Genes on the move
Quantifying the adaptive power of the mobile integron



Célia Souque

Worcester College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2019

Abstract

Mobile integrons are widespread and clinically relevant drivers of antibiotic resistance. Acting as genetic platforms, they allow bacteria to capture, express and re-order mobile antibiotic resistance gene cassettes, whose expression levels depend on their distance from the promoter, located at the integration site. Cassettes can be integrated and excised by the integrase enzyme, produced by the bacteria in time of stress. It has been hypothesized that integrons allow bacteria to adapt quickly to changing antibiotic pressures by reshuffling the cassette order, and therefore their resistance levels, such that integrons provide 'adaptation on demand'.

To test this 'adaptation on demand' hypothesis, I used a custom three-cassette integron system transformed in both *E.coli* and *P. aeruginosa*. I first investigated the effect of cassette position on cassette expression at the phenotypic, transcriptional and translational level and confirmed a transcriptional origin of the cassette expression gradient, which was either strengthened or reduced by additional translational effects. I then combined experimental evolution and whole genome sequencing to quantify the evolvability benefits of cassette shuffling against increasing concentrations of antibiotics. By comparing strains with and without a functional integrase and focusing on two different cassettes and antibiotics, I observed striking differences in population survival rates and evolutionary pathways depending on the cassette and on the antibiotic. On one hand, with the antibiotic gentamicin and the *aadB* cassette, I saw an increase in evolvability and identified extensive cassettes rearrangements as well as cassette duplications in the populations with a functional integrase. On the other hand, with piperacillin and bla_{VEB-1}, integrase activity was linked with decreased evolvability and a much lower rearrangement frequency, but I observed unexpected integron-mediated recombinations in the plasmid conjugation machinery.

These results highlight the key impact of antibiotics and cassette characteristics on the potential integrase shuffling benefits and provide novel insights on the dynamics controlling cassettes and integrase prevalence in the environment. In my experiments, high cassette mobility was necessary to compensate for the high fitness cost of integrase expression, or the integrase was a burden for bacteria survival. Moreover, I showed integrase activity was not limited to cassette shuffling or cassette acquisition but may also promote cassette duplication and beneficial off-target recombination, with far-reaching consequences on plasmid and antibiotic resistance evolution.

Acknowledgements

I would like to acknowledge the numerous people who made this work possible: to paraphrase a popular saying, it also takes a village to write a thesis !

I would like to thank my supervisors, Craig MacLean and Kevin Foster, for their constant support and encouragements, and for how much I learnt in the past four years, both on the science of evolutionary biology and on how to become a well-rounded researcher. Many thanks to Jose Antonio Escudero for introducing me to the wonderful world on integrons, and without whom this project could not have existed.

I am very grateful to all the formidable members of both the MacLean and Foster lab I have had the pleasure to work with during this thesis. They truly were one of the best aspects of this PhD. I want to thank the great friends I made in Thailand: Mo Yin, Ricardo, Tom and Fai for their patient help with the boardgame project, and Ben Cooper for hosting me for the favourite part of my PhD.

Thank you to all the Worcester MCR members, friends and housemates I had the joy to meet and who made this Oxford experience so memorable. I thank the DTP crew, Charly, Liam, Chris and all the others, for the fun-filled evenings and outings. I would like to acknowledge my program, the Doctoral Training Partnership, its program directors, and my funder, the BBSRC, for giving me the freedom to become the biologist I always wanted to be.

Last but not the least, I would like to thank James, my parents, and my brother, not only for their heroic proof-reading efforts and writing encouragements, but also for their unconditional love and support all through this PhD. None of this could have been possible without you.

Contributions

I declare that this thesis is substantively my own work.

Contributions by others are listed below:

- **Chapter 2**: Jose Antonio Escudero built the plasmids in *E. coli* and performed part of the MICs in *E. coli*.
- **Chapter 3**: Jose Antonio Escudero created the WT and Δint plasmids. Library preparation and Next-Generation sequencing was performed by the Oxford Genomics Centre at the Wellcome Centre for Human Genetics.
- **Chapter 4**: Jose Antonio Escudero created the WT and Δint plasmids. Library preparation and Next-Generation sequencing was performed by the Oxford Genomics Centre at the Wellcome Centre for Human Genetics.

Contents

List of Abbreviations **viii**

1 Introduction **1**
1.1 The modern rise of antibiotic resistance . . . 1
1.2 Antibiotics & antibiotic resistance mechanisms . . . 3
1.2.1 Antibiotics mode of actions . . . 3
1.2.2 Evolution of antibiotic resistance . . . 4
1.2.3 Two problematic bacteria species: *E. coli* and *P. aeruginosa* 10
1.3 Integrons, genetic shuffling devices . . . 10
1.3.1 Integrons history and epidemiology . . . 10
1.3.2 Integron structure & mechanism . . . 11
1.3.3 Integrons & AMR . . . 15
1.4 Thesis structure . . . 17
1.4.1 Study system . . . 17
1.4.2 Experimental approach . . . 17
1.4.3 Thesis outline . . . 18

2 Cassette position & integron expression in *P. aeruginosa* and *E. coli* **20**
2.1 Abstract . . . 20
2.2 Introduction . . . 21
2.3 Material & Methods . . . 23
2.4 Results . . . 29
2.4.1 Resistance levels across arrays . . . 29
2.4.2 Cassette transcription levels . . . 31
2.4.3 *In-silico* array analysis . . . 33
2.4.4 Differences in integrase induction . . . 36
2.5 Discussion . . . 36
2.6 Acknowledgements . . . 43
2.7 Supplementary material . . . 43
2.7.1 MIC values . . . 43
2.7.2 ANOVA & Coefficients estimates . . . 45
2.7.3 BPROM & RBS Calculator simulations . . . 50

3 Experimental evolution with *aadB*: Shuffling and cassette duplication 53
3.1 Abstract . . . 53
3.2 Introduction . . . 54
3.3 Material & Methods . . . 56
3.4 Results . . . 61
3.4.1 Integrase activity can increase evolvability . . . 61
3.4.2 Extensive duplications and rearrangements mediated by the integrase . . . 63
3.4.3 x4 MIC time-point . . . 72
3.4.4 Duplications in pAMBL plasmids . . . 75
3.5 Discussion . . . 76
3.6 Acknowledgements . . . 82
3.7 Supplementary materials . . . 83
3.7.1 MIC in experimental conditions . . . 83
3.7.2 Recombinations identification by PCR . . . 84
3.7.3 SNP & recombinations at x1024 MIC . . . 85
3.7.4 SNPs & recombinations at x4 MIC . . . 90
3.7.5 pAMBL experimental evolution . . . 95
3.7.6 Cassette duplications in INTEGRALL . . . 97

4 Experimental evolution with *blaVEB-1*: costly integrase and plasmid rearrangements 98
4.1 Abstract . . . 98
4.2 Introduction . . . 99
4.3 Materials & Methods . . . 100
4.4 Results . . . 105
4.4.1 Integrase activity can decrease survivability . . . 105
4.4.2 Extensive rearrangements of the plasmid backbone & mutations in the conjugation machinery . . . 106
4.4.3 Recombinations and potential off-target effects of the integrase 111
4.4.4 Experimental evolution with *E. coli* . . . 114
4.5 Discussion . . . 116
4.6 Supplementary materials . . . 121
4.6.1 MIC in experimental conditions . . . 121
4.6.2 Loss of the plasmid in the absence of antibiotics . . . 122
4.6.3 PCR primers binding site . . . 123
4.6.4 Genomic variants . . . 124
4.6.5 Experimental evolution with MG1655:A5 . . . 128

5 Conclusion **130**
5.1 Summary 130
5.2 Future directions 133

Bibliography **141**

List of Abbreviations

HGT . . . . . Horizontal Gene Transfer

MIC . . . . . Minimum Inhibitory Concentration

PCR . . . . . Polymerase Chain Reaction

RT-qPCR . . . Quantitative reverse transcription PCR

LB . . . . . Luria-Bertani (broth)

MH . . . . . Mueller-Hinton (broth)

EUCAST . . . European Committee on Antimicrobial Susceptibility Testing

WGS . . . . . Whole Genome Sequencing

ORF . . . . . Open-reading frame

SD . . . . . Shine-Dalgarno (sequence)

IC50 . . . . . Half maximal inhibitory concentration

SNP . . . . . Single nucleotide polymorphism

T4SS . . . . . Type 4 Secretion System

ssDNA . . . . single-stranded DNA

CI . . . . . Chromosomal integron

MI . . . . . Mobile integron

1 Introduction

Integrons are bacterial gene capture and shuffling devices heavily involved in the spread of antibiotic resistance. Their powerful recombination system has been hypothesized to provide a high adaptive potential to their host, but their contribution to the evolution of antibiotic resistance has never been experimentally quantified. In this chapter, I will first introduce briefly the processes involved in the evolution of antibiotic resistance, with an emphasis on the mechanisms relevant for this thesis. I will then focus on the integron system and present our current understanding of the link between integrons and antibiotic resistance, before outlining the structure and aims of this thesis.

1.1 The modern rise of antibiotic resistance

At the start of the 20th century the discovery and development of the first antibiotics, Salvarsan (for treating syphilis), Protonsil (the first sulfonamide) and the most famous, penicillin, revolutionized medicine and created the foundations of modern pharmaceutical research (Aminov, 2010). Antibiotics, 'magic bullets' (as coined by Paul Ehrlich (Schwartz, 2004)) which selectively target microbes but spare the host, changed the course of human history by lessening the burden of one of the leading causes of human mortality: bacterial infections. They opened the

door for complex surgeries and medical treatments, from cancer chemotherapy to organ transplantations, where infections are otherwise unavoidable and became a cornerstone of modern medicine (O'Neill, 2014).

This antibiotic revolution is now threatened by the rise of antibiotic-resistant bacteria and especially the so-called ESKAPE pathogens (Rice, 2008) (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa*, and *Enterobacter* species), which are responsible for the majority of nosocomial infections (Rice, 2008) and are becoming increasingly multi- and pan- antibiotic resistant Boucher *et al.* (2009). Alexander Fleming, when he received his Nobel Prize for the discovery of penicillin, had already predicted this: *If you use penicillin, use enough* (Fleming, 1964). And indeed, penicillin resistance could already be observed in hospitals by 1947 (Barber, 1947). Since then, humanity has initiated an arms race against bacteria: as we discover and use new families of antibiotics, resistance to these antibiotics arises in bacteria as well (Davies and Davies, 2010). In 2014 a Review on Antimicrobial Resistance commissioned by the UK Prime Minister to economist Jim O'Neill estimated that in 2050, 10 million more people should die from antimicrobial resistant infections per year than currently do if antimicrobial resistance (or AMR) is left unchecked (O'Neill, 2014). This prediction must be taken with a grain of salt owing to its strong assumptions, due to our lack of knowledge about even the *current* public health burden of AMR (de Kraker *et al.*, 2016). Nevertheless, multiple reports have shown the rise of resistant bacteria around the globe in the last decade (Van Boeckel *et al.* (2019),Magiorakos *et al.* (2013),Frieden (2013)).

Relying on a constant supply of novel antibiotics to outrun resistance is not sustainable. After a 'golden age' of antibiotic development between the 1950s and 1970s, no new family of antibiotics has been discovered since 1987 (Conly and Johnston, 2005). This may hopefully change with the recent discovery of potential new antibiotic classes, such as Teixobactin (Ling *et al.*, 2015) and malacidins (Hover

et al., 2018) (targeting Gram-positive bacteria) as well as darobactin (against Gram-negatives) (Imai *et al.*, 2019). However bringing new antibiotics to the market is riddled with challenges, as seen with the disengagement from antibiotic research of leading firms in the pharmaceutical industry (Clift, 2019) and the high number of bankruptcies of antibiotic biotech companies (Melinta, in December 2019, being the latest to date (CIDRAP, 2019)). Developing new antibiotics is notoriously unprofitable, as the use of any novel antibiotic will first be kept voluntarily low by practitioners to limit the apparition of resistance, which does not make for an attractive business model. Novel incentives are being considered to fix the antibiotic market, but will require time and extensive investments (Clift, 2019).

We therefore need to better manage the antibiotics we already have, or will develop in the future, to limit the emergence of resistance. Such an endeavour will only be achievable through an in-depth understanding of the mechanisms and forces governing the evolution and spread of AMR.

1.2 Antibiotics & antibiotic resistance mechanisms

1.2.1 Antibiotics mode of actions

Antibiotics work by interfering with key mechanim of bacteria cell processes (Figure 1.1) including DNA replication (fluoroquinolones antibiotics, such as ciprofloxacin), DNA synthesis (trimethoprim and sulphonamides), RNA synthesis (rifamycins), cell wall synthesis (β-lactams, glycopeptides and lipopeptides) and protein synthesis (macrolides, amphenicols, aminoglycosides and tetracyclines) (Kohanski *et al.*, 2010). Antibiotics can either induce cell death or stop cell replication depending on the bacteria species and the antibiotic concentration (Pankey and Sabath, 2004).

In particular, aminoglycosides (such as gentamicin, which we will use in Chapter 3) affect protein synthesis by binding to the 16S unit of the 30S ribosome, causing codon misreading and the creation of mistranslated proteins (Krause *et al.*, 2016).

The mistranslated proteins then penetrate into the bacteria membrane, disrupting its integrity and leading to the entry of even more antibiotics (Kohanski *et al.*, 2010).

β-lactams, including penicillin, piperacillin and meropenem (used in Chapter 4), are the most widely used class of antibiotics (Bush and Bradford, 2016a). They target specific transpeptidases, also called penicillin-binding proteins, which promote the cross-linking of peptidoglycans, the building blocks of the bacteria cell wall. By inhibiting the action of the transpeptidases, β-lactams ultimately induce cell lysis (Kohanski *et al.*, 2010).

The actual pathway through which antibiotics induce cell death remains hotly debated, but a common pathway has been suggested for aminoglycosides, β-lactams and potentially fluoroquinolones antibiotics. The disruption of the cell membrane, either due to mistranslated proteins (aminoglycoside) or inhibition of peptidoglycans cross-linking (β-lactams), has been hypothesised to stimulate the envelope and redox-responsive two-component systems. This perturbs the cell metabolism and the membrane potential, causing the formation of lethal hydroxyl radicals. These radicals then ultimately induce cell death through damages to the cell DNA, lipids and proteins (Kohanski *et al.*, 2010). Although this model for antibiotic induced cell death has proven to be controversial (Van Acker and Coenye, 2017), it is now widely recognised that antibiotics have complex effects on cellular physiology that go well beyond the inhibition of the targeted molecules (Dwyer *et al.*, 2015).

1.2.2 Evolution of antibiotic resistance

Bacteria resistance to antibiotics is not a novel phenomenon (D 'Costa *et al.*, 2011). Numerous micro-organisms naturally produce antibiotics, and have been doing so for millions of years, either to fight competitors or as potential signalling molecules (Sengupta *et al.*, 2013). Bacteria have evolved mechanisms to resist these compounds long before their use by humans: a highly diverse collection of genes encoding for resistance against various antibiotics was found in 30 000-year-old permafrost samples (D 'Costa *et al.*, 2011). Along with *de novo* mutations, the

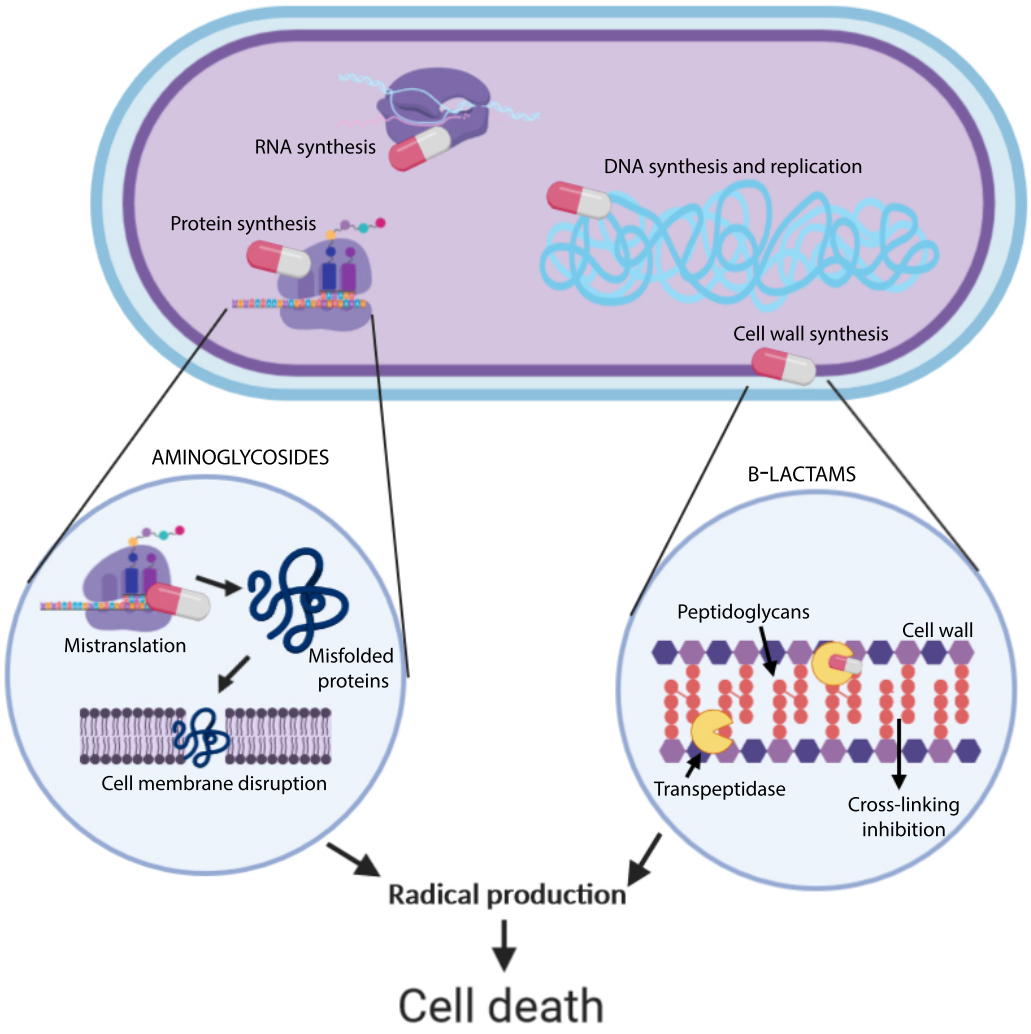


Figure 1.1: Diversity of antibiotic mode of action: antibiotics may target DNA replication and synthesis (fluoroquinolones, trimethoprim and sulphonamides), RNA synthesis (rifamycins), cell wall synthesis (β-lactams, glycopeptides and lipopeptides) and protein synthesis (macrolides, amphenicols, aminoglycosides and tetracyclines). In particular, aminoglycoside antibiotics target bacterial ribosomes, inducing the production of misfolded proteins which can then disturb the cell membrane, while β-lactams antibiotics inhibit the transpeptidase enzymes which ensure the cross-linking between peptidoglycans, blocking cell-wall synthesis. According to the 'reactive oxygen species hypothesis' (Van Acker and Coenye, 2017), the membrane's disruption then leads to the production of hydroxyl radicals, which induces cell death through damage to essential cell components. Figure created with BioRender.com .

recent spread of antibiotic resistance relies heavily on the mobilisation of such ancient genes, like beta-lactamase enzymes (Hall and Barlow, 2004), but also on fundamental bacterial systems, like stress responses, horizontal gene transfer (HGT) and mobile genetic elements (MGE) (Hughes and Datta, 1983).

Antibiotic resistance mechanisms

To counter the effects of antibiotics, bacteria have evolved complex mechanisms which can be categorized into a few broad strategies (Lin *et al.*, 2015) (Figure 1.2.A).

First, bacteria may directly target the antibiotic molecule itself through dedicated enzymes. β-lactamases and aminoglycoside modifying enzymes are a prime example of this strategy and are the leading cause of resistance against β-lactams and aminoglycoside antibiotics (Lin *et al.*, 2015). As we will see in the next section, these enzymes can either be encoded on mobile plasmids or are permanent feature of the bacteria genome such as the β-lactamase AmpC, located on the chromosome of *P. aeruginosa* (Lodge *et al.*, 1990a).

Second, bacteria may reduce antibiotic penetration either by limiting the entry of antibiotics within the cell or by actively exporting them through machinery called efflux pumps (Lin *et al.*, 2015). In Gram-negative bacteria, antibiotics usually enter through channels called porins, and mutations reducing porin production or enhancing efflux pump activity have been shown to frequently arise during therapy (Breidenstein *et al.*, 2011).

Third, bacteria may alter the target site of the antibiotics. A classical example is the development of rifampin resistance through mutation of the *rpoB* gene, which generates high-level of rifampin resistance through a single point-mutation (Qi *et al.*, 2014). Bacteria may also become resistant through the acquisition of a resistant copy of the target gene, such as the *dfr* genes which provide resistance to trimethoprim through the production of a trimethoprim resistant dihydrofolate reductase (Amyes *et al.*, 1990).

Resistance mechanisms have been shown to engender a biological cost, impacting the rate at which bacteria grow and multiply (Andersson and Levin, 1999). This cost can originate from the metabolic burden associated with the carriage of a mobile genetic element (San Millan and MacLean, 2017) or from the disruption of normal physiological processes by the resistance mechanism (Yang *et al.*, 2017). However such costs can often be alleviated through compensatory mutations which adapt the cell metabolism to the presence of the resistance mechanism (Händel *et al.*, 2013).

