



Modeling endonuclease colicin-like bacteriocin operons as ‘genetic arms’ in plasmid-genome conflicts

Pavithra Anantharaman Sudhakari¹ · Bhaskar Chandra Mohan Ramisetty¹

Received: 19 October 2021 / Accepted: 6 March 2022 / Published online: 23 March 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Plasmids are acellular propagating entities that depend on bacteria, as molecular parasites, for propagation. A ‘tussle’ between bacteria and plasmid ensues; bacteria for riddance of the plasmid and plasmid for persistence within a live host. Plasmid-maintenance systems such as endonuclease Colicin-Like Bacteriocins (CLBs) ensure plasmid propagation within the population; (i) the plasmid-cured cells are killed by the CLBs; (ii) damaged cells lyse and release the CLBs that eliminate the competitors, and (iii) the released plasmids invade new bacteria. Surprisingly, endonuclease CLB operons occur on bacterial genomes whose significance is unknown. Here, we study genetics, eco-evolutionary drive, and physiological relevance of genomic endonuclease CLB operons. We investigated plasmidic and genomic endonuclease CLB operons using sequence analyses from an eco-evolutionary perspective. We found 1266 genomic and plasmidic endonuclease CLB operons across 30 bacterial genera. Although 51% of the genomes harbor endonuclease CLB operons, the majority of the genomic endonuclease CLB operons lacked a functional lysis gene, suggesting the negative selection of lethal genes. The immunity gene of the endonuclease CLB operon protects the plasmid-cured host, eliminating the metabolic burden. We show mutual exclusivity of endonuclease CLB operons on genomes and plasmids. We propose an anti-addiction hypothesis for genomic endonuclease CLB operons. Using a stochastic hybrid agent-based model, we show that the endonuclease CLB operons on genomes confer an advantage to the host genome in terms of immunity to the toxin and elimination of plasmid burden. The conflict between bacterial genome and plasmids allows the emergence of ‘genetic arms’ such as CLB operons that regulate the ecological interplay of bacterial genomes and plasmids.

Keywords Colicin-like bacteriocins · Anti-addiction hypothesis · Plasmid-genome competition · Quasi-cooperation · Agent-based model

Introduction

Plasmids, like viruses, are acellular propagating entities that are common in the bacterial domain. Propagation of plasmids is obligately dependent on their bacterial hosts, sometimes conferring advantage (mutualistic) and other times as molecular parasites (Kado 1998). Unless these propagating entities confer an advantage in the contextual habitat, it is a burden with costs on the fitness relative to the peers. At the

molecular level and a gene-centered view (Dawkins 1989), the existence and propagation of the plasmid at the cost of the bacterial genome is parasitism (MacLean and San Millan 2015). Curing the plasmid is a disadvantage to the plasmid propagation because the cured bacteria may out-compete the plasmid-harboring bacteria under non-selective conditions. That conflict between propagating entities (here, the bacterial genome and plasmid) is never-ending genetic warfare that drives the emergence of ingenious genetic systems (Eberhard 1990). Bacterial curing can occur by passive mechanisms wherein one daughter cell may not inherit the plasmid (Spengler et al. 2006).

The obligatory dependence of plasmids in bacterial cells drives the emergence of multiple ingenious mechanisms that ensure plasmid propagation within the bacterial population. Some genetic systems ensure the dissemination of the plasmid to both the daughter cells. Mechanisms such as (i) high

Communicated by Dr. Martine Collart.

✉ Bhaskar Chandra Mohan Ramisetty
ramisettybcm@biotech.sastra.edu; bcmr25@gmail.com

¹ Laboratory of Molecular Biology and Evolution, 312@ASK1, School of Chemical and Biotechnology, SAstra Deemed University, Thanjavur, India

intracellular copy number of plasmids and (ii) stable partitioning systems, such as the *parMRC* system in low copy number plasmids, allow the spatial distribution of the plasmid copies within the bacterium to increase the probability of dissemination to both the daughter cells (Million-Weaver and Camps 2014). Some others, referred to here as ‘Genetic arms’, are genetic elements that confer an advantage to the associated plasmid/genome by competitively eliminating those that do not carry their homologs. In conflicts between plasmids and genome, some genes confer an advantage by inducing ‘addiction’, a phenomenon where the cell cured of the plasmid is killed by the plasmid-encoded genes (Ramisetty and Santhosh 2016; Fraikin et al. 2020). These are typically composed of a toxic component and an antidote that can neutralize the toxic activity of the toxic component. Plasmids encode strategic genetic arms, exemplified by Colicin-Like Bacteriocin (CLB) operons (Cascales et al. 2007), restriction-modification systems (RMs) (Kulakauskas et al. 1995), and toxin–antitoxin systems (TAs) for stable maintenance in the host population (Van Melderen and Saavedra De Bast 2009). The strategy for plasmid maintenance is by eliminating the daughter cells that may not have inherited a plasmid, the phenomenon of post-segregational killing (Tsang 2017; Harms et al. 2018). A significant proportion of plasmid-bearing bacteria in a population increases the probability of spatiotemporal propagation of that plasmid. The endonuclease CLB operons are similar ‘genetic arms’ that enhance the propagation of the associated plasmid (bearer) by forced maintenance of the plasmid within the host (plasmid addiction) and eliminating the plasmid-free cells from the population. They do so by encoding a toxin–antidote pair whose expression results in the release of toxins that are active against the cells that do not encode them (as they lack cognate antidote to neutralize the toxin). Thus, they alter the bacterial competition. However, endonuclease CLBs are different from other toxin–antidote pairs such as TAs with respect to the mode of toxin release.

This study explored the interplay of endonuclease CLB operons as genetic arms in the conflicts between bacterial genome and plasmids. The endonuclease CLBs are the genetic arms as they enhance the propagation of the bearer by eliminating or manipulating the competition. The bacteriocins such as colicin molecules bind to a specific outer membrane receptor followed by translocation via Tol or TonB machinery to its target. Colicins are of two groups, group A is translocated by the Tol system, and group B is translocated by the TonB system. Group A includes colicins A, E1 to E9, K, N, U, L, Y, and S4 and group B includes colicins B, H, Ia and b, M, D, 5, and 10. Group A colicins are associated with small plasmids and encode lysis protein that aid in colicin release into the medium upon host cell lysis. Group B colicins are associated with large plasmids and do not encode lysis protein; colicins are not released

into the medium (Cascales et al. 2007). Based on the targets, colicins are classified as endonuclease and pore-forming colicins. Endonuclease colicins degrade the nucleic acids such as tRNA, rRNA, and DNA, whereas pore-forming colicins form tiny pores in the inner membrane, causing membrane perturbations (Cascales et al. 2007). Endonuclease colicin operons are one of the well-characterized systems comprising the genes required for colicin production (colicin gene), colicin neutralization (immunity gene) and colicin release (cell lysis gene) from the cell. A typical endonuclease colicin operon comprises colicin activity gene (*cx*a), immunity gene (*cx*i), and lysis gene (*cx*l) (Cascales et al. 2007; Heng et al. 2007). The *cx*a encodes colicin that cleaves nucleic acids; *cx*i encodes immunity protein that inactivates its cognate colicin; *cx*l encodes a lysis protein that lyses the host cell aiding in the release of the colicin (Masaki and Ohta 1985; Kleantous et al. 1999). Among the endonuclease colicins, the targets of each type differ. For example, E3 specifically targets the 16S rRNA gene. Colicin E2, E7, E8, and E9 cleave DNA at a precise site. Colicin E4 and E6 hydrolyze rRNA, and D cleaves tRNA (Cascales et al. 2007). Mutations in the lysis gene hamper host cell lysis and colicin release (Cavard et al. 1985). The immunity protein makes protein–protein interaction with the colicin, both the endogenous and exogenous colicin (cognate colicins released by other colicin-producing bacteria), thereby protecting the host cell (Cascales et al. 2007). Colicins are predominantly encoded on plasmids, rendering horizontal transfers in the population (Hardy et al. 1973; Gordon and O’Brien 2006). Plasmidic colicin operons are implicated in plasmid maintenance (Inglis et al. 2013), bacterial suicide (Granato and Foster 2020), and other ecological phenomena such as biofilms, virulence, etc. (Lin et al. 2004; Bucci et al. 2011; Sharp et al. 2017; Weiss et al. 2020). However, the occurrence of colicin operons on bacterial genomes is intriguing because of the threats of genome degradation (by colicins) and the lysis of the host (by lysis protein).

With a gene-centered perspective, we see plasmids and the “purpose” of their genes is to enhance the propagation of the plasmids, directly or indirectly. Therefore, we consider plasmids and genomes as two independent entities and plasmids as obligate molecular parasites on bacteria that confer advantages to the host under specific conditions. In this study, we investigate the dynamics of CLB operons (more specifically, the endonuclease encoding CLB operons) in the propagation of plasmids and bacterial genomes. We aim to rationalize the role of plasmidic and genomic endonuclease CLB operons using sequence analyses and theoretical modeling from an eco-evolutionary perspective. Using nucleotide sequence homology and distribution patterns, we show that the occurrence of identical endonuclease CLB operons is mutually exclusive on genomes and plasmids, meaning that the CLB operon is present either on the plasmid or on

the genome but never both of the same cell. Similar to TAs, we propose an anti-addiction hypothesis (Saavedra De Bast et al. 2008; Ramisetty and Santhosh 2016) for endonuclease CLB operons on genomes. Here, we provide a model integrating the endonuclease CLBs as genetic arms, anti-addiction hypothesis for genomic endonuclease CLB operons and the sequence of events in the CLB operon-mediated interplay of genomes and plasmids by simulating the competition between endonuclease CLB-plasmids and bacterial genomes.

