Molecular kinetics dictate population dynamics in CRISPR-based plasmid defense

Luke Richards †, Danna Lee †, Jakub Wiktor, Johanna Cederblad, Daniel Jones*

Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 Box 596, 75124, Uppsala, Sweden.

*Corresponding author(s). E-mail(s): daniel.jones@icm.uu.se;
†These authors contributed equally to this work.

Abstract

Understanding and manipulating the complexity of microbial community diversity, including the mobile genetic elements contained within, represents a great challenge that has wide-ranging potential benefits across synthetic biology, agriculture and medicine. An important component to this complexity is the acquisition of genetic material via conjugative plasmids, which can encode new functions such as antibiotic resistance, and the degradation of those plasmids by CRISPR interference. In this work we use single-cell tracking of *E. coli* populations in microfluidic devices coupled with high resolution fluorescence microscopy to characterize plasmid dynamics at the single-cell and population level. On a population level we find that the ability of cells to clear plasmids is highly dependent on both the number of spacer targets present and the defense expression level. Additionally we assess the impact that counter-defense mechanisms such as plasmid addiction have on plasmid population dynamics and show that CRISPR may be an ineffective method to counter plasmids with such mechanisms. By leveraging single-cell tracking, we were also able to report conjugation rate per neighboring donor cell, estimate the latent period between plasmid uptake and subsequent onward transmission, and corroborate existing estimates of cascade search time. This synthesis of population and single-cell measurements suggests that plasmids are the subject of a dynamic tug-of-war between defense expression, spacer distribution, neighboring cell identity and plasmid cost-benefit tradeoffs. The use of imaging and analysis techniques used here and subsequent multi-scale measurements will facilitate the disentanglement of how these factors coordinate to realize community-wide plasmid dynamics in diverse contexts.

Keywords: CRISPR, plasmid conjugation, population dynamics, microfluidics, single-cell tracking

1 Introduction

The importance and ubiquity of microbial communities across society cannot be understated. As well as being disease-causing agents, they are also responsible for important nutrient cycling functions in soils [1], and can be a valuable industrial resource for the engineering of synthetic communities. Microbial communities have an extremely complex taxonomic, spatial and functional composition [2, 3], with an added level of complexity driven by their propensity to share genetic material (and associated functions) via horizontal gene transfer (HGT). Plasmids, often transferred via Type IV secretion systems (T4SS) [4], are a form of HGT and can bestow beneficial functions on their hosts. They can encode new metabolic functions [5, 6] or, perhaps most notably, antibiotic resistance. The projected death toll incurred by antibiotic resistance is in the millions and climbing [7].

In contrast to the potentially beneficial features plasmids can confer to their hosts, conjugative plasmids often impose a fitness cost on host cells[8–11]. Thus, harboring conjugative plasmids is not purely beneficial to host cells, and bacteria have consequently evolved an array of defense systems against conjugative plasmids or mobile genetic elements (MGEs) more generally [12]. Equally, those mobile elements have evolved strategies to evade such defense systems [13]. The cost of plasmid acquisition and maintenance must be balanced against the cost of defending against plasmid acquisition and the current - or indeed potential future - benefit of the encoded functions.

CRISPR defense is widely recognized as a method for microbes to protect against invading genetic material [14] with CRISPR systems present in around 75% of archeal isolates and around 36% of bacteria [15] but the evidence for CRISPR-Cas systems' effectiveness is mixed [16]. In general, CRISPR appears to be used by other bacteria to target plasmids for degradation although perhaps less often than phages; in a survey of Pseudomonas species, 30.52% spacers were found to target phages and 5.61% plasmids [17]. CRISPR demonstrably provides defense against phage in *P. aeruginosa*[18]. Contrastingly, a genetic study across many bacterial taxa found a more significant negative correlation between CRISPR genes and plasmid absence than between CRISPR genes and phage absence, suggesting a stronger effect against plasmids [19]. Another study found no connection for other measures of HGT such as presence of prophage genes and recently acquired genes [20] but a negative correlation between multidrug resistance (often transferred by HGT) and CRISPR defense has been shown in *Enterococci sp.*[21].

Wild-type *E. coli* K-12 contains a type I-E CRISPR-Cas system, comprised of a protein complex called Cascade which performs RNA-guided search for target DNA; upon target recognition, Cascade recruits the helicase-nuclease protein Cas3 to degrade target DNA. *E. coli* K-12 generates very low levels of CRISPR-RNA suggesting that

this system is perhaps not active [22] under known environmental conditions, and indeed, in the case of *E. coli*, CRISPR interference seems ineffective against mobile genetic elements [23]. Nonetheless, K12 Cascade is effective in generating CRISPR-RNAs when Cascade and Cas3 are expressed [22] meaning that this model organism remains an important tool for understanding CRISPR-MGE dynamics. On the other hand, experimental measurements performed in *E. coli* with this system have confirmed a picture whereby plasmids can persist over many generations in cell lineages despite being targeted for CRISPR interference [24]. This raises questions about the functionality and effectiveness of CRISPR-based defense against conjugative plasmids but could be the product of an adaptive mechanism to maintain plasmid frequency at low levels as a way of 'bet-hedging'. This low-frequency feature of broad host range plasmids has been reviewed and discussed as a method for dealing with environmental uncertainty [25].

Studies artificially inducing and manipulating CRISPR components show a clear relationship between the level of induction of Cascade complexes (responsible for searching for and identifying CRISPR target sequences) and the clearance of plasmids from the population; leading to an estimated search time per Cascade complex of around 90 minutes [26]. Stochasticity in the single-cell expression level of Cascade has also been shown to affect CRISPR effectiveness [27], while highly expressing CRISPR systems also carries costs in terms of expression burden, DNA damage and replication interference [28, 29]. In fact, growth rate of cells in and of itself has been implicated in cells' ability to degrade plasmids [27]. Taken together these studies indicate a complex interplay between, defense expression, growth, clearance and cost benefit trade-offs at a single cell level which play out to become population level changes.

Theoretical studies of plasmid conjugation and population dynamics often take the form of "mass action" ordinary diferential equation (ODE) models describing the interactions of well-mixed bacteria in homogeneous environments[30–32]. However, real-life bacterial habitats are often characterized by spatial constraints, physical inhomogeneity, small population sizes, and significant cell-to-cell variability. These factors are key contributors to population dynamics and demand an understanding of spatially-resolved single-cell interactions. A recent work identified small, confined environments as a key contributor to plasmid persistence under circumstances under which it would theoretically not persist[33]. Furthermore, confinement has been shown to increase conjugation rates[34]. Cell-to-cell variability (or "noise") is an inescapable feature of gene expression which may have potential adaptive functions in the dynamics of mobile genetic element defense mechanisms[27]. Such variability also allows many "experiments within an experiment"; enabling single-cell characterization of CRISPR interference and cell growth across a range of Cascade expression levels.

Here, we investigate the dynamics of plasmid conjugation and CRISPR interference at both the population and the single-cell level in *E. coli*, using microfluidics coupled with time-lapse fluorescence microscopy. Use of these techniques allows spatiotemporal-resolution and classification of every cell and conjugation event within populations, facilitating measurement of key biophysical factors in the conjugation - CRISPR dynamic. We characterize these dynamics at both the population and single-cell level. At the population level we find a clear increase in resistance of the

population to the conjugative plasmid when CRISPR target number increases on the plasmid. This also has profound consequences for the population depending the presence of a plasmid addiction system. With the help of single-cell tracking and classification, we are able to report, for the first time, conjugation rates per neighboring plasmid-containing cell and estimate the latent period between plasmid acquisition and subsequent onward transmission. We also find that cell-to-cell variability in CRISPR-Cascade expression significantly affects dynamics at the single-cell level, inducing a growth cost at high expression levels where the clearance rate is highest. The insights presented here are vital for the improvement of our understanding, and potential future manipulation, of conjugation dynamics across a range of natural and synthetic microbial communities. This can have potentially far-reaching applications across medicine, industry and agriculture.

2 Results

2.1 Population dynamics across CRISPR interference levels

2.1.1 Increasing target number effectively increases the effectiveness of CRISPR targeting

In order to visualize the single-cell dynamics of conjugation and CRISPR interference, we engineered *E. coli* K-12 MG1655 to place fluorescent labels on the CRISPR Cascade complex and the conjugative plasmid RP4 (Fig. 1A) for time-lapse imaging in microfluidic devices (Supplementary Fig. S1). Plasmid donor cells harbored the RP4 conjugative plasmid, modified to contain a *parSMT1* site as well as 0, 1, or 18 protospacers from the *E. coli* Type I-E CRISPR arrays. Recipient cells were modified by genomic insertion of mCherry-ParB into the *gtrA* locus, under control of a constitutive promoter. (See Methods for detailed information about cloning and strain construction.) Upon conjugation into a recipient cell, mCherry-ParB binds cooperatively to the *parSMT1* site in the modified RP4 plasmid, forming a bright fluorescent focus (Fig. 1A). Recipient cells were also modifed by fusing the fluorescent protein sYFP2 to Cas8e in order to visualize Cascade levels in single cells. Finally, the wild-type Pcascade and Pcas3 promoters were exchanged for araBp8 and lacUV5 promoters, respectively, enabling induction of Cascade and Cas3 expression by arabinose and IPTG, respectively [35]. These modifications enabled classification of detected cells as donor, recipient, and transconjugant depending on their fluorescent properties (Fig. 1A&B). RP4 also contains a plasmid addiction mechanism: the *parDE* toxin antitoxin system [36]. In order to assess the dynamics of CRISPR interference in the absence of such a confounding factor the ParE protein was also removed in the modified RP4.

To start experiments, a 50/50 mixed population of donor and recipient cells was loaded into the microfluidic device and time-lapse imaging was performed in phase contrast, mCherry, and sYFP2 channels for 6 hours. Observing the fraction of the total population corresponding to each of these groups presents an overview of population dynamics over time for each CRISPR interference level (Fig. 1C). For the non-targeted RP4-0-target variant, the population is rapidly and comprehensively

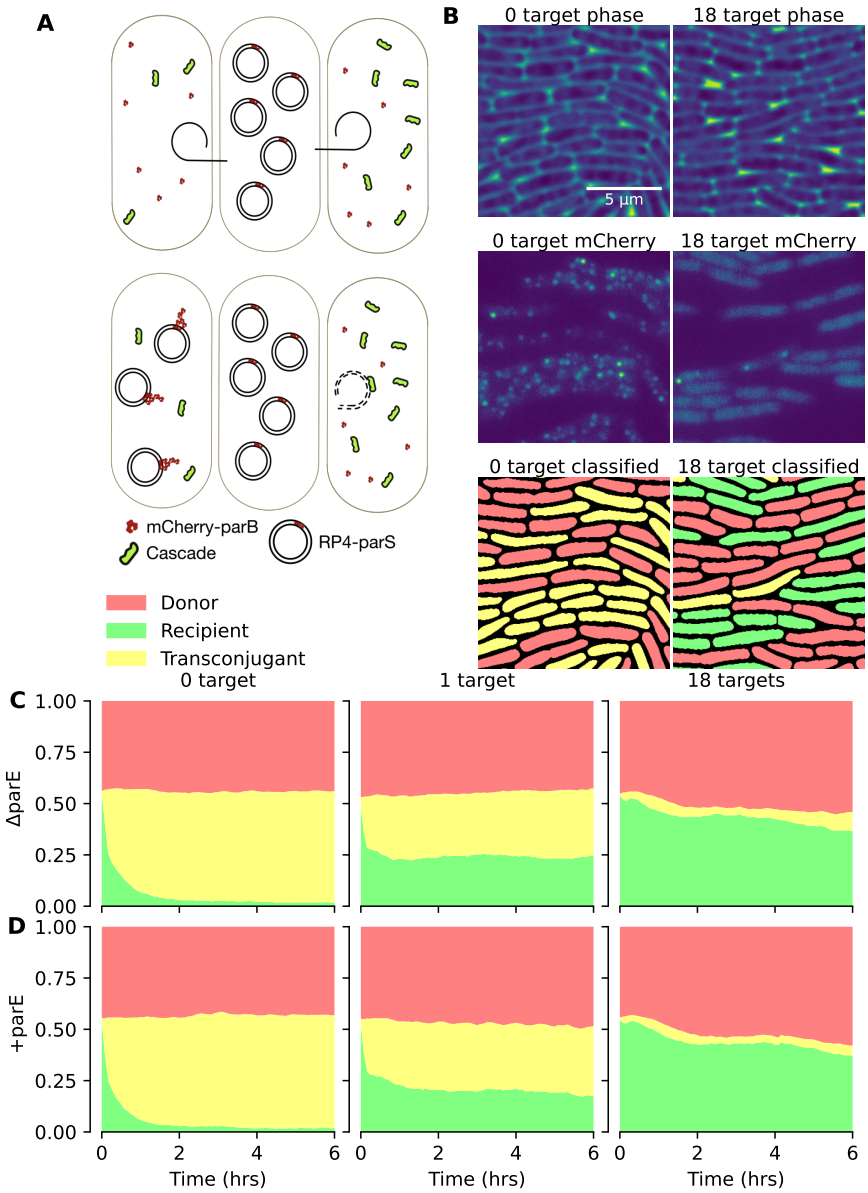

Fig. 1 A) Donor cells containing parSMT1-containing RP4 plasmids transfer plasmids via conjugation into fluorescently labeled recipient cells. Recipient cells contain mCherry-tagged ParB and sYFP2-tagged Cascade to allow tracking of conjugation and Cascade expression levels. B) Corresponding phase contrast (top), mCherry fluorescence (middle), and classified segmented cell (bottom) images for 0 target and 18 target RP4 variant experiments. C) Mean fraction of the cell population classified as donor, recipient or transconjugant across the course of microfluidic experiments with RP4 plasmid variants. At least 8000 individual cells were observed for each time point.

converted to transconjugants, reflecting RP4's high conjugation rate. For the RP4-1-target variant, there is a moderate increase in plasmid clearance, resulting in a higher steady-state fraction of transconjugants and recipients. Finally, the RP4-18-target variant containing all 18 protospacers displays a high susceptibility to targeting by CRISPR interference, with only a very small proportion of the population succumbing to conjugation.

