plasmids, almost half of which encode a Upf31 isoform very similar to that of pBP136. We show here that both this common isoform and Upf31 of pBP136 destabilize the archetypal IncP-1β plasmid R751, which has a truncated version of *upf31* and is otherwise stable in *E. coli*. In pBP136Km, the presence of Upf31 was associated with changes in the expression of plasmid backbone regulatory genes and chromosomal genes. In silico analyses suggested a physical interaction between Upf31 and plasmid regulator KorB at its C-terminal dimerization domain (CTD), which was shown to be essential for the destabilization phenotype. Our study highlights that adaptation of broad-host-range self-transmissible plasmids to new hosts can involve fine-tuning their intricate gene regulatory systems through mutations in a single gene.

Results

Plasmid-Encoded *upf31* Results in Poor Plasmid Persistence in *E. coli* Hosts

We first sought to confirm a previous finding (Sota and Top 2008) that plasmid pBP136Km shows poor persistence in various *E. coli* hosts. Using *E. coli* K-12 MG1655 (hereafter K-12), we performed a plasmid persistence assay entailing daily passage of triplicate cultures in nonselective conditions (i.e. without antibiotics). This initial assay with the ancestral host–plasmid pair showed poor plasmid persistence (Fig. 1a in blue). In line with the previous study, we randomly isolated one plasmid-bearing clone from each of the triplicate populations on day 5 and tested plasmid persistence in these clones. The persistence of the plasmid in these clones was markedly improved (clones A, B, and C in Fig. 1a in red, green, and magenta, respectively).

The three isolated clones were sequenced to identify the genetic changes that may explain the observed change in plasmid persistence (supplementary table S1, Supplementary Material online). We found deletions within pBP136Km consistently targeting the region often associated with accessory genes between the *tra* and *trb* operons; the most stable clones B and C had identical plasmids (Fig. 1b). Additionally, no genetic changes were observed in the chromosomes of these three clones. As the plasmid deletions in all three clones at least included *upf30.5* and *upf31.0*, we suspected the presence of one or both genes to be responsible for the poor plasmid persistence in *E. coli*.

We also routinely sequenced clones recovered after transferring the ancestral plasmid between strains of *E. coli* by conjugation (i.e. matings) or electroporation. Among these clones were several with deletions either internal to *upf31* or in its putative promoter (supplementary table S1, Supplementary Material online). To test whether deletion of *upf31* alone could explain improved plasmid persistence, we compared the persistence of one of these recovered variants, named pBP136KmΔ*upf31*, to that of the ancestral plasmid pBP136Km in the same host, *E. coli* K-12 (Fig. 2a). This selected plasmid contained a 168 bp deletion internal to the open reading frame of *upf31*, whereas *upf30.5* was intact (Fig. 1b and supplementary table S1, Supplementary Material online). Given that the plasmid with inactivated *upf31* was persistent for at least 10 d, this gene of unknown

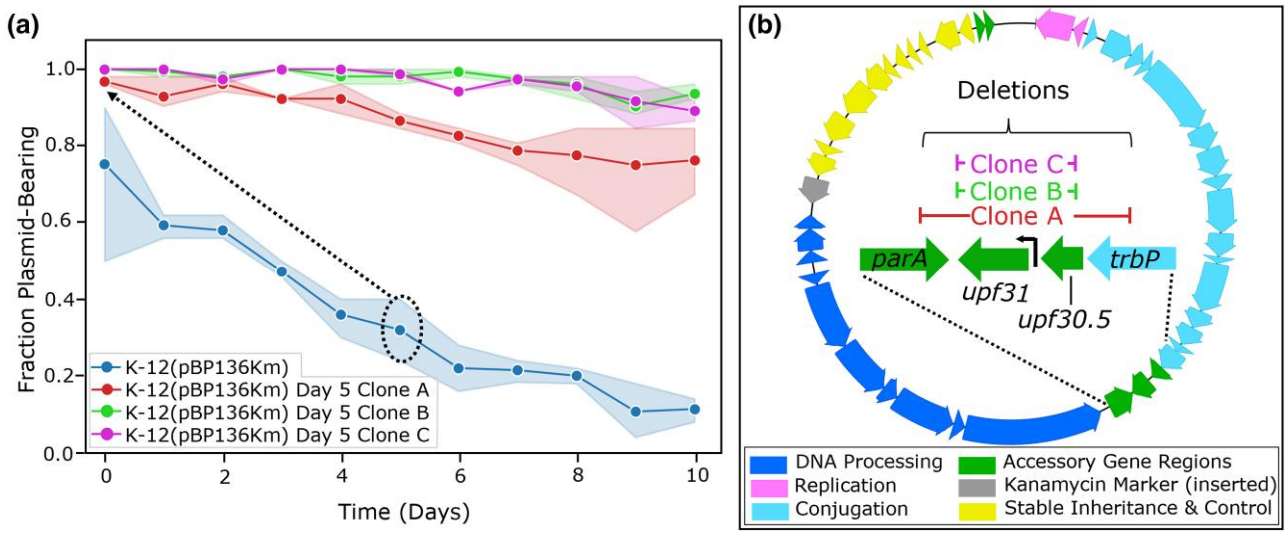

Fig. 1. Plasmid evolution rapidly improved pBP136Km persistence. a) Persistence of the ancestral plasmid pBP136Km (blue) and three populations (red, green, magenta) inoculated with clones that were isolated on Day 5 of the ancestral plasmid assay (see dotted ellipse and arrow). Clones isolated on Day 5 had greatly improved persistence compared with the ancestor. Lighter shades indicate 95% confidence interval. b) Genomic map of pBP136Km in evolved day 5 clones revealed deletions in the accessory region (green) between the *trb-tra* operons (light and dark blue, respectively).

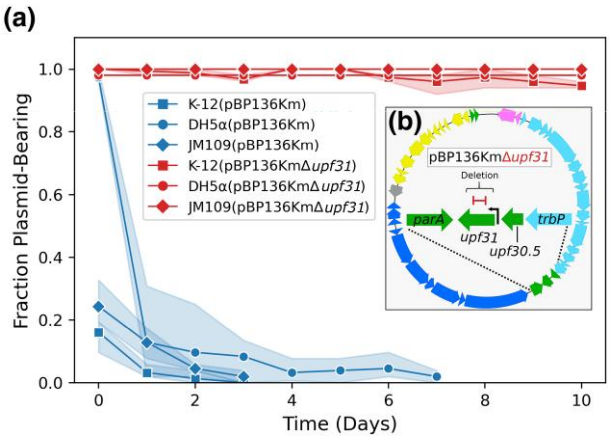

Fig. 2. *Upf31* inactivation alone explained improved persistence in *E. coli* hosts. a) The poor persistence of plasmid pBP136Km (blue lines) in *E. coli* strains K-12 MG1655, DH5α, and JM109 (square, circle, and diamond respectively) was markedly improved after an internal 168-bp deletion in plasmid gene *upf31* (red lines). Lighter shades indicate 95% confidence interval. b) A map of pBP136KmΔ*upf31* showing the 168 bp deletion (red) within the coding sequence of *upf31*.

function must be responsible for the very poor persistence of pBP136 (compare red vs. blue squares of Fig. 2a). To confirm that this result was not an exception in *E. coli* K-12 MG1655, we also compared plasmid persistence in *E. coli* K-12 derivatives DH5α and JM109. Despite being considered good hosts for cloning vectors, these two strains also showed poor plasmid persistence and a drastic improvement after inactivation of *upf31* (Fig. 2a). Importantly, the clone containing pBP136KmΔ*upf31* was isolated from an overnight mating followed by plating for transconjugants on selective media (supplementary table S1, Supplementary Material online). This provides an example of very rapid plasmid evolution that results in remarkably improved plasmid persistence (i.e. <48 h, see Hall et al. 2020). We hereafter refer to the negative effect of Upf31 on plasmid persistence as its "destabilization phenotype."

Host Fitness was Lowered in the Presence of *upf31* and pBP136Km

To investigate whether the improved plasmid persistence associated with *upf31* inactivation was due to amelioration of plasmid fitness costs, we employed two approaches: (i) a comparison of the growth dynamics between the two plasmid–host pairs, with a focus on the maximum growth rate and carrying capacity, and (ii) competition assays involving plasmid-bearing and plasmid-free strains. First, we observed that K-12-containing pBP136KmΔ*upf31* had a significantly higher maximum growth rate and reached a higher final density in comparison with K-12-containing ancestral pBP136Km (Fig. 3a). These results demonstrate that the presence of *upf31* encoded on pBP136Km imposes a substantial fitness cost on its host.

Next, we confirmed the results of the growth rate observations by estimating the cost of the ancestral and evolved plasmid in competition assays. This was done by individually competing each plasmid-bearing strain against the plasmid-free K-12. We note that direct competition assays between identical hosts with and without a highly transmissible conjugative plasmid, such as pBP136Km, can be confounded by plasmid transfer during the assays. While IncP-1 plasmids are less efficient at transferring in liquid than on surfaces, they have been shown to transfer at detectable rates (Zhong et al. 2010). To overcome this limitation, we developed a novel low-density competition assay which allows competing two strains at densities too low for appreciable plasmid transfer to occur (see Materials and Methods). Indeed, as conjugation requires cell contact, decreasing initial densities results in decreasing cell collisions and thus plasmid transfer events (Kosterlitz et al. 2022). The results (Fig. 3b) show a fitness of K-12 (pBP136Km) relative to K-12 of only 0.51, suggesting the plasmid caused a 49% reduction in fitness. In contrast, the fitness of K-12 (pBP136KmΔ*upf31*) was statistically indistinguishable from that of K-12. We conclude that the persistence differences in Fig. 2 can be attributed to a very large fitness cost conferred by pBP136Km, which was alleviated by the inactivation of one plasmid gene, *upf31*.

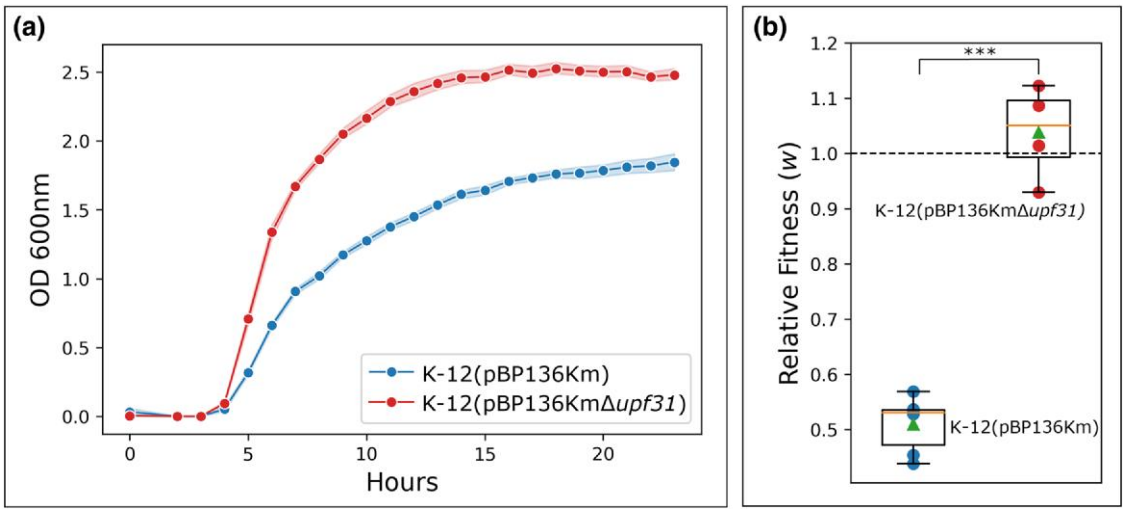

Fig. 3. The presence of *upf31* encoded on pBP136Km lowered host fitness. a) Bacterial growth in batch culture was slower (blue) when carrying pBP136Km with ancestral *upf31* compared with the evolved plasmid pBP136KmΔ*upf31* with inactivated *upf31* (red). Lighter shades indicate 95% confidence interval (two sample independent *t*-test of max growth rates, $n = 10$, $P = 2.35 \times 10^{-3}$). b) With the fitness of plasmid-free K-12 normalized to 1 (dashed line), the relative fitness (w) of *E. coli* with ancestral pBP136Km (in blue) was 0.51, i.e. a ~49% cost (two sample independent *t*-test, $n = 5$, $P = 1.70 \times 10^{-6}$). The relative fitness of evolved genotype K-12 (pBP136KmΔ*upf31*) to plasmid-free K-12 was statistically indistinguishable (two sample independent *t*-test, $n = 5$, $P = 0.40$). Box is interquartile range, green triangle is mean, and orange line is median, $^{***}P \leq 0.001$.

