1 **The distribution of plasmid fitness effects explains plasmid persistence in bacterial**

2 **communities**

3 Aida Alonso-del Valle[1], Ricardo León-Sampedro[1,2], Jerónimo Rodríguez-Beltrán[1,2], Javier

4 DelaFuente[1], Marta Hernández-García[1,3], Patricia Ruiz-Garbajosa[1,3], Rafael Cantón[1,3], Rafael

5 Peña-Miller[4,*], Álvaro San Millán[1,2,5]*.

6

7 [1] *Servicio de Microbiología. Hospital Universitario Ramón y Cajal* and *Instituto Ramón y Cajal de*

8 *Investigación Sanitaria*. Madrid, Spain.

9 [2] *Centro de Investigación Biológica en Red. Epidemiología y Salud Pública*, *Instituto de Salud Carlos*

10 *III*. Madrid. Spain.

11 [3] *Red Española de Investigación en Patología Infecciosa. Instituto de Salud Carlos III*. Madrid. Spain.

12 [4] Center for Genomic Sciences, *Universidad Nacional Autónoma de México*, Cuernavaca, Mexico.

13 [5] Current address: *Centro Nacional de Biotecnología–CSIC*, Madrid, Spain.

14 * Correspondence: Álvaro San Millán, alvsanmillan@gmail.com ORCID: 0000-0001-8544-0387 and

15 Rafael Peña-Miller, rafael.penamiller@gmail.com ORCID: 0000-0002-2767-0640

16

17 **Introductory paragraph**

18 Plasmid persistence in bacterial populations is strongly influenced by the fitness effects associated with

19 plasmid carriage. However, plasmid fitness effects in wild-type bacterial hosts remain largely

20 unexplored. In this study, we determined the distribution of fitness effects (DFE) for the major antibiotic

21 resistance plasmid pOXA-48 in wild-type, ecologically compatible enterobacterial isolates from the

22 human gut microbiota. Our results show that although pOXA-48 produced an overall reduction in

23 bacterial fitness, the DFE was dominated by quasi-neutral effects, and beneficial effects were observed

24 in several isolates. Incorporating these data into a simple population dynamics model revealed a new

25 set of conditions for plasmid stability in bacterial communities, with plasmid persistence increasing with

26 bacterial diversity and becoming less dependent on conjugation. Moreover, genomic results showed a

27 link between plasmid fitness effects and bacterial phylogeny, helping to explain pOXA-48 epidemiology.

28 Our results provide a simple and general explanation for plasmid persistence in natural bacterial

29 communities.

Introduction

Plasmids are extra-chromosomal mobile genetic elements able to transfer between bacteria through conjugation[1]. Plasmids carry accessory genes that help their hosts to adapt to a myriad of environments and thus play a key role in bacterial ecology and evolution[2]. A key example of the importance of plasmids in bacterial evolution is their central role in the spread of antibiotic resistance mechanisms among clinical pathogens over recent decades[3,4]. Some of the most clinically relevant resistance genes, such those encoding carbapenemases (ß-lactamase enzymes able to degrade carbapenem antibiotics), are carried on conjugative plasmids that spread across high-risk bacterial clones[5,6].

Despite the abundance of plasmids in bacterial populations and the potential advantages associated with their acquisition, these genetic elements generally produce physiological alterations in their bacterial hosts that lead to a reduction in fitness[7–9]. These fitness costs make it difficult to explain how plasmids are maintained in bacterial populations over the long-term in the absence of selection for plasmid-encoded traits, a puzzle known as "the plasmid-paradox"[10]. Different solutions to this paradox have been proposed. For example, compensatory evolution contributes to plasmid persistence by alleviating the costs associated with plasmid-carriage, and a high conjugation rate can promote the survival of plasmids as genetic parasites[11–18].

Over the past decades, many studies have investigated the existence conditions for plasmids in bacterial populations[14,18–23]. However, understanding of plasmid population biology is held in check by limitations of the model systems used for its study. First, most experimental reports of fitness costs have studied arbitrary associations between plasmids and laboratory bacterial strains[7,24]. These examples do not necessarily replicate plasmid fitness effects in natural bacterial hosts, which remain largely unexplored. Second, studies tend to analyse the fitness effects of a single plasmid in a single bacterium. However, plasmid fitness effects can differ between bacteria[25–28], and this variability may impact plasmid persistence in bacterial communities (for a relevant example see[29]). Third, most mathematical models of plasmid population biology study clonal or near-clonal populations. However, bacteria usually live in complex communities in which conjugative plasmids can spread between different bacterial hosts[30–32]. To fully understand plasmid persistence in natural bacterial populations, it will be necessary to address these limitations.

2

59 In this study, we provide the first description of the distribution of fitness effects (DFE) of a plasmid in

60 wild-type bacterial hosts. We used the clinically relevant carbapenem-resistance conjugative plasmid

61 pOXA-48 and 50 enterobacteria strains isolated from the gut microbiota of patients admitted to a large

62 tertiary hospital in Madrid. Incorporation of the experimentally determined DFE into a population biology

63 model provides new key insights into the existence conditions of plasmids in bacterial communities.

64 **Results**

65 *Construction of a pOXA-48 transconjugant collection*

66 We studied the DFE of the plasmid pOXA-48 in a collection of ecologically compatible bacterial hosts.

67 pOXA-48 is an enterobacterial, broad-host-range, conjugative plasmid that is mainly associated with *K.*

68 *pneumoniae* and *Escherichia coli*[33–35]. pOXA-48 encodes the carbapenemase OXA-48 and is

69 distributed worldwide, making it one of the most clinically important carbapenemase-producing

70 plasmids[6,34]. The gut microbiota of hospitalised patients is a frequent source of enterobacteria clones

71 carrying pOXA-48[6]. In recent studies, we described the in-hospital epidemiology of pOXA-48 in a large

72 collection of extended-spectrum ß-lactamase (ESBL)- and carbapenemase-producing enterobacteria

73 isolated from more than 9,000 patients in our hospital over a period of two years (R-GNOSIS collection,

74 see methods)[31,36–38]. pOXA-48-carrying enterobacteria were the most frequent carbapenemase-

75 producing enterobacteria in the hospital, with 171 positive isolates, and they colonised 1.13% of the

76 patients during the study period (105/9,275 patients). In this study we focused on plasmid pOXA-

77 48_K8, which is a recently described pOXA-48-like plasmid isolated from a *K. pneumoniae* in our

78 hospital[31] (Figure 1a, for simplicity we will refer to pOXA-48_K8 and pOXA-48-like plasmids as pOXA-

79 48 throughout the text).

80 To study the DFE of pOXA-48, we selected 50 isolates from the R-GNOSIS collection as bacterial

81 hosts. Our criteria were to select (i) pOXA-48-free isolates, to avoid selecting clones in which

82 compensatory evolution had already reduced plasmid-associated costs; (ii) isolates from the most

83 frequent pOXA-48-carrying species, *K. pneumoniae* and *Escherichia coli*; and (iii) strains isolated from

84 patients located in wards in which pOXA-48-carrying enterobacteria were commonly reported[31]. The

85 underlying rationale was to select clones which were naïve to pOXA-48 but ecologically compatible with

86 it (*i.e.* isolated from patients coinciding on wards with others who were colonised with pOXA-48-carrying

87 clones). We selected 25 *K. pneumoniae* and 25 *E. coli* isolates that are representative of the R-

3

88   GNOSIS study and cover the *K. pneumoniae* and *E. coli* phylogenetic diversity in the collection (see

89   methods, Figure 1b and Supplementary Table 1). It is important to note that, because of the nature of

90   the R-GNOSIS collection, the isolates used in this study produce ESBLs. However, ESBL-producing

91   enterobacteria are widespread not only in hospitals but also in the community[39], and most pOXA-48-

92   carrying enterobacteria isolated in our hospital also produce ESBLs[36].

93   pOXA-48 was introduced into the collection of recipient strains by conjugation (see Methods), and the

94   presence of the plasmid was confirmed by PCR and antibiotic susceptibility testing (Supplementary

95   Table 2). The presence of the entire pOXA-48 plasmid was confirmed by sequencing the complete

96   genomes of the 50 transconjugant clones, which also revealed the genetic relatedness of the isolates

97   (Figure 1b). In line with previous studies[31,40], the sequencing results revealed that a subset of isolates

98   initially identified as *K. pneumoniae* in fact belonged to the species *Klebsiella quasipneumoniae* (n= 4)

99   and *Klebsiella variicola* (n= 1). These species are also pOXA-48 hosts in our hospital[31] and so were

100   maintained in the study (Figure 1b).
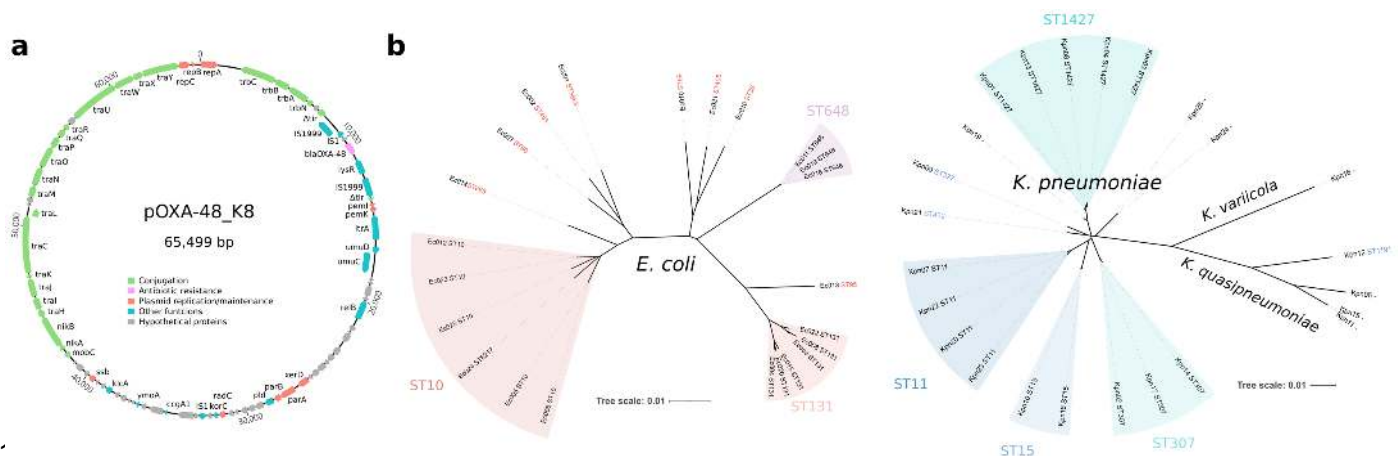
Figure 1. Experimental model system. Representation of pOXA-48 plasmid and the enterobacteria strains used in this study. (a) pOXA-48_K8 (accession number MT441554). Reading frames are shown as arrows, indicating the direction of transcription. Colours indicate gene function classification (see legend). The bla_{OXA-48} gene is shown in pink. (b) Unrooted phylogeny of whole-genome assemblies from *E. coli* clones (left) and *Klebsiella* spp. clones (right). Branch length gives the inter-assembly mash distance (a measure of k-mer similarity). The grouping of multi-locus sequence types (ST) is also indicated (*E. coli* ST6217 belongs to the ST10 group). Note that the sequencing results revealed that a

109   subset of isolates initially identified as *K. pneumoniae* were in fact *Klebsiella quasipneumoniae* (n= 4)

110   and *Klebsiella variicola* (n= 1).

---

111   *Measuring pOXA-48 fitness effects*

112   To measure pOXA-48 fitness effects, we performed growth curves and competition assays for all the

113   plasmid-carrying and plasmid-free clones in the collection. We first performed growth curves in pure

114   cultures to calculate maximum growth rate (μ_{max}) and maximum optical density (OD_{max}), which can be

115   used to estimate the intrinsic population growth rate (r) and carrying capacity (K), respectively

116   (Supplementary Figure 1). We also measured the area under the growth curve (AUC), which integrates

117   information about r and K. To estimate plasmid-associated fitness effects, we compared these

118   parameters between each plasmid-carrying and plasmid-free pair of isogenic isolates (Figure 2a). The

119   results showed that, as expected, pOXA-48 produced an overall decrease in the parameters extracted

120   from the growth curves. However, in a substantial subset of clones, plasmid acquisition was not

121   associated with a reduction in these parameters (Figure 2a).

122   Competition assays allow measurement of the relative fitness (w) of two bacteria competing for

123   resources in the same culture[41]. Competition between otherwise isogenic plasmid-carrying and

124   plasmid-free clones thus provides a quantitative assessment of the fitness costs associated with

125   plasmid carriage. For the competition assays, we used flow cytometry; strains were labelled using an

126   in-house developed small, non-conjugative plasmid vector, called pBGC, that encodes an inducible

127   green fluorescent protein (GFP) (Supplementary Figure 2). pBGC was introduced into the wild-type

128   isolate collection by electroporation, and all pOXA-48-carrying and pOXA-48-free clones were

129   competed against their pBGC-carrying parental strain. We were unable to introduce pBGC into eight of

130   the isolates; in those cases, for the competitor, we used *E. coli* strain J53 carrying the pBGC vector

131   (see Methods for details). Data from the competition assays were used to calculate the competitive

132   fitness of pOXA-48-carrying clones relative to their plasmid-free counterparts (Figure 2b). There were

133   no significant differences between the fitness effects of pOXA-48 in *Klebsiella* spp. and *E. coli* isolates

134   (ANOVA effect of Species x Plasmid interaction; F=0.088, df=1, P=0.767).

135   To validate our results, we compared the values obtained from growth curves and competition assays.

136   This analysis revealed a significant correlation between relative fitness values and the parameters

137   extracted from the growth curves (Supplementary Figure 3).

5
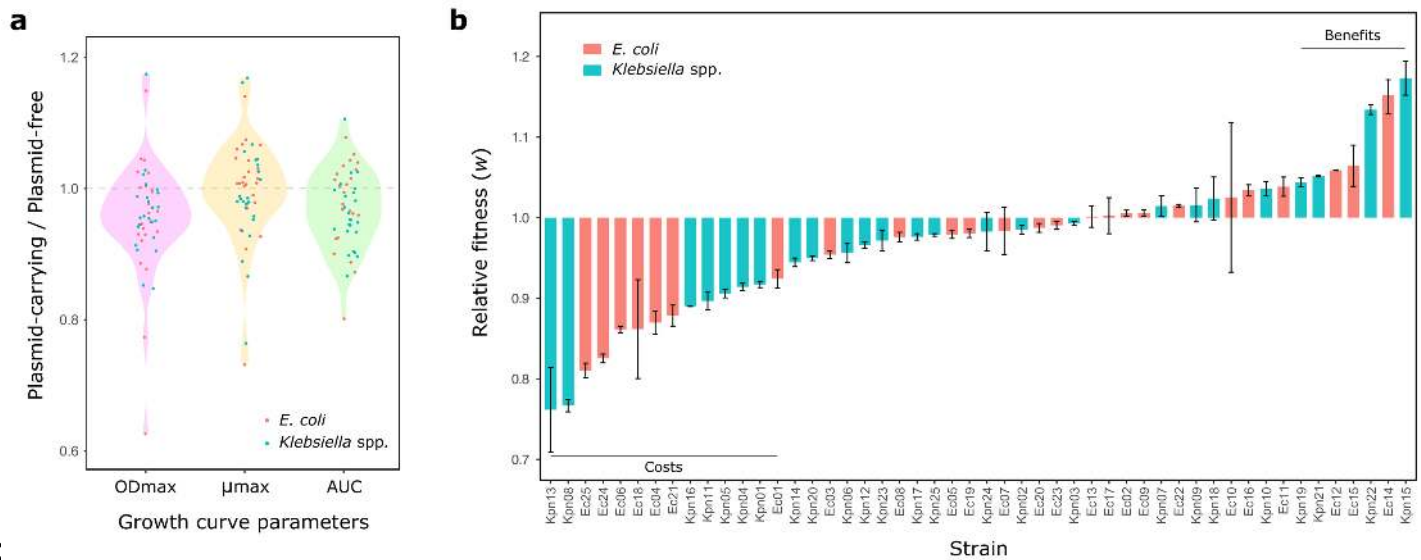
Figure 2. pOXA-48 fitness effects in a set of ecologically compatible wild-type enterobacteria. (a) Relative values of growth-curve parameters (plasmid-carrying/plasmid-free isogenic clones): maximum optical density (OD_{max}, pink), maximum growth rate (μ_{max}, yellow), and area under the curve (AUC, green). Dots represent each relative value (red, *E. coli*; blue, *Klebsiella* spp.). Values below 1 indicate a reduction in these parameters associated with plasmid-acquisition. Five biological replicates were performed for each growth curve. (b) Relative fitness (*w*) of plasmid-carrying clones compared with plasmid-free clones obtained by competition assays (red, *E. coli*; blue, *Klebsiella* spp.). Values below 1 indicate a reduction in *w* due to plasmid acquisition; values above 1 indicate an increase in *w*. Bars represent the mean of five independent experiments, and error bars represent the standard error of the mean. Two horizontal lines indicate those clones showing significant costs or benefits associated with carrying pOXA-48 plasmid.

The distribution of pOXA-48 fitness effects

Results from the competition assays showed that the overall effect of pOXA-48 was a small but significant reduction in relative fitness (mean *w*= 0.971, ANOVA effect of plasmid; F=70.04, df=1, P=1.02×10^{-15}). However, plasmid fitness effects varied greatly between the isolates in the collection, producing a normal distribution ranging from a >20% reduction to almost a 20% increase in relative fitness (Figure 2b and 3a; Shapiro-Wilk normality test, P= 0.14). Indeed, plasmid acquisition was associated with a significant fitness decrease in only 14 strains, and 7 isolates showed a significant increase in fitness (Bonferroni corrected two sample t-test, *P*< 0.05). These results revealed a highly

6

158   dynamic scenario in which a plasmid produces a wide distribution of fitness effects in different bacterial

159   hosts, ranging from costs to benefits.