Materials and methods

Prevalence of endonuclease CLB operon

We analyzed the distribution of endonuclease CLB operons in completely sequenced bacterial genomes and plasmids listed in the NCBI database as of July 2020. We took sequences of nine endonuclease colicins (E2, E3, E4, E6, E7, E8, E9, D, and cloacin DF13) as reference queries in nucleotide homology search (Supplementary figure S1). The above-mentioned colicin sequences were taken as reference sequences to perform a nucleotide homology search against all bacterial genomes and plasmids (Supplementary Sheet S1). The sequence similarity between the reference sequences is shown using a distance tree (Supplementary figure S2). Since multiple CLBs share similar genetic features and detectable sequence homology (Riley 1998; Cascales et al. 2007), we searched for genetic modules similar to the query reference sequences showing > 50% sequence similarity. We included all the hits harboring endonuclease CLB operons irrespective of the CLB type or annotation. Since CLBs share considerably high sequence similarity, nucleotide homology searches with lower sequence identity values allow homologs from other colicin types other than the nine endonuclease colicins. First, the endonuclease colicins themselves share high sequence homology with respect to their colicin, immunity, or lysis gene. Second, few colicin operons like ColE6 harbors the E8 immunity gene in addition to its own E6 immunity gene. We then narrowed our search specifically to the endonuclease colicins by setting higher threshold values for sequence identity (~ > 80%) and coverage (~ > 90%). The first hit list was created by filtering only “complete sequence”, “chromosome”, and “genome” in MS-excel. Other insignificant hits such as draft genomes and contigs are removed from the hit list. The second hits list was created by performing individual TBLASTN for colicin, immunity, and lysis genes separately. Hits were analyzed manually for each of the gene targets. However, the hits were considered to harbor endonuclease CLB operon only if they were present in the hit list of all the three gene targets. We compared both lists to make a master list (Supplementary

Sheet S2). Our analysis weighs on the sequence homology search across all bacterial genomes and plasmids irrespective of the genome annotations and colicin type. Hits on genomes and plasmids were sorted separately and plotted as graphs. We then analyzed the endonuclease CLB operons on genomes and plasmids for conservation, prevalence, and multiplicity across genera. Genetic locus was examined for the conservation of operon (Fig. S1). The distribution of endonuclease CLB operons on both genomes and plasmids was plotted as graphs. In each genus, the hits harbouring endonuclease CLB operons either on genome or plasmid and on both genome and plasmid were grouped to verify their statistical significance. We performed a two-tailed student's t-test to test the statistical significance of the proposed mutual exclusivity hypothesis.

Conservation of endonuclease CLB operon

Using the NCBI graphics of the hit sequences, genetic loci were checked for the completeness of the operon (Supplementary Sheet S3). The analysis was based on the sequence annotations of the respective strains submitted to NCBI. In the case of annotation, artifacts for colicin activity gene, such as “HNH domain-containing gene” encoding < 200 amino acids, were considered insignificant and were not included in this study. We analyzed the sequence similarity between the endonuclease CLB operons for strains harboring the operons on both genomes and their plasmid using the MAFFT alignment method in Jalview version 2.11.0 (Waterhouse et al. 2009). For visualization and presentation of the alignments, NCBI multiple sequence alignment viewer 1.16.2 (<https://www.ncbi.nlm.nih.gov/projects/msviewer/>).

Model description

We adopted a hybrid model to understand the role of genomic endonuclease CLB operons in plasmid and genome evolution (Fig. 1a). Traditional agent-based models have “agents” with specific characteristics, and at each time step, the agents evolve individually depending on their local demography. This approach can be computationally expensive and time-demanding (Guo et al. 2008). On the other hand, modeling the complete system analytically can be complex. Therefore, established parameters such as logistic growth are analytically modeled with stochasticity, and an agent-based approach is applied to infer the spatial dependence on the population dynamics. We simulated the competition outcome between five cell types: (i) sensitive to endonuclease CLB (S); (ii) carrying endonuclease CLB operon on the plasmid (a small low-copy number plasmid-containing endonuclease colicin operon) (C); and (iii) carrying endonuclease CLB operon on the genome (C_g). Within

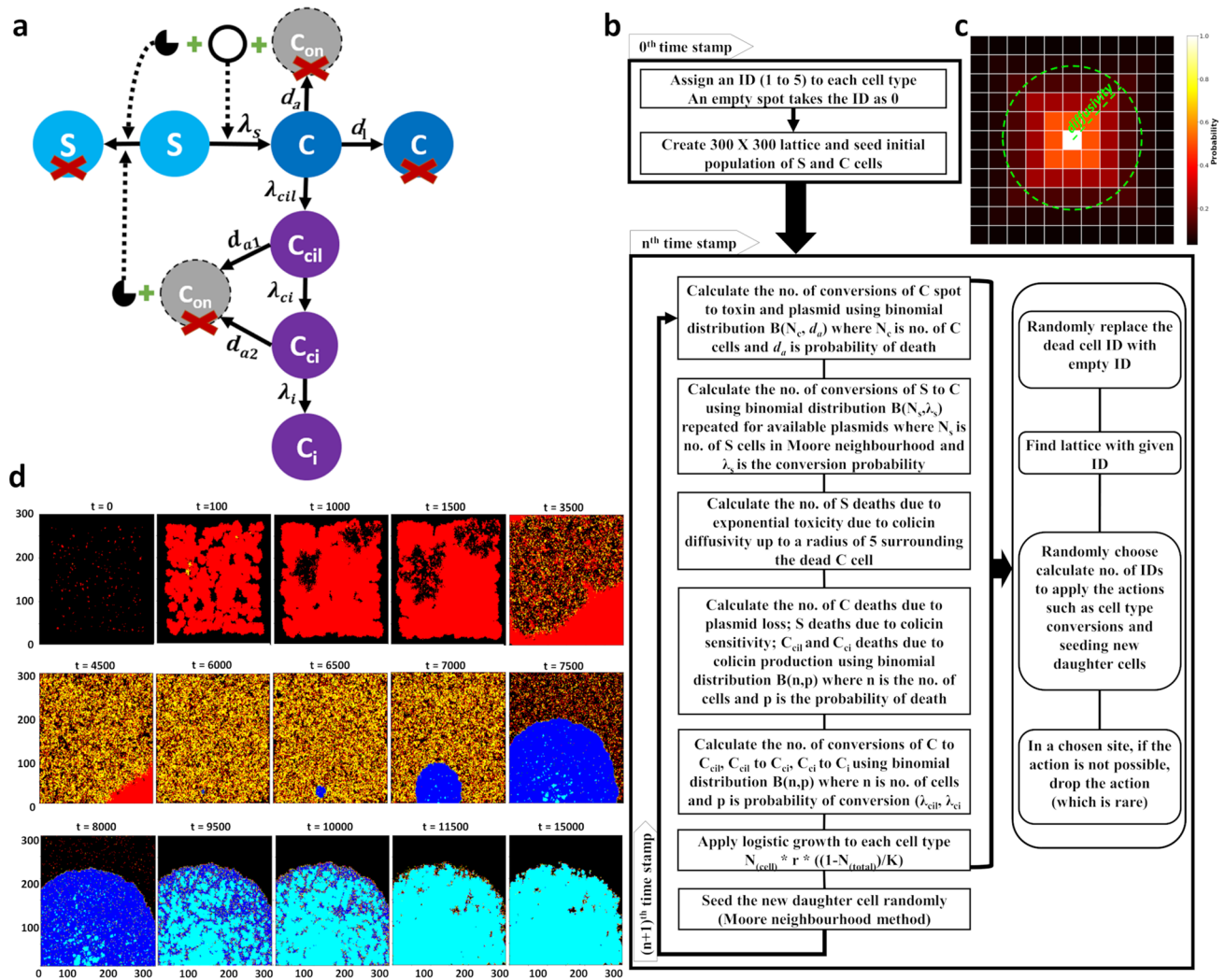


Fig. 1 Theoretical model representing the interactions between the colicinogenic and sensitive strains. **a** Upon plasmid invasion, sensitive (S) cells transform into colicinogenic (C) cells producing endonuclease CLBs. Once the endonuclease CLB operon integrates onto the genome (C_{cil} cells) at a low frequency, the cells survive upon plasmid loss. However, with the accumulation of mutations in the operon, the degenerated operon retaining only the immunity genes is selected (C_i cells). The cells containing complete operon (C_{cil}) and operon containing CLB activity and immunity genes (C_{ci}) are at the risk of operon activation (C_{on} cell) followed by death. Hence, selection works against the C_{cil} and C_{ci} cells. λ_s -Switching rate of S to C; λ_{cil} -Switching rate of C to C_{cil} ; λ_{ci} -Switching rate of C_{cil} to C_{ci} ; λ_i -Switching rate of C_{ci} to C_i ; d_a -death rate of C due to operon activation; d_{a1} -death rate of C_{cil} due to operon activation; d_{a2} -death rate of C_{ci} due to operon activation. **b** Simulation flowchart. Step-wise work of the simulation is detailed. **c** Exponential toxicity of CLB released. The central white grid represents the CLB released by the dead colicinogenic cell, which is now the toxic spot to the surrounding grids up to a radius of 5. With the increasing radius, the toxicity (or the CLB-mediated death probability) decreases by the factor $1/2^{\text{radius}}$. **d** Lattice showing population dynamics. The lattice structure is shown at specific time points (t in hours). The colors represent black-empty spots; red-S cells; yellow-C cells; dark blue- C_{ci} cells; cyan- C_i cells

the C_g population, we show three variations: the population that (i) has complete operon (C_{cil}); (ii) has endonuclease CLB activity and immunity genes but not lysis gene (C_{ci}); and (iii) has only immunity gene (C_i). As a proof-of-concept, the framework of the model is adapted (with modifications) from a previously published dataset from in vitro competitive experiments with *E. coli* colicinogenic strains (von Bronk et al. 2017; Weiss et al. 2020). We assigned

the model parameters for simulation from the experimental values. Initial communities were seeded in a 375:5 (S: C) ratio in 300 × 300 lattice. Five different agents were used in this model: sensitive S cells, colicinogenic C, and C_g (C_{cil} , C_{ci} , and C_i) cells. Each cell type is considered an individual characterized by a particular status behavior. The model includes (i) reproduction of viable S, C, and C_g cells, (ii) lysis of C cells due to operon activation (λ_c) and subsequent

release of endonuclease CLBs and plasmids, (iii) death of C cells due to plasmid loss (Million-Weaver and Camps 2014), (iv) switching of C to C_g (C_{cil} or C_{ci} or C_i at equal probability) cells (λ_g), (vi) lysis of C_{cil} and C_{ci} cells with a release of endonuclease CLB. The cell state with activated endonuclease CLB operon is represented as a C_{on} cell that eventually dies and releases endonuclease CLB plasmids and endonuclease CLBs. The switching frequency between the cell types depends on the varying probability (through parameter sweeping), availability of the primary cell type and the number of endonuclease CLB plasmid (released by dead C cells). We have not considered the endonuclease CLB operon degeneration on plasmids because of complications of the copy number and relatively high probability of mutations and recombination. We expect that the cost incurred by the endonuclease CLB plasmid, irrespective of the completeness of the operon, renders plasmid elimination upon the emergence of genomic endonuclease CLB operons.