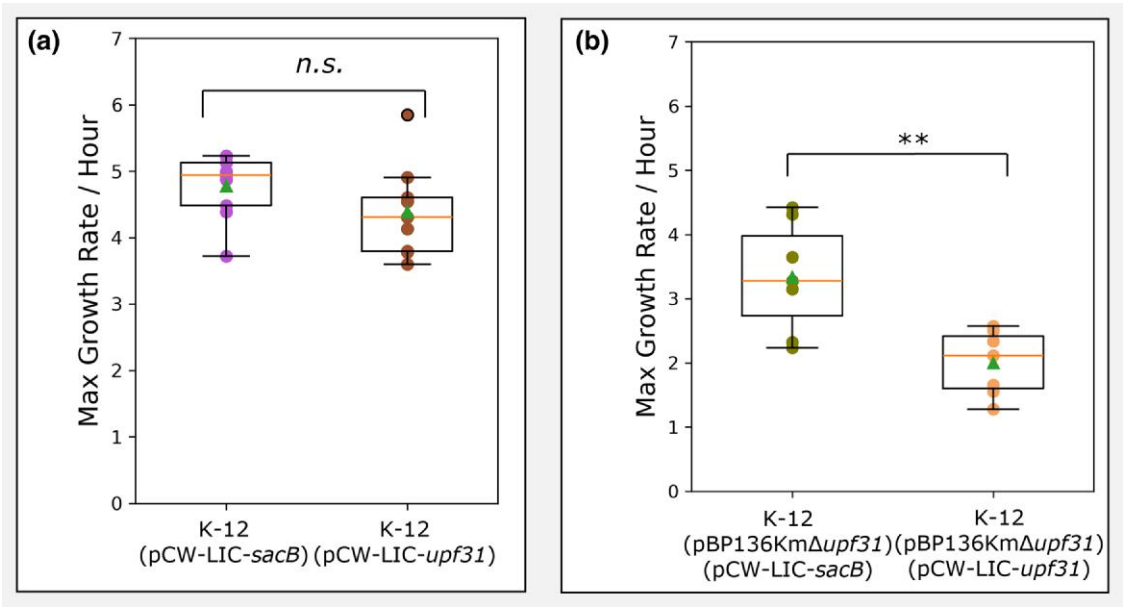

Fig. 4. Upf31 requires the presence of pBP136Km to significantly reduce K-12 growth rate. a) In the absence of plasmid pBP136Km there was no significant difference in maximum growth rate between *E. coli* K-12 expressing control *sacB* (purple) and *upf31* (brown) (two sample independent *t*-test, $n = 10$, $P = 0.46$). b) K-12 (pBP136KmΔ*upf31*) with *upf31* expressed in trans (orange) showed a lower maximum growth rate than K-12 (pBP136KmΔ*upf31*) with control *sacB* expressed in trans (green) (two sample independent *t*-test, $n = 10$, P-value $= 4.23 \times 10^{-3}$). Box is interquartile range, green triangle is mean, and orange line is median, $^{n.s.}P > 0.05$, $^{**}P \leq 0.01$.

The Fitness Cost Imposed by *upf31* Requires the Presence of pBP136Km

We next sought to understand whether the expression of *upf31* directly imposed a fitness cost on K-12 in the absence of plasmid pBP136Km. This was done by expressing *upf31* within K-12 in trans from expression vector pCW-LIC-*upf31*. Production of protein Upf31 was verified via induction of pCW-LIC-*upf31* and observation of the associated 25.4 kDa product (supplementary fig. S1, Supplementary Material online). As a negative control, we used the nonmodified vector pCW-LIC-*sacB*. The maximum growth rates of K-12 with or without the Upf31 protein were not significantly different (Fig. 4a). This suggests that Upf31 does not impose a significant fitness cost on K-12 in the absence of pBP136Km, where it is naturally encoded. In contrast, when *upf31* was expressed in trans in the presence of pBP136KmΔ*upf31*, the maximum growth rate was significantly reduced compared with the same strain with the control vector (Fig. 4b). Thus, we demonstrated that Upf31 did not directly impose a fitness cost on K-12 but rather the interaction of Upf31 with pBP136Km was responsible for the host fitness cost.

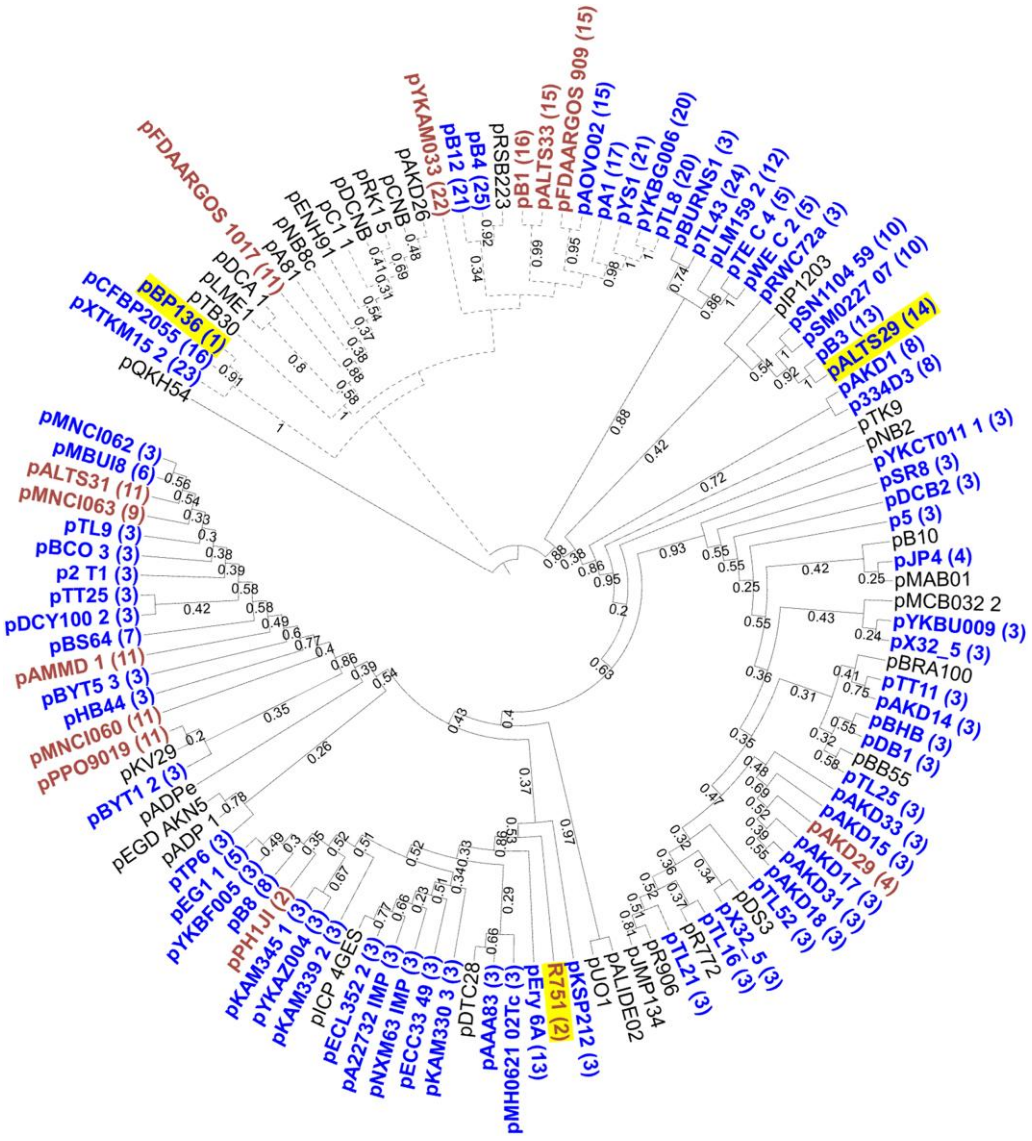

Fig. 5. Phylogeny of 115 IncP-1β plasmids with those containing *upf31* homologs annotated in color (blue for full-length alleles, red for truncated alleles). Plasmids in this study highlighted in yellow. The distinct *upf31* allele numbers are shown in parentheses after the plasmid name. Dashed branches are IncP-1βI, solid branches are IncP-1βII, bootstrap values >0.16 are shown. Summary phylogeny based on nine backbone genes (see Materials and Methods). Some IncP-1β plasmids are not shown as they have missing or incomplete backbone genes (supplementary data S1, Supplementary Material online). The IncP-1γ plasmid pQKH54 was used as the outgroup.

Homologs of *upf31* are Mostly Encoded on Plasmids of the IncP-1β Subgroup

Given the *upf31*-associated destabilization phenotype observed with pBP136Km, we next searched for *upf31* homologs in the ~60,000 plasmids of the PLSDB database (see Materials and Methods). We found 99 plasmids that contained alleles showing sequence similarity to *upf31*/pBP136. Most of these plasmids (94) belong to the incompatibility group IncP-1β, and *upf31* homologs are present in ~80% of all sequenced IncP-1β plasmids (bold in Fig. 5). Strikingly, no *upf31* homologs were found in any of the other 13 IncP-1 subgroups (α, γ, δ, ϵ, ζ, η, θ, ι, κ, λ, μ, ο, ρ), with a single exception of one IncP-1ι plasmid. Additionally, *upf31* homologs were encoded on two putative IncF plasmids and three unclassified plasmids (supplementary data S1, Supplementary Material online). Of the 99 plasmids with *upf31* matches, 73 encode full-length homologs with a CDS of the exact same length (675 bp) as *upf31* of plasmid pBP136 (hereafter named *upf31*/pBP136). The remaining 26 plasmids show similarity to just part of full-length *upf31*/pBP136, such as those with apparent 3′-end truncations, partial similarity within a larger CDS, etc.

Among the 99 plasmids, we distinguished 25 distinct *upf31* alleles, generally numbered in order of sequence similarity to *upf31*/pBP136 (allele numbers are in parentheses in Fig. 5). Seventeen of these 25 alleles share the same length as *upf31*/pBP136 and are translated into 13 unique protein isoforms (see supplementary data S2, Supplementary Material online). In Fig. 5, we show the phylogeny of 115 IncP-1β plasmids, excluding those that have missing or incomplete backbone genes. Of these plasmids shown in Fig. 5, 70 encode full-length *upf31* homologs (blue in Fig. 5) and 13 partial-length homologs (red in Fig. 5). Figure 5 shows that the *upf31* alleles are not conserved within the clades present in the IncP-1β plasmid backbone gene summary phylogeny.

The pBP136 *upf31* Allele Destabilizes Other IncP-1β Plasmids

The Upf31 homologs encoded by two IncP-1β plasmids were tested for their ability to destabilize their respective plasmids, as seen with pBP136Km. First, the archetype IncP-1β plasmid R751 (Thorsted et al. 1998) encodes a *upf31* homolog with a 156 bp deletion at the 3′ end compared with pBP136Km. The high persistence of R751 in *E. coli* K-12 suggests that this 3′-truncated allele does not negatively affect *E. coli* fitness (supplementary fig. S2, Supplementary Material online). We tested whether *upf31*/pBP136 expressed in trans would alter the persistence dynamics of R751 in *E. coli*, i.e. K-12 (R751). Using the same pCW-LIC-*upf31* expression vector as above, K-12 (R751) (pCW-LIC-*upf31*) dynamics were compared with control K-12 (R751; pCW-LIC-*sacB*). Plasmid R751 showed lower persistence and resulted in a statistically significant lower host growth rate (supplementary fig. S2, Supplementary Material online) when *upf31*/pBP136 was present. This suggests that the longer *upf31* allele of pBP136, whose protein product includes a longer C-terminus, has a negative effect on R751 cost and persistence.

We also tested the *upf31* homolog encoded on pALTS29, a multidrug resistance IncP-1β plasmid isolated from the biosolids of a wastewater treatment plant (Law et al. 2021; supplementary fig. S3, Supplementary Material online). A persistence assay of K-12(pBP136KmΔ*upf31*) (pCW-LIC-*upf31*/pALTS29) showed high persistence (~80% plasmid bearing after 10 d, supplementary fig. S3, Supplementary Material online). This demonstrates that *upf31*/pALTS29 did not destabilize pBP136KmΔ*upf31* as seen for *upf31*/pBP136. Conversely, we tested whether *upf31*/pBP136 would destabilize pALTS29, like it did with R751, and indeed found that K-12 (pALTS29) (pCW-LIC-*upf31*/pBP136) rapidly lost pALTS29 (supplementary fig. S4, Supplementary Material online). One or more of the 13 residues that differ between these two IncP-1β plasmids must thus be critical to Upf31's destabilization phenotype (supplementary data S2, Supplementary Material online). In conclusion, some but not all *upf31* homologs of IncP-1β plasmids cause increased cost of plasmid carriage and poor plasmid persistence in *E. coli* K-12.