160   To place our results in context with previous reports, we compared the DFE for pOXA-48 with the

161   results from a recent meta-analysis of plasmid fitness effects by Vogwill and MacLean[24] (Figure 3).

162   These authors recovered data for 50 plasmid-bacterium pairs from 16 studies. The DFE constructed

163   from those reports showed a higher mean plasmid cost (mean w= 0.91) and differed significantly from

164   the DFE we report here for pOXA-48 in wild-type enterobacteria (Wilcoxon signed rank test, V= 922, P=

165   0.006). The discrepancy between these distributions may, at least in part, reflect the different nature of

166   plasmid-bacterium associations considered in the different studies. Although the plasmids studied in

167   earlier reports were isolated from natural sources, they were introduced into laboratory bacterial strains,

168   and the detected fitness effects may not be fully representative of wild-type plasmid-bacterium

169   associations. Our study, on the other hand, analysed the fitness effects of pOXA-48 in ecologically

170   compatible bacterial hosts. Taken together, the data suggest that the distribution of plasmid fitness

171   effects is likely influenced by the ecological compatibility between plasmids and their bacterial hosts.

172
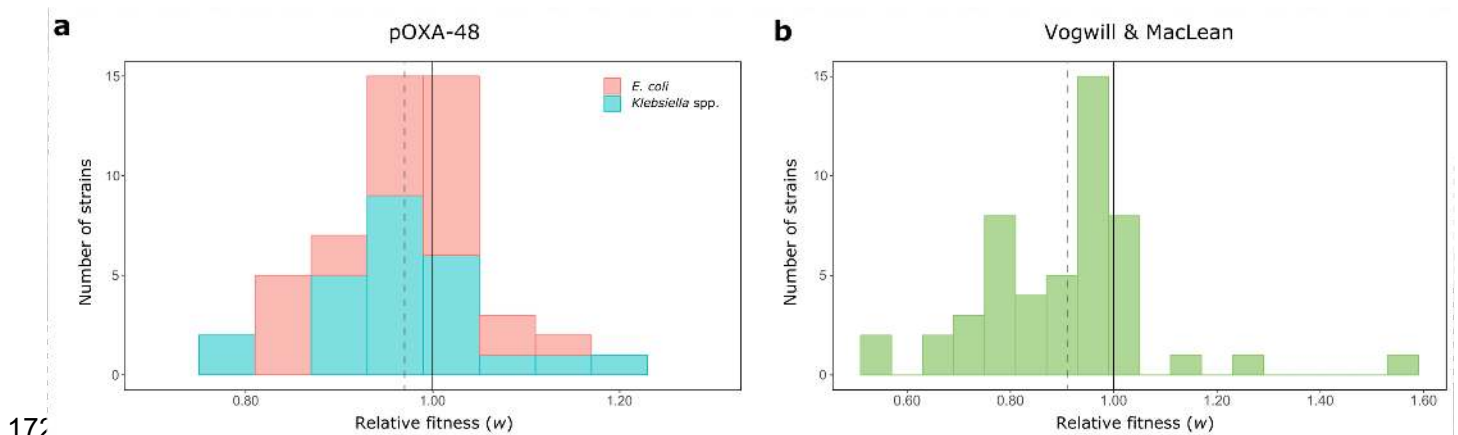
173   Figure 3. Distribution of plasmid fitness effects. Comparison between the DFE obtained in this study

174   and the DFE from previous studies. (a) DFE for pOXA-48 in the ecologically compatible collection of

175   enterobacteria isolates. Bars indicate the number of *E. coli* (red) and *Klebsiella* spp. (blue) strains in

176   each relative fitness category. The grey dotted line indicates the mean relative fitness of the population.

177   Note that relative fitness values are normally distributed (w= 0.971, var= 0.0072). (b) DFE for plasmids

178   in bacterial hosts obtained in a previous meta-analysis[24]. Most of the included studies were based on

179   arbitrary associations between plasmids and laboratory strains. Bars indicate the number of plasmid-

180   bacterium associations in each relative fitness category. The grey dotted line indicates the mean

181   relative fitness across studies. Relative fitness values are not normally-distributed (w= 0.91, var= 0.029;

182   Shapiro-Wilk normality test, P= 0.0006).

183   *pOXA-48 fitness effects across bacterial phylogeny*

184   A key limit to the prediction of plasmid-mediated evolution is the inability to anticipate plasmid fitness

185   effects in new bacterial hosts. This is particularly relevant to the evolution of antibiotic resistance

186   because some of the most concerning multi-resistant clinical pathogens arise from very specific

187   associations between resistance plasmids and high-risk bacterial clones[4,6,42]. Interestingly, a recent

188   study in an important pathogenic *E. coli* lineage (ST131) showed that the acquisition and maintenance

189   of resistance plasmids is associated with specific genetic signatures[43]. Pursuing this idea, we analysed

190   the DFE for pOXA-48 across the whole-genome phylogeny of our isolates, with the aim of determining if

191   genetic content could help to predict plasmid fitness effects (Figure 4). We calculated the genetic

192   relatedness of *Klebsiella* spp. and *E. coli* isolates by reconstructing their core genome phylogeny

193   (Figure 4a). Plasmid fitness effects can also be strongly influenced by the accessory genome. For

194   example, the presence of further mobile genetic elements can deeply impact the costs of plasmids[44,45].

195   Therefore, we also constructed trees from the distance matrix of the accessory gene network[46], which

196   includes plasmid content (Figure 4b).

197   For each group of isolates, we scanned the fitness effects of pOXA-48 across the core and accessory

198   genome using the local indicator of phylogenetic association index[47,48] (LIPA, see Supplementary

199   Figure 4, Supplementary Table 3, and methods for the complete analysis). For the *E. coli* isolates, the

200   results showed no association of pOXA-48 fitness effects with the core or accessory phylogenies (LIPA,

201   P> 0.1). In contrast, for *Klebsiella* spp., LIPA indices revealed a significant phylogenetic signal in four

202   clones in which pOXA-48 produced a high fitness cost, all of them belonging to ST1427 (Kpn01, Kpn04,

203   Kpn08, and Kpn13, accounting for 4 of the 5 ST1427 clones analysed in this study; LIPA, P< 0.05).

204   Three of these ST1427 clones also produced a significant signal in the analysis of fitness effects across

205   the accessory genome (Kpn01, Kpn08, and Kpn13; LIPA, P< 0.05). The results thus reveal that pOXA-

206   48 tended to produce a high cost in *K. pneumoniae* clones belonging to ST1427. Interestingly, although

207   *K. pneumoniae* ST1427 is relatively common in our hospital (4.8% of ESBL-producing *K.*

208   *pneumoniae*[38]), none of the 103 pOXA-48-carrying *K. pneumoniae* isolates recovered in the R-GNOSIS

209   collection belong to this ST[31] (Fisher's exact test for count data, 8/166 vs 0/103, P= 0.025). These

210 results suggest that the high cost associated with plasmid acquisition in this clade may limit in-hospital

211 spread of pOXA-48-carrying *K. pneumoniae* ST1427. Conversely, pOXA-48 is commonly associated

212 with *K. pneumoniae* ST11 in our hospital[31,36], and in the four ST11 clones tested in this study, pOXA-48

213 produced neutral (Kpn07, Kpn20, Kpn23) or even beneficial fitness effects (Kpn22, Figure 4A) (pOXA-

214 48 fitness effects in ST1427 [n=5] vs. in ST11 [n=4], Welch's unequal variances two-tailed t-test, t= -

215 2.39, df= 7, P= 0.048).
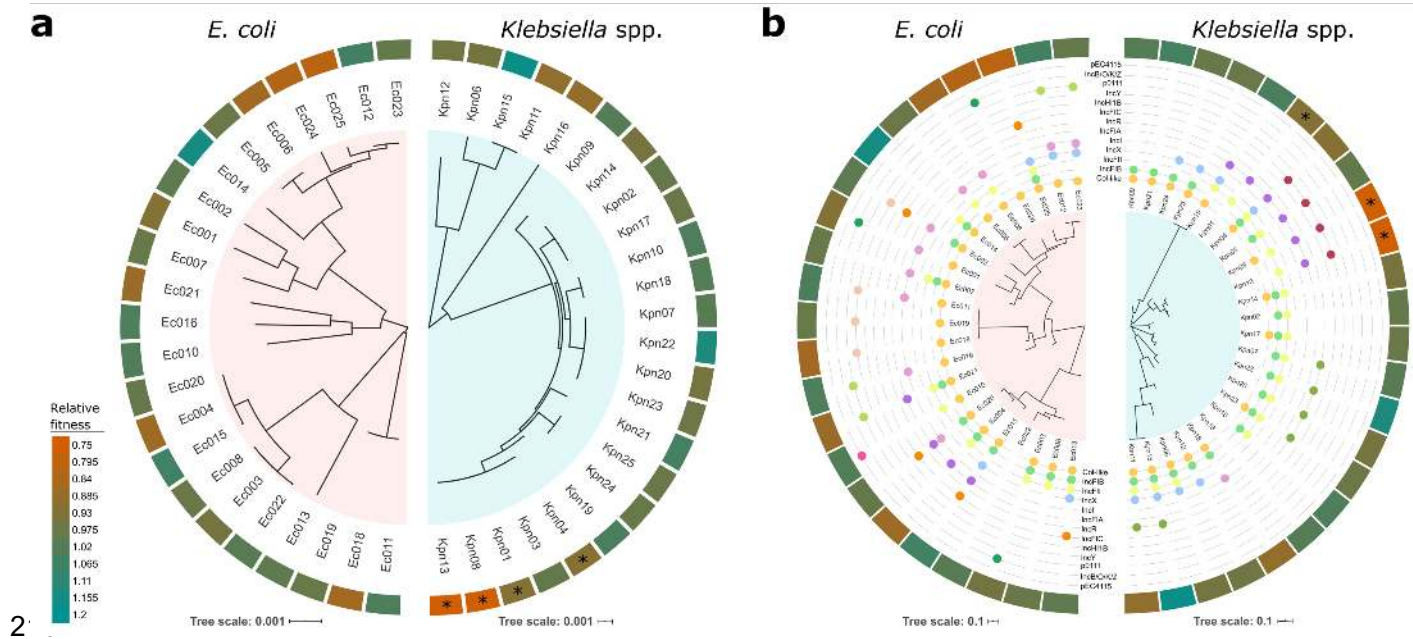
Figure 4. Fitness effects of pOXA-48 across bacterial genome content. An association was found between pOXA-48 fitness effects and bacterial host genomic content for four *K. pneumoniae* ST1427 isolates. (a) Core genome relationships among *E. coli* (left) and *Klebsiella* spp. (right). Tree construction is based on polymorphisms in the core genome. The outer circle indicates the relative fitness of pOXA-48-carrying bacterial hosts (see legend for colour code; red indicates fitness costs and green indicates fitness benefits associated with pOXA-48 carriage). Asterisks denote clones with a phylogenetic signal associated with plasmid fitness effects (LIPA, P< 0.05). (b) Accessory genome relationships among *E. coli* (left) and *Klebsiella* spp. (right). Tree construction is based on the distance matrix of the accessory gene network of each group. The outermost circle indicates relative fitness as in (a). The intermediate circles indicate presence/absence of plasmids belonging to the different plasmid families named in the figure. Asterisks denote clones with a significant phylogenetic signal associating accessory genome composition with pOXA-48 fitness effects (LIPA, P< 0.05).

229 *Modelling the role of DFE in plasmid stability*

230 In general, mathematical models of plasmid population biology consider a clonal population in which the

231 plasmid produces a constant reduction in growth rate[14,18–23]. These models usually include the rate of

232 plasmid loss through segregation[49,50] and the rate of horizontal plasmid transfer by conjugation[19,20,51],

233 and some of them also incorporate a rate of compensatory mutations that alleviate plasmid fitness

234 costs over time[14,23]. Our results show that plasmids produce a wide DFE in naturally compatible

235 bacterial hosts, and this distribution could strongly influence plasmid stability in polyclonal bacterial

236 communities. To assess the effect of the DFE on plasmid stability in bacterial communities, we

237 developed a simple mathematical model based on Stewart and Levin's pioneering work on plasmid

238 existence conditions[19].

239 The model describes the population dynamics of multiple subpopulations competing for a single

240 exhaustible resource in well-mixed environmental conditions, assuming that transition between plasmid-

241 bearing and plasmid-free cells is driven by segregation events. The growth rate of each subpopulation

242 is determined by a substrate-dependent Monod term that depends on the extracellular resource

243 concentration, and therefore each strain can be described by two structurally identifiable parameters[49]:

244 the resource conversion rate (ρ) and the specific affinity for the resource (V_{max}/K_m). These parameters

245 were estimated from the optical densities of each strain growing in monoculture (with and without

246 plasmids) using a Markov chain Monte Carlo (MCMC) method with a Metropolis-Hastings sampler (See

247 Methods, Figure 5a and Supplementary Figure 5).

248 By solving the system of differential equations (described in Methods), we were able to evaluate the

249 final frequency of plasmid-bearing cells in an experiment of duration *T* time units and quantify the

250 fitness effect of the plasmid on the strain. Figure 5b shows the DFE obtained after performing *in silico*

251 pair-wise competition experiments between plasmid-bearing and plasmid-free subpopulations (with

252 parameter values shown in Supplementary Table 5, Supplementary Figure 6), resulting in a theoretical

253 DFE (*w*= 0.985, var= 0.0070) that is consistent with the experimentally measured DFE presented in

254 Figure 3 (*w*= 0.971, var= 0.0072). Moreover, comparison of model predictions with relative fitness

255 values obtained by flow cytometry are consistent (R^2= 0.603; Figure 5c), showing that the population

256 dynamics model can accurately predict the outcome of a competition experiment from the individual

257 growth dynamics.

10

258   Previous studies showed that the probability of plasmid fixation is correlated with the rate of horizontal

259   transmission[19,20,49]. As previous models, we consider horizontal transmission of plasmids as a function

260   of the densities of donor and recipient cells, with conjugation events occurring at a constant rate.

261   Competition experiments for a range conjugation rates are illustrated in Figure 5d; while at low

262   horizontal transmission rates plasmid-free cells outcompete plasmid-bearing cells, at higher conjugative

263   rates, plasmid-bearing cells increase in frequency.

264
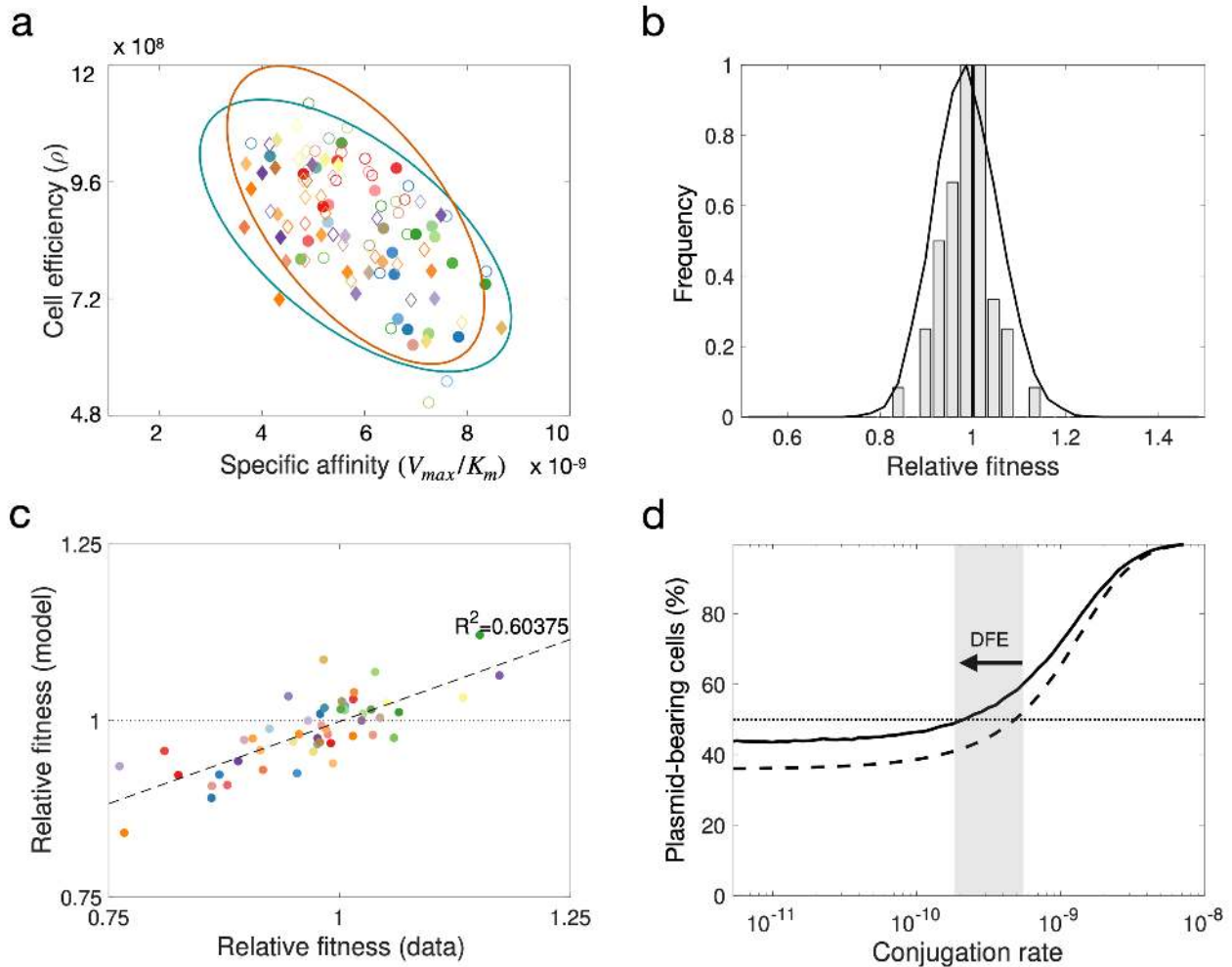
Figure 5. Modelling the DFE for pOXA-48. (a) Distribution of parameter values obtained using Bayesian

inference to estimate growth kinetic parameters from OD measurements obtained for each strain in

isolation. Diamonds represent *Klebsiella* spp. strains and circles *E. coli* clones; filled symbols denote

plasmid-bearing strains and empty symbols plasmid-free cells. The ellipses represent standard

deviations of best-fit Normal distributions (green for plasmid-bearing strains and orange for plasmid-free

cells). (b) Bars represent a DFE obtained from *in silico* competition experiments with parameter values

determined from experimental growth curves. The solid curve represents the computationally estimated

11

272   DFE obtained by randomly sampling wild-type and transconjugant parameter distributions obtained

273   using the MCMC algorithm and numerically solving the model to evaluate the relative fitness associated

274   with plasmid carriage. (c) Comparison of relative fitness values obtained experimentally and using the

275   population dynamics model (R^2= 0.603). (d) Fraction of plasmid-bearing cells as a function of the rate of

276   horizontal transfer for random plasmid-host associations sampled from the MCMC parameter

277   distribution. The dotted line illustrates the mean of 10^4 pair-wise competition experiments under the

278   assumption plasmid-bearing is associated with a constant reduction in fitness in different clones (*w*=

279   0.985, var= 0), while the solid line is obtained by considering a wide DFE (*w*= 0.985, var= 0.0070). The

280   arrow denotes the difference in the conjugation threshold that positively selects for plasmids in the

281   population, supporting the tenet that the DFE maintains plasmids in the population at lower conjugation

282   rates.