To investigate the potential effectiveness of CRISPR interference in defending against plasmids making use of addiction systems, we also performed microfluidic experiments in parallel using the RP4 variants with ParE (Fig. 1D). We hypothesized that, in this scenario, cleared cells will suffer a fitness cost and be unable to persist as effectively in the population. For *parE*-containing plasmids we see broadly similar dynamics across CRISPR interference levels. In these experiments specifically, the level of Cascade induction is high and immediate; conjugation is never allowed to proceed in the absence of CRISPR defense. We hypothesized that, under these conditions, the toxin may not act effectively, owing to a lack of time to accumulate inside cells before clearance occurs.

2.1.2 Toxin-antitoxin systems can be effective counter-defense against CRISPR interference.

To assess the impact of plasmids being established in cells before CRISPR interference is induced, 18-target donor cells were introduced into microfluidic experiments mixed with non-induced recipients cells. For the first hour of imaging the conjugation was allowed to proceed in the absence of CRISPR induction at which point the media-in channel was switched to media containing inducers (Fig. 2). Around the 70-80 minute time point the clearing of transconjugants is apparent in both the +*parE* (right) and Δ*parE* (left) experiments. Additionally in the case of +*parE* there is also a marked decrease in the fraction of the population made up of recipients (Fig. 2A). We also observe the expected increase in Cascade expression concomitantly with the onset of clearing after inducer introduction (Fig. 2). In Fig. 2C we show representative images of segmented and classified cells at the end of the experiments. Experiments with *parE* present, show elongated, dying recipient (cleared transconjugant) cells whereas without *parE* the cells clear and remain healthy in appearance. Control experiments were also performed where the recipient cells had CRISPR arrays deleted. In these control experiments cells were never cleared of plasmids (Fig. S2). Taken together, these observations lead to the corollary that CRISPR interference may be an ineffective mechanism to protect against unwanted conjugative plasmids that make use of *parDE*, and indeed wider toxin-antitoxin, systems. The results also highlight the important of timing of MGE defence system expression [37].

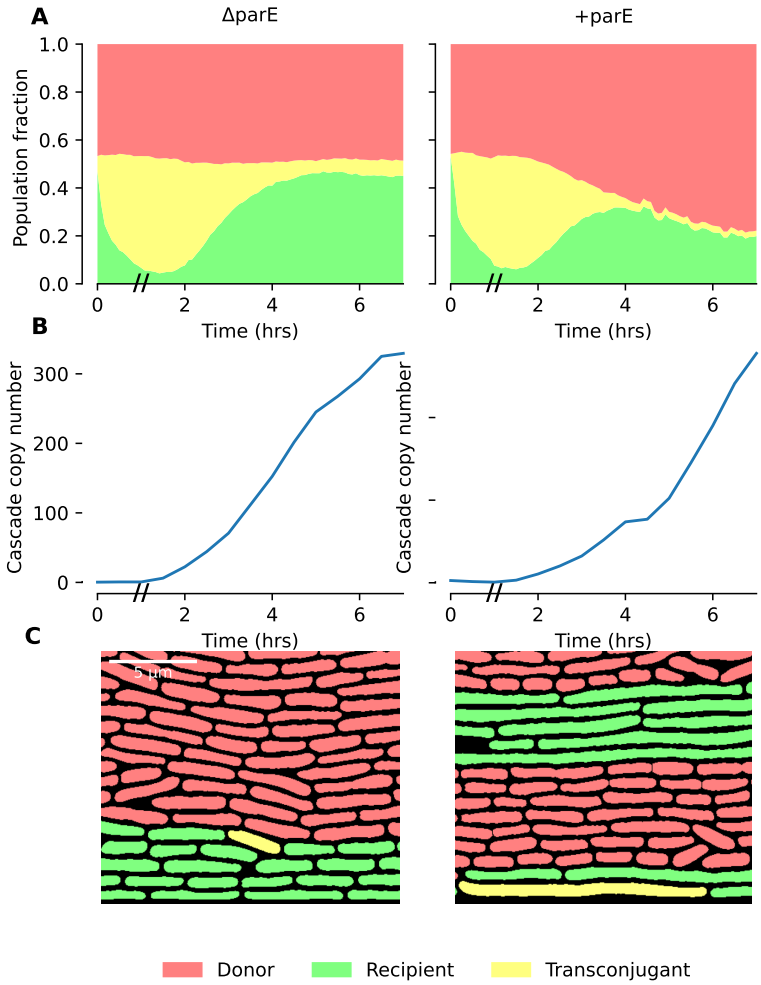

Fig. 2 A) Mean fraction of the cell population classified as donor, recipient or transconjugant across the course of microfluidic experiments with delayed Cascade induction with 18-target RP4 plasmids, with and without *parE*. B) Mean Cascade copy number per cell across the course of the experiment. Inducers were added to the media at 60 minutes, indicated with // on axis. C) Representative segmented, classified cell images at the end of +*parE* (right) and Δ*left* (left) experiments. At least 5000 individual cells were observed for each time point.

2.2 Single-cell characterization of conjugation kinetics

2.2.1 Conjugation proceeds with a single rate-limiting step at a rate of two conjugation events/hour/donor neighbor

The population dynamics shown in Figs. 1 and 2 reflect the interplay of the opposing forces of plasmid conjugation and CRISPR interference. Next, we sought to characterize each of these processes separately at the single-cell level, focusing first on conjugation. In Fig. 3, we focus our attention on how quickly recipient cells are "converted" to transconjugants as a function of how many donor neighbors they have. In particular, in Fig. 3A, we plot the the number of transconjugants T divided by the number of transconjugants T plus the number of recipients R. This quantity, $T/(T+R)$, is the fraction of recipients that have been converted to tranconjugants. We plot this fraction for different subsets of the population based on how many donor neighbors each recipient or transconjugant has. Neighbors were computed using morphological operations on segmentation masks (see Methods). Perhaps unsurprisingly, recipients with more donor neighbors are conjugated more rapidly than recipients with fewer donor neighbors. As seen in Figs. 3A and 3B, the presence or absence of *parE* does not appear to affect conjugation dynamics.

To further characterize the quantitative dynamics of conjugation, we focused our attention on RP4-0-target donors, with and without *parE*, which are not targeted for CRISPR interference. For each of the curves with one or more donor neighbors, we fit an exponential function of the form

$$\frac{T}{T+R} = 1 - e^{-k_{conj}t} \tag{1}$$

where k_{conj} represent the conjugation rate for a given number of donor neighbors. In Fig. 3B, we plot k_{conj} for donor neighbor counts from 1 to 5 and find that conjugation rate is linearly proportional to donor neighbor count until about four donor neighbors, at which point the increase appears to saturate. Thus, conjugation from donor neighbors appears to be additive, at least for roughly one to four donor neighbors. For single donor neighbors, we obtained a value of $k_{conj} = 2.4\,\mathrm{h}^{-1}$, which implies that recipients with a single donor neighbor will be conjugated after $1/2.4 \approx 0.42\,\mathrm{h}$ or 25 minutes on average.

2.2.2 New transconjugants experience an hour-long latent period before performing conjugation

In Fig. 3A we also plotted $T/(T+R)$ for cells with zero donor neighbors. These curves generally do not show an immediate increase from $t = 0$, but eventually start to rise after a delay of approximately 30 minutes. We hypothesized that this delay reflects the latent period required for newly-conjugated transconjugants to produce conjugation machinery and become able to perform conjugation themselves. However, as these plots are based on the number of donor-neighbors in a particular frame only, we cannot rule out the possibility that some of these cells may have transiently come into contact

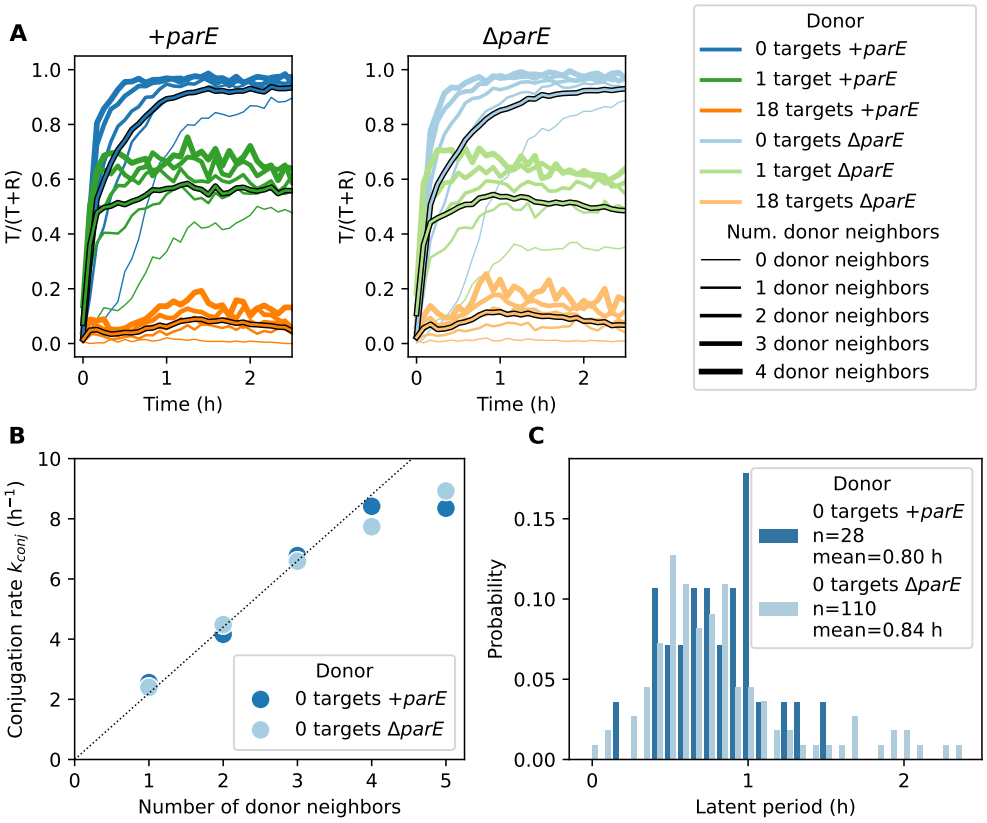

Fig. 3 Single-cell characterization of conjugation dynamics. (A) Fraction of recipients that have been converted to transconjugants, as a function of time and number of donor neighbors (different line thicknesses). This fraction is given by the ratio $T/(T+R)$ where T and R are the numbers of transconjugants and recipients, respectively. As expected, the fewer target sites on the donor plasmid, the more likely recipients are to become transconjugants (donor plasmid denoted by color as indicated in legend). Similarly, cells with more donor neighbors are more likely to become transconjugants. At least 1595 cells were analyzed for each donor and timepoint (2975 cells on average). **(B)** Dependence of conjugation rate on number of donor neighbors. Conjugation rates were determined by fitting exponential functions (Equation 1) to the 0 target donor curves shown in (A) for nonzero donor neighbors. The conjugation rate is directly proportional to the number of donor neighbors, highlighting the effects of local spatial inhomogeneities. The mean numbers of donor neighbors during the first hour of the experiment were 2.0 and 1.9 for $+parE$ and $\Delta parE$ RP4-0-target donors, respectively. **(C)** Latent period for conjugation, defined as the length of time from when a recipient becomes a transconjugant until the new transconjugant itself performs conjugation.

with donors in previous frames, meaning that 30 minutes should be considered a lower bound for the latent period.

To more accurately characterize the latent period, we turned to lineage-based analysis based on tracking individual cells and their daughters over time [38]. We sought to identify lineages of recipient cells in which conjugation is performed by newly-conjugated transconjugant cells rather than donor cells. In particular, we selected

recipient cell lineages based on the following criteria: 1. no donor neighbors prior to conjugation 2. no transconjugant neighbors for at least the first 10 minutes of the lineage (to ensure that conjugation is performed by "new" transconjugants undergoing the latent period) 3. starting from the first frame with a transconjugant neighbor, a transconjugant neighbor must be present in at least 80% of frames until conjugation. In Fig. 3C, we plot the times from the first appearance of a transconjugant neighbor until the recipient is conjugated and find a mean latent period of approximately 50 minutes, roughly consistent with the picture in Fig. 3A.

2.3 Cell-to-cell variability in Cascade expression determines fates of recipient lineages

To investigate the function of expression stochasticity in potential bet-hedging strategies we now turn to analyzing the apparent variations in Cascade expression between single cells (Fig. 4A). In Fig. 4B, we plot a histogram of Cascade copy number per cell for recipients and transconjugants, combined from all experiments and donor strains. On average, recipient or transconjugants contained 600 Cascade complexes per cell. Absolute measurements of CRISPR expression are relatively few, but Jones *et al* measured the wild-type expression of *Streptococcus pyogenes* Cas9 (SpCas9) at 300 Cas9/cell in *S. pyogenes* [39], while Vink *et al* measured 120 Cascade per cell in *E. coli* from the wild-type Cascade promoter [26]. Thus, the expression here appears to be around 2-5 times higher than wild-type expression. With single-cell Cascade copy number estimations in hand, we evaluated the extent to which Cascade expression is costly for cell growth. In Fig. 4B, we also plot the growth rate as a function of binned Cascade expression (blue dots; see right y-axis). We find that growth is largely unaffected up to about 1000 Cascade complexes per cell. However, the highest-expressing Cascade bin ($\approx$ 1800 per cell) exhibits a decrease of a few percent. Evidently, in this relatively rich growth medium, Cascade expression does not impose an excessive burden on the cells.

Next, we turned our attention to characterizing the dynamics of CRISPR interference. We performed cell tracking to build a database of lineages and trajectories of plasmid copy numbers over time; illustrated in Fig. 4D. By summing over this database, we can compute the fraction of lineages that have cleared the plasmid as a function of the elapsed time since the initial conjugation. Note that this can involve multiple generations of cells along a lineage. This clearance fraction is shown in Figs. 4E and F for *+parE* and Δ*parE* plasmid variants, respectively. For each plasmid, we plot the clearance fraction for all cells (lines with black outline) as well as clearance fraction broken out by Cascade expression quintile of the cells included in the analysis. Considering plasmid variants first, we find rapid and complete clearance for RP4-18-target variants, somewhat slower and incomplete clearance for RP4-1-target variants, and low albeit nonzero clearance for RP4-0-target variants. The latter should have essentially zero clearance and is therefore representative of a false negative dot-detection rate. After about 20-30 minutes, the RP4-0-target clearance curves plateau around 0.2, indicating that plasmid copy number has built up sufficiently that the probability of false negatives is small.

