Knowledge Gaps:

This list provides a brief overview of the major knowledge gaps for this project, how they were resolved and where to find the information.

Knowledge Gap	Resolved By	Information is located	Date resolved
CRISPR efficiency boosters and limitations	[Became irrelevant, switched topics]		
Evolution Mechanisms in Bacteria Restrictions	[Became irrelevant, switched topics]		

Literature Search Parameters:

These searches were performed between (Start Date of reading) and XX/XX/2019.
List of keywords and databases used during this project.

Database/search engine	Keywords	Summary of search

Tags:

Tag Name	
#Introduction	#MLM
#Body	

Article #1 Notes: Title

Article notes should be on separate sheets

KEEP THIS BLANK AND USE AS A TEMPLATE

Source Title	
Source citation (APA Format)	
Original URL	
Source type	
Keywords	
#Tags	
Summary of key points + notes (include methodology)	
Research Question/Problem/ Need	
Important Figures	
VOCAB: (w/definition)	
Cited references to follow up on	
Follow up Questions	

Article #1 Notes: Harnessing CRISPR interference to resensitize laboratory strains and clinical isolates to last resort antibiotics

Source Title	Nature
Source citation (APA Format)	Chiacchiera, A. F., Casanova, M., Bellato, M., Piazza, A., Migliavacca, R., Batt, G., Magni, P., & Pasotti, L. (2025). Harnessing CRISPR interference to resensitize laboratory strains and clinical isolates to last resort antibiotics. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-024-81989-5
Original URL	https://doi.org/10.1038/s41598-024-81989-5
Source type	Scientific Article
Keywords	Antimicrobial resistance, Antibiotic re-sensitization, CRISPR array, *Escherichia coli* clinical isolates, bla_{NDM}-type $bla_{ctx\text{-}M}$-type *mcr-1*

#Tags	
Summary of key points + notes (include methodology)	Antimicrobial resistance is a growing global issue, exacerbated by numerous factors, such as the lack of drug production and the rapid spread of Antibiotic-Resistant Genes (ARG) among bacteria populations. The purpose of this article was to test the capabilities of CRISPRi to resensitize E. Coli to antibiotics meropenem, colistin, and cefotaxime via individual and multigene targeting of ARGs. All CRISPR edited E. Coli strains had singular or multiple ARGs that allowed them increased resistance to tested antibiotics. Multiple sets of experiments were performed testing different variables, all of which had three test strains: sensitive strain, resistant strain and a specific CRISPRi transformant (sCRISPRi), bearing the ARG-carrying plasmid(s) and a CRISPRi platform. Minimum inhibitory concentration (MIC), growth delays (Δt) and inhibitory concentration were tested for all experiments. All edited strains demonstrated increased susceptibility, though the results were variable based on strains, media and growth conditions. Unsurprisingly, TOP10 F' had a stronger resensitization ability than MG1655, highlighting the importance of plasmid uptake, as well as other potential factors (antibiotic sensitivity of the background hosts, growth rate, and expression levels of dCas9 and gRNA) for

	efficiency. Resensitization was achieved in E. coli grown in human urine samples, potentially leading to future *in vivo* studies and use. Escaper analysis confirmed the lack of new mutations in target sequences, tested in both liquid and solid assays to accommodate for collective antibiotic tolerance. This suggests a lowered risk of mutations that have been associated with previous CRISPR experiments. Future testing with other bacteria, in more complex models, are needed. Notes: **Classification** • **Paper type:** (**Empirical** / Review / Methods / Design / Other) **Key methods** • Three test strains were tested: sensitive strain, without ARG-carrying plasmids; resistant strain, with the target ARGs, specific CRISPRi transformant, bearing the ARG-carrying plasmids and a CRISPRi platform targeting 1-2 ARGs o The strains were characterized via microplate and agar plate assays to quantify antibiotic resensitization in terms of minimum inhibitory

	concentration, growth delays, and inhibitory concentration • A two-plasmid CRISPRi platform targeting single ARGs that relied on a low-copy plasmid carrying the gRNA module under the control of an IPTG-inducible PLlacO1 promoter, and a medium-copy plasmid, carrying the dCas9 module under the PJ23116 weak constitutive promoter, previously evaluated as a high-repression efficiency and low-burden dCas9 expression cassette39. A gRNA module blocked tetA transcription by targeting its CDS. The same platform was used to test Ampicillin re-sensitization. o Another set of TC- and AMP-resistant control strains, with the system, were evaluated to ensure the system did not cause relevant burden to the bacteria. • To test whether the growth in liquid or solid media could affect the rescue of escaper cells, liquid assays were done.

	• The platform was then extended to address the simultaneous targeting of tetA and blaTEM−116.
	  o To overcome competition between gRNAs, an N-3-oxohexanoyl-L-homoserine lactone (HSL)-inducible dCas9 module was placed in the same plasmid.
	• The one-plasmid system with CRISPRi array was reprogrammed to inhibit the expression of two t ARGs, blaNDM and mcr, conferring resistance to meropenem and colistin (last-resort drugs).
	• The platform was tested in clinical isolates.
	  o For delivery, a trans-conjugative platform consisting of pTA-Mob helper plasmid, bearing the RK2 conjugative machinery53, and a mobilizable vector, carrying the CRISPRi circuitry and an origin of transfer was used.
	  o The platform was further tested with different isolates and different antibiotics to evaluate potential.

	Main findings • The resistance to four antibiotics was significantly reduced by individual and multi-gene targeting of ARGs in E. coli. o Escaper analysis confirmed absence of mutations in targeted strands. • Meropenem, colistin and cefotaxime susceptibility was successfully increased (MIC up to > 4-fold) and growth delay (up to 11 h). • For Tetracycline re-sensitization in microplate assays by single ARG targeting, MIC of the resistant strain was more than 10-fold higher than that of the sensitive strain. tetA inhibition in the sCRISPRi strain increased in the time needed for population recovery in presence of sub-inhibitory TC concentrations. • blaTEM−116 inhibition by a gRNA targeting bla promoter led to a MIC of 5000 μg/ml & a growth delay of 6–7 h for AMP between 10 and 1000 μg/ml.