---

283   *Community complexity promotes plasmid persistence*

284   To explore how plasmid stability is affected by increasing community complexity and rates of horizontal

285   transmission, we randomly sampled *N*= 10^4 plasmid-free cells from the distribution of growth

286   parameters estimated using the MCMC algorithm. These random communities were used to study the

287   population dynamics of plasmids transmitting vertically and horizontally in multi-strain communities. The

288   fitness cost (or benefit) of bearing plasmids was modelled as a random variable that modifies the wild-

289   type (plasmid-free) growth rate by a factor σ, such that if σ= 0, the DFE has zero variance (Figure 6a),

290   but if σ> 0, the resulting DFE is a symmetrical heavy-tailed distribution with a right-hand tail expanding

291   towards positive fitness effects (Figure 6b), indicating the existence of plasmid-host associations in

292   which plasmid carriage produces a fitness benefit.

293   To assess how DFE influences plasmid persistence in polymicrobial communities, we extended the

294   model to consider populations composed of subsets of 1, 2, 3, 4, ..., *M* ≤ *N* cell types sampled randomly

295   from the wild-type parameter distribution (see Methods and Supplementary Figure 7 and 8). This

296   enabled us to estimate the relative frequency of plasmid-bearing cells at the end of a long-term

297   experiment and evaluate the stability of the plasmid in multi-strain communities with different population

298   structures. Initial bacterial densities were determined by first running the system forward (with all strains

299   initially present at equal densities and carrying pOXA-48) for *T*= 24 time units, and then clearing all

300   plasmid-free cells from the population. This assumption is akin to patients receiving an antibiotic

301   therapy that clears all plasmid-free cells from the bacterial community. The results obtained after 5,000

302   computer simulations over a range of conjugation rates and numbers of cell types in the community are

303   shown in Figure 6. The simulations either assumed an identical fitness cost for all strains (w= 0.985,

304   var= 0; Figure 6a,c) or allowed plasmid fitness effects to vary according to the experimentally

305   determined DFE (w= 0.985, var= 0.0070; Figure 6b,d).

306   Although the mean fitness cost was the same in both conditions, the results of the computational

307   experiments suggest that allowing fitness effects to vary between members of the population markedly

308   increases the chances of plasmid persistence, especially at low conjugation rates. More importantly, in

309   the numerical simulations, plasmid frequency decreased as a function of the number strains in the

310   community when plasmid acquisition was associated with a constant fitness cost, but increased with

311   community complexity for DFEs with larger variance (Figure 6c,d,e). The explanation for this effect is

312   that if plasmid fitness cost is identical for all community members, diversity simply means extra

313   competition for plasmid-carrying cells, and plasmid persistence becomes more dependent on a high

314   conjugation rate. In contrast, if the fitness effects vary, a larger number of available bacterial hosts in

315   the population increases the probability of the plasmid arriving to a host in which it produces a neutral

316   or beneficial fitness effect. This is an important result because it implies that increasing bacterial

317   community complexity could increase the probability of plasmid persistence in natural environments.

318   Given that most natural microbiota are complex and plasmids can usually conjugate and replicate in

319   different clones, this may explain the high prevalence of plasmids in nature. Our results also indicate

320   that the threshold conjugation rate for plasmid persistence may be lower than previously thought. In

321   fact, once plasmids are present in multiple members of a community, they may be able to persist even

322   in the absence of conjugation (Figure 6d).
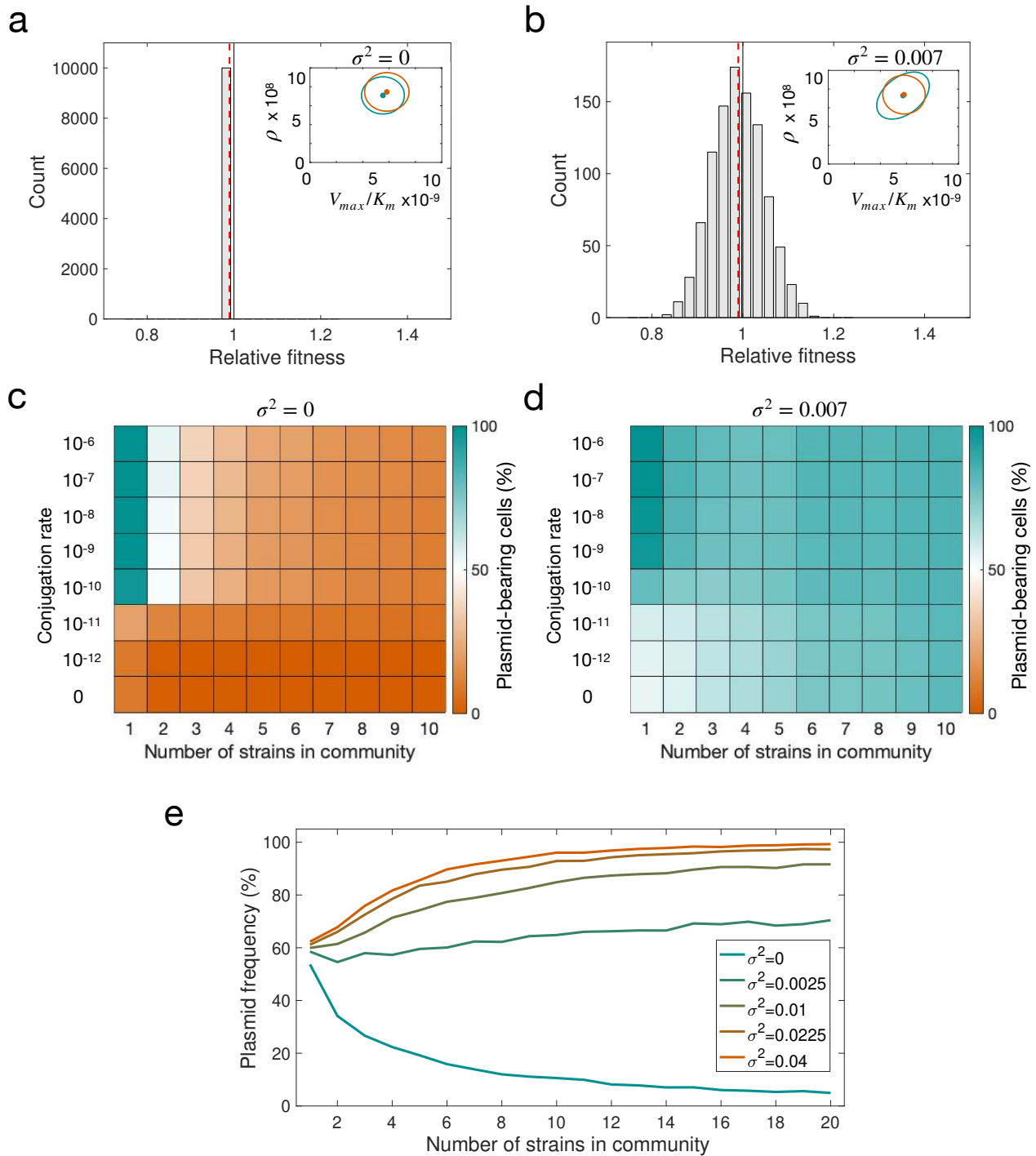
Figure 6. Modelling plasmid persistence in polymicrobial communities, assuming fixed (a,c) or variable (b,d) plasmid fitness effects. (a-b) Relative fitness histogram obtained by randomly sampling 10^4 parameter values from the parameter distribution shown in the inset plot (points illustrate the expected values of each distribution and ellipses their standard deviation; green, plasmid-bearing bacteria; orange, plasmid-free bacteria). The green ellipse in b is larger as a consequence of considering that the cost of plasmid-bearing is normally-distributed with variance 0.007. As a result, the DFE also has higher variance, with a considerable fraction of plasmid-host associations producing a benefit to the host.

14

331   Dotted red lines indicate mean relative fitness of plasmid carrying cells. (c-d) Colour gradient represents

332   the percentage of cells carrying plasmids at the end of 5,000 stochastic simulations; orange indicates a

333   population without plasmids and green a community composed of plasmid-carrying cells. If plasmid-

334   bearing is associated with a fixed fitness cost for all members of the community, plasmid maintenance

335   requires a high conjugation rate. The increased proportion of plasmid-bearing cells in d indicates that a

336   DFE with high variance reduces the critical conjugation rate needed to maintain plasmids in the

337   population, enabling plasmids to persist at low conjugation rates. (e) Mean fraction of plasmid-bearing

338   cells as a function of the number of strains in the community with a conjugation rate $\gamma = 1.5 \times 10^{-11}$. If

339   the plasmid always produces a reduction in host fitness (mean $w < 1$ and low variance), plasmid

340   frequency decreases as the number of strains in the community increases (green line). In contrast, for

341   higher variance at the same mean w, the fraction of plasmid-bearing cells increases with community

342   complexity (orange line).

---

343   **Discussion**

344   The DFE for new mutations is a central concept in genetics and evolutionary biology, with implications

345   ranging from population adaptation rates to complex human diseases[52]. The fitness effects of new

346   spontaneous mutations in bacteria follow a heavy-tailed distribution dominated by quasi-neutral

347   mutations with infrequent strongly deleterious mutations[53,54]. Horizontally acquired genes also produce

348   a distribution of fitness effects in new bacterial hosts[55,56]. However, horizontal gene transfer in bacteria

349   is frequently mediated by entire mobile genetic elements, such as plasmids, that carry multiple genes.

350   Numerous studies have measured the fitness effects of individual plasmids in a bacterial host[24], but the

351   DFE of a plasmid in multiple, ecologically compatible bacterial hosts had not been reported before.

352   Here, we determined the DFE of a carbapenem resistance plasmid in wild-type enterobacteria

353   recovered from the human gut microbiota. Unsurprisingly, the DFE of pOXA-48 differed from the DFE of

354   spontaneous mutations. As with spontaneous mutations, the pOXA-48 DFE was also dominated by

355   quasi-neutral effects and was slightly shifted towards fitness costs; however, instead of a single heavy

356   tail of deleterious effects, it had a symmetrical shape, with tails expanding both towards negative and

357   positive fitness effects (Figure 3a).

358   Two key implications of the experimentally determined DFE in this study are that, according to our

359   simple mathematical model, the probability of plasmid persistence becomes less dependent on a high

15

360   conjugation rate and increases with the number of bacterial strains in the population. The complex and

361   multi-clonal nature of most natural bacterial communities attests the likely relevance of our findings to

362   the extremely high prevalence of plasmids in bacterial populations[57]. The human gut microbiota, for

363   example, includes a great variety of bacteria from hundreds of species[58], including several strains from

364   the *Enterobacterales* order alone[59]. Our experimental system is in fact inspired by the dynamics of

365   pOXA-48 in the gut microbiota of hospitalised patients. In a recent study, we observed that once

366   patients are colonised by a pOXA-48-carrying clone, the plasmid spreads through conjugation to other

367   resident enterobacteria present in the gut microbiota[31]. Crucially, pOXA-48 usually persists in the gut of

368   patients throughout the hospital stay and can be detected in subsequent hospital admissions months or

369   years later, and not necessarily in the original colonizing strain[31]. Our results indicate that the pOXA-48

370   DFE could explain the long-term persistence of this and other plasmids in the human gut microbiota.

371   Another interesting result of this study is that pOXA-48 produced a particularly elevated cost in *K.*

372   *pneumoniae* isolates belonging to ST1427 (Figure 4A). ST1427 is under-represented among the pOXA-

373   48-carrying *K. pneumoniae* isolates in our hospital, which are dominated by ST11[31,36]. Remarkably, in

374   the four *K. pneumoniae* ST11 clones tested in this study, pOXA-48 produced neutral (Kpn07, Kpn20,

375   Kpn23) or even beneficial fitness effects (Kpn22, Figure 4A). Therefore, despite the small number of *K.*

376   *pneumoniae* clones analysed, our results suggest that phylogeny might influence fitness compatibility

377   between plasmids and bacteria at the clonal level, dictating the epidemiology of plasmid-bacterium

378   associations in clinical settings. Further analysis of a larger sample of *K. pneumoniae* isolates from the

379   different STs will be needed to elucidate the genetic basis underlying these specific interactions

380   between bacterial phylogeny and pOXA-48 fitness effects.

381   The main experimental limitation of our study is that plasmid fitness effects were determined *in vitro*,

382   using planktonic cultures in LB medium. This is the standard practise in the field, and previous studies

383   have shown that plasmid fitness effects measured in laboratory conditions correlate with those

384   measured in animal models[24]; however, our results may not be fully representative of pOXA-48 fitness

385   effects in the human gut. Future studies will need to explore more complex *in vitro* systems[60], as well as

386   *in vivo* animal models[61]. Another important limitation of our study is that we modelled bacterial

387   communities with a simple resource competition model that does not consider spatial structure[62],

388   complex ecological interactions between community members[63], plasmid-host co-evolution[64], or

16

389 differential rates of horizontal transmission[28]. Although more complex models[65] will be needed to

390 integrate bacterial community complexity and plasmid fitness effects, consideration of diverse

391 polymicrobial populations with complex spatiotemporal interactions would likely only increase DFE

392 variance, therefore promoting plasmid stability.

393 **Methods**

394 *Strains, pOXA-48 plasmid, and culture conditions*

395 We selected 50 representative ESBL-producing clones form the R-GNOSIS collection (Supplementary

396 Table 1). This collection was constructed in our hospital as part of an active surveillance-screening

397 program for detecting patients colonised by ESBL/carbapenemase-producing enterobacteria, from

398 March 4th, 2014, to March 31st, 2016 (R-GNOSIS-FP7-HEALTH-F3-2011-282512, www.r-gnosis.eu/,

399 approved by the Ramón y Cajal University Hospital Ethics Committee, Reference 251/13)[36,38]. The

400 screening included a total of 28,089 samples from 9,275 patients admitted at 4 different wards

401 (gastroenterology, neurosurgery, pneumology and urology) in the Ramon y Cajal University Hospital

402 (Madrid, Spain). The characterisation of samples was performed during the R-GNOSIS study

403 period[36,66]; rectal swabs were plated on Chromo ID-ESBL and Chrom-CARB/OXA-48 selective agar

404 media (BioMérieux, France) and bacterial colonies able to grow on these media were identified by

405 MALDI-TOF MS (Bruker Daltonics, Germany) and further characterized by pulsed-field gel

406 electrophoresis (PFGE). For the present study, we selected 25 *E. coli* and 25 *K. pneumoniae* ESBL-

407 producing isolates from the R-GNOSIS collection. The strains were representative of *E. coli* and *K.*

408 *pneumoniae* diversity in the R-GNOSIS collection (randomly chosen form the most common pulsed-

409 field gel electrophoresis profiles[38]), they did not carry any carbapenemase gene and they were

410 recovered from patients not colonised by other pOXA-48-carrying clones. To construct the

411 transconjugants, we used the most common pOXA-48 plasmid variant from the R-GNOSIS collection in

412 our hospital, according to plasmid genetic sequence (pOXA-48_K8, accession number MT441554)[30].

413 Bacterial strains were cultured in lysogeny broth (LB) at 37°C in 96-well plates with continuous shaking

414 (250 rpm) and on LB agar plates at 37°C.