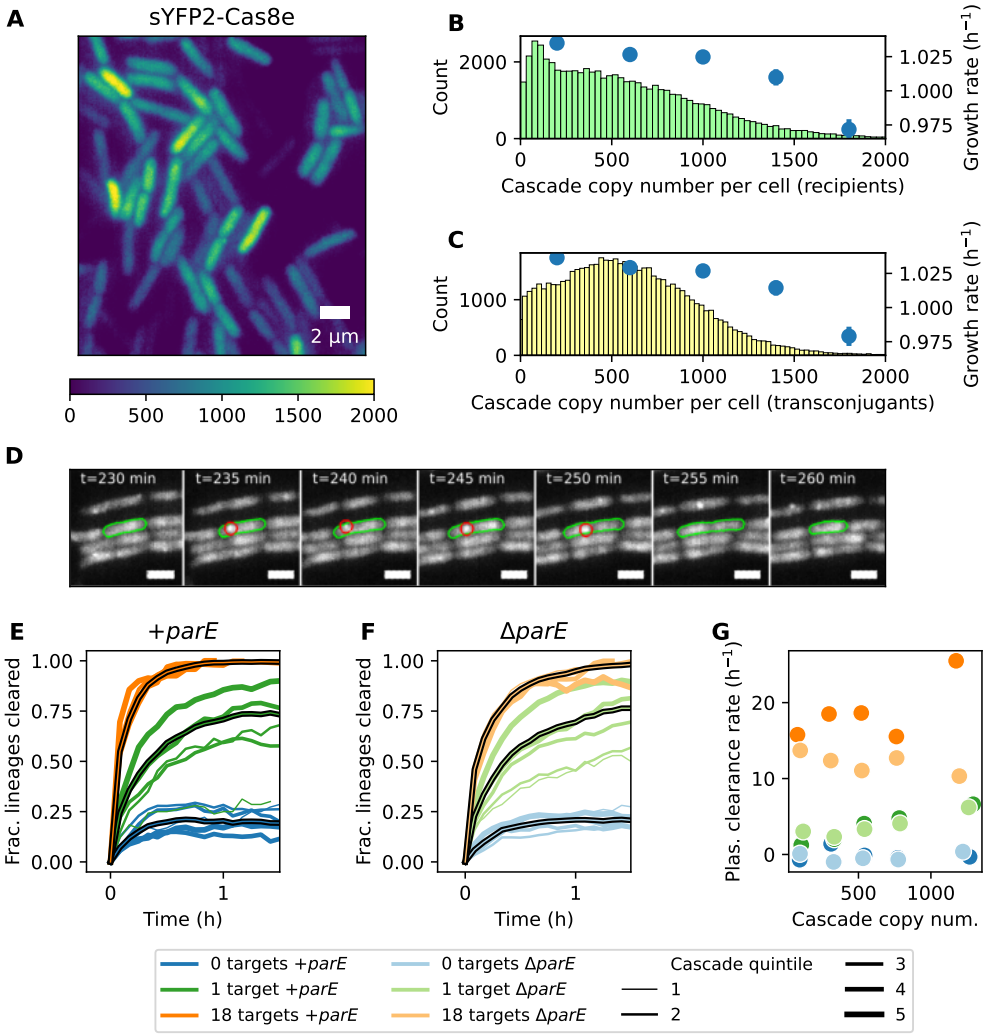

Fig. 4 Cell-to-cell variability in CRISPR interference. (A) Representative fluorescence image acquired in the Venus channel. The image has been rescaled so that the colormap indicates the corresponding Cascade copy number in an average-sized cell. Dark areas correspond to donor cells, which do not express sYFP2-Cas8e. **(B)** and **(C)**. Estimated Cascade copy number per cell and dependence of growth rate on (binned) Cascade copy number for recipient and transconjugant cells. **(D)** Time-lapse images in mCherry-ParB channel showing a recipient cell (green outline) that receives a plasmid (dot outlined in red) at $t = 235$ min before clearing the plasmid at $t = 255$ min. **(E)** and **(F)**. Fraction of cell lineages that acquired and subsequently cleared the plasmid as a function of time, CRISPR target number (color), and Cascade expression quintile (line thickness). Initial plasmid acquisition is defined as $t = 0$ in these plots. More CRISPR targets and higher Cascade expression generally yield faster clearance (see text for caveats). Black-outlined lines show clearance fraction for all cells in all Cascade expression quintiles. At least 1189 unique plasmid infection events were analyzed for each donor. **(G)** Plasmid clearance rate for all plasmid variants as a function of Cascade expression. Clearance rate increases linearly for 1-target plasmid but has no clear trend for 18-target plasmids, which we attribute to Cas3 cleavage becoming rate-limiting. See text and Appendix A for calculation of clearance rate.

Next considering Cascade expression, we binned "infection events" (any interval from a recipient cell lineage during which a plasmid is present *e.g.* Fig. 4D) into quintiles according to mean Cascade expression during the infection. For each RP4 variant, we plot the clearance curve for each Cascade expression quintile (Fig. 4EF; Cascade quintile represented by line thickness). For RP4-0-target, Cascade expression had no clear relationship with clearance, as expected. For RP4-1-target, Cascade expression strongly affected clearance, with the highest quintile exhibiting clearance close to that of RP4-18-target. Somewhat surprisingly, for RP4-18-target, Cascade expression again had limited effect on clearance. We hypothesize that this is due to saturation of the Cas3 protein, normally recruited by the Cascade complex to perform degradation.

Finally, we estimate the plasmid clearance rates for each plasmid variant and Cascade expression quintile. To do so, we use the clearance fraction at $t = 5$ min after initial infection. Assuming plasmid replication time > 5 min, this allows us to avoid the confounding effects of plasmid replication after conjugation. We assume exponentially distributed times to clearance, while correcting for false negative plasmid measurements and for the possibility of additional conjugation events during the 5 minute interval. A complete derivation can be found in Appendix A. The estimated clearance rates are shown in Fig. 4G as a function of Cascade copy number, and are also reproduced in Supplementary Table S1. If we compare with Fig. 3B (conjugation rates), we see few transconjugants when the clearance rate is significantly greater than the conjugation rate (RP4-18-target), and around 50% transconjugants when the clearance rate is comparable to the conjugation rate (RP4-1-target), which seems reasonable. In the Discussion we compare these rates to previous biophysical measurements and find reasonable agreement.

2.4 Bioinformatics: 'Wild' toxin-antitoxin and CRISPR spacer relationship

Given our earlier finding that *parDE*-containing plasmids are able to effectively counter CRISPR clearance and cause cell death, we hypothesized that spacers against *parDE*-containing plasmids should be underrepresented in CRISPR arrays found in bacterial populations. Effective CRISPR targeting and *parDE* should be evolutionarily incompatible. However, the general applicability of our earlier finding is still under question owing to the controlled experimental set-up we use, insofar as Cascade expression levels (and timing thereof) are not necessarily representative of wild bacterial populations. Moreover, we found that effect of the toxin-antitoxin defense is only obviously apparent in the situation where cells are allowed time to accumulate toxin. A negative correlation between the presence of *parDE* and CRISPR targets in available sequence databases would provide stronger evidence for the applicability of this mechanism in wider bacterial populations.

We sourced a collection of 72556 plasmid sequences from the NCBI [40] and a collection of 353377 spacer sequences [41]. Searching for spacer targets in plasmids with blastn yielded 30485 plasmids with spacer target sequences. We then used the RP4 parE (XCD26522.1) and parD (XCD26521.1) amino acid sequences to run local tblastn against our library of plasmid sequences to identify those with both parE and parD sequences. We identified 96 plasmid sequences containing parE and parD

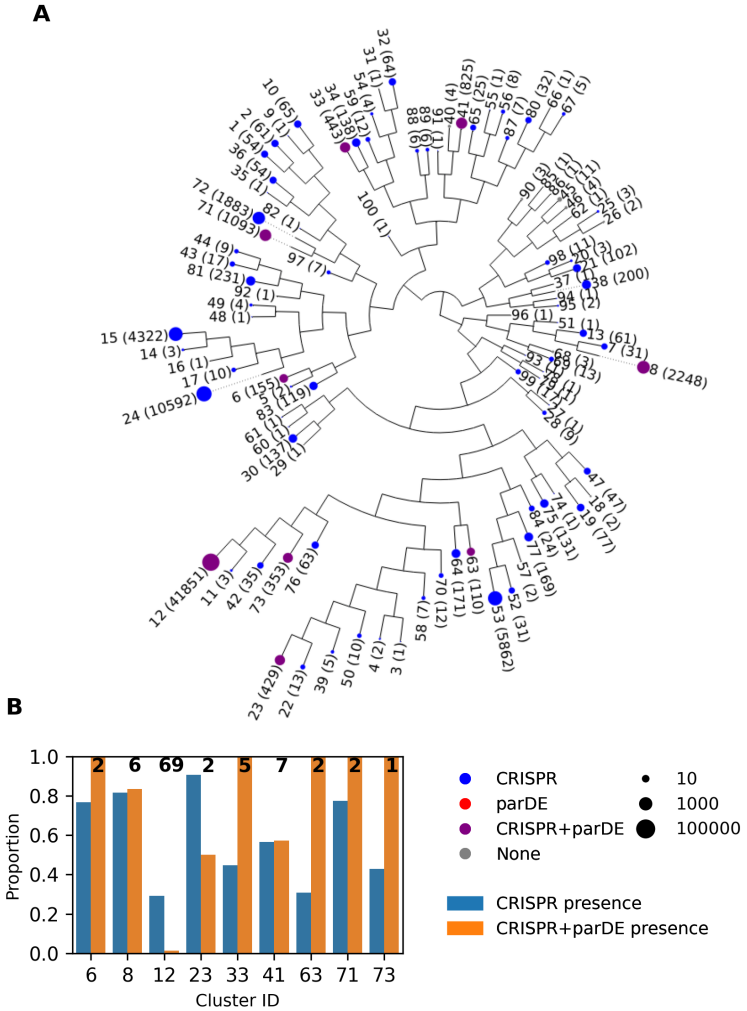

Fig. 5 Genetic co-occurrence of *parDE* and CRISPR spacers. A) Phylogenetic tree depicting clustering of plasmid sequences analysed. Only relationships between clusters are shown for visualisation ease. Plasmid cluster ID number is indicated at the end of nodes, number of plasmids contained within each cluster is indicated in brackets. Circles at the nodes are color coded for the spacer and or *parDE* content and size is scaled according to the number plasmids in each cluster. Scale of size in legend is only approximate. **B)** Proportion of plasmids with CRISPR spacer targets and proportion of *parDE*-containing plasmids that have both *parDE* and CRISPR spacer target, for all clusters that contain some plasmids with both *parDE* and CRISPR spacer targets. Bold numbers above bars indicate the raw number of *parDE* containing plasmids .

sequences and 23 sequences with both *parDE* sequences and CRISPR spacer targets. Of the plasmids that contained spacer sequences, the maximum and minimum number of spacers was 539 and 1 respectively. The mean and median numbers were 6.1 and 3, respectively. If all plasmids are considered those values drop to 2.6 and 0. Naïvely, the fraction of plasmids with spacer sequences 30485/72556 (42%) compared to the fraction of *parDE*-containing plasmids with spacer sequences 23/96 (24%), is an under representation of CRISPR spacers in *parDE*-containing plasmids.

In order to test this under representation statistically we constructed a generalized linear mixed model (GLMM) where CRISPR presence/absence is the response variable and, plasmid length and *parDE* presence/absence are fixed effects. To account for phylogenetic relationships, we clustered plasmids based on sequence similarity (4) into 100 groups (Fig. 5A). Qualitatively, we can see that the presence of clusters containing both *parDE* and CRISPR is not restricted to any one phylogenetic lineage and that CRISPR spacers are extremely prevalent among lineages. These cluster assignments were then included in our GLMM as a random effect. Both plasmid size and *parDE* presence display a significant effect on the presence of CRISPR spacers in a plasmid; the effect of plasmid size was very small (Table S2). If we consider the difference in ratios on a cluster-by-cluster basis (Fig. 5B) we can see that the relationship is likely driven by one very large cluster (ID number 12) which has a very large under representation of *parDE* and CRISPR containing plasmids. Taken together, these observations indicated that generally plasmids are targeted by CRISPR spacers and there is a bias toward not targeting plasmids with *parDE* sequences, but many *parDE*-containing plasmids are nonetheless targeted. The effectiveness of such targeting is likely highly context dependent.

3 Discussion

Here, we have characterized plasmid population dynamics at both the population and single-cell level, demonstrating their dependence on spacer target number, Cascade expression level and toxin-antitoxin (TA) system presence. We also report conjugation rates per neighboring plasmid-containing cell, plasmid transmission latency and refined Cascade search time estimates; together, these comprise a synthesis of multi-scale measurements describing CRISPR - plasmid conjugation dynamics.

Overall, we found that increasing the number of spacer targets on plasmids was an effective method of increasing CRISPR interference besides increasing the expression of Cascade complexes. A recent study showed that plasmids have on average low numbers of spacer targets [42] and indeed our bioinformatic findings suggest that spacer target number per plasmid is often low, in the single digits range. The wild-type level of CRISPR-RNA production in *E. coli* is very low [22, 43]. Nonetheless, the wild-type Cascade constituents are capable of providing CRISPR defense when expressed [22] or when the, normally repressive, activity of *HNS* is lifted [43]. Likewise, when target number and expression level were both low, the *E. coli* Type I-E CRISPR system used in our study had only limited effectiveness in targeting RP4. When the plasmid was targeted by all spacers, on the other hand, interference was efficient, holding the percentage of transconjugants at less than 10% of recipients. This is relevant

for our broader understanding of conjugation and clearance dynamics, development of simulation models and for synthetic biology applications. The latter includes *e.g.* use in biotechnology applications safeguarding against plasmid uptake and antibiotic acquisition [44].

