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. 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.

39 ## Introduction

40 Plasmids are genetic elements that colonize prokaryotic cells where they replicate
41 independently of the chromosome. They are considered as a major driving force in
42 prokaryotic ecology and evolution as they can be transferred between cells, making them
43 potent agents of lateral gene transfer (Lederberg and Tatum 1946; Thomas and Nielsen
44 2005) and microbial warfare (Czárán et al. 2002; Majeed et al. 2011). Plasmids may
45 contribute to the evolution of their host via the supply of genes encoding for new functions
46 that are beneficial under growth limiting conditions (e.g., antibiotics resistance (Porse et al.
47 2016), resistance to heavy metals (Gullberg et al. 2014; Dziewit et al. 2015)) or during
48 adaptation to new habitats (e.g., catabolic functions (Schmidt et al. 2011)). Nonetheless, the
49 importance of plasmids goes beyond microbial evolution as they are widely used as vectors
50 for genetic engineering (Simon et al. 1983; Bevan 1984) as well as for applications in
51 biotechnology (Ullrich et al. 1977) and synthetic biology (Shetty et al. 2008). Therefore,
52 understanding plasmid biology and evolution is of major interest for basic microbiology
53 research and biological applications.

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

66 has the potential to elevate the probability for the emergence of novel mutations in the gene
67 open reading frame. Consequently, the mutational supply of plasmid-encoded genes is
68 increased for multicopy plasmids. Indeed, it has been previously shown that multicopy
69 plasmids have the potential to enable rapid evolution of plasmid-encoded antibiotic
70 resistance genes under strong selective conditions for the antibiotic resistance (San Millan et
71 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 and Kuroiwa 1991)). Thus, the mutational supply on the mitochondrial genome is
90 larger than the mitochondrial genome size by several orders of magnitude. The combination
91 of high mutational supply on the mitochondrion genome and random segregation of the
92 mitochondria during cell division lead to genetic heterogeneity of the mitochondrial genomes
93 within a 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.,
96 (He et al. 2010)). The consequences of segregational drift to the evolution of multicopy
97 plasmids have been so far overlooked.

98 Population genetics theory postulates that the effect of random genetic drift on allele
99 dynamics is inversely correlated with the effective population size. In a larger population, the
100 impact of genetic drift on the dynamics of allele frequency is reduced (see (Lanfear et al.
101 2014) for review). Thus, the implications of segregational drift to plasmid genome evolution
102 are expected to be dependent on the plasmid population size within the cell, i.e., the plasmid
103 copy number. Here we investigate plasmid genome evolution in light of the segregational drift
104 using experimental evolution and *in silico* simulation of allele 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 replicate in a broad host range in
112 the absence of 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 environmental conditions. To vary the mutation
124 rate, the plasmids were evolved in two host genotypes, *E. coli* K-12 substrain MG1655 (wild-
125 type, wt) 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 assessed 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 mapped to the reference *E. coli* MG1655 genome and to the two plasmids.
162 The plasmid copy number within the sequenced plasmid host population was estimated from
163 the average sequencing coverage of the plasmids relative to that of the chromosome. The
164 sequencing coverage varied among the populations and the replicons: the chromosomal
165 coverage ranged between 104x and 300x, while the average coverage on the plasmids was
166 between 325x and 28,054x (Supplementary Table S1). The high coverage on the plasmid
167 genomes increases the detection sensitivity for rare and low-frequency variants. Nonetheless,
168 in order to compare the number of mutations across replicons and populations, the
169 sequencing reads of each replicon (i.e., chromosome and plasmids) were subsampled to
170 match the lowest-coverage dataset unless stated otherwise. Furthermore, variants occurring
171 in both ancestral and evolved populations were excluded.

172 A comparative genomic analysis of the subsampled sequencing data revealed two
173 point mutations on the pLC plasmid (Table 1). Both mutations are located in an intergenic
174 region and their allele frequency (AF) is lower than 0.10 (Supplementary Table S2). The

175 paucity of plasmid variants should be evaluated against the background of variants on the
176 host chromosome, where we observed a total of 1,394 point mutations. We calculated the
177 expected frequency of point mutations on the two plasmids using the chromosomal
178 substitution frequency, the plasmid genome size and the plasmid copy number. This shows
179 that mutations are unlikely to occur on either of the plasmid types evolved in the wild-type
180 host due to the combination of low substitution rate and the plasmid small genome size
181 (Table 1). In contrast, the combination of high genetic polymorphism in the hypermutator host
182 with the high copy number of pHC are likely to result in observable genetic variants in the
183 pHC plasmid within the time span of our experiment (Table 1 and Supplementary Table S1).
184 Nonetheless, the subsampled sequencing data for pHC did not reveal any point mutations
185 occurring on that plasmid. Notably, we detected two single nucleotide insertions in pHC
186 intergenic regions with an allele frequency of AF = 0.3 and AF = 0.35. In order to increase
187 the resolution for the detection of presumably low-frequency point mutations, we applied the
188 same variant calling pipeline to the sequencing read datasets without subsampling (i.e., full
189 sequencing coverage). This analysis revealed one point mutation on plasmid pHC at an
190 intergenic position with a sequencing depth of ~9,600 and a low allele frequency of AF = 0.03
191 (Supplementary Table S2). In contrast, considering the full sequencing depth for plasmid
192 pLC, which is on average ~10x lower than that of plasmid pHC, we observed two additional
193 intergenic point mutations (Supplementary Table S2). Thus, despite the high sequencing
194 depth, the observed number of point mutations on plasmid pHC does not reach the expected
195 frequency of point mutations, even when the full sequencing coverage is included in the
196 analysis. On the other hand, the number of point mutations detected on plasmid pLC is
197 higher than what we expected (Table 1).

198 The number of point mutations on the pHC plasmid was further validated by the
199 analysis of the total pHC population sequencing data (i.e., samples were prepared without
200 selection for plasmid hosts). Considering the full sequencing coverage, we detected one
201 additional point mutation on the plasmid. The intergenic point mutation was observed with a
202 low allele frequency of AF = 0.04 (Supplementary Table S2) in one of the pHC populations

203 having the hypermutator host genotype. In addition, we compared the number of point
204 mutations observed on the chromosome in the pooled pHC host populations to that of the
205 total population sequencing data (Supplementary Table S1). The comparison shows that the
206 wild-type populations include, on average, two additional point mutations per population in
207 comparison to the corresponding data of pooled plasmid host colonies. The comparison of
208 hypermutator populations shows that the total population sequencing data includes, on
209 average, 14 point mutations per population less than the pooled colonies sequencing data.
210 Overall, the above comparison between the two sequencing approaches demonstrates that
211 the number of point mutations observed in the pooled host colonies data is representative
212 both for the number of variants on the pHC plasmid and the chromosome.

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

224 ***Plasmid population dynamics.***

225 What are the theoretical expectations for the impact of segregational drift on the dynamics of
226 plasmid alleles in comparison to chromosomal alleles? To address this question we
227 simulated plasmid and chromosome allele dynamics. For this purpose, we applied the
228 standard haploid Wright-Fisher model to simulate forward-time evolution of a population of
229 cells that is subject to random genetic drift (Gillespie 2010). In the simulation, the cell

230 population size remains constant throughout evolution. To simulate plasmid evolution, we
231 added an additional layer of random genetic drift by applying a Wright-Fisher model to the
232 plasmid inheritance (similarly to mitochondria simulated evolution (Peng and Kimmel 2005)).
233 Thus, during cell division, plasmid alleles are selected randomly into the daughter cells of the
234 next generation. In terms of population genetics, the plasmid population in the simulated
235 evolution has a substructure within the population. This implies that plasmid allele dynamics
236 have two hierarchical levels: within the cell and within the total population.

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

258 the chromosome of the respective population size (Fig. 2B). This demonstrates that the
259 increase in allele frequency of plasmid mutations is generally slower in comparison to the
260 mutations emerging on a haploid chromosome.

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

284 Performing the simulations where the allele has a positive selection advantage (s =
285 0.01) reveals a higher loss rate of the mutant allele in p10x10^5 in comparison to p100x10^5,
286 yet the median fixation time is lower for p10x10^5 than for p100x10^5. Notably, the median
287 fixation time of plasmid mutant alleles that are not lost is smaller in comparison to
288 chromosome populations of comparable total replicon numbers when the initial frequency is
289 low albeit the majority of plasmid populations lost the mutant allele under these conditions
290 (Fig. 3, s = 0.01, f = 0.0001; Supplementary Table S4). The median fixation time of plasmid
291 mutant alleles is similar to the chromosomal one when the initial frequency is increased (Fig.
292 3, s = 0.01, f = 0.001; Supplementary Table S4). With a high selection coefficient (s = 0.1),
293 mutant allele loss is smaller than for s = 0.01 and the median fixation time of p10x10^5 is
294 similar to the chromosome populations. Notably, the median fixation time in p100x10^5 is
295 substantially higher in comparison to all other replicons under these conditions (Fig. 3, s =
296 0.1, f = 0.0001 and f = 0.001; Supplementary Table S4).