Regulatory responses

Alongside genetic changes, bacteria can also adapt to the presence of antibiotics through regulatory mechanisms that alter cellular physiology (Poole, 2012). These responses may enhance the expression of multidrug efflux pumps (Fetar *et al.*, 2011) or induce the formation of biofilms which reduce sensitivity to antibiotics (Stewart, 2002). An example of such a regulatory mechanism is the SOS response, a global response to DNA damage (McKenzie *et al.*, 2000) (Figure 1.2B). In the presence of single-stranded DNA (ssDNA) the RecA protein catalyzes the auto-proteolysis of the repressor LexA, which regulates the expression of genes belonging to the SOS regulon. The number and identity of genes contained in this regulon vary from species to species, but these genes are usually involved in DNA repair (Baharoglu and Mazel, 2014b). Beyond DNA repair, the SOS response is linked to several bacteria adaptive processes, such as biofilm formation (Gotoh *et al.*, 2010), persistence (Dörr *et al.*, 2009) and expression of virulence factors (Quinones *et al.*, 2005). The SOS response also enhances genome plasticity due to induction of stress-induced mutagenesis, increase of intra-chromosomal recombinations (Cirz *et al.*, 2006) and integron integrase expression (Guerin *et al.*, 2009), as we will see in the next section.

Several mechanisms can lead to the presence of ssDNA and therefore the induction of the SOS response in a bacteria cell, even in the absence of stressors: replication stalling (Hishida *et al.*, 2004), transcription replication collision (blockage of the DNA polymerase by transcription complexes) (Dutta *et al.*, 2011) and R-loop formation (RNA-DNA hybrids created by stalling of the RNA polymerase)

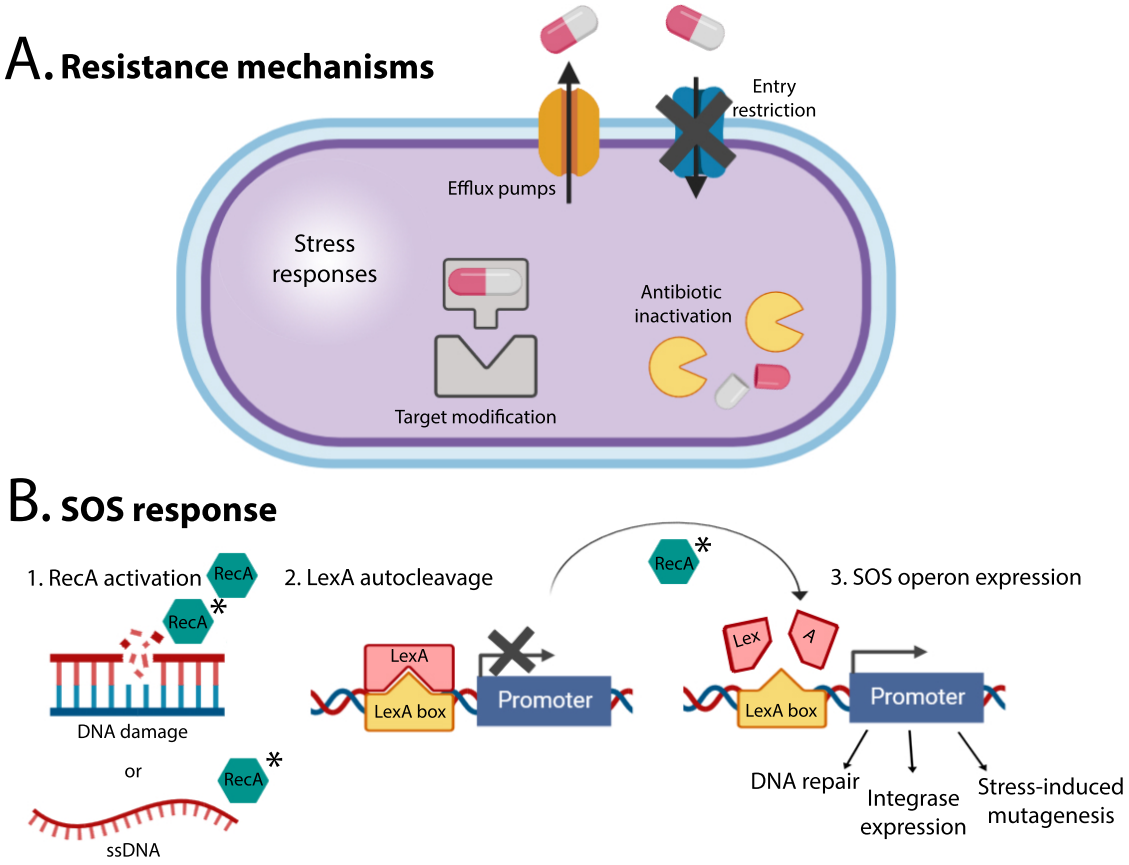


Figure 1.2: A. **Diversity of antibiotic resistance mechanisms**: bacteria may counter the action of antibiotics through direct inactivation of the antibiotic, modification or supplementation of the antibiotic target site, reduction of antibiotic intracellular concentration through efflux pumps or reduced porin expression and adaptation of the cell physiology through various stress responses. B. **SOS response mechanism**: in presence of DNA damage or ssDNA from HGT, the RecA protein is activated and catalyses the cleavage of the repressor LexA, leading to increased expression of the genes from the SOS operon. Figure created with BioRender.com

(Wimberly *et al.*, 2013). Direct attacks to the DNA integrity are potent inducers, such as UV radiation or reactive oxygen species (Jena, 2012). A number of antibiotics also induces the SOS response, such as fluoroquinolones (Ysern *et al.*, 1990), aminoglycosides (Baharoglu and Mazel, 2011a) and β-lactams (Miller *et al.*, 2004b). A final inducer of the SOS response is horizontal gene transfer, due to the the entry of ssDNA during transformation (Baharoglu *et al.*, 2012) and conjugation (Baharoglu *et al.*, 2010).

Horizontal gene transfer & Mobile genetic elements

As we have seen in the previous paragraph, bacteria can evolve resistance either through *de novo* mutations or through the acquisition of resistance genes by horizontal gene transfer (HGT). HGT allows the transfer of resistant determinants between bacteria through several mechanisms (Figure 1.3), the most common being conjugation (transfer of DNA from one cell to the other through conjugative pili), transformation (uptake of naked DNA by the cell as ssDNA) and transduction (transfer of DNA through bacteriophages) (Von Wintersdorff *et al.*, 2016).

Of these mechanisms, conjugation is considered the most relevant for the spread of resistance (Getino and de la Cruz, 2014). Conjugative plasmids as well as integrative and conjugative elements rely on two sets of genetic components for conjugation (Getino and de la Cruz, 2014): a mobility component for DNA processing which includes the origin of transfer (oriT) and the relaxase, and a mating-pair formation component which encodes for a Type 4 Secretion System (T4SS).

Intra-genomic mobility between plasmids and chromosome is also possible through promiscuously recombining loci called transposons, which use site-specific recombinase to insert themselves into the host DNA (Frost *et al.*, 2005) and have been heavily involved in the spread of antibiotic resistance (Liebert *et al.*, 1999). Another site-specific recombination system is the integron, which acts as a genetic platform for the acquisition and re-ordering of small mobile genetic elements called gene cassettes, which often encode for antibiotic resistance genes.

1.2.3 Two problematic bacteria species: *E. coli* and *P. aeruginosa*

This work will focus on two clinically relevant bacterial species, *E. coli* and *P. aeruginosa*, both Gram-negative negative facultative anaerobes.

Workhorse of microbiologists, *E. coli* is a common and fast-growing gut bacteria found mainly mostly in the mammalian gut microbiome, but also in soil and water (Blount, 2015). It is a member of the Enterobacteriaceae family, which

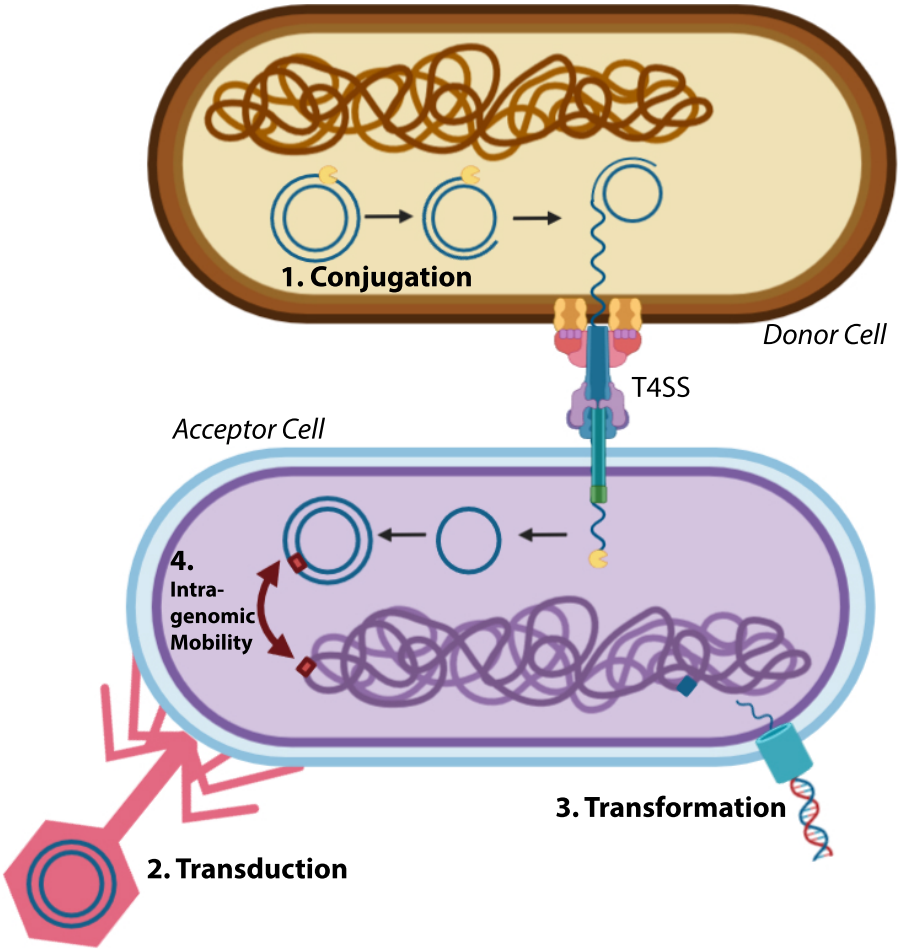


Figure 1.3: Inter- and intra-cellular DNA transfer mechanisms: 1. Conjugation: the DNA processing happening during conjugation is represented here. The relaxase (in yellow) cleaves and binds to the plasmid DNA and is then recruited by the T4SS, where it acts as pilot protein to translocate the ssDNA into the recipient cell. The ssDNA is then re-circularized and replicated to generate a new copy of the original plasmid (Getino and de la Cruz, 2014). 2. Transduction: transfer of plasmids through bacteriophages. 3. Transformation: bacteria-mediated uptake of environmental DNA. 4. Intra-genomic mobility: mobile genetic elements, such as transposons, may jump from plasmids to chromosomes, and vice-versa. Figure created with BioRender.com .

contains other pathogens such as *Klebsiella* and *Shigella*. *E. coli* is a very diverse species, containing both harmless commensal and highly pathogenic strains, as the enterohaemorrhagic *E. coli* O157:H7 (Lim *et al.*, 2010). *E. coli* is characterized by extensive genomic plasticity, with a typical *E. coli* genome containing around 80% of accessory genes (Lim *et al.*, 2010). It is intrinsically susceptible to most antibiotic classes but can easily acquire resistance by horizontal gene transfer (Poirel *et al.*, 2018). Antibiotic resistance in *E. coli* is now a growing concern (Poirel *et al.*, 2018)

with the rise of extended-spectrum β-lactamase producer *E. coli* strains in Europe (ECDC/EMEA, 2009) and the emergence and spread of plasmid-encoded colicin resistance in *E. coli*, a last resort antibiotic (Liu *et al.*, 2016).

By comparison *P. aeruginosa* is not a common member of the human flora but can become an opportunistic pathogen (Lister *et al.*, 2009). It is one of the leading causes of nosocomial infections (Weinstein *et al.*, 2005) and is especially prevalent in the lungs of patients with cystic fibrosis (Cystic Fibrosis Foundation, 2013). Recent studies have reported a dramatic increase in the prevalence of resistant *P. aeruginosa* strains, linked with an increase in mortality and hospital length of stay (Weinstein *et al.*, 2005).

P. aeruginosa has a 'mosaic' genome consisting of a highly conserved core genome inter-spaced with strain-specific genomic segments which provide genomic plasticity (Mathee *et al.*, 2008). Inversely to *E. coli*, the *P. aeruginosa* core genome represents 88% of an average genomeValot *et al.* (2015). It has the highest proportion of predicted regulatory genes of all sequenced genomes, with 8.4% of *P. aeruginosa* genes involved in regulation (Stover *et al.*, 2000), helping the bacteria easily to various environments and nutrient sources (Frimmersdorf *et al.*, 2010).

P. aeruginosa is characterized by a high level of intrinsic antibiotic resistance. Its outer-membrane permeability is 12 to 100-fold lower than *E. coli*'s (Yoshimura and Nikaido, 1982). This is due to the replacement of general diffusion porins (as can be found in *E. coli* or other Enterobacteriaceae) by specific porins specialised in the uptake of various nutrients (Chevalier *et al.*, 2017), which reduces the intake of antibiotics within the cell. *P. aeruginosa* also actively exports antibiotics through four clinically relevant efflux pumps (Valot *et al.*, 2015). Moreover, the *P. aeruginosa* core genome contains inducible resistance genes encoding for antibiotic-modifying enzymes, such as the cephalosporinase AmpC (Lodge *et al.*, 1990b) or the kanamycin-specific phosphotransferase APH(3$\acute{}$)-II (Okii *et al.*, 1983). This basal resistance level can easily be increased through mutations leading to AmpC or efflux pumps over-expression, or down-regulation of porins expression (Moradali *et al.*, 2017).

1.3 Integrons, genetic shuffling devices

1.3.1 Integrons history and epidemiology

Integrons are a prime example of the recruitment by bacteria of ancient bacterial structures to counter the human use of antibiotics. Consisting of an integrase gene followed by an array of (usually) promoterless gene cassettes (Figure 1.4, structure described in detail in the next section), integrons are common features of bacteria chromosomes and have been identified in 10% of bacterial sequenced genomes (Cury *et al.*, 2016), most commonly in γ-proteobacteria from either marine or soil/freshwater environments (Mazel, 2006). The integrase is believed to have first evolved hundred of millions of years ago, as its species-specific clustering indicates an origin pre-dating speciation of these genera (Mazel, 2006)(Rowe-Magnus *et al.*, 2001).

Now integrons are often found in conjunction with antibiotic resistance, to the point that the prevalence of the *IntI1* integron integrase has been proposed as an indicator for the presence of resistance genes in environmental studies (Ma *et al.* (2017),Lucassen *et al.* (2019)). Integron involvement with clinical antibiotic resistance started early on: integrons were retrospectively identified in some of the first multi-drug resistant bacteria ever reported, in an outbreak in Japan in 1960 (Nakaya *et al.*, 1960)(Mazel, 2006). The prevalence of integrons detected in Gram-negative clinical isolates now reaches up to 60% of isolates in some studies (Escudero *et al.*, 2015) and integrons are found at similar frequencies in pathogens and commensal bacteria from live-stock (Gillings *et al.*, 2008). This spread is thought to have been initiated by the mobilization of chromosomal integrons by other mobile genetic elements such as conjugative plasmids and transposons (Gillings *et al.*, 2008). The most commonly found class of integrons, called class 1, is thought to have arisen through repeated association with transposons, during which it gained its 3'-conserved segment containing *sul1*, a sulfonamide resistance gene and *qacE*Δ, a partially deleted quaternary ammonium resistance cassette (Gillings *et al.*, 2008).

From this evolutionary history emerges a separation between two types of integrons (Mazel, 2006): on one side the sedentary chromosomal integrons, or 'super-integrons', found on the chromosome of environmental bacteria, and on the other the so-called mobile integrons due to their association with transposons and conjugative plasmids, which are involved with antibiotic resistance. Mobile integrons are then further separated into different classes (1 to 5) depending on the sequence of their integrase gene. Each class is thought to have arisen separately through mobilisation with different transposases (Cambray *et al.*, 2010a). Chromosomal and mobile integrons differ by their structure and content (Escudero *et al.*, 2015): chromosomal integrons can contain more than 200 cassettes, with a wide variety of functions, while mobile integron arrays are much shorter, with 1 to 8 cassettes which often encode for at least one antibiotic resistance determinant (Zhang *et al.*, 2018). Cassette exchange between chromosomal and mobile integrons has been reported (Rowe-Magnus *et al.*, 2002), and is thought to be a powerful genetic source for the acquisition of novel resistance genes in mobile integrons. Overall, more than 130 integron cassettes encoding for resistance determinants have been identified (Partridge *et al.*, 2009), providing resistance against almost all antibiotic families (Escudero *et al.*, 2015).

The rest of this work will focus on mobile, and especially class 1, integrons.

1.3.2 Integron structure & mechanism

All integrons share a common structure and are composed of a stable platform followed by a variable cassette array (Escudero *et al.*, 2015).

Stable platform: integrase & attI site

This platform is made of the gene encoding for the integrase, *IntI*, followed by a recombination site, called *attI*, which enables cassette insertion at the start of the array (Hall *et al.*, 1991). The integron integrase belongs to the family of Y-recombinases and its expression is ensured by a promoter *Pint* under the control of

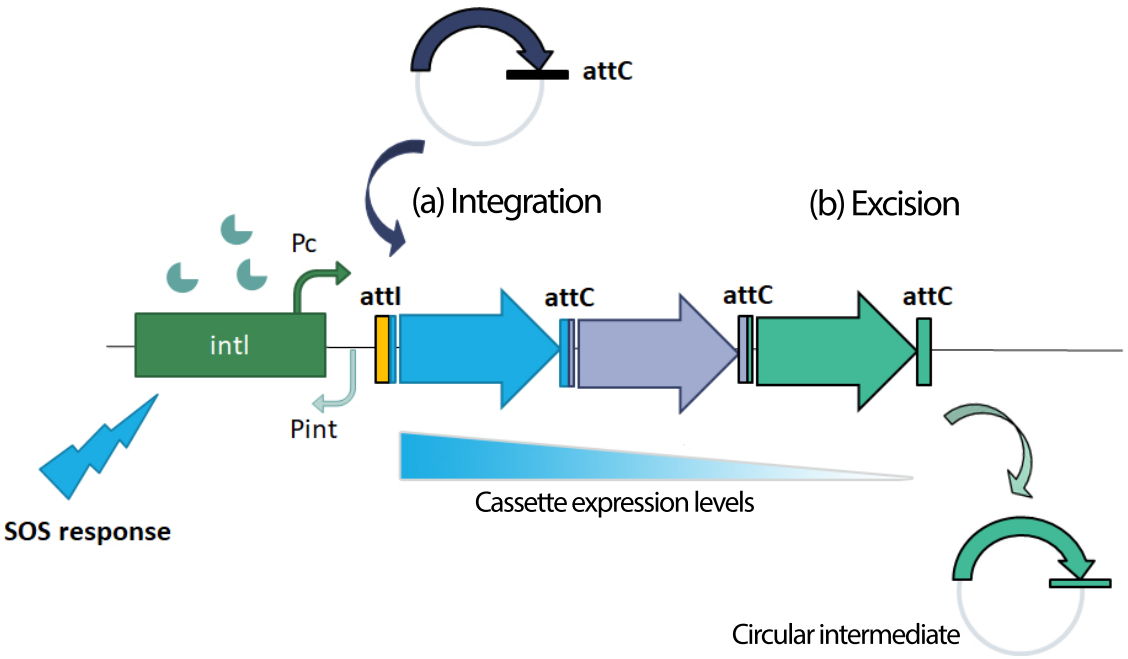


Figure 1.4: Organisation of the integron
The integron consists of an integrase gene, *intI*, followed by an array of promoterless gene cassettes (represented here by arrows). When under stress the bacteria produces the integrase enzyme which promotes cassette integration (a) or excision (b). Integration happens through recombination of the *attI* site of the integrase and the *attC* site of a free circular cassette, while recombination between a cassette *attC* site and the *attC* of the preceding cassette causes the cassette excision into its circular form. Cassettes are expressed by the Pc promoter present at the beginning of the integron. Only cassettes closest to the integrase are usually expressed, while the others can serve as a 'reservoir' of genes.

the SOS response (Guerin *et al.*, 2009). Little homology can be found between the *attI* sites of different integrons but these sites usually contain a core site made of two integrase binding sites, L and R (Figure 1.5 A). The recombination point during cassette insertion is found at a 5'-GTT-3' triplet within the R box, with cleavage taking place between the A and the C of the complementary strand (Hall *et al.*, 1991). Two additional integrase binding sites characterized by GTTR consensus sequences can be found in the *attI* site of class 1 integrase, but these sites are not strictly necessary for recombination (Gravel *et al.*, 1998).

Variable array: cassettes & attC sites

Integron cassettes are small mobile genetic elements usually consisting of a promoterless open reading frame (ORF) terminated by a recombination site called the *attC*

site (Recchia and Hall, 1995a).

Transcription of cassettes is guaranteed by the promoter Pc located within the integrase at the start of the array (Collis and Hall, 1995), while a minority of mobile integron cassettes contain their own promoter (Escudero *et al.*, 2015). Two ORFs located within or just upstream of the *attI* site have also been shown to improve cassette translation for the 20% of class 1 integron cassettes which lack a translation initiation region (Hanau-Berçot *et al.*, 2002) (Papagiannitsis *et al.*, 2017). Cassette expression has been shown to depend on its distance from the start of the array (Collis and Hall, 1995). The origin of this gradient is debated: *attC* sites could act as Rho-independent terminators to limit transcription (Collis and Hall, 1995) or impede ribosome progression and translation (Jacquier *et al.*, 2009). An inverse correlation has been found between the strengths of the different Pc and Pint variants: the stronger the Pc promoter and the expression of the cassettes, the less the integrase is induced. This is due to the impact of variations in Pc on the sequence of the integrase protein, affecting its recombination properties (Jové *et al.*, 2010), but also to transcriptional interference between some Pc variants (PcS) and Pint (Guérin *et al.*, 2011).

