UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE VETERINARIA

TESIS DOCTORAL

Chassis de captura de cassettes de integron

Integron Cassette capture chassis

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Filipa Moutinho Trigo da Roza

DIRIGIDA POR

José Antonio Escudero García-Calderón

Madrid

Universidad Complutense de Madrid
Facultad de Veterinaria

CHASSIS DE CAPTURA DE CASSETTES DE INTEGRON

INTEGRON CASSETTE CAPTURE CHASSIS

Filipa Moutinho Trigo da Roza

Tesis Doctoral

Bajo la dirección de:

José Antonio Escudero García-Calderón

2024

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE VETERINARIA

PROGRAMA DE DOCTORADO EN VETERINARIA

TESIS DOCTORAL

CHASSIS DE CAPTURA DE CASSETTES DE INTEGRON

INTEGRON CASSETTE CAPTURE CHASSIS

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

FILIPA MOUTINHO TRIGO DA ROZA

DIRIGIDA POR

JOSÉ ANTONIO ESCUDERO GARCÍA-CALDERÓN

2024

“If you can dream it, you can do it.”

Walt Disney

Aos meus avós Mamelha, Manu, Fernando e António

Acknowledgments

One of the most rewarding parts of writing a PhD thesis is looking back and appreciating the journey—the challenges faced, the lessons learned, and the profound personal growth along the way. But even more fulfilling is remembering those who stood by our side throughout it all. As I was told (and, of course, had to see to believe), pursuing a PhD can often feel like a lonely effort—battling doubts, mistakes, and endless moments of "I should have known". Yet, I was lucky. I was surrounded by people who believed in me more than I ever did, and who constantly reminded me that failing is not losing but a step towards success.

While I will attempt to express my gratitude in a few words, I know I will probably fail (again!). And because saying thank you is deeply personal, I'll take the liberty of switching languages when needed.

Somos inevitavelmente um pouco do que nos fazem e por isso quero agradecer primeiro à minha família. Obrigada em especial à minha Mãe, ao meu Pai, e aos meus irmãos Marta, António e Luisinha por me fazerem sentir que não importa para onde vá ou o que seja (veterinária de cavalos, podadora de bonsais, doutorada em microbiologia ou formar a primeira girlsband de garagem?), tenho-vos sempre comigo. Obrigada à Lara e ao Tio Zé por me mimarem tanto, e aos tios e primos e às estrelas que temos a olhar por nós ("ursas, cisnes e leões...") por darem luz ao meu caminho.

Todo esto empezó con una gráfica en la pizarra, una moneda lanzada al aire, y una decisión que ya estaba tomada. Con un veterinario que cometió la locura de creer en otra veterinaria que no había hecho una PCR en su vida. Gracias jefecito por la confianza, por enseñarme el rigor científico, la pasión por cuestionar más y mejor, la obsesión por no dejar nada por explicar, y la capacidad de inspirar con gráficas hechas a mano (jaja). Aunque hayamos sufrido los dos, ha sido un placer crecer en este laboratorio y verte crecer como jefe, pero sobre todo saber que he ganado un amigo para toda la vida!

A los grandes MBAitas & compañía que han cruzado camino, les debo las risas, el apoyo en los días de lágrimas, y el hecho de haber llegado hasta aquí. Sin vosotros no sería posible. Prima, gracias por hacerme sentir en casa en una cuidad lejos de la mía!

Por las comidas compartidas (de Espe, claro!), los karaokes improvisados, los masajes imprescindibles y sobre todo, por obligarme a pedir ayuda. Alberto, gracias por hacer las mejores bromas de la historia de los laboratorios y por compartir el peso de ser primerizos del labo! Carlotinha, gracias por entenderme desde el minuto uno! Luchía, gracias por tener un consejo sabio y un dicho siempre a mano, como buena Mary Poppins :) Paulaner, gracias por escucharme sin juzgar, hacerme correr más rápido, imaginar que podríamos ganar un Nobel, llevarme a los mejores spots en el barrio y ser la vecina más cool de la historia (con Niccolo)! Francisco Manué, gracias por hacerme reír hasta llorar, por traer un poco de Sevilla a Madrid, y por enseñar que sigue existiendo gente honestamente buena en este mundo. Monsieur Kieffer, merci pour m'écouter, me conseiller et donner les abrazos les plus forts du monde. Pequeña Amalia, gracias por ser la felicidad del labo y tener siempre una sonrisa puesta! Laurita (o Doña Hortensia para los amigos) gracias por poner orden en este sitio, por hacerme reír y tener siempre una respuesta a todas mis dudas. Y gracias Patri, por la buena energía y ganas de aprender!

Ao melhor MBAita de todos (lo siento chicos), André: ainda não sei bem a quem agradecer por (finalmente!) te teres cruzado no meu caminho. Mas a ti quero agradecer-te por em tão pouco tempo te tornares no melhor companheiro científico que podia imaginar, no meu melhor amigo, e na minha pessoa favorita. Por me lembrares que a vida é bonita, por sorrires sempre que olhas para mim, por me ouvires durante horas a fio, por fazeres piadas iguais às minhas, saberes sempre o que penso, e me fazeres sentir que sou capaz de tudo. O melhor ainda está por escrever!

A la comunidad de microbiólogos de Madrid, gracias por acogerme y hacerme sentir una de vosotros. A Bruno, mi tutor de tesis, gracias por todo el apoyo en estos años, y también a todos de ARU, por ayudarnos a arrancar como grupo y asegurar que nos sentíamos parte del vuestro! A los grupos de Álvaro, Jero y Teresa, muchas gracias por tantas discusiones científicas estimulantes. Y claro, muchas gracias a Modesto, por toda la ayuda y colaboración en esta tesis.

Viver longe de casa não é fácil. Mas umas das coisas mais bonitas de o fazer é encontrar pessoas que acabam a ser "a família longe da família". Querida Xin Xen, não tenho muitas palavras para agradecer a tua amizade durante estes anos, o bom que foi sermos flatmates, conhecer o Luís (tu não largues esse homem! Hehehe. Obrigada

querido Buzón!) e depois a Vera (bebé preferida de todos os tempos), e agora o Tomás! E o quanto me ajudaste a chegar até aqui!! Obrigada também querida Alicia, por tanto em tão pouco tempo! Por demonstrares de que são feitos os verdadeiros amigos. Great things are coming! Obrigada queridos Chica e Vasco por tantos jantares, conversas boas e garrafas de Pleno! E obrigada à mais recente aquisição portuguesa de Madrid, João, pela boa companhia, pelas melhores recomendações de sci-fi e pela revisão metódica desta tese!

I will never forget my three months in Lausanne, as they came at a delicate moment in both my professional and personal life, and were a cure to (almost!) all my problems. First of all, thank you, Melanie, for having me in your lab, for teaching me, and for sharing your expertise in the *Vibrio* world! Thank you Grazia, for being the best lab neighbor (!!); thank you, Alexis and Céline, for taking me on early morning runs with always great conversations; David, for your crucial tips; and Lauriane and Sandrine, for all the help in the lab. And to the whole Blokesch Block, you know you rock! Beyond the lab, thank you to dear Doro and Paola for having me in your beautiful house and showing me the wonders of Switzerland. Y mil gracias Bego, porqué probablemente hayas salvado mi vida, dos veces!!

Obrigada aos que de alguma uma maneira me inspiraram e motivaram a fazer este doutoramento que estava para lá dos mais ambiciosos sonhos: ao Prof. Luís Lamas, à Prof. Manuela, à Natacha, à Teresa, à Mariana, e ao Hugo Pissarra. Gracias Erley por ser mi hermano científico mayor, siempre disponible a escucharme! Y gracias Olalla, por enseñarme que soy capaz y no estoy sola nunca.

Obrigada aos amigos de sempre, à Rita sobretudo, por seres o meu apoio incondicional e o espelho de realidade que às vezes preciso, mas também ao João, à Ana Sofia, ao Luís, à Carlota, à Carolina e à Mafalda. E aos novos amigos, Marco e Mariana, Vera, Gonçalo, Leonor e Ritinha, porque me mostrarem que a vida nunca nos deixa de trazer coisas boas!

Y por fin, gracias Madrid, por tanta magia, tanta vida, por tanto que he aprendido. Por hacerme dudar en qué sentido de la A5 está mi casa ♡

The work performed in this thesis was supported by a PhD scholarship from Fundação para a Ciência e Tecnologia (https://doi.org/10.54499/SFRH/BD/144108/2019) and an EMBO Scientific Exchange Grant.

Index

Cercedilla, 2020

Index

Acknowledgments **9**

Index **17**

Abstract **29**

Resumen **31**

Introduction **37**

1 – Antimicrobial Resistance 37

1.1 – Historical perspectives and definitions 37

1.2 – Resistance mechanisms 39

1.3 – Mobile genetic elements 42

1.3.1 – Forms of DNA exchange 43

1.3.2 – Types of elements 44

2 – Integrons 48

2.1 – A unique recombination system 48

2.1.1 – Stable platform 48

2.1.2 – Variable array 52

2.1.3 – External recombination sites 53

2.1.4 – Recombination reactions 54

2.2 – Types of Integrons 58

2.2.1 – Chromosomal integrons 58

2.2.2 – Mobile integrons 58

2.3 – Detection systems 60

2.4 – The integron as a biotechnological tool 62

3 – *Vibrio cholerae* 64

3.1 – An environmental bacteria that led to eight pandemics 64

3.2 – Superintegron 65

3.2.1 – Organization and importance 65

3.2.2 – 19 toxin-antitoxin systems 66

3.3 – Natural competence 68

Objectives 75

Materials and methods 79

Strains, plasmids, and primers used 79
Growth conditions 79
Plasmid constructions 79
E. coli chromosome constructions 80
V. cholerae chromosome constructions 81
Constructions by natural transformation 81
Clean-mutant generation 81
Killing assays 82
Classical recombination assays 82
Library generation 84
Library sequencing and analysis 85
Library sequencing and read processing 85
Cassette identification and quantification 85
Recombination through conjugation 86
SeqDelTA 87
Last deletion step 87
ΔSI strain sequencing 88
Mutation correction 89
Competition assays 89
Natural transformation assays 90
Recombination through natural transformation 90
Statistical analysis 91

Results 95

Chapter 1: Establishment of a platform to select for integron recombination events 95

1.1 – Characterization of different promoters in *E. coli* and *V. cholerae* 97
1.1.1 – Characterization of constitutive promoters 97
1.1.2 – Characterization of inducible promoters 98
1.1.3 – Characterization of inducible promoters through integron recombination 101
1.2 – CcdB/CcdA type II toxin-antitoxin system as a reporter 103

1.2.1 – Plasmid-based tool using CcdB/CcdA TA system as a reporter 104
1.2.2 – Chromosome-based tool using CcdB/CcdA type II TA system as a reporter ... 108
1.3 – SacB as a reporter 110
Chapter 2: The cassette harvester 117
2.1 – Plasmid setup 117
2.2 – Creating a cassette library from *V. cholerae* N16961 superintegron 118
2.3 – Library analysis 119
Chapter 3: A sentinel bacterium 124
3.1 – *E. coli* setup 124
3.2 – Conjugation of clinical plasmids 125
Chapter 4: *Vibrio cholerae* as a chassis to test directly DNA samples 128
4.1 – Creating the perfect chassis 128
4.1.1 – Superintegron deletion 128
4.1.2 – *V. cholerae*$_{\Delta SI}$ phenotypic characterization 130
4.1.3 – Characterization of *V. cholerae*$_{\Delta SI}$ natural competence 131
4.1.4 – Enhancing *V. cholerae*$_{\Delta SI}$ natural competence for integron recombination 133
4.1.5 – Testing recombination through natural transformation 135
4.2 – Integron cassette capture chassis 139
4.2.1 – *V. cholerae* setup 139
4.2.2 – Testing DNA samples 139

Discussion 147

Establishment of a platform to select for integron recombination events 147
The cassette harvester 149
A sentinel bacterium 151
Vibrio cholerae as a chassis to test directly DNA samples 152
Concluding remarks 154

Conclusions 159

Bibliography 163

Supplementary Information 187

Strains 187
Plasmids 191

Primers 193
Reagents 199
SeqDelTA 200
Supplementary figures 209

Figure index

Figure 1. Life expectancy 37
Figure 2. Bacterial resistance mechanisms to antimicrobials 42
Figure 3. Mobile genetic elements and forms of DNA exchange 44
Figure 4. Organization of integrons 49
Figure 5. Integron recombination sites 51
Figure 6. *attI* × *attC* reaction 55
Figure 7. *attC* × *attC* reaction 56
Figure 8. *attI* × *attI* reaction 57
Figure 9. Organization of the 19 toxin-antitoxin cassettes in the *Vibrio cholerae* N16961 superintegron 67
Figure 10. Regulatory network of natural competence in *Vibrio cholerae* 69
Figure 11. Killing assay 82
Figure 12. Classical recombination assay 84
Figure 13. Integron cassette capture platform 96
Figure 14. Characterization of constitutive promoters in *E. coli* and *V. cholerae* 98
Figure 15. Characterization of the P_{BAD} promoter in *E. coli* and *V. cholerae* 99
Figure 16. Characterization of the P_{VanM} promoter in *E. coli* and *V. cholerae* 100
Figure 17. Characterization of the P_{TetM} promoter in *E. coli* and *V. cholerae* 101
Figure 18. Characterization of the inducible promoters in *E. coli* through integron-mediated recombination 103
Figure 19. Predicted 3D structures of $CcdB_{wt}$ and CcdB::attI fusion protein 104
Figure 20. Plasmid setup and killing assay *ccdB::attI1* 106
Figure 21. Killing assays of *ccdB::attI1::cassette* 107
Figure 22. Classical recombination assay of P_{Lac} *ccdB::attI1* in a plasmid setup 108
Figure 23. Chromosome setup and killing assay of *ccdB::attI1* 109
Figure 24. Classical recombination assay of *ccdB::attI1* in a chromosome setup 110
Figure 25. Predicted 3D structures of $SacB_{wt}$ and SacB::*attI1* fusion proteins 112
Figure 26. Survivors in a killing assay testing the three SacB::*attI1* versions 113
Figure 27. Chromosome setup and killing assay *sacB::attI1* 114
Figure 28. Classical recombination assay of *sacB::attI1* in a chromosomal setup 115
Figure 29. Plasmid setup and cassette library generation 118

Figure 30. Library generation from *V. cholerae* N16961 superintegron 119
Figure 31. Library coverage across the chromosomes of *V. cholerae* N16961 120
Figure 32. Analysis of superintegron representation in the plasmid library 122
Figure 33. Correlations between cassette representation and cassette features 123
Figure 34. *E. coli* as a sentinel bacterium to detect integrons in mobile plasmids 125
Figure 35. Integron detection by the sentinel *E. coli* 126
Figure 36. SeqDelTA 129
Figure 37. Growth curves and competitions of *V. cholerae*$_{\Delta SI}$ vs *V. cholerae*$_{WT}$ 131
Figure 38. Classical natural competence test of *V. cholerae* N16961$_{\Delta SI}$ 133
Figure 39. Classical natural competence test of *V. cholerae* N16961$_{\Delta SI}$ generated mutants 134
Figure 40. Chitin-independent natural transformation assay for integron recombination ... 136
Figure 41. Chitin-independent natural transformation assay for integron recombination of *V. cholerae* N16961$_{\Delta SI}$ generated mutants 137
Figure 42. Classical natural transformation assay for integron recombination of selected *V. cholerae*$_{\Delta SI}$ mutants 138
Figure 43. Integron cassette capture chassis 139
Figure 44. Chitin-independent natural transformation assay for integron recombination of the integron cassette capture chassis 140

Supplementary Figure 1. Chosen locations to introduce the *attI1* site in the counter-selective markers 209
Supplementary Figure 2. Initial *ccdB::attI1* tests 210
Supplementary Figure 3. Nanoplot report on read length and quality from ΔSI library sequencing 211
Supplementary Figure 4. Analysis of reads that mapped to 0 cassettes 212

Table index

Supplementary Table 1. Strains used and generated in this thesis 187
Supplementary Table 2. Plasmids used and generated in this thesis 191
Supplementary Table 3. Primers used in this thesis 193
Supplementary Table 4. Antibiotics and reagents used in this thesis 199
Supplementary Table 5. Strains generated with SeqDelTA 200
Supplementary Table 6. Strains generated with pMP7 201
Supplementary Table 7. Primers used for SeqDelTA 201
Supplementary Table 8. PCR and sanger sequencing of the SI captured cassettes in the cassette harvester and the integron cassette capture chassis 203

List of abbreviations

aHJ: atypical Holliday junction
AR: antimicrobial resistance
bp: base pairs
bs: bottom strand
CALIN: clusters of *attC* sites lacking integron-integrases
carb^R: carbenicillin-resistant
CDS: coding sequence
CFU: colony forming unit
cm^R: chloramphenicol-resistant
CSM: counter-selectable marker
DAP: diaminopimelic acid
DASW: defined artificial seawater
°C: degrees Celsius
DNA: desoxyribonucleic acid
DR: direct repeats
dsDNA: double-strand DNA
ΔSI: superintegron deleted
dT: thymidine
EHB: extra-helical bases
***gcu*:** gene cassette of unknown function
gDNA: genomic DNA
GFP: green fluorescent protein
HGT: horizontal gene transfer
HJ: Holliday junction
IC: integron cassette
I3C: Integron Cassette Capture Chassis
ICE: integrative conjugative element
IPTG: isopropyl β-d-1-thiogalactopyranoside
IS: insertion sequence
kana^R: kanamycin-resistant
kb: kilobases
KO: knocked-out
LB: lysogeny broth
MDR: multi-dug resistant
mg: milligram
MGE: mobile genetic element
MI: mobile integron
µg: microgram
µL: microliter
µM: micromolar
mL: milliliter
nm: nanometer
nt: nucleotide
OD_{600}: optical density at 600 nm
ORF: open reading frame
PCR: polymerase chain reaction
RNA: ribonucleic acid
rpm: rotations per minute
SALIN: single *attC* sites lacking integron-integrases
SCI: sedentary chromosomal integron
SD: standard deviation
SI: superintegron
SOE-PCR: splicing by overlap extension polymerase chain reaction
spec^R: spectinomycin-resistant
***spp.*:** species
SSB: single-strand DNA-binding protein
ssDNA: single-strand DNA
TA: toxin-antitoxin

tDNA: transformed DNA

Tn: transposon

ts: top strand

UCS: uncoupled central spacer

VTS: variable terminal structure

WGS: whole genome sequencing

WHO: World Health Organization

WT: wild-type

zeo^{R}: zeocin-resistant

Abstract

Monte Clérigo, 2020

Abstract

Integrons are genetic elements that capture, rearrange, and disseminate integron cassettes (IC), including those responsible for antimicrobial resistance (AR). These genetic platforms have significantly contributed to the rise and global spread of multidrug-resistant pathogens by recruiting more than 170 different AR genes. However, despite their importance, detecting integron cassettes remains challenging. Conventional approaches, such as PCR, often introduce biases, while more advanced techniques, like deep sequencing, are not yet practical for routine use in laboratories due to their cost and complexity.

This thesis presents the development of a novel biotechnological tool to detect integron cassettes in a sequence-independent manner, a breakthrough that aims to overcome the limitations of traditional methods. The tool is based on re-engineering a class 1 integron to act as a capture platform by **embedding an integron integration site (*attI1*) inside a counter-selectable marker (CSM)**. We have tested two CSMs: the ***ccdB*** toxin, part of a type II toxin-antitoxin system, and ***sacB***, a gene encoding levansucrase, which, in the presence of sucrose, degrades it into levan, a toxic compound for bacteria. The underlying rationale is that when a cassette is inserted into the *attI1* site, it disrupts the CSM, allowing bacterial cells to survive under otherwise lethal conditions. Thus, survival serves as a readout for successful cassette capture. Prior to selection, growth under survival conditions ensures bacterial viability.

During the development phase, we tested several tool versions, optimizing them to ensure that the generated CSM::*attI1* fusion protein is functional. We succeeded and obtained a low escape mutant rate across the different setups. This broad detection range made the tool sensitive and versatile. As an initial validation, we demonstrated that both counter-selectable markers function effectively with a classical recombination assay, reliably reporting cassette capture events independently of their phenotype. We then expanded the application of this platform into three distinct and innovative uses.

We first developed a plasmid-based tool—the **cassette harvester**—to create cassette libraries by capturing cassettes from large sedentary chromosomal integrons

(SCIs). We tested this tool on the *Vibrio cholerae* superintegron (SI) and, by transforming and inducing the plasmid-based tool, we obtained approximately 10^5 colonies, each representing an individual capture event from the SI. The resulting cassette library was sequenced and analyzed. Our analysis confirmed the high specificity of the method and a satisfactory representation of the cassettes. This demonstrates the tool's efficiency in capturing cassettes from SCIs and opens the possibility of studying integron cassettes of unknown functions using a high-throughput method.

Second, we implemented a chromosome-based version of the tool to create a **sentinel *Escherichia coli*** strain. This strain was designed to detect integron cassettes from naturally occurring mobile plasmids in a co-culture system, where *E. coli* acted as the recipient of plasmids from donor bacteria. The sentinel strain efficiently identified integrons and provided valuable insights into the cassette content of the tested plasmids, making it a powerful tool for monitoring the presence and composition of these genetic platforms in microbial communities.

Third, we developed the final tool, the **Integron Cassette Capture Chassis (I3C)**. We applied our system to the naturally competent bacterium *Vibrio cholerae* to capture cassettes directly from exogenous DNA samples. To eliminate interference from the superintegron, we deleted it from the strain's genome. Additionally, we optimized a natural transformation protocol to enhance integron recombination. By combining this genetically modified strain with a DNA enrichment step, we were able to detect cassettes from various DNA samples, showcasing the tool's potential for environmental and clinical applications.

Overall, the work performed in this thesis introduces a groundbreaking detection platform that unveils the integron content of complex samples with high specificity without relying on prior knowledge of sequence information. The versatility and robustness of this tool make it a promising candidate for diagnostic purposes and for advancing the study of integrons and the functions encoded in large chromosomal arrays.

Resumen

Los integrones son elementos genéticos capaces de captar, reorganizar y diseminar cassettes de integrones (IC), incluyendo ICs que confieren resistencia a antimicrobianos. Estas plataformas genéticas han contribuido significativamente al aumento y dispersión global de patógenos multirresistentes al reclutar más de 170 genes de resistencia frente a la mayoría de las familias de antibióticos. Sin embargo, a pesar de su importancia, la detección de cassettes de integrón sigue siendo un desafío. Los métodos convencionales, como la PCR, suelen introducir sesgos, mientras que las técnicas más avanzadas, como la secuenciación masiva, aún no están establecidas en el uso rutinario de los laboratorios de diagnóstico debido a su coste y complejidad.

Esta tesis presenta el desarrollo de una herramienta biotecnológica novedosa para detectar cassettes de integrón independientemente de su secuencia, una innovación que busca superar las limitaciones de los métodos tradicionales. La herramienta se basa en la modificación de un integrón de clase 1 para que actúe como una plataforma de captura, mediante la **inserción de un sitio de integración de integrón (*attI1*) dentro de un marcador de contraselección (CSM)**. Para ello, probamos dos CSMs: la toxina ***ccdB***, parte de un sistema toxina-antitoxina de tipo II, y ***sacB***, un gen que codifica la levansucrasa, una enzima que degrada la sacarosa en levan, un compuesto tóxico para las bacterias. Asimismo, cuando un cassette se inserta en el sitio ***attI1***, interrumpiendo el CSM, permite que las células bacterianas sobrevivan en condiciones letales. De esta forma, la supervivencia sirve como un indicador de la captura de ICs. Antes de la selección, el crecimiento bajo condiciones de supervivencia asegura la viabilidad bacteriana.

Durante la fase de desarrollo, probamos varias versiones de la herramienta, optimizándolas para asegurar que la fusión CSM::*attI1* generada fuera funcional. Obtuvimos una baja tasa de mutantes de escape en los diferentes sistemas probados, generando un amplio rango para la detección de ICs. Como validación inicial, demostramos la funcionalidad de ambos marcadores de contraselección con un ensayo de recombinación clásica, reportando los eventos de captura de cassettes

independientemente de su fenotipo. A continuación, extendimos la aplicación de esta plataforma a tres usos distintos e innovadores.

Primero, desarrollamos una herramienta basada en plásmidos—el **cassette harvester**—para crear librerías de ICs capturándolos de integrones cromosómicos sedentarios (SCIs). Probamos esta herramienta en el superintegrón (SI) de *Vibrio cholerae* y, al transformar e inducir la herramienta basada en plásmidos, obtuvimos aproximadamente 10^5 colonias, cada una representando un evento de captura individual del SI. Al secuenciar y analizar la librería generada, confirmamos la alta especificidad del método y una representación satisfactoria de los cassettes del SI. Esto demuestra la eficiencia de la herramienta en la captura de cassettes de SCIs y abre la posibilidad de estudiar ICs de funciones desconocidas utilizando un método de alto rendimiento.

En segundo lugar, implementamos una versión de la herramienta basada en el cromosoma para crear una cepa ***Escherichia coli* centinela**. Esta cepa fue diseñada para detectar ICs de plásmidos movilizables en un sistema de co-cultivo, donde *E. coli* actúa como receptor de plásmidos de bacterias donadoras. El *E. coli* centinela identificó eficazmente integrones en los plásmidos testados, y también permitió la caracterización del contenido de cassettes, convirtiéndose en una herramienta poderosa para monitorear la presencia y composición de estas plataformas genéticas en comunidades microbianas.

Finalmente, desarrollamos la herramienta final, el **Chasis de Captura de Cassettes de Integrón (I3C)**, aplicando nuestro sistema a la bacteria naturalmente competente *Vibrio cholerae* para capturar cassettes directamente de muestras de ADN exógeno. Para eliminar la interferencia del superintegrón, lo eliminamos del genoma de la cepa con un método innovador, el SeqDelTA. Además, optimizamos un protocolo de transformación natural para mejorar la recombinación de integrones. Al combinar esta cepa genéticamente modificada con un paso de enriquecimiento de ADN, pudimos detectar cassettes de varias muestras de ADN, mostrando el potencial de la herramienta para aplicaciones ambientales y clínicas.

El trabajo realizado en esta tesis introduce una plataforma de detección revolucionaria que revela, con alta especificidad, el contenido de integrones en

muestras complejas sin depender del entorno genético de los mismos. La versatilidad y robustez de esta herramienta la convierten en una candidata prometedora para fines de diagnóstico, pero también para avanzar en el estudio de integrones y descubrir nuevas funciones codificadas en grandes integrones cromosómicos.

MIGUEL SOUSA TAVARES

vinha de lado algum; mulatas de dentes impressionantemente
brancos rodando como piões sobre as ancas, e policias
que se guardam de ar feroz mas que viravam a cabeça à
passagem das mulatas.
Digo sorrio, olhando a cena. Convenci-me de que,
no mundo inteiro, não havia lugar algum para onde um
homem pudesse fugir em busca de liberdade. Tornado
território de odios, extremismos, guerras ou preparo
de guerras, o mundo parecia entregue a uma casta
de visionários dementes ou desumanos – Hitlers, Mus-
solinis, Francos, Estalines, Salazares, Getulios – cuja
ambição era pisar, mandar, humilhar, e cuja identidade
comum era o desprezo pela liberdade. Deleitavam-se em
discursos, berros, ameaças, comícios, paradas militares,
fardas, condecorações, e no terror que inspiravam a um
séquito de seguidores obsequiosos e temerosos ou a resi-
tentes anónimos, arrastados pelas prisões e pelos campos
de extermínio, silenciados, torturados, reduzidos à inha-
ção humanas. Por cada tirano humano, havia milhares, milhões,
da Pátria ou do genero humano, lembrara, fechados em
de vitimas. Inumeras Olgas Benários e muitos outros e
outras cujos nomes já ninguem lembrara, fechados em
celas minimas, alimentados com comida de cão, despoja-
dos do espaço, da luz, da paz, reduzidos ao mais miserá-
vel da condição humana, para expiarem o mais inolerável
dos crimes: terem ousado pensar diferente, terem divi-
do da sabedoria, da justeza, da magnificência desses
omnividentes. Sim, o Brasil não era diferente. Havia
em cada esquina, vigiando tudo e mais alguma
simples gesto de caminhar na rua, de
to. Os jornais estavam cheios de

RIO DAS FLORES

noticias que não pronunciavam nada de bom, os generais
multiplicavam as ameaças, julgando que a farda era sinal
de razão e de direito, símbolo da Pátria e dos valores,
e o enviado do Destino, Getúlio Vargas, meditava todos
os dias na inevitabilidade de um golpe constitucional ou
militar que poupasse o Brasil à tragédia de ficar órfão
da sua protecção. Sim, o Brasil à tragédia de ficar órfão
grandes tiranos ou tiranetes se equivaliam na substância
das suas justificações, na cobardia das suas defesas. Mas
havia uma diferença, contudo, como se as ditaduras se
tornassem ridículas nos trópicos; o povo era demasiado
alegre, o mar era demasiado envolvente, a comida não era
suave, a cachaça não era inocente, nenhuma ordem apagaria o cheiro a
guita calar à musica, nenhuma ordem apagaria o cheiro a
verde da mata e os gritos dos papagaios, os policias conti-
nuariam sempre a virar a cabeça à passagem das mulatas,
e, por mais que gritassem, que mandassem, que ameaçassem,
ninguem, jamais, conseguiria calar para sempre
aquela absurda alegria, aquela devastadora liberdade
ele sentia ali.
Eancou as pernas e suspirou profundamente.
Sol descera já no horizonte, demasiado cedo
tece no inverno dos trópicos. A luz tremeluzia
nos candeeiros da Avenida Atlântica, bruxuleando
vencer a escuridão que rapidamente
que era natural, era brasileiros, a luz
estrangeira, como tudo o resto
nha inglesa Light. Lá
rmavam-se mais
ados, e o ruid
as ruas

Introduction

Galicia, 2021

Introduction

1 – Antimicrobial Resistance

1.1 – Historical perspectives and definitions

An antimicrobial is any substance that can kill microorganisms or stop their activity. It can be of natural origin or synthetic. Antimicrobials have existed long before humans, shaping the evolution and establishment of microbial communities in nature (Aminov & Mackie, 2007). Since the discovery of penicillin by Alexander Fleming in 1928 (Fleming, 1929), we have manufactured them to satisfy our needs. Undeniably, antibiotics have revolutionized healthcare in a world where infectious diseases were the primary cause of death. They have played a crucial role in the increase of life expectancy (from less than 50 years in the 19th century to more than 80 years in developed countries nowadays) (**Figure 1**), by enabling, together with other relevant discoveries, the development of advanced treatment techniques, such as preventive surgeries, cancer chemotherapy, organ transplantation, among many others. As an example, *Vibrio cholerae*-associated mortality can rise to 50% if not treated with antibiotics (Sack et al., 2004) (as throughout the first pandemic), and can be as low as 1% when the appropriate care is available (European Centre for Disease Prevention and Control data).

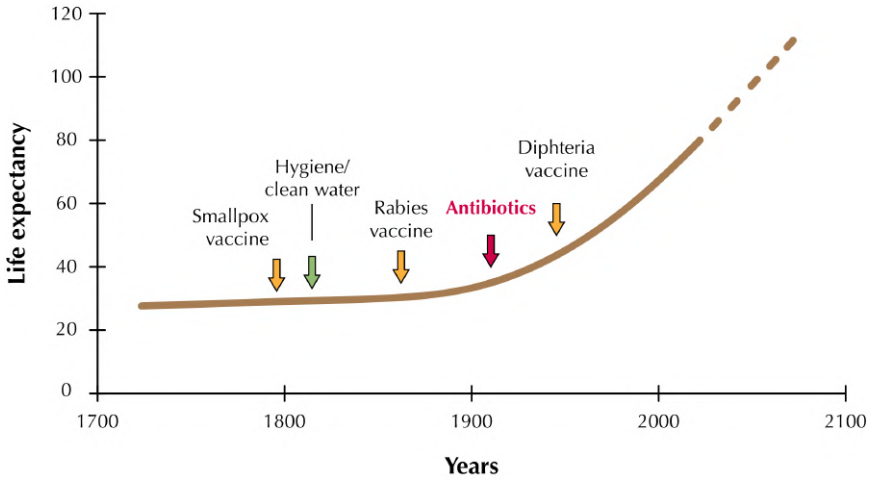


Figure 1. Life expectancy

Schematic representation of the life expectancy from 1700 to 2100. Five significant events that revolutionized modern medicine are represented in arrows, including the discovery of antibiotics. Data source: https://ourworldindata.org/life-expectancy

The decades that followed the discovery of penicillin were crucial in the breakthrough of new antimicrobial compounds and families. Although some resistance was observed, multi-drug resistance (MDR) was not an expected phenomenon in such a short period of time since the stepwise appearance of multiple mutations was considered to be beyond the potential of any bacterial population, especially due to the cost it could entail. In the following years, the consumption of antibiotics grew in both human and veterinary care, to the point of its usage as growth promoters in animals (Van Boeckel et al., 2017). In sum, mortality caused by infectious diseases was thought to be a problem surmounted.

Yet, bacteria proved to have surprising new means of adapting and acquiring multiple resistances to antimicrobials. For instance, the horizontal transfer of genetic elements, such as integrons, transposons, and plasmids—sources of extreme evolvability—was not anticipated at the beginning of the antibiotic era (MacLean & San Millan, 2019). Besides, the entry into the new millennium was characterized by a tremendous void in the discovery of new antimicrobial compounds.

The World Health Organization (WHO) defined antimicrobial resistance (AR) as a microorganism's ability to resist a compound that was once able to kill the pathogen in question. In the XXIth century, AR has been extensively reported in all countries in the world, and it has entered the top priorities of the World Health Organization. In 2019, AR was associated with five million deaths (Murray et al., 2022), and experts believe that this number can rise to 10 million in 2050, becoming the first cause of death in the world (O'Neill, 2016).

Antimicrobial resistance is a multifaceted situation that affects all of society; indeed, the economic burden of controlling MDR bacteria is already worrisome, with increasing losses due to longer hospitalization times and higher mortality rates. In low- and medium-income countries, the problem is even more severe. The inadequate access to antibiotics among people with low incomes, coupled with the overconsumption among the middle classes, and a lack of separation between prescribing and provider entities, could lead to a catastrophic outcome. In these countries, resistance is predicted to increase 4-7 times faster (Hofer, 2019; Laxminarayan et al., 2016).

Antimicrobial resistance is now well-studied. It is known that the strength of selective pressure—dependent on antibiotic usage—is directly related to the appearance and increase of resistance. Actually, the majority of international organizations' healthcare policies are focused on awareness and responsible intake of these critical compounds. Such is the case of the World Antibiotic Awareness Week, an initiative from the WHO that began in 2015.

However, it is also acknowledged that this problem requires much more than awareness and consumption measures. Researchers have focused their efforts on tackling antibiotic resistance from an epidemiological point of view, but also from a mechanistic and functional perspective in *in vitro* models. Still, there is a significant gap in the communication and interconnections between these two approaches (MacLean & San Millan, 2019). In the light of these facts, it is clear that single, isolated interventions have limited impact. Coordinated actions are required to minimize the emergence and spread of antimicrobial resistance (Calvo-Villamañán et al., 2022; Hernando-Amado et al., 2019). In September 2016, AR became the fourth health issue—after HIV, noncommunicable diseases, and Ebola—to be discussed by the United Nations General Assembly (Laxminarayan et al., 2016). There is growing pressure to create a global action plan, such as the Paris Agreement for climate change, to keep antimicrobial resistance under control.

1.2 – Resistance mechanisms

Bacteria have achieved countless ways to escape the activity of antimicrobials. Inevitably, as the frequency of resistance increases, new mechanisms and new genes are discovered.

Resistance can be classified as intrinsic or acquired (Blair et al., 2015). The intrinsic resistance of a bacterial species to a particular antimicrobial is a consequence of inherent or functional characteristics (or a lack of them). The simplest example of intrinsic resistance results from the absence of a specific antibiotic target. For instance, the biocide triclosan has extensive efficiency against Gram-positive bacteria and plenty of Gram-negative bacteria, but it is unable to inhibit the growth of *Pseudomonas aeruginosa*. Indeed, this species possesses an insensitive allele of *fabI*, the target of

triclosan (Zhu et al., 2010). By contrast, acquired resistance implies the gain or the change of a function. Then, depending on the target and action of the antimicrobial, several means can be responsible for the lack of activity of an antimicrobial (**Figure 2**).

A. **Reduced permeability:** Gram-negative bacteria are naturally less permeable than Gram-positive bacteria; nevertheless, some hydrophilic antimicrobials are able to surpass the thick membrane of these microorganisms by using their porins. In *Escherichia coli*, OmpF and OmpC are the major unspecific porins. A typical resistance mechanism is the reduction of the synthesis of these porins and a shift to more specific channels that limit the entry of antimicrobials (Blair et al., 2015).

B. **Efflux:** efflux pumps are significant contributors to the intrinsic resistance of bacteria. Overexpressing some of these pumps, or acquiring genes encoding such proteins, can lead to very high resistance rates, especially in Gram-negatives. Some pumps can only accommodate a few substrates, but many have a broad substrate range, being called multi-drug resistance efflux pumps (Blanco et al., 2016).

C. **Drug inactivation:** bacteria can modify or degrade the antimicrobial and thus prevent its action. One way to achieve this is by hydrolyzing the compound. Since the discovery of penicillin, β-lactamases (enzymes that hydrolyze β-lactams such as penicillin) have been a major cause of antimicrobial resistance (Abraham & Chain, 1940). Thousands of these enzymes are known, targeting all antibiotics in this class, and some of them have a large spectrum, conferring resistance to a high number of different compounds within the β-lactams, dubbed extended-spectrum β-lactamases (ESBL) (Naas et al., 2017).

   Another common mechanism of drug inactivation is the addition of a chemical group to the antimicrobial. Aminoglycosides are particularly susceptible to this (Llano-Sotelo et al., 2002), as demonstrated by the high levels of resistance conferred by aminoglycoside modifying-enzymes. These

enzymes can be classified in three main groups, according to the group that they transfer: acetyltransferases, phosphotransferases, and nucleotidyl/adenyltransferases (Vakulenko & Mobashery, 2003).

D. **Target protection:** an alternative mechanism that allows bacteria to evade the activity of antimicrobials is the protection of the drug target trough binding of a protein or the addition of a chemical group. Proteins such as those from the Erm family protect the ribosome from the attachment of macrolides and lincosamides, through methylation of the targeted region (Leclercq, 2002). Qnr proteins mimic DNA structure, and so, they interfere allosterically with the binding of fluoroquinolones, conferring protection to their targets (Tran et al., 2005).

E. **Target modification:** point mutations in the target can also alter the affinity of some compounds. Mutations in *gyrA* and *parC* genes dramatically increase resistance to fluoroquinolones. As another example, *S. pneumoniae* and *S. aureus* can acquire resistance to linezolid by mutating one of the copies of the 23S rRNA ribosomal subunit, followed by recombination at high frequency between homologous alleles, which rapidly leads to a resistant population (Billal et al., 2011; Gao et al., 2010).

F. **Expression changes:** many researchers identify antimicrobial targets by searching for amplified or overexpressed genes, which leads to the natural conclusion that these phenomena lead to a status of resistance. Indeed, this is the case for numerous drugs, such as the increase of folate synthesis, the target of trimethoprim (Palmer & Kishony, 2014).

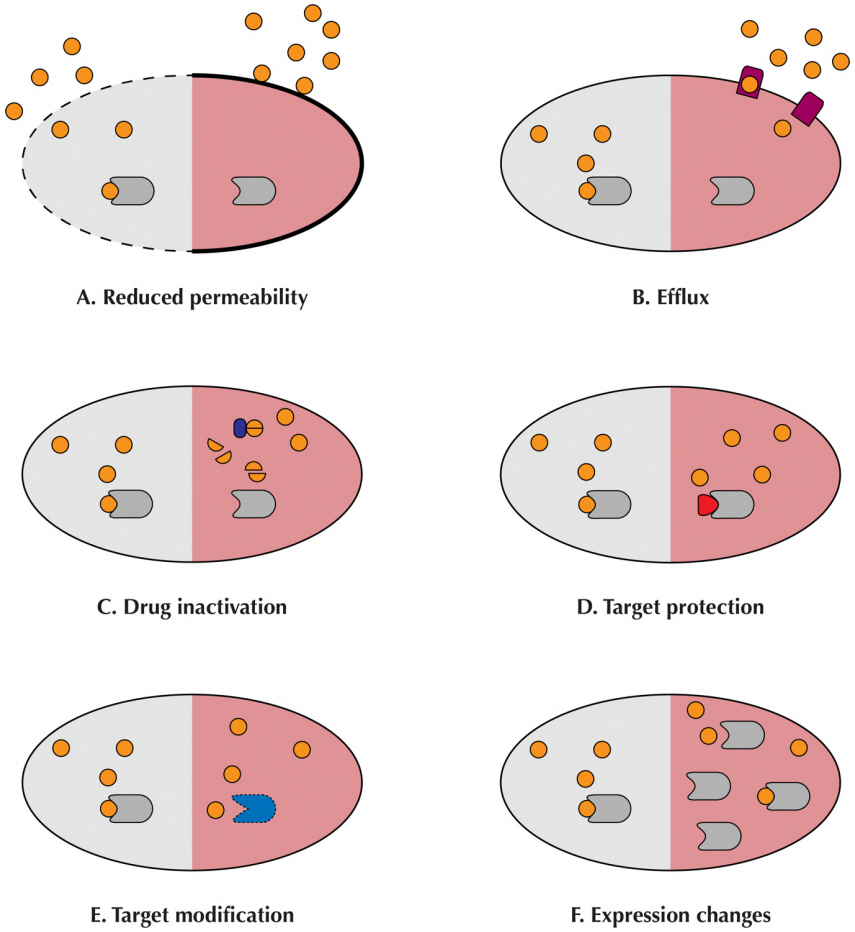


Figure 2. Bacterial resistance mechanisms to antimicrobials

Schematic representation of six major resistance mechanisms. Each half cell represents sensibility (grey) and resistance (pink). Antibiotics are depicted by orange circles and targets are represented by a grey rod shape. **A.** Reduced permeability is intrinsic to Gram-negative bacteria, but some changes to the membrane can also occur, such as the downregulation of porins. **B.** Efflux is usually mediated by pumps that export antibiotics out of the cell. **C.** Drug inactivation is usually mediated by enzymes that degrade or modify the antibiotic, such as β-lactamases or aminoglycoside modifying enzymes. **D.** Target protection occurs when a protein that protects the target binds to the target site, preventing the antibiotic from attaching. **E.** Target modification typically occurs when a mutation or an enzyme modifies the target of the antibiotic thus preventing its binding. **F.** Expression changes of a target can also neutralize or decrease the effect of an antibiotic.

1.3 – Mobile genetic elements

Following Darwin's principles on evolution, it was assumed that mutations led to the appearance of new phenotypes and, if the phenotype conferred an advantage to the host—as in the case of antimicrobial resistance, in the presence of selective pressure—the mutation was preserved. This model suggested that antimicrobial resistance was possible, but multi-drug resistance was very unlikely to appear in such a short amount of time due to its high cost. However, as history has shown, bacteria have surprised us by demonstrating a superior layer of control of evolution: horizontal gene transfer (HGT). Rather than mutation accumulation, the acquisition of new genes became the leading source of antimicrobial resistance, and mobile genetic elements

(MGE) have significantly contributed to the rapid appearance of MDR strains (Partridge et al., 2018).

1.3.1 – Forms of DNA exchange

MGEs are segments of DNA encoding genes (whether for antimicrobial resistance or other purposes) and specific proteins that allow them to move within genomes—**intracellular movement**—, or between bacterial cells—**intercellular movement** (Frost et al., 2005) (**Figure 3**).

The first intercellular mechanism of DNA exchange to be discovered was **transformation** (Griffith, 1928). This type of genetic material movement is the natural capacity that some bacterial species have developed to acquire external DNA from the outside of the cell via a dedicated uptake apparatus. Such ability can serve different purposes, such as using DNA for nutrients, helping in the genome repair, or being acquired by the chromosome through homologous recombination (Dubnau, 1999). **Conjugation** consists of the transfer of genetic material by direct contact between donor and recipient cells, where the donor cell synthesizes a multiprotein apparatus, such as a pilus. It involves a conjugative element, either self-replicating such as conjugative plasmids, or located in the chromosome, as integrative conjugative elements (ICE) (Wozniak & Waldor, 2010). Conjugation also entails proteins that enable the transfer, which are usually encoded within the MGE. This process can be very promiscuous, even allowing trans-kingdom DNA exchange (Heinemann & Sprague, 1989). New findings suggest alternative routes for plasmid dissemination beyond traditional methods. Certain plasmids have been observed to employ specialized vesicles for their transfer, a phenomenon termed **vesiduction**, believed to be particularly efficient within biofilm communities (Abe et al., 2020). Moreover, **transduction** is the form of intercellular DNA transfer mediated by independent bacterial viruses, called bacteriophages. Its primary purpose is the transfer of the DNA or RNA of the phage, although some large segments of the host DNA can also be packed and transmitted to the following infected cells (Chiang et al., 2019). Finally, new "non-canonical" mechanisms of intercellular DNA exchange have been described, such as nanotubes and GTAs (Arnold et al., 2022), further emphasizing the complexity of intercellular DNA exchange.

Intracellular movement can take several forms: homologous recombination, transposons exchanged between different plasmids or the chromosome, and even integron rearrangements by the acquisition of cassettes contained in other integrons. In sum, both intercellular and intracellular movements are dominated by the exchange of MGEs.

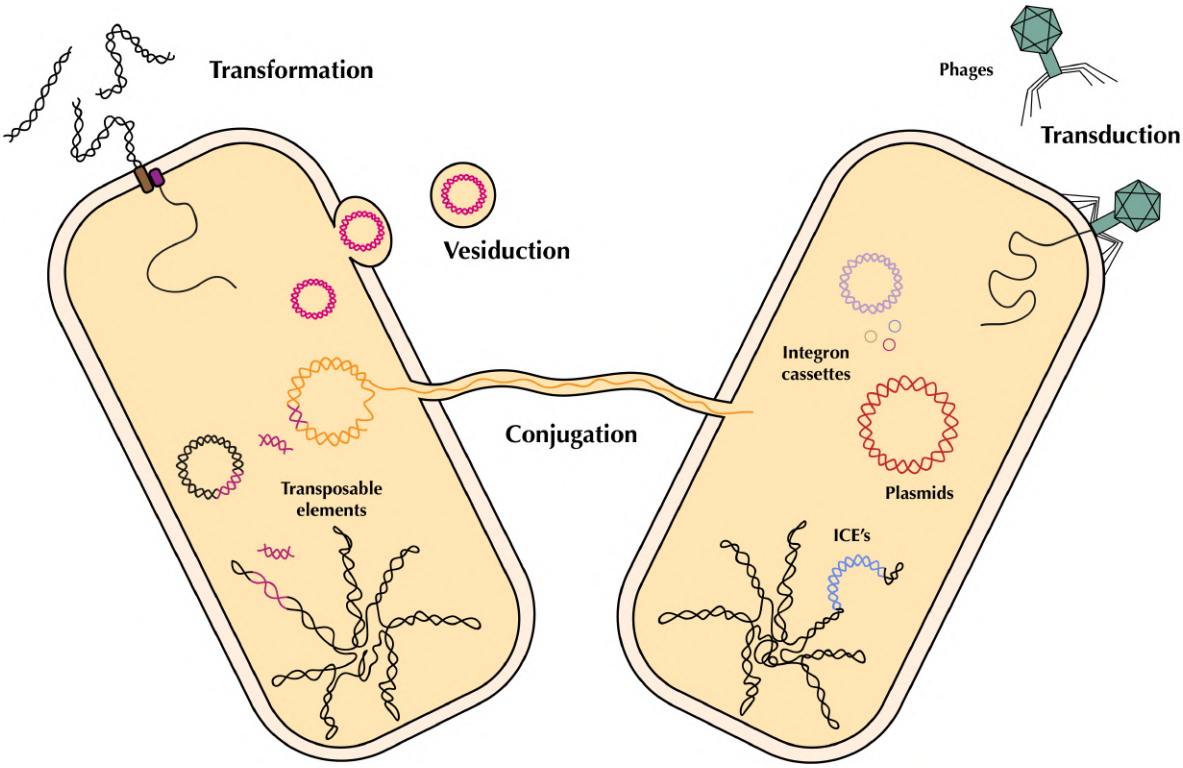


Figure 3. Mobile genetic elements and forms of DNA exchange

Schematic representation of MGEs and HGT between two bacterial cells. The four major forms of DNA exchange are represented. On the left, transformation of exogenous DNA and the uptake pilus; further down, vesicles being formed and carrying plasmids outside the cell; in the center, the connection between the two cells represents a pilus facilitating conjugation, a direct transfer of DNA material, which often includes plasmids; and on the left transduction, that allows the entrance and exit of DNA mediated by phages. Mobile genetic elements are also represented: plasmids in both bacterial cells, transposable elements, capable of excising and integrating into the genome and plasmids, are shown as DNA segments; integron cassettes, are exhibited within the right cell; phages carrying their own genetic material and potentially bacterial DNA also in the cell; and finally, an ICE being excised from the chromosome in the bottom right cell.

1.3.2 – Types of elements

Several MGEs are known and well characterized. **Plasmids** are self-replicative molecules of double-stranded DNA, usually covalently closed and circular (Frost et al., 2005), though linear plasmids have been found in an increasing number of species (Hawkey et al., 2022; Hinnebusch & Tilly, 1993). Generally, they encode core genes in a so-called “backbone”, where conjugation genes can be present. They can also encode a variable collection of accessory genes that are considered non-essential to plasmids. These DNA molecules must maintain a tight regulation of their replication, copy

number, and ensure the inheritance to the daughter-cells, in a process called partitioning. It is known that plasmids with identical replication mechanisms cannot stably co-exist in a given cell, which gave light to their initial classification: the incompatibility (Inc) system (Carattoli et al., 2005), still used nowadays. In addition, plasmids can also be categorized on the basis of mobility proteins, of their genetic similarity, or even by their size (Redondo-Salvo et al., 2021; Rodríguez-Beltrán et al., 2021). While some plasmids are only maintained in a few, closely related species, others are easily maintained and replicated in quite diverse genera (Jain & Srivastava, 2013; Redondo-Salvo et al., 2020).

Plasmids have been identified in a significant number of species, from Gram-negatives to Gram-positives, having an enormous contribution to antimicrobial resistance dissemination (Rodríguez-Beltrán et al., 2021). In clinical isolates, they can be substantially large, and associated with many other MGEs, such as insertion sequences, transposons, or integrons. By contrast, those found in environmental isolates are usually small and cryptic as they mainly encode genes of unknown functions (see for example references Attéré et al., 2017; Fogarty et al., 2024). Nonetheless, the importance of small plasmids in the dissemination of AR has also been demonstrated (Ares-Arroyo et al., 2018). Another interesting feature of conjugative plasmids is that they can also be responsible for the transfer of chromosomal DNA as well as other nonconjugative plasmids (Reimmann & Hass, 1993).

In recent years, it has also become clear that conjugation could also be achieved by chromosomally encoded elements: **integrative and conjugative elements (ICE)**. These elements contain regulatory systems, as well as genes that allow conjugation, and their integration/excision within the chromosome. After their excision, they are replicated by the cell machinery (their main difference to plasmids), and conjugated to a new cell where they can be integrated in another chromosome. An integrase mediates the integration and excision reactions, and determines the frequency as well as the site of integration (Wozniak & Waldor, 2010). ICEs can encode antimicrobial resistance genes and have proven to be of substantial clinical relevance (Waldor et al., 1996; Wozniak et al., 2009).

Transposable elements are also one of the main players in horizontal gene transfer. The simplest transposable elements, insertion sequences (IS), are small mobile elements that usually carry little more than the transposase gene necessary to its mobilization. The edges of the IS typically reveal inverted repeats that are named IR_L (left) and IR_R (right). The transposition of the IS involves the binding of the transposase to the inverted repeats, a set of DNA cleavages, and strand transfer reactions that move these ends into a target DNA molecule. This often generates target site duplications on insertion. However, some IS may not contain these features, according to its type of transposition mechanism and enzyme family (Partridge et al., 2018; Siguier et al., 2014).

ISs were not thought to be involved in the dissemination of genes; nonetheless, the discovery of composite transposons came to undermine this assumption. Indeed, composite transposons can carry "passenger" genes, flanked by the same or two related IS, that move as a single unit. Many of these elements contain a strong promoter that ensures the expression of the contained genes (Kamruzzaman et al., 2015; Mahillon & Chandler, 1998). Unit transposons differ from composite transposons as they are not flanked by IS but by inverted repeats. They can carry several genes as well as a transposase that ensures its mobility (Partridge et al., 2018).