	• For multi-gene targeting systems, TC-treatment was the same as for single-gene. For AMP a major improvement over the previous platform was observed. • The data shows that the one-plasmid CRISPRi platform resulted in similar performance between individual guide RNA and two-spacer systems when targeting ARGs present in vLC and MC plasmids. • The data showed a clear effect on colistin re-sensitization even using only one gRNA integrated in a double-spacer CRISPRi array. o Escapers with an inactivated CRISPRi system still occur at sub-lethal MER- and COL-concentrations. • The proportion of AMP-, MER- and COL-resistant transconjugants effectively re-sensitized to the respective target antibiotic ranged from 71 to 99.9% o Growth Inhibition Assays ensured that susceptibility to both MER and COL was

	successfully increased • After testing the platform in multiple media, they determined that antibiotic sensitivity increased from nutrient composition. • The CRISPRi platform successfully increased the sensitivity to the tested antibiotics. However, re-sensitization varied and dependent on different factors such as strains, media and growth conditions. **Limitations / biases** • Limited scope of experiment; few antibiotics, resistance genes, and clinical isolates were tested. • Further studies are needed to test delivery to other bacteria, and test whether conjugation efficiency can be improved.
Research Question/Problem/ Need	Can CRISPR interference effectively inhibit Antibiotic-Resistant Genes in E. Coli?

Important Figures	https://www.nature.com/articles/s41598-024-81989-5/figures/3
VOCAB: (w/definition)	CRISPR interference: a type of CRISPR that inhibits gene expression without damaging target DNA. TOP10 F': A strain of chemically competent E. coli bacteria, treated to be more receptive in taking up plasmids. MG1655: A strain of E. coli with minimal genetic editing. IC_{99} values: The concentration of a substance needed to inhibit a specific biological process by 99%.
Cited references to follow up on	Neil, K., Allard, N., Roy, P., Grenier, F., Menendez, A., Burrus, V., & Rodrigue, S. (2021). High-efficiency delivery of CRISPR-Cas9 by engineered probiotics enables precise microbiome editing. *Molecular systems biology*, *17*(10), e10335. https://doi.org/10.15252/msb.202110335 Li, Q., Sun, M., Lv, L., Zuo, Y., Zhang, S., Zhang, Y., & Yang, S. (2023). Improving the Editing Efficiency of CRISPR-Cas9 by Reducing the Generation of Escapers Based on the Surviving Mechanism. *ACS synthetic biology*, *12*(3), 672–680. https://doi.org/10.1021/acssynbio.2c00619
Follow up Questions	How can plasmid uptake be increased to boost efficiency of CRISPR methods?

Article #2 Notes: PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures

Source Title	RNA Biology
Source citation (APA Format)	Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., & Randau, L. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA biology*, *16*(4), 504–517. https://doi.org/10.1080/15476286.2018.1504546
Original URL	https://doi.org/10.1080/15476286.2018.1504546
Source type	Scientific Article
Keywords	Protospacer-adjacent motifs (PAM), CRISPR, DNA recognition, Cas proteins
#Tags	
Summary of key points + notes (include methodology)	Protospacer-adjacent motifs are typically short sequences usually required for Cas protein DNases and RNases activity in bacteria. However, potentially due to the exposure to viral mutation,

	modifications and anti-CRISPR proteins, PAM recognition sequences have developed to become in terms of recognition, requirements, flexibility, length and structure. The variety in PAM recognition systems has proven to be useful for genomic editing, as well as for bacteria's own survival.
	Notes:
	Classification
	• **Paper type:** (Empirical / **Review** / Methods / Design / Other)
	Main Themes / Topics Covered
	• Identification of PAM Sequences
	  ○ Web tools were created to extract spacer sequences and to identify potential target sequences for *in silico*
	  ○ Plasmid depletion assays involve the insertion of a randomized DNA sequence adjacent to a target sequence within a plasmid that is transformed into a host with an active CRISPR-Cas system, but this is experimental.
	  ○ An *in vivo* method called PAM-SCANR involves adding catalytically dead Cas9 variant (dCas9) to a target library
	• PAM recognition by adaptation modules

	○ The naïve adaptation process integrates sequences that have not been encountered before. ○ Primed adaptation results in the integration of sequences that partially match pre-existing spacers. ▪ Cas1 and Cas2 proteins are essential for both types, as well as Cas4. ▪ Cas9 has an important PAM recognition mechanism. • PAM recognition by type i crispr-cas effector complexes ○ Type 1 CRISPR uses Cascade (CRISPR associated complex for antiviral defense) for target identification, with 8 currently described variations. ○ The type I-E Cascade has multiple structural differences, with more promiscuous PAM recognition in comparison to other type I or type II systems. ○ Subtype I-F also has structural differences, and targets foreign DNA with a PAM element with two consecutive G-C base pairs. ▪ A minimal subtype called type I-Fv has some structural differences from I-F but

still recognizes the two consecutive G-C base pairs.

- Other type I systems are not as well known.

- PAM recognition by type II effector proteins
  - Type II CRISPR-Cas systems employ a single Cas9 effector protein and have been reengineered as a gene editing tool.
    - The Cas9 protein is responsible for PAM recognition. However, multiple attempts have been made to mitigate or broaden the targeting of PAM sequences.
    - Anti-CRISPR proteins have also been identified and suggested to allow for modulation of Cas9 editing approaches.
- PAM recognition by type v effector proteins
  - Type V systems are defined by the protein Cas12 and have 5 subtypes. They have similar architecture to Cas9, but different mechanisms. For instance, while Cas9 is only base read-out, Cas12 is base & shape.
    - Cas12 can be used for genome editing, with it's benefits of small size, tracrRNA independency and asymmetric cleavage

sites.