In circumstances where CRISPR interference was effective when the plasmid was targeted by all spacers and Cascade expression was relatively high, clearance could prove something of a pyrrhic victory. RP4 encodes the *parDE* plasmid addiction mechanism which targets gyrase and ultimately decreases cell viability [36, 45]. Recipient cells in which the plasmid was established and subsequently cleared exhibited elongated phenotypes typical of cells with DNA damage and ultimately the combined transconjugant-recipient-cleared transconjugant population of cells suffered. Given the potential cost of expressing CRISPR defense [46–48], it is possibly an unrealistic expectation that CRISPR defense is strongly expressed at all times in wild populations. It may be more likely that expression is switched on as required or that cells utilize expression variation between cells to take advantage - at the population level - of a few cells with high expression. Such an assumption that most cells have low CRISPR expression most of the time would make toxin accumulation in cells before plasmid clearance much more likely and increase the countering capabilities of TA systems. Indeed, our bioinformatic assessment suggests that CRISPR targeting is used less often than expected for plasmids that contain *parDE* TA systems but many *parDE*-containing plasmids do indeed still have spacer targets and the trend is not consistent across all plasmid clades, indicating that CRISPR targeting may not be completely redundant in such cases.

Cell-to-cell variability and its effect on population dynamics has aroused interest in recent years[24, 25, 27, 49]. The limited plasmid clearance capability of CRISPR at the "wild-type-like" end of the spectrum paints a questionable picture of CRISPR efficacy against plasmids. Indeed, only the cells with the highest levels of Cascade expression can effectively clear the 1-target RP4 plasmid. Such high expression levels may be costly; studies have highlighted the potential costs associated with CRISPR immunity although the effect is often small [46, 47]. In our study, we observed only a small reduction in growth rate at the very highest Cascade expression levels. Many reports cited here, and indeed our study, look at CRISPR dynamics in an *in vitro* setting with relatively high resource availability [8, 24, 26, 27] - plasmids and CRISPR interference are unlikely to induce a cost or induce a very small cost. Other studies take a genetic outlook [17, 19–21, 23] and therefore do not assess the capability of cells to expel plasmids when it becomes advantageous to do so. Previous experimental work [8] has shown a growth cost of plasmid maintenance and acquisition in some circumstances and theoretical models show that costly plasmids can be driven to extinction by CRISPR interference in populations [50]. Expressing a range of Cascade expression levels, and therefore clearance capabilities, may allow CRISPR to subtly provide the necessary nudge to clear a few cells in circumstances where plasmid presence is detrimental, even perhaps in the face of toxin-induced stress in extreme circumstances. Equally it would also maintain a subset of plasmid-containing cells when beneficial.

Acquiring more spacer targets for any single plasmid increases the clearance capability without increasing expression burden. In an ecological sense, however, this has

a countervailing effect, effectively diluting the presence of other spacers present in the genome. This increases the expression burden required to clear MGEs targeted by other spacers in the genome. The concept of diluting targeting spacers by having more spacers genetically encoded has been explored and discussed previously in the context of modeling viral attack and spacer mutation and binding efficiency dynamics, leading to predictions of an optimal spacer number [51]. The "nudging" capability of Cascade expression variation could be fine-tuned based on a cell and population's ecological niche requirements by modifying spacer numbers for different MGEs. This kind of fine-tuning can be achieved in response to phage on relatively short time-scales (days) [18].

A key variable in calculating the efficiency of CRISPR interference, in the context of varying Cascade expression levels (copy number) and spacer targets, is the Cascade complex's search time. Vink *et al* estimated that a single Cascade complex takes about 90 minutes to find and degrade its target sequence [26]. We find, for RP4-1target cells with a mean Cascade copy number of 600 complexes/cell, an interference rate of $k_{clear} = 6.0\ \mathrm{h}^{-1} = 0.1\ \mathrm{min}^{-1}$. Assuming that 10% of Cascade complexes (*i.e.* 60 complexes) are loaded with the correct guide RNA yields a per-complex search time of

$$T_{\mathrm{search}} = \frac{1}{0.1\ \mathrm{min}^{-1}} \times 60 = 600\ \mathrm{min}$$

Given the various assumptions and uncertainties this is not a glaring discrepancy. For example, the estimate of 90 mins is directly proportional to the estimated time of replication for their plasmid of interest (3 mins). An underestimate of this value, caused for example by a delay in replication upon entry into cells or a delay between replication cycles (eclipse time), would cause a proportional increase in the estimated search time. The eclipse time was estimated as 3 minutes for the R1 plasmid in *E. coli* in the absence of replication control but with such systems in place the time was as high as 27 minutes ([52]), easily then including 600 mins in the possible range of Cascade search times.

By leveraging single cell tracking and neighbor analysis, we were also able to measure conjugation rate and conjugation rate per neighboring donor cell. We found that conjugation rate increases linearly with the number of donor neighbors, up to four donors, giving a rate of 2.4 conjugations per hour per donor neighbor. It is worth highlighting in its own right the fact that conjugation rate curves are well-fit by an exponential function. In general, the waiting time for an event (e.g. a chemical reaction) to occur is described by an exponential when the reaction has a single rate-limiting step. This implies that the speed of conjugation is limited by a single rate-limiting step. One plausible candidate is the pilus' search for recipient cells which is used for plasmid transfer [53]. Additionally, we were able to quantify the latent period between plasmid acquisition and subsequent transfer of the plasmid to be around 50 minutes; much slower than rate of conjugation per donor (a single donor-contacted cell would be conjugated in around 25 minutes). Measurements accounting for such spatially influenced parameters are important in understanding and manipulating plasmid dynamics. For example, cells can be tightly packed and not segregated (e.g. biofilm),

sparse and non segregated (marine/ aquatic environments/ blood) or dense and segregated (soil). Previous work has highlighted the influence of spatial confinement in bacterial populations; influencing conjugation rates [33, 34] and motility [54].

Taken together, our results suggest that individual cells are the subject of a delicate trade-off between plasmid cost, plasmid benefit, defense expression, the cost of plasmid clearance, spacer acquisition and spatial considerations; neighboring cells that can influence the rate of plasmid acquisition. If CRISPR defense is expressed too little, there is no interference. With moderate expression, and sufficient spacers, interference can be efficient. At high expression, interference is efficient, even with few targets, but the burden of expression may become relevant in certain environments. Furthermore, the ability of plasmids to fight back with toxin-antitoxin systems means that plasmid acquisition may be especially risky given that the cost of clearing that plasmid may be greater than the burden of defense expression or plasmid maintenance. Maintaining a range of Cascade expression levels, coupled with varied spacer acquisition may maintain cells at all points along this continuum and subsequently be prepared for all eventualities. In the future, it would be interesting to benchmark the performance of CRISPR-Cas against other systems and track plasmid population dynamics under high-pressure selective conditions.

4 Methods

4.1 Cloning and strain construction

4.1.1 RP4 variant construction

RP4 variants were constructed by first performing the following modifications to plasmid pMM441 (a mobilizable plasmid containing an RSF1010 origin of replication and oriT, Cas9, and a guide RNA): (1) replacement of Cas9 with SCFP3A; (2) replacement of Cas9 crRNA with the pMT1 *parS* site; (3) insertion of CRISPR-Cascade targets. The part of the resulting plasmid containing a chloramphenicol resistance gene, pMT1*parS*, and Cascade targets was then PCR amplified and inserted into RP4 using recombineering.

Each of these steps will now be described in further detail. To replace Cas9 with SCFP3A, pMM441 was digested with restriction enzymes NotI and XhoI (Thermo Fisher) which flank the Cas9 coding sequence. pMM441 was a gift from Timothy Lu (Addgene plasmid #61271). At the same time, SCFP3a from plasmid p15a-Cas9deg-SOS-amp (gift from Elf lab) was PCR amplified using primers DJ001 and DJ002 which contained NotI and XhoI sites in their 5' ends . The resulting PCR product (containing the linearized plasmid minus Cas9) was digested with NotI and XhoI, ligated into the digested pMM441 vector using T4 DNA ligase (Thermo Fisher), purified, and transformed into chemically-competent TOP10 cells, resulting in plasmid pDJ301. To replace the Cas9 crRNA with pMT1parS, pDJ301 was PCR amplified with primers DJ009 and DJ010, and a *parS* site was PCR amplified from plasmid pBlueDSBarms-parSMT1-BglII-frtCmR (gift from Elf lab) using primers DJ011 and DJ012. The resulting parS fragment was inserted into the vector using the HiFi DNA Assembly

Kit (New England Biolabs), resulting in plasmid pDJ302. Next, three different derivatives were constructed from pDJ302, containing 0, 1, or 18 targets for Cascade. To create plasmids pDJ310 and pDJ311 (containing 0 and 1 target, respectively), pDJ302 was PCR amplified with primer pairs DJ003/DJ004 and DJ005/DJ006, respectively. Primers DJ003/DJ004 contained 5' "floppy ends" such that, when the resulting linear product was circularized by ligation, a CTT PAM site adjacent to either a 32 bp randomized DNA sequence would be inserted into the plasmid. Similarly, primers DJ005/006 were designed to insert a protospacer corresponding to spacer 1 from the E. coli CRISPR1 array.

At the parS/ParB-mCherry system exhibited good functionality in identifying plasmid-carrying cells, using CFP expression to identify plasmid-carrying cells was not necessary, and CFP could be deleted. To delete SCFP3a from pDJ310/311, the plasmids were PCR amplified with primers DJ075/DJ076, which were designed not to amplify the region containing SCFP3a and its promoter. The resulting linear products were phosphorylated using T4 PNK, blunt end ligated with T4 DNA Ligase, cleaned up, and transformed into TOP10 chemically competent cells (Thermo Fisher), thus creating plasmids pDJ320 (no targets) and pDJ321 (1 target). To create plasmid pDJ323 (containing all 18 targets from the E. coli CRISPR arrays 1 and 2), pDJ320 was PCR amplified and linearized using primers DJ019/DJ115 to create a vector backbone. At the same time, plasmid pTarget (gift from Brouns lab), containing all 18 protospacers from E. coli CRISPR array 1 and 2, was amplified using primers DJ022/116. The primers were designed with overlaps appropriate to HiFi DNA Assembly. The two fragments were assembled using HiFI DNA Assembly, purified, and transformed into TOP10 chemically competent cells (Thermo Fisher Scientific).

To create the *wt parE* RP4 variants used in this work, a cassette containing (1) chloramphenicol resistance (2) pMT1parS (3) 0, 1, or 18 CRISPR targets was amplified from plasmids pDJ320, pDJ321, and pDJ323, respectively, using primers DJ066/DJ067 with AccuPrime Pfx DNA Polymerase (Thermo Fisher). These primers contained homology arms appropriate to lambda red recombination into RP4, upon which the cassette replaced the divergently transcribed tetA-tetR genes in RP4. The cassettes were integrated into RP4 using lambda red recombination and restreaked three times to ensure segregation of recombineered RP4 variants. Clones were streaked on tetracycline plates to ensure complete segregation (since segregated clones are sensitive to tetracycline). The resulting strains were EL4120, EL4121, and EL4122.

To create the Δ*parE* RP4 variants used in this work, we chose to back up a few steps and start over from *wt* RP4. Our approach was first to delete *parE* from *wt* RP4 using the DIREX method [55], and then re-introduce the cassette containing CRISPR targets and the *parS* site. To delete parE, we amplified the *AcatsacA* cassette from genomic DNA isolated from strain DA42825 (gift of Andersson lab) with primers DJ128/DJ062 and DJ063/DJ129 which contain appropriate overhangs to replace *parE* with the *AcatsacA* cassette while creating 40 bp direct repeats immediately outside of the cassette. The *AcatsacA* cassette contains the *cat* gene for chloramphenicol resistance as well as the *sacB* gene imparting sucrose sensitivity. The cassette also contains two copies of the *amilCP* blue chromoprotein as inverted repeats at either end of the cassette. It is necessary to amplify the cassette in two separate reactions due to the

inverted *amilCP* repeats; the two fragments contain a 277 bp overlap in the *cat* gene which is subsequently recombined *in vivo* under lambda red.

The amplified cassette was eletroporated into strain EL3316 which contains *wt* RP4 as well the pSIM5-tet lambda RED recombinase plasmid. The resulting unstable intermediate RP4-parE::AcatsacA has parE replaced by the AcatsacA cassette. The cassette can be maintained by propagation in chloramphenicol-containing media. To resolve the unstable intermediate and create the $\Delta parE$ RP4, cells were streaked on sucrose-containing plates to select for cells in which recombination occurred. The designed direct repeats as well as the inverted *amilCP* repeats recombine to yield a scarless deletion of *parE* (see ref. [55] for a pedagogical explanation of this technique). Primers were designed to perform an in-frame deletion with the first 10 and last 10 amino acids of ParE left intact.

With the RP4-ΔparE plasmid in hand, the cassettes from plasmids pDJ320, pDJ321, and pDJ323, were amplified as before and integrated into RP4-ΔparE using lambda RED recombineering (also as before), yielding strains EL4519-4522 with 0, 1, and 18 CRISPR targets, respectively. The three RP4 versions were then conjugated into strain EL4009 to create strains EL4533-4535 used as donors in the paper.

DNA cleanup steps were performed with the Monarch PCR and DNA Cleanup Kit (New England Biolabs) unless otherwise specified. PCR steps were performed with Q5 DNA Polymerase unless otherwise specified (New England Biolabs).

4.1.2 Donor strain construction

As E. coli will opportunistically use arabinose as a carbon source depending on the presence of alternative carbon sources, we elected to delete the *araBAD* operon in experimental strains to ensure that arabinose would function only as an inducer and not as a carbon source. The *araBAD* operon was deleted and replaced with a kanamycin resistance marker in strain MG1655 using lambda red recombination, yielding strain EL4008. The araBAD deletion was subsequently transferred to strain EL904 (MG1655 rph+) using P1 transduction, yielding strain EL4009. Finally, the three *wt parE* RP4 variants containing 0, 1 and 18 targets were respectively conjugated from strains EL4067, EL4068, and EL4069, into strain EL4009, yielding strains EL4020, EL4021, and EL4022. The $\Delta parE$ variants were similarly conjugated into strain EL4009 yielding strains EL4533, EL4534, and EL4535.

4.1.3 Recipient strain construction

Creation of the recipient strain EL4012 from strain MG1655 involved a number of steps: (1) creating an N-terminal fusion of SYFP2 with Cas8e; (2) P1 phage transduction of mCherry-ParB into the *gtrA* locus; (3) replacement of the Pcas3 promoter with the Placuv5 promoter; (4) replacement of the Pcascade promoter with the araBp8 promoter (5) deletion of araBAD operon. In replacing the *cas3* and Cascade promoters, we are following the appproach of Datsenko *et al* [35]. For recipient strain El4014, we also deleted the CRISPR1 and CRISPR2 arrays.