Simulation of the competition

We simulated the competition in Python 3.7. We took a hybrid approach to realize the simulation. We simulated the model in time steps of 1 with a total number of time steps of 15,000. Each cell is assigned a unique ID in the simulation. The competition is modeled using a 2D lattice, representing the width and height of the bacterial niche. We adapted an agent-based stochastic model to track the fate of each cell type (agents) as a function of space and time.

The entire simulation is summarized (Fig. 1b) below.

Simulation time step 0:

- A Grid (300×300) is initially generated. All of the Grids are initialized with empty spots ($ID=0$).
- Random spots are chosen and are initialized with S ($ID=1$) and C ($ID=2$) cells in the ratio of 375:5 (S:C). This step is an agent-based approach that models the spatial dependence of the cells.

Simulation from time step 1 to 15,000 (the following steps are repeated in a loop till the end of the time stamp):

- The number of cells with their unique IDs is calculated. Here, the simulation dynamics are governed by the number of cells calculated.
- To model the consequences of the death of C cells, we calculated the number of conversions of C cells to toxins and endonuclease CLB-plasmids using binomial distribution $B(N_c, d_a)$, where N_c is the number of C cells, and d_a is its death probability. Here, the C cells are randomly chosen for death. Such randomness induces spatial stochasticity in the simulation.

The number of endonuclease CLB-plasmids released by dead C cells is assumed between 10 and 15 copies per cell.

- When a C cell dies and releases toxins, it kills S cells based on the Moore neighborhood (Gray 2003) of C. This simulation governs the toxic interactions between the S cells and the toxic lattice sites by assuming exponential toxicity (Weber et al. 2014). The likelihood of No of S cells that die due to the C cells exponentially decreases as per the Moore neighborhood of the dead C cell and toxicity (tox). For exponential toxicity, the population of S cells within the radius of 5 is considered (exponential factor is $1/2^{\text{radius}}$); reducing the death probability exponentially ($prob(0-1) < (tox)^r$) with the increasing radius (r) (Fig. 1c).
- The endonuclease CLB-plasmids released by the dead C cell can invade the S cells in the vicinity and convert S cells to C cells. To model the dependence of conversions of S cells on the available plasmids, we calculated the number of S to C cells using binomial distribution $B(N_s, \lambda_s)$ repeated for the number of available endonuclease CLB-plasmids. Here, N_s is the number of S cells in the Moore neighborhood of a dead C cell and λ_s is the conversion probability. This is another instance where spatial stochasticity is modeled using an agent-based approach.
- C cells also convert to C_{cil} cells. C_{cil} cells convert to C_{ci} , and C_{ci} converts to C_i cells. These are modeled using a random number of cells based on binomial distribution with mean corresponding to respective conversion rates. Again, these analytically generated numbers (with stochasticity from the binomial random number generation) are applied in the grid using the agent-based approach.
- The C_{cil} and C_{ci} also have death probability, and the released toxin is toxic to the S cells in their vicinity.
- Finally, the total number of surviving cells are grown logistically as,

$$\frac{dn}{dt} = rn \left(1 - \frac{N}{K} \right)$$

where r is the growth rate, n is the number of each cell type, N is the total population size, and K is the carrying capacity. The multiplication of a cell is modeled using a Moore neighborhood (eight nearest neighbors) method where a new cell takes a random spot in its nearest neighbors. Each cell type is allowed to grow logistically yet limited by the total cell density.

The simulation is repeated with varying parameters (Table 1). A parameter sweeping of the parameters was done to trace out boundary parameters for which C_i can dominate the simulation. As both spatial stochasticity and analytical stochasticity are modeled, the whole simulation is iterated ten times to analyze the simulation's behavior statistically.

The dominating cells (population of a cell type > 90%) at the end of the simulation are recorded for each iteration. A given simulation setting is stable when the same cell type dominates more than 50% of the iterations. The lack of experimental data is compensated by testing the competition outcome for a range of reaction rates. We have tested the model with low probability values. For example, the plasmid invasion probability is as low as 5 out of 10,000 S cells take up the endonuclease CLB plasmid and get converted to C cells. Visualization of one of the iterations is shown (supplementary video, Fig. 1d).

Results

Distribution of endonuclease CLB operons across bacterial genera

To examine the prevalence of genomic and plasmidic endonuclease CLB operons, we performed a nucleotide homology search taking nine endonuclease colicins (E2, E3, E4, E6, E7, E8, E9, D, and cloacin DF13) as reference sequence against all bacterial genomes and plasmids. Endonuclease

colicins are translocated into the cytoplasm and degrade the nucleic acids. The colicin E2, E7–E9 targets the DNA, E3, E4, E6, and CloDF13 targets the rRNA while D cleaves the tRNA (Cascales et al. 2007). Within the limitations of available sequences (as of July 2020), we obtained 639 genomes and 627 plasmids hit across 30 bacterial genera harboring partial to complete operons (Fig. 2a). All 30 genera are Gram-negative *Gammaproteobacteria* (belonging to families including *Enterobacteriaceae*, *Morganellaceae*, *Yersiniaceae*, *Pseudomonadaceae*, *Pectobacteriaceae*, *Vibrionaceae*, *Erwiniaceae*, and *Pasteurellaceae*) and *Betaproteobacteria* (family *Burkholderiaceae*). The total occurrence of endonuclease CLB operons (both genomes and plasmids) is high in *Klebsiella* (412 operons) followed by *Salmonella* (198 operons), *Pseudomonas* (186 operons), and *Escherichia* (163 operons).

Plasmidic endonuclease CLB operons were prevalent among the strains of *Klebsiella* (61%) and *Escherichia* (26%). Genomic endonuclease CLB operons were prevalent among the strains of *Pseudomonas* (30%), *Salmonella* (28%), and *Yersinia* (15%) (Table S1). The strains of *Klebsiella*, *Citrobacter*, *Salmonella*, and *Enterobacter* harbored endonuclease CLB operons on both genome and plasmid.

Table 1 Model parameters and values used in the study

Global parameters				
Lattice dimension	300 X 300			
Time step Δt [hours]	1			
Total simulation time T	15,000			
Iterations	10			
Carrying capacity, K	60,000			
Copy number of Colicin-like Bacteriocin plasmid	20			
Diffusivity of toxin	5			
Mean variation (to account for uncertainty)	0.25			
Simulation parameters (includes range of tested values)				
For S				
Initial abundance	375			
Growth rates	0.0729			
Switching rates to C (λ_s)	0.0005	0.001	0.005	0.01
For C				
Initial abundance	5			
Growth rates	0.0729			
Operon activation rates (C to C_{on}) (d_a)	0.008			
Switching rates to C_{cil} (λ_{cil})	10^{-8}	10^{-7}	10^{-6}	10^{-5}
Death rate due to plasmid loss (d_l)	0.001			
For C_{cil} , C_{ci} , and C_i				
Initial abundance	0			
Growth rates	0.0729			
Death rate of C_{cil} due to colicin activation (d_{a1})	0.006			
Death rate of C_{ci} due to colicin activation (d_{a2})	0.002			
Switching rates of C_{cil} to C_{ci} (λ_{ci})	10^{-6}	10^{-5}	10^{-4}	
Switching rates of C_{ci} to C_i (λ_i)	10^{-6}	10^{-5}	10^{-4}	

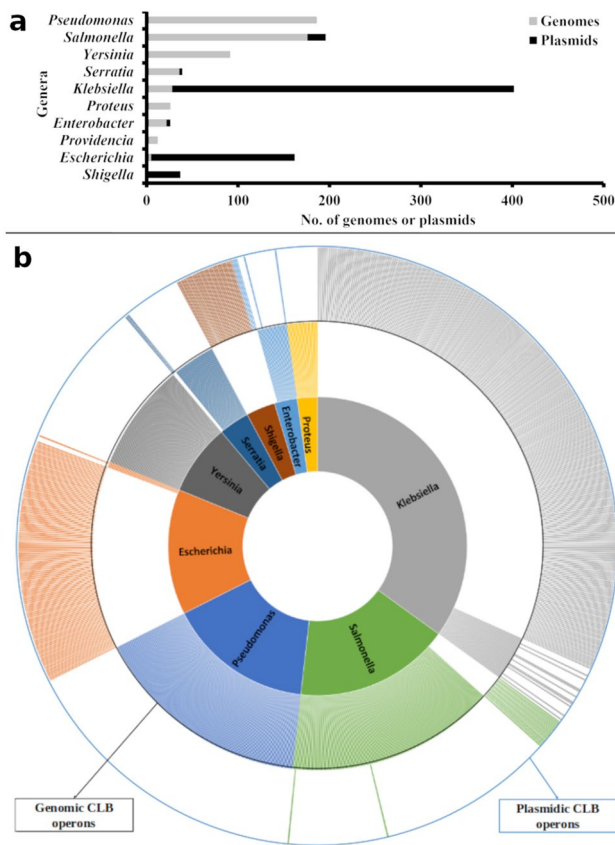


Fig. 2 **a** Distribution of CLB across bacterial genera. We obtained 1266 genome and plasmid hits (from the genera *Enterobacteriaceae*, *Morganellaceae*, *Yersiniaceae*, *Pseudomonadaceae*, *Pectobacteriaceae*, *Vibrionaceae*, *Erwiniaceae*, *Pasteurellaceae*, and *Burkholderiaceae*) (supplementary table 1). Here, we represent the distribution for genera comprising more than ten colicinogenic strains. **b** Mutual exclusivity of CLB on genomes and plasmids. The hit list was sorted to distinguish whether strains carried endonuclease CLBs on the genome or plasmid. The data visualization is done using a sunburst graph. Different color codes are given to each segment specific to genera (labeled). The outer segment represents the plasmids for each genus, and the inner segment represents the genomes. We performed a two-tailed student's *t* test to test the statistical significance of the endonuclease CLB operons occurring mutually exclusive on genomes and plasmids. Of the observed number of strains (hits), the occurrence of endonuclease CLB operon either on genome or plasmid is statistically higher than its occurrence on both genome and plasmid [*p* value (two-tail) is 0.017]