Responsible for cassette mobility, *attC* sites are extremely diverse in their sequences but share a palindromic organization with two inverted regions, R"-L" and L'-R' (Figure 1.5 A), which create a common single-stranded and hairpin-shaped DNA secondary structure (Francia *et al.* (1999) , Bouvier *et al.* (2005))(Figure 1.5 B). Upon folding, this structure resembles the canonical R and L boxes of the *attI* site and is the substrate recognized and recombined by the integrase (Francia *et al.* (1999) , Johansson *et al.* (2004)). Similarly to the *attI* site, recombination happens at a 5'-GTT-3' triplet located in the R' box on the bottom strand (Bouvier *et al.*, 2005), which is part of the only conserved motif between *attC* sites: a GTTRRRY sequence matched by its RYYYAAC inverse in the R" box. Several additional features can be distinguished in the structure of the *attC* sites: the presence of bases without complement in the R"-L" region which become extrahelical bases extruding from the hairpin structure, a non-complementary spacer region between the R"-R'

and L"-L' boxes (called an unpaired central spacer) and a variable terminal structure located at the end of the hairpin stem (Figure 1.5 C). These structural features have been found to influence recombination frequency and play an important role in the determination of the bottom strand as the recombinogenic strand, which is necessary to ensure the insertion of cassettes in the correct orientation relative to the *attI* site and the Pc promoter (Bouvier *et al.* (2009), Nivina *et al.* (2016)).

Integrase-mediated recombinations

Through its atypical sequence-independent selectivity the integrase can recognize and recombine both the *attI* and *attC* sites. Recombination between a cassette *attC* site and the *attC* site of the upstream cassette lead to the excision of the cassette as a circular intermediate (Collis and Hall, 1992a). These excised cassettes are then preferably reinserted at the start of the array through recombination between their *attC* site and the *attI* site. Due to the recombination pathway, which includes a replicative resolution of the atypical Holliday junction formed during recombination (Loot *et al.*, 2012), both *attC* x *attC* and *attI* x *attC* reactions are semi-conservative and a copy of the ancestral array (before excision/integration) is always created. In the case of mobile integrons, where arrays may be present on multi-copy plasmids, inter-molecular *attC* x *attC* and *attI* x *attI* reactions are also possible, leading to the creation of plasmids co-integrates. Recombination may also happen between *attI* or *attC* sites and non-specific secondary sites containing the GTT sequence in other parts of the genome. These non-canonical reactions can lead to the insertion of a cassette outside of the integron (Francia *et al.*, 1999) as well as to the potential formation of deleterious chromosome dimers (Cambray *et al.*, 2010a).

Costly integrase

The expression of the integrase has been shown to carry a strong biological cost: in species without an SOS response where the integrase is constantly expressed, the integrase strongly reduces host fitness and is quickly inactivated through mutations (Starikova *et al.*, 2012). The biological basis for this cost remains unclear, but the fitness cost of a given integrase has been shown to correlate with its catalytic

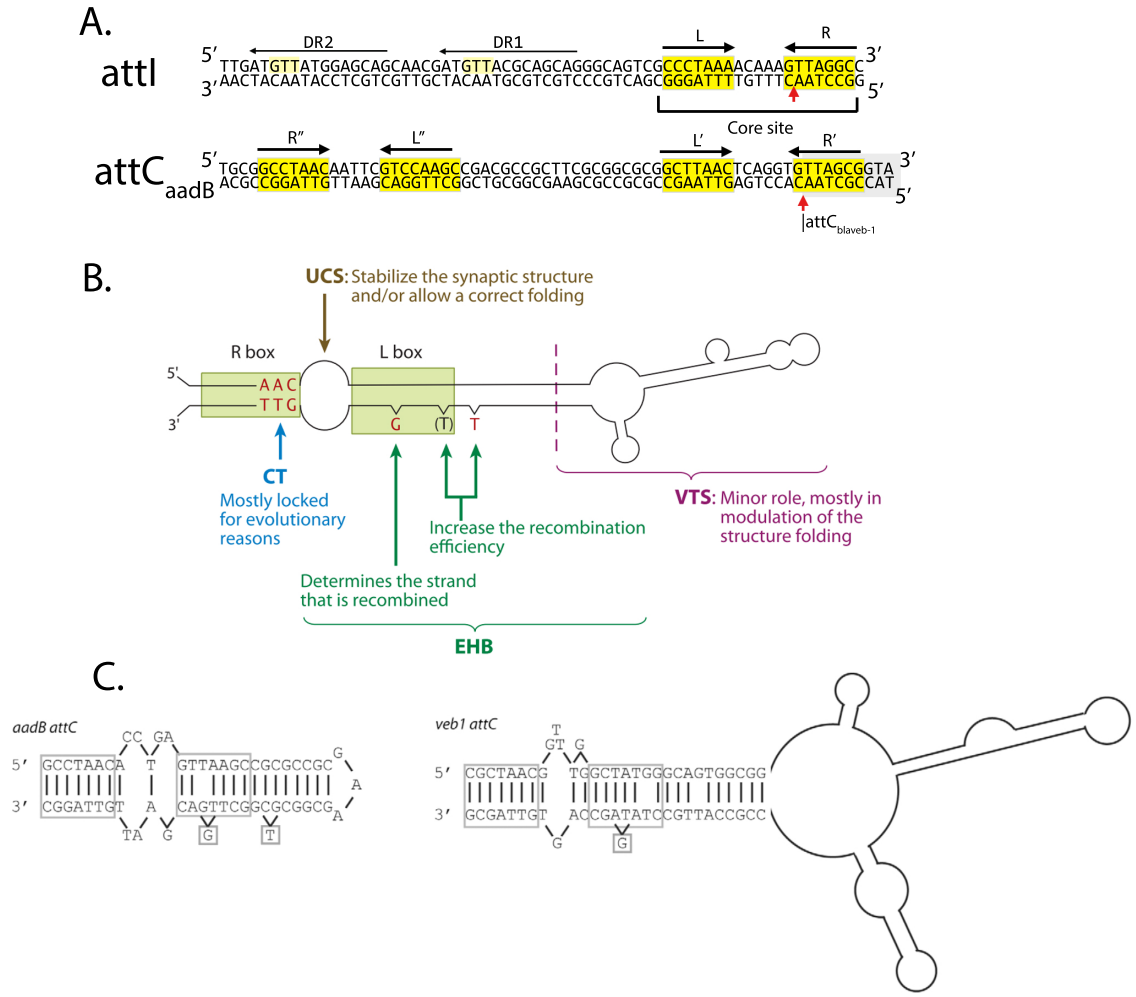


Figure 1.5: Sequence and structure of attI and attC sites A. Sequence of the *attI* site from R388 and the *attC* site of the *aadB* cassette. The core site boxes are highlighted in yellow, with the recombination point indicated by a red arrow. The GTT triplets of the direct repeats (DR) bound by the integrase are indicated in pale yellow. When inserted into an array the *attC* site creates a hybrid hairpin with the *attC* site of the following cassette, the sequence of which is here indicated in grey (Escudero *et al.*, 2015). **B.** Schematic representation of the folded structure of an *attC* site. CT: conserved triplet. UCS: unpaired central spacer. EHB: extrahelical base. VTS: variable terminal structure. Figure adapted from (Cambray *et al.*, 2010a). **C.** Representation of the *attC* sites of the *aadB* (left) and *blaVEB-1* cassettes used later in this study. Figure adapted from (Aubert *et al.*, 2012).

activity, with integrases with a higher activity being more costly (Lacotte *et al.*, 2017). This cost is hypothesized to originate from reduced chromosomal stability due to off-target recombinations between chromosomal secondary sites leading to deleterious deletions (Harms *et al.*, 2013). In species with an SOS response, this cost is alleviated through SOS regulation, making integrons low-cost and efficient genetic platforms (Lacotte *et al.*, 2017). This guarantees the stability of the integron array in the absence of stressors but allows cassette recombination in stressful environments or during HGT.

1.3.3 Integrons & AMR

Given our knowledge of integron biology, how can we explain its evolutionary success in environments with a high selection pressure for antibiotic resistance? Based on its mechanism, the integron may increase bacteria evolvability against antibiotics in two ways.

The integron is a useful platform for gene acquisition in association with horizontal gene transfer through the broad specificity of the integrase (MacDonald *et al.*, 2006), which does not require homologous sequences or specific motifs for cassette acquisition. While not mobile by itself, the integron is a great vector for the spread of resistance genes across species in conjugation with other mobile genetic elements.

However, HGT is not required to explain the integron's success. Through the combination of the cassettes' expression gradient along the array and the recombination abilities of the integrase, the integron has the potential to modulate cassette expression and to reduce the cost of resistance. Cassettes located at the end of the array can theoretically be kept at a low expression level, and therefore create a bank of low cost genes, but remain available to be brought forward for maximal expression when required by cassette shuffling, so that integrons provide 'adaptation on demand'.

The adaptive power of the integron's shuffling has been demonstrated theoretically: a mathematical model of a population of integrons showed the modulation of cassette expression through rearrangements provided a selective advantage from integrase activity for a wide range of parameters (Engelstädter *et al.*, 2016). Experimentally, we know cassette rearrangements, including cassette duplication, can result from integrase activity (Collis and Hall, 1992b). Moreover unregulated expression of the integrase has been shown to lead to the evolution of a higher level of antibiotic resistance (Barraud and Ploy, 2015), but with strong constraints on the resulting cassette rearrangements. However, the adaptive power of integron shuffling has never been experimentally quantified, which will be the aim of this work.

Finally, while we focus here on the benefits of cassette shuffling related to antibiotic resistance, the adaptive advantages provided by the integron shuffling potentially go beyond clinical settings and antibiotics. Chromosomal integrons act as hotspots for genetic diversity for numerous species, notably the Vibrionaceae(Rowe-Magnus *et al.*, 2003), *Xanthomonas* species (Gillings *et al.*, 2005) or *Pseudmonoas stutzeri* (Holmes *et al.*, 2003). While we do not know the function of most of these integron cassettes metagenome, these integrons arrays are often highly dynamic, with less than 10% of shared cassettes between cultivar of the same species Koenig *et al.* (2011). It has therefore been hypothesized that integrons are especially useful for niche adaptation and the creation of genetic diversity within bacteria species (Koenig *et al.*, 2011).

1.4 Thesis structure

1.4.1 Study system

In this work, I will investigate the evolvability benefits of cassette shuffling through a custom three cassette integron system spanning all possible cassette arrangements, placed in both *E. coli* and *P. aeruginosa*.

Our experimental system consists of six arrays containing the *aadB*, *dfrA5* and *blaVEB-1* cassettes. Each of these cassettes provides resistance to a different antibiotic family: *aadB* encodes for an aminoglycoside modifying enzyme (Ramirez and Tolmasky, 2010), *blaVEB-1* for an extended-spectrum β-lactamase (Naas *et al.*, 2008) and *dfrA5* for trimethoprim resistant dihydrofolate reductase (Sundström *et al.*, 1988).

These three antibiotic families are routinely used in the clinic and are considered either 'critically important' (aminoglycosides and some β-lactams, including piperacillin) or 'highly important' antimicrobials by the WHO (trimethoprim and remaining β-lactams) (Group *et al.*). In particular, gentamicin (which we will use in Chapter 3) is a commonly used aminoglycoside discovered in 1963 (Becker and Cooper, 2013) and used for the treatment of sepsis (Gonzalez-Padilla *et al.*, 2015), ventilator-associated pneumonia (Ioannidou *et al.*, 2007) and zoonotic diseases (Jackson *et al.*, 2013). Piperacillin (used in Chapter 4) is a broad-spectrum β-lactam derived from penicillin (Schoonover *et al.*, 1995). It is especially potent against *P. aeruginosa* and often used in combination with tazobactam, a β-lactamase inhibitor, to treat nosocomial infections (Bush and Bradford, 2016b). Finally, trimethoprim is commonly used to treat uncomplicated urinary tract infections (Brogden *et al.*, 1982).

We use the R388 plasmid as a backbone, a well-studied IncW broad-host-range plasmid which naturally contains an integron (Fernández-López *et al.*, 2006). We replaced this integron by our custom arrays and transformed these plasmids in both *Escherichia coli* and *Pseudomonas aeruginosa*. Mobile integrons have been found at high frequency in clinical settings in both these species (Rao *et al.* (2006), Xu *et al.* (2009) Fonseca *et al.* (2005)). As we saw previously, these species differ in regulatory responses and intrinsic antibiotic resistance, which may impact the effect of the integrase. Moreover, the levels of resistance provided by the same plasmid resistance gene can vary between two species, with the example of plasmid-borne *aacC3* gene providing gentamicin resistance in *P. aeruginosa*, but not in *E. coli* (Kato *et al.*

(1982), Van Boxtel and Van De Klundert (1998)). By using two different host backgrounds, we will be able to investigate the potential differences in resistance and evolvability provided by the same integron between bacteria, improving our understanding of the dynamics shaping integrons prevalence.

1.4.2 Experimental approach

In this work, I will investigate integrase activity using experimental evolution. Consisting in the propagation of multiple populations of organisms, such as bacteria, for thousands of generations in a controlled environment, experimental evolution allows empirical testing of evolutionary hypotheses and theories (Buckling *et al.*, 2009). The flagship experiment of this approach is the *E. coli Long Term Evolution Experiment*, or LTEE (Lenski and Travisano, 1994): since February 1988, 12 populations of *E. coli* are transferred each day with samples regularly taken and frozen. Running now for more than 70 000 generations, the LTEE has provided great insights on the dynamics of adaptation (Lenski and Travisano, 1994). The recent technological developments in whole genome sequencing have magnified the power of experimental evolution studies, by allowing researchers to investigate both phenotypic and genomic dynamics underlying evolution (Bruger and Marx, 2018).

When applied to antibiotic resistance, this approach creates powerful and elegant studies such as the MEGA plate experiment, which tracks the evolution of trimethoprim resistant lineages of *E. coli* across a physical antibiotic landscape on a gigantic agar plate and shows how the spread of resistant mutants may be spatially constrained by sensitive clones (Baym *et al.*, 2016). Experimental evolution has been used successfully by scientists to answer relevant questions on the evolution of resistance, such as the influence of environmental structure (Santos-Lopez *et al.*, 2019), historical contingency (Card *et al.*, 2019) or stress responses (Torres-Barceló *et al.*, 2015) on antibiotic resistance evolution.

1.4.3 Thesis outline

First, I will use this system to study in depth the extent and origins of the cassette expression gradient (Chapter 2). Through a combination of experimental and *in-silico* tools, I will investigate the effect of cassette position and array composition at the phenotypic, transcriptional and translational level, in both *E. coli* and *P. aeruginosa*. We can hypothesize that the integron shuffling benefits will be proportional to the range of cassette expression levels attainable by cassette shuffling: too small of a range and the adaptive benefit provided by cassette rearrangement may not be enough to compensate for the cost of cassette rearrangement. Understanding the impact of cassette order and array composition on cassette expression is therefore a required step to understand the potential benefits of cassette shuffling.

Second, I will use experimental evolution combined with next-generation sequencing to measure the impact of integrase activity on resistance evolution. I will directly measure the adaptive benefits of bringing a distant cassette closer to the promoter by comparing the evolvability of strains with and without a functional integrase containing the relevant resistance cassette in last position of the array. We can hypothesize that the benefits of cassette shuffling will depend on the characteristics of both the cassette and the stressor involved. I will therefore study cassette shuffling for two cassettes: first *aadB*, a highly mobile cassette which expression relies heavily on position, against gentamicin (Chapter 3), then *blaVEB-1*, a cassette with low recombinogenic potential and constant expression, against piperacillin (Chapter 4).

Finally, I will summarise the results of this work and highlight future directions and challenges of integron biology (Chapter 5).

2

Cassette position & integron expression in *P. aeruginosa* and *E. coli*

2.1 Abstract

Heavily involved in the spread of multi-drug resistance, integrons are powerful systems allowing bacteria to adapt rapidly to changing environments through the acquisition or shuffling of promoterless gene cassettes. Cassette expression levels depend on their position from the promoter located upstream of the integron cassette array, allowing bacteria to modulate cassette expression through integrase-mediated cassette rearrangements. The potential benefits of this cassette shuffling will depend on the differential in cassette expression between positions, which needs to be important enough to compensate for the cost of integrase expression, and whose mechanistic origins remain debated. In this study we thoroughly investigate the effect of position on cassette expression in a three-cassette system in both *E. coli* and *P. aeruginosa* at the phenotypic, transcriptional and translational level, and confirm a transcriptional origin of the cassette expression gradient, with additional translational effects either reducing or strengthening the effect of position.

2.2 Introduction

Integrons are genetic structures acting as recombination platforms, allowing the acquisition, excision and reordering of (usually) promoterless genes, called genes cassettes (Escudero *et al.*, 2015). Integrons found on plasmids, called mobile integrons, have been heavily involved in the spread of antibiotic resistance and are now widespread in both the environment and clinical settings (Gillings, 2014). Mobile integrons all consist of an integrase gene named *IntI* followed by an array of gene cassettes, which often encode for antibiotic resistance genes (Recchia and Hall, 1995a). The expression of those genes is guaranteed by the presence of a promoter Pc located upstream of the *IntI* gene (Stokes and Hall, 1989), and has been shown to depend on the position of the cassette in the array (Collis and Hall, 1995): the closer a cassette is to the promoter, the more it will be expressed. Integron cassettes are characterized by the presence of a specific hairpin structure at their 3' end called the *attC* site (Hall *et al.*, 1991). Through the action of the integrase enzyme, bacteria can excise cassettes through site-specific *attC* x *attC* reactions, and also integrate new cassettes through recombination with another attachment site called the *attI* site, near the 5' end of the integrase gene (Collis and Hall, 1992b). This mechanism increases bacteria adaptability in two ways: through the acquisition of new genes by horizontal gene transfer and inter-integron exchange of cassettes (Rowe-Magnus *et al.*, 2002), but also through regulation of the cassettes expression within the integron. By reshuffling the cassette order, integrons can modulate the cassettes expression level: useful cassettes can be brought forward to ensure maximal expression, while unnecessary cassettes can be kept at the end of the array at a lower cost, ready to be brought forward when needed (Escudero *et al.*, 2015). Thanks to its regulation by the SOS response (Guerin *et al.*, 2009) this reshuffling mechanism has been hypothesised to provide 'adaptation on demand' for bacteria, helping its host to adapt quickly to changing environments (Escudero *et al.*, 2015).

Integrase expression carries an important fitness cost, possibly due to reduced genomic stability (Harms *et al.*, 2013). The benefits of the 'adaptation on demand'

mechanism will be directly linked to the gradient in cassette expression along the array. If the difference in cassette expression is too small, the cost of the integrase will surpass any fitness increase provided by the cassettes reshuffling, potentially leading to loss of the integrase, as shown by a mathematical model (Engelstädter *et al.*, 2016).

Cassette expression has been shown to be modulated by both transcriptional and translational effects. For example, the Pc promoter at the start of the array can be of variable strength, which is inversely correlated to the excision activity of the integrase (Jové *et al.*, 2010). Several possible mechanisms have been proposed for the gradient in cassette expression itself. While initially thought to be due to *attC* sites acting as transcription terminators (Collis and Hall, 1995), a translational origin for the expression gradient has also been offered, with *attC* sites reducing cassette expression by impeding ribosome progression (Jacquier *et al.*, 2009). Position is however not the only effect known to impact cassette expression. Some cassettes have been shown to contain their own promoters (Bissonnette *et al.*, 1991), ensuring their own expression independently of their position, and even helping the expression of downstream or upstream cassettes (Tansirichaiya *et al.*, 2019), and *attC* sites can modify the translation of the downstream cassette(Jacquier *et al.*, 2009). These effects may directly impact the potential benefits of cassette shuffling by modifying the expression gradient along the array.

In order to understand the potential evolvability benefits of the 'adaptation on demand' mechanism and its involvement in the integron's evolutionary success, we investigated the effect of position in mobile integrons through a three-cassette integron system, covering all six possible array orders. We used a combination of experimental and computational tools to systematically test the effect of cassette position and array composition at the phenotype, transcription and translation level in both *E. coli* and *P. aeruginosa*. We were able to determine the importance of transcriptional and translational effects impacting our system: a weak cassette expression gradient is due to the attenuation of transcription, but translational effects specific to certain cassettes play an important role in integron cassette

expression and can either strengthen or reduce the importance of position. The evolvability benefits of cassette shuffling may therefore be strongly specific to the cassettes and antibiotics considered.

2.3 Material & Methods

Bacteria and growth conditions

A complete list of strains and plasmids can be found in the final Appendice Table 5.4. Unless stated, bacteria cultures were grown in LB Miller broth (Sigma Aldrich). Media was supplemented with 100 mg/L of carbenicillin for *E. coli* and with 100 mg/L of ceftazidime for *P. aeruginosa* strains.

Strains construction

Arrays construction

Six integron arrays (Table 2.1) covering all possible cassette orders were created using plasmid R388 (Avila and de la Cruz, 1988) as backbone and three resistance cassettes: *aadB*, *bla-VEB1* and *dfrA5*. The *bla-VEB1* and *aadB* cassettes were amplified from the integron of the *E. coli* MG-1 clinical isolate (Poirel *et al.*, 1999), while the *dfrA5* cassette was obtained from an enteroinvasive *E. coli* strain isolated in Senegal (Gassama *et al.*, 2004). These cassettes were then assembled into custom integron arrays using Gibson assembly and inserted into the plasmid R388 while replacing its original dhfr - orf9 - qacEΔ1 - sul1 integron array (Fernández-López *et al.*, 2006). The original R388 strong PcS promoter variant (high cassette expression but low integrase activity (Jové *et al.*, 2010)) was replaced by the weaker PcW promoter to guarantee high integrase activity. The final arrays were then transformed into chemically competent *E. coli* MG1655.