Creation of N-terminal SYFP2-Cas8e fusion. To create the fusion, we used the Dup-In method to which enables creation of markerless, scarless fusions

[55, 56]. This technique involves first creating a 500 bp duplication of *sfyp2* flanking a selectable/counter-selectable *Acatsac1* cassette, which contains the *cat* gene for chloremphenicol resistance and the *sacB* gene from *Bacillus subtilus* conferring sucrose sensitivity. This construct has previously been created in the Dan Andersson lab as strain DA43403. We amplified the *syfp2* duplication-*Acatsac1* cassette from genomic DNA isolated from strain DA43403 (gift of Andersson lab) using primers with homology such that the product would recombine in the N-terminal position relative to *cas8e*. Due to the duplicated *syfp2* segments, the entire construct cannot be amplified at once. Rather, two separate PCRs were performed using primers DJ062/DJ065 and DJ063/DJ064 respectively. Primer DJ065, containing homology to the N-terminus of *cas8e*, was designed such that *sfyp2* would lack a stop codon, with an 18bp linker sequence (as in Vink *et al* [26]), and deleting the start codon of *cas8e*. The two PCR products overlap by 277 bp in the *cat* gene such that they will recombine to form the correct, complete construct using lambda red recombination (see Figure 4 in [56]).

The result of this lambda red recombination was strain EL3177. As long as this strain is propagated in chloramphenicol-containing media, the unstable duplication will be maintained. The construct was then transferred to MG1655 *rph+* using P1 phage transduction. After isolation of a successful transductant, the strain was grown on sucrose-containing plates to select for colonies in which the flanking *syfp2* duplications have recombined, eliminating the *Acatsac1* cassette, and yielding an N-terminal fusion as desired (strain EL3178).

P1 phage transduction of mCherry-ParB. A cassette containing an mCherry-ParB fusion protein expressed from a medium-strength constitutive promoter, as well as a spectinomycin resistance marker, was transduced from strain EL851 (gift of Elf lab) to strain EL3178 to create strain EL3202. The cassette is integrated in the *gtrA* locus.

Replacement of Pcas3 promoter with Placuv5 promoter. The Pcas3 promoter was first replaced with the *Acatsac1* cassette (described above). To accomplish this, the *Acatsac1* cassette was PCR amplified with primers DJ079/DJ080 which contain appropriate homology flanking Pcas3. This product was integrated into the chromosome using lambda red recombineering, thus replacing Pcas3 with the *Acatsac1* cassette, and yielding strain EL3654. The PlacUV5 promoter was ordered as a gBlock (Integrated DNA Technologies) also containing homology flanking the now-deleted Pcas3 promoter. This gBlock was then integrated into the genome again using lambda red. Successful clones were identified by sucrose counter-selection (as correct clones will delete the *Acatsac1* cassette), yielding strain EL3658.

Replacement of Pcascade promoter with araBp8 promoter. A similar two-step approach was employed to replace the Pcascade promoter. First, the Pcascade promoter was replaced by the *Acatsac1* cassette using lambda red recombination, yielding strain EL3661. Second, the araBp8 promoter was PCR amplified from strain KD263 (gift from Severinov lab) using primers DJ092/DJ093 which also contain homology flanking the deleted Pcascade promoter. This PCR product was integrated into strain EL3661 using lambda red recombination and correct clones were isolated using sucrose counter-selection, yielding strain EL3701.

At this point we had achieved the desired construct, and several rounds of lambda red recombination had been performed. To mitigate the risk of off-target mutations accumulating elsewhere in the genome, we performed a final step of transducing the modifications in the vicinity of the Cascade operon into a clean background strain. To create a clean background strain, we transduced the *Pcas::Acatsac1* construct from strain EL3654 into strain EL3202 using P1 phage transduction, yielding strain EL3946. This strain has not had any lambda red recombinations performed in it, only P1 phage transductions. Finally, the modifications from strain EL3701 were transduced into strain EL3946 using P1 transduction, which was possible since clones with the correct transduction could be idenfitied by sucrose counter-selection (only successful transductants can grow on sucrose). This yielded strain EL3957.

Deletion of *araBAD* operon. Again to ensure that arabinose would be used only as an inducer and not a substrate for growth, the *araBAD::kanR* deletion was transduced from strain EL4008 into strain EL3957, yielding strain EL4012 used as recipient in the manuscript.

Deletion of CRISPR1 and CRISPR2 arrays. For deletion of the CRISPR arrays, we again used the DIREX method to create markerless, scarless deletions [55]. For each array, we kept the leader and terminator sequences intact while deleting all spacers from the array. To delete the CRISPR1 array, we amplified the *AcatsacA* cassette from genomic DNA isolated from strain DA42825 (gift of Andersson lab) with primers DJ098/DJ062 and DJ063/DJ099 which contain appropriate overhangs to replace CRISPR1 with the *AcatsacA* cassette while creating 40 bp direct repeats immediately outside of the cassette, and also creating a 277 bp overlap in the *cat* gene for recombination of the two fragments as described above. The products were integrated into strain EL3701 using lambda red recombination, and after confirmation of the correct clone, the unstable intermediate was resolved by growth on sucrose-containing media yielding the desired deletion of CRISPR1. An identical procedure was followed for CRISPR2, with the *AcatsacA* cassette being amplified by primer pairs DJ100/DJ062 and DJ063/DJ101, yielding strain EL3898 with both arrays deleted.

To avoid accumulation of off-target mutations due to multiple lambda red rounds, we transduced this construct into strain EL3946 using P1 phage transduction. Since the CRISPR1 and CRISPR2 arrays are separated by around 20 kb, and P1 transduction transfers approximately 100kb at a time, it is possible to transduce both mutations at once, subject to sequence verification. Finally, the *araBAD::kanR* deletion was transduced from strain EL4008 to create strain EL4014, the desired final construct.

4.2 Microfluidic assays

Strains were grown overnight in LB broth (tryptone 10 g/L, yeast extract 5 g/L, NaCl 10 g/L) from single colony with appropriate antibiotic addition (spectinomycin 15 μg/L for recipient strains or chloramphenicol 12.5 μg/L for donor strains) shaking 200 rpm 37°C. Overnight cultures were diluted 20 μL into 10 mL M9 media in 50 mL falcon tubes. Final M9 media composition is as follows (Na_2HPO_4 12.8 g/L, KH_2PO_4 3 g/L, NaCl 0.5 g/L, NH_4Cl 1 g/L, $CaCl_2$ 100 μM, $MgSO_4$ 2 mM, x1 RPMI 1640 amino acid solution 1640 (Sigma R7513), glucose 0.4%, pluronic F-127 85 mg/L (Sigma-Aldrich

P2443)), where inducers were included in the media these are final concentration (IPTG (isopropyl β-D-1-thiogalactopyranoside) 1 mM, arabinose 0.2%). Diluted cells were returned to shaking incubation for three hours; when cells were in exponential growth cells. At this point cells were mixed 1:1 (donor:recipient) by volume and loaded into microfluidic chips for imaging. The device used has two parallel row of traps separated by a central barrier. This allows loading of two different strain combinations in each experiment. For most experiments +parE and -parE variants of the same donor strain were run concurrently. For experiments with delayed induction, the +/- CRISPR array-containing recipient strains were run concurrently with one donor strain.

The microfluidic device used here is a modified version of the device used in Baltekin et al 2017[57] with a 1000 μm channel height and large rectangular cell traps measuring 48 μm x 60 μm. Cells were loaded into these traps immediately after mixing under 0.4-0.6 bar pressure. After cells were loaded fresh M9 media was flown into the device at 0.2 bar and continued for the duration of each experiment. Imaging setup was composed of a Nikon Ti2 equipped with Teledyne Kinetix camera, 100x Ph3 objective (CFI Plan Apo lambda 100X/1.45 oil. MRD01905), 1.5x internal magnification selected (for a total magnification of 150x), Lumencore SPECTRA light engine and filters configured for mCherry (No excitation filter, Chroma ET645/75M emission filter and Chroma ZT594rdc dichroic mirror) and SYFP2 (Chroma Z514/10 excitation filter, Semrock FF01-542/27M emission filter and Semrock BrightLine Di02/R514 dichroic mirror) fluorescence imaging. The microscope stage was encased in Okolab H201-ENCLOSURE hood and H201-T-UNIT-BL temperature control unit to maintain a constant 30°C throughout the experiments. Cells were imaged in phase contrast (100 ms exposure), mCherry fluorescence (200 ms), SYFP2 fluorescence (100 ms) channels over a period of 6 hours. Phase, mCherry and SYFP2 images were taken at 1.25, 5, and 30 min intervals, respectively.

4.3 Image analysis

The image analysis performed in this work is based on the following foundational steps: segmentation, cell tracking, dot (i.e., plasmid) detection, and cell classification. These steps are followed by downstream analyses such as plasmid clearance rates, cell death rates, growth rate computations, etc. Foundational steps as well as downstream analyses were integrated into a custom image analysis pipeline based on the pipeline described in ref. [58]. All code used to perform the analyses is available upon request.

Segmentation was performed on phase contrast images using the Omnipose software package [59], which uses a U-net-based neural net architecture to perform morphology-independent cell segmentation. Custom neural nets were trained for this project using manually-curated segmentation of cells grown in microfluidic devices as ground truth. Cell tracking was performed using the Baxter algorithm [38]. Dot detection was performed using the radial symmetry algorithm; detected dots were subsequently refined by fitting a Gaussian to each detected dot.

To classify cells as donors, recipients, and transconjugants, a two-step process was used. Recipients and transconjugants express ParB-mCherry, whereas donors have no signal in the mCherry channel. Cells were first classified as "mCherry positive" or "mCherry negative" by thresholding, where the threshold was chosen to minimize the

standard deviation of fluorescence signal intensity within the respective "mCherry positive" and "mCherry negative" groups. "mCherry negative" cells can then immediately be identified as donors. "mCherry positive" cells containing detected mCherry dots (corresponding to ParB-mCherry-tagged RP4 plasmids) are classified as transconjugants, whereas "mCherry positive" cells not containing detected mCherry dots are classified as recipients. For analyses involving cell tracking across multiple frames, transconjugant cells must have two consecutive frames without detected dots in order to revert to recipient status. This is to avoid false clearance events resulting from failure to detect plasmids in a single frame (which can plausibly occur due to RP4's relatively low copy number and the fact that fluorescence imaging takes place in a single z-axis plane, so that plasmids may be undetected due to being out of focus).

Cell neighbors were determined using morphological operations on segmentation masks. For each segmented cell, the segmentation mask was cropped to the bounding box of the cell plus 15 pixels in each dimension. (Pixel size is $6.5\,\mu\text{m}/150 \approx 0.043\,\mu\text{m}$ per pixel). The cell of interest was dilated by six pixels, and the other cells within the cropped region were subsequently dilated by six pixels, one at a time. When the central cell's dilated mask overlapped by 40 pixels or more with another cell's dilated mask, the cells were determined to be neighbors. We experimented with different thresholds and found empirically that 40 pixels was a reasonable threshold to match our judgements by eye of which cells were neighbors.

4.4 Cascade copy number calculations

E. coli strain EL4269 was used as a basis for the calculation of Cascade copy number. This strain contains SYFP2 fused to lacI along with a single genomic lacI-binding site Osym. The strain was induced and prepared in the same way as cells used in experiments. After growth to exponential growth phase cells were mounted on agarose pads (1.5% agarose, M9 + inducers as described above). Cells were imaged on the same microscope set-up as described above but with varying exposure times and light intensity settings (1000 ms, 10% intensity and 500 ms 40% intensity). Circular ROIs were drawn around visible dots within cells (representative of a single bound Cascade molecule) and a background measurement taken inside the same cell away from the dot position with the same size and shape ROI. By averaging background subtracted measurements over biological and technical repetitions we obtained a value of 6382 and 2222 fluorescence units per Cascade molecule under (1000 ms and 500 ms) imaging conditions, respectively. Next we used constitutive SYFP2 expressor, strain EL2788 imaged under the same conditions as the LacI fusion strain. Manual cell counting, automated masking with Fiji auto-threshold function and intensity measurements allowed us to calculate the copy number as approximately 23 Cascade copies per cell across all replicates. We than imaged the constitutive expressor in microfluidic traps under experimental imaging conditions. Given the now known copy number per cell we were able to calculate 526 total arbitrary fluorescence units per Cascade-SYFP2 molecule under experimental imaging conditions. This value was then used to calculate copy number from background subtracted fluorescence values in experiments.

4.5 Bioinformatics analysis

CRISPR spacer sequences[41] and *parDE* sequences (XCD26522.1, XCD26521.1) were used to search plasmid sequences[40] using local blastn and tblastn respectively. For spacer sequence hits those with 100% sequence identity over 100% of the sequence length were retained. For *parDE* sequences, hits with an e-value below 1E-30 were taken forwards and considered *parDE*-like sequences for our analysis. To account for plasmid phylogenetic relatedness we first sub-sampled plasmid sequences by sketching plasmid sequences with Mash (v2.3)[60], taking 1000 21bp long fragments. We then used Mash to compute a distance matrix for all plasmid sequences based on Jaccard distance. This distance matrix was then read into Matlab and a linkage tree calculated. K-means clustering and the elbow method was used to determine a suitable number of clusters to represent the phylogenetic relationships between plasmids. These cluster assignments were used as a random effect in a generalized linear mixed model with *parDE* presence/ absence and plasmid length included as fixed effects and CRISPR presence/ absence as the response variable.

Supplementary figures and tables. Further supplementary information can be found in the Appendices in this document.

Acknowledgements. We are grateful to the Brouns lab for the gift of the pTarget plasmid, the Semenova and Severinov labs for the gift of strain KD267, the Andersson lab for the gift of strains DA43403 and DA42825, the Elf lab for the gift of strain EL851, and the Lu lab for the gift of plasmid pMM441. Useful discussions and feedback were provided by Anna Knöppel, Jimmy Larsson, Johan Elf, and Nathalie Balaban. DJ acknowledges support from the Swedish Research Council (grant number 2020-05137). LR acknowledges support from the Carl Tryggers Foundation (grant number CTS 21:1334). Computation and data management were enabled by resources provided by the Swedish National Infrastructure for Computing at UPPMAX.

