Segregational drift and the interplay between plasmid copy number and evolvability

Judith Ilhan[1], Anne Kupczok[1], Christian Woehle[1], Tanita Wein[1], Nils F. Hülter[1], Philip Rosenstiel[2], Giddy Landan[1], Itzhak Mizrahi[3], Tal Dagan[1*]

[1] Institute of Microbiology and [2] Institute of Clinical Molecular Biology, Kiel University, 24118 Kiel, Germany. [3] The Department of Life Sciences & the National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, Israel.

[*] Corresponding author

Abstract

The ubiquity of plasmids in all prokaryotic phyla and habitats and their ability to transfer between cells marks them as prominent constituents of prokaryotic genomes. Many plasmids are found in their host cell in multiple copies. This leads to an increased mutational supply of plasmid-encoded genes and genetically heterogeneous plasmid genomes. Nonetheless, the segregation of plasmid copies into daughter cells during cell division is considered to occur in the absence of selection on the plasmid alleles. Here, we investigate the implications of random genetic drift of multicopy plasmids during cell division – termed here *segregational drift* – to plasmid evolution. Performing experimental evolution of low- and high-copy non-mobile plasmids in *Escherichia coli*, we find that the evolutionary rate of multicopy plasmids does not reflect the increased mutational supply expected according to their copy number. In addition, simulated evolution of multicopy plasmid alleles demonstrates that segregational drift leads to increased loss frequency and extended fixation time of plasmid mutations in comparison to haploid chromosomes. Furthermore, an examination of the experimentally evolved hosts reveals a significant impact of the plasmid type on the host chromosome evolution. Our study demonstrates that segregational drift of multicopy plasmids interferes with the retention and fixation of novel plasmid variants. Depending on the selection pressure on newly emerging variants, plasmid genomes may evolve slower than haploid chromosomes, regardless of their higher mutational supply. We suggest that plasmid copy number is an important determinant of plasmid evolvability due to the manifestation of segregational drift.

Introduction

Plasmids are genetic elements that colonize prokaryotic cells where they replicate independently of the chromosome. They are considered as a major driving force in prokaryotic ecology and evolution as they can be transferred between cells, making them potent agents of lateral gene transfer (Lederberg and Tatum 1946; Thomas and Nielsen 2005) and microbial warfare (Czárán et al. 2002; Majeed et al. 2011). Plasmids may contribute to the evolution of their host via the supply of genes encoding for new functions that are beneficial under growth limiting conditions (e.g., antibiotics resistance (Porse et al. 2016), resistance to heavy metals (Gullberg et al. 2014; Dziewit et al. 2015)) or during adaptation to new habitats (e.g., catabolic functions (Schmidt et al. 2011)). Nonetheless, the importance of plasmids goes beyond microbial evolution as they are widely used as vectors for genetic engineering (Simon et al. 1983; Bevan 1984) as well as for applications in biotechnology (Ullrich et al. 1977) and synthetic biology (Shetty et al. 2008). Therefore, understanding plasmid biology and evolution is of major interest for basic microbiology research and biological applications.

Plasmids differ from the chromosomal genetic component in several key aspects, including a relatively small genome size and a variable ploidy level. Plasmids are often found in multiple copies in the cell where the plasmid copy number (PCN) depends on the replicon type, the host genetics, and the environmental conditions (Nordström 2006; Santos-Lopez et al. 2017). The copy number of natural plasmids ranges between 1-15 for low-copy plasmids (Bazaeal and Helinski 1968) and can reach 200 copies for high-copy plasmids (Projan et al. 1987). Plasmids used for biotechnological applications are often modified to have a higher copy number, e.g., 500-700 (Vieira and Messing 1982; Rosano and Ceccarelli 2014), in order to maximize the protein expression level of focal genes. Notably, genes encoded on multicopy plasmids residing within a single host cell may comprise multiple alleles due to independent emergence of mutations in the plasmid copies (Bedhomme et al. 2017; Rodriguez-Beltran et al. 2018). An increased plasmid copy number has the potential to

67 elevate the probability for the emergence of novel mutations in the gene open reading frame.
68 Consequently, the mutational supply of plasmid-encoded genes is increased for multicopy
69 plasmids. Indeed, it has been previously shown that multicopy plasmids have the potential to
70 enable rapid evolution of plasmid-encoded antibiotic resistance genes under strong selective
71 conditions for the antibiotic resistance (San Millan et al. 2016).

72 The inheritance of multicopy plasmids to daughter cells depends on their segregation
73 mechanism. Some plasmids are equipped with an active partition system. Such systems
74 comprise plasmid-encoded proteins that function in the translocation of plasmid copies
75 toward the cell poles prior to cell division, thereby ensuring stable inheritance (e.g., *parABS*
76 system (Abeles et al. 1985; Baxter and Funnell 2014)). In the absence of a partition system,
77 plasmid segregation into daughter cells depends on the physical distribution of plasmid
78 copies in the cell during cell division (Wang 2017). Notably, plasmid segregation into
79 daughter cells in both routes is considered to occur in the absence of selection on the
80 plasmid allele pool. Consequently, the dynamics of multicopy plasmid alleles in the
81 population has a constant component of random genetic drift. In other words, the distribution
82 of plasmid alleles into daughter cells is independent of the allele impact on the host fitness,
83 also under selective conditions for a plasmid-encoded trait. Here we term this phenomenon
84 *segregational drift*. Previous studies suggested that segregational drift plays a role in the
85 evolution of eukaryotic organelles – mitochondria and plastids (Birky 2001). Similar to
86 multicopy plasmids, the mitochondrion organelle is found in multiple copies in the cell (e.g.,
87 ~1,700 in mammalian cells (Robin and Wong 1988)), and in addition, each mitochondrion
88 organelle may harbour multiple copies of the mitochondrial genome (e.g., up to 10 copies
89 (Satoh 1991)). Thus, the mutational supply on the mitochondrial genome is larger than the
90 mitochondrial genome size by several orders of magnitude. The combination of high
91 mutational supply on the mitochondrion genome and random segregation of the mitochondria
92 during cell division lead to genetic heterogeneity of the mitochondrial genomes within a
93 single cell (termed 'mitochondrion heteroplasmy' (Birky 2001)). Notably, many of the
94 mitochondrial alleles are found in very low frequency, thus the mitochondrial genetic

95 heterogeneity is observed only with the application of deep-sequencing approaches (e.g., He
96 et al. 2010). The consequences of segregational drift to the evolution of multicopy plasmids
97 have been so far overlooked. Population genetics theory postulates that the effect of random
98 genetic drift on allele dynamics is inversely correlated with the effective population size. Thus
99 in a larger population, the impact of genetic drift on the dynamics of allele frequency is
100 reduced (see Lanfear et al. 2014 for review). Thus, the implications of segregational drift to
101 plasmid genome evolution are expected to be dependent on the plasmid population size
102 within the cell, i.e., the plasmid copy number. Here we investigate plasmid genome evolution
103 in light of the segregational drift using experimental evolution and *in silico* simulation of allele
104 dynamics in the population.

105 **Results**

106 ***Experimental evolution.***

107 In our study we compare the rate of evolution between low-copy and high-copy plasmid
108 replicons using an experimental evolution approach. In the experiment, two model plasmids
109 having low- or high-copy number were evolved in an *E. coli* host under conditions selecting
110 for the plasmid presence. The low-copy model plasmid (pLC, 5.2 kb; Fig. 1) is derived from
111 plasmid pBBR1 (Antoine and Locht 1992) that is known to stably replicate in a broad host
112 range despite lacking an active partitioning mechanism. The plasmid encodes an ampicillin-
113 resistance gene (*bla*; β-lactamase) and is non-mobile (i.e., it is not transmissible via
114 conjugation). The average pLC copy number in the wild-type *E. coli* host was 8.0 ± 0.78 (*SD*,
115 *n* = 12; using qPCR), which is within the range calculated previously for pBBR1 (Jahn et al.
116 2016). The accessory part of the high copy number model plasmid (pHC, 4.8 kb; Fig. 1) is
117 identical to that of plasmid pLC, but the backbone comprises a pUC origin of replication
118 (Vieira and Messing 1982), that is a derivative of the *ori* of plasmid ColE1 (Hershfield et al.
119 1974). The copy number of plasmid pHC in the wild-type *E. coli* host was approximately 80-
120 fold higher than that of plasmid pLC (715.56 ± 27.4 *SD*, *n* = 12; using qPCR). This is within
121 the typical copy number range of the pUC plasmid backbone (Rosano and Ceccarelli 2014).

122 In addition to PCN, our experimental evolution setup accounts for two known determinants of
123 evolutionary dynamics: mutation rate and selection pressure. To vary the mutation rate, the
124 plasmids were evolved in two host genotypes, *E. coli* K-12 substrain MG1655 (wild-type, wt)
125 and a hypermutator derivative of the wild-type strain (Δ*mutS*) in which mutations are
126 expected to arise faster in comparison to the wild-type (e.g., 33-fold (Arjan et al. 1999)). The
127 populations were evolved under two different environmental regimes: at 37°C and at 42°C.
128 Growth of *E. coli* at 42°C is expected to entail an acclimation to the elevated temperature
129 (Guyot et al. 2014), and it has been shown to involve genomic adaptation (Deatherage et al.
130 2017). The experiment was conducted with six replicates for each combination of the three
131 experimental evolution factors – plasmid replicon type, host genotype and growth
132 temperature. The populations were evolved for 800 generations in continuous culture under
133 selective conditions for ampicillin-resistance. At the end of the experiment, samples of
134 ancestral and evolved populations were sequenced using a population sequencing approach.
135 Notably, the two model plasmids differ in their stability within the host. While the pLC replicon
136 remained unaltered compared to its pBBR1 ancestor, the high copy number of plasmid pHC
137 comes at the cost of lower plasmid stability due to deletions of stability factors during
138 construction of the pUC replicon (e.g., lack of the multimer resolution site *cer* originally
139 present in plasmid ColE1). We assesed the stability of both plasmids in the *E. coli* wild-type
140 host using the same selective constraints for the plasmid presence as in the evolution
141 experiment. No plasmid-free segregants of pLC-wt were observed after overnight growth (n =
142 5), whereas the proportion of pHC-free segregants ranged between 1% and 39% (n = 5; see
143 Methods). Plasmid loss during the evolution experiment has the potential to result in a
144 proportion of plasmid-free cells in the total population. Such plasmid-free cells have a
145 collective resistance to ampicillin thanks to the presence of β-lactamase secreted from
146 plasmid carrying hosts in the environment (Vega and Gore 2014). Therefore, we quantified
147 the proportion of plasmid-hosts in the evolved populations using replica plating of single
148 colonies. This showed that pLC-hosts accounted for 89% to 100% of the populations in all
149 conditions and replicates (Supplementary Table S1). The proportion of pHC-hosts in the

150 evolved wild-type populations ranged between 1% and 78%, while the range of plasmid
151 carrying hosts in the evolved hypermutator populations was between 35% and 98%
152 (Supplementary Table S1). In order to increase the resolution of the existing plasmid allele
153 pool, we sequenced the genomes of each pHC population using two different sampling
154 strategies. In addition to the sequencing of total populations (i.e., plasmid host cells and
155 plasmid-free segregants), the genome of each pHC population was sequenced from pooled
156 samples of 100 colonies of plasmid carrying hosts (see Methods). In the following, the
157 resulting genetic variants reported for pHC populations are based on the pooled colonies
158 data unless stated otherwise.