- Prevention of autoimmunity in type III crispr-cas systems
    - Type III is found in both Archaea and Bacteria. Type III-A systems usually carry an adaptation module. Most III-B , III-C, or III-D systems do not, depending on others to incorporate new spacers.
    - Interference in type III diverged from type I in that they both require target sequence recognition for DNA degradation.
        - type III-A and B avoid self-targeting by Cas10 as long as the 5′-end of the crRNA is complementary to the 3′-flank of the target RNA protospacer
- Target recognition by type VI rnases
    - The type VI system has a single-effector RNA-guided RNases that has been classified into 4 subtypes

Key Findings or Consensus

- The actions of Cas RNases/Dnases require the absence of complementarity between crRNA tags and the protospacer.

	• CRISPR-Cas systems vary widely with structure and mechanisms used to ensure target specificity.
Research Question/Problem/ Need	How do PAM sequences differ across CRSIPR systems?
Important Figures	https://pmc.ncbi.nlm.nih.gov/articles/PMC6546366/figure/F0001/
VOCAB: (w/definition)	R-loop: A three-stranded nucleic acid where a DNA-RNA hybrid forms, displacing one strand of the DNA Glutamine wedge: Component of the CRISPR-associated protein Cse1 that recognizing PAM region of DNA. Haloarchaeal: Class of archaea
Cited references to follow up on	Jones, D. L., Leroy, P., Unoson, C., Fange, D., Ćurić, V., Lawson, M. J., & Elf, J. (2017). Kinetics of dCas9 target search in *Escherichia coli*. *Science*, *357*(6358), 1420–1424. https://doi.org/10.1126/science.aah7084
Follow up Questions	Can we utilize Cas protein's scanning abilities to recognize different sequences other than PAM?

Article #3 Notes: CRISPR technologies and the search for the PAM-free nuclease.

Source Title	Nature Communications
Source citation (APA Format)	Collias, D., & Beisel, C. L. (2021). CRISPR technologies and the search for the PAM-free nuclease. *Nature Communications*, *12*(1), 555. https://doi.org/10.1038/s41467-020-20633-y
Original URL	https://doi.org/10.1038/s41467-020-20633-y
Source type	Scientific Article
Keywords	PAM, Protospacer Adjacent Motif, CRSIPR, antibiotic resistance, bacteria, ortholog mining, protein engineering
#Tags	
Summary of key points + notes (include methodology)	Protospacer adjacent motifs (PAMs) are essential for the recognition of self-versus non-self in bacteria CRISPR systems but consequently limit CRISPR editing to target sequences that begin with a PAM sequence and are thus recognizable. However,

	many research attempts have been initiated recently to relax PAM recognition needs, namely natural ortholog mining and protein engineering. While PAM-free systems have not yet been found, attempts have uncovered more diverse PAMs than previously thought, and several attempts have significantly relaxed the recognition, potentially making future CRISPR editing a significantly more flexible tool. **Notes:** **Classification** • **Paper type:** (Empirical / **Review** / Methods / Design / Other) **Main Themes / Topics Covered** • A growing need for flexible targeting with Cas nucleases o While original CRISPR editing did not require PAM relaxation but instead introduced insertions or deletions (indels) through nonhomologous end joining, more recent CRISPR editing prompts the need for relaxed PAM. • Mining natural Cas orthologs for altered PAM recognition o In previous years, multiple Cas9 nucleases from model bacteria were being characterized for editing with little consideration for PAM diversity. o Many different PAM sequencing studies reveal PAMs are not solely a consensus sequence or a motif and instead represent a landscape of sequences with different extents of recognition.

- One study screened 70 Cas9 orthologs and uncovered an assortment of PAM profiles, including variants recognizing C-rich (, T-rich and A-rich PAMs

- Amino-acid identity analyses comparing PID of SpyCas9 and other Streptococci Cas9 nucleases led to the identification of Streptococcus canis Cas9, one of the most relaxed profiles observed so far in nature.

- PAM profiles for HkCas12a and PiCas12a suggest that further mining of Cas12a may identify additional and highly diverse PAMs.

- The Cas13 single effectors from Type VI systems provide a foundation on which to obtain PAM-free nucleases for other CRISPR-based applications.

- The known diversity of Cas nucleases supports PAM recognition as a flexible feature that can be altered with few mutations.

- Applying protein engineering to alter PAM recognition
  - Protein engineering has proven to be a powerful means to alter PAM recognition starting from individual CRISPR nucleases, and can be used to drive PAM sequences to more technologically useful tools.
  - The most relaxed PAM preference to date was evolved from a SpyCas9, making PAM-free SpyCas9 almost within reach.
    - There have been attempts to engineer nonnatural PAM preferences for SpyCas9

	○ Similar engineering approaches are also being applied to other Cas9 orthologs. ○ The natural diversity of PAM preferences can be exploited to meld engineering approaches and create variants. ○ Attempts have also been made to leverage PAM recognition by Cas12a. • Anticipated trade-offs with a PAM-free nuclease ○ PAM-free nuclease has the ability to target any sequence, greatly simplifying the selection of sites with high on-target but low off-target activity, generating predictable disruptive indels, or placing the base-editing window directly over the target nucleotide. ○ Self-targeting of this DNA would be immediate and unavoidable ○ Nucleases with no PAM requirements would have to interrogate every sequence in the genome, causing extended timescales and an increased propensity for off-targeting. **Key Findings or Consensus** • Engineering efforts for SpyCas9 have relaxed it to one of two bases at a single position. • Several methods for PAM realization through natural ortholog mining and protein engineering have shown success in relaxing PAM requirements.

