Modeling Colicin Operons as Genetic Arms in Plasmid-genome Conflicts

Pavithra Anantharaman Sudhakari

SASTRA University School of Chemical and Biotechnology: Shanmugha Arts Science Technology and Research Academy School of Chemical and Biotechnology

Bhaskar Chandra Mohan Ramisetty (✉ bcmr25@gmail.com)

SASTRA Deemed University <https://orcid.org/0000-0003-2511-4162>

Research Article

Keywords: Colicins, anti-addiction hypothesis, plasmid-genome competition, quasi-cooperation, agent-based model

Posted Date: September 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-853467/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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. The conflict between bacterial genome and plasmids allows the emergence of 'genetic arms' such as colicin (Col) operons. Endonuclease Col operons encode three proteins; an endonuclease colicin, an immunity protein, and lysis protein. Col operons are plasmid-maintenance systems; (i) the plasmid-cured cells are killed by the colicins; (ii) damaged cells lyse and release the colicins that eliminate the competitors, and (iii) the released plasmids invade new bacteria. Surprisingly, some bacterial genomes have Col operons. The eco-evolutionary drive and physiological relevance of genomic Col operons are unknown. We investigated plasmidic and genomic Col operons using sequence analyses from an eco-evolutionary perspective. We found 1,248 genomic and plasmidic colicins across 30 bacterial genera. Although 51% of the genomes harbor colicins, the majority of the genomic colicins lacked a functional lysis gene, suggesting the negative selection of lethal genes. The immunity gene of the Col operon protects the plasmid-cured host; eliminating the metabolic burden. We show mutual exclusivity of Col operons on genomes and plasmids. We propose an anti-addiction hypothesis for genomic colicins. Using a stochastic agent-based model, we show that the genomic colicins confer an advantage to the host genome in terms of immunity to the toxin and elimination of plasmid burden. Col operons are 'genetic arms' that regulate the ecological interplay of bacterial genomes and plasmids.

Introduction

Plasmids are one of the simplest propagating entities which are made of DNA. Plasmids, like viruses, are acellular propagating entities that are common in the bacterial domain. Plasmids live off of their bacterial hosts, sometimes conferring advantage and most of the time as molecular parasites. Unless these propagating entities confer an advantage in the contextual habitat, it is a burden with costs on the fitness relative to the peers. Even in isolated populations, the bacterial cells devoid of plasmids (non-contributing) are advantaged over plasmid-bearing clones. 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 propagation of the plasmid because the cured bacteria may outcompete the plasmid-harboring bacteria under non-selective conditions. That conflict between propagating entities (the bacterial genome and plasmid) is a never-ending genetic warfare that drives the emergence of ingenious genetic systems (Eberhard 1990). Bacterial curing can occur by passive mechanisms wherein plasmid may not be inherited by one daughter cell (Spengler, et al. 2006). The obligatory dependence of plasmids on bacterial cells drives the emergence of multiple ingenious mechanisms that ensure the propagation of the plasmid within the bacterial population. A significant proportion of plasmid-bearing bacteria in a population increases the probability of spatiotemporal propagation of that plasmid.

Some genetic systems ensure the dissemination of the plasmid to both the daughter cells. Mechanisms such as (i) high intracellular copy number of plasmids and (ii) stable partitioning systems, such as the *parM* system of R-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). The other strategy is by eliminating the daughter cells that may not have inherited a plasmid. Ingenious genetic systems have evolved that are generally referred to as the genetic arms. 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 (Col) 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 Melder and Saavedra De Bast 2009).

In this study, we explored the interplay of colicins operons in bacterial genome and plasmids evolution. Two major colicins are endonucleases (toxin cleaves the DNA) and pore-forming (toxin permeabilizes the membrane by forming pores). Colicins are classified into two groups based on cross-resistance and mechanism of toxin release. Group A comprises colicins that are translocated by the Tol system and are released by killing the host. Group B comprises colicins that are translocated by the TonB system and are not released into the extracellular medium. The colicin proteins can be taken up by bacteria from the surroundings and result in death or inhibition of vital activities. Endonuclease colicin operons are one of the well-characterized systems comprising of the genes required for colicin production (colicin gene), colicin neutralization (immunity gene) and colicin release (cell lysis gene) from the cell. A typical endonuclease Col operon comprises colicin activity gene (*cx*), immunity gene (*cx*), and lysis gene (*cx*) (Cascales, et al. 2007, Heng, et al. 2007). The *cx* encodes colicin that cleaves nucleic acids; *cx* encodes immunity protein that inactivates its cognate colicin, and *cx* encodes a lysis protein that lyses the host cell aiding in the release of the colicin (Kleanthous, et al. 1999, Masaki and Ohta 1985). Mutations in the lysis gene were shown to 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 colicins (cognate colicins released by other colicin-producing bacteria), thereby protecting the host cell (Cascales, et al. 2007).