415 *Construction of transconjugants collection*

416   We performed an initial conjugation round to introduce pOXA-48_K8 plasmid from wild type *E. coli*

417   C609 strain[31], into *E. coli* β3914[67], a diaminopimelic acid (DAP) auxotrophic laboratory mutant of *E. coli*

418   K-12 (kanamycin, erythromycin and tetracycline resistant, Supplementary Table 1), which was used as

419   the common counter-selectable donor. The pOXA-48-carrying wild type *E. coli* C609 and *E. coli* β3914

420   were streaked from freezer stocks onto solid LB agar medium with ertapenem 0.5 µg/ml and DAP 0.3

421   mM, respectively, and incubated overnight at 37°C. Donor and recipient colonies were independently

422   inoculated in 2 ml of LB in 15-ml culture tubes and incubated overnight. After growth, donor and

423   recipient cultures were collected by centrifugation (15 min, 1,500 g) and cells were re-suspended in

424   each tube with 300 µl of sterile NaCl 0.9%. Then, the suspensions were mixed in a 1:1 proportion,

425   spotted onto solid LB medium with DAP 0.3 mM and incubated at 37ºC overnight. Transconjugants

426   were selected by streaking the conjugation mix on LB with ertapenem (0.5 µg/ml), DAP 0.3 mM,

427   tetracycline (15 µg/ml), and kanamycin (30 µg/ml). The presence of pOXA-48 was checked by PCR,

428   using primers for bla_{OXA-48} gene and for the replication initiation protein gene *repC* (Supplementary

429   Table 4).

430   We used the counter-selectable *E. coli* β3914/pOXA-48_K8 donor to conjugate plasmid pOXA-48 in the

431   50 wild type strains. We used the same protocol described above, but the final conjugation mix was

432   plated on LB with no DAP (to counter-select the donor) and with amoxicillin-clavulanic acid (to select for

433   transconjugants). The optimal concentration of amoxicillin-clavulanic acid was experimentally

434   determined for each isolate in the collection and ranged from 64 µg/ml to 384 µg/ml. The presence of

435   pOXA-48 in the transconjugants was checked by PCR, as described above, and by antibiotic

436   susceptibility testing and whole genome sequencing (see below). To test the stability of plasmid pOXA-

437   48 in the transconjugants we propagated cultures in LB with no antibiotic selection (two consecutive

438   days, 1:10,000 dilution) and plated cultures on LB agar. After ON incubation at 37ºC, 100 independent

439   colonies of each transconjugant were replicated both on LB agar and LB agar with amoxicillin-

440   clavulanic acid to identify pOXA-48-carrying colonies (including negative controls of plasmid-free wild

441   type clones). Results showed that the plasmid was overall stable in the transconjugants; 100% stable in

442   43 isolates, and ≥ 90% stable in the 7 remaining isolates.

443   *Antibiotic susceptibility testing*

18

444 Antibiotic susceptibility profiles were determined for every wild-type and transconjugant strain by the

445 disc diffusion method following the EUCAST guidelines (www.eucast.org) (Supplementary Table 2). We

446 used the following antimicrobials agents: imipenem (10 µg), ertapenem (10 µg), amoxicillin-clavulanic

447 acid (20/10 µg), rifampicin (30 µg), streptomycin (300 µg), chloramphenicol (30 µg) and amikacin (30

448 µg) (Bio-Rad, CA, USA). pOXA-48-carrying and pOXA-48-free strains were pre-cultured in Müller-

449 Hinton (MH) broth at 37 ºC in 15 ml test tubes with continuous shaking (250 rpm), and disc diffusion

450 antibiograms were performed on MH agar plates (BBL, Becton Dickinson, MD, USA).

451 *Growth curves*

452 Pre-cultures of plasmid-free and plasmid-carrying strains (5 replicates of each) were prepared by

453 inoculating single independent colonies into LB broth and overnight incubation at 37 ºC with continuous

454 shaking (250 rpm). Overnight cultures were diluted 1:1,000 into fresh LB in 96-well plates, which were

455 incubated during 22 h at 37 ºC with shaking (250 rpm) in a plate reader (Synergy HTX Multi-Mode

456 Reader, BioTek Instruments, Inc, VT, USA). Optical densities (OD) were measured every 15 minutes

457 during the incubation. The maximum growth rate (μ_{max}), maximum optical density (OD_{max}), and area

458 under the growth curve (AUC) were determined using Gen5™ Microplate Reader and Imager Software

459 and the *growthrates* package in R. We calculated the relative OD_{max}, μ_{max}, and AUC, by dividing the

460 average value of each parameter for the pOXA-48-carrying isolate between that of the pOXA-48-free

461 isolate using the follow formula:

462 $$Relative_{ODmax,Vmax,AUC} = \frac{Plasmid-carrying\ _{ODmax,Vmax,AUC}}{Plasmid-free\ _{ODmax,Vmax,AUC}}$$

463 *Construction of pBGC, a GFP-expressing non-mobilizable plasmid.*

464 To fluorescently label the wild type isolates for competition assays using flow cytometry, we constructed

465 the pBGC plasmid, a non-mobilizable version of the *gfp*-carrying small plasmid pBGT[68] (Supplementary

466 Figure 2, accession number MT702881). The pBGT backbone was amplified, except for the region

467 including the oriT and *bla*TEM1 gene, using the pBGC Fw/Rv primers. The *gfp* terminator region was

468 independently amplified using the GFP-Term Fw/Rv primers (Supplementary Table 4). PCR

469 amplifications were made with Phusion Hot Start II DNA Polymerase at 2 U/µL (ThermoFisher

470 Scientific, MA, USA), and PCR products were digested with DpnI to eliminate plasmid template before

471 setting up the assembly reaction (New England BioLabs, MA, USA). Finally, pBGC was constructed by

19

472   joining the amplified pBGT backbone and the *gfp* terminator region using the Gibson Assembly Cloning

473   Kit (New England BioLabs, MA, USA). Resulting reaction was transformed by heat shock into NEB 5-

474   alpha Competent *E. coli* (New England BioLabs, MA, USA), following manufacturer's instructions.

475   Transformation product was plated on LB agar with arabinose 0.1% and chloramphenicol 30 µg/ml, and

476   incubated overnight at 37 ºC. Plasmid-bearing colonies were selected by green fluorescence. The *gfp*

477   gene in pBGC is under the control of the P_{BAD} promoter, so GFP production is generally repressed and

478   induced by the presence of arabinose. pBGC was completely sequenced using primers described in

479   Supplementary Table 4. We confirmed that neither pOXA-48, nor helper plasmid pTA-Mob[69], could

480   mobilized pBGC by conjugation using the conjugation protocol described above, confirming that pBGC

481   plasmid is not mobilizable. Finally, pBGC plasmid was introduced into our isolate collection by

482   electroporation (Gene Pulser Xcell Electroporator, BioRad, CA, USA). Of note, we were not able to

483   obtain pBGC-carrying transformants in eight of the isolates due to a pre-existing high chloramphenicol

484   resistance phenotype.

485   *Competition assays using flow cytometer*

486   We performed competition assays[41], using flow cytometry, to obtain the relative fitness of pOXA-48-

487   carrying isolates compared to their pOXA-48-free parental counterparts. We used the collection of

488   pBGC transformed wild type isolates as competitors against their isogenic pOXA-48-carrying and

489   pOXA-48-free isolates. Specifically, two sets of competitions were performed for each isolate: pOXA-

490   48-free vs. pBGC-carrying, and pOXA-48-carrying vs. pBGC-carrying. Five biological replicates of each

491   competition were performed. Pre-cultures were incubated overnight in LB in 96-well plates at 225 rpm

492   an 37ºC, then mixed 1:1 and diluted 10,000-fold in 200 µl of fresh LB in in 96-well plates. Mixtures were

493   competed for 24 h in LB at 37ºC and 250 rpm (the low initial cell density and the strong shaking hinders

494   pOXA-48 conjugation, see control experiment below). To determine the initial proportions, initial 1:1

495   mixes were diluted 2,000-fold in 200 µl of NaCl 0.9 % with L-arabinose 0.1 %, and incubated at 37 ºC at

496   250 rpm during 1.5 h to induce *gfp* expression. The measurements were performed via flow cytometry

497   using a CytoFLEX Platform (Beckman Coulter Life Sciences, IN, US) with the following parameters: 50

498   µl min^{-1} flow rate, 22 µm core size, and 10,000 events recorded per sample (Supplementary Figure 9).

499   After 24 hours of incubation, final proportions were determined as described above, after 2,000-fold

20

500   dilution of the cultures. The fitness of each strain relative to its pBGC-carrying parental isolate was

501   determined using the formula:

502     $$w = \frac{\ln(N_f / N_i)}{\ln(N_{f,pBGC+} / N_{i,pBGC+})}$$

503   where w is the relative fitness of the pOXA-48-carrying ($w_{pOXA-48+}$) or pOXA-48-free ($w_{pOXA-48-}$) isolates

504   compared to the pBGC-bearing parental clone, N_i and N_f are the number of cells of the pBGC-free

505   clone at the beginning and end of the competition, and $N_{i,\,pBGC}$ and $N_{f,\,pBGC}$ are the number of cells of the

506   pBGC-carrying clone at the beginning and end of the competition, respectively. The fitness of the

507   pOXA-48-carrying isolates relative to the pOXA-48-free parental isolates were calculated with the

508   formula, $w_{pOXA-48+}$ / $w_{pOXA-48-}$ to correct for the fitness effects of pBCG (see Supplementary Figure 10 for

509   pBGC fitness effects), and the error propagation method was used to calculate the standard error of the

510   resulting value. Note that the fitness effects of pBGC did not correlate with those form pOXA-48

511   (Pearson's correlation, R= 0.11, t= 0.66, df= 39, P= 0.51). For the 8 strains where pBGC plasmid could

512   not be introduced (Ec13, Kpn10, Kpn11, Kpn19-Kpn23), pOXA-48-carrying and pOXA-48-free isolates

513   were competed against a pBGC-carrying $E.\ coli$ J53[70] (a sodium azide resistant laboratory mutant of $E.$

514     $coli$ K-12), following the same protocol described above. In general, we prefer to perform competitions

515   assays between isogenic bacteria to avoid interactions between clones that may affect the outcome of

516   the competition for reasons beyond the presence of the plasmid under study (such as bacteriocin

517   production). However, we did not observe any evidence of growth inhibition between the 8 wild type

518   isolates and $E.\ coli$ J53 in the flow cytometry data, and the relative fitness results obtained with these

519   competitions were comparable to those obtained in the isogenic competitions (two-tailed t-test, t= 1.64,

520   df= 11.2, P= 0.13). To confirm that the isogenic competitions and those against $E.\ coli$ J53/pBGC

521   produced similar results, we selected 10 random isolates from the 42 isolates with fitness data

522   calculated from isogenic competitions, and repeated their competitions against $E.\ coli$ J53/pBGC

523   (Supplementary Figure 11). Results showed that relative fitness values calculated with isogenic

524   competitions and those using $E.\ coli$ J53/pBGC presented a good correlation (Pearson's correlation, R=

525   0.81, t= 3.96, df= 8, P= 0.004, Supplementary Figure 11). Finally, we performed controls to test for the

526   potential conjugative transfer of pOXA-48 during head-to-head competitions by plating the final time

527   points of the competition assays on amoxicillin-clavulanic acid (with the adequate concentration for

528   each isolate), and chloramphenicol (30 µg/ml). No transconjugants were detected in these controls,

529   showing that the low initial inoculum size we used in the competitions (10,000-fold dilution), and the

530   vigorous shaking of the liquid cultures prevented pOXA-48 conjugation.

531   *DNA extraction and genome sequencing*

532   Genomic DNA of all the pOXA-48 bearing strains was isolated using the Wizard genomic DNA

533   purification kit (Promega, WI, USA), and quantified using the QuantiFluor dsDNA system (Promega, WI,

534   USA), following manufacturers' instructions. Whole genome sequencing was conducted at the

535   Wellcome Trust Centre for Human Genetics (Oxford, UK), using the Illumina HiSeq4000 platform with

536   125 base pair (bp) paired-end reads and at MicrobesNG (Birmingham, UK), using Illumina platforms

537   (MiSeq or HiSeq2500) with 250 bp paired-end reads.

538   *Bioinformatic analyses*

539   The Illumina sequence reads were trimmed using the Trimmomatic v0.33 tool[71]. SPAdes v3.9.0[72] was

540   used to generate *de novo* assemblies from the trimmed sequence reads with the –cov-cutoff flag set to

541   'auto'. QUAST v4.6.0[73] was used to generate assembly statistics. Three genomes were dropped from

542   the analysis because of the poor quality of the sequences (2 *E. coli* [Ec09, Ec17] and 1 *K. pneumoniae*

543   [Kpn05]). All the *de novo* assemblies used reached enough quality including total size of 5–7 Mb, and

544   the total number of contigs over 1 kb was lower than 200. Prokka v1.5[74] was used to annotate the *de*

545   *novo* assemblies with predicted genes. The seven-gene ST of all the isolates was determined using the

546   multilocus sequence typing (MLST) tool (https://github.com/tseemann/mlst). The plasmid content of

547   each genome was characterised using PlasmidFinder 2.1[75], and the antibiotic resistance gene content

548   was characterised with ResFinder 3.2[76] (Supplementary Table 1).

549   In order to confirm the presence of the entire pOXA-48_K8 plasmid, the sequences belonging to pOXA-

550   48 plasmid in the transconjugants were mapped using as reference the complete sequence of plasmid

551   from the donor strain, which had been previously sequenced by PacBio[31] (from *K. pneumoniae* k8 –

552   GenBank Accession Number MT441554). Snippy v3.1 (https://github.com/tseemann/snippy) was used

553   to check that no SNPs or indels accumulated in pOXA-48_K8 during strain construction . Coding

554   sequences in pOXA-48 were predicted and annotated using Prokka 1.14.6 software[74]. Plasmid

555   annotation was complemented with the National Center for Biotechnology Information (NCBI)

556   Prokaryotic Genome Annotation Pipeline[77].

557 To determine distances between genomes we used Mash v2.0[78] with the raw sequence reads, and a

558 phylogeny was constructed with mashtree v0.33[79]. For the analysis of the core genome we calculated

559 the genetic relatedness of isolates belonging to *Klebsiella* spp. and to *E. coli* by reconstructing their

560 core genome phylogeny with an alignment of the SNPs obtained with Snippy v3.1

561 (https://github.com/tseemann/snippy). A maximum-likelihood tree was generated using IQ-TREE with

562 automated detection of the best evolutionary model[80]. The tree was represented with midpoint root

563 using the phylotools package in R (https://github.com/helixcn/phylotools) and visualised using the iTOL

564 tool[81]. We also constructed a distance matrix of the accessory gene network to analyse the accessory

565 genome. To this end, we used AccNET, a tool that allows to infer the accessory genome from the

566 proteomes and cluster them based on protein similarity[46]. The set of representative proteins was used

567 to build a binary matrix (presence/absence of proteins in the accessory genome) in the R-environment

568 and a cladogram to classify the strains according to the accessory genomes. The Euclidean distance

569 was calculated by the 'dist' function and a hierarchical clustering was performed with UPGMA using the

570 'hclust' function in the R environment. This cladogram was represented with midpoint root using the

571 phylotools package in R (https://github.com/helixcn/phylotools) and visualised using the iTOL tool[81].

572 *Analysis of plasmid fitness effects across bacterial phylogeny*

573 We tested for the presence of phylogenetic signal in core and accessory genomes of *E. coli* and *K.*

574 *pneumoniae* using several statistical tests available in the *phylosignal* R package[47]. In essence, these

575 analyses are designed to identify statistical dependence between a given continuous trait (relative

576 fitness) and the phylogenetic tree of the taxa from which the trait is measured. Therefore, a positive

577 phylogenetic signal indicates that there is a tendency for related taxa to resemble each other[82]. Several

578 indices have been proposed to identify phylogenetic signal, but the choice among them is not

579 straightforward[83]. We first assayed the methods implemented in the *phyloSignal* function, which

580 produce global measures of phylogenetic signal (*i.e.* across the whole phylogeny). The methods

581 employed were Abouheif's C_{mean}, Moran's I index, Bloomberg's K and K*, and Pagel's λ[47]. All methods

582 except Pagel's λ detected a marginally significant phylogentic signal in the *K. pneumoniae* core genome

583 (Supplementary table 3 [first tab]; 0.11>P>0.02). Abouheif's C_{mean} and Moran's I (but not Bloomberg's K

584 and K*, and Pagel's λ) also detected a marginally significant signal in the *K. pneumoniae* accessory

585 genome tree (Supplementary table 3 [first tab]; P<0.056 for both cases). Intrigued by these results, we

23

586 used the Local Indicator of Phylogenetic Association (LIPA) based on local Moran's I, which is meant to

587 detect local hotspots of phylogenetic signal[47,48]. LIPA, implemented in the *lipaMoran* function, computes

588 local Moran's I indexes for each tip of the phylogeny and a non-parametric test to ascertain statistical

589 significance (Supplementary Figure 4 and Supplementary table 3 [second tab]).

590 *Plasmid population dynamics model*

591 We used a simple mathematical model of microbial growth under resource limitation to study the role of

592 the DFE in the ecological dynamics of a plasmid spreading in a bacterial population[14]. Bacterial growth

593 rate was modelled as a saturating function of the environmental resource concentration, R,

594 $G(R) = \rho \cdot \frac{V_{max} R}{K_m + R} = \rho \cdot u(R),$

595 where ρ denotes the cell's efficiency to convert resource molecules into biomass and $u(R)$ a resource

596 uptake function that depends on the maximum uptake rate (V_{max}) and a half-saturation constant (K_m).