The Computationally Predicted Methylase Function of Upf31 was not Supported by Experimental Results

The original pBP136 reference genome included an automated annotation of *upf31* as a DNA methylase, based on protein similarity, as noted in the original annotation of archetype IncP-1β plasmid R751 (Thorsted et al. 1998). To provide further insight into the possible function(s) of Upf31, we performed a bioinformatic analysis using InterProScan (Jones et al. 2014) which found homology to protein family N6 adenine-specific DNA methyltransferase (supplementary data S3, Supplementary Material online). An alignment of Upf31 showed it contains the specific motif order of group β N6 adenine-specific SAM-dependent methylases (Malone et al. 1995; supplementary data S4, Supplementary Material online). To try to determine the specific target recognition region, we also used Phyre2, which predicts secondary and tertiary structure and identifies matches to known structures (Kelley et al. 2015). The output strongly predicted Upf31 to be a DNA adenine methylase (Dam) homolog with 70% of the amino acids modeling with 100% confidence (supplementary data S5, Supplementary Material online), including the putative target recognition region. Dam is a key regulatory methylase in Gammaproteobacteria that methylates the N6 position of adenines in GATC sequences (e.g. GA^{m}TC; Marinus and Løbner-Olesen 2014). In summary, InterProScan, Phyre2, and our alignment all predicted Upf31 to be an N6 adenine methylase, with Phyre2 additionally showing Upf31 may target GATC sequences in similar fashion to the Dam protein.

To test whether Upf31 had Dam-like specificity, we digested genomic DNA extracted from *dam*$^{-}$ strain *E. coli* JM110 with and without pBP136Km and *dam*$^{+}$ K-12, using restriction enzymes that digest only methylated or nonmethylated "GATC" DNA sequences, respectively. The results suggest that Upf31 is not a functional Dam homolog as the restriction profiles of DNA from *dam*$^{-}$ JM110 (pBP136Km) were not the same as those of the *dam*$^{+}$ strain K-12. In particular, the genomic DNA of JM110 (pBP136Km) was cut by *Mob*I (which targets "GATC" sequence) and was *not* cut by *Dpn*I (which targets "GA^{m}TC"). This shows that pBP136 with *upf31* did not restore *dam* "GA^{m}TC" methylation patterns under our experimental conditions (supplementary fig. S5, Supplementary Material online).

Next, to determine whether Upf31 more broadly encoded a methyltransferase, we used the base pair modification detection available via Single Molecule Real-Time (SMRT) sequencing (Flusberg et al. 2010). This method has the benefit of not requiring a priori knowledge of the base pair modification chemistry or target DNA sequence of the putative methylase. The output showed no discernible kinetic patterns associated with methylated base pair incorporation. This suggests that Upf31 does not function as a methylase under our experimental conditions, in contrast to the computational predictions.

The Presence of *upf31* Alters the Expression of Global Plasmid Regulators and Associated Operons

To understand how *upf31* and pBP136Km interact to influence *E. coli* fitness, we next performed RNA-seq on K-12-containing pBP136Km and K-12-containing pBP136KmΔ*upf31*. We considered K-12 (pBP136Km) to be the treatment (presence of the intact gene *upf31*), and K-12 (pBP136KmΔ*upf31*) with the internal 168 bp deletion to be the control (absence of a functional *upf31*). The presence of *upf31* led to >2-fold expression changes in 654 of the ~4,400 chromosomal genes, with 412 of those upregulated and 242 downregulated (adjusted *P*-value <0.05, supplementary data S6, Supplementary Material online). The gene ontology tool DAVID (Dennis et al. 2003) predicted downregulation of cellular pathways related to nitrogen and sugar metabolisms, while biosynthesis of siderophores, ABC transporters, and sulfur metabolism pathways were all upregulated (supplementary data S7, Supplementary Material online).

Most interestingly, K-12 (pBP136Km) showed >2-fold expression increase in 11 of its 46 plasmid genes relative to the *upf31* deletion variant (adjusted *P*-value < 0.05, Fig. 6; supplementary data S6, Supplementary Material online). These genes were primarily part of the stable inheritance and control region made up of regulons controlled by proteins KorA, KorB, and KorC (Fig. 6a in yellow). Strikingly, K-12 (pBP136Km) showed a relative decrease in expression in only one gene, *upf31* itself. This relative comparison of gene expression was made possible by pBP136KmΔ*upf31* not having a complete *upf31* deletion but rather an internal 168 bp

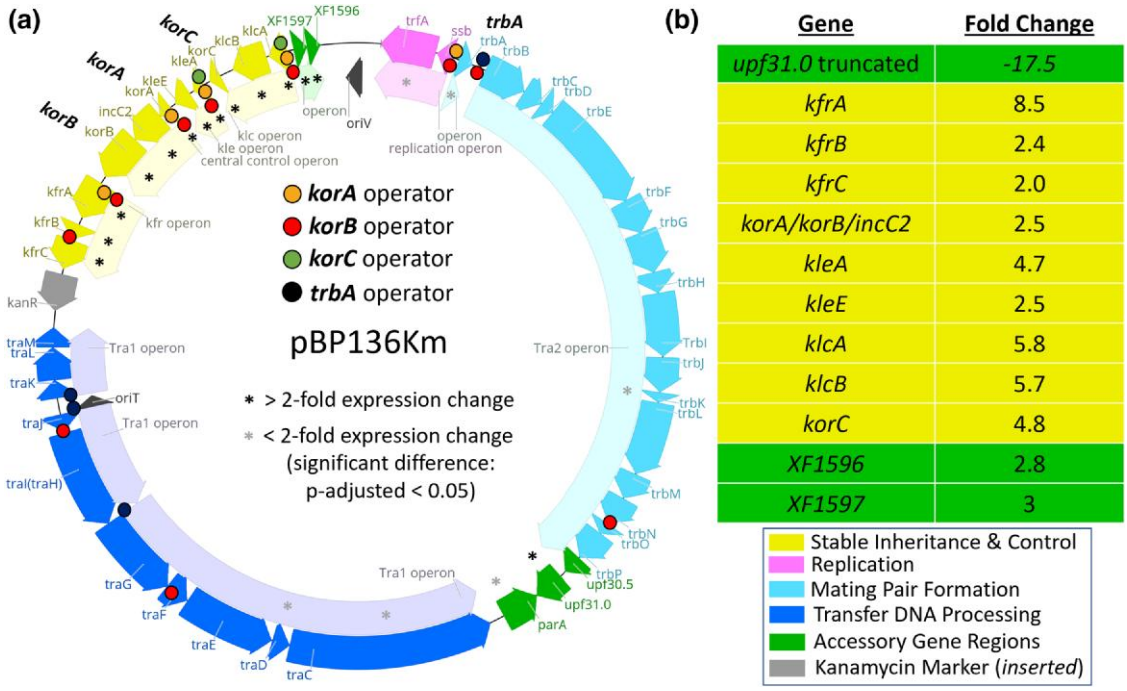

Fig. 6. Expression of global plasmid regulators and their associated operons was increased in the presence of ancestral *upf31* compared with the deletion mutant. a) Genetic map of pBP136Km with operons and operator sites annotated for global regulators KorA, KorB, KorC, and TrbA. Asterisks denote differentially expressed plasmid genes in the presence of *upf31*. b) Table of plasmid-encoded genes with at least 2-fold differential expression (black asterisks in a) in the presence of *upf31*. Note that the expression of the ancestral *upf31* is 17.5-fold lower than that of the truncated *upf31* knockout.

deletion that left an inactivated 507 bp *upf31* product to be expressed (Fig. 2b). The differences in *upf31* gene length were accounted for within the DESeq2 model before calculations of fold change (see Materials and Methods).

The effect of Upf31 to cause at least 2-fold increased expression of all four of the canonical IncP-1 regulatory genes (Adamczyk and Jagura-Burdzy 2003)—*korA*, *korB*, *korC*, and *trbA*—along with the other genes in the operons they regulate (Fig. 6a in bold) is notable. These regulators are known to autogenously repress their own expression (Bingle and Thomas 2001) by binding cognate operators near their respective promoters (orange, red, green, and black dots in Fig. 6a). They are also known to repress expression from cognate promoters within their respective regulons (yellow region in Fig. 6a). The increased repressor gene expression should therefore decrease rather than increase expression of all other regulated operons as the expression of the repressor genes increases. This suggests that Upf31 has a general effect of diminishing the repressive ability of Kor regulatory proteins, both in terms of their autogenous control and repression of their respective regulons.

Upf31 Represses its Own Transcription

A possible mechanistic explanation for the decrease in *upf31* transcription in the presence of Upf31 is that Upf31 represses its own transcription, like the IncP-1 plasmid regulators KorA, KorB, KorC, and TrbA previously studied (Pansegrau et al. 1994). Testing this hypothesis requires first verifying the putative promoter upstream of *upf31* (i.e. upf31p) and then measuring expression from this promoter with and without Upf31 present in the cells. To test the functionality of the putative *upf31* promoter upf31p, a 68 bp region upstream of *upf31* was inserted into *xylE* reporter plasmid pGCMT1 to create pGCMTupf31p (supplementary figs. S6 and S7, Supplementary Material online). The *xylE* assay confirmed that this region possesses significant promoter activity with 7.4 XylE units compared with the 0.09 units for the empty vector (supplementary table S2, Supplementary Material online).

Since Akhtar (2002) had previously failed to demonstrate strong repressor activity from the 3′-end truncated *upf31* in R751, we tested whether adding the C-terminus of pBP136's Upf31 to R751's Upf31 restored the protein into a functional repressor. To do this, we constructed an R751-pBP136 hybrid *upf31* gene (aa 1–172 from R751 and aa 173–225 from pBP136) and inserted this hybrid gene into pBR322 to create pBR*upf31*, with *upf31* transcribed from the *tet* promoter in the vector. Expression from the *upf31* promoter in pGCMTupf31p was then measured in the presence and absence of pBR*upf31*. The full-length hybrid Upf31 expressed from pBR*upf31* caused repression of upf31p, resulting in an average of 0.09 XylE units compared with 4.76 XylE units with pBR322 as negative control (supplementary table S2, Supplementary Material online). This confirmed that *upf31* expression is likely autoregulated by Upf31 binding near upf31p.

Upf31 is Predicted to Bind KorB via a CTD

A hypothesis to explain our transcriptome results emerges from the fact that the high level of repression associated with the stable inheritance and control region depends on the cooperative interaction of KorA (and TrbA) with KorB (Zatyka et al. 1997; Kostelidou et al. 1999; Bingle et al. 2008). If Upf31 were able to disrupt this cooperativity, then it might decrease repression while allowing the concentration of the repressors to rise. To explore the hypothesis that Upf31 disrupts the cooperative interaction of KorA (and TrbA) with KorB, we used AlphaFold2 (Jumper et al. 2021) and ColabFold (Mirdita et al. 2022) to predict the structures of every individual pBP136 coding region in complex with Upf31. We subsequently ranked these interactions using a self-

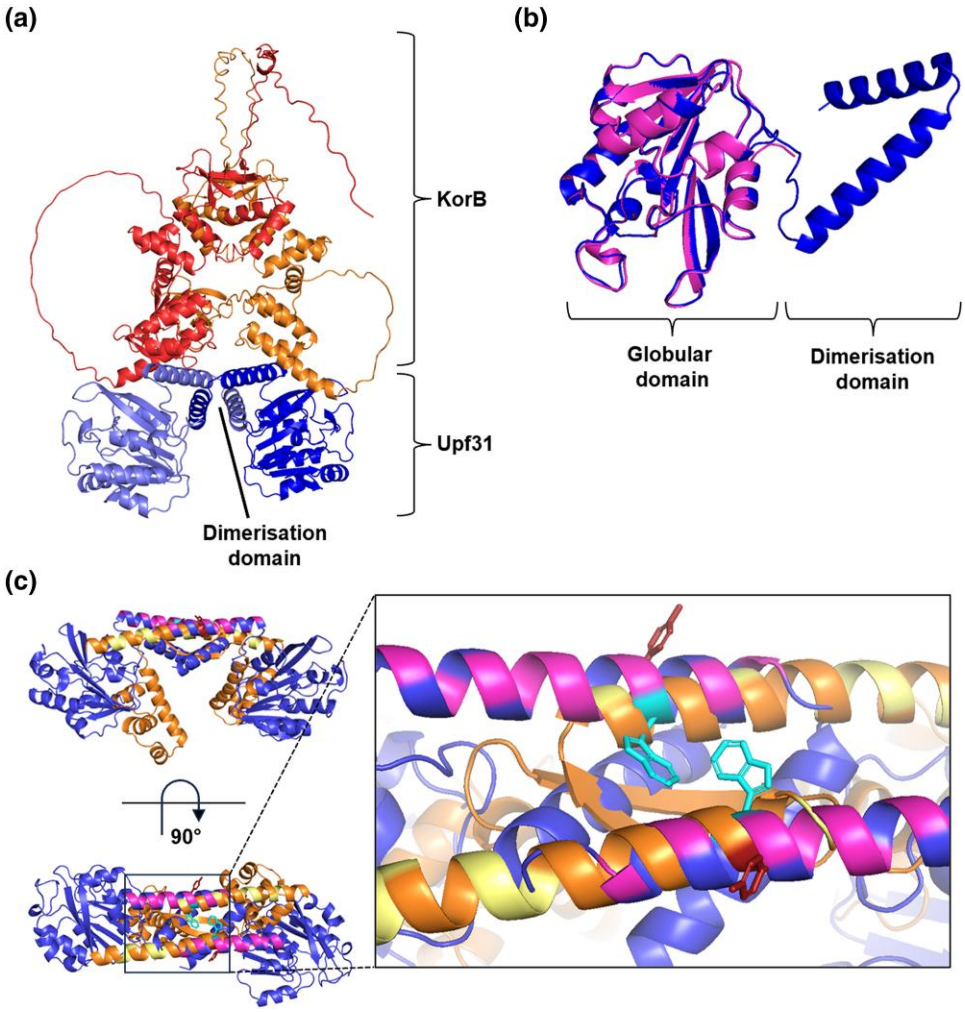

Fig. 7. AlphaFold2 modeling reveals the potential interaction between pBP136 Upf31 and KorB is mediated by the CTD. a) AlphaFold2 modeling of the pBP136 Upf31 homodimer (light and dark blue) with the pBP136 KorB homodimer (red and orange). b) Superimposed AlphaFold2 models of pBP136 Upf31 (dark blue) and R751 Upf31 (magenta) highlighting the additional CTD of pBP136 Upf31. c) Overlay of the AlphaFold2 model of pBP136 Upf31 homodimer (dark blue) with the crystal structure of the RK2 KorA homodimer (PDB: 5CM3, DNA duplex removed). Basic residues in the dimerization domain are highlighted in magenta or yellow. KorA Y84 is highlighted in deep red, Upf31 W198 is highlighted in cyan.