Author Contributions. LR performed experiments, analyzed data, and wrote analysis code. DL performed experiments, and designed and constructed strains. JC developed experimental protocols. JW wrote analysis code. DJ analyzed data, wrote analysis code, and conceived and supervised the project.

References

[1] Hartmann, M., Six, J.: Soil structure and microbiome functions in agroecosystems. Nature Reviews Earth & Environment **4**(1), 4–18 (2023)

[2] Van Der Heijden, M.G., Bardgett, R.D., Van Straalen, N.M.: The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecology letters **11**(3), 296–310 (2008)

[3] Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., Knight, R.: Diversity, stability and resilience of the human gut microbiota. Nature **489**(7415), 220–230 (2012)

[4] Thomas, C.M., Nielsen, K.M.: Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nature reviews microbiology **3**(9), 711–721 (2005)

[5] Palomino, A., Gewurz, D., DeVine, L., Zajmi, U., Moralez, J., Abu-Rumman, F., Smith, R.P., Lopatkin, A.J.: Metabolic genes on conjugative plasmids are highly prevalent in Escherichia coli and can protect against antibiotic treatment. The ISME Journal **17**(1), 151–162 (2023)

[6] Schmid, K., Schupfner, M., Schmitt, R.: Plasmid-mediated uptake and metabolism of sucrose by Escherichia coli K-12. Journal of Bacteriology **151**(1), 68–76 (1982)

[7] GBD 2021 Antimicrobial Resistance Collaborators: Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. The Lancet **404**(10459), 1199–1226 (2024) https://doi.org/10.1016/S0140-6736(24)01867-1 . Publisher: Elsevier. Accessed 2025-03-04

[8] Prensky, H., Gomez-Simmonds, A., Uhlemann, A.-C., Lopatkin, A.J.: Conjugation dynamics depend on both the plasmid acquisition cost and the fitness cost. Molecular systems biology **17**(3), 9913 (2021)

[9] Subbiah, M., Top, E.M., Shah, D.H., Call, D.R.: Selection pressure required for long-term persistence of bla cmy-2-positive inca/c plasmids. Applied and Environmental Microbiology **77**(13), 4486–4493 (2011)

[10] Diaz Ricci, J.C., Hernández, M.E.: Plasmid effects on Escherichia coli metabolism. Critical reviews in biotechnology **20**(2), 79–108 (2000)

[11] Carroll, A.C., Wong, A.: Plasmid persistence: costs, benefits, and the plasmid paradox. Canadian journal of microbiology **64**(5), 293–304 (2018)

[12] Rocha, E.P., Bikard, D.: Microbial defenses against mobile genetic elements and viruses: Who defends whom from what? PLoS biology **20**(1), 3001514 (2022)

[13] Niault, T., Houte, S., Westra, E., Swarts, D.C.: Evolution and ecology of anti-defence systems in phages and plasmids. Current Biology **35**(1), 32–44 (2025)

[14] Horvath, P., Barrangou, R.: CRISPR/Cas, the immune system of bacteria and archaea. Science **327**(5962), 167–170 (2010)

[15] Pourcel, C., Touchon, M., Villeriot, N., Vernadet, J.-P., Couvin, D., Toffano-Nioche, C., Vergnaud, G.: CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. Nucleic acids research **48**(D1), 535–544 (2020)

[16] Westra, E.R., Levin, B.R.: It is unclear how important CRISPR-Cas systems are

for protecting natural populations of bacteria against infections by mobile genetic elements. Proceedings of the National Academy of Sciences **117**(45), 27777–27785 (2020)

[17] Wheatley, R.M., MacLean, R.C.: CRISPR-Cas systems restrict horizontal gene transfer in Pseudomonas aeruginosa. The ISME journal **15**(5), 1420–1433 (2021)

[18] Watson, B.N., Capria, L., Alseth, E.O., Pons, B.J., Biswas, A., Lenzi, L., Buckling, A., Houte, S., Westra, E.R., Meaden, S.: CRISPR-Cas in Pseudomonas aeruginosa provides transient population-level immunity against high phage exposures. The ISME journal **18**(1), 039 (2024)

[19] O'Meara, D., Nunney, L.: A phylogenetic test of the role of CRISPR-Cas in limiting plasmid acquisition and prophage integration in bacteria. Plasmid **104**, 102418 (2019)

[20] Gophna, U., Kristensen, D.M., Wolf, Y.I., Popa, O., Drevet, C., Koonin, E.V.: No evidence of inhibition of horizontal gene transfer by CRISPR–Cas on evolutionary timescales. The ISME journal **9**(9), 2021–2027 (2015)

[21] Palmer, K.L., Gilmore, M.S.: Multidrug-resistant enterococci lack CRISPR-cas. MBio **1**(4), 10–1128 (2010)

[22] Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., Van Der Oost, J.: Small CRISPR RNAs guide antiviral defense in prokaryotes. Science **321**(5891), 960–964 (2008)

[23] Touchon, M., Charpentier, S., Pognard, D., Picard, B., Arlet, G., Rocha, E.P., Denamur, E., Branger, C.: Antibiotic resistance plasmids spread among natural isolates of Escherichia coli in spite of CRISPR elements. Microbiology **158**(12), 2997–3004 (2012)

[24] Mamontov, V., Martynov, A., Morozova, N., Bukatin, A., Staroverov, D.B., Lukyanov, K.A., Ispolatov, Y., Semenova, E., Severinov, K.: Persistence of plasmids targeted by CRISPR interference in bacterial populations. Proceedings of the National Academy of Sciences **119**(15), 2114905119 (2022)

[25] Heuer, H., Abdo, Z., Smalla, K.: Patchy distribution of flexible genetic elements in bacterial populations mediates robustness to environmental uncertainty. FEMS microbiology ecology **65**(3), 361–371 (2008)

[26] Vink, J.N., Martens, K.J., Vlot, M., McKenzie, R.E., Almendros, C., Bonilla, B.E., Brocken, D.J., Hohlbein, J., Brouns, S.J.: Direct visualization of native CRISPR target search in live bacteria reveals cascade dna surveillance mechanism. Molecular cell **77**(1), 39–50 (2020)

[27] McKenzie, R.E., Keizer, E.M., Vink, J.N., Van Lopik, J., Büke, F., Kalkman, V., Fleck, C., Tans, S.J., Brouns, S.J.: Single cell variability of CRISPR-cas interference and adaptation. Molecular Systems Biology **18**(4), 10680 (2022)

[28] Killelea, T., Hawkins, M., Howard, J.L., McGlynn, P., Bolt, E.L.: Dna replication roadblocks caused by cascade interference complexes are alleviated by recg dna repair helicase. RNA biology **16**(4), 543–548 (2019)

[29] Cho, S., Choe, D., Lee, E., Kim, S.C., Palsson, B., Cho, B.-K.: High-level dcas9 expression induces abnormal cell morphology in Escherichia coli. ACS synthetic biology **7**(4), 1085–1094 (2018)

[30] Levin, B.R., Stewart, F.M., Rice, V.A.: The kinetics of conjugative plasmid transmission: fit of a simple mass action model. Plasmid **2**(2), 247–260 (1979)

[31] Simonsen, L., Gordon, D., Stewart, F., Levin, B.R.: Estimating the rate of plasmid transfer: an end-point method. Microbiology **136**(11), 2319–2325 (1990)

[32] Hernández-Beltrán, J.C.R., San Millán, A., Fuentes-Hernández, A., Peña-Miller, R.: Mathematical models of plasmid population dynamics. Frontiers in microbiology **12**, 606396 (2021)

[33] Weiss, A., Wang, T., You, L.: Promotion of plasmid maintenance by heterogeneous partitioning of microbial communities. Cell systems **14**(10), 895–905 (2023)

[34] Sun, H., Radosevich, M., Sun, Y., Millet, L., Qian, S., Zhuang, J.: Pore confinement enhances but surface adhesion reduces bacterial cell-to-cell conjugation. Biology and Fertility of Soils **60**(7), 901–910 (2024)

[35] Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., Semenova, E.: Molecular memory of prior infections activates the crispr/cas adaptive bacterial immunity system. Nature communications **3**(1), 945 (2012)

[36] Roberts, R.C., Ström, A.R., Helinski, D.R.: The parde operon of the broad-host-range plasmid rk2 specifies growth inhibition associated with plasmid loss. Journal of molecular biology **237**(1), 35–51 (1994)

[37] Arias, C.F., Acosta, F.J., Bertocchini, F., Herrero, M.A., Fernández-Arias, C.: The coordination of anti-phage immunity mechanisms in bacterial cells. Nature communications **13**(1), 7412 (2022)

[38] Magnusson, K.E., Jaldén, J., Gilbert, P.M., Blau, H.M.: Global linking of cell tracks using the viterbi algorithm. IEEE transactions on medical imaging **34**(4), 911–929 (2014)

[39] Jones, D., Unoson, C., Leroy, P., Curic, V., Elf, J.: Kinetics of dcas9 target search

in Escherichia coli. Biophysical Journal **112**(3), 314 (2017)

[40] Galata, V., Fehlmann, T., Backes, C., Keller, A.: Plsdb: a resource of complete bacterial plasmids. Nucleic acids research **47**(D1), 195–202 (2019)

[41] Grissa, I., Vergnaud, G., Pourcel, C.: CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic acids research **35**(suppl_2), 52–57 (2007)

[42] Androsiuk, L., Maane, S., Tal, S.: CRISPR spacers acquired from plasmids primarily target backbone genes, making them valuable for predicting potential hosts and host range. Microbiology Spectrum **12**(12), 00104–24 (2024)

[43] Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L., Severinov, K.: Transcription, processing and function of CRISPR cassettes in Escherichia coli. Molecular microbiology **77**(6), 1367–1379 (2010)

[44] Lee, D., Muir, P., Lundberg, S., Lundholm, A., Sandegren, L., Koskiniemi, S.: A CRISPR-cas9 system protecting e. coli against acquisition of antibiotic resistance genes. Scientific Reports **15**(1), 1545 (2025)

[45] Tu, C.-H., Ruan, S., Holt, M., Bourne, C.R.: Broken beyond repair: Ta system pare toxins mediate effective gyrase inhibition without driving resistance. Journal of Bacteriology, 00416–24 (2025)

[46] Liu, Z.-L., Hu, E.-Z., Niu, D.-K.: Investigating the relationship between CRISPR-Cas content and growth rate in bacteria. Microbiology spectrum **11**(3), 03409–22 (2023)

[47] Vale, P.F., Lafforgue, G., Gatchitch, F., Gardan, R., Moineau, S., Gandon, S.: Costs of CRISPR-Cas-mediated resistance in streptococcus thermophilus. Proceedings of the Royal Society B: Biological Sciences **282**(1812), 20151270 (2015)

[48] Zaayman, M., Wheatley, R.M.: Fitness costs of CRISPR-Cas systems in bacteria. Microbiology **168**(7), 001209 (2022)

[49] Erickson, K.E., Otoupal, P.B., Chatterjee, A.: Gene expression variability underlies adaptive resistance in phenotypically heterogeneous bacterial populations. ACS infectious diseases **1**(11), 555–567 (2015)

[50] Siedentop, B., Losa Mediavilla, C., Kouyos, R.D., Bonhoeffer, S., Chabas, H.: Assessing the role of bacterial innate and adaptive immunity as barriers to conjugative plasmids. Molecular Biology and Evolution **41**(10), 207 (2024)

[51] Martynov, A., Severinov, K., Ispolatov, I.: Optimal number of spacers in CRISPR arrays. PLoS computational biology **13**(12), 1005891 (2017)

[52] Olsson, J.A., Berg, O.G., Dasgupta, S., Nordström, K.: Eclipse period during replication of plasmid r1: contributions from structural events and from the copy-number control system. Molecular microbiology **50**(1), 291–301 (2003)

[53] Goldlust, K., Ducret, A., Halte, M., Dedieu-Berne, A., Erhardt, M., Lesterlin, C.: The f pilus serves as a conduit for the dna during conjugation between physically distant bacteria. Proceedings of the National Academy of Sciences **120**(47), 2310842120 (2023)

[54] Zhang, H., Wetherington, M.T., Ko, H., FitzGerald, C.E., Munro, E.M., Nirody, J.A.: Bacterial motility patterns adapt smoothly in response to spatial confinement and disorder. bioRxiv, 2024–09 (2024)

[55] Näsvall, J.: Dup-in and direx: Techniques for single-step, scar-free mutagenesis with marker recycling. In: Reisch, C.R. (ed.) Recombineering: Methods and Protocols, pp. 85–104. Humana Press, New York (2022). Chap. 7

[56] Näsvall, J., Knöppel, A., Andersson, D.I.: Duplication-insertion recombineering: a fast and scar-free method for efficient transfer of multiple mutations in bacteria. Nucleic acids research **45**(5), 33–33 (2017)

[57] Baltekin, Ö., Boucharin, A., Tano, E., Andersson, D.I., Elf, J.: Antibiotic susceptibility testing in less than 30 min using direct single-cell imaging. Proceedings of the National Academy of Sciences **114**(34), 9170–9175 (2017)

[58] Camsund, D., Lawson, M.J., Larsson, J., Jones, D., Zikrin, S., Fange, D., Elf, J.: Time-resolved imaging-based CRISPRi screening. Nature methods **17**(1), 86–92 (2020)

[59] Cutler, K.J., Stringer, C., Lo, T.W., Rappez, L., Stroustrup, N., Brook Peterson, S., Wiggins, P.A., Mougous, J.D.: Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. Nature methods **19**(11), 1438–1448 (2022)

[60] Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., Phillippy, A.M.: Mash: fast genome and metagenome distance estimation using minhash. Genome biology **17**, 1–14 (2016)