Creation of Pseudomonas strains

Pseudomonas arrays bearing strains were created through filter mating using the previous *E. coli* strains as donors and *Pseudomonas* PAO1 as recipient. Bacteria were incubated overnight in Luria-Bertani (LB) broth with 100mg/L of carbenicillin

Array name	1st position	2nd position	3rd position
A1	dfrA5	blaVEB	aadB
A2	blaVEB	aadB	dfrA5
A3	blaVEB	dfrA5	aadB
A4	dfrA5	aadB	blaVEB
A5	aadB	blaVEB	dfrA5
A6	aadB	dfrA5	blaVEB

Table 2.1: Arrays composition

at 37°C for the donors and in LB Miller broth without antibiotic at 42°C for the recipient. The next day cells were spun down, washed, and re-suspended in LB broth, before mixing in a 1:4 donor to acceptor ratio. The mix, as well as pure donor and acceptor controls, were put on filters placed on LB agar without antibiotic and incubated at 37°C overnight. Afterwards, filters were placed in tubes containing LB media and agitated. The resulting supernatants were plated on LB agar supplemented with 50 mg/L of kanamycin and 25 mg/L of ceftazidime and incubated for 48h. As *P. aeruginosa PAO1* has a higher innate resistance to kanamycin (41.9 mg/L, see Table 2.3)) than *E. coli MG1655* (1.5 mg/L) due to a chromosomally encoded phosphotransferase (Okii *et al.*, 1983), kanamycin was used to select against the *E. coli* donors, while ceftazidime was used to select for the presence of the plasmid in the *P. aeruginosa* transconjugants. The final colonies were controlled by PCR for the presence of the plasmid and the absence of *E. coli* DNA.

Antibiotic susceptibility testing

The minimum inhibitory concentrations (MIC) for each antibiotic were determined in Mueller-Hinton (MH) broth for *E. coli* strains and in cation-adjusted Mueller-Hinton Broth 2 (MH2) for *Pseudomonas* strains, following the broth microdilution method from the Clinical and Laboratory Standards Institute guidelines (CLSI, 2017). Briefly 5x10^5 c.f.u bacteria inocula were prepared using individual colonies grown on selective agar (*P.aeruginosa*) or diluted overnight culture (*E. coli*) in doubling concentrations of antibiotics and incubated for 20h in a shaking incubator at 37°C. The next day, plates' optical density (OD) was read using a Biotek Synergy

2 plate reader. Wells were considered empty when the overall OD_{600} was under 0.1 and the MIC for each assay was defined as the minimal concentration in which growth was inhibited in all three technical replicates (separate wells, but grown on the same day from the same inoculum). When differences in MICs between the different arrays were small, interlaced 2-fold-increasing concentration scales were used to more finely determine the MICs. The final MICs values are the average of three to four replicate assays (from separately prepared inocula, on different days) for *P. aeruginosa*, and one to two replicates for *E. coli*.

Cassette expression by RT-qPCR

RNA & DNA extractions

Each bacterial strain was inoculated in MH2 medium supplemented with antibiotic and grown overnight at 37°C with constant shaking (225 rpm). The overnight cultures were diluted 1:50 in fresh MH2 without antibiotic and incubated until they reached an OD_{600} between 0.5 and 0.6. Both RNA and DNA were extracted for each sample. Half of each culture was mixed with RNAprotect Bacteria Reagent (Qiagen) according to the manufacturer instruction. Total RNA extraction was performed using the RNeasy Mini kit (Qiagen) on the QIAcube extraction machine (Qiagen). The other halves were treated with RNase and used to extract total gDNA using the DNeasy Blood & Tissue Kit on the QiaCube (Qiagen). Each strain was extracted three times from cultures started on different days.

Plasmid copy number

Plasmid copy number was determined for each gDNA sample using the approach described in (San Millan *et al.*, 2015a): all samples were first digested in order to linearize the plasmid using the restriction enzyme BamHI (BamHI FastDigest, Thermo Fisher Scientific) according to the manufacturer instruction for gDNA digestion. Linearizing the plasmid increases DNA template accessibility and therefore prevents from underestimating the plasmid copy number (Providenti *et al.*, 2006). The amplified regions were controlled for the absence of BamHI

restriction sites. The *orf9* gene was used as the R388 plasmid target and the mono-copy *rpoD* and *ssrA* genes were used as chromosomal target for *P. aeruginosa* and *E. coli* respectively (Primers given in Appendice Table 5.1). Amplifications were carried out using the Luna Universal Probe qPCR Master Mix (New England Biolabs) for the *P. aeruginosa* samples and iTaq Universal SYBR Green Supermix (BioRad) for the *E. coli* samples. Thermal cycling protocol consisted of 1s/3s at 95°C (Denaturation), 20s/30s at 60°C (Annealing/Extension) for 40 cycles for *P. aeruginosa*/*E. coli*. Melting curve analysis was included for samples detected without probes. 4-fold dilution standard curves were included to control for differences in primer efficiencies. Plasmid copy number was calculated as the ratio between the plasmid and chromosomal target DNA quantities.

Reverse transcription and qPCR

All RNA samples were treated with the TURBO DNA-free Kit (ThermoFisher) to eliminate genomic DNA. Concentration of the RNA samples was quantified using the Quantifluor RNA system (Promega). cDNA was synthesized from 100ng of RNA templates using random primers from the GoScript Reverse Transcription Mix (Promega). qPCR was carried out on the StepOnePlus Real-time PCR platform (Applied Biosystems) using the iTaq Universal SYBR Green Supermix. Each cassette, as well as two reference genes (*actpA* and *acp* for *P. aeruginosa* samples, *ssrA* and *rrsA* for *E. coli* samples), were amplified using the primers described in Appendice Table 5.1 in two technical replicates for each extraction. Standard curves for each pair of cassette primers were included in each PCR using restriction enzyme-digested gDNA as template, and used to quantify the amount of target cDNA in each sample. This allowed us to control for inter-run variations, compensate for differences in primers efficiency and compare the differences in transcription level between cassettes, since we assume our standard, being genomic DNA, has equal number of copies of each cassette. Melting curve analysis were included after each run to test for non-specific amplification products. For each biological replicate the cassettes transcript levels were normalized based on the geometric means of

the two internal reference genes, using the first array A1 as a reference, before division by its plasmid copy number.

Integrase expression by RT-qPCR

Culture conditions & RNA extraction

To measure integrase expression under antibiotic treatment, overnight cultures in selective media of PAO1:A4 and MG1655:A4 were diluted 1:50 in fresh MH2 without antibiotic and incubated until they reached an OD_{600} between 0.5 and 0.6. These cultures were then diluted 1:20 in fresh MH2 supplemented with antibiotics at a concentration corresponding to half the minimum inhibitory concentration, determined specifically for these conditions, alongside a control culture diluted without antibiotic. As this is not the standard starting inoculum for MIC determination and given MICs can be heavily influenced by the size of the initial inoculum (Brook, 1989) in an antibiotic-dependant manner, we determined the minimal inhibitory concentrations at this inoculum size in order to be able to compare antibiotic treatments. We measured the growth curves of inoculates prepared in the same conditions, and incubated in doubling concentrations of antibiotics for 20h. The minimal inhibitory concentrations were determined as the concentrations for which the final OD_{600} were less than the initial density. The values obtained for *E. coli* were 0.025, 4 and 16 mg/L for ciprofloxacin, gentamicin and piperacillin respectively, and 0.2, 96 and 256 mg/L for *P. aeruginosa*. After incubation for 2h at 37°C with shaking, RNA extraction was performed following the protocol mentioned earlier.

Reverse transcription and qPCR

Extracted RNAs were treated and transcribed in cDNA as described previously. For each treatment one chromosomal reference gene (*acp* for *P.aeruginosa* and *qssrA* for *E. coli*), one plasmid reference gene (*dfrA5*) and one to two target genes (*Int1* and *LexA* for *P. aeruginosa*, *Int1* only for *E. coli*) were amplified in three technical replicates from three biological replicates. Standard curves from digested gDNA for the plasmid and target genes were included in each amplification. Cassette and target

transcripts were first normalized using the chromosomal reference gene and sample grown without antibiotic as a reference. As cassettes have been shown to be stably expressed in R388 (Fernández-Lópe *et al.*, 2014), for each sample we divided the quantity of integrase transcripts by the *dfrA5* cassette transcription level to control for differences in plasmid copy number or plasmid transcription between samples.

Statistical analysis

All statistical analyses were conducted using R (version 3.6.1) and RStudio (Version 1.2.5).

Multivariate linear model

For all linear models diagnostics plots were used to control for normality of the data, residuals distribution, the assumption of equal variance, and outliers influence.

Resistance levels: The following multivariate linear model was first applied for each species/antibiotic family combination: $log_2(MIC) \sim Antibiotic + Cassette\ Position + Antibiotic * Cassette\ Position$. Cassette position was considered a continuous variable, from 1 to 3. Any non-significant terms were then removed from the model until the best fitting model was obtained. For the β-lactams resistance in *E. coli* dataset an extra model was fitted, $log_2(MIC) \sim Antibiotic + Previous\ cassette$, where *Previous cassette* is a categorical variable representing the identity of the previous cassette. To account for the repeated model testing, a Bonferroni correction was applied to the resulting anova p-values by multiplying by the number of models tested (here five). Similarly within each model, the p-value for each antibiotic was adjusted for the number of antibiotics tested for this species/antibiotic family combination.

Transcription levels : The linear model applied to *E. coli* and *P. aruginosa* transcript levels is the following: $Cassette\ transcript\ /Plasmid\ copynumber \sim Extraction + Gene + Cassette\ Position + Gene * Cassette\ Position$. To account for the repeated model testing, a Bonferroni correction was applied to the

resulting anova p-values by multiplying by the number of species tested (here two). Similarly within each model, the p-value for each gene coefficient was adjusted by multiplying by the number of genes tested (here three)

Dunnett tests

Post-hoc Dunnett tests were carried out in R using the *multcomp* package (Therneau, 2015) to correct for multiple comparison when comparing integrase expression under antibiotic treatment to the no antibiotic control.

In-silico analysis

Putative promoters were detected using the web-based software BPROM (Solovyev and Salamov, 2011) (Softberry programs) by considering the sequences of each cassette and integrase separately. Only results with an overall LDF score over 2 were considered.

Translation initiation rate at the cassettes start codons were predicted with the RBS Predictor software from De Novo DNA (version 2.1) (Espah Borujeni *et al.*, 2017) using a mRNA sequence starting at the PcW and covering the whole array. Both *E. coli* and *P. aeruginosa* were used as host organism and provided identical results, due to a common 16S rRNA 3' end.

2.4 Results

2.4.1 Resistance levels across arrays

Using the broad host range conjugative plasmid R388, we created six three-cassette integron arrays (Table 2.1), with each cassette providing resistance to a different family of antibiotics: *dfrA5* (encoding for a trimethoprim resistant dihydrofolate reductase (Sundström *et al.*, 1988)), *bla-VEB1* (an extended-spectrum β-lactamase (Naas *et al.*, 2008)) and *aadB* (an aminoglycoside-2'-adenylyltransferase, providing resistance to gentamicin, kanamycin and tobramycin (Ramirez and Tolmasky, 2010)). We transformed these plasmids into both *P. aeruginosa* and *E.coli*. All three cassettes are widely distributed and clinically relevant: *aadB* was detected in

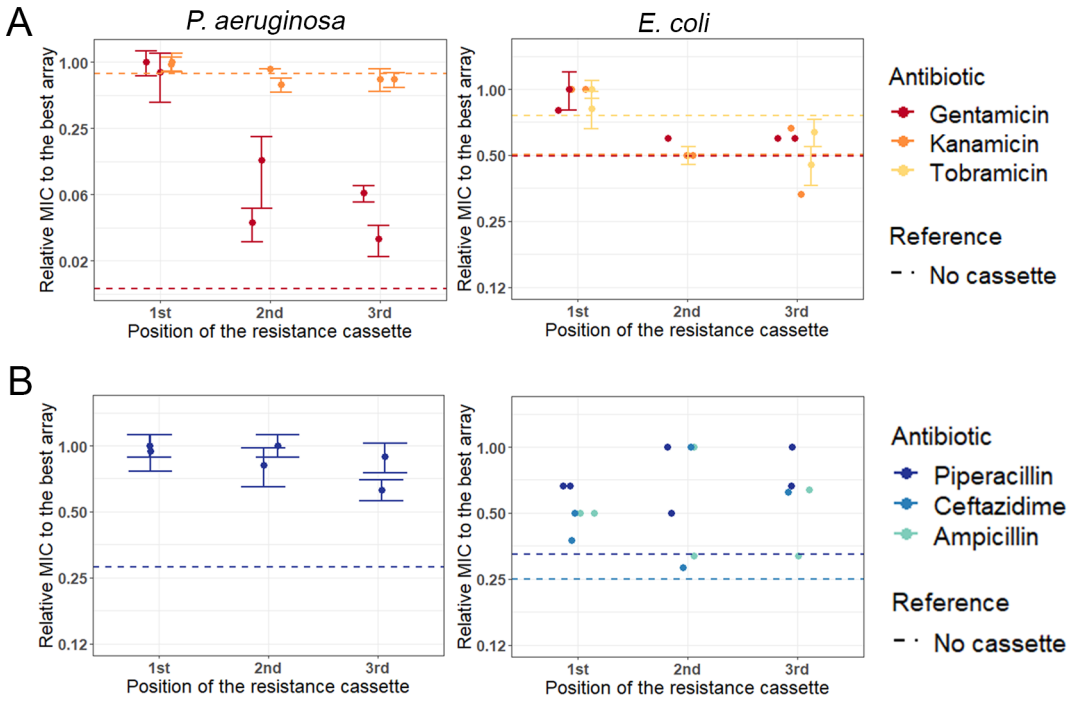


Figure 2.1: Levels of antibiotic resistance for the different arrays depending on the position of the resistance cassette for aminoglycosides (A) and β-lactams (B). Only antibiotics for which an effect of position is visible are represented. MICs are normalised relative to the MIC of the array with the highest concentration. Reference lines indicate the MICs of the strains without arrays, respectively PAO1 and MG1655. Error bars represent standard errors. Note the different scale used for the aminoglycoside figure for *P. aeruginosa*.

81% of all aminoglycoside-resistant enterobacteria in the US (Miller *et al.*, 1997), *blaVEB-1* has been linked to outbreaks in Vietnam and Thailand but also France, Belgium and Korea (Naas *et al.*, 2008) while integron-associated *dfrA5* and similar *dfr* genes are found in more than 60% of trimethoprim-resistant isolates in Europe (Blahna *et al.*, 2006).

We first measured the MICs for each of these arrays. The overall resistance levels as well as the impact of cassette position were heavily dependent on the cassettes and on the antibiotics considered. High levels of resistance independent of position are common: *dfrA5* makes both species virtually resistant to any concentration of trimethoprim while in *Pseudomonas aeruginosa*, *blaVEB-1* provides high levels of

resistance to ceftazidime, ticarcillin and ampicillin in all arrays (Table 2.3).

We first considered the impact of position in *P. aeruginosa*. A strong effect is visible with gentamicin resistance, with a 10-fold decrease in resistance between arrays with *aadB* in first and second position, followed by a less dramatic decrease between second and third position (Figure 2.1 A). This effect is visible, but is much weaker with kanamycin, another aminoglycoside, with only a 50% reduction in resistance between first and second position. *aadB* provides an overall very weak increase in kanamycin resistance as less resistant arrays are actually more sensitive to kanamycin than the wild-type PAO1. Similarly, a weak effect of position is visible with *blaVEB-1* in *P. aeruginosa* and the antibiotic piperacillin, where we can see a 50% decrease in resistance between arrays containing the *blaVEB-1* cassette in first and last position (Figure 2.1 B). We used a linear model to quantify the effect of position for each family of antibiotics (Table 2.5 and Table 2.4): position was an overall strong predictor of MICs levels for aminoglycoside resistance (Bonferroni corrected p-value $>$ 1.79e-06, ANOVA adjusted R-squared: 0.5973, F-statistic: 21.76 on 3 and 39 degrees of freedom (DF), overall corrected p-value: 9.53e-08), with a significant antibiotic x position interaction (corrected p-value $>$2.83e-05): the model predicts a negative correlation between MIC and position for both antibiotics (kanamycin: -0.24, gentamicin: -2.11), but the effect of position is significant only for gentamicin. The effect of position for piperacillin resistance was non significant (corrected p-value p=0.6145, ANOVA adjusted R-squared: 0.089, F-statistic 2.652 on 1 and 16 DF).

The resistance profiles provided by the same arrays show differences depending on the host, with final resistance levels much higher in *P. aeruginosa* than in *E. coli*. The strong positional effect visible with gentamicin resistance is much reduced in *E.coli*, with a 2-fold decrease in MIC between first and second position for gentamicin, kanamycin and tobramycin resistance, but the effect remains significant (Bonferroni corrected p-value $>$1.712e-03, ANOVA ajusted R-squared: 0.5665, F-statistic: 11.45 on 4 and 28 DF, overall corrected p-value: 6.18e-05 - Table 2.6). Surprisingly, in *E. coli* the best array for resistance against all three β-lactams

measured (ampicillin, ceftazidime and piperacillin) is A1, followed by A6, both of which do not have *blaVEB-1* in first position. Instead, in both of these arrays *blaVEB-1* is located directly after *dfrA5*, suggesting an effect of *dfrA5* on the subsequent cassette. Indeed there was no significant effect of position (corrected p-value >1 - Table 2.7) but the identity of the previous cassette was a good predictor of resistance levels (corrected p-value: 0.03179, ANOVA adjusted R-squared: 0.9362, F-statistic 70.68 on 4 and 15 DF, overall corrected p-value: 7.51e-09 - Table 2.8), with a positive effect of *dfrA5* on the subsequent cassette (0.646, std error=0.2867). Surprisingly this effect is only visible in *E.coli*.

In all but one antibiotic/host combination the effect of position is subtle and hard to detect using micro-broth dilution, given that two-fold variation in MIC values are often considered negligible by EUCAST standards. Overall the resistance levels given by the different arrays produce a more complex picture than expected, with variations in the effect of cassette position between species and antibiotics.

2.4.2 Cassette transcription levels

We then investigated the effect of position at the transcriptional level and measured cassette transcripts by RT-qPCR in all six arrays and for each species. To compensate for potential differences in plasmid copy numbers which could impact cassette expression, plasmid copy number was also measured for each sample, and transcript levels were adjusted accordingly. The array A2 in *P. aeruginosa* showed reduced levels of cassette expression as well as disproportionate integrase expression when measured, and is therefore excluded from this analysis until verification of the plasmid sequence by whole genome sequencing.

The effect of cassette position is visible in the transcript levels, but again subject to variations between cassettes and species (Figure 2.2). In *P. aeruginosa* the effect of position is most visible with the *aadB* cassette, where cassette transcript is decreased 3-fold between the first and the last position, while this effect is much weaker with the *blaVEB-1* and *dfrA5* cassettes. Surprisingly, the overall transcript level of the

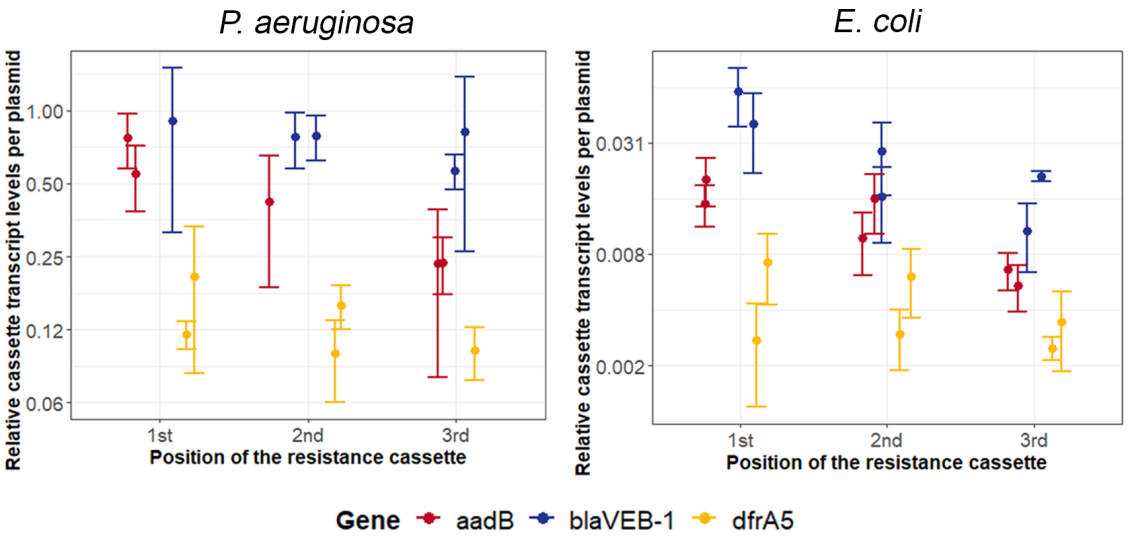


Figure 2.2: Relative transcript levels for each cassette depending on its position. Error bars represent the standard error of three independent biological replicates.

blaVEB-1 cassette is the highest, while *dfrA5* transcript levels are weakest, even after correcting for differences in primers efficiencies. We first carried out a multiple regression analysis using the *P. aeruginosa* results (Table 2.9: Adjusted R-squared: 0.6189, F-statistic: 11.21 on 7 and 37 DF, Bonferroni corrected p-value: 3.2e-07) and verified that both cassette identity (corrected p-value <9.96e-07) and cassette position (corrected p-value <.0171) are strong predictors of the final transcript levels, with a strong negative correlation between cassette position and transcript level for the *aadB* cassette (estimate = -0.23, standard error (std) = 0.07) but weaker for *dfrA5* (estimate = -0.02, std = 0.08) and *blaVEB-1* (estimate = -0.07, std = 0.08).

