Evolution of a Plasmid Regulatory Circuit Ameliorates Plasmid Fitness Cost

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

[1] Bioinformatics and Computational Biology Program, University of Idaho, Moscow, Idaho, USA.

[2] Institute for Interdisciplinary Data Sciences, University of Idaho, Moscow, Idaho, USA.

[3] Department of Biological Sciences, University of Idaho, Moscow, Idaho, USA.

[4] Department of Molecular Microbiology, John Innes Centre, Norwich, UK.

[5] Biology Department, University of Washington, Seattle, Washington, USA.

[6] School of Biosciences, University of Birmingham, Edgbaston, Birmingham, UK.

Corresponding Author: Eva M. Top

evatop@uidaho.edu

Abstract

Plasmids play a major role in bacterial evolution and rapid adaptation by facilitating the horizontal transfer of diverse genes. Understanding how plasmids are transferred and maintained in bacterial populations is important, especially given the increasing plasmid-mediated spread of antibiotic-resistance genes to human pathogens. We investigated why broad-host range plasmid pBP136, originally isolated from clinical samples of *Bordetella pertussis,* quickly became extinct in laboratory *Escherichia coli* populations. We found that the inactivation of a previously uncharacterized plasmid gene, *upf31*, drastically improved long-term maintenance of the plasmid in *E. coli*. Loss of this single gene was associated with decreased transcription of numerous genes in the plasmid *korA*, *korB* and *korC* regulons, as well as changes in many chromosomal genes primarily related to metabolism. This change in transcriptome is likely initiated by Upf31 interacting with one of these major plasmid regulators, KorB. Expression of *upf31 in trans* not only negatively affected the persistence of a pBP136 *upf31* deletion mutant, but also of the closely related archetype IncPβ plasmid R751, which is stable in *E. coli* and natively encodes an internally truncated *upf31* allele. This suggests that whereas the *upf31* allele in pBP136 might advantageously modulate gene expression in its original host, *B. pertussis*, the same function can have harmful effects in *E. coli*. Thus, using multiple hosts to study the effects of knockouts in broad-host-range plasmid genes of unknown function may reveal unexpected mechanisms that determine the fate of that plasmid in bacterial communities.

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[1,2]. 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[3,4]. Explaining how plasmids evolve to successfully transfer and maintain themselves in bacterial populations and communities is thus a critical step to understanding and 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 strains[5]. 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[6]. Several studies have shown compensatory mutations that arise during serial batch cultivation of plasmid-host pairs which ameliorate plasmid fitness cost and thereby restore plasmid persistence. Some of the important early studies[7,8] showed that such cost amelioration could occur through genetic changes on the plasmid but were unable to explore the underlying molecular mechanisms. With the advent of next generation sequencing, compensatory mutations have been identified in specific plasmid genes, including those encoding accessory functions[9,10], replication inititation[11–

62 [13], and conjugation machinery[14–16]. Additionally, chromosomal evolution can stabilize
63 host-plasmid pairs through amelioration of plasmid cost via mutations including global
64 regulators[9,17] and putative helicases[18,19]. Thus, compensatory mutations that increase
65 plasmid persistence can occur on the plasmid, chromosome, or both[20].

66 Some plasmid types, such as those of the incompatibility group IncP used in this study,
67 are known to be highly persistent across many Gram-negative bacteria[21–26], with few
68 exceptions[5,27]. This remarkable persistence is aided by several conserved regions[28–30]
69 of backbone genes which ensure the fidelity of critical functions like replication, stable
70 inheritance, mating pair formation and conjugative DNA transfer. These backbone
71 genes are controlled by a complex genetic regulatory circuit with four main regulators
72 located chiefly within the stable maintenance and control region[31,32]. The regulators are
73 part of a negative feedback loop (referred to as autogenous control[33]) that limits their
74 own expression and that of the operons they control. This in turn allows fine-tuned and
75 temporal plasmid gene expression to minimize cost of plasmid carriage to the host and
76 prevent potential conflicts with host cellular machinery. Collectively, this regulatory
77 network enables IncP plasmids to replicate, transfer, and be stably maintained in
78 multiple classes of the *Proteobacteria.*

79 Given the known broad host range of IncP plasmids, we sought to examine the
80 previously reported but unexplained poor persistence of subgroup IncPβ plasmid
81 pBP136Km in several *Escherichia coli* strains[34]*.* This plasmid was previously obtained
82 by inserting a kanamycin resistance gene in pBP136[35], a medically relevant plasmid
83 found within a *Bordetella pertussis* strain isolated from a lethal infection in an infant in
84 Japan[36].

We report a single genetic change in pBP136Km that occurred within 50 generations of evolution in *E. coli* in the absence of antibiotic selection and drastically increased the plasmid's persistence. This improved persistence was explained by a decrease in the plasmid fitness cost due to partial deletion of a previously uncharacterized plasmid-encoded gene, *upf31*. The high cost of plasmid carriage caused by this *upf31* allele was shown to extend to the archetype IncPβ plasmid R751[28], which is otherwise stable in *E. coli*. In pBP136Km the *upf31* deletion was associated with changes in expression of plasmid backbone regulatory genes and chromosomal genes. While previous studies found that evolution of chromosomally encoded regulators can improve plasmid persistence, here plasmid persistence was improved by a single mutation in a plasmid gene that widely affects the transcription of plasmid backbone regulatory genes as well as chromosomal genes.

Results

Plasmid-encoded *upf31* results in poor plasmid persistence in *E. coli* hosts.

We first sought to confirm a previous finding that plasmid pBP136Km shows poor persistence in various *E. coli* hosts[34]. Using *E. coli* K-12 MG1655 (hereafter K-12), we performed a plasmid persistence assay entailing daily passage of triplicate cultures in non-selective conditions (*i.e.,* without antibiotics). This initial assay with the ancestral host-plasmid pair showed poor plasmid persistence (Figure 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 used these clones to initiate new persistence

107 assays. The persistence of the plasmid in these clones were markedly improved (clones
108 A, B, and C in Figure 1A in red, green, magenta).

109 To identify genetic changes in the evolved lineages we used whole genome sequencing
110 of the three isolated clones (Supplemental Table S1). We found deletions within
111 pBP136Km consistently targeting the region often associated with accessory genes
112 between the *tra* and *trb* operons (Figure 1B). Additionally, no genetic changes were
113 observed in the chromosomes of these three clones, and clones B and C had identical
114 plasmids (Figure 1B in green and magenta). As the plasmid deletions in all three clones
115 at least included *upf30.5* and *upf31.0*, we suspected the presence of one or both genes
116 to be responsible for the poor plasmid persistence in *E. coli*.

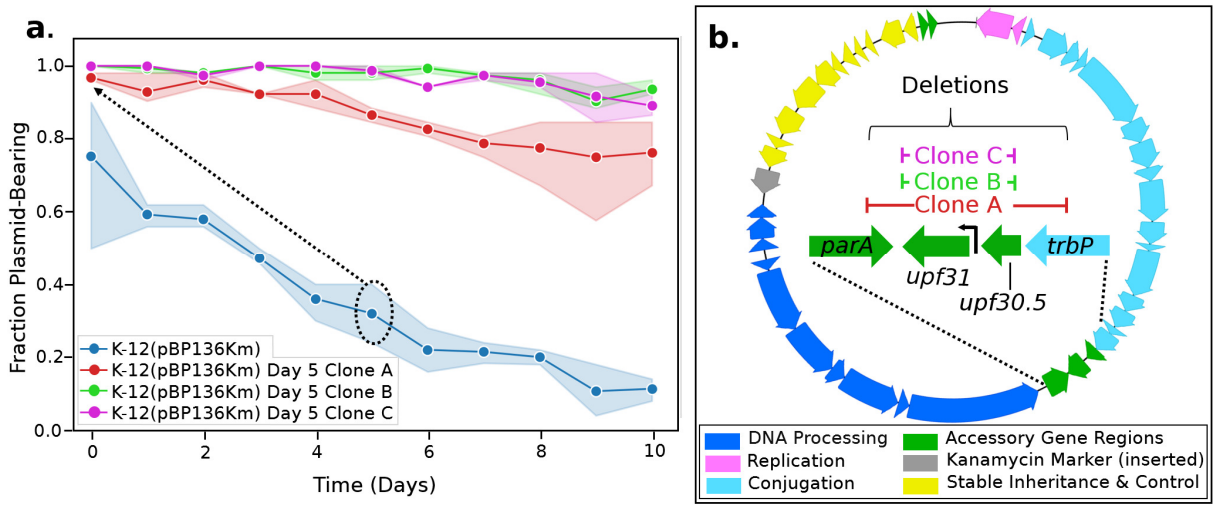


Fig 1. Plasmid evolution rapidly improved pBP136Km persistence. (a) Persistence of the ancestral plasmid pBP136Km (blue) and three populations (red, green, magenta) inoculated with clones that were isolated on Day 5 of the ancestral plasmid assay (see dotted ellipse and arrow). Clones isolated on Day 5 had greatly improved persistence compared to 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).