Research Question/Problem/ Need	How are natural ortholog mining and protein engineering currently being used to relax PAM requirements of CRISPR ?
Important Figures	https://www.nature.com/articles/s41467-020-20633-y/figures/4
VOCAB: (w/definition)	**Phylogenetic: relating to the evolutionary development and diversification of a species or group of organisms, or of a particular feature of an organism.** **Orthologous: genes in different species that evolved from a common ancestral gene through speciation events.**
Cited references to follow up on	Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X., Makarova, K. S., Koonin, E. V., Sharp, P. A., & Zhang, F. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, *520*(7546), 186–191. https://doi.org/10.1038/nature14299
Follow up Questions	The article states that PAM sequences are significantly more diverse than previously thought, albeit with preference. Given this, how is it possible that none of the PAM sequences occur naturally in bacteria DNA?

Article #4 Notes: CRISPR-Cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa*

Source Title	The ISME Journal
Source citation (APA Format)	Wheatley, R. M., & MacLean, R. C. (2021). CRISPR-Cas systems restrict horizontal gene transfer in *Pseudomonas aeruginosa*. *The ISME Journal*, *15*(5), 1420–1433. https://doi.org/10.1038/s41396-020-00860-3
Original URL	https://doi.org/10.1038/s41396-020-00860-3
Source type	Scientific Journal
Keywords	CRISPR-Cas, HCT, Horizontal Gene Transfer,
#Tags	#Introduction #Body
Summary of key points + notes (include methodology)	CRISPR-Cas systems are adaptive immunity systems that can uptake DNA for later recognition of hostile or harmful organisms. Increasing evidence shows DNA adaption in P.

	aeruginosa bacteria causes CRISPR-Cas systems to consequently target components of horizontal gene transfer. Bacteria with active CRISPR-Cas systems are associated with a lower genome size, a higher GC content, and a significant percentage of spacers that target either integrative conjugative elements or the conserved conjugative transfer apparatus. Thus, it is logical to conclude that CRISPR systems inhibit HGT, though why CRISPR-Cas systems target these is debated. **Notes:** **Classification** • **Paper type:** (**Empirical** / Review / Methods / Design / Other) **Key methods** • P. aeruginosa was used due to the diversity and plasticity of their genomes, and because of high variability of CRISPR-cas presence. • The study analyzed 300 high-quality assembled genomes (including 201 complete genomes). o ICE sequences were downloaded from the ICEberg 2.0 database, containing 552 sequences o Plasmid sequences were downloaded from a curated database of plasmid sequences containing 10,892 complete plasmid sequences o Conjugative transfer gene sequences were downloaded from annotated P. aeruginosa genes.

	• CRISPRCasFinder was used to predict the presence of CRISPR arrays and cognate Cas proteins.
	  ○ CRISPRCasFinder was used to identify and type Cas systems in genomes with predicted CRISPR loci
	• Abundance of ICE and prophage in P. aeruginosa genomes was the chosen target of the study
	• The influence of phage and ICE was quantified via a systematic search in the genomes with intra-ST CRISPR variability.
	Main findings
	• Acr genes were found in 20/149 genomes with a CRISPR-Cas system, leaving 129 genomes that were predicted to encode functional CRISPR-Cas systems.
	• CRISPR-Cas systems were associated with smaller P. aeruginosa genome size.
	  ○ Genomes with both CRISPR-Cas systems and Acr genes were significantly larger than CRISPR genomes with no Acr genes.
	• The article found that the presence of CRISPR-Cas was associated with higher genomic GC content.
	• A large proportion of CRISPR array spacers (30.52%) were predicted to target phage DNA.
	  ○ (5.61%) were predicted to target ICE, plasmids and conjugative transfer genes.

	○ The remaining had no matches in the system • The study identified self-targeting spacers in 70/300 genomes. ○ Self-targeting spacers were common in genomes predicted to be inactivated due to presence of Acr genes (17/20 genomes) ○ Self-targeting spacers were also found in 43/129 genomes predicted to contain a functional CRISPR-Cas system. • Spacers targeting ICE or conjugative transfer system genes were widespread, occurring in 111/129 CRISPR(+) genomes • The study found CRISPR(−) genomes were larger than their CRISPR(+) counterparts **Limitations / biases** • Small dataset
Research Question/Problem/ Need	Do CRISPR-Cas systems restrict horizontal gene transfer in P. aeruginosa?
Important Figures	https://www.nature.com/articles/s41396-020-00860-3/figures/1
VOCAB: (w/definition)	GC content: Percentage of Guanine of Cytosine in a molecule Integrative conjugative elements: Mobile genetic elements that can integrate into a host genome via conjugation.

Cited references to follow up on	Gophna, U., Kristensen, D. M., Wolf, Y. I., Popa, O., Drevet, C., & Koonin, E. V. (2015). No evidence of inhibition of horizontal gene transfer by CRISPR-Cas on evolutionary timescales. *The ISME journal*, *9*(9), 2021–2027. https://doi.org/10.1038/ismej.2015.20
Follow up Questions	Does an increase in frequency of HGT-targeting spacers lead to a proportionally lower HGT rate? Does CRISPR-Cas limit HGT long-term?

Article #5 Notes: Advances in the Study of Bacterial Toxins, Their Roles and Mechanisms in Pathogenesis

Source Title	The Malaysian Journal of Medical Sciences
Source citation (APA Format)	Ghazaei, C. (2022). Advances in the study of bacterial toxins, their roles and mechanisms in pathogenesis. *The Malaysian Journal of Medical Sciences (MJMS)*, *29*(1), 4–17. https://doi.org/10.21315/mjms2022.29.1.2
Original URL	https://doi.org/10.21315/mjms2022.29.1.2
Source type	Academic Journal
Keywords	Bacteria, toxin, pathogenesis, mechanism,
#Tags	#Introduction
Summary of key points + notes (include methodology)	In this review, the different types and classifications of bacterial toxins, as well as their self-helping benefits and harmful effects to the host are detailed. Exotoxins are classified into three types, superantigens, membrane disrupting toxins, and A-B toxins, all three classifications primarily damage host cells or systems. Endotoxins are present on the outer portion of the cell wall of Gram-negative bacteria and play a diverse role; for instance,

toxin-antitoxin modules are responsible for bacteria's stagnant state when surviving in low nutrient environments, the self-suicide of bacterial cells when population is too high, and biofilm formation. There are six types of TA modules. Finally, Cytolethal Distending Toxins, which act as genotoxins are also produce by bacteria. Interesting, some studies indicate their genotoxicity may aid the progression of certain cancers.