Colicins are predominantly encoded on plasmids rendering the possibilities of horizontal transfers in the population (Gordon and O'Brien 2006, Hardy, et al. 1973). 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. (Bucci, et al. 2011, Lin, et al. 2004, Sharp, et al. 2017, Weiss, et al. 2020). However, the occurrence of colicin on bacterial genomes is intriguing because of the threats of genome degradation (by colicin) and the lysis of the host (by lysis protein).

Further, carrying identical colicin on both genomes and plasmids is energetically expensive and irrational for the propagation of the plasmid. Therefore, the occurrence of Col operon on the genome seems counterproductive to the propagation of the genome; during colicin release, the cell is lysed, and then the genome is degraded.

The release of colicin 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 colicin release, the concentration of colicin release, the initial abundance of the population, range of toxicity, stress conditions, resource availability, plasmid copy number, the concentration of lysis protein (Mader, et al. 2015, Weber, et al. 2014, Weiss, et al. 2020), etc., which determines the competition outcome. Plasmids may confer advantages such as metabolizing novel molecules and antibiotic resistance. In such cases, several groups refer to them as mutualism between plasmid and genomes. However, one could say mutualism only in the contextual environment wherein the plasmid-mediated advantages are evident. For example, antibiotic resistance plasmid within bacterium in media/location without the antibiotic is not mutualism. We do not and should not attribute foresight of an impending antibiotic exposure unless such exposure is a repetitive phase in the host's life cycle. Therefore, transient advantages do not amount to mutualism in the case of molecular parasites.

In this study, we explored the dynamics of Col operons (more specifically, the endonuclease encoding Col operons) in the propagation of plasmids and bacterial genomes. For simplicity, we consider plasmids as autonomous propagating entities that are parasites on bacteria similar to viruses. We see plasmids and the "purpose" of their genes is to enhance the propagation of the plasmids, directly or indirectly. In our view, plasmids do not form a part of the "self" of the host. Therefore, we consider plasmids and genomes as two independent entities and plasmids as "obligate molecular parasites" on bacteria. We aim to rationalize the impact of plasmidic and genomic Col 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 colicin is mutually exclusive on genomes and plasmids, meaning that colicin operon is present either on the plasmid or on the genome but never both. Similar to toxin-antitoxin systems, we propose an anti-addiction hypothesis for genomic colicins. Here, we provide a proof-of-concept that rationalizes the anti-addiction hypothesis for genomic colicins by simulating the competition between Col plasmids and genomes.

Materials And Methods

Prevalence of Col operon

We analyzed nine nuclease colicins (E2, E3, E4, E6, E7, E8, E9, D, and cloacin DF13) (Supplementary file S1) in completely sequenced bacterial genomes and plasmids listed in the NCBI database as of July 2020. 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). We narrowed our search to specific colicins for which threshold values for sequence identity ($\sim >80\%$) and coverage ($\sim >90\%$) were set distinctively for each of the nine colicin types. The first hit list was created by filtering only "complete sequences," "chromosome," and "genomes" in MS-excel. 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 Col 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). Hits on genomes and plasmids were sorted separately and plotted as graphs. We then analyzed the colicin operons on genomes and plasmids for their conservation, prevalence, and multiplicity across genera. Genetic locus was examined for the conservation of operon (fig. S1). The distribution of colicins on both genomes and plasmids was plotted as graphs. In each genus, the hits harbouring Col operons either on genome or plasmid and on both genome and plasmid were grouped to verify their statistical significance. We performed a t-Test (two-tail) to test the statistical significance of our mutual exclusivity hypothesis.

Conservation of Col 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 colicins for strains harboring colicins on both genome 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.

Model description

We adopted a stochastic agent-based model to simulate the competition outcome between the strains: (i) sensitive to colicin (S); (ii) carrying colicin on the plasmid (C); and (iii) carrying colicin on the genome (C_g). Within the C_g population, we show three variations: the population that (i) has complete operon (C_{cil}); (ii) has colicin 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 our 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 100:5 (S: C) ratio in 250X250 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 as 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 colicins 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 colicin. We consider a small low-copy number (20 plasmids) plasmid containing endonuclease colicin operon.