117

118 We also routinely performed sequencing of clones recovered after transferring the
119 ancestral plasmid between strains of *E. coli* by conjugation (*i.e.,* matings) or
120 electroporation. Among these recovered clones we discovered several with deletions
121 either internal to *upf31* or in its putative promoter (Supplemental Table S1). To test if
122 deletion of *upf31* alone could explain improved plasmid persistence, we compared the
123 persistence of one of these recovered variants, named pBP136KmΔ*upf31,* to that of the
124 ancestral plasmid in the same host, *E. coli* K-12 (Fig 2A).

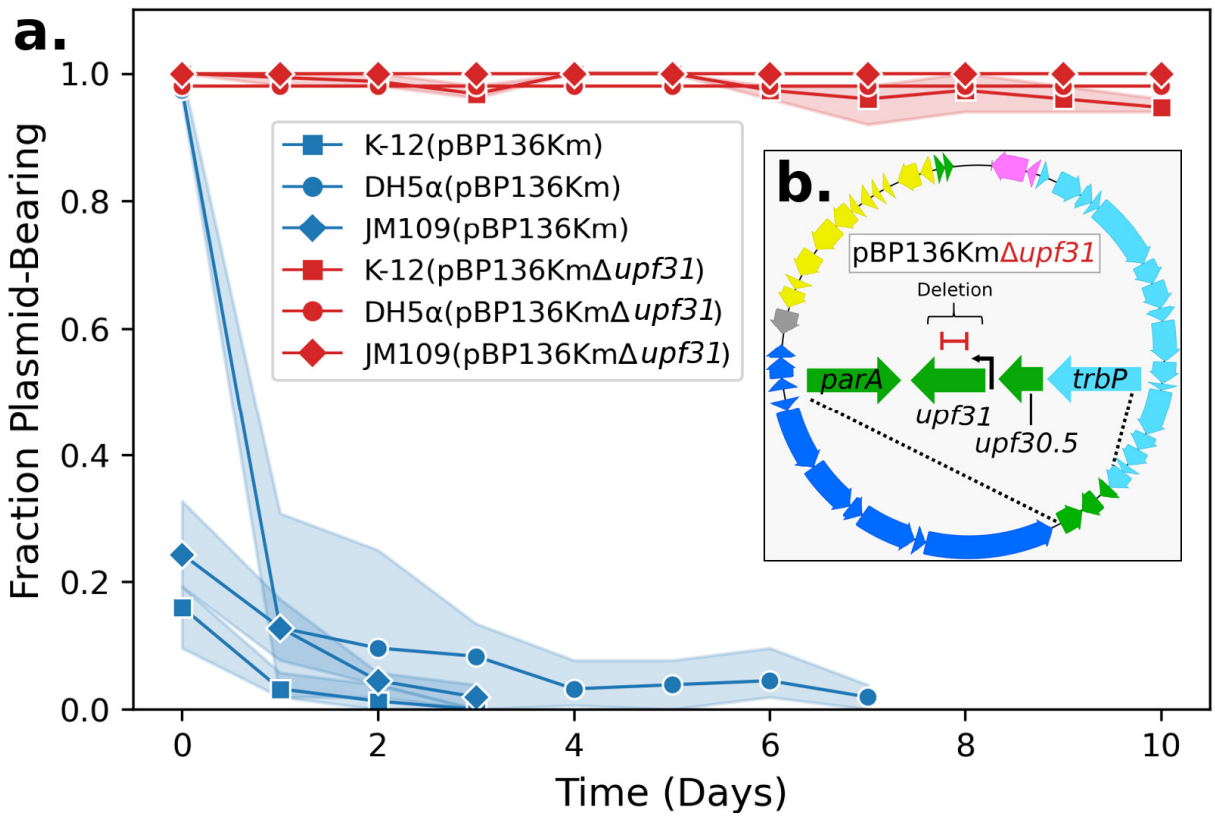


Fig 2. *upf31* deletion alone explained improved persistence in *E. coli* hosts. **(a)** The poor persistence of plasmid pBP136Km (blue lines) in *E. coli* strains K-12 MG1655, DH5α, and JM109 (square, circle, and diamond respectively) was markedly improved after an internal 168-bp deletion in plasmid 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.*

125

126

127 This selected plasmid contained a 168-base pair (bp) deletion internal to the open
128 reading frame of *upf31* whereas *upf30.5* was intact (Fig 2B). Given that the plasmid with
129 inactivated *upf31* was persistent for at least 10 days, this gene of unknown function
130 must be responsible for the very poor persistence of pBP136 (compare red versus blue
131 squares of Fig 2A). The improvement in plasmid persistence after inactivation of *upf31*
132 was also seen in other well-studied strains of *E. coli* including DH5α and JM109 (Fig
133 2A). Importantly, the evolution from ancestral pBP136Km into pBP136KmΔ*upf31*
134 appeared during an overnight mating followed by plating for transconjugants on
135 selective media (Supplemental Table S1). This provides an example of very rapid
136 plasmid evolution that results in remarkably improved plasmid persistence (*i.e.,* < 48
137 hours).

138 **Host fitness was lowered in the presence of *upf31* and pBP136Km.**

139 To investigate if the improved plasmid persistence associated with *upf31* inactivation
140 was due to amelioration of plasmid fitness costs, we employed two approaches: *1)* a
141 comparison of the growth dynamics between the two plasmid-host pairs, with a focus on
142 the maximum growth rate and carrying capacity, and *2)* competition assays involving
143 plasmid-bearing and plasmid-free strains. First, we observed that K-12 containing
144 pBP136KmΔ*upf31* had a significantly higher maximum growth rate and reached a
145 higher end-point density in comparison to K-12 containing ancestral pBP136Km (Figure
146 3A). These results demonstrate that the presence of *upf31* encoded on pBP136Km
147 imposes a substantial fitness cost on its host.

148 Next, we confirmed the results of the growth rate observations by estimating the cost of
149 the ancestral and evolved plasmid in competition assays. This was done by individually
150 competing each plasmid-bearing strain against the plasmid-free K-12. We note that
151 direct competition assays between hosts containing a highly transmissible conjugative
152 plasmid such as pBP136Km can be confounded by plasmid transfer during liquid growth
153 assays. While IncP plasmids are less efficient at transferring in liquid than on surfaces,
154 they have been shown to transfer at detectable rates[37]. To overcome this limitation, we
155 developed a novel low-density competition assay which allows competing two strains at
156 densities too low for appreciable plasmid transfer to occur (see Methods). Indeed, as
157 conjugation requires cell contact, decreasing initial densities result in decreasing cell
158 collisions and thus plasmid transfer events[38]. The results (Fig 3B) show a fitness of K-12
159 (pBP136Km) relative to K-12 of only 0.51, suggesting the plasmid caused a 49%
160 reduction in fitness. By contrast, the fitness of K-12 (pBP136KmΔ*upf31*) was statistically
161 indistinguishable from that of K-12. We conclude that the persistence differences in
162 Figure 2 can be attributed to a very large fitness cost conferred by pBP136Km, which

163 was alleviated by the inactivation of one plasmid gene, *upf31.*

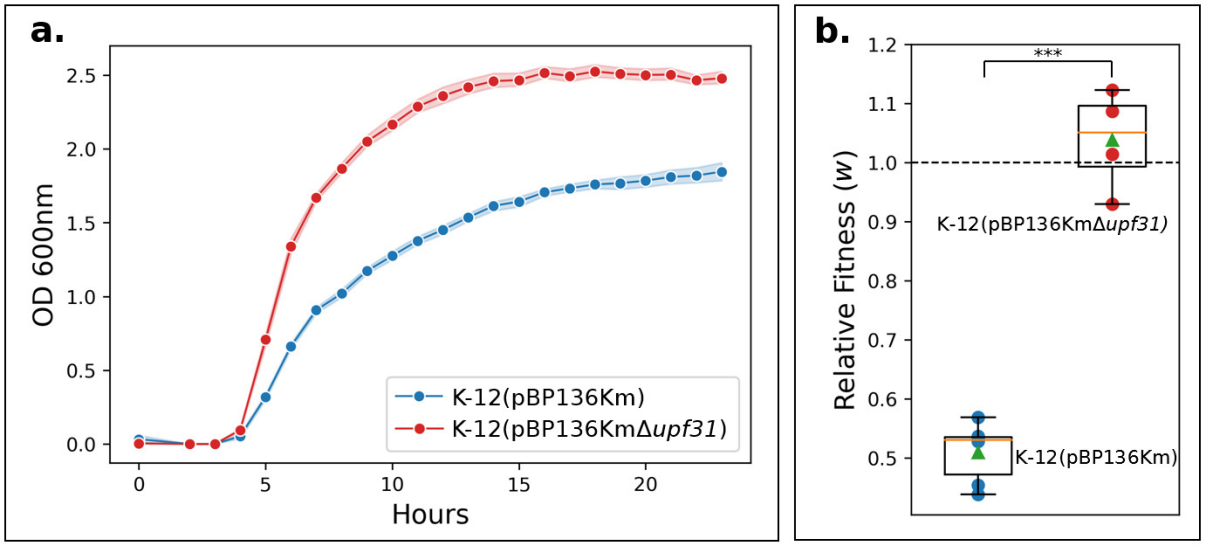


Figure 3. The presence of *upf31* encoded on pBP136Km lowered host fitness. (a) Bacterial growth in batch culture was slower (blue) when carrying pBP136Km with ancestral *upf31* compared to the evolved plasmid pBP136KmΔ*upf31* with inactivated *upf31* (red). Lighter shades indicate 95% confidence interval. (Two sample independent T-test of max growth rates, *n*=10, p-value= 2.35×10^{-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-value=1.70×10^{-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-value=0.40). Box is interquartile range, green triangle is mean, orange line is median, ***=p ≤ 0.001.