159 ***Plasmid genome evolution.***

160 For the identification of emerging mutations, sequencing reads of the ancestral and evolved
161 populations were aligned to the reference *E. coli* MG1655 genome and to the two plasmids.
162 Sequencing coverage varied among the populations and the replicons: the chromosomal
163 coverage ranged between 104x and 300x, while the average coverage on the plasmids was
164 between 325x and 28,054x (Supplementary Table S1). The high coverage on the plasmid
165 genomes increases the detection sensitivity for rare and low-frequency variants. Nonetheless,
166 in order to compare the number of mutations across replicons and populations, the
167 sequencing reads of each replicon (i.e., chromosome and plasmids) were subsampled to
168 match the lowest-coverage dataset unless stated otherwise. Furthermore, variants occurring
169 in both ancestral and evolved populations were excluded.

170 A comparative genomic analysis of the subsampled sequencing data revealed two
171 point mutations on the pLC plasmid (Table 1). Both mutations are located in an intergenic
172 region and their allele frequency (AF) is lower than 0.10 (Supplementary Table S2). The
173 paucity of plasmid variants should be evaluated against the background of variants on the
174 host chromosome, where we observed a total of 1,394 point mutations. We calculated the
175 expected frequency of point mutations on the two plasmids using the chromosomal
176 substitution frequency, the plasmid genome size and the plasmid copy number. This shows

177 that mutations are unlikely to occur on either of the plasmid types evolved in the wild-type
178 host due to the combination of low substitution rate and the plasmid small genome size
179 (Table 1). In contrast, the combination of high genetic polymorphism in the hypermutator host
180 with the high copy number of pHC are likely to result in observable genetic variants in the
181 pHC plasmid within the time span of our experiment (Table 1 and Supplementary Table S1).
182 Nonetheless, the subsampled sequencing data for pHC did not reveal any point mutations
183 occurring on that plasmid. Notably, we detected two single nucleotide insertions in pHC
184 intergenic regions with an allele frequency of AF = 0.3 and AF = 0.35. In order to increase
185 the resolution for the detection of presumably low-frequency point mutations, we applied the
186 same variant calling pipeline to the sequencing read datasets without subsampling (i.e., full
187 sequencing coverage). This analysis revealed one point mutation on plasmid pHC at an
188 intergenic position with a sequencing depth of ~9,600 and a low allele frequency of AF = 0.03
189 (Supplementary Table S2). In contrast, considering the full sequencing depth for plasmid
190 pLC, which is on average ~10x lower than that of plasmid pHC, we observed two additional
191 intergenic point mutations (Supplementary Table S2). Thus, despite the high sequencing
192 depth, the observed number of point mutations on plasmid pHC does not reach the expected
193 frequency of point mutations, even when the full sequence coverage is included in the
194 analysis. On the other hand, the number of point mutations detected on plasmid pLC is
195 higher than what we expected (Table 1).

196 The number of point mutations on the pHC plasmid was further validated by the
197 analysis of the total pHC population sequencing data (i.e., samples were prepared without
198 selection for plasmid hosts). Considering the full sequencing coverage, we detected one
199 additional point mutation on the plasmid. The intergenic point mutation was observed with a
200 low allele frequency of AF = 0.04 (Supplementary Table S2) in one of the pHC populations
201 having the hypermutator host genotype. In addition, we compared the number of point
202 mutations observed on the chromosome in the pooled pHC host populations to that of the
203 total population sequencing data (Supplementary Table S1). The comparison shows that the
204 wild-type populations include, on average, two additional point mutations per population in

205 comparison to the corresponding data of pooled plasmid host colonies. The comparison of
206 hypermutator populations shows that the total population sequencing data includes, on
207 average, 14 point mutations per population less than the pooled colonies sequencing data.
208 Overall, the above comparison between the two sequencing approaches demonstrates that
209 the number of point mutations observed in the pooled host colonies data is representative
210 both for the number of variants on the pHC plasmid and the chromosome.

211 This discrepancy between the observed and expected frequency of point mutations
212 on the pHC plasmid can be explained under the segregational drift hypothesis as due to the
213 rapid loss of most plasmid mutations. Our analysis further reveals that the allele frequency of
214 plasmid variants is considerably lower in comparison to that of chromosomal variants. Of the
215 total chromosomal point mutations, 30% reached fixation within the population (AF ≥ 0.9),
216 whereas point mutations on plasmids remain at a low frequency (AF ≤ 0.09; Supplementary
217 Fig. S2). The stark difference in the allele frequency distribution between the plasmid and
218 chromosomal alleles indicates that the increase in plasmids allele frequency is considerably
219 slower in comparison to chromosomal alleles. This observation is in agreement with our
220 hypothesis according to which mutations that emerge on multicopy plasmid genomes are
221 repeatedly 'diluted' in the population due to segregational drift.

222 ***Plasmid population dynamics.***

223 What are the theoretical expectations for the impact of segregational drift on the dynamics of
224 plasmid alleles in comparison to chromosomal alleles? To address this question we
225 simulated plasmid and chromosome allele dynamics. For this purpose, we applied the
226 standard haploid Wright-Fisher model to simulate forward-time evolution of a population of
227 cells that is subject to random genetic drift (Gillespie 2010). In the simulation, the cell
228 population size remains constant throughout evolution. To simulate plasmid evolution we
229 added an additional layer of random genetic drift by applying a Wright-Fisher model to the
230 plasmid inheritance (similarly to mitochondria simulated evolution (Peng and Kimmel 2005)).
231 Thus, during cell division, plasmid alleles are selected randomly into the daughter cells of the

232 next generation. In terms of population genetics, the plasmid population in the simulated
233 evolution has a substructure within the population. This implies that plasmid allele dynamics
234 have two hierarchical levels: within the cell and within the total population.

235 We compared the allele frequency dynamics of neutral variants emerging in a
236 population of $N = 10^5$ cells on a haploid chromosome (termed $Cx10^5$) and two multicopy
237 plasmids: one plasmid has a copy number of 10 plasmids per cell (termed $p10x10^5$), while
238 the second plasmid has 100 copies per cell (termed $p100x10^5$). The plasmid copy number
239 remains constant throughout the simulation, hence, a population of 10^5 cells harbours a total
240 of 10^6 p10 replicons. The total number of p100 replicon in such a population is 10^7. In order
241 to examine the effect of the increased total number of plasmid replicons, we simulated in
242 addition the allele dynamics of a haploid chromosome with two additional population sizes. In
243 the first population, the number of cells equals the total p10 replicons, i.e., $N = 10^6$ (termed
244 $Cx10^6$). In the second population the number of cells matches the total $p100x10^5$ replicons,
245 i.e., $N = 10^7$ (termed $Cx10^7$). In addition, we eliminate the effect of replicon size on the
246 probability of mutation by defining the plasmids and chromosomes to have an equal size of
247 500 loci. The evolution of all replicons was simulated over 1,000 generations in 100 replicate
248 populations. The simulation results reveal that the number of accumulated mutations on the
249 plasmids is higher in comparison to $Cx10^5$ and lower in comparison to chromosomes with the
250 respective population size (i.e., $p10x10^5$ vs. $Cx10^6$ and $p100x10^5$ vs. $Cx10^7$; Fig. 2A). These
251 results demonstrate that the number of accumulating plasmid mutations is lower than the
252 expected number according to plasmid mutational supply (i.e., total plasmid replicons). This
253 discrepancy indicates that most plasmid mutations are lost rapidly after their emergence.
254 Further comparison of allele frequency distribution among replicons shows that $p10x10^5$ and
255 $p100x10^5$ median AF is smaller in comparison to $Cx10^5$, and in comparison to the
256 chromosome of the respective population size (Fig. 2B). This demonstrates that the increase
257 in allele frequency of plasmid mutations is generally slower in comparison to the mutations
258 emerging on a haploid chromosome.

259 To compare the dynamics of mutant alleles residing on the plasmid or chromosome
260 while excluding the effect of mutational supply, we simulated the inheritance of a mutant
261 allele already present in the population at a specified initial frequency (*f*). Thus, the allele
262 dynamics in the population are determined by the combination of the initial mutant allele
263 frequency, random genetic drift and the mutant allele selection coefficient (*s*). The simulation
264 of an initially low-frequency mutant allele (f = 0.0001) that is neutral to the fitness of the cell
265 (i.e., selection coefficient s = 0) shows that the majority of mutant alleles on plasmids are lost
266 quickly in comparison to to mutant alleles on chromosomes (Fig. 3A). For example, the loss
267 of the mutant allele on the plasmid in 20% of the replicates was observed at generation 22
268 for p10x10^5 and at generation 76 for p100x10^5, but much later in the chromosomal replicons:
269 generation 128 in Cx10^5, generation 139 in Cx10^6 and generation 1,136 in Cx10^7 (Fig. 3A).
270 Increasing the initial mutant allele frequency 10-fold (f = 0.001) results in a similar rate of
271 allele loss between the plasmids and Cx10^5 populations. Nonetheless, loss rates of plasmid
272 mutant alleles remain higher in comparison to chromosomal mutant alleles in simulations
273 with the equivalent total number of replicons (Fig. 3A; Supplementary Table S4). Overall, our
274 results indicate that the loss dynamics of plasmid alleles due to segregational drift depend on
275 the initial mutant allele frequency in the total population and the distribution of the mutant
276 allele among plasmid hosts. Notably, the initial allele frequencies used in the simulation may
277 be encountered upon the acquisition of a new plasmid variant in the population, e.g., by
278 conjugation or transformation. In the case of *de novo* plasmid mutations the initial allele
279 frequency of the mutant plasmid is expected to be extremely low. Hence, according to our
280 simulations, in the absence of a strong benefit for the cell, a mutation on a plasmid will be
281 quickly lost.

282 Performing the simulations where the allele has a positive selection advantage (s =
283 0.01) reveals a higher loss rate in p10x10^5 in comparison to p100x10^5 populations, but the
284 median fixation time is lower for p10x10^5 than for p100x10^5. Notably, the median fixation time
285 of plasmid alleles (when not lost) is smaller in comparison to chromosome populations of
286 comparable total replicons when the initial frequency is low (f = 0.0001) albeit the majority of

287 plasmid populations lost the allele under these conditions (Fig. 3A; Supplementary Table S4).
288 The median fixation time of plasmid alleles is similar to the chromosomal one when the initial
289 frequency is increased (*f* = 0.001) (Fig. 3A; Supplementary Table S4). With high selection
290 coefficient (*s* = 0.1), allele loss is smaller than for *s* = 0.01 with a similar median fixation time
291 among p10x10^5 and the chromosome populations. Notably, the median fixation time in
292 p100x10^5 is substantially higher in comparison to all other replicons under these conditions
293 (Figure 3A; Supplementary Table S4). The longer fixation time of p100x10^5 alleles under
294 these simulated conditions are well explained by the effect of segregational drift on the rate
295 of plasmid allele fixation within the host cell. The high selection coefficient acts in favour of
296 cells harbouring at least one plasmid mutant, such that those cells rise quickly to fixation
297 within the population. In comparison, the increase of the plasmid allele in the total population
298 is slower due to the effect of random genetic drift of plasmid alleles during cell division, which
299 hinders the increase of plasmid allele frequency in daughter cells (Fig. 3B). In other words,
300 when the plasmid copy number is high, plasmid alleles within the cell remain polymorphic
301 (i.e., heterogeneous) over a longer time scale, also when the mutant allele is beneficial.
302 Notably, a plasmid copy number of 100 is within the range of PCNs observed in natural
303 isolates (Projan et al. 1987; Zhong et al. 2011), hence, the extended fixation time observed
304 for p100x10^5 is likely to occur also under natural conditions. Since our simulations do not
305 include intra-host plasmid recombination, the observed differences in allele dynamics
306 between plasmids and chromosomes are solely due to inheritance and segregational drift.
307 Overall, the results of the simulation demonstrate that segregational drift of multicopy
308 plasmids has a considerable impact on plasmid allele dynamics in the population by
309 increasing the loss rate and fixation time of plasmid alleles.