297 In the presence of balancing selection for multiple plasmid alleles (e.g., alleles
298 supplying resistance for different antibiotics), plasmid loci may remain polymorphic over a
299 long time-scale. Previous studies showed that when the balancing selection is replaced by
300 selection for only one of the plasmid alleles, alternative plasmid alleles may remain
301 polymorphic in the population for a certain time (Bedhomme et al. 2017; Rodriguez-Beltran et
302 al. 2018). We hypothesize that the persistence of alternative plasmid alleles in the population
303 in the absence of selection for the allele, is affected by segregational drift. To test our
304 hypothesis, we performed simulations of plasmid allele dynamics post balancing selection.
305 Thus, the initial mutant allele frequency is set to exactly f = 0.5 (i.e., each cell has equal
306 proportions of the ancestral and mutant allele). Here we compare the dynamics of plasmid
307 alleles to chromosomal alleles with the same total replicons as in the above simulations. Our
308 results show that in the absence of selection for the mutant allele, both the ancestral and the
309 mutant allele remain in the population in equilibrium as expected (Fig. 3, s = 0, f = 0.5; Fig.
310 4A, s = 0, f = 0.5). When the mutant allele is beneficial (s = 0.01 or s = 0.1), the ancestral
311 allele persists in the population when it is on the plasmid for a longer time scale in

312 comparison to a chromosomal allele of the same total replicon population (Fig. 3, s=0.01 and
313 s=0.1, f=0.5). Nonetheless, the proportion of mutant cells in the population declines at the
314 beginning of the simulation, hence, the beneficial mutant allele is tends to be lost in cells –
315 also when the mutant allele is beneficial (Fig. 4A, s = 0.01 and s = 0.1, f = 0.5). Furthermore,
316 the consequences of segregational drift to the loss of the mutant (beneficial) allele are more
317 pronounced in the low copy plasmid (p10) simulations. The simulations of the high copy
318 plasmid (p100) with s = 0.1 show that the ancestral allele may persist in the population four-
319 fold longer than the expected for a chromosomal allele with the same total number of
320 replicons (i.e., $Cx10^7$; Fig. 3; Fig. 4B).

321 The longer fixation time of $p100x10^5$ alleles under the simulated conditions of strong
322 selection (s=0.1) are well explained by the effect of segregational drift on the rate of plasmid
323 allele fixation within the host cell. The high selection coefficient acts in favour of cells
324 harbouring at least one plasmid mutant, such that those cells rise quickly to fixation within the
325 population. In comparison, the increase of the plasmid allele in the total population is slower
326 due to the effect of random genetic drift of plasmid alleles during cell division, which hinders
327 the increase of plasmid allele frequency in daughter cells. In other words, when the plasmid
328 copy number is high, plasmid alleles within the cell remain polymorphic (i.e., heterogeneous)
329 over a longer time scale, also when the mutant allele is beneficial. This is also observed for
330 the post balancing selection simulation where the loss of the non-mutant (less beneficial)
331 alleles is prolonged for higher copy numbers.

332 Notably, a plasmid copy number of 100 is within the range of PCNs observed in
333 natural isolates (Projan et al. 1987; Zhong et al. 2011), hence, the extended fixation (and
334 loss) time observed for $p100x10^5$ is likely to occur also under natural conditions. Since our
335 simulations do not include intra-host plasmid recombination, the observed differences in
336 allele dynamics between plasmids and chromosomes are solely due to inheritance and
337 segregational drift. Overall, the results of the simulation demonstrate that segregational drift

338 of multicopy plasmids has a considerable impact on plasmid allele dynamics in the
339 population by increasing the loss rate and fixation time of plasmid alleles.

340 ***Evolution of the host chromosome.***

341 The comparison between the evolved and ancestral populations revealed that most of the
342 evolved genetic variants are found on the chromosome (Table 1 and Supplementary Table
343 S3). The majority of the chromosomal variants (90%) in our experiment are observed in the
344 hypermutator populations where the substitution rate is, on average, 10-fold higher in
345 comparison to the wild-type populations (Table 1). Genetic variants are observed in 638
346 protein coding genes and ncRNA genes, of which 173 (27%) genes accumulated genetic
347 variants in more than one population (Fig. 5A). Of the mutated genes, 60 genes (35%) are
348 specific to populations belonging to one of the eight experimental factor combinations,
349 indicating that the combination of host genotype, plasmid replicon type and temperature had
350 an impact on the evolution of these genes. Furthermore, a total of 50 mutated genes (29%)
351 are shared among hypermutator populations having the same plasmid type, regardless of the
352 growth temperature (Fig. 5A). Hypermutator-pLC populations share, on average, 19.62 ±
353 2.09 (*SD*) mutated genes, whereas hypermutator-pHC populations show a lower level of
354 parallelism with 14.45 ± 3.30 shared genes on average (SD). Furthermore, two small clusters
355 of parallel mutated genes are observed in the wild-type hosts, the first cluster comprises the
356 pLC populations and the second cluster includes the pHC populations evolved at 42°C.

357 At the resolution of specific mutations, 144 (15%) variants were detected in more than
358 one of the 48 evolved populations. The allele frequency of variants occurring in more than
359 one population is significantly higher than the allele frequency of population specific variants
360 ($P < 0.001$, using two-sample Kolmogorov-Smirnov test, $n_{parallel} = 720$, $n_{specific} = 841$). To
361 quantify the contribution of the three experimental evolution factors to parallel evolution of the
362 host chromosome, we compared the allele frequency of parallel genetic variants among all
363 populations. We observed a significant effect of plasmid type ($P = 1.1\text{x}10^{-69}$) and temperature
364 ($P = 3.9\text{x}10^{-46}$) on the allele frequency variation (using ANOVA on aligned rank transformed

365 allele frequencies (Wobbrock et al. 2011); detailed statistics in Supplementary Table S5). In
366 the wild-type populations, there is a strong interaction between plasmid type and temperature,
367 whereas in the hypermutator populations, the same interaction is observed along with a main
368 effect of the plasmid type (Fig. 5B). Our results thus reveal that the effect plasmid replicon
369 type on the chromosomal allele dynamics is equal or larger in comparison to the growth
370 temperature. Elevated growth temperature has an impact on *E. coli* physiology and
371 consequences to its evolution (Deatherage et al. 2017). A recent study showed that plasmid
372 acquisition may have a physiological effect on its *Pseudomonas aeruginosa* host (San Millan
373 et al. 2018). Our results suggest that the effect of plasmids on the host metabolism and
374 evolution can be equal or larger than that of the environmental temperatures.

375 To further identify variants that are characteristic for specific factor combinations, we
376 tested for the effect of the factors and their interaction on the allele frequency of each parallel
377 variant (using ANOVA on aligned rank transformed allele frequencies and FDR). This results
378 in 37 (2.7%) variants whose allele frequency is significantly different among the factor
379 combinations (Fig. 5C; Supplementary Table S6). A single variant occurred in all wt-pLC
380 populations evolved at 37°C, where a nonsense mutation of Glutamine into a stop codon
381 occurred in the *rpoS* gene. Hence, that gene is likely non-functionalized in these populations.
382 The *rpoS* encodes the RNA polymerase σ^S-subunit that is the master regulator of the general
383 stress response in *E. coli* (e.g., (Patten et al. 2004)). Loss of RpoS function was previously
384 reported to occur rapidly in nutrient-limited chemostat populations (Notley-McRobb et al.
385 2002). A cluster including two non-synonymous substitutions is observed for wt-pHC
386 populations evolved at 42°C. The substitutions occurred in a gene encoding for a tRNA
387 modification enzyme (*miaB*) and a DNA polymerase I gene (*polA*). Overall, we identified 11
388 different *polA* mutations in 16 pHC-populations and none in pLC-populations (Supplementary
389 Table S3). Mutations in *polA* have been previous reported to decrease the copy number of
390 high copy number ColE1 plasmid mutants in *E. coli* (Yang and Polisky 1993). Most of the
391 *polA* mutations in our set are located in the same region (Klenow region), similarly to the

392 previously reported mutants. This suggests that the *polA* mutations observed in the wt-pHC
393 populations evolved as a response to the pHC high copy number.

394 Mutations that are characteristic to hypermutator-pLC populations (Fig. 5C) are found
395 in genes that belong to four main functional categories, including carbohydrate metabolism,
396 fatty acid metabolism, biofilm formation and stress response. Example genes where amino
397 acid replacements have occurred are propionyl-CoA synthetase (*prpE*), a part of the
398 propionate metabolism pathway (Textor et al. 1997), and a fimbrial-like adhesin protein
399 (*sfmH*) that is known to promote adhesion in *E. coli* (Korea et al. 2010). Two additional non-
400 synonymous substitutions in genes encoding for an integrase (*intA*) and a glycosyl
401 transferase (*wcaA*) are common to all six hypermutator-pLC populations evolved at 42°C (Fig.
402 5C). The evolution of hypermutator-pHC populations is characterized by a majority of non-
403 synonymous point mutations. These variants are fixed (i.e., AF > 0.9) in all hypermutator-
404 pHC populations. The mutated genes are classified into three main categories including
405 carbohydrate metabolism, stress response and DNA replication. Example mutated genes are
406 DNA polymerase II (*polB*) and DNA gap repair protein (*recF*). Taken together, the differential
407 acquisition of mutations in hypermutator populations carrying plasmid pLC or pHC suggests
408 that the plasmid replicon type had a significant impact on the host chromosome evolution.