164

165 **The fitness cost imposed by *upf31* requires the presence of pBP136Km.**

166 We next sought to understand if the expression of *upf31* directly imposed a fitness cost

167 on K-12 in the absence of plasmid pBP136Km. This was done by expressing *upf31*

168 within K-12 *in trans* from expression vector pCW-LIC-*upf31.* Production of protein Upf31

169 was verified via induction of pCW-LIC-*upf31* and observation of the associated 25.4 kDa

170 product (Supplemental Figure S1). As a negative control we used the non-modified

171 vector pCW-LIC (hereafter named pCW-LIC-*sacB* for clarity, as *sacB* is present in this

172 vector but lost during insertion of *upf31* to make pCW-LIC-*upf31*). The maximum growth
173 rates of K-12 with or without the Upf31 protein were not significantly different (Fig 4A).
174 This suggests that Upf31 does not impose a significant fitness cost on K-12 in the
175 absence of the plasmid it is encoded on. In contrast, when *upf31* was expressed *in*
176 *trans* in the presence of pBP136KmΔ*upf31,* the maximum growth rate was significantly
177 reduced compared to the same strain with the control vector (Fig 4B). Thus, we
178 demonstrated that Upf31 did not directly impose a fitness cost on K-12 but rather the
179 interaction of Upf31 with pBP136Km was responsible for the host fitness cost.

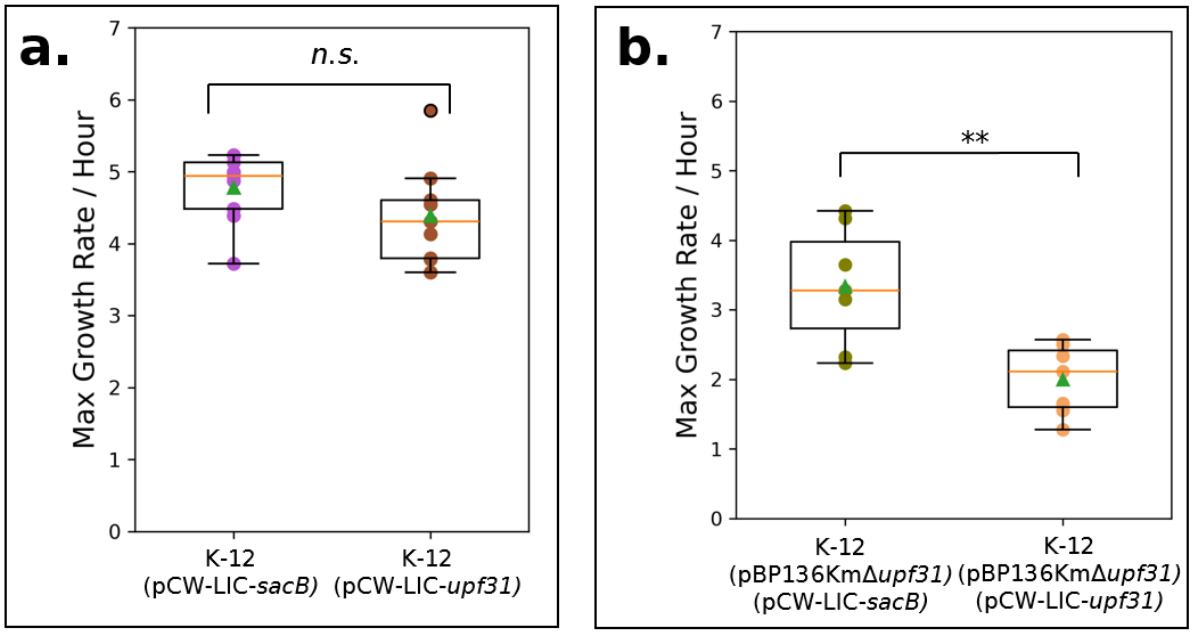


Figure 4. Upf31 requires the presence of pBP136Km to significantly reduce K-12 growth rate. (a) In the absence of plasmid pBP136Km there was no significant difference in maximum growth rate between *E. coli* K-12 expressing control *sacB* (purple) and *upf31* (brown) (Two sample independent T-test, *n*=10, p-value=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=4.23×10^{-3}). Box is interquartile range, green triangle is mean, orange line is median, *n.s.* = $p > 0.05$, **=$p \leq 0.01$.

180

181 **Gene *upf31* also affects plasmid persistence and host fitness for archetype IncPβ**
182 **plasmid R751 when complemented *in trans*.**

183 We examined if the high cost of plasmid carriage observed with *upf31* extends to other
184 IncPβ plasmids. The archetype IncPβ plasmid R751[28] encodes a *upf31* homolog with a
185 66 base pair internal deletion compared to pBP136Km. The high persistence of R751 in
186 *E. coli* suggests that this allele does not negatively affect *E. coli* fitness (Supplemental
187 Figure S2). To test if *upf31* from pBP136Km affects K-12 (R751) when expressed *in*
188 *trans*, we compared K-12 (R751) (pCW-LIC-*upf31*) to control K-12 (R751) (pCW-LIC-
189 *sacB*) in terms of plasmid persistence and growth rate. Plasmid R751 showed lower
190 persistence and resulted in a statistically significant lower host growth rate
191 (Supplemental Figure S2) when pBP136Km's *upf31* is present.

192 Forty-nine IncPβ plasmids were identified in NCBI that contain a gene with full-length
193 homology to the ancestral *upf31* and associated promoter in pBP136, although none
194 have an identical protein sequence (Supplemental Data S1). We tested the persistence
195 of two of these plasmids, pAKD1, isolated from forest soil[39], and the wastewater
196 associated pALTS29[40]. Whereas plasmid pAKD1 showed poor persistence, similar to
197 pBP136Km, pALTS29 demonstrated high persistence (Supplemental Figure S3).
198 Plasmid pAKD1 and pALTS29 had three amino acid changes compared to Upf31 in
199 pBP136 (D76A, A92V, R124G), whilst pALTS29 had a unique Q82P change. Plasmid
200 pALTS29 may be worth further analysis, especially since the environment in which it
201 was found could be one that has selected variants that can persist better in a range of
202 Enterobacteriaceae.

203 **The computationally predicted methylase function of Upf31 is not supported by**
204 **experimental results.**

205 To provide further insight into the possible function(s) of Upf31 we performed a
206 bioinformatic analysis using Phyre2. The output strongly predicted Upf31 to be a Dam
207 methylase homolog with 70% of the amino acids modeling with 100% confidence
208 (Supplemental Data S2). This is consistent with the pBP136 reference genome
209 including a computationally automated annotation of *upf31* as encoding a DNA
210 methylase based on protein similarity. To test if Upf31 is a Dam homolog, we digested
211 genomic DNA extracted from *dam*$^{-}$ strain *E. coli* JM110 with and without pBP136Km
212 and *dam*$^{+}$ K-12, using restriction enzymes that digest only methylated or non-
213 methylated 'GATC' DNA sequences. The results suggest that Upf31 is not a functional
214 Dam homolog as the restriction profiles of JM110 (pBP136Km) were opposite of those
215 of the *dam+* strain K-12 (Supplemental Figure S4).

216 Next, to determine if Upf31 more broadly encoded a methyltransferase, we used the
217 base pair modification detection available via Single Molecule Real Time (SMRT)
218 sequencing. This method has the benefit of not requiring *a priori* knowledge of the base
219 pair modification chemistry or target DNA sequence of the putative methylase. Briefly,
220 the methylation-free *E. coli* ER2796[41] was used with *upf31* expressed from vector pCW-
221 LIC-*upf31* or pCW-LIC-*sacB* as the vector control. Genomic DNA extracts of both
222 strains were analyzed using the base modification analysis on a Pacbio Sequel II, which
223 measures the kinetic incorporation of base pairs to detect patterns distinctive from non-
224 methylated base pairs. The output showed no discernible kinetic patterns associated
225 with methylated base pair incorporation, indicating that Upf31 does not function as a

226 methylase under our experimental conditions, in contrast to the computational
227 predictions.