310 ***Evolution of the host chromosome.***

311 The comparison between the evolved and ancestral populations revealed that most of the
312 evolved genetic variants are found on the chromosome (Table 1 and Supplementary Table
313 S3). The majority of the chromosomal variants (90%) in our experiment are observed in the
314 hypermutator populations where the substitution rate is, on average, 10-fold higher in

315 comparison to the wild-type populations (Table 1). Genetic variants are observed in 638
316 protein coding genes and ncRNA genes, of which 173 (27%) genes accumulated genetic
317 variants in more than one population. The distribution of common mutated genes reveals a
318 strong signal of parallel evolution among populations evolved under similar settings (Fig. 4A).
319 Of the parallel mutated genes, 60 genes (35%) are specific to populations belonging to one
320 of the eight experimental factor combinations, indicating that the combination of host
321 genotype, plasmid replicon type and temperature had an impact on the evolution of these
322 genes. Furthermore, a total of 50 mutated genes (29%) are shared among hypermutator
323 populations having the same plasmid type, regardless of the growth temperature (Fig. 4A).
324 Hypermutator-pLC populations share, on average, 19.62 ± 2.09 (*SD*) mutated genes,
325 whereas hypermutator-pHC populations show a lower level of parallelism with 14.45 ± 3.30
326 shared genes on average (SD). Furthermore, two small clusters of parallel mutated genes
327 are observed in the wild-type hosts, the first cluster comprises the pLC populations and the
328 second cluster includes the pHC populations evolved at 42°C.

329 A strong signal of parallel evolution is observed also at the resolution of specific
330 mutations where 144 (15%) variants were detected in more than one of the 48 evolved
331 population. The allele frequency of parallel variants is significantly higher than the allele
332 frequency of population specific variants ($P < 0.001$, using two-sample Kolmogorov-Smirnov
333 test, $n_{parallel} = 720$, $n_{specific} = 841$). To quantify the contribution of the three experimental
334 evolution factors to parallel evolution of the host chromosome, we compared the allele
335 frequency of parallel genetic variants among all populations. We observed a significant effect
336 of plasmid type ($P = 1.1\text{x}10^{-69}$) and temperature ($P = 3.9\text{x}10^{-46}$) on the allele frequency
337 variation (using ANOVA on aligned rank transformed allele frequencies (Wobbrock et al.
338 2011); detailed statistics in Supplementary Table S5). In the wild-type populations, there is a
339 strong interaction between plasmid type and temperature, whereas in the hypermutator
340 populations, the same interaction is observed along with a main effect of the plasmid type
341 (Fig. 4B). Our results thus reveal that plasmid replicon type is a strong determinant of host
342 evolution and it has an equal or larger effect in comparison to growth temperature.

343 To further identify variants that are characteristic for specific factor combinations, we
344 tested for the effect of the factors and their interaction on the allele frequency of each parallel
345 variant (using ANOVA on aligned rank transformed allele frequencies and FDR). This results
346 in 37 variants whose allele frequency is significantly different among the factor combinations
347 (Fig. 4C; Supplementary Table S6). A single variant occurred in all wt-pLC populations
348 evolved at 37°C, where a nonsense mutation of Glutamine into a stop codon occurred in the
349 *rpoS* gene. Hence, that gene is likely non-functionalized in these populations. The *rpoS*
350 encodes the RNA polymerase σ^S-subunit that is the master regulator of the general stress
351 response in *E. coli* (e.g., Patten et al. 2004). Loss of RpoS function was previously reported
352 to occur rapidly in nutrient-limited chemostat populations (Notley-McRobb et al. 2002). A
353 cluster including two non-synonymous substitutions is observed for wt-pHC populations
354 evolved at 42°C. The substitutions occurred in a gene encoding for a tRNA modification
355 enzyme (*miaB*) and a DNA polymerase I gene (*polA*). Overall, we identified 11 different *polA*
356 mutations in 16 pHC-populations and none in pLC-populations (Supplementary Table S3).

357 Mutations that are characteristic to hypermutator-pLC populations (Fig. 4C) are found
358 in genes that belong to four main functional categories, including carbohydrate metabolism,
359 fatty acid metabolism, biofilm formation and stress response. Example genes where amino
360 acid replacements have occurred are propionyl-CoA synthetase (*prpE*), a part of the
361 propionate metabolism pathway (Textor et al. 1997), and a fimbrial-like adhesin protein
362 (*sfmH*) that is known to promote adhesion in *E. coli* (Korea et al. 2010). Two additional non-
363 synonymous substitutions in genes encoding for an integrase (*intA*) and a glycosyl
364 transferase (*wcaA*) are common to all six hypermutator-pLC populations evolved at 42°C (Fig.
365 4C). The evolution of hypermutator-pHC populations is characterized by a majority of non-
366 synonymous point mutations. These variants are fixed (i.e., AF > 0.9) in all hypermutator-
367 pHC populations. The mutated genes are classified into three main categories including
368 carbohydrate metabolism, stress response and DNA replication. Example mutated genes are
369 DNA polymerase II (*polB*) and DNA gap repair protein (*recF*). Taken together, the differential

370 acquisition of mutations in hypermutator populations carrying plasmid pLC or pHC suggests
371 that the plasmid replicon type had a significant impact on the host chromosome evolution.

372 **Discussion**

373 Previous experimental evolution studies of plasmid-host adaptation noted that evolved
374 variants are mostly observed in the host chromosome rather than the plasmid genome
375 (Bouma and Lenski 1988; San Millan et al. 2014; Loftie-Eaton et al. 2017), and this is indeed
376 the case also for our study. Naturally, the mutational supply of a chromosome of the size of *E.*
377 *coli* (4.6 Mb) is higher than the mutational supply of our model plasmid genomes.
378 Consequently, more mutations are likely to arise on the chromosome than on the plasmid.
379 Notably, in most studies of plasmid-host coevolution (including ours), the selection pressure
380 is imposed on the maintenance of the plasmid in the host population rather than a specific
381 plasmid allele (e.g., as in San Millan et al. 2016). Because plasmids depend on the host
382 machinery for their replication and persistence, the selection for plasmid maintenance is, in
383 practice, imposed on the host machinery rather than on the plasmid itself. Moreover, we
384 have to keep in mind that the potential adaptive responses to the selective pressure for the
385 plasmid maintenance are numerous in the complex host chromosomes (e.g., San Millan et al.
386 2014; Harrison et al. 2015), but are very limited in the genetically compact plasmid genome.
387 This, in combination with the higher mutational supply of the chromosome, implies that most
388 adaptive variants are expected to occur on the chromosome. This can explain the high
389 frequency of genetic variants observed on the chromosome in our experiment, as well as
390 others (e.g., San Millan et al. 2014; Harrison et al. 2015; Loftie-Eaton et al. 2017).

391 The significant effect of plasmid type on the chromosomal allele frequency observed
392 in our experiment (Figs. 4A,C) suggests that evolution of the host chromosome depends on
393 the replicon type of our model plasmids, with the origin of replication being a prominent
394 characteristic thereof. The evolutionary scenario portrayed here is highly relevant for
395 plasmid-mediated evolution of antibiotics resistant strains (e.g., Sandegren et al. 2011;

396 Jechalke et al. 2014; Stoesser et al. 2016) as in such incidents, the antibiotics impose a
397 selection pressure for the plasmid maintenance in the population. Our findings suggest that
398 the evolution of plasmid-mediated antibiotics resistance is expected mostly to involve the
399 host adaptation to the plasmid replicon type rather than evolution of the plasmid encoded
400 trait.

401 The results from our experimental evolution of model plasmids demonstrate that the
402 low-copy replicon has accumulated more point mutations in comparison to the high-copy
403 replicon (Table 1). This observation is in contrary to our expectation according to the
404 mutational supply on the plasmid genome. This deviation from the expectation can be
405 explained by the inverse correlation between population size and the impact of genetic drift
406 on allele dynamics in the population. The low-copy plasmid constitutes a small replicon
407 population within the cell. The dynamics of allele frequency due to drift in small populations
408 are expected to have large fluctuations such that novel alleles are quickly lost or fixed
409 (Bodmer and Cavalli-Sforza 1976). In contrast, the high copy plasmid constitutes a large
410 replicon population within the cell where fluctuations in allele frequency over time are
411 expected to be small (Bodmer and Cavalli-Sforza 1976). The impact of segregational drift on
412 plasmid alleles is best understood in comparison to the chromosomal alleles. While novel
413 chromosomal alleles emerging on a haploid chromosome are inherited to all daughter cells in
414 the population, the inheritance of plasmid alleles depends on the segregation pattern of
415 plasmid copies. Indeed, our simulations demonstrate that the frequency of neutral mutations
416 emerging on a multicopy plasmid is higher in comparison to a haploid chromosome (Fig. 2),
417 which is in agreement with the higher mutational supply of multicopy plasmids (San Millan et
418 al. 2016). Nevertheless, the median allele frequency of mutations arising on multicopy
419 plasmids is lower than that of chromosomal mutations in a comparable population size (i.e.,
420 number of replicons). Hence, the allele frequency of mutations residing on multicopy
421 plasmids is increasing slower in comparison to chromosomal alleles. Following the
422 emergence of a novel plasmid allele, plasmid segregation during cell division has three
423 potential outcomes for the plasmid genotype of the daughter cells: 1) loss of the novel

424 plasmid allele, 2) the presence of polymorphic plasmid alleles in the cell, or 3) fixation of the
425 novel plasmid allele in the cell (see illustration in Fig. 5). The expected effect of plasmid copy
426 number on the allele dynamics is best observed in the simulated evolution of beneficial
427 alleles (i.e., $s > 0$): on the low-copy plasmid (p10), the loss of plasmid alleles occurs more
428 frequently in comparison to the chromosome, but the median fixation time is nearly the same
429 as that of the chromosome. Indeed, previous studies showed that under strong selective
430 conditions, beneficial mutations on multicopy plasmids can rise quickly to fixation (San Millan
431 et al. 2016). Notably, our simulation of beneficial alleles reveals that alleles on the high copy
432 plasmid (p100) remain polymorphic over a longer time scale and have a higher median
433 fixation time (Fig. 3). For neutral alleles, segregational drift of multicopy plasmids hinders the
434 increase of plasmid allele frequency; it leads to frequent and rapid mutant allele loss on the
435 one hand, and to longer fixation time on the other hand. Altogether, our results demonstrate
436 that the effect of segregational drift on plasmid allele dynamics counteracts the effect of
437 higher mutational supply accompanying increased copy number. Consequently, the
438 evolvability of low-copy plasmids is higher than that of high copy plasmids despite the latter
439 higher mutational supply.

440 The impact of segregational drift is not expected to be limited to non-mobile plasmids
441 as mobile (or conjugative) plasmids are also found in multicopy state (e.g., Figurski and
442 Helinski 1979). Plasmid copy number may be heterogeneous within the population (e.g.,
443 Münch et al. 2015), especially in the absence of an active partition mechanism, and
444 furthermore plasmid copy number can be variable during evolution (e.g., San Millan et al.
445 2016; Santos-Lopez et al. 2017). The evolution of multicopy plasmids entails time spans of
446 intra-cellular heterogeneity, which a situation also encountered by mitochondrial genomes
447 within single cells. Our study thus reveals parallel evolutionary concepts between plasmids
448 and eukaryotic organelles. In summary, a realistic portrayal and modelling of plasmid
449 genome evolution requires deeper understanding of segregational drift.