409 **Discussion**

410 Previous experimental evolution studies of plasmid-host adaptation noted that evolved
411 variants are mostly observed in the host chromosome rather than the plasmid genome
412 (Bouma and Lenski 1988; San Millan et al. 2014; Loftie-Eaton et al. 2017), and this is indeed
413 the case also for our study. Naturally, the probability for a mutation in a chromosome of the
414 size of *E. coli* (4.6 Mb) is higher than that of our model plasmid genomes. Consequently,
415 more mutations are likely to arise on the chromosome than on the plasmid. Notably, in most
416 studies of plasmid-host coevolution (including ours), the selection pressure is imposed on the
417 maintenance of the plasmid in the host population rather than a specific plasmid allele (e.g.,

418 as in (San Millan et al. 2016)). Because plasmids depend on the host machinery for their
419 replication and persistence, the selection for plasmid maintenance is, in practice, imposed on
420 the host machinery rather than on the plasmid itself. Moreover, the potential adaptive
421 responses to the selective pressure for the plasmid maintenance are numerous in the
422 complex host chromosomes (e.g., (San Millan et al. 2014; Harrison et al. 2015)), but are very
423 limited in the genetically compact plasmid genome (e.g., (Porse et al. 2016)). This, in
424 combination with the higher probability for mutations on the chromosome, implies that most
425 variants - adaptive or neutral - are expected to occur on the chromosome. This can explain
426 the high frequency of genetic variants observed on the chromosome in our experiment, as
427 well as in other studies (e.g., (San Millan et al. 2014; Harrison et al. 2015; Loftie-Eaton et al.
428 2017)).

429 The significant effect of plasmid type on the chromosomal allele frequency observed
430 in our experiment (Figs. 5A,C) suggests that evolution of the host chromosome depends on
431 the replicon type of our model plasmids, with the origin of replication being a prominent
432 characteristic thereof. The evolutionary scenario portrayed here is highly relevant for
433 plasmid-mediated evolution of antibiotics resistant strains (e.g., (Sandegren et al. 2011;
434 Jechalke et al. 2014; Stoesser et al. 2016)) as in such incidents, the antibiotics impose a
435 selection pressure for the plasmid maintenance in the population. Our findings suggest that
436 the evolution of plasmid-mediated antibiotics resistance is expected mostly to involve the
437 host adaptation to the plasmid replicon type rather than evolution of the plasmid encoded
438 trait.

439 The results from our experimental evolution of model plasmids demonstrate that the
440 low-copy replicon has accumulated more point mutations in comparison to the high-copy
441 replicon (Table 1). This observation contrasts with an expectation derived solely from
442 mutation rates and genome size. This deviation from the naive expectation can be explained
443 by the inverse correlation between population size and the impact of genetic drift on allele
444 dynamics in the population. The low-copy plasmid constitutes a small replicon population

445 within the cell. The dynamics of allele frequency due to drift in small populations are
446 expected to have large fluctuations such that novel alleles are quickly lost or fixed (Bodmer
447 and Cavalli-Sforza 1976). In contrast, the high copy plasmid constitutes a large replicon
448 population within the cell where fluctuations in allele frequency over time are expected to be
449 small (Bodmer and Cavalli-Sforza 1976). The impact of segregational drift on plasmid alleles
450 is best understood in comparison to the chromosomal alleles. While novel chromosomal
451 alleles emerging on a haploid chromosome are inherited to all daughter cells in the
452 population, the inheritance of plasmid alleles depends on the segregation pattern of plasmid
453 copies. Indeed, our simulations demonstrate that the frequency of neutral mutations
454 emerging on a multicopy plasmid is higher in comparison to a haploid chromosome (Fig. 2),
455 which is in agreement with the higher mutational supply of multicopy plasmids (San Millan et
456 al. 2016). Nevertheless, the median allele frequency of mutations arising on multicopy
457 plasmids is lower than that of chromosomal mutations in a comparable population size (i.e.,
458 number of replicons). Hence, the allele frequency of mutations residing on multicopy
459 plasmids is increasing slower in comparison to chromosomal alleles. Following the
460 emergence of a novel plasmid allele, plasmid segregation during cell division has three
461 potential outcomes for the plasmid genotype of the daughter cells: 1) loss of the novel
462 plasmid allele, 2) the presence of polymorphic plasmid alleles in the cell, or 3) fixation of the
463 novel plasmid allele in the cell (see illustration in Fig. 6). The expected effect of plasmid copy
464 number on the allele dynamics is best observed in the simulated evolution of beneficial
465 alleles (i.e., $s > 0$): on the low-copy plasmid (p10), the loss of plasmid alleles occurs more
466 frequently in comparison to the chromosome, but the median fixation time is nearly the same
467 as that of the chromosome. Indeed, previous studies showed that under strong selective
468 conditions, beneficial mutations on multicopy plasmids can rise quickly to fixation (San Millan
469 et al. 2016). Notably, our simulation of beneficial alleles reveals that alleles on the high copy
470 plasmid (p100) remain polymorphic over a longer time scale and have a higher median
471 fixation time (Fig. 3). For neutral alleles, segregational drift of multicopy plasmids hinders the
472 increase of plasmid allele frequency; it leads to frequent and rapid mutant allele loss on the

473 one hand, and to longer fixation time on the other hand. Altogether, our results demonstrate
474 that the effect of segregational drift on plasmid allele dynamics counteracts the effect of
475 higher mutational supply accompanying increased copy number. Consequently, the
476 evolvability of low-copy plasmids is higher than that of high copy plasmids despite the latter
477 higher mutational supply.

478

479 The impact of segregational drift is not expected to be limited to non-mobile plasmids
480 as mobile (or conjugative) plasmids are also found in multicopy state (e.g., (Figurski and
481 Helinski 1979)). Moreover, plasmid copy number may be heterogeneous within the
482 population, especially for plasmids that were modified towards a high copy number such as
483 the pUC origin (e.g., (Münch et al. 2015)). Furthermore plasmid copy number can be variable
484 during evolution (e.g., (San Millan et al. 2016; Santos-Lopez et al. 2017)). In our simulation
485 we demonstrate that segregational drift during the evolution of plasmids having a constant
486 copy number has an effect on their evolutionary rate. This scenario is relevant for naturally
487 occurring plasmids having a tightly controlled copy number and segregation mechanisms.
488 Variable plasmid copy number within the population, and over time, is expected to have
489 implications for the effect of segregational drift on plasmid evolution and may thus lead to
490 evolutionary rate heterogeneity in plasmid genomes. The evolution of multicopy plasmids
491 entails time spans of intra-cellular heterogeneity, which is a situation also encountered by
492 mitochondrial genomes within single cells. Our study thus highlights the operation of
493 analogous evolutionary processes in bacterial plasmids and eukaryotic organelles. In
494 summary, a realistic portrayal and modelling of plasmid genome evolution requires deeper
495 understanding of segregational drift.

496

497 ## Methods

498 ***Bacterial strains, growth conditions and DNA techniques***

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

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

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

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

534 ***Construction of plasmids pLC and pHC***

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

550 ***Experimental design and setup***

551 The evolution experiment was set up using a full factorial design of 2 plasmid types × 2 host
552 genotypes × 2 growth temperatures. Six biological replicates were used for each of the eight
553 experimental factor groups, resulting in a total of 48 populations that were evolved for 800
554 generations. The evolution experiment was conducted in chemostats made from off-the-shelf
555 materials similar in design as previously described for the cultivation of yeast (Miller et al.
556 2013). The chemostat system was modified to enable bacterial culturing over longer time
557 periods. Briefly, air breaks were introduced upstream of culturing vessels to prevent bacterial
558 growth into medium supply lines and culture vessels were equipped with additional effluent
559 ports to ensure constant efflux in case of clogging caused by biofilm-formation (additional
560 information provided in Supplementary Fig. S3). Prior to the evolution experiment, plasmid
561 pLC and pHC, respectively, were transferred into the wild-type and the Δ*mutS* strain and a
562 single colony of each type (wt-pLC, wt-pHC, Δ*mutS*-pLC, and Δ*mutS*-pHC) was used to
563 found a population that was serially passaged (1:1000) daily for about 100 generations in
564 M9/amp$_{100}$ at 37°C to acclimate the populations to the media conditions. At the onset of the
565 evolution experiment, 50 µl overnight culture of each of the four populations (termed here
566 ancestral populations) was used to inoculate 6 replicate chemostat cultures with a working
567 volume of 20 ml at 37°C and 42°C each. The 48 chemostat cultures were grown to stationary
568 densities overnight (i.e., without addition of media) and then continuously incubated at a
569 dilution rate of 0.27 h^{-1}. Hence, the growth rates are homogeneous among all populations
570 through the growth in the chemostat. The bacterial concentration of chemostat cultures
571 grown at steady-state was stable over time and among replicates (M = 1.92 × 10^8 cells per
572 ml, SD = 1.11 × 10^7 cells per ml, n = 432). Nevertheless, in order to ensure the stable
573 maintenance of flow rates during the course of the experiment, chemostat units and tubings
574 were replaced every three weeks (approximately 200 generations). Meanwhile, chemostat
575 populations were stored at -80°C in 25% (v/v) glycerol. From these frozen stocks, the
576 chemostat cultures were restarted as follows: 10 ml of the frozen stock culture were pelleted
577 by low-speed centrifugation (5000×g) for removal of glycerol and the cells were harvested in