Simulation of the competition

Python 3.7 scripts were written to simulate the model. Libraries such as Matplotlib, Pandas, and NumPy were used. The competition is modeled using a 2D lattice, representing the width and height of the

bacterial niche. Each grid in the lattice can contain one bacterial cell. We adapted an agent-based stochastic model to track the fate of each cell type (agents) as a function of space and time. The agents reproduce at specific growth rates. The multiplication of a cell is modeled using a Moore neighborhood (8 nearest neighbors) method. Each cell type is allowed to grow logistically yet limited by the carrying capacity and the total population at each instant. The logistic growth equation is calculated 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. We performed a synchronous simulation where at each time step, the reaction conditions were applied for the corresponding population and not for the individual bacterial cell. The C , C_{cil} , and rarely C_{cil} cells release toxin followed by their lysis. Once lysis occurred, the lattice site is now toxic to the nearby S cells. Our simulation governs the toxic interactions between the S cells and the toxic lattice sites by assuming exponential toxicity (Weber, et al. 2014). 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 with the increasing radius. The exponential toxicity is calculated as,

$$\text{prob}(0 - 1) < (\text{tox})^r$$

The state conversion probabilities are calculated binomially (binomial (n, p, N)), representing the binomial distribution with probability p and local trials n repeated N times). Model parameters, simulation conditions, and reaction rates are provided in the supplementary table with appropriate references (fig. 3b). The lack of experimental data is compensated by testing the competition outcome for a range of reaction rates. Uncertainty in the model is introduced through binomial probabilities for every parameter. The simulations with parameter sweeping were run ten times for 15,000 minutes. The competition winner is assigned if they constitute more than 90% of the population in more than 5 out of 10 iterations. If none of the cell types attain 90% of the population, then the population is considered 'disproportionately draw.' Our model predicts the relative success of plasmids in competitive environments. The success of plasmids and bacterial genomes is measured as the percentage of Col plasmid-containing cells (C) and genomic Colicin-containing cells (C_g) respectively in the population as a function of time.

Results

Distribution of Colicin operons across bacterial genera

To examine the prevalence of genomic and plasmidic Col operons, we performed a nucleotide homology search for nine nuclease colicins (E2, E3, E4, E6, E7, E8, E9, D, and cloacin DF13) against all bacterial genomes and plasmids. 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 Col operons (fig. 1a). All 30 genera are gram-negative *Gammaproteobacteria* and *Betaproteobacteria*. Plasmidic colicins

were prevalent among the strains of *Klebsiella* (61%) and *Escherichia* (26%). Genomic colicins were prevalent among the strains of *Pseudomonas* (30%), *Salmonella* (28%), and *Yersinia* (15%) (table S1). The strains of *Klebsiella*, *Citrobacter*, *Salmonella*, and *Enterobacter* harbored Col operons on both genome and plasmid. By setting up the threshold values for sequence identity (>90%), we narrowed our analysis to nine 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 Col operon on the genome. Endonuclease colicins E2, E8, and CloDF13 were found only on the genomes of *E. coli* and *Klebsiella* but not other genera. Very few (3%) strains of *Shigella* harbored nuclease Col plasmids (Supplementary Sheet S1).

Mutual exclusivity of plasmidic and genomic colicin operons

To investigate if the colicins are present on both genomes and plasmids, we analyzed their distribution pattern across all the genera. We found that some genera had Col operons only on genomes, and some had only on plasmids (fig. 1b). For example, we observed only genomic colicins in *Pseudomonas* strains while only plasmidic colicins in *Shigella*. Whereas in strains of *Klebsiella* (10 strains), *Escherichia* (1 strain), *Salmonella* (2 strains), and *Enterobacter* (2 strains), both genomic and plasmidic colicins were present (Supplementary Sheet S4). To rationalize the multiplicity of colicins within a strain, we examined the operon sequence similarity and conservation. We observed that no strains encoded multiple copies of identical colicins (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 CloDF13-like operon on its plasmid. We also observed 85 strains harboring multiple Col 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 klebicin B and ColE3 on its plasmids pNY9_1 and pNY9_4, respectively. The occurrence of Col operons either on genomes or plasmids but not on both is statistically significant ($p = 0.017$). Thus, the occurrence of Col operons on either genome or plasmid is greater than its occurrence on both genome and plasmid. No bacterial strain had identical Col operons on genomes as well as plasmids.

Degeneration of genomic Col operon

Of the 620 genomic colicin operons, five were complete operons, 614 lacked lysis gene, and 323 lacked the colicin activity gene, whereas the immunity gene was conserved in almost all operons. Our comparative gene conservation analysis found that most genomes either lacked lysis genes or harbored only immunity genes (fig. 2). To account for the prevalence of genomic colicins, as harboring lysis genes is strategically inefficient, we compared the conservation of genomic and plasmidic operons. Colicin 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 roggenkampii* R11 harbors an S-type pyocin operon on its plasmid and an orphan immunity gene on its genome. Exceptionally, *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 Col operons lacking functional lysis gene are either group B colicins that do not release the colicins or group A colicins with truncated lysis gene. We observed high conservation of the immunity gene on genomes (e.g., *Mixta theicola* SRCM103227) compared to colicin and lysis genes. Several truncations in the operon genes could account for the intermediates in the degeneration process. Conceivably, we did not find any Col operon lacking immunity gene on the genomes, as the absence of immunity protein is detrimental to the host.