Notes:

Classification

- **Paper type:** (Empirical / **Review** / Methods / Design / Other)

Main Themes / Topics Covered

- Types of Exotoxins
    - Superantigens (Type I Toxins)
        - Do not enter cell but provokes an immune response that has a range of side effects.
    - Membrane Disrupting Toxins (Type II Toxins)
        - Disrupt the plasma membrane of the host cell, thus causing cell lyses.
        - Either kill cell or enter the cell.
            - Two further types: one forms protein channels in the plasma membrane; the

other disrupts the integrity of membrane phospholipids.

- A-B Toxins (Type III Toxins)
  - A-subunit is an enzyme with toxic activity, and B-subunit is the binding component, binding exotoxin to the human cell membrane.
    - Efficient as little exposure to the host's immune system.
- Toxin-Antitoxin Modules
  - Inhibit their own cell growth or death during overexpression and take part in epigenetic regulatory mechanisms in bacteria.
    - Can also comprise a protein toxin that inhibits cell growth by interfering with vital processes and a protein or RNA antitoxin that protects its own cell by sequestering the toxin's action.
- Types of toxin-antitoxin modules
  - The method used to inhibit the toxin classifies the antitoxin into six types: TA modules of types II, IV, V and VI are protein-natured antitoxins, and types I and III are small regulatory RNAs.
- Toxin-antitoxin and their role in bacterial pathogenesis

	○ TA modules can be involved in pathogenesis and mechanisms. • Toxin-antitoxin modules and bacterial persistence ○ TA systems create antibiotic resistance ○ TA modules are stress responders that slow down microbe's physiological activities by inhibiting processes. ○ TA modules can mimic host's defense mechanisms that the pathogen encounters during infections in humans ○ • Cytolethal Distending Toxins — Genotoxins ○ CDTs target and modulate the eukaryotic cell cycle by inducing DNA lesions. • Cytolethal Distending Toxins Act Like Enzymes ○ CDTB Acts as DNase ▪ Studies show CDTB subunit has nuclear localisation signal (NLS) sequences, leading to the nuclear localisation of CDTB subunit ▪ CDTs can induce DNA damage response (DDR), leading to double-stranded breaks

	in DNA ▪ • CDT Targets Host Defence Systems and Play Role Invirulence o CDT targets host cells at three levels. ▪ promotion of infection ▪ promotion of inflammatory responses ▪ impairment of acquired immunity **Key Findings or Consensus** • In pathogenic bacteria, TA plays a beneficial role to the bacteria, helping downplay and avoid detection, influence the system, and act as a best-programmed cell death. • CDTs cause damage to the cells of a host.
Research Question/Problem/ Need	How do toxins produce by bacteria aid the species and harm the host?
Important Figures	No figures
VOCAB: (w/definition)	Genotoxicity: the property of chemical agents that damage the genetic information within a cell. Gram-negative bacteria: bacteria with a thinner peptidoglycan

	wall, and are of major concern due to their high resistance.
Cited references to follow up on	Wen, Y., Behiels, E., & Devreese, B. (2014). Toxin-Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity. *Pathogens and disease*, *70*(3), 240–249. https://doi.org/10.1111/2049-632X.12145
Follow up Questions	Would the inhibition of TA modules lead to decreased survival rates outside of a host?

Article #6 Notes: Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids

Source Title	Nucleic Acids Research
Source citation (APA Format)	Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R. A., Randau, L., Sørensen, S. J., & Shah, S. A. (2019). Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Research*, *48*(4), 2000–2012. https://doi.org/10.1093/nar/gkz1197
Original URL	https://doi.org/10.1093/nar/gkz1197
Source type	Academic Journal
Keywords	CRISPR-cas, plasmids, Type IV
#Tags	
Summary of key points + notes (include methodology)	CRISPR-cas systems are divided into two groups, Class 1 and 2, and further divided into types I, III, IV, and II, V, VI. This article studies class 1 CRISPR-cas system IV. The major takeaway is

	that type IV is mainly found within plasmids, and targets other plasmids. This contrasts with the usual purpose of CRISPR-cas systems to target viral DNA. Currently, the mechanistic basis of the bias remains unclear. A current theory argues that similar entities more commonly compete for similar needs, thus creating competition between plasmids and potentially explaining the evolutionary need for these described Type IV systems. **Notes:** **Classification** • **Paper type:** (**Empirical** / Review / Methods / Design / Other) **Key methods** • Bacterial and archaeal complete and draft genomes were obtained from genebank • A neighbour-joining tree was constructed and used to pick diverse representative type IV systems, which were then annotated manually using PSI-BLAST. • Out of 883 detected Csf2 proteins… 69 diverse representatives were selected for further analysis. • An aggregate protein similarity tree was then generated including all proteins. • 481 type IV and 535 non-type IV arrays were identified. **Main findings** • Phylogenetic analysis revealed a richness of type IV gene arrangements and a complex evolutionary relationship.