assessment ranking score (ranking_confidence: 0.2pTM + 0.8ipTM, see supplementary data S8, Supplementary Material online). The structure prediction for Upf31 (Fig. 7) is that it possesses a globular N-terminal domain (NTD; residues 1 to ~180) and a CTD (residues ~180 to 225). The NTD encodes a predicted helix-turn-helix motif, which is consistent with DNA binding (as also suggested for KorA previously, Kostelidou and Thomas 2002), and Upf31's autogenous regulation of its own expression as shown earlier. However, the most notable result of this analysis is the structural similarity between the CTD of Upf31 and that of KorA. Modeling of Upf31 with KorB gave a robust ranking_confidence score of 0.64. Removal of the CTD from Upf31 drastically reduced the predicted interaction between KorB and Upf31 to 0.274. When pBP136 KorB was modeled with Upf31/R751, which natively lacks a CTD, a similarly low score of 0.252 was predicted. Given the globular domain sequence similarity between the Upf31 of pBP136Km and R751, these results are highly suggestive of pBP136Km Upf31 interacting with KorB via its CTD (Fig. 7b). Although the Upf31 CTD does not line up with KorA CTD in the way that the CTD of TrbA does, one can detect similarities like the aromatic residue tryptophan protruding at position 198 and the patch of predominantly basic amino acids (Fig. 7c). This predicted interaction of Upf31 with KorB might allow Upf31 to reduce available KorB either when bound to DNA or free, since we previously observed that KorA and KorB could be copurified in the absence of DNA (Bingle et al. 2008). This would in turn explain the increased transcription of the IncP-1 regulatory genes as well as the operons they regulate.

If Upf31 and KorB physically interact, co-evolution of the respective genes *ufp31* and *korB* would be expected. To investigate this further, we generated a co-phylogeny of IncP-1 encoded *upf31* alleles and their respective *korB* alleles to examine whether phylogenetic signals are consistent with co-evolution between these two genes (Fig. 8). Of note are two similarly distinct monophyletic clades in each gene tree with no discordance or paraphyly (e.g. vertical/crisscrossing lines) between the groups. Instead, discordance is only observed between poorly resolved nodes (e.g. short branch lengths) within each respective clade, which lack phylogenetic signal. These observations agree with gene co-evolution and are thus consistent with the Upf31–KorB interaction hypothesis.

The phylogenetic congruence between *upf31* and *korB* could be explained by other factors like shared genetic drift. To examine this, we created similar co-phylogenies of two

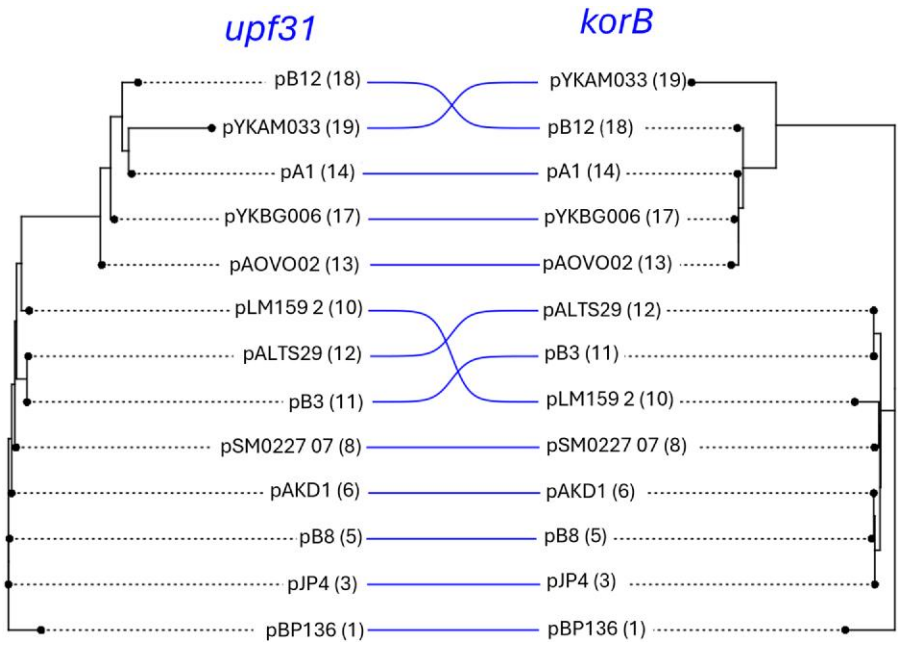

Fig. 8. Co-phylogenies of *upf31* and *korB* homologs. Plasmid representatives from the IncP-1β plasmids that encode the 17 distinct full-length *upf31* alleles are shown on the left with their respective *korB* gene on the right.

controls, *korA-korB* as a positive control (since Bingle et al. [2008] showed their protein products physically interact), and *trfA1-korB* as a negative control (given their protein products do not physically interact). Like *upf31-korB*, both control phylogenies resulted in two distinct monophyletic clades (supplementary fig. S8, Supplementary Material online). Furthermore, both phylogenies showed no discordance between the two main clades. Considering the nonbinding properties of TrfA1-KorB, this result suggests the limitation of this phylogenetic approach to the detection of physically interacting proteins. Alternatively, the tightly regulated interplay of backbone genes may collectively co-select for maintenance of the entire system. In either case, a finding of noncongruence would have been evidence of a lack of co-evolution and thereby lessen the likelihood of physical protein interaction. Instead, our finding of congruence in the gene co-phylogeny of *upf31-korB* does not conflict with, but rather is in agreement with, the Upf31–KorB interaction hypothesis.

The CTD of Upf31 is Essential for the Plasmid Destabilization Phenotype

Additional experiments were performed to test the computational prediction that Upf31 requires a CTD to cause poor plasmid persistence. Plasmid R751 encodes a truncated Upf31 allele lacking a CTD and the plasmid persisted very well in *E. coli* K-12 but was rapidly lost in the presence of Upf31/pBP136Km expressed in trans (supplementary fig. S2, Supplementary Material online). To test whether Upf31/R751 was otherwise functional, aside from the missing CTD, the hybrid Upf31 previously used to test the *upf31* promoter was utilized. This hybrid combines the truncated Upf31/R751 with the CTD region of Upf31/pBP136 and was expressed in trans via vector pBR*upf31*. Strikingly, the transformation of pBR*upf31* into competent *E. coli* C600 (R751) resulted in hundreds of pBR*upf31* transformants on media selecting for just pBR*upf31* but fewer than 10 transformants on media selecting for both pBR*upf31* and R751. This experiment demonstrated that adding the CTD of Upf31/pBP136 to the truncated Upf31/R751 restores the destabilization phenotype.

We next performed experiments to test whether certain residues in the Upf31 CTD might be critical for the destabilization phenotype. It was previously shown that cooperativity between KorB and IncP-1 regulator KorA requires KorA to possess an aromatic residue (typically Y at position 84), which protrudes from the CTD-encoded α-helix structures (Bingle et al. 2008). Since the CTD of KorA and Upf31 share structural similarity (Fig. 7c), and the predicted complex between KorB and Upf31 involves the Upf31 CTD, we scanned this region for aromatic residues (Y, F, and W) that may be essential for cooperativity with KorB, analogous to KorA Y84. This identified a primary Upf31 candidate residue, W198, that is near the Y84 residue KorA when the two proteins are overlaid (Fig. 7c). Two alternative Upf31 candidate aromatic residues in the same vicinity were also identified, F185 and Y180. We individually mutated each of these (Y180A, F185A, and W198A) within the hybrid Upf31 to test whether R751 would still be displaced, as described above for the wild-type (WT) hybrid Upf31f. While pBRupf31 behaved as expected, transformation with any of the three mutant derivatives resulted in similar numbers of colonies on both media (>200), indicating that Upf31's ability to displace R751 was lost in each case.

We also tested whether the three CTD mutations (Y180A, F185A, and W198A) would alter Upf31's function as an autorepressor, using the pGCMTupf31p *xylE* reporter system described above. In contrast to the WT hybrid Upf31, the three mutant proteins did not show repressor function, each showing at least a ~42-fold increase in XylE units relative to the WT hybrid Upf31 (supplementary table S3, Supplementary Material online). Since these mutants each knockout both repression and interference with R751 maintenance, they do not identify a single aromatic residue critical for KorB-binding analogous to KorA's Y84. However, our

computational and experimental findings demonstrate very clearly that the presence of CTD is necessary for the plasmid destabilization phenotype.

Finally, since the R751-pBP136 hybrid we had created differs from the most common Upf31 isoform (Upf31 isoform 3, supplementary data S1, Supplementary Material online) at just two positions (aa174 and aa202), we tested whether these differences would affect the destabilization phenotype. Starting from the WT hybrid in pBRupf31, two additional mutations were created: K174E at the NTD-CTD border and I202L in the CTD, singly and together, the double mutant encoding a Upf31 protein identical to isoform 3. Transformation of the starting WT pBRupf31 alongside the two single mutants and the double mutant into *E. coli* C600 (R751) showed the destabilization phenotype remaining (supplementary data S9, Supplementary Material online). Furthermore, pGCMTupf31p *xylE* reporter gene tests showed full retention of repressor activity by the three derivatives (supplementary table S4, Supplementary Material online). Since Upf31 isoform 3 is encoded on 50 (~44%) of the sequenced IncP-1β plasmids (Fig. 5) by alleles 3 and 5 (which differ by synonymous SNP C456T), these results indicate that the observations reported here likely apply to many of the IncP-1β plasmids.

Discussion

This work sheds new light on the biology and evolution of an intensively studied conjugative plasmid group, the broad-host-range IncP-1 plasmids. We demonstrate that the poor persistence in *E. coli* of pBP136, a conjugative IncP-1β plasmid from *B. pertussis*, could rapidly improve even in the absence of selection for the plasmid. The associated drastic amelioration of the plasmid fitness cost to *E. coli* was due to the inactivation of a previously uncharacterized plasmid gene, *upf31*, which affected the expression of the plasmid backbone regulatory circuit including all four major regulators as well as hundreds of chromosomal genes. This change in transcriptome may have been initiated by Upf31 interacting with one of these major plasmid regulators, KorB. Plasmids of the incompatibility group IncP-1 are notably persistent in many Proteobacteria, and those from the IncP-1β subgroup are particularly diverse and widespread (Schmidhauser and Helinski 1985; De Gelder et al. 2007; Shintani et al. 2010; Yano et al. 2012, 2013; Jain and Srivastava 2013; Klümper et al. 2015; Kottara et al. 2018; Hayakawa et al. 2022). However, we show here that different alleles of plasmid genes like *upf31* can cause striking differences in plasmid persistence in particular hosts. Understanding how this gene acts and why it may be an advantage in certain hosts will expand our understanding of the genetic toolbox plasmids can exploit for their success.