578 the final culturing volume of 20 ml and transferred into chemostat vessels. Each population
579 was evolved for 90 days at steady-state resulting in approximately 800 generations. On a
580 regular basis, all chemostat populations were reciprocally tested for cross-contamination
581 using qPCR. To this end, each culture was tested for the presence of the *mutS* gene and for
582 the presence of both plasmids using the following primer pairs: q_pBB_rep_F/R
583 complementary to the *rep* gene on plasmid pLC, q_pUC_ori_F/R complementary to the *ori*
584 on plasmid pHC, and primer q_mutS_F/R complementary to *mutS* gene in the *E. coli*
585 MG1655 background. The proportion of plasmid bearing cells in evolved and ancestral
586 populations was assessed by appropriately diluting samples of chemostat cultures with 1x
587 phosphate-buffered saline, spreading on M9 agar plates and incubating at 37°C or 42°C for
588 14 hours. At least 100 single colonies of each chemostat sample were transferred to
589 M9/amp$_{100}$ and to M9 agar plates. Colonies that grew only on non-selective media were
590 counted as plasmid-free segregants. To verify the absence of the plasmids, 1-2 randomly
591 chosen colonies per population were chosen for testing the presence of the plasmid using
592 qPCR (q_bla_F/R for the plasmids and q_idnT_F/R for the chromosome). The proportion of
593 plasmid-bearing cells was determined from the comparison of the number of colonies grown
594 on ampicillin plates to the total number of transferred colonies grown on non-selective
595 medium.

596 ***Plasmid copy number and loss frequency***

597 Plasmid copy numbers were determined using quantitative PCR (qPCR) as described
598 previously (Skulj et al. 2008), but with primers complementary to *idnT* on the *Escherichia coli*
599 MG1655 chromosome (q_idnT_F/R) and to the *bla* gene (q_bla_F/R) on plasmid pLC and
600 pHC. Samples of overnight cultures were heated at 98°C for 10 min and then frozen at -20°C
601 to disrupt the cells. The qPCR reactions were carried out in a total volume of 10 µl containing
602 1x iTaq Universal SYBR Green Supermix (Bio-Rad Laboratories), 100 nM of each primer
603 (final concentration) and 1 µl sample. All qPCR reactions, including positive and non-
604 template controls, were performed in technical replicates on a CFX Connect Real-Time PCR
605 Detection System (Bio-Rad Laboratories) using the following cycling conditions: 95°C for 3

606 min, and 40 cycles of 10 s at 95°C and 1 min at 59°C. Primer specificities and efficiencies
607 were determined using standard-curve and melt-curve analyses. The ratio between the
608 number of plasmid specific amplicons and chromosome specific amplicons is defined as the
609 plasmid copy number (Skulj et al. 2008) and was here calculated considering the
610 amplification efficiencies of both primer pairs.

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

618 ***Whole genome sequencing***

619 The ancestral populations of the wild-type and the Δ*mutS* strain carrying the designated
620 plasmids (wt-pHC, wt-pLC, Δ*mutS*-pHC, and Δ*mutS*-pLC) and the 48 populations evolved for
621 800 generations were sequenced using Illumina sequencing technology. Population
622 sequencing was used to enable the detection of rare variants and to asses the frequency of
623 alleles within and across populations. Total DNA (genomic DNA plus plasmid DNA) was
624 isolated from archived samples stored at -80°C using the Wizard Genomic DNA Purification
625 kit (Promega). The preparation of DNA for sequencing of the evolved populations differed
626 between the two plasmids due to the difference in their stability within the host. For pLC-
627 populations, 1 ml of cell suspension from thawed samples was used for DNA extraction. The
628 frequent loss of pHC led to the presence of plasmid-free segregants in the evolved
629 populations. Therefore, in addition to the sequencing of DNA extracted from total pHC-
630 populations, DNA was sequenced from pooled colonies of plasmid-carrying cells. The deep
631 sequencing of plasmid-carrying hosts allows the detection of rare variants from a population
632 of cells with a potentially heterogeneous plasmid allele pool. DNA from total pHC populations

633 was isolated from archived samples as described for pLC-populations. The preparation of
634 DNA from plasmid-carrying cells included a preliminary stage of selection for plasmid-hosts.
635 Thus, frozen chemostat culture samples were first plated on LB plates supplemented with
636 ampicillin (100 µg ml^{-1}) and grown overnight at 37°C or 42°C, to select plasmid-carrying cells.
637 Then, 100 colonies per pHC-population were pooled for extraction of total DNA (i.e., 4x600
638 colonies in total). Sample libraries for Illumina sequencing were prepared using either the
639 Nextera or Nextera XT library preparation kit. All samples were quantified on a Qubit
640 fluorometer (Invitrogen by Life Technologies) and DNA fragment length distribution was
641 assessed on a TapeStation (Agilent Technologies). Libraries were sequenced on either a
642 HiSeq 2500 platform with 2x125 bp reads or a NextSeq 500 platform with 2x150 bp reads
643 (see Supplementary Table S1 for details). Due to low initial coverage, some libraries of the
644 wt-pLC and wt-pHC populations were re-sequenced on a HiSeq 2500 platform.

645 ***Variant detection***

646 Sequencing reads were trimmed to remove both Illumina specific adaptors and low quality
647 bases using the program Trimmomatic v.0.35 (Bolger et al. 2014) with these parameters:
648 `ILLUMINACLIP:NexteraPE-PE.fa:2:30:10 CROP:150/125` (NextSeq/HiSeq) `HEADCROP:5`
649 `LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:36`. Quality of sequencing reads
650 was inspected before and after trimming using FastQC v.0.11.5 (Andrews). To confirm
651 plasmid maps derived from partial Sanger-sequencing after plasmid construction, plasmid
652 genomes of the four ancestral populations were assembled from trimmed paired end
653 sequencing reads using plasmidSPAdes (SPAdes v.3.9.0 (Nurk et al. 2013)). To improve
654 assembly speed and accuracy of the high copy plasmid pHC samples, BBNorm (BBMap tool
655 suite v.35.82 (https://sourceforge.net/projects/bbmap/)) was used to normalize paired end
656 read data to an average coverage of 200x. The assemblies resulted in contigs with
657 overhangs at each end. There was no variation between plasmid assemblies of the same
658 type. Therefore, only one assembly of plasmid pLC and pHC was used as a reference for
659 variant detection. Plasmid genome assemblies were annotated using SnapGene software
660 v.2.4 (GLS Biotech). The chromosomal insertion sequence art3 (described above) was

661 included in the reference genome sequence of *E. coli* MG1655 (GenBank acc. no.
662 NC_000913.3). Reads were then mapped to this reference genome and to the assembled
663 plasmid genomes using BWA-MEM v.0.7.5a-r405 (Li and Durbin 2009). Corresponding
664 mapping files of libraries that were sequenced twice were merged using PICARD tools
665 v.2.7.1 (http://broadinstitute.github.io/picard/). Mapping statistics were retrieved using
666 BAMStats v.1.25 (https://sourceforge.net/projects/bamstats/files/). Then, indexing, local
667 realignment of sequencing reads, removal of read duplicates and ambiguous aligned reads
668 were performed using PICARD tools, SAMtools v.0.1.19 (Li et al. 2009) and GATK v.3.6
669 (McKenna et al. 2010) retaining only paired mapped reads with a minimum mapping quality
670 of 20. To enable a between-sample and between-replicon comparison of measures deduced
671 from the sequencing reads (e.g., number of mutations) that is not biased by differences in
672 sequencing coverage, subsampling of reads was performed. Therefore, mapping files were
673 first split by contig (i.e., plasmid and chromosome) using Bamtools v.2.4.2 (Barnett et al.
674 2011). Reads of each chromosomal and plasmid BAM file were then subsampled to the
675 minimal mean coverage of 104x using SAMtools (seed 1). Subsampled and non-subsampled
676 datasets were then used to call variants. Short indels and SNPs were called using LoFreq
677 v.2.1.2 (Wilm et al. 2012) and VarScan v.2.4.3 (Koboldt et al. 2009). Pindel version 0.2.5b9
678 (Ye et al. 2009) was used with options `-v 30 -M 2` to detect indels. The maximum number
679 of reads used to call variants with LoFreq and to call variants with VarScan by using
680 SAMtools mpileup was set to 15,000,000 for non-subsampled datasets. Variants that were
681 supported by alternative reads in one direction only were removed. Furthermore, only
682 variants that were concordantly identified by two or more of the variant calling tools were
683 retained using BEDTools' intersect utility v2.17.0 (Quinlan and Hall 2010) with parameters `-f`
684 `1.0 -r -u`. BEDTools' intersect utility was then used with parameters `-f 1.0 -r -v` to
685 keep only variants present exclusively in evolved but not in the ancestral samples. A custom
686 python script using pysam (https://github.com/pysam-developers/pysam) and R version 3.4.3
687 (R 2017) were then used to retain only variants that met the following criteria: allele
688 frequency ≥ 0.02, depth of coverage ≥ 10, ratio of the numbers of forward and reverse reads

689 supporting the variant ≥ 0.25, a minimum medium mapping quality of 40, a minimum medium
690 base quality of 30 and a minimum median length of 30 nt to the end of the reads supporting
691 the variant. Indels within an homopolymer stretch of ≥ 8 nucleotides and indels with less than
692 10 reads supporting the variant were excluded. Variants whose read support at the
693 corresponding position in ancestor samples was not sufficient to detect a variant (i.e., < 10x)
694 were excluded. Variant genomic location and the type of mutation, i.e., intergenic,
695 synonymous, non-synonymous, or nonsense, were annotated using an in-house Perl script.
696 Plasmid copy numbers were inferred prior to the de-duplication as the ratio of plasmid to
697 chromosomal mean coverage. The number of intergenic and synonymous sites from the
698 annotated reference genome sequences that were used for the mapping of sequencing
699 reads. Intergenic positions were extracted from the reference sequences and the total
700 number of intergenic sites was calculated as the sum of intergenic positions whose coverage
701 was ≥ 10x. Potential synonymous sites were determined for each position within the
702 reference sequences. The total number of synonymous sites were then calculated for each
703 mapping file using the Nei-Gojobori method (Nei and Gojobori 1986) considering only those
704 positions with a minimum sequencing coverage of 10x.