450

451 ## Methods

452 ### *Bacterial strains, growth conditions and DNA techniques*

453 *Escherichia coli* K-12 substr. MG1655 (DSM No. 18039, DSMZ) and a mismatch-repair
454 deficient *E. coli* MG1655 derivative strain (Δ*mutS*) were used in the evolution experiment.
455 *Escherichia coli* DH5α (Hanahan 1983) was used during construction of plasmids. All strains
456 were routinely grown at 37°C in LB medium supplemented with 100 mg L^{-1} ampicillin when
457 required. The evolution experiment was performed in M9 medium (M9/amp_{100}) prepared as
458 described (Sambrook and Russell 2001), supplemented with 0.4% (w/v) glucose, 0.1% (w/v)
459 casamino acids, 20 mg L^{-1} uracil, 0.5 mg L^{-1} thiamine, 0.005% (v/v) Antifoam 204 (Sigma-
460 Aldrich), and 100 mg L^{-1} ampicillin. Phusion polymerase (Thermo Fisher Scientific) was used
461 for amplification of DNA-fragments for cloning and plasmid constructs were verified by
462 restriction analysis and confirmed by sequencing. Linear DNA fragments and plasmids were
463 introduced into *E. coli* strains by electroporation using a Bio-Rad Gene Pulser device with
464 parameters reported previously (Dower et al. 1988). PCR-primers are listed in
465 Supplementary Table S7.

466 ### *Construction of the host strains (wild-type and Δ*mutS*)*

467 A marker-free Δ*mutS* deletion strain was constructed using the λ Red/ET Quick & Easy *E.*
468 *coli* Gene Deletion Kit (GeneBridges) according to the manufacturers protocol. Briefly, the
469 *mutS* gene was replaced with a PCR generated DNA fragment (primer pair mutS_KO_F/R)
470 containing a kanamycin resistance marker gene flanked by FRT sites. One kanamycin
471 resistant clone, in which the replacement of *mutS* with the marker-containing cassette was
472 verified by PCR (primer pair PGK_F and mutS_test_R), was chosen for the removal of the
473 kanamycin resistance marker through expression of the site specific FLP recombinase
474 carried on plasmid 707-FLPe (GeneBridges). One resulting kanamycin-sensitive clone that
475 had lost 707-FLPe was chosen and the deletion of the marker-containing cassette was
476 verified by PCR using primer pair mutS_test_F/R.

477 In order to potentially compare mutations occurring within non-coding segments of DNA, both
478 plasmids and host strains were equipped with segments of randomly composed, non-coding
479 DNA. Such a 600 bp-stretch of non-coding DNA (termed art3) was commercially synthesized
480 (GeneAart Strings Service, Thermo Fisher Scientific) and inserted into the Tn*7*-specific
481 *attTn7* site of the wild-type and Δ*mutS* strain following the method described by(McKenzie
482 and Craig 2006). Briefly, the art3 fragment was amplified by using primer pair art3_F/R and
483 cloned into the *Not*I restriction site present in the mini-Tn*7*-transposon carried on the
484 transposition vector pGRG25 (GenBank acc. no. DQ460223). The resulting plasmid was
485 delivered into *E. coli* wild-type and the Δ*mutS* strain and a clone carrying the chromosomal
486 *attTn7::miniTn7(art3)* insertion was verified by PCR targeting the Tn*7*-insertion site (primer
487 pair attTn7_F/R).

488 ***Construction of plasmids pLC and pHC***

489 Plasmid pBBR1MCS-5 (GenBank accession no. U25061) was used to construct plasmid pLC.
490 A 1,944-bp fragment comprising the gentamicin resistance gene and the *lac*Zα gene
491 fragment was excised with *BstB*I and *Bsa*I. The resulting plasmid backbone fragment (2824
492 bp) was blunted and ligated to a blunted *Kpn*I-*Sac*I fragment containing *bla* with its promoter
493 obtained from pBluescript SK(+) into which the *bla* fragment had been inserted into the
494 *Eco*RV site as a blunted *Cla*I-*Not*I fragment from plasmid pKD4 (GenBank acc. no.
495 AY048743). Next, 600 bp long stretches of random DNA (designated as art1 and art2) were
496 inserted into the *Eco*RI site (art1) and into the *Xho*I site (art2) of pBBR1-bla, giving plasmid
497 pLC (GenBank acc. no. MH238456). The high copy plasmid pHC (GenBank acc. no.
498 MH238457) was constructed as plasmid pLC, but the *oriV* and the *repA* gene were then
499 replaced by a pUC origin of replication. This was achieved by amplifying the plasmid except
500 for the region comprising the *oriV* and the *repA* gene (using primer pair inv_pLC_F/R). The
501 resulting fragment (2,735 bp) was treated with T4-Polynucleotide Kinase (New England
502 Biolabs) and ligated to a PCR fragment (916 bp) comprising the pUC-origin of replication
503 amplified from plasmid pCR4-TOPO (Thermo Fisher Scientific) with primer pair pUC-ori_F/R.

504 ***Experimental design and setup***

505 The evolution experiment was set up using a full factorial design of 2 plasmid types × 2 host
506 genotypes × 2 growth temperatures. Six biological replicates were used for each of the eight
507 experimental factor groups, resulting in a total of 48 populations that were evolved for 800
508 generations. The evolution experiment was conducted in chemostats made from off-the-shelf
509 materials similar in design as previously described for the cultivation of yeast (Miller et al.
510 2013). The chemostat system was modified to enable bacterial culturing over longer time
511 periods. Briefly, air breaks were introduced upstream of culturing vessels to prevent bacterial
512 growth into medium supply lines and culture vessels were equipped with additional effluent
513 ports to ensure constant efflux in case of clogging caused by biofilm-formation (additional
514 information provided in Supplementary Fig. S3). Prior to the evolution experiment, plasmid
515 pLC and pHC, respectively, were transferred into the wild-type and the Δ*mutS* strain and a
516 single colony of each type (wt-pLC, wt-pHC, Δ*mutS*-pLC, and Δ*mutS*-pHC) was used to
517 found a population that was serially passaged (1:1000) daily for about 100 generations in
518 M9/amp$_{100}$ at 37°C to acclimate the populations to the media conditions. At the onset of the
519 evolution experiment, 50 µl overnight culture of each of the four populations (termed here
520 ancestral populations) was used to inoculate 6 replicate chemostat cultures with a working
521 volume of 20 ml at 37°C and 42°C each. The 48 chemostat cultures were grown to stationary
522 densities overnight (i.e., without addition of media) and then continuously incubated at a
523 dilution rate of 0.27 h^{-1}. The bacterial concentration of chemostat cultures grown at steady-
524 state was stable over time and among replicates (M = 1.92 × 10^8 cells per ml, SD = 1.11 ×
525 10^7 cells per ml, n = 432). Nevertheless, in order to ensure the stable maintenance of flow
526 rates during the course of the experiment, chemostat units and tubings were replaced every
527 three weeks (approximately 200 generations). Meanwhile, chemostat populations were
528 stored at -80°C in 25% (v/v) glycerol. From these frozen stocks, the chemostat cultures were
529 restarted as follows: 10 ml of the frozen stock culture were pelleted by low-speed
530 centrifugation (5000×g) for removal of glycerol and the cells were harvested in the final
531 culturing volume of 20 ml and transferred into chemostat vessels. Each population was

532 evolved for 90 days at steady-state resulting in approximately 800 generations. On a regular
533 basis, all chemostat populations were reciprocally tested for cross-contamination using
534 qPCR. To this end, each culture was tested for the presence of the *mutS* gene and for the
535 presence of both plasmids using the following primer pairs: q_pBB_rep_F/R complementary
536 to the *rep* gene on plasmid pLC, q_pUC_ori_F/R complementary to the *ori* on plasmid pHC,
537 and primer q_mutS_F/R complementary to *mutS* gene in the *E. coli* MG1655 background.
538 The proportion of plasmid bearing cells in evolved and ancestral populations was assessed
539 by appropriately diluting samples of chemostat cultures with 1x phosphate-buffered saline,
540 spreading on M9 agar plates and incubating at 37°C or 42°C for 14 hours. At least 100 single
541 colonies of each chemostat sample were transferred to M9/amp$_{100}$ and to M9 agar plates.
542 Colonies that grew only on non-selective media were counted as plasmid-free segregants.
543 To verify the absence of the plasmids, 1-2 randomly chosen colonies per population were
544 chosen for testing the presence of the plasmid using qPCR (q_bla_F/R for the plasmids and
545 q_idnT_F/R for the chromosome). The proportion of plasmid-bearing cells was determined
546 from the comparison of the number of colonies grown on ampicillin plates to the total number
547 of transferred colonies grown on non-selective medium.

548 ***Plasmid copy number and loss frequency***

549 Plasmid copy numbers were determined using quantitative PCR (qPCR) as described
550 previously (Skulj et al. 2008), but with primers complementary to *idnT* on the *Escherichia coli*
551 MG1655 chromosome (q_idnT_F/R) and to the *bla* gene (q_bla_F/R) on plasmid pLC and
552 pHC. Samples of overnight cultures were heated at 98°C for 10 min and then frozen at -20°C
553 to disrupt the cells. The qPCR reactions were carried out in a total volume of 10 µl containing
554 1x iTaq Universal SYBR Green Supermix (Bio-Rad Laboratories), 100 nM of each primer
555 (final concentration) and 1 µl sample. All qPCR reactions, including positive and non-
556 template controls, were performed in technical replicates on a CFX Connect Real-Time PCR
557 Detection System (Bio-Rad Laboratories) using the following cycling conditions: 95°C for 3
558 min, and 40 cycles of 10 s at 95°C and 1 min at 59°C. Primer specificities and efficiencies
559 were determined using standard-curve and melt-curve analyses. The ratio between the

560 number of plasmid specific amplicons and chromosome specific amplicons is defined as the
561 plasmid copy number (Skulj et al. 2008) and was here calculated considering the
562 amplification efficiencies of both primer pairs.

563 To determine the segregational loss of plasmid pLC and pHC, five single colonies of
564 each strain (i.e., wt-pLC, wt-pHC) were picked from M9/amp$_{100}$ agar plates and grown
565 overnight in 2 ml M9/amp$_{100}$ with constant shaking at 37°C for approximately 4.4 ± 0.42 (*SD*)
566 generations. Appropriate dilutions of stationary cell cultures were plated on M9 agar,
567 incubated overnight at 37°C, and 100 colonies per replicate were streaked onto M9/amp$_{200}$,
568 M9/amp$_{100}$, and M9 plates. The number of plasmid-free segregants was determined from the
569 comparison of colonies grown on selective and non-selective plates.

570 ***Whole genome sequencing***

571 The ancestral populations of the wild-type and the Δ*mutS* strain carrying the designated
572 plasmids (wt-pHC, wt-pLC, Δ*mutS*-pHC, and Δ*mutS*-pLC) and the 48 populations evolved for
573 800 generations were sequenced using Illumina sequencing technology. Population
574 sequencing was used to enable the detection of rare variants and to asses the frequency of
575 alleles within and across populations. Total DNA (genomic DNA plus plasmid DNA) was
576 isolated from archived samples stored at -80°C using the Wizard Genomic DNA Purification
577 kit (Promega). The preparation of DNA for sequencing of the evolved populations differed
578 between the two plasmids due to the difference in their stability within the host. For pLC-
579 populations, 1 ml of cell suspension from thawed samples was used for DNA extraction. The
580 frequent loss of pHC led to the presence of plasmid-free segregants in the evolved
581 populations. Therefore, in addition to the sequencing of DNA extracted from total pHC-
582 populations, DNA was sequenced from pooled colonies of plasmid-carrying cells. The deep
583 sequencing of plasmid-carrying hosts allows the detection of rare variants from a population
584 of cells with a potentially heterogeneous plasmid allele pool. DNA from total pHC populations
585 was isolated from archived samples as described for pLC-populations. The preparation of
586 DNA from plasmid-carrying cells included a preliminary stage of selection for plasmid-hosts.