	• A set of archaeal type IV modules were found to cluster as a clear outgroup. • Type IV systems had a strong targeting bias towards plasmids ○ Targeted plasmids tend to be relatively large **Limitations / biases** • Spacer match limitations • PAM recognition limitations.
Research Question/Problem/ Need	What are Type IV CRISPR-cas system, their role, and their mechanisms?
Important Figures	https://academic.oup.com/view-large/figure/199624154/gkz1197fig3.jpg
VOCAB: (w/definition)	Spacer acquisition machinery: the molecular components (primarily Cas1-Cas2) responsible for incorporating short DNA fragments into CRISPR locus DinG helicase: member of the superfamily 2 DNA helicases Phylogenetic analysis: a method that reconstructs the evolutionary relationships between a set of biological entities, such as species, genes, or proteins, by analyzing their shared traits.

Cited references to follow up on	Li, W., & Godzik, A. (2006). CD-HIT: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158
Follow up Questions	Would a higher amount of plasmids with Type IV lead to lesser amounts of other types of plasmids, thus limiting HGT?

Article #7 Notes: Using Machine Learning to Predict Antimicrobial Resistance—A Literature Review

Source Title	Antibiotics
Source citation (APA Format)	Sakagianni, A., Koufopoulou, C., Feretzakis, G., Kalles, D., Verykios, V. S., Myrianthefs, P., & Fildisis, G. (2023). Using machine learning to predict antimicrobial resistance: A literature review. *Antibiotics*, *12*(3), 452. https://doi.org/10.3390/antibiotics12030452
Original URL	https://doi.org/10.3390/antibiotics12030452
Source type	Scientific Journal
Keywords	Machine learning, artificial intelligence, antimicrobial resistance, AMR, antibiotic stewardship, clinical decision support tools
#Tags	
Summary of key points +	The use of Artificial Intelligence is becoming increasingly more

notes (include methodology)	diverse, with the application of AI to the antibiotic resistance crisis being one example, as well as being of the "most crucial areas of interest". This review article covers the many different applications of machine learning algorithms to this crucial issue. The most popular types are currently, "supervised machine learning algorithms. . . used with linear and logistic regression, k-nearest neighbors (k-NN), support vector machine (SVM), decision tree (DT), random forest (RF). . . gradient boosting machine (GBM). . . neural networks and deep learning approaches." These models can have vast implications from diagnosis of AMR, reducing time but increasing cost, prediction of AMR, prescription, clinical decisions, and even environmental predictions. While the majority of these have promising results and while some have been supported by following studies, there still remain signifigant challenges. As are all AI models, they are limited by current available data. Additionally, there are concerns regarding accuracy and reliability. It is emphasized that while AI is promising within the field, clinical judgement is still held above. **Classification** • **Paper type:** (Empirical / **Review** / Methods / Design / Other) **Main Themes / Topics Covered**

Machine Learning (ML) Applications in the Field of AMR

- Diagnosis of AMR
    - The implementation of ML methods has reduced the time of bacterial susceptibility profiling < three hours for the flow-cytometry AST and only 30 min for the infrared spectrometry
        - This is costly
    - ML-based MALDI algorithms are available for micro-organism identification
    - ML-driven predictive models of AMR could bridge specimen collection and results from molecular and genotypic susceptibility analysis, allowing for time-sensitive options.
- Prediction of AMR
    - ML tools have been used by several researchers to predict antibiotic susceptibility patterns of pathogens, allowing for the selection of the most appropriate treatment.
        - One such study used a decision tree for the prediction of extended-spectrum β-lactamase production in *E. coli* and *Klebsiella spp*. bacteremia based on patient epidemiological and microbiological data
        - Another study, an open-source machine learning algorithm trained in predicting antibiotic resistance for three Gram-negative bacteria isolated from hospital

	patients' blood and urine performed better than medical staff. ▪ Another model train on Electronic Health Records successfully predicted resistance in a patient culture. ▪ ML algorithms can accelerate the workflow • Machine-Learning-Assisted Antibiotic Prescription o One model reduced inappropriate drug recommendations to 5% in comparison to 9% by medical staff. ▪ Most of these data-driven models lack generalizability and require more diverse datasets. • Machine Learning-Assisted Clinical Decision Support Systems (ML-CDSS) o MLM Clinical Decision Support Systems have been developed, but they can be wrong which could have severe impacts. o MLM for CDSS could help maintain antimicrobial stewardship and thus limit spread. o Some models can be accessible in impoverished areas, which are often large contributors to AMR. o Deep learning models have demonstrated to have the best results.

	• Prediction of AMR in the Environment Employing AI/ML ○ Machine learning models have been used for the prediction of ARGs in various environments, such as in beaches, soil, and wastewater. **Key Findings or Consensus** • MLM can offer flexible solutions to AMR in healthcare, but do not yet prove to consistently outperform typical statistics.
Research Question/Problem/ Need	How can Machine-Learning Algorithms be used for the Antimicrobial Resistance Crisis?
Important Figures	N/A
VOCAB: (w/definition)	Logistic regression: a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome K-nearest neighbors: a supervised machine learning algorithm that classifies a new data point by finding the K most similar data points (neighbors) from a training dataset and assigning the new point to the majority class among those neighbors. Support vector machine: supervised machine learning algorithm primarily used for classification and regression tasks. It operates by finding an optimal hyperplane that separates data points into different classes.
Cited references to	Su, M., Satola, S. W., & Read, T. D. (2019). Genome-Based

follow up on	Prediction of Bacterial Antibiotic Resistance. *Journal of clinical microbiology*, *57*(3), e01405-18. https://doi.org/10.1128/JCM.01405-18
Follow up Questions	Can ML or another form of AI be used to predict mutations?