The last decade has seen a flood of seminal studies regarding compensatory evolution of plasmid fitness cost, including multiple studies showing that changes in chromosomal regulatory systems can improve plasmid–host relations (Brockhurst and Harrison 2022). One example is mutations of the *fur* transcriptional regulator involved in iron uptake (Troxell and Hassan 2013; Stalder et al. 2017), while the evolution of the chromosomal *gacA/S* regulators (Harrison et al. 2015) was observed to lower the fitness cost of plasmid carriage within 48 h of plating on agar (Hall et al. 2020). A few studies have found compensatory evolution via putative chromosomally encoded helicases (Loftie-Eaton et al. 2017), in one case largely restoring gene expression levels of the plasmid–host pair to that of the plasmid-free bacteria (San Millan et al. 2015). Amid these discoveries, the complex and tightly controlled plasmid transcriptional regulatory systems have perhaps been overlooked as a key factor in plasmid–host relations. One other study showed that a compensatory mutation in a plasmid-encoded DNA-binding protein resulted in increased expression of the plasmid's *parAB* operon, which in turn ameliorated the plasmid's fitness cost by an unknown mechanism (Hall et al. 2021). Similarly to our findings, this compensatory mutation only showed a fitness effect on the host in the presence of the plasmid and not in its absence, consistent with upregulation of plasmid-encoded genes. However, unlike here, there was no general shift in the expression of genes in the plasmid's stable inheritance and control region. Our work shows that the plasmid-encoded Upf31 inactivated during evolution in *E. coli* regulates its own transcription as well as that of the primary plasmid regulators associated with the stable inheritance and control region. Thus, fine-tuning the transcriptional regulatory systems of *both* plasmids and chromosomes should be included in the growing list of evolutionary pathways that can decrease the fitness cost of a plasmid to its new host.

Critical IncP-1 plasmid functions like replication, partitioning, mating pair formation, and conjugative DNA transfer are controlled by a complex genetic regulatory circuit with four repressors: KorA, KorB, KorC, and TrbA. Of these, KorA and TrbA have been shown to interact with KorB via their CTDs (Kostelidou et al. 1999; Kostelidou and Thomas 2000; Zatyka et al. 2001; Bingle et al. 2008; McLean et al. 2025). We hypothesize that the CTD of Upf31 also interacts with KorB, as predicted by the AlphaFold2 modeling and shown experimentally via mutations in Upf31 that restore stability of IncP-1 plasmids. Moreover, the high concordance of the *upf31* and *korB* gene phylogenies is consistent with coevolution through physical protein interaction (Fig. 8). This Upf31–KorB association may override the other KorB regulator interactions with concomitant derepression of gene expression, which is associated with poor *E. coli* growth.

In addition to the predicted interaction with KorB, the Upf31 NTD appeared to be essential for Upf31 autorepression, likely by binding to DNA in the *upf31* promoter region. Dimerization of monomers containing helix-turn-helix motifs, like that of the Upf31 NTD, is characteristic for binding DNA (Brennan and Matthews 1989). Similarly, WT KorA is a dimer when it binds to DNA and when interacting with KorB (McLean et al. 2025). It therefore appears that three of the mutations tested in the Upf31 CTD disrupted Upf31 dimerization, which would account for both the loss of Upf31 autorepression and the plasmid destabilization that may be due to Upf31–KorB interaction. To test for Upf31–KorB interaction, future studies should take biochemical analyses further, including protein pull-down experiments with purified proteins. Additionally, mutational screenings of Upf31's CTD could identify residue(s) that are essential for KorB interaction if they (i) induce a destabilization phenotype but (ii) retain Upf31 dimerization as manifested by high repression of the *upf31* promoter (i.e. ability to bind DNA).

We showed three protein isoforms that possess a destabilization phenotype: (i) full-length Upf31/pBP136, (ii) the hybrid Upf31 that combines the naturally truncated Upf31/R751 with the CTD of Upf31/pBP136, and (iii) Upf31 isoform 3, which is encoded on ~44% of currently sequenced IncP-1β plasmids. Given that the CTD of protein isoform 3 is identical to nine other full-length Upf31 isoforms (supplementary data

S2, Supplementary Material online), it seems likely that well over half of all IncP-1β plasmids encode a Upf31 protein with a destabilizing phenotype like Upf31/pBP136 in this study. Thus, the majority of IncP-1β plasmids are likely poorly maintained in several *E. coli* strains, and the truncation observed in R751 and other mutations in IncP-1β plasmids may represent an adaptation to hosts like *E. coli*. While we demonstrated the necessity of the CTD region for the destabilization phenotype, additional data showed that mutations in the DNA-binding NTD can also disrupt the phenotype. This is best demonstrated by the stability observed with Upf31/pALTS29 (supplementary fig. S3, Supplementary Material online), whose amino acid sequence differs from Upf31 isoform 3 by a single L78R change. Further studies are warranted to investigate the remaining protein isoforms, which will aid in identifying residues critical for the destabilization phenotype.

The Upf31-associated derepression of global plasmid repressors triggers a secondary effect of large-scale differential expression of both plasmid and chromosomal genes. Of note is that four of the plasmid-encoded genes that were overexpressed (*kleA*, *kleE*, *klcA*, and *klcB*) were previously shown to be harmful to *E. coli* growth when not repressed by the global Kor protein regulators (Figurski et al. 1982; Kornacki et al. 1993; Larsen and Figurski 1994). Furthermore, we also observed a 2.5-fold higher expression of *incC2* and *KorB* with the WT *upf31*. A previous study demonstrated simultaneous overexpression of these IncP-1 plasmid partitioning genes caused toxicity and poor *E. coli* growth (Rosche et al. 2000). Finally, the differential expression of many chromosomal genes in the presence of Upf31 suggests direct or indirect plasmid-chromosome "cross-talk," as shown in other studies (Vial and Hommais 2020; Hall et al. 2021; Marincola et al. 2021; Thompson et al. 2023). We have shown that Upf31 only exerted a significant negative fitness effect on *E. coli* K-12 when plasmid pBP136 was also present, and not in its absence. This suggests that Upf31 does not have a direct effect on the chromosome transcriptome with measurable fitness consequences. However, it must have an indirect effect on the chromosomal transcriptome, likely through its effect on the plasmid regulators. Whether the observed poor growth of *E. coli* in our study is caused by differential expression of particular gene(s) from the plasmid, chromosome, or both, is a question for future studies.

Upf31 might be an important clue to understanding the four groups of conserved inverted repeats (IR: 5′-CAGCATCG-3′) that were previously described (Thorsted et al. 1998) in IncP-1β plasmids but without any explanation of purpose or function. The *upf31* gene itself is embedded in one of the four IR regions and appears to autoregulate its own expression by binding to one of the IR regions, as it is the most obvious operator-like sequence in the promoter region. The striking conservation of these IRs across IncP-1β plasmids, coupled with the nearly exclusive presence of *upf31* in this subgroup, suggests that Upf31 may bind similar sequences across IncP-1β plasmids. Notably, *klcB* of pBP136 is one of the most highly overexpressed genes in the presence of Upf31 (Fig. 6b) and contains three of these IRs within its coding sequence. If Upf31 binds to these IRs, it could result in a local concentration of the protein, which would reduce KorB availability and accentuate expression.

We also report two important lessons learned during this study. First, we frequently observed rapid loss of *upf31* during routine laboratory cultivation due to various deletions (supplementary table S1, Supplementary Material online), prompting us to always sequence the region of *upf31* to confirm its presence. This suggests that rapid evolution of costly plasmid genes may often go unrecognized during cultivation and plasmid transfer between strains by matings and electroporation. Given that many IncP-1β plasmids were transferred into *E. coli* to generate sufficient high-quality plasmid DNA for sequencing (Heuer et al. 2004; Schlüter et al. 2005; Sen et al. 2011; Brown et al. 2013), it is possible that some of the *upf31* genes mutated during those steps and thus no longer reflect the natural diversity of *upf31* genes (Fig. 5). Future work should investigate how even brief cultivation may select for mutations in plasmids that may distort our understanding of their original host range. These findings also align with previously shown rapid (<48 h) compensatory evolution in plasmid–host pairs, even without known selection for the plasmid (Hall et al. 2020). Second, our work showed that Upf31 is likely not a functional DNA methylase as suggested by automated annotations in many plasmid genomes. Even with the experiments we conducted, the possibility that Upf31 methylates still exists, for example, if the methylation would target RNA molecules. However, it also highlights the argument that plasmid genes *without* experimental data on their function should be assigned locus tags rather than putative function-related gene names (Thomas et al. 2017).

Despite decades of great progress in plasmid biology and the pivotal role of conjugative plasmids in antibiotic resistance dissemination, a substantial portion of plasmids harbor multiple genes with unknown functions. Our study highlights the importance of one such gene, which, when inactivated, led to a lasting host–plasmid pair, effectively rescuing a plasmid from potential extinction in a bacterial population (Gomulkiewicz and Holt 1995). The investigation of this specific gene revealed that this evolutionary rescue is closely linked to alterations in the plasmids' regulatory circuit. These findings underscore the significance of identifying and experimentally validating uncharacterized plasmid genes to understand how evolution underwrites the spread and persistence of plasmids in bacterial communities.

Materials and Methods

Bacterial Strains, Plasmids, and Media

IncP-1β plasmid pBP136 was discovered in a strain of *B. pertussis* isolated from a lethal case of infant whooping cough (Kamachi et al. 2006). A kanamycin gene was later inserted to generate pBP136Km (NCBI accession number NZ_OR146256.1), which provided a selectable marker in the originally cryptic plasmid (Sota et al. 2007). The expression vector pCW-LIC-*upf31* was derived from pCW-LIC-*sacB* as described below in the cloning section.

Escherichia coli K-12 MG1655 (F^-, λ^-, *ilvG*, *rfb*-50, *rph*-1) or "K-12" was derived from an isolate within the stool of a diphtheria patient in 1925 (Bachmann 1972), and *E. coli* JM109 and DH5α (Anton and Raleigh 2016) are later derivatives with useful attributes relevant to cloning (*end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17 [r_k^-, m_k^+], *rel*A1, *sup*E44, Δ(*lac-pro*AB), [*F′ tra*D36, *pro*AB, *lac*I^qZΔM15] for JM109, and F^- ϕ80*lacZ*ΔM15 Δ(*lacZ*YA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 λ^-*thi*-1 *gyr*A96 *rel*A1 for DH5α). *Escherichia coli* C600 is also a K-12 derivative (*thr*-1, *leuB*6, *thi*-1, *lacY*1, *glnV*44, *rfbD*1, and *fhuA*21). We generated rifampicin and nalidixic resistant clones of *E. coli*

K-12 MG1655 by recovering resistant mutants after plating on agar with the respective antibiotic. New constructs and initial, intermediate, and final populations of the plasmid persistence assays were all archived at −70 °C in 30% glycerol.

All bacteria in this study were grown at 37 °C in lysogeny broth (LB) shaken at 200 rpm or on LB agar (LBA) plates. Strains were grown with 50 μg/mL kanamycin (km) or 100 μg/mL ampicillin (amp) when appropriate for plasmid maintenance. Expression vectors were induced with 500 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) when induction is mentioned in the text.

Plasmid Persistence Assay

All plasmid persistence assays were performed in triplicate and first grown overnight (O/N) with kanamycin to select for initial plasmid maintenance (time T0). Daily serial transfers of 4.9 μL of overnight culture into 5 mL of sterile broth *without* antibiotics were then performed for 10 more days (times T1 to T10). Each day the cultures were diluted 10^{-6} in 1× phosphate-buffered saline and 100 μL was spread onto dilution plates. The daily plasmid-containing fraction of the population was determined by replica plating 52 randomly chosen colonies from the dilution plates onto LB plates with and without kanamycin.

Growth Assays

We performed growth assays in batch culture to measure the effect of Upf31 on growth rates and final densities. Strains were grown from freezer archives for 24 h and diluted 1:100, then grown another 24 h and finally diluted 1:1,000 in sterile media. From these ten biological replicates each were loaded at a volume of 200 μL into the respective wells of a 96-well plate. A SPECTROstar Nano plate reader measured optical density at 600 nm every 10 min for 23 h with incubation at 37 °C and 500 rpm orbital shaking. The R package GrowthRates (Hall et al. 2014) v. 0.8.4 was used to calculate maximum growth rate using default parameters.

Low-Density Competition Assays

The low-density competition assay was developed in this study to avoid plasmid transfer during the assay, as plasmid pBP136 transfers very efficiently, with almost all recipients containing the plasmid after typical overnight liquid matings. It contains two distinct steps: (i) determining the likely "conjugation-free" time window for specific low donor and recipient densities, and (ii) performing a traditional competition assay at low densities within the established time window.

To find the likely window of time before density is high enough for conjugation to occur, we used K-12 MG1655 strains with spontaneous mutations conferring resistance to nalidixic acid (Nal) and rifampicin (Rif) (donor K-12N(pBP136Km) and recipient K-12R, respectively). Donor and recipient were mixed 1:1 at densities of 10^2/mL into 2.5 mL LB shaken 200 rpm at 37 °C. To select for possible transconjugants, the cultures were mixed hourly with another 2.5 mL of LB with Km and Rif to create a 5 mL mixture with 50 μg/mL Km and 50 μg/mL Rif. These were grown 24 h after which the earliest time point showing turbidity was associated with the emergence of transconjugants K-12R(pBP136Km). The first turbid culture emerged after 6 h, and therefore a 5 h window of likely conjugation-free growth was chosen for the subsequent competition assays.