705 ***Simulation***

706 Forward-time simulations based non non-overlapping generations with a constant cell
707 population size were performed with python3.4 and the module simuPOP v1.1.8.3 (Peng and
708 Kimmel 2005). Two types of cells were simulated independently. Cells with chromosomes
709 have `L` haploid loci (`ploidy=1`). Cells with plasmids have a predefined plasmid copy number
710 `cp` that stays constant throughout the simulation. Each plasmid has `L` loci, resulting in a total
711 number of `L×cp` loci in each cell. This was implemented based on `cp` customized
712 chromosomes (`chromTypes=[sim.CUSTOMIZED]*cp`), each with `L` haploid loci. Alleles were
713 simulated as binary (`alleleType='binary'`). Inheritance mode for the chromosome
714 corresponds to a Wright-Fisher model (operator `RandomSelection()`). Inheritance mode for

715 the plasmid corresponds to a Wright Fisher model with mitochondrial inheritance of
716 organelles (operator `RandomSelection(ops=[sim.MitochondrialGenoTransmitter()])`).

717 For the simulations with mutation, initial populations (t=0) were homogenous, i.e., all
718 loci had state 0. The mutation rate in the simulation was constant with 2×10^{-9} per locus per
719 generation (operator `SNPmutator(u=2e-9)`). For the simulated evolutionary dynamics of
720 mutant alleles, we simulated the evolution of one binary locus (`L=1`) having a mutant allele
721 already present in the initial population at a specified initial frequency (*f*) and no new
722 mutation is applied during simulated evolution. Initial populations were randomly generated
723 with a particular mutant frequency `f` (parameter `freq=[1-f,f]` for the method
724 `initGenotype`). Note that this assigns the mutant allele randomly to each locus,
725 independent of the state of the other loci in the same cell for plasmid simulations. For the
726 simulated evolution with selection, the probability to be inherited to the next generation is
727 proportional to the fitness value of the cell. For cells with a chromosome, a fitness value of
728 1+s was assigned to allele state 1 and fitness value 1 was assigned to allele state 0. For
729 cells with plasmids, a fitness value of 1+s was assigned if at least one plasmid has allele
730 state 1, otherwise the fitness is 1 (i.e., dominant selection mode). Selection was applied
731 using a fitness field (`infoFields="fitness"`) and the operator `PySelector`. Simulations
732 were run for a particular number of generations using the `evolve` method.

733 Data availability

734 Sequence reads of evolved and ancestral populations are found in SRA (SRA acc.
735 SRP141152). Plasmid sequences are found in GenBank (pLC: MH238456, pHC: MH238457).

736 References

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

740 Andrews S. *FASTQC. A quality control tool for high throughput sequence data.* Available

741 from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

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

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

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

749 Baxter JC, Funnell BE. 2014. Plasmid Partition Mechanisms. Microbiol. Spectr. 2.

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

752 Becker L, Steglich M, Fuchs S, Werner G, Nübel U. 2016. Comparison of six commercial kits
753 to extract bacterial chromosome and plasmid DNA for MiSeq sequencing. Sci Rep
754 6:28063.

755 Bedhomme S, Perez Pantoja D, Bravo IG. 2017. Plasmid and clonal interference during post
756 horizontal gene transfer evolution.Giraud T, Koskella B, Laine A-L, editors. Mol. Ecol.
757 26:1832–1847.

758 Bevan M. 1984. Binary *Agrobacterium* vectors for plant transformation. Nuc. Acids Res.
759 12:8711–8721.

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

762 Bodmer WF, Cavalli-Sforza LL. 1976. Genetics, Evolution, and Man. W H Freeman &
763 Company

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

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

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

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

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

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

778 Figurski DH, Helinski DR. 1979. Replication of an origin-containing derivative of plasmid RK2
779 dependent on a plasmid function provided in *trans*. Proc. Natl. Acad. Sci. U.S.A.
780 76:1648–1652.

781 Gillespie JH. 2010. Population Genetics. JHU Press

782 Gullberg E, Albrecht LM, Karlsson C, Sandegren L, Andersson DI. 2014. Selection of a
783 multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. mBio
784 5:e01918–14.

785 Guyot S, Pottier L, Hartmann A, Ragon M, Hauck Tiburski J, Molin P, Ferret E, Gervais P.
786 2014. Extremely rapid acclimation of *Escherichia coli* to high temperature over a few
787 generations of a fed-batch culture during slow warming. Microbiologyopen 3:52–63.

788 Hanahan D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol.
789 166:557–580.

790 Harrison E, Guymer D, Spiers AJ, Paterson S, Brockhurst MA. 2015. Parallel compensatory
791 evolution stabilizes plasmids across the parasitism-mutualism continuum. Curr. Biol.
792 25:2034–2039.

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

796 Hershfield V, Boyer HW, Yanofsky C, Lovett MA, Helinski DR. 1974. Plasmid ColE1 as a
797 Molecular Vehicle for Cloning and Amplification of DNA. Proc. Natl. Acad. Sci. U.S.A.
798 71:3455–3459.

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

802 Jechalke S, Heuer H, Siemens J, Amelung W, Smalla K. 2014. Fate and effects of
803 veterinaryantibiotics in soil. Trends Microbiol. 22:536–545.

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

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

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

812 Lederberg J, Tatum EL. 1946. Gene recombination in Escherichia coli. 158:558.

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

815 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
816 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map
817 format and SAMtools. Bioinformatics 25:2078–2079.

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

821 Majeed H, Gillor O, Kerr B, Riley MA. 2011. Competitive interactions in *Escherichia coli*

822 populations: the role of bacteriocins. ISME J. 5:71–81.

823 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,
824 Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce
825 framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–
826 1303.

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

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

832 Münch KM, Müller J, Wienecke S, Bergmann S, Heyber S, Biedendieck R, Münch R, Jahn D.
833 2015. Polar Fixation of Plasmids during Recombinant Protein Production in *Bacillus*
834 *megaterium* Results in Population Heterogeneity.Pettinari MJ, editor. Applied Environ.
835 Microbiol. 81:5976–5986.

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

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

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

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

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

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

850 Porse A, Schønning K, Munck C, Sommer MOA. 2016. Survival and Evolution of a Large
851 Multidrug Resistance Plasmid in New Clinical Bacterial Hosts. Mol. Biol. Evol. 33:2860–
852 2873.

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

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

858 R DCT. 2017. R: A language and environment for statistical computing. R Foundation for
859 Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/

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

862 Rodriguez-Beltran J, Hernandez-Beltran JCR, DelaFuente J, Escudero JA, Fuentes-

863 Hernandez A, MacLean RC, Peña-Miller R, San Millan A. 2018. Multicopy plasmids allow
864 bacteria to escape from fitness trade-offs during evolutionary innovation. Nat. Ecol. Evol.
865 2:873–881.

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

868 Sambrook J, Russell DW. 2001. Molecular Cloning. 3rd ed. CSHL Press

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

871 San Millan A, Peña-Miller R, Toll-Riera M, Halbert ZV, McLean AR, Cooper BS, MacLean RC.
872 2014. Positive selection and compensatory adaptation interact to stabilize non-
873 transmissible plasmids. Nat. Commun. 5:5208.

874 San Millan A, Toll-Riera M, Qi Q, Betts A, Hopkinson RJ, McCullagh J, MacLean RC. 2018.
875 Integrative analysis of fitness and metabolic effects of plasmids in *Pseudomonas*
876 *aeruginosa* PAO1. ISME J. 3:679.

877 Sandegren L, Linkevicius M, Lytsy B, Melhus Å, Andersson DI. 2011. Transfer of an
878 *Escherichia coli* ST131 multiresistance cassette has created a *Klebsiella pneumoniae*-
879 specific plasmid associated with a major nosocomial outbreak. J. Antimicrob. Chemother.
880 67:74–83.

881 Santos-Lopez A, Bernabe-Balas C, Ares-Arroyo M, Ortega-Huedo R, Hoefer A, San Millan A,
882 Gonzalez-Zorn B. 2017. A Naturally Occurring Single Nucleotide Polymorphism in a
883 Multicopy Plasmid Produces a Reversible Increase in Antibiotic Resistance. Antimicrob.
884 Agents Chemother. 61.