Article #8 Notes: Repurposing CRISPR-Cas systems as DNA-based smart antimicrobials

Source Title	Cell and Gene Therapy Insights
Source citation (APA Format)	Barrangou, R., & Ousterout, D. G. (2017). Repurposing CRISPR-Cas systems as DNA-based smart antimicrobials. *Cell and Gene Therapy Insights*, *3*(1), 63–72. https://doi.org/10.18609/cgti.2017.008
Original URL	https://doi.org/10.18609/cgti.2017.008
Source type	Journal Article
Keywords	CRISPR-cas, phages, antimicrobials
#Tags	#Intro
Summary of key points + notes (include methodology)	CRISPR-cas was discovered in the 1980s and has since be repurposed as a gene editing tool outside of the bacterium. However, the use of the CRISPR arrays naturally occurring in bacteria has been significantly underused. This mini-review

	covers the few articles that had explored the repurposing of CRISPR-cas systems to be self-targeting at the time of writing. There are two described methods, the insertion of self-targeting spacers, and the delivery of complete CRISPR arrays to bacteria. Delivery is often done with lytic or lysogenic phages. All studies were successful in triggering autoimmunity, and due to the nature of CRISPR, there are very little risks for targeting of the infection's host. Still, there remains challenges with delivery, and this method involves extensive knowledge of the given bacteria. **Classification** • **Paper type:** (Empirical / Review / Methods / Design / **Other**) (Expert Opinion) **Opinion:** **CRISPR-Cas systems can be repurposed as specific and programmable antimicrobials by directing Cas nucleases to target bacterial DNA.** **Main arguments:** • Antibiotics cause resistance and disrupt healthy microbiomes, creating need for targeted alternatives. o Furthermore, the antimicrobials cannot harm the host.

	• CRISPR-Cas systems provide sequence-specific DNA targeting, enabling self-targeting antibiotics. • Engineered bacteriophages can deliver the CRISPR payloads. **Limits:** • Preclinical, needs further validation
Research Question/Problem/ Need	Can CRISPR-cas be repurposed as self-targeting microbial?
Important Figures	N/A
VOCAB: (w/definition)	Seminal: New or groundbreaking Lysogenic: a virus that integrates its DNA into the bacterial host Lytic: a virus that use's bacteria cell machinery to replicate it's own DNA
Cited references to follow up on	Gomaa, A. A., Klumpe, H. E., Luo, M. L., Selle, K., Barrangou, R., & Beisel, C. L. (2014). Programmable Removal of Bacterial Strains by Use of Genome-Targeting CRISPR-Cas Systems. *mBio*, *5*(1), e00928-13. https://doi.org/10.1128/mBio.00928-13
Follow up Questions	Can mutations be inserted preemptively to trigger targeting and inhibition of mutations?

Article #9 Notes: Is Smaller Better? A Proposal to Use Bacteria For Neuroscientific Modeling

Source Title	Frontiers in Computational Neuroscience
Source citation (APA Format)	Ram, A., & Lo, A. W. (2018). Is smaller better? A proposal to use bacteria for neuroscientific modeling. *Frontiers in Computational Neuroscience*, *12*, 7. https://doi.org/10.3389/fncom.2018.00007
Original URL	https://doi.org/10.3389/fncom.2018.00007
Source type	Journal Article
Keywords	Quorum sensing, neural networks (computer), *Bacillus subtilis*, cell-cell communication, network models
#Tags	https://pmc.ncbi.nlm.nih.gov/articles/PMC5829041/figure/F1/
Summary of key points + notes (include methodology)	Bacteria and neurons have many common or similar features, notable of which is quorum sensing, a form of bacterial cell-cell signaling. The similarities may indicate a common evolutionary origin of that of eukaryotic cell-cell signaling we see in neurons.

	Additionally, it adds basis to the hypothesis that bacteria, through horizontal gene transfer, played a role in eukaryotic cell-cell signaling evolution. These similarities are so important that the author of the paper is arguing for the use of bacteria to model neurons, as a simpler, faster, and cheaper method. Indeed, the author can list several notable commonalities, including ion-based communication and quorum sensing vs. neuron communication. Essentially, the author argues the similarities between bacteria and neurons may be a beneficial means to study neurological diseases, though admittedly not without the limitation of the bacteria being, plainly, not a neuron. **Classification** • **Paper type:** (Empirical / Review / Methods / Design / **Other**) (Hypothetical/Theory) **Theory:** **Bacteria would serve as an accurate, cheaper and easier model for neurons due to their similarities between multiple functions and similar structures.** **Main arguments:** • Bacteria and neurons share similar structures. o Ion channels o Membrane o Receptors • Bacteria are easier and cheaper for laboratory tests than eukaryotic cells, especially neuronal.

	• Bacteria and neurons have multiple similar functions (e.g. quorum sensing and neuron communication).
	  ○ operate in series and in parallel
	  ○ able to develop into a hierarchical multi-circuit system
	  ○ provide both excitatory and inhibitory signals
	• There is a known link between microbiota composition and the central nervous system.
	  ○ These links are observable in many different scenarios.
	  ▪ Bacteria could provide more complete models of neurological disorders.
	• Bacteria biofilms display a gradient action potential like action.
	• Some behavioral similarities, like chemotaxis, can be observed.
	
	Limitations / biases:
	• Theoretical
	• Significant differences.

Research Question/Problem/ Need	Can bacteria be used to model neurons?
Important Figures	https://pmc.ncbi.nlm.nih.gov/articles/PMC5829041/figure/F1/
VOCAB: (w/definition)	Chemotaxis: the directed movement of a cell or organism in response to a chemical gradient

	Ameliorating: to improve something that was previously considered bad.
Cited references to follow up on	Snyder, S. H., & Innis, R. B. (1979). Peptide neurotransmitters. *Annual Review of Biochemistry*, *48*, 755–782. https://doi.org/10.1146/annurev.bi.48.070179.003543 Bassler B. L. (2002). Small talk. Cell-to-cell communication in bacteria. *Cell*, *109*(4), 421–424. https://doi.org/10.1016/s0092-8674(02)00749-3
Follow up Questions	Would drugs typically used on neurons influence bacteria, too?