By setting up the threshold values for sequence identity (> 90%), we narrowed our analysis to nine endonuclease type-specific colicins (E2, E3, E4, E6, E7, E8, E9, D, and cloacin DF13). Of the 240 strains encoding endonuclease colicins, cloacin DF13 is highly prevalent (78%) and distributed among the *Klebsiella* genus (fig. S3). We rarely found endonuclease colicin operon on the genome. Endonuclease colicins E2, E8, and CloDF13 were found on the genomes of *Escherichia*, *Shigella*, *Klebsiella*, and *Citrobacter*. Very

few (3%) strains of *Shigella* harbored endonuclease colicin operons on plasmids (Supplementary Sheet S1).

Mutual exclusivity of plasmidic and genomic endonuclease CLB operons

To investigate if the endonuclease CLB operons are present on both genomes and plasmids, we analyzed their distribution pattern across all the genera. We found that some genera had endonuclease CLB operons only on genomes, and some had only on plasmids (Fig. 2b). For example, we observed only genomic endonuclease CLB operons in *Pseudomonas* strains while only plasmidic endonuclease CLB operons in *Shigella*. Whereas in strains of *Klebsiella* (10 strains), *Escherichia* (1 strain), *Salmonella* (2 strains), and *Enterobacter* (2 strains), both genomic and plasmidic endonuclease CLB operons were present (Supplementary Sheet S4). We examined the operon sequence similarity and conservation to rationalize the multiplicity of endonuclease CLB operons within a strain. We found no strains encoding multiple copies of identical endonuclease CLB operons on the genome and plasmids (Fig. S4). For example, *E. coli* BR10-DEC harbors colicins B and M on the genome, but colicin E1 on its plasmid. *Citrobacter koseri* AR_0024 harbors incomplete colicin D on the genome but a CloDF13-like operon on its plasmid. Few strains harbored more than one colicin type. For example, the genome of *E. coli* 15RDA-Livestock feces-ECO087 harbors both colicin E2 and E8. We also observed 85 strains harboring multiple endonuclease colicin plasmids but varied in the type of colicins they encode (Supplementary Sheet S5). For example, *E. coli* ACN001 harbors colicins K, E9-like, and E2-like operons on its plasmids pACN001-E, pACN001-D, and pACN001-B, respectively. *K. pneumoniae* NY9 harbors klebicine B and Cole3 on its plasmids pNY9_1 and pNY9_4, respectively. The occurrence of endonuclease CLB operons either on genomes or plasmids but not on both is statistically significant ($p=0.017$). Thus, the occurrence of endonuclease CLB operons on either genome or plasmid is greater than its occurrence on both genome and plasmid. Except for *K. pneumoniae* subsp. *pneumoniae* KPNIH24, which harbors CloDF13 on both its genome and plasmid, no other bacterial strain had identical endonuclease CLB operons on its genomes and also on plasmids. The genome of *K. pneumoniae* subsp. *pneumoniae* KPNIH24 harbors two copies of CloDF13, but the colicin gene encodes 273 aa (lacks 5' end of protein sequences) in contrast to its reference sequence that encodes 561 aa.

Degeneration of genomic endonuclease CLB operon

The genetics of endonuclease CLB operon showed, out of 620 genomic endonuclease CLB operons, five were

complete operons, 614 lacked lysis gene, and 323 lacked the CLB activity gene. In contrast, the immunity gene was conserved in almost all operons. Upon comparing the gene conservation on plasmids and genomes, we found that most genomes lacked lysis genes or harbored only immunity genes (Fig. 3). To account for the occurrence of genomic endonuclease CLB operons, we compared the conservation of genomic and plasmidic operons. Endonuclease CLB operon on genomes is highly degenerative. For example, in *K. pneumoniae* TK421, the CloDF13 activity gene on the genome is replaced by a transposase gene, whereas its plasmid contains a complete CloDF13 operon. *Enterobacter rogenkampii* R11 harbors an S-type pyocin operon on its plasmid and an orphan immunity gene on its genome. An exception is that *K. pneumoniae* subsp. *pneumoniae* KPNIH24 harbors CloDF13 on its genome and plasmid (fig. S5). However, the integrity and functionality of these operons are unknown. The endonuclease CLB operons lacking functional lysis gene represent the truncation of lysis gene or other colicin types (group B colicin that typically does not encode lysis gene). We observed high conservation of the immunity gene on genomes (e.g., *Mixta theicola* SRCM103227) compared to CLB activity and lysis genes. We also compared the specific loci with and without the endonuclease CLB operon within a species (Fig. 3a). Similar comparisons can trace the ancestral loci prior to the integration of endonuclease CLB operon and the putative mechanisms of integration. Several truncations in the operon genes could account for the intermediates in the degeneration process. We did not find any endonuclease CLB operon lacking immunity gene on the genomes, as the absence of immunity protein is detrimental to the host.

The competitive advantage of genomic endonuclease CLB operons

Considering plasmids and genomes as competing entities, we developed a stochastic agent-based model (Kerr et al. 2002; Majeed et al. 2011) to simulate the competition outcome between plasmids and genomes regulated by the endonuclease CLB operon (Fig. 1). As a proof-of-concept, we adapted the model's framework from a previously published dataset from in vitro competitive experiments with *E. coli* colicinogenic strains (von Bronk et al. 2017; Weiss et al. 2020). Initially, we considered three cell types with the identical genetic background but differed in: (i) absence of endonuclease CLB operon (S-cell); (ii) carrying endonuclease CLB operon on a low-copy number plasmid (C-cell); and (iii) carrying endonuclease CLB operon on the genome (C_g cells) (Fig. 1a). Stochastic fluctuations in the parameters strongly affect the competition outcome (Shimoni et al. 2009). We have incorporated the uncertainty into the model by testing a range of values for theoretical parameters

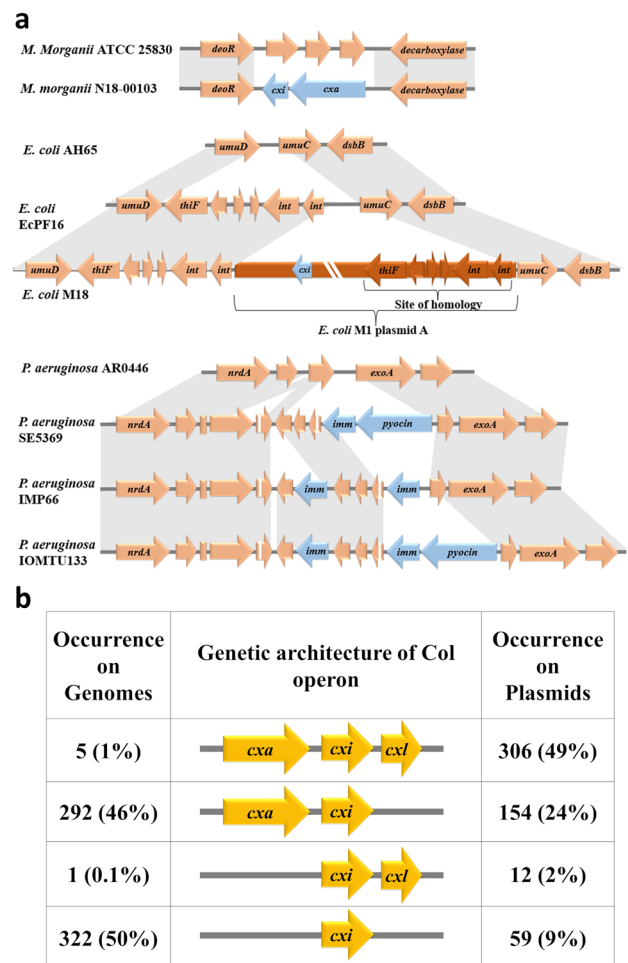


Fig. 3 Genetic architecture of the operon. **a** Integration of endonuclease CLB operons on genomes. In *Morganella Morganii* N18-00103, the CLB operon is flanked by *deoR* (deoxyribose operon repressor) and α -decarboxylase genes. We found *Morganella* strains lacking the operon (e.g., ATCC 25830) instead harboring three hypothetical genes showing no sequence similarity with the CLB operon. Likely that the operon integrated by replacing the region between the flanking genes. In *Escherichia coli* M18, the complete CLB plasmid (*E. coli* M1 plasmid A) has been integrated at the homology site between the DNA polymerase V core genes (*umuD* and *umuC*). In *Pseudomonas aeruginosa* IOMTU133, two CLB operons (pyocin-like operon and Col E3-like gene) have independently integrated with proximity. We found the integration of multiple genes between *nrdA* (ribonucleoside-diphosphate reductase subunit alpha) and *exoA* (exotoxin A). For example, *P. aeruginosa* SE5369 harbors only pyocin-like operons. Whereas *P. aeruginosa* IMP66 has only Col E3-like gene but lacks a pyocin-like operon. **b** Degeneration of endonuclease CLB operon on genomes. Here, we compare the different genetic organizations of the operons on genomes and plasmids. On the majority of the genomes, the immunity gene is orphaned. On the majority of the plasmids, the endonuclease CLB operon is complete, including the lysis gene. We observed 49% of plasmidic endonuclease CLBs and nearly 1% of genomic endonuclease CLBs are complete operons. Endonuclease CLB operon on genomes is highly degenerative. Of the genomes carrying incomplete operon, 614 lacked a functional lysis gene