Bacteriophages are viruses that infect bacteria, and they are known to be the most abundant biological form on earth due to their high prevalence and their rapid replication (Clokie et al., 2011). Some of these elements use a recombinase to introduce their genetic material inside the chromosome, either DNA or retro-transcribed RNA, and they take over the cell machinery to replicate their genes, allowing the formation of new virus particles that leave the host-cell (Penadés et al., 2015).

Transduction happens when chromosomal DNA is packed inside the virus capsid and transmitted to another bacteria. Interestingly, some other MGEs have manipulated phages to hijack their packaging system and ensure their own transfer (Fillol-Salom et al., 2019). Additionally, phages can recombine between them, as well as with other MGEs, rendering transduction as one of the most effective mechanisms of horizontal gene transfer.

Integrons are ancient genetic platforms that allow the acquisition, stockpiling, and reshuffling of **integron cassettes** (Escudero et al., 2015). We will review them in-depth in the following chapter. Mobile integrons are tightly associated with transposons and plasmids, showing the complexity of mobile genetic elements and the powerful means that bacteria have gathered to ensure their own survival.

2 – Integrons

Bacteria have evolved a myriad of tools which allow them to rapidly adapt to challenging conditions, including the presence of antimicrobials. The integron, an ancient structure that has shaped their evolution for hundreds of millions of years (Rowe-Magnus et al., 2001), is one of the most rousing examples of these tools.

Integrons are genetic platforms that allow the **acquisition**, **stockpiling**, **excision**, and **re-ordering** of open reading frames (ORF) found in **integron cassettes** (**IC**) (Escudero et al., 2015; Mazel, 2006). These platforms have found their way to the clinical environment through transposons and conjugative plasmids (Gillings et al., 2008), having played an unquestionable role in the establishment of antimicrobial resistance 60 years ago (Mitsuhashi et al., 1961), which further led to their discovery in the late 1980s (Stokes & Hall, 1989). Nowadays, integrons have become commonplace among Gram-negative pathogenic bacteria.

2.1 – A unique recombination system

An integron is composed of a stable platform, that allows the capture of genes, and a variable array, formed by the arrangement of integron cassettes (Escudero et al., 2015) (**Figure 4**).

2.1.1 – Stable platform

All integrons contain a stable platform composed of three key elements:

- the **integrase gene**, encoding the DNA recombinase that governs all recombination reactions (***intI***);
- the **integration site** (***attI***) where new integron cassettes are incorporated;
- and a **cassette promoter** ($\mathbf{P_c}$) that ensures the expression of integron cassettes.

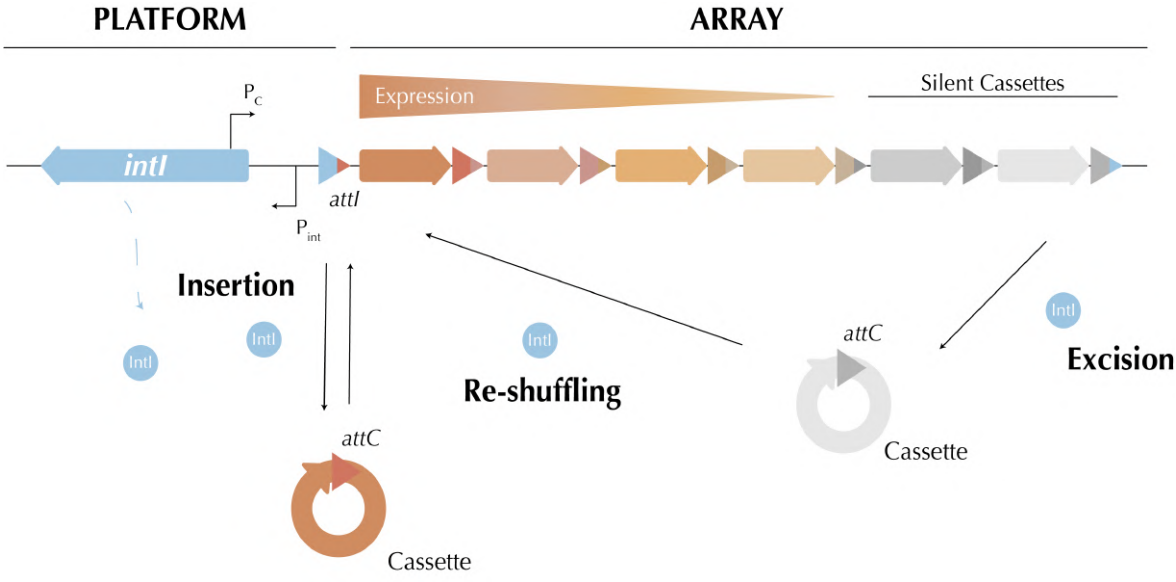


Figure 4. Organization of integrons

Schematic representation of an integron. The functional platform is depicted on the left, composed by the integrase-encoding gene (*intI1*), the cassette promoter (P_C), the integrase promoters (P_{int}), and the integron insertion site (*attI*). The variable array of integron casettes is represented on the right. Hybrid *attI* and *attC* sites are indicated with corresponding colors. Arrows inside the cassettes indicate the direction of the open reading frame. Their expression level is reflected by the color intensity of each arrow. The two main reactions catalyzed by the IntI that lead to cassette insertion (*attC* × *attI*), excision (*attC* × *attC*) and reshuffling are also represented.

Integron integrases belong to the family of Y-recombinases (Nunes-Düby et al., 1998), forming a specific cluster within it. These enzymes mediate recombination between specific DNA sites, recognized by sequence, and allow a variety of biological functions. The catalytic domain of this family is highly conserved: indeed, integron integrases possess, as all Y-recombinases, the patches I and II, boxes I and II, and the active residues RKHRHY (Grindley et al., 2006; Nunes-Düby et al., 1998). The main difference among integron integrases (hereafter integrases) is a 19 amino-acid long extra-domain, called the I2, that forms an α-helix (Messier & Roy, 2001). This domain enables integrases to recognize specific structural features (instead of a particular sequence for other Y-recombinases) of the cassette attachment sites and ensures that ssDNA recombination is possible (Demarre et al., 2007; MacDonald et al., 2006). In all circumstances, a single Y-recombinase monomer can only cleave one strand, with four monomers being necessary to perform the recombination reaction (Grindley et al., 2006).

The continuous expression of the integrase can lead to random recombination events with deleterious effects for the cell and induce premature death (Harms et al., 2013; Starikova et al., 2012), suggesting that this particular Y-recombinase must be subjected to tight regulation. Indeed, a careful analysis of the upstream region of the

intI gene reveals a promoter (P_{int}) overlapped in its -10 box by a LexA binding motif (Guerin et al., 2009). LexA is the master repressor of the SOS response, a universal regulatory system that allows repairing or bypass of DNA damage, and so, in normal conditions *intI* is not expressed. On the contrary, *intI* expression is increased by 4.5-fold in the class 1 integron and by 37-fold in the *Vibrio cholerae* superintegron when LexA is inactivated (Cambray et al., 2011).

New genes enter the integron at a specific site called the **integron attachment site** (***attI***). This region is composed of two integrase binding sites, the R and L sites, forming its core. Generally, they are hard to identify: the L domain is always degenerate compared to the R site, and the central region can change significantly among *attI* sites. Recombination takes place in the R box, within the 5'-GTT-3' conserved triplet, and cleavage occurs between the A and C of the complementary strand (Gravel et al., 1998) (**Figure 5A**).

IntI1 and *attI1* are, respectively, the specific integrase and integration site of the class 1 integron (the most studied and well-known, reviewed in-depth later in this chapter). In this class, four integrase monomers bind to the integron recombination site: two within the R and L boxes, as mentioned, and two on imperfect direct repeats, named DR1 and DR2. These additional sequences do not seem to be a common feature of other *attI* sites; while they enhance recombination in class 1 integrons, they are not essential (Gravel et al., 1998; Partridge et al., 2000).

The reaction between same-class integron attachment sites and integrases (*attI*s and their IntIs) seems to be very specific. Although some activity is possible in some cross-recombination assays, its efficiency is 100 times lower (Biskri et al., 2005).

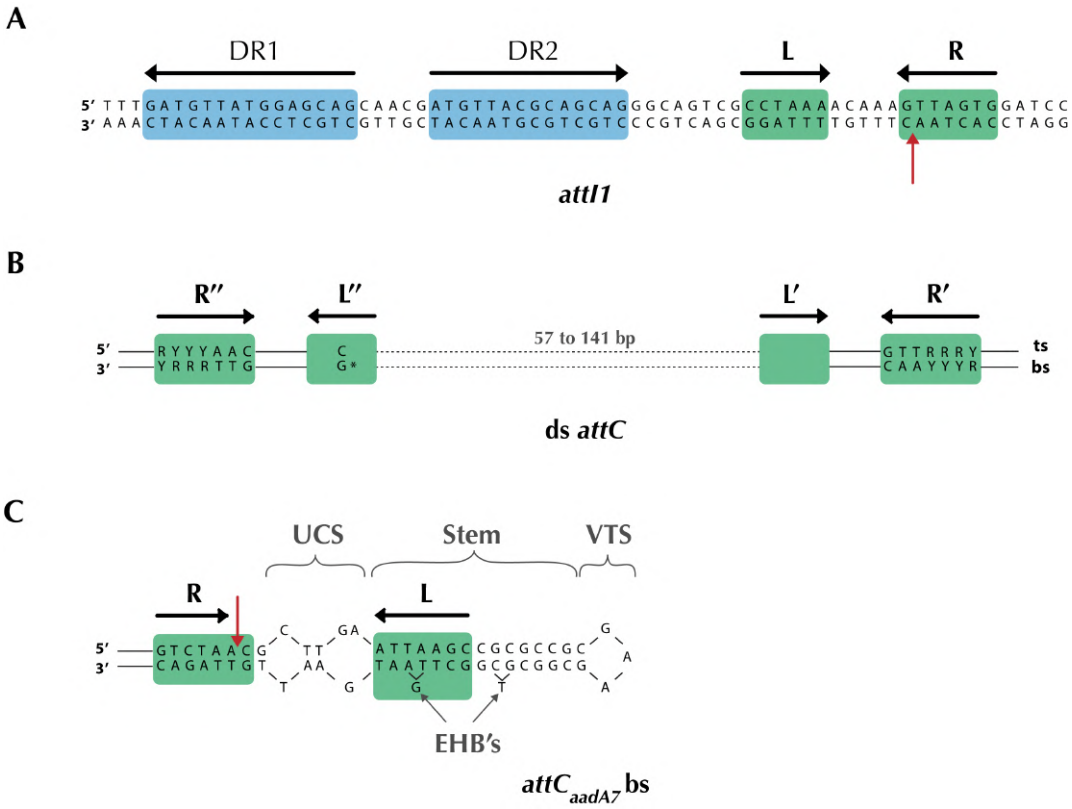


Figure 5. Integron recombination sites

Schematic representation of the two main integron recombination sites. Highlighted in green and blue are the domains potentially recognized by IntI1. Cleavage sites are denoted by red arrows. **A.** The *attI1* site structure is depicted, with its inverted (R and L) and direct (DR1 and DR2) repeats. **B.** Diagram of *attC* sites in double-stranded form; inverted repeats (R″, L″, L′, R′) are shown with gray arrows. Central variable regions are illustrated with dotted lines, while key conserved nucleotides are identified. Labels for the top (ts) and bottom strands (bs) are provided. **C.** Hypothetical secondary structure for the $attC_{aadA7}$. The structural details of *attC* sites are represented including the R and L boxes, the Unpaired Central Spacer (UCS), the Extra Helical Bases (EHBs), the stabilizing stem, and the Variable Terminal Structure (VTS).

Finally, the stable part of the integron platform possesses a resident **cassette promoter** (**P_c**), that ensures the expression of the integrated integron cassettes. In class 1 integrons, eight variants of the P_c are known, and their relative strength has been measured (Jové et al., 2010; Lévesque et al., 1994). Occasionally, the P_c promoter can be associated with a secondary promoter, located in the *attI* (Papagiannitsis et al., 2009).

Cassette expression in integrons is generally thought to be affected by their distance to the P_C promoter (**Figure 4**), as in any other operon-like structure. The first cassettes of the array would have maximal expression, and on the contrary, the ones found too far away from the P_C would be silent. However, the polar expression of cassettes from the resident promoter was recently challenged. Indeed, the translation efficiency of each cassette influences the expression of the genes that follow. On one hand, if a cassette is poorly translated, it can lead to lower expression of the

downstream cassettes. On the other hand, a higher translation rate can lead to an expression even above what was described for the first position of the array (Carvalho et al., 2024).

In class 1 integrons, P_{IntI} and P_C face each other in the integron array, being subjected to transcription interference. This translates to a trade-off between cassette expression and integrase expression (Guerin et al., 2009; Krin et al., 2014). Nonetheless, the higher prevalence of the weaker versions of the cassette promoter suggests that a flexible array is more adaptive than a higher expression of those genes (Moura et al., 2012; Vinué et al., 2011).

2.1.2 – Variable array

Integron cassettes, shortly called cassettes, usually consist of a single, promoterless, **open reading frame** followed by a **cassette recombination site** (***attC***) (Recchia & Hall, 1995). These are circular non-replicative elements that need to be within the integron to be functional. The serial integration of cassettes in the *attI* site leads to the formation of an array of adaptive functions, as integron cassettes can harbor genes for a variety of purposes, and the majority of the functions are yet to be discovered (Boucher et al., 2007).

The cassette attachment site (*attC*) is necessary for the integration of cassettes inside the integron and its unique features are essential for single-stranded DNA (ssDNA) recombination. The size of *attC* sites varies considerably among different cassettes, from 57 to 141 bp. Their conserved sequences consist of 4 regions of inverted homology, R''-L'' and L'-R' (**Figure 5B**), that allow the formation of a hairpin structure and the appearance of the canonical core binding sites R and L (Stokes et al., 1997). It is the folding of the ssDNA, and in particular, of the bottom-strand (bs), that forms the substrate recognized by the integrase (Bouvier et al., 2005; Francia et al., 1999) (**Figure 5C**). *attC* sites show almost no sequence conservation, limited to the inverted triplets 5'-AAC-3' and 5'-GTT-3' located in the R'' and R' boxes, respectively. Instead, they exhibit a remarkable palindromic organization that forms the secondary structure necessary for recombination.

The 3-dimensional structure of the *attC* sites ensures a new layer of information and regulation mainly detained by three structural features (Bouvier et al., 2009):

- The **extra-helical bases** (**EHB**s) arise from the R''-L'' arm and have no complementary nucleotides on the R'-L' arm. Depending on the *attC* sites, two or three EHBs can be present, usually within the L box. They are essential to recombination as they dictate which strand is going to be exchanged, they serve to stabilize the synaptic complex, and they avoid the second cut of the integrase by pushing the active tyrosine residue bound to the *attC* L box (MacDonald et al., 2006).
- The **unpaired central spacer** (**UCS**) between the R and L boxes results from this zone's lack of complementarity. This configuration is fundamental to attaining a high recombination rate (Nivina et al., 2016).
- Ultimately, the **variable terminal structure** (**VTS**) has a crucial function in the formation of the hairpin (Loot et al., 2010). It corresponds to the sequence located at the end of the stem, and its length can significantly vary between *attC* sites.

The tight regulation of hairpin formation and its structural features provides the tools for integrase recognition and recombination of cassettes with different attC sites. It also inserts the gene in its correct position, assuring its expression. Top-strand (ts) recombination would introduce the promoterless ORF in an antisense orientation (Nivina et al., 2016).

2.1.3 – External recombination sites

A new integron recombination site was recently described by Loot and collaborators (Loot et al., 2024). In this study, they demonstrate that ***attG*** sites (for integron genomic attachment sites) are capable of capturing cassettes from other integrons if an integrase is present. Interestingly, if inserted near a promoter, these cassettes can be expressed. The *attG* sites are widely distributed in bacterial genomes and exhibit some specific—and familiar—features, such as the conservation of a few bases (either the GTT or GAT triplet where the integrase cuts, as in an *attI*, site between the G-T/A). There is no evidence of a secondary structure formation, and its recombinogenic form is dsDNA. The existence of such sites raises the likelihood that,

rather than being traces of integrons, some small CALINs (clusters of *attC* sites lacking integron-integrases) and SALINs (single *attC* sites lacking integron-integrases) could result from the capture of integron cassettes in *attG* sites.

2.1.4 – Recombination reactions

The integron uniqueness comes from the fact that it mediates not only recombination of double-stranded DNA (dsDNA), but also, and more significantly, of ssDNA.

attI* × *attC: the *attI* × *attC* reaction achieves the integration of a cassette in the first position inside the integron. Despite recombining ssDNA with dsDNA, it is the most efficient reaction performed by the integrase, having a recombination rate of $\sim 10^{-2}$ in a classical recombination assay (Bouvier et al., 2005). The first cut performed happens within the R box of the *attC* site, separating the AA from the C in the 5'-AAC-3' triplet (Hall et al., 1991). As previously mentioned, the EHBs avoid the second cut by pulling the catalytic tyrosine apart from the phosphate link (MacDonald et al., 2006). This reaction forms an atypical Holliday Junction (aHJ), contrasting with the "regular" Holliday Junction (HJ) resolution of Y-recombinases (Grindley et al., 2006): the aHJ stops in the first strand exchange because the second one would lead to a linear product, abortive to the cell. It is replication that resolves this junction (Loot et al., 2012), delivering two products: the original platform, and the new integron that has acquired a new cassette. This illustrates the potential of adaptation of the integron: as recombination is semi-conservative, even if the new gene is toxic or too costly, the host can always survive (**Figure 6**).

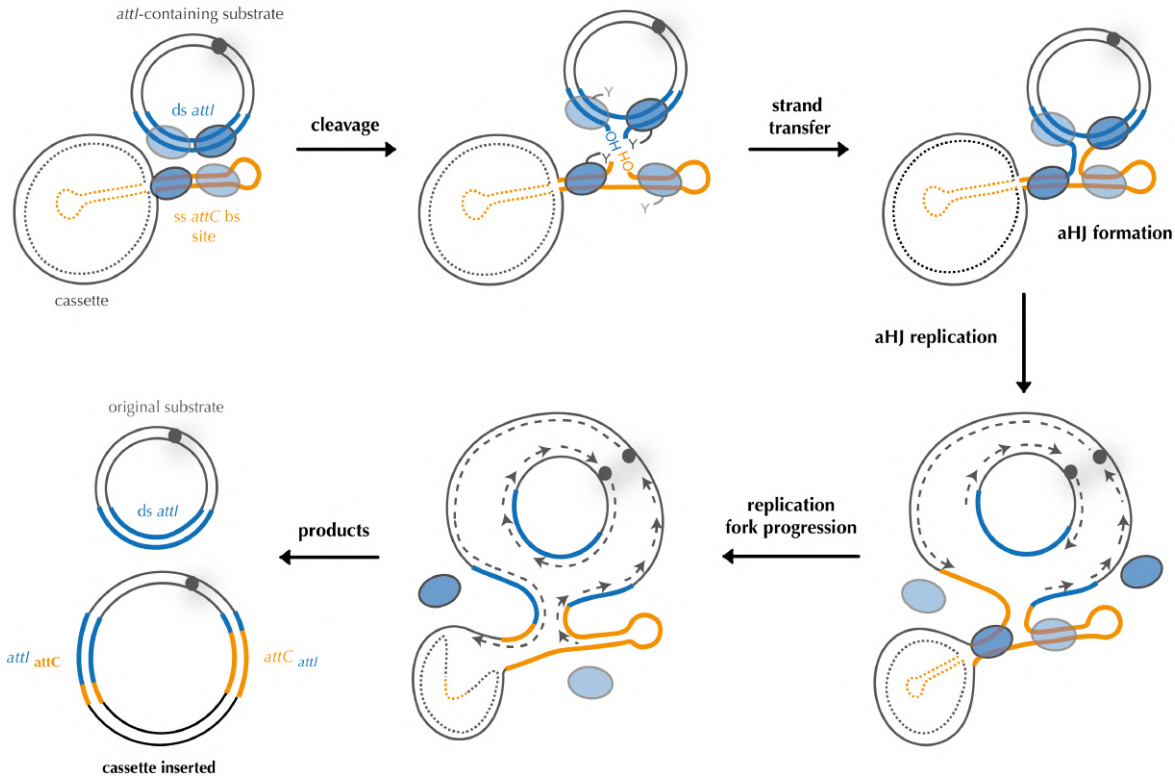


Figure 6. *attI* × *attC* reaction

Illustration of the detailed mechanism of recombination between an *attI* site (bold blue lines) and a single-stranded bottom *attC* site (bold orange lines) of a gene cassette, leading to the integration of the cassette. The nature of the top strand of the *attC* site is depicted as a dotted line to represent uncertainty regarding its form (single or double-stranded). The integrase enzyme is shown as four protomers (blue ovals) forming a synaptic complex that enables the recombination process. Within this complex, catalytic cleavage and strand transfer by two activated integrase protomers result in the formation of an aHJ. The classical resolution would lead to the formation of abortive molecules with closed ends, while non-abortive resolution involves a replication step, represented by the dark grey circle (origin of replication) and the newly synthesized DNA strands (dashed lines). The figure portrays the semi-conservative nature of integron recombination, as it generates two products: the original substrate produced from the replication of the top strand, and the novel composite molecule generated from bottom strand replication, now incorporating the gene cassette. Key recombination intermediates such as hybrid *attC*/*attI* sites are indicated, emphasizing the dynamic and intricate nature of integron recombination. Adapted from Escudero et al., 2015.

attC* × *attC: for cassette excision, the integrase must bind to two folded ssDNA *attC* sites and recombine them. The rate of this reaction is dependent on the size of the cassette and the position of the *attC* sites in the lead or lag strand, but can reach the same levels of the *attI* × *attC* reaction (Loot et al., 2017). The cassette is then extracted in the form of a covalently closed circle (Collis et al., 1993; Collis & Hall, 1992). Once again, after the binding of the integrase and the cut of bs of both *attC* sites, an aHJ is formed. The same replicative resolution of the aHJ has been proposed, conceiving a consistent semi-conservative model also for this reaction (Escudero et al., 2015). Knowing that ssDNA is unstable inside the cell, and even if the replication converts it to dsDNA, the outcome of the excised cassette is dependent on the insertion on the first position (reshuffling) or its loss (excision). Once more, the semi-conservative model of

the integron allows the bacteria to keep a backup of the original setup, gaining a new layer of control (**Figure 7**).

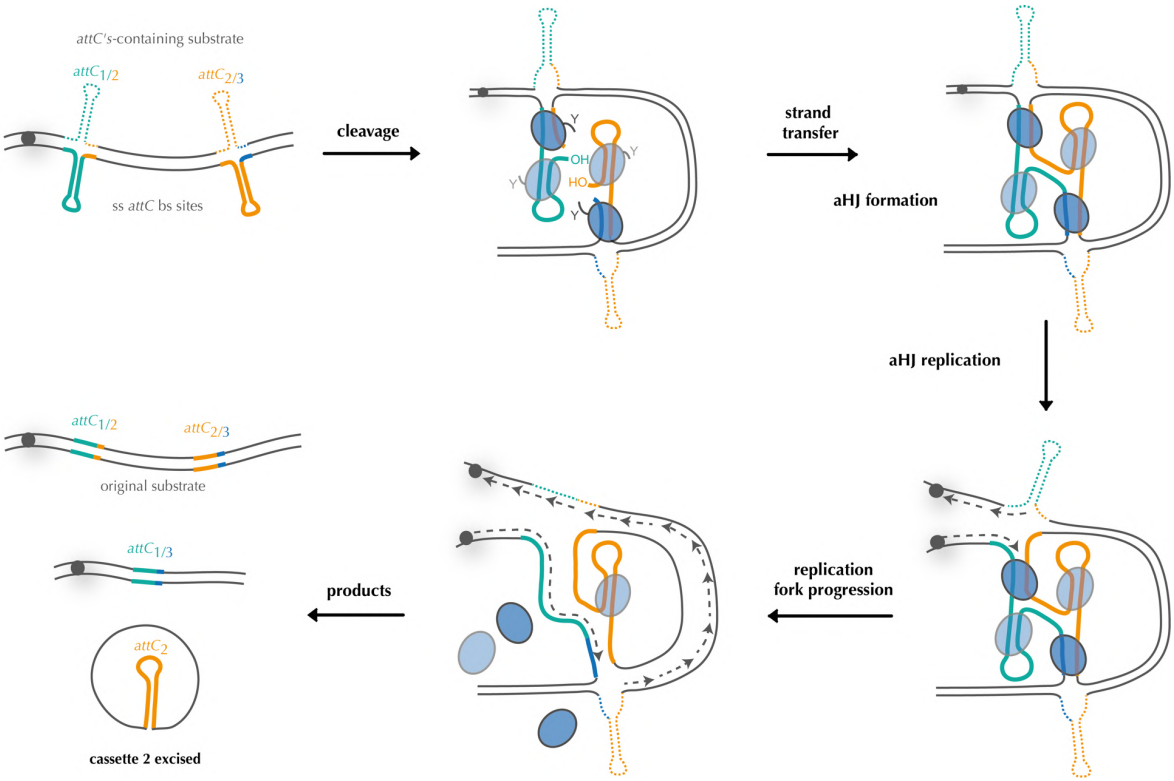


Figure 7. *attC* × *attC* reaction

Illustration of the detailed mechanism of recombination between two single-stranded bottom *attC* sites (depicted in bold green and orange), leading to cassette excision. The ambiguity regarding whether the top strands are single or double-stranded is represented by dotted lines. At the core of the process is a synaptic complex consisting of two DNA duplexes bound by four integrase protomers, with the active sites represented by dark blue ovals. Recombination is initiated through cleavage and strand transfer by the activated protomers, forming an aHJ. This junction's resolution is believed to mirror the *attC* × *attI* recombination reaction, incorporating a crucial replication step indicated by the grey circle symbolizing the origin of replication, and the nascent leading and lagging strands (dashed lines). The diagram shows the resulting two recombination products: the reconstituted substrate following the replication of the top strand, alongside the separate excised integron cassette and the cassette-less molecule—both outcomes of the bottom strand replication. Adapted from Escudero et al., 2015.

attI* × *attI: Recombination between two *attI* sites has been described (Collis et al., 2001; Hansson et al., 1997) but seems to be undertaken at a much lower rate, approximately 10^{-5}. In contrast with the other reactions, it involves a pair of dsDNA substrates. A recent in-depth characterization of the *attI* × *attI* recombination has proven that both classical resolution of the HJ and replicative resolution are possible (Escudero et al., 2016), illustrating the adaptive potential of the IntI (**Figure 8**). However, there is still no obvious biological explanation for this reaction in the integron, especially in chromosomal integrons, where the formation of dimers would have to be resolved before cell division. Nevertheless, in mobile integrons, the *attI* × *attI* reactions would lead to an exchange of the array. One could also hypothesize that

the exchange between entire arrays could also contribute to diversity due to the potential differences in P_C promoter strengths (**Figure 8**).

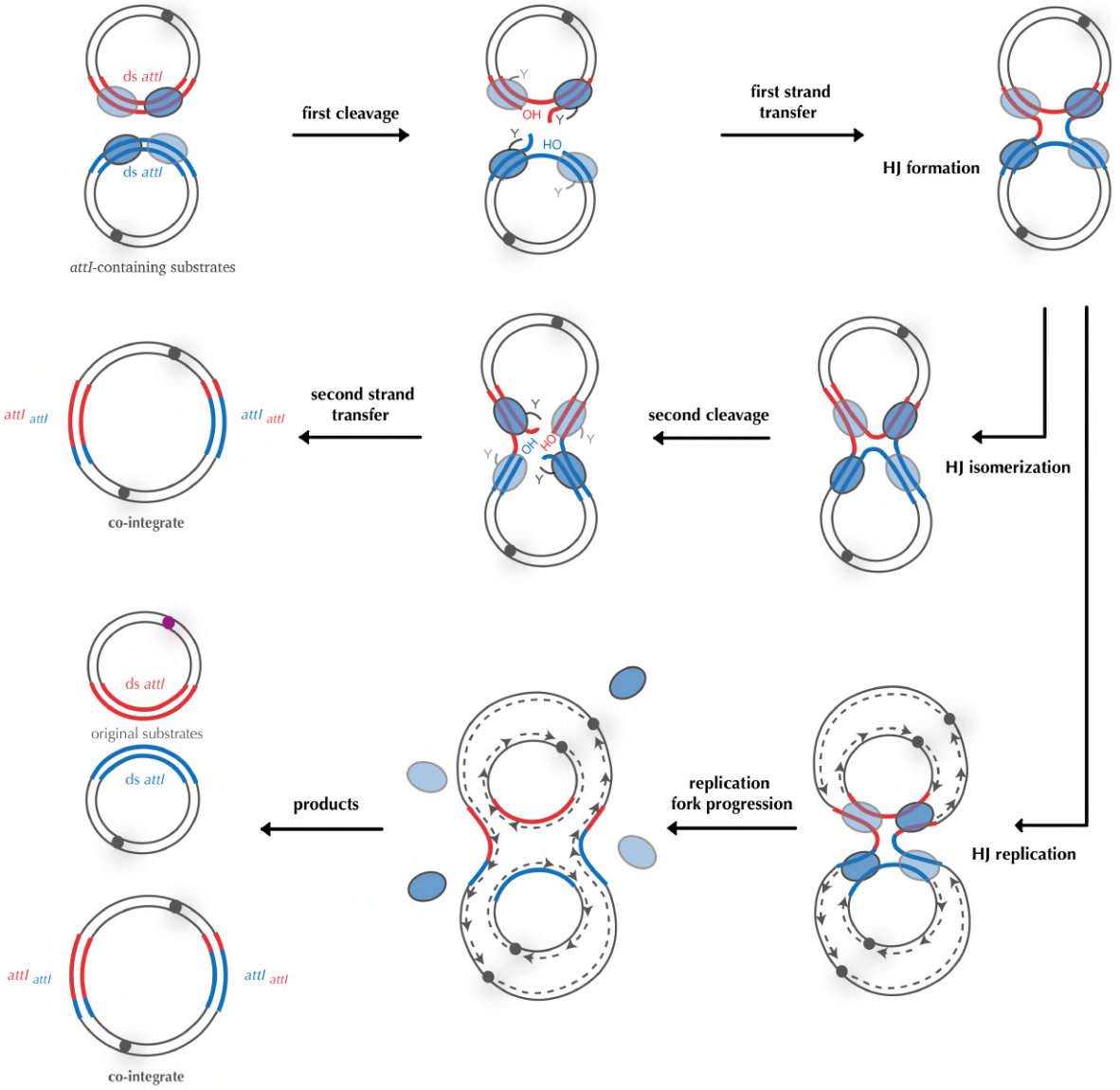


Figure 8. *attI* × *attI* reaction

Illustration of two potential resolution pathways for the recombination event between two double-stranded *attI* sites (illustrated with bold blue and red lines). The first pathway aligns with the classical resolution mechanism observed in site-specific recombination mediated by Y-recombinases. Here, the synaptic complex is composed of two DNA duplexes, each bound by four recombinase protomers, with the initial pair of activated protomers represented by dark blue ovals. Recombination proceeds with the cleavage and transfer of one strand from each duplex to form a HJ. The isomerization of the HJ shifts catalytic activity between two pairs of protomers (dark and light-blue ovals), facilitating the second strand exchange and leading to the creation of a recombined DNA molecule (co-integrate). The second proposed pathway involves the resolution of the HJ through replication, denoted by a grey circle for the origin of replication and the emerging leading and lagging strands indicated by dashed lines. The products are detailed as two initial substrates originating from top strand replication, alongside a co-integrate formed as a result of bottom strand replication. Hybrid *attI* sites, integral to the recombination process, are highlighted within the figure. Adapted from Escudero et al., 2015.

2.2 – Types of Integrons

2.2.1 – Chromosomal integrons

The coexistence of integrons and bacteria started more than a hundred million years ago (Rowe-Magnus et al., 2001). To date, integrons are present in the chromosomes of more than 17% of available sequenced genomes (Cambray et al., 2010). As a result, they are called sedentary chromosomal integrons (SCI), contrasting with their homologs, "mobile" integrons, in a few aspects. The most glaring one is their size: chromosomal integrons can be very large. Such is the case of a few genera that harbor the so-called "superintegrons" (SI), very large SCIs that can contain more than 200 cassettes (Mazel, 2006). SCIs are present in their vast majority in γ-Proteobacteria phylum, however they can be found in other groups such as β-Proteobacteria, ∂/ε subgroups and Spirochetia (Boucher et al., 2007; Néron et al., 2022; Rowe-Magnus et al., 2001). The analysis of the phylogenetic tree of their integrases suggests that they have evolved together with their hosts (Nemergut et al., 2008; Rowe-Magnus et al., 2001). Nonetheless, some incongruences have been observed. This can be explained by a certain degree of horizontal transfer; indeed, it seems that in SCIs this phenomenon exists at a lower rate and occurs between bacteria sharing the same environment (Gillings et al., 2008). Another feature that supports the collective evolution of SCIs and their hosts is that *attC* sites are almost identical at the species level, suggesting a tight relationship between the array and the host (Rowe-Magnus et al., 2001, 2003).

Chromosomal integrons encode a very heterogeneous set of integron cassettes. Although some of the ORFs have been characterized (Barker et al., 1994; Guérout et al., 2013; Ogawa & Takeda, 1993), most of their functions remain unknown. Such is the case of the *Vibrio* genus, where cassettes of their SIs have some degree of homology with known proteins, but the majority have never been identified before (Boucher et al., 2007). We will review the *Vibrio cholerae* superintegron in the next chapter.

2.2.2 – Mobile integrons

It is now broadly accepted that chromosomal integrons found their way to the clinical setting through transposons and conjugative plasmids (Boucher et al., 2007; Cambray et al., 2010; Escudero et al., 2015; Mazel, 2006). In the class 1 case, an identical platform and integrase gene can be found in the chromosomes of the nonpathogenic

soil and freshwater class β-Proteobacteria (Gillings et al., 2008). Integrons have expanded their gene pool by recruiting genes from chromosomes (Rowe-Magnus et al., 2002), and possibly from other origins. They are called "mobile" as they are present within other MGEs and have acquired the capacity to move rapidly between species. Nowadays, mobile integrons (MI) have become commonplace among clinically relevant Gram-negative bacteria, being present in 30 to over 50% of clinical isolates (Escudero et al., 2015).

A few characteristics distinguish them from their counterparts in the chromosome:

- as mentioned before, MIs are small, with arrays containing up to 11 cassettes (Naas et al., 2001);
- they are mobilizable;
- their *attC* sites are very different within and among them, suggesting different origins for the cassettes (Rowe-Magnus et al., 2001);
- and their cassette content is quite "homogenous", as they encode mainly antibiotic resistance genes (Partridge et al., 2009).

Based on the integrase sequence, six classes of mobile integrons have been described. Class 1 is the most spread and clinically relevant and also the best-studied, having become the experimental model of mobile integrons. With classes 2 and 3, they form the most critical MIs involved in antimicrobial resistance (Escudero et al., 2015). Classes 1 and 3 are linked with the Tn402 (Arakawa et al., 1995; Collis et al., 2002; Xu et al., 2007), whereas class 2 has mainly been found associated with the Tn7 (Ramírez et al., 2010). With the appearance of trimethoprim resistance in *Vibrio* spp., classes 4 and 5 were discovered (Hochhut et al., 2001; Sorum et al., 1992). And finally, class 6 was recently encountered in a mobile element within the chromosome of a MDR *Escherichia coli* (Jové et al., 2019).

The six classes of mobile integrons encode 170 unique integron cassettes that confer resistance virtually to all classes of antibiotics, except to tetracyclines (Hipólito et al., 2022, 2023; Partridge et al., 2009). This constitutes clear evidence that integrons are crucial players in the resistance field.

2.3 – Detection systems

In the clinical environment, where mobile integrons predominate, it is clear that a capturing system such as the integron ensures an adaptive advantage to the pathogen. But despite their importance, routine detection of integron cassettes in clinically relevant bacteria remains challenging (Cury et al., 2016).

The bulk of efforts to identify AR genes has been focused on plasmids and transposons; however, only a few approaches to identifying integrons and their cassette content are available to date.

The classical methods consist of polymerase chain reactions (PCRs) that target and amplify different compounds of the integron.

The first PCR approach arose from the fact that the class 1 integrons are frequently linked with Tn402, and their extremities can be reasonably conserved. The 5' primer targets the integrase, and the 3' primer targets the *qacE*Δ*sul1* genes, which are very often present at the end of the class 1 arrays (Lévesque et al., 1995). This approach is the most commonly used, but it has been demonstrated that it can miss many arrays where the 3' conserved sequence is not present (Amos et al., 2018; Rao et al., 2008). Another similar technique has been developed for class 1 integrons (using different primers) (White et al., 2000) and class 2 integrons (White et al., 2001). Again, the main limitation of these methods is the bias toward a specific set of integrons and conserved sequences that are not present in all classes of MIs.

A second version of a PCR methodology, targeting *attC* sites, has been claimed to be the gold standard on cassette detection, as it is "sequence-independent". It consists of a pair of primers containing wobble bases to encompass the maximum amount of *attC* sequences (Stokes et al., 2001). A recent work from the same laboratory depicts the identification of integron cassettes in a variety of environments, isolating more than 4,000 to 18,000 cassettes per sample. However, the sensitivity of this method is very debatable, as *attC* sites are defined primarily for not having any sequence conservation, as well as its specificity, as the technique doesn't recover the full *attC* site, being very difficult to prove if the gene corresponds to an integron cassette (Ghaly et al., 2019).

Whole-genome sequencing (WGS) has revolutionized our knowledge of bacterial genomics, and technological advances are providing evidence that next-generation sequencing (NGS) is a powerful surveillance tool (Boolchandani et al., 2019). The gap between research and clinical laboratories is being filled by new bioinformatic tools that can help with the analysis of the generated data (Florian Fricke & Rasko, 2014). Yet, identifying integrons has been hindered by the difficulty of recognizing integron cassettes.

Despite all this, a few tools to attempt the detection of integrons have been developed. The software XXR identifies *attC* sites in *Vibrio* spp. superintegrons using pattern-matching techniques (Rowe-Magnus et al., 2003). The programs ACID (no longer available) (Joss et al., 2009) and ATTACCA (now part of RAC) (Tsafnat et al., 2009) were intended to explore classes 1, 2, and 3 of mobile integrons. As all of them rely on sequence conservation, they are inadequate to identify nor align distantly related *attC* sites, hampering the detection of new gene cassettes and integron arrays. HattCI is a method that rapidly identifies *attC* sites in DNA sequences using a hidden Markov model (HMM). By breaking down the *attC* site motifs (R and L boxes, spacers, and loops) into separate states, HattCI models both their sequence characteristics and internal dependencies, based on a reference set of *attC* sites from class 1 and 2 integrons (Pereira et al., 2016). More recently, a new program was developed – the Integron Finder – that used the same approach to detect integrons and their most distinctive constituents. It uses an HMM profile to identify integrases; it relies on covariance models to recognize *attC* sites, thus identifying CALINs and SALINs; and it annotates known *attI* sites, P_{IntI} and P_c, and any pre-existing ORFs in gene cassettes, such as antimicrobial resistance genes (Cury et al., 2016; Néron et al., 2022). A curated database for integrons is also available— INTEGRALL—(Moura et al., 2009), accounting for more than 11,000 entries from 460 bacterial species.

Lastly, Rowe-Magnus (Rowe-Magnus, 2009) suggested to couple the capability of the integron platform to acquire new cassettes, and the achievement of sequence-independent recovery of integron cassettes from genomic libraries. The particular interest of this type of tool lies in the ability to gain access to a pool of cassettes from unculturable organisms. The importance of their retrieval is supported by a recent

study, suggesting that the integron cassette metagenome contains a repertoire of genes belonging to new and currently uncharacterized protein families (Sureshan et al., 2013).

The integron platform is of the utmost importance, and the lack of quick, reliable, and efficient tools to detect them has hampered our battle against antimicrobial resistance. If no significant efforts are made, the adaptative potential of integrons will continue to ensure that attempts to harness or control MDR bacteria are destined to fail.

2.4 – The integron as a biotechnological tool

On the bright side, the integron system is full of potential that could be explored in the most different sectors of biology. As we have seen, these genetic platforms can be used to perform a "self" identification, but other biotechnological applications have been suggested or described.

Gestal and collaborators (Gestal et al., 2011) have developed an integron-built cloning technique based on the delivery of synthetic cassettes into the genetic backbone of both MIs and SCIs via electroporation or natural transformation. This method doesn't need the classical antibiotic pressure for clone selection and allows the creation of an unlimited amount of genetic organizations. Indeed, a synthetic version of an integron to create a maximum number of arrangements between genes has also been described. Its purpose was to optimize the metabolic pathway of the tryptophan operon in *Escherichia coli*, and it was achieved with only six recombination events (Bikard et al., 2010a).

In a recent study, Nivina and collaborators (Nivina et al., 2020) developed an algorithm to generate highly efficient synthetic *attC* sites embedded in protein sequences. The authors demonstrated that these sites can be introduced in genes without disrupting their function. This allows for “à la carte” evolution of proteins, potentially combining different modules and generating large-scale libraries of multidomain or multi-modular proteins.

Both the specificity and flexibility of site-specific recombination render the potential for integron applications wide-ranging. Indeed, the self-identification of integrons could be further exploited to create a tool to fully identify integron cassettes from any origin. The idea of harnessing integrons for self-identification could evolve into a revolutionary diagnostic instrument capable of identifying integron cassettes from diverse sources. By pairing this capability with bacteria that exhibit natural competence - particularly those attuned to integron functionality - we could develop an efficient detection system. Notably, *Vibrio cholerae*, renowned for its role in advancing our understanding of chromosomal integrons, is also prone to uptake DNA from its surroundings. This positions it as an ideal candidate to serve as a chassis for such an innovative tool.

3 – *Vibrio cholerae*

3.1 – An environmental bacteria that led to eight pandemics

Vibrio cholerae is a halotolerant bacterium of the Vibrionaceae family that thrives in the saline waters of Asia's coastal deltas, most notably the Ganges. It is infamously linked to cholera, a devastating diarrheal disease that, without timely intervention, can lead to fatal dehydration (Clemens et al., 2017). This species is usually classified on the basis of its lipopolysaccharide antigen (O antigen), and more than 200 serovars have been characterized (Faruque et al., 1998). Historically, most of the eight recorded pandemics have been linked to the O1 serotype, with the exception of the 8th pandemic, due to serotype 0139 (Albert et al., 1993; Bik et al., 1995). In fact, we are still in the midst of the seventh, and this bacterium still kills more than 95,000 people annually, the majority being children (Escudero & Mazel, 2017). The virulence of this human pathogen is primarily due to two well-characterized factors: the **cholera toxin** (Mekalanos et al., 1983), which is prophage-encoded (Waldor & Mekalanos, 1996), and the **toxin-coregulated pilus** (Herrington et al., 1988), essential for bacterial colonization.

However, the human intestine is not the primary habitat of this bacterium, which can colonize a variety of different organisms. The natural habitats of *V. cholerae* range from open oceanic expanses to the smaller-scale ecosystems found on the exoskeleton of marine invertebrates (Sack et al., 2004; Tamplin et al., 1990). Here, *V. cholerae* not only survives but flourishes, using chitin as a nutrient source and as a signal to enter a state primed for gene exchange (Meibom et al., 2005), a testament to its evolutionary mastery of resource utilization.

V. cholerae is also known as the model organism for studying chromosomal integrons (Rowe-Magnus et al., 1999). It harbors a gigantic structure containing more than 170 cassettes that can vary dramatically between strains. This proficiency for genetic exchange contributes to its versatility, allowing it to occupy diverse ecological niches and adapt to various hosts, from single-celled protozoans to complex multicellular organisms like plants and mammals.

3.2 – Superintegron

3.2.1 – Organization and importance

The first chromosomal integron described was the *V. cholerae* superintegron (SI), an extremely large structure found on the second chromosome of this species. In strain N16916 (belonging to the El Tor biotype of the serotype O1 and responsible for an outbreak of the seventh pandemic), the superintegron contains 179 cassettes and spans 126 kilobases, accounting for 3% of the total DNA of the cell (Mazel et al., 1998). In other *V. cholerae* strains, the variability of this region is very significant, to the point of being used in the genetic classification of isolates (Chun et al., 2009; Labbate et al., 2007). The *V. cholerae* SI is the best studied and the paradigm in the field of chromosomal integrons.

As in all integrons, a stable platform is present, where a specific integrase gene *intIA* and its *attIA* site work together to capture gene cassettes. However, the most striking feature of the SI is its cassette content. The known genes of the superintegron encode diverse functions such as virulence, restriction-modification enzymes, acetyltransferases, promoter-encoding cassettes, toxin-antitoxin systems, and phage-defense systems (Blanco et al., 2024; Boucher et al., 2007, 2011; Darracq et al., 2024; Iqbal et al., 2015; Rapa & Labbate, 2013). Still, 65% of cassettes have no homologs in the databases, and 13% had homologs of unknown function (Boucher et al., 2007).

Our current understanding of the specific functions that these cassettes encode is still very limited, yet we can infer their significant role in ecological adaptation. This is evidenced by the distinctive SI content found in isolates of the same species from different regions, contrasted by the striking similarity observed among superintegrons of different *Vibrio* species inhabiting identical ecological niches (Boucher et al., 2011). Such observations indicate a selective advantage for organisms carrying specific cassettes and highlight the intense exchange of integron genes within local environments. Recently, we have showed that deleting the SI does not pose a cost for *V. cholerae* in a panoply of different conditions, which has strong implications regarding the functions one could encounter in these chromosomal platforms (Blanco, Trigo da Roza, et al., 2024), and that finding the specific utility of these cassettes will be a laborious task.

3.2.2 – 19 toxin-antitoxin systems

One of the most compelling questions in the field of superintegrons is how the stability of such large yet neutral structures is achieved. Indeed, 19 toxin-antitoxin systems contained in gene cassettes were described in the *V. cholerae* N16961 SI (**Figure 9**) (Iqbal et al., 2015; Krin et al., 2023).

Toxin-antitoxin (TA) systems consist of a persistent toxin that disrupts a vital cellular function, and an antitoxin, with a shorter half-life, that neutralizes the toxin. If the expression of these genes is stopped, the faster-degrading antitoxin is rapidly lost, leaving the toxin active and leading to either cell death or a halt in growth. Three types of TA systems are known: type I consists of a hydrophobic protein as a toxin and a small RNA (sRNA) as an antitoxin. Type II TA systems are the most represented among known bacterial genomes, where both toxin and antitoxin are proteins. Finally, type III TA systems are composed of a protein-RNA TA pair (for reviews, see, for example, Gerdes et al., 2005; Melderen & Bast, 2009). Discovered in 1983 within *Escherichia coli*'s plasmid-F (Ogura & Hiraga, 1983), TA systems have been recognized for ensuring plasmid inheritance via a process known as post-segregational killing, which operates independently from replication and partition strategies. Beyond plasmids, TA systems are frequently encountered within the bacterial genome. A variety of theories have been put forward regarding the function of chromosomal TAs, suggesting roles in regulating cellular processes such as programmed cell death in response to an external stress (Aizenman et al., 1996), modulating cell growth in starvation conditions (Gerdes et al., 2005), and the appearance of persistent cells capable of surviving stressful environments (Germain et al., 2013; Korch & Hill, 2006). Another interesting hypothesis proved experimentally by Guérout and collaborators (Guérout et al., 2013), is that TA systems can also modulate the acquisition of MGE if cross-interaction between addiction modules occurs. TA systems can also provide bacteria with a defense mechanism against phages in an abortive infection process. The phages' highjack of the transcriptional and translation machineries limits the production of the anti-toxin, leading to cell death (Laub & Typas, 2024). Finally, similar to their role in plasmid maintenance, TA systems have been proposed to stabilize large regions of the chromosome, preventing fortuitous deletions in unstable regions such as chromosomal MGE (Rowe-Magnus et al., 2003; Szekeres et al., 2007; Wozniak & Waldor, 2009).

TA systems within superintegrons appear to protect against the loss of cassettes by blocking large-scale chromosomal changes, thereby preserving the array. However, this alone does not fully explain the substantial size of superintegrons. In stress conditions, when the integrase is active, there is a critical need for precise control over the rate at which cassettes are excised. If the balance tips too far towards cassette excision, it would ultimately result in the loss of the mostly dormant cassettes. Indeed, a recent study showed that a substantial increase in the excision rate of TA cassettes results in cell death (Richard et al., 2024).

The majority of TA systems in the SI belong to type II TA systems, with the only exception of VCA0495 gene and its two associated antisense non-coding RNAs, that form a type I system. Interestingly, four TA pairs are redundant: *parD/E*-1 and *parD/E*-3, *relB/E*-2 and *relB/E*-4, and finally both N-acetyltransferase/*ataR* (Iqbal et al., 2015; Krin et al., 2023).

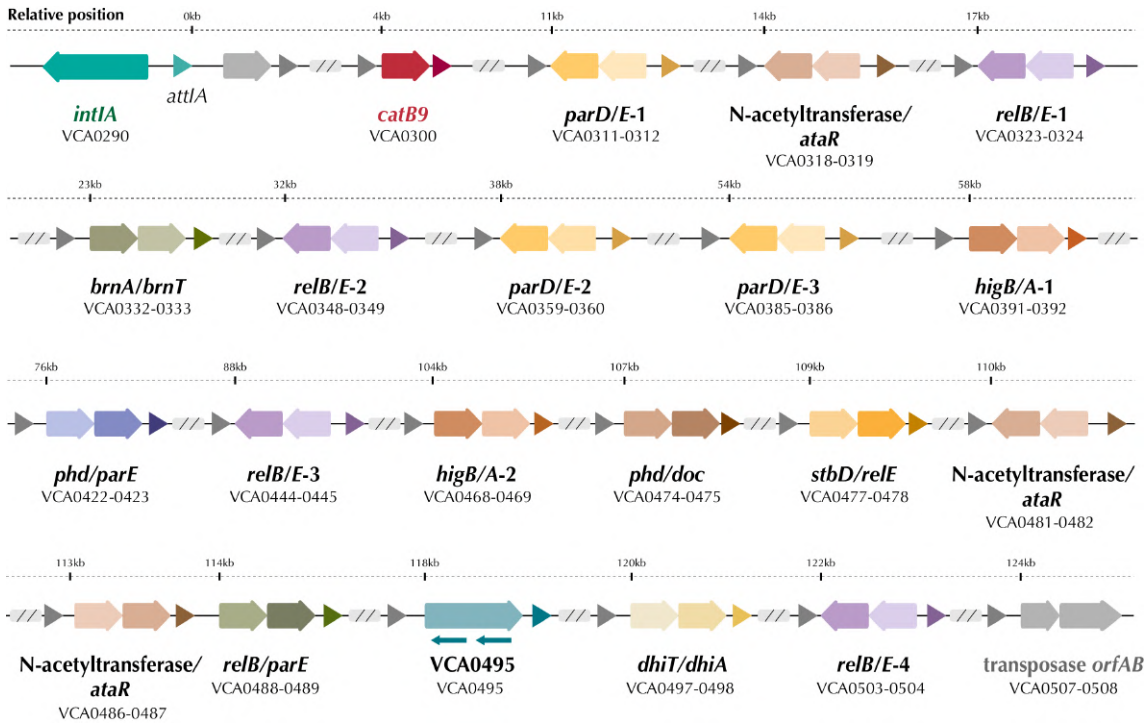


Figure 9. Organization of the 19 toxin-antitoxin cassettes in the *Vibrio cholerae* N16961 superintegron

Diagram of the organization of the TA systems within the superintegron cassette array. Individual genes contained within cassettes are depicted as arrows, while the *attC* sites are denoted by triangles. Toxin-antitoxin (TA) genes are represented by colored arrows with darker arrows representing toxins and lighter arrows representing antitoxins. The orientation and position of each arrow indicate the transcriptional direction of the genes. Additionally, small blue arrows point regions corresponding to non-coding RNAs.

The environmental cache of gene cassettes contained within the superintegrons of varying strains, their stability, and the low cost of such gigantic structures suggest that chromosomal integrons can be a vast and diverse repository of unique protein families and functions that hold immeasurable biotechnological promise.

3.3 – Natural competence

The evolvability of *V. cholerae* is heavily associated with its proficiency in horizontal gene transfer, a common phenomenon across many bacterial species. Perhaps a compelling illustration of HGT within this species is the cholera toxin being phage-borne (Waldor & Mekalanos, 1996). Yet, the exchange of genetic material in *V. cholerae* extends far beyond, as we have seen with the example of the superintegron.