885 Satoh M, Kuroiwa T. 1991. Organization of multiple nucleoids and DNA molecules in
886 mitochondria of a human cell. Exp. Cell Res. 196:137–140.

887 Schmidt R, Ahmetagic A, Philip DS, Pemberton JM. 2011. Catabolic Plasmids. In: eLS. eLS.

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

890 Simon R, Priefer U, Puhler A. 1983. A Broad Host Range Mobilization System for *In Vivo*
891 Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. Nat. Biotech.
892 1:784–791.

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

896 Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson
897 LW, Kasarskis A, Batty EM, et al. 2016. Evolutionary History of the Global Emergence of
898 the *Escherichia coli* Epidemic Clone ST131. mBio 7:e02162.

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

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

904 Ullrich A, Shine J, Chirgwin J, Pictet R, Tischer E, Rutter WJ, Goodman HM. 1977. Rat
905 insulin genes: construction of plasmids containing the coding sequences. 196:1313–
906 1319.

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

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

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

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

916 Wobbrock JO, Findlater L, Gergle D, Higgins JJ. 2011. The aligned rank transform for
917 nonparametric factorial analyses using only anova procedures. New York, New York,
918 USA: ACM

919 Yang YL, Polisky B. 1993. Suppression of ColE1 high-copy-number mutants by mutations in
920 the *polA* gene of *Escherichia coli*. J. Bacteriol. 175:428–437.

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

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

927

928

929 **Acknowledgments**

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

935 **Author contributions**

936 J.I, I.M., and T.D. designed the experiment. J.I. established and performed the experimental
937 evolution study and the comparative genomic analysis. A.K. performed the simulated
938 evolution. P.R. sequenced the genomes. T.W. and N.H. took part in the experimental work.
939 C.W., G.L, A.K., and T.D. took part in the computational analysis. All authors interpreted the
940 results and drafted the manuscript.

941 **Competing interests**

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

943

944 **Tables**

945 Table 1. Number of intergenic and synonymous point mutations detected in the subsampled
946 read data (104x). A detailed information used for the calculation of the mean and standard
947 deviation for each of the eight experimental factor groups is shown in Supplementary Table
948 S1. Additional low frequency point mutations were detected on the pLC plasmid considering
949 the full sequencing coverage (see variant details in Supplementary Table S2). A detailed list
950 of all observed variants on the chromosome is given in Supplementary Table S3. The variant
951 distribution across the replicons is depicted in Supplementary Fig. S1. We note that plasmid
952 copy number estimate from the relative sequencing coverage of small plasmids may differ
953 from the estimated PCN using qPCR due to the DNA extraction method as well as the
954 sequencing approach (Becker et al. 2016). The large-scale estimates of PCN from
955 sequencing results are lower than those obtained with qPCR, hence our estimate of the
956 plasmid mutational supply is conservative. The ancestral PCN estimated from the
957 sequencing results are as following: pLC-wt: 2, pLC-hypermutator: 13, pHC-wt: 86, pHC-
958 hypermutator: 116. A comparison of the PCN distribution in the evolved populations reveals
959 that the PCN is significantly different from that of the ancestor plasmid in several of the
960 experimental evolution settings. However, the direction of change (i.e., PCN increase or
961 decrease) is heterogeneous (see supplementary table S1 for details).

			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	20.9 ± 4.9	$16.3 \times 10^{-2} \pm 5.0 \times 10^{-2}$	0
		42	16	2.4 ± 1.8	4.8 ± 3.2	$4.5 \times 10^{-2} \pm 4.8 \times 10^{-2}$	0
	Hypermutator	37	148	22.4 ± 5.2	7.1 ± 2.2	$5.0 \times 10^{-1} \pm 1.7 \times 10^{-1}$	1
		42	87	13.2 ± 3.3	4.1 ± 1.4	$1.8 \times 10^{-1} \pm 1.1 \times 10^{-1}$	1
pHC	Wild-type	37	6	0.87 ± 0.6	605.0 ± 501.3	2.0 ± 2.0	0
		42	6	0.87 ± 0.8	294.0 ± 105.8	$7.9 \times 10^{-1} \pm 7.2 \times 10^{-1}$	0
	Hypermutator	37	98	14.3 ± 3.8	154.5 ± 63.4	8.0 ± 5.0	0
		42	147	21.4 ± 11.5	252.6 ± 106.3	19.1 ± 16.2	0

962

963 [(1)] The chromosomal substitution rate is calculated as $R_{Chr} = N_{obs} \times (N_{sites} \times N_{gen})^{-1}$, where N_{obs}
964 is the number of observed point mutations, N_{sites} is the number of intergenic and synonymous

965 sites on the chromosome having a coverage ≥10 and N_{gen} is the number of generations in
966 our experiment.

967 (2) Plasmid copy number (PCN) was estimated from the relative coverage of plasmid and
968 chromosome replicons (detailed list in Supplementary Table S1).

969 (3) The expected number of point mutations on the plasmid is calculated as $N_{exp} = R_{Chr} \times N_{sites}$
970 $\times$ PCN $\times$ N_{gen}, where N_{sites} is the number of intergenic and synonymous sites on the plasmid
971 having a coverage ≥10.

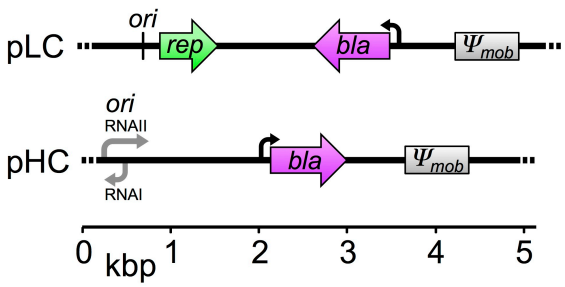
972

973

974

975 **Figures**

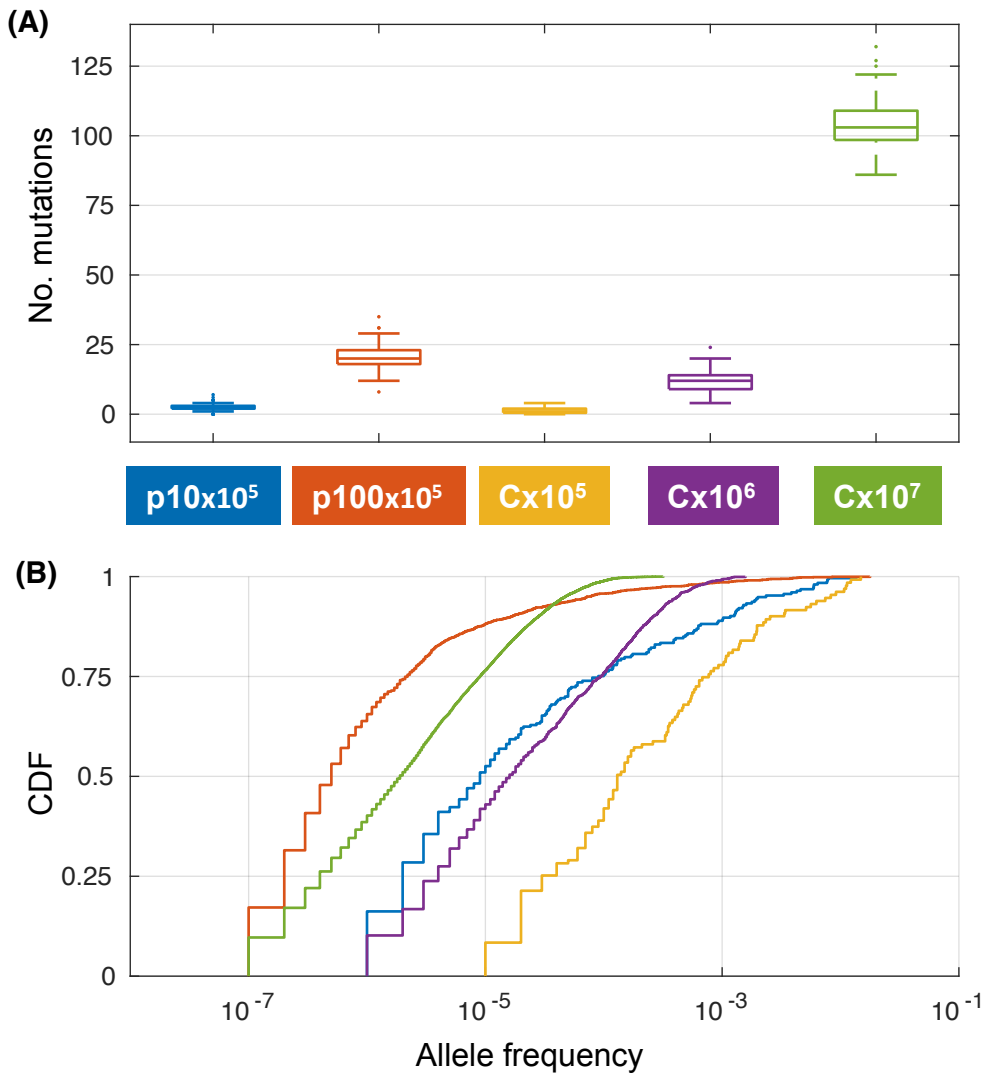
976 **Figure 1.** Genetic maps of plasmid pLC and pHC. The plasmids are of comparable genome
977 size and differ in the origin of replication. The *mob* gene originally found in pBBR1 was non-
978 functionalized by truncation during plasmid construction (denoted as Ψ_{mob}).