(with binomial probabilities) (Table 1, Fig. 1b, c). Parameter sweeping tests the competition outcome for a range of parameter values abiding by Monte Carlo realizations, allowing statistical evaluation of the model (Fig. 1b). Our model depicts the interactions between the Colicinogenic and sensitive population and the competition outcome upon invasion by endonuclease CLB plasmids. In the theoretical model, the switching rate of S to C represents the plasmid invasion stage, the death rate of C due to plasmid loss represents the plasmid addiction stage, and the switching rate of C to C_g represents the integration stage. We did parameter sweeps to check the influence of invasion rate of endonuclease CLB plasmid (λ_s), death of C due to plasmid loss rate, operon activation rate in C (λ_c), acquisition rate of endonuclease CLB operon on the genome (λ_g), and CLB toxicity (tox) on the competition outcome. We tested our model output for the otherwise possible range of values for each parameter to avoid biased observations and interpretations. The reduction in the sensitive population represents the plasmid invasion where the plasmid is advantaged at the cost of the genome. The winning population of C represents the quasi-cooperation (endonuclease CLB-plasmids benefit only the associated genomes by providing immunity to the toxin), where both plasmids and genomes are benefited (Fig. 4). Thus, genomic endonuclease CLB operons are selected as a counter-strategy that eliminates the burden of endonuclease CLB plasmid imposed via addiction. In an evolutionary sense, genomic endonuclease CLB operons are no longer required after anti-addiction but for protection against the exogenous endonuclease CLBs. We then incorporated three genomic endonuclease CLB operon variations in the model: C_g population with (i) complete operon (C_{cil}); (ii) CLB activity and immunity genes but not lysis gene (C_{ci}); and (iii) only immunity gene (C_i) (Fig. 1a). Using line graphs, we visualized the dynamics of total number of endonuclease CLB plasmids and genomes at each time point in the simulation (Fig. 5). In coherence to in-silico analysis showing the high conservation of the immunity gene over CLB activity and lysis genes, the model predicts the outcome of the competition among endonuclease CLB sensitive, plasmidic endonuclease CLB containing population and genomic endonuclease CLB containing population (C_{cil} , C_{ci} and C_i) in varied conditions (Figs. 1d, 4).

Gene-centered model for endonuclease CLBs in plasmid–genome conflicts

The eco-evolutionary cycle of endonuclease CLB plasmids and host genomes includes plasmid invasion, plasmid addiction, quasi-cooperation between the plasmids and genomes, plasmid loss via anti-addiction (Fig. 6). During plasmid invasion, the plasmid is advantaged as it propagates along with its host genome. With the establishment of plasmid

addiction, only plasmid-bearing cells are selected in the population against plasmid-free clones despite the metabolic burden incurred by the plasmid, because the plasmid-free cells would be eliminated by the secreted endonuclease CLBs. The CLB proteins induce lethality in plasmid-free cells resulting in the elimination of competition. The plasmid-containing cells are resistant to exogenous CLBs due to the neutralizing effects of the immunity protein. A daughter cell that fails to inherit a copy of the plasmid becomes sensitive to the lethality of exogenous CLBs as there is no source for the production of cognate immunity protein (plasmid addiction). Through the elimination of cured cells and plasmid-free cells, plasmids increase their maintenance within the population favoring genome propagation along with plasmid propagation (Quasi-cooperation). The endonuclease CLB operon has two promoters, an SOS promoter upstream to the CLB activity gene and a constitutive promoter upstream to the immunity gene (Cascales et al. 2007). Upon DNA damage, the operon produces two mRNA transcripts as there are two terminators. The minor mRNA corresponds to the complete operon, including the lysis gene. The major mRNA corresponds to the CLB activity gene along with its immunity gene. The translation of the major mRNA transcript produces the CLB activity and immunity proteins which then associate to form dimeric complexes lacking the enzymatic activity. The constitutive promoter of the immunity gene allows constant production of immunity protein, ensuring no free CLBs inside the host (Cavard et al. 1985; Dawkins 1989). The CLB promoters are activated at low-level DNA damage leading to the production of bacteriocin in a large proportion of the population (Mavridou et al. 2018). In conditions of host DNA damage and the induction of SOS response, the increased expression from plasmid-encoded CLBs causes lethality, without which the bacteria would have grown better (Gillor et al. 2008). The lysis gene expression results in the host cell lysis and the release of CLB proteins and plasmids. The released plasmids have an opportunity to be taken up by other cells. Occasionally, the CLB operon may integrate into the chromosome through non-specific means associated with transposons or complete plasmid integration. In similar cases, the host is relieved from plasmid addiction. The chromosomal CLB operons can act as anti-addiction systems by providing immunity protein to protect the host from exogenous CLBs. Here, the host genome is advantaged because the cured daughter cells are resistant to the endogenous and exogenous CLBs, and there is a reduction in the metabolic burden incurred by the plasmid. However, the CLB activity and lysis genes are costly to maintain because of lethality through DNA degradation and cell membrane lysis, respectively. The immunity protein is essential as long as the exogenous CLBs are present in the niche. Therefore, the lysis and colicin genes are lost from most chromosomal CLB operons. The weapons

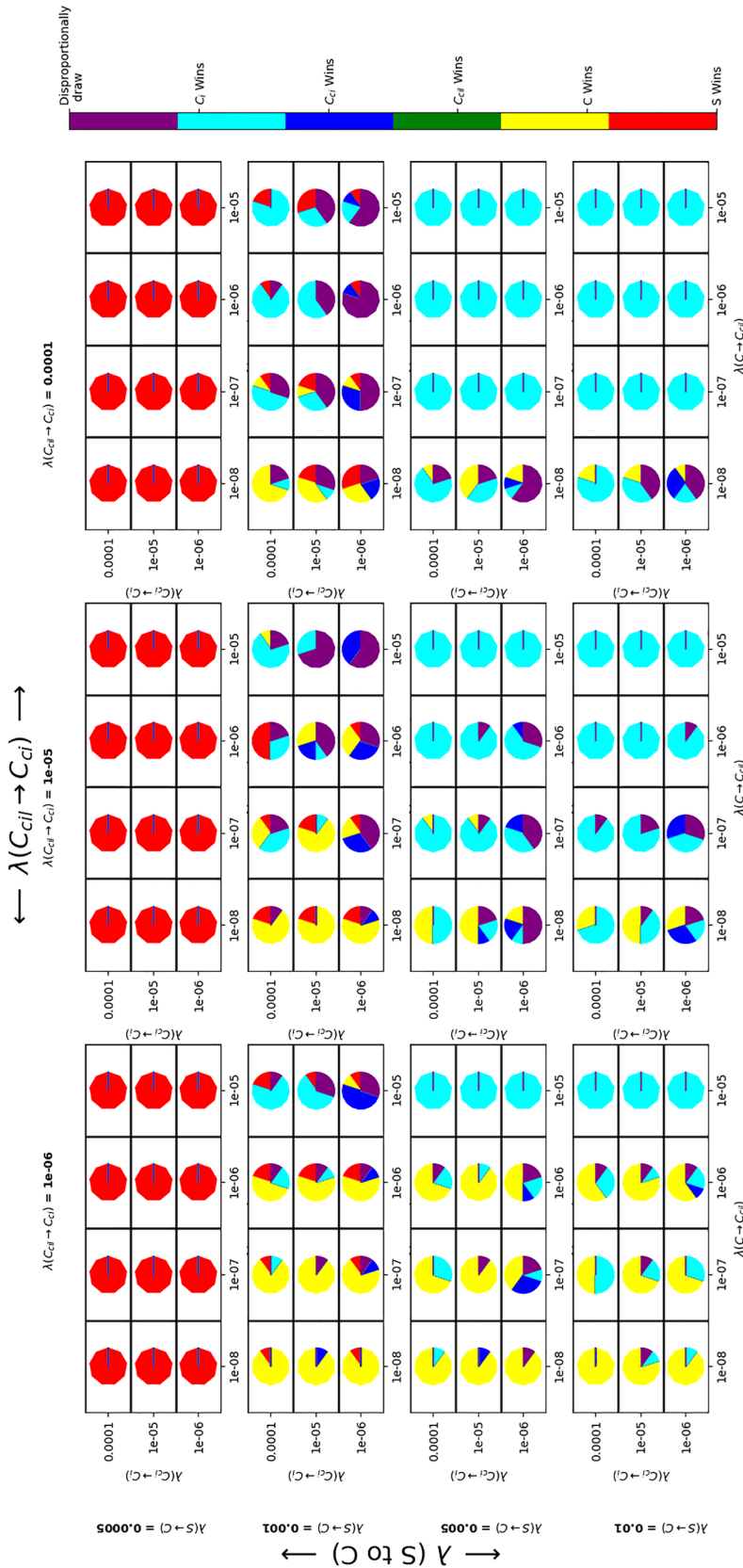


Fig. 4 Simulation outcome of the competition between the cell types. A combined pie chart shows the overall simulation outcome with four test parameters (with sweeping). The four test parameters include conversion of S to C, C_{cil} to C_{ci} , C to C_{ci} , and C_{ci} to C. The outermost Y-axis represents the S to C conversion with increasing probabilities. The outermost X-axis represents the C_{cil} to C_{ci} conversion with increasing probabilities. Each subplot contains X and Y axes, representing C to C_{cil} and C_{ci} to C_i conversions with increasing probabilities. Stochasticity in the model is introduced through binomial probabilities for every parameter. The simulations with parameter sweeping were iterated ten times for 15,000 h. The color code represents the respective cell type winning. A cell type is considered a ‘winner’ if it constitutes more than 90% of the population in 5 out of 10 iterations. The population is considered ‘disproportionately draw’ if none reaches 90% proportion