228 **The presence of *upf31* alters expression of global plasmid regulators and**
229 **associated operons.**

230 To understand how *upf31* and pBP136Km interact to influence *E. coli* fitness, we next
231 performed RNA-seq on K-12 containing pBP136Km and K-12 containing
232 pBP136KmΔ*upf31*. We considered K-12 (pBP136Km) to be the treatment (presence of
233 the intact gene *upf31*), and K-12 (pBP136KmΔ*upf31*) with the internal 168-bp deletion
234 to be the control (absence of a functional *upf31*). The presence of *upf31* led to greater
235 than two-fold expression changes in 654 of the ~4,400 chromosomal genes, with 412 of
236 those upregulated and 242 downregulated (adjusted P-value < 0.05, Supplemental Data
237 S3). The gene ontology tool DAVID[42] predicted downregulation of cellular pathways
238 related to nitrogen and sugar metabolisms, while biosynthesis of siderophores, ABC
239 transporters, and sulfur metabolism pathways were all upregulated (Supplemental Data
240 S4).

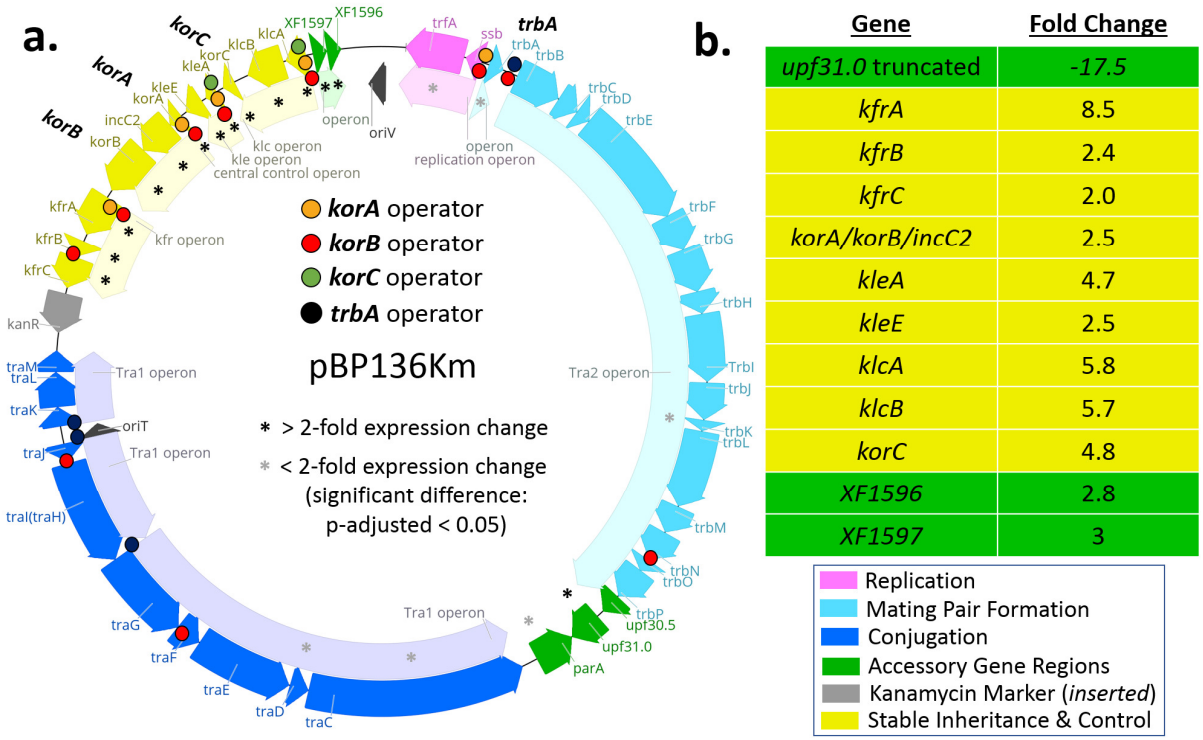
241 Most interestingly, K-12 (pBP136Km) showed greater than two-fold expression increase
242 in 11 of its 46 plasmid genes relative to the *upf31* deletion variant (adjusted P-value <
243 0.05, Supplemental Data S3). These genes were primarily part of the stable inheritance
244 and control region made up of regulons controlled by proteins KorA, KorB and KorC
245 (Figure 5A in yellow). Strikingly, K-12 (pBP136Km) showed a relative decrease in
246 expression in only one gene, *upf31* itself. This relative comparison of gene expression
247 was made possible by pBP136KmΔ*upf31* not having a complete *upf31* deletion but

248 rather an internal 168-bp deletion that left an inactivated 507 bp *upf31* product to be

249 expressed (Figure 5B). A possible mechanistic explanation for this decrease is that

250 Upf31 represses its own transcription, like IncP plasmid regulators KorA, KorB, and

251 KorC[43]. A putative promoter lies just upstream of *upf31* (Supplemental Figure S5).



Gene	Fold Change
upf31.0 truncated	*-17.5*
kfrA	8.5
kfrB	2.4
kfrC	2.0
korA/korB/incC2	2.5
kleA	4.7
kleE	2.5
klcA	5.8
klcB	5.7
korC	4.8
XF1596	2.8
XF1597	3

Figure 5. Expression of global plasmid regulators and their associated operons was increased in the presence of ancestral *upf31* compared to 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 the presence of *upf31.* Note expression of the ancestral *upf31* is ~16-fold lower than that of the truncated *upf31* knockout.

252

253 The effect of Upf31 to cause at least two-fold increased expression of all four of the

254 canonical IncP regulatory genes[44]– *korA, korB, korC,* and *trbA* – along with the other

255 genes in the operons they regulate (Figure 5A in bold) is not so easy to explain. These

256 regulators are known to autogenously repress their own expression[33] by binding
257 cognate operators near their respective promoters (orange, red, green, and black dots
258 in Figure 5A). They are also known to repress promoters for the other operons in their
259 regulons. The increased repressor gene expression should therefore decrease rather
260 than increase expression of all other regulated operons as the expression of the
261 repressor genes increases (yellow in Figure 5A). This suggests that *upf31* has a general
262 effect of diminishing the repressive ability of *kor* regulatory genes, both in terms of their
263 autogenous control and repression of their respective regulons.

264 **Upf31 is computationally predicted to bind KorB via a C-terminal dimerization**
265 **domain.**

266 A hypothesis to explain these results emerges from the fact that the high level of
267 repression observed with these promoters depends on the cooperative interaction of
268 KorA (and TrbA) with KorB[45–47]. If Upf31 were able to disrupt this cooperativity then it
269 might decrease repression while allowing the concentration of the regulators to rise. To
270 explore the hypothesis that Upf31 disrupts the cooperative interaction of KorA (and
271 TrbA) with KorB we used Alphafold2[48] and ColabFold[49] to predict the structures of every
272 individual pBP136 coding region in complex with Upf31. We subsequently ranked these
273 interactions using a self-assessment ranking score (ranking_confidence: 0.2pTM +
274 0.8ipTM). The full set of data generated in this way is in Supplemental Data S5. The
275 structure prediction for Upf31 is that it possesses a globular DNA binding domain
276 (residues 1 to 174) and a C-terminal dimerization domain (CTD) which is consistent with
277 its proposed autogenous regulation of its own expression. However, the most exciting
278 result of this analysis is that it predicts that the C-terminal dimerization domain mediates

279 interaction with KorB (Figure 6A) in a way that is very similar to KorA. Modeling of Upf31
280 with KorB gave a robust ranking_confidence score of 0.64. Removal of the CTD from
281 Upf31 drastically reduced the predicted interaction between KorB and Upf31 to 0.274.
282 When pBP136 KorB was modelled with R751Upf31, which natively lacks a C-terminal
283 dimerization domain, a similarly low score of 0.252 was predicted. Given the globular
284 domain sequence similarity between pBP136 and R751 these results are highly
285 suggestive of pBP136 Upf31 interacting with KorB via the CTD (Figure 6B). Although
286 the Upf31 CTD does not line up with KorA CTD in the way that the CTD of TrbA does,
287 one can detect similarities like the Tyrosine at position 180 and the patch of
288 predominantly basic amino acids jumps. This predicted interaction might allow Upf31 to
289 reduce available KorB either when bound to DNA or free since we previously observed
290 that KorA and KorB could be copurified in the absence of DNA (McLean *et al.*, pending
291 submission).