Competition assays were then performed using K-12 versus K-12(pBP136Km) and K-12 versus K-12(pBP136KmΔ*upf31*). Each experiment began with 1:1 mixtures at densities of 10^2/mL in 5 mL LB without antibiotics in shaken test tubes. Plating onto LBA occurred at the start time and after 5 h (after appropriate dilutions), and the relative increase in cell numbers (cfu/mL) was used as a proxy for relative fitness as follows (Wiser and Lenski 2015):

$$w = \frac{\ln\left(\frac{A_f}{A_i}\right)}{\ln\left(\frac{B_f}{B_i}\right)}$$

where w is the Malthusian relative fitness term, A_i and A_f are the initial and final plasmid-containing K-12 populations, respectively, and B_i and B_f are the initial and final plasmid-free K-12 populations, respectively.

Identification of *upf31* Homologs and Phylogenetic Analysis

Plasmid-encoded homologs of *upf31* were identified by a local BLASTN search of the plasmid database PLSDB (Galata et al. 2019) and compiled into a spreadsheet (supplementary data S2, Supplementary Material online). The pBP136/*upf31* sequence and the immediate upstream promoter region was queried against PLSDB version 2023_11_03_v2 with percent identity (perc_identity) > 60%, qcov_hsp_perc > 12% (i.e. ~100 bp), and $e < 0.0001$. The resulting matches were aligned with MAFFT v7.505 using default values (Katoh and Standley 2013).

A phylogeny of 115 IncP-1β plasmids was constructed using nine IncP-1 backbone genes (kleE, korA, korB, traG, traI, trbA, trbC, trbG, and trfA1). These genes were previously identified as good candidates for inferring vertical history (Sen et al. 2013). The IncP-1β plasmids were identified by the PLSDB plasmid database and a literature search. We included the IncP-1γ plasmid pQKH54 as an outgroup (Sen et al. 2013). We inferred vertical evolutionary history for each backbone gene by estimating maximum-likelihood phylogenies using iqtree v2.2.5 (Minh et al. 2020). Models of sequence evolution were identified under Bayesian information criteria using ModelFinder (Kalyaanamoorthy et al. 2017) as implemented within iqtree v2.2.5. Gene phylogenies were estimated with 1,000 ultra-fast bootstrap replicates (Hoang et al. 2018). We then estimated a summary lineage tree of the nine backbone gene phylogenies using ASTRAL-III v5.7.3 (Zhang et al. 2018). The resulting lineage tree was rooted on IncP-1γ as outgroup. The information regarding encoded homologs of *upf31* described above (supplementary data S1, Supplementary Material online) was then overlayed on the ASTRAL lineage tree using Interactive Tree Of Life (iTOL) v6.9.1 (Letunic and Bork 2024).

The *ufp31-korB* co-phylogeny was produced by identifying 13 unique *upf31* alleles in the BLASTN output (see above) and obtaining the cognate *korB*. Only alleles ±3 bp in length from the 675 bp *upf31*/pBP136 were utilized. Multisequence alignments for each respective gene were produced for each unique allele with MAFFT v7.505. Respective phylogenies were then produced with iqtree v2.2.5 using the same protocol described above and visualized across from each other using the cophylo function within phylotools v2.3-0 (Revell 2024). The *korA-korB* and *trfA1-korB* co-phylogenies were produced in a similar manner.

Methylation Detection

To test whether *upf1* encodes a methyltransferase as computationally predicted, we utilized the base modification analysis provided by the Pacbio Sequel II DNA sequencer. The methylase-free *E. coli* strain ER2796 (Anton et al. 2015) with expression vector pCW-LIC-*upf31* was tested against ER2796 (pCW-LIC-*sacB*) for differential base modification (Blow et al. 2016) on a Pacbio Sequel II at the Arizona Genomic Institute in Tucson, AZ, USA. Both strains were grown with antibiotic selection and 500 μM IPTG induction before genomic extraction using the Sigma GeneElute Bacterial kit. Base Modification Analysis was run on SMRT Link v. 10.2.0.133434 with a reference made by merging the *E. coli* assembly with the respective plasmids and enabling "Find Modified Base Motifs" and "Consolidate Mapped BAMs for IGV" options. All the other options were left as default.

RNA-Seq

To understand how *upf31* and pBP136Km interact to influence *E. coli* fitness, we performed RNA-seq. Strains *E. coli* K-12 MG1655 (pBP136Km) and *E. coli* K-12 MG1655 (pBP136KmΔ*upf31*) were grown overnight in 5 mL LB with Km and centrifuged. The cell pellets were sent to Zymo Research for Total RNA-Seq Service. Sequencing libraries were constructed from total RNA samples and were prepared using the Zymo-Seq RiboFree Total RNA Library Prep Kit (cat. #: R3000), according to the manufacturer's instructions (https://www.zymoresearch.com/products/zymo-seq-ribofree-total-rna-library-kit). RNA-Seq libraries were sequenced on an Illumina NovaSeq to a sequencing depth of at least 30 million read pairs (2 × 150 bp reads) per sample.

A custom bioinformatics pipeline began with TrimGalore! v. 0.67 to remove adapters and low-quality reads. Bowtie2 (Langmead and Salzberg 2012) v. 2.4.5 then mapped reads to merged chromosomal and plasmid references. These were counted using featureCounts (Liao et al. 2014) v. 2.0.1, the output of which was entered into DESeq2 (Love et al. 2014) v. 1.38.1, for normalization of gene length before final calculations of relative expression.

DNA Sequencing

Whole genome sequencing was performed by SeqCenter, LLC in Pittsburg, PA, USA. Genomic DNA extractions were performed using Sigma GeneElute Bacterial kit. The libraries were prepared using the Illumina DNA Prep kit with IDT 10 bp UDI indices and sequenced on an Illumina NextSeq 2000 with 2 × 150 bp reads. Demultiplexing, quality control and adapter trimming was initially performed using Illumina's bcl-convert v. 3.9.3. TrimGalore! v. 0.6 was additionally used to remove remaining adapters and low-quality reads. Breseq (Barrick et al. 2014) v. 0.36.0 was used to identify mutations between ancestral and evolved strains.

Cloning

pCW-LIC (hereafter pCW-LIC-*sacB*) was a gift from Cheryl Arrowsmith (Addgene plasmid #26098; http://n2t.net/addgene:26098; RRID:Addgene26098). Plasmid pCW-LIC-*sacB* was double cut at the *Nde*I and *Hin*dIII restrictions sites, which created a 4,962 bp linearized backbone and a 2,302 bp linear fragment containing *sacB* and its promoter. These were separated and the linearized backbone recovered from an agarose gel using a Thermo Scientific (TS) GeneJET Gel Extraction Kit. Next, gene *upf31* was amplified from pBP136Km miniprep using upstream primer 5′-*GGTGGT*CATATGTCCAGGAAGAAGGCCAT GAG-3′ and downstream primer 5′-*GGTGG*TAAGC TTCTACTCGGCCGCCTCTAG-3′ (flanking sequences for restriction enzymes in italics, *Nhe*I and *Hin*dIII sites underlined, respectively). The ends of the *upf31* amplicon were then double digested at the *Nde*I and *Hin*dIII sites, and the small digested ends removed using a TS GeneJET PCR Purification Kit. The *upf31* segment and linearized pCW-LIC backbone were joined using T4 DNA ligase and the now-circularized pCW-LIC-*upf31* was electroporated into DH5α using standard methods (Sambrook and Russell 2001).

To construct the upf31p-*xylE* reporter gene, fusion complementary 68-nt oligomers were designed to create a double stranded fragment with *Xba*I and *Eco*RI sticky ends for insertion into *xylE* reporter plasmid pGCMT1 to create pGCMT-upf31p as shown in supplementary fig. S6, Supplementary Material online. The 68 bp region upstream of *upf31* was from plasmid R751.

Construction of the R751-pBP136 hybrid *upf31* was done in three steps. The ribosome-binding site plus codons 1 to 172 were amplified by PCR with primers 5′-TGCAA GCTTTAATGCGGTAGCCAAGTCCCGATTTACTCCAG-3′ and 5′-CCGGCTCGGGGTAGTTCATC-3′ from R751 template. Codons 173 to 224 were made by annealing and extension using overlapping synthetic oligomers (104 and 117 nt), designed from the sequence of pBP136. These two segments were then joined and amplified by Spliced Overlap Extension PCR (Horton et al. 1990) with primers 5′-TGCAAGCTTTAA TGCGGTAGCCAAG-3′ and 5′-TCGGTCGACGCAGGC GTGAC-3′. After cutting with restriction enzymes *Hin*dIII and *Sal*I, this amplicon was inserted into pBR322 (accession J01749.1) so that the full-length hybrid *upf31* is transcribed from the *tetA* promoter. Mutations Y180A, F185A, and W198A were then introduced by PCR on this template using divergent back-to-back pairs of mutagenic primers (Y180AF 5′-CCCTCGGCAGCAACTTTCGTG-3′, Y180AR 5′-CGCG CAAATCGTGCAGCTTG-3′; F185AF 5′-CCGTGAGCGCG AGCGCATCAAG-3′, F185AR 5′-GCGTTGCTGCCGAGGT AGCGC-3′; W198AF 5′-CGCAAGCGAAAATCGCCAAG C-3′, and W198AR 5′-CGCGGGCTTTCTTCCGCTTG-3′). The amplified, linearized plasmids were phosphorylated, circularized by ligation and then transformed into DH5α before sequencing to check the mutations.

The pBRupf31/isoform3 was created as above with the following primers: K174E 5′-AGCTGCACGATTTGCGCTAC-3′, K174E 3′-CGGCTGGCTCGGGGTAGTTCATC-5′; I202L 5′-TCGCCAAGCTCGATCCGCTG-3′, and I202L 3′-GTTTCGCTTGCCAGCGGGCTTTC-5′.

xylE Reporter Assay

Overnight LB cultures were inoculated from single colonies of strains with reporter plasmid pGCMT1 or its derivative pGCMTupf31p alone (with just kanamycin selection) or with a second plasmid vector (pBR322) or vector plus the hybrid *upf31* gene (with both kanamycin and ampicillin) and incubated for 16 h. Bacteria from 1 mL aliquots were pelleted, resuspended in 0.5 mL sonication buffer, sonicated with three bursts of 5 s, the cell debris cleared with 10 min centrifugation at maximum speed in a microfuge at 4 °C and then measured quantitatively for XylE activity and protein concentration, as described previously (Zukowski et al. 1983).

Upf31 Protein Expression

Induction cultures of pCW-LIC-*upf31* in *E. coli* K-12 M1655 were started from an overnight liquid culture of LB (Lennox) with 100 µg/mL ampicillin (LB Amp). Twenty-five milliliters of LB Amp cultures were initiated at OD_{600} 0.05 using the liquid overnight culture to serve as the induction cultures. The 25 mL cultures were grown at 37 °C while shaking at 200 rpm. When the 25 mL cultures reached $OD_{600} > 1.0$, 1 mL of culture was removed, and the cells were harvested by centrifugation for 4 min at 16,000×*g* at room temperature in an Eppendorf 5425 centrifuge. This material served as the preinduction (or 0 h) time point for each culture. IPTG was then added to each induction culture to a final concentration of 0.1, 0.5, or 1 mM and growth was monitored at OD_{600}. At the time points indicated in supplementary fig. S2, Supplementary Material online, 1 mL of culture was removed from each culture and cells were isolated as above. Cells were resuspended in 1× Tris-Glycine running buffer (25 mM Tris-Cl, pH 8.3, 192 mM glycine, and 0.1% SDS) at 100 µL per $1.0 = OD_{600}$ cells. Laemmli buffer from Bio-Rad was added to the resuspended cell pellets to a final concentration of 1× and heated 10 min at >98 °C prior to gel loading.

Upf31 SDS–PAGE Gel Electrophoresis

Heated, denatured cellular extract samples (see previous section) were loaded on a Tris-glycine gel with a 12% acrylamide resolving gel and a 4% acrylamide stacking gel. Each sample lane contained 0.1 OD_{600} of cells. The molecular weight standards used were unstained broad range protein standards (NEB) and then the gel was run at 100 volts until the dye front was 1 cm from the bottom of the gel and then stained 1 h in Coomassie blue stain (50% methanol, 10% acetic acid, 0.25% Coomassie brilliant blue R-250) followed by destaining (5% methanol, 7% acetic acid) overnight.