This bacterium is known for its inherent competence, actively uptaking environmental DNA, a process known as natural transformation. This process was first described by Frederick Griffith in 1928 in *Streptococcus pneumoniae* (Griffith, 1928), and since then, it has been extensively explored in various bacterial species (Antonova & Hammer, 2015; Dubnau, 1999; Seitz & Blokesch, 2013a). With the exception of *Neisseria* and *Helicobacter* genus, in which bacteria constitutively take up DNA, in most microorganisms, the genes for competence are expressed only when specific environmental signals are present (Dubnau & Blokesch, 2019). For *V. cholerae*, several inducers of DNA uptake have been identified, highlighting that its competence system is coordinated by specific extracellular signals **exclusive to marine settings**, which are not found in the human host environment.

Chitin is the most prevalent polysaccharide in aquatic environments, as it is present in the exoskeleton of marine invertebrates, and the prime carbon source for marine bacteria (Aluwihare et al., 2005). In *V. cholerae*, chitin activates a cascade of genes that induce the natural competent state (Meibom et al., 2005). Since the discovery of the chitin-induced natural competence state, several authors have characterized the regulatory network of the DNA uptake process (**Figure 10**).

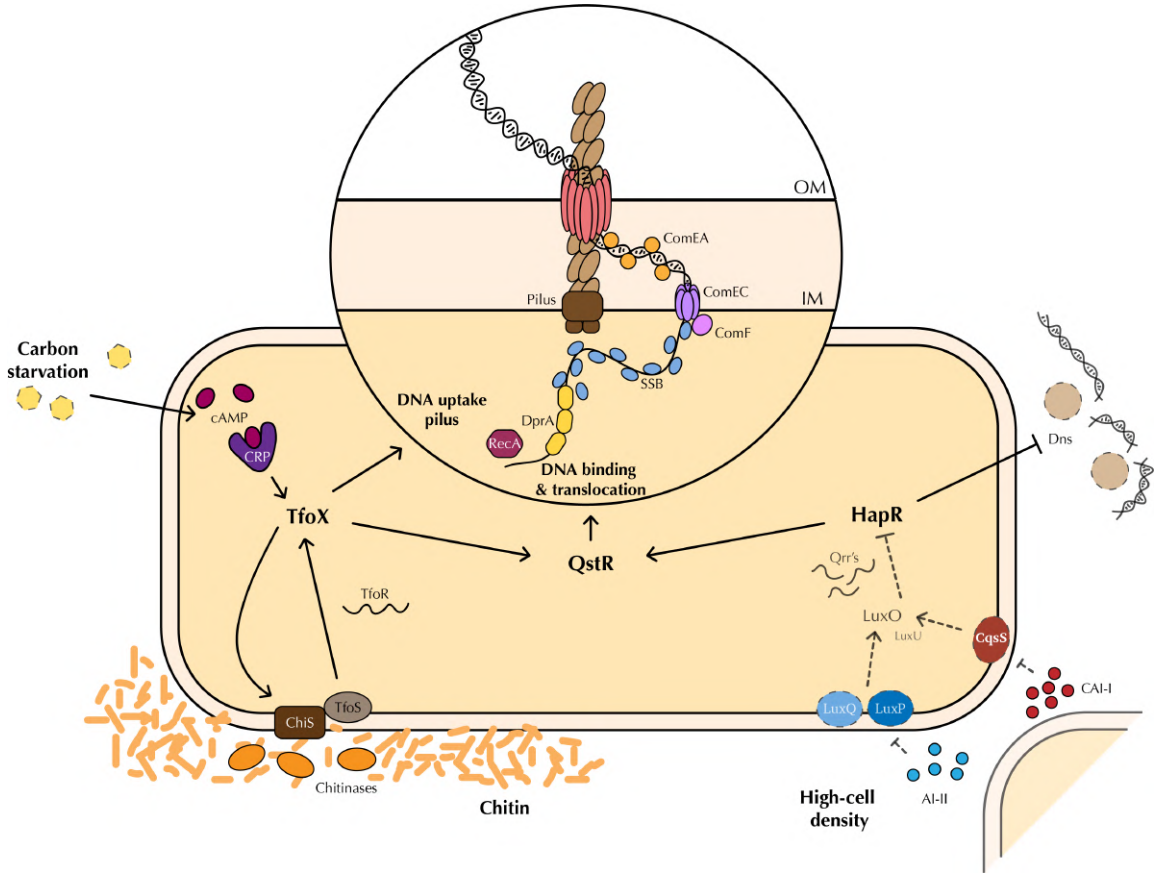


Figure 10. Regulatory network of natural competence in *Vibrio cholerae*

Schematic representation of the regulatory cascade of the natural competent state in *V. cholerae*. Central to this process, and represented on the left side of the cell, is TfoX, the master regulator of natural transformation. Interaction with chitin initiates the activation sequence, where the transmembrane regulators ChiS and TfoS, in concert with the small RNA TfoR, promote TfoX expression. TfoX then stimulates both chitin degradation and the assembly of the DNA-uptake pilus. Additionally, TfoX activates QstR, prompting the expression of genes essential for DNA binding and translocation into the cell. Carbon starvation, depicted on the left, also influences the competent state through the CRP-cAMP complex, which enhances TfoX expression by binding to its promoter. Quorum sensing (QS) via HapR modulation is represented on the right side of the cell. HapR leads to the expression of competence genes and represses the nuclease Dns to preserve transformed DNA (tDNA). HapR levels are regulated by the cell density-dependent QS system, involving autoinducers CAI-I and AI-II and their membrane sensors, CqsS and LuxP/Q, respectively. At high cell density, a decrease in phosphorylated LuxO allows HapR expression. At the top of the cell, tDNA uptake is depicted. Upon entry in the periplasm via a dedicated pilus, ComEA, ComEC, and ComF participate in the binding and translocation of tDNA through the inner membrane. The tDNA enters the cell presumably as ssDNA, as the cell employs SSB to protect it, and then DprA recruits RecA, potentially invoking the SOS response.

TfoX is the master regulator of natural transformation (Meibom et al., 2005). Indeed, when cells are attached to chitin, ChiS and TfoS, two transmembrane regulators, together with TfoR, a small RNA, induce the expression of TfoX (A. B. Dalia, et al., 2014a; Yamamoto et al., 2011, 2014). The regulator continues the cycle of activation of chitin degradation but also ensures the assembly of the DNA-uptake pilus (Seitz & Blokesch, 2013b) that can itself bind to chitin (Adams et al., 2019). Through the activation of QstR, TfoX is also responsible for the expression of DNA binding and translocation genes. Blokesch and collaborators demonstrated that when overexpressed, TfoX can efficiently induce chitin-independent transformation.

Moreover, it was also demonstrated that **carbon starvation** plays an important part in the expression of TfoX as the CRP-cAMP complex binds directly to its promoter to initiate transcription (Blokesch, 2012). Indeed, the presence of glucose in the medium is sufficient to repress natural transformation.

Another important layer of regulation in natural transformation is the response to **quorum sensing**, coordinated by HapR (Mukherjee & Bassler, 2019). Similar to TfoX, through the activation of QstR, HapR induces a subset of essential competence genes, such as *comEA*, *comEC*, and *comF*, involved in DNA binding and translocation through the inner membrane (Dubnau & Blokesch, 2019). It also represses the periplasmatic and extracellular nuclease Dns (Blokesch & Schoolnik, 2008), avoiding the degradation of transformed DNA (tDNA). The expression of HapR is controlled by quorum sensing: in this bacterial species, two main autoinducers are produced, CAI-I and AI-II. At low cell density, the unbound CAI-I and AI-II membrane sensors, CqsS and LuxP/Q, respectively, induce the phosphorylation of LuxU, which in turn donates its phosphor group to LuxO. Phosphorylated LuxO represses HapR by inducing the production of quorum regulatory RNAs, that directly bind to the RBS of HapR. On the contrary, at high cell density, CAI-I and AI-II are attached to its membrane sensors, and consequently, LuxO is not phosphorylated, thus not active, allowing the expression of HapR (Antonova & Hammer, 2015; Mukherjee & Bassler, 2019).

It is assumed, as in other species, that the DNA enters the *V. cholerae* cell in ssDNA form, inducing the SOS response. The single-strand DNA-binding protein (SSB) is immediately deployed by the cell to protect the DNA from degradation, and then DprA binds to the transformed DNA to recruit RecA (Sharma et al., 2023). Here, a cascade of events is determined by the utility of the tDNA that was acquired.

The reasons behind the development of natural competence in bacteria, such as *V. cholerae*, to incorporate extracellular DNA, have intrigued scientists since Griffith's initial work. It is commonly assumed that this process exists as a source of nutrients in conditions of low carbon sources (Finkel & Kolter, 2001; Palchevskiy & Finkel, 2006), as well as for DNA repair from kin neighbors (Matthey & Blokesch, 2016). The state of natural competence is highly relevant in marine settings, where vibrios frequently encounter zooplankton with chitinous structures, providing ample opportunities for

HGT among dense microbial consortia. Compelling evidence of *V. cholerae's* strategic genetic acquisition is its use of the type VI secretion system in competence conditions (Jaskólska et al., 2018) to outcompete neighboring bacteria, thus appropriating their DNA. Indeed, contrarily to other species, Vibrio's do not appear to take up only species-specific DNA, being able to acquire genetic material from other bacteria (Suckow et al., 2011).

The complexity of the regulation and the lack of a complete understanding of natural competence render this process challenging to control in the laboratory. For instance, even the chitin origin itself can be critical to the full induction of the competence state in laboratory conditions. Nevertheless, it remains an effective and predominantly employed method for genetic modifications in *V. cholerae*, with several studies working on improving its efficiency and underlying its utility and importance (A. B. Dalia, et al., 2014b; T. N. Dalia et al., 2017; Marvig & Blokesch, 2010).

It is evident that *V. cholerae* has evolved a sophisticated set of genetic tools, enabling it to leverage HGT and thrive by acquiring new and advantageous functions from its microbial neighbors. This versatility not only underscores the organism's adaptability but also marks it as an ideal chassis for a diverse collection of biotechnological applications. As a fast-growing biosafety level 2 organism with a natural competency for DNA uptake, *V. cholerae* stands out as uniquely suitable for working with various MGEs, including integrons.

The central aim of this thesis is to develop a novel tool that will use the genetic strengths of *V. cholerae*, coupled with the biotechnological potential of integrons, to efficiently capture and identify integron cassettes. Targeting this objective, the research presented here seeks to contribute with significant advancements to the biotechnological domain, particularly in the area of genetic engineering.

Objectives

Lisboa, 2022

Objectives

Integrons are genetic platforms responsible for the emergence and establishment of antimicrobial resistance. However, detecting integron cassettes remains challenging, as conventional methods such as PCR are often biased, and deep sequencing is still not practical for routine analysis.

In this doctoral thesis, we aim to develop a biotechnological tool that harnesses the biotechnological potential of integrons with different approaches to detect integron cassettes independently of their genetic context, phenotype, or background. The specific objectives of this thesis are:

- Developing an integron cassette capture platform: by re-engineering a class 1 integron, using its broad specificity, we aim to use it as a cassette capture device. The readout is given by a counter-selectable marker, allowing for the blind selection of cassette acquisition, i.e., independently of the phenotype conferred by the integron cassette.

- Creating the chromosomal cassette harvester: we want to generate cassette libraries from SCIs by adapting the capture device to a plasmid setup.

- Generating a sentinel bacterium capable of detecting integrons: by inserting our capture device into a bacterium optimized for plasmid uptake by conjugation and using it to assess its ability to detect integron cassettes circulating among mobile plasmids.

- Optimizing *Vibrio cholerae* as a chassis to test DNA samples directly: first, eliminating the *V. cholerae* superintegron will prevent interference with our capture platform. Then, by enhancing its natural competence to create a hyper-competent strain, we will form the foundation of our integron cassette-capturing tool to test DNA samples directly.

Materials and methods

Porto Côvo, 2021

Materials and methods

Strains, plasmids, and primers used

All strains, plasmids, and primers used in this thesis are listed in **Supplementary Table 1**, **Supplementary Table 2**, and **Supplementary Table 3**.

Growth conditions

E. coli strains were grown in LB Lennox plates (BD, France) supplemented with agar at 1.5% (Condalab, Spain) and incubated at 37°C unless stated otherwise. *V. cholerae* strains were grown in LB Miller agar plates (Sigma-Aldrich, USA) supplemented with agar at 1.5% (Condalab, Spain) at 30°C unless stated otherwise. Liquid bacterial cultures were incubated at 37°C/30°C with continuous shaking at 200 rpm. If necessary, the medium was supplemented with the corresponding antibiotic for the selection of the bacterial strain of interest. All the antibiotics and reagents used are listed in **Supplementary Table 4**.

Plasmid constructions

Plasmid constructions were made using either an in-house Gibson Assembly cloning master mix (Gibson et al., 2009) or Fast-Digest restrictions enzymes (ThermoFisher Scientific, USA). For the Gibson assembly, the primers used for insert amplification and vector linearization were designed to have 20 complementary base pairs at their ends for successful hybridization. If needed, DNA fragments were synthesized *in vitro* (IDT, USA). PCRs for constructions were carried out using the Phusion Green High-Fidelity (HF) Master Mix (ThermoFisher Scientific, USA). PCR products were subsequently purified with GeneJet PCR Purification Kit (ThermoFisher Scientific, USA) or High-Pure PCR Product Purification Kit (Roche, Switzerland). For the digested fragments, reactions were prepared and incubated following the manufacturer's instructions. The DNA fragments were then ligated with Gibson Assembly enzymes by incubating the reactions at 50°C for 30 minutes or with the T4 DNA ligase (ThermoFisher Scientific, USA) by incubating the reactions for 2h at 22°C. Finally, the cloning mix was transformed into either *E. coli* DH5α electro-competent

cells or *E. coli* π3813 if the construction bore a functional *ccdB* gene. Relevant genes and sequences were then verified through colony-PCR using the DreamTaq Green Master Mix (ThermoFisher Scientific, USA) and sequenced by Sanger sequencing (Macrogen, South Korea; or Eurofins, France). Plasmids were then extracted using the Plasmid GeneJet Plasmid Miniprep Kit (ThermoFisher Scientific, USA) and electroporated in the desired recipient.

E. coli chromosome constructions

Chromosome constructions in *E. coli* MG1655 were cloned in the *attB* site of the λ phage recurring to the thermo-sensitive plasmid, pKOBEG (Derbise et al., 2003). This plasmid encodes the *red* operon of the same λ phage, allowing homologous recombination between a transformed DNA fragment and the targeted region of the chromosome, and a modified pSC101 origin of replication that renders it sensitive to temperature. To generate the DNA fragments transformed, the target construction was amplified from a previously built plasmid template with 20 bp of homology with the homology regions. The homology regions were then amplified from the lysate of the targeted cell, and the three fragments were ligated through SOE-PCR. SOE-PCRs were performed using the Phusion Green High-Fidelity (HF) Master Mix (ThermoFisher Scientific, USA) and the following thermocycle conditions: an initial stage of 30 seconds at 98°C, followed by 20 cycles of amplification (10 seconds at 98°C, 15 seconds at 55°C, and 30 seconds/kb at 72°C), followed by 10 cycles of amplification (10 seconds at 98°C, 15 seconds at 55°C, and 30 seconds/kb + 20s per cycle at 72°C), and a final stage of 10 min at 72°C. To prepare the electrocompetent cells, the *E. coli* MG1655 strain containing the pKOBEG was grown with the proper antibiotics and arabinose 0.2% to induce the *red* operon. After reaching an OD_{600} of 1.0, electrocompetent cells were prepared as usual. The PCR product obtained was then electroporated in the recipient cell, and cells were incubated for 2h at 37°C. Transformants were plated in LB agar with the appropriate antibiotics for the correct selection of the construction, and incubated at 42°C to induce the loss of the pKOBEG plasmid. The correct acquisition of the construction was verified through external colony-PCR and sequenced by Sanger sequencing (Macrogen, South Korea; or Eurofins, France). The loss of the pKOBEG plasmid was verified through phenotype and colony-PCR.

V. cholerae chromosome constructions

Constructions by natural transformation

To clone the different tested constructions in *V. cholerae,* we recurred to natural transformation. To do so, we generated fragments with either 500 or 1,000 bp of homology with the targeted region in the genome. The DNA fragments transformed were generated as described above for *E. coli* MG1655. Natural transformation was performed as previously described in (Stutzmann & Blokesch, 2020) with some modifications. Briefly, strains were inoculated from frozen stocks into liquid LB medium and grown overnight at 30°C. Overnight cultures were then diluted 1:100 and grown to an OD_{600} of 1.0. Cells were washed and resuspended in DASW (defined artificial seawater) (Instant Ocean, USA) 0.5x and inoculated in a final volume of 1mL DASW with 150 µl of crab chitin (Apollo Scientific, UK) slurry. Cells were incubated in static conditions at 30°C for ~18 h. Then, 550 µL of supernatant was removed from the reactions, and the purified transforming DNA (tDNA) was added. DNA and cells were incubated for ~18 hours in static conditions at 30°C. When needed, reactions were outgrown by adding 1 ml LB broth and shaking at 30°C for 2h. Cells were plated in LB agar with the corresponding antibiotic. The correct introduction of the construction was verified by external colony-PCR and sequenced by Sanger Sequencing (Macrogen, South Korea; or Eurofins, France).

Clean-mutant generation

Clean deletions were generated essentially as described (Meibom et al., 2004). The suicidal vectors used (pGP704-Sac28) are listed in **Supplementary Table 1**. These plasmids' replication is dependent on the chromosome-encoded π protein, and they contain a *sacB* gene that acts as a counter-selectable marker. Briefly, the selected plasmid was introduced into *V. cholerae* N16961 via biparental mating, and the successful integration was selected with ampicillin 100 µg/mL in thiosulfate citrate bile salts sucrose (TCBS) medium (Sigma-Aldrich, USA) (*Vibrio* selective agar). Isolated colonies were grown in LB agar plates with ampicillin to ensure the correct integration of the plasmid. Double-crossover was subsequently favored by growth in liquid LB medium for ~6h x 2. Sucrose counterselection was performed in LB (ØNaCl) + sucrose 10% agar plates incubated at 20°C. The targeted deletion was then confirmed by colony-PCR.

Killing assays

Killing assays were performed to (i) assess the efficiency of the CSM::attI1 fusion proteins, (ii) establish the limit of detection of the generated tool, and iii) characterize the escape mutants of the system (**Figure 11**). Briefly, cells were grown in a liquid medium overnight in survival conditions, i.e., in the case of the CcdB/CcdA TA system, in the presence of the appropriate reagents to induce the production of *ccdA* and repress the induction of *ccdB*, and in the case of the *sacB* counter-selectable marker in LB medium. The next day, cells were diluted up to 10^{-7} and plated in 5 µL spots both in survival conditions, as previously described, or killing conditions. In the latter, for the CcdB/CcdA TA system, cells were plated in the presence of the appropriate reagents to repress the production of *ccdA* and induce the production of *ccdB*. In the case of the *sacB* counter-selectable marker, cells were plated in LB (ØNaCl) + sucrose 10% medium at either 30°C for *E. coli* strains or 20°C for *V. cholerae* strains. If necessary, 100 µL of the overnight cultures were also plated in killing conditions. After this, the number of survivors in killing conditions over the total colony forming-units (CFU) was calculated, allowing us to measure the escape mutant frequency. If relevant, a colony-PCR was performed to characterize the escape mutants, and PCR products were sequenced by Sanger sequencing (Macrogen, Korea).

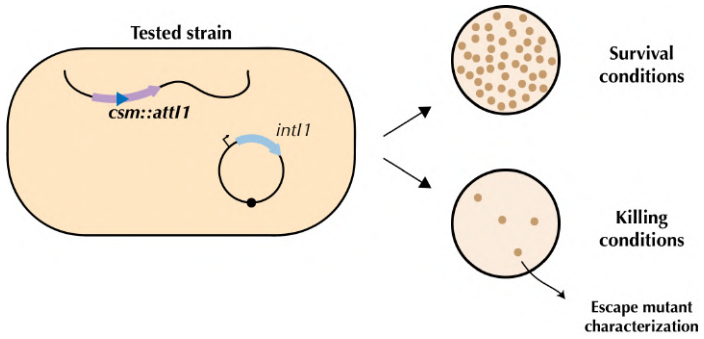


Figure 11. Killing assay

Schematic representation of a killing assay to test the *csm::attI1* constructions. The recipient is cultured under "survival conditions" and then plated either in survival conditions to count the total CFUs and "killing" conditions to count the escape mutants. This allows us to count the escape mutant frequency and the range of action of the tool. Furthermore, we can also characterize the escape mutants by PCR and Sanger sequencing.

Classical recombination assays

Classical recombination assays were performed to measure the efficiency of the *csm::.attI1* constructions in integron recombination (Figure 12). These experiments were essentially performed as previously described (Biskri et al., 2005; Bouvier et al.,

2005) with some modifications. This assay simulates the natural conditions of cassette delivery through horizontal gene transfer by conjugating a suicidal plasmid (pA123) that contains the $attC_{aadA7}$ site. This plasmid carries an RP4 origin of transfer ($oriT_{RP4}$) oriented in such a way as to deliver the bottom strand of the *attC* recombination site. This ensures the transfer of the *attC* site in its recombinogenic form, thus facilitating integron recombination. The recipient strain cannot sustain the replication of the suicidal plasmid; therefore, the only way for the vector pA123 to persist in the recipient cells is for it to recombine with the *attI1* site contained in the recipient strain. The cassette capture efficiency is measured by calculating the recombination frequency.

Briefly, the donor strain A123 was cultured overnight in LB medium supplemented with chloramphenicol 25 µg/mL and diaminopimelic acid (DAP) 0.3 mM. The recipient-tested strains were cultured overnight in LB medium supplemented with appropriate antibiotics for its correct selection, and reagents for integrase repression. The next day, the donor culture was diluted 1:100 in LB with chloramphenicol and DAP, as indicated above, and the recipient strains were diluted 1:100 in LB with the same antibiotics and the appropriate reagents to *intI1*. These cultures were incubated until an OD_{600} of 0.7. Next, after washing the cultures, the recipient and donor were mixed in a total volume of 1 mL at a ratio of 4:1, centrifuged for 2 min at 8,000 rpm, and resuspended in 100 µL of LB. This volume was spread on a conjugation membrane (Millipore mixed cellulose ester membrane, 47 mm diameter, 0.45 µm pore size) on LB agar + DAP 0.3 mM + the appropriate reagent to induce *intI1*. Plates were incubated overnight (~18 hours) to allow conjugation and recombination to occur. The following day, the cells contained in the membrane were resuspended in 5 mL of LB, after which 1:10 serial dilutions up to 10^{-7} were performed. 5 µL of each dilution was plated on LB medium supplemented with the appropriate antibiotics and reagents to recover the total recipient bacteria, and, to select for recombinant bacterial cells, dilutions were plated on LB medium with chloramphenicol either 25 or 2.5 µg/mL (for *E. coli* or *V. cholerae* respectively), as well as other antibiotics and reagents needed. The ratio between recombinants growing in chloramphenicol and the total recipients allowed us to measure the **recombination frequency by phenotype selection**.

To count recombinant bacteria by survival from insertion within the chosen counter-selectable marker, cells were either plated in the appropriate reagents to

induce the expression of *ccdB* and repress the expression of *ccdA*, or in the case of *sacB*, cells were plated in LB medium (ØNaCl) + sucrose 10%, and incubated at 30°C in the case of *E. coli* and 20°C in the case of *V. cholerae*. The ratio between the number of bacteria growing in killing conditions over the total recipients allowed us to measure the **recombination frequency by blind selection**.

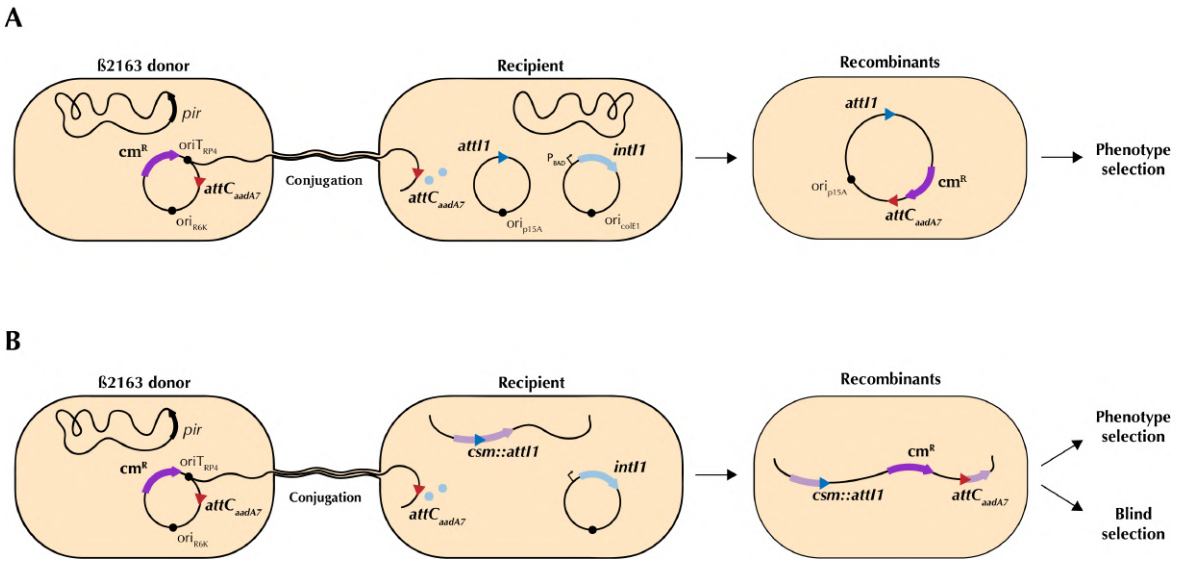


Figure 12. Classical recombination assay

Schematic representation of a classical recombination assay. A donor conjugates a suicidal plasmid that contains an *attC* site, thus mimicking an excised cassette. A recipient, either (**A**) a control with only an *attI1* or (**B**) the tested tool with the *csm::attI1* construction, receives the plasmid, and recombinants can be selected either by phenotype, in both cases, or blind selection in the tested *csm::attI1* recipient.

In all cases, a minimum of eight colonies were verified by PCR with either MFD/SWbeg primers or sacB::attI check R/SWbeg primers. If necessary, recombination frequencies were corrected with the PCR results.

Library generation

To create a library from *V. cholerae* N16961 superintegron, we first electroporated this strain with plasmid pB746, which contains *ccdA* and *intI1* (see **Figure 29A**). Next, the cells were cultured in the presence of arabinose 0.2% to ensure the production of *ccdA*, and they were electroporated with plasmid pB824, which contains the *ccdB::attI1* construction (see **Figure 29A**). The next day, at least three transformant colonies were selected and grown in liquid LB medium in the presence of zeocin 50 µg/mL, carbenicillin 100 µg/mL, arabinose 0.2%, and vanillate 25 µM (to induce the expression of the integrase). After 3 hours of incubation, serial 1:10

dilutions of the liquid culture were performed up to a 10^{-7} dilution, and 5µL of each dilution was plated on LB medium supplemented with various antibiotics and reagents. To count the total recipient cells, the culture was plated on zeocin 50 µg/mL, carbenicillin 100 µg/mL, and arabinose 0.2%. To count recombinants, the culture was plated on 50 zeocin µg/mL, carbenicillin 100 µg/mL, glucose 1%, and IPTG 200 µg/mL. The recombination frequency was calculated as the proportion of recombinant CFUs with respect to the total number of recipient CFUs. The cassette content of recombinants was amplified with primers MDF and MRV II. Cassettes were then identified through Sanger sequencing (Macrogen, Korea).

Library sequencing and analysis

Library sequencing and read processing

The plasmid library was extracted from a pool of approximately 10^5 cells. A PCR of the plasmid extraction using primers MFD and MRV II was performed, and the resulting purified PCR product was sequenced by Plasmidsaurus using Oxford Nanopore Technology. Reads were processed with a custom pipeline using software packages from Bioconda, CRAN or Bioconductor public repositories, as follows. Firstly, the quality of sequencing reads was assessed using Nanoplot (version 1.42.0), and low-quality reads (filtered to Q >10 and length >100 nt) and adapters were trimmed using NanoFilt (version 2.8.0) and PoreChop (version 0.2.4), respectively. After quality control, the reads were mapped to the reference genome *Vibrio cholerae* O1 biovar El Tor str. N16961 (Refseq GCF_000006745.1) using Minimap2 (version 2.28) with the *-ax map-ont* option, allowing for secondary alignments to account for multiple integron recombination events. Mapping coverage was analyzed and visualized using Bedtools (version 2.31.1) and BAMdash (version 0.2.4).

Cassette identification and quantification

The structure of the N16961 superintegron was annotated with IntegronFinder (version 2.0.5, with options *--local-max --evalue-attc 4 --calin-threshold 1*). 176 *attC*s were identified. Furthermore, two *attC* sites were manually added (93.02% and 96.72% identity with other *attC* sites from the SI using Geneious software), as it was previously described that the *V. cholerae* N16961 SI possesses 179 cassettes (Rowe-

Magnus et al., 1999). Nevertheless, we were able to only identify 178. Cassettes were defined as sequences located between consecutive *attC* sites, or between the *attI* site and the first *attC* site. The integron annotation was subsequently combined with Bakta (software version 1.9.2, Database v. 5.1) genome annotation to obtain a final updated annotated genome. Then, read counts per feature (CDS, *attC*, cassette) and features per read were determined using *featureCounts* function from the Rsubread package (version 2.18.0), with the following options: *isPairedEnd=FALSE, countMultiMappingReads=TRUE, allowMultiOverlap=TRUE*.

Data handling, statistical analysis and plots were carried out with R packages circlize, ggplot2, ggplotly, ggpubr, and Biocircos, as detailed in the custom R script, as described in the GitHub Repository (https://github.com/mredrejo/cassette_harvester).

Recombination through conjugation

To detect integrons from conjugative or mobilizable plasmids, we performed a conjugation assay similar to the classical recombination assay but with some modifications. In this case, the recipient strain, D106, was cultured overnight in LB medium supplemented with zeocin 50 µg/mL, carbenicillin 100 µg/mL, and glucose 1%. The donor strain A074 was cultured overnight in MH medium supplemented with trimethoprim 50 µg/mL, and the donor strain A075 was cultured overnight with carbenicillin 100 µg/mL. The next day, the donor cultures were diluted 1:100 in the same conditions as indicated above, and the recipient strain was diluted 1:100 in LB with the zeocin 50 µg/mL, carbenicillin 100 µg/mL, and arabinose 0.2%. These cultures were incubated until an OD_{600} of 0.7. Next, after washing the cultures, the recipient and each donor were mixed in a total volume of 1 mL at a ratio of 4:1, centrifuged for 2 min at 8,000 rpm, and resuspended in 100 µL of LB. This volume was spread on a conjugation membrane (Millipore mixed cellulose ester membrane, 47 mm diameter, 0.45 µm pore size) on LB agar + arabinose 0.2%. Plates were incubated overnight (~18 hours) to allow conjugation and recombination to occur. The following day, the cells contained in the membrane were resuspended in 5 mL of LB, after which 1:10 serial dilutions up to 10^{-7} were performed. 5 µL of each dilution was plated on LB medium supplemented with zeocin 100 µg/mL to recover the total recipient bacteria. To

recover recombinants through blind selection, cultures were plated in LB medium (ØNaCl) + sucrose 10% + zeocin 12.5 µg/mL and incubated at 30°C. To characterize the captured cassettes, we performed a colony PCR to 24 recombinants from each donor using the primers sacB::attI check F and sacB::attI check R.

SeqDelTA

All strains, plasmids, and primers for the deletion of the SI are listed in **Supplementary Table 5**, **Supplementary Table 6**, and **Supplementary Table 7**. To generate the ΔSI mutant of *V. cholerae* N16961, we developed SeqDelTA, as depicted in **Figure 36**. After identifying the 19 toxin-antitoxin (TA) systems contained in the *V. cholerae* N16961 superintegron, we prepared successive constructs designed with homology regions and a selective marker. A left homology region (LHR) was designed and maintained throughout the procedure. This region corresponded to the *catB9* gene (VCA0300) and was amplified with primers VCA 299 F and LRHI VCA 300 R. For the successive deletions, different right homology regions (RHD) were designed to maintain the antitoxin gene but to eliminate part of the corresponding toxin gene, ensuring that the event was not lethal to the bacteria. As selective markers, three different antibiotic resistance genes (Zeo^R, Cm^R, and $Carb^R$) were used and amplified with the same primers, facilitating the consequent design and assembly of the HRs. These three fragments (LHR, resistance marker, and RHR) were then assembled by SOE-PCR, placing the resistance marker between the LHR and the RHR. Each successive construct was introduced by natural transformation (described above) into the *Vibrio cholerae* strain obtained from the previous TA system deletion. This strategy allowed the removal of the resistance marker introduced in the previous step while advancing the deletion of TA systems by changes in the RHR. When necessary, due to gene orientation and/or gene order in the TA system, alternative LHRs were designed, and two resistance markers were maintained in consecutive allelic replacements before the new step allowed the deletion of both.

Last deletion step

The last deletion was performed using the R6K origin of replication-plasmid pMP7 as described in (Val et al., 2012). pMP7 replication is dependent on the chromosome-encoded π protein and encodes the *ccdB* toxin under the control of the

inducible promoter P_{BAD}. 500 bp of each side of the superintegron were amplified with primers RHI link pMP7 F and LRHI+D R for the LHR, and RHD F and RHD link pMP7 R for the RHR. The two fragments were then joined by SOE-PCR and digested with enzymes NaeI and HindIII. Finally, the insert was ligated with the empty pMP7 backbone, giving rise to pMP7_Δ*intIA attIA* zeo^R. We started the cloning steps by transforming the pMP7_Δ*intIA attIA* zeo^R plasmid into *ccdB*-resistant *E. coli* π3813 (thy::*pyr* +) competent cells. Transformants were selected in LB agar plates containing chloramphenicol (Cm) 25 µg/mL, thymidine 0.3 mM, and glucose 1%. For conjugal transfer of pMP7_Δ*intIA attIA* zeoR into *V. cholerae*, the plasmid was first transformed into the donor strain *E. coli* β3914 (dap::pir +). Transformants were selected in LB agar plates supplemented with chloramphenicol 25 µg/ml, DAP 0.3 mM, and glucose 1%. Conjugation assay was performed by growing both donor (*E. coli* β3914) and recipient (*V. cholerae* A066) strains to an OD_{600} of 0.2. Then, cells were mixed in a ratio of 1/10 (donor/recipient) and transferred to a mating filter placed on an LB agar plate supplemented with DAP 0.3 mM and glucose 1%, and conjugation was incubated overnight at 37˚C. Integration of the pMP7_Δ*intIA attIA* zeo^R into the A066 genome was selected by washing the filter in 5 mL of LB and plating serial dilutions on Cm 2.5 µg/mL and glucose 1% plates, but lacking DAP. Afterward, Cm^R colonies were grown in liquid LB medium and plated in LB agar plates supplemented with arabinose 0.2% to express *ccdB* and select for the second crossover that implies the excision of the pMP7 backbone. After this process, we amplified the superintegron region with external primers and selected a *V. cholerae* ΔSI strain (strain A101, **Supplementary Table 5**).

ΔSI strain sequencing

Genomic DNA of the obtained *V. cholerae* ΔSI strain (A101) was extracted using the DNeasy Blood and Tissue Kit (QIAGEN, Germany) following the manufacturer's protocol. The DNA concentration was determined using a Qubit (Thermo Scientific, USA). DNA sequencing was performed at Institut Pasteur using Illumina paired-end reads, at MiGS (Pittsburgh, US), now SeqCoast Genomics (Portsmouth, US), and we performed in-house long-read sequencing using MinION (Oxford Nanopore Technologies). A hybrid assembly was generated using Unicycler (version 0.4.9). Data analysis was accomplished with Geneious Prime software, and the genetic variants

were identified by mapping the generated reads to the *Vibrio cholerae* O1 biovar El Tor str. N16961 (Refseq GCF_000006745.1).

Mutation correction

Three mutations were found in *V. cholerae*$_{\Delta SI}$ strain (A101): *rocS* (VC0653), *cry2* (VC01392), and *rpoS* (VC0534). While *rocS* and *rpoS* mutations led to frameshifts, *cry2* contained a non-synonymous mutation. To revert the mutations to the WT variant, allelic exchanges were performed using the pMP7, as previously described. The three plasmids used are listed in **Supplementary Table 6**: pMP7_*rocS*, then pMP7_*rpoS*, and finally, pMP7_*cry2*. After each step, we Sanger-sequenced the targeted gene and selected a *V. cholerae* strain corrected for *rocS*+ (A677), *rocS*+, *rpoS*+ (A684), and finally *rocS*+, *rpoS*+, *cry2*+ (B522). We verified by Illumina whole-genome sequencing that the *V. cholerae*$_{\Delta SI}$ (B522) did not contain any other unintended mutations.

Competition assays

Competition assays were performed by flow cytometry to measure the fitness values of *V. cholerae* N16961 WT and the ΔSI mutant, relative to an *E. coli* DH5α strain as a common competitor carrying a pSU38 plasmid with the *gfp* under the control of the constitutive promoter P_CS (strain A370, **Supplementary Table 1**). The competition procedure was performed as described in (DelaFuente et al., 2020) with some modifications. Briefly, overnight cultures were prepared by inoculating single colonies from an LB agar plate in 200 µl of liquid LB in a 96-well plate (Nunc, Thermo Scientific, USA). After 22h of growth at 37˚C with 250 rpm shaking, cultures were mixed at 1:1 proportion and diluted 1:400 in fresh medium. In order to confirm the initial proportions of *E. coli* carrying *gfp* and the non-fluorescent *V. cholerae* strains, cells were diluted 1:400 in NaCl 0.9%, and 30,000 events per sample were recorded using a CytoFLEX S flow cytometer (Beckman Coulter). Bacterial mixtures were co-cultured for 22h at 37˚C with 200 rpm shaking, and final proportions were determined as stated above. The fitness values of both *V. cholerae* WT and ΔSI, relative to *E. coli*/pSU38::P_CS *gfp* were calculated using the formula: $w=\ln(N_{final,gfp-}/N_{initial,gfp-})/\ln(N_{final,gfp+}/N_{initial,gfp+})$, where w is the relative fitness of the non-GFP-tagged *V. cholerae* strains, $N_{initial,gfp-}$ and $N_{final,gfp-}$ are the numbers of non-GFP-tagged *V. cholerae* at the beginning and end of the competition, and $N_{inital,gfp+}$ and $N_{final,gfp+}$ are the numbers of *E. coli*/pSU38::P_CS *gfp* cells

at the beginning and end of the competition, respectively. Six biological replicates were performed for each competition experiment, and significant differences were assessed by performing an unpaired t-test.

Natural transformation assays

To quantify the natural competence of the *V. cholerae* N16961 mutants, assays were performed as described above. Cells were incubated with either 2 µg of a PCR product carrying a kanamycin resistance cassette flanked by 550 bp/1 kb/2 kb long homology regions targeting the *lacZ* gene, or the gDNA extracted from strain #135. Primers and strains used for this assay are listed in **Supplementary Table 1** and **Supplementary Table 3**. In these assays, cells were incubated for 8h with tDNA before plating in selective medium (LB plates supplemented with kanamycin 75 µg/mL). Transformation frequencies were calculated as the number of CFUs growing in kanamycin over the total CFUs.

Recombination through natural transformation

Recombination assays through natural transformation were either performed following a classical natural transformation assay, as previously described with some modifications, or with a chitin-independent transformation assay. In the classical natural transformation assay, we introduced a step where the integrase was induced for four hours.

In the chitin-independent assay, the protocol was adapted from Chlebek et al., 2019. Strains were inoculated from frozen stocks into liquid LB medium with glucose 1% and grown overnight at 30°C. Overnight cultures were then diluted 1:100 and grown in the appropriate antibiotics and arabinose 0.2% for four hours. Cells were washed and resuspended in DASW (Instant Ocean, USA) 0.5x and 7 µL were inoculated in a final volume of 350 µL DASW (if needed, 150 µL of chitin were also added). 2 µg of tDNA were added (except for the Phi29 amplified product, where we inoculated the whole amplified product) and cells were incubated in static conditions at 30°C for ~18 h. The following day, 1:10 serial dilutions up to 10^{-7} were performed, and 5 µL of each dilution was plated on LB medium supplemented with the appropriate antibiotics and glucose 1% to recover the total recipient bacteria. To recover transformants or

recombinants through phenotype selection, cells were plated in either chloramphenicol 2.5 µg/mL + glucose 1% or kanamycin 75 µg/mL + glucose 1%. To recover recombinants through blind selection, cultures were plated in LB medium (ØNaCl) + sucrose 10% + zeocin 12.5 µg/mL and incubated at 20°C. To characterize the captured cassettes, we performed a colony PCR to the recombinants using the primers sacB::attI check F and sacB::attI check R.

Concerning the transformed DNA samples, the three tested PCR products from the pMBA collection were amplified with primers IntI seq F II and GFP seq R II. The integron content of the R388 plasmids was amplified with primers IntI seq F II and sul1 R. Finally, the *Vibrio cholerae* WT gDNA (strain A096 **Supplementary Table 1**) was amplified with the EquiPhi29 polymerase (Thermo Scientific, USA) and the GTT R* primer (RYYYA*A*C, where * represents a phosphorothioate bond) (**Supplementary Table 3**). The amplification was performed as recommended by the manufacture with some modifications. We first annealed the primer with the DNA by heating the pool at 95°C for 5 min and letting the temperature drop gradually. Next, the amplification mix was performed as described by the manufacturer with the EquiPhi29 polymerase (Thermo Scientific, USA) and incubated at 37°C for 16h. The resulting amplified product was transformed into competent cells (approximately 10 µg).

Statistical analysis

Statistical analysis was performed using GraphPad Prism 10.1.1 (GraphPad Software). Significant differences were determined either using the Mann-Whitney U test or the Kruskal-Wallis test, with multiple comparisons corrected by Dunnett's test. A significance level of 0.05 was applied in all cases. Significance levels were indicated as follows: $p \geq 0.05$ (ns), $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), and $p < 0.0001$ (****).

Results

Results

Chapter 1: Establishment of a platform to select for integron recombination events

Integrons, agents of evolution and adaptation, have co-existed and evolved with bacteria for millions of years. Their importance for antimicrobial resistance is undeniable, and the reservoir of genes that they encode is potentially infinite. However, as previously reviewed (see page 60), our means to detect integrons are hampered due to the bias of PCRs, the variable nature of integron arrays, as well as the impracticality of deep sequencing methods for routine use.

The first challenge of this PhD thesis was creating a platform that captures integron cassettes (IC) regardless of the phenotype conferred. To achieve this, we aimed to redesign a class 1 integron that serves as a capture device. The readout is provided by a counter-selectable maker (CSM) where the *attI* is embedded (CSM::*attI*), ensuring only the viability of recombinants (**Figure 13**). Prior to recombination, the survival of the cells is granted by the growth in non-toxic conditions (which can vary depending on the CSM used). The choice of the class 1 integron lies in the broad cassette specificity of the IntI1 (Biskri et al., 2005) – capable of recognizing *attC*s from all integron classes.

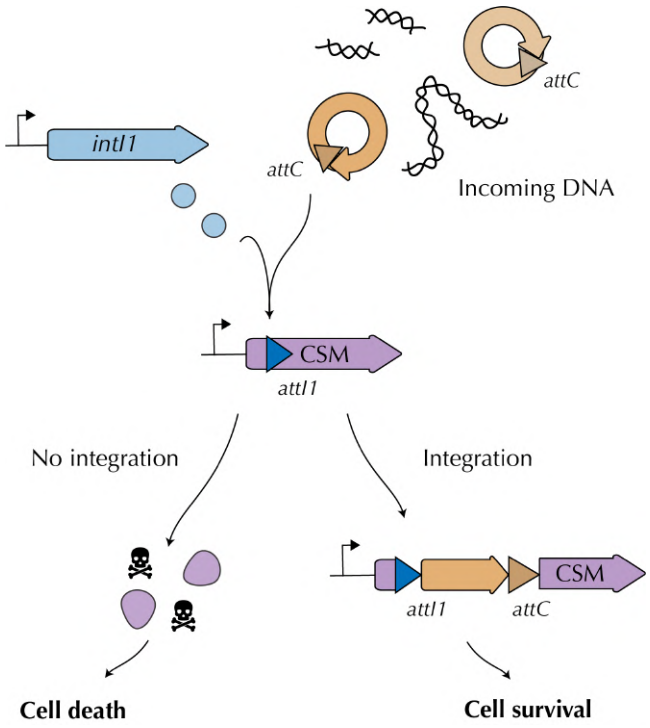


Figure 13. Integron cassette capture platform

Schematic representation of the cassette capture platform. We will re-engineer a class 1 integron to serve as a capture device using a counter-selectable marker (CSM) as a blind reporter. The rationale is that, by embedding an integration site (*attI1*) within the CSM gene, and upon expression of the integrase (*intI1*), incoming cassettes will disrupt the CSM gene so that only recombinants will survive.

This first chapter is divided into three main sections: **1**) the characterization of different promoters in both *Vibrio cholerae* and *Escherichia coli*, as we aimed to design a capture platform that not only works in our main chassis but also in other commonly used bacterial species for biotechnology, such as *E. coli*; **2** and **3**) the creation of two platforms based on two distinct CSMs, the TA system CcdB/CcdA, and the SacB levansucrase. Importantly, we demonstrate the efficiency of these systems by performing two experiments, a **killing assay** and a **classical recombination assay**, thoroughly explained in the methods section (pages 82-86, **Figure 11** and **Figure 12**):

- The **killing assay** allowed us to characterize three parameters: (i) the effectiveness of the CSM::*attI1* fusion proteins, (ii) the limit of detection of the generated tool, and iii) the escape mutants from the system.
- The **classical recombination assay** allows us to ensure that the CSM::*attI1* platform captures cassettes and that we can select for recombination blindly.

1.1 – Characterization of different promoters in *E. coli* and *V. cholerae*

We first needed to understand the activity of both inducible and constitutive promoters in *V. cholerae* and *E. coli*. This is crucial as we plan to use both types of promoters to control the *intI1* gene and the counter-selectable marker, paving the way for the design and finetuning of the capture platform.

1.1.1 – Characterization of constitutive promoters

We decided to characterize three different constitutive promoters: the P_{Lac}, which, although not classified as a constitutive promoter, behaves as one when used without its repressor LacI; and two integron cassette promoters, a weaker version ($P_C W$) and stronger version ($P_C S$). To do so, we designed a plasmid with a medium copy number (p15A origin, 10-15 copies) containing a *gfp* gene as a reporter. The gene was cloned in such a way that the ribosome binding site (RBS) is conserved, with the only differing element being the promoter. The *gfp* gene encodes a green fluorescent protein that allows us to measure fluorescence by flow cytometry. We transformed these plasmids into two *E. coli* strains (DH5α and MG1655), as well as *V. cholerae* N16961, selecting three transformants per strain to serve as biological replicates. We measured fluorescence in stationary phase and included the wild-type (WT) strain to set the baseline value for fluorescence to all strains. We observed that the $P_C W$ was the weakest promoter and $P_C S$ was the strongest in all strains, while P_{Lac} had an intermediate strength (except in *E. coli* MG1655, where the presence of the chromosomal repressor LacI could be interfering). To address this interference and improve the P_{Lac} expression in MG1655, we added IPTG at 200 µg/mL, retrieving a similar expression level as in DH5α (**Figure 14**).

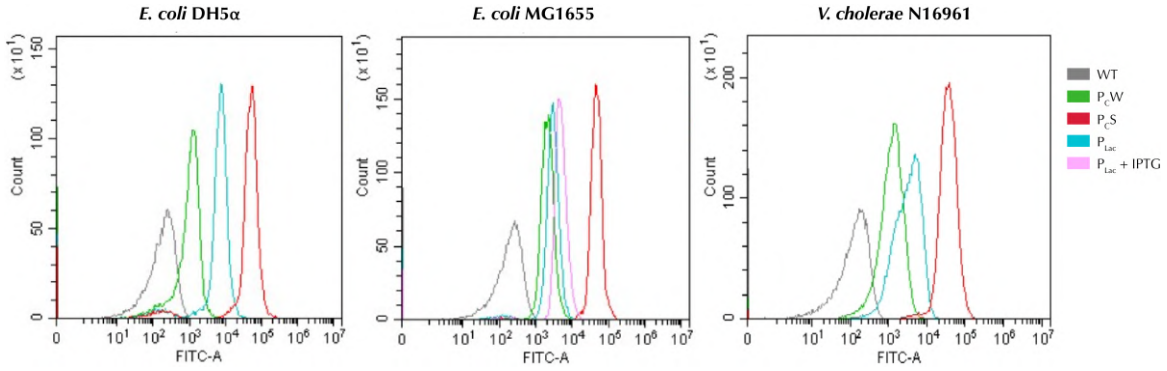


Figure 14. Characterization of constitutive promoters in *E. coli* and *V. cholerae*

Histograms of GFP fluorescence (assessed by flow cytometry) controlled by different constitutive in *E. coli* DH5α, *E. coli* MG1655, and *V. cholerae* N16961. The x-axis represents the GFP fluorescence intensity (FITC-A, arbitrary units), and the y-axis represents the cell count. In grey, the fluorescence intensity of the WT strain is shown; in green, a P_CW *gfp* construction; in red, a P_CS *gfp*; in blue, a P_{Lac} gfp; and in pink, a P_{Lac} *gfp* grown with IPTG 200 µg/mL. Three biological replicates were analyzed and a representative replicate is depicted.

In general, all three promoters seem to work properly in the two targeted species.

1.1.2 – Characterization of inducible promoters

Inducible promoters allow the control of gene expression in response to specific environmental or chemical stimuli, providing a versatile tool for studying gene function and regulation, but also for biotechnological purposes such as ours.

We aimed to characterize the P_{BAD} promoter, a well-known inducible promoter from the arabinose operon of *E. coli*. The expression from the P_{BAD} promoter is triggered in the presence of arabinose and repressed by glucose. It is regulated by the AraC protein, which binds to the promoter region and changes conformation in the presence of arabinose, facilitating RNA polymerase binding and transcription initiation (Guzman et al., 1995). To characterize the P_{BAD} promoter, we used a plasmid that possesses the same origin of replication (p15a) (San Millan et al., 2016) and would allow us to compare the results of the P_{BAD} to the results of the constitutive promoters. GFP fluorescence was measured in inducing conditions, using different concentrations of arabinose: 0.1%, 0.2%, and 0.4%; and in repressive conditions: glucose 1% and LB without any supplements. Again, we measured the fluorescence in three biological replicates. Our results show that, in *E. coli,* glucose is necessary for an efficient repression. In *V. cholerae*, glucose at 1% does not seem to confer the same level of repression, as we observed a modest escape in *gfp* expression. Furthermore, different

levels of arabinose lead to different levels of expression, especially in strain *E. coli* DH5α, where the effects are more marked (**Figure 15**).

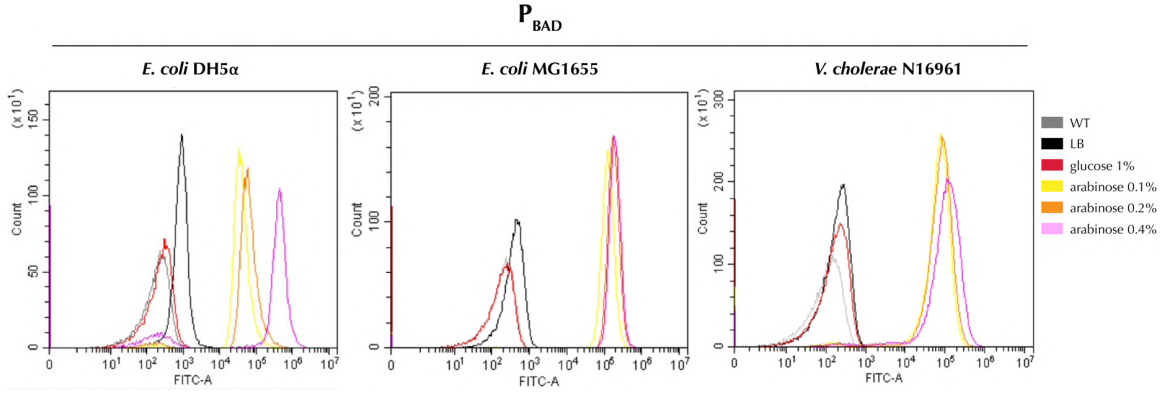


Figure 15. Characterization of the P_{BAD} promoter in *E. coli* and *V. cholerae*

Histograms of GFP fluorescence (assessed by flow cytometry) controlled by the P_{BAD} promoter in *E. coli* DH5α, *E. coli* MG1655, and *V. cholerae* N16961. The x-axis represents the GFP fluorescence intensity (FITC-A, arbitrary units), and the y-axis represents the cell count. In grey, the fluorescence intensity of the WT strain is shown; in black, cells with the pBAD::*gfp* plasmid are grown in LB without any supplements; in red, with glucose 1%; in pink, with arabinose 0.1%; in yellow, with arabinose 0.2%; and in orange, with arabinose 0.4%. Three biological replicates were analyzed and a representative replicate is depicted.