In *E. coli* the effect of position is similarly visible, this time stronger with *blaVEB-1* and weaker with *dfrA5* and *aadB*. We found the same overall differences between cassettes, with *blaVEB-1* globally more expressed than *dfrA5*. Using a similar statistical model (Table 2.10: Adjusted R-squared: 0.57, F-statistic: 10.55 on 7 and 43 DF, corrected p-value: 2.46e-07) we could again identify cassette position (corrected p-value<8.6e-03) and cassette identity (corrected p-value< 2.06e-06) as strong predictors of cassette transcript levels, and find a significant interaction between cassette position and the nature of the cassette (corrected p-value<0.01288) due to the weaker effect of position on *dfrA5* (estimate = -0.0001, std = 0.0035) and *aadB* (estimate =-0.006, std = 0.0037) transcripts.

Though harder to see at the resistance level, we could confirm and properly quantify the negative correlation between cassette transcription and distance to the promoter. This effect holds true between species and cassettes, but can heavily vary in intensity. The stark difference between global cassette transcripts, higher levels of transcripts of *blaVEB-1* than *dfrA5* - even with *blaVEB-1* in last and *dfrA5* in first position within the same array - was however unexpected, given cassettes in the same array are theoretically expressed through the same promoter, which is located at the start of the array. Moreover differences at the transcriptional levels do not seem to be able to explain the strong boost in resistance given by the first position in *aadB* in *Pseudomonas* (the difference in *aadB* transcripts between 1st and 2nd position is similar to the difference between 2nd and 3rd), nor the effect of *dfrA5* on downstream *blaVEB-1* cassettes (no increase in *blaVEB-1* transcripts is visible in cassettes located after *dfrA5*). In order to gather more insights into these effects, we used *in-silico* tools to identify potential factors affecting cassette expression.

2.4.3 *In-silico* array analysis

We first tried to identify potential promoters within the cassettes. While integron cassettes are supposed to be promoterless (Collis and Hall, 1995), studies have found cassettes containing their own promoters (da Fonseca and Vicente, 2012) which can even influence the expression of downstream cassettes (Tansirichaiya *et al.*, 2019). We used BPROM (Solovyev and Salamov, 2011) to identify putative promoters within the cassettes and their localisation relative to the RT-qPCR primers (**Table** 2.14, **Figure** 2.4 Supplementary material). BPROM returned several high-scored putative promoters on both strands, notably in the *blaVEB-1* cassette. All of these promoters are located within coding regions of the cassettes, whose transcripts would not lead to expression of the entire cassette. Those partial transcripts however could help explain the higher copies of *blaVEB-1* transcripts seen by RT-qPCR, given they are predominantly located within this cassette. These promoters could also impact the transcription of downstream (or even upstream) cassettes but, as seen previously, cassette transcript levels do not seem to be strongly impacted by

Cassette	Shine-Dalgarno sequence	Distance to the start codon
blaVEB-1	AGGAG	9bp
aadB	AGG	4bp
dfrA5	GGA	9bp

Table 2.2: Cassettes SD sequences
Translation initiation is maximal when the SD sequence is closest to the AGGAGGU consensus and the spacing between the SD sequence and the start codon is closest to 5bp (Salis *et al.*, 2009). The SD sequences for *dfrA5* and *blaVEB-1* are in agreement with (Hanau-Berçot *et al.*, 2002) but they determined *aadB* to be lacking an SD box.

the identity of the previous cassette. Experimental evidence would be needed to confirm the existence of those putative promoters as the accuracy of the BPROM algorithm is lower within coding regions.

Second, we looked into differences at the translational level between cassettes. It has been shown that modulation in cassette expression can happen at the translational level (Hanau-Berçot *et al.*, 2002) (Jacquier *et al.*, 2009), especially in cassettes missing Shine-Dalgarno sequences (Hanau-Berçot *et al.*, 2002). Shine-Dalgarno sequences are complementary to the 3' end of the 16S rRNA and allow ribosome recruitment and the start of translation. In *E. coli* the SD consensus sequence is AGGAGGU and ideally located between 2bp and 15bp before the start codon (Schmidt, 1984). We identified an SD sequence with a strong consensus at the start of *blaVEB-1* but only minimal SD sites at the start of *dfrA5* and *aadB* (Table 2.2). It has been shown that a short open reading frame contained within the *attI* site can substantially enhance the translation of an SD-deficient cassette when located in first position (Hanau-Berçot *et al.*, 2002). This short ORF is also present in our arrays and could explain the increase in resistance observed when *aadB* is in first position.

We then used the RBS Calculator program (Salis *et al.*, 2009) to predict translation initiation rate for the cassettes start codons in each array mRNA (Table 2.15). RBS Calculator uses a thermodynamic model to quantify the strength

of the molecular interaction between the mRNA and the 30S ribosome complex and takes into account the impact of RNA secondary structures. As expected, we see a striking difference between the predicted translation rate of the *blaVEB-1* cassette on one side and the *aadB* and *dfrA5* cassette on the other side, due to the difference in ΔG_{mRNA_rRNA}, which corresponds to the strength of the mRNA/rRNA complex and is directly linked to the Shine Dalgarno sequence (Espah Borujeni *et al.*, 2017). Moreover we can also detect an impact of *attC* sites on the translation of downstream cassettes. Especially, we can see an increase in the predicted translation initiation rate of the *blaVEB-1* cassette in array A1 and A6, where *blaVEB-1* is located after *dfrA5*, mainly due to differences in $\Delta G_{standby}$ which represents the energy needed to unfold secondary structures limiting the access to the ribosome standby-site upstream of the SD sequence (Espah Borujeni *et al.*, 2017). Care should be taken with the results of these simulations as RBS Calculator assumes that the mRNA returns to an equilibrium after translation initiation (Salis, 2011), which may not happen if long hairpins or complex structures are present, which is the case in our system with its *attC* sites hairpins. However this result is in line with the impact of *attC* sites on the translation of a downstream cassette experimentally proven by (Jacquier *et al.*, 2009), even if the proposed mechanism differs, and provides an explanation to the increase in *blaVEB-1* expression observed after *dfrA5*.

2.4.4 Differences in integrase induction

We then looked at the induction of the integrase by different antibiotics, in both *E. coli* and *P. aeruginosa*. We measured the differences in integrase transcripts after incubation with ciprofloxacin, gentamicin and piperacillin at half the MIC concentration (Figure 2.3 A), which should provide close to maximal integrase expression, as the SOS response has been shown to be maximally induced by ciprofloxacin at around 2/3 of the MIC (Torres-Barceló *et al.*, 2015). In both species the level of integrase transcription when the integrase is repressed is around 10% of the transcription of the first cassette, showing the repression by LexA is not total. The maximal increase in integrase transcription is between 3- and 4- fold

in both *E. coli* and *P. aeruginosa* but varies depending on the species/antibiotic combination. In *E. coli* all three antibiotics induced a similar but weak level of integrase transcription, which was not significantly different from the no antibiotic control (p= 0.25, p=0.24 and p =0.31 for ciprofloxacin, gentamicin and piperacillin respectively, 1-tailed Dunnett test, Table 2.13). In *P. aeruginosa* however, we detected a much stronger increase in integrase transcription relatively to the non-antibiotic sample with gentamicin, and only a 50% increase after treatment with ciprofloxacin, while no induction by piperacillin was visible. We then measured the level of *LexA* transcripts to confirm this increase in integrase transcripts is effectively linked to the SOS response, as *LexA* transcription is auto-regulated (Brent, 1982) (Figure 2.3 B). Only the induction by gentamicin provided transcript levels significantly different from the un-induced control (Table 2.11 and Table 2.12).

2.5 Discussion

Integrons have been hypothesised to provide 'adaptation on demand' (Escudero *et al.*, 2015) for bacteria by linking a modulation of cassette expression levels based on cassette position to a recombination mechanism regulated by the SOS response, allowing bacteria to quickly adapt to changing environments by reshuffling the cassette array, therefore changing its expression levels. The potential benefits of integron shuffling will be heavily dependent on the differential in cassette expression between the different positions of the array. A strong difference in cassette expression can allow for the creation of a low-cost but easily mobilisable reservoir of genes in distant cassettes. Too small a difference, however, and the shuffling benefits may not be able to compensate for the hefty cost of integrase expression (Lacotte *et al.*, 2017). While the effect of cassette position has been known for a long time (Collis and Hall, 1995), the mechanisms behind it remain controversial (Escudero *et al.*, 2015).

In order to investigate the effect of position on cassette expression in mobile integrons we created a 3 cassette integron model system, made of six arrays covering

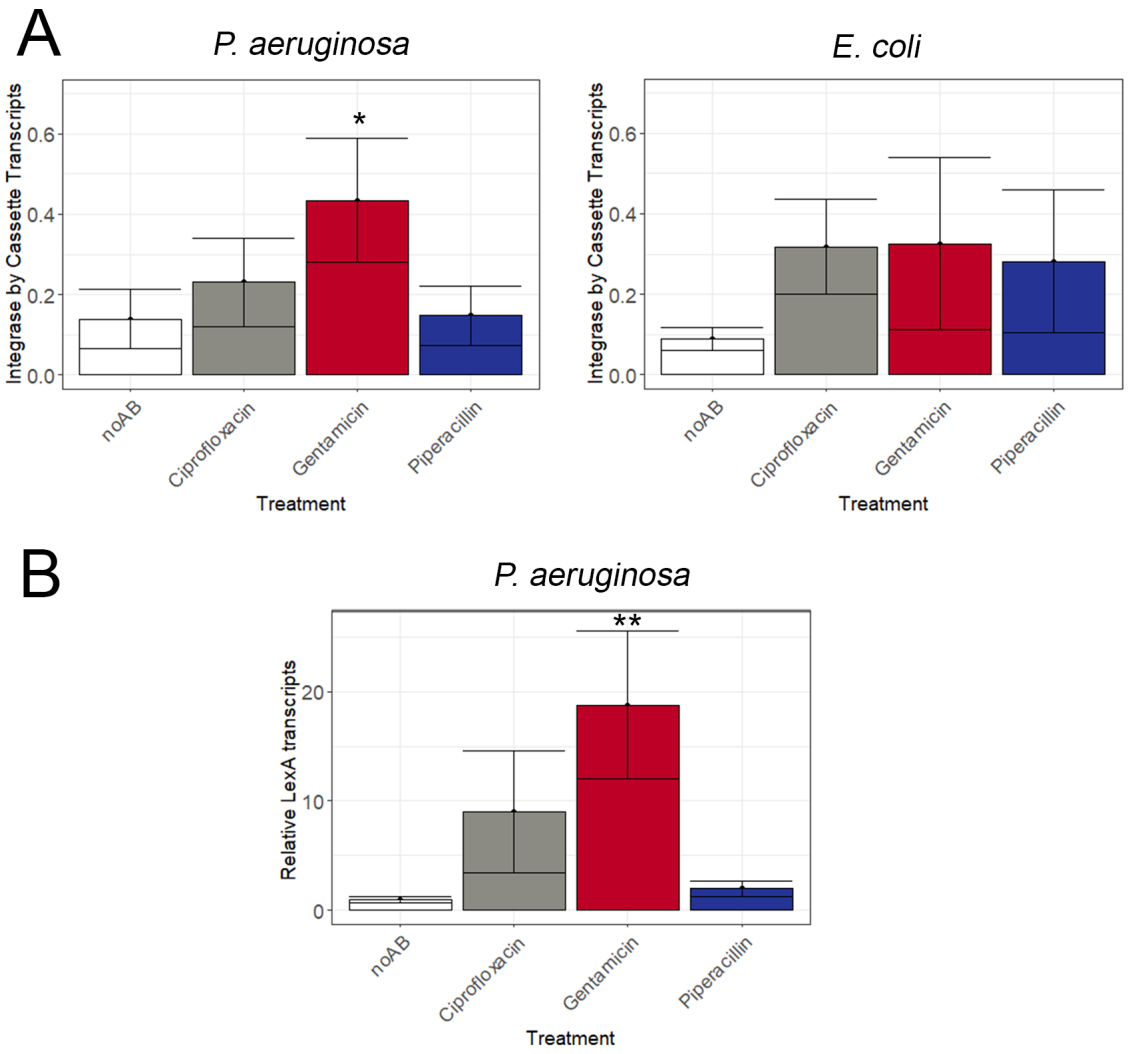


Figure 2.3: **A** Integrase transcript levels, normalized by the level of cassette transcripts in both *P. aeruginosa* and *E. coli*. **B** Relative levels of LexA transcripts in comparison to the treatment without antibiotic. Error bars represent standard errors. ** $p < 0.01$, * $p < 0.05$ - from 1-tailed Dunnett test (Supplementary material 2.6.2)

all possible cassette configurations. We measured the resistances and transcription levels for each of these arrays in both *E. coli* and *P. aeruginosa*, as well as the induction of the integrase by several antibiotics.

We detected differing effects of position at the phenotypic level, which were heavily dependent on the cassette and on the antibiotic considered. In *P. aeruginosa* the effect of position was either dramatic (gentamicin resistance, with a 10-fold decrease in resistance between first and second position), weak (piperacillin and kanamycin resistance, with only a 50% decrease in resistance), or invisible under

reasonable concentrations of antibiotics, due to high levels of resistance (trimethoprim, ceftazidime and ticarcillin). Different effects of position were visible in *E. coli*, with only a weak effect of position on gentamicin resistance and surprising higher levels of resistance against β-lactams when the *blaVEB-1* cassette was following the *dfrA5* cassette, independently of position.

In (Collis and Hall, 1995), the first paper showing the impact of cassette position in integrons, the decrease in IC_{50} in gentamicin and streptomycin resistance was between 2- and 10-fold from one position to the next with the *aadA2* and *aacA4* cassettes. Apart from the unique case of *aadB* and gentamicin resistance in *P. aeruginosa* the effect of position we detected is much weaker in comparison, showing that the effect of position may not be the same for all cassettes. These two systems have different Pc variants, strong for (Collis and Hall, 1995) and weak in our experiment, and may play a role in those differences.

The explanation for the decrease in cassette expression along the array proposed by (Collis and Hall, 1995) involved *attC* sites acting as Rho-independent transcription terminators due to their hairpin shape. To investigate the effect of position on transcription, we measured cassette transcription by RT-qPCR and saw a decrease in transcript levels the further away cassettes were from Pc promoter, both in *E. coli* and in *P. aeruginosa*, but with varying intensity depending on the cassettes. Surprisingly, this effect was only visible when comparing the transcription of the same cassette between different arrays, and not when comparing the transcription of different cassettes within the same array, as *blaVEB-1* was consistently over-expressed and *dfrA5* under-expressed in comparison to *aadB*, independently of position - which may be due to the presence of internal promoters or differences in primers efficiencies we did not manage to capture. This decrease in transcript levels alongside position is in line with previous results: Northern blots from (Collis and Hall, 1995),(Coyne *et al.*, 2010) where micro-arrays were used to look at the expression along a 5-cassette array in *Acinetobacter baumannii*, and (Barraud and Ploy, 2015) which looked at the transcription of *catB9* within arrays of different

sizes. On the other side (Jacquier *et al.*, 2009) showed no difference in transcription between the bla_{IMP-8} cassette in first position and *aac(6')-Ib*$_7$ in second position, and therefore ruled out a transcriptional basis for the effect of position, but may have been misled by comparing different cassettes within the same array, where the effect of position was not visible in our study. The exact mechanism involved in the attenuation of transcription along the array remains however unclear: Rho-independent transcription terminators consist of a hairpin structure followed by a T-rich region (Wilson and von Hippel, 1995), which are not commonly present after *attC* sites, and (Jacquier *et al.*, 2009) saw no difference in cassette transcription after partially deleting the upstream *attC* site and destabilising the hairpin.

As transcriptional effects did not explain the impact of first position in gentamicin resistance, nor the effect of *dfrA5* on the following *blaVEB-1* cassette, we then investigated cassette translation using an *in-silico* approach and found a translational explanation for both those phenomena.

We first identified differences in the quality of the Shine-Dalgarno sequence of the cassettes, which were confirmed after using RBS Calculator to predict the translation initiation rate of each cassette. The poor translation initiation region of *aadB* may explain why its expression levels are heavily dependent on position. (Hanau-Berçot *et al.*, 2002) showed the translation of a SD-deficient cassette relies on a short open reading frame overlapping the *attI* site, and can explain why the expression of *aadB* is strongly enhanced when *aadB* is in first position.

We also predicted different translation rates depending on the upstream *attC* site due to differences in the local RNA secondary structure, and showed an increased translation rate when *blaVEB-1* was located after *dfrA5*. (Jacquier *et al.*, 2009) showed previously how different *attC* sites led to different translation rates of the downstream cassette and how the presence of a small ORF in the *attC* site could enhance the expression of the subsequent cassette through translational coupling. We observed a similar translational enhancement with *dfrA5* but from a different origin as no ORF is identifiable in the cassette *attC* site. Interestingly, a short ORF

is present in *blaVEB-1 attC* site, without visible effect on the expression of the downstream cassettes. According to (Jacquier *et al.*, 2009) the hairpins of the *attC* sites also regulate cassette expression by impeding ribosome progression, therefore limiting translational coupling between the two cassettes, since partial deletion of the *attC* site increases the translation of the downstream cassette (Jacquier *et al.*, 2009). Our simulations provide another explanation for this effect: depending on their structure and closeness to the start codon of the subsequent cassette, *attC* sites can favor or disfavor the binding of the ribosome upstream of the Shine-Dalgarno sequence, modulating the expression of the cassette downstream. This explanation removes the need for translation coupling between the cassettes to explain the impact of the *attC* site on translation, which does not seem supported by our data: if translation coupling played an important role in cassette expression we would expect a cassette to be more strongly expressed when it is located after a well-translated cassette than after a poorly translated one, but there is no visible increase in *aadB* expression when it is located after *blaVEB-1*, a cassette with a strong SD sequence. However translation coupling has been shown to be much stronger in the presence of SD sequence (Spanjaard and van Duin, 1989), and the poor SD sequence in *aadB* could limit its impact.

Why these translational effects impact *E. coli* and *P. aeruginosa* differently remains to be determined. These two species have a common 16S rRNA 3'end and we would not expect differences in translation initiation rate. However, how differences in cassette expression correlates with differences in resistances levels may differ between species, based on interactions with the cell physiology or other resistance mechanisms, such as efflux pumps.

We also investigated the induction of the integrase by different antibiotics at high concentration in *E. coli* and *P. aeruginosa*. The basal level of integrase expression was surprisingly high at around 10% the number of transcripts of the first cassette in both species and the range of integrase induction we observed

after antibiotic treatment, up to 4-fold, was similar to previous studies (Hocquet *et al.*, 2012)(Guerin *et al.*, 2009).

The levels of induction of the SOS response by gentamicin were surprising. In *P. aeruginosa* the higher level of SOS induction by gentamicin than by ciprofloxacin, an antibiotic targeting DNA and known SOS inducer, was unexpected and had not been reported before. Previous studies have shown SOS induction by aminoglycoside in *Vibrio cholerae* but also *Klebsiella pneumoniae* and *Photorhabdus luminescens*, through the formation of reactive oxygen species (Baharoglu *et al.*, 2013) and translation stalling which produces single strand DNA R-loops (Baharoglu and Mazel, 2014b). This however may not represent the levels of integrase present in the cells, as aminoglycosides impact translation and could therefore limit the production of the integrase enzyme. In *E. coli*, integrase induction by gentamicin, if confirmed by further experiments, is also surprising, as aminoglycosides have been shown to not induce the SOS response in this species (Baharoglu and Mazel, 2011b)(Baharoglu and Mazel, 2014a) due to high levels of *RpoS*, which protects against oxydative stress (Baharoglu *et al.*, 2013). An important difference is the antibiotic concentration used between our studies: (Baharoglu and Mazel, 2014a) studies induction at 1/100 the MIC, while we used much higher concentration where the oxidative stress response may become overwhelmed, even in *E.coli*. Finally, we were surprised not to see induction of the SOS response by piperacillin in *P. aeruginosa*. In *E. coli* a well characterized pathway links cell wall stress to the SOS response (Miller *et al.*, 2004b), but our observations are in line with (Hocquet *et al.*, 2012), where ceftazidime did not induce the SOS response in *P. aeruginosa*.

With three cassettes, our system only aims to represent the effect of position in mobile integrons. Chromosomal integrons, such as the integrons found in the *Vibrio* genus, can contain more than 200 cassettes (Escudero *et al.*, 2015), with complex mechanisms regulating cassette expression in response to environmental cues(Michael and Labbate, 2010). As mobile integrons can contain up to 8 cassettes (Ploy *et al.* (1998), Naas *et al.* (2001)) the effect of position may be stronger in longer arrays, but

analysis of a recently created integron dataset (Zhang *et al.*, 2018) showed that more than 70% of functional integrons present on plasmid have three cassettes or less.

A limitation of our study is also the use of MICs to measure cassette expression: MICs are notoriously crude measurements which may struggle to capture subtle differences between arrays. Measuring the fitness of the different arrays in the presence of antibiotics would give a finer picture of the array's resistance profile. Using resistance levels as a proxy for cassette expression levels also has its pitfalls, as the relationship between cassette expression and resistance levels may not be linear and depends on the resistance mechanism. Resistance levels from an array may also be impacted by cross-resistance or collateral sensitivity between cassettes, or interactions with other resistance mechanisms such a efflux pumps, and may explain the differences in resistances levels seen between bacteria species since they express the same arrays. Measuring directly the levels of a cassette's protein products would allow us to overcome this limitation, and confirm the results of our simulations.