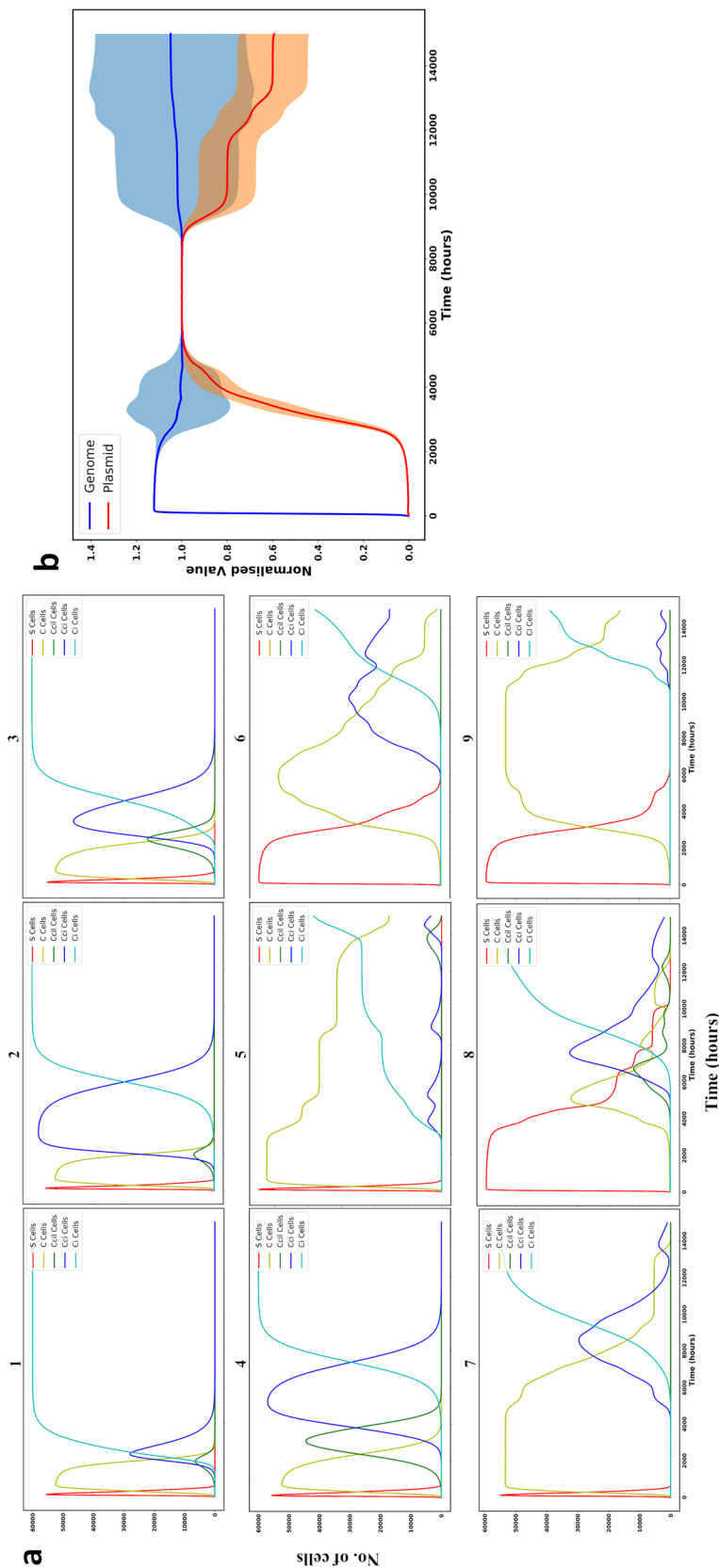


Fig. 5 Modelling and simulating the implications of CLB in plasmids-genomes conflict. **a** Line graph depicting the population dynamics of each cell type. Each line graph is plotted by varying the values of each of the four tested parameters. The value of the parameters used in each graph is given in the order: S to C conversion, C death due to plasmid loss, C to C_{ci}, C_{ci} to C_{cil}, and C_{cil} to C_i conversions, respectively. a.1 10^{-2} , 10^{-3} , 10^{-5} , 10^{-4} , 10^{-4} , 0.5; a.2 10^{-2} , 10^{-3} , 10^{-5} , 10^{-4} , 10^{-6} , 0.5; a.3 10^{-2} , 10^{-3} , 10^{-5} , 10^{-5} , 10^{-3} , 10^{-5} , 10^{-6} , 0.5; a.4 10^{-2} , 10^{-3} , 10^{-5} , 10^{-5} , 10^{-6} , 0.5; a.5 10^{-2} , 10^{-3} , 10^{-6} , 10^{-6} , 10^{-4} , 0.5; a.6 10^{-3} , 10^{-4} , 10^{-8} , 10^{-4} , 10^{-8} , 10^{-5} , 10^{-5} , 0.5; a.7 10^{-2} , 10^{-4} , 10^{-8} , 10^{-4} , 10^{-5} , 10^{-5} , 10^{-5} , 0.5; a.8 10^{-3} , 10^{-3} , 10^{-5} , 10^{-5} , 10^{-8} , 10^{-8} , 10^{-4} , 0.5. **b** The dynamics of the number of plasmids and genomes over the period are plotted as a line chart. The total number of S, C, C_{ci}, C_{cil}, and C_i cells was taken to represent the total number of genomes and the number of C cells alone to represent the total number of plasmids. Simulations were iterated ten times, considering the parameter values that show uncertainty in the game outcome, and the statistical summary of the output is shown as a line. The additional lines above and below the mean line represent the variability (mean variation of 25%) in the data at each plotted point. Two subtle regions (between 2000 and 4000 h and after 9000 h) in the graph represent the uncertainty wherein genomes or plasmids get equal chances to outcompete the other. Parameters and values used: S to C conversion is 10^{-3} , C death due to plasmid loss is 10^{-4} , C to C_{ci} conversion is 10^{-8} , C_{ci} to C_{cil} conversion is 10^{-5} , C_{cil} to C_i conversion is 10^{-5} , and CLB toxicity is 50%

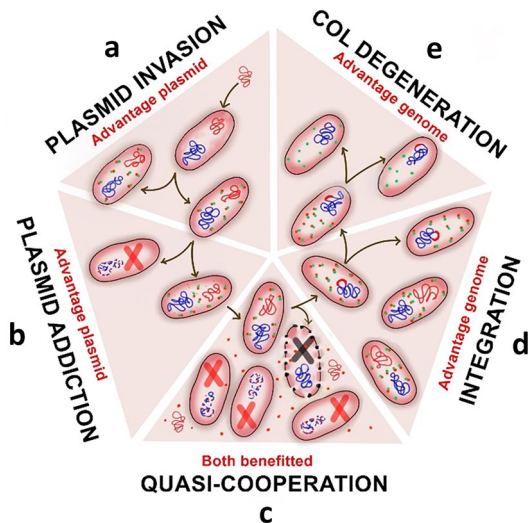


Fig. 6 The eco-evolutionary cycle of CLB plasmids and host genomes. The ecological cycle can be divided into five theoretical stages. **a** Plasmid invasion. The plasmid is advantaged by its association with the host genome. **b** Plasmid addiction. Cells cured of the plasmid are eliminated from the population. The plasmids are advantaged at the cost of host cells. **c** Quasi-cooperation. The plasmid-containing cells are resistant to exogenous endonuclease CLBs due to the neutralizing effects of the immunity protein. They have increased genome propagation along with plasmid propagation. It is ‘quasi’ because the plasmid-free clones of the host cell are also killed. **d** Chromosomal integration. The chromosomal endonuclease CLB operons can act as anti-addiction systems by providing immunity protein to protect the host from exogenous endonuclease CLBs. The host genome is advantaged because the cured daughter cells are no longer killed by the plasmid-encoded endonuclease CLBs and reduce the metabolic burden of harboring the plasmid. **e** Endonuclease CLB operon degeneration. Selection against the lethal gene (lysis gene) results in stabilizing the genomic CLB operons in the population. The genome is in blue, and the plasmid is in red. The red and green dots represent CLB and its cognate immunity protein, respectively. The red cross represents the death of CLBs-sensitive cells by exogenous CLBs, and the black cross represents the lysis of colicinogenic cells following the endonuclease CLB operon activation

(CLB operons) will determine the winners and outcomes in a conflict (between plasmids and genomes) and by the end of which begins the process of disarmament to reduce the burdens and costs of the weapons (degeneration of genomic CLB operons).

Discussion

The endonuclease CLBs occur predominantly on bacterial plasmids and genomes. However, the presence of endonuclease CLB operons in a given bacterium will be either on the plasmid or the genome, but rarely both. The genomic endonuclease CLB operons are highly degenerated except for the immunity gene which is highly conserved. The causes and consequences of the genomic CLBs and their degeneration

is of evolutionary interest in genome–plasmids conflicts. Our simulation study shows that with the emergence of cells harbouring genomic CLBs in the population, the CLB plasmids no longer impose addiction. Furthermore, cells encoding the immunity protein (but not) outcompetes all other counterparts during the course of time.