587 Thus, frozen chemostat culture samples were first plated on LB plates supplemented with
588 ampicillin (100 µg ml^{-1}) and grown overnight at 37°C or 42°C, to select plasmid-carrying cells.
589 Then, 100 colonies per pHC-population were pooled for extraction of total DNA (i.e., 4x600
590 colonies in total). Sample libraries for Illumina sequencing were prepared using either the
591 Nextera or Nextera XT library preparation kit. All samples were quantified on a Qubit
592 fluorometer (Invitrogen by Life Technologies) and DNA fragment length distribution was
593 assessed on a TapeStation (Agilent Technologies). Libraries were sequenced on either a
594 HiSeq 2500 platform with 2x125 bp reads or a NextSeq 500 platform with 2x150 bp reads
595 (see Supplementary Table S1 for details). Due to low initial coverage, some libraries of the
596 wt-pLC and wt-pHC populations were re-sequenced on a HiSeq 2500 platform.

597 ***Variant detection***

598 Sequencing reads were trimmed to remove both Illumina specific adaptors and low quality
599 bases using the program Trimmomatic v.0.35 (Bolger et al. 2014) with these parameters:
600 `ILLUMINACLIP:NexteraPE-PE.fa:2:30:10 CROP:150/125` (NextSeq/HiSeq) `HEADCROP:5`
601 `LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:36`. Quality of sequencing reads
602 was inspected before and after trimming using FastQC v.0.11.5 (Andrews 2010). To confirm
603 plasmid maps derived from partial Sanger-sequencing after plasmid construction, plasmid
604 genomes of the four ancestral populations were assembled from trimmed paired end
605 sequencing reads using plasmidSPAdes (SPAdes v.3.9.0 (Nurk et al. 2013)). To improve
606 assembly speed and accuracy of the high copy plasmid pHC samples, BBNorm (BBMap tool
607 suite v.35.82 (https://sourceforge.net/projects/bbmap/)) was used to normalize paired end
608 read data to an average coverage of 200x. The assemblies resulted in contigs with
609 overhangs at each end. There was no variation between plasmid assemblies of the same
610 type. Therefore, only one assembly of plasmid pLC and pHC was used as a reference for
611 variant detection. Plasmid genome assemblies were annotated using SnapGene software
612 v.2.4 (GLS Biotech). The chromosomal insertion sequence art3 (described above) was
613 included in the reference genome sequence of *E. coli* MG1655 (GenBank acc. no.
614 NC_000913.3). Reads were then mapped to this reference genome and to the assembled

615 plasmid genomes using BWA-MEM v.0.7.5a-r405 (Li and Durbin 2009). Corresponding
616 mapping files of libraries that were sequenced twice were merged using PICARD tools
617 v.2.7.1 (http://broadinstitute.github.io/picard/). Mapping statistics were retrieved using
618 BAMStats v.1.25 (https://sourceforge.net/projects/bamstats/files/).Then, indexing, local
619 realignment of sequencing reads, removal of read duplicates and ambiguous aligned reads
620 were performed using PICARD tools, SAMtools v.0.1.19 (Li et al. 2009) and GATK v.3.6
621 (McKenna et al. 2010) retaining only paired mapped reads with a minimum mapping quality
622 of 20. To enable a between-sample and between-replicon comparison of measures deduced
623 from the sequencing reads (e.g., number of mutations) that is not biased by differences in
624 sequencing coverage, subsampling of reads was performed. Therefore, mapping files were
625 first split by contig (i.e., plasmid and chromosome) using Bamtools v.2.4.2 (Barnett et al.
626 2011). Reads of each chromosomal and plasmid BAM file were then subsampled to the
627 minimal mean coverage of 104x using SAMtools (seed 1). Subsampled and non-subsampled
628 datasets were then used to call variants. Short indels and SNPs were called using LoFreq
629 v.2.1.2 (Wilm et al. 2012) and VarScan v.2.4.3 (Koboldt et al. 2009). Pindel version 0.2.5b9
630 (Ye et al. 2009) was used with options `-v 30 -M 2` to detect indels. The maximum number
631 of reads used to call variants with LoFreq and to call variants with VarScan by using
632 SAMtools mpileup was set to 15,000,000 for non-subsampled datasets. Variants that were
633 supported by alternative reads in one direction only were removed. Furthermore, only
634 variants that were concordantly identified by two or more of the variant calling tools were
635 retained using BEDTools' intersect utility v2.17.0 (Quinlan and Hall 2010) with parameters `-f`
636 `1.0 -r -u`. BEDTools' intersect utility was then used with parameters `-f 1.0 -r -v` to
637 keep only variants present exclusively in evolved but not in the ancestral samples. A custom
638 python script using pysam (https://github.com/pysam-developers/pysam) and R version 3.4.3
639 (R Development Core Team 2017) were then used to retain only variants that met the
640 following criteria: allele frequency ≥ 0.02, depth of coverage ≥ 10, ratio of the numbers of
641 forward and reverse reads supporting the variant ≥ 0.25, a minimum medium mapping quality
642 of 40, a minimum medium base quality of 30 and a minimum median length of 30 nt to the

643 end of the reads supporting the variant. Indels within an homopolymer stretch of ≥ 8
644 nucleotides and indels with less than 10 reads supporting the variant were excluded.
645 Variants whose read support at the corresponding position in ancestor samples was not
646 sufficient to detect a variant (i.e., < 10x) were excluded. Variant genomic location and the
647 type of mutation, i.e., intergenic, synonymous, non-synonymous, or nonsense, were
648 annotated using an in-house Perl script. Plasmid copy numbers were inferred from the de-
649 duplicated, non-subsampled BAM files as the ratio of plasmid to chromosomal mean
650 coverage. The number of intergenic and synonymous sites from the annotated reference
651 genome sequences that were used for the mapping of sequencing reads. Intergenic
652 positions were extracted from the reference sequences and the total number of intergenic
653 sites was calculated as the sum of intergenic positions whose coverage was ≥ 10x. Potential
654 synonymous sites were determined for each position within the reference sequences. The
655 total number of synonymous sites were then calculated for each mapping file using the Nei-
656 Gojobori method (Nei and Gojobori 1986) considering only those positions with a minimum
657 sequencing coverage of 10x.

658 ***Simulation***

659 Forward-time simulations based non non-overlapping generations with a constant cell
660 population size were performed with python3.4 and the module simuPOP v1.1.8.3 (Peng and
661 Kimmel 2005). Two types of cells were simulated independently. Cells with chromosomes
662 have `L` haploid loci (`ploidy=1`). Cells with plasmids have a predefined plasmid copy number
663 `cp` that stays constant throughout the simulation. Each plasmid has `L` loci, resulting in a total
664 number of `L×cp` loci in each cell. This was implemented based on `cp` customized
665 chromosomes (`chromTypes=[sim.CUSTOMIZED]*cp`), each with `L` haploid loci. Alleles were
666 simulated as binary (`alleleType='binary'`). Inheritance mode for the chromosome
667 corresponds to a Wright-Fisher model (operator `RandomSelection()`). Inheritance mode for
668 the plasmid corresponds to a Wright Fisher model with mitochondrial inheritance of
669 organelles (operator `RandomSelection(ops=[sim.MitochondrialGenoTransmitter()])`).

For the simulations with mutation, initial populations (t=0) were homogenous, i.e., all loci had state 0. The mutation rate in the simulation was constant with 2×10^{-9} per locus per generation (operator `SNPmutator(u=2e-9)`). For the simulated evolutionary dynamics of mutant alleles, we simulated the evolution of one binary locus (`L=1`) having a mutant allele already present in the initial population at a specified initial frequency (*f*) and no new mutation is applied during simulated evolution. Initial populations were randomly generated with a particular mutant frequency `f` (parameter `freq=[1-f,f]` for the method `initGenotype`). Note that this assigns the mutant allele randomly to each locus, independent of the state of the other loci in the same cell for plasmid simulations. For the simulated evolution with selection, the probability to be inherited to the next generation is proportional to the fitness value of the cell. For cells with a chromosome, a fitness value of 1+s was assigned to allele state 1 and fitness value 1 was assigned to allele state 0. For cells with plasmids, a fitness value of 1+s was assigned if at least one plasmid has allele state 1, otherwise the fitness is 1 (i.e., dominant selection mode). Selection was applied using a fitness field (`infoFields="fitness"`) and the operator `PySelector`. Simulations were run for a particular number of generations using the `evolve` method.

Data availability

Sequence reads of evolved and ancestral populations are found in SRA (SRA acc. SRP######). Plasmid sequences are found in GenBank (pLC: MH######, pHC: MH######).

References

Abeles AL, Friedman SA, Austin SJ. 1985. Partition of unit-copy miniplasmids to daughter cells. III. The DNA sequence and functional organization of the P1 partition region. J. Mol. Biol. 185:261–272.

Andrews S. 2010. FASTQC. A quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Antoine R, Locht C. 1992. Isolation and molecular characterization of a novel broad-host-range plasmid from *Bordetella bronchiseptica* with sequence similarities to plasmids from gram-positive organisms. Mol. Microbiol. 6:1785–1799.

698 Arjan JA, Visser M, Zeyl CW, Gerrish PJ, Blanchard JL, Lenski RE. 1999. Diminishing
699 returns from mutation supply rate in asexual populations. Science 283:404–406.

700 Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. 2011. BamTools: a C++
701 API and toolkit for analyzing and managing BAM files. Bioinformatics 27:1691–1692.

702 Baxter JC, Funnell BE. 2014. Plasmid partition mechanisms. Microbiol. Spectr. 2,
703 10.1128/microbiolspec.PLAS-0023-2014

704 Bazaeal M, Helinski DR. 1968. Circular DNA forms of colicinogenic factors E1, E2 and E3
705 from *Escherichia coli*. J. Mol. Biol. 36:185–194.

706 Bedhomme S, Perez Pantoja D, Bravo IG. 2017. Plasmid and clonal interference during post
707 horizontal gene transfer evolution. Mol. Ecol. 26:1832–1847.

708 Bevan M. 1984. Binary *Agrobacterium* vectors for plant transformation. Nuc. Acids Res.
709 12:8711–8721.

710 Birky CW. 2001. The inheritance of genes in mitochondria and chloroplasts: laws,
711 mechanisms, and models. Annu. Rev. Genet. 35:125–148.

712 Bodmer WF, Cavalli-Sforza LL. 1976. Genetics, Evolution, and Man. San Francisco, CA: W
713 H Freeman & Company.

714 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
715 data. Bioinformatics 30:2114–2120.

716 Bouma JE, Lenski RE. 1988. Evolution of a bacteria/plasmid association. Nature 335:351–
717 352.

718 Czárán TL, Hoekstra RF, Pagie L. 2002. Chemical warfare between microbes promotes
719 biodiversity. Proc. Natl. Acad. Sci. U.S.A. 99:786–790.

720 Deatherage DE, Kepner JL, Bennett AF, Lenski RE, Barrick JE. 2017. Specificity of genome
721 evolution in experimental populations of *Escherichia coli* evolved at different
722 temperatures. Proc. Natl. Acad. Sci. U.S.A. 114:E1904–E1912.