Supplementary Information

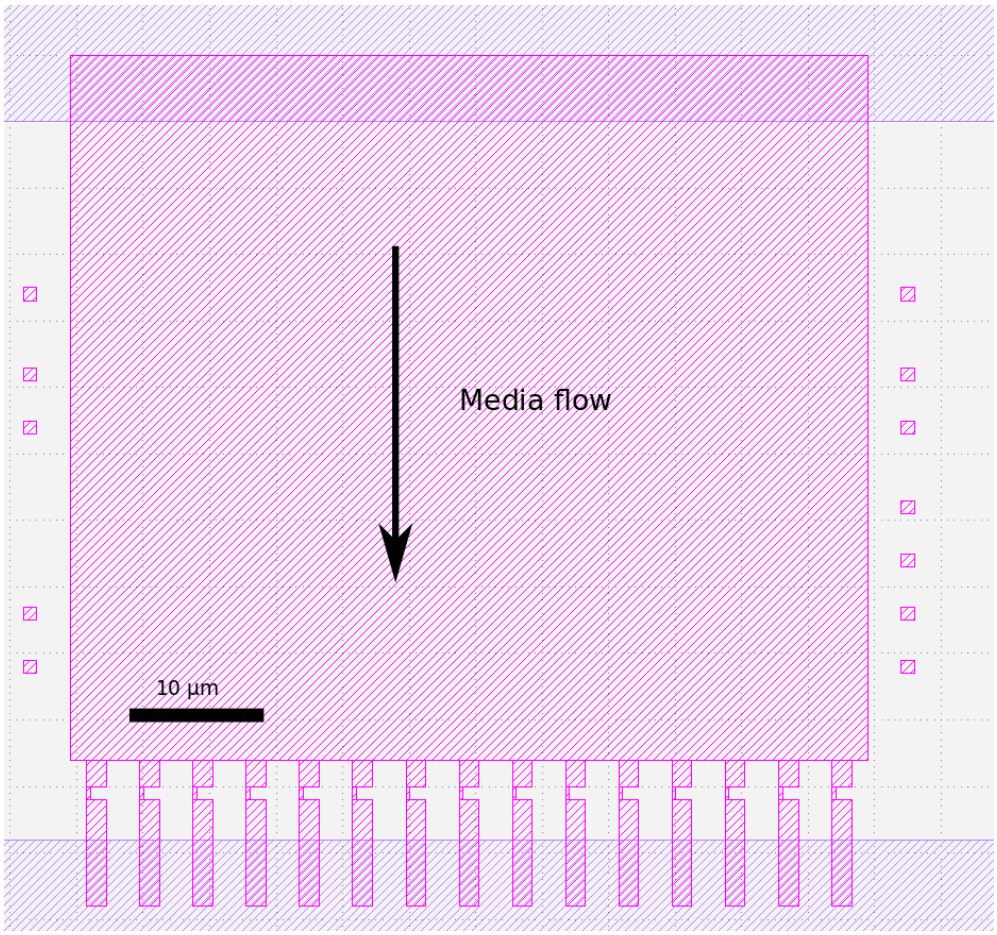

Fig. S1 Schematic of of a single microfluidic trap/chamber. Cells occupy the rectangular area which measures 48x60 μm. The channels at the bottom are 1.5 μm wide with 0.25 μm constrictions, enabling media to flow over the cells as indicated while preventing cells from passing through. Each chip contains 200 such traps.

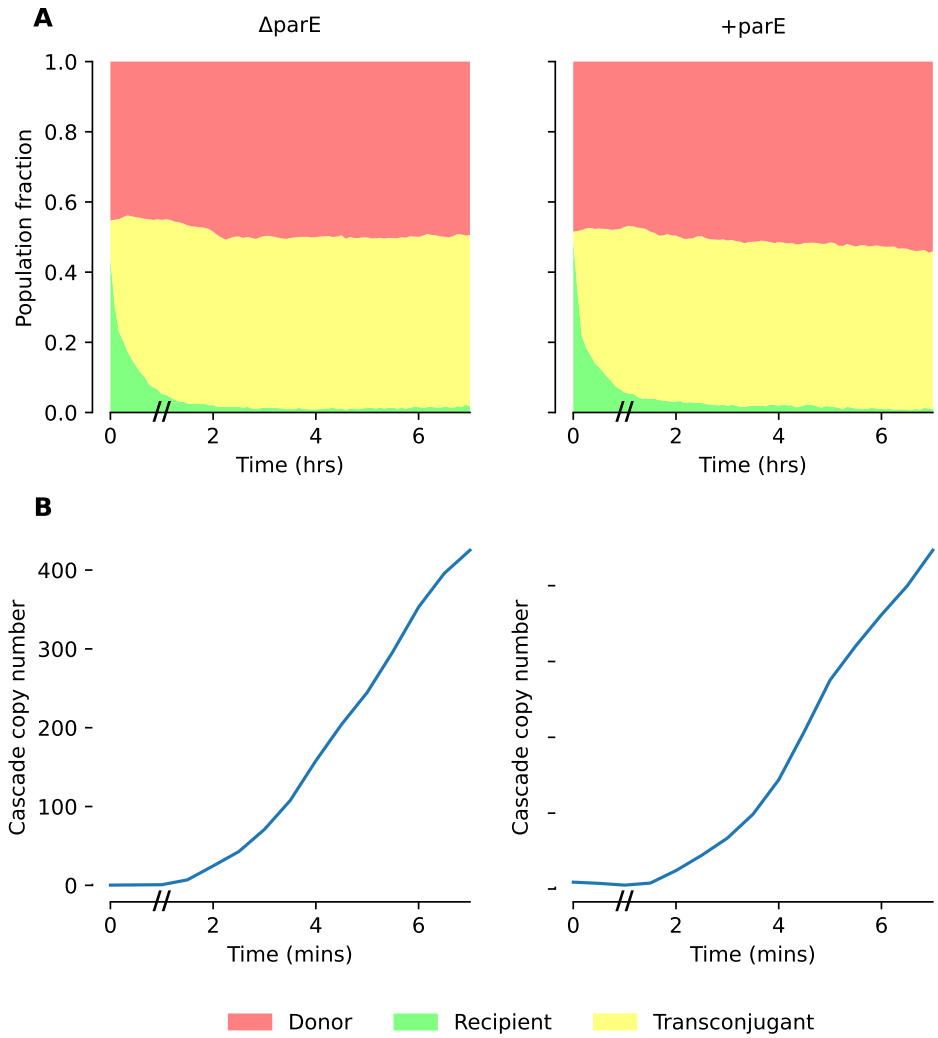

Fig. S2 A) Mean fraction of the cell population classified as donor, recipient or transconjugant across the course of microfluidic experiments with delayed Cascade induction with 18-target RP4 plasmids, with and without parE, for recipient *E. coli* strains with genomic CRISPR arrays deleted. B) Mean Cascade copy number per cell across the course of the experiment. Inducers were added to the media at 60 minutes, indicated with // on axis.

Table S1 Inferred plasmid clearance rates k_{clear}, computed both naively and accounting for conjugation, false negatives, etc (see Appendix A). Two consecutive plasmid-free frames were required to call a clearance event. Rates computed over all cells (average copy number 600 Cascade per cell). To obtain approximate rate per Cascade, divide by 600. To obtain corresponding rates for "wild-type" expression levels ($\approx$ 125 Cascade per cell), divide rates by 5. Corrected rates are higher because the non-negligible probability of an additional conjugation event during the two required plasmid-free frames means that the apparent clearance rate will be lower than the true clearance rate (Appendix A).

Plasmid	Naive k_{clear} (h^{-1})	Corrected k_{clear} (h^{-1})
RP4-0-target	0.78	0.0
RP4-1-target	3.54	5.4
RP4-18-target	9.66	16.8
RP4-0-target$\Delta parE$	0.72	0.0
RP4-1-target$\Delta parE$	3.12	4.56
RP4-18-target$\Delta parE$	7.26	12.3

Table S2 Statistical summary of generalized linear mixed model fitting CRISPR spacer target presence to the size of plasmids and the presence or absence of parDE. Displayed, for each model component estimate is, standard error, test statistic, degrees of freedom, p value, and 95% confidence intervals.

	Estimate	SE	tStat	DF	pValue	95% confidence intervals	
						Lower	Upper
Intercept	0.37331	0.19279	1.9363	51091	0.052833	-0.0045652	0.75119
Size	6.93E-07	5.46E-08	12.69	51091	7.59E-37	5.86E-07	8.01E-07
parDE	-1.4447	0.38591	-3.7435	51091	0.00018166	-2.201	-0.68827

Table S3 List of strains used in this work.

Strain	Genotype
DA42825	MG1655 rph(+)<AcatsacA>rph(+)
DA43403	MG1655 ΔaraBAD::syfp2-Acatsac1-syfp2
KD263	K-12 F+, lacUV5-cas3 araBp8-cse1, CRISPR I: repeat-G8-repeat, ΔCRISPR II
EL851	MG1655 rph+ Δlac ΔmalI::frt gtrA::P58-mCherry:parB-SpR
EL904	MG1655 rph+
EL2788	MG1655 gtrA::P58-SYFP2-parB-frt-kan-frt
EL3177	BW25993 rph+ Δphi80 sYFP2(dup645bp)-Acatsac1-sYFP2(dup645bp)-cas8e
EL3178	MG1655 rph+ sYFP2-cas8e
EL3202	MG1655 rph+ sYFP2-cas8e gtrA::P58-mCherry:parB-SpR
EL3316	MG1655 rph+ gtrA::P58-mCherry:parB-SpR /pSIM5-Tet /RP4
EL3654	MG1655 rph+ Pcas3::acatsac1 sYFP2-cas8e gtrA::P58-mCherry:parB-SpR
EL3658	MG1655 rph+ Pcas3::Placuv5 sYFP2-cas8e gtrA::P58-mCherry:parB-SpR /pSIM5-te
EL3661	MG1655 rph+ Pcas3::Placuv5 sYFP2-cas8e gtrA::P58-mCherry:parB-SpR Pcascade::acatsac1 /pSIM5-tet
EL3701	MG1655 rph+ gtrA::P58-mCherry:parB-SpR Pcas3::Placuv5 Pcascade::araBp8 sYFP2-cas8e
EL3898	MG1655 rph+ Pcas3::Placuv5 sYFP2-cas8e gtrA::P58-mCherry:parB-SpR Pcascade::araBp8 ΔCRISPR1 ΔCRISPR2 /pSIM5-tet
EL3946	MG1655 rph+ sYFP2-cas8e gtrA::P58-mCherry:parB-SpR Pcas3::acatsac1
EL3957	MG1655 rph+ gtrA::P58-mCherry:parB-SpR Pcas3::Placuv5 Pcascade::araBp8 sYFP2-cas8e
EL4008	MG1655 araBAD::kan
EL4009	MG1655 rph+ araBAD::kan
EL4012	MG1655 rph+ araBAD::kan gtrA::P58-mCherry:parB-SpR Pcas3::Placuv5 Pcascade::araBp8 sYFP2-cas8e
EL4014	MG1655 rph+ araBAD::kan gtrA::P58-mCherry:parB-SpR Pcas3::Placuv5 Pcascade::araBp8 sYFP2-cas8e ΔCRISPR1 ΔCRISPR2
EL4067	MG1655 lacA-cat ompF::tet /RP4-0-target
EL4068	MG1655 lacA-cat ompF::tet /RP4-1-target
EL4069	MG1655 lacA-cat ompF::tet /RP4-18-target
EL4120	MG1655 rph+ araBAD::kan /RP4-0-target
EL4121	MG1655 rph+ araBAD::kan /RP4-1-target
EL4122	MG1655 rph+ araBAD::kan /RP4-18-target
EL4269	BW25993 LacI-SYFP2 Δ(Plac-lacZ)::term DELphi80 DEL(ACACTTTCTTAAAT;ycil; =phi80 att site)::lacI(Q55N) aslA-glmZ intergenic::cat-Osym ::Kan_p138_RBS
EL4516	MG1655 rph+ gtrA::P58-mCherry:parB-SpR /RP4-parE::AcatsacA /pSIM5-Tet
EL4517	MG1655 rph+ gtrA::P58-mCherry:parB-SpR /RP4-ΔparE /pSIM5-Tet
EL4519	MG1655 rph+ gtrA::P58-mCherry-parB-SpR /RP4-ΔparE-0-target
EL4520	MG1655 rph+ gtrA::P58-mCherry-parB-SpR /RP4-ΔparE-1-target
EL4521	MG1655 rph+ gtrA::P58-mCherry-parB-SpR /RP4-ΔparE-18-target
EL4533	MG1655 rph+ araBAD::kan /RP4-ΔparE-0-target
EL4534	MG1655 rph+ araBAD::kan /RP4-ΔparE-1-target
EL4535	MG1655 rph+ araBAD::kan /RP4-ΔparE-18-target

Table S4 List of primers used in this work.