The competitive advantage of genomic Colicins

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 Col operon. As a proof-of-concept, we adapted our 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 Col operon (S-cell); (ii) carrying colicin on the plasmid (C-cell); and (iii) carrying colicin on the genome (C_g -cells) (fig. 3a). Stochastic fluctuations in the parameters strongly affect the competition outcome (Shimoni, et al. 2009). We have incorporated such uncertainty into the model by testing a range of values for theoretical parameters (with binomial probabilities). Parameter sweeping tests the competition outcome for a range of parameter values abiding by Monte Carlo realizations, allowing statistical evaluation of the model (fig. 3b). Our model depicts the interactions between the Colicinogenic and sensitive population and the competition outcome upon invasion by Col 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 Col plasmid (λ_s), death of C due to plasmid loss rate, operon activation rate in C (λ_c), acquisition rate of Col operon on the genome (λ_g), and colicin 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 (Col plasmids benefit only the associated genomes by providing immunity to the toxin), where both plasmids and genomes are benefitted (fig. 4, 6). Thus, genomic colicins are selected as a counter-strategy that eliminates the burden of Col plasmid imposed via addiction. In an evolutionary sense, genomic colicins are no longer required after anti-addiction but for protection against the exogenous colicins. We then incorporated three genomic Col operon variations in the model: C_g population with (i) complete operon (C_{cil}); (ii) Colicin activity and immunity genes but not lysis gene (C_{ci}); and (iii) only immunity gene (C_i) (fig. 3a). In coherence to in-silico analysis showing the high conservation of the immunity gene over Colicin activity and lysis genes, our model predicts the outcome of the competition among Colicin sensitive, plasmidic Col containing population and genomic Col containing population (C_{cil} , C_{ci} and C_i) in varied conditions (fig. 4).

Discussion

Plasmids, like viruses and viroid, are acellular propagating entities that are obligately dependent on cellular propagating entities for propagation. The successful existence and propagation of plasmids depend on the host's availability and the ability to persist within the host. Plasmid interactions with their respective host are less detrimental, unlike the lytic phages. Plasmids with genetic elements that ensure successful dissemination of the plasmid to the offspring will prevail better. Col operons are such genetic elements that mediate the interplay between the plasmids and genomes. This study investigated the abundance of Col operons on genomes and proposed an anti-addiction phenomenon for genomic colicins.

We first analyzed the abundance of Col operons on plasmids as colicins are predominantly plasmidic (Cascales, et al. 2007). The observed prevalence of genomic colicins (51%) across the bacterial genera could be a sample bias because there are a greater number of completely sequenced bacterial genomes compared to sequenced plasmids, which again could be the degree of their respective availability. Interestingly, most endonuclease colicin operons are plasmidic, indicating that either endonuclease colicin operons have recently evolved (and yet to integrate into genomes) or that a genome with integrated endonuclease colicin operon on the genome is inviable. With available data, we think that the latter is the case because even one double-strand break of the genome could be very costly. Typically, double-stranded breaks in bacteria are repaired through non-homologous end joining, which is highly erroneous, especially when a template is unavailable for homologous recombination (Ayora, et al. 2011, Shuman and Glickman 2007).

Plasmidic Col operons impose 'addiction' (Inglis, et al. 2013), a phenomenon where the cured daughter cell dies due to the toxic effects of the Col proteins. In such scenarios, a bacterium harboring a Col plasmid has limited possibilities; (i) to bear with the plasmid due to 'addiction' with a relative risk of being outcompeted by plasmid-bearing bacteria, (ii) to risk death upon loss of plasmid due to the lack of immunity protein and (iii) attain immunity by acquiring Col operon on to the genome. 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 theory, all infected bacteria have to either retain the plasmid or die due to colicins. 'Forced' plasmid maintenance (addiction) is similar to being parasitized by phages like M13 (Shapiro, et al. 2016). In nature, the probability of Col plasmid-harboring bacterium being cured of that plasmid is low because of addiction: a choice of plasmid maintenance or death. The presence of Col 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. Based on the above rationales, we devised two premises. (i) A bacterium harboring Col plasmid would not have a Col operon on its genome, and (ii) a bacterium with genomic Col operon would not harbor a plasmid-encoding identical Col operon. We found our mutual exclusivity data for the occurrence of Col operon fulfilling both the premises supporting the anti-addiction hypothesis (fig. 1), corroborating the observations in the anti-addiction hypothesis of TAs (Ramisetty and

Santhosh 2016). The prevalence and mutual exclusivity of genomic colicins indicate a selective advantage of carrying colicin operons on genomes.