In conclusion, our study shows a transcriptional origin for the gradient in cassette expression seen along integron arrays and highlights the importance of translational effects, which can either strengthen or reduce the impact of position. The variability of the effect of position between cassettes may have far-reaching consequences in terms of arrays composition: cassettes without SD sequences require being in first position for strong expression within the same array, and could therefore not be co-expressed simultaneously. However, this also creates a stronger gradient of expression and potentially allows the creation of a low-cost reservoir of genes, which would not be possible based only on the transcriptional effect. On the other hand, cassettes where position matters less assure constant expression - but also constant cost. The impact of a cassette and its *attC* site on the downstream cassettes may also influence the prevalence of a cassette: by promoting the expression of the subsequent resistance cassette the upstream cassette can be co-selected. This may be the case with the *dfrA5* cassettes, as prevalence of *dfr* genes has been reported unchanged in clinical isolates despite a 85% reduction in trimethoprim prescription

(Brolund *et al.*, 2010). On the opposite side, cassettes reducing the expression of a downstream cassette may be at a higher risk of being excised, if stronger expression of the downstream cassette is advantageous. Finally, the difference in optimal array order between species may also give an additional benefit to the SOS induction and integrase activation happening during conjugation (Baharoglu *et al.*, 2010), as it allows cassette reshuffling and selection of the best array order for its new environment, but also its new host.

2.6 Acknowledgements

We thank Laurent Poirel and Marie-Cecile Ploy for the gift of the strains containing the cassettes used in the construction of the integrons arrays.

2.7 Supplementary material

2.7.1 MIC values

	1st position		**2nd position**		**3rd position**		
Aminoglycosides	A5	A6	A2	A4	A1	A3	Reference
P. aeruginosa							
Gentamicin	341.3	277.3	12.0	44.0	22.0	8.5	3.0
Kanamicin	51.7	53.9	46.9	33.6	37.5	37.8	41.9
E. coli							
Gentamicin	2.0	2.5	1.5	1.5	1.5	1.5	1.3
Kanamicin	3.0	3.0	1.5	1.5	1.0	2.0	1.5
Tobramicin	1.5	1.8	0.9	TBD	1.2	0.8	1.4

Beta-lactams	A2	A3	A1	A5	A4	A6	Reference
P. aeruginosa							
Ampicillin	1500.0	1524.0	1682.7	1682.7	1524.0	1524.0	591.3
Piperacillin	78.9	83.8	68.2	83.8	74.6	52.6	23.4
Ceftazidime	TBD	>1024	>1024	>1024	>1024	>1024	100.0
Ticarcillin	TBD	TBD	>1024	>1024	>1024	>1024	48.0
E. coli							
Ampicillin	100	100	200	64	64	128	<0.25
Ceftazidime	256	192	512	144	TBD	320	128
Piperacillin	8	8	12	6	8	12	4

Trimethoprim	A1	A4	A3	A6	A2	A5	Reference
P. aeruginosa							
	>1500	>1500	>1500	>1500	TBD	1500	192
E. coli							
	>2000	TBD	>2000	>2000	>2000	>2000	64<

Table 2.3: MICs of the different arrays in *P. aeruginosa* and *E. coli*. Arrays are ordered by the position of the relevant resistance cassette. *TBD* = To Be Determined (missing measurement)

2.7.2 ANOVA & Coefficients estimates

Significance codes: '***' < 0.001, '**' < 0.01, '*' < 0.05, '.' < 0.1

Resistances levels

P. aeruginosa

Aminoglycosides

Model: $log_2(MIC) \sim Antibiotic + Cassette\ Position + Antibiotic * Cassette\ Position$

Multiple R-squared: 0.626, Adjusted R-squared: 0.5973

F-statistic: 21.76 on 3 and 39 DF, p-value: 1.906e-08, Bonferroni corrected p-value: 9.53e-08

Analysis of variance table								Coefficients						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Corrected			Estimate	Std. Error	t- value	Pr(>|t|)	Corrected	
Antibiotic	1	0.316	0.316	0.3278	0.5702	2.851		(Intercept)	5.8291	0.5283	11.034	1.47E-13		***
Cassette Position	1	36.024	36.024	37.4015	3.58E-07	1.79E-06	***	AntibioticGentamicin	3.7648	0.7969	4.724	2.98E-05	5.96e-05	***
Interaction	1	26.548	26.548	27.5626	5.67E-06	2.83E-05	***	Cassette Position	-0.2441	0.2383	-1.024	0.312		
Residuals	39	37.564	0.963					Interaction	-1.868	0.3558	-5.25	5.67E-06		***

Table 2.4: Linear model of the impact of position on aminoglycoside resistance levels in *P. aeruginosa*

Beta-lactams

Model: $log_2(MIC) \sim Cassette\ Position$

Multiple R-squared: 0.1422, Adjusted R-squared: 0.08859

F-statistic: 2.652 on 1 and 16 DF, p-value: 0.1229, Bonferroni corrected p-value: 0.6145

Analysis of variance table						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Corrected
Cassette Position	1	0.4056	0.4056	2.6525	0.1229	0.6145
Residuals	16	2.4466	0.15291			

Table 2.5: Linear model of the impact of position on beta-lactam resistance levels in *P. aeruginosa*

E. coli

Aminoglycosides

Model: $log_2(MIC) \sim Antibiotic + Cassette\ Position$

Multiple R-squared: 0.6207, Adjusted R-squared: 0.5665

F-statistic: 11.45 on 4 and 28 DF, p-value: 1.236e-05, Bonferroni corrected p-value: 6.18e-05

Analysis of variance table								Coefficients						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Corrected			Estimate	Std. Error	t- value	Pr(>|t|)	Corrected	
Antibiotic	3	4.899	1.633	9.7347	1.447e-04	7.235e-04	***	(Intercept)	1.55968	0.23442	6.653	3.21E-07		***
Cassette Position	1	2.7872	2.78717	16.615	3.425e-04	1.712e-03	**	AntibioticAmikacin	-1.05599	0.24801	-4.258	2.1e-04	8.4e-04	***
Residuals	28	4.697	0.16775					AntibioticGentamicin	-0.09141	0.22817	-0.401	0.6917	2.767	
								AntibioticTobramicin	-0.66732	0.19784	-3.373	0.00219	0.00876	**
								Cassette Position	-0.33485	0.08215	-4.076	0.000342		***

Table 2.6: Linear model of the impact of position on aminoglycoside resistance levels in *E. coli*

Beta-lactams

Model: $log_2(MIC) \sim Antibiotic + Cassette\ Position$

Multiple R-squared: 0.9021, Adjusted R-squared: 0.8838

F-statistic: 49.17 on 3 and 16 DF, p-value: 2.682e-08, Bonferroni corrected p-value: 1.341e-07

Analysis of variance table								Coefficients						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Corrected			Estimate	Std. Error	t- value	Pr(>|t|)	Corrected	
Antibiotic	2	77.204	38.602	73.6669	8.48E-09	4.24e-08	***	(Intercept)	6.48928	0.50013	12.975	6.58E-10		***
Cassette Position	1	0.089	0.089	0.1692	0.6863	3.4315		AntibioticCeftazidime	1.12635	0.39175	2.875	0.011	0.033	*
Residuals	16	8.384	0.524					AntibioticPiperacillin	-3.52945	0.41793	-8.445	2.73E-07	8.19e-07	***
								Cassette Position	0.08299	0.20174	0.411	0.686		

Table 2.7: Linear model of the impact of position on beta-lactam resistance levels in *E. coli*

Model: $log_2(MIC) \sim Antibiotic + Previous_Cassette$

Multiple R-squared: 0.9496, Adjusted R-squared: 0.9362

F-statistic: 70.68 on 4 and 15 DF, p-value: 1.502e-09, Bonferroni corrected p-value: 7.51e-09

Analysis of variance table								Coefficients						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Corrected			Estimate	Std. Error	t- value	Pr(>|t|)	Corrected	
Antibiotic	2	77.204	38.602	134.1421	2.69E-10	1.345e-09	***	(Intercept)	6.6	0.276	23.911	2.34E-13		***
previous_cassette	2	4.156	2.078	7.2215	0.006358	0.03179	*	AntibioticCeftazidime	1.049	0.2905	3.61	0.00257	0.00771	**
Residuals	15	4.317	0.288					AntibioticPiperacillin	-3.5294	0.3097	-11.396	8.73E-09	2.619e-08	***
								previous_cassetteaadB	-0.4806	0.2994	-1.606	0.12921	0.25842	
								previous_cassettedfrA	0.6466	0.2867	2.255	0.03951	0.0792	.

Table 2.8: Linear model of the impact of the previous cassette on beta-lactam resistance levels in *E. coli*

Transcription levels

In the following results all contrasts use *aadB* as the reference.

P. aeruginosa

Model: *Cassette transcript /Plasmid copynumber* $\sim$ *Extraction* + *Gene* + *Cassette Position* + *Gene* $*$ *Cassette Position.*

Multiple R-squared: 0.6795, Adjusted R-squared: 0.6189

F-statistic: 11.21 on 7 and 37 DF, p-value: 1.614e-07, Bonferroni corrected p-value: 3.228e-07

Analysis of Variance Table							
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Corrected	
Extraction	2	1.18973	0.59486	11.0546	0.0001722		***
Gene	2	2.37175	1.18588	22.0376	4.98E-07	9.96e-07	***
Position_cassette	1	0.41448	0.41448	7.7025	0.0085	0.0171	*
Gene:Position_cassette	2	0.24574	0.12287	2.2833	0.1161	0.2322	
Residuals	37	1.99102	0.05381				

Coefficients						
	Estimate	Std. Error	t value	Pr(>|t|)	Corrected	
(Intercept)	0.73698	0.15465	4.766	2.90E-05		***
Extraction2	0.35759	0.0847	4.222	0.000151		***
Extraction3	0.0269	0.0847	0.318	0.752602		
GeneqdfrA1	-0.70898	0.21417	-3.31	0.002086	0.006258	**
GeneqblaVEB1	-0.0255	0.23689	-0.108	0.914867	2.744601	
Position_cassette	-0.22672	0.06696	-3.386	0.001694		**
GeneqdfrA1:Position_cassette	0.20894	0.10436	2.002	0.052643	.	
GeneqblaVEB1:Position_cassette	0.15709	0.10436	1.505	0.140746		

Table 2.9

E. coli

Model: *Cassette transcript /Plasmid copynumber* $\sim$ *Extraction* + *Gene* + *Cassette Position* + *Gene* $*$ *Cassette Position*.

Multiple R-squared: 0.632, Adjusted R-squared: 0.5721

F-statistic: 10.55 on 7 and 43 DF, p-value: 1.232e-07, Bonferroni corrected p-value: 2.464e-07

Analysis of variance table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Corrected	
Extraction	2	0.0013643	0.00068214	4.6289	0.0151115	0.030223	*
Gene	2	0.0056958	0.00284788	19.3254	1.03e-06	2.06e-06	***
Position_cassette	1	0.002145	0.002145	14.5557	4.304e-03	8.608e-03	***
Gene:Position_cassette	2	0.0016759	0.00083794	5.6862	0.00644	0.01288	*
Residuals	43	0.0063367	0.00014736				

Coefficients

	Estimate	Std. Error	t value	Pr(>|t|)	Corrected	
(Intercept)	0.030556	0.00838	3.646	0.000713		***
Extraction2	-0.012119	0.004046	-2.995	0.004539		**
Extraction3	-0.008858	0.004258	-2.08	0.043507		*
GeneqblaVEB1	0.040142	0.01123	3.575	0.002646	0.007938	**
GeneqdfrA1	-0.017564	0.011126	-1.579	0.36531	1.09593	
Position_cassette	-0.005803	0.003678	-1.578	0.121917		
GeneqblaVEB1:Position_cassette	-0.011947	0.005213	-2.292	0.026867		*
GeneqdfrA1:Position_cassette	0.004813	0.00508	0.947	0.348679		

Table 2.10

Integrase expression

P. aeruginosa

Model: *Integrase transcript /Cassette transcript* $\sim$ *Extraction* + *Antibiotic* .

Multiple R-squared: 0.7723, Adjusted R-squared: 0.5826

F-statistic: 4.07 on 5 and 6 DF, p-value: 0.05856

Analysis of Variance Table						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Extraction	2	0.17916	0.089582	5.2305	0.04843	*
Antibiotic	3	0.16939	0.056462	3.2967	0.09958	.
Residuals	6	0.10276	0.017127			

Multiple Comparisons of Means: Dunnett Contrasts						
	Estimate	Std. Error	t value	Pr(>|t|) - double tailed	Pr(>|t|) - single tailed	
Ciprofloxacin - noAB == 0	0.092038	0.106855	0.861	0.7336	0.3668	
Gentamicin - noAB == 0	0.295199	0.106855	2.763	0.0767	0.03835	*
Piperacillin - noAB == 0	0.008899	0.106855	0.083	0.9995	0.49975	

Table 2.11

Model: *LexA transcripts* $\sim$ *Extraction* + *Antibiotic* .

Multiple R-squared: 0.8405, Adjusted R-squared: 0.7077

F-statistic: 6.326 on 5 and 6 DF, p-value: 0.02199

Analysis of Variance Table						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Extraction	2	3.4961	1.748	3.541	0.09648	.
Antibiotic	3	12.1176	4.0392	8.1823	0.0153	*
Residuals	6	2.9619	0.4937			

Multiple Comparisons of Means: Dunnett Contrasts						
	Estimate	Std. Error	t value	Pr(>|t|) - double tailed	Pr(>|t|) - single tailed	
Ciprofloxacin - noAB == 0	1.1331	0.5737	1.975	0.2101	0.10505	
Gentamicin - noAB == 0	2.5123	0.5737	4.379	0.0114	0.0057	**
Piperacillin - noAB == 0	0.1361	0.5737	0.237	0.9899	0.49495	

Table 2.12

E. coli

Model: *Integrase transcript /Cassette transcript* $\sim$ *Extraction* $+$ *Antibiotic* .

Multiple R-squared: 0.539, Adjusted R-squared: 0.1548

F-statistic: 1.403 on 5 and 6 DF, p-value: 0.3426

Analysis of variance table					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Extraction	2	0.24506	0.122532	2.4013	0.1713
Antibiotic	3	0.11288	0.037626	0.7374	0.567
Residuals	6	0.30616	0.051027		

Multiple Comparisons of Means: Dunnett Contrasts					
	Estimate	Std. Error	t value	Pr(>|t|) - double tailed	Pr(>|t|) - single tailed
Ciprofloxacin - noAB == 0	0.2303	0.1844	1.249	0.505	0.2525
Gentamicin - noAB == 0	0.2381	0.1844	1.291	0.482	0.241
Piperacillin - noAB == 0	0.1931	0.1844	1.047	0.621	0.3105

Table 2.13

2.7.3 BPROM & RBS Calculator simulations

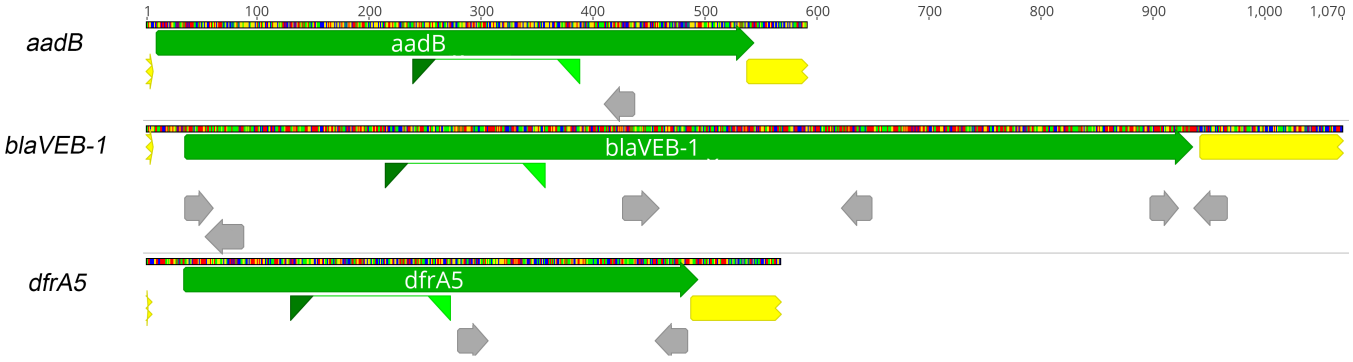


Figure 2.4: Promoters detected (grey arrows) by BPROM within each cassette and qPCR primers localisation (green chevrons). Cassettes ORFs are represented by the green arrows while *attC* sites are represented in yellow. Figure created using Geneious version 10.2 created by Biomatters

Gene	Strand	Position within the cassette	-10 box	-35 box	Score		
					-10 box	-35 box	Overall score[1]
Int1	+	930 (PcW)	TCGTAAGCT	TGGACA	47	29	3.1
	-	1029 (Pint)	GCATAGACT	TTGCTG	38	47	4.06
blaVEB-1	+	68	GGATATTAT	ATGAAA	65	30	5.69
	+	466	TGCTAAAAT	GCGACA	81	-4	4.38
	+	930	AATTACTAT	TTGCAA	44	55	6.83
	-	46	TAATAATAT	TTGTAA	47	47	6.83
	-	614	TCTTATTAT	TTTTTA	65	26	6.34
	-	929	CGGTAGTTT	TTGCCT	39	56	5.7
dfrA5	+	313	CGTTATAGT	TGGCTG	65	10	2.24
	-	445	AGTTAATGT	TTGCCA	43	61	4.61
aadB	-	402	CGGCATAGT	TCCACT	40	4	2.8

Table 2.14: Putative promoters detected by BPROM
[1] Overall score = Linear discriminant function (LDF). Can be approximated by $log(< likelihood\ of\ a\ site\ being\ a\ promoter > / < likelihood\ of\ a\ site\ not\ being\ a\ promoter >)$

Array	Cassette	Tir	ΔG_{total}	ΔG_{mRNA_rRNA}	$\Delta G_{spacing}$	$\Delta G_{stacking}$	$\Delta G_{standby}$	ΔG_{start}	ΔG_{mRNA}
A1	dfrA5	3951.6	-2.6	-22.4	0.3	0.0	0.0	-0.4	-20.5
	blaVEB-1	25848.1	-6.8	-25.3	0.7	0.0	0.0	-2.8	-20.6
	aadB	25.6	8.6	-20.1	1.5	0.0	1.5	-2.8	-28.3
A2	blaVEB-1	12412.0	-5.1	-20.6	0.7	0.0	3.1	-2.8	-14.5
	aadB	25.6	8.6	-20.1	1.5	0.0	1.5	-2.8	-28.3
	dfrA5	259.6	3.5	-33.9	0.3	0.0	1.7	-0.4	-36.3
A3	blaVEB-1	12412.0	-5.1	-20.6	0.7	0.0	3.1	-2.8	-14.5
	dfrA5	142.0	4.8	-20.5	0.3	0.0	2.6	-0.4	-23.3
	aadB	211.8	3.9	-20.2	1.5	0.0	0.0	-2.8	-25.3
A4	dfrA5	3951.6	-2.6	-22.4	0.3	0.0	0.0	-0.4	-20.5
	aadB	211.8	3.9	-20.2	1.5	0.0	0.0	-2.8	-25.3
	blaVEB-1	11394.8	-4.9	-32.2	0.7	0.0	1.8	-2.8	-27.6
A5	aadB	181.5	4.3	-21.0	1.5	0.0	2.0	-2.8	-24.4
	blaVEB-1	11394.8	-4.9	-32.2	0.7	0.0	1.8	-2.8	-27.6
	dfrA5	142.0	4.8	-20.5	0.3	0.0	2.6	-0.4	-23.3
A6	aadB	181.5	4.3	-21.0	1.5	0.0	2.0	-2.8	-24.4
	dfrA5	259.6	3.5	-33.9	0.3	0.0	1.7	-0.4	-36.3
	blaVEB-1	25848.1	-6.8	-25.3	0.7	0.0	0.0	-2.8	-20.6

Table 2.15: Translation initiation rates and Gibbs free energy predicted by RBS Calculator for each array mRNA.
Terms explanation, according to (Espah Borujeni *et al.*, 2017):
Tir = Translation Initiation Rate (au)
ΔG_{total}= Total Gibbs Free Energy Change ($\Delta G_{total} = \Delta G_{mRNA_rRNA} + \Delta G_{spacing} + \Delta G_{stacking} + \Delta G_{standby} + \Delta G_{start} - \Delta G_{mRNA}$)
ΔG_{mRNA_rRNA}= Gibbs Free Energy for mRNA:rRNA complex hybridization.
$\Delta G_{spacing}$ = Gibbs Free Energy Penalty for non-optimal Spacing
$\Delta G_{stacking}$ = Gibbs Free Energy change due to stacked nucleotides in spacer region
$\Delta G_{standby}$ = Gibbs Free Energy Penalty for unfolding any secondary structures sequestering the ribosome standby site
ΔG_{start}= Gibbs Free Energy Change when the tRNA binds to the start codon
ΔG_{mRNA} = Gibbs Free Energy Change for unfolding the mRNA secondary structure

3

Experimental evolution with *aadB*: Shuffling and cassette duplication

3.1 Abstract

Integrons have been hypothesized to provide 'adaptation on demand' for bacteria through the modulation of cassette expression levels via genetic reshuffling. The adaptive benefits of this reshuffling have been proven theoretically but have never been experimentally quantified in the context of antibiotic resistance evolution. In this chapter, we investigate the evolutionary advantage provided by the integrase against increasing concentrations of gentamicin. Using custom integron arrays containing a cassette encoding aminoglycoside resistance in distant position, we measure the evolvability benefits of the integrase by comparing strains with and without a functional integrase. We show through the combination of experimental evolution and whole genome sequencing that integrase activity can increase bacteria adaptability and induce extensive cassette rearrangements, including 'cut-and-paste' recombinations bringing the resistance cassette to first position. We identify extensive cassette duplications as a result of integrase activity, both in our custom integron and in a clinical plasmid, highlighting cassette duplication as an additional integrase mechanism to modulate cassette expression.

3.2 Introduction

Through the activity of the integrase enzyme, bacteria can integrate, excise and reorder gene cassettes within the integron. This cassette shuffling has been shown to happen both in the lab (Barraud and Ploy, 2015) and in clinical settings (Hocquet *et al.*, 2012) and is predicted to play an important part in the spread of integrons among clinically relevant bacteria (Escudero *et al.*, 2015). As integron cassettes expression usually depends on their position within the array, as seen previously, cassette shuffling modulates cassette expression and is therefore expected to help bacteria adapt faster to changing environments.