723 Dower WJ, Miller JF, Ragsdale CW. 1988. High efficiency transformation of *E. coli* by high
724 voltage electroporation. Nuc. Acids Res. 16:6127–6145.

725 Dziewit L, Pyzik A, Szuplewska M, Matlakowska R, Mielnicki S, Wibberg D, Schlüter A,
726 Pühler A, Bartosik D. 2015. Diversity and role of plasmids in adaptation of bacteria
727 inhabiting the Lubin copper mine in Poland, an environment rich in heavy metals. Front.
728 Microbiol. 6:152.

729 Figurski DH, Helinski DR. 1979. Replication of an origin-containing derivative of plasmid RK2
730 dependent on a plasmid function provided in *trans*. Proc. Natl. Acad. Sci. U.S.A.
731 76:1648–1652.

732 Gillespie JH. 2004. Population genetics: a concise guide. Baltimore, Maryland: John Hopkins
733 University Press.

734 Gullberg E, Albrecht LM, Karlsson C, Sandegren L, Andersson DI. 2014. Selection of a
735 multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. mBio
736 5:e01918–14.

737 Guyot S, Pottier L, Hartmann A, Ragon M, Hauck Tiburski J, Molin P, Ferret E, Gervais P.

738 2014. Extremely rapid acclimation of *Escherichia coli* to high temperature over a few
739 generations of a fed-batch culture during slow warming. Microbiologyopen 3:52–63.

740 Hanahan D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol.
741 166:557–580.

742 Harrison E, Guymer D, Spiers AJ, Paterson S, Brockhurst MA. 2015. Parallel compensatory
743 evolution stabilizes plasmids across the parasitism-mutualism continuum. Curr. Biol.
744 25:2034–2039.

745 He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz LA,
746 Kinzler KW, Vogelstein B, Papadopoulos N. 2010. Heteroplasmic mitochondrial DNA
747 mutations in normal and tumour cells. Nature 464:610–614.

748 Hershfield V, Boyer HW, Yanofsky C, Lovett MA, Helinski DR. 1974. Plasmid ColE1 as a
749 molecular vehicle for cloning and amplification of DNA. Proc. Natl. Acad. Sci. U.S.A.
750 71:3455–3459.

751 Jahn M, Vorpahl C, Hübschmann T, Harms H, Müller S. 2016. Copy number variability of
752 expression plasmids determined by cell sorting and Droplet Digital PCR. Microb. Cell
753 Fact. 15:211.

754 Jechalke S, Heuer H, Siemens J, Amelung W, Smalla K. 2014. Fate and effects of veterinary
755 antibiotics in soil. Trends Microbiol. 22:536–545.

756 Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson
757 RK, Ding L. 2009. VarScan: variant detection in massively parallel sequencing of
758 individual and pooled samples. Bioinformatics 25:2283–2285.

759 Korea C-G, Badouraly R, Prevost M-C, Ghigo JM, Beloin C. 2010. *Escherichia coli* K-12
760 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface
761 specificities. Env. Microbiol. 12:1957–1977.

762 Lanfear R, Kokko H, Eyre-Walker A. 2014. Population size and the rate of evolution. Trends
763 Ecol. Evol. 29:33–41.

764 Lederberg J, Tatum EL. 1946. Gene recombination in *Escherichia coli*. Nature 158:558.

765 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
766 transform. Bioinformatics 25:1754–1760.

767 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
768 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map
769 format and SAMtools. Bioinformatics 25:2078–2079.

770 Loftie-Eaton W, Bashford K, Quinn H, Dong K, Millstein J, Hunter S, Thomason MK, Merrikh
771 H, Ponciano JM, Top EM. 2017. Compensatory mutations improve general
772 permissiveness to antibiotic resistance plasmids. Nat. Ecol. Evol. 1:1354–1363.

773 Majeed H, Gillor O, Kerr B, Riley MA. 2011. Competitive interactions in *Escherichia coli*
774 populations: the role of bacteriocins. ISME J. 5:71–81.

775 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
776 Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce
777 framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–
778 1303.

779 McKenzie GJ, Craig NL. 2006. Fast, easy and efficient: site-specific insertion of transgenes
780 into enterobacterial chromosomes using Tn7 without need for selection of the insertion
781 event. BMC Microbiol. 6:39.

782 Miller AW, Befort C, Kerr EO, Dunham MJ. 2013. Design and use of multiplexed chemostat
783 arrays. J. Vis. Exp. 72:e50262.

784 Münch KM, Müller J, Wienecke S, Bergmann S, Heyber S, Biedendieck R, Münch R, Jahn D.
785 2015. Polar fixation of plasmids during recombinant protein production in *Bacillus*
786 *megaterium* results in population heterogeneity. Applied Environ. Microbiol. 81:5976–
787 5986.

788 Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and
789 nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3:418–426.

790 Nordström K. 2006. Plasmid R1—replication and its control. Plasmid 55:1–26.

791 Notley-McRobb L, King T, Ferenci T. 2002. *rpoS* mutations and loss of general stress
792 resistance in *Escherichia coli* populations as a consequence of conflict between
793 competing stress responses. J. Bacteriol. 184:806–811.

794 Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD,
795 Pyshkin A, Sirotkin A, Sirotkin Y, et al. 2013. Assembling single-cell genomes and mini-
796 metagenomes from chimeric MDA products. J. Comput. Biol. 20:714–737.

797 Patten CL, Kirchhof MG, Schertzberg MR, Morton RA, Schellhorn HE. 2004. Microarray
798 analysis of RpoS-mediated gene expression in *Escherichia coli* K-12. Mol. Genet.
799 Genomics 272:580–591.

800 Peng B, Kimmel M. 2005. simuPOP: a forward-time population genetics simulation
801 environment. Bioinformatics 21:3686–3687.

802 Porse A, Schønning K, Munck C, Sommer MOA. 2016. Survival and evolution of a large
803 multidrug resistance plasmid in new clinical bacterial hosts. Mol. Biol. Evol. 33:2860–
804 2873.

805 Projan SJ, Monod M, Narayanan CS, Dubnau D. 1987. Replication properties of pIM13, a
806 naturally occurring plasmid found in *Bacillus subtilis*, and of its close relative pE5, a
807 plasmid native to *Staphylococcus aureus*. J. Bacteriol. 169:5131–5139.

808 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
809 features. Bioinformatics 26:841–842.

810 R Development Core Team. 2017. R: A language and environment for statistical computing.
811 Vienna, Austria: the R foundation for statistical computing. Available from: https://www.R-
812 project.org/

813 Robin ED, Wong R. 1988. Mitochondrial DNA molecules and virtual number of mitochondria
814 per cell in mammalian cells. J. Cell. Physiol. 136:507–513.

815 Rodriguez-Beltran J, Hernandez-Beltran JCR, DelaFuente J, Escudero JA, Fuentes-
816 Hernandez A, MacLean RC, Peña-Miller R, San Millan A. 2018. Multicopy plasmids allow
817 bacteria to escape from fitness trade-offs during evolutionary innovation. Nat. Ecol. Evol.
818 2:873–881.

819 Rosano GL, Ceccarelli EA. 2014. Recombinant protein expression in *Escherichia coli*:
820 advances and challenges. Front. Microbiol. 5:172.

821 Sambrook J, Russell DW. 2012. Molecular Cloning. Cold Spring Harbor, N.Y.: Cold Spring
822 Harbor Laboratory Press.

823 San Millan A, Escudero JA, Gifford DR, Mazel D, MacLean RC. 2016. Multicopy plasmids
824 potentiate the evolution of antibiotic resistance in bacteria. Nat. Ecol. Evol. 1:10.

825 San Millan A, Peña-Miller R, Toll-Riera M, Halbert ZV, McLean AR, Cooper BS, MacLean RC.
826 2014. Positive selection and compensatory adaptation interact to stabilize non-
827 transmissible plasmids. Nat. Commun. 5:5208.

828 Sandegren L, Linkevicius M, Lytsy B, Melhus Å, Andersson DI. 2011. Transfer of an
829 *Escherichia coli* ST131 multiresistance cassette has created a *Klebsiella pneumoniae*-
830 specific plasmid associated with a major nosocomial outbreak. J. Antimicrob. Chemother.
831 67:74–83.

832 Santos-Lopez A, Bernabe-Balas C, Ares-Arroyo M, Ortega-Huedo R, Hoefer A, San Millan A,
833 Gonzalez-Zorn B. 2017. A naturally occurring single nucleotide polymorphism in a
834 multicopy plasmid produces a reversible increase in antibiotic resistance. Antimicrob.
835 Agents Chemother. 61.

836 Satoh M. 1991. Organization of multiple nucleoids and DNA molecules in mitochondria of a
837 human cell. Exp. Cell Res. 196:137–140.

838 Schmidt R, Ahmetagic A, Philip DS, Pemberton JM. 2011. Catabolic plasmids. In: eLS, (Ed.).

839 Shetty RP, Endy D, Knight TF. 2008. Engineering BioBrick vectors from BioBrick parts. J.
840 Biol. Eng. 2:5.

841 Simon R, Priefer U, Puhler A. 1983. A broad host range mobilization system for *in vivo*
842 genetic engineering: transposon mutagenesis in gram negative bacteria. Nat. Biotech.
843 1:784–791.

844 Skulj M, Okrslar V, Jalen S, Jevsevar S, Slanc P, Strukelj B, Menart V. 2008. Improved
845 determination of plasmid copy number using quantitative real-time PCR for monitoring
846 fermentation processes. Microb. Cell Fact. 7:6.

847 Spencer DH, Tyagi M, Vallania F, Bredemeyer AJ, Pfeifer JD, Mitra RD, Duncavage EJ.
848 2014. Performance of common analysis methods for detecting low-frequency single
849 nucleotide variants in targeted next-generation sequence data. J. Mol. Diagn. 16:75–88.

850 Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson
851 LW, Kasarskis A, Batty EM, et al. 2016. Evolutionary history of the global emergence of
852 the *Escherichia coli* epidemic clone ST131. mBio 7:e02162.

853 Textor S, Wendisch VF, De Graaf AA, Muller U, Linder MI, Linder D, Buckel W. 1997.
854 Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in
855 bacteria. Arch. Microbiol. 168:428–436.

856 Thomas CM, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer
857 between bacteria. Nat. Rev. Mol. Cell Biol. 3:711–721.

858 Ullrich A, Shine J, Chirgwin J, Pictet R, Tischer E, Rutter WJ, Goodman HM. 1977. Rat
859 insulin genes: construction of plasmids containing the coding sequences. 196:1313–
860 1319.

861 Vega NM, Gore J. 2014. Collective antibiotic resistance: mechanisms and implications. Curr.
862 Opin. Microbiol. 21:28–34.

863 Vieira J, Messing J. 1982. The pUC plasmids, an M13mp7-derived system for insertion
864 mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268.

865 Wang Y. 2017. Spatial distribution of high copy number plasmids in bacteria. Plasmid 91:2–8.

866 Wilm A, Aw PPK, Bertrand D, Yeo GHT, Ong SH, Wong CH, Khor CC, Petric R, Hibberd ML,
867 Nagarajan N. 2012. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for
868 uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nuc.
869 Acids Res. 40:11189–11201.

870 Wobbrock JO, Findlater L, Gergle D, Higgins JJ. 2011. The aligned rank transform for
871 nonparametric factorial analyses using only ANOVA procedures. Proceedings of the
872 SIGCHI Conference on Human Factors in Computing Systems. 143-146.

873 Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to
874 detect break points of large deletions and medium sized insertions from paired-end short
875 reads. Bioinformatics 25:2865–2871.