While the integration of the Col operon on the genome may have allowed the curing, harboring genomic col operons is still not the best-case scenario, especially in cases where there is no longer a threat of exogenous colicins (the selection pressure). If we compare isogenic bacterial cells with and without Col plasmid, the Col operon-harboring cell is disadvantaged in growth conditions without selection. This is because of the probability of genomic double-strand breaks due to colicin proteins. In terms of overall fitness, the loss of toxic genes such as the cell lysis gene and the colicin gene should be favored. We observed 49% of plasmids and nearly 1% of genomes harbor complete Col operons. The comparative conservation analyses highlight the degeneration of the Col operon. We can interpret that the early loss of the immunity gene from the genome is detrimental to bacterial survival. In contrast, the early loss of lysis and colicin genes benefits the host strain (Papadakos, et al. 2012). In an evolutionary sense, genomic colicins are no longer required after anti-addiction but for protection against the exogenous colicins. Of the three Col operon genes, only immunity protein is helpful in anti-addiction and protection against exogenous colicins. The colicin protein is harmful but can be antagonized by the immunity protein. The lysis protein is detrimental to the cell. Colicin and lysis genes have no evolutionary benefit to the cell. Therefore, the lysis and colicin genes should be less conserved relative to immunity genes. Evolution works at the level of genes whose cumulative expression determines the ecological success of the associated replicon (Dawkins 1989). Phenomena like suicide and altruism can be used appropriately when terms like "self" are defined clearly. These terms can be attributed only to cellular entities and not to viruses and plasmids because they do not have much to "sacrifice" other than their physical genome itself. Phenomena directly consequential of the plasmids and their genes should be attributed to the plasmid alone and not to the host harboring that plasmid. This area is a matter of philosophical debate and, at times, is highly subjective. Some view at the level of cells, while others view at the level of genes (Ramisetty and Sudhakari 2020). Some questions for such contentions in relevance to the topic of this study are as follows. Is the plasmid a part of the genome's "self"? Are two daughter bacterial cells, one with a plasmid and one cured, kin or competitors? Are the genomic Col operons (or other genetic elements like prophages, integrated plasmids, and other horizontally acquired genes) "self" or not yet? These questions are philosophical in nature and beyond the scope of this work.

Our theoretical analysis unravels the probable advantage of genomic colicins and their selection over plasmidic colicin. 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. The preferential conservation of the immunity gene over Col and lysis genes allows the host to dispose of the burden imposed by Col plasmids and the cost incurred by Col 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 (fig. 5, 6).

The Plasmid-Genome conflict

Plasmid maintenance systems such as Col operons exploit the Plasmid-Genome conflict. The eco-evolutionary cycle of colicin 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 propagated 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 (Hayes and Sauer 2003, Silva, et al. 2012). A daughter cell that fails to inherit a copy of the plasmid becomes sensitive to the lethality of exogenous colicins as there is no source for the production of cognate immunity protein. The colicin operon acts as 'genetic arms' to induce the addiction to the plasmid. Through the elimination of cured cells, plasmids increase their maintenance within the population. Therefore, plasmids are advantaged at the cost of the host genome. But, in conditions of host DNA damage and the induction of SOS response, the plasmid-encoded colicins are the stress 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 colicin proteins and plasmids. The released plasmids have an opportunity to be taken up by other cells. The colicin proteins induce lethality in plasmid-free cells resulting in the elimination of competition. The plasmid-containing cells are resistant to exogenous colicins due to the neutralizing effects of the immunity protein. They have increased genome propagation along with plasmid propagation. The ultimate destiny of the Col plasmid-containing cell is to either remain parasitized by the plasmid, lyse due to lysis protein or die due to damage. Occasionally, the colicin operon may integrate into the chromosome through non-specific means associated with transposons or complete plasmid integration (fig. 2a, 5). In such cases, the host is relieved from plasmid addiction. The chromosomal colicin operons can act as anti-addiction systems by providing immunity protein to protect the host from exogenous colicins. Here, the host genome is advantaged because the cured daughter cells are resistant to the endogenous and exogenous colicins, and there is a reduction in the metabolic burden incurred by the plasmid. Interestingly, two of the three genes, the colicin gene and lysis gene, are costly to maintain because of their respective lethal DNA degradation and cell membrane lysis. The immunity gene-encoded protein is essential as long as the colicin gene is encoded and immunity against the exogenous colicins. The lysis gene and the colicin genes are lost from most chromosomal colicin operons. Therefore, the genome is advantaged.