Finally, we also characterized P_{Van} and P_{Tet} promoters in *E. coli* and *V. cholerae*, two well-characterized inducible systems commonly used in bacterial gene expression studies. The P_{Van} promoter is derived from the vanillate operon in *Caulobacter crescentus* and is activated by the presence of vanillate, a derivative of lignin. In the absence of vanillate, the VanR regulator binds to the promoter, preventing transcription. Upon addition of vanillate, VanR undergoes a conformational change that allows RNA polymerase to bind and initiate transcription, resulting in gene expression (Thanbichler et al., 2007). The P_{Tet} promoter, on the other hand, is controlled by the tetracycline repressor protein (TetR) and is responsive to tetracycline or its analogs, such as anhydrotetracycline (aTc). In the absence of aTc, TetR binds to the operator sequences (*tetO*) of the P_{Tet} promoter, blocking transcription. When aTc is present, it binds to TetR, causing it to release from the operator site, thereby allowing the RNA polymerase to access the promoter and initiate gene expression (Orth et al., 2000). We decided to characterize two ameliorated versions of the P_{Tet} and the P_{Van} promoters—called "Marionette" promoters—recently developed in Prof. Voigt's lab (Meyer et al., 2019). In this case, we used the plasmids described in the reference, replacing the *yfp* gene for our *gfp* but maintaining the RBS of the original plasmid. This plasmid also

possesses a p15A origin of replication, allowing us to compare the results between all promoters.

For the $P_{VanM,}$ we grew the cells in repressive conditions (in LB without any supplement) and inducing conditions: 12.5 µM, 25 µM, 50 µM, and 100 µM of vanillate. We measured fluorescence in stationary phase as previously described. The P_{VanM} promoter allowed for a strong induction in all strains, and the different concentrations of vanillate allowed a gradual increase in fluorescence. Still, the repression was not complete in all three strains tested (**Figure 16**).

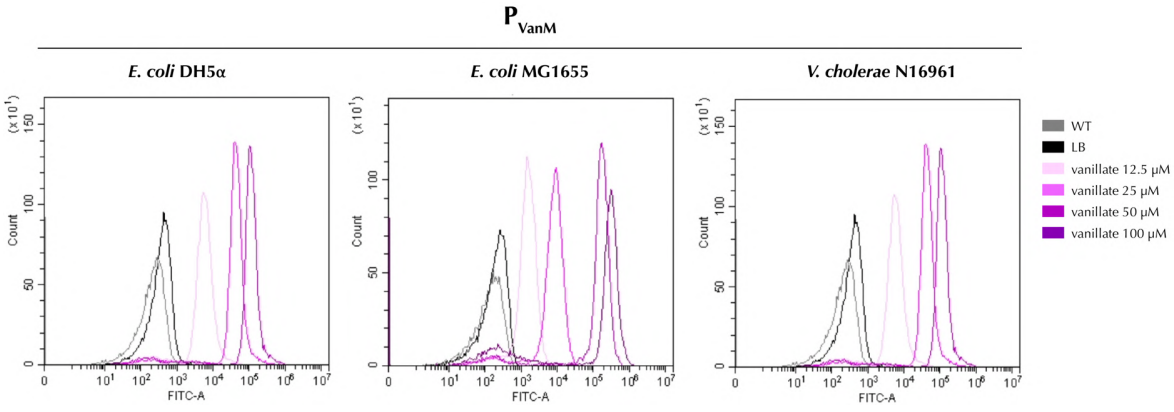


Figure 16. Characterization of the P_{VanM} promoter in *E. coli* and *V. cholerae*

Histograms of GFP fluorescence (assessed by flow cytometry) controlled by the P_{VanM} promoter in *E. coli* DH5α, *E. coli* MG1655, and *V. cholerae* N16961. The x-axis represents the GFP fluorescence intensity (FITC-A, arbitrary units), and the y-axis represents the cell count. In grey, the fluorescence intensity of the WT strain is shown; in black, cells with the pVanM::*gfp* plasmid are grown in LB without any supplement; in light pink, with vanillate 12.5 µM; in dark pink, with vanillate 25 µM; in light purple, with vanillate 50 µM; and in dark purple, with vanillate 100 µM. Three biological replicates were analyzed and a representative replicate is depicted.

For the $P_{TetM,}$ we grew the cells in repressive conditions (in LB without any supplement) and inducible conditions: 100 ng/mL and 200 ng/mL of anhydrotetracycline (aTc). GFP fluorescence was measured as before, in this case only in *E. coli* MG1655 and *V. cholerae* N16961. The P_{TetM} promoter seemed to confer strong repression and an intermediate induction, also increasing with higher concentrations of aTc (**Figure 17**).

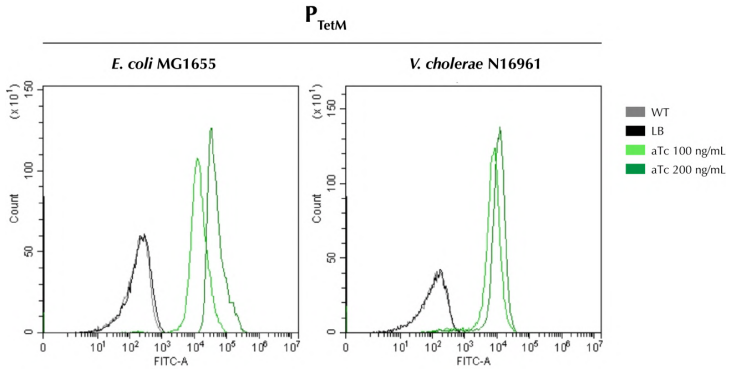


Figure 17. Characterization of the P_{TetM} promoter in *E. coli* and *V. cholerae*

Histograms of GFP fluorescence (assessed by flow cytometry) controlled by the P_{TetM} promoter in *E. coli* MG1655, and *V. cholerae* N16961. The x-axis represents the GFP fluorescence intensity (FITC-A, arbitrary units), and the y-axis represents the cell count. In grey, the fluorescence intensity of the WT strain is shown; in black, cells with the pTetM::*gfp* plasmid are grown in LB without any supplement; in light green, with anhydrotetracycline (aTc) 100 ng/mL; and in dark green, with aTc 200 ng/mL. Three biological replicates were analyzed and a representative replicate is depicted.

In *E. coli*, both P_{BAD} and P_{TetM} promoters seem to work efficiently. In *V. cholerae*, only the P_{TetM} seems to confer an effective repression and induction. The P_{VanM}, although ensuring a high expression in both species, remains the poorer in repression.

1.1.3 – Characterization of inducible promoters through integron recombination

To deepen the inducible promoter characterization, we decided to test integron recombination. This assay is crucial not only because the integrase can become toxic if not properly repressed (and thus, the assay would provide more insights into the correct repression of the promoters) but also because we will need to employ one of these inducible promoters to control *intI1* in our tool (see **Figure 13**).

For this experiment, we used a plasmid with a ColE1 origin of replication containing the *intI1* gene controlled by the P_{BAD} promoter. Concerning the P_{VanM} and P_{TetM}, we replaced the *yfp* for the *intI1* gene, maintaining all the other characteristics of the previously described plasmids. We then transformed these three plasmids in an *E. coli* Top10 strain containing a plasmid with an *attI1* site, generating our three recipient strains, and performed a classical recombination assay (Bouvier et al., 2005) (see methods, **Figure 12**). This assay utilizes conjugation, which involves ssDNA transfer, to deliver the *attC* site in single-stranded form to the tested recipient cells that express the *intI1* and contains the *attI1* recombination site. The *attC* site transferred via conjugation is carried on a pSW plasmid, which cannot replicate in the recipient cell, as

it is *pir*-dependent (Demarre et al., 2005) (**Figure 18A**). This system allows the delivery of an *attC* site ssDNA form and the maintenance of the *attI* site in dsDNA form, ensuring a maximal efficiency of integron recombination. Recombinants are selected by phenotype, as the integrated plasmid contains a chloramphenicol resistance gene. The recombination frequency is calculated as the ratio of the colonies growing in chloramphenicol and the total colony-forming units (CFU). We obtained recombination frequencies of 2.3×10^{-1} for the P_{BAD} promoter in induced conditions (arabinose 0.2%) and 1.3×10^{-6} in repressed conditions (glucose 1%). These results are consistent with those previously described in the literature (Demarre et al., 2007) and show that the P_{BAD} promoter is very efficient in the induction and repression of the *intI1* gene. For the P_{VanM} promoter, we obtained a recombination frequency of 1.1×10^{-1} in inducing conditions (12.5 µM) and 1.9×10^{-2} in repressing conditions (no supplement added). These results parallel those obtained with flow cytometry and suggest that this promoter is not very efficient for gene repression but still efficient for induction. Finally, the P_{TetM} promoter led to a recombination frequency of 10^{-4} in both inducing conditions (aTc 200 ng/mL) and repressing conditions (no supplement added). These results are not compatible with our characterization by flow cytometry. Even after re-sequencing the plasmid, we were unable to explain why both induction and repression were so poorly achieved with this promoter (**Figure 18B**). At this point, we could not characterize the activity of the integrase in *V. cholerae* N16961 due to the potential interference of the superintegron (for more details, see Chapter 4: *Vibrio cholerae* as a chassis to test directly DNA samples).

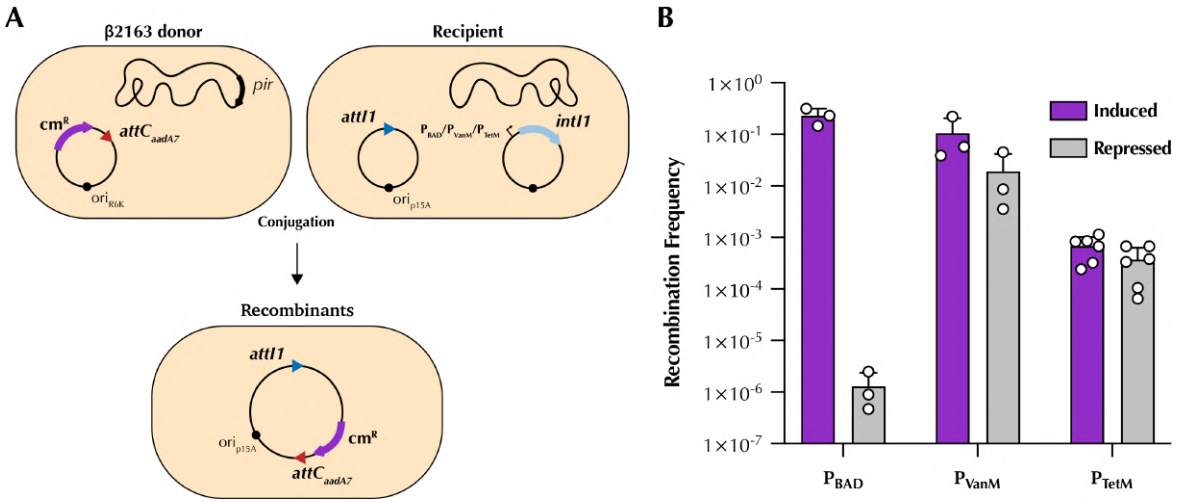


Figure 18. Characterization of the inducible promoters in *E. coli* through integron-mediated recombination

A. Schematic representation of the classical recombination assay. The pSW23T::$attC_{aadA7}$ suicide vector is delivered to the recipient strains containing an integrase-expressing vector and a pSU38::*attI1*. As pSW23T (*pir* dependent) cannot replicate in the recipient strains, recombinant clones can be selected with chloramphenicol. **B.** Classical recombination assay testing different promoters (P_{BAD}, P_{VanM}, and P_{TetM}) controlling the *intI1* gene in induced and repressed conditions. Recombinants were selected by phenotype (chloramphenicol resistance). The recombination frequency was calculated as the ratio of the number of CFUs growing in chloramphenicol and the total CFUs. Bar charts show the mean of the recombination frequency ± s.d. from biological replicates (individual dots).

1.2 – CcdB/CcdA type II toxin-antitoxin system as a reporter

After the successful characterization of promoters in the two targeted species as a chassis for our tool, we sought to find a suitable counter-selectable marker for integron recombination. The first CSM we evaluated was the CcdB/CcdA toxin-antitoxin system. This system belongs to the type II family of TA systems, where CcdB is the toxin and CcdA the antitoxin. CcdB binds to the DNA gyrase A subunit, binding to the cleaved complex and inhibiting the passage of the DNA and RNA polymerase, thus inducing cell death by DNA damage. CcdA, the labile antitoxin, when present, binds to CcdB, altering its conformation and preventing its binding to the DNA gyrase. In its original host, the F-plasmid of *Escherichia coli*, this system ensures plasmid stability by killing cells that lose the plasmid, thereby preventing thereby preventing the emergence of plasmid-free cells (Bernard & Couturier, 1992). CcdB has been widely used in biotechnological applications (Bernard, 1996; Szpirer & Milinkovitch, 2005; Val et al., 2012; Zhang et al., 2017) and displays a low frequency of spontaneous escape mutants.

Embedding a 61 base pair (bp) *attI1* site within the small *ccdB* gene (composed of 318 bp) could be considered a meticulous task. However, we took advantage of a

recent study (López-Igual et al., 2019), where several toxins were strategically split to introduce an intein. An intein is a protein segment that can excise itself from a larger protein while joining the remaining portions (exteins) together through a precise splicing reaction (Perler, 2002). Conversely, in our approach, the *attI1* site has to be translated and folded as well. Knowing that *ccdB* from *Vibrio fischeri* is functional in both *E. coli* and *V. cholerae* (López-Igual et al., 2019), we inserted the *attI1* site sequence in frame with the *V. fischeri ccdB* gene, incorporating 21 amino acids between K49 and P50 (**Supplementary Figure 1**). Utilizing AlphaFold3, we predicted the structures of the WT protein (**Figure 19A**) and the CcdB::*attI* fusion protein (**Figure 19B**). Notably, the *attI1* sequence was inserted within a substantial linking region. Nevertheless, our folding predictions did not confidently forecast the folding pattern of the fusion protein.

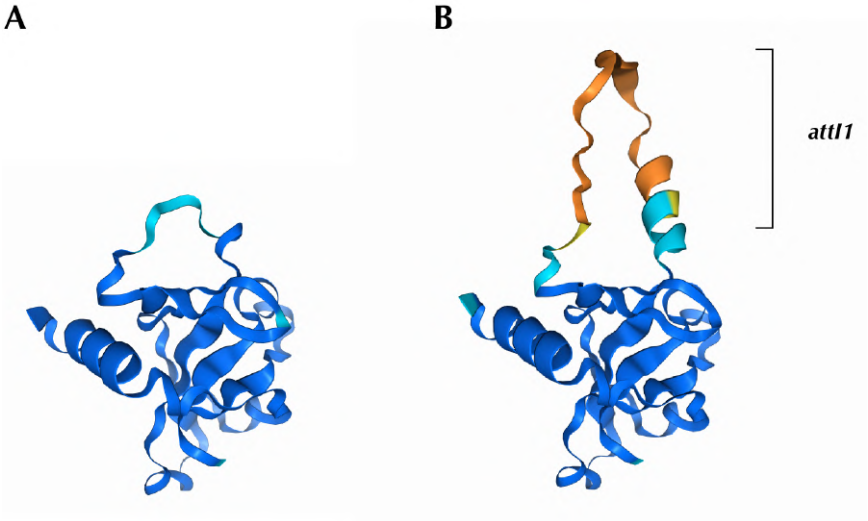


Figure 19. Predicted 3D structures of $CcdB_{wt}$ and CcdB::attI fusion protein
A. Predicted structure of wild-type CcdB protein from *Vibrio fischeri*. **B.** Predicted structure of CcdB::attI fusion, highlighting the integration site and altered folding. Models were generated using AlphaFold3. Blue indicates high confidence in folding prediction, and red indicates low confidence.

1.2.1 – Plasmid-based tool using CcdB/CcdA TA system as a reporter

To assess the functionality of the fusion protein, we synthesized (IDT, USA) and cloned *ccdB::attI1* in a pColE1 plasmid controlled by the P_{BAD} promoter. Despite the successful cloning, we quickly noticed the appearance of high numbers of escape mutants (**Supplementary Figure 2**), as previously described (López-Igual et al., 2019). This first experiment showed us two important factors: *ccdB::attI* is toxic, and the addition of the antitoxin (*ccdA*) is crucial, thus adding a layer of complexity to the control of the platform. To test this hypothesis, we cloned *ccdB::attI1* under the control

of either the P_{Lac} or the $P_{C}W$ in a p15a plasmid. We chose to control *ccdA* by having it under the regulation of a P_{BAD} in a pColE1 plasmid. Finally, in this setup, the integrase (*intI1*) was cloned with the P_{VanM} promoter, which, although not very repressive, showed good performance in the induction of integron recombination (see previous results, **Figure 18**). These constructions were cloned in *E. coli* DH5α (**Figure 20A**). To assess both the impact of adding an *attI* into the *ccdB* sequence and the potential background noise of the system, we performed a killing assay. Briefly, cells were grown overnight in "survival" conditions, in this case with arabinose 0.2%. The next day, cells were diluted and plated in both "survival" conditions, to count the total CFUs and in "killing" conditions, with glucose 1%, to count the number of escape mutants. We then calculated the escape mutant frequency and compared it to the WT toxin, cloned in an identical setup. We obtained escape mutant frequencies of 6.8×10^{-6} when *ccdB::attI1* was controlled by the P_{Lac} and 9.0×10^{-5} when *ccdB::attI1* was controlled by the $P_{C}W$ (**Figure 20B**). The frequencies were also similar between both *ccdB::attI1* and *ccdB*, showing that adding an *attI* site in this particular position does not affect the efficacy of the toxin. The escape mutants were analyzed by PCR and Sanger sequencing. The vast majority were due to ISs insertions in the *ccdB* gene (92%). With this experiment, we show that the CcdB/CcdA TA system as a CSM provides a broad range to detect integron cassettes.

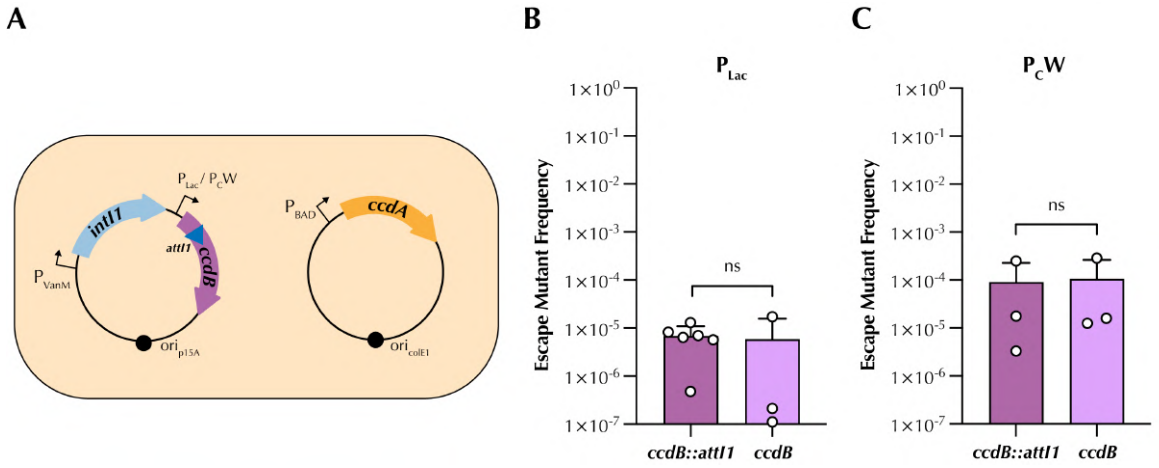


Figure 20. Plasmid setup and killing assay *ccdB::attI1*

A. Schematic representation of the integron cassette capture platform set up in a plasmid. *ccdB* was cloned either under the control of a P_{Lac} or P_CW in a p15a plasmid, together with *intI1*, controlled by the P_{VanM} promoter. *ccdA* is cloned in a pColE1 plasmid under the control of the P_{BAD} promoter. **B and C.** Killing assays of the two different plasmid setups containing the integron cassette capture platform (either with *ccdB* controlled by the P_{Lac} promoter or the P_CW) in *E. coli* DH5α compared to the WT gene. Cells were grown in either survival conditions (with arabinose 0.2% that induces the expression of *ccdA*) or killing conditions (with glucose 1% that represses the expression of *ccdA*). The escape mutant frequency was calculated the ratio of the CFUs, growing in glucose, and the total CFUs, growing in arabinose. Bar charts show the mean of the escape mutant frequency ± s.d. from biological replicates (individual dots). Statistical significance was determined using the Mann-Whitney U test; ns, not significant.

Next, we sought to demonstrate that the insertion of an integron cassette is sufficient to disrupt the *ccdB* gene. To do so, we cloned eight different cassettes in *ccdB::attI1*, mimicking an integration event. We chose cassettes of different sizes, conferring resistance to different antibiotic families, and we included cassettes that left the *ccdB::attI1::cassette* resulting fusion protein in frame. As before, we performed a killing assay, plating strains in both survival and killing conditions. Our results show that all cassettes, when interrupting the *ccdB* gene, are enough to arrest the activity of the toxin (**Figure 21**).

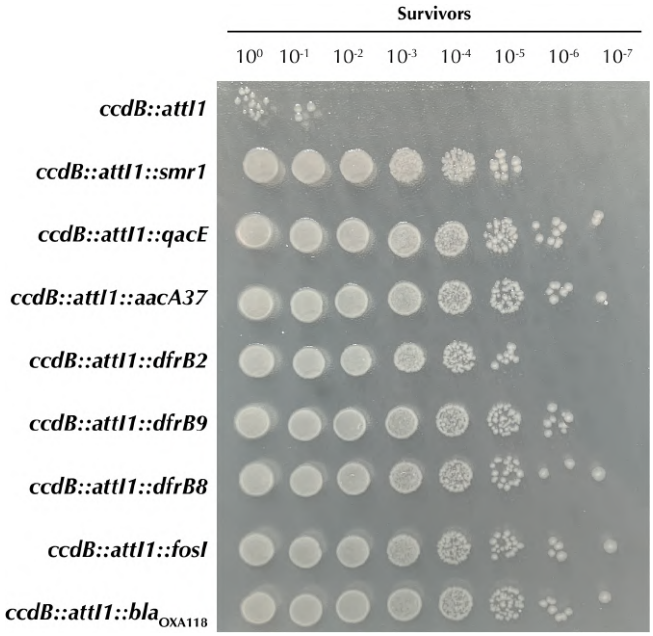


Figure 21. Killing assays of *ccdB::attI1::cassette*

The strains containing the various *ccdB::attI1::cassette* combinations were plated in 5 µL spots in dilutions up to 10^{-7}, in killing conditions (LB + carbenicillin + kanamycin + glucose 1%).

Having established the basis for the tool based on the CcdB/CcdA TA system, we proceeded to characterize the platform's ability to **detect recombination events independently of the phenotype conferred by the cassette**. For that, we performed a classical recombination assay (see Methods page 84) using the *E. coli* DH5α containing the P_{Lac} *ccdB::attI1* platform (**Figure 22A**) as a recipient. The assay was performed using the same Cm^R suicide plasmid mimicking an integron cassette, but recombinants were selected either by phenotype or blindly, plating cells in "killing" conditions (i.e., in glucose 1%). As a control, we used a knocked-out (KO) version of *ccdB::attI1*, where recombinants could only be selected by phenotype. When selecting by chloramphenicol resistance, we obtained recombination frequencies of 2.2×10^{-1} for the KO version (*ccdB$_{KO}$::attI1*) and 5.8×10^{-2} for *ccdB::attI1*. However, the recombination frequency dropped to 1.0×10^{-3} when selecting blindly, in "killing" conditions (**Figure 22B**). This was probably attributable to the dominance (Rodríguez-Beltrán et al., 2020) of *ccdB*: due to the multicopy nature of the plasmid carrying the platform, only a few *attI1* integrate the cassette. This means that in the majority of the cells (~80%), empty copies of *ccdB::attI1* still induce death. As the resistance to chloramphenicol can be conferred by just a single copy of the gene, this selection is more efficient. Nevertheless, we demonstrate that the CcdB/CcdA TA system as a CSM

allows for the blind selection of integron recombination events, although not 100% efficient.

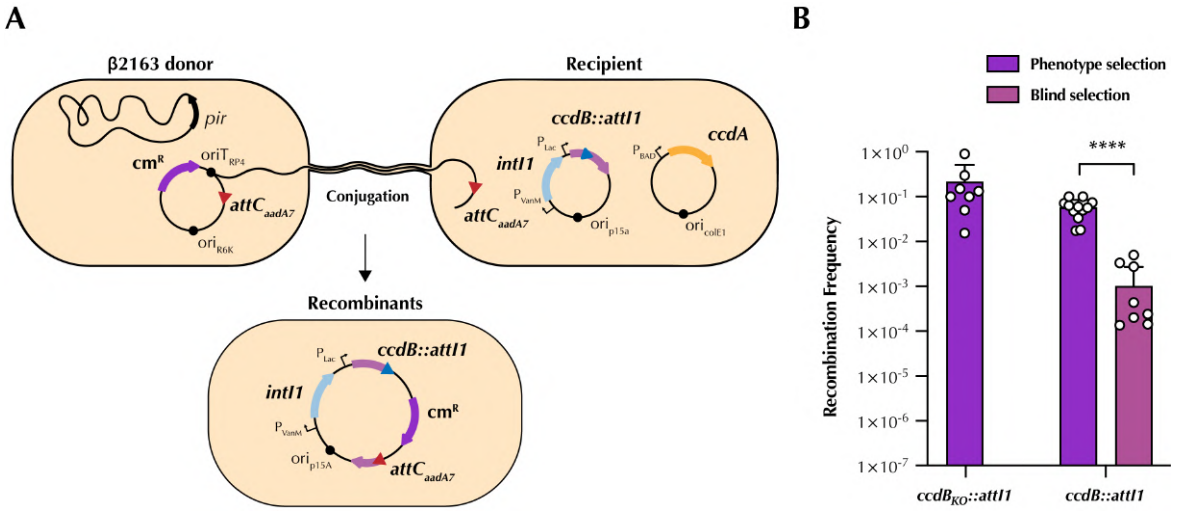


Figure 22. Classical recombination assay of P_{Lac} *ccdB::attI1* in a plasmid setup

A. Schematic representation of the classical recombination assay testing the plasmid setup containing the integron cassette capture platform in *E. coli* DH5α compared to the KO version. **B.** Recombination frequencies. Recombinants were selected by phenotype (chloramphenicol resistance) or blind selection (growth in glucose). The recombination frequency was calculated as the ratio of the number of recombinants, and the total CFUs, growing arabinose. Bar charts show the mean of the recombination frequency ± s.d. from biological replicates (individual dots). Statistical significance was determined using the Mann-Whitney U test; ns, not significant. Statistical significance was determined using the Mann-Whitney U test; ****, $p < 0.0001$.

1.2.2 – Chromosome-based tool using CcdB/CcdA type II TA system as a reporter

We proved that the CcdB/CcdA TA system works as a blind integron capture platform but has limitations in a high-copy number plasmid setup. We deemed it essential to try a monocopy system by cloning *ccdB::attI1* in the chromosome. Using the same setup, we introduced both P_{BAD} *ccdA* and P_{Lac} *ccdB::attI1* in the *attB* site of the λ phage located in the *E. coli* MG1655 chromosome (**Figure 23A**). We chose this location as it assumed that it is innocuous for the cell. By performing a killing assay as previously described, we proved that the setup works in the chromosome and has a low escape mutant frequency of 2.8 x 10^{-5}, comparable to the WT toxin (4.7 x 10^{-5}) (**Figure 23B**). Here, the dominant effect of *ccdB* can also be seen: the escape mutant frequency is higher than in a plasmid, as the cell only needs one mutation to neutralize the *ccdB* gene.

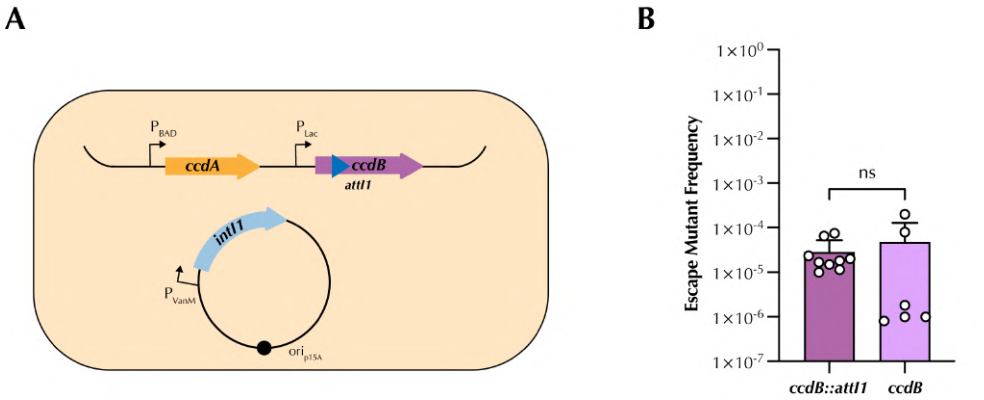


Figure 23. Chromosome setup and killing assay of *ccdB::attI1*

A. Schematic representation of the integron cassette capture platform set up in the chromosome. *ccdB* was cloned under the control of a P_{Lac}. *ccdA* is cloned under the control of the P_{BAD} promoter. The *intI1* was cloned in a p15a plasmid controlled by the P_{VanM} promoter. **B.** Killing assay of the chromosome construction cloned in *E. coli* MG1655 compared to the WT gene. Cells were grown in either survival conditions (with arabinose 0.2% that induces the expression of *ccdA*) or killing conditions (with glucose 1% that represses the expression of *ccdA*). The escape mutant frequency was calculated the ratio of the CFUs, growing in glucose, and the total CFUs, growing in arabinose. Bar charts show the mean of the escape mutant frequency ± s.d. from biological replicates (individual dots). Statistical significance was determined using the Mann-Whitney U test; ns, not significant.

We performed the same classical recombination assay testing the chromosome construction (**Figure 24A**). We compared it to an "*attI*1" control strain that contains the integron recombination site cloned in a plasmid and the *intI1* also controlled by a P_{VanM} promoter. When selecting by phenotype, we show that both strains present a recombination frequency of 10^{-1}. When selecting blindly, we achieve 8.4 x 10^{-2} recombination frequency, comparable to selecting by phenotype (**Figure 24B**). Indeed, a monocopy version of the platform is essential to achieve full efficiency of the blind selection. Surprisingly, it does not seem to entail any loss in the recombination frequency, suggesting that one copy of the *attI1* is enough.

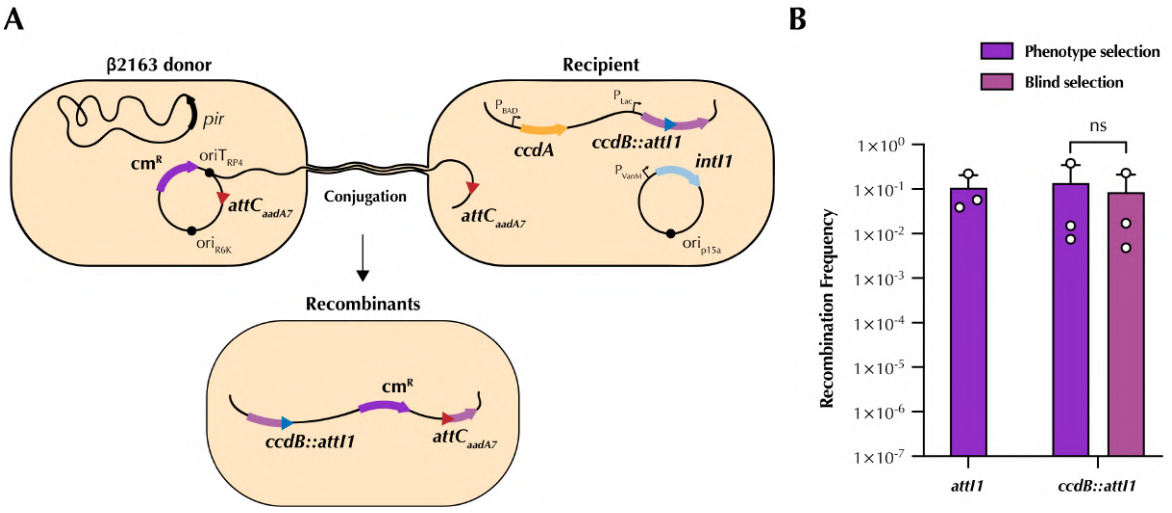


Figure 24. Classical recombination assay of *ccdB::attI1* in a chromosome setup

A. Schematic representation of the classical recombination assay testing the chromosome setup containing the integron cassette capture platform in *E. coli* MG1655 compared to an *attI* version. **B.** Recombination frequencies. Recombinants were selected by phenotype (chloramphenicol resistance) or blind selection (growth in glucose). The recombination frequency as the ratio of the number of recombinants, and the total CFUs, growing arabinose. Bar charts show the mean of the recombination frequency ± s.d. from biological replicates (individual dots). Statistical significance was determined using the Mann-Whitney U test; ns, not significant.

The *ccdB::attI1* platform, both chromosome- and plasmid-based, is protected by patent ES 2 969 666 A1.

Despite our efforts to use the CcdB/CcdA TA system as a counter-selectable marker in *Vibrio cholerae*, we faced significant challenges that prevented it from working as intended. We tried several constructions, including those characterized in *E. coli*, but none worked as expected, either due to the inability to clone them or to the high frequency of escape mutants (above 10^{-3}). These difficulties were likely a result of the lower efficiency of the inducible promoters in this species (see previous section). In another hand, this showed us the need to find and test other counter-selectable markers. Therefore, we decided to explore alternative systems, to ensure the efficiency and reliability of our tool in our main targeted chassis, *V. cholerae*.

1.3 – SacB as a reporter

The second counter-selectable marker we aimed to evaluate for our tool was the *sacB* gene from *Bacillus subtilis*. *sacB* encodes a levansucrase, an enzyme that catalyzes the hydrolysis of sucrose to produce levan, a fructose polymer (Meng & Fütterer, 2003). In *B. subtilis*, levansucrase plays a role in the levan synthesis, which, by forming a capsule, potentially plays a role in protecting against environmental stresses and in

favoring nutrient uptake (Paton, 1960). However, in *Vibrio cholerae*, *Escherichia coli,* and other Gram-negative bacteria, the expression of *sacB* in the presence of sucrose leads to the accumulation of levan, whose toxicity leads to cell death. This toxicity is due to the disruption of cellular metabolism and osmotic balance. *sacB* has been extensively employed in genetic manipulation and other molecular biology applications (Blomfield et al., 1991; Gay et al., 1985; Marx, 2008; Schweizer, 1992) due to its high efficiency as a counter-selectable marker and low frequency of spontaneous escape mutants. One of the advantages of using *sacB* is that this counter-selectable marker is usually expressed from its own constitutive promoter, diminishing the complexity of the tool. Also the toxicity is reduced, due to the fact that in the LB medium, there is no sucrose, and therefore the production of levansucrase does not pose a problem in such conditions.

Using the same approach as with CcdB, we predicted the structures of the WT protein (**Figure 25A**) and three potential SacB::*attI1* fusion proteins: SacB::*attI1*$_{300}$ (**Figure 25B**), SacB::*attI1*$_{203}$ (**Figure 25C**), and SacB::*attI1*$_{403}$ (**Figure 25D**). As before, we tried to introduce the *attI1* sequence within substantial linking regions, away from the catalytic residues (the center of the protein) (**Supplementary Figure 1**).

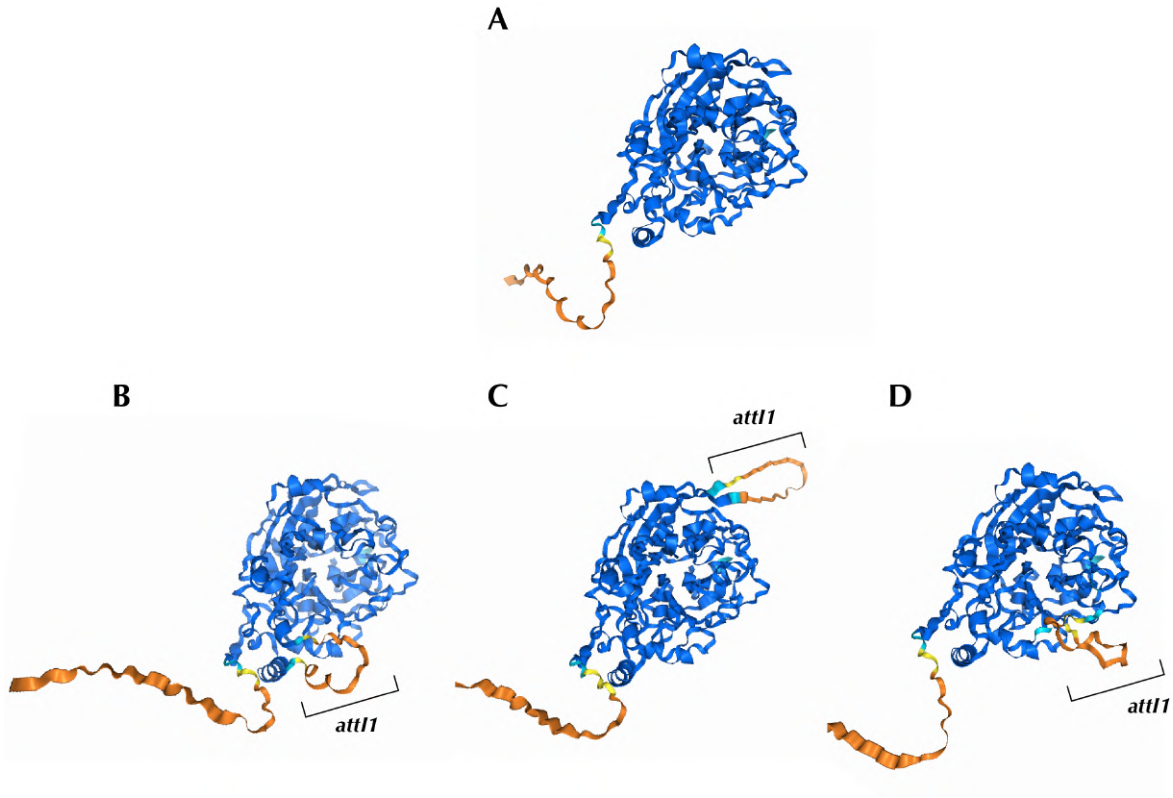


Figure 25. Predicted 3D structures of SacB$_{wt}$ and SacB::*attI1* fusion proteins
A. Predicted structure of wild-type SacB protein from *Bacillus subtillis*. **B to D.** Predicted structures of SacB::*attI* fusions, SacB::*attI1*$_{300}$ (A), SacB::*attI1*$_{203}$ (B), and SacB::*attI1*$_{403}$ (C), highlighting the integration *attI1* site and altered folding. Models were generated using AlphaFold3. Blue indicates high confidence in folding prediction, and red indicates low confidence.

Following the results obtained with the CcdB/CcdA TA system as a counter-selectable marker, we introduced by the *attI1* site SOE-PCR in the three selected locations of the *sacB* gene and cloned them in the chromosome 1 of *V. cholerae*, choosing the Tn7 region of this strain. Apart from being a commonly used region to clone genes to be expressed in *trans* in *V. cholerae*, the rationale for selecting this particular position is that it possesses an *attI* site within the miniTn7 cloned to reconstitute the knocked-out *hapR* gene (Meibom et al., 2005). The *attI1* site was introduced as part of the miniTn7 resistance marker (as it is an integron cassette), but we also reasoned that we could use it as a control for recombination. So, in this case, we substituted the *attI1* and the resistance marker with each of our three *sacB::attI1* versions. We performed a killing assay to test the three potential versions of the *sacB::attI1* fusion. Briefly, cells were grown overnight in "survival" conditions, in this case, LB medium. The next day, cells were plated in serial dilutions in both "survival" conditions to count the total CFUs and in "killing" conditions, LB without NaCl (ØNaCl) with sucrose at 10%, incubated at 20°C (Fullner & Mekalanos, 1999; Jaskólska et al.,

2022; Meibom et al., 2004). The SacB::*attI1*$_{300}$ and the SacB::*attI1*$_{203}$ showed a low number of survivors. The SacB::*attI1*$_{403}$ version had no observable killing, suggesting that the enzyme's activity is impaired by the introduction of the *attI1* in this position (**Figure 26**). Based on these results, we decided to proceed with the *sacB::attI1*$_{300}$ version, henceforth termed *sacB::attI1*, for the sake of simplicity.

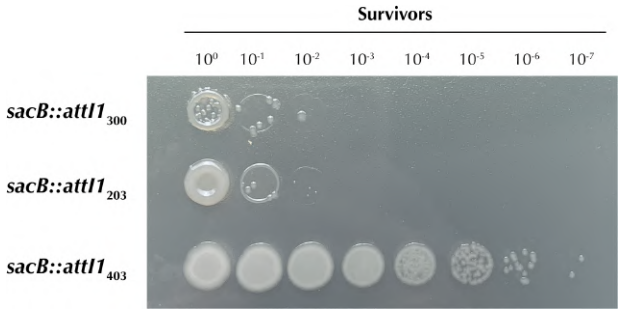


Figure 26. Survivors in a killing assay testing the three SacB::*attI1* versions
The three *sacB::attI1* three versions were plated in 5µL spots in dilutions up to 10^{-7}, in killing conditions in LB (ØNaCl) + sucrose 10%.

To further characterize *sacB::attI1* and ensure that the selective medium is not toxic *per se*, we also performed the same experiment with a WT strain, i.e., a strain that does not contain any *sacB* construct. In this setup, we introduced the *intI1* controlled by the P_{BAD} promoter in a pColE1 plasmid (**Figure 27A**). We show that cells achieve the same levels of counts in both media and that the decrease in *sacB::attI1* strain is ≥5 logs in LB (ØNaCl) + sucrose 10%, with low levels of escape mutants (1.5 x 10^{-6}), comparable to the WT *sacB* (3.1 x 10^{-6}) (**Figure 27B**). When we sequenced the escape mutants, we found a variable collection of mutations that led to the inactivation of *sacB,* such as insertions, deletions, and point mutations, leading to frameshifts and the premature appearance of stop codons, as well as promoter deletions.

Having this system working in *V. cholerae*, we wondered whether we could also establish it for *E. coli* MG1655. We cloned the first version of *sacB::attI1* in the *attB* site and performed the same experiments described above. In *E. coli*, cells plated in LB (ØNaCl) + sucrose 10% were incubated at 30°C (Blomfield et al., 1991). We show that the system is also effective in *E. coli*, with escape mutant frequencies of 2.9 x 10^{-5} (**Figure 27C**), similar to what we described for the CcdB/CcdA TA system. We also showed that the selective medium is not toxic to the WT cells.

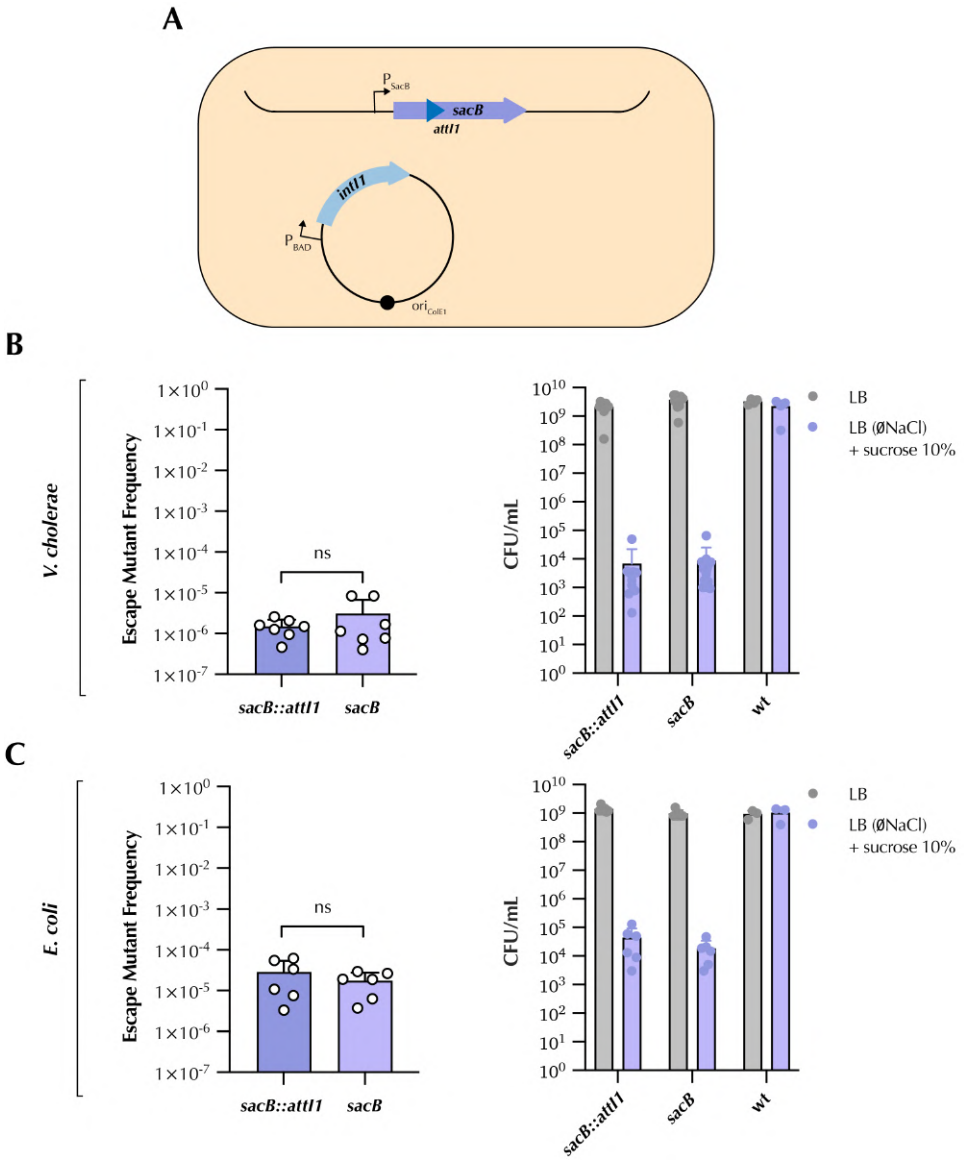


Figure 27. Chromosome setup and killing assay *sacB::attI1*

A. Schematic representation of the integron cassette capture platform set up in the chromosome. *sacB* was cloned with its own regulatory region (promoter + RBS). The *intI1* was cloned in a pColE1 plasmid controlled by the P_{BAD} promoter. **B.** CFUs counts of the *V. cholerae* Tn7::*sacB::attI* mutant compared to the WT strain in both LB medium and LB (ØNaCl) + sucrose 10%. Bar charts show the mean of the CFUs ± s.d. from biological replicates (individual dots). **C and D.** Killing assays of the chromosome constructions cloned in *V. cholerae* N16961 and *E. coli* MG1655 compared to the WT gene. Cells were grown in either survival conditions (standard LB medium) or killing conditions (LB (ØNaCl) + sucrose 10%). The escape mutant frequency was calculated as the ratio of the number of CFUs growing in LB (ØNaCl) + sucrose 10%) and the total CFUs, growing in LB. Bar charts show the mean of the escape mutant frequency ± s.d. from biological replicates (individual dots). Statistical significance was determined using the Mann-Whitney U test; ns, not significant.

After establishing *sacB::attI1* as an efficient counter-selectable marker in both *V. cholerae* and *E. coli*, we tested if blind selection for integron recombination was possible. We performed the same classical recombination assay as for the CcdB/CcdA TA system (see **Figure 18A**) using *sacB::attI1* strains as recipients (**Figure 27A**). As previously mentioned, *V. cholerae* possesses a superintegron in its chromosome 2. This platform was deleted before these experiments to avoid interference (see chapter 4).

We show that in both species, blind selection is slightly less efficient than phenotype selection, with levels of $10^{-3}/10^{-4}$. In fact, when we performed confirmation PCRs, we had more false positive results when selecting by phenotype than when selecting blindly. Although the recombination frequencies shown are corrected with the PCRs results, this could explain the statistical difference between them. Moreover, the recombination frequencies are lower than what was described before with the CcdB/CcdA TA system, but similar to the *attI1* control (**Figure 28**).

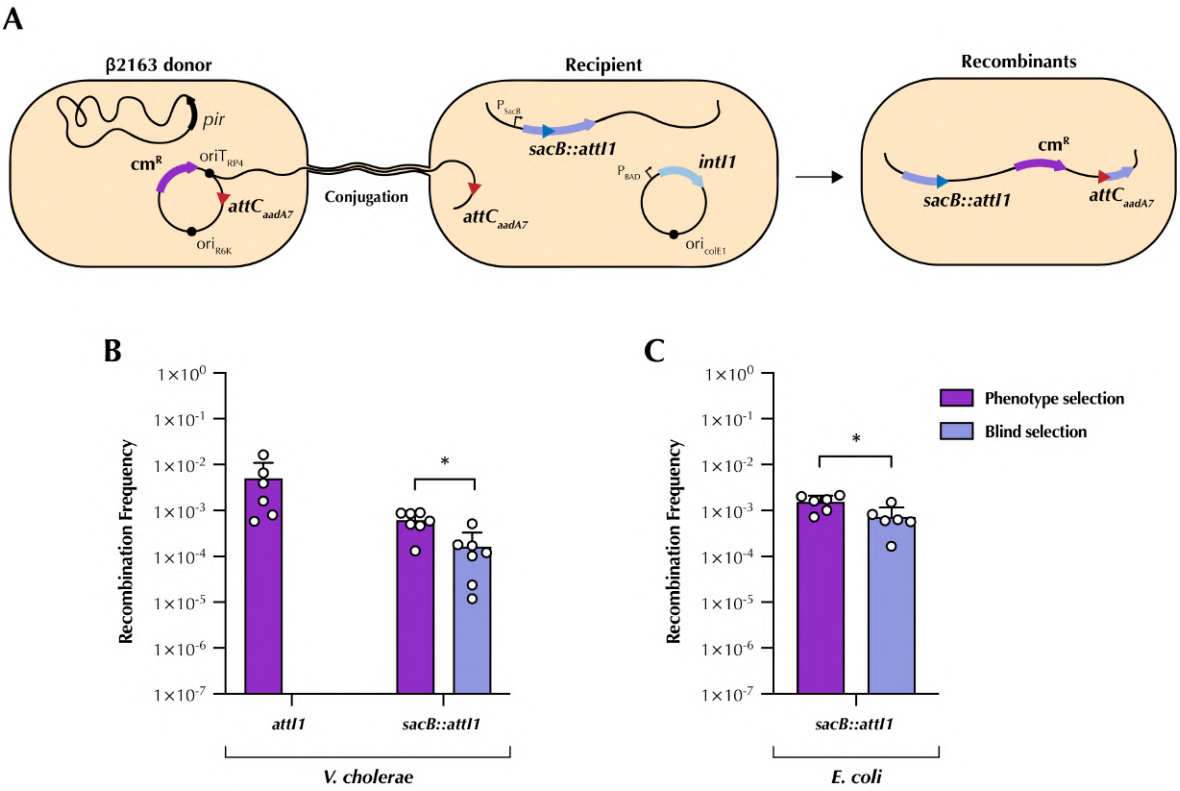


Figure 28. Classical recombination assay of *sacB::attI1* in a chromosomal setup

A. Classical recombination of the plasmid setup containing the integron cassette capture platform in *V. cholerae* N16961 or *E. coli* MG1655. **B.** Recombination frequencies. In both cases, recombinants were selected by phenotype (chloramphenicol resistance) or blind selection (growth in sucrose). In *V. cholerae*, we compared the results to an *attI* version of the same construction, where we can only select by phenotype. The recombination frequency was calculated as the ratio of the number of recombinants and the total CFUs, growing in LB. Bar charts show the mean of the recombination frequency ± s.d. from biological replicates (individual dots). Statistical significance was determined using the Mann-Whitney U test; *, $p < 0.05$.

The *sacB::attI1* platform protection is protected by patent ES 2 991 744 A1.