The frequency and propagation patterns of genes are dependent on the consequences of the gene product activity, directly or indirectly. For example, it was recently shown that the relative gene frequency on plasmids is dependent on the contribution of the gene product to the fitness of the host cell, the time of acquisition and horizontal gene transfer rates (Lehtinen et al. 2021). Most genes confer direct survival or growth advantage by increasing the fitness of the bearers. The genetics of such genes is simple; the conservation or propagation of those genes is proportional to the degree of fitness conferred by the genes. Some genes, like endonuclease CLB operons, confer advantages indirectly by eliminating or modulating the competitors. Thus, endonuclease CLB operons are ‘genetic arms’ that encode a toxin–antidote pair and a lysis gene, mediating the conflicts between plasmids and genomes. Therefore, the frequency and distribution patterns of endonuclease CLB operons are deviant from the expected trends and are subject to eco-evolutionary dynamics and contexts in terms of plasmid–genome conflicts. The objective of this work is to decipher the rationales that operate in the propagation of endonuclease CLB operons. This study is more relevant in growth conditions where there is no selection for plasmid-encoded traits or plasmids with no beneficial traits to the host.

Endonuclease CLB genes are abundant on plasmids (about 627 out of 1266 total colicinogenic strains) and equally on bacterial genomes, indicating efficient propagation and potential for horizontal gene transfer. Endonuclease CLB genes are found predominantly in *Gammaproteobacteria* (Fig. 2a), indicating a form of specificity which could be a restriction due to their association with certain plasmids and their respective host range. The host range of the plasmids is restricted by the ability of the replication initiation of the plasmids within the host, i.e., the origin of replication should be recognized by the host cell replication machinery (Jain and Srivastava 2013). The abundance of endonuclease CLB operons implies that the consequences of the operon products as a unit are beneficial for the propagation of the bearer. Being associated with plasmids increases their probability of transfer between bacterial cells and across replicons. Plasmidic endonuclease CLB operons impose ‘addiction’ (Inglis et al. 2013), a phenomenon where the cured daughter cell is killed due to the toxic effects of the endonuclease CLB proteins. Despite the metabolic burden on the host, addiction ensures the maintenance of plasmid-bearing cells in the population that seems like a mutualistic interaction but essentially not in conditions non-selective for plasmidic

traits. Conceivably, a bacterium harboring a endonuclease CLB plasmid has limited possibilities; (i) to bear with the plasmid due to ‘addiction’ with a relative risk of being out-competed, (ii) to risk death upon loss of plasmid due to the lack of immunity protein and (iii) attain immunity by acquiring endonuclease CLB operon on to the genome. Those bacteria that lose the endonuclease CLB plasmid are eliminated from the population. The survivors are those that either harbor the plasmid or gain the endonuclease CLB operon on the genome and are thus abundant.

We found a distribution pattern of endonuclease CLB operons among the plasmids and genomes of *Gammaproteobacteria*. Among several genera, the endonuclease CLB operons are predominant either on the genome or on the plasmids (Fig. 2b). In *Pseudomonas* and *Salmonella*, endonuclease CLB operons are predominant on genomes, while in *Klebsiella* and *Escherichia*, endonuclease CLB operons are predominant on plasmids. We further explored this distribution pattern at the level of individual endonuclease CLB operons in relation to their functions. In theory, all infected bacteria have to either retain the plasmid or die due to CLBs. ‘Forced’ plasmid maintenance (addiction) is similar to being parasitized by phages (Shapiro et al. 2016). Phages are parasites that prey on their host through the lytic cycle or lysogeny for maintenance and propagation in the bacterial population. Bacterium cured of lysogenic prophage is lysed by the mature phage and is at imminent risk of secondary phage infection. In nature, the probability of CLB plasmid-harboring bacteria being cured of that plasmid is low because of addiction: a choice of plasmid maintenance or death. The presence of CLB operon on the genome will allow plasmid curing through anti-addiction, a phenomenon characterized by protection of the cured host by the genome encoded immunity protein. We devised two premises to test the anti-addiction hypothesis. Premise 1: a bacterium harboring CLB plasmid would not have an identical CLB operon on its genome. Premise 2: a bacterium with genomic CLB operon would not harbor a plasmid-encoding identical CLB operon. Within the limits of the available nucleotide sequences in the NCBI database, we found only one instance where there is an identical sequence on the plasmid and two copies on the genome in *Klebsiella pneumoniae* subsp. *pneumoniae* KPNIH24. Statistically, our findings are quite significant, supporting the hypothesis that the genomic endonuclease CLB operons serve as anti-addiction modules. These observations were similar to the anti-addiction hypothesis of Toxin-antitoxin systems (Ramisetty and Santhosh 2016; Horesh et al. 2020).

Integration of the CLB operon on the genome may allow the plasmid curing and provide immunity from exogenous CLB proteins. However, harboring genomic CLB operons is risky, especially in cases where there is no longer a threat of exogenous CLBs (the selection pressure). The immunity

gene alone is sufficient for protection against exogenous CLBs. On the other hand, the activity of endonuclease CLBs (double-strand breaks) and the lysis proteins (cell membrane damage) could be lethal and increase the risk during stress conditions. If we compare isogenic bacterial cells with and without CLB plasmid, the CLB operon-harboring cell is disadvantaged in non-selective conditions. Interestingly, most endonuclease CLB operons are plasmidic, indicating that either endonuclease CLB operons have recently evolved (and yet to integrate into genomes) or that a genome with complete operon is inviable. We think the latter is the case based on the available data because a double-strand break of the genome due to the toxin could be lethal, and the host is lysed due to lysis protein. Typically, double-stranded breaks in bacteria are repaired through non-homologous end joining (NHEJ), which is highly erroneous.

To understand the counterintuitive possibility of CLB activity and lysis genes on the genome, we took a closer look at the genomic endonuclease CLB operons in terms of their conservation among various genera. Out of all the endonuclease CLB operon hits, we observed 49% of plasmids and nearly 1% of genomes harbor complete endonuclease CLB operons (colicin, immunity and lysis genes). The majority of the genomic colicin operons in our analysis either lacked intact lysis genes or lacked both colicin and lysis genes (Fig. 3b). Our conservation analyses highlight the degeneration of genomic endonuclease CLB operons relative to that of the plasmidic operons. We can reason that the early loss of the immunity gene from the genome is detrimental to bacterial survival. In contrast, the early loss of lysis and CLB activity genes benefits the host strain (Papadakos et al. 2012). Evolution works at the level of genes whose cumulative expression determines the ecological success of the associated replicon. Losing detrimental genes and retaining beneficial genes is how the genomes are continuously optimized.

The expression of toxin–antidote pairs, such as endonuclease CLB operons, greatly influences the bacterial population dynamics, especially during the competition. The release of CLB determines the composition and stability of microbial populations (Escalante and Travisano 2017; Gonze et al. 2018). Simulation-based studies have elaborated on the factors such as time-point of CLB release, the concentration of CLB release, the initial abundance of the population, range of toxicity, stress conditions, resource availability, plasmid copy number, the concentration of lysis protein (Weber et al. 2014; Mader et al. 2015; Weiss et al. 2020), etc., which determines the competition outcome. Our theoretical analysis depicts the probable advantage of genomic endonuclease CLB operons and their selection. In our model, the competition outcome is influenced by a range of parameters such as plasmid

invasion rate, degree of plasmid addiction, cost of lysis gene, chromosomal integration rate, and operon conservation rate (Fig. 4). As we see that the number of parameters that were swept is a multidimensional problem, the multi-objective optimization (for example, low plasmid invasion, low chromosomal integration, rare beneficial mutations, etc.) will determine the critical parameters that allow the emergence and maintenance of the C_i cell population. The preferential conservation of the immunity gene over CLB activity and lysis genes allows the host to dispose of the burden imposed by CLB plasmids and the cost incurred by CLB activity and lysis genes. Interpretation of the model in terms of plasmid–genome conflict gives a better picture of the eco-evolutionary cycle that governs the dynamicity in the population. In the conflict between plasmids and genomes, dynamic time points exist wherein genomes or plasmids get equal chances to outcompete the other or exist in cooperation (Figs. 5, 6). Introducing cell types containing the degenerative operons on plasmids, which is beyond the scope of this study, might influence the population dynamics from a different perspective. Obtaining empirical evidence for these events is technically difficult at this point and resources. The anti-addiction hypothesis for endonuclease CLBs can be proved using methodology followed for toxin–antitoxin systems (Saavedra De Bast et al. 2008). The degeneration of the genomic endonuclease CLB operons is methodologically challenging. The anti-addiction strategy evolved with the genomic CLBs could be adapted as a potential strategy to eradicate the CLB plasmids containing antibiotic resistance genes. A non-pathogenic sensitive strain with genomic immunity gene (of CLB) could outcompete the CLB plasmid-containing pathogens or antibiotic resistant strains. This strategy could augment the probiotics-based therapeutic approaches. Experimental competition assays should be explored to provide insights and possibilities of using genomic CLBs for treating bacterial infections.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00438-022-01884-4>.

Acknowledgements This work was supported by Prof T. R. Rajagopalan grants (internal funding) and Central Research Facility sanctioned by SASTRA Deemed University, Thanjavur. PAS is supported through Junior Research Fellowship by the University Grants Commission (479/CSIR-UGC NET JUNE2019), Government of India. This manuscript has been released as a Pre-Print at bioRxiv (Sudhakari and Ramisetty 2021). We acknowledge the support from the System Analyst, L. Thiagarajan, and School of Computing at SASTRA Deemed University, for providing computing resources to run the simulations. We acknowledge the suggestions from Mr. Karthik Suresh, Department of Physics, University of Regina, Canada, towards modeling and simulation.

Author contribution BCMR has conceived the idea, designed the work, interpreted the results, and wrote the manuscript. PAS has designed

the work, performed the analyses, interpreted the results, and wrote the manuscript.

Funding This work was supported by Prof T. R. Rajagopalan grants (Internal Fund).

Code availability The python Source code can be found at https://github.com/Pavi31/Plasmid-Genome_conflicts.git.

Declarations

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent to publish Not applicable.