This study highlights the prevalence of colicins on the genomes and plasmids of bacterial genera. Mutual exclusivity of colicins on genomes and plasmids supports the 'anti-addiction' hypothesis (Ramisetty and Santhosh 2017). Colicins are the 'genetic arms,' which influence the 'arms race' between genomes and plasmids (Inglis, et al. 2013). Plasmids occasionally provide an ecological advantage to the host; nevertheless, they are parasites and a metabolic burden. Colicin on plasmids ensures plasmid propagation via host addiction. This study highlights the prevalence of colicins on the genomes and plasmids of bacterial genera. Genomic colicins enable anti-addiction to plasmids, thereby eliminating the metabolic burden of plasmids on the host. Empirical evidence is required to confirm these observations. Colicin operons play a significant role in the ecology and evolution of bacterial genomes and plasmids. The antagonistic role of colicins promotes genome diversity in bacteria (Kirkup and Riley 2004). Studying

the phenomenon of anti-addiction may help mitigate the plasmid-mediated virulence and antibiotic resistance. The weapons (Col operons) will determine the winners and outcomes in a conflict (between plasmids and genomes), by the end of which begins the process of disarmament to reduce the burdens and costs of the weapons (degeneration of genomic Col operons).

Declarations

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

Conflict of interests: 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.

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

Authors' contribution: Bhaskar Chandra Mohan Ramisetty has conceived the idea, designed the work, interpreted the results, and wrote the manuscript. Pavithra Anantharaman Sudhakari has designed the work, performed the analyses, interpreted the results, and wrote the manuscript.

Acknowledgments: This work was supported by Prof T. R. Rajagopalan grants and Central Research Facility sanctioned by SASTRA Deemed University, Thanjavur. PAS is supported through Junior Research Fellowship by University Grants Commission (UGC), Government of India. This manuscript has been released as a Pre-Print at bioRxiv (Sudhakari and Ramisetty 2021).

Ethics approval: Not applicable

Consent to participate: Not applicable

Consent to publication: Not applicable

References

Ayora S, Carrasco B, Cardenas PP, Cesar CE, Canas C, Yadav T, Marchisone C, Alonso JC (2011) Double-strand break repair in bacteria: a view from *Bacillus subtilis*. *FEMS Microbiol Rev* 35: 1055-1081 doi: 10.1111/j.1574-6976.2011.00272.x

Bucci V, Nadell CD, Xavier JB (2011) The evolution of bacteriocin production in bacterial biofilms. *The American naturalist* 178: E162-173 doi: 10.1086/662668

Cascales E, Buchanan SK, Duche D, Kleanthous C, Lloubes R, Postle K, Riley M, Slatin S, Cavard D (2007) Colicin biology. *Microbiology and molecular biology reviews* : *MMBR* 71: 158-229 doi: 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. *Molecular & general genetics* : MGG 199: 95-100 doi: 10.1007/BF00327516

Dawkins R 1989 *The selfish gene* New edition. Oxford ; New York : Oxford University Press, 1989., pp. Pages.

Eberhard WG (1990) Evolution in bacterial plasmids and levels of selection. *Q Rev Biol* 65: 3-22 doi: 10.1086/416582

Escalante AE, Travisano M (2017) Editorial: Conflict and Cooperation in Microbial Societies. *Front Microbiol* 8: 141 doi: 10.3389/fmicb.2017.00141

Gillor O, Vriezen JAC, Riley MA (2008) The role of SOS boxes in enteric bacteriocin regulation. *Microbiology* 154: 1783-1792 doi: 10.1099/mic.0.2007/016139-0

Gonze D, Coyte KZ, Lahti L, Faust K (2018) Microbial communities as dynamical systems. *Current Opinion in Microbiology* 44: 41-49 doi: <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 doi: 10.1099/mic.0.28690-0

Granato ET, Foster KR (2020) The Evolution of Mass Cell Suicide in Bacterial Warfare. *Current biology* : CB 10.1016/j.cub.2020.05.007

Hardy KG, Meynell GG, Dowman JE, Spratt BG (1973) Two major groups of colicin factors: Their evolutionary significance. *Molecular and General Genetics MGG* 125: 217-230 doi: 10.1007/BF00270744

Hayes CS, Sauer RT (2003) Toxin-antitoxin pairs in bacteria: killers or stress regulators? *Cell* 112: 2-4 doi: 10.1016/s0092-8674(02)01282-5

Heng NCK, Wescombe PA, Burton JP, Jack RW, Tagg JR (2007) The Diversity of Bacteriocins in Gram-Positive Bacteria. *Bacteriocins: Ecology and Evolution*: 45-92 doi: 10.1007/978-3-540-36604-1_4

Inglis RF, Bayramoglu B, Gillor O, Ackermann M (2013) The role of bacteriocins as selfish genetic elements. *Biology letters* 9: 20121173 doi: 10.1098/rsbl.2012.1173