In this chapter, we successfully establish a platform for selecting integron recombination events, addressing the significant challenge of detecting integron cassettes beyond the phenotype conferred and, potentially, its genetic context. We characterized the CcdB/CcdA toxin-antitoxin system as a counter-selectable marker for

this capture platform and identified its limitations in *V. cholerae*. Despite these challenges, the CcdB/CcdA system remains a valuable tool in *E. coli*, where it can be effectively utilized for detecting ICs. This led us to explore the *sacB* gene as an alternative counter-selectable marker, demonstrating its potential through preliminary tests. These foundational experiments pave the way for developing robust and reliable biotechnological tools for integron cassette capture. In the following chapters, we will describe the various applications we tested with these systems, highlighting their practical utility and effectiveness in diverse biological contexts.

Chapter 2: The cassette harvester

After the successful establishment of a platform to select for integron recombination events, we aimed to tackle one of the most challenging problems in the field of integrons—studying the cassettes encoded in sedentary chromosomal integrons (SCI). Understanding the functions encoded in cassettes of SCIs is essential for comprehending bacterial adaptation and survival mechanisms. These integrons contain numerous gene cassettes that can encode a variety of functions, many of which are still unknown. The variability of genes encountered in these sedentary platforms is such that it is commonly assumed they contribute significantly to the bacteria's ability to adapt to different environments (for a more comprehensive review, see pages 58 and 65).

Although modern techniques would allow the synthesis and cloning of all the genes encoded in a specific genome, having a simple and affordable tool capable of producing cassette libraries would be of extreme utility. This section details the construction and analysis of a chromosomal integron cassette library, aiming to overcome the limitations of current methodologies. By creating a comprehensive library from the *Vibrio cholerae* N16961 superintegron, this application of our tool seeks to provide a robust resource for studying the diversity and functionality of these genetic elements.

2.1 – Plasmid setup

To establish a cassette library of a given SCI of interest, we took advantage of the plasmid-based tool developed in Chapter 1 (see **Figure 20** and **Figure 22**). The reasoning behind this selection is the possibility of using the plasmidic library in other genetic contexts. The system was modified so that the library would be usable to study the captured cassettes without interference from the *intI1* and *ccdA* genes. To avoid this interference, we cloned *ccdB::attI1* in a p15A plasmid (where the library will be created), and the *intI1* and *ccdA* in a pColE1 plasmid, which is then easily removed (**Figure 29A**). The rationale of this experiment is that these two plasmids can be transformed by electroporation in the targeted bacteria containing the SCI of interest. Upon the induction of the *intI1* gene, the tool is activated, i.e., *ccdA* is repressed and

ccdB induced. The expected result will be the survival of bacteria containing a cassette inserted in the *ccdB::attI1* reporter. This plasmid can be extracted from the survival population, treated as a library of SCI cassettes, and further transformed in other genetic backgrounds (**Figure 29B**).

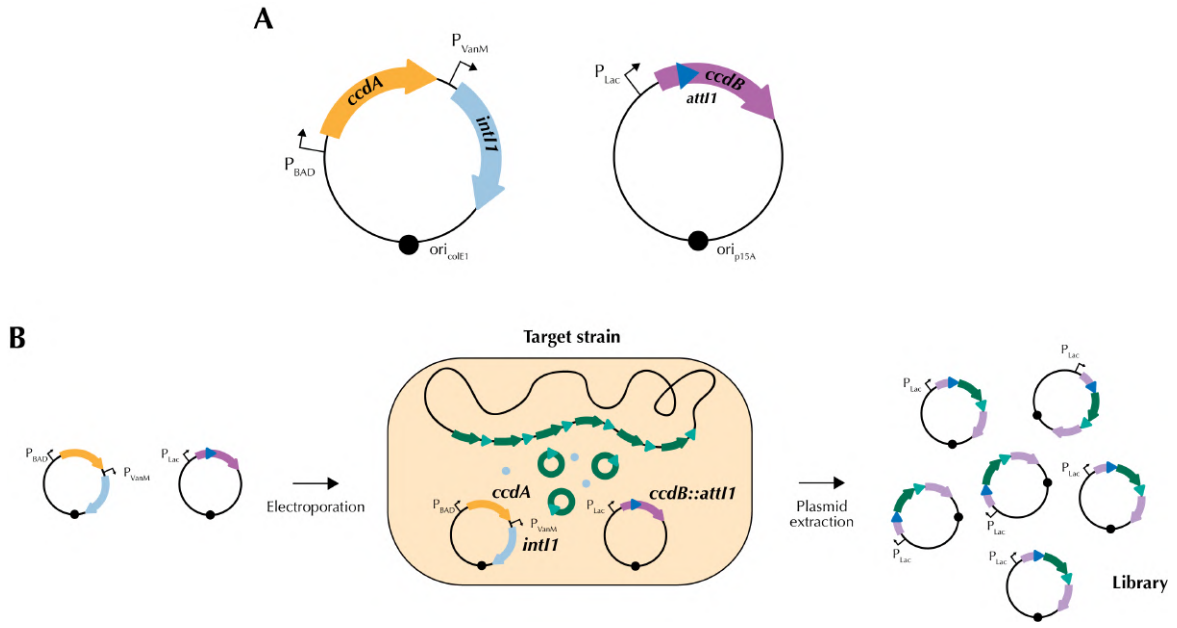


Figure 29. Plasmid setup and cassette library generation

A. Schematic representation of the plasmid setup used for the chromosomal IC library creation. *ccdB* was cloned under the control of a P_{Lac} promoter in a p15a plasmid. *ccdA* is cloned in a pColE1 plasmid under the control of the P_{BAD} promoter, together with *intI1*, controlled by the P_{VanM} promoter. **B.** Schematic representation of the protocol to generate the libraries. After the electroporation of both plasmids, we induce the expression of *intI1* and maintain the expression of *ccdA*. After growing cells for 4h in these conditions, cells are plated in killing conditions. Only cells that have captured cassettes in our *ccdB::attI1* bearing plasmid can survive. Plasmids are then extracted and treated as a library.

2.2 – Creating a cassette library from *V. cholerae* N16961 superintegron

We set out to demonstrate the efficiency of this tool in *V. cholerae* N16961. As previously mentioned, this bacterium possesses a very large superintegron that contains 179 cassettes, spans 126 kb, and is considered the paradigm of SCIs (for more details refer to page 65). To do so, we transformed the plasmid-based tool (**Figure 29A**) in *V. cholerae* N16961 by electroporation. As a control, we also transformed a version of the ColE1 plasmid devoid of *intI1.* The next day, we selected independent transformants that were inoculated in liquid media containing vanillate 12.5 µM and arabinose 0.2% to induce the expression of *intI1* and *ccdA*. After incubating bacteria for 4h in these conditions, cells were plated in glucose 1% to stop the production of *ccdA*. Control cells were instead plated in arabinose 0.2% to count the total CFUs. Our results show that we can capture cassettes from *V. cholerae* N16961 chromosome, as we

obtained recombination frequencies of 10^{-3}, compared to 10^{-6} when the *intI1* was not present (i.e., escape mutants) (**Figure 30A**). To conduct a preliminary verification of the library, we performed a PCR of the insert to determine if any cassette had been captured. The expected size of the PCR is 667 bp if empty and higher if containing a cassette. We recovered 100% (81/81) of positive samples in each assay performed (**Figure 30B**), and we sequenced by Sanger-sequencing the PCR products of 79 different colonies. Among them, we identified 65 different cassettes, suggesting our library contains a large proportion of SI cassettes. Details on the prevalence of the different cassettes can be found in **Supplementary Table 8**. These results show that this tool is a reliable method to capture cassettes from SCI's.

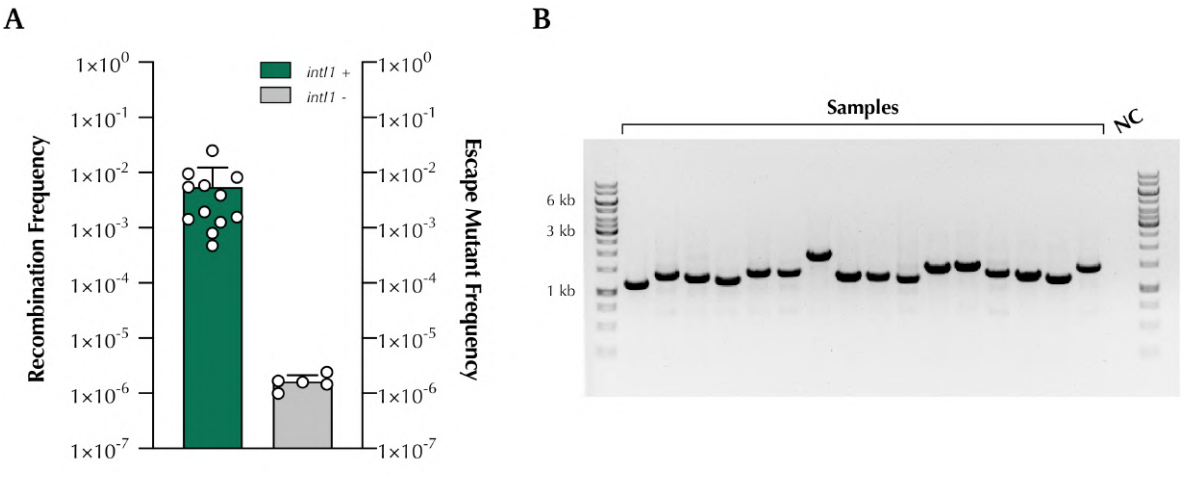


Figure 30. Library generation from *V. cholerae* N16961 superintegron

A. Recombination frequencies of the library generation from *V. cholerae* N16961 superintegron. The escape mutant frequencies, depicted in grey (*intI1* -), correspond to a strain where we transformed a Δ*intI1* version of the *ccdA* plasmid. The recombination frequencies were calculated as the ratio of the number of recombinants and the total CFUs, growing in arabinose. Bar charts show the mean of the recombination frequency ± s.d. from biological replicates (individual dots). **B.** Colony PCR results of the captured cassettes in the plasmid vector. NC, negative control.

The cassette harvester setup and method are protected by patent ES 2 969 666 A1.

2.3 – Library analysis

To better characterize the specificity of our tool, we sought to perform a deeper analysis of the generated library. For that, we recovered approximately 10^5 CFUs and extracted the plasmid content. We then performed a PCR of the insert and sequenced the obtained PCR product using long-read sequencing (Oxford Nanopore). This generated ~45,000 reads and 55 Mb with a mean Phred quality score of 15.2 and a

mean read length of 1200 bp (**Supplementary Figure 3**). The reads were then trimmed (to remove the *ccdB* sequence) and mapped to *V. cholerae* N16961 genome (RefSeq GCF_000006745.1), allowing secondary alignments. 76,857 alignments were produced. The initial coverage per nucleotide (nt) allowed us to measure the specificity of this technique: more than 99.9% of the reads were mapped within the superintegron, showing that the method is highly specific (**Figure 31**). Indeed, only 25 reads were mapped exclusively outside the SI and had no detectable pattern in the surrounding sequences. Overall, these results were expected as it is known that integron recombination is highly specific.

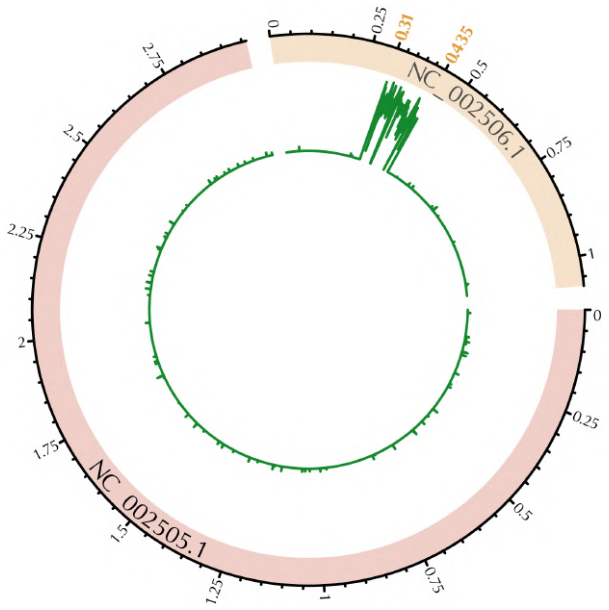


Figure 31. Library coverage across the chromosomes of *V. cholerae* N16961
Average coverage per kilobase (green plot) given in log10 reads of the two chromosomes of *V. cholerae* N16961 (RefSeq GCF_000006745.1). The sector NC_002505.1 corresponds to chromosome 1 and NC_002506.1 to chromosome 2. Chromosome coordinates are marked in the outside circle in black, and the superintegron coordinates are marked in orange.

After this, we used a *featureCounts* modified script to convert the mapping bam file to counts per feature (CDS, *attC,* or Cassette; for more details, see Methods page 85). 72,118 alignments mapped into annotated features (93.8%). We show that all cassettes were represented in our library with one exception: cassette 68—corresponding to an IS and two phage-defense systems—contains five ORFs and spans almost 4 kb, rendering its excision very unlikely. In all other cassettes, counts varied from 1 to up to 2,000, with a mean of 419 counts per cassette (**Figure 32A**). Remarkably, the less represented cassettes are either phage-defense systems (cassettes 17 and 69) (Darracq et al., 2024), or an IS (cassette 156). Three cassettes with less than 10 reads had no relevant features annotated. Interestingly, TA systems were not among the less-

represented cassettes and were homogeneously distributed. An interactive version of **Figure 32A** can be found in https://github.com/mredrejo/cassette_harvester. The great majority of the reads (95%) mapped to only one cassette, while 4.5% mapped to 2 cassettes. We found reads that aligned with up to five cassettes (**Figure 32B**). The reads that mapped to 0 cassettes were very short (mapping to *attC* sites) and/or of low quality (**Supplementary Figure 4**). It is of note that 23 cassettes are duplicated in the SI (>95% of identity). This could also explain why 4.5% mapped to >2 cassettes. Interestingly, the same analysis, but counting the number of CDS per read, revealed that 22% mapped to zero CDS, 72% mapped to 1 CDS, and 6% mapped to 2 CDS. A few reads mapped to 3 or 4 CDS (**Figure 32C**). This can be due to the number of CDS contained in each cassette: a great number only possess one CDS, but a substantial amount can have zero CDS (either by incorrect annotation or by being gene-less cassettes (Blanco, Hipólito, et al., 2024)) or two CDSs (e.g. TA systems). Moreover, the appearance of more than 3 CDSs could also be explained by the capture of more than one cassette.

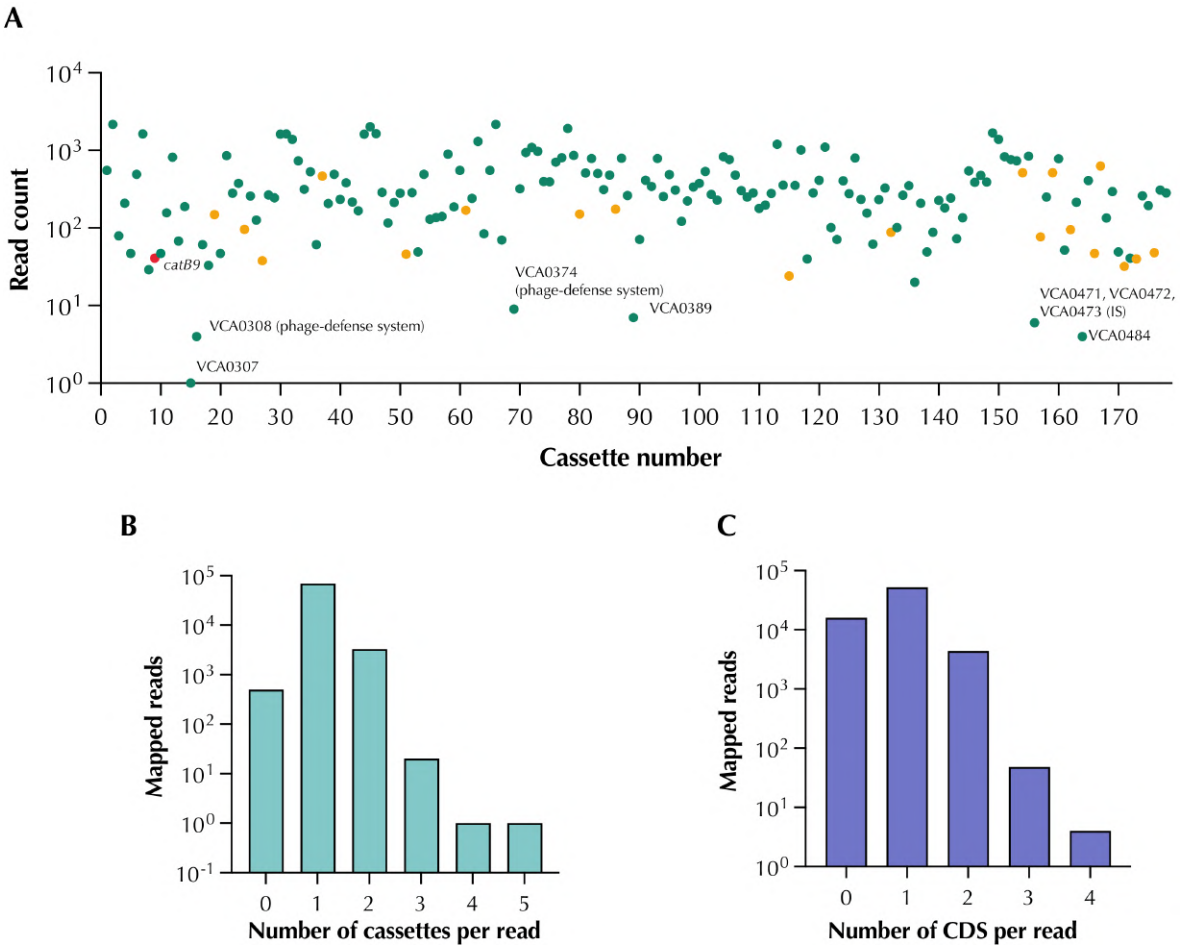


Figure 32. Analysis of superintegron representation in the plasmid library

A. Read count of the 178 cassette features. *catB9* cassette is marked in red and the 19 TA cassettes are marked in orange. Cassettes with lower (>10) read counts are marked with their ORFs as well as their identified function. **B.** Number of reads that mapped to 0-5 cassettes. **C.** Number of reads that mapped to 0-4 CDS.

Additionally, we attempted to find an explanation for the variations in the cassette representation in our library. The excision of a cassette is dependent on the folding probability of its own *attC* site but also on the previous one, as it is the *attC* × *attC* reaction that leads to the excision of a cassette (see **Figure 7**). We analyzed the e-values (folding probability) of the surrounding *attC*s of each cassette, and correlated the mean of the e-values to the counts of the *attC* feature using a Spearman correlation. We found no correlation (R = 0.054; p = 0.48), suggesting that the folding probability is not important in this case (**Figure 33A**). Another important feature is the size of cassettes—or the distance between the two *attC*s sites. This is critical as the two folded ssDNA recombination sites must interact with each other to produce an excision event, thus, their distance is crucial in the probability of excision. In fact, it has been demonstrated that shorter cassettes are more recombinogenic and that the size of CDSs

in the SI is considerably smaller compared to the rest of the chromosome 2 of *V. cholerae* N16961 (Loot et al., 2017). Therefore, we correlated the size of each cassette and its counts, and we found a moderate negative correlation (R = -0.54; *p* = 1.3e-14), showing that the higher the size of the cassette, the lower the count (**Figure 33B**). It is of note that we cannot rule out the influence of the PCR amplification step.

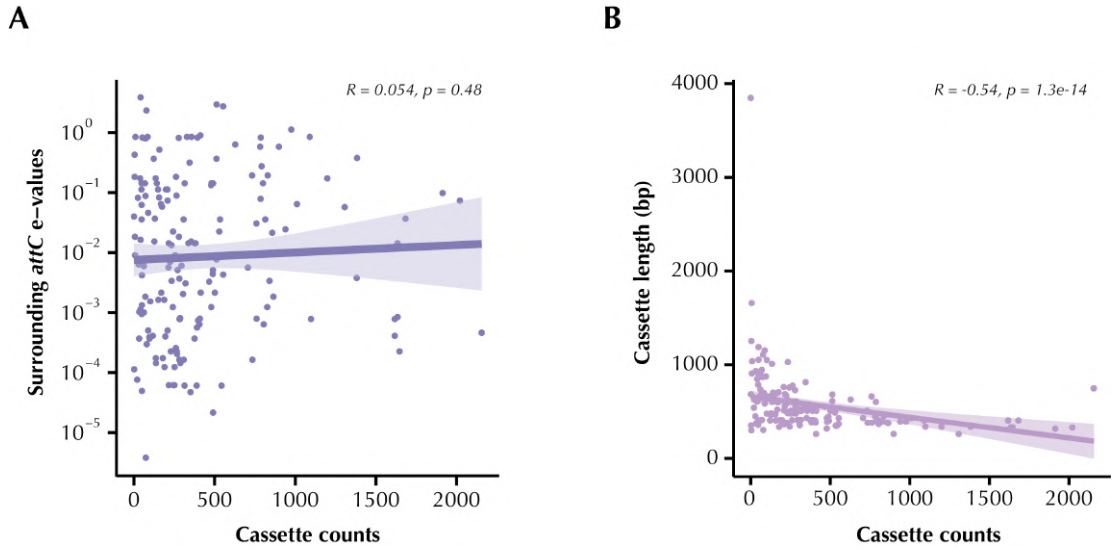


Figure 33. Correlations between cassette representation and cassette features

A. Correlation of the mean of the surrounding *attC* site e-values and the number of counts of the corresponding cassette. **B.** Correlation of the cassette length and the number of counts of the same cassette. Both correlations were calculated using a Spearman correlation.

Altogether, these results mark a significant achievement in the development of this thesis and in the field of integrons. Our tool has demonstrated immediate practical value by generating a comprehensive library of the *V. cholerae* N16961 superintegron with extreme specificity and sensitivity. This application of the integron capture chassis will likely enable further exploration of the functions of each cassette as well as the creation of libraries from other SCIs.

Chapter 3: A sentinel bacterium

Integrons are pivotal in the transmission of antimicrobial resistance among Gram-negative bacteria. Indeed, these platforms are present in over 50% of these clinical isolates (refer to page 58 for more details), especially in conjugative plasmids. It is known that conjugation serves as a primary mechanism through which bacteria transfer genetic material, making it paramount to detect integrons carried in such efficient vehicles. Interestingly, conjugation delivers a ssDNA product, facilitating the *attC* folding and, thus, the excision of cassettes.

In this chapter, our goal was to adapt our tool to create a sentinel bacterium with the ability to detect integrons through conjugation, thereby providing a tool to explore the dynamics of horizontal gene transfer of these genetic elements, and effectively detecting the dissemination of resistance genes across diverse environments.

3.1 – *E. coli* setup

E. coli has been the workhorse of molecular biology (Ruiz & Silhavy, 2022). In fact, this species has been extensively used to study conjugation since its discovery (Couturier et al., 2023; DelaFuente et al., 2022; Lederberg & Tatum, 1946) as well as other HGT platforms such as integrons (Biskri et al., 2005; Hipólito et al., 2023; Loot et al., 2024). Additionally, the commonly used *E. coli* strains have also been adapted as an *in vivo* tool (López-Igual et al., 2019), showcasing the versatility of choosing such species to create a sentinel bacterium. Hence, to achieve this goal, we decided to use the laboratory strain *E. coli* MG1655 as a chassis and combine it with the SacB-based tool. This would also facilitate the adaptation to a more complex environment, as there would be no need to implement tight control of CSM as in the case of the CcdB/CcdA TA system. In this setup, the sentinel *E. coli* containing the *sacB::attI1* construct in the chromosome and the pBAD::*intI1* plasmid would be co-cultured with the tested bacteria to allow both conjugation and integrase expression and then plated in "killing" conditions. This would only allow the growth of transconjugants where integron cassettes had been excised and captured in our *sacB::attI1* platform (**Figure 34**).

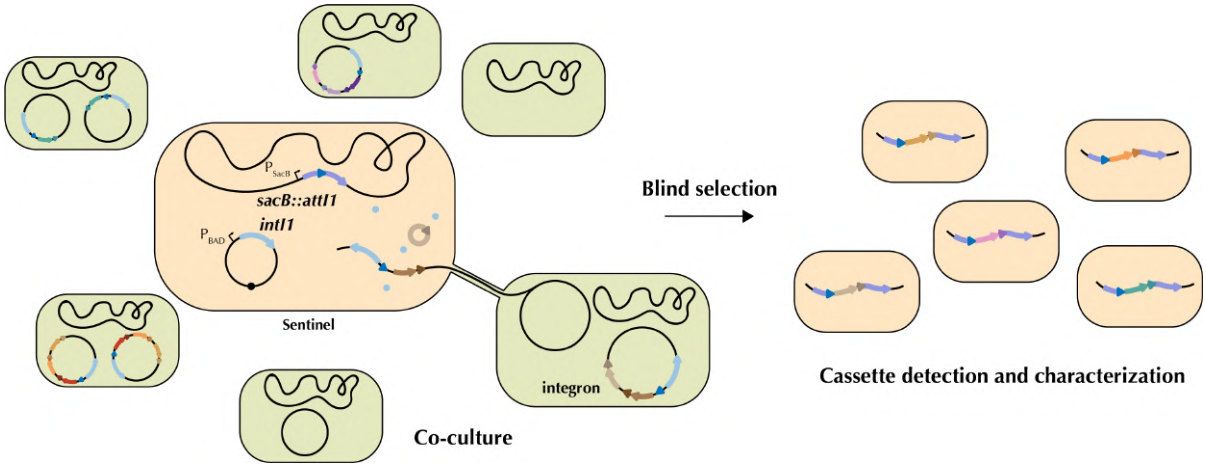


Figure 34. *E. coli* as a sentinel bacterium to detect integrons in mobile plasmids
Schematic representation of the setup used to create a sentinel bacterium. In this system, the tested sample is co-cultured with our sentinel bacterium that contains a SacB-based version of our tool. Upon conjugation, represented in the lower right corner, the expressed IntI1 can mobilize the integron cassettes inside the *sacB::attI1* platform. After, the selection of recombinants is achieved through growth in sucrose.

3.2 – Conjugation of clinical plasmids

As this had already been demonstrated in Chapter 1 (see page 115) with a suicidal plasmid mimicking an integron, we sought to prove that the tool was also usable with clinical plasmids, such as the widely known R388 (Avila & de la Cruz, 1988). This plasmid belongs to the IncW family and has been extensively used in the study of conjugation and also as a molecular biology tool. It possesses an integron that contains a trimethoprim resistance gene, *dfrB2*, and a gene cassette of unknown function (*gcuA*). To extend the number of cassettes tested, as well as their different *attC* sites, we also selected a modified version of the R388 plasmid that includes a *dfrA5* gene, conferring resistance to trimethoprim, a $bla_{VEB\text{-}1}$ gene, encoding an extended-spectrum β-lactamase, and *aadB*, conferring resistance to aminoglycosides (**Figure 35A**) (Souque et al., 2021). As controls, we included the previously tested suicidal plasmid mimicking an integron cassette, pSW23T $attC_{aadA7}$, as well as a sample with no plasmid donor as a negative control. The sentinel *E. coli* was co-cultured individually with each of the donor strains (*E. coli* MG1655 carrying the tested plasmids) in filters to favor conjugation. This was performed in the presence of arabinose to induce the expression of *intI1*. The next day, the mixture was plated in LB (ØNaCl) + sucrose 10% to select for recombinants, and zeocine, to only allow the growth of the sentinel. We recovered bacteria at a frequency of 10^{-3} for the R388 and 10^{-4} for the modified version. These

frequencies were comparable to the suicidal plasmid control and well above the negative control (10^{-5}) (**Figure 35B**).

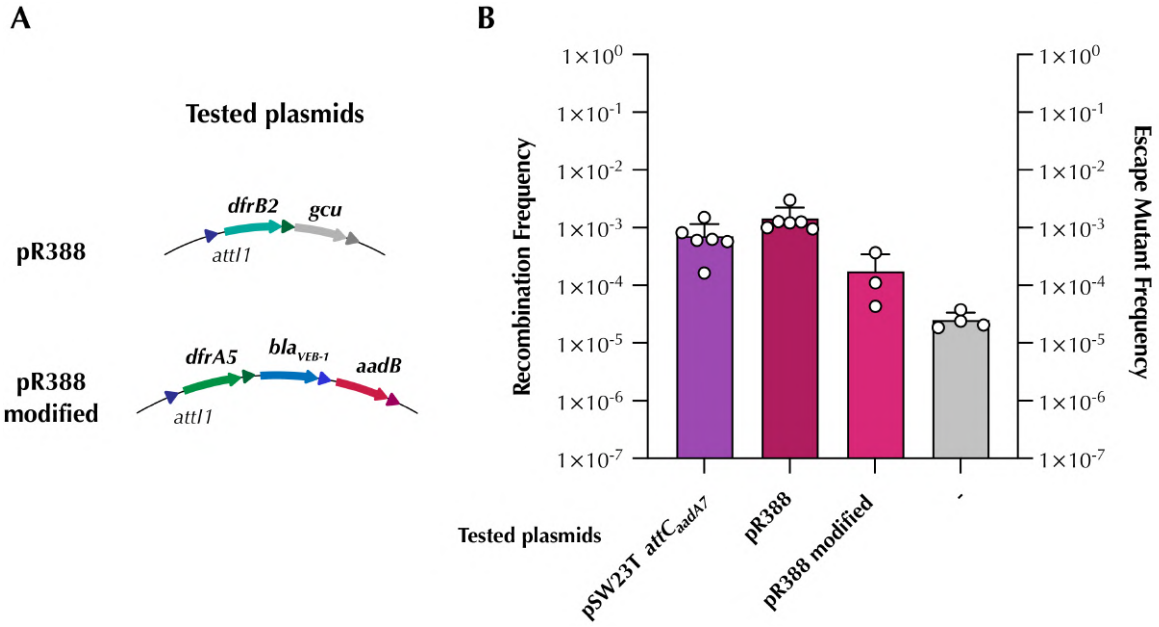


Figure 35. Integron detection by the sentinel *E. coli*

A. Schematic representation of the integron arrays contained in the tested plasmids. Arrows represent genes, and triangles represent recombination sites. **B.** Recombination frequencies of the tested samples. The recombination frequencies were calculated as the ratio of the number of recombinants and the total CFUs. The escape mutant frequencies (represented in grey) correspond to the - sample where no donor was co-cultured with the sentinel *E. coli*. Bar charts show the mean of the recombination frequency ± s.d. from biological replicates (individual dots).

We performed a PCR to amplify the cassette content of the recombinants in order to analyze the efficiency of this tool. Concerning the conjugation of the original R388, out of 32 colonies tested, 29 were cassette capture events. The size of the PCR product allowed us to infer the detection of the *gcuA* cassette in almost all the positive samples but one, corresponding to the *dfrB2* cassette. Sanger sequencing of three PCR products confirmed these results. In the conjugation of the modified R388, we detected 22 cassettes out of 24 colonies tested. Interestingly, the majority of colonies had the *aadB* cassette (the smaller cassette), but we also found the bla_{VEB-1}, and the combination of both *aadB* and bla_{VEB-1}. Finally, we also identified the *dfrA5* cassette. This shows that this method is sensitive and specific enough to detect integrons through conjugation, and even allows the characterization of the integron array.

In this Chapter, we show that our tool is highly versatile and can be adapted as a sentinel bacterium capable of detecting integrons contained in conjugative plasmids. This sentinel bacterium not only provides a novel means to study horizontal gene

transfer dynamics, potentially *in vivo*, but could also offer a practical solution for real-time monitoring of resistance gene spread, ultimately contributing to more effective strategies for managing antimicrobial resistance.

Chapter 4: *Vibrio cholerae* as a chassis to test directly DNA samples

The final aim of this thesis was to adapt our tool to detect integrons directly from DNA samples. This is crucial for advancing our knowledge and understanding of these genetic elements, as it would provide a rapid and easy-to-use tool to monitor antimicrobial resistance and, importantly, a detection method applicable to non-culturable organisms. *Vibrio cholerae* provides a unique advantage as a chassis for such applications due to its natural competence and compatibility with integron activity.

This final chapter explores the potential of *V. cholerae* as a tool for screening integrons in DNA samples, the genetic modifications performed to achieve this goal, and a novel approach to tracking the dissemination of integrons and identifying new genes contained in these platforms.

4.1 – Creating the perfect chassis

4.1.1 – Superintegron deletion

V. cholerae is an interesting microorganism for establishing a platform to capture gene cassettes. However, as previously discussed, it possesses an enormous chromosomal integron, dubbed the superintegron (SI) (for more details refer to page 65 and **Figure 9**). Using the tool directly in the wild-type strain would result in interference, as demonstrated by our results when establishing chromosomal libraries in Chapter 2.

Deleting the SI in *V. cholerae* is a longstanding milestone in the field, not only for biotechnology purposes but also to study these chromosomal platforms with the same depth as the class 1 integron. Previous efforts to delete the SI were hampered by unknown TA systems. The recent complete characterization of TA systems in the SI (Iqbal et al., 2015; Krin et al., 2023) allowed us to develop the Sequential Deletion of TA systems (SeqDelTA) (Blanco & Trigo da Roza et al., 2024), a multi-step approach that allows the inactivation of TA systems while deleting, at each step, the cassette cargo between two TAs (**Figure 36**).

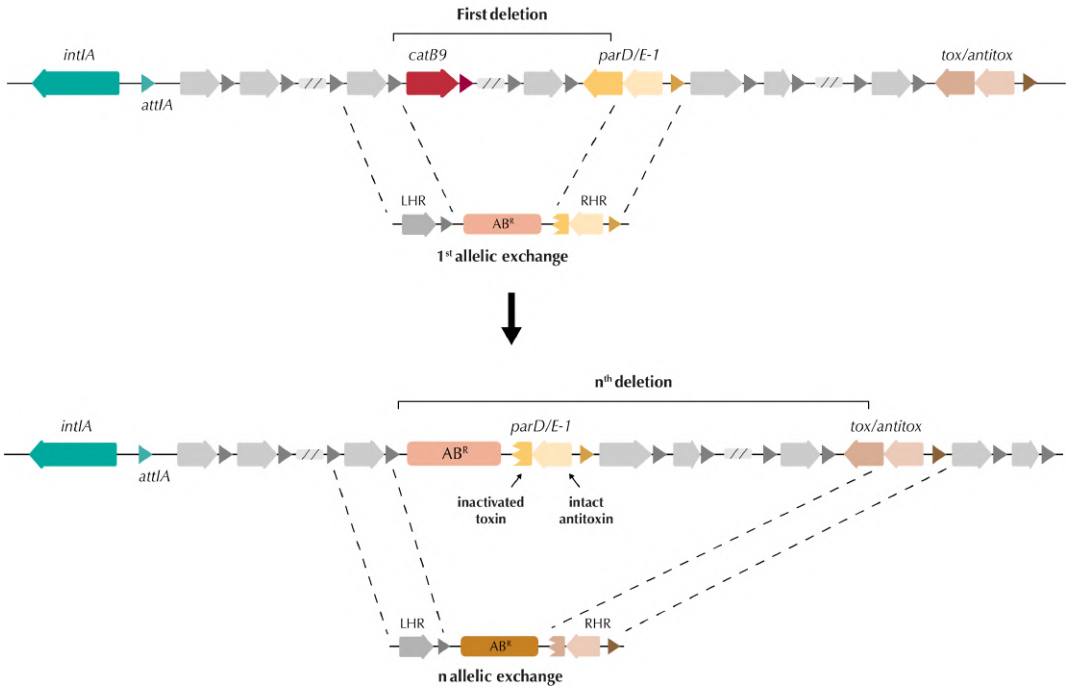


Figure 36. SeqDelTA

Schematic representation of SeqDelTA, a method employed to achieve the deletion of the superintegron. The superintegron and two TA systems are represented, as well as their two deletions (first deletion and n^{th} deletion). Below, the fragments assembled by PCR are characterized: the left homology region (LHR) targets the cassette located upstream the *catB9* gene and is conserved, while the right homology region (RHR) is constructed at each step, ensuring that the antitoxin is intact while deleting the toxin. The resistance markers are depicted here in light pink and brown and can be recycled every two steps.

This approach exploits *V. cholerae*'s natural competence and homologous recombination to deliver allelic exchanges. To do so, we transformed linear DNA segments that include a resistance marker flanked by left and right homology regions (LHR and RHR). These regions are designed to match specific sequences within the superintegron, facilitating allelic replacement from gene VCA0291 (the integrase) to VCA0506 in 16 deletion steps. For allelic exchanges, we consistently used a specific LHR that targets the cassette located upstream of the *catB9* chloramphenicol resistance cassette (8^{th} position of the array (VCA0299). This choice was initially made to avoid interference with the P_C promoter, as the strain is susceptible to chloramphenicol (leading us to suspect that this part of the array is silent), although it was recently demonstrated that the *catB9* gene is transcriptionally active (Blanco, Hipólito, et al., 2024). In contrast, the RHR was changed at each step to target the subsequent TA system. They were specifically designed to ensure that the allelic replacement preserved the antitoxin gene while removing the associated toxin. By maintaining a constant LHR and varying the RHR, we could efficiently recycle resistance markers, removing each marker introduced in the preceding step. This deletion strategy, termed

SeqDelTA, remained generally consistent unless gene orientation or order within the TA systems required adjustments (for example, the *phd*/*parE* TA system where the toxin is positioned after the anti-toxin (see **Figure 9**)); in these cases, we employed alternative LHRs and kept two resistance markers for consecutive replacements, followed by a third replacement to eliminate both markers. After each replacement, colonies that grew on selective media were screened via PCR and phenotypic checks for loss of the previous resistance marker. We also analyzed the growth curves of PCR-validated clones to ensure there were no hitchhiking mutations with deleterious effects. A clone demonstrating normal growth was selected to proceed with further deletions. After completing the series of replacements, we removed all cassettes from VCA0300 (*catB9*) up to VCA0503 (the last toxin). Finally, we employed a counter-selectable integrative vector (pMP7) (Val et al., 2012) to perform a clean deletion of the remaining sections of the superintegron, including the integrase, the first eight cassettes, and the last two cassettes, along with a transposase just after the final cassette (VCA0291 to VCA0508).

We first verified the complete removal of the superintegron by PCR. Then, the generated strain was sequenced with both Nanopore and Illumina sequencing. The combination of small and long reads allowed us to obtain a high-quality genome sequence, in which we could exclude genomic rearrangements while detecting unintended single nucleotide polymorphisms (SNPs). The detected SNPs (*rocS* in the 3rd deletion step, *rpoS* in the 12th, and *cry2* in the 16th) were corrected using the same counter-selectable integrative vector as mentioned above.

The SeqDelTA method and the *V. cholerae* N16961 ΔSI strain are protected by patent ES 2 970 040 A1.

4.1.2 – *V. cholerae*$_{\Delta SI}$ phenotypic characterization

The deletion of the SI entails the loss of 3% (126 kb) of the genome. This substantial genomic reduction could potentially impact growth or fitness. To investigate this, we conducted growth curves of 24 independent *V. cholerae*$_{\Delta SI}$ colonies in comparison to the wild-type N16961 strain in LB medium. As illustrated in **Figure 37A**, both strains exhibited similar growth patterns over 16 hours, with no significant differences observed. To enhance the sensitivity of our fitness assessment, we

performed competition assays using flow cytometry, employing a fluorescent *E. coli* strain as a common competitor. The results showed no significant fitness differences between the *V. cholerae* WT and ΔSI strains (**Figure 37A**).

As reviewed in depth in Chapter 3 of the Introduction section (page 64), *V. cholerae* thrives in various environments outside the human host, particularly in brackish or marine waters with temperatures ranging from 12 to 30°C. Indeed, natural transformation assays are usually performed at 30°C, as temperature can enhance this process. Therefore, we also assessed the impact of SI deletion on growth under these conditions. As shown in **Figure 37B**, the growth of both WT and ΔSI strains was similar at 30°C and 20°C. Overall, these findings suggest that the deletion of the SI does not significantly affect bacterial growth or fitness.

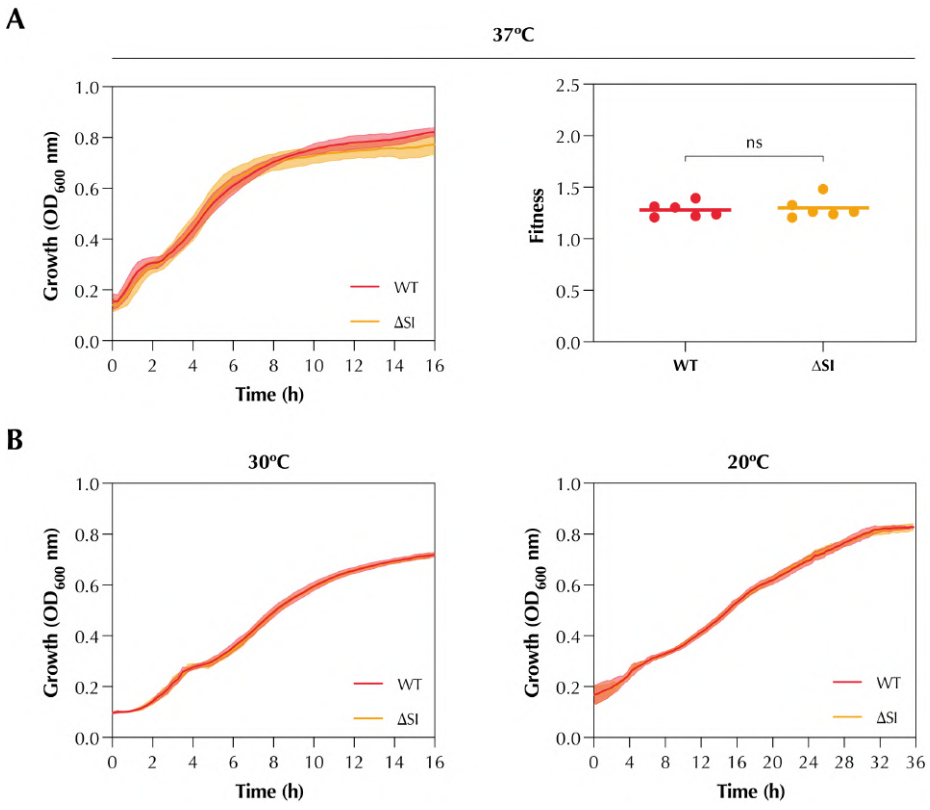


Figure 37. Growth curves and competitions of *V. cholerae*$_{\Delta SI}$ vs *V. cholerae*$_{WT}$

A. Growth curves and fitness of *V. cholerae*$_{\Delta SI}$ (orange) and *V. cholerae* WT (red) at 37°C. The relative fitness was performed by competing both strains against a common competitor - *E. coli* DH5α /pSU38::P$_{tet}$S *gfp*. Fitness values were calculated with 6 independent experiments from flow cytometry. Statistical significance was determined using the Mann-Whitney U test; ns, not significant. **B.** Growth of *V. cholerae*$_{\Delta SI}$ (orange) and *V. cholerae* WT (red) at 30°C and 20°C of 10 independent colonies.

4.1.3 – Characterization of *V. cholerae*$_{\Delta SI}$ natural competence

After deleting the superintegron, we sought to test if *V. cholerae*$_{\Delta SI}$ was as competent as other isolates well characterized in the field. Integrons are often

described as adaptive memories for bacteria, and the superintegron encodes more than 179 cassettes. One could then hypothesize that the deletion of the superintegron could entail a cost for natural transformation, as an important gene of the regulatory competence network could be lost. Moreover, strain N16961 was initially characterized as non-competent due to a frameshift in the *hapR* gene. The phenotype was corrected by introducing *hapR* in a miniTn7, generating the parental strain of *V. cholerae*$_{\Delta SI}$ (Meibom et al., 2005). Thus, it seemed essential to fully characterize the natural competence of our strain.

For this, we compared *V. cholerae* A1552, *V. cholerae* N16961 (the ΔSI parental strain, *hapR*+), and *V. cholerae*$_{\Delta SI}$ (**Figure 38**). We performed a classical natural competence assay as previously described (Marvig & Blokesch, 2010), and tested four different DNA products: i) gDNA A1552 Δ*lacZ*: a gDNA extraction from a *V. cholerae* A1552 Δ*lacZ* mutant, where the *lacZ* gene is replaced by a kanamycin resistance gene; ii) PCR product 500 bp: a PCR amplification of the kanamycin gene using 500 bp of homology with the surrounding *lacZ* gene areas; and iii) and iv) the same PCR using 1 kb and 2 kb of homology, respectively. 2 µg of each product were provided. Note that the genetic context of the *lacZ* gene is identical in both A1552 and N16961 strains.

We obtained transformation frequencies ~10^{-4} for the gDNA products, 10^{-5} for the 500bp of homology PCR product, 10^{-4} for 1 kb of homology, and 10^{-3} for 2 kb of homology (**Figure 38**). The only statistically significant difference is between A1552 and N16961 when transforming the 2 kb product. However, these variations are expected in natural competence. Our results show that the *V. cholerae* N16961$_{\Delta SI}$ strain is as competent as A1552, showing that the deletion of the superintegron does not affect transformation. Moreover, the transformation frequencies obtained in this test are similar to what was obtained before in Prof. Blokesch's Lab (Stutzmann & Blokesch, 2020).

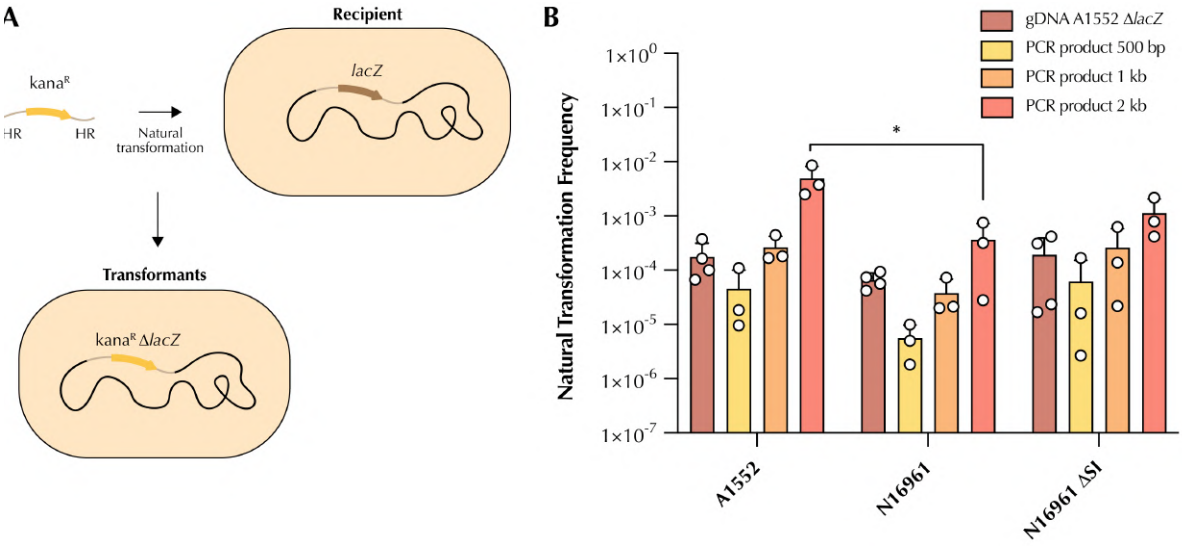


Figure 38. Classical natural competence test of *V. cholerae* N16961$_{\Delta SI}$

A. Schematic representation of the classical natural competence assay. Above, a *V. cholerae* recipient is depicted with the *lacZ* gene and its adjacent regions. The transformed DNA corresponds to a kanamycin resistance gene (kana^R) and its promoter merged to the adjacent regions of the target (in light brown), dubbed homology regions (HR). After transformation, the kana^R gene substitutes the *lacZ* gene by homologous recombination, allowing for the selection of transformants. **B.** Transformation frequencies are given on the Y-axis and were calculated as the ratio of the number of CFUs, growing in kanamycin, and the total CFUs. Bar charts show the mean of the recombination frequency ± s.d. from biological replicates (individual dots). Statistical significance was determined using the Kruskal-Wallis test with Dunnett's correction for multiple comparisons; *, $p < 0.05$.

4.1.4 – Enhancing *V. cholerae*$_{\Delta SI}$ natural competence for integron recombination

After ensuring that the *V. cholerae*$_{\Delta SI}$ strain was naturally competent, we sought to improve this trait and optimize it for **integron recombination**. A previous report showed that delivering DNA through natural transformation for *intIA*-mediated recombination is not efficient enough (recombination frequencies of ~10^{-7}) to achieve our goals (Vit et al., 2021), as it would fall below our background noise frequency (see escape mutant frequencies in Chapter 1). However, we aimed to overcome these limitations by generating a series of specific mutants that would be more efficient in both natural competence and integron recombination.

Recently, two clusters of genes, *ddmABC* and *ddmDE*, were characterized as phage-defense systems (Jaskólska et al., 2022) that also influence plasmid maintenance in *V. cholerae*. DdmDE degrades small plasmids, leading to their elimination, while DdmABC enhances this process by either actively participating in degradation or facilitating plasmid clustering. Additionally, the latter phage-defense system employs an abortive infection-like mechanism to eliminate large conjugative plasmids. As this application for our tool involves the creation of a small plasmid containing the integron recombinase gene—*intI1*—we deemed it essential to delete the *ddm* system for

optimization. To achieve this, we generated a marker-less **Δ*ddm*** strain, deleting the gene clusters using a counter-selectable integrative vector (pGP704-SacB) as previously described (Meibom et al., 2004).

Using the same technique, we generated a series of mutants of the ΔSI strain: i) a **Δ*dns*** strain, characterized as one of the most transformable *V. cholerae* mutants, as the periplasmatic and extracellular nuclease (Dns) is inactivated (Blokesch & Schoolnik, 2008); ii) a **Δ*luxO*** strain, described in a few studies as more efficient in transformation (Haycocks et al., 2019; Simpson et al., 2019), as LuxO can repress HapR, another important regulator of natural competence; and iii) a **Δ*dprA*** strain, as DprA, a DNA binding protein, is known to bind to the uptaken DNA, and recruit RecA together with SSB (Sharma et al., 2023). We hypothesized that DprA could be competing with IntI1 and negatively affect recombination frequencies. For more details on the natural competence regulation of *V. cholerae*, refer to page 68 and **Figure 10**.

We then tested these strains in a classical natural competence assay, as described above, only testing the 1 kb PCR product. When comparing to the ΔSI strain, our results show no differences between these mutants, as they all present transformation frequencies around 10^{-4} (**Figure 39**). The only exception is the Δ*dprA* strain, which yielded no transformants. This was expected since DprA is essential to recruit RecA and, hence, to perform homologous recombination.

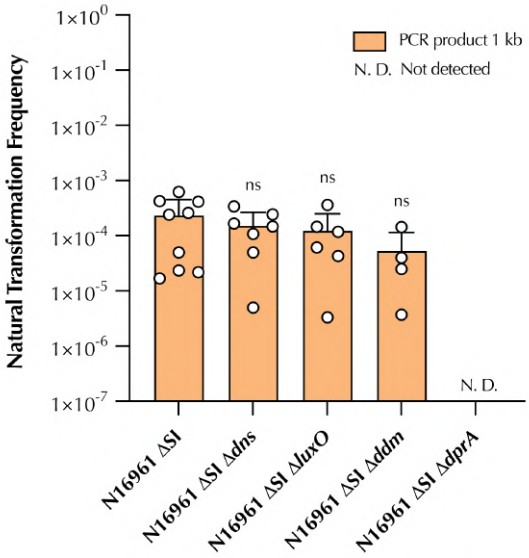


Figure 39. Classical natural competence test of *V. cholerae* N16961$_{\Delta SI}$ generated mutants
Transformation frequencies are given on the Y-axis and were calculated as the ratio of the number of CFUs growing in kanamycin, and the total CFUs. Bar charts show the mean of the recombination frequency ± s.d. from biological replicates (individual dots). Statistical significance between each mutant and the ΔSI control was determined using the Kruskal-Wallis test with Dunnett's correction for multiple comparisons; ns, not significant.

4.1.5 – Testing recombination through natural transformation

We then aimed to fine-tune a previously existing chitin-independent transformation assay (Chlebek et al., 2019) and adapt it for capturing integron cassettes. The goal was to achieve faster detection times for the capture of cassettes and have the master regulator of natural competence – TfoX (Meibom et al., 2005) – controlled by an inducible promoter, thus increasing the efficiency of transformation. For that, we constructed a pBAD::*intI1-tfoX* plasmid, cloning both genes under the control of the P_{BAD} promoter that was introduced in all ΔSI-generated mutants. Furthermore, we took advantage of a pre-existing *attI1* site cloned within the chromosome 1 of our strain, inserted in the miniTn7 (as it contains a gentamicin resistance marker that derives from an integron) when the *hapR* gene was reconstructed in *V. cholerae*$_{\Delta SI}$ parental strain (Meibom et al., 2005).