597 If we denote with B_p the density of plasmid-bearing cells and with B_0 the density of plasmid-free cells

598 (each with its own growth kinetic parameters and growth functions denoted $G_p(R)$ and $G_0(R)$,

599 respectively), then the density of each subpopulation can be described by a system of ordinary

600 differential equations:

601 $\frac{dR}{dt} = -u_p(R) - u_0(R) - dR,$

602 $\frac{dB_p}{dt} = (1 - \lambda) \, G_p(R) B_p + \gamma B_0 B_p - dB_p,$

603 $\frac{dB_0}{dt} = G_0(R) B_0 + \lambda \, G_p(R) B_p - \gamma B_0 B_p - dB_0.$

604 where λ represents the rate of segregational loss rate and d a dilution parameter. Moreover, we

605 represent with γ the rate of conjugative transfer, and therefore we model plasmid conjugation as a

606 function of the densities of donor and recipient cells. By numerically solving the system of equations

607 (using standard differential equations solvers in Matlab), we obtain the final density of each bacterial

608 type in an experiment of $T = 24$ units of time with $d = 0$ (to replicate the batch culture conditions used

609 to estimate the DFE experimentally).

610 Growth kinetic parameters were determined with a Markov chain Monte Carlo method (MCMC; scripts

611 coded in R and available at http://www.github.com/esb-lab/pNUK73/) applied to growth curves of each

24

612   strain growing in isolation, with and without plasmids. This data fitting algorithm implements a

613   Metropolis-Hastings sampler with a burn-in parameter of 0.2 and executed for 5×10^6 iterations, or until

614   achieving convergence of the Markov chains (see Supplementary Figure 5 for an example and

615   Supplementary Table 5 for parameters values estimated for each strain).

616   *Stochastic simulations of polymicrobial communities*

617   Numerical experiments were performed by randomly sampling $N = 1 \times 10^4$ cells from the parameter

618   distribution obtained after applying the MCMC algorithm to all 50 strains and fitting a bivariate Normal

619   distribution. We then assembled 5,000 synthetic communities composed of a random subset of $M < N$

620   different strains sampled from this distribution, and solved a multi-strain extension of the population

621   dynamics model. For each numerical experiment, the total density of strain i would be $B^i(t) = B_p^i(t) +$

622     $B_0^i(t)$, where B_p^i and B_0^i denote, respectively, the densities of plasmid-bearing and plasmid-free cells of

623   type $1 \leq i \leq M$. To model the fitness effects of bearing plasmids, we introduced a parameter, σ, such

624   that when $\sigma = 0$, the fitness difference between B_p^i and B_0^i corresponds to a fixed reduction in growth

625   rate (corresponding to a DFE with variance 0 and mean $w = 0.985$). Conversely, if $\sigma > 0$, then growth

626   kinetic parameters for each plasmid-bearing strain in the community were determined by sampling s_i

627   from a Normal distribution, $N(0, \sigma^2)$, and multiplying both V_{max}^i and ρ^i by a factor of $(1 + s_i)$.

628   As with the single-strain model, we consider segregational loss as a transition from B_p^i to B_0^i occurring

629   at a rate λ, but now we also consider that plasmid-free cells can acquire plasmids via conjugation from

630   any plasmid-bearing strain in the community, at a constant rate γ, and with equal probability of

631   transferring between different bacterial hosts. Therefore, we obtain a system of $2M + 1$ differential

632   equations that can be written, for each strain i, as follows:

633     $$\frac{dB_p^i}{dt} = (1 - \lambda)\, G_p^i(R)\, B_p^i \;+ \gamma \sum_{j=1}^{M} B_p^j B_0^i \;-\; dB_p^i,$$

634     $$\frac{dB_0^i}{dt} = \; G_0^i(R)\, B_0^i \;+\; \lambda G_p^i(R)\, B_p^i \;-\; \gamma \sum_{j=1}^{M} B_p^j B_0^i \;-\; dB_0^i.$$

635   Furthermore, if $\hat{R}$ represents the input of resource into the system, then

636     $$\frac{dR}{dt} = -\sum_{i=1}^{M} \left(u_p^i(R) + u_0^i(R) \right) - d\left(R - \hat{R} \right).$$

25

637   Initial bacterial densities were determined by first running the system forward (with all strains initially

638   present at equal densities) for $T = 24$ time units, and then clearing all plasmid-free cells from the

639   population. This assumption is consistent with patients receiving antimicrobial therapy that clears all

640   susceptible (plasmid-free) cells from the microbiota or, in an experimental microcosm, to a round of

641   growth in selective media after an overnight culture. As we are interested in the long-term population

642   dynamics, we ran each simulation starting from the aforementioned initial condition until the plasmid

643   fraction was below a threshold $\epsilon > 0$ (i.e. plasmid extinction), the plasmid fraction was near 100% and

644   the total plasmid-free density was below ϵ (i.e. plasmid fixation), or wild-type and transconjugant sub-

645   populations appeared to co-exist indefinitely in the population (either in equilibrium or exhibiting

646   oscillatory behaviour, as illustrated in Supplementary Figures 7 and 8).

647   *Statistical analyses*

648   The statistical tests used are indicated in the text. Analyses were performed using R (v. 3.5.0).

649   **References**

650   1.   LEDERBERG, J. Cell genetics and hereditary symbiosis. *Physiol. Rev.* **32**, 403–430 (1952).

651   2.   Frost, L. S., Leplae, R., Summers, A. O. & Toussaint, A. Mobile genetic elements: The agents of
652   open source evolution. *Nature Reviews Microbiology* vol. 3 722–732 (2005).

653   3.   Partridge, S. R., Kwong, S. M., Firth, N. & Jensen, S. O. Mobile genetic elements associated with
654   antimicrobial resistance. *Clinical Microbiology Reviews* vol. 31 (2018).

655   4.   San Millan, A. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. *Trends in*
656   *Microbiology* (2018) doi:10.1016/j.tim.2018.06.007.

657   5.   Bonomo, R. A. *et al.* Carbapenemase-Producing Organisms: A Global Scourge. *Clinical Infectious*
658   *Diseases* vol. 66 1290–1297 (2018).

659   6.   David, S. *et al.* Epidemic of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by
660   nosocomial spread. *Nat. Microbiol.* **4**, 1919–1929 (2019).

661   7.   San Millan, A. & MacLean, R. C. Fitness Costs of Plasmids: a Limit to Plasmid Transmission.
662   *Microbiol. Spectr.* (2017) doi:10.1128/microbiolspec.MTBP-0016-2017.

663   8.   Baltrus, D. A. Exploring the costs of horizontal gene transfer. *Trends in Ecology and Evolution* vol. 28
664   489–495 (2013).

665   9.   San Millan, A. *et al.* Integrative analysis of fitness and metabolic effects of plasmids in Pseudomonas
666   aeruginosa PAO1. *ISME J.* **12**, 3014–3024 (2018).

667   10.   Harrison, E. & Brockhurst, M. A. Plasmid-mediated horizontal gene transfer is a coevolutionary
668   process. *Trends in Microbiology* (2012) doi:10.1016/j.tim.2012.04.003.

669   11.   Bouma, J. E. & Lenski, R. E. Evolution of a bacteria/plasmid association. *Nature* **335**, 351–352
670   (1988).

671   12.   Dahlberg, C. & Chao, L. Amelioration of the Cost of Conjugative Plasmid Carriage in Eschericha coli

672 K12. *Genetics* **165**, 1641–1649 (2003).

673 13. Dionisio, F., Conceiçã, I. C., Marques, A. C. R., Fernandes, L. & Gordo, I. The evolution of a
674   conjugative plasmid and its ability to increase bacterial fitness. *Biol. Lett* **1**, 250–252 (2005).

675 14. San Millan, A. *et al.* Positive selection and compensatory adaptation interact to stabilize non-
676   transmissible plasmids. *Nat. Commun.* **5**, 5208 (2014).

677 15. Harrison, E., Guymer, D., Spiers, A. J., Paterson, S. & Brockhurst, M. A. Parallel Compensatory
678   Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum. *Curr. Biol.* **25**, 2034–2039
679   (2015).

680 16. Loftie-Eaton, W. *et al.* Compensatory mutations improve general permissiveness to antibiotic
681   resistance plasmids. *Nat. Ecol. Evol.* (2017) doi:10.1038/s41559-017-0243-2.

682 17. Wein, T., Hülter, N. F., Mizrahi, I. & Dagan, T. Emergence of plasmid stability under non-selective
683   conditions maintains antibiotic resistance. *Nat. Commun.* **10**, 1–13 (2019).

684 18. Lopatkin, A. J. *et al.* Persistence and reversal of plasmid-mediated antibiotic resistance. *Nat.*
685   *Commun.* **8**, 1689 (2017).

686 19. Stewart, F. M. & Levin, B. R. The Population Biology of Bacterial Plasmids: A PRIORI Conditions for
687   the Existence of Conjugationally Transmitted Factors. *Genetics* **87**, 209–28 (1977).

688 20. Bergstrom, C. T., Lipsitch, M. & Levin, B. R. Natural selection, infectious transfer and the existence
689   conditions for bacterial plasmids. *Genetics* **155**, 1505–1519 (2000).

690 21. Zwanzig, M. *et al.* Mobile Compensatory Mutations Promote Plasmid Survival. *mSystems* **4**, (2019).

691 22. Lili, L. N., Britton, N. F. & Feil, E. J. The persistence of parasitic plasmids. *Genetics* **177**, 399–405
692   (2007).

693 23. Ponciano, J. M., De Gelder, L., Top, E. M. & Joyce, P. The population biology of bacterial plasmids: A
694   hidden Markov model approach. *Genetics* **176**, 957–968 (2007).

695 24. Vogwill, T. & Maclean, R. C. The genetic basis of the fitness costs of antimicrobial resistance: A
696   meta-analysis approach. *Evol. Appl.* **8**, 284–295 (2015).

697 25. De Gelder, L., Ponciano, J. M., Joyce, P. & Top, E. M. Stability of a promiscuous plasmid in different
698   hosts: no guarantee for a long-term relationship. *Microbiology* **153**, 452–463 (2007).

699 26. Humphrey, B. *et al.* Fitness of Escherichia coli strains carrying expressed and partially silent IncN and
700   IncP1 plasmids. *BMC Microbiol* **12**, 53 (2012).

701 27. Di Luca, M. C. *et al.* Low biological cost of carbapenemase-encoding plasmids following transfer from
702   Klebsiella pneumoniae to Escherichia coli. *J. Antimicrob. Chemother.* **72**, 85–89 (2017).

703 28. Gama, J. A., Kloos, J., Johnsen, P. J. & Samuelsen, Ø. Host dependent maintenance of a bla NDM-
704   1-encoding plasmid in clinical Escherichia coli isolates. *Sci. Rep.* **10**, 1–7 (2020).

705 29. Li, L. *et al.* Plasmids persist in a microbial community by providing fitness benefit to multiple
706   phylotypes. *ISME J.* **14**, 1170–1181 (2020).

707 30. Klümper, U. *et al.* Broad host range plasmids can invade an unexpectedly diverse fraction of a soil
708   bacterial community. *ISME J.* **9**, 934–945 (2015).

709 31. Leon-Sampedro, R. *et al.* Dissemination routes of the carbapenem resistance plasmid pOXA-48 in a
710   hospital setting. *bioRxiv* 2020.04.20.050476 (2020) doi:10.1101/2020.04.20.050476.

711 32. Dahlberg, C. *et al.* Interspecies bacterial conjugation by plasmids from marine environments
712   visualized by gfp expression. *Mol. Biol. Evol.* **15**, 385–390 (1998).

713 33. Poirel, L., Héritier, C., Tolün, V. & Nordmann, P. Emergence of Oxacillinase-Mediated Resistance to

714 Imipenem in Klebsiella pneumoniae. *Antimicrob. Agents Chemother.* **48**, 15–22 (2004).

715 34. Pitout, J. D. D., Peirano, G., Kock, M. M., Strydom, K.-A. & Matsumura, Y. The Global Ascendency of
716 OXA-48-Type Carbapenemases. *Clin. Microbiol. Rev.* **33**, (2020).

717 35. Ledda, A. *et al.* Hospital outbreak of carbapenem-resistant Enterobacteriales associated with an
718 OXA-48 plasmid carried mostly by Escherichia coli ST399. *bioRxiv* 2020.06.15.148189 (2020)
719 doi:10.1101/2020.06.15.148189.

720 36. Hernández-García, M. *et al.* Characterization of carbapenemase-producing Enterobacteriaceae from
721 colonized patients in a university hospital in Madrid, Spain, during the R-GNOSIS project depicts
722 increased clonal diversity over time with maintenance of high-risk clones. *J. Antimicrob. Chemother.*
723 **73**, 3039–3043 (2018).

724 37. Diaz-Agero, C. *et al.* Local prevalence of extended-spectrum beta-lactamase (ESBL) producing
725 Enterobacteriaceae intestinal carriers at admission and co-expression of ESBL and OXA-48
726 carbapenemase in Klebsiella pneumoniae: A prevalence survey in a Spanish University Hospital.
727 *BMJ Open* **9**, (2019).

728 38. Maechler, F. *et al.* Contact isolation versus standard precautions to decrease acquisition of extended-
729 spectrum β-lactamase-producing Enterobacterales in non-critical care wards: a cluster-randomised
730 crossover trial. *Lancet Infect. Dis.* **20**, 575–584 (2020).

731 39. Karanika, S., Karantanos, T., Arvanitis, M., Grigoras, C. & Mylonakis, E. Fecal Colonization with
732 Extended-spectrum Beta-lactamase-Producing Enterobacteriaceae and Risk Factors among Healthy
733 Individuals: A Systematic Review and Metaanalysis. *Clin. Infect. Dis.* **63**, 310–318 (2016).

734 40. Holt, K. E. *et al.* Genomic analysis of diversity, population structure, virulence, and antimicrobial
735 resistance in Klebsiella pneumoniae, an urgent threat to public health. *Proc. Natl. Acad. Sci. U. S. A.*
736 **112**, E3574–E3581 (2015).

737 41. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. *Long-Term Experimental Evolution in*
738 *Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. Source: The American*
739 *Naturalist* vol. 138 (1991).

740 42. Stoesser, N. *et al.* Evolutionary history of the global emergence of the Escherichia coli epidemic clone
741 ST131. *MBio* **7**, (2016).

742 43. McNally, A. *et al.* Combined Analysis of Variation in Core, Accessory and Regulatory Genome
743 Regions Provides a Super-Resolution View into the Evolution of Bacterial Populations. *PLOS Genet.*
744 **12**, e1006280 (2016).

745 44. San Millan, A., Heilbron, K. & MacLean, R. C. Positive epistasis between co-infecting plasmids
746 promotes plasmid survival in bacterial populations. *ISME J.* **8**, 601–612 (2014).

747 45. San Millan, A., Toll-Riera, M., Qi, Q. & MacLean, R. C. Interactions between horizontally acquired
748 genes create a fitness cost in Pseudomonas aeruginosa. *Nat. Commun.* **6**, (2015).

749 46. Lanza, V. F., Baquero, F., De La Cruz, F. & Coque, T. M. AcCNET (Ac cessory Genome C
750 onstellation Net work): Comparative genomics software for accessory genome analysis using
751 bipartite networks. *Bioinformatics* **33**, 283–285 (2017).

752 47. Keck, F., Rimet, F., Bouchez, A. & Franc, A. Phylosignal: An R package to measure, test, and
753 explore the phylogenetic signal. *Ecol. Evol.* **6**, 2774–2780 (2016).

754 48. Anselin, L. Local Indicators of Spatial Association-LISA. *Geogr. Anal.* **27**, 93–115 (2010).

755 49. Cooper, N. S., Brown, M. E. & Caulcott, C. A. A mathematical method for analysing plasmid stability

28

756   in micro-organisms. *J. Gen. Microbiol.* **133**, 1871–1880 (1987).

757 50.   Paulsson, J. & Ehrenberg, M. Noise in a minimal regulatory network: Plasmid copy number control.

758   *Quarterly Reviews of Biophysics* vol. 34 1–59 (2001).

759 51.   Simonsen, L., Gordon, D. M., Stewart, F. M. & Levin, B. R. Estimating the rate of plasmid transfer: An

760   end-point method. *J. Gen. Microbiol.* **136**, 2319–2325 (1990).

761 52.   Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev.* **8**,

762   608–618 (2007).

763 53.   Bataillon, T. & Bailey, S. F. Effects of new mutations on fitness: insights from models and data. *Ann.*

764   *N. Y. Acad. Sci.* **1320**, 76–92 (2014).

765 54.   Robert, L. *et al.* Mutation dynamics and fitness effects followed in single cells. *Science (80-.).* **359**,

766   1283–1286 (2018).

767 55.   Vos, M., Hesselman, M. C., te Beek, T. A., van Passel, M. W. J. & Eyre-Walker, A. Rates of Lateral

768   Gene Transfer in Prokaryotes: High but Why? *Trends in Microbiology* vol. 23 598–605 (2015).

769 56.   Sorek, R. *et al.* Genome-wide experimental determination of barriers to horizontal gene transfer.

770   *Science (80-.).* **318**, 1449–1452 (2007).

771 57.   Acman, M., van Dorp, L., Santini, J. M. & Balloux, F. Large-scale network analysis captures biological

772   features of bacterial plasmids. *Nat. Commun.* **11**, 2452 (2020).

773 58.   Almeida, A. *et al.* A new genomic blueprint of the human gut microbiota. *Nature* **568**, 499–504 (2019).