The molecular mechanisms of integron cassette shuffling have been extensively studied. The integrase, a tyrosine recombinase encoded by the *IntI*, recognises two types of integron specific recombination sites, called the *attI* and *attC* sites. The *attC* sites are palindromic structures located at the end of the cassettes. Highly diverse in length and sequence, they conserve a specific cruciform secondary structure, which is used as substrate by the integrase (MacDonald *et al.*, 2006). The *attI* site, located upstream of the *IntI* gene, diverges between types of integrases, with integrases preferentially recognizing their corresponding *attI* site. Reaction between two *attC* sites located within the same array leads to the excision of the cassettes in covalently closed circles (Collis and Hall, 1992a). These circular cassettes can then be inserted into first position through the *attI* × *attC* reaction (Hall *et al.*, 1991). The recognition of extrahelical bases present in the *attC* sites by the integrase ensures the integration of the promoterless cassettes in the correct orientation relative to the integrase, and therefore guarantees cassette expression through the Pc promoter located at the start of the array (Francia *et al.* (1999), Johansson *et al.* (2004), Bouvier *et al.* (2005). Cassette insertion may also happen in the middle of the array through *attC* × *attC* reactions, but these reactions are less favourable due to the need for both sites to be present in ssDNA conformation. While the integrase reacts preferentially with the integron *attI* and *attC* sites, it has also be shown to recombine with non-canonical secondary sites (Francia *et al.*

(1993),Recchia *et al.* (1994)), potentially creating genomic instability, which has been hypothesized to be the origin of the integrase fitness cost (Harms *et al.*, 2013).

While the molecular mechanisms of integron shuffling are known in details, few studies have experimentally quantified the evolutionary benefits of the integrase and the trade-off between integrase cost and adaptability advantage. Previous studies showed that integrase activity is quickly lost after integron introduction in *Acinetobacter baylyi*, a species lacking an SOS response (Starikova *et al.*, 2012), and that over-expression of the integrase sped up chloramphenicol resistance evolution (Barraud and Ploy, 2015), but none have investigated the benefits of integrase activity under its natural promoter and its associated LexA regulation.

Moreover, a deeper understanding of the landscape of rearrangements available from one array is required if we want to properly determine the benefits of cassette shuffling. As shown by (Engelstädter *et al.*, 2016) through their simulations of integron-bearing populations of bacteria, parameters such the re-insertion rate of excised cassettes can have a strong impact on the adaptation benefits of the integrase. (Barraud and Ploy, 2015) showed that excision of previous cassettes, instead of 'cut-and-paste' cassette re-insertion, is preferred to bring the *catB9* cassette into first position under chloramphenicol selection pressure, but more studies are needed to know if this effect is generalizable to all cassettes.

In this chapter we aim to quantify the evolvability benefits and the diversity of rearrangements produced by the reshuffling of the *aadB* cassette using the synthetic arrays introduced previously. *aadB* encodes for a clinically relevant aminoglycoside nucleotidyltransferase, which deactivates aminoglycoside antibiotics through the addition of an AMP group (Cox *et al.*, 2015). As seen in the previous chapter, *aadB* is strongly sensitive to cassette position due to its weak Shine-Dalgarno sequence and its expression level drops when separated from the open reading frame located within the *attI* site. We therefore expect integrase activity to be strongly beneficial as it can bring the *aadB* cassette closer to the start of the array.

Using experimental evolution combined with next generation sequencing, we investigated the effect of integrase activity on *P. aeruginosa* bacteria containing the *aadB* cassette in distant position from the promoter under increasing concentration of gentamicin. We find integrase activity can increase survival rates and gentamicin resistance evolution through extensive cassette duplications and rearrangements. Arising from an early time-point, these duplications happen through frequent insertion of the *aadB* cassette into first position. Moreover, we show using a clinical plasmid that these cassette duplications are not limited to the *aadB* cassette. Cassette duplication may therefore play an unexpected role in the integron's evolutionary success.

3.3 Material & Methods

Strains and growth conditions

The strains and plasmids used in this chapter are listed in **Table** 3.1. Strains were grown in LB Miller broth or agar supplemented with 100 mg/L of ceftazidime.

Strain/Plasmid	Description	Source
Strain		
PAO1	*P. aeruginosa* strain PAO1	
Plasmids		
WTA3	Custom integron array *intI1 - blaVEB-1 - dfrA5 - aadB* on R388 plasmid backbone	Previous chapter
ΔintA3	Array WTA3 with a 948bp deletion of the integrase intI1	This study
WTA4	Custom integron array *intI1 - dfrA5 - aadB - blaVEB1* on R388 plasmid backbone.	Previous chapter
ΔintA4	Array WTA4 with a 948bp deletion of the integrase intI1	This study
pAMBL1	Clinical plasmid containing a *blaVIM - aadB* array	(San Millan *et al.*, 2015b)
pAMBL2	Clinical plasmid containing a *blaVIM - aacA4 - blaVIM - blaVIM - aadA1* array	(San Millan *et al.*, 2015b)

Table 3.1: Strains used in this chapter

Δ*int* construction

Δ*int* mutants of the custom arrays previously described (Chapter 2.3) were created by introducing a 948 bp deletion of the integrase *IntI1* gene, deleting most of

the integrase open reading frame but conserving both the Pint and Pc promoters. These plasmids were then conjugated into *P. aeruginosa* as described previously in Chapter 2.

Experimental evolution

Determination of the minimum inhibitory concentration in experimental conditions

As antibiotics' minimum inhibitory concentrations vary depending on the size of the starting inoculum (Brook, 1989), we determined MICs in the conditions of the experimental evolution experiment. Overnight cultures inoculated from 2-3 morphologically similar colonies grown on selective agar were incubated for 20h with shaking in MH2 media with antibiotics. These overnight cultures were then diluted 1/10 000 and supplemented with doubling concentrations of gentamicin in three replicates. MICs were determined the next day after 20h of incubation. This process was repeated twice. In these conditions the MICs for PAO1:WTA3, PAO1:ΔintA3, PAO1:WTA4 and PAO1:ΔintA4 were identical at 24 mg/L.

Ramping experiment

At the start of the experiment 90 individual colonies grown on selective agar of each strain (PAO1:WTA3, PAO1:ΔintA3, PAO1:WTA4 and PAO1:ΔintA4) were inoculated in 200 μL of MH2 media supplemented with gentamicin at a concentration of 1/8 MIC. WT and Δint strains were placed in a chequerboard pattern by interlacing the different genotypes to control easily for cross-contamination. Wells at the edge of every plate were kept bacteria-free to avoid edge effects and identify contaminations. These 90 populations were passaged every day with a 1/10 000 dilution factor and the antibiotic concentration was doubled, starting at 1/8 MIC until reaching a concentration of 1024x MIC. Alongside these 90 populations per strain which were transferred in increasing antibiotic concentrations, 30 populations per strain were transferred as controls in constant conditions: 15 without antibiotic and 15 at a constant concentration of 1/8 MIC. Each population's OD_{600} was measured each day and a population was considered extinct when its OD fell

below 0.1 after 20h of incubation. All populations were frozen in 15% glycerol every two days.

DNA extraction

Liquid cultures were grown from the frozen stock of all surviving PAO1:WTA3 and PAO1:ΔintA3 populations at x1024 MIC in LB Miller media supplemented with gentamicin at x128 MIC, and were incubated for 24h with shaking. Six populations were inoculated from each control treatment in either LB Miller supplemented with a concentration of 1/8MIC or with no antibiotic. For the x4 MIC time point 26 populations of each PAO1:WTA3 and PAO1:ΔA3 genotype were regrown in x2 MIC concentration of gentamicin. Ancestral PAO1:WTA3, PAO1:ΔintA3, PAO1:WTA4, and PAO1:ΔintA4 populations were incubated in 100 mg/L of ceftazidime from the initial frozen stock. DNA extractions of the whole populations were carried out using the DNeasy Blood & Tissue Kit (Qiagen) on the QiaCube extraction platform (Qiagen) combined with RNAse treatment. DNA concentrations were determined using the Quantifluor dsDNA system (Promega).

PCR controls

At the x1024 and x4 MIC transfers all surviving populations were controlled for cross-contamination by verifying the size of the integrase by PCR (Primers given in Appendice Table 5.2). Starting materials were 2 μL of extracted DNA for the x1024 MIC time point and 2 μL of inoculate previously incubated for 24h then boiled for 10 minutes at 105°C for the x4 MIC time point populations. PCR reactions were carried out using the GoTaq G2 DNA mastermix (Promega) and the following protocol: 30s at 95°C, 30s at 55°C, 3 minutes at 72°C for 30 cycles. Plate misshandling during the transfers resulted in the contamination of 40 and 34 wells out of 180 ramping populations for each array. Areas of the plates where cross-contamination was detected in several wells in close proximity were excluded from the rest of the analysis, for a final population number of 65 per strain. The final populations at x1024 were also analysed by PCR to determine the position of

the *aadB* cassette relative to the start and the end of the array as well as identify any *aadB* duplications or inversions and deletions of the plasmid backbone.

Next Generation Sequencing

Library preparation and Next-Generation sequencing using the NovaSeq 6000 Sequencing System (Illumina) was carried out at the Oxford Genomics Centre at the Wellcome Centre for Human Genetics. 22 PAO1:WTA3 and 6 PAO1:ΔintA3 populations were sequenced from the x1024 MIC time point. For each control treatment, six populations were pooled together and sequenced as one. For the x4 MIC time point 26 PAO1:WTA3 and 26 PAO1:ΔintA3 were sequenced.

Bioinformatic pipeline

Quality control & artefact removal

PCR duplications and optical artifacts were removed using MarkDuplicates (Picard) (Institute, 2018) then low quality bases and adaptors were trimmed from the sequencing reads using Trimmomatic v0.39 (Bolger *et al.*, 2014). Finally overall read quality control was performed using FastQC (Andrews *et al.*, 2012) and multiQC (Ewels *et al.*, 2016). During this process one PAO1:WTA3 sample from the x1024 MIC time-point was removed due to the presence of non-*Pseudomonas* DNA.

Variant calling

Single-nucleotide polymorphisms (SNP), point insertion and deletion identification was performed using the *breseq* (Deatherage and Barrick, 2014) pipeline in polymorphism mode. For each population reads were mapped to the *P. aeruginosa* PAO1 complete genome NC_ 002516.2 and the predicted sequence of WTA3. The pipeline output was then further processed in MATLAB (MathWorks). Variants present in the un-evolved ancestor populations at any frequency were filtered out. We also excluded variants reaching a frequency of less than 30% within a population. A 5% threshold was applied to the pooled controls, which allows the detection of any variant present in more than 30% of a single population. Final results were exported in table format and processed for visualisation using Circos (Krzywinski

et al., 2009) and Geneious (Biomatters). Apart from the expected *IntI* deletion, the PAO1:ΔintA3 genome was shown to differ from PAO1:WTA3 and PAO1 by two SNPs (Table 3.2), likely to have arisen during the conjugation process. *PA3734* is predicted to be a lipase involved in cell-wall metabolism (Dettman *et al.*, 2015) and may be involved in quorum-sensing (Levesque, 2006), while *phzM* and *phzA1* are involved in the production of pyocyanins (Higgins *et al.*, 2018). No literature linking those genes to aminoglycoside resistance was identified. There were no unexpected genomic differences between PAO1:WTA4 and PAO1:ΔintA4.

Four PAO1:ΔintA3 samples and one PAO1:WTA3 sample from the 4x MIC time point were removed from the analysis due to a wrong or mixed genotype from potential mislabelling or mishandling during DNA processing, leading to a final genomic dataset of 22 PAO1:ΔintA3 and 25 PAO1:WTA3 populations at 4x MIC. All samples from the 1024x MIC time point were of the correct genotype.

Gene name	Effect	Product
phzM/phzA1	intergenic (-175/-522)	phenazine-specific methyltransferase / phenazine biosynthesis protein
PA3734	non synonymous	hypothetical protein

Table 3.2: SNPs present in the PAO1:ΔintA3 but not in the PAO1:WTA3 genome

Rearrangement identification

Potential new junctions between distant regions of the reference genome were identified through the *breseq* software on the plasmid and on the chromosome (Barrick *et al.*, 2014). Copy number variants were identified with CNOGpro (Brynildsrud, 2018) and used to confirm potential duplications or large scale deletions. Finally, *de novo* assembly of the plasmids using plasmidSPAdes (Antipov *et al.*, 2016) was carried out to provide additional evidence for the cassette rearrangements and visualised using Bandage (Wick *et al.*, 2015).

Statistical analysis

Statistical analysis were carried out using R (version 3.6.1) and RStudio (Version 1.2.5). Survival analysis using the log-rank test was performed using the *survminer* (survminer) and *survival* (Therneau, 2015) packages to compare survival rates between populations with and without a functional integrase.

3.4 Results

3.4.1 Integrase activity can increase evolvability

We selected two arrays in *P. aeruginosa* where the *aadB* cassette is placed either in second (A4) and third (A3) position and where we expect reshuffling to lead to an increase in gentamicin resistance by bringing the *aadB* cassette into first position. For each array we built a Δ*int* mutant containing a 948bp deletion of the active site of the integron integrase but conserving the *Pc* and *Pint* promoters. No significant difference in gentamicin MICs were visible between the WT arrays and the mutants Δ*int* (Supplementary Figure 3.10). Both the PAO1:A3 and PAO1:A4 strains had similar levels of gentamicin resistance (MIC = 24mg/L) when using an inoculum size mimicking the conditions of the following evolution experiment.

Using experimental evolution, we compared the evolvability of each WT and Δ*int* strain under increasing concentration of gentamicin. We followed 65 populations of each strain, starting at 1/8 MIC and doubling the concentration of gentamicin each day until reaching 1024 times the initial MIC (Figure 3.1 A). As a control, 15 populations of each strain were transferred without antibiotic (no selection pressure) while 15 populations were kept at a constant sub-MIC concentration of antibiotic (minimal selection pressure ensuring plasmid maintenance).

We observed the evolution of higher levels of gentamicin resistance in all four strains: after four transfers all populations survived at the initial MIC concentration and populations from each genotype were alive at the final time point of x1024 MIC (Figure 3.1 B). Populations with a functional integrase had a higher survival

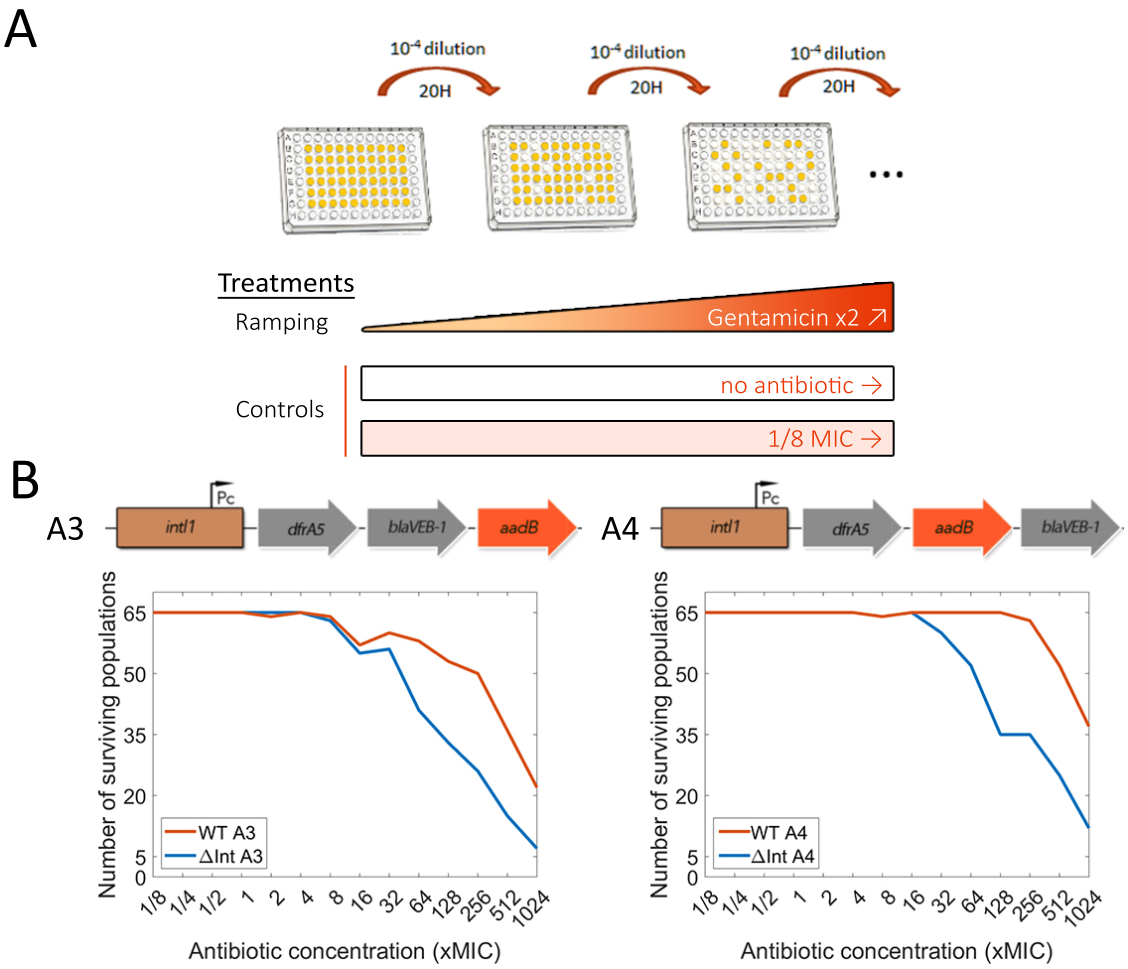


Figure 3.1: **A** Experimental evolution protocol. **B** Survival curves. The diagrams above represent the cassettes organisation for each array, with the *aadB* cassette providing gentamicin resistance highlighted in color.

rate than the Δint mutants throughout the experiment for both A3 and A4 arrays. This difference in survival rate between strains with and without a functional integrase was significant for both arrays (log-rank test: A3 arrays: Chisq= 17.7, df=1,p = 3e-05. A4 arrays: Chisq= 31.3, df=1, p= 2e-08) and shows the integrase can increase evolvability for antibiotic resistance. Interestingly, both WT and Δ*int* strains with the *aadB* cassette in second position (array A4) had a higher survival rate than the corresponding strains with *aadB* in third position (array A3) despite a similar initial MIC.

3.4.2 Extensive duplications and rearrangements mediated by the integrase

We aimed to quantify the involvement of cassette shuffling in the differences in survival rate using whole genome sequencing of the entire populations. We focused first on the A3 populations at the final x1024 MIC time-point (21 WTA3 and 6 Δ*int*A3 populations) and searched for cassette rearrangements by PCR (Supplementary Figure 3.11), showing extensive cassette rearrangements. We then sequenced these populations through short-read sequencing and identified several mechanisms potentially leading to increased levels of gentamicin resistance, both on the plasmid and on the chromosome.

Cassette rearrangements

We first aimed to quantify the extent of cassette rearrangements. Rearrangements in each population were identified from the short-read data through a combination of bioinformatic approaches (Supplementary Table 3.7): we first identified new junctions between distant parts of the reference plasmid using *breseq*, then determined cassette copy number using CNOGpro and assembled the evolved plasmid *de novo* using plasmidSPAdes. Finally, we cross-referenced the predicted recombinations from the sequencing data with the results from the PCR screen: all predicted recombinations matched the bands of the PCR screen, and only two extraneous bands could not be explained in three populations (Supplementary Figure 3.11), confirming the robustness of our bioinformatic analysis.

We observed extensive rearrangements in the PAO1:WTA3 populations (Figure 3.2). At the final time-point only 10% of the surviving PAO1:WTA3 populations still contained the ancestral integron array, with the *aadB* cassette located in first position in 19 out of 21 populations (Figure 3.3). No rearrangements were found in any of the PAO1:ΔintA3 populations nor in the 1/8 MIC WT and Δint pooled controls. The plasmid was lost in all populations transferred without antibiotic.

Apart from one mixed population, all rearranged arrays swept through the populations and were present at 100% frequency (Supplementary Table 3.7).