References

- Bucci V, Nadell CD, Xavier JB (2011) The evolution of bacteriocin production in bacterial biofilms. *Am Nat* 178:E162–173. <https://doi.org/10.1086/662668>
- Cascales E, Buchanan SK, Duche D, Kleanthous C, Lloubes R, Postle K, Riley M, Slatin S, Cavard D (2007) Colicin biology. *Microbiol Mol Biol Rev* 71:158–229. <https://doi.org/10.1128/MMBR.00036-06>
- Cavard D, Lloubes R, Morlon J, Chartier M, Lazdunski C (1985) Lysis protein encoded by plasmid ColA-CA31. Gene sequence and export. *Mol Gen Genet* 199:95–100. <https://doi.org/10.1007/BF00327516>
- Dawkins R (1989) *The selfish gene*. New. Oxford University Press, Oxford, p 1989
- Eberhard WG (1990) Evolution in bacterial plasmids and levels of selection. *Q Rev Biol* 65:3–22. <https://doi.org/10.1086/416582>
- Escalante AE, Travisano M (2017) Editorial: conflict and cooperation in microbial societies. *Front Microbiol* 8:141. <https://doi.org/10.3389/fmicb.2017.00141>
- Fraikin N, Goormaghtigh F, Van Melderen L (2020) Type II toxin-antitoxin systems: evolution and revolutions. *J Bacteriol*. <https://doi.org/10.1128/JB.00763-19>
- Gillor O, Vriezen JAC, Riley MA (2008) The role of SOS boxes in enteric bacteriocin regulation. *Microbiology* 154:1783–1792. <https://doi.org/10.1099/mic.0.2007/016139-0>
- Gonze D, Coyte KZ, Lahti L, Faust K (2018) Microbial communities as dynamical systems. *Curr Opin Microbiol* 44:41–49. <https://doi.org/10.1016/j.mib.2018.07.004>
- Gordon DM, O'Brien CL (2006) Bacteriocin diversity and the frequency of multiple bacteriocin production in *Escherichia coli*. *Microbiology* 152:3239–3244. <https://doi.org/10.1099/mic.0.28690-0>
- Granato ET, Foster KR (2020) The evolution of mass cell suicide in bacterial warfare. *Curr Biol*. <https://doi.org/10.1016/j.cub.2020.05.007>
- Gray LF (2003) A mathematician looks at Wolfram's new kind of science. *Notices of the AMS*, vol 50, no 2
- Guo Z, Sloot PM, Tay JC (2008) A hybrid agent-based approach for modeling microbiological systems. *J Theor Biol* 255:163–175. <https://doi.org/10.1016/j.jtbi.2008.08.008>

- Hardy KG, Meynell GG, Dowman JE, Spratt BG (1973) Two major groups of colicin factors: their evolutionary significance. *Mol Gen Genet* 125:217–230. <https://doi.org/10.1007/BF00270744>
- Harms A, Brodersen DE, Mitarai N, Gerdes K (2018) Toxins, targets, and triggers: an overview of toxin-antitoxin biology. *Mol Cell* 70:768–784. <https://doi.org/10.1016/j.molcel.2018.01.003>
- Heng NCK, Wescombe PA, Burton JP, Jack RW, Tagg JR (2007) The diversity of bacteriocins in gram-positive bacteria. *Bacteriocins*. https://doi.org/10.1007/978-3-540-36604-1_4
- Horesh G, Fino C, Harms A, Dorman MJ, Parts L, Gerdes K, Heinz E, Thomson NR (2020) Type II and type IV toxin-antitoxin systems show different evolutionary patterns in the global *Klebsiella pneumoniae* population. *Nucleic Acids Res* 48:4357–4370. <https://doi.org/10.1093/nar/gkaa198>
- Inglis RF, Bayramoglu B, Gillor O, Ackermann M (2013) The role of bacteriocins as selfish genetic elements. *Biol Lett* 9:20121173. <https://doi.org/10.1098/rsbl.2012.1173>
- Jain A, Srivastava P (2013) Broad host range plasmids. *FEMS Microbiol Lett* 348:87–96. <https://doi.org/10.1111/1574-6968.12241>
- Kado CI (1998) Origin and evolution of plasmids. *Antonie Van Leeuwenhoek* 73:117–126. <https://doi.org/10.1023/A:1000652513822>
- Kerr B, Riley MA, Feldman MW, Bohannan BJ (2002) Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* 418:171–174. <https://doi.org/10.1038/nature00823>
- Kleanthous C, Kuhlmann UC, Pommer AJ, Ferguson N, Radford SE, Moore GR, James R, Hemmings AM (1999) Structural and mechanistic basis of immunity toward endonuclease colicins. *Nat Struct Biol* 6:243–252. <https://doi.org/10.1038/6683>
- Kulakauskas S, Lubys A, Ehrlich SD (1995) DNA restriction-modification systems mediate plasmid maintenance. *J Bacteriol* 177:3451–3454. <https://doi.org/10.1128/jb.177.12.3451-3454.1995>
- Lehtinen S, Huisman JS, Bonhoeffer S (2021) Evolutionary mechanisms that determine which bacterial genes are carried on plasmids. *Evol Lett* 5:290–301. <https://doi.org/10.1002/evl3.226>
- Lin YH, Liao CC, Liang PH, Yuan HS, Chak KF (2004) Involvement of colicin in the limited protection of the colicin producing cells against bacteriophage. *Biochem Biophys Res Commun* 318:81–87. <https://doi.org/10.1016/j.bbrc.2004.03.184>
- MacLean RC, San Millan A (2015) Microbial evolution: towards resolving the plasmid paradox. *Curr Biol* 25:R764–767. <https://doi.org/10.1016/j.cub.2015.07.006>
- Mader A, von Bronk B, Ewald B, Kesel S, Schnetz K, Frey E, Opitz M (2015) Amount of colicin release in *Escherichia coli* is regulated by lysis gene expression of the colicin E2 operon. *PLoS One* 10:e0119124. <https://doi.org/10.1371/journal.pone.0119124>
- Majeed H, Gillor O, Kerr B, Riley MA (2011) Competitive interactions in *Escherichia coli* populations: the role of bacteriocins. *ISME J* 5:71–81. <https://doi.org/10.1038/ismej.2010.90>
- Masaki H, Ohta T (1985) Colicin E3 and its immunity genes. *J Mol Biol* 182:217–227. [https://doi.org/10.1016/0022-2836\(85\)90340-7](https://doi.org/10.1016/0022-2836(85)90340-7)
- Mavridou DAI, Gonzalez D, Kim W, West SA, Foster KR (2018) Bacteria use collective behavior to generate diverse combat strategies. *Curr Biol* 28(345–355):e344. <https://doi.org/10.1016/j.cub.2017.12.030>
- Million-Weaver S, Camps M (2014) Mechanisms of plasmid segregation: have multicopy plasmids been overlooked? *Plasmid* 75:27–36. <https://doi.org/10.1016/j.plasmid.2014.07.002>
- Papadakos G, Wojdyla JA, Kleanthous C (2012) Nuclease colicins and their immunity proteins. *Q Rev Biophys* 45:57–103. <https://doi.org/10.1017/S0033583511000114>
- Ramisetty BC, Santhosh RS (2016) Horizontal gene transfer of chromosomal type II toxin-antitoxin systems of *Escherichia coli*. *FEMS Microbiol Lett*. <https://doi.org/10.1093/femsle/fnv238>
- Riley MA (1998) Molecular mechanisms of bacteriocin evolution. *Annu Rev Genet* 32:255–278. <https://doi.org/10.1146/annurev.genet.32.1.255>
- Saavedra De Bast M, Mine N, Van Melderen L (2008) Chromosomal toxin-antitoxin systems may act as antiaddiction modules. *J Bacteriol* 190:4603–4609. <https://doi.org/10.1128/JB.00357-08>
- Shapiro JW, Williams ES, Turner PE (2016) Evolution of parasitism and mutualism between filamentous phage M13 and *Escherichia coli*. *PeerJ* 4:e2060. <https://doi.org/10.7717/peerj.2060>
- Sharp C, Bray J, Housden NG, Maiden MCJ, Kleanthous C (2017) Diversity and distribution of nuclease bacteriocins in bacterial genomes revealed using Hidden Markov Models. *PLoS Comput Biol* 13:e1005652. <https://doi.org/10.1371/journal.pcbi.1005652>
- Shimoni Y, Altuvia S, Margalit H, Biham O (2009) Stochastic analysis of the SOS response in *Escherichia coli*. *PLoS One* 4:e5363. <https://doi.org/10.1371/journal.pone.0005363>
- Spengler G, Molnar A, Schelz Z, Amaral L, Sharples D, Molnar J (2006) The mechanism of plasmid curing in bacteria. *Curr Drug Targets* 7:823–841
- Sudhakari PA, Ramisetty BCM (2021) Modeling colicin operons as genetic arms in plasmid-genome conflicts. *bioRxiv*. <https://doi.org/10.1101/2021.05.15.444200>
- Tsang J (2017) Bacterial plasmid addiction systems and their implications for antibiotic drug development. *Postdoc J* 5:3–9
- Van Melderen L, Saavedra De Bast M (2009) Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genet* 5:e1000437. <https://doi.org/10.1371/journal.pgen.1000437>
- von Bronk B, Schaffer SA, Gotz A, Opitz M (2017) Effects of stochasticity and division of labor in toxin production on two-strain bacterial competition in *Escherichia coli*. *PLoS Biol* 15:e2001457. <https://doi.org/10.1371/journal.pbio.2001457>
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Weber MF, Poxleitner G, Heibisch E, Frey E, Opitz M (2014) Chemical warfare and survival strategies in bacterial range expansions. *J R Soc Interface* 11:20140172. <https://doi.org/10.1098/rsif.2014.0172>
- Weiss AS, Gotz A, Opitz M (2020) Dynamics of ColicinE2 production and release determine the competitive success of a toxin-producing bacterial population. *Sci Rep* 10:4052. <https://doi.org/10.1038/s41598-020-61086-z>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.