Computational Modeling of Interactions Between Upf31 and pBP136-Encoded Proteins

An "in silico pull-down" of Upf31 was carried out using the LazyAF pipeline (McLean 2024). In short, the pBP136 genome (AB237782) was retrieved from NCBI GenBank as a FASTA protein coding sequence. The Upf31 protein coding sequence was used to generate individual concatenated FASTA files with Upf31 and each pBP136 coding sequence. ColabFold v1.5.5: AlphaFold2 w/MMseqs2 BATCH was then run on Google Colaboratory using the High-RAM A100 GPU with the following settings: msa_mode: MMseqs2 (UniRef + Environmental), num_models: 5, num_recycles: 3, stop_at_score: 100. Subsequently the JSON files were analyzed to retrieve the pTM and ipTM scores for each top ranked model and calculated the ranking_confidence score (0.2 pTM + 0.8 ipTM).

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

The authors acknowledge Tùng Lê at the John Innes Centre (UK) for his work on the in silico Upf31 pull-down. At the University of Washington, Dr Ben Kerr helped O.K. design the low-density competition assay. We are indebted for the laboratory efforts of undergraduate researchers Luke Hoover, Morgan Sower, Sue Winger, Katlyn Schafer, Courtney Stattner, Mattie Hagestad, Audrey Dingel, and Alexandra Gal and research technician Jack Millstein at the University of Idaho. The authors also thank Salvador ("Chava") Castaneda Barba for improving this manuscript. The project was made possible with the skilled sequencing efforts of IIDS Genomics and Bioinformatics Resources Core (GBRC) at the University of Idaho as well as the Arizona Genomics Institute at the University of Arizona.

Funding

E.M.T. and C.A.E. received support for this work via the National Institute of Allergy and Infectious Diseases grant no. R01 AI084918 from the National Institutes of Health (NIH) and National Science Foundation (NSF) Environmental Biology Division grant number 2142718. C.A.E. was also supported by the NSF Graduate Research Fellowship (GRFP) grant no. DGE-2019265372 as well as the Bioinformatics and Computational Biology (BCB) Fellowship and the Paul Joyce Memorial BCB Fellowship Endowment at the University of Idaho. O.K. was supported by the NSF GRFP grant no. DGE-1762114 and the NSF Postdoctoral Research Fellowships in Biology Program grant no. DBI-2305907. Further support was provided by the Wellcome Trust Investigator grant 221776/Z/2/Z to Tùng Lê that supported T.C.M., and by the Biotechnology and Biological Sciences Research Council (BBSRC) Institute Strategic Program "Harnessing Biosynthesis for Sustainable Food and Health" (HBio) (BB/X01097X/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

A data repository is available at https://github.com/clinte14/evo_plasmid_reg_supplemental. This repository includes all raw experimental data, sequencing, and statistical data as well as plasmid vector sequences.

References

Adamczyk M, Jagura-Burdzy G. Spread and survival of promiscuous IncP-1 plasmids. *Acta Biochim Pol*. 2003:50(2):425–453. https://doi.org/10.18388/abp.2003_3696.

Akhtar P. Phylogeny and genomic motifs of the beta sub-family of IncP-1 plasmids [PhD thesis]. England: University of Birmingham; 2002.

Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet*. 2022:399(10325):629–655. https://doi.org/10.1016/S0140-6736(21)02724-0.

Anton BP, Mongodin EF, Agrawal S, Fomenkov A, Byrd DR, Roberts RJ, Raleigh EA. Complete genome sequence of ER2796, a DNA methyltransferase-deficient strain of *Escherichia coli* K-12. *PLoS One*. 2015:10(5):e0127446. https://doi.org/10.1371/journal.pone.0127446.

Anton BP, Raleigh EA. Complete genome sequence of NEB 5-alpha, a derivative of *Escherichia coli* K-12 DH5α. *Genome Announc*. 2016:4(6):e01245-16. https://doi.org/10.1128/genomeA.01245-16.

Bachmann BJ. Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol Rev*. 1972:36(4):525–557. https://doi.org/10.1128/br.36.4.525-557.

Barrick JE, Colburn G, Deatherage DE, Traverse CC, Strand MD, Borges JJ, Knoester DB, Reba A, Meyer AG. Identifying structural

variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics*. 2014:15(1):1039. https://doi.org/10.1186/1471-2164-15-1039.

Bingle LEH, Rajasekar KV, Muntaha St, Nadella V, Hyde EI, Thomas CM. A single aromatic residue in transcriptional repressor protein KorA is critical for cooperativity with its co-regulator KorB. *Mol Microbiol*. 2008:70(6):1502–1514. https://doi.org/10.1111/j.1365-2958.2008.06498.x.

Bingle LEH, Thomas CM. Regulatory circuits for plasmid survival. *Curr Opin Microbiol*. 2001:4(2):194–200. https://doi.org/10.1016/S1369-5274(00)00188-0.

Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, Froula J, Kang DD, Malmstrom RR, Morgan RD, *et al*. The epigenomic landscape of prokaryotes. *PLoS Genet*. 2016:12(2):e1005854. https://doi.org/10.1371/journal.pgen.1005854.

Bottery MJ, Wood AJ, Brockhurst MA. Adaptive modulation of antibiotic resistance through intragenomic coevolution. *Nat Ecol Evol*. 2017:1(9):1364–1369. https://doi.org/10.1038/s41559-017-0242-3.

Bouma JE, Lenski RE. Evolution of a bacteria/plasmid association. *Nature*. 1988:335(6188):351–352. https://doi.org/10.1038/335351a0.

Brennan RG, Matthews BW. The helix-turn-helix DNA binding motif. *J Biol Chem*. 1989:264(4):1903–1906. https://doi.org/10.1016/S0021-9258(18)94115-3.

Brockhurst MA, Harrison E. Ecological and evolutionary solutions to the plasmid paradox. *Trends Microbiol*. 2022:30(6):534–543. https://doi.org/10.1016/j.tim.2021.11.001.

Brown CJ, Sen D, Yano H, Bauer ML, Rogers LM, Van der Auwera GA, Top EM. Diverse broad-host-range plasmids from freshwater carry few accessory genes. *Appl Environ Microbiol*. 2013:79(24):7684–7695. https://doi.org/10.1128/AEM.02252-13.

Dahlberg C, Chao L. Amelioration of the cost of conjugative plasmid carriage in *Escherichia coli* K12. *Genetics*. 2003:165(4):1641–1649. https://doi.org/10.1093/genetics/165.4.1641.

De Gelder L, Ponciano JM, Joyce P, Top EM. Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. *Microbiology (Reading)*. 2007:153(Pt 2):452–463. https://doi.org/10.1099/mic.0.2006/001784-0.

De Gelder L, Williams JJ, Ponciano JM, Sota M, Top EM. Adaptive plasmid evolution results in host-range expansion of a broad-host-range plasmid. *Genetics*. 2008:178(4):2179–2190. https://doi.org/10.1534/genetics.107.084475.

Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol*. 2003:4(9):R60. https://doi.org/10.1186/gb-2003-4-9-r60.

Figurski DH, Pohlman RF, Bechhofer DH, Prince AS, Kelton CA. Broad host range plasmid RK2 encodes multiple kil genes potentially lethal to *Escherichia coli* host cells. *Proc Natl Acad Sci U S A*. 1982:79(6):1935–1939. https://doi.org/10.1073/pnas.79.6.1935.

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*. 2010:7(6):461–465. https://doi.org/10.1038/nmeth.1459.

Galata V, Fehlmann T, Backes C, Keller A. PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Res*. 2019:47(D1):D195–D202. https://doi.org/10.1093/nar/gky1050.

Gomulkiewicz R, Holt RD. When does evolution by natural selection prevent extinction? *Evolution*. 1995:49(1):201–207. https://doi.org/10.2307/2410305.

Hall BG, Acar H, Nandipati A, Barlow M. Growth rates made easy. *Mol Biol Evol*. 2014:31(1):232–238. https://doi.org/10.1093/molbev/mst187.

Hall JPJ, Wright RCT, Guymer D, Harrison E, Brockhurst MA. Extremely fast amelioration of plasmid fitness costs by multiple functionally diverse pathways. *Microbiology*. 2020:166(1):56–62. https://doi.org/10.1099/mic.0.000862.

Hall JPJ, Wright RCT, Harrison E, Muddiman KJ, Wood AJ, Paterson S, Brockhurst MA. Plasmid fitness costs are caused by specific genetic conflicts enabling resolution by compensatory mutation. *PLoS Biol*. 2021:19(10):e3001225. https://doi.org/10.1371/journal.pbio.3001225.

Harrison E, Guymer D, Spiers AJ, Paterson S, Brockhurst MA. Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. *Curr Biol*. 2015:25(15):2034–2039. https://doi.org/10.1016/j.cub.2015.06.024.

Hayakawa M, Tokuda M, Kaneko K, Nakamichi K, Yamamoto Y, Kamijo T, Umeki H, Chiba R, Yamada R, Mori M, *et al*. Hitherto-unnoticed self-transmissible plasmids widely distributed among different environments in Japan. *Appl Environ Microbiol*. 2022:88(18):e01114-22. https://doi.org/10.1128/aem.01114-22.

Heuer H, Szczepanowski R, Schneiker S, Pühler A, Top EM, Schlüter A. The complete sequences of plasmids pB2 and pB3 provide evidence for a recent ancestor of the IncP-1β group without any accessory genes. *Microbiology (Reading)*. 2004:150(Pt 11):3591–3599. https://doi.org/10.1099/mic.0.27304-0.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018:35(2):518–522. https://doi.org/10.1093/molbev/msx281.

Horton RM, Cai ZL, Ho SN, Pease LR. Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *Biotechniques*. 1990:8(5):528–535. https://doi.org/10.2144/000114017.

Hughes JM, Lohman BK, Deckert GE, Nichols EP, Settles M, Abdo Z, Top EM. The role of clonal interference in the evolutionary dynamics of plasmid-host adaptation. *mBio*. 2012:3(4):e00077-12. https://doi.org/10.1128/mBio.00077-12.

Jain A, Srivastava P. Broad host range plasmids. *FEMS Microbiol Lett*. 2013:348(2):87–96. https://doi.org/10.1111/1574-6968.12241.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliamH, Maslen J, Mitchell A, Nuka G, *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014:30(9):1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Jordt H, Stalder T, Kosterlitz O, Ponciano JM, Top EM, Kerr B. Coevolution of host–plasmid pairs facilitates the emergence of novel multidrug resistance. *Nat Ecol Evol*. 2020:4(6):863–869. https://doi.org/10.1038/s41559-020-1170-1.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021:596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017:14(6):587–589. https://doi.org/10.1038/nmeth.4285.

Kamachi K, Sota M, Tamai Y, Nagata N, Konda T, Inoue T, Top EM, Arakawa Y. Plasmid pBP136 from *Bordetella pertussis* represents an ancestral form of IncP-1β plasmids without accessory mobile elements. *Microbiology (Reading)*. 2006:152(Pt 12):3477–3484. https://doi.org/10.1099/mic.0.29056-0.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013:30(4):772–780. https://doi.org/10.1093/molbev/mst010.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015:10(6):845–858. https://doi.org/10.1038/nprot.2015.053.

Klümper U, Riber L, Dechesne A, Sannazzarro A, Hansen LH, Sørensen SJ, Smets BF. Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. *ISME J*. 2015:9(4):934–945. https://doi.org/10.1038/ismej.2014.191.

Kornacki JA, Chang CH, Figurski DH. kil-kor regulon of promiscuous plasmid RK2: structure, products, and regulation of two operons that constitute the kilE locus. *J Bacteriol*. 1993:175(16):5078–5090. https://doi.org/10.1128/jb.175.16.5078-5090.1993.

Kostelidou K, Jones AC, Thomas CM. Conserved C-terminal region of global repressor KorA of broad-host-range plasmid RK2 is required for co-operativity between KorA and a second RK2 global regulator, KorB. *J Mol Biol*. 1999:289(2):211–221. https://doi.org/10.1006/jmbi.1999.2761.

Kostelidou K, Thomas CM. The hierarchy of KorB binding at its 12 binding sites on the broad-host-range plasmid RK2 and modulation of this binding by IncC1 protein. *J Mol Biol*. 2000:295(3):411–422. https://doi.org/10.1006/jmbi.1999.3359.

Kostelidou K, Thomas CM. DNA recognition by the KorA proteins of IncP-1 plasmids RK2 and R751. *Biochim Biophys Acta*. 2002:1576(1-2):110–118. https://doi.org/10.1016/S0167-4781(02)00306-8.

Kosterlitz O, Muñiz Tirado A, Wate C, Elg C, Bozic I, Top EM, Kerr B. Estimating the transfer rates of bacterial plasmids with an adapted Luria–Delbrück fluctuation analysis. *PLoS Biol*. 2022:20(7):e3001732. https://doi.org/10.1371/journal.pbio.3001732.