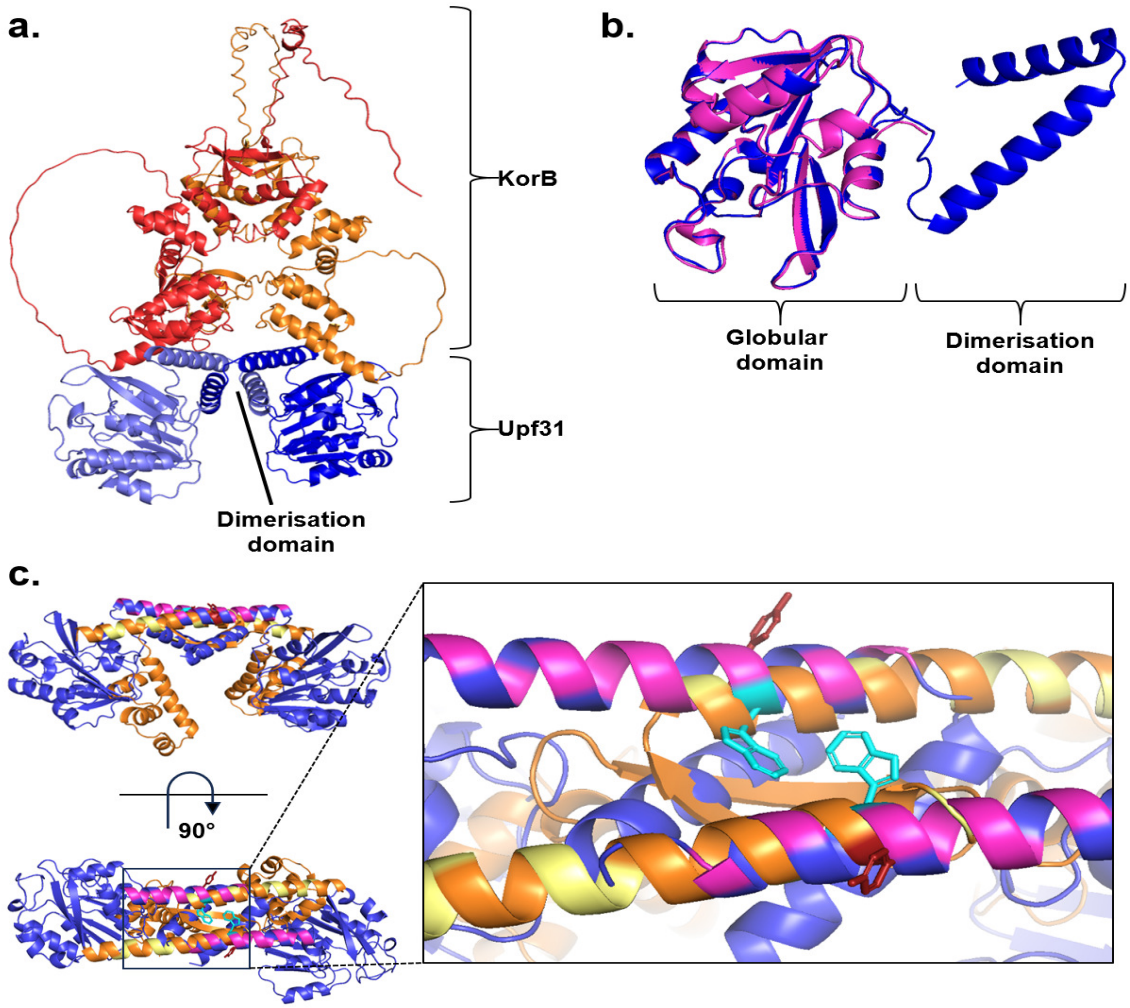


Figure 6. AlphaFold2 modeling reveals the potential interaction between pBP136 Upf31 and KorB is mediated by the C-terminal dimerization domain. (a) AlphaFold2 modelling of the pBP136 Upf31 homodimer (light and dark blue) with the pBP136 KorB homodimer (red and orange). **(b)** Superimposed AlphaFold2 models of pBP136 Upf31 (dark blue) and R751 Upf31 (magenta) highlighting the additional C-terminal dimerization domain 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.

295 ## Discussion

296 This work sheds new light on the biology and evolution of conjugative plasmids. We
297 demonstrate that the persistence of a conjugative broad-host-range plasmid from
298 *Bordetella pertussis* rapidly improved in *E. coli,* even in the absence of selection for the
299 plasmid. The associated drastic amelioration of the plasmid fitness cost to this host was
300 due to the inactivation of a previously uncharacterized plasmid gene*, upf31,* which
301 affected the expression of the plasmid backbone regulatory circuit including all four
302 major regulators as well as hundreds of chromosomal genes. This change in
303 transcriptome is likely initiated by Upf31 interacting with one of these major plasmid
304 regulators, KorB. Even though plasmids like pBP136 of the incompatibility group IncP
305 are notably persistent in many Proteobacteria[5,21–27], we show that different alleles of
306 plasmid genes like *upf31* may be more or less suited to plasmid persistence in particular
307 hosts. Understanding how this gene acts and why it is an advantage in some hosts will
308 expand our understanding of the genetic toolbox plasmids can exploit for their success.

309 The last decade has seen a flood of seminal studies regarding compensatory evolution
310 of plasmid fitness cost, including multiple studies showing that changes in chromosomal
311 regulatory systems can improve plasmid-host relations. One example is mutations of
312 the *fur* transcriptional regulator involved in iron uptake[9,50], while evolution of the
313 chromosomal *gacA/S* regulators[17] was observed to lower the fitness cost of plasmid
314 carriage within 48 hours of plating on agar[51]. A few studies have found compensatory
315 evolution via putative chromosomally-encoded helicases[18], in one case largely restoring
316 gene expression levels of the plasmid-host pair to that of the plasmid-free bacteria[52].
317 Amid these discoveries, the complex and tightly regulated plasmid transcriptional

318 regulatory systems have perhaps been overlooked as a key factor in plasmid-host
319 relations. One exception was a study which found that a plasmid improved persistence
320 via differential expression of a *parAB* operon, with backbone genes and plasmid
321 regulators otherwise remaining unchanged[20]. Our work shows that the regulatory
322 systems of *both* plasmids and chromosomes should be included in the growing list of
323 evolutionary pathways that can decrease the fitness cost of a plasmid.

324 The exact mechanism of Upf31 remains experimentally unvalidated but we hypothesize
325 a C-terminal dimerization domain interaction with plasmid regulator KorB. This is
326 supported by our use of AlphaFold2 to predict structures and interactions between
327 Upf31 and all other pBP136-encoded proteins, which identified KorB as a highly likely
328 target for Upf31 (Figure 6A). Furthermore, the CTD of Upf31 modeled very similar to the
329 CTD of KorA, which was previously shown to interact with KorB[47]. KorB is known to
330 interact co-operatively with the other plasmid backbone repressors[45,53,54]. Therefore, a
331 KorB-Upf31 interaction that reduces KorB's ability to interact with other repressors could
332 explain the observed weakening of KorB autogenous control as well as weaker operon
333 repression. It would also align with our result that plasmid R751 with a *upf31* allele that
334 natively lacks a CTD, was very persistent in *E. coli*. In sum, our study suggests that just
335 like KorA, Upf31 may also belong to the family of proteins that can interact with KorB,
336 and that its interaction may override the other interactions with concomitant
337 derepression of gene expression.

338 The Upf31-associated derepression of global plasmid regulators triggers a secondary
339 effect of large-scale differential gene expression. Thus, the next question becomes
340 which of these differentially expressed genes explains poor *E. coli* fitness. As this

341 negative fitness effect of *upf31* was only observed in the presence of IncPβ plasmids,
342 we focus here on changes in the plasmid transcriptome. First, of the three plasmids in
343 this study that contained full-length *upf31* homologs, only pALTS29 showed high
344 persistence in *E. coli,* which may suggest low fitness cost. Notably, this plasmid lacked
345 a toxin-antitoxin (TA) system found in nearly all IncPβ plasmids. This suggests that
346 overexpression of TA genes may explain poor *E. coli* fitness. However, the higher
347 persistence of pALTS29 could equally be explained by the Q82P amino acid difference
348 in Upf31 of pALTS29. Another possible explanation for the negative effect of Upf31 on
349 *E. coli* fitness is a set of four genes in the control region that were upregulated in the
350 presence of Upf31 (*kleA, kleE, klcA, klcB,* or '*kil' genes*). These genes have been
351 shown to have harmful effects on the growth of *E.coli* strains when not repressed by the
352 global Kor regulators[55–57], with *klcB* found harmful only in the presence of an IncP
353 plasmid[58]. Future studies will have to determine which of the overexpressed genes
354 have such a negative effect on *E coli* fitness and by which mechanism. Conversely, it is
355 left to be investigated if *upf31* may have presented an advantage in the original *B.*
356 *pertussis* host.

357 We next wish to report two important lessons we learned during this study. First, we
358 frequently observed rapid loss of *upf31* during routine laboratory cultivation due to
359 various deletion, prompting us to always sequence the region of *upf31* to confirm its
360 presence. This suggests that rapid evolution of costly plasmid genes may often go
361 unrecognized during cultivation and plasmid transfer between strains by matings and
362 electroporation. Our work aligns with previous findings of rapid (<48 hours)
363 compensatory evolution between plasmid-host pairs, even without known selection[51].