774 59.   Martinson, J. N. V. *et al.* Rethinking gut microbiome residency and the Enterobacteriaceae in healthy

775   human adults. *ISME J.* **13**, 2306–2318 (2019).

776 60.   Stalder, T. *et al.* Evolving Populations in Biofilms Contain More Persistent Plasmids. *Mol. Biol. Evol.*

777   (2020) doi:10.1093/molbev/msaa024.

778 61.   Gumpert, H. *et al.* Transfer and persistence of a multi-drug resistance plasmid in situ of the infant gut

779   microbiota in the absence of antibiotic treatment. *Front. Microbiol.* **8**, (2017).

780 62.   Nadell, C. D., Drescher, K. & Foster, K. R. Spatial structure, cooperation and competition in biofilms.

781   *Nature Reviews Microbiology* vol. 14 589–600 (2016).

782 63.   Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: Networks, competition, and

783   stability. *Science (80-.).* **350**, 663–666 (2015).

784 64.   Jordt, H. *et al.* Coevolution of host–plasmid pairs facilitates the emergence of novel multidrug

785   resistance. *Nat. Ecol. Evol.* **4**, (2020).

786 65.   Lopatkin, A. J. & Collins, J. J. Predictive biology: modelling, understanding and harnessing microbial

787   complexity. *Nature Reviews Microbiology* (2020) doi:10.1038/s41579-020-0372-5.

788 66.   Hernández-García, M. *et al.* Intestinal co-colonization with different carbapenemase-producing

789   Enterobacterales isolates is not a rare event in an OXA-48 endemic area. *EClinicalMedicine* **15**, 72–

790   79 (2019).

791 67.   Le Roux, F., Binesse, J., Saulnier, D. & Mazel, D. Construction of a Vibrio splendidus mutant lacking

792   the metalloprotease gene vsm by use of a novel counterselectable suicide vector. *Appl. Environ.*

793   *Microbiol.* **73**, 777–784 (2007).

794 68.   San Millan, A., Escudero, J. A., Gifford, D. R., Mazel, D. & MacLean, R. C. Multicopy plasmids

795   potentiate the evolution of antibiotic resistance in bacteria. *Nat. Ecol. Evol.* (2016)

796   doi:10.1038/s41559-016-0010.

797 69.   Strand, T. A., Lale, R., Degnes, K. F., Lando, M. & Valla, S. A New and Improved Host-Independent

798   Plasmid System for RK2-Based Conjugal Transfer. *PLoS One* **9**, e90372 (2014).

799   70.   Matsumura, Y., Peirano, G. & Pitout, J. D. D. Complete genome sequence of Escherichia coli J53, an

800   azide-resistant laboratory strain used for conjugation experiments. *Genome Announc.* **6**, (2018).

801   71.   Bolger, M. A., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data |

802   Bioinformatics | Oxford Academic. *Bioinformatics (Oxford, England)* vol. 30 2114–2120

803   https://academic.oup.com/bioinformatics/article/30/15/2114/2390096 (2014).

804   72.   Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell

805   sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

806   73.   Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome

807   assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

808   74.   Seeman, T. Prokka: rapid prokaryotic genome annotation | Bioinformatics | Oxford Academic.

809   *Bioinformatics* **30**, 2068–2069 (2014).

810   75.   Carattoli, A. *et al.* In Silico detection and typing of plasmids using plasmidfinder and plasmid

811   multilocus sequence typing. *Antimicrob. Agents Chemother.* **58**, 3895–3903 (2014).

812   76.   Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.*

813   **67**, 2640–2644 (2012).

814   77.   Haft, D. H. *et al.* RefSeq: An update on prokaryotic genome annotation and curation. *Nucleic Acids*

815   *Res.* **46**, D851–D860 (2018).

816   78.   Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash.

817   *Genome Biol.* **17**, 132 (2016).

818   79.   Katz, L. S. *et al.* Mashtree: a rapid comparison of whole genome sequence files. (2019)

819   doi:10.21105/joss.01762.

820   80.   Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective

821   stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274

822   (2015).

823   81.   Letunic, I. & Bork, P. Interactive Tree of Life v2: Online annotation and display of phylogenetic trees

824   made easy. *Nucleic Acids Res.* **39**, (2011).

825   82.   Blomberg, S. P., Garland, T. & Ives, A. R. Testing for phylogenetic signal in comparative data:

826   Behavioral traits are more labile. *Evolution (N. Y).* **57**, 717–745 (2003).

827   83.   Münkemüller, T. *et al.* How to measure and test phylogenetic signal. *Methods Ecol. Evol.* **3**, 743–756

828   (2012).

829

Acknowledgements

30

847 **Author Contributions**

848 ASM, AAdV and RPM conceived the study. RC designed and supervised sampling and collection

849 of R-GNOSIS bacterial isolates. MHG, PRG collected the bacterial isolates and performed

850 bacterial characterization. AAdV performed the experimental work with help from JRB and JdLF.

851 AAdV, JRB and JdLF analysed experimental results. RLS and JRB performed the

852 bioinformatic/phylogenetic analyses. RPM developed the mathematical model and computer

853 simulations. ASM supervised the study. ASM, AAdV and RPM wrote the initial draft of the

854 manuscript and all the authors contributed to the final version of the manuscript and approved it.

855 **Competing Interests statement**

856 Authors declare no competing interests.

857 **Data availability**

858 The sequences generated and analysed during the current study are available in the Sequence Read

859 Archive (SRA), BioProject ID: PRJNA641166, https://www.ncbi.nlm.nih.gov/sra/PRJNA641166.

860 **Code availability**

861 The code generated during the current study is available in GitHub, http://www.github.com/esb-lab/

862 pOXA48/

1 **Supplementary Information**

2

3 **The distribution of plasmid fitness effects explains plasmid persistence in bacterial**

4 **communities**

5 Aida Alonso-del Valle[1], Ricardo León-Sampedro[1,2], Jerónimo Rodríguez-Beltrán[1,2], Javier

6 DelaFuente[1], Marta Hernández-García[1,3], Patricia Ruiz-Garbajosa[1,3], Rafael Cantón[1,3], Rafael

7 Peña-Miller[4,*], Álvaro San Millán[1,2,5]*.

8

9 [1] *Servicio de Microbiología. Hospital Universitario Ramón y Cajal* and *Instituto Ramón y Cajal de*

10 *Investigación Sanitaria*. Madrid, Spain.

11 [2] *Centro de Investigación Biológica en Red. Epidemiología y Salud Pública*, *Instituto de Salud Carlos*

12 *III*. Madrid. Spain.

13 [3] *Red Española de Investigación en Patología Infecciosa. Instituto de Salud Carlos III*. Madrid. Spain.

14 [4] Center for Genomic Sciences, *Universidad Nacional Autónoma de México*, Cuernavaca, Mexico.

15 [5] Current address: *Centro Nacional de Biotecnología–CSIC*, Madrid, Spain.

16 * Correspondence: Álvaro San Millán, alvsanmillan@gmail.com ORCID: 0000-0001-8544-0387 and

17 Rafael Peña-Miller, rafael.penamiller@gmail.com ORCID: 0000-0002-2767-0640

18

19

20   Supplementary Figure 1. Growth curves of wild-type isolates and pOXA-48-carrying transconjugants.
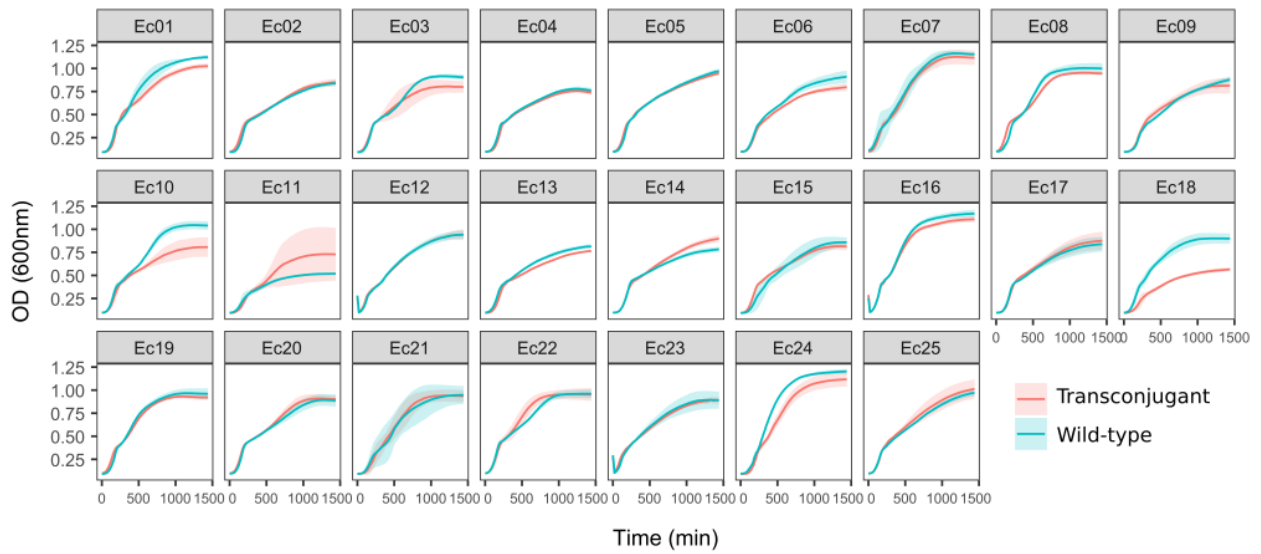
23   Growth curves of pOXA48-free (wild-type, blue) and pOXA-48-carrying (transconjugant, red) for

24   every (a) *E. coli* and (b) *Klebsiella* spp. analysed in this study. The lines represent the average of

25   five biological replicates and the shaded area indicates 95% confidence intervals.

2

26   Supplementary Figure 2. Construction of plasmid pBGC.
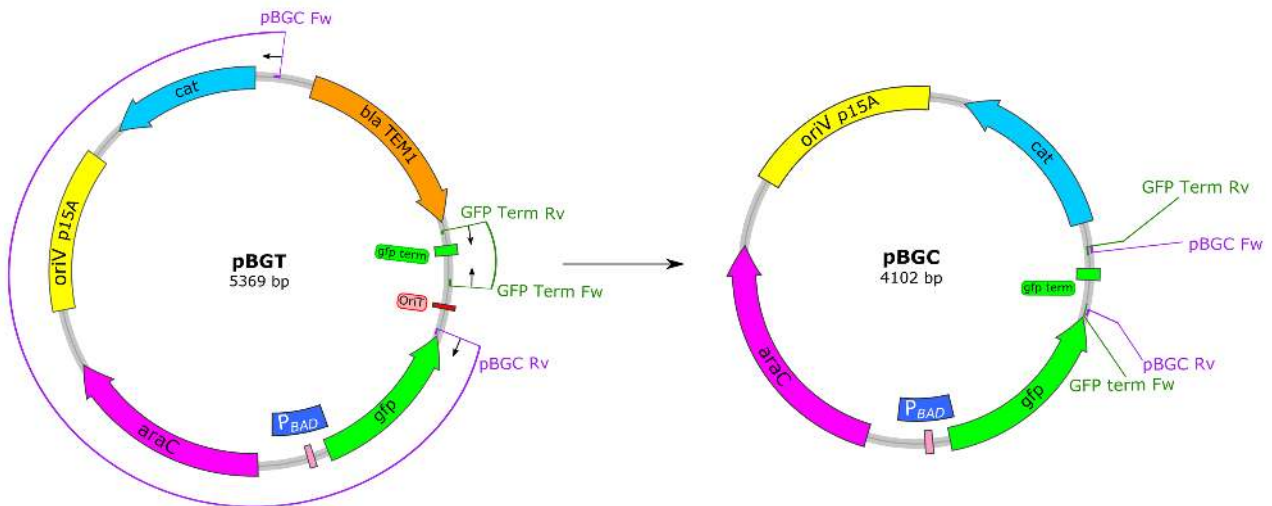
27

28   Schematic representation of the construction of plasmid pBGC (accession number MT702881) form

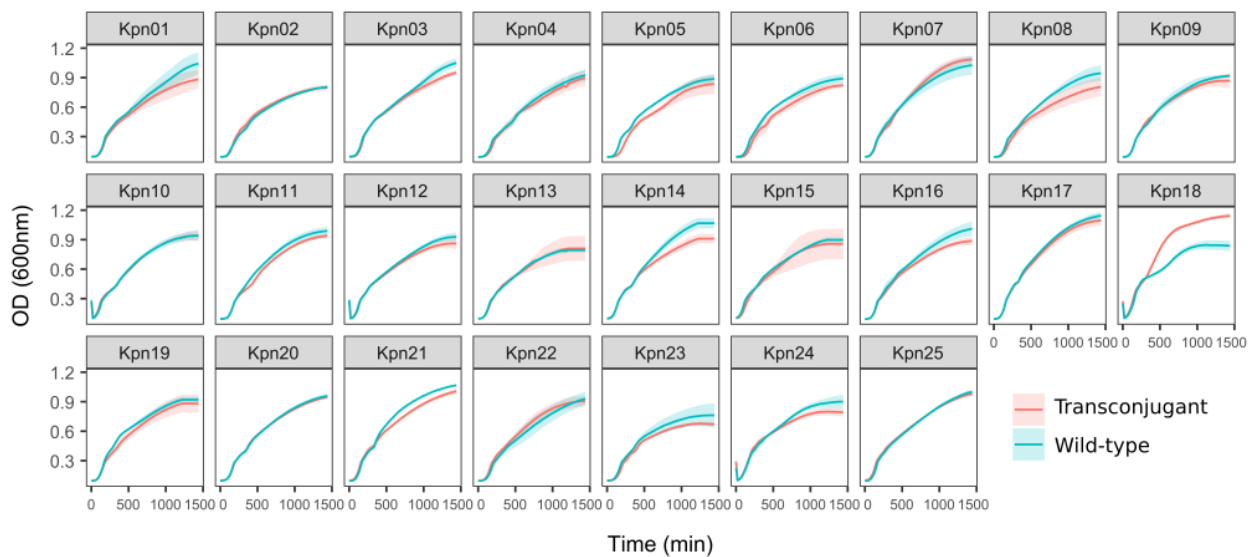29   plasmid pBGT[68]. Two segments of pBGT were amplified using primers with added cohesive ends

30   (pBGC Fw/Rv and GFP Term Fw/Rv, Supplementary Table 4). pBGC plasmid resulted from the Gibson

31   assembly of the amplified fragments. The reading frames for genes are shown as arrows, with the

32   direction of transcription indicated by the arrowhead. The origin of replication (oriV), origin of transfer

33   (oriT), and P$_{BAD}$ promoter are also indicated.

3

34 Supplementary Figure 3. Correlation between relative growth curve parameters and relative fitness.
35
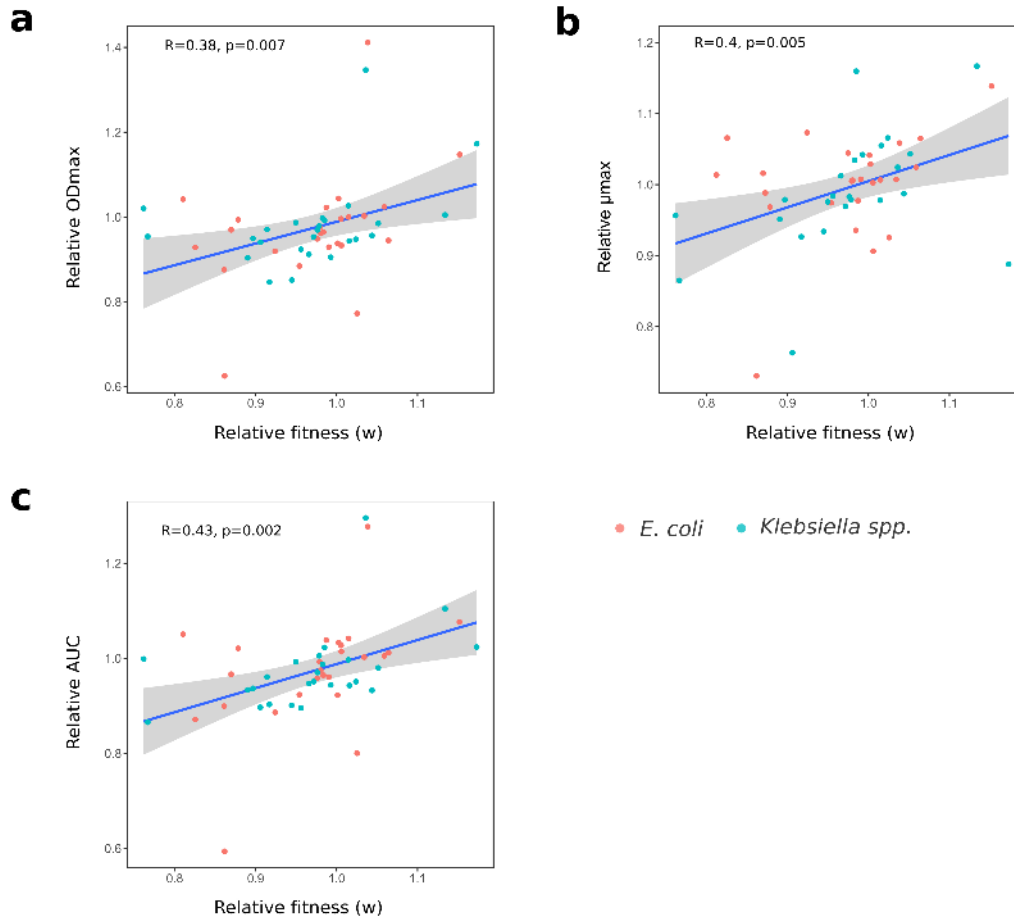
36

37 Correlation between relative growth curve parameters (a) maximum optical density (OD_{max}), (b)

38 maximum growth rate (μ_{max}), and (c) area under the growth curve (AUC), and relative fitness values

39 obtained from competition assays for each strain. The blue line represents the linear regression model

40 and the grey shading represents 95% confidence intervals. Points represent each relative value (red, *E.*

41 *coli* and blue, *Klebsiella* spp.). Pearson's correlation (R) and p-value are indicated. As expected,

42 maximum optical density, maximum growth rate and area under the growth curve are positively

43 correlated with relative fitness.