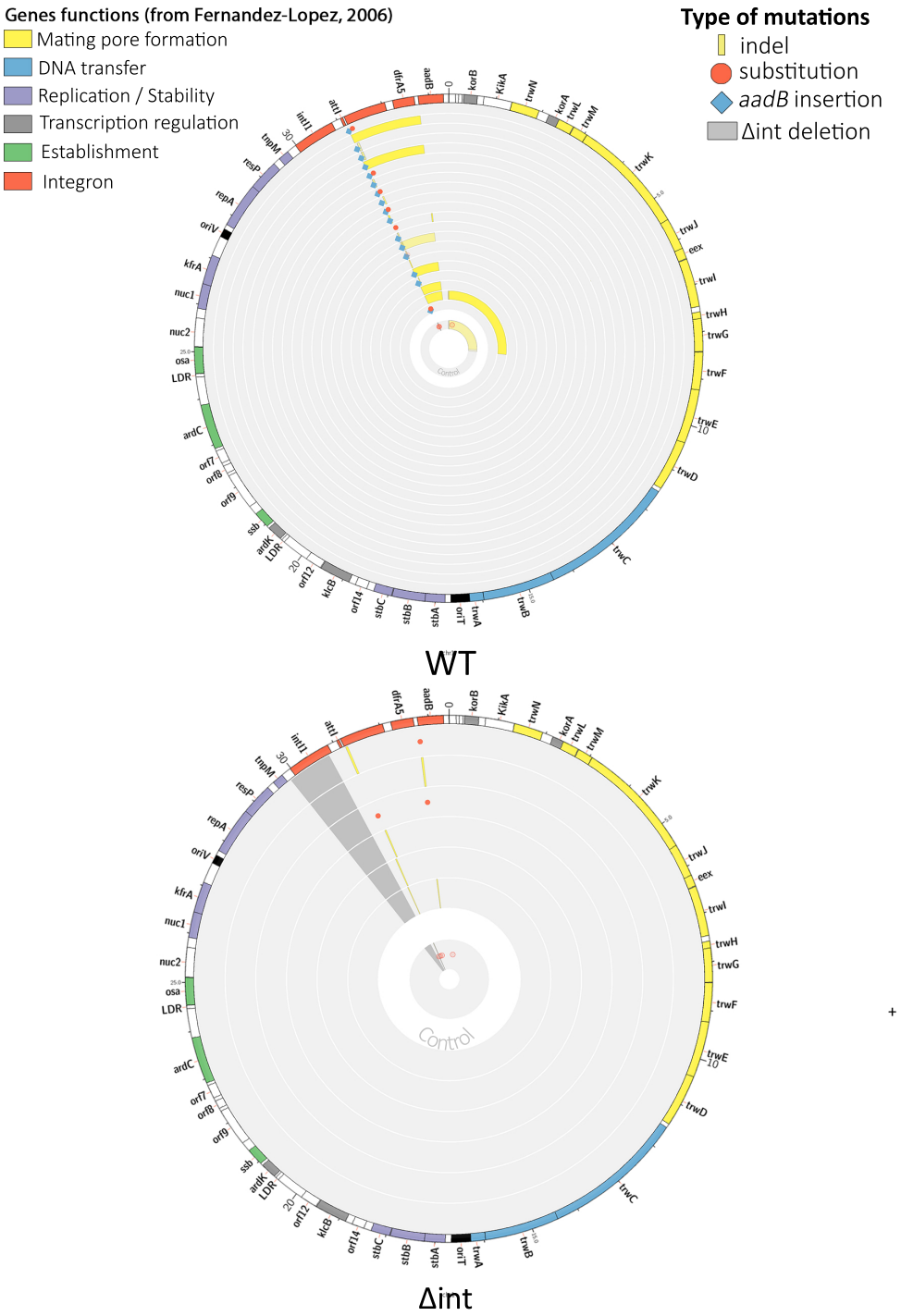


Figure 3.2: Representation of the plasmid mutations and rearrangements in the surviving PAO1:WTA3 (top - 21 populations) and PAO1:ΔintA3 (bottom - 6 populations) populations at x1024 MIC, mapped to the plasmid reference sequence. Each circle represents a separate population, with the inner circle representing the variants present in an equimolar pool of six 1/8 MIC control populations. Indels are represented in yellow and single nucleotide substitutions in red. *aadB* insertions are represented by blue lozenge. The color intensity represents the frequency of the corresponding mutation. The dark grey area in the PAO1:ΔintA3 populations represents the location of the *IntI1* deletion. The function of each R388 gene as described in (Fernández-López *et al.*, 2006) is indicated by a specific color in the outer circle.

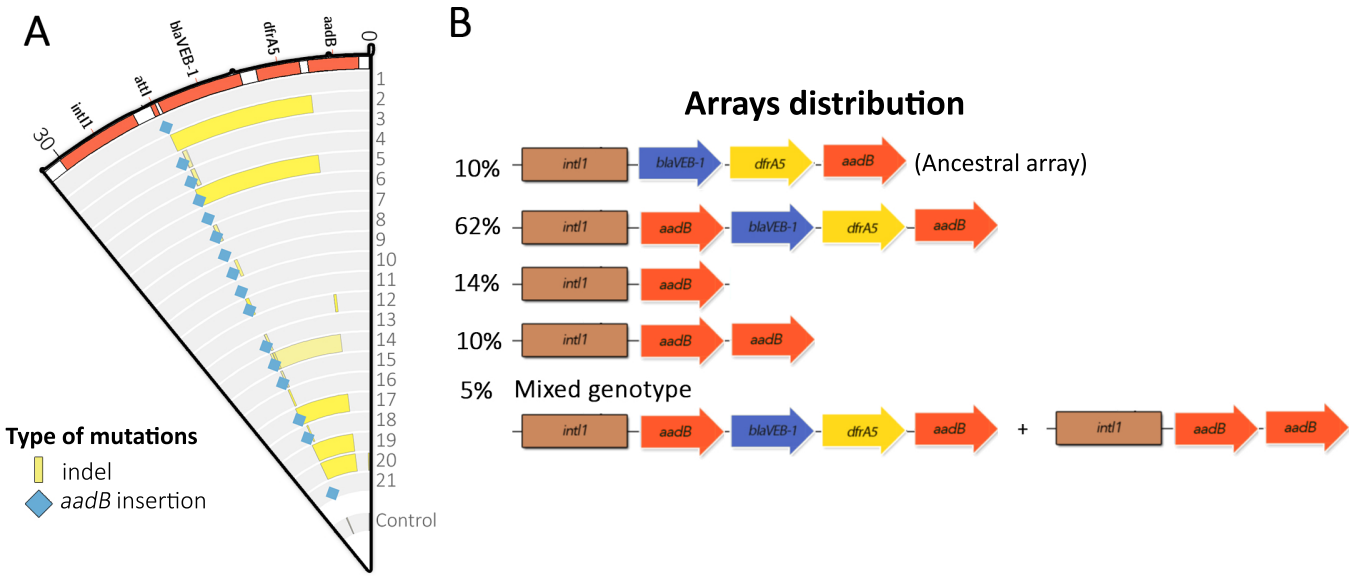


Figure 3.3: A) Close-up of the rearrangements within the WT population integrons. The sample number is indicated on the right B) Schematic representation of the different arrays identified within the WT populations. Apart from the 'Mixed genotype' population which contained two types of arrays, all populations contained only one type of array which reached fixation. Percentages represent the prevalence of these arrays across all populations.

Interestingly, we never identified any *aadB-blaVEB-1-dfrA5* arrays, which would be the result of an *aadB* excision followed by reintegration in first position within the same array. Instead we identified *aadB* duplications and deletions of the two *dfrA5* and *blaVEB-1* cassettes in 75% and 25% of populations respectively. The junction sites for both the *aadB* insertion and the *blaVEB-1-dfrA5* deletions were consistent with integrase activity: recombination happened at the 5'-GTT-3'triplet of the R box of the *attI*, *aadB attC* and *dfrA5 attC* site. The mixed WT14 population, where *aadB* is inserted in first position in 100% of the population and *blaVEB-1 - dfrA5* is deleted only partially, offers a potential hypothesis for the succession of events leading to the *aadB - aadB* array: after insertion of the *aadB* cassette in first position through *attC*x*attI* reaction, deletion of other cassettes is facilitated, either through integrase activity or homologous recombination between identical *aadB* cassettes.

blaVEB-1 mutations

We observed widespread polymorphisms at the beginning of the *blaVEB-1* cassette, both in Δ*int*A3, WTA3 populations (when *blaVEB-1* is still present) and in

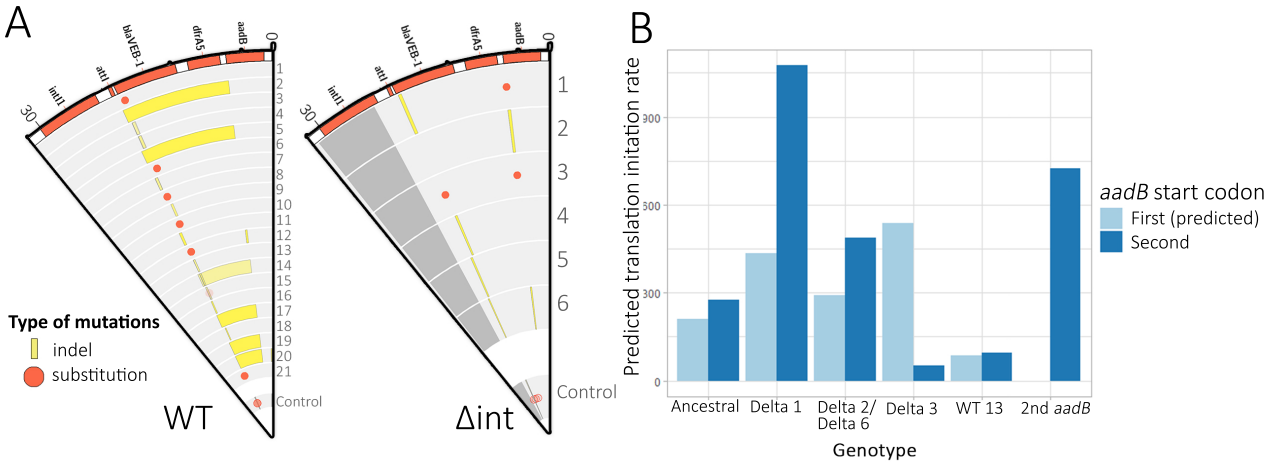


Figure 3.4: A) Close-up of substitutions and indels in the integrons of both WT and Δ*int*A3 populations. B) Predicted translation initiation rate by RBS Calculator (Espah Borujeni *et al.*, 2017) of first and second *aadB* start codon for both ancestral and mutant transcripts.

1/8 MICs controls (Figure 3.4). Overall, *blaVEB-1* is either mutated or deleted in nearly all populations in both genotypes and in both treatments (Table 3.3). These mutations vary in position and nature (indels or substitution) but are all predicted to be non-synonymous and point toward a high cost of *blaVEB-1* expression (Supplementary Table 3.8).

aadB intergenic mutations

We detected mutations in the intergenic region between the *attC* site and the start codon of the *aadB* cassette in 4 out of 6 of ΔintA3 populations. Using RBS Calculator (Espah Borujeni *et al.*, 2017) on the evolved integron transcripts, we predicted a higher translation initiation rate at either the *aadB* start codon or at the following start codon, located 21bp downstream (Figure 3.4 B). Interestingly, the translation initiation rate of a second *aadB* cassette is also predicted to increase when following another *aadB attC* site, as in the *aadB* - *aadB* arrays. Finally we observed a 41 bp deletion within the *dfrA5 attC* site of a WTA3 population which was not predicted to increase translation initiation rate but may increase translational coupling with *dfrA5* or lead to the creation of a fused protein with part of the previous cassette.

Gene name	Ramping				1/8 MIC	
	Delta (freq)	Delta (n)	WT (freq)	WT (n)	Delta	WT
blaVEB-1	0.83	5	0.96	27	0.96	0.97
dfrA5/aadB gene	0.67	4	0.10	2	0.00	0.00
dfrA5	0.00	0	0.27	6	0.00	0.00
eex	0.00	0	0.05	1	0.00	0.38
KikA	0.00	0	0.05	1	0.00	0.38
korA	0.00	0	0.05	1	0.00	0.38
korB	0.00	0	0.05	1	0.00	0.38
trwF	0.00	0	0.05	1	0.00	0.38
trwG	0.00	0	0.05	1	0.00	0.38
trwH	0.00	0	0.05	1	0.00	0.38
trwI	0.00	0	0.05	1	0.00	0.38
trwJ	0.00	0	0.05	1	0.00	0.38
trwK	0.00	0	0.05	1	0.00	0.38
trwL	0.00	0	0.05	1	0.00	0.38
trwM	0.00	0	0.05	1	0.00	0.38
trwN	0.00	0	0.05	1	0.00	0.38
hypothetical protein	0.00	0	0.00	0	0.16	0.12

Table 3.3: List of mutated plasmid genes and their average mutation frequency at x1024 MIC. (freq) represents the cumulative frequencies of mutations found within each mutated plasmid gene, normalized by the number of samples, across each genotype. (n) represents the number of mutations present in this gene across populations from this genotype. (n) may be greater than the number of populations if populations contain multiple mutations. Cumulative mutation frequency for these genes in the 1/8 MIC pooled populations is indicated as reference.

Other plasmid mutations & rearrangements

While most of the recombinations and mutations visible within the plasmid were localised to the integron array, we observed an extended deletion in one WTA3 population. This deletion starts from the *aadB attC* site and finishes within the *trwF* gene, effectively deleting most of the genes involved in mating pore formation (Figure 3.5). Two similar deletions are visible in the 1/8 MIC WTA3 pooled control, both starting at the same point of the *aadB attC* site but finishing at different locations in the *trwF* gene. These deletions are potentially the result of an integrase mediated recombination between the *attC* site of *aadB* and a secondary site. The

deletion start matches the 5'-GTT-3'recombination site of the *attC aadB* site. On the other end for both deletions the junction happens within a 5'-GAT-3'triplet and a motif close to GTTRRRY. While this is not the canonical 5'-GTT-3'triplet, it has been shown that the identity of the second nucleotide in the triplet does not influence the efficiency of the integrase binding, and that 5'-GAT-3'is the favoured triplet for recombination with secondary sites (Francia *et al.*, 1993). These deletions may therefore be the result of off-target activity of the integrase.

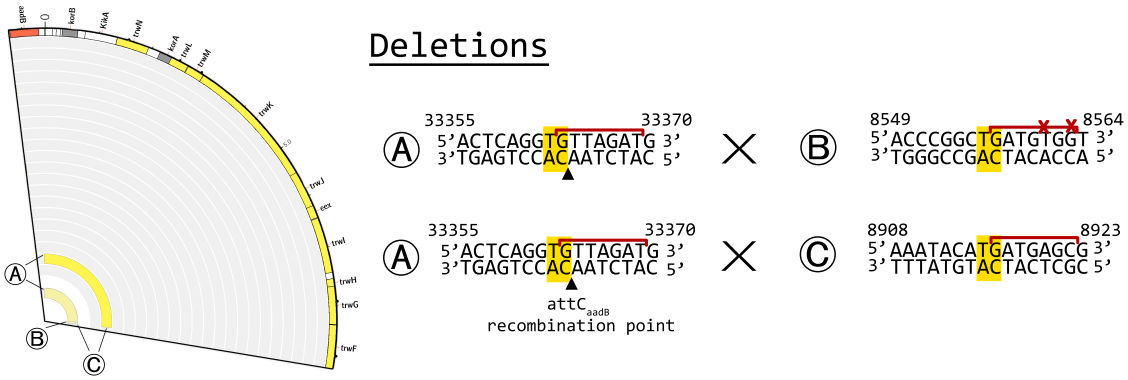


Figure 3.5: Left: Rearrangements in the WT populations. Each junction site is indicated by a letter (deletion). Right: junction sequences for each rearrangement. Junction site is indicated in yellow. When the crossover point is unclear due to sequence homology between each junction the entire homology is highlighted. Motifs close to the GTTRRRY integron secondary sequence are indicated by a red line, with red crosses indicating the potential mismatch. The recombination point of the *aadB attC* site is indicated by a black arrow.

Chromosomal mutations

Finally, we also observed numerous mutations in the chromosome, with a very similar average cumulative frequency of 2.66 and 2.64 SNPs per population for WTA3 and Δ*int*A3 genotypes respectively. With mutations present in more than 40% of populations, the most frequent targets were *amgS*, *rluD* and *PA2020* (*mexZ*) genes (Table 3.4). *mexZ* is a repressor gene of the *mexXY* multidrug efflux system operon and 8 out of 12 mutations we observed in *mexZ* were frameshift mutations (Supplementary Table 3.9). Mutations inactivating *mexZ* have been widely identified in aminoglycoside resistant *P. aeruginosa* isolates found in cystic fibrosis patients (Vogne *et al.*, 2004), and upregulation of *mexXY* has been associated with a 2- to

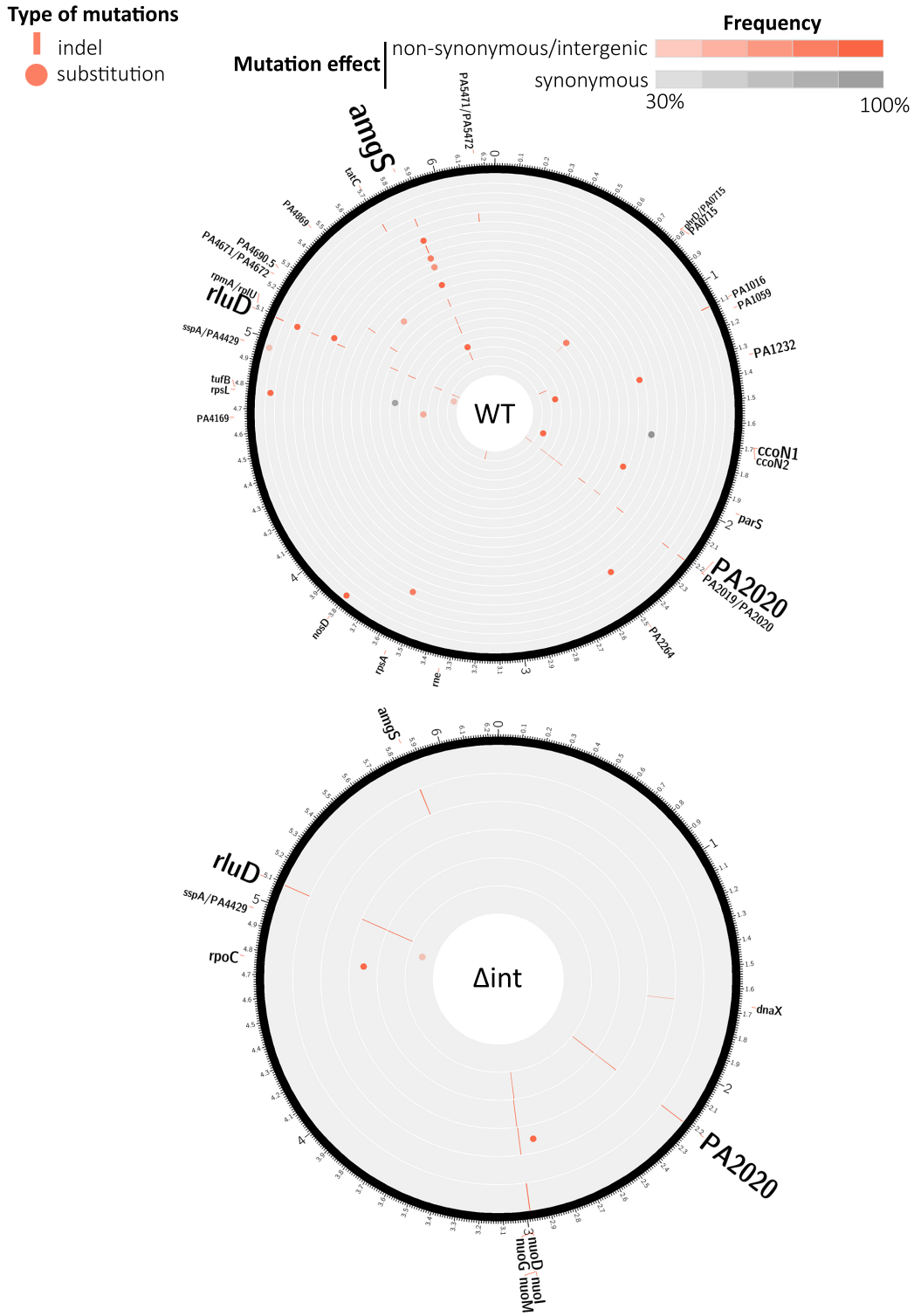


Figure 3.6: Chromosomal mutations at the final time-point. Each circle represents a different population. The color of the mutation marker represents its frequency within a population. The size of the gene labels on the outer ring represents the overall cumulative frequency across all populations from this genotype.

16- fold increase in aminoglycoside resistance (Vogne *et al.*, 2004). *AmgS* is part of an envelope stress-responsive two-component system called AmgRS (Lau *et al.*, 2013), and *amgS* mutations can also upregulate the multidrug efflux system *mexXY* in the presence of aminoglycosides (Lau *et al.*, 2015). Every insertion and deletion observed within this gene kept the reading frame intact, hinting to gain instead of loss of function. *rluD* is a pseudouridine synthase which introduces pseudouridine modifications in helix 69 of the 23S rRNA, a known aminoglycoside binding site (Wang *et al.*, 2012). In our experiment 8 out of 11 mutations introduced a frame-shift, and *rluD*(-) strains were shown to have reduced aminoglycoside binding (Sakakibara and Chow, 2017).

Other less frequent mutations are also plausibly linked with aminoglycoside resistance: *rpmH, rpSA, PA4671, rplU, rpsL* are ribosomal proteins and *rpsL* and *gidB* mutations have shown to confer high level of streptomycin resistance (Wistrand-Yuen *et al.*, 2018), while *ccoN*, *cycH*, *nosD*, *tatC*, *PA4429* and the *nuo* operon were identified in a tobramycin resistance screen (Wang *et al.*, 2012). Interestingly mutations in the nuo pathways were only observed in the Δ*intA3* populations (Table 3.4).

We observed no wide-scale rearrangement in either WTA3 or Δ*intA3* populations: the largest recombination event was a 298bp deletion in the *gidB* gene of one ΔintA3 population (Supplementary Table 3.9). The overall mutations distribution between deletions, insertions and substitutions was roughly similar between the two genotypes (Supplementary Figure 3.12), and no significant impact of integrase activity could be observed. Excluding one WTA population, the rate of synonymous mutation was extremely low in the ramping populations, showing strong sign of selection.

In conclusion, we identified extensive cassette rearrangements and duplications within the populations with a functional integrase, which explain the observed increase in evolvability. Apart from cassette rearrangements and mutations in the intergenic region of *aadB* and the *nuo* pathway, both WT and Δ*int* genotypes shared similar evolutionary trajectories with wide-spread mutations in the *blaVEB-1* cassette and in a few chromosomal genes involved in aminoglycoside resistance.

Gene name	Ramping				1/8 MIC		no Antibiotic		
	Delta (freq)	Delta (n)	WT (freq)	WT (n)	Delta	WT	Delta	WT	Gene product
PA2020	0.50	3	0.43	9	0.00	0.00	0.00	0.00	transcriptional regulator
rluD	0.50	3	0.38	8	0.00	0.00	0.00	0.00	pseudouridine synthase
amgS	0.17	1	0.50	11	0.00	0.00	0.00	0.00	protein AmgS
gidB	0.17	1	0.00	0	0.00	0.00	0.00	0.00	protein gidB
nuoD	