364 Second, our work showed that Upf31 is likely not a functional methylase in contrast to
365 automated annotations and computational predictions. This highlights the argument that
366 plasmid genes *without* experimental data should be assigned locus tags rather than
367 putative function-related gene names[59].

368 In conclusion, despite decades of great progress in plasmid biology and the pivotal role
369 of conjugative plasmids in antibiotic resistance dissemination, a substantial portion of
370 plasmids harbor multiple genes with unknown functions. Our study highlights the
371 importance of one such gene, which when inactivated led to a lasting host-plasmid pair,
372 effectively rescuing a plasmid from potential extinction[60] in a bacterial population. The
373 investigation of this specific gene revealed that this evolutionary rescue is closely linked
374 to alterations in the plasmids regulatory circuit. These findings underscore the
375 significance of identifying and experimentally validating uncharacterized plasmid genes
376 to understand how evolution underwrites the spread and persistence of plasmids in
377 bacterial communities.

378

379 ## Materials and Methods

380 ### Bacterial strains, plasmids and media

381 IncPβ plasmid pBP136 was discovered in a strain of *Bordetella pertussis* isolated from
382 a lethal case of infant whooping cough[35]. A kanamycin gene was later inserted to
383 generate pBP136Km (NCBI accession number NZ_OR146256.1), which provided a
384 selectable marker in the originally cryptic plasmid[56]. The expression vector pCW-LIC-
385 *upf31* was derived from pCW-LIC-*sacB* as described below in the cloning section.

386 *Escherichia coli* K-12 MG1655 was derived[57] from an isolate within the stool of a
387 diphtheria patient in 1925, and *E. coli* JM109 and DH5α[58] are later derivatives of this
388 strain with useful attributes relevant to cloning. We generated rifampicin and nalidixic
389 resistant clones of *E. coli* K-12 MG1655 by recovering resistant mutants after plating on
390 agar with the respective antibiotic. New constructs and initial, intermediate and final
391 populations of the plasmid persistence assays were all archived at -70 °C in 30%
392 glycerol.

393 All bacteria in this study were grown at 37°C in lysogeny broth (LB) shaken at 200 RPM
394 or on LB agar (LBA) plates. Strains were grown with 50 µg ml^{-1} kanamycin (km) or 100
395 µg ml^{-1} ampicillin (amp) when appropriate for plasmid maintenance. Expression vectors
396 were induced with 500 µM Isopropyl β- d-1-thiogalactopyranoside (IPTG) when
397 induction is mentioned in the text.

398 **Plasmid Persistence Assay**

399 All plasmid persistence assays were performed in triplicate and first grown overnight
400 (O/N) with kanamycin to select for initial plasmid maintenance (time T0). Daily serial
401 transfers of 4.9 µL of overnight culture into 5 mL of sterile broth *without* antibiotic was
402 then performed for ten more days (times T1-T10). Each day the cultures were diluted
403 10^{-6} in 1x phosphate buffered saline (PBS) and 100 uL was spread onto dilution plates.
404 The daily plasmid-containing fraction of the population was determined by replica
405 plating 52 randomly chosen colonies from the dilution plates onto LB plates with and
406 without kanamycin.

407 **Growth Assays**

408 We performed growth assays in batch culture to measure the effect of Upf31 on growth
409 rates and final densities. Strains were grown from freezer archives for 24 hours and
410 diluted 1:100, then grown another 24 hours and finally diluted 1:1000 in sterile media.
411 From these ten biological replicates each were loaded at a volume of 200uL into the
412 respective wells of a 96-well plate. A SPECTROstar Nano plate reader measured
413 optical density at 600nm every ten minutes for 23 hours with incubation at 37C and 500
414 RPM orbital shaking. The R package GrowthRates[61] v. 0.8.4 was used to calculate
415 maximum growth rate using default parameters.

416 **Low-Density Competition Assays**

417 The low-density competition assay was developed in this study to avoid plasmid transfer
418 during the assay. It contains two distinct steps: (*i*) determining the likely "conjugation-
419 free" time window for specific low donor and recipient densities, and (*ii*) performing a
420 traditional competition assay at low densities within the established time window.

421 To find the likely window of time before density is high enough for conjugation to occur,
422 we used K-12 MG1655 strains[34] with spontaneous mutations conferring resistance to
423 nalidixic acid (Nal) and rifampicin (Rif) (donor K-12N(pBP136Km) and recipient K-12R,
424 respectively). Donor and recipient were mixed1:1 at densities of 10^2/ml into 2.5 mL LB
425 shaken 200 RPM at 37°C. To select for possible transconjugants, the cultures were
426 mixed on hourly timepoints with another 2.5 mL of LB with Km and Rif to create a 5mL
427 mixture with 50 µg ml^{-1} Km and 50 µg ml^{-1} Rif. These were grown 24 hours after which
428 the earliest timepoint with turbidity was associated with the emergence of
429 transconjugants K-12R(pBP136Km). The first turbid culture emerged at hour six, and

430 therefore a five-hour window of likely conjugation-free growth was chosen for the
431 subsequent competition assays.

432 Traditional competition assays were then performed using K-12 vs. K-12(pBP136Km)
433 and K-12 vs. K-12(pBP136KmΔ*upf31*). Each experiment began with 1:1 mixture at
434 densities of 10^2/ml in LB without antibiotics and were shaken. Plating onto LBA occurred
435 at the start time and hour five endpoint (after an appropriate dilution), and the relative
436 increase in cell numbers (cfu/ml) was used as a proxy for relative fitness as follows:[62]

437 $$w = \frac{ln\left(\frac{A_f}{A_i}\right)}{ln\left(\frac{B_f}{B_i}\right)}$$

438 where w is the Malthusian relative fitness term, A_I and A_F are the initial and final plasmid
439 containing K-12 populations, respectively, and B_I and B_F are the initial and final plasmid-
440 free K-12 populations, respectively.

441 **Methylation Detection**

442 To test if *upf1* encodes a methyltransferase as computationally predicted we utilized the
443 base modification analysis provided by the Pacbio Sequel II DNA sequencer. The
444 methylase-free *E. coli* strain ER2796[43] with expression vector pCW-LIC-*upf31* was
445 tested against ER2796 (pCW-LIC-*sacB*) for differential base modification[36] on a Pacbio
446 Sequel II at the Arizona Genomic Institute in Tucson, AZ, USA. Both strains were grown
447 with antibiotic selection and IPTG induction before genomic extraction using the Sigma
448 GeneElute Bacterial kit. Base Modification Analysis was run on SMRT Link v.
449 10.2.0.133434 with a reference made by merging the *E. coli* assembly with the

450 respective plasmids and enabling “Find Modified Base Motifs” and “Consolidate Mapped
451 BAMs for IGV” options. All the other options were left as default.

452 **RNA-Seq**

453 To understand how *upf31* and pBP136Km interact to influence *E. coli* fitness, we
454 performed RNA-seq. Strains K-12WT (pBP136Km) and K-12WT (pBP136KmΔ*upf31)*
455 were grown overnight in LB with Km and pelleted down. The samples were sent to
456 Zymo Research for Total RNA-Seq Service. Sequencing libraries were constructed from
457 total RNA sample and were prepared using the Zymo-Seq RiboFree Total RNA Library
458 Prep Kit (Cat # R3000) according to the manufacturer’s instructions
459 (https://www.zymoresearch.com/products/zymo-seq-ribofree-total-rna-library-kit). RNA-
460 Seq libraries were sequenced on an Illumina NovaSeq to a sequencing depth of at least
461 30 million read pairs (150 base paired-end sequencing) per sample.
462 A custom bioinformatics pipeline began with TrimGalore![63] v. 0.67 to remove adapters
463 and low-quality reads. Bowtie2[64] v. 2.4.5 then mapped reads to merged chromosomal
464 and plasmid references. These were counted using featureCounts[65] v. 2.0.1, the output
465 of which was input unto DESeq2[66] v. 1.38.1, for final calculations of relative expression.

466 **DNA Sequencing**

467 Whole genome sequencing was performed by SeqCenter, LLC in Pittsburg, PA, USA.
468 Genomic DNA extractions were performed using Sigma GeneElute Bacterial kit. The
469 libraries were prepared using the Illumina DNA Prep kit with IDT 10bp UDI indices and
470 sequenced on an Illumina NextSeq 2000 with 2x151bp reads. Demultiplexing, quality
471 control and adapter trimming was initially performed using Illumina’s bcl-convert v.

472 3.9.3. TrimGalore! v. 0.6 was additionally used to remove remaining adapters and low-
473 quality reads. Breseq[67] v. 0.36.0 was used to identify mutations between ancestral and
474 evolved strains.