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 doi: 10.1038/nature00823

Kirkup BC, Riley MA (2004) Antibiotic-mediated antagonism leads to a bacterial game of rock-paper-scissors in vivo. *Nature* 428: 412-414 doi: 10.1038/nature02429

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:

Kulakauskas S, Lubys A, Ehrlich SD (1995) DNA restriction-modification systems mediate plasmid maintenance. *Journal of bacteriology* 177: 3451-3454 doi: 10.1128/jb.177.12.3451-3454.1995

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 doi: 10.1016/j.bbrc.2004.03.184

MacLean RC, San Millan A (2015) Microbial Evolution: Towards Resolving the Plasmid Paradox. *Current biology* : CB 25: R764-767 doi: 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 doi: 10.1371/journal.pone.0119124

Majeed H, Gillor O, Kerr B, Riley MA (2011) Competitive interactions in *Escherichia coli* populations: the role of bacteriocins. *The ISME journal* 5: 71-81 doi: 10.1038/ismej.2010.90

Masaki H, Ohta T (1985) Colicin E3 and its immunity genes. *J Mol Biol* 182: 217-227 doi: 10.1016/0022-2836(85)90340-7

Million-Weaver S, Camps M (2014) Mechanisms of plasmid segregation: have multicopy plasmids been overlooked? *Plasmid* 75: 27-36 doi: 10.1016/j.plasmid.2014.07.002

Papadakos G, Wojdyla JA, Kleanthous C (2012) Nuclease colicins and their immunity proteins. *Quarterly reviews of biophysics* 45: 57-103 doi: 10.1017/S0033583511000114

Ramisetty BC, Santhosh RS (2016) Horizontal gene transfer of chromosomal Type II toxin-antitoxin systems of *Escherichia coli*. *FEMS Microbiol Lett* 363:10.1093/femsle/fnv238

Ramisetty BCM, Santhosh RS (2017) Endoribonuclease type II toxin-antitoxin systems: functional or selfish? *Microbiology* 163: 931-939 doi: 10.1099/mic.0.000487

Ramisetty BCM, Sudhakari PA (2020) 'Bacterial Programmed Cell Death': Cellular Altruism or Genetic Selfism? *FEMS Microbiol Lett* 10.1093/femsle/fnaa141

Shapiro JW, Williams ES, Turner PE (2016) Evolution of parasitism and mutualism between filamentous phage M13 and *Escherichia coli*. *PeerJ* 4: e2060 doi: 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 doi: 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 doi: 10.1371/journal.pone.0005363
- Shuman S, Glickman MS (2007) Bacterial DNA repair by non-homologous end joining. *Nature reviews Microbiology* 5: 852-861 doi: 10.1038/nrmicro1768
- Silva F, Queiroz JA, Domingues FC (2012) Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*. *Biotechnology advances* 30: 691-708 doi: 10.1016/j.biotechadv.2011.12.005
- Spengler G, Molnar A, Schelz Z, Amaral L, Sharples D, Molnar J (2006) The mechanism of plasmid curing in bacteria. *Current drug targets* 7: 823-841 doi:
- Sudhakari PA, Ramisetty BCM (2021) Modeling colicin operons as genetic arms in plasmid-genome conflicts. *bioRxiv*: 2021.2005.2015.444200 doi: 10.1101/2021.05.15.444200
- Van Melder L, Saavedra De Bast M (2009) Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS genetics* 5: e1000437 doi: 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 doi: 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 doi: 10.1093/bioinformatics/btp033
- Weber MF, Poxleitner G, Heibisch E, Frey E, Opitz M (2014) Chemical warfare and survival strategies in bacterial range expansions. *Journal of the Royal Society, Interface* 11: 20140172 doi: 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. *Scientific reports* 10: 4052 doi: 10.1038/s41598-020-61086-z

Figures

Figure 1

(a) Distribution of colicins across bacterial genera. We obtained 1248 genome and plasmid hits across 30 bacterial genera (belonging to families including Enterobacteriaceae, Morganellaceae, Yersiniaceae, Pseudomonadaceae, Pectobacteriaceae, Vibrionaceae, Erwiniaceae, and Pasteurellaceae, and Burkholderiaceae) (supplementary table 1). Here, we represent the distribution for genera comprising more than 10 colicinogenic strains. Plasmidic colicins were prevalent among the strains of *Klebsiella*

(61%), *Escherichia* (26%), *Shigella* (6%), and *Salmonella* (3%). Genomic Colicins were prevalent among the strains of *Pseudomonas* (30%), followed by *Salmonella* (28%), *Yersinia* (15%), *Serratia* (6%), and *Klebsiella* (4%). (b) Mutual exclusivity of colicins on genomes and plasmids. Hit list was sorted to distinguish whether strains carried colicins on the genome or plasmid. The data visualization is done using a sunburst graph. Different color code is given to each segment specific to genera (as labeled). For each genus, the outer segment represents the plasmids, and the inner segment represents the genomes. We performed a t-Test (two-tail) to test the statistical significance of the Col operons occurring mutually exclusive on genomes and plasmids. Of the observed number of strains (hits), the occurrence of Col operon either on genome or plasmid is statistically higher than its occurrence on both genome and plasmid (P-value (two-tail) is 0.017).