44   Supplementary Figure 4. Local Indicator of Phylogenetic Association (LIPA) analyses.
45
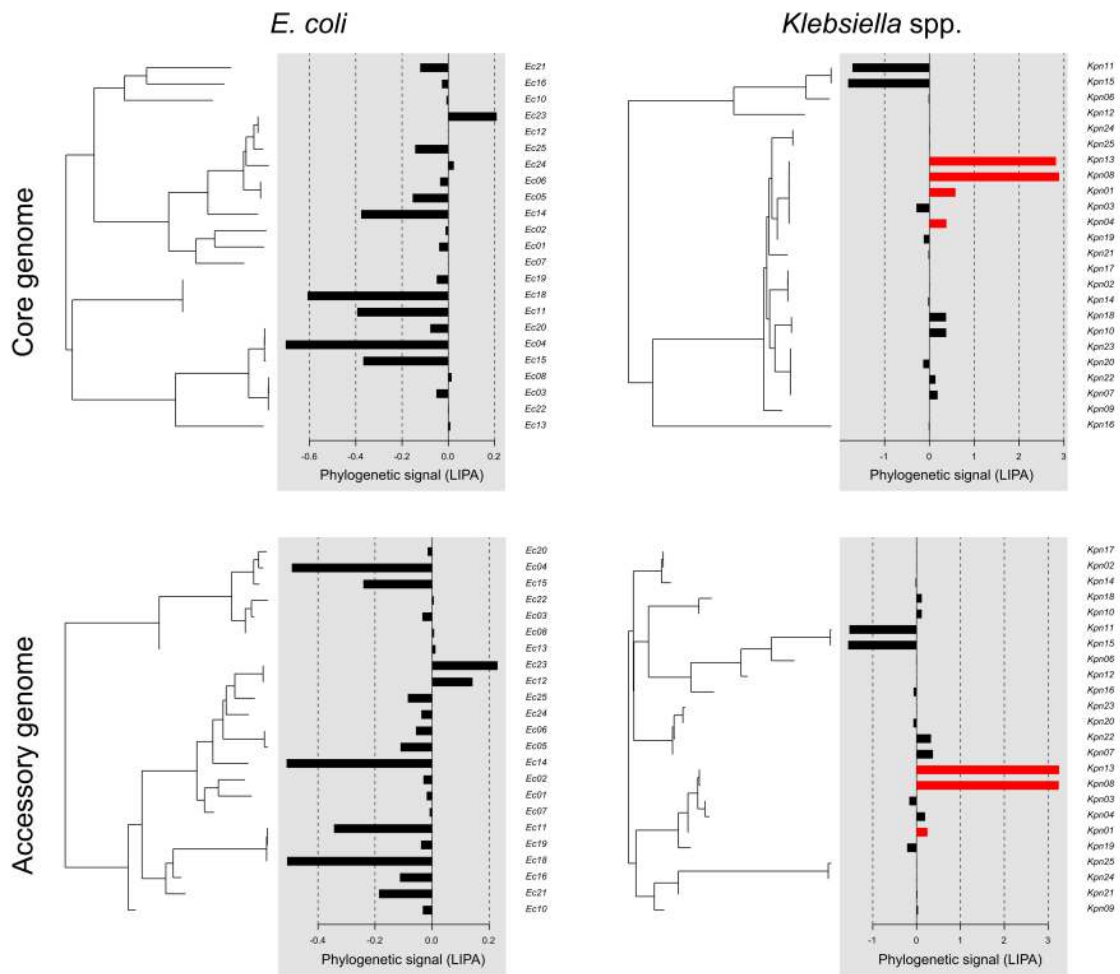
46
47

48   Phylogenetic trees for core (upper panels) and accessory (lower panels) genomes obtained for *E. coli*

49   (left) and *Klebsiella* spp. (right). Bar plots show the LIPA score associated with each tip of the

50   phylogeny, with higher values representing a stronger phylogenetic signal. Red colour indicates

51   statistically significant LIPA scores (*i.e.* phylogenetic signal, Supplementary Table 3).

52   Supplementary Figure 5. Model parametrization.
53
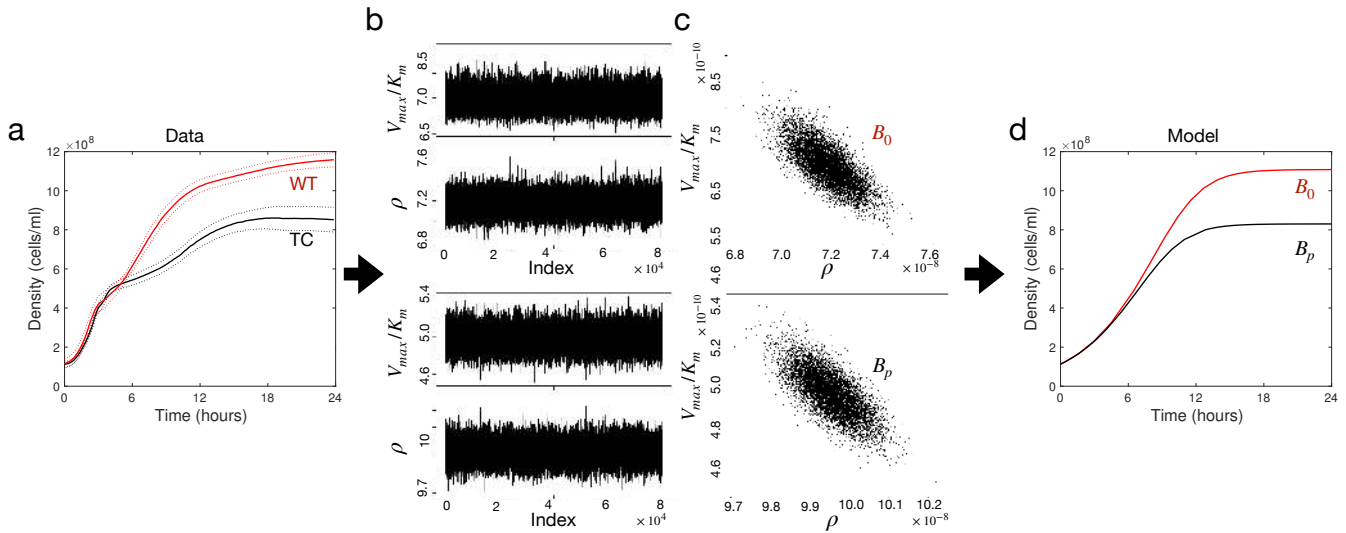
54

55   (a) Bacterial density of strain Kpn18 as a function of time obtained from the optical density of wild-type

56   (red) and transconjugant (black) strains growing in isolation. (b) Traces of chains for parameters

57   V_{max}/K_m (above) and ρ (below) obtained by fitting a simple Monod model to growth curve data using a

58   Metropolis-Hastings Markov chain Monte Carlo method (MCMC). (c) 2-dimensional posterior

59   distributions obtained for each strain (top: B_0, bottom: B_p). (d) Numerical solutions of the model using

60   parameters selected randomly from the posterior distribution and with initial conditions determined from

61   the experimental growth curves.

62   Supplementary Figure 6. *In silico* competition experiments.

63

64   Each box represents a theoretical pair-wise competition experiment between plasmid-free (open

65   circles) and plasmid-bearing cells (filled circles). The diameter of each circle is proportional to the

66   relative fraction of the population, a value estimated by numerically solving the model for $T = 24$ with

67   parameter values obtained from the posterior distribution of each strain. Horizontal axis represents the

68   specific affinity (V_{max}/K_m) and the vertical axis the cell's resource conversion rate (ρ).

69   Supplementary Figure 7. Effect of conjugation and community complexity in plasmid population

70   dynamics in the absence of a DFE.
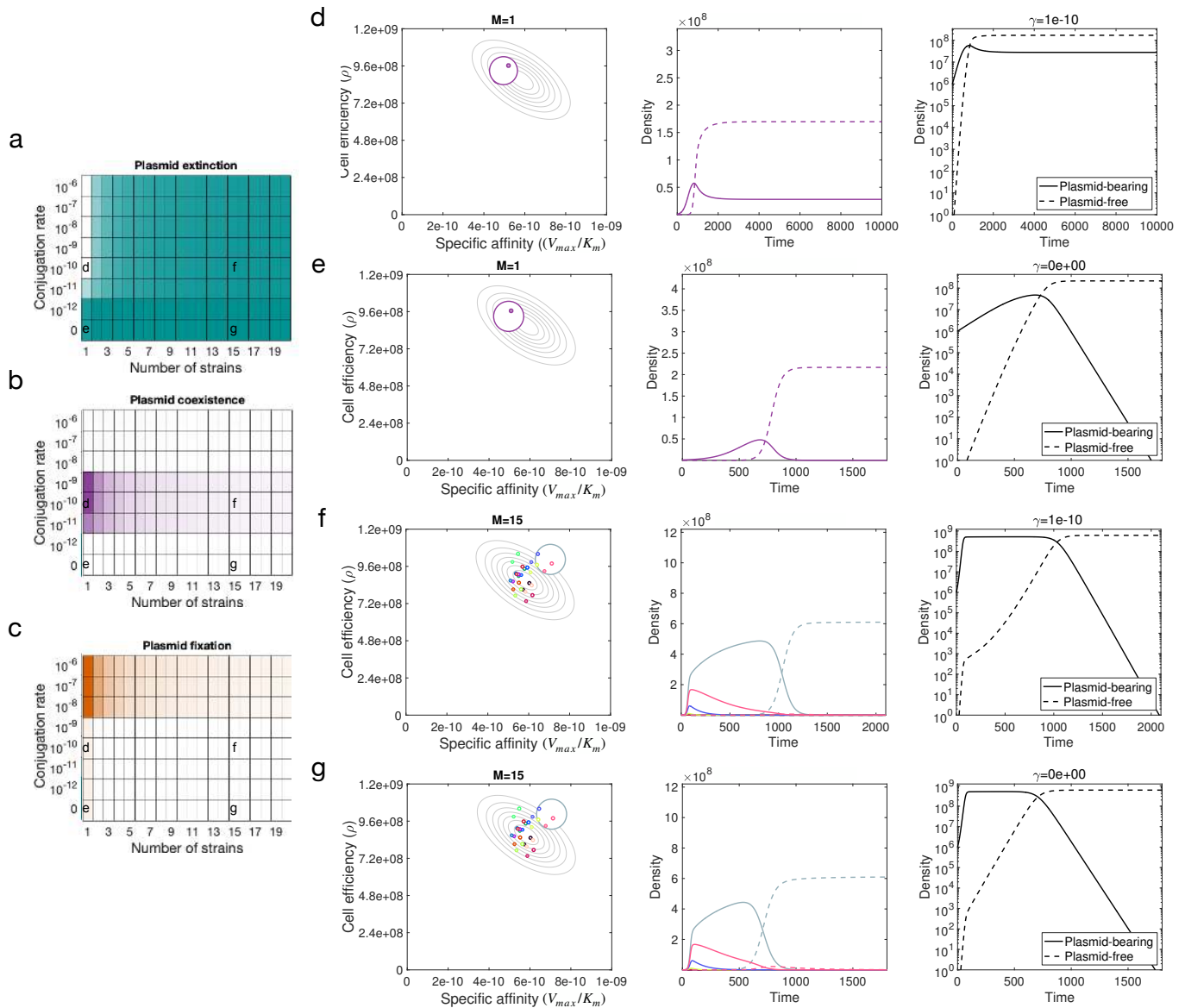
71

72   Numerical simulations of the population dynamics model performed over a range of conjugation rates

73   and number of strains in the community. In this case, all strains exhibit a reduction of fitness when

74   carrying the plasmid (a DFE with mean w= 0.985 and variance 0). The colour of each box in the grid

75   corresponds to the percentage of 5,000 random communities that exhibited: a) plasmid extinction (total

76   plasmid frequency was below a threshold), b) plasmid-bearing and plasmid-free cells co-exist in the

77   population, and c) every cell in the population carries the plasmid at the end of the experiment. (d-g)

78   Example of relative abundances over time for a range of conjugation rates in a community composed of

79   1 (d,e) and 15 (f,g) strains, with segregation rate $\lambda = 1 \times 10^{-8}$ and conjugation rate $\gamma = 10^{-10}$ (d,f) or

80    $\gamma = 0$ (e,g). The left-hand column illustrates the growth kinetic parameters for each strain (empty

8

81   circles denote plasmid-free cells and filled circles plasmid-bearing cells, with diameters proportional to

82   their final relative abundances). Middle column shows the density of each subpopulation as a function

83   of time (dotted lines denote plasmid-free strains and solid lines subpopulations carrying the

84   plasmid). Right-hand column shows semilog plots with the total fraction of cells with and without

85   plasmids (solid and dotted lines, respectively). As plasmid-bearing is associated with a fitness cost,

86   then the plasmid is only maintained in the population at high conjugation rates.

87 Supplementary Figure 8. Effect of conjugation and community complexity in plasmid population

88 dynamics in the presence of a DFE.

89

90 Numerical simulations of the population dynamics model performed over a range of conjugation rates

91 and number of strains in the community. This case corresponds to a wide DFE (mean $w = 0.985$ and

92 variance 0.007). The colour of each box in the grid corresponds to the percentage of 5,000 random

93 communities that exhibited: a) plasmid extinction (total plasmid frequency was below a threshold), b)

94 plasmid-bearing and plasmid-free cells co-exist in the population, and c) every cell in the population

95 carries the plasmid at the end of the experiment. d-g) Example of relative abundances over time for a
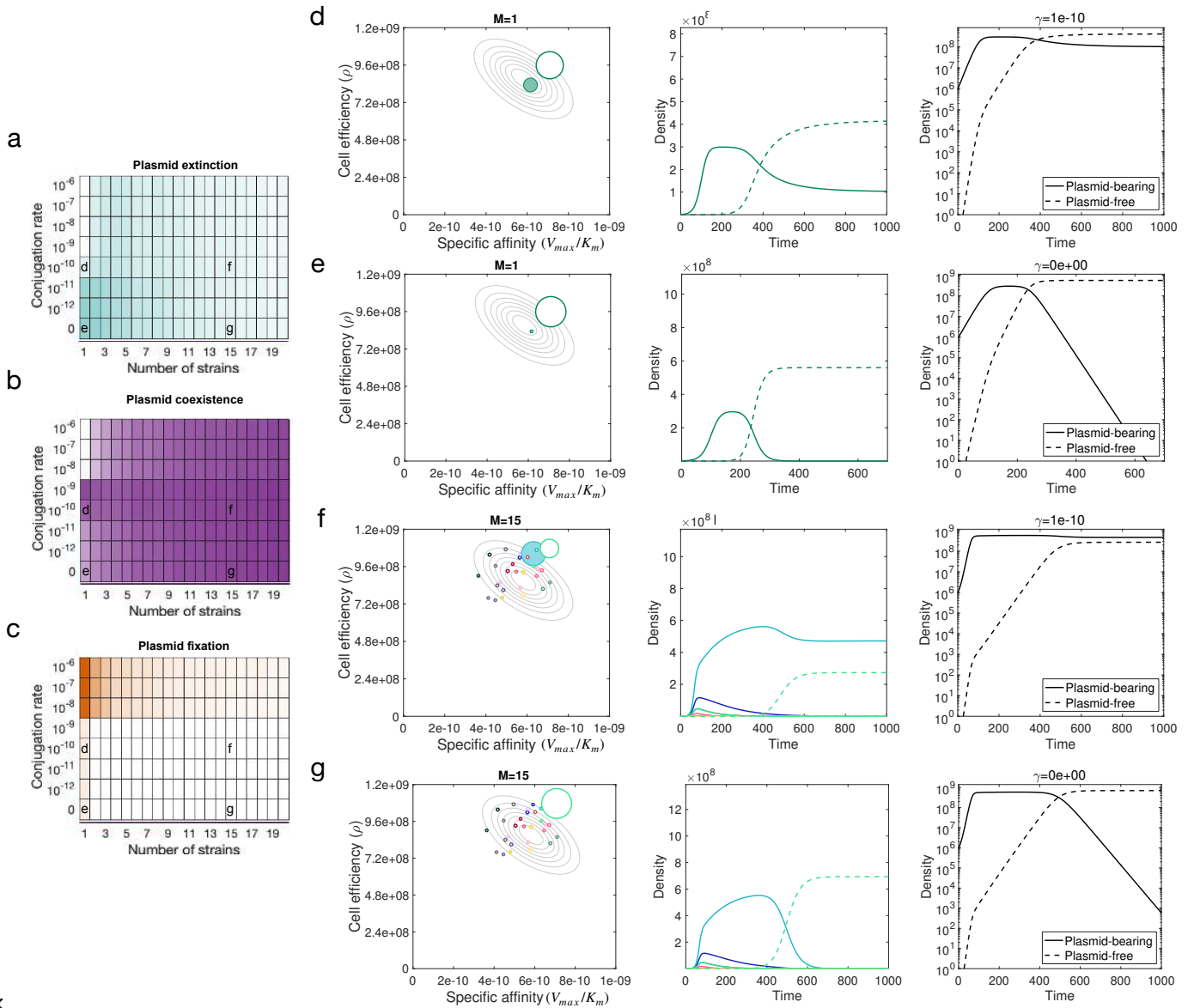
96 range of conjugation rates in a community composed of 1 (d,e) and 15 (f,g) strains, with segregation