Number	Name	Sequence
DJ00001	NotI-sCFP-fwd	TAAGCA GCGGCCGC ATGGTGAGCAAGGGCGA
DJ00002	XhoI-sCFP-rev	TGCTTA CTCGAG TCACAATTCCTTGTACAGCTCGTC
DJ00003	notrac-randtarget-fwd	CACTCTGGGCAA CTT GGTACC CAGCCATAAAACAATACTTAATAC
DJ00004	notrac-randtarget-rev	CGGTATCTAGAACGGTCAGA CCATGG AGGGTGAAGACGAAAGG
DJ00005	notrac-target1-fwd	CACCGTCGTTGC CTT GGTACC CAGCCATAAAACAATACTTAATAC
DJ00006	notrac-target1-rev	AGATTTCACGCCTGACGCTG CCATGG AGGGTGAAGACGAAAGG
DJ00009	pDJ301_amp_fwd	gacagacgtcatactagatatcaag
DJ00010	pDJ301_amp_rev	gttctgcagggtgctagg
DJ00011	parSMT1-fwd	gacctagcaccctgcagaacaggatgccgaagagcatc
DJ00012	parSMR1-rev	tatctagtatgacgtctgtccgaaattatgagtcacgaagag
DJ00019	pDJ312_vec_fwd	cagccataaaacaatacttaatactatagaatg
DJ00022	pTarget_ins_rev	agtattaagtattgttttatggctgggtaccaagctttcgcagacg
DJ00062	cat-midF	CGACGATTTCCGGCAGTTTC
DJ00063	cat-midR2	GCCGACATGGAAGCCATCAC
DJ00064	casA-syfp2-DupIn-F	gaaatgcaggagaaatgtcggagcatatgaaggagaacaa ATGGTTAGCAAGGGCGAAGAAC
DJ00065	casA-syfp2-DupIn-R	ttcgcgggcgtacagggatccagttatcaataagcaaatt agccatgagcgccccggg TTTATACAGCTCATCCATACCCAG
DJ00066	RP4-ins-fwd	cgagcgcggccgtgaggcttggacgctagggcagggcatg CTTCCAGATGTATGCTCTTC
DJ00067	RP4-ins-rev	cttgccggaagtcgccttgacccgcatggcataggcctat GTCGCTTGATATCTAGTATGACG
DJ00075	pDJ310_delCFP_fw	CCTAGCACCCTGCAGAAC
DJ00076	pDJ310_delCFP_rv	ACTGATAGGGAATCGATCCC
DJ00115	pDJ323_vec_rev_2.0	ccgttctgtc CCATGGAGGGTGAAGACG
DJ00116	pDJ323_ins_fwd_2.0	ccctccatgg GACAGAACGGCCTCAGTAG
DJ00128	DELparE_F_P1	GGGATCGCGCTTGACGGCCTACATCCTCACGGCTGAGGCC ACGCGACTTGCCGAC GTGTAGGCTGGAGCTGCTTC
DJ00129	DELparE_R_P1	GCGGCTGAAATCAGCCCTTGAGCCTGTCGGCAAGTCGCGT GGCCTCAGCCGTGAG GTGTAGGCTGGAGCTGCTTC
DJ00079	Pcas3_Acatsac1_up	TGGGTTACACGAAGGGTAAATATTGCCGGACAAATTAAAC GTGTAGGCTGGAGCTGCTTC
DJ00080	Pcas3_Acatsac1_down	TTTAAAAGGTTCCATTAATAGCCTCCCTGTTTTTTTAGTA CATATGAATATCCTCCTTAGTTCC
DJ00081	Pcas3_to_Placuv5_gblock	TGGGTTACACGAAGGGTAAATATTGCCGGACAAATTAAAC AGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCA CTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGT GTGGAATTGTGAGCGGATAACAATTTCACACAC
DJ00098	DELCRISPR1_F_P1	TACTAAAAAAACAGGGAGGCTATTAATGGAACCTTTTAAA ATACTTTTAGA GAGTTCCCCGCGCCAGCGGGGATAAACCG GGCGCACTGGATGCG GTGTAGGCTGGAGCTGCTTC
DJ00099	DELCRISPR1_R_P1	GGGAACTCCAAGTGATATCCATCATCGCATCCAGTGCGCC CGGTTTATCCCCGCT GTGTAGGCTGGAGCTGCTTC
DJ00100	DELCRISPR2_F_P1	TTATGTTTAGA GTGTTCCCCGCGCCAGCGGGGATAAACCG CCATATAACCCGTTA GTGTAGGCTGGAGCTGCTTC
DJ00101	DELCRISPR2_R_P1	CTAATATAAAAACTTGAGAAAGAGATAACGGGTTATATGG CGGTTTATCCCCGCT GTGTAGGCTGGAGCTGCTTC

Appendix A Estimation of plasmid degradation / CRISPR interference rates

A.1 Purpose

We want to estimate clearance/plasmid degradation rates from time-lapse fluorescence microscopy observations of plasmid copy number. The fact that our observations take place at discrete time intervals of 5 minutes requires some consideration, as well as the possibility of additional conjugation events taking place during each 5 minute interval.

Thoughout, we will assume that we are starting with an observation of one plasmid at $t = 0$. We will also assume that 5 minutes is a sufficiently short time interval that plasmid replication cannot occur within this interval. All times are in minutes, and all rates are in units of min^{-1}.

A.2 Naive approach

If there is one plasmid at $t = 0$, and plasmids are degraded at rate k_d, the probability of zero plasmids at $t = 5$ is given by the cumulative distribution function (cdf) of the exponential distribution with rate k_d:

$$P(0, t = 5|1, t = 0) = 1 - e^{-5k_d}. \tag{A1}$$

This can easily be solved for k_d:

$$k_d = -\frac{1}{5}\log(1 - P_0) \tag{A2}$$

where we defined $P_0 = P(0, t = 5|1, t = 0)$ for less notational clutter. **NB: P_0 is what we actually observe experimentally.**

A.3 Requirement of no additional conjugation events

One problem with this approach is that it ignores the possibility that another conjugation event may have occurred in this 5 minutes. The probability of 0 plasmids at $t = 5$ is implicitly the probability of (degradation within 5 minutes) AND (no additional conjugation within 5 minutes). The probability of no additional conjugation is given by 1 minus the cdf of the exponential distribution with rate k_c where k_c is the rate at which plasmids are conjugated into the cell: $P(\text{no conjugation}) = e^{-5k_c}$. (If e.g. $k_c = 0.07$ this probability becomes 0.70). We assume that degradation and conjugation are independent, so that

$$P_0 = (1 - e^{-5k_d}) \times e^{-5k_c} \tag{A3}$$

which yields

$$k_d = -\frac{1}{5}\frac{\log(1 - P_0)}{e^{-5k_c}} \tag{A4}$$

Since $e^{-5k_c} < 1$, this will lead to a larger estimated value for k_d than with the naive approach. For instance, if $k_c = 0.07$, then $1/e^{-5k_c} \approx 1.42$.

A.4 Possibility of additional conjugation and degradation events

In reality, there are numerous "trajectories" that could lead to observation of 0 plasmids at $t = 5$ starting from one plasmid at $t = 0$. For instance the original plasmid could have been degraded AND any number of additional conjugation and degradation events could have occurred as long as the number of additional degradation events matches the number of additional conjugation events. So to correctly compute P_0, we need to sum over all trajectories yielding 0 plasmids after 5 minutes. Let T denote plasmid copy number trajectories. Then

$$P_0 = \sum_T P(0, T) = \sum_T P(0|T)P(T)$$

But $P(0|T) = 1$ for "valid" trajectories and 0 otherwise, so

$$P_0 = \sum_{\text{valid } T} P(T)$$

Valid trajectories are trajectories where (the initial plasmid is degraded) AND (any subsequently conjugated plasmids are degraded). Let d_i represent degradation of the original plasmid, and c and d represent conjugation and degradation of any subsequent plasmids.

$$P_0 = P(d_i = 1, c = 0, d = 0) + P(d_i = 1, c = 1, d = 1) + P(d_i = 1, c = 2, d = 2) + ...$$

$$P_0 = P(d_i = 1, c = 0) + \sum_{n=1}^{\infty} P(d_i = 1, c = n, d = n)$$

$$P_0 = P(d_i = 1, c = 0) + \sum_{n=1}^{\infty} P(d_i = 1, c = n, d = n)$$

$$P_0 = P(d_i = 1, c = 0) + P(d_i = 1) \sum_{n=1}^{\infty} P(c = n, d = n)$$

where we assumed that degradation d_i of the initial plasmid is statistically independent of conjugation and degradation of any subsequent plasmids. If we further also assume that any future conjugation/degradation paired events are statistically independent from one another then we get

$$P_0 = P(d_i = 1, c = 0) + P(d_i = 1) \sum_{n=1}^{\infty} (P(c = 1, d = 1))^n$$

In equation A1 we found that $P(d_i = 1) = 1 - e^{-5k_d}$ and in equation A3 we found that $P(d_i = 1, c = 0) = (1 - e^{-5k_d}) \times e^{-5k_c}$. So thus far we have

$$P_0 = (1 - e^{-5k_d}) \times e^{-5k_c} + (1 - e^{-5k_d}) \sum_{n=1}^{\infty} (P(c = 1, d = 1))^n. \tag{A5}$$

A.4.1 Calculation of $P_{GL} = P(c = 1, d = 1)$

What is the probability of one conjugation and one degradation event occuring within the next 5 minutes? Let's call this P_{GL} (for "gain and loss"). It is the (probability of conjugation at time t') times (the probability of no additional conjugation events within $5 - t'$) times (the probability of degradation within $5 - t'$) summed over all t' from 0 to 5 minutes. (Strictly speaking the first parenthesis should say: probability of conjugation between t' and $t' + dt'$). Or

$$P_{GL} = P(c = 1, d = 1) = \int_0^5 P(\text{gain at time } t') \times P(\text{no additional gains within } 5 - t') \times P(\text{degrade within } 5 - t') dt'$$

$$P_{GL} = \int_0^5 k_c e^{-k_c t'} \times e^{-(5-t')k_c} \times (1 - e^{-(5-t)k_d}) dt'$$

where used the fact that the first term in the integral (gain at time t') is given by the exponential distribution with rate k_c, the second term (no additional gain within $5 - t'$) is 1 minus the cdf of the exponential distribution with rate k_c, and the third term (degrade within $5 - t'$) is the cdf of the exponential distribution with rate k_d over the interval $5 - t'$. After some algebra we find that

$$P_{GL} = 5k_c e^{-5k_c} - \frac{k_c}{k_d} e^{-5k_c} (1 - e^{-5k_d}). \tag{A6}$$

The first term can be straightforwardly interpreted as the probably that exactly one conjugation occurs during the 5 minutes. The second term is slightly less transparent but should reflect the probability that the plasmid does not get degraded in the time after conjugation. Looking at some limits, if $5k_c >> 1$ (fast conjugation) the probability goes to zero since the chance of only one conjugation is small. If $k_d >> 1$ (fast degradation) then the second term goes to zero and the probability reduces to the first term i.e. the probability of having one conjugation event. Makes sense.

A.4.2 Putting it together

Plugging these results back into equation A5 we obtain

$$P_0 = (1 - e^{-5k_d}) e^{-5k_c} + (1 - e^{-5k_d}) \sum_{n=1}^{\infty} P_{GL}^n$$

$$P_0 = (1 - e^{-5k_d})e^{-5k_c} + (1 - e^{-5k_d}) \sum_{n=1}^{\infty} P_{GL}^n$$

$$P_0 = (1 - e^{-5k_d})(e^{-5k_c} + \sum_{n=1}^{\infty} P_{GL}^n)$$

$$P_0 = (1 - e^{-5k_d})(e^{-5k_c} + P_{GL} \sum_{n=0}^{\infty} P_{GL}^n).$$

Finally, using the formula for the sum of geometric series and the fact that $P_{GL} < 1$ we find

$$P_0 = (1 - e^{-5k_d}) \left(e^{-5k_c} + \frac{P_{GL}}{1 - P_{GL}} \right) \tag{A7}$$

where

$$P_{GL} = 5k_c e^{-5k_c} - \frac{k_c}{k_d} e^{-5k_c}(1 - e^{-5k_d}).$$

This will need to be solved numerically for k_d, assuming that k_c is known from non-CRISPR-targeting experiments. Note that if $P_{GL} << 1$ this reduces to equation A3. As an example, if $k_c = 0.07$ and $k_d = 0.2$, we obtain $P_{GL} = 0.09$. Which seems reasonable.

A.5 Requiring two consecutive plasmid-free frames

We might choose to require that two consecutive plasmid-free frames are required to call a clearance event, in order to cut down on false negatives (or false positive for clearance, depending on how you look at it). In this case P_0 becomes the probability of 0 plasmids at $t = 5$ AND $t = 10$ given that one plasmid was observed at $t = 0$. Or

$$\begin{aligned} P_0 = P(&\text{degrade original plasmid within 5 min) AND} \\ &P(\text{no new conj. OR all new conj. degraded between 0 and 5 min) AND} \\ &P(\text{no new conj. OR all new conj. degraded between 5 and 10 min}) \end{aligned} \tag{A8}$$

In other words we just need to add another factor of the second term in equation A7:

$$P_0 = (1 - e^{-5k_d}) \left(e^{-5k_c} + \frac{P_{GL}}{1 - P_{GL}} \right)^2 \tag{A9}$$

A.6 Correcting for false negatives

So far we have mostly avoided considering the possibility of false negatives for plasmid presence. We can estimate the probability of false negatives (no plasmid is detected when plasmid is present) from experiments with non-targeted plasmids, where we assume that all plasmid loss in this case is due to false negatives. (This may not be 100% true if e.g. cell division occurs in a newly-conjugated cell before the plasmid has a chance to replicate, but it's probably still a good-enough assumption since few divisions will occur within 5 minutes).

What we have been calling P_0 - the probability of zero plasmids at $t = 5$ given one plasmid as $t = 0$ - is really the probability of *observing* zero plasmids which may not be the true probability. We can write the total probability of observing zero plasmids as a sum of the probabilities of observing zero plasmids given that true number of plasmids is either zero or greater than zero:

$$\begin{aligned} P_{obs}(0) &= P_0 = P_{obs}(0, 0 \text{ plas.}) + P_{obs}(0, > 0 \text{ plas.}) \\ P_{obs}(0) &= P_{obs}(0|0 \text{ plas.})P_{true}(0 \text{ plas.}) + P_{obs}(0| > 0 \text{ plas.})P_{true}(> 0 \text{ plas.}) \end{aligned} \tag{A10}$$

We assume that dot detection parameters have been tuned such that the probability of false positives is negligable and hence $P_{obs}(0|0 \text{ plas.}) = 1$. We define the probability of false negatives $P_{FN} = P_{obs}(0| > 0 \text{ plas.})$ and use the fact that there must be either zero or greater than zero plasmids so that $P_{true}(0) + P_{true}(> 0) = 1$ to obtain

$$P_{obs}(0) = P_{true}(0) + P_{FN}(1 - P_{true}(0))$$

which after rearranging yields

$$P_{true}(0) = \frac{P_{obs}(0) - P_{FN}}{1 - P_{FN}}. \tag{A11}$$

This seems reasonable because we essentially subtract the probability of false negatives and then normalize to the size of the interval remaining after subtracting P_{FN} so that P_{true} again lies between 0 and 1.

To use this expression we simply substitute P_{true} into e.g. equations A7 or A9, using the appropriate estimated P_{FN}. For context we typically see P_{FN} values around 0.07 when two plasmid-free frames are required and 0.17 when one plasmid-free frame is required.

A.7 Limitations

This all becomes much harder if plasmid replication is included. This assumption should be revisited; however, it is not unreasonable to assume that a 60 kb plasmid takes longer than 5 minutes to replicate. Also, RP4 carries the entry exclusion gene *trbK* which prevents conjugation into cells already harboring RP4. In the preceding analysis, we asssume that TrbK protein concentration does not build up sufficiently in 5 minutes to confer the entry exclusion phenotype.