979

980

981 **Figure 2**. Simulated evolution of plasmids and chromosomes under neutrality with mutation
982 (500 loci, mutation rate 2 × 10^{-9} per locus per generation). (**A**) Distribution of the number of
983 mutations after 1,000 generations in 100 replicates. The color of the boxes corresponds to
984 the combination of replicon and population size as in the legend. (**B**) Joint cumulative
985 distribution function (CDF) of the allele frequency of all mutations present in the replicates
986 after 1,000 generations.



987

988

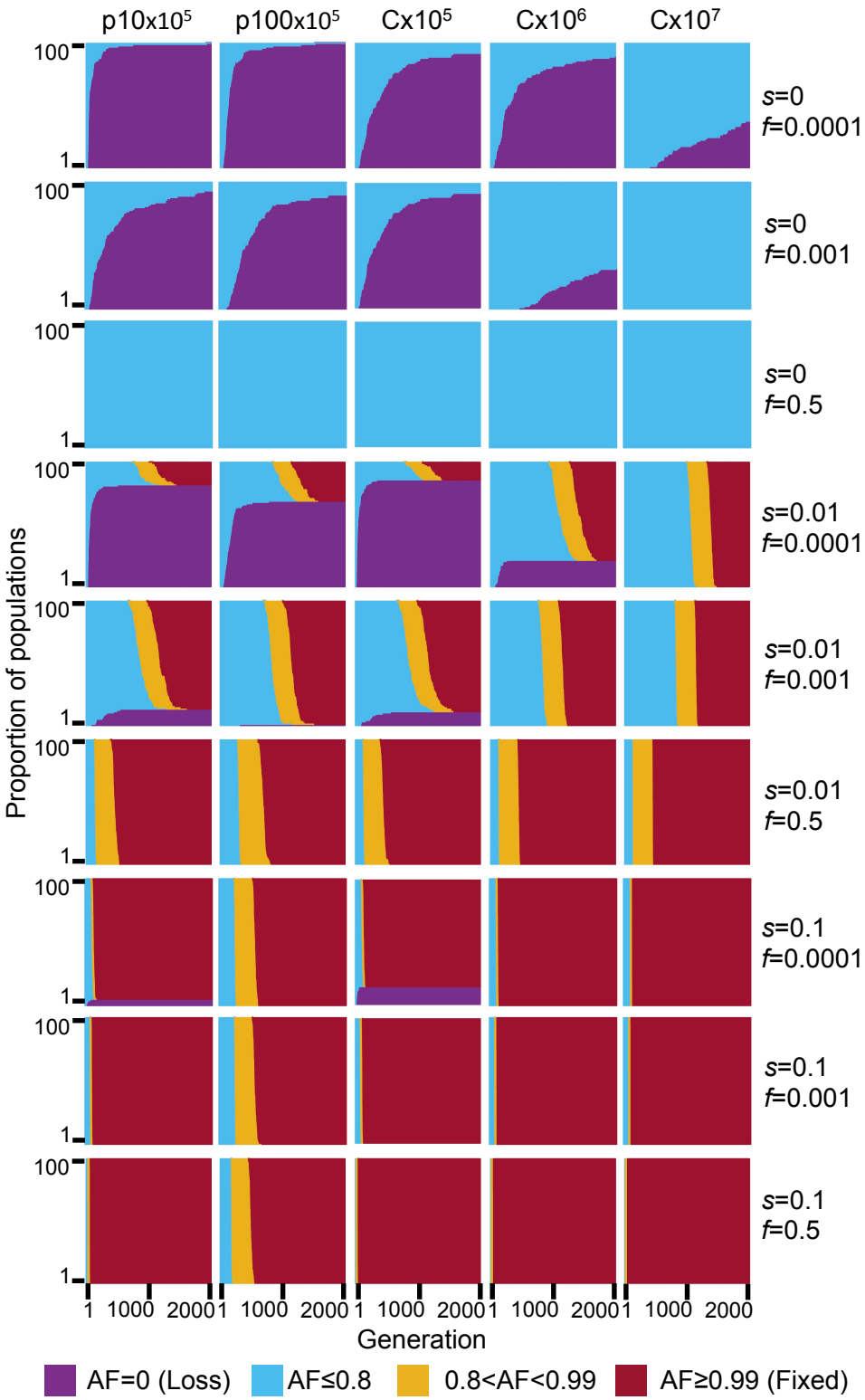
989

990 **Figure 3**. Simulated evolutionary dynamics of mutant alleles on plasmids and chromosomes.
991 Allele frequency dynamics over 2,000 generations in 100 replicate simulated populations. (**A**)
992 Columns are the five combinations of simulated replicon type and population size, and rows
993 are the different combinations of selective coefficient (*s*) and initial allele frequency (*f*).
994 Generations are on the x-axis, proportion of AF of replicates are stacked along the y-axis,
995 population allele frequency is colour coded. For f=0.001 and f=0.0001, the starting frequency
996 of mutant alleles is randomly sampled with the respective frequencies. For f=0.5, the starting
997 frequency of the mutant allele is exactly 0.5 for the chromosome and exactly 0.5 inside each
998 cell for the plasmids. Note that, the initial frequency of cells with a mutant allele in the
999 population is higher in the plasmid simulations compared to chromosome simulations (see
1000 Supplementary Table S4 for expected number of mutant cells).

1001

1002

1003



1004

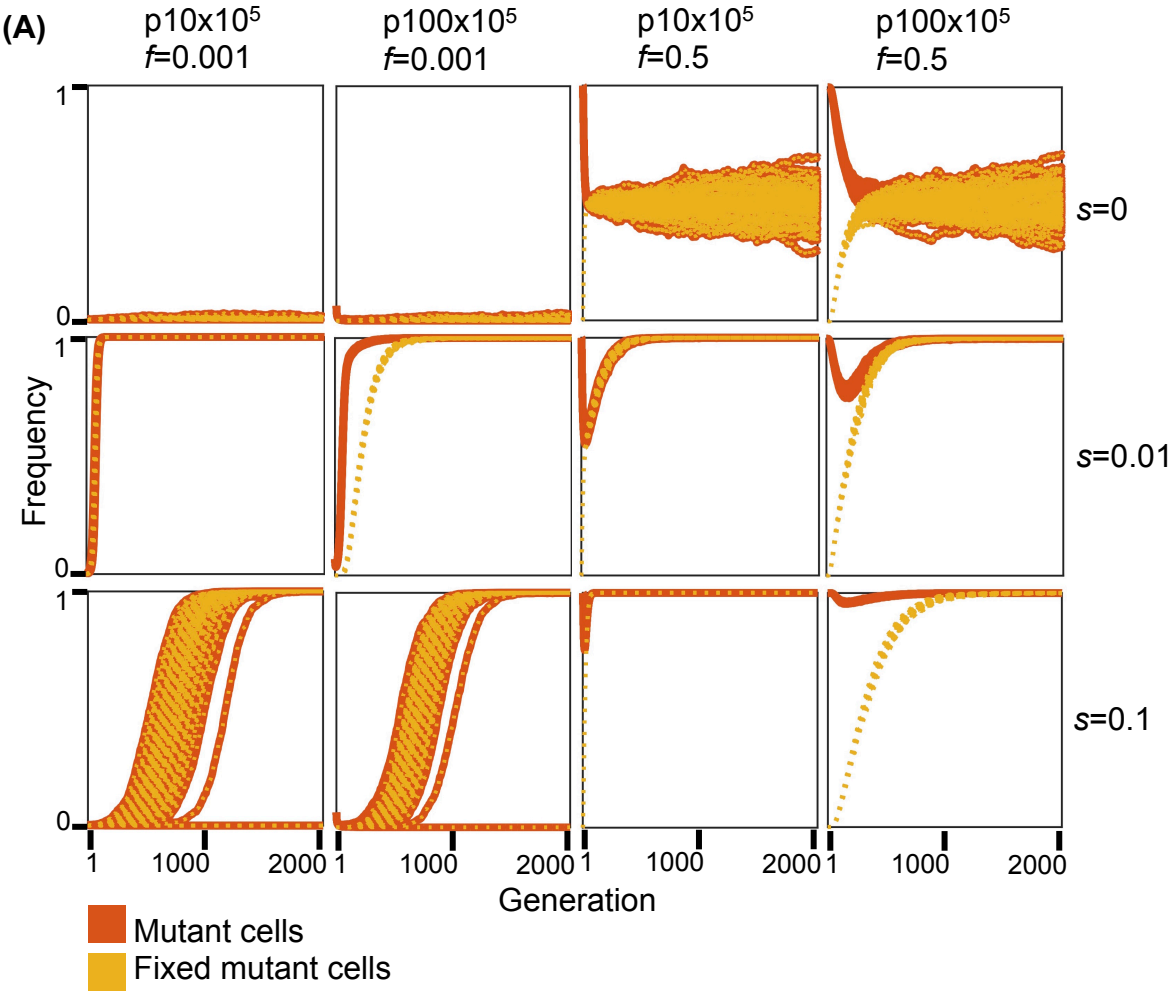
1005

1006

1007

1008 **Figure 4**. Frequency trajectories of plasmid alleles per cell in the plasmid simulations. (A)
1009 Data shown for with initial allele frequencies f = 0.001, and f = 0.5. The frequency of mutant
1010 cells is calculated as the proportion of cells with at least one mutant plasmid allele within the
1011 population. The frequency of fixed mutant cells is the proportion of cells in which all plasmids
1012 loci contain the mutant allele (i.e., the plasmid allele is fixed in the cell). Starting with a low
1013 initial allele frequency (f = 0.001), the difference between mutant plasmid allele dynamics
1014 and mutant cells in the population is most pronounced with s = 0.01. The results of the
1015 simulation where s = 0.1 are presented for comparison; in this setting, the dynamics of
1016 mutant plasmid alleles and mutant cells are similar. In the simulation of plasmid allele
1017 dynamics post balancing selection (f = 0.5) the ancestral and mutant alleles are set to an
1018 equal ratio within all cells At the beginning of the simulation. (B) Ancestral allele loss is
1019 calculated as the median number of generation where the ancestral allele was lost across all
1020 simulated populations.