We first tested two strains (ΔSI and ΔSI Δ*ddm* strains) in our chitin-independent natural competence assay for integron recombination to set the basis for further experiments. In these assays, strains were cultured overnight in glucose to repress both *intI1* and *tfoX*, followed by a 2.5-hour induction with arabinose before transferring them to defined artificial seawater (DASW), as described in the reference. This step differs from the original protocol, where *tfoX* is expressed overnight, as our prior data suggested that overnight expression of the *intI1* gene led to lower cell counts and altered morphologies (not shown), indicating potential deleterious effects. Additionally, we changed the tested DNA, as we had previously tested capturing a suicidal plasmid that mimics a cassette (Bouvier et al., 2005), with no success (data not shown). For this experiment, we amplified three products corresponding to three different lab-generated integron arrays based on our pMBA collection (Carvalho et al., 2024). These amplified arrays are composed of two cassettes, a *catB3* cassette following a different cassette in the first position (*aacA54* for PCR product pC137, *aacA8* for PCR product pC155 and *aacA61* for PCR product pC167) (**Figure 40A**). 2 µg of each product were transformed. In this experiment, we only tested for the capture of the second cassette, selecting for recombination events in the presence of chloramphenicol, as we know that the excision of a cassette in the first position is less frequent (see Chapter 2, page 125) and, therefore, can be below our limit of detection. We obtained a recombination frequency of 10^{-5} for the ΔSI strain and 10^{-4} for the ΔSI

Δ*ddm* strain (**Figure 40B**). This supports previous observations (Gestal et al., 2011) that bacteria can excise and integrate a single cassette from a free linear fragment of DNA uptaken by natural competence. Importantly, in these samples, the first cassette (and thus the *attC*) identity does not influence the recombination frequency of the second one (all *p* values were > 0.9999 using a Kruskal-Wallis test with Dunnett's correction for multiple comparisons). This allowed us to compare strains (ΔSI vs. ΔSI Δ*ddm*), combining the data from all samples. This shows that the ΔSI Δ*ddm* strain is more efficient in integron recombination when receiving DNA by natural transformation.

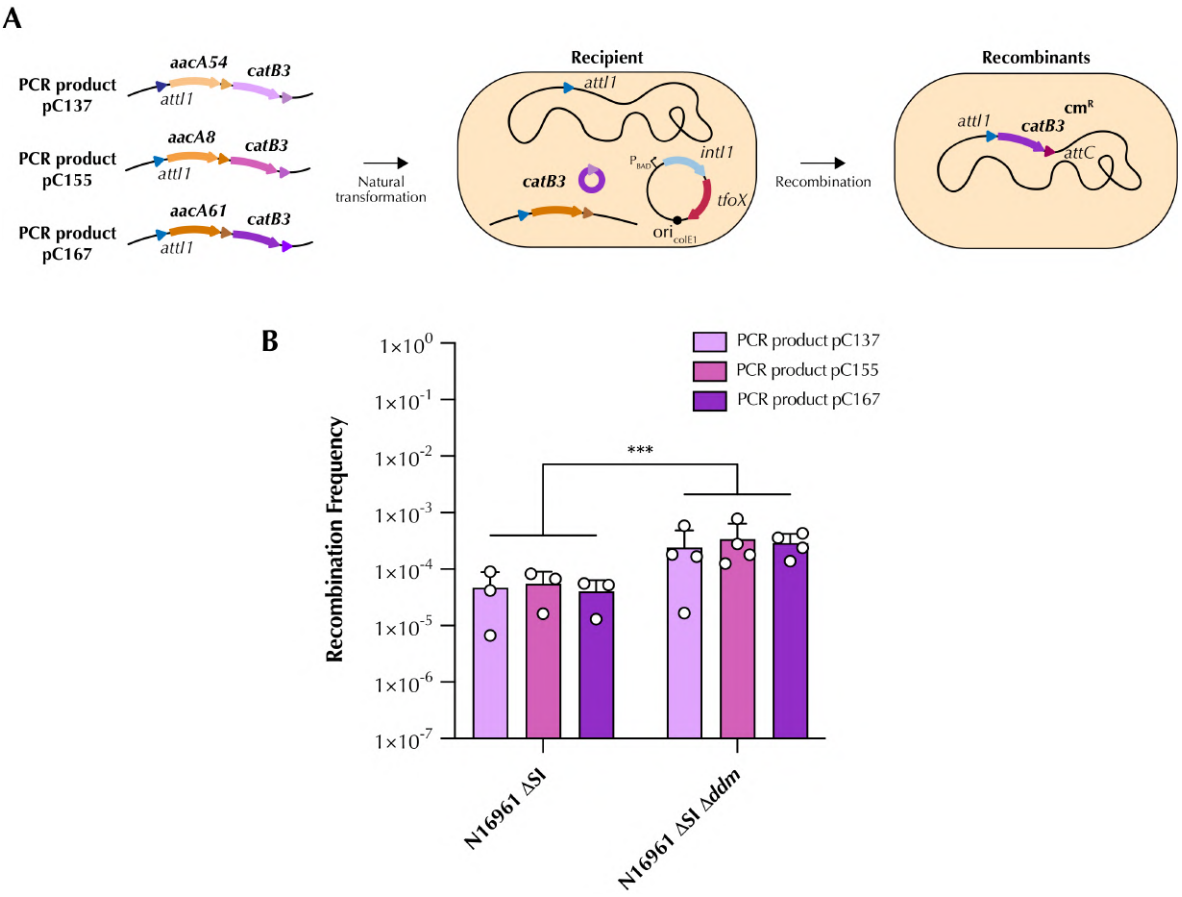


Figure 40. Chitin-independent natural transformation assay for integron recombination
Strains N16961 ΔSI and N16961 ΔSI Δ*ddm* were tested in a chitin-independent transformation assay for integron recombination upon *tfoX* and *intI1* induction with arabinose at 0.2%. **A.** Schematic representation of the recombination assay. Three different integron arrays were transformed into the recipients that contain an *attI1* site in the chromosome. The capture of the second cassette (*catB3*) was selected with chloramphenicol. **B.** Recombination and frequencies are given on the Y-axis and were calculated as the ratio of the number of CFUs growing in chloramphenicol or kanamycin and the total CFUs. Bar charts show the mean of the recombination frequency ± s.d. from biological replicates (individual dots). Statistical significance was determined using the Mann-Whitney U test; ***, $p < 0.001$.

Having established a basis for integron recombination using natural transformation as a method to deliver DNA, we then tested all our ΔSI-generated mutants in the same chitin-independent assay. As our three different products did not

show any difference in their excision and capture efficiency, we decided to use only pC167. As a control for transformation, we used the same 1 kb homology Δ*lacZ* kana^R PCR product. Again, 2 µg of each product were provided. We obtained recombination frequencies and natural transformation frequencies that ranged from 10^{-5} to 10^{-4} (**Figure 41A**), and 10^{-6} to 10^{-5} (**Figure 41B**) respectively, except for the Δ*dprA* mutant where we could not detect any recombinants nor transformants. This suggests that DprA is essential—not only to homologous recombination—but also to the uptake of DNA inside the cell. Perhaps not surprisingly, as part of our tool is plasmid-based, we observe the same slight superiority of the Δ*ddm* strain as before, although it is just statistically significant when measuring integron recombination. We also observe that the Δ*dns* strain seems to confer an advantage in both transformation and recombination, though not statistically significant. The Δ*luxO* strain does not confer any advantage in our experiments. However, in general, this chitin-independent protocol seems to be less efficient when compared to the classical natural transformation protocol. Indeed, when comparing natural transformation frequencies, it seems less effective, and the most efficient mutants only seem to reach the expected WT chitin competence state (see **Figure 41B** vs. **Figure 39)**.

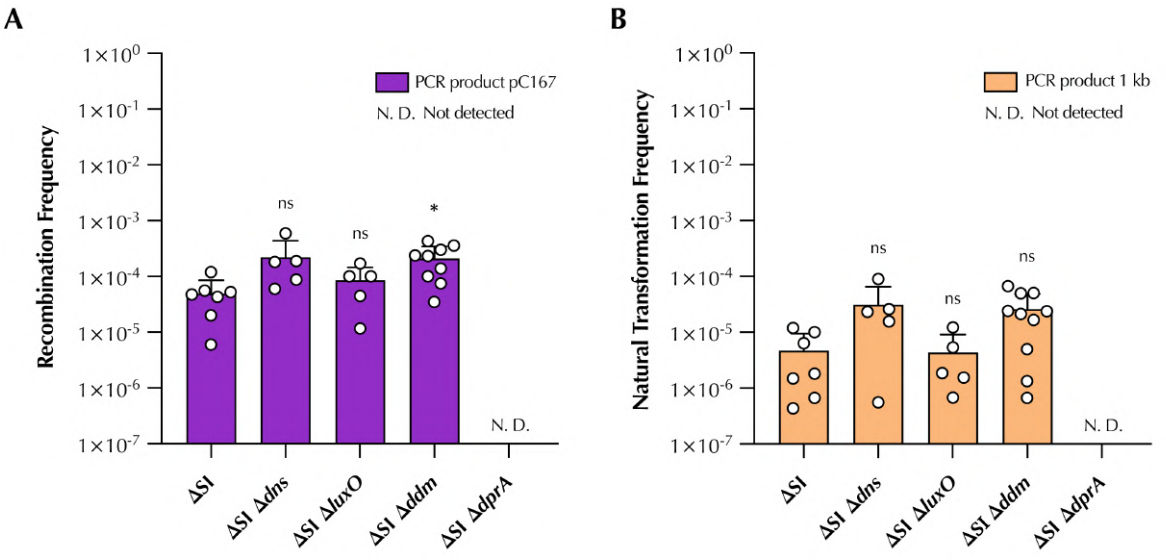


Figure 41. Chitin-independent natural transformation assay for integron recombination of *V. cholerae* N16961$_{\Delta SI}$ generated mutants

Mutants derived from strain N16961 ΔSI were tested in a chitin-independent assay upon *tfoX* and *intI1* induction with arabinose at 0,2%. Recombination (**A**) and transformation (**B**) frequencies are given on the Y-axis and were calculated as the ratio of the number of CFU's growing in chloramphenicol or kanamycin and the total CFUs. Bar charts show the mean of the recombination frequency ± s.d. from biological replicates (individual dots). Statistical significance between each mutant and the ΔSI control was determined using the Kruskal-Wallis test with Dunnett's correction for multiple comparisons; ns, not significant, *, $p < 0.05$.

As it seemed that inducing the *tfoX* master regulator was not as efficient as inducing competence by growth on chitin, we decided to test integron recombination using a classical recombination assay. For that, we transformed the *V. cholerae*$_{\Delta SI}$ Δ*ddm* mutant and the newly created double mutant Δ*ddm* Δ*dns* (as both individual mutants demonstrated a slight superiority in the last experiment) with a pBAD::*intI1* plasmid. The protocol was similar to the experiment illustrated in both **Figure 38** and **Figure 39**, with the exception of the *intI1* induction with arabinose for 4h before putting the cells in DASW with chitin (usually cells are grown in LB for ~2h incubation with chitin). We used the same PCR product to test recombination, pC167, this time also separately selecting for kanamycin resistance, thus first cassette integration events. For the natural transformation control, we used the same 1 kb homology PCR product. Intriguingly, recombination frequencies fell below 10^{-4}, suggesting that this protocol does not pose an advantage regarding integron recombination events (**Figure 42A**). On a positive note, we were able to detect first-cassette integration events, although at a lower frequency than the second. Natural transformation frequencies were recovered to normal values (**Figure 42B**) as previously demonstrated (**Figure 39**) and much higher than in our adapted chitin-independent assay (**Figure 41B**).

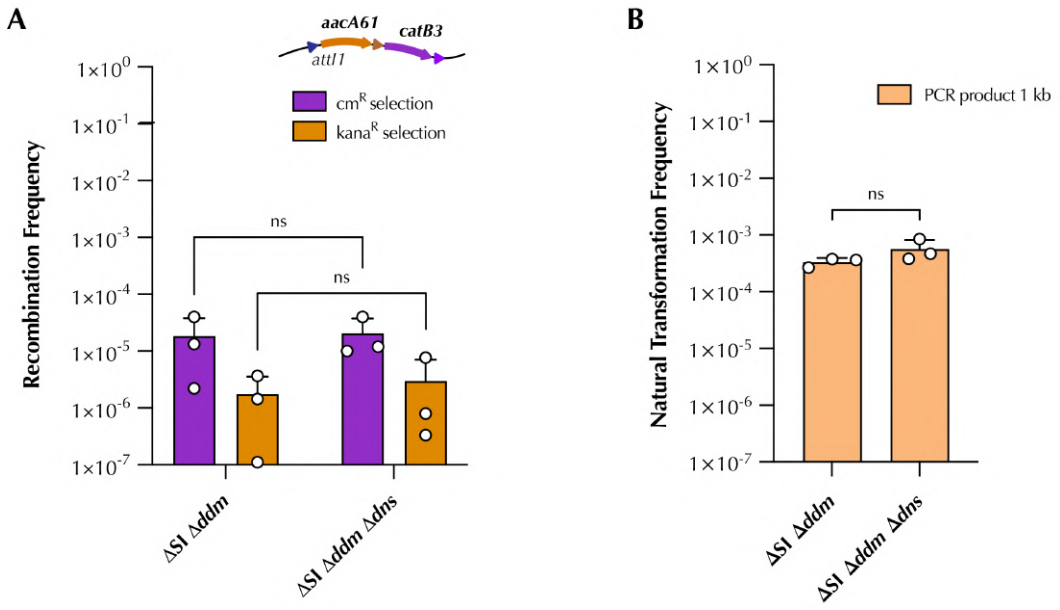


Figure 42. Classical natural transformation assay for integron recombination of selected *V. cholerae*$_{\Delta SI}$ mutants

Mutants derived from strain N16961 ΔSI were tested in a chitin-dependent assay upon *intI1* induction with arabinose at 0.2%. Recombination (**A**) and transformation (**B**) frequencies are given on the Y-axis and were calculated as the ratio of the number of CFU's growing in chloramphenicol for the selection if the first cassette, or kanamycin for the selection of the second cassette, and the total CFUs. Bar charts show the mean ± s.d. from biological replicates (individual dots). Statistical significance was determined using the Mann-Whitney U test; ns, not significant.

However—and as our goal is not homologous recombination, but integron recombination—we concluded that the chitin-independent protocol is superior. Importantly, as the statistical significance of the superiority in natural competence of the Δ*dns* or the double mutant Δ*ddm* Δ*dns* could not be proven, we chose the *V. cholerae* ΔSI Δ*ddm* to continue with our experiments.

4.2 – Integron cassette capture chassis

4.2.1 – *V. cholerae* setup

The final aim of the thesis was to integrate the re-engineered *V. cholerae* ΔSI Δ*ddm* for integron recombination through natural competence with our SacB-based tool, creating the integron cassette capture chassis (I3C). In this setup, the I3C containing the pBAD::*intI1-tfoX* plasmid would be co-cultured with the tested DNA sample, inducing the *intI1* and the *tfoX* genes to allow both natural transformation and integron recombination. Upon sucrose selection, only the transformants where integron cassettes had been excised and captured in our *sacB::attI1* platform would grow (**Figure 43**).

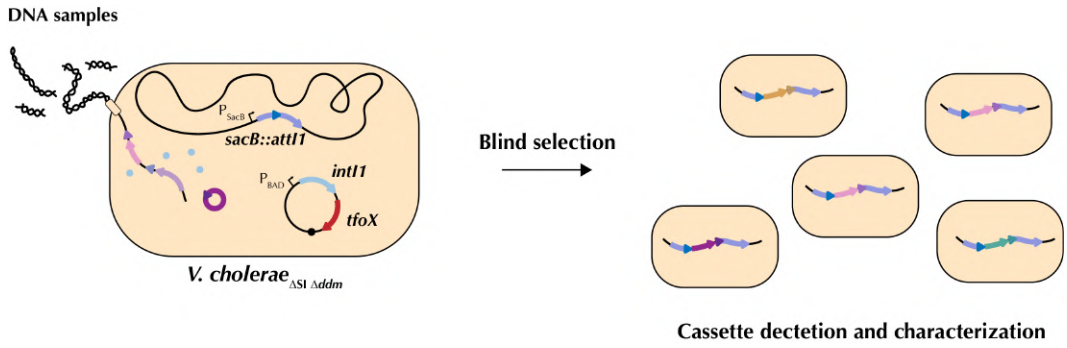


Figure 43. Integron cassette capture chassis

Schematic representation of the setup used to create the integron cassette capture chassis. In this system, the tested DNA sample is uptaken by natural transformation—upon induction of transformation by the master regulator TfoX—inside the *V. cholerae* mutant that contains a SacB-based version of our tool. After this, the expressed IntI1 can mobilize these integron cassettes inside the *sacB::attI1* platform. Then, the selection of recombinants is achieved through growth in sucrose.

4.2.2 – Testing DNA samples

We sought to prove that the I3C could detect integron cassettes directly from DNA samples. For this, we used the same PCR-amplified product pC167, containing the *aacA64* cassette in first position and the *catB3* cassette in second position. We also tested two additional PCR-amplified DNA samples: the integron array from the R388 plasmid, which contains a trimethoprim resistance gene, *dfrB2*, and a gene cassette of

unknown function (*gcuA*) (Avila & de la Cruz, 1988), as well as the modified version that includes a *dfrA5* gene, conferring resistance to trimethoprim, a bla_{VEB-1} gene, encoding an extended-spectrum β-lactamase, and *aadB*, conferring resistance to aminoglycosides (Souque et al., 2021). Additionally, we sought to test a more "realistic" sample. By extracting genomic DNA from the wild-type *V. cholerae* N16961, we aimed to test if our tool could capture cassettes from the superintegron. This had to be coupled with a previous step of amplification with the Phi29 polymerase and a small primer targeting the *attC* sites (**Figure 44A**). For more details on this amplification step, refer to page 90 of the Materials and Methods section. This step was essential as we were not able to detect cassettes from transforming the gDNA directly (data not shown). Finally, as a negative control, we also included a sample with no tDNA.

Upon induction of the *intI1* and *tfoX* genes, as previously described, the I3C and the tDNA were co-cultured overnight in DASW with chitin. After this, the mixture was plated in LB (ØNaCl) + sucrose 10% to select for recombinants. We recovered bacterial colonies at a frequency of 10^{-5} for the pC167 DNA sample and 10^{-6} for all the other samples, including the negative control (**Figure 44A**).

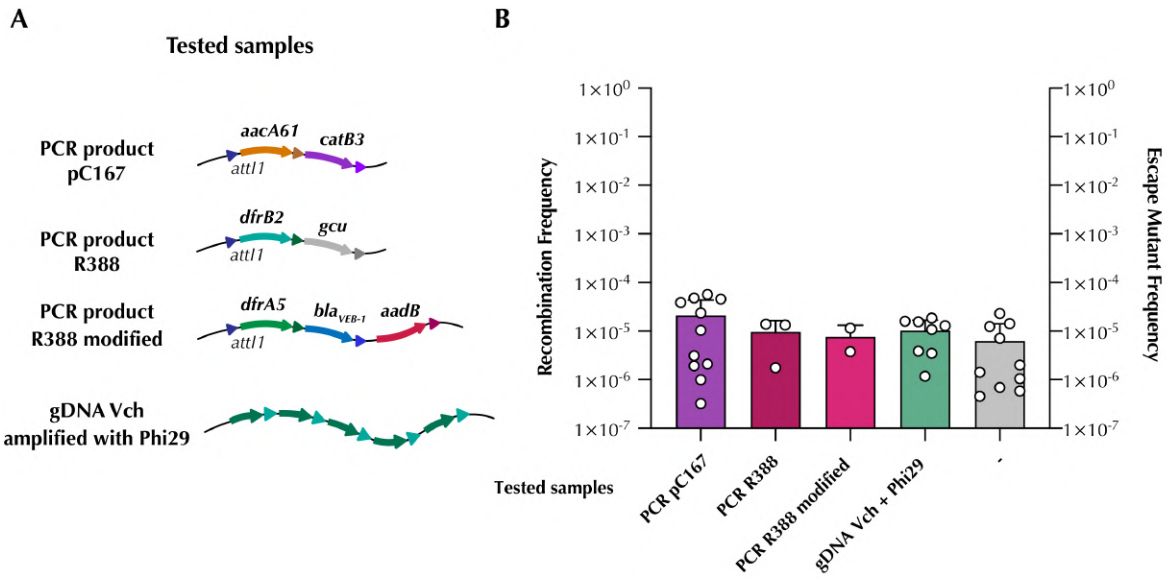


Figure 44. Chitin-independent natural transformation assay for integron recombination of the integron cassette capture chassis

A. Schematic representation of the integrons arrays contained in the tested DNA samples. Arrows represent genes, and triangles represent recombination sites. **B.** Recombination frequencies of the tested samples. The recombination frequencies were calculated as the ratio of the number of recombinants and the total CFUs. The escape mutant frequencies (depicted in grey) correspond to the sample where no DNA was co-cultured with the integron cassette capture chassis. Bar charts show the mean of the recombination frequency ± s.d. from biological replicates (individual dots).

Our results suggested that we could only detect integron cassettes from the PCR product pC167, and that the other were either negative or too close to the limit of detection. To confirm this, we performed a PCR amplifying the cassette content of *sacB::attI1*. We found that out of 40 colonies tested for the pC167 PCR sample, 24 were positive, i.e., 60%. For the R388 PCR samples, 14/16 tested colonies were positive, i.e, 87.5%. For the modified R388 PCR product, 1/8 were positive, i.e., 12.5%. Finally, 24/108 colonies were positive for the *Vibrio cholerae* WT gDNA sample, i.e., 22.2%. Details on the prevalence of the superintegron cassettes found in this experiment are described in **Supplementary Table 8**, comparing them to the cassette harvester frequencies. These results prove that our tool is capable of detecting integron cassettes from DNA samples, although further improvement in the efficiency of transformation and recombination is needed. It is of note that, in this case, we could not compare blind selection with phenotype selection. Indeed, when growing our control sample in chloramphenicol to detect the integration of the *catB3* cassette, we could not select for recombinants. As in this setup the *attI1* site is located ~1100 bp away from the P_{SacB} promoter, we assumed that the expression when the *sacB* gene is not translated is too low to confer a resistance phenotype. In fact, it would be interesting to compare the results to the previous experiments, where we obtained recombination frequencies of 10^{-4} using the same PCR-amplified samples (see **Figure 41**). Although the lack of expression limits the comparison to phenotype-selection results, this could confer an advantage when capturing potentially toxic cassettes.

The *sacB::attI1* platform optimized for *V. cholerae* do detect ICs directly from DNA samples is currently being evaluated for patent protection with the reference number P202430387.

In this Chapter, we achieved the successful deletion of the superintegron in *Vibrio cholerae*, a long-awaited milestone in the field that can unblock the study of sedentary chromosomal integrons. After the deletion of more than 126 kb of the genome, we showed that the natural competence of the ΔSI mutant is unaltered. Through the optimization of this process for integron recombination, we demonstrate, for the first time, that this bacterium can excise and capture a single integron cassette

from linear DNA. Finally, we combined the generated *V. cholerae* ΔSI Δ*ddm* with our SacB-based tool, demonstrating that we can efficiently capture and detect integron cassettes directly from DNA samples. The development of this tool marks a significant advancement in the rapid and direct identification of integron cassettes, expanding the potential for antimicrobial resistance tracking and the discovery of novel genetic elements in complex environments.

Discussion

Discussion

Integrons are genetic elements that facilitate the capture, integration, and expression of integron cassettes, playing a key role in the spread of antimicrobial resistance through horizontal gene transfer (Escudero et al., 2015). Beyond the antimicrobial resistance crisis, integrons have coexisted with bacteria in their chromosomes for eons, encoding genes with a panoply of diverse functions—many of which remain unexplored—including virulence, metabolism, phage resistance, and environmental adaptation (Blanco et al., 2024; Boucher et al., 2007, 2011; Darracq et al., 2024; Iqbal et al., 2015; Rapa & Labbate, 2013). Their ability to stockpile, shuffle, and disseminate these genes into the clinical setting has significantly shaped bacterial evolution and poses a challenge in combating multidrug-resistant bacteria.

Despite their prevalence and importance, detecting and studying integron cassettes remains difficult due to the limitations of traditional methods like PCRs (Sandoval-Quintana et al., 2023). These challenges call for innovative approaches that can capture integron cassettes independently of their genetic context, origin, or phenotype conferred.

In this thesis, three novel biotechnological tools were developed to address these challenges, each offering distinct applications in cassette detection. These tools not only advance the practical ability to study integrons but also provide mechanistic insights into integron recombination.

The following sections will discuss the key findings, compare these tools with existing methods, and explore the broader implications and future perspectives of this work in understanding the implication of integrons in AR and more.

Establishment of a platform to select for integron recombination events

To establish an efficient platform to detect integron recombination events, we first evaluated both constitutive and inducible promoters to assess their efficiency in *Escherichia coli* and *Vibrio cholerae*, the two main chassis for our integron detection tool. This characterization allowed us to better select the promoters controlling the

different parts of the capture platform, specifically the counter-selectable marker and the integrase gene.

The tool was constructed under the premise that integron recombination is very specific; indeed, integron integrases can recognize a wide variety of *attC* sites but not other non-integron recombination sites. Utilizing a recovery system based on the recombination activity of integron integrases not only increases the chances of recovering integron cassettes from samples—especially when compared to traditional PCR detection methods—but also allows the correct retrieval of cassettes within their *true* biological limits (i.e., their *attC* sites). We decided to use a class 1 integron, as it has been demonstrated that the IntI1 has a broad-spectrum activity for other integron classes (Biskri et al., 2005).

In order to establish a reliable method for selecting integron recombination, we carefully chose a counter-selectable marker. Positive selection markers, such as antibiotic resistance, are commonly used because they provide a selective advantage to the host organism. However, this type of selectable marker was not suitable for our needs. Negative or counter-selectable markers kill or inhibit the growth of the host organism upon selection. If interrupted or lost, bacteria are able to grow. We used two different negative CSMs: the CcdB/CcdA type II toxin-antitoxin system and the SacB levansucrase enzyme.

In this thesis, we demonstrate that the introduction of an integron recombination site (*attI*) in large linking regions of different CSMs does not affect the toxicity of the fusion CSM::*attI1* protein. Furthermore, we enabled blind selection of integron recombination events by disruption of the CSM. While interrupting a CSM is a strategy widely employed in molecular biology, with examples such as the pTOPO plasmid, where a *ccdB* gene avoids the relegation of the empty backbone, it had never been adapted to integron recombination. This, together with other pioneer studies in the field of integrons (Bikard et al., 2010b; Gestal et al., 2011; Nivina et al., 2020; Rowe-Magnus, 2009), highlights the biotechnological potential of these platforms. The efficiency in blind selection was slightly lower in a plasmid-based context. This was due to the high copy number of plasmids compared to the monocopy nature of chromosomal DNA, which can result in inconsistent recombination events across

plasmid copies. Nevertheless, we proved that both are useful and applicable to different purposes in the following chapters.

Despite the success of this platform, expanding the range of CSMs—such as testing additional TA systems (López-Igual et al., 2019) or employing widely used markers like TetA that, apart from conferring resistance to tetracycline, causes sensitivity to fusaric acid (Li et al., 2013)—could offer greater flexibility in diverse bacterial species. Moreover, the integration of an additional selection layer, such as CRISPR-Cas9, to target non-recombined cells, may enhance the precision and range of cassette capture. By expanding the CSM repertoire and incorporating such advanced methods, future iterations of this platform could significantly broaden its applicability. This would be especially valuable in environments where integron detection is challenging, such as soil or wastewater samples with high microbial diversity and low DNA concentration, as well as in clinical diagnostics where rapid and accurate monitoring of integron-borne resistance genes is critical.

Ultimately, the use of these reporters allowed for the precise detection of cassette insertion with a classical recombination assay, both in plasmid- and chromosome-based contexts. This platform played a crucial role in creating ideal conditions for high-throughput detection of integron cassettes, offering a dependable method for detecting integrons in various applications.

The cassette harvester

This chapter focused on constructing chromosomal integron cassette libraries capturing cassettes from the *Vibrio cholerae* superintegron, the paradigm of SCIs (Mazel, 2006). The setup used is a modified version of the plasmid-based tool developed in Chapter 1 to avoid the presence of the *intI1* gene in the final library.

The tool was highly efficient in the capture of integron cassettes since more than 10^5 bacteria were recovered, each corresponding to a recombination event. The background noise was neglectable: we calculated the ratio of the number of escape mutants and the recombined cells, and it only represents 0.1% of the library. By sequencing and analyzing the created library, we were able to confirm the high specificity of the tool, as 99.9% of the reads mapped to the superintegron. Furthermore, we explored the frequency of each cassette within the library. Not surprisingly, some

cassettes were more represented than others, with a 1,000-fold difference. This is attributable to two main factors: the first one is **the size of the cassette**, as it is known that this could influence the frequency of excision (Loot et al., 2017). The distance between two *attC* sites likely imposes structural constraints during recombination, which could explain why smaller cassettes are excised more frequently. Indeed, our data supports this hypothesis (**Figure 33B**). Also, we cannot discard the bias introduced in the PCR amplification step, as it tends to favor smaller or more easily amplifiable fragments, which could also explain the differential representation observed in the library. The second factor is **the toxicity of the gene**, as some underrepresented cassettes could be lethal if expressed, such as the recently discovered phage-defense systems (Darracq et al., 2024). This highlights the need for a promoterless version of the tool, where the CSM can be induced but not the gene captured—such as in the I3C— to ensure that all cassettes are more fairly represented. Another interesting finding is that in the majority of the cases, only one cassette was captured, as 95% of the reads were mapped to only one cassette. This supports our previous observation on the size of the cassettes excised, i.e., the excision of two adjacent cassettes is unlikely. It also suggests that integron recombination is inherently more efficient at capturing individual cassettes, possibly due to the natural constraints of the process or the time limit imposed by the 4-hour *intI1* induction step. Moreover, the high success rate in capturing cassettes from another integron class demonstrates the versatility of our system, further validating the efficiency of class 1 integrons as a broad-spectrum platform for IC capture.

Additionally, adapting the tool to other species, such as changing the origin of replication and the resistance markers—and potentially combining all the tool elements in one plasmid—could be particularly valuable in producing libraries from other SCIs. Our method, both experimental and analytical, ensures that the libraries produced are extremely reliable.

The creation of this cassette library serves as a valuable resource for studying the potential functions encoded in SCIs. Beyond *V. cholerae*, adapting the tool to other bacterial species could unveil the functional diversity of integrons across different environments. By refining the method to control for gene toxicity and bias, future ameliorations of this tool could ensure even greater representativity of cassettes,

paving the way for the discovery of novel functions that contribute to bacterial adaptability and resistance.

A sentinel bacterium

Here, we developed and optimized a sentinel bacterium approach using *E. coli* as a host to detect integrons present in mobile clinical plasmids. By engineering *E. coli* with the SacB-based tool to serve as a recipient for conjugative plasmids, we were able to monitor the transfer and capture of integron cassettes from these plasmids.

This method relies on conjugation, where clinical plasmids are transferred into the *E. coli* chassis, and their integron content is detected through blind selection in sucrose. We have demonstrated that this application is highly efficient (100% of detection) by detecting integrons from clinical plasmids, namely the R388 (Avila & de la Cruz, 1988), a plasmid that contains resistance cassettes but also integron cassettes of unknown functions (*gcuA*). Furthermore, and because our tool recovered the total content of the analyzed integron arrays, we were also able to characterize the integron content of the tested strain by sequencing the captured cassettes in the obtained sentinel *E. coli* recombinants.

One noteworthy mechanistic insight from our experiment is that we were able to capture the first cassette of the integron array. This is particularly interesting given the nature of the recombination sites, where *attI* needs to be in dsDNA form and *attC* in ssDNA form for the *attI* × *attC* reaction to be favored—conditions that are met when the cassettes are in their free form. When both sites are on the same molecule, they are more likely to be in the same DNA form (either both in dsDNA or both in ssDNA), making excision via the *attI* × *attC* reaction less likely. However, our results show that capturing the first cassette is indeed possible, which solves a potential limitation of our tool: not being able to capture the first cassette. Another hypothesis to consider is that rearrangement could be occurring in the free DNA fragment, allowing the first cassette to be rearranged to another position, and then excised and captured. This could be tested by including a single-cassette integron array in future experiments. Such investigations could raise interesting questions about integron cassette mobility and recombination dynamics.

A limitation in these experiments is the relatively small number of plasmids and strains tested. Expanding our testing to include additional conjugative and mobilizable plasmids, as well as clinical strains, will provide further insights into the tool's diagnostic potential.

This sentinel bacterium approach provides a robust system for identifying integrons in clinical samples, enabling controlled and efficient surveillance of multidrug-resistant plasmids. One promising application of this sentinel *E. coli* could be its usage as a quick and inexpensive tool for detecting antimicrobial resistance, such as in the food industry, where timely surveillance is critical. Additionally, this tool could be applied in clinical settings for rapid screening of resistant pathogens or even in environmental monitoring to track resistance gene dissemination and as a proxy for human pollution (Gillings et al., 2015).

Vibrio cholerae as a chassis to test directly DNA samples

The final chapter of the results focused on the development of *Vibrio cholerae* as a chassis for testing environmental and clinical DNA samples for integron cassette capture. To create the ideal chassis, we engineered a strain of *V. cholerae* with its superintegron deleted (ΔSI). This is a long-awaited milestone in the field, as previous attempts were hampered by the presence of uncharacterized TA systems. The obtention of a ΔSI mutant has deepened the study of these large chromosomal integrons; indeed, we have showed in our lab that they are functionally and genetically isolated from the rest of the genome (Blanco & Trigo da Roza et al., 2024).

Moreover, we attempted to enhance *V. cholerae* natural competence for integron recombination. While we succeeded in the adaptation of the protocol for integron recombination—having demonstrated for the first time in *V. cholerae* that linear DNA can be uptaken by the cell and a single cassette can be excised and captured—we could not improve the transformation efficiency of our strain. This could be due to several factors. It is indeed hard to control the competence regulatory network, as much is still unknown. Moreover, if given the appropriate source of chitin, cells achieve their maximum competence capacity. In fact, we could not couple this with integron recombination, as the competence state is only reached after 24 hours of growth on chitin. At this point, the induced integrase is probably degraded. One of the potential

ameliorations for this could be the decoupling of *tfoX* and *intI1* expression. If *tfoX* is induced overnight, a very high transformation rate can be achieved (Chlebek et al., 2019). After this, the integrase can be induced for a few hours before the incubation with DNA. Future perspectives include the testing of this hypothesis.

Nevertheless, using this platform, we successfully tested various DNA samples and demonstrated that we can select blindly for integron recombination using directly DNA samples. The efficiencies were variable, from 12.5% to almost 100%. This is due to the proximity of the recombination frequencies to the background noise. Again, the finetuning of the setup could dramatically improve the tool. Likewise, the potential identification, deletion of other novel defense/restriction systems (Vizzarro et al., 2024), and further testing of mutants could be valuable to increase the efficiency of the tool, as much of the DNA might be degraded upon entry into the cell. Another important point is that we had to incorporate an amplification step for the gDNA sample tested. While this may appear similar to the biased approach used in traditional PCR detection methods, our strategy minimizes bias by using primers that specifically target the conserved region of the *attC* sites—the R box sequence, RYYYAAC. This ensures that *attC* sites from all integron classes are targeted for amplification, reducing the potential for selective bias. A similar PCR approach was developed in previous works (Ghaly et al., 2019; Ghaly et al., 2022; Stokes et al., 2001). However, the primers used in these studies present more constraints, and the recovered products are directly sequenced and analyzed. In contrast, our integron capture platform ensures that only “true” cassettes are retrieved after the amplification step.

Finally, as previously discussed in the results section, we believe that the location of the *attI1* site inside the *sacB* gene is too far from the promoter to ensure a correct expression, rendering the cassette silent. While this is a limitation in the current experiment to compare blind selection to phenotype selection, it could be extremely valuable to avoid a toxicity bias when capturing unknown cassettes, such as in the cassette harvester.

In this chapter, we demonstrate the potential of this system for detecting integron cassettes in diverse genetic backgrounds, especially from unculturable microorganisms. This work establishes the I3C as a powerful biotechnological tool for

capturing and analyzing integron cassettes directly from environmental and clinical DNA samples.

Concluding remarks

In the context of antimicrobial resistance detection, the tools developed in this thesis offer distinct advantages over existing methods. Traditional PCR and sequencing techniques, while rapid and efficient at identifying known resistance genes, are limited by their dependence on conserved sequences that are not present in all types of integrons (for an in-depth review, see page 60 of the introduction section). Similarly, phenotypic methods like disk diffusion and automated susceptibility testing, although reliable, require longer times for bacterial growth and can miss integron-mediated resistance that has yet to be expressed. In contrast, the **sentinel *E. coli*** overcomes this by detecting integron cassettes in less than 48 hours, making it suitable for rapid diagnostics in clinical and environmental settings. Moreover, the tools developed in this thesis allow for the direct capture of integron cassettes through recombination, enabling the detection and analysis of unknown or uncharacterized cassettes that might otherwise be missed by molecular techniques. This is particularly crucial in identifying novel resistance mechanisms that are not included in standard detection panels. Similarly, the **I3C** allows direct detection from DNA samples, making the detection of resistance possible in unculturable microorganisms.

Additionally, the **cassette harvester** adds to the resourcefulness of field integron research by enabling high-throughput library construction from sedentary chromosomal integrons. The **I3C** is also applicable to environmental samples, especially in the discovery of new cassettes, understanding their ecology, and potentially gaining new insight into the origin of integron cassettes. These tools are not only complementary but provide a multi-faceted approach to studying integron dynamics, moving beyond AR detection methods by offering mechanistic insights and broadening the scope of integron research.

In conclusion, the work presented in this thesis has introduced three powerful biotechnological tools that advance our understanding of integron dynamics and antimicrobial resistance. The **cassette harvester** enables the high-throughput capture

and analysis of integron cassettes, allowing for detailed studies of cassette frequency and function in *Vibrio cholerae* and other species. The **sentinel *E. coli* system** offers an efficient method for detecting integrons in mobile clinical plasmids, providing a valuable platform for tracking resistance gene transfer in clinical and environmental settings. Finally, the **I3C** enhances our ability to monitor integron cassettes directly from DNA samples, potentially offering new insights into the mobility and ecology of integron-borne genes. Together, these tools represent a significant leap forward in both the detection and functional study of integrons.

Conclusions

Lisboa, 2021

Conclusions

Through the development of diverse platforms and exploration of integron dynamics, this work introduces three distinct biotechnological applications of an integron cassette capture system, highlighting their impact on identifying integron cassettes and understanding recombination mechanisms.

1. The development of a platform for integron recombination has provided a versatile tool to study recombination using as a chassis both *Escherichia coli* and *Vibrio cholerae,* and two different counter-selectable markers: the CcdB/CcdA toxin-antitoxin system and the SacB levansucrase encoding gene. This platform allows for controlled identification of recombination events, offering an alternative to detect these platforms that is not reliant on the phenotype conferred by the cassette.

2. The construction of a comprehensive cassette library from the *Vibrio cholerae* superintegron has produced a rich repository of genetic diversity. This tool demonstrates the potential of an integron-based tool to harvest and characterize integron cassettes, opening avenues for future biotechnological applications, including gene discovery and functional analysis.

3. The utilization of *Escherichia coli* as a sentinel bacterium has proven effective in detecting integrons within mobile plasmids. This approach underscores the utility of *E. coli* as a biosensor for the identification and tracking of these genetic platforms, contributing to efforts in monitoring the dissemination of integrons and antimicrobial resistance in clinical and environmental contexts.

4. The optimization of *Vibrio cholerae* as a chassis for testing DNA samples via natural transformation assays has significantly enhanced the application of integron systems. This work establishes *V. cholerae* as a highly efficient tool for direct DNA environmental sampling and testing, furthering the potential for integron-based biotechnological interventions.

Bibliography

Segovia, 2021

Bibliography

Abe, K., Nomura, N., & Suzuki, S. (2020). Biofilms: Hot spots of horizontal gene transfer (HGT) in aquatic environments, with a focus on a new HGT mechanism. *FEMS Microbiology Ecology*, *96*(5), fiaa031. https://doi.org/10.1093/femsec/fiaa031

Abraham, E. P., & Chain, E. (1940). An enzyme from bacteria able to destroy penicillin. *Nature*, *146*, 837.

Adams, David. W., Stutzmann, S., Stoudmann, C., & Blokesch, M. (2019). DNA-uptake pili of Vibrio cholerae are required for chitin colonization and capable of kin recognition via sequence-specific self-interaction. *Nature Microbiology*, *4*(9), 1545–1557. https://doi.org/10.1038/s41564-019-0479-5

Aizenman, E., Engelberg-Kulka, H., & Glaser, G. (1996). An Escherichia coli chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: A model for programmed bacterial cell death. *Proceedings of the National Academy of Sciences*, *93*(12), 6059–6063. https://doi.org/10.1073/pnas.93.12.6059

Albert, M. J., Siddique, A. K., Islam, M. S., Faruque, A. S., Ansaruzzaman, M., Faruque, S. M., & Sack, R. B. (1993). Large outbreak of clinical cholera due to Vibrio cholerae non-O1 in Bangladesh. *Lancet (London, England)*, *341*(8846), 704. https://doi.org/10.1016/0140-6736(93)90481-u

Aluwihare, L. I., Repeta, D. J., Pantoja, S., & Johnson, C. G. (2005). Two Chemically Distinct Pools of Organic Nitrogen Accumulate in the Ocean. *Science*, *308*(5724), 1007–1010. https://doi.org/10.1126/science.1108925

Aminov, R. I., & Mackie, R. I. (2007). Evolution and ecology of antibiotic resistance genes. *FEMS Microbiology Letters*, *271*, 147–161. https://doi.org/10.1111/j.1574-6968.2007.00757.x

Amos, G. C. A., Ploumakis, S., Zhang, L., Hawkey, P. M., Gaze, W. H., & Wellington, E. M. H. (2018). The widespread dissemination of integrons throughout bacterial communities in a riverine system. *ISME Journal*, *12*, 681–691. https://doi.org/10.1038/s41396-017-0030-8

Antonova, E. S., & Hammer, B. K. (2015). Genetics of natural competence in Vibrio cholerae and other Vibrios. *Microbiology Spectrum*, *3*(3), VE-0010-2014. https://doi.org/10.1128/microbiolspec.ve-0010-2014

Arakawa, Y., Murakami, M., Suzuki, K., Ito, H., Wacharotayankun, R., Ohsuka, S., Kato, N., & Ohta, M. (1995). A novel integron-like element carrying the metallo-β-lactamase gene blaIMP. *Antimicrobial Agents and Chemotherapy*, *39*(7), 1612–1615. https://doi.org/10.1128/AAC.39.7.1612

Ares-Arroyo, M., Bernabe-Balas, C., Santos-Lopez, A., Baquero, M. R., Prasad, K. N., Cid, D., Martin-Espada, C., San Millan, A., & Gonzalez-Zorn, B. (2018). PCR-based analysis of ColE1 plasmids in clinical isolates and metagenomic samples reveals their importance as gene capture platforms. *Frontiers in Microbiology*, *9*, 469. https://doi.org/10.3389/fmicb.2018.00469

Arnold, B. J., Huang, I. T., & Hanage, W. P. (2022). Horizontal gene transfer and adaptive evolution in bacteria. *Nature reviews. Microbiology*, *20*(4), 206–218. https://doi.org/10.1038/s41579-021-00650-4

Attéré, S. A., Vincent, A. T., Paccaud, M., Frenette, M., & Charette, S. J. (2017). The role for the small cryptic plasmids as moldable vectors for genetic innovation in Aeromonas salmonicida subsp. Salmonicida. *Frontiers in Genetics*, *8*, 211. https://doi.org/10.3389/fgene.2017.00211

Avila, P., & de la Cruz, F. (1988). Physical and genetic map of the IncW plasmid R388. *Plasmid*, *20*(2), 155–157. https://doi.org/10.1016/0147-619X(88)90019-4

Barker, A., Clark, C. A., & Manning, P. A. (1994). Identification of VCR, a repeated sequence associated with a locus encoding a hemagglutinin in Vibrio cholerae O1. *Journal of Bacteriology*, *176*(17), 5450–5458. https://doi.org/10.1128/jb.176.17.5450-5458.1994

Bernard, P. (1996). Positive selection of recombinant DNA by CcdB. *BioTechniques*, *21*(2), 320–323. https://doi.org/10.2144/96212pf01

Bernard, P., & Couturier, M. (1992). Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *Journal of Molecular Biology*, *226*(3), 735–745. https://doi.org/10.1016/0022-2836(92)90629-x

Bik, E. M., Bunschoten, A. E., Gouw, R. D., & Mooi, F. R. (1995). Genesis of the novel epidemic Vibrio cholerae O139 strain: Evidence for horizontal transfer of genes involved in polysaccharide synthesis. *The EMBO Journal*, *14*(2), 209–216. https://doi.org/10.1002/j.1460-2075.1995.tb06993.x

Bikard, D., Julié-Galau, S., Cambray, G., & Mazel, D. (2010a). The synthetic integron: An in vivo genetic shuffling device. *Nucleic Acids Research*, *38*(15), e153. https://doi.org/10.1093/nar/gkq511

Bikard, D., Julié-Galau, S., Cambray, G., & Mazel, D. (2010b). The synthetic integron: An in vivo genetic shuffling device. *Nucleic Acids Research*, *38*(15), e153. https://doi.org/10.1093/nar/gkq511

Billal, D. S., Feng, J., Leprohon, P., Légaré, D., & Ouellette, M. (2011). Whole genome analysis of linezolid resistance in Streptococcus pneumoniae reveals resistance and compensatory mutations. *BMC Genomics*, *12*, 512. https://doi.org/10.1186/1471-2164-12-512

Biskri, L., Bouvier, M., Guérout, A.-M., Boisnard, S., & Mazel, D. (2005). Comparative

Study of Class 1 Integron and Vibrio cholerae Superintegron Integrase Activities. *Journal of Bacteriology*, *187*(5), 1740–1750. https://doi.org/10.1128/JB.187.5.1740

Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. V. (2015). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, *13*, 42–51. https://doi.org/10.1038/nrmicro3380

Blanco, P., Hernando-Amado, S., Reales-Calderon, J. A., Corona, F., Lira, F., Alcalde-Rico, M., Bernardini, A., Sanchez, M. B., & Martinez, J. L. (2016). Bacterial Multidrug Efflux Pumps: Much More Than Antibiotic Resistance Determinants. *Microorganisms*, *4*(1), 14. https://doi.org/10.3390/microorganisms4010014

Blanco, P., Hipólito, A., García-Pastor, L., Trigo da Roza, F., Toribio-Celestino, L., Ortega, A. C., Vergara, E., San Millán, Á., & Escudero, J. A. (2024). Identification of promoter activity in gene-less cassettes from Vibrionaceae superintegrons. *Nucleic Acids Research*, gkad1252. https://doi.org/10.1093/nar/gkad1252

Blanco, P., Trigo da Roza, F., Toribio-Celestino, L., García-Pastor, L., Caselli, N., Morón, A., Ojeda, F., Darracq, B., Vergara, E., Amaro, F., San Millán, Á., Skovgaard, O., Mazel, D., Loot, C., & Escudero, J. A. (2024). Chromosomal integrons are genetically and functionally isolated units of genomes. *Nucleic Acids Research*, 1–17. https://doi.org/10.1093/nar/gkae866

Blokesch, M. (2012). Chitin colonization, chitin degradation and chitin-induced natural competence of Vibrio cholerae are subject to catabolite repression. *Environmental Microbiology*, *14*(8), 1898–1912. https://doi.org/10.1111/j.1462-2920.2011.02689.x

Blokesch, M., & Schoolnik, G. K. (2008). The extracellular nuclease Dns and its role in natural transformation of Vibrio cholerae. *Journal of Bacteriology*, *190*(21), 7232–7240. https://doi.org/10.1128/JB.00959-08

Blomfield, I. C., Vaughn, V., Rest, R. F., & Eisenstein, B. I. (1991). Allelic exchange in Escherichia coli using the Bacillus subtilis sacB gene and a temperature-sensitive pSC101 replicon. *Molecular Microbiology*, *5*(6), 1447–1457. https://doi.org/10.1111/j.1365-2958.1991.tb00791.x

Boolchandani, M., D'Souza, A. W., & Dantas, G. (2019). Sequencing-based methods and resources to study antimicrobial resistance. *Nature Reviews Genetics*, *20*(6), 356–370. https://doi.org/10.1038/s41576-019-0108-4

Boucher, Y., Cordero, O. X., Takemura, A., Hunt, D. E., Schliep, K., Bapteste, E., Lopez, P., Tarr, C. L., & Polz, M. F. (2011). Local mobile gene pools rapidly cross species boundaries to create endemicity within global Vibrio cholerae populations. *mBio*, *2*(2), e00335-10. https://doi.org/10.1128/mBio.00335-10

Boucher, Y., Labbate, M., Koenig, J. E., & Stokes, H. W. (2007). Integrons: Mobilizable

platforms that promote genetic diversity in bacteria. *Trends in Microbiology*, *15*(7), 301–309. https://doi.org/10.1016/j.tim.2007.05.004

Bouvier, M., Demarre, G., & Mazel, D. (2005). Integron cassette insertion: A recombination process involving a folded single strand substrate. *EMBO Journal*, *24*, 4356–4367. https://doi.org/10.1038/sj.emboj.7600898

Bouvier, M., Ducos-Galand, M., Loot, C., Bikard, D., & Mazel, D. (2009). Structural features of single-stranded integron cassette attC sites and their role in strand selection. *PLoS Genetics*, *5*(9), e1000632. https://doi.org/10.1371/journal.pgen.1000632

Calvo-Villamañán, A., San Millán, Á., & Carrilero, L. (2022). Tackling AMR from a multidisciplinary perspective: A primer from education and psychology. *International Microbiology*, *26*(1), 1–9. https://doi.org/10.1007/s10123-022-00278-1

Cambray, G., Guerout, A.-M., & Mazel, D. (2010). Integrons. *Annual Review of Genetics*, *44*, 141–166. https://doi.org/10.1146/annurev-genet-102209-163504

Cambray, G., Sanchez-Alberola, N., Campoy, S., Guerin, É., Da Re, S., González-Zorn, B., Ploy, M. C., Barbé, J., Mazel, D., & Erill, I. (2011). Prevalence of SOS-mediated control of integron integrase expression as an adaptive trait of chromosomal and mobile integrons. *Mobile DNA*, *2*, 6. https://doi.org/10.1186/1759-8753-2-6

Carattoli, A., Bertini, A., Villa, L., Falbo, V., Hopkins, K. L., & Threlfall, E. J. (2005). Identification of plasmids by PCR-based replicon typing. *Journal of Microbiological Methods*, *63*(3), 219–228. https://doi.org/10.1016/j.mimet.2005.03.018

Carvalho, A., Hipólito, A., Trigo da Roza, F., García-Pastor, L., Vergara, E., Buendía, A., García-Seco, T., & Escudero, J. A. (2024). *The expression of integron arrays is shaped by the translation rate of cassettes* (p. 2024.03.26.586746). bioRxiv. https://doi.org/10.1101/2024.03.26.586746

Chiang, Y. N., Penadés, J. R., & Chen, J. (2019). Genetic transduction by phages and chromosomal islands: The new and noncanonical. *PLoS Pathogens*, *15*(8), e1007878. https://doi.org/10.1371/journal.ppat.1007878

Chlebek, J. L., Hughes, H. Q., Ratkiewicz, A. S., Rayyan, R., Wang, J. C.-Y., Herrin, B. E., Dalia, T. N., Biais, N., & Dalia, A. B. (2019). PilT and PilU are homohexameric ATPases that coordinate to retract type IVa pili. *PLOS Genetics*, *15*(10), e1008448. https://doi.org/10.1371/journal.pgen.1008448

Chun, J., Grim, C. J., Hasan, N. A., Lee, J. H., Choi, S. Y., Haley, B. J., Taviani, E., Jeon, Y.-S., Kim, D. W., Lee, J.-H., Brettin, T. S., Bruce, D. C., Challacombe, J. F., Detter, J. C., Han, C. S., Munk, A. C., Chertkov, O., Meincke, L., Saunders, E., … Colwell, R. R.