876 Zhong C, Peng D, Ye W, Chai L, Qi J, Yu Z, Ruan L, Sun M. 2011. Determination of plasmid
877 copy number reveals the total plasmid DNA amount is greater than the chromosomal
878 DNA amount in *Bacillus thuringiensis* YBT-1520. PLoS ONE 6:e16025.

879

880

881 **Acknowledgments**

882 We thank Sarah Kovarik, Mathias Krüger and Rabea Plohr for their assistance in the
883 experimental work. We thank Maxime Godfroid and Devani Romero Picazo for critical
884 comments on the manuscript. We thank Wolfgang Streit for providing the pBBR1 plasmid.
885 The study was supported by the European Research Council (Grant No. 281357 awarded to
886 TD).

887 **Author contributions**

888 J.I, I.M. & T.D. designed the experiment. J.I. established and performed the experimental
889 evolution study and the comparative genomic analysis. A.K. performed the simulated
890 evolution. P.R. sequenced the genomes. T.W and N.H took part in the experimental work.
891 C.W., G.L, A.K. and T.D. took part in the computational analysis. All authors interpreted the
892 results and drafted the manuscript.

893 **Competing interests**

894 The authors declare no competing financial interests and no conflict of interest.

895

896 **Tables**

897 Table 1. Number of intergenic and synonymous point mutations detected in the subsampled
898 read data (104x). Due to the inherent difficulties in the robust inference of insertions and
899 deletions (indels; e.g., Spencer et al. 2014), indels are excluded from the counts of observed
900 point mutations. Hence, our estimate of mutation frequency is conservative. A detailed list of
901 the reported statistics for each of the six biological replicates belonging to one of the eight
902 experimental factor groups is shown in Supplementary Table S1. We note that additional low
903 frequency point mutations were detected on the pLC plasmid considering the full sequencing
904 coverage (see variant details in Supplementary Table S2). A detailed list of all observed
905 variants on the chromosome is given in Supplementary Table S3. The variant distribution
906 across the replicons is depicted in Supplementary Fig. S1.

			Chromosome		Plasmid		
Plasmid type	Host genotype	T (°C)	Observed number of point mutations	Substitution per site per generation[1] $\times 10^{-9}$ (*Mean ± SD*)	PCN[2] (*Mean ± SD*)	Expected number of point mutations[3] (*Mean ± SD*)	Observed number of point mutations
pLC	Wild-type	37	17	2.5 ± 0.7	13.0 ± 2.1	$10.2 \times 10^{-2} \pm 2.8 \times 10^{-2}$	0
		42	16	2.4 ± 1.8	4.0 ± 2.2	$3.6 \times 10^{-2} \pm 3.6 \times 10^{-2}$	0
	Hypermutator	37	148	22.4 ± 5.2	6.0 ± 1.6	$4.3 \times 10^{-1} \pm 1.3 \times 10^{-1}$	1
		42	87	13.2 ± 3.3	3.6 ± 1.2	$1.6 \times 10^{-1} \pm 9.1 \times 10^{-2}$	1
pHC	Wild-type	37	6	0.87 ± 0.6	111.3 ± 66.2	$3.7 \times 10^{-1} \pm 3.5 \times 10^{-1}$	0
		42	6	0.87 ± 0.8	99.7 ± 23.2	$2.9 \times 10^{-1} \pm 2.7 \times 10^{-1}$	0
	Hypermutator	37	98	14.3 ± 3.8	64.8 ± 20.2	3.3 ± 1.7	0
		42	147	21.4 ± 11.5	89.4 ± 27.5	6.5 ± 4.6	0

907

908 [1] The chromosomal substitution rate is calculated as $R_{Chr} = N_{obs} \times (N_{sites} \times N_{gen})^{-1}$, where N_{obs}
909 is the number of observed point mutations, N_{sites} is the number of intergenic and synonymous
910 sites on the chromosome having a coverage ≥10 and N_{gen} is the number of generations in
911 our experiment.

912 [2] Plasmid copy number (PCN) was estimated from the coverage of plasmid and
913 chromosome replicons (detailed list in Supplementary Table S1).

914 [3] The expected number of point mutations on the plasmid is calculated as $N_{exp} = R_{Chr} \times N_{sites}$
915 $\times$ PCN $\times N_{gen}$, where N_{sites} is the number of intergenic and synonymous sites on the plasmid
916 having a coverage ≥10.

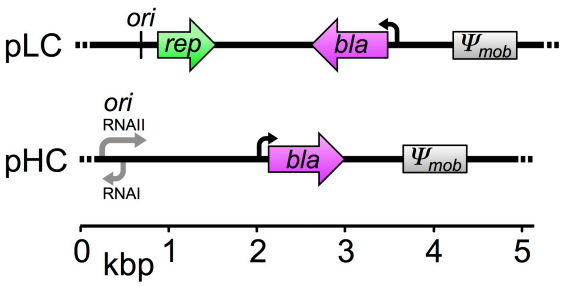
917

918

919

920 **Figures**

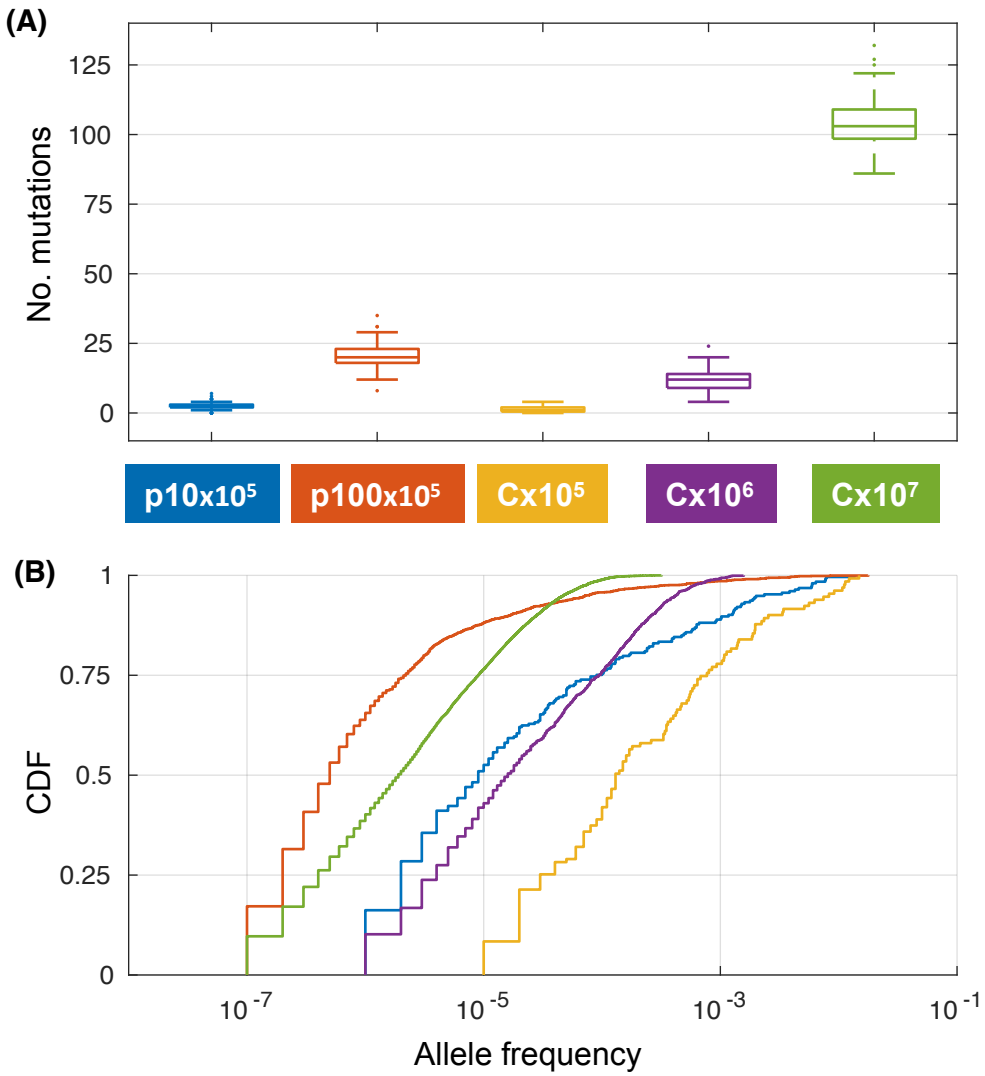
921 **Figure 1.** Genetic maps of plasmid pLC and pHC. The plasmids are of comparable genome
922 size and differ in the origin of replication. The *mob* gene originally found in pBBR1 was non-
923 functionalized by truncation during plasmid construction (denoted as Ψ_{mob}).



924

925

926 **Figure 2**. Simulated evolution of plasmids and chromosomes under neutrality with mutation
927 (500 loci, mutation rate 2 × 10^{-9} per locus per generation). (**A**) Distribution of the number of
928 mutations after 1,000 generations in 100 replicates. The color of the boxes corresponds to
929 the combination of replicon and population size as in the legend. (**B**) Joint cumulative
930 distribution function (CDF) of the allele frequency of all mutations present in the replicates
931 after 1,000 generations.