a

b

Occurrence on Genomes	Genetic architecture of Col operon	Occurrence on Plasmids
5	<i>cxl</i> → <i>cxl</i> → <i>cxl</i>	306
292	<i>cxl</i> → <i>cxl</i>	154
1	<i>cxl</i> → <i>cxl</i>	12
322	<i>cxl</i>	59

Figure 2

Genetic architecture of the operon. (a) Integration of Col operons on genomes. In *Morganella Morganii* N18-00103, the Col 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 Col operon. Likely that the operon integrated by replacing the region between the flanking genes. In *Escherichia coli* M18, the complete Col 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 colicin 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 operon. Whereas *P. aeruginosa* IMP66 has only Col E3-like gene but lacks a pyocin-like operon. (b) Degeneration of Col 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 Col operon is complete, including the lysis gene. We observed 49% of plasmidic colicins and nearly 1% of genomic colicins are complete operons. Colicin 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. *Enterobacter roggenskampi* R11 harbors an orphan immunity gene on its genome. However, the integrity and functionality of the operon are unknown. Of the genomes carrying incomplete operon, 614 lacked a functional lysis gene. The Col operons lacking functional lysis gene are group B colicins or group A colicins with truncated lysis gene. We observed high conservation of the immunity gene on genomes (e.g., *Mixta theicola* SRCM103227) compared to colicin activity and lysis genes. Interestingly, we did not find any Col operon lacking immunity gene, as the absence of immunity protein is detrimental to the host.

Figure 3

Theoretical model representing the interactions between the colicinogenic and sensitive strains. (a) Upon plasmid invasion, sensitive (S) cells transform into colicinogenic (C) cells producing colicins. Once the

Col operon integrates onto the genome (Ccil cells), though 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 (Ci cells). λ_S -Switching rate of S to C; λ_{cil} -Switching rate of C to Ccil; λ_{ci} -Switching rate of Ccil to Cci; λ_i -Switching rate of Cci to Ci; λ_{da} -death rate of C due to operon activation; λ_{dl} -death rate of C due to plasmid loss; λ_{da1} -death rate of Ccil due to operon activation; λ_{da2} -death rate of Cci due to operon activation. (b) Model parameters and reaction rates. Parameter sweeps tested for each cell type are provided.

Figure 4

Simulation outcome of the competition between the cell types. Parameter sweeping for a range of values allows the uncertainty in the simulation outcome. Here, we tested the simulation outcome based on seven parameters, for each of which range of values was tested. Uncertainty in the model is introduced through binomial probabilities for every parameter. The simulations with parameter sweeping were run 10 times for 15,000 minutes. This combined pie chart shows the competition outcome for each of the tested parameter values. The colour code represents the respective cell type winning. A cell type is considered 'winner' if they constitute more than 90% of the population in more than 5 out of 10 iterations. If none reaches 90% proportion, then the population is considered as 'disproportionately draw'.

Figure 5

Modelling and simulating the implications of colicin in plasmids-genomes conflict. The dynamics of the number of plasmids and genomes over the period are plotted as a line chart. The total number of S, C, Ccil, Cci, and Ci 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 2000-4000 hours and after 9000 hours) 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 Ccil conversion is 10^{-8} , Ccil to Cci conversion is 10^{-5} , Cci to Ci conversion is 10^{-5} , and colicin toxicity is 50%.

Figure 6

The eco-evolutionary cycle of colicin plasmids and host genomes. The ecological cycle can be divided into five theoretical stages. (i) Plasmid invasion. The plasmid is advantaged by its association with the host genome. (ii) Plasmid addition. Cells cured of the plasmid are eliminated from the population. The plasmids are advantaged at the cost of host cells. (iii) Quasi-cooperation. The plasmid-containing cells are resistant to exogenous colicins 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. (iv) Chromosomal integration. The chromosomal colicin operons can act as anti-addiction systems by providing immunity protein to protect the host from exogenous colicins. The host genome is advantaged because the cured daughter cells are no longer killed by the plasmid-encoded colicins and reduce the metabolic burden of harboring the plasmid. (v) Colicin operon degeneration. Selection against the lethal gene (lysis gene) results in stabilizing the genomic colicins in the population.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfigures.docx](#)
- [SupplementaryfileS1.docx](#)
- [Supplementarysheet.xlsx](#)