97 rate $\lambda = 1 \times 10^{-8}$ and conjugation rate $\gamma = 10^{-10}$ (d,f) or $\gamma = 0$ (e,g). The left-hand column illustrates

98 the growth kinetic parameters for each strain (empty circles denote plasmid-free cells and filled circles

99   plasmid-bearing cells, with diameters proportional to their final relative abundances). Middle column

100   shows the density of each subpopulation as a function of time (dotted lines denote plasmid-free strains

101   and solid lines subpopulations carrying the plasmid). Right-hand column shows semilog plots with the

102   total fraction of cells with and without plasmids (solid and dotted lines, respectively). Note how a wide

103   DFE allows plasmids to persist, even at very low conjugation rates.

11

104   Supplementary Figure 9. Determination of different cells types in competition assays using flow

105   cytometry.
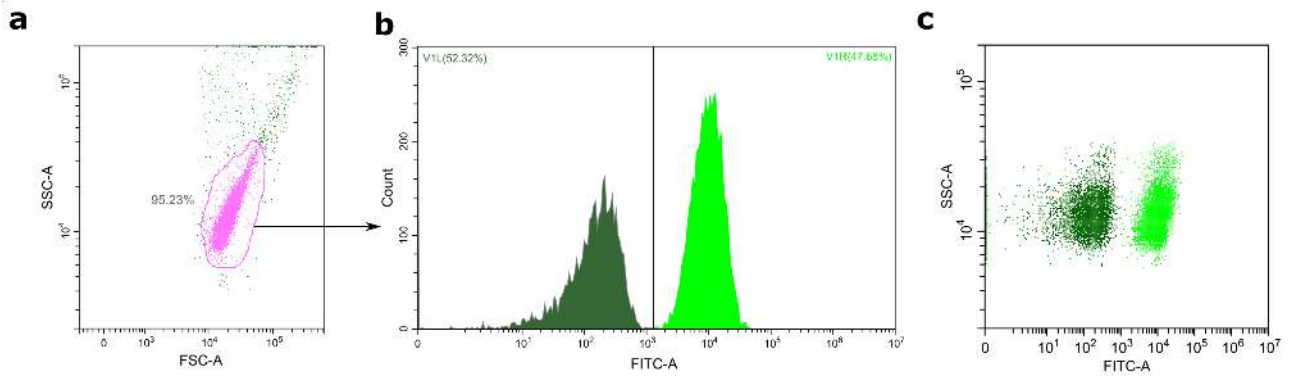
106

107   We used flow cytometry to differentiate between GFP-producing and -non-producing cells. (a) we used

108   forward versus side scatter (FSC vs SSC) gating to identify bacterial cells in the sample. (b-c) GFP-

109   producing (bright green) and -non-producing (dark green) cells were differentiated using the FITC-A

110   (fluorescein isothiocyanate) channel, allowing us to measure the proportion of each competitor in the

111   mix.

112   Supplementary Figure 10. Distribution of pBGC fitness effects.

113
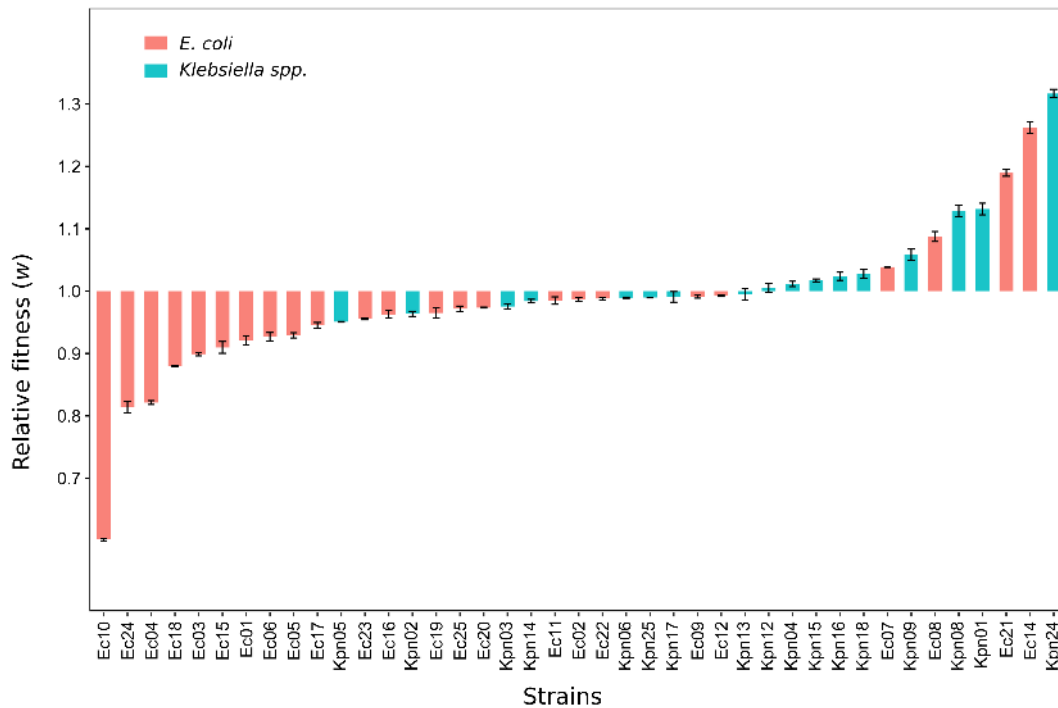
114

115

116   Relative fitness (w) of pBGC-carrying clones compared to plasmid-free clones, obtained from

117   competition assays (red, *E. coli* and blue, *Klebsiella* spp.). Values below 1 indicate a reduction in w and

118   values above 1 indicate an increase in w due to pBGC acquisition. Bars represent the average of five

119   independent experiments and error bars represent the standard error of the mean. Note that the fitness

120   effects of pBGC did not correlate with those form pOXA-48 (Pearson's correlation, R= 0.11, t= 0.66, df=

121   39, *P*= 0.51).

13

122   Supplementary Figure 11. Correlation between relative fitness values calculated in competitions vs. *E.*

123   *coli* J53/pBGC or isogenic clones with pBGC.

124
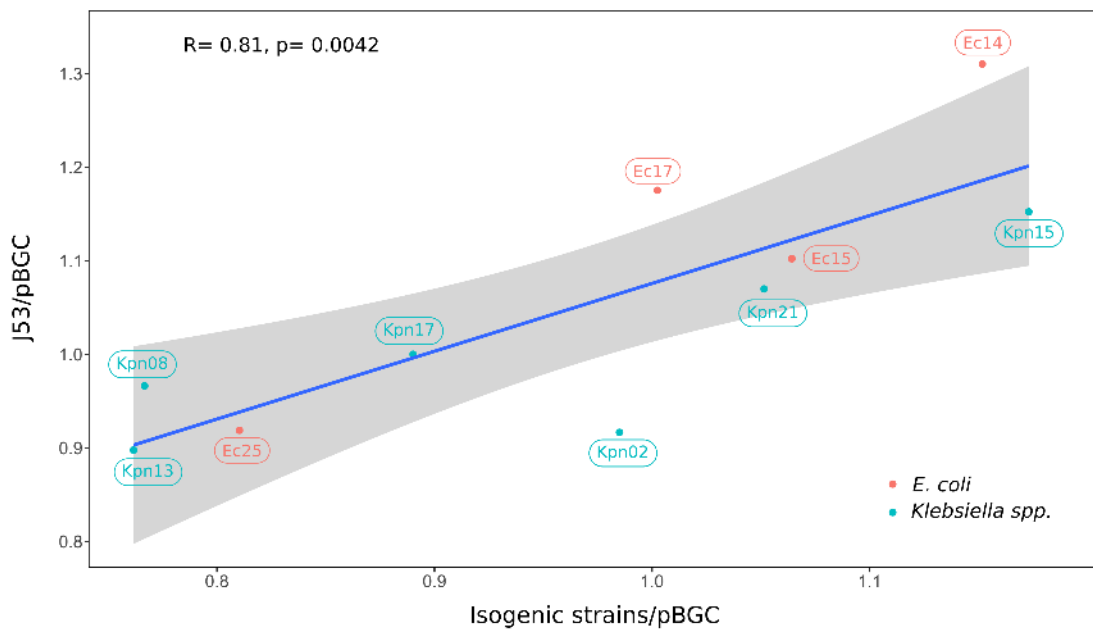
125
126

127   Correlation between relative fitness values obtained from competitions assays using pBGC-carrying

128   isogenic isolates and pBGC-carrying *E.* coli J53 for ten different isolates. The blue line represents the

129   linear regression model and the grey shading represents 95% confidence intervals. Blue points

130   correspond to *Klebsiella* spp. isolates and red points to *E. coli* isolates. Labels indicate isolates names.

131   Pearson's correlation (R) and p-value are indicated.

14

132   Supplementary Table 4. Primers used in this study.

Primers		
Name	**Sequence 5'->3'**	**Use**
Oxa-48 Fw	TTGGTGGCATCGATTATCGG	Amplification of *bla*$_{TEM1}$ gene
Oxa-48 Rv	GAGCACTTCTTTTGTGATGGC	
IncL Fw	CGGAACCGACATGTGCCTACT	Amplification of *repC* gene
IncL Rv	GAACTCCGGCGAAAGACCTTC	
pBGC Fw	CGTTGATCGGCACGTAAG	Amplification of pBGC backbone for Gibson cloning (see Supplementary Figure 2)
pBGC Rv	GCTGTCTAGACTATTTGTATAGTTCATCCATGC	
GFP-Term Fw	atacaaatagtctagacagcGGGAATCCTGCTCTGCGAG[1]	
GFP-Term Rv	ctcttacgtgccgatcaacgGGGTTATTGTCTCATGAGCGG[1]	
pBGC_Seq1_Fw	AGTTAAAAGGTATTGATTTTAA	pBGC sequencing (3503-3524)[2]
pBGC_Seq1_Rv	GCCACATCTTGCGAATA	pBGC sequencing (464-480)[2]
pBGC_Seq2_Fw	ATAAGATCACTACCGGGC	pBGC sequencing (44-61)[2]
pBGC_Seq2_Rv	ACCCGACAGGACTATAAAGATA	pBGC sequencing (1285-1306)[2]
pBGC_Seq3_Fw	GAGGTAACTGGCTTGGAGG	pBGC sequencing (931-949)[2]
pBGC_Seq3_Rv	GTCGCGTCTGTCACATCT	pBGC sequencing (2124-2141)[2]
pBGC_Seq4_Fw	GTTTCCCGACTGGAAAGC	pBGC sequencing (1703-1720)[2]
pBGC_Seq4_Rv	CTTTGGTCCCGCTTTGTTAC	pBGC sequencing (2894-2913)[2]
pBGC_Seq5_Fw	GTCGGTCGATAAAAAAATCGAG	pBGC sequencing (2650-2671)[2]
pBGC_Seq5_Rv	ATGTGGTCTCTCTTTTCGTTGG	pBGC sequencing (3764-2671)[2]

133

134   1. Lower case nucleotides correspond with the added cohesive ends.

135   2. Numbers correspond with primers binding sites in pBGC (according to sequence available with

136   GenBank Accession Number MT702881).

15

137 Supplementary Table 5. Growth kinetic parameters of each strain obtained using the MCMC algorithm.

Strain	Wild-type		Transconjugant	
	V_{max}/K	ρ	V_{max}/K	ρ
Ec01	4.895×10^{-10}	9.844×10^{8}	5.283×10^{-10}	8.776×10^{8}
Ec02	7.603×10^{-10}	5.504×10^{8}	6.648×10^{-10}	6.786×10^{8}
Ec03	7.597×10^{-10}	8.900×10^{8}	6.538×10^{-10}	8.151×10^{8}
Ec04	8.373×10^{-10}	7.764×10^{8}	6.576×10^{-10}	7.702×10^{8}
Ec05	6.299×10^{-10}	7.727×10^{8}	7.831×10^{-10}	6.416×10^{8}
Ec06	6.852×10^{-10}	9.518×10^{8}	6.839×10^{-10}	6.567×10^{8}
Ec07	3.786×10^{-10}	1.040×10^{9}	4.154×10^{-10}	1.013×10^{9}
Ec08	5.304×10^{-10}	1.050×10^{9}	5.053×10^{-10}	9.900×10^{8}
Ec09	6.610×10^{-10}	9.195×10^{8}	7.311×10^{-10}	8.694×10^{8}
Ec10	5.660×10^{-10}	1.071×10^{9}	7.368×10^{-10}	8.472×10^{8}
Ec11	7.248×10^{-10}	5.064×10^{8}	7.248×10^{-10}	6.483×10^{8}
Ec12	5.200×10^{-10}	8.042×10^{8}	4.752×10^{-10}	8.016×10^{8}
Ec13	6.831×10^{-10}	8.528×10^{8}	7.711×10^{-10}	7.932×10^{8}
Ec14	6.515×10^{-10}	6.589×10^{8}	8.356×10^{-10}	7.500×10^{8}
Ec15	6.322×10^{-10}	9.106×10^{8}	6.998×10^{-10}	8.528×10^{8}
Ec16	4.908×10^{-10}	1.122×10^{9}	5.550×10^{-10}	1.040×10^{9}
Ec17	6.089×10^{-10}	8.298×10^{8}	6.366×10^{-10}	8.647×10^{8}
Ec18	6.655×10^{-10}	8.963×10^{8}	6.939×10^{-10}	6.250×10^{8}
Ec19	6.080×10^{-10}	9.812×10^{8}	6.200×10^{-10}	9.424×10^{8}
Ec20	5.033×10^{-10}	1.024×10^{9}	5.296×10^{-10}	9.133×10^{8}
Ec21	5.546×10^{-10}	1.021×10^{9}	4.897×10^{-10}	8.386×10^{8}
Ec22	6.148×10^{-10}	9.736×10^{8}	6.619×10^{-10}	9.886×10^{8}
Ec23	5.443×10^{-10}	9.635×10^{8}	5.186×10^{-10}	9.094×10^{8}
Ec24	6.010×10^{-10}	1.008×10^{9}	4.808×10^{-10}	9.768×10^{8}
Ec25	6.784×10^{-10}	9.237×10^{8}	5.465×10^{-10}	1.004×10^{9}
Kpn01	4.841×10^{-10}	8.746×10^{8}	3.652×10^{-10}	8.673×10^{8}
Kpn02	6.638×10^{-10}	7.905×10^{8}	6.352×10^{-10}	7.952×10^{8}
Kpn03	4.802×10^{-10}	9.639×10^{8}	3.686×10^{-10}	9.975×10^{8}
Kpn04	4.844×10^{-10}	9.288×10^{8}	4.303×10^{-10}	8.936×10^{8}
Kpn05	5.151×10^{-10}	9.303×10^{8}	5.155×10^{-10}	8.513×10^{8}
Kpn06	5.240×10^{-10}	8.952×10^{8}	5.664×10^{-10}	7.742×10^{8}
Kpn07	4.498×10^{-10}	8.682×10^{8}	3.785×10^{-10}	9.458×10^{8}
Kpn08	7.163×10^{-10}	8.211×10^{8}	4.334×10^{-10}	7.188×10^{8}
Kpn09	6.207×10^{-10}	8.064×10^{8}	7.302×10^{-10}	7.772×10^{8}
Kpn10	4.833×10^{-10}	7.990×10^{8}	4.467×10^{-10}	7.967×10^{8}
Kpn11	5.380×10^{-10}	9.710×10^{8}	5.620×10^{-10}	8.490×10^{8}
Kpn12	5.620×10^{-10}	8.490×10^{8}	5.620×10^{-10}	8.490×10^{8}

Kpn13	7.086×10^{-10}	9.195×10^{8}	7.364×10^{-10}	7.206×10^{8}
Kpn14	4.162×10^{-10}	8.994×10^{8}	5.828×10^{-10}	7.306×10^{8}
Kpn15	6.243×10^{-10}	8.854×10^{8}	7.489×10^{-10}	8.917×10^{8}
Kpn16	5.386×10^{-10}	8.523×10^{8}	4.359×10^{-10}	8.465×10^{8}
Kpn17	4.144×10^{-10}	1.037×10^{9}	4.002×10^{-10}	9.786×10^{8}
Kpn18	6.906×10^{-10}	7.173×10^{8}	4.977×10^{-10}	9.954×10^{8}
Kpn19	5.578×10^{-10}	8.319×10^{8}	6.083×10^{-10}	7.739×10^{8}
Kpn20	4.857×10^{-10}	1.022×10^{9}	4.287×10^{-10}	1.046×10^{9}
Kpn21	4.674×10^{-10}	1.077×10^{9}	5.487×10^{-10}	9.948×10^{8}
Kpn22	4.708×10^{-10}	1.006×10^{9}	5.224×10^{-10}	1.006×10^{9}
Kpn23	7.897×10^{-10}	6.709×10^{8}	7.201×10^{-10}	6.328×10^{8}
Kpn24	5.736×10^{-10}	7.568×10^{8}	8.670×10^{-10}	6.601×10^{8}
Kpn25	4.873×10^{-10}	9.630×10^{8}	4.255×10^{-10}	9.896×10^{8}

138