932

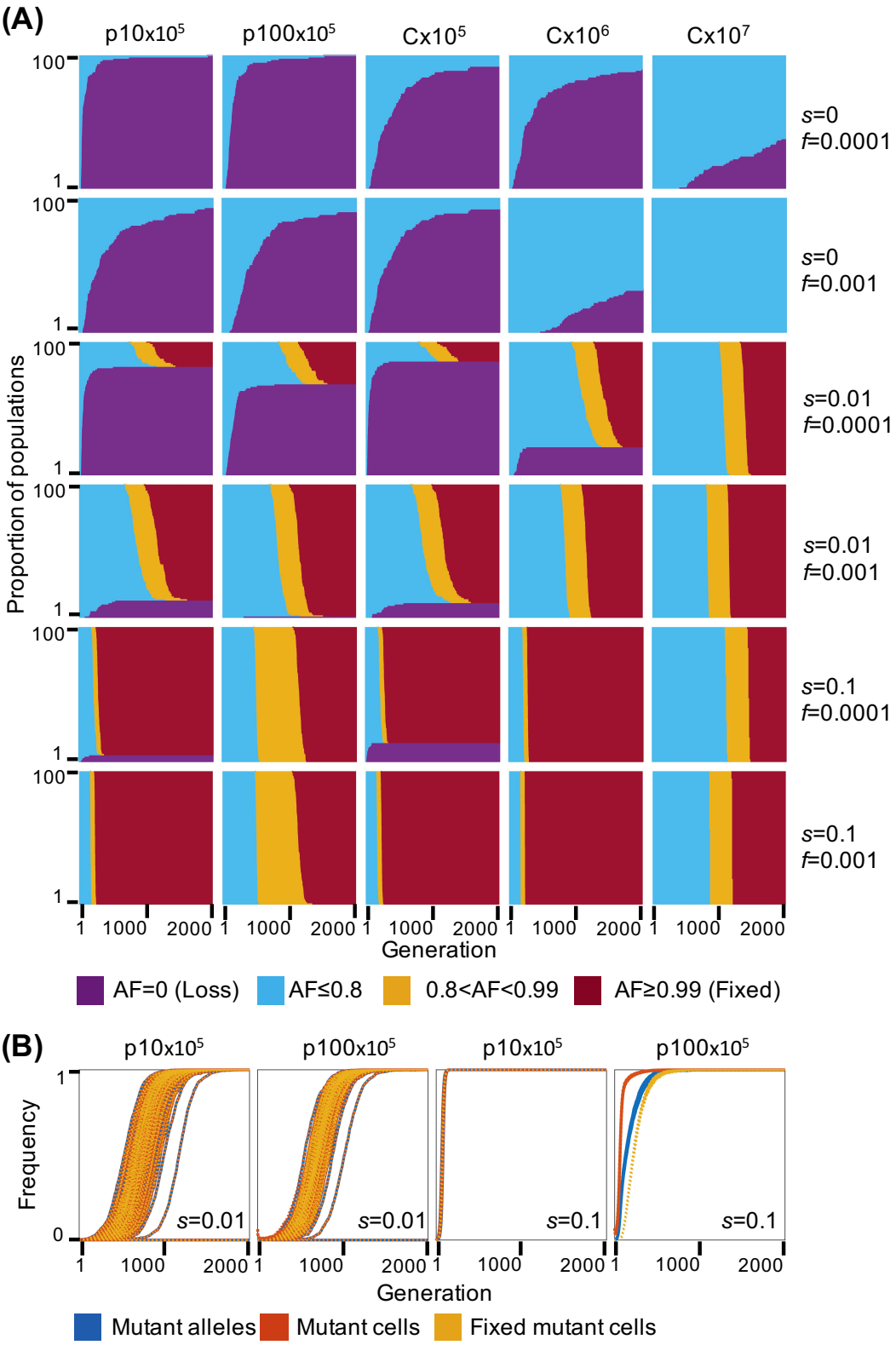
933

934

935 **Figure 3**. Simulated evolutionary dynamics of mutant alleles on plasmids and chromosomes.
936 Allele frequency dynamics over 2,000 generations in 100 replicate simulated populations. (**A**)
937 Columns are the five combinations of simulated replicon type and population size, and rows
938 are the different combinations of selective coefficient (*s*) and initial allele frequency (*f*).
939 Generations are on the x-axis, proportion of AF of replicates are stacked along the y-axis,
940 population allele frequency is colour coded. Note that the starting frequency of plasmid
941 alleles is implemented as the frequency of mutant alleles in the total plasmid population.
942 Thus, the initial frequency of cells with a mutant allele in the population is higher in the
943 plasmid simulations compared to chromosome simulations (see Supplementary Table S4 for
944 expected number of mutant cells). (**B**) Frequency trajectories for plasmid simulations with
945 initial allele frequency f = 0.001. The frequency of mutant alleles is calculated as the number
946 of mutant plasmid alleles in the total plasmid population. The frequency of mutant cells is
947 calculated as the proportion of cells with at least one mutant plasmid allele within the
948 population. The frequency of fixed mutant cells is the proportion of cells in which all plasmids
949 loci contain the mutant allele (i.e., the plasmid allele is fixed in the cell). Simulation results
950 are presented for s = 0.1 where the difference between mutant plasmid allele dynamics and
951 mutant cells in the population is pronounced. The results of the simulation where s = 0.01 are
952 presented for comparison; in this setting, dynamics of mutant plasmid alleles and mutant
953 cells are similar.

954

955


(A)
p10x10^5
p100x10^5
Cx10^5
Cx10^6
Cx10^7
s=0
f=0.0001
s=0
f=0.001
s=0.01
f=0.0001
s=0.01
f=0.001
s=0.1
f=0.0001
s=0.1
f=0.001
Proportion of populations
100
1
1
1000
2000
Generation
AF=0 (Loss)
AF≤0.8
0.8<AF<0.99
AF≥0.99 (Fixed)
(B)
p10x10^5
p100x10^5
p10x10^5
p100x10^5
Frequency
1
0
s=0.01
s=0.01
s=0.1
s=0.1
1
1000
2000
Generation
Mutant alleles
Mutant cells
Fixed mutant cells

961 **Figure 4**. Parallel evolution of the chromosome. (**A**) A heatmap representing the number of
962 shared mutated genes among the 48 evolved populations. The number of genes mutated in
963 population pair *x*, *y* is calculated as the number of genes in which genetic variants (single
964 base substitutions or indels) are observed in both populations. Cells along the diagonal
965 present the number of genes where a mutation was detected in the corresponding population
966 and at least one additional population. Annotation bars at the top and left right present the
967 experimental factors (color-coded as shown in (C)). (**B**) An interaction plot of the main factors
968 in the experiment as calculated from the complete set of allele frequencies of parallel
969 variants (included in Supplementary Table S3). Dots represent means and error bars
970 represent the standard error of the mean. ANOVA demonstrated significant effects for
971 plasmid replicon type and temperature along with significant interactions between the three
972 factors used in the experiment (statistics in Supplementary Table S5). (**C**) Color-coded matrix
973 of variants whose allele frequency is significantly different among the main factors or their
974 combination (using ANOVA on aligned rank transformed allele frequencies and FDR;
975 statistics in Supplementary Table S6). Intragenic variants (11 synonymous and 22 non-
976 synonymous) are indicated by the gene symbol and, in case of a single base substitution,
977 coded as [ancestral amino acid][amino acid position][alternative amino acid]. The two
978 insertion mutations are indicated by a gene symbol and coded as [nucleotide position in
979 gene][+ inserted nucleotide]. [*] indicates a non-sense codon. The two intergenic variants are
980 indicated by [i] coded as [ancestral nucleotide][genomic position][alternative nucleotide].

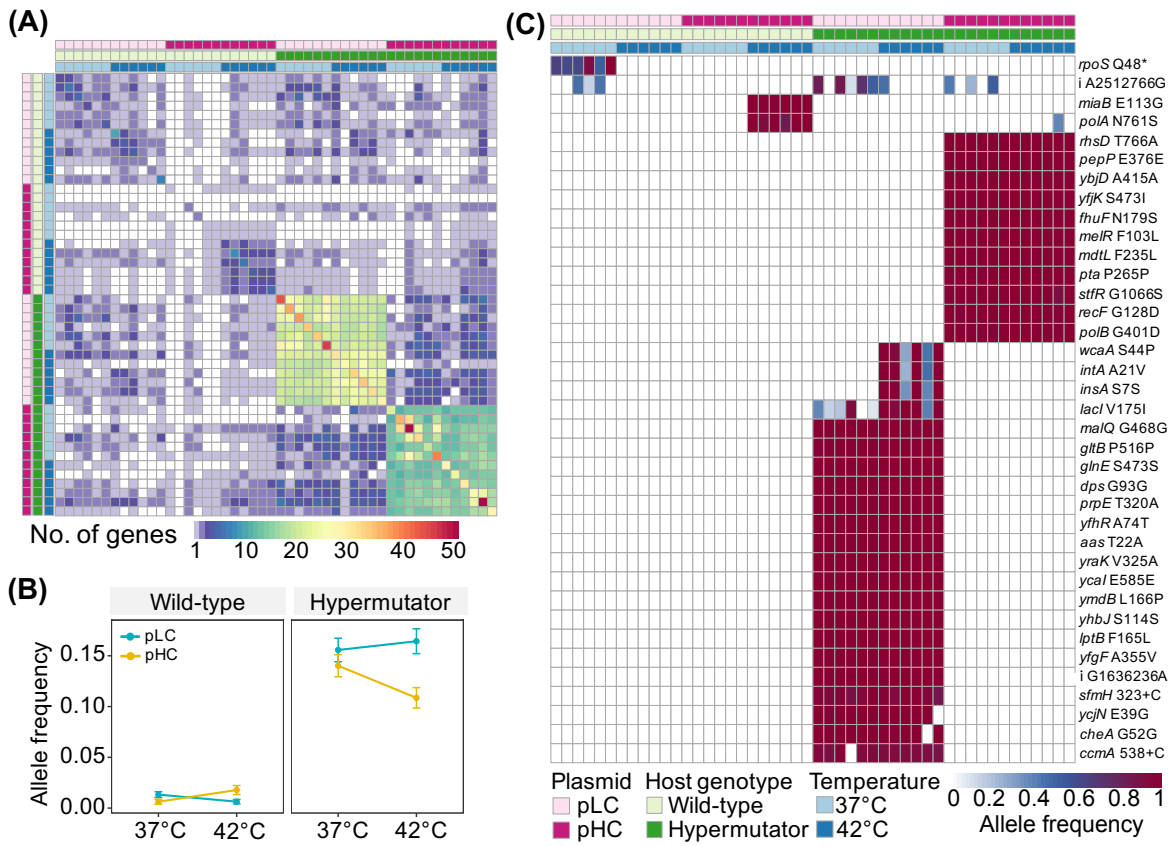
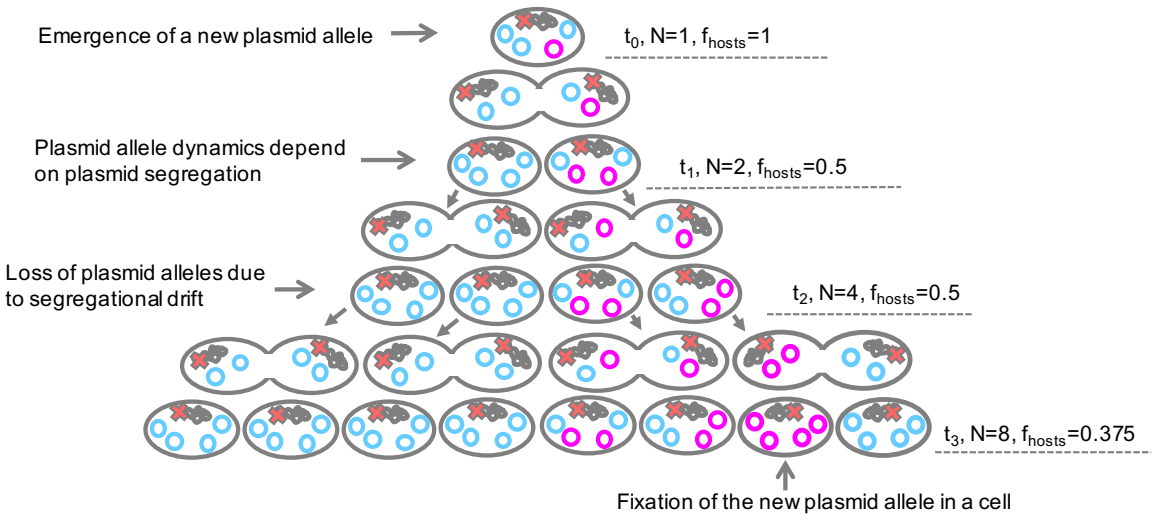


Figure 5. An illustration of the consequences of segregational drift to plasmid allele dynamics. The illustrated plasmid copy number is four and its segregation into daughter cells is balanced (i.e., the daughter cells inherit an equal number of plasmids). In the depicted scenario, one allele emerges on the chromosome (red X) and another allele emerges on the plasmid (magenta) at t_0. Both alleles are considered neutral (i.e., s = 0). Changes in population size (N) and the frequency of hosts (f_{hosts}) over three generations are presented. The chromosomal allele is inherited into all daughter cells, such that it is present in the total population. Random segregation of the plasmid genotypes leads to decrease in the number of hosts. At t_3, only a small minority in the population harbours the new plasmid allele. In the absence of selection for the plasmid allele presence, the new plasmid allele is deemed to remain at a very low frequency within the population (or be lost). Note: this is a simplistic example that does not include loss of host cells due to drift and it does not include consideration of recombination among plasmids.


Emergence of a new plasmid allele
t_0, N=1, f_{hosts}=1
Plasmid allele dynamics depend on plasmid segregation
t_1, N=2, f_{hosts}=0.5
Loss of plasmid alleles due to segregational drift
t_2, N=4, f_{hosts}=0.5
t_3, N=8, f_{hosts}=0.375
Fixation of the new plasmid allele in a cell