Kottara A, Hall JPJ, Harrison E, Brockhurst MA. Variable plasmid fitness effects and mobile genetic element dynamics across Pseudomonas species. *FEMS Microbiol Ecol*. 2018:94(1):fix172. https://doi.org/10.1093/femsec/fix172.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012:9(4):357–359. https://doi.org/10.1038/nmeth.1923.

Larsen MH, Figurski DH. Structure, expression, and regulation of the kilC operon of promiscuous IncP alpha plasmids. *J Bacteriol*. 1994:176(16):5022–5032. https://doi.org/10.1128/jb.176.16.5022-5032.1994.

Law A, Solano O, Brown CJ, Hunter SS, Fagnan M, Top EM, Stalder T. Biosolids as a source of antibiotic resistance plasmids for commensal and pathogenic bacteria. *Front Microbiol*. 2021:12:606409. https://doi.org/10.3389/fmicb.2021.606409.

Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res*. 2024:52(W1):W78–W82. https://doi.org/10.1093/nar/gkae268.

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014:30(7):923–930. https://doi.org/10.1093/bioinformatics/btt656.

Loftie-Eaton W, Bashford K, Quinn H, Dong K, Millstein J, Hunter S, Thomason MK, Merrikh H, Ponciano JM, Top EM. Compensatory mutations improve general permissiveness to antibiotic resistance plasmids. *Nat Ecol Evol*. 2017:1(9):1354–1363. https://doi.org/10.1038/s41559-017-0243-2.

Loftie-Eaton W, Yano H, Burleigh S, Simmons RS, Hughes JM, Rogers LM, Hunter SS, Settles ML, Forney LJ, Ponciano JM, *et al*. Evolutionary paths that expand plasmid host-range: implications for spread of antibiotic resistance. *Mol Biol Evol*. 2016:33(4):885–897. https://doi.org/10.1093/molbev/msv339.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014:15(12):550. https://doi.org/10.1186/s13059-014-0550-8.

Malone T, Blumenthal RM, Cheng X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J Mol Biol*. 1995:253(4):618–632. https://doi.org/10.1006/jmbi.1995.0577.

Marincola G, Jaschkowitz G, Kieninger A-K, Wencker FDR, Feßler AT, Schwarz S, Ziebuhr W. Plasmid-chromosome crosstalk in staphylococcus aureus: a horizontally acquired transcription regulator controls polysaccharide intercellular adhesin-mediated biofilm formation. *Front Cell Infect Microbiol*. 2021:11:660702. https://doi.org/10.3389/fcimb.2021.660702.

Marinus MG, Løbner-Olesen A. DNA methylation. *EcoSal Plus*. 2014:6(1). https://doi.org/10.1128/ecosalplus.ESP-0003-2013.

McLean TC. LazyAF, a pipeline for accessible medium-scale in silico prediction of protein-protein interactions. *Microbiology (Reading)*. 2024:170(7):001473. https://doi.org/10.1099/mic.0.001473.

McLean TC, Balaguer-Pérez F, Chandanani J, Thomas CM, Aicart-Ramos C, Burick S, Olinares PDB, Gobbato G, Mundy JEA, Chait BT, *et al*. Korb switching from DNA-sliding clamp to repressor mediates long-range gene silencing in a multi-drug resistance plasmid. *Nat Microbiol*. 2025:10(2):448–467. https://doi.org/10.1038/s41564-024-01915-3.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020:37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015.

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods*. 2022:19(6):679–682. https://doi.org/10.1038/s41592-022-01488-1.

Modi RI, Adams J. Coevolution in bacterial-plasmid populations. *Evolution*. 1991:45(3):656–667. https://doi.org/10.2307/2409918.

Norman A, Hansen LH, Sørensen SJ. Conjugative plasmids: vessels of the communal gene pool. *Philos Trans R Soc Lond B Biol Sci*. 2009:364(1527):2275–2289. https://doi.org/10.1098/rstb.2009.0037.

Pansegrau W, Lanka E, Barth PT, Figurski DH, Guiney DG, Haas D, Helinski DR, Schwab H, Stanisich VA, Thomas CM. Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. *J Mol Biol*. 1994:239(5):623–663. https://doi.org/10.1006/jmbi.1994.1404.

Porse A, Schønning K, Munck C, Sommer MOA. Survival and evolution of a large multidrug resistance plasmid in new clinical bacterial hosts. *Mol Biol Evol*. 2016:33(11):2860–2873. https://doi.org/10.1093/molbev/msw163.

Revell LJ. Phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ*. 2024:12:e16505. https://doi.org/10.7717/peerj.16505.

Rosche TM, Siddique A, Larsen MH, Figurski DH. Incompatibility protein IncC and global regulator KorB interact in active partition of promiscuous plasmid RK2. *J Bacteriol*. 2000:182(21):6014–6026. https://doi.org/10.1128/JB.182.21.6014-6026.2000.

Sambrook J, Russell DW. *Molecular cloning: a laboratory manual*. 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2001. WorldCat.org.

San Millan A. Evolution of plasmid-mediated antibiotic resistance in the clinical context. *Trends Microbiol*. 2018:26(12):978–985. https://doi.org/10.1016/j.tim.2018.06.007.

San Millan A, MacLean RC. Fitness costs of plasmids: a limit to plasmid transmission. *Microbiol Spectr*. 2017:5(5):65–79. https://doi.org/10.1128/microbiolspec.mtbp-0016-2017.

San Millan A, Peña-Miller R, Toll-Riera M, Halbert ZV, McLean AR, Cooper BS, MacLean RC. Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids. *Nat Commun*. 2014:5(1):5208. https://doi.org/10.1038/ncomms6208.

San Millan A, Toll-Riera M, Qi Q, MacLean RC. Interactions between horizontally acquired genes create a fitness cost in *Pseudomonas aeruginosa*. *Nat Commun*. 2015:6(1):6845. https://doi.org/10.1038/ncomms7845.

Schlüter A, Heuer H, Szczepanowski R, Poler SM, Schneiker S, Pühler A, Top EM. Plasmid pB8 is closely related to the prototype IncP-1β plasmid R751 but transfers poorly to *Escherichia coli* and carries a new transposon encoding a small multidrug resistance efflux protein. *Plasmid*. 2005:54(2):135–148. https://doi.org/10.1016/j.plasmid.2005.03.001.

Schmidhauser TJ, Helinski DR. Regions of broad-host-range plasmid RK2 involved in replication and stable maintenance in nine species of gram-negative bacteria. *J Bacteriol*. 1985:164(1):446–455. https://doi.org/10.1128/jb.164.1.446-455.1985.

Sen D, Brown CJ, Top EM, Sullivan J. Inferring the evolutionary history of IncP-1 plasmids despite incongruence among backbone gene trees. *Mol Biol Evol*. 2013:30(1):154–166. https://doi.org/10.1093/molbev/mss210.

Sen D, Van der Auwera GA, Rogers LM, Thomas CM, Brown CJ, Top EM. Broad-host-range plasmids from agricultural soils have IncP-1

backbones with diverse accessory genes. *Appl Environ Microbiol*. 2011:77(22):7975–7983. https://doi.org/10.1128/AEM.05439-11.

Shintani M, Takahashi Y, Yamane H, Nojiri H. The behavior and significance of degradative plasmids belonging to Inc groups in Pseudomonas within natural environments and microcosms. *Microbes Environ*. 2010:25(4):253–265. https://doi.org/10.1264/jsme2.me10155.

Sota M, Top EM. Host-specific factors determine the persistence of IncP-1 plasmids. *World J Microbiol Biotechnol*. 2008:24(9):1951–1954. https://doi.org/10.1007/s11274-008-9653-2.

Sota M, Tsuda M, Yano H, Suzuki H, Forney LJ, Top EM. Region-specific insertion of transposons in combination with selection for high plasmid transferability and stability accounts for the structural similarity of IncP-1 plasmids. *J Bacteriol*. 2007:189(8):3091–3098. https://doi.org/10.1128/JB.01906-06.

Sota M, Yano H, Hughes JM, Daughdrill GW, Abdo Z, Forney LJ, Top EM. Shifts in the host range of a promiscuous plasmid through parallel evolution of its replication initiation protein. *ISME J*. 2010:4(12):1568–1580. https://doi.org/10.1038/ismej.2010.72.

Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. *Nat Rev Genet*. 2015:16(8):472–482. https://doi.org/10.1038/nrg3962.

Stalder T, Cornwell B, Lacroix J, Kohler B, Dixon S, Yano H, Kerr B, Forney LJ, Top EM. Evolving populations in biofilms contain more persistent plasmids. *Mol Biol Evol*. 2020:37(6):1563–1576. https://doi.org/10.1093/molbev/msaa024.

Stalder T, Rogers LM, Renfrow C, Yano H, Smith Z, Top EM. Emerging patterns of plasmid-host coevolution that stabilize antibiotic resistance. *Sci Rep*. 2017:7(1):4853. https://doi.org/10.1038/s41598-017-04662-0.

Thomas CM, Thomson NR, Cerdeño-Tárraga AM, Brown CJ, Top EM, Frost LS. Annotation of plasmid genes. *Plasmid*. 2017:91:61–67. https://doi.org/10.1016/j.plasmid.2017.03.006.

Thompson CMA, Hall JPJ, Chandra G, Martins C, Saalbach G, Panturat S, Bird SM, Ford S, Little RH, Piazza A, *et al*. Plasmids manipulate bacterial behaviour through translational regulatory crosstalk. *PLoS Biol*. 2023:21(2):e3001988. https://doi.org/10.1371/journal.pbio.3001988.

Thorsted PB, MacArtney DP, Akhtar P, Haines AS, Ali N, Davidson P, Stafford T, Pocklington MJ, Pansegrau W, Wilkins BM, *et al*. Complete sequence of the IncPβ plasmid R751: implications for evolution and organisation of the IncP backbone. *J Mol Biol*. 1998:282(5):969–990. https://doi.org/10.1006/jmbi.1998.2060.

Troxell B, Hassan HM. Transcriptional regulation by Ferric Uptake Regulator (Fur) in pathogenic bacteria. *Front Cell Infect Microbiol*. 2013:3:59. 10.3389/fcimb.2013.00059.

Vial L, Hommais F. Plasmid-chromosome cross-talks. *Environ Microbiol*. 2020:22(2):540–556. https://doi.org/10.1111/1462-2920.14880.

Wiser MJ, Lenski RE. A comparison of methods to measure fitness in *Escherichia coli*. *PLoS One*. 2015:10(5):e0126210. https://doi.org/10.1371/journal.pone.0126210.

Yang J, Wu R, Xia Q, Yu J, Yi L-X, Huang Y, Deng M, He W-Y, Bai Y, Lv L, *et al*. The evolution of infectious transmission promotes the persistence of mcr-1 plasmids. *mBio*. 2023:14(4):e00442-23. https://doi.org/10.1128/mbio.00442-23.

Yano H, Deckert GE, Rogers LM, Top EM. Roles of long and short replication initiation proteins in the fate of IncP-1 plasmids. *J Bacteriol*. 2012:194(6):1533–1543. https://doi.org/10.1128/JB.06395-11.

Yano H, Rogers LM, Knox MG, Heuer H, Smalla K, Brown CJ, Top EM. Host range diversification within the IncP-1 plasmid group. *Microbiology (Reading)*. 2013:159(Pt 11):2303–2315. https://doi.org/10.1099/mic.0.068387-0.

Yano H, Wegrzyn K, Loftie-Eaton W, Johnson J, Deckert GE, Rogers LM, Konieczny I, Top EM. Evolved plasmid-host interactions reduce plasmid interference cost. *Mol Microbiol*. 2016:101(5):743–756. https://doi.org/10.1111/mmi.13407.

Zatyka M, Bingle L, Jones AC, Thomas CM. Cooperativity between KorB and TrbA repressors of broad-host-range plasmid RK2. *J Bacteriol*. 2001:183(3):1022–1031. https://doi.org/10.1128/JB.183.3.1022-1031.2001.

Zatyka M, Jagura-Burdzy G, Thomas CM. Transcriptional and translational control of the genes for the mating pair formation apparatus of promiscuous IncP plasmids. *J Bacteriol*. 1997:179(23):7201–7209. https://doi.org/10.1128/jb.179.23.7201-7209.1997.

Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*. 2018:19(Suppl 6):153. https://doi.org/10.1186/s12859-018-2129-y.

Zhong X, Krol JE, Top EM, Krone SM. Accounting for mating pair formation in plasmid population dynamics. *J Theor Biol*. 2010:262(4):711–719. https://doi.org/10.1016/j.jtbi.2009.10.013.

Zukowski MM, Gaffney DF, Speck D, Kauffmann M, Findeli A, Wisecup A, Lecocq JP. Chromogenic identification of genetic regulatory signals in Bacillus subtilis based on expression of a cloned Pseudomonas gene. *Proc Natl Acad Sci U S A*. 1983:80(4):1101–1105. https://doi.org/10.1073/pnas.80.4.1101.