(2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic Vibrio cholerae. *Proceedings of the National Academy of Sciences*, *106*(36), 15442–15447. https://doi.org/10.1073/pnas.0907787106

Clemens, J. D., Nair, G. B., Ahmed, T., Qadri, F., & Holmgren, J. (2017). Cholera. *The Lancet*, *390*(10101), 1539–1549. https://doi.org/10.1016/S0140-6736(17)30559-7

Clokie, M. R., Millard, A. D., Letarov, A. V., & Heaphy, S. (2011). Phages in nature. *Bacteriophage*, *1*(1), 31–45. https://doi.org/10.4161/bact.1.1.14942

Collis, C. M., Grammaticopoulos, G., Briton, J., Stokes, H. W., & Hall, R. M. (1993). Site-specific insertion of gene cassettes into integrons. *Molecular Microbiology*, *9*(1), 41–52. https://doi.org/10.1111/j.1365-2958.1993.tb01667.x

Collis, C. M., & Hall, R. M. (1992). Gene cassettes from the insert region of integrons are excised as covalently closed circles. *Molecular Microbiology*, *6*(19), 2875–2885. https://doi.org/10.1111/j.1365-2958.1992.tb01467.x

Collis, C. M., Kim, M. J., Partridge, S. R., Stokes, H. W., & Hall, R. M. (2002). Characterization of the class 3 integron and the site-specific recombination system it determines. *Journal of Bacteriology*, *184*(11), 3017–3026. https://doi.org/10.1128/JB.184.11.3017-3026.2002

Collis, C. M., Recchia, G. D., Kim, M. J., Stokes, H. W., & Hall, R. M. (2001). Efficiency of recombination reactions catalyzed by class 1 integron integrase IntI1. *Journal of Bacteriology*, *183*(8), 2535–2542. https://doi.org/10.1128/JB.183.8.2535-2542.2001

Couturier, A., Virolle, C., Goldlust, K., Berne-Dedieu, A., Reuter, A., Nolivos, S., Yamaichi, Y., Bigot, S., & Lesterlin, C. (2023). Real-time visualisation of the intracellular dynamics of conjugative plasmid transfer. *Nature Communications*, *14*(1), 294. https://doi.org/10.1038/s41467-023-35978-3

Cury, J., Jové, T., Touchon, M., Néron, B., & Rocha, E. P. (2016). Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Research*, *44*(10), 4539–4550. https://doi.org/10.1093/nar/gkw319

Dalia, A. B., Lazinski, D. W., & Camilli, A. (2014a). Identification of a Membrane-Bound Transcriptional Regulator That Links Chitin and Natural Competence in *Vibrio cholerae*. *mBio*, *5*(1), e01028-13. https://doi.org/10.1128/mBio.01028-13

Dalia, A. B., McDonough, E., & Camilli, A. (2014b). Multiplex genome editing by natural transformation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(24), 8937–8942. https://doi.org/10.1073/pnas.1406478111

Dalia, T. N., Yoon, S. H., Galli, E., Barre, F.-X., Waters, C. M., & Dalia, A. B. (2017).

Enhancing multiplex genome editing by natural transformation (MuGENT) via inactivation of ssDNA exonucleases. *Nucleic Acids Research*, *45*(12), 7527–7537. https://doi.org/10.1093/nar/gkx496

Darracq, B., Littner, E., Brunie, M., Bos, J., Kaminski, P.-A., Depardieu, F., Slesak, W., Debatisse, K., Touchon, M., Bernheim, A., Bikard, D., Roux, F. L., Mazel, D., Rocha, E. P. C., & Loot, C. (2024). *Sedentary chromosomal integrons as biobanks of bacterial anti-phage defence systems* (p. 2024.07.02.601686). bioRxiv. https://doi.org/10.1101/2024.07.02.601686

DelaFuente, J., Rodriguez-Beltran, J., & Millan, A. S. (2020). Methods to Study Fitness and Compensatory Adaptation in Plasmid-Carrying Bacteria. *Methods in Molecular Biology*, *2075*, 371–382. https://doi.org/10.1007/978-1-4939-9877-7_26

DelaFuente, J., Toribio-Celestino, L., Santos-Lopez, A., León-Sampedro, R., Alonso-Del Valle, A., Costas, C., Hernández-García, M., Cui, L., Rodríguez-Beltrán, J., Bikard, D., Cantón, R., & San Millan, A. (2022). Within-patient evolution of plasmid-mediated antimicrobial resistance. *Nature Ecology & Evolution*, *6*(12), 1980–1991. https://doi.org/10.1038/s41559-022-01908-7

Demarre, G., Frumerie, C., Gopaul, D. N., & Mazel, D. (2007). Identification of key structural determinants of the IntI1 integron integrase that influence attC × attI1 recombination efficiency. *Nucleic Acids Research*, *35*(19), 6475–6489. https://doi.org/10.1093/nar/gkm709

Demarre, G., Guérout, A.-M., Matsumoto-Mashimo, C., Rowe-Magnus, D. A., Marlière, P., & Mazel, D. (2005). A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalpha) conjugative machineries and their cognate Escherichia coli host strains. *Research in Microbiology*, *156*(2), 245–255. https://doi.org/10.1016/j.resmic.2004.09.007

Derbise, A., Lesic, B., Dacheux, D., Ghigo, J. M., & Carniel, E. (2003). A rapid and simple method for inactivating chromosomal genes in Yersinia. *FEMS Immunology and Medical Microbiology*, *38*(2), 113–116. https://doi.org/10.1016/S0928-8244(03)00181-0

Dubnau, D. (1999). DNA uptake in bacteria. *Annual Review of Microbiology*, *53*, 217–244.

Dubnau, D., & Blokesch, M. (2019). Mechanisms of DNA Uptake by Naturally Competent Bacteria. *Annual Review of Genetics*, *53*(1), 217–237. https://doi.org/10.1146/annurev-genet-112618-043641

Escudero, J. A., Loot, C., Parissi, V., Nivina, A., Bouchier, C., & Mazel, D. (2016). Unmasking the ancestral activity of integron integrases reveals a smooth evolutionary transition during functional innovation. *Nature Communications*, *7*, 10937. https://doi.org/10.1038/ncomms10937

Escudero, J. A., & Mazel, D. (2017). Genomic Plasticity of Vibrio cholerae. *International Microbiology*, *20*(3), 138–148. https://doi.org/10.2436/20.1501.01.295

Escudero, J. A., Loot, C., Nivina, A., & Mazel, D. (2015). The Integron: Adaptation On Demand. *Microbiology Spectrum*, *13*, MDNA3-0019–2014. https://doi.org/10.1128/microbiolspec

Faruque, S. M., Albert, M. J., & Mekalanos, J. J. (1998). Epidemiology, genetics, and ecology of toxigenic Vibrio cholerae. *Microbiology and Molecular Biology Reviews*, *62*(4), 1301–1314. https://doi.org/10.1128/mmbr.62.4.1301-1314.1998

Fillol-Salom, A., Bacarizo, J., Alqasmi, M., Ciges-Tomas, J. R., Martínez-Rubio, R., Roszak, A. W., Cogdell, R. J., Chen, J., Marina, A., & Penadés, J. R. (2019). Hijacking the hijackers: Escherichia coli pathogenicity islands redirect helper phage packaging for their own benefit. *Molecular Cell*, *75*, 1020–1030. https://doi.org/10.1016/j.molcel.2019.06.017

Finkel, S. E., & Kolter, R. (2001). DNA as a nutrient: Novel role for bacterial competence gene homologs. *Journal of Bacteriology*, *183*(21), 6288–6293. https://doi.org/10.1128/JB.183.21.6288-6293.2001

Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *British Journal of Experimental Pathology*, *10*(3), 226.

Florian Fricke, W., & Rasko, D. A. (2014). Bacterial genome sequencing in the clinic: Bioinformatic challenges and solutions. *Nature Reviews Genetics*, *15*(1), 49–55. https://doi.org/10.1038/nrg3624

Fogarty, E. C., Schechter, M. S., Lolans, K., Sheahan, M. L., Veseli, I., Moore, R. M., Kiefl, E., Moody, T., Rice, P. A., Yu, M. K., Mimee, M., Chang, E. B., Ruscheweyh, H.-J., Sunagawa, S., Mclellan, S. L., Willis, A. D., Comstock, L. E., & Eren, A. M. (2024). A cryptic plasmid is among the most numerous genetic elements in the human gut. *Cell*, *187*(5), 1206-1222.e16. https://doi.org/10.1016/j.cell.2024.01.039

Francia, M. V., Zabala, J. C., De La Cruz, F., & García Lobo, J. M. (1999). The IntI1 integron integrase preferentially binds single-stranded DNA of the attC site. *Journal of Bacteriology*, *181*(21), 6844–6849. https://doi.org/10.1128/jb.181.21.6844-6849.1999

Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: The agents of open source evolution. *Nature Reviews Microbiology*, *3*(9), 722–732. https://doi.org/10.1038/nrmicro1235

Fullner, K. J., & Mekalanos, J. J. (1999). Genetic Characterization of a New Type IV-A Pilus Gene Cluster Found in Both Classical and El Tor Biotypes ofVibrio cholerae. *Infection and Immunity*, *67*(3), 1393–1404.

https://doi.org/10.1128/iai.67.3.1393-1404.1999

Gao, W., Chua, K., Davies, J. K., Newton, H. J., Seemann, T., Harrison, P. F., Holmes, N. E., Rhee, H. W., Hong, J. I., Hartland, E. L., Stinear, T. P., & Howden, B. P. (2010). Two novel point mutations in clinical Staphylococcus aureus reduce linezolid susceptibility and switch on the stringent response to promote persistent infection. *PLoS Pathogens*, *6*(6), e1000944. https://doi.org/10.1371/journal.ppat.1000944

Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T., & Kado, C. I. (1985). Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *Journal of Bacteriology*, *164*(2), 918–921. https://doi.org/10.1128/jb.164.2.918-921.1985

Gerdes, K., Christensen, S. K., & Løbner-Olesen, A. (2005). Prokaryotic toxin–antitoxin stress response loci. *Nature Reviews Microbiology*, *3*(5), 371–382. https://doi.org/10.1038/nrmicro1147

Germain, E., Castro-Roa, D., Zenkin, N., & Gerdes, K. (2013). Molecular Mechanism of Bacterial Persistence by HipA. *Molecular Cell*, *52*(2), 248–254. https://doi.org/10.1016/j.molcel.2013.08.045

Gestal, A. M., Liew, E. F., & Coleman, N. V. (2011). Natural transformation with synthetic gene cassettes: New tools for integron research and biotechnology. *Microbiology*, *157*, 3349–3360. https://doi.org/10.1099/mic.0.051623-0

Ghaly, T. M., Geoghegan, J. L., Alroy, J., & Gillings, M. R. (2019). High diversity and rapid spatial turnover of integron gene cassettes in soil. *Environmental Microbiology*, *21*(5), 1567–1574. https://doi.org/10.1111/1462-2920.14551

Ghaly, T. M., Penesyan, A., Pritchard, A., Qi, Q., Rajabal, V., Tetu, S. G., & Gillings, M. R. (2022). Methods for the targeted sequencing and analysis of integrons and their gene cassettes from complex microbial communities. *Microbial genomics*, *8*(3), 000788. https://doi.org/10.1099/mgen.0.000788

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Gillings, M. R., Boucher, Y., Labbate, M., Holmes, A., Krishnan, S., Holley, M., & Stokes, H. W. (2008). The evolution of class 1 integrons and the rise of antibiotic resistance. *Journal of Bacteriology*, *190*(14), 5095–5100. https://doi.org/10.1128/JB.00152-08

Gillings, M. R., Gaze, W. H., Pruden, A., Smalla, K., Tiedje, J. M., & Zhu, Y. G. (2015). Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME Journal*, *9*(6), 1269–1279. https://doi.org/10.1038/ismej.2014.226

Gravel, A., Fournier, B., & Roy, P. H. (1998). DNA complexes obtained with the integron

integrase Intl1 at the attl1 site. *Nucleic Acids Research*, *26*(19), 4347–4355. https://doi.org/10.1093/nar/26.19.4347

Griffith, F. (1928). The Significance of Pneumococcal Types. *The Journal of Hygiene*, *27*(2), 113–159. https://doi.org/10.1017/s0022172400031879

Grindley, N. D. F., Whiteson, K. L., & Rice, P. A. (2006). Mechanisms of site-specific recombination. *Annual Review of Biochemistry*, *75*, 567–605. https://doi.org/10.1146/annurev.biochem.73.011303.073908

Guerin, É., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Da Re, S., Gonzalez-Zorn, B., Barbé, J., Ploy, M. C., & Mazel, D. (2009). The SOS response controls integron recombination. *Science*, *324*, 1034. https://doi.org/10.1126/science.1172914

Guérout, A. M., Iqbal, N., Mine, N., Ducos-Galand, M., Van Melderen, L., & Mazel, D. (2013). Characterization of the phd-doc and ccd toxin-antitoxin cassettes from Vibrio superintegrons. *Journal of Bacteriology*, *195*(10), 2270–2283. https://doi.org/10.1128/JB.01389-12

Guzman, L. M., Belin, D., Carson, M. J., & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology*, *177*(14), 4121–4130. https://doi.org/10.1128/jb.177.14.4121-4130.1995

Hall, R. M., Brookes, D. E., & Stokes, H. W. (1991). Site-specific insertion of genes into integrons: Role of the 59-base element and determination of the recombination cross-over point. *Molecular Microbiology*, *5*(8), 1941–1959. https://doi.org/10.1111/j.1365-2958.1991.tb00817.x

Hansson, K., Sköld, O., & Sundström, L. (1997). Non-palindromic attl sites of integrons are capable of site-specific recombination with one another and with secondary targets. *Molecular Microbiology*, *26*(3), 441–453. https://doi.org/10.1046/j.1365-2958.1997.5401964.x

Harms, K., Starikova, I., & Johnsen, P. J. (2013). Costly class-1 integrons and the domestication of the the functional integrase. *Mobile Genetic Elements*, *3*(2), e24774. https://doi.org/10.4161/mge.24774

Hawkey, J., Cottingham, H., Tokolyi, A., Wick, R. R., Judd, L. M., Cerdeira, L., de Oliveira Garcia, D., Wyres, K. L., & Holt, K. E. (2022). Linear plasmids in Klebsiella and other Enterobacteriaceae. *Microbial Genomics*, *8*(4), 000807. https://doi.org/10.1099/mgen.0.000807

Haycocks, J. R. J., Warren, G. Z. L., Walker, L. M., Chlebek, J. L., Dalia, T. N., Dalia, A. B., & Grainger, D. C. (2019). The quorum sensing transcription factor AphA directly regulates natural competence in *Vibrio cholerae*. *PLOS Genetics*, *15*(10), e1008362. https://doi.org/10.1371/journal.pgen.1008362

Heinemann, J. A., & Sprague, G. F. (1989). Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature*, *340*(6230), 205–209. https://doi.org/10.1038/340205a0

Hernando-Amado, S., Coque, T. M., Baquero, F., & Martínez, J. L. (2019). Defining and combating antibiotic resistance from One Health and Global Health perspectives. *Nature Microbiology*, *4*(9), 1432–1442. https://doi.org/10.1038/s41564-019-0503-9

Herrington, D. A., Hall, R. H., Losonsky, G., Mekalanos, J. J., Taylor, R. K., & Levine, M. M. (1988). Toxin, toxin-coregulated pili, and the toxR regulon are essential for Vibrio cholerae pathogenesis in humans. *The Journal of Experimental Medicine*, *168*(4), 1487–1492. https://doi.org/10.1084/jem.168.4.1487

Hinnebusch, J., & Tilly, K. (1993). Linear plasmids and chromosomes in bacteria. *Molecular Microbiology*, *10*(5), 917–922. https://doi.org/10.1111/j.1365-2958.1993.tb00963.x

Hipólito, A., García-Pastor, L., Blanco, P., Trigo da Roza, F., Kieffer, N., Vergara, E., Jové, T., Álvarez, J., & Escudero, J. A. (2022). The expression of aminoglycoside resistance genes in integron cassettes is not controlled by riboswitches. *Nucleic Acids Research*, *50*(15), 8566–8579. https://doi.org/10.1093/nar/gkac662

Hipólito, A., García-Pastor, L., Vergara, E., Jové, T., & Escudero, J. A. (2023). Profile and resistance levels of 136 integron resistance genes. *Npj Antimicrobials and Resistance*, *1*(1), 1–12. https://doi.org/10.1038/s44259-023-00014-3

Hochhut, B., Lotfi, Y., Mazel, D., Faruque, S. M., Woodgate, R., & Waldor, M. K. (2001). Molecular analysis of antibiotic resistance gene clusters in vibrio cholerae O139 and O1 SXT Constins. *Antimicrobial Agents and Chemotherapy*, *45*(11), 2991–3000. https://doi.org/10.1128/AAC.45.11.2991

Hofer, U. (2019). The cost of antimicrobial resistance. *Nature Reviews*, *17*, 2019.

Iqbal, N., Guérout, A. M., Krin, E., Le Roux, F., & Mazel, D. (2015). Comprehensive functional analysis of the 18 Vibrio cholerae N16961 toxin-antitoxin systems substantiates their role in stabilizing the superintegron. *Journal of Bacteriology*, *197*, 2150–2159. https://doi.org/10.1128/JB.00108-15

Jain, A., & Srivastava, P. (2013). Broad host range plasmids. *FEMS Microbiology Letters*, *348*(2), 87–96. https://doi.org/10.1111/1574-6968.12241

Jaskólska, M., Adams, D. W., & Blokesch, M. (2022). Two defence systems eliminate plasmids from seventh pandemic Vibrio cholerae. *Nature*, *604*(7905), 323–329. https://doi.org/10.1038/s41586-022-04546-y

Jaskólska, M., Stutzmann, S., Stoudmann, C., & Blokesch, M. (2018). QstR-dependent regulation of natural competence and type VI secretion in Vibrio cholerae. *Nucleic Acids Research*, *46*(20), 10619–10634.

https://doi.org/10.1093/nar/gky717

Joss, M. J., Koenig, J. E., Labbate, M., Polz, M. F., Gillings, M. R., Stokes, H. W., Doolittle, W. F., & Boucher, Y. (2009). ACID: Annotation of cassette and integron data. *BMC Bioinformatics*, *10*, 118. https://doi.org/10.1186/1471-2105-10-118

Jové, T., Cerdeira, L., Jesus, T., Lincopan, N., & Ploy, M. C. (2019). Functional characterization of a novel class of mobile integrons. *FEMS Congress*.

Jové, T., Da Re, S., Denis, F., Mazel, D., & Ploy, M. C. (2010). Inverse correlation between promoter strength and excision activity in class 1 integrons. *PLoS Genetics*, *6*(1), e1000793. https://doi.org/10.1371/journal.pgen.1000793

Kamruzzaman, M., Patterson, J. D., Shoma, S., Ginn, A. N., Partridge, S. R., & Iredell, J. R. (2015). Relative strengths of promoters provided by common mobile genetic elements associated with resistance gene expression in Gram-negative bacteria. *Antimicrobial Agents and Chemotherapy*, *59*(8), 5088–5091. https://doi.org/10.1128/AAC.00420-15

Korch, S. B., & Hill, T. M. (2006). Ectopic Overexpression of Wild-Type and Mutant hipA Genes in Escherichia coli: Effects on Macromolecular Synthesis and Persister Formation. *Journal of Bacteriology*, *188*(11), 3826–3836. https://doi.org/10.1128/jb.01740-05

Krin, E., Baharoglu, Z., Sismeiro, O., Varet, H., Coppée, J.-Y., & Mazel, D. (2023). Systematic transcriptome analysis allows the identification of new type I and type II Toxin/Antitoxin systems located in the superintegron of Vibrio cholerae. *Research in Microbiology*, *174*(1–2), 103997. https://doi.org/10.1016/j.resmic.2022.103997

Krin, E., Cambray, G., & Mazel, D. (2014). The superintegron integrase and the cassette promoters are co-regulated in Vibrio cholerae. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0091194

Labbate, M., Boucher, Y., Joss, M. J., Michael, C. A., Gillings, M. R., & Stokes, H. W. (2007). Use of chromosomal integron arrays as a phylogenetic typing system for Vibrio cholerae pandemic strains. *Microbiology (Reading, England)*, *153*(Pt 5), 1488–1498. https://doi.org/10.1099/mic.0.2006/001065-0

Laub, M. T., & Typas, A. (2024). Principles of bacterial innate immunity against viruses. *Current Opinion in Immunology*, *89*, 102445. https://doi.org/10.1016/j.coi.2024.102445

Laxminarayan, R., Sridhar, D., Blaser, M., Wang, M., & Woolhouse, M. (2016). Achieving global targets for antimicrobial resistance. *Science*, *352*, 874–875. https://doi.org/10.1126/science.aaf9286

Leclercq, R. (2002). Mechanisms of resistance to macrolides and lincosamides: Nature of the resistance elements and their clinical implications. *Clinical Infectious*

Diseases, *34*, 482–492. https://doi.org/10.1086/324626

Lederberg, J., & Tatum, E. L. (1946). Gene Recombination in Escherichia Coli. *Nature*, *158*(4016), 558–558. https://doi.org/10.1038/158558a0

Lévesque, C., Brassard, S., Lapointe, J., & Roy, P. H. (1994). Diversity and relative strength of tandem promoters for the antibiotic-resistance genes of several integrons. *Gene*, *142*(1), 49–54. https://doi.org/10.1016/0378-1119(94)90353-0

Lévesque, C., Piche, L., Larose, C., & Roy, P. H. (1995). PCR mapping of integrons reveals several novel combinations of resistance genes. *Antimicrobial Agents and Chemotherapy*, *39*(1), 185–191. https://doi.org/10.1128/aac.39.1.185

Li, X., Thomason, L. C., Sawitzke, J. A., Costantino, N., & Court, D. L. (2013). Positive and negative selection using the tetA-sacB cassette: Recombineering and P1 transduction in Escherichia coli. *Nucleic Acids Research*, *41*(22), e204. https://doi.org/10.1093/nar/gkt1075

Llano-Sotelo, B., Azucena, E. F., Kotra, L. P., Mobashery, S., & Chow, C. S. (2002). Aminoglycosides modified by resistance enzymes display diminished binding to the bacterial ribosomal aminoacyl-tRNA site. *Chemistry and Biology*, *9*, 455–463. https://doi.org/10.1016/S1074-5521(02)00125-4

Loot, C., Bikard, D., Rachlin, A., & Mazel, D. (2010). Cellular pathways controlling integron cassette site folding. *EMBO Journal*, *29*(15), 2623–2634. https://doi.org/10.1038/emboj.2010.151

Loot, C., Ducos-Galand, M., Escudero, J. A., Bouvier, M., & Mazel, D. (2012). Replicative resolution of integron cassette insertion. *Nucleic Acids Research*, *40*(17), 8361–8370. https://doi.org/10.1093/nar/gks620

Loot, C., Millot, G. A., Richard, E., Littner, E., Vit, C., Lemoine, F., Néron, B., Cury, J., Darracq, B., Niault, T., Lapaillerie, D., Parissi, V., Rocha, E. P. C., & Mazel, D. (2024). Integron cassettes integrate into bacterial genomes via widespread non-classical attG sites. *Nature Microbiology*, *9*(1), 228–240. https://doi.org/10.1038/s41564-023-01548-y

Loot, C., Nivina, A., Cury, J., Escudero, J. A., Ducos-Galand, M., Bikard, D., Rocha, E. P. C., & Mazel, D. (2017). Differences in Integron Cassette Excision Dynamics Shape a Trade-Off between Evolvability and Genetic Capacitance. *mBio*, *8*(2), e02296-16. https://doi.org/10.1128/mBio.02296-16

López-Igual, R., Bernal-Bayard, J., Rodríguez-Patón, A., Ghigo, J.-M., & Mazel, D. (2019). Engineered toxin–intein antimicrobials can selectively target and kill antibiotic-resistant bacteria in mixed populations. *Nature Biotechnology*, *37*(7), 755–760. https://doi.org/10.1038/s41587-019-0105-3

MacDonald, D., Demarre, G., Bouvier, M., Mazel, D., & Gopaul, D. N. (2006). Structural

basis for broad DNA-specificity in integron recombination. *Nature*, *440*, 1157–1162. https://doi.org/10.1038/nature04643

MacLean, R. C., & San Millan, A. (2019). The evolution of antibiotic resistance. *Science*, *365*(6458), 1082–1083. https://doi.org/10.1126/science.aax3879

Mahillon, J., & Chandler, M. (1998). Insertion Sequences. *Microbiology and Molecular Biology Reviews*, *62*(3), 725–774. https://doi.org/10.1128/mmbr.62.3.725-774.1998

Marvig, R. L., & Blokesch, M. (2010). Natural transformation of *Vibrio cholerae* as a tool Optimizing the procedure. *BMC Microbiology*, *10*, 155.

Marx, C. J. (2008). Development of a broad-host-range sacB-based vector for unmarked allelic exchange. *BMC Research Notes*, *1*, 1. https://doi.org/10.1186/1756-0500-1-1

Matthey, N., & Blokesch, M. (2016). The DNA-Uptake Process of Naturally Competent *Vibrio cholerae*. *Trends in Microbiology*, *24*(2), 98–110. https://doi.org/10.1016/j.tim.2015.10.008

Mazel, D. (2006). Integrons: Agents of bacterial evolution. *Nature Reviews Microbiology*, *4*(8), 608–620. https://doi.org/10.1038/nrmicro1462

Mazel, D., Dychinco, B., Webb, V. A., & Davies, J. (1998). A distinctive class of integron in the Vibrio cholerae genome. *Science*, *280*, 605–608. https://doi.org/10.1126/science.280.5363.605

Meibom, K. L., Blokesch, M., Dolganov, N. A., Wu, C.-Y., & Schoolnik, G. K. (2005). Chitin Induces Natural Competence in *Vibrio cholerae*. *Science*, *310*(5755), 1824–1827. https://doi.org/10.1126/science.1120096

Meibom, K. L., Li, X. B., Nielsen, A. T., Wu, C.-Y., Roseman, S., & Schoolnik, G. K. (2004). The Vibrio cholerae chitin utilization program. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(8), 2524–2529. https://doi.org/10.1073/pnas.0308707101

Mekalanos, J. J., Swartz, D. J., Pearson, G. D., Harford, N., Groyne, F., & de Wilde, M. (1983). Cholera toxin genes: Nucleotide sequence, deletion analysis and vaccine development. *Nature*, *306*(5943), 551–557. https://doi.org/10.1038/306551a0

Melderen, L. V., & Bast, M. S. D. (2009). Bacterial Toxin–Antitoxin Systems: More Than Selfish Entities? *PLOS Genetics*, *5*(3), e1000437. https://doi.org/10.1371/journal.pgen.1000437

Meng, G., & Fütterer, K. (2003). Structural framework of fructosyl transfer in Bacillus subtilis levansucrase. *Nature Structural Biology*, *10*(11), 935–941. https://doi.org/10.1038/nsb974

Messier, N., & Roy, P. H. (2001). Integron integrases possess a unique additional domain

necessary for activity. *Journal of Bacteriology*, *183*(22), 6699–6706. https://doi.org/10.1128/JB.183.22.6699-6706.2001

Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J., & Voigt, C. A. (2019). Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. *Nature Chemical Biology*, *15*(2), 196–204. https://doi.org/10.1038/s41589-018-0168-3

Mitsuhashi, S., Harada, K., Hashimoto, H., & Egawa, R. (1961). On the drug-resistance of enteric bacteria. 4. Drug-resistance of Shigella prevalent in Japan. *The Japanese Journal of Experimental Medicine*, *31*, 47–52.

Moura, A., Jové, T., Ploy, M. C., Henriques, I., & Correia, A. (2012). Diversity of gene cassette promoters in class 1 integrons from wastewater environments. *Applied and Environmental Microbiology*, *78*(15), 5413–5416. https://doi.org/10.1128/AEM.00042-12

Moura, A., Soares, M., Pereira, C., Leitão, N., Henriques, I., & Correia, A. (2009). INTEGRALL: A database and search engine for integrons, integrases and gene cassettes. *Bioinformatics*, *25*(8), 1096–1098. https://doi.org/10.1093/bioinformatics/btp105

Mukherjee, S., & Bassler, B. L. (2019). Bacterial quorum sensing in complex and dynamically changing environments. *Nature Reviews. Microbiology*, *17*(6), 371–382. https://doi.org/10.1038/s41579-019-0186-5

Murray, C. J. L., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., McManigal, B., … Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: A systematic analysis. *The Lancet*, *399*(10325), 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0

Naas, T., Mikami, Y., Imai, T., Poirel, L., & Nordmann, P. (2001). Characterization of In53, a class 1 plasmid- and composite transposon-located integron of Escherichia coli which carries an unsual array of gene cassettes. *Journal of Bacteriology*, *183*(1), 235–249. https://doi.org/10.1128/JB.183.1.235

Naas, T., Oueslati, S., Bonnin, R. A., Dabos, M. L., Zavala, A., Dortet, L., Retailleau, P., & Iorga, B. I. (2017). Beta-lactamase database (BLDB) – structure and function. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *32*(1), 917–919. https://doi.org/10.1080/14756366.2017.1344235

Nemergut, D. R., Robeson, M. S., Kysela, R. F., Martin, A. P., Schmidt, S. K., & Knight, R. (2008). Insights and inferences about integron evolution from genomic data. *BMC Genomics*, *9*, 1–12. https://doi.org/10.1186/1471-2164-9-261

Néron, B., Littner, E., Haudiquet, M., Perrin, A., Cury, J., & Rocha, E. P. C. (2022).

IntegronFinder 2.0: Identification and Analysis of Integrons across Bacteria, with a Focus on Antibiotic Resistance in Klebsiella. *Microorganisms*, *10*(4), Article 4. https://doi.org/10.3390/microorganisms10040700

Nivina, A., Escudero, J. A., Vit, C., Mazel, D., & Loot, C. (2016). Efficiency of integron cassette insertion in correct orientation is ensured by the interplay of the three unpaired features of attC recombination sites. *Nucleic Acids Research*, *44*(16), 7792–7803. https://doi.org/10.1093/nar/gkw646

Nivina, A., Grieb, M. S., Loot, C., Bikard, D., Cury, J., Shehata, L., Bernardes, J., & Mazel, D. (2020). Structure-specific DNA recombination sites: Design, validation, and machine learning–based refinement. *Science Advances*, *6*(30), eaay2922. https://doi.org/10.1126/sciadv.aay2922

Nunes-Düby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T., & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Research*, *26*(2), 391–406. https://doi.org/10.1093/nar/26.2.391

Ogawa, A., & Takeda, T. (1993). The gene encoding the heat-table Enterotoxin of Vibrio cholerae Is Flanked by 123-Base Pair Direct Repeats. *Microbiology and Immunology*, *37*(8), 607–616. https://doi.org/10.1111/j.1348-0421.1993.tb01683.x

Ogura, T., & Hiraga, S. (1983). Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proceedings of the National Academy of Sciences*, *80*(15), 4784–4788. https://doi.org/10.1073/pnas.80.15.4784

O'Neill, J. (2016). Tackling drug-resistant infections globally: Final report and recommendations. *The Review on Antimicrobial Resistance*. https://doi.org/10.4103/2045-080x.186181

Orth, P., Schnappinger, D., Hillen, W., Saenger, W., & Hinrichs, W. (2000). Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nature Structural Biology*, *7*(3), 215–219. https://doi.org/10.1038/73324

Palchevskiy, V., & Finkel, S. E. (2006). Escherichia coli competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient. *Journal of Bacteriology*, *188*(11), 3902–3910. https://doi.org/10.1128/JB.01974-05

Palmer, A. C., & Kishony, R. (2014). Opposing effects of target overexpression reveal drug mechanisms. *Nature Communications*, *5*, 4296. https://doi.org/10.1038/ncomms5296

Papagiannitsis, C. C., Tzouvelekis, L. S., & Miriagou, V. (2009). Relative strengths of the class 1 integron promoter hybrid 2 and the combinations of strong and hybrid 1 with an active P2 promoter. *Antimicrobial Agents and Chemotherapy*, *53*(1),

277–280. https://doi.org/10.1128/AAC.00912-08

Partridge, S. R., Kwong, S. M., Firth, N., & Jensen, S. O. (2018). Mobile genetic elements associated with antimicrobial resistance. *Clinical Microbiology Reviews*, *31*(4), 1–61. https://doi.org/10.1128/CMR.00088-17

Partridge, S. R., Recchia, G. D., Scaramuzzi, C., Collis, C. M., Stokes, H. W., & Hall, R. M. (2000). Definition of the attl1 site of class 1 integrons. *Microbiology*, *146*, 2855–2864. https://doi.org/10.1099/00221287-146-11-2855

Partridge, S. R., Tsafnat, G., Coiera, E., & Iredell, J. R. (2009). Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiology Reviews*, *33*, 757–784. https://doi.org/10.1111/j.1574-6976.2009.00175.x

Paton, A. M. (1960). The Role of Pseudomonas in Plant Disease. *Journal of Applied Bacteriology*, *23*(3), 526–532. https://doi.org/10.1111/j.1365-2672.1960.tb00224.x

Penadés, J. R., Chen, J., Quiles-Puchalt, N., Carpena, N., & Novick, R. P. (2015). Bacteriophage-mediated spread of bacterial virulence genes. *Current Opinion in Microbiology*, *23*, 171–178. https://doi.org/10.1016/j.mib.2014.11.019

Pereira, M. B., Wallroth, M., Kristiansson, E., & Axelson-Fisk, M. (2016). HattCI: Fast and Accurate attC site Identification Using Hidden Markov Models. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, *23*(11), 891–902. https://doi.org/10.1089/cmb.2016.0024

Perler, F. B. (2002). InBase: The Intein Database. *Nucleic Acids Research*, *30*(1), 383–384. https://doi.org/10.1093/nar/30.1.383

Ramírez, M. S., Piñeiro, S., & Centrón, D. (2010). Novel insights about class 2 integrons from experimental and genomic epidemiology. *Antimicrobial Agents and Chemotherapy*, *54*(2), 699–706. https://doi.org/10.1128/AAC.01392-08

Rao, S., Maddox, C. W., Hoien-Dalen, P., Lanka, S., & Weigel, R. M. (2008). Diagnostic accuracy of class 1 integron PCR method in detection of antibiotic resistance in Salmonella isolates from swine production systems. *Journal of Clinical Microbiology*, *46*(3), 916–920. https://doi.org/10.1128/JCM.01597-07

Rapa, R. A., & Labbate, M. (2013). The function of integron-associated gene cassettes in Vibrio species: The tip of the iceberg. *Frontiers in Microbiology*, *4*, 385. https://doi.org/10.3389/fmicb.2013.00385

Recchia, G. D., & Hall, R. M. (1995). Gene cassettes: A new class of mobile element. *Microbiology*, *141*(12), 3015–3027. https://doi.org/doi.org/10.1099/13500872-141-12-3015

Redondo-Salvo, S., Bartomeus-Peñalver, R., Vielva, L., Tagg, K. A., Webb, H. E., Fernández-López, R., & de la Cruz, F. (2021). COPLA, a taxonomic classifier of plasmids. *BMC Bioinformatics*, *22*(1), 390. https://doi.org/10.1186/s12859-

021-04299-x

Redondo-Salvo, S., Fernández-López, R., Ruiz, R., Vielva, L., de Toro, M., Rocha, E. P. C., Garcillán-Barcia, M. P., & de la Cruz, F. (2020). Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nature Communications*, *11*(1), 3602. https://doi.org/10.1038/s41467-020-17278-2

Reimmann, C., & Hass, D. (1993). Mobilization of Chromosomes and Nonconjugative Plasmids by Cointegrative Mechanisms. In D. B. Clewell (Ed.), *Bacterial Conjugation*. Springer.

Richard, E., Darracq, B., Littner, E., Vit, C., Whiteway, C., Bos, J., Fournes, F., Garriss, G., Conte, V., Lapaillerie, D., Parissi, V., Rousset, F., Skovgaard, O., Bikard, D., Rocha, E. P. C., Mazel, D., & Loot, C. (2024). Cassette recombination dynamics within chromosomal integrons are regulated by toxin-antitoxin systems. *Science Advances*, *10*, eadj3498.

Rodríguez-Beltrán, J., DelaFuente, J., León-Sampedro, R., MacLean, R. C., & San Millán, Á. (2021). Beyond horizontal gene transfer: The role of plasmids in bacterial evolution. *Nature Reviews Microbiology*, *19*(6), 347–359. https://doi.org/10.1038/s41579-020-00497-1

Rodríguez-Beltrán, J., Sørum, V., Toll-Riera, M., de la Vega, C., Peña-Miller, R., & San Millán, Á. (2020). Genetic dominance governs the evolution and spread of mobile genetic elements in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(27), 15755–15762. https://doi.org/10.1073/pnas.2001240117

Rowe-Magnus, D. A. (2009). Integrase-directed recovery of functional genes from genomic libraries. *Nucleic Acids Research*, *37*(17), 31–33. https://doi.org/10.1093/nar/gkp561

Rowe-Magnus, D. A., Guerout, A. M., Biskri, L., Bouige, P., & Mazel, D. (2003). Comparative analysis of superintegrons: Engineering extensive genetic diversity in the Vibrionaceae. *Genome Research*, *13*, 428–442. https://doi.org/10.1101/gr.617103

Rowe-Magnus, D. A., Guérout, A. M., & Mazel, D. (1999). Super-integrons. *Research in Microbiology*, *150*, 641–651. https://doi.org/10.1016/S0923-2508(99)00127-8

Rowe-Magnus, D. A., Guerout, A. M., & Mazel, D. (2002). Bacterial resistance evolution by recruitment of super-integron gene cassettes. *Molecular Microbiology*, *43*(6), 1657–1669. https://doi.org/10.1046/j.1365-2958.2002.02861.x

Rowe-Magnus, D. A., Guerout, A. M., Ploncard, P., Dychinco, B., Davies, J., & Mazel, D. (2001). The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *PNAS*, *98*(2), 652–657. https://doi.org/10.1073/pnas.98.2.652

Ruiz, N., & Silhavy, T. J. (2022). How Escherichia coli Became the Flagship Bacterium of Molecular Biology. *Journal of Bacteriology*, *204*(9), e00230-22. https://doi.org/10.1128/jb.00230-22

Sack, D. A., Sack, R. B., Nair, G. B., & Siddique, A. (2004). Cholera. *The Lancet*, *363*, 223–233. https://doi.org/10.1016/S0140-6736(03)15328-7

San Millan, A., Escudero, J. A., Gifford, D. R., Mazel, D., & MacLean, R. C. (2016). Multicopy plasmids potentiate the evolution of antibiotic resistance in bacteria. *Nature Ecology & Evolution*, *1*(1), 10. https://doi.org/10.1038/s41559-016-0010

Sandoval-Quintana, E., Lauga, B., & Cagnon, C. (2023). Environmental integrons: The dark side of the integron world. *Trends in Microbiology*, *31*(5), 432–434. https://doi.org/10.1016/j.tim.2022.01.009

Schweizer, H. P. (1992). Allelic exchange in Pseudomonas aeruginosa using novel ColE1-type vectors and a family of cassettes containing a portable oriT and the counter-selectable Bacillus subtilis sacB marker. *Molecular Microbiology*, *6*(9), 1195–1204. https://doi.org/10.1111/j.1365-2958.1992.tb01558.x

Seitz, P., & Blokesch, M. (2013a). Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental Gram-negative bacteria. *FEMS Microbiology Reviews*, *37*(3), 336–363. https://doi.org/10.1111/j.1574-6976.2012.00353.x

Seitz, P., & Blokesch, M. (2013b). DNA-uptake machinery of naturally competent *Vibrio cholerae*. *Proceedings of the National Academy of Sciences*, *110*(44), 17987–17992. https://doi.org/10.1073/pnas.1315647110

Sharma, D. K., Misra, H. S., Bihani, S. C., & Rajpurohit, Y. S. (2023). Biochemical Properties and Roles of DprA Protein in Bacterial Natural Transformation, Virulence, and Pilin Variation. *Journal of Bacteriology*, *205*(2), e00465-22. https://doi.org/10.1128/jb.00465-22

Siguier, P., Gourbeyre, E., & Chandler, M. (2014). Bacterial insertion sequences: Their genomic impact and diversity. *FEMS Microbiology Reviews*, *38*, 865–891. https://doi.org/10.1111/1574-6976.12067

Simpson, C. A., Podicheti, R., Rusch, D. B., Dalia, A. B., & van Kessel, J. C. (2019). Diversity in Natural Transformation Frequencies and Regulation across Vibrio Species. *mBio*, *10*(6), e02788-19. https://doi.org/10.1128/mBio.02788-19

Sorum, H., Roberts, M. C., & Crosa, J. H. (1992). Identification and cloning of a tetracycline resistance gene from the fish pathogen Vibrio salmonicida. *Antimicrobial Agents and Chemotherapy*, *36*(3), 611–615. https://doi.org/10.1128/AAC.36.3.611

Souque, C., Escudero, J. A., & Maclean, R. C. (2021). Integron activity accelerates the evolution of antibiotic resistance. *eLife*, *10*, 1–47.

https://doi.org/10.7554/eLife.62474

Starikova, I., Harms, K., Haugen, P., Lunde, T. T. M., Primicerio, R., Samuelsen, Ø., Nielsen, K. M., & Johnsen, P. J. (2012). A Trade-off between the fitness cost of functional integrases and long-term stability of integrons. *PLoS Pathogens*, *8*(11), e1003043. https://doi.org/10.1371/journal.ppat.1003043

Stokes, H. W., Gorman, D. B. O., Recchia, G. D., Parsekhian, M., & Hall, R. M. (1997). Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Molecular Microbiology*, *26*(4), 731–745.

Stokes, H. W., & Hall, R. M. (1989). A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: Integrons. *Molecular Microbiology*, *3*(12), 1669–1683. https://doi.org/10.1111/j.1365-2958.1989.tb00153.x

Stokes, H. W., Holmes, A. J., Nield, B. S., Holley, M. P., Nevalainen, K. M. H., Mabbutt, B. C., & Gillings, M. R. (2001). Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Applied and Environmental Microbiology*, *67*(11), 5240–5246. https://doi.org/10.1128/aem.67.11.5240-5246.2001

Stutzmann, S., & Blokesch, M. (2020). Comparison of chitin-induced natural transformation in pandemic *Vibrio cholerae* O1 El Tor strains. *Environmental Microbiology*, *22*(10), 4149–4166. https://doi.org/10.1111/1462-2920.15214

Suckow, G., Seitz, P., & Blokesch, M. (2011). Quorum Sensing Contributes to Natural Transformation of Vibrio cholerae in a Species-Specific Manner. *Journal of Bacteriology*, *193*(18), 4914–4924. https://doi.org/10.1128/JB.05396-11

Sureshan, V., Deshpande, C. N., Boucher, Y., Koenig, J. E., Genomics, Mc., Stokes, H. W., Harrop, S. J., Curmi, P. M. G., & Mabbutt, B. C. (2013). Integron gene cassettes: A repository of novel protein folds with distinct interaction sites. *PLoS ONE*, *8*(1), e52934. https://doi.org/10.1371/journal.pone.0052934

Szekeres, S., Dauti, M., Wilde, C., Mazel, D., & Rowe-Magnus, D. A. (2007). Chromosomal toxin–antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Molecular Microbiology*, *63*(6), 1588–1605. https://doi.org/10.1111/j.1365-2958.2007.05613.x

Szpirer, C. Y., & Milinkovitch, M. C. (2005). Separate-component-stabilization system for protein and DNA production without the use of antibiotics. *BioTechniques*, *38*(5), 775–781. https://doi.org/10.2144/05385RR02

Tamplin, M. L., Gauzens, A. L., Huq, A., Sack, D. A., & Colwell, R. R. (1990). Attachment of Vibrio cholerae serogroup O1 to zooplankton and phytoplankton of Bangladesh waters. *Applied and Environmental Microbiology*, *56*(6), 1977–1980. https://doi.org/10.1128/aem.56.6.1977-1980.1990

Thanbichler, M., Iniesta, A. A., & Shapiro, L. (2007). A comprehensive set of plasmids for

vanillate- and xylose-inducible gene expression in Caulobacter crescentus. *Nucleic Acids Research*, *35*(20), e137. https://doi.org/10.1093/nar/gkm818

Tran, J. H., Jacoby, G. A., & Hooper, D. C. (2005). Interaction of the plasmid-encoded quinolone resistance protein Qnr with Escherichia coli DNA gyrase. *Antimicrobial Agents and Chemotherapy*, *49*(1), 118–125. https://doi.org/10.1128/AAC.49.7.3050-3052.2005

Tsafnat, G., Coiera, E., Partridge, S. R., Schaeffer, J., & Iredell, J. R. (2009). Context-driven discovery of gene cassettes in mobile integrons using a computational grammar. *BMC Bioinformatics*, *10*, 281. https://doi.org/10.1186/1471-2105-10-281

Vakulenko, S. B., & Mobashery, S. (2003). Versatility of aminoglycosides and prospects for their future. *Clinical Microbiology Reviews*, *16*(3), 430–450. https://doi.org/10.1128/CMR.16.3.430-450.2003

Val, M.-E., Skovgaard, O., Ducos-Galand, M., Bland, M. J., & Mazel, D. (2012). Genome Engineering in Vibrio cholerae: A Feasible Approach to Address Biological Issues. *PLoS Genetics*, *8*(1), e1002472. https://doi.org/10.1371/journal.pgen.1002472

Van Boeckel, T. P., Glennon, E. E., Chen, D., Gilbert, M., Robinson, T. P., Grenfell, B. T., Levin, S. A., Bonhoeffer, S., & Laxminarayan, R. (2017). Reducing antimicrobial use in food animals. *Science*, *357*(6358), 1350–1352. https://doi.org/10.1126/science.aao1495

Vinué, L., Jové, T., Torres, C., & Ploy, M. C. (2011). Diversity of class 1 integron gene cassette Pc promoter variants in clinical Escherichia coli strains and description of a new P2 promoter variant. *International Journal of Antimicrobial Agents*, *38*(6), 526–529. https://doi.org/10.1016/j.ijantimicag.2011.07.007

Vit, C., Richard, E., Fournes, F., Whiteway, C., Eyer, X., Lapaillerie, D., Parissi, V., Mazel, D., & Loot, C. (2021). Cassette recruitment in the chromosomal Integron of *Vibrio cholerae*. *Nucleic Acids Research*, *49*(10), 5654–5670. https://doi.org/10.1093/nar/gkab412

Vizzarro, G., Lemopoulos, A., Adams, D. W., & Blokesch, M. (2024). Vibrio cholerae pathogenicity island 2 encodes two distinct types of restriction systems. *Journal of Bacteriology*, *206*(9), e0014524. https://doi.org/10.1128/jb.00145-24

Waldor, M. K., & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science (New York, N.Y.)*, *272*(5270), 1910–1914. https://doi.org/10.1126/science.272.5270.1910

Waldor, M. K., Tschäpe, H., & Mekalanos, J. J. (1996). A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in Vibrio cholerae O139. *Journal of Bacteriology*, *178*(14), 4157–4165. https://doi.org/10.1128/jb.178.14.4157-4165.1996

White, P. A., Iver, C. J. M. C., & Rawlinson, W. D. (2001). Integrons and Gene Cassettes in the Enterobacteriaceae Downloaded from http://aac.asm.org/ on December 30 , 2015 by TERKKO NATIONAL LIBRARY OF HEALTH SCIENCES. *Antimicrobial Agents and Chemotherapy*, *45*(9), 2658–2661. https://doi.org/10.1128/AAC.45.9.2658

White, P. A., McIver, C. J., Deng, Y. M., & Rawlinson, W. D. (2000). Characterization of two new gene cassettes, aadA5 and dfrA17. *FEMS Microbiology Letters*, *182*, 265–269. https://doi.org/10.1016/S0378-1097(99)00600-X

Wozniak, R. A. F., Fouts, D. E., Spagnoletti, M., Colombo, M. M., Ceccarelli, D., Garriss, G., Déry, C., Burrus, V., & Waldor, M. K. (2009). Comparative ICE Genomics: Insights into the Evolution of the SXT/R391 Family of ICEs. *PLOS Genetics*, *5*(12), e1000786. https://doi.org/10.1371/journal.pgen.1000786

Wozniak, R. A. F., & Waldor, M. K. (2009). A Toxin–Antitoxin System Promotes the Maintenance of an Integrative Conjugative Element. *PLOS Genetics*, *5*(3), e1000439. https://doi.org/10.1371/journal.pgen.1000439

Wozniak, R. A. F., & Waldor, M. K. (2010). Integrative and conjugative elements: Mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nature Reviews Microbiology*, *8*, 552–563. https://doi.org/10.1038/nrmicro2382

Xu, H., Davies, J., & Miao, V. (2007). Molecular characterization of class 3 integrons from Delftia spp. *Journal of Bacteriology*, *189*(17), 6276–6283. https://doi.org/10.1128/JB.00348-07

Yamamoto, S., Izumiya, H., Mitobe, J., Morita, M., Arakawa, E., Ohnishi, M., & Watanabe, H. (2011). Identification of a Chitin-Induced Small RNA That Regulates Translation of the tfoX Gene, Encoding a Positive Regulator of Natural Competence in Vibrio cholerae ▿ . *Journal of Bacteriology*, *193*(8), 1953–1965. https://doi.org/10.1128/JB.01340-10

Yamamoto, S., Mitobe, J., Ishikawa, T., Wai, S. N., Ohnishi, M., Watanabe, H., & Izumiya, H. (2014). Regulation of natural competence by the orphan two-component system sensor kinase ChiS involves a non-canonical transmembrane regulator in Vibrio cholerae. *Molecular Microbiology*, *91*(2), 326–347. https://doi.org/10.1111/mmi.12462

Zhang, Q., Yan, Z., Xu, Y., Sun, J., & Shang, G. (2017). Characterization of Inducible ccdB Gene as a Counterselectable Marker in Escherichia coli Recombineering. *Current Microbiology*, *74*(8), 961–964. https://doi.org/10.1007/s00284-017-1273-3

Zhu, L., Lin, J., Ma, J., Cronan, J. E., & Wang, H. (2010). Triclosan resistance of Pseudomonas aeruginosa PAO1 is due to FabV, a triclosan-resistant enoyl-acyl carrier protein reductase. *Antimicrobial Agents and Chemotherapy*, *54*(2), 689–698. https://doi.org/10.1128/AAC.01152-09

Supplementary information

Lausanne, 2023

Supplementary Information

Strains

Supplementary Table 1. Strains used and generated in this thesis

Name	Description	Reference
A001	*V. cholerae* N16961 *Tn7::hapR+*	Meibom *et al.*, 2005
A074	*E. coli* MG1655 /pR388	Laboratory collection
A075	*E. coli* MG1655 /pR388:: *dfrB5 blaVEB aadB*	Laboratory collection
A093	*E. coli* DH5*α*	Laboratory collection
A094	*E. coli* MG1655	Laboratory collection
A116	*E. coli* β3914 donor for conjugation (F-) RP4-2-Tc::Mu Δ*dapA*::(*erm-pir*) [Km^R Em^R resistant to CcdB (B410gyrA462zei::Tn10)	Le Roux *et al.*, 2007
A118	*E. coli* π3813 resistant to CcdB (B410gyrA462zei::Tn10)	Le Roux *et al.*, 2007
A123	*E. coli* β2163 donor for conjugation (F-) RP4-2-Tc::Mu Δ*dapA*::(*erm-pir*) [Km^R Em^R] /pSW23T::$attC_{aadA7}$	Nivina *et al.*, 2006
A129	*E. coli* MG1655 /pKOBEG	Laboratory collection
A186	*E. coli* DH5α /pBAD::*gfp*	San Millan *et al.*, 2016
A199	*E. coli* DH5α /pBAD18::*ccdB::attI1*	This study
A206	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdB::attI1*	This study
A208	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdBKO::attI1*	This study
A220	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdB*	This study
A277	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdB::attI1::smr1*	This study
A278	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdB::attI1::smr2*	This study
A279	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdB::attI1::qacE*	This study
A280	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdB::attI1::aacA37*	This study