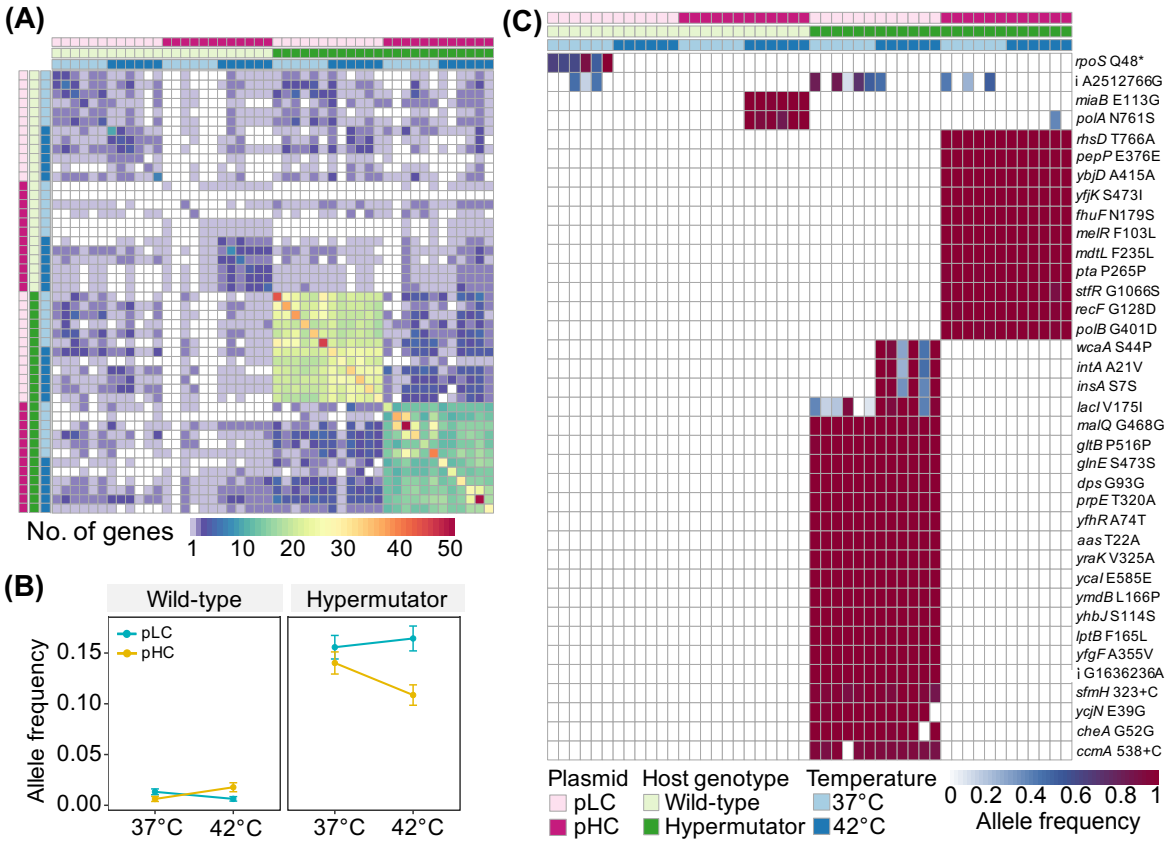
(A)



(B) Loss of the ancestral allele (median no. generations; *f*=0.5)

s	**p10x10^5**	**p100x10^5**	**Cx10^5**	**Cx10^6**	**Cx10^7**
0	–	–	–	–	–
0.01	811	1086	775	1019	1285
0.1	131	876	108	129	157

1023 **Figure 5**. Parallel evolution of the chromosome. (**A**) A heatmap representing the number of
1024 shared mutated genes among the 48 evolved populations. The number of genes mutated in
1025 population pair *x*, *y* is calculated as the number of genes in which genetic variants (single
1026 base substitutions or indels) are observed in both populations. Cells along the diagonal
1027 present the number of genes where a mutation was detected in the corresponding population
1028 and at least one additional population. Annotation bars at the top and left right present the
1029 experimental factors (color-coded as shown in (C)). (**B**) An interaction plot of the main factors
1030 in the experiment as calculated from the complete set of allele frequencies of parallel
1031 variants (included in Supplementary Table S3). Dots represent means and error bars
1032 represent the standard error of the mean. ANOVA demonstrated significant effects for
1033 plasmid replicon type and temperature along with significant interactions between the three
1034 factors used in the experiment (statistics in Supplementary Table S5). (**C**) Color-coded matrix
1035 of variants whose allele frequency is significantly different among the main factors or their
1036 combination (using ANOVA on aligned rank transformed allele frequencies and FDR;
1037 statistics in Supplementary Table S6). Intragenic variants (11 synonymous and 22 non-
1038 synonymous) are indicated by the gene symbol and, in case of a single base substitution,
1039 coded as [ancestral amino acid][amino acid position][alternative amino acid]. The two
1040 insertion mutations are indicated by a gene symbol and coded as [nucleotide position in
1041 gene][+ inserted nucleotide]. [*] indicates a non-sense codon. The two intergenic variants are
1042 indicated by [i] coded as [ancestral nucleotide][genomic position][alternative nucleotide].



1043

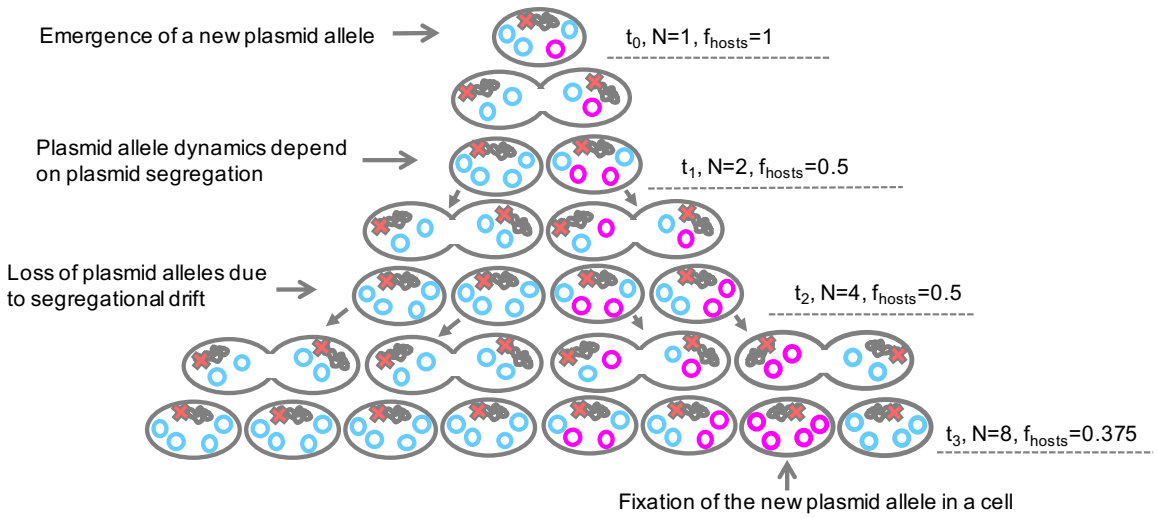
1044

1045

1046 **Figure 6**. An illustration of the consequences of segregational drift to plasmid allele
1047 dynamics. The illustrated plasmid copy number is four and its segregation into daughter cells
1048 is balanced (i.e., the daughter cells inherit an equal number of plasmids). In the depicted
1049 scenario, one allele emerges on the chromosome (red X) and another allele emerges on the
1050 plasmid (magenta) at t_0. Both alleles are considered neutral (i.e., s = 0). The population size
1051 (N) and the frequency of hosts (f_{hosts}) over three generations are presented. The
1052 chromosomal allele is inherited into all daughter cells, such that it is present in the total
1053 population. Random segregation of the plasmid genotypes leads to decrease in the number
1054 of hosts. At t_3, only a small minority in the population harbours the new plasmid allele. In the
1055 absence of selection for the plasmid allele presence, the new plasmid allele is deemed to
1056 remain at a very low frequency within the population (or be lost). Note: this is a simplistic
1057 example that does not include loss of host cells due to drift and it does not include
1058 consideration of recombination among plasmids.

1059

1060



1061