475 **Cloning**

476 pCW-LIC (hereafter pCW-LIC-*sacB*) was a gift from Cheryl Arrowsmith (Addgene
477 plasmid #26098; http://n2t.net/addgene:26098; RRID:Addgene_26098). Plasmid pCW-
478 LIC-*sacB* was double cut at the *NdeI* and *HindIII* restrictions sites, which created a
479 4,962 base pair (bp) linearized backbone and a 2,302 bp linear fragment containing
480 *sacB* and its promoter. These were separated and the linearized backbone recovered
481 from an agarose gel using a Thermo Scientific (TS) GeneJET Gel Extraction Kit. Next,
482 gene *upf31* was amplified from pBP136Km miniprep using upstream primer 5'-
483 *GGTGGT*CATATGTCCAGGAAGAAGGCCATGAG-3' and downstream primer 5'-
484 *GGTGG*TAAGCTTCTACTCGGCCGCCTCTAG-3' (flanking sequences for restriction
485 enzymes in italics, *NheI* and *HindIII* sites underlined, respectively). The ends of the
486 *upf31* amplicon were then double digested at the *NdeI* and *HindIII* sites, and the small
487 digested ends removed using a TS GeneJET PCR Purification Kit. The *upf31* segment
488 and linearized pCW-LIC backbone were joined using T4 DNA ligase and the now-
489 circularized pCW-LIC-*upf31* was electroporated into DH5α using standard methods[68].

490 **Upf31 protein expression**

491 Induction cultures of pCW-LIC-upf31 in *E. coli* K-12 M1655 were started from an
492 overnight liquid culture of LB (Lennox) with 100 µg/ml ampicillin (LBAmp). Twenty-five
493 milliliter LB Amp cultures were initiated at OD600 0.05 using the liquid overnight culture

494 to serve as the induction cultures. The 25 milliliter cultures were grown at 37° C while
495 shaking at 200 rpm. When the 25 milliliter cultures reached OD600 > 1.0, one milliliter of
496 culture was removed, and the cells were harvested by centrifugation for four minutes at
497 16,000 x g at room temperature in an Eppendorf 5425 centrifuge. This material served
498 as the pre-induction (or 0 hour) time point for each culture. ITPG was then added to
499 each induction culture to a final concentration of 0.1 mM, 0.5 mM, or 1 mM and growth
500 was monitored at OD600. At the time points indicated in Figure S2, one milliliter of
501 culture was removed from each culture and cells were isolated as above. Cells were
502 resuspended in 1x Tris-Glycine running buffer (25 mM Tris-Cl, pH 8.3, 192 mM glycine,
503 and 0.1% SDS) at 100 microliters per 1.0 = OD600 cells. Laemmli buffer was added to
504 the resuspended cell pellets to a final concentration of 1x and heated 10 minutes at >
505 98° C prior to gel loading.

506 **Upf31 SDS-Page Gel**

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

514 **Computational Modeling of Interactions between Upf31 and pBP136-encoded**
515 **Proteins**

An '*in silico* pulldown' of Upf31 was carried out using the LazyAF pipeline[69]. 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).

Acknowledgements

We acknowledge Tùng Lê at the John Innes Centre (UK) for his work on the *in silico* Upf31 pulldown. 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. We 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.

538 ## Funding Resources

539 E.M.T. received support for this work from National Institute of Allergy and Infectious
540 Diseases Extramural Activities grant no. R01AI084918 from the National Institutes of
541 Health. C.A.E. was supported by the National Science Foundation Graduate Research
542 Fellowship grant no. DGE-2019265372 as well as the Bioinformatics and Computational
543 Biology (BCB) Fellowship and the Paul Joyce Memorial BCB Fellowship Endowment at
544 the University of Idaho. O.K. was supported by the National Science Foundation
545 Graduate Research Fellowship grant no. DGE-1762114. This work is supported by the
546 Wellcome Trust Investigator grant 221776/Z/2/Z to Tùng Lê that supported T.C.M, and
547 by the BBSRC funded Institute Strategic Program Harnessing Biosynthesis for
548 Sustainable Food and Health (HBio) (BB/X01097X/1). The funders had no role in study
549 design, data collection and analysis, decision to publish, or preparation of the
550 manuscript.

551

552 ## References

553 1. Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nature*
554 *Reviews Genetics* **16**, 472–482 (2015).
555 2. Norman, A., Hansen, L. H. & Sørensen, S. J. Conjugative plasmids: vessels of the communal gene
556 pool. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **364**,
557 2275–89 (2009).
558 3. San Millan, A. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. *Trends in*
559 *Microbiology* **26**, 978–985 (2018).

560 4. Murray, C. J. L. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic
561 analysis. *The Lancet* **399**, 629–655 (2022).
562 5. De Gelder, L., Ponciano, J. M., Joyce, P. & Top, E. M. Stability of a promiscuous plasmid in different
563 hosts: No guarantee for a long-term relationship. *Microbiology* **153**, 452–463 (2007).
564 6. Modi, R. I. & Adams, J. Coevolution in bacterial-plasmid populations. *Evolution* **45**, 656–667 (1991).
565 7. Bouma, J. E. & Lenski, R. E. Evolution of a bacteria/plasmid association. *Nature* **335**, 351–352 (1988).
566 8. Dahlberg, C. & Chao, L. Amelioration of the Cost of Conjugative Plasmid Carriage in Eschericha coli
567 K12. *Genetics* **165**, 1641–1649 (2003).
568 9. Stalder, T. *et al.* Emerging patterns of plasmid-host coevolution that stabilize antibiotic resistance.
569 *Scientific Reports* **7**, 1–10 (2017).
570 10. Bottery, M. J., Wood, A. J. & Brockhurst, M. A. Adaptive modulation of antibiotic resistance through
571 intragenomic coevolution. *Nat Ecol Evol* **1**, 1364–1369 (2017).
572 11. Sota, M. *et al.* Shifts in the host range of a promiscuous plasmid through parallel evolution of its
573 replication initiation protein. *ISME Journal* **4**, 1568–1580 (2010).
574 12. Hughes, J. M. *et al.* The role of clonal interference in the evolutionary dynamics of plasmid-host
575 adaptation. *mBio* **3**, e00077-12 (2012).
576 13. Yano, H. *et al.* Evolved plasmid-host interactions reduce plasmid interference cost. *Molecular*
577 *Microbiology* **101**, 743–756 (2016).
578 14. De Gelder, L., Williams, J. J., Ponciano, J. M., Sota, M. & Top, E. M. Adaptive Plasmid Evolution
579 Results in Host-Range Expansion of a Broad-Host-Range Plasmid. *Genetics* **178**, 2179–2190 (2008).
580 15. Porse, A., Schønning, K., Munck, C. & Sommer, M. O. A. Survival and Evolution of a Large Multidrug
581 Resistance Plasmid in New Clinical Bacterial Hosts. *Molecular Biology and Evolution* **33**, 2860–2873
582 (2016).

583 16. Jordt, H. *et al.* Coevolution of host–plasmid pairs facilitates the emergence of novel multidrug
584 resistance. *Nat Ecol Evol* **4**, 863–869 (2020).
585 17. Harrison, E., Guymer, D., Spiers, A. J., Paterson, S. & Brockhurst, M. A. Parallel Compensatory
586 Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum. *Current Biology* **25**,
587 2034–2039 (2015).
588 18. Loftie-Eaton, W. *et al.* Compensatory mutations improve general permissiveness to antibiotic
589 resistance plasmids. *Nature Ecology and Evolution* **1**, 1354–1363 (2017).
590 19. Millan, A. S. *et al.* Positive selection and compensatory adaptation interact to stabilize non-
591 transmissible plasmids. *Nature Communications* **5**, 5208 (2014).
592 20. Hall, J. P. J. *et al.* Plasmid fitness costs are caused by specific genetic conflicts enabling resolution by
593 compensatory mutation. *PLOS Biology* **19**, e3001225 (2021).
594 21. Schmidhauser, T. J. & Helinski, D. R. Regions of broad-host-range plasmid RK2 involved in replication
595 and stable maintenance in nine species of gram-negative bacteria. *J Bacteriol* **164**, 446–455 (1985).
596 22. Shintani, M., Takahashi, Y., Yamane, H. & Nojiri, H. The behavior and significance of degradative
597 plasmids belonging to Inc groups in Pseudomonas within natural environments and microcosms.
598 *Microbes Environ* **25**, 253–265 (2010).
599 23. Yano, H., Deckert, G. E., Rogers, L. M. & Top, E. M. Roles of long and short replication initiation
600 proteins in the fate of IncP-1 Plasmids. *Journal of Bacteriology* **194**, 1533–1543 (2012).
601 24. Jain, A. & Srivastava, P. Broad host range plasmids. *FEMS Microbiology Letters* **348**, 87–96 (2013).
602 25. Yano, H. *et al.* Host range diversification within the IncP-1 plasmid group. *Microbiology (United*
603 *Kingdom)* **159**, 2303–2315 (2013).
604 26. Klümper, U. *et al.* Broad host range plasmids can invade an unexpectedly diverse fraction of a soil
605 bacterial community. *ISME J* **9**, 934–945 (2015).