Article #10 Notes: DRAMMA: a multifaceted machine learning approach for novel antimicrobial resistance gene detection in metagenomic data

Source Title	Microbiome
Source citation (APA Format)	Rannon, E., Shaashua, S., & Burstein, D. (2025). DRAMMA: a multifaceted machine learning approach for novel antimicrobial resistance gene detection in metagenomic data. *Microbiome*, *13*(1). https://doi.org/10.1186/s40168-025-02055-4
Original URL	https://doi.org/10.1186/s40168-025-02055-4
Source type	Journal
Keywords	Machine Learning Model, Antibiotic Resistance Genes, Metagenomes,
#Tags	#MLM
Summary of key points + notes (include methodology)	**Summary:** Antibiotic resistance is a significant global issue, that is exacerbated by the slow recognition of new antibiotic-resistant genes. While previous studies have attempted to design bioinformatic tools that could recognize ARGs from a genome, the genomes they were based upon were incomplete and current, rendering the tools unable to classify novel ARGs. Similarly, further works were able to define ARGs without the need for a predefined database but lack biological knowledge further than the sequence properties. Here, DRAMMA, a Random-Forest learning model was used to attempt to identify novel ARGs within a sequence. The model was trained on protein coding genes from 22,241 metagenomes and had ROC-AUC scores of 0.98 and 0.938 for the training set and mean PR-AUC scores of 0.857 and 0.668 for metagenomic five-fold cross validation and taxonomic. The model was

noted to have several limitations, including its weaknesses regarding Archaea and the inherent bias within its training set.

Notes:

Classification

- **Paper type:** (**Empirical** / Review / Methods / Design / Other)

Key methods

- Protein coding genes in large contigs were compiled from 22,241 metagenomes
    - -Genomes taken from NCBI WGS, BioProject PRJEB27054 and EBI's Mgnify
    - 34,311,250 proteins were collected
- 4 categories were extracted: amino acid patterns, amino acids properties, HGT signals and genomic context.
- Random forest model was chosen for its favorable tradeoff between predictive accuracy and computational efficiency.
- DRAMMA's accuracy was retested with another dataset to ensure that the high performance was not due to data leakage between groups that shared similar properties.
- The number of novel ARGs identified in different groups were normalized to account for bias of data sets.
- Known ARGs were labeled using DRAMMA-HMM-DB

	• To identify the novel predicted ARGs, the proteins were filtered from the original 649 million: o Proteins were further filtered by proteins with model score higher than threshold for 95% precision->1.28M, for unknown function (according to KEGG)->113.7k, clustered based on sequence similarity->3.26k, unknown function (according to BLAST) ->381, unknown (according to HH-suit search)->213. **Main findings** • DRAMMA had mean ROC-AUC scores of 0.98 and 0.938 for the training set and mean PR-AUC scores of 0.857 and 0.668 for metagenomic five-fold cross validation and taxonomic. • ARGs were highly enriched in human and animal microbiomes. • ARGs candidates most commonly provided resistance to beta-lactam antibiotics, which aligns with beta-lactam antibiotics being the most prescribed antibiotics. • ARGs encoding efflux pumps were more easily recognized **Limitations / biases** • The model appears to be much less adept at predicting mutations originating from Archaea. • Inherent biases within the dataset likely hindered the model's ability regarding new resistance families. • The ARGs were labeled using HMM profiles which may lead to false positives. • The model was trained using large protein contigs despite short contigs being able to code for resistance genes. • Sensitivity and accuracy of bioinformatics tools used likely impacted downstream analysis.

Research Question/Problem / Need	Can a machine learning model be used to predict antibiotic resistant genes that have no sequence similarities to any known resistance gene?
Important Figures	https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-025-02055-4/figures/1
VOCAB: (w/definition)	ROC-AUC Score: A single summary metric for classification of a model's performance across different thresholds. PR-AUC Score: A single summary metric used to evaluate the performance of a model, particularly for imbalanced datasets.
Cited references to follow up on	Li, Y., Xu, Z., Han, W., Cao, H., Umarov, R., Yan, A., Fan, M., Chen, H., Duarte, C. M., Li, L., Ho, P. L., & Gao, X. (2021). HMD-ARG: hierarchical multi-task deep learning for annotating antibiotic resistance genes. *Microbiome*, *9*(1), 40. https://doi.org/10.1186/s40168-021-01002-3
Follow up Questions	If it is possible to identify ARGs without current matches, is it possible to project future ARGs?

Article #11 Notes: An Introduction to Machine Learning for Clinicians

Source Title	Journal of the Association of American Medical Colleges
Source citation (APA Format)	Rowe M. (2019). An Introduction to Machine Learning for Clinicians. *Academic medicine : journal of the Association of American Medical Colleges*, *94*(10), 1433–1436. https://doi.org/10.1097/ACM.0000000000002792
Original URL	https://doi.org/10.1097/ACM.0000000000002792
Source type	Journal
Keywords	Machine Learning, Clinicians, AI
#Tags	#MLM
Summary of key points + notes (include methodology)	As machine learning is rapidly integrated into healthcare, in fields ranging from surgery to diagnosis, an understanding of the basics of machine learning, it's challenges, and its implications become imperative. Machine learning describes the use of software algorithms to identify patterns in datasets and make statistical guesses as to what the most likely outcome would be. It is important to note this is not the "correct" answer, but what would follow previously observed patterns. Furthermore, MLM have several limitations. While data is the most crucial aspect of their training, clinical data is limited. Additionally, the data sets provided can have bias, or class imbalance. This skews the predictive ability of the model, and the bias may lead to social inequality and disparity. Models have also been known to find patterns regarding the wrong target- e.g. a model trained to identify skin cancer correlated the cancer to pictures with rulers (measuring the cancer). In the context of healthcare, MLM are unable to identify causation, only correlation. Finally, as MLM are becoming more complex, it becomes significantly more difficult for untrained clinicians to validate output. These risks with MLMs point to the importance of clinician input during