A281	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdB::attI1::dfrB2*	This study
A282	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdB::attI1::dfrB9*	This study
A283	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdB::attI1::dfrB8*	This study
A284	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdB::attI1::fosI*	This study
A285	*E. coli* DH5α /pBAD43::*ccdA* /pSU38::*ccdB::attI1::blaOXA118*	This study
A288	*E. coli* DH5α /pBAD43::*ccdA* /pSU38:: PcW *ccdB::attI1*	This study
A289	*E. coli* DH5α /pBAD43::*ccdA* /pSU38:: PcW *ccdB*	This study
A344	*E. coli* DH5α /pVanM::*gfp*	This study
A369	*E. coli* DH5α /pSU38::*gfp*	This study
A370	*E. coli* DH5α /pSU38:: P_CS *gfp*	This study
A371	*E. coli* DH5α /pSU38:: P_CW *gfp*	This study
A401	*E. coli* DH5α /pBAD43::*ccdA* /pVanM:: *intI1* P_{Lac} *ccdB::attI1*	This study
A404	*E. coli* DH5α /pBAD43::*ccdA* /pVanM:: *intI1* P_{Lac} *ccdB*	This study
A407	*E. coli* DH5α /pBAD43::*ccdA* /pVanM:: *intI1* P_{Lac} *ccdBKO::attI1*	This study
A410	*E. coli* DH5α /pBAD43::*ccdA* /pVanM:: *intI1* P_CW *ccdB::attI1*	This study
A413	*E. coli* DH5α /pBAD43::*ccdA* /pVanM:: *intI1* P_CW *ccdB*	This study
A527	*V. cholerae* N16961 *Tn7::hapR+* /pVanM::*gfp*	This study
A601	*E. coli* MG1655/pBGT	This study
A602	*V. cholerae* N16961 *Tn7::hapR+* /pBGT	This study
A629	*V .cholerae* N16961 *Tn7::hapR+* /pBAD43::*ccdA* /pSU38::*ccdB::attI1*	This study
A681	*E. coli* MG1655 /pVanM::*gfp*	This study
A782	*E. coli* MG1655 attB:: P_{BAD} *ccdA* P_{Lac} *ccdB::attI1*	This study
A785	*E. coli* Top10 /pSU38::*attI1* /pBAD18::*intI1*	This study

A786	*E. coli* Top10 /pSU38::*attI1* /pVanM::*intI1*	This study
A787	*E. coli* MG1655 *attB*:: P_{BAD} *ccdA* P_{Lac} *ccdB::attI1* /pVanM::*intI1*	This study, strain CECT30915
A811	*E. coli* MG1655 *attB*:: P_{BAD} *ccdA* P_{Lac} *ccdB*	This study
A832	*E. coli* MG1655 *attB*:: P_{BAD} *ccdA* P_{Lac} $ccdB_{KO}$*::attI1*	This study
A839	*E. coli* MG1655 /pTetM::*gfp*	This study
A880	*V. cholerae* N16961 *Tn7::hapR+* /pTetM::*gfp*	This study
A981	*E. coli* Top10 /pSU38::*attI1* /pTetM::*intI1*	This study
B522	*V. cholerae* N16961 *Tn7::hapR+* ΔSI	This study, strain CECT30916
B948	*V. cholerae* N16961 *Tn7::hapR+* /pSU38::*gfp*	This study
B949	*V. cholerae* N16961 *Tn7::hapR+* /pSU38:: PcW *gfp*	This study
B950	*V. cholerae* N16961 *Tn7::hapR+* /pSU38:: PcS *gfp*	This study
C275	*V. cholerae* N16961 *Tn7::hapR+* ΔSI */pBAD18::intI1-tfoX*	This study, derived from CECT30916
C522	*V. cholerae* A1552	Blokesch Lab collection
#135	*V. cholerae* A1552 Δ*lacZ* $kana^R$	Blokesch Lab collection
C524 (#9727)	*E. coli* S17-1 λ pir /pGP704::*sacB* Δ*ddmABC*	Blokesch Lab collection
C525 (#9728)	*E. coli* S17-1 λ pir /pGP704::*sacB* Δ*ddmDE*	Blokesch Lab collection
C526 (#923)	*E. coli* S17-1 λ pir /pGP704::*sacB* Δ*dns*	Blokesch Lab collection
#1036	*E. coli* S17-1 λ pir /pGP704::*sacB* Δ*luxO*	Blokesch Lab collection
#1049	*E. coli* S17-1 λ pir /pGP704::*sacB* Δ*dprA*	Blokesch Lab collection
C531	*V. cholerae* N16961 *Tn7::hapR+* ΔSI Δ*dprA*	This study, derived from CECT30916
C532	*V. cholerae* N16961 *Tn7::hapR+* ΔSI Δ*dns*	This study, derived from CECT30916
C533	*V. cholerae* N16961 *Tn7::hapR+* ΔSI Δ*luxO*	This study, derived from CECT30916
C535	*V. cholerae* N16961 *Tn7::hapR+* ΔSI Δ*dns* Δ*luxO*	This study, derived from CECT30916

C536	*V. cholerae* N16961 *Tn7::hapR*+ ΔSI Δ*ddmABC* Δ*ddmDE*	This study, derived from CECT30916
C539	*V. cholerae* N16961 *Tn7::hapR*+ ΔSI Δ*dns* /pBAD18*::intI1-tfoX*	This study, derived from CECT30916
C542	*V. cholerae* N16961 *Tn7::hapR*+ ΔSI Δ*luxO* /pBAD18::*intI1-tfoX*	This study, derived from CECT30916
C545	*V. cholerae* N16961 *Tn7::hapR*+ ΔSI Δ*dprA* /pBAD18::*intI1-tfoX*	This study, derived from CECT30916
C549	*V. cholerae* N16961 *Tn7::hapR*+ ΔSI Δ*ddmABC* Δ*ddmDE /pBAD18::intI1-tfoX*	This study, derived from CECT30916
C650	*V. cholerae* N16961 *Tn7:: sacB hapR*+ ΔSI Δ*ddm*	This study, derived from CECT30916
C869	*V. cholerae* N16961 *Tn7:: sacB ::attI1*(version 1) *hapR*+ ΔSI Δ*ddm*	This study, derived from CECT30916
C870	*V. cholerae* N16961 *Tn7:: sacB ::attI1*(version 2) *hapR*+ ΔSI Δ*ddm*	This study, derived from CECT30916
C872	*V. cholerae* N16961 *Tn7:: sacB ::attI1*(version 3) *hapR*+ ΔSI Δ*ddm*	This study, derived from CECT30916
C874	*V. cholerae* N16961 *Tn7:: sacB ::attI1*(version 1) *hapR*+ ΔSI Δ*ddm* /pBAD18::*intI1-tfoX*	This study, strain CETC31061
C918	*V. cholerae* N16961 *Tn7::hapR*+ ΔSI Δ*ddmABC* Δ*ddmDE* Δ*dns*	This study, derived from CECT30916
D050	*V. cholerae* N16961 *Tn7:: sacB ::attI1*(version 1) *hapR*+ ΔSI Δ*dns*	This study, derived from CECT30916
D052	*V. cholerae* N16961 *Tn7:: sacB ::attI1*(version 1) *hapR*+ ΔSI Δ*ddm* Δ*dns*	This study, derived from CECT30916
D056	*E. coli* MG1655 *attB*:: *sacB ::attI1*(version 1)	This study
D106	*E. coli* MG1655 *attB*:: *sacB ::attI1*(version 1) /pBAD18::*intI1*	This study

Plasmids

Supplementary Table 2. Plasmids used and generated in this thesis

Name	Description	Characteristics	Reference
pA120	pBAD18::*intI1*	oriV$_{ColE1}$; [Carb^R]	Demarre *et al.*, 2007
pA121	pSU38	oriV$_{p15A}$; [Km^R]	Biskri *et al.*, 2005
pA123	pSW23T::*attCaadA7*	oriV$_{R6K\gamma}$, oriT$_{RP4}$; [Cm^R]	Nivina *et al.*, 2016
pA128	pKOBEG Carb^R	oriV$_{pSC101}$; [Carb^R]	Derbise *et al.*, 2003
pA129	pKOBEG Cm^R	oriV$_{pSC101}$; [Cm^R]	Derbise *et al.*, 2003
pA186	pBAD::*gfp*	oriV$_{p15A}$; oriT$_{RP4}$ [Carb^R] [Cm^R]	San Millan *et al.*, 2016
pA198	pBAD18::*ccdB*::*attI1*	oriV$_{ColE1}$; [Carb^R]	This study
pA202	pBAD18::*ccdA*	oriV$_{ColE1}$; [Carb^R]	Laboratory collection
pA204	pSU38::*ccdB*::*attI1*	oriV$_{p15A}$; [Km^R]	This study
pA205	pSU38::*ccdB$_{KO}$*::*attI1*	oriV$_{p15A}$; [Km^R]	This study
pA218	pSU38::*ccdB*	oriV$_{p15A}$; [Km^R]	This study
pA286	pSU38::P$_C$W *ccdB*	oriV$_{p15A}$; [Km^R]	This study
pA329	pSU38::*gfp*	oriV$_{p15A}$; [Km^R]	This study
pA344	pVanM::*gfp*	oriV$_{p15A}$; [Km^R]	This study
pA346	pVanM::*intI1*	oriV$_{p15A}$; [Zeo^R]	This study
pA348	pVanM::*intI1* + P$_{Lac}$ *ccdB*::*attI*	oriV$_{p15A}$; [Zeo^R]	This study
pA370	pSU38::P$_C$S *gfp*	oriV$_{p15A}$; [Km^R]	Blanco & Trigo da Roza et al., 2024
pA371	pSU38::P$_C$W *gfp*	oriV$_{p15A}$; [Km^R]	This study
pA404	pVanM::*intI1* + P$_{Lac}$ *ccdB*	oriV$_{p15A}$; [Zeo^R]	This study
pA407	pVanM::*intI1* + P$_{Lac}$ *ccdBKO*::*attI*	oriV$_{p15A}$; [Zeo^R]	This study
pA410	pVanM::*intI1* + P$_C$W *ccdB*::*attI*	oriV$_{p15A}$; [Zeo^R]	This study
pA413	pVanM::*intI1* + P$_C$W *ccdB*	oriV$_{p15A}$; [Zeo^R]	This study
pA498	pVanM::*intI1*	oriV$_{p15A}$; [Km^R]	This study
pA839	pTetM::*gfp*	oriV$_{p15A}$; [Km^R]	This study
pA877	pTetM::*intI1*	oriV$_{p15A}$; [Km^R]	This study
pA896	pTetM::*intI1* Zeo^R	oriV$_{p15A}$; [Zeo^R]	This study
pB629	pTetM:: ccdB::attI1::NPU	oriV$_{p15A}$; [Km^R]	This study
pB746	pBAD::ccdA + P$_{VanM}$ intI1	oriV$_{ColE1}$; [Carb^R]	This study
pB824	pSU38::*ccdB*::*attI1*	oriV$_{p15A}$; [Zeo^R]	This study

pB907	pBAD18::*intI1* Spec^R	oriV_{ColE1}; [Spec^R]	This study
pC137	pMBA::*aacA54 catB3*	oriV_{p15A}; [Zeo^R]	Carvalho *et al.*, 2024
pC155	pMBA::*aacA8 catB3*	oriV_{p15A}; [Zeo^R]	Carvalho *et al.*, 2024
pC167	pMBA::*aacA61 catB3*	oriV_{p15A}; [Zeo^R]	Carvalho *et al.*, 2024
pC274	pBAD18::*intI1-tfoX*	oriV_{ColE1}; [Spec^R]	This study

Primers

Supplementary Table 3. Primers used in this thesis

Name	Sequence (5′ → 3′)	Description
MRV	AGCGGATAACAATTTCACACAGGA	pSU38 insert amplification
MRV II	GGTTTCCCGACTGGAAAGCG	
MFD	GCCAGGGTTTTCCCAGTCAC	
bb p929 F	CCGTCGTTGCCTGATGGATC	Amplification of pSU38 backbone
bb p929 R	GGCTTGCTCCTTTGTTAGAGC	
p929 GFP F	GCTCTAACAAAGGAGCAAGCCATG AGTAAAGGAGAAGAAC	*gfp* amplification to clone in pSU38
p929 GFP R	GATCCATCAGGCAACGACGGCGCT GTCTAGACTATTTGTA	
Pc F	GCAGTGAGCGCAACGCAATTACCG TGGAAACGGATGAAGG	P_CW/P_CS amplification to clone in pSU38
Pc R	CTCGAATTCGTAATCATGGTTCTGG ACCAGTTGCGTGAGC	
Pc INV F	GCTCACGCAACTGGTCCAGAACCA TGATTACGAATTCGAG	Amplification of pSU38 backbone Δprom
Pc INV R	CCTTCATCCGTTTCCACGGTAATTG CGTTGCGCTCACTGC	
Marionette F II	TCCCGCTTAACGATCGTTGG	Marionette insert amplification
Marionette R	GGTTTTGCACCATTCGATGG	
bb Marionette F	CAAATTCCAGAAAAGAGGCC	Amplification of Marionette plasmid backbone
bb Marionette R	CTAGTATTTCCCCTCTTTCTC	
Marionette GFP F	GAGAAAGAGGGGAAATACTAGAT GAGTAAAGGAGAAGAAC	*gfp* amplification to clone in Marionette
Marionette GFP R	GGCCTCTTTTCTGGAATTTGCGCTG TCTAGACTATTTGTA	
IntI1 Marionette F	GAGAAAGAGGGGAAATACTAGAT GAAAACCGCCACTGCGCC	*intI1* amplification to clone in Marionette
IntI1 Marionette R	GGCCTCTTTTCTGGAATTTGGTACC GAGCTACCTCTCACT	
SWbeg	CCGTCACAGGTATTTATTCGGCG	pSW23T insertion verification
pBAD F	AGATTAGCGGATCCTACCTG	pBAD18/43 insert amplification
pBAD R	TTAATCTGTATCAGGCTGAA	
pBAD EcoRI F	TTTTGGGCTAGCGAATTCGAGC	

pBAD BamHI XbaI R	TGCAGGTCGACTCTAGAGGAT	*ccdB*$_{KO}$*::attI* amplification from IDT synthetized product
ccdB in frame F	GATAAGAAACGATGTTATGGAGCAGCAACG	SNP introduction in *ccdB*$_{KO}$*::attI1* that renders *ccdB::attI1* in frame
ccdB in frame R	TCCATAACATCGTTTCTTATCAAGTAGTTCG	
ccdBΔattI F	TGATAAGAAAGCACCAAGTCACCTTTGTCC	*attI1* deletion to recover the WT *ccdB* gene
ccdBΔattI R	GACTTGGTGCTTTCTTATCAAGTAGTTCGA	
bb Marionette res F	AACACCCCTTGTATTACTG	Amplification of Marionette plasmid backbone without resistance marker
bb Marionette res R	GGTAACTGTCAGACCAAG	
Zeo Marionette F	AACTTGGTCTGACAGTTACCTCTTAATCCTGTTCTTCGGC	Zeo^{R} marker amplification to clone in Marionette
Zeo Marionette R	CAGTAATACAAGGGGTGTTAAATGGCCAAACTGACCAGC	
MarLac F	ACGAATCAGACAATTGACGG	Marionette second insert verification
MarLac R	AATATTTGCTCATGAGCCCG	
bb Mar ccdb F	GGCATCACCGGCGCCACAG	Marionette backbone amplification to introduce a second insert
bb Mar ccdB R	GGCCACGATGCGTCTGGCGT	
MRV II pVan F	ACGCCAGACGCATCGTGGCCGGTTTCCCGACTGGAAAGCG	pSU38 prom + insert amplification to clone in Marionette plasmid as second insert
MFD pVan R	CTGTGGCGCCGGTGATGCCCGCCAGGGTTTTCCCAGTCAC	
IntI R bb	CTTTGTTTTAGGGCGACTGC	pSU38 backbone amplification to insert a cassette inside *ccdB::attI*
p929 ccdB::attI F	GGGCACCAAGTCACCTTTGTCC	
gblock F	GCAGTCGCCCTAAAACAAAG	Amplification of integron cassette from IDT synthetized DNA
smr1 R	ACAAAGGTGACTTGGTGCCCGCCTAACGCCTGAATTAAGC	
smr2 R	ACAAAGGTGACTTGGTGCCCGCCTAACGCCCGAATTAAGC	
qacE R	ACAAAGGTGACTTGGTGCCCGCCTAACGCCGAGTTCAGCG	

aacA37 R	ACAAAGGTGACTTGGTGCCCGCCT AACGCGTGAATTAAGC	
dfrB2 R	ACAAAGGTGACTTGGTGCCCGCCT AACGATGAAGCTCAGC	
dfrB8 R	ACAAAGGTGACTTGGTGCCCGCCT AACGAATTAGCTCAGC	
dfrB9 R	ACAAAGGTGACTTGGTGCCCGCCT AACGTTCGAGGTGAGG	
fosI R	ACAAAGGTGACTTGGTGCCCGCCT AACGCTTGAGTTAAGC	
blaOXA118 R	ACAAAGGTGACTTGGTGCCCGCCT AACGTTGAAATCAGCG	
ybhC ext F	TTTGTGACCAGAAGACCGCA	*attB* site cloning verification
ybhB ext R	CTCATCAGTAACGATCTGCG	
ybhC F	CCTGTACCGTACAGAGTAAT	*attB* site RHR amplification
attB Universal R	GTATAAAAAAGCAGGCTTCA	
attB Universal F	TTATACTAACTTGAGCGAAACGGG	*attB* site LHR amplification
ybhB R	TGGCGATAATATTTCACCGC	
PBAD ccdB chrom F	AAATAGAAAAATGAATCCGTTGAA GCCTGCTTTTTTATACTCATCGATGC ATAATGTGCC	pBAD18/43 insert + resistance marker amplification for *attB* cloning
PBAD ccdB chrom R	TTTGTCTTTTTACCTTCCCGTTTCGC TCAAGTTAGTATAACACTACGTGAA CCATCACCC	
ccdB zeo ccdA chrom F	CAAAAAAAGGCCCCCCGTTAGGGA GGCCTTCAATAATTGGGGTTTCCCG ACTGGAAAGCG	pSU38 insert +resistance amplification for *attB* cloning to substitute carb^{R} of pBAD18 insertion in *attB*
ccdB zeo chrom R	TTTGTCTTTTTACCTTCCCGTTTCGC TCAAGTTAGTATAAGCGGATTTGTC CTACTCAGG	
RHE Tn7 ext F II	GATGCACCAAAAAACATGGC	Tn7 site cloning verification
RHD Tn7 ext R II	TGCCCGTCGTATTAAAGAGG	
RHE Tn7 F II	AATGGCCGCATTCAAATCCG	Tn7 site RHR amplification
RHE Tn7 R II	TAGCTTACGACGCTACACCC	
RHD Tn7 F	GCTTGCTCAATCAATCACCG	Tn7 site LHR amplification
RHD Tn7 R II	CTCAGTATCGCTGACTTTGG	

sacB Tn7 F	GGGTGTAGCGTCGTAAGCTAGAATTCGGCTTAACCCATGC	*sacB* amplification to clone in Tn7 with zeo^R marker
sacB zeo R	CCTCTTACGTGCCGACAATTATTCGTTCAAGCCGAGATCG	
zeo sacB F	CGATCTCGGCTTGAACGAATAATTGTCGGCACGTAAGAGG	Zeo^R marker amplification to clone with *sacB* in Tn7
zeo Tn7 R	CGGTGATTGATTGAGCAAGCCGGATTTGTCCTACTCAGG	
sacB attI1 F I	GATAAAAAACGCACGGCTGA	*sacB::attI1*$_{300}$ construct
sacB attI1 R I	GGCTTTGCAGAAGTTTTTGAC	
attI1 sacB F I	GTCAAAAACTTCTGCAAAGCCGATGTTATGGAGCAGCAAC	
attI1 sacB R I	TCAGCCGTGCGTTTTTTATCCCAACTTTGTTTTAGGGCGAC	
sacB attI1 F II	GAGCTCTTTGAACATCAACGG	*sacB::attI1*$_{203}$ construct
sacB attI1 R II	GGTCTGATGCTGATACGTTAAC	
attI1 sacB F II	TTAACGTATCAGCATCAGACCGATGTTATGGAGCAGCAAC	
attI1 sacB R II	CCGTTGATGTTCAAAGAGCTCCAACTTTGTTTTAGGGCGAC	
sacB attI1 F III	GCCTAACGATGTAACCTTTAC	*sacB::attI* $_{403}$ construct
sacB attI1 R III	GATCAAGATCCATTTTTAACACAAGG	
attI1 sacB F III	TGTTAAAAATGGATCTTGATCGATGTTATGGAGCAGCAAC	
attI1 sacB R III	TAAAGGTTACATCGTTAGGCCAACTTTGTTTTAGGGCGAC	
SacB attB F	TGAAGCCTGCTTTTTTATACGAATTCGGCTTAACCCATGC	*sacB* amplification to clone in *attB* with zeo^R marker
zeo attB R II	TTTCGCTCAAGTTAGTATAACGGATTTGTCCTACTCAGGA	
sacB::attI check F	ACACTGGAACTGAAGATGGC	Cassette insertion verification
sacB::attI check R	GTGAACAGGTACCATTTGCC	
bb pSU38 res F II	CACCGTTTCTGCGGACTGGC	Amplification of pSU38 plasmid backbone without resistance marker
bb pSU38 res R II	CGGAATCGTTTTCCGGGACG	

Zeo pSU38 F II	CGTCCCGGAAAACGATTCCGTCTTAATCCTGTTCTTCGGC	Zeo^R marker amplification to clone in pSU38
Zeo pSU38 R II	GCCAGTCCGCAGAAACGGTGAATTGTCGGCACGTAAGAGG	
PVan IntI1 ccdA F	CGCTCTCCTGAGTAGGACAACCAATTATTGAAGGCCTCCC	Marionette prom + insert amplification to clone in pBAD18 as second insert
PVan IntI1 ccdA R	CGCAACGTTCAAATCCGCTCAGTCAGTGAGCGAGGAAGC	
pBAD ins F	GAGCGGATTTGAACGTTGCG	pBAD18 amplification to introduce a second insert
XylS Pm F	TTGTCCTACTCAGGAGAGCG	
KanR-500flank-up	GCGCTTTATCAACACGCTGAATTGC	Kana^R amplification + 500 bp of homology to Δ*lacZ*
KanR-500flank-down	ACGCGAAGATGGTCACATTCCACAC	
LacZ-FRT1	ATGCGCAACTTCTCCGATATTCTTCTTAGCC	Kana^R amplification + 1 kb of homology to Δ*lacZ*
LacZ-FRT4	GAGATACCACTTATCGCCCGCCACCAACTCG	
Tfm-II-gDNA-1000	AAGCTTCCTGCTTGGAAGAAATGGC	Kana^R amplification + 2 kb of homology to Δ*lacZ*
Tfm-II-gDNA+1000	CGGTGTATCTGTGGCAACGGTTTC	
ddmA check up	TTGATCGCAATACCCACGGTTTAGG	Δ*ddmABC* confirmation PCR
ddmC check dw	TTGCGCGGAATATCCTTGAACAGTC	
VC1771 check F	CGCTGGCGTTAAGTTGATGTAC	Δ*ddmDE* confirmation PCR
VC1770 check R	TAATGCGCTCAGATTGCTCACC	
VC0048 check start	ATGAAAGATCAGGATTTAGCGGCATGG	Δ*dns* confirmation PCR
VC0048 check end	ATACATCAAAATATCCATCATCATAGC	
ΔluxO check F	ATTGGAATGGTGTGAACGCG	Δ*luxO* confirmation PCR
ΔluxO check R	AAGTAGCTGAAGAATCGGCG	
VC0470 check start	ATGAAAGATCAGGATTTAGCGGCATGG	Δ*dprA* confirmation PCR
VC0470 check end	ATACATCAAAATATCCATCATCATAGC	
bb res p3938 F	GGTAACTGTCAGACCAAG	Amplification of pBAD18 plasmid backbone without resistance marker
bb res p3938 R	ACTCTTCCTTTTTCAATATTATTGAAGC	

aadA1 p3938 F	AATATTGAAAAAGGAAGAGTATGAGGGAAGCGGTGATCGC	$Spec^R$ marker amplification to clone in pBAD18
specR pBAD R	CTTGGTCTGACAGTTACCTTATTTGCCGACTACCTTGG	
bb pBAD operon F	CGGATTCACTACCCCTGCG	Amplification of pBAD18 plasmid backbone to introduce a second gene as operon
bb pBAD operon R	CCAGGATTGACTTGCGCTGC	
TfoX pBAD F	GCAGCGCAAGTCAATCCTGGAGCGGATAACAATTTCACACAGGA	*tfoX* amplification to clone in pBAD18
TfoX pBAD R	GCGCAGGGGTAGTGAATCCGATCCGCCAAAACAGCCAAGC	
IntI seq F II	TGGAAACGGATGAAGGCACG	pMBA integron array amplification
GFP seq R II	AACTGCGCTGGTCAGTTTGG	
sul1 R	TCTCATCGAAGAAGGAGTCC	R388 integron array amplification
GTT R*	RYYYA*A*C	Phi29 integron cassette enrichment

Reagents

Supplementary Table 4. Antibiotics and reagents used in this thesis

Reagent	Medium final concentration (*E. coli* / *V. cholerae*)	Reference
Carbenicillin	100 µg/mL	BioChemica, A1491
Ampicillin	100 µg/mL	Sigma-Aldrich, A9518
Kanamycin	25 µg/mL / 75 µg/mL	Sigma-Aldrich, K4000
Spectinomycin	50 µg/mL / 200 µg/mL	Sigma-Aldrich, S4014
Zeocin	50 µg/mL	InvivoGen, Ant-Zn
Trimethoprim	50 µg/mL	Sigma-Aldrich, T7883
Chloramphenicol	25 µg/mL / 2.5 µg/mL	Sigma-Aldrich, C0378
DAP	0.3 mM	Sigma-Aldrich, 33240
dT	0.3 mM	Sigma-Aldrich, T9250
Arabinose	0.20%	Sigma-Aldrich, W325501
Glucose	1%	Sigma-Aldrich, 158968
Vanillate	12.5 µM	Sigma-Aldrich, 94770
IPTG	200 µg/mL	Sigma-Aldrich, I6758
aTc	200 ng/mL	Sigma-Aldrich, 37919
Chitin	8 µg/mL	Apollo Scientific, OR2403

SeqDelTA

Supplementary Table 5. Strains generated with SeqDelTA

Name	Description	Primers LHR	Primers RHR	Resistance marker
A400	ΔVCA300-311	VCA 299 F – LRHI VCA 300 R	LRHD VCA 311 F – VCA 312 R	[Zeo^R]
A003	ΔVCA300-320	VCA 299 F – LRHI VCA 300 R	LRHD VCA 318 F – VCA 320 R	[Cm^R]
A004	ΔVCA300-325	VCA 299 F – LRHI VCA 300 R	LRHD VCA 323 F – VCA 325 R	[Carb^R]
A005	ΔVCA300-332	VCA 299 F – LRHI VCA 300 R	LRHD + prom VCA 332 F – prom VCA 332 R VCA 333 F – VCA 334 R	[Zeo^R]
A006	ΔVCA300-350	VCA 299 F – LRHI VCA 300 R	LRHD VCA 348 F – VCA 350 R	[Carb^R]
A007	ΔVCA300-360	VCA 299 F – LRHI VCA 300 R	LRHD VCA 359 F – VCA 360 R	[Cm^R]
A009	ΔVCA300-385	VCA 299 F – LRHI VCA 300 R	LRHD VCA 311 F – VCA 386 R	[Zeo^R]
A023	ΔVCA300-391	VCA 299 F – LRHI VCA 300 R	LRHD VCA 391 F – VCA 393 R	[Carb^R]
A024	ΔVCA300-391 + VCA422-444	VCA 421 F – LRHI VCA 422 R	LRHD VCA 443 F – VCA 447 R	[Carb^R, Zeo^R]
A029	ΔVCA300-470	VCA 299 F – LRHI VCA 300 R	LRHD + prom VCA 469 F – prom VCA 469 R VCA 469 F – VCA 470 R	[Cm^R]
A035	ΔVCA300-470 + VCA474-483	VCA 474 F – link AT/TA R / link AT/TA F – LRHI VCA 478 R	LRHD VCA 318 F – VCA 483 R	[Cm^R, Zeo^R]
A041	ΔVCA300-483	VCA 299 F – LRHI VCA 300 R	LRHD VCA 318 F – VCA 483 R	[Carb^R]
A047	ΔVCA300-483 + ΔVCA495-497	VCA 494 F – LRHI VCA 495 R	LRHD + prom VCA 497 F – prom VCA 497 R VCA 498 F – VCA 499-500 R	[Carb^R, Cm^R]
A054	ΔVCA300-483 + ΔVCA487-505	VCA 486 F – LRHI VCA 487 R	LRHD VCA 348 F – VCA 505 R	[Carb^R, Zeo^R]
A066	ΔVCA300-VCA505	VCA 299 F – LRHI VCA 300 R	LRHD VCA 348 F – VCA 505 R	[Zeo^R]

Supplementary Table 6. Strains generated with pMP7

Name	Description	pMP7
A101	*V. cholerae* N16961 ΔSI	pMP7_Δ *intIA attIA zeo*^{R}
A677	*V. cholerae* N16961 ΔSI *rocS*$^{+}$	pMP7_*rocS*$^{+}$
A684	*V. cholerae* N16961 ΔSI *rocS*$^{+}$ *rpoS*$^{+}$	pMP7_*rpoS*$^{+}$
B522	*V. cholerae* N16961 ΔSI *rocS*$^{+}$ *rpoS*$^{+}$ *cry2*$^{+}$	pMP7_*cry2*$^{+}$

Supplementary Table 7. Primers used for SeqDelTA

Name	Sequence (5′ → 3′)
3083	TCTAGGGCGGCGGATTTGTC
3327	TGCAATTGTCGGCACGTAAG
VCA 299 F	GGGCGTTAGAGCTTTATTGG
LRHI VCA 300 R	GACAAATCCGCCGCCCTAGAGAGCTTTATTTACTCGGACG
LRHD VCA 311 F	CTTACGTGCCGACAATTGCAGGTAAACGCTCAATCAAAGG
VCA 312 R	CTTTATTACGACAGCCATCGC
LRHD VCA 318 F	CTTACGTGCCGACAATTGCATGAAACGGTTCCTATCGTGC
VCA 320 R	CTGCCAAATCAGCATCAAGC
LRHD VCA 323 F	CTTACGTGCCGACAATTGCACAATGATGTCTGAAATCGGC
VCA 325 R	ACCTATGAGCTGACCAATGC
LRHD + prom VCA 332 F	CTTACGTGCCGACAATTGCAAAATATCACCTAACAAGCGC
prom VCA 332 R	CTGAGATGATCCTGACAATACCGTGCTTTTCGAG
VCA 333 F	AAAAGCACGGTATTGTCAGGATCATCTCAGTTCGG
VCA 334 R	TCATTGACCTCATCAAGACC
LRHD VCA 348 F	CTTACGTGCCGACAATTGCAATTGAGAGATGCTTCTTCCC
VCA 350 R	GAAACCTTCAGATAGCCGTC
LRHD VCA 359 F	CTTACGTGCCGACAATTGCATCACGTAAATCGGCTTTGGC
VCA 360 R	TGACCCTGCTCATTTCTTGC
LRHD VCA 311 F	CTTACGTGCCGACAATTGCAGGTAAACGCTCAATCAAAGG
VCA 386 R	ATACGATGGACGCCATAAGC
LRHD VCA 391 F	CTTACGTGCCGACAATTGCAGACTTAAGAATTCCACCGGG
VCA 393 R	TCAAGGTTAAGTTGGCCACG
VCA 421 F	GGTGATGCGACTCAAAAAGC
LRHI VCA 422 R	GACAAATCCGCCGCCCTAGATTGTAACGCTAGAGGTGACC
LRHD VCA 443 F	CTTACGTGCCGACAATTGCATTGAACTGCTGTTGGAGTGG
VCA 447 R	ATCGAGGGAAACGCATAACC

LRHD + prom VCA 469 F	CTTACGTGCCGACAATTGCAGCTCTTATCTGAACTAATTCTTGCC
prom VCA 469 R	GCCGTAAATGGTTAGCAAATCACTGAGATATTCATCTCGG
VCA 469 F	CCGAGATGAATATCTCAGTGATTTGCTAACCATTTACGGC
VCA 470 R	TGGTATAGAAGTCCTGTGCC
VCA 474 F	GGGCGTTATAACCTAATGGG
link AT/TA R	GGTAGTTGTTTTTTTGCGACAGCGGCCTTGGGAATATCAACGGCTCC
link AT/TA F	GGAGCCGTTGATATTCCCAAGGCCGCTGTCGCAAAAAAACAACTACC
LRHI VCA 478 R	GACAAATCCGCCGCCCTAGAACTCCATTCCTTTTGCGCAG
LRHD VCA 318 F	CTTACGTGCCGACAATTGCATGAAACGGTTCCTATCGTGC
VCA 483 R	TTCCTGCCAATACTTGCTACC
VCA 494 F	TCGATGCTGCTTAACGGTGC
LRHI VCA 495 R	GACAAATCCGCCGCCCTAGAATCATGGCCATCAACTCTCC
LRHD + prom VCA 497 F	CTTACGTGCCGACAATTGCACGGGCGTTATATGCTTATAGG
prom VCA 497 R	GGATTTTCACCATTCACCGCATAGGGCATCAATCTCTGGC
VCA 498 F	GCCAGAGATTGATGCCCTATGCGGTGAATGGTGAAAATCC
VCA 499-500 R	TTGTGAGACTAAGCTGACCG
VCA 486 F	CCTATAATCGCTCAACTGACGG
LRHI VCA 487 R	GACAAATCCGCCGCCCTAGAACGGATTAGACTGACCTTTCC
VCA 505 R	CCTACACCGACAATGAAACC
RHI link pMP7 F	TCTGCGAGGCTGGCCGGCGTCCGTCAGCTGCGTCCAAACG
LRHI+D R	GCGACACTTACTCAAGCTGTCATGGGTTCTTTGCGAAATC
RHD F	ACAGCTTGAGTAAGTGTCGC
RHD link pMP7 R	TCAAGCTTATCGATACCGTCATTTGCTTTATGACTCGCGC
***rocS*_pMP7 F**	CGATAAGCTTGATATCGAATTCCAGAAAAACCCTTGGTGTGC
***rocS*_pMP7 R**	CAGACAATTGACGGCTCTAGAATCAATCAGTTGAGGATTGC
***rpoS*_pMP7 F**	CGATAAGCTTGATATCGAATTCCGACGCGCAAGCACTTCTTT
***rpoS*_pMP7 R**	CAGACAATTGACGGCTCTAGAGTCAGCAATACCGTAACCA
***cry2*_pMP7 F**	GATAAGCTTGATATCGAATTCGTGAATCATAAGCGCGGTT
***cry2*_pMP7 R**	CAGACAATTGACGGCTCTAGAAACTCGGTAATAGGCAGAA
pMP7_bb_Gibson F	TCTAGAGCCGTCAATTGTCTG
pMP7_bb_Gibson R	GAATTCGATATCAAGCTTATCG

Supplementary Table 8. PCR and sanger sequencing of the SI captured cassettes in the cassette harvester and the integron cassette capture chassis

Cassette number	VC IDs	Feature	PCR count library	PCR count I3C
Cassette 1	VCA0292		0	0
Cassette 2	VCA0293		3	1
Cassette 3	NA		0	0
Cassette 4	VCA0294		1	0
Cassette 5	NA		2	0
Cassette 6	VCA0297		1	0
Cassette 7	VCA0298		0	0
Cassette 8	VCA0299		0	1
Cassette 9	VCA0300	catB9	1	0
Cassette 10	VCA0301		1	0
Cassette 11	VCA0302		1	1
Cassette 12	VCA0303		0	1
Cassette 13	VCA0304		0	0
Cassette 14	VCA0306		1	0
Cassette 15	VCA0307		0	0
Cassette 16	VCA0308	PD system	0	0
Cassette 17	VCA0309		0	0
Cassette 18	VCA0310		2	0
Cassette 19	VCA0311, VCA0312	TA	0	0
Cassette 20	VCA0313		1	0
Cassette 21	VCA0314, VCA0315		2	0
Cassette 22	VCA0316		1	0
Cassette 23	VCA0317		1	1
Cassette 24	VCA318, VCA319	TA	0	0
Cassette 25	NA		0	0
Cassette 26	VCA0322	PD system	0	1
Cassette 27	VCA0323, VCA0324	TA	0	0
Cassette 28	VCA0325		0	2
Cassette 29	NA		0	1
Cassette 30	VCA0328		0	0
Cassette 31	VCA0329		0	0
Cassette 32	VCA0329a		0	0

Cassette 33	VCA0329b		2	5
Cassette 34	NA		1	0
Cassette 35	VCA0330		0	1
Cassette 36	VCA0331		0	0
Cassette 37	VCA0332, VCA0333	TA	0	0
Cassette 38	VCA0334		0	0
Cassette 39	VCA0336		0	2
Cassette 40	VCA0337		0	0
Cassette 41	VCA0338		0	0
Cassette 42	VCA0339		1	0
Cassette 43	VCA0340		1	0
Cassette 44	VCA0341		1	0
Cassette 45	VCA0342		1	0
Cassette 46	VCA0343		0	0
Cassette 47	VCA0344		0	0
Cassette 48	VCA0345		0	0
Cassette 49	VCA0346		1	0
Cassette 50	VCA0347		0	0
Cassette 51	VCA0348, VCA0349	TA	0	0
Cassette 52	VCA0350		0	0
Cassette 53	VCA0351		0	0
Cassette 54	VCA0353		0	0
Cassette 55	VCA0354		0	1
Cassette 56	VCA0355		1	0
Cassette 57	VCA0356	PD system	1	0
Cassette 58	VCA0356a		1	0
Cassette 59	VCA0357		0	0
Cassette 60	VCA0358		0	1
Cassette 61	VCA0359, VCA0360	TA	0	0
Cassette 62	VCA0361		0	0
Cassette 63	VCA361a, VCA0362		3	1
Cassette 64	VCA0363		0	0
Cassette 65	VCA0364		0	0
Cassette 66	VCA0365		0	1
Cassette 67	VCA0366	PD system	0	0

Cassette 68	VCA0367, VCA368, VCA0369, VCA0370, VCA0371, VA0372	PD system/IS	0	0
Cassette 69	VCA0374	PD system	0	0
Cassette 70	VCA0375		3	0
Cassette 71	VCA0376		0	0
Cassette 72	VCA0378		1	0
Cassette 73	VCA0379		0	0
Cassette 74	VCA0380		0	0
Cassette 75	VCA0381		1	0
Cassette 76	VCA0382		0	0
Cassette 77	VCA0382a		3	0
Cassette 78	NA		1	1
Cassette 79	NA		2	0
Cassette 80	VCA0385, VCA0386	TA	0	0
Cassette 81	VCA0387		3	0
Cassette 82	VCA0388		0	0
Cassette 83	VCA0389		2	0
Cassette 84	NA		0	0
Cassette 85	NA		0	0
Cassette 86	VCA0391, VCA0392	TA	1	0
Cassette 87	NA		0	0
Cassette 88	VCA0395		1	0
Cassette 89	VCA0396	PD system	0	0
Cassette 90	VCA0397		0	0
Cassette 91	VCA0398		0	0
Cassette 92	VCA0399	PD system	0	0
Cassette 93	VCA0400		1	0
Cassette 94	VCA0401		0	0
Cassette 95	VCA0402		0	0
Cassette 96	VCA0403		0	0
Cassette 97	NA		0	0
Cassette 98	VCA0405		1	0
Cassette 99	VCA0406		0	0
Cassette 100	VCA0407		1	2
Cassette 101	VCA0408		1	0

Cassette 102	VCA0409	PD system	2	0
Cassette 103	VCA0410	PD system	0	0
Cassette 104	VCA0412		0	0
Cassette 105	VCA0413		1	0
Cassette 106	VCA0414		0	0
Cassette 107	VCA0415		0	0
Cassette 108	VCA0416		0	0
Cassette 109	VCA0417		0	0
Cassette 110	NA		0	0
Cassette 111	VCA0418		1	0
Cassette 112	VCA0419	PD system	0	0
Cassette 113	VCA0420		1	0
Cassette 114	VCA0421		0	0
Cassette 115	VCA0422, VCA0423	TA	0	0
Cassette 116	VCA0424		0	0
Cassette 117	VCA0425		1	0
Cassette 118	VCA0426, VCA0427		1	0
Cassette 119	VCA0428		1	0
Cassette 120	NA		0	0
Cassette 121	VCA0431		0	0
Cassette 122	VCA0432		0	0
Cassette 123	VCA0433		0	0
Cassette 124	VCA0435		1	0
Cassette 125	VCA0436		0	0
Cassette 126	NA		0	1
Cassette 127	VCA0439		0	0
Cassette 128	VCA0440		0	0
Cassette 129	VCA0441	PD system	0	0
Cassette 130	VCA0442		1	0
Cassette 131	VCA0443		0	0
Cassette 132	VCA0444, VCA0445	TA	0	0
Cassette 133	VCA0446, VCA0447	PD system	0	0
Cassette 134	VCA0448		1	0
Cassette 135	VCA0449		0	0
Cassette 136	VCA0450	PD system	0	0
Cassette 137	VCA0451		0	0

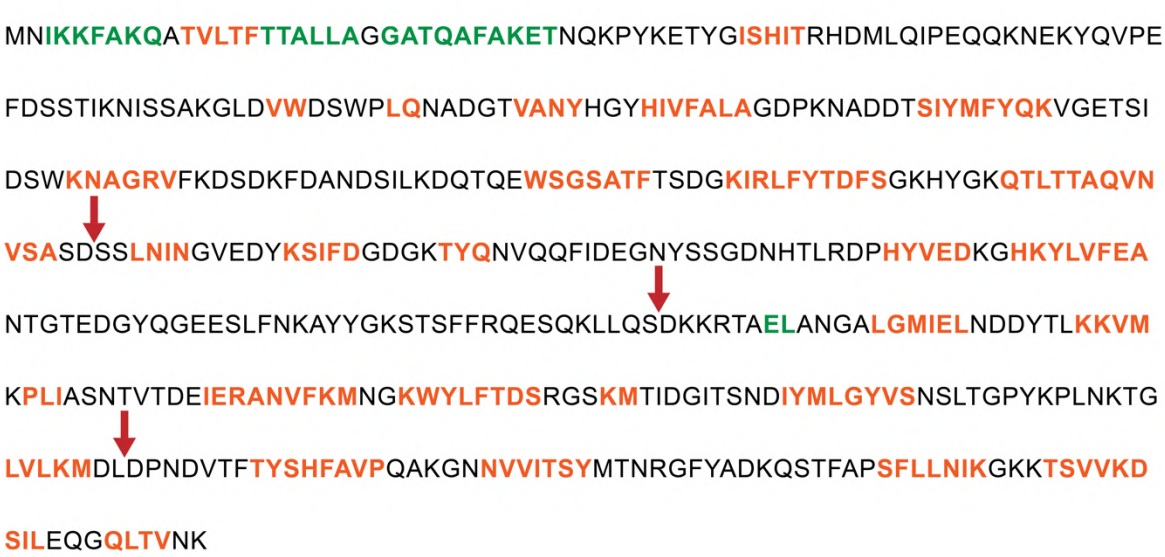
Cassette 138	VCA0453		1	0
Cassette 139	VCA0454		0	0
Cassette 140	VCA0455		1	0
Cassette 141	NA		0	0
Cassette 142	VCA0457	PD system	1	0
Cassette 143	VCA0458	PD system	0	0
Cassette 144	VCA0459		0	0
Cassette 145	VCA0460		0	0
Cassette 146	NA		1	0
Cassette 147	NA		0	0
Cassette 148	NA		1	0
Cassette 149	VCA0463		2	0
Cassette 150	VCA0464		0	0
Cassette 151	VCA0465		0	0
Cassette 152	VCA0466		0	1
Cassette 153	VCA0467		0	0
Cassette 154	VCA0468, VCA0469	TA	1	0
Cassette 155	VCA0470		2	1
Cassette 156	VCA0471, VCA0472, VCA0473	IS	0	0
Cassette 157	VCA0474, VCA0475	TA	0	0
Cassette 158	VCA0476		1	0
Cassette 159	VCA0477, VCA0478	TA	1	0
Cassette 160	VCA0479		0	0
Cassette 161	VCA0480		0	0
Cassette 162	VCA0481, VCA0482	TA	2	0
Cassette 163	VCA0483		1	0
Cassette 164	VCA0484		0	0
Cassette 165	VCA0485		0	0
Cassette 166	VCA0486, VCA487	TA	0	0
Cassette 167	VCA0488, VCA489	TA	2	0
Cassette 168	VCA0490		1	0
Cassette 169	VCA0491		1	0
Cassette 170	VCA0492, VCA0493, VCA0494		0	0
Cassette 171	VCA0495	TA	0	0

Cassette 172	VCA0496		1	0
Cassette 173	VCA0497, VCA0498	TA	0	0
Cassette 174	NA		0	0
Cassette 175	VCA0501		1	0
Cassette 176	VCA0503, VCA0504	TA	0	0
Cassette 177	VCA0505		1	0
Cassette 178	VCA0506		0	0

Supplementary figures

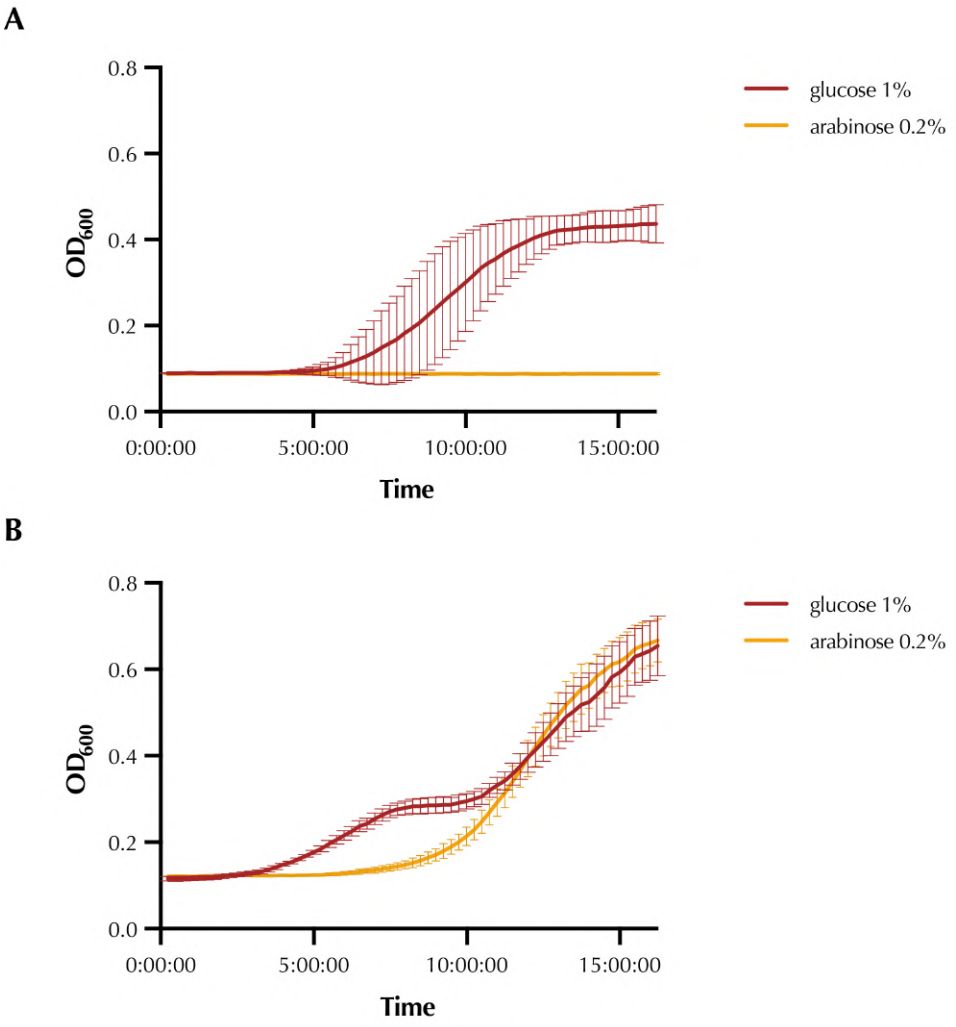
CcdB (*Vibrio fischeri*)

SacB (*Bacillus subtillis*)



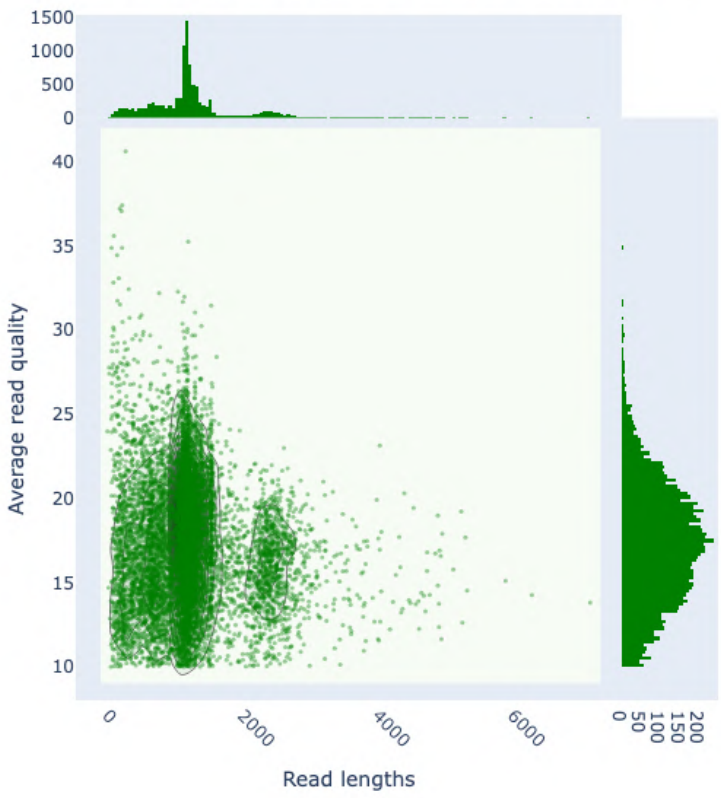
Supplementary Figure 1. Chosen locations to introduce the *attI1* site in the counter-selective markers

Aminoacidic sequences of the two chosen CSMs: *ccdB* from *V. fischeri* and *sacB* from *B. subtillis*. Orange represents β-sheets and green represents helixes. These features were predicted using Phyre2. Red arrows illustrate the locations where the *attI1* site was introduced.



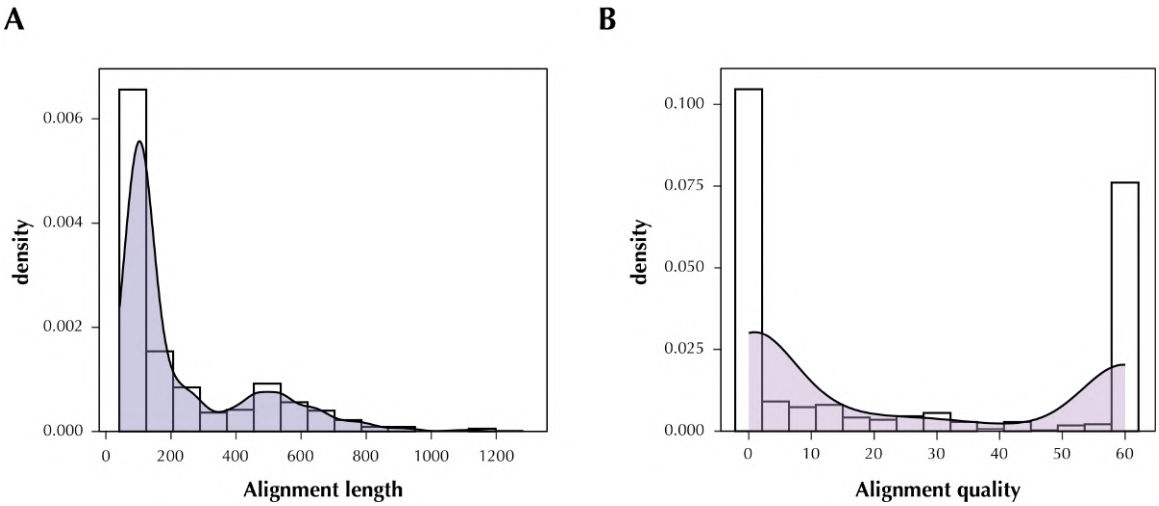
Supplementary Figure 2. Initial *ccdB::attI1* tests

Growth curves of strain *E. coli* DH5α containing a pBAD18::*ccdB::attI1* plasmid. Strains were grown in either LB + glucose 1% to repress *ccdB* expression or arabinose 0.2% to induce death. Data represents mean ± s.d of at least three biological replicates. **A.** Overnight growth curves immediately after plasmid transformation. **B.** Overnight growth curves after -80°C storage.



Supplementary Figure 3. Nanoplot report on read length and quality from ΔSI library sequencing

Average read quality and read length of the *V. cholerae* N16961 ΔSI library. Each dot represents one read. Two clades can be identified, one around 1,000 bp of read length (corresponding to the insertion of one cassette), and one around 2,000 bp (corresponding to the insertion of two cassetes).



Supplementary Figure 4. Analysis of reads that mapped to 0 cassettes

A. Density of alignment lengths. The majority of reads that mapped to zero cassettes are small (>200 bp). **B.** Density of alignment quality. Two peaks can be seen, one of very low quality (reads that can be discarded) and one of reads of high quality that only map to *attC* sites.