606 27. Kottara, A., Hall, J. P. J., Harrison, E. & Brockhurst, M. A. Variable plasmid fitness effects and mobile
607 genetic element dynamics across Pseudomonas species. *FEMS microbiology ecology* **94**, (2018).
608 28. Thorsted, P. B. *et al.* Complete sequence of the IncPβ plasmid R751: Implications for evolution and
609 organisation of the IncP backbone. *Journal of Molecular Biology* **282**, 969–990 (1998).
610 29. Thomas, C. M. Paradigms of plasmid organization. *Molecular Microbiology* **37**, 485–491 (2000).
611 30. Sen, D., Brown, C. J., Top, E. M. & Sullivan, J. Inferring the evolutionary history of INCP-1 plasmids
612 despite incongruence among backbone gene trees. *Molecular Biology and Evolution* **30**, 154–166
613 (2013).
614 31. Rajasekar, K. V. *et al.* Flexibility of KorA, a plasmid-encoded, global transcription regulator, in the
615 presence and the absence of its operator. *Nucleic Acids Research* **44**, 4947–4956 (2016).
616 32. Bingle, L. E. H. *et al.* Flexibility in repression and cooperativity by KorB of broad host range IncP-1
617 plasmid RK2. *Journal of Molecular Biology* **349**, 302–316 (2005).
618 33. Bingle, L. E. H. & Thomas, C. M. Regulatory circuits for plasmid survival. *Current Opinion in*
619 *Microbiology* **4**, 194–200 (2001).
620 34. Sota, M. & Top, E. M. Host-specific factors determine the persistence of IncP-1 plasmids. *World*
621 *Journal of Microbiology and Biotechnology* **24**, 1951–1954 (2008).
622 35. Sota, M. *et al.* Region-specific insertion of transposons in combination with selection for high
623 plasmid transferability and stability accounts for the structural similarity of IncP-1 plasmids. *Journal*
624 *of Bacteriology* **189**, 3091–3098 (2007).
625 36. Kamachi, K. *et al.* Plasmid pBP136 from Bordetella pertussis represents an ancestral form of IncP-1β
626 plasmids without accessory mobile elements. *Microbiology* **152**, 3477–3484 (2006).
627 37. Zhong, X., Kro˙l, Jarosław E., Top, E. M. & Krone, S. M. Accounting for mating pair formation in
628 plasmid population dynamics. *Journal of Theoretical Biology* **262**, 711–719 (2010).

629 38. Kosterlitz, O. *et al.* Estimating the transfer rates of bacterial plasmids with an adapted Luria–
630 Delbrück fluctuation analysis. *PLoS Biol* **20**, e3001732 (2022).
631 39. Drønen, A. K., Torsvik, V., Goksøyr, J. & Top, E. M. Effect of mercury addition on plasmid incidence
632 and gene mobilizing capacity in bulk soil. *FEMS Microbiology Ecology* **27**, 381–394 (1998).
633 40. Law, A. *et al.* Biosolids as a Source of Antibiotic Resistance Plasmids for Commensal and Pathogenic
634 Bacteria. *Front Microbiol* **12**, 606409 (2021).
635 41. Anton, B. P. *et al.* Complete genome sequence of ER2796, a DNA methyltransferase-deficient strain
636 of Escherichia coli K-12. *PLoS ONE* **10**, (2015).
637 42. Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome*
638 *Biology* **4**, R60 (2003).
639 43. Pansegrau, W. *et al.* Complete Nucleotide Sequence of Birmingham IncPα Plasmids: Compilation
640 and Comparative Analysis. *Journal of Molecular Biology* **239**, 623–663 (1994).
641 44. Adamczyk, M. & Jagura-Burdzy, G. Spread and survival of promiscuous IncP-1 plasmids. *Acta*
642 *Biochim Pol* **50**, 425–453 (2003).
643 45. Kostelidou, K., Jones, A. C. & Thomas, C. M. Conserved C-terminal region of global repressor KorA of
644 broad-host-range plasmid RK2 is required for co-operativity between KorA and a second RK2 global
645 regulator, KorB11Edited by J. Karn. *Journal of Molecular Biology* **289**, 211–221 (1999).
646 46. Zatyka, M., Jagura-Burdzy, G. & Thomas, C. M. Transcriptional and translational control of the genes
647 for the mating pair formation apparatus of promiscuous IncP plasmids. *J Bacteriol* **179**, 7201–7209
648 (1997).
649 47. Bingle, L. E. H. *et al.* A single aromatic residue in transcriptional repressor protein KorA is critical for
650 cooperativity with its co-regulator KorB. *Mol Microbiol* **70**, 1502–1514 (2008).
651 48. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589
652 (2021).

653 49. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682
654 (2022).
655 50. Hassan, H. & Troxell, B. Transcriptional regulation by Ferric Uptake Regulator (Fur) in pathogenic
656 bacteria. *Frontiers in Cellular and Infection Microbiology* **3**, (2013).
657 51. Hall, J. P. J., Wright, R. C. T., Guymer, D., Harrison, E. & Brockhurst, M. A. Extremely fast
658 amelioration of plasmid fitness costs by multiple functionally diverse pathways. *Microbiology* **166**,
659 56–62 (2020).
660 52. Millan, A. S., Toll-Riera, M., Qi, Q. & MacLean, R. C. Interactions between horizontally acquired
661 genes create a fitness cost in Pseudomonas aeruginosa. *Nature Communications* **6**, 6845 (2015).
662 53. Kostelidou, K. & Thomas, C. M. The hierarchy of KorB binding at its 12 binding sites on the broad-
663 host-range plasmid RK2 and modulation of this binding by IncC1 protein11Edited by J. Karn. *Journal*
664 *of Molecular Biology* **295**, 411–422 (2000).
665 54. Zatyka, M., Bingle, L., Jones, A. C. & Thomas, C. M. Cooperativity between KorB and TrbA repressors
666 of broad-host-range plasmid RK2. *J Bacteriol* **183**, 1022–1031 (2001).
667 55. Figurski, D. H., Pohlman, R. F., Bechhofer, D. H., Prince, A. S. & Kelton, C. A. Broad host range
668 plasmid RK2 encodes multiple kil genes potentially lethal to Escherichia coli host cells. *Proc. Natl.*
669 *Acad. Sci. U.S.A.* **79**, 1935–1939 (1982).
670 56. Kornacki, J. A., Chang, C. H. & Figurski, D. H. kil-kor Regulon of promiscuous plasmid RK2: Structure,
671 products, and regulation of two operons that constitute the kilE locus. *Journal of Bacteriology* **175**,
672 5078–5090 (1993).
673 57. Larsen, M. H. & Figurski, D. H. Structure, expression, and regulation of the kilC operon of
674 promiscuous IncP alpha plasmids. *J Bacteriol* **176**, 5022–5032 (1994).

675 58. Bhattacharyya, A. & Figurski, D. H. A small protein-protein interaction domain common to KlcB and
676 global regulators KorA and TrbA of promiscuous IncP plasmids11Edited by M. Belfort. *Journal of*
677 *Molecular Biology* **310**, 51–67 (2001).
678 59. Thomas, C. M. *et al.* Annotation of plasmid genes. *Plasmid* **91**, 61–67 (2017).
679 60. Gomulkiewicz, R. & Holt, R. D. When does Evolution by Natural Selection Prevent Extinction?
680 *Evolution* **49**, 201–207 (1995).
681 61. Hall, B. G., Acar, H., Nandipati, A. & Barlow, M. Growth Rates Made Easy. *Molecular Biology and*
682 *Evolution* **31**, 232–238 (2014).
683 62. Wiser, M. J. & Lenski, R. E. A Comparison of Methods to Measure Fitness in Escherichia coli. *PLOS*
684 *ONE* **10**, e0126210 (2015).
685 63. TrimGalore! https://github.com/FelixKrueger/TrimGalore
686 64. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359
687 (2012).
688 65. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning
689 sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
690 66. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq
691 data with DESeq2. *Genome Biology* **15**, 550 (2014).
692 67. Barrick, J. E. *et al.* Identifying structural variation in haploid microbial genomes from short-read
693 resequencing data using breseq. *BMC Genomics* **15**, 1039 (2014).
694 68. Sambrook, Joseph. & Russell, D. W. Molecular Cloning : A Laboratory Manual. *Cold Spring Harbor*
695 *Laboratory Press*, Cold Spring Harbor, N.Y., (2001).
696 69. McLean, T. C. LazyAF, a pipeline for accessible medium-scale in silico prediction of protein-protein
697 interactions. 2024.01.29.577767 Preprint at https://doi.org/10.1101/2024.01.29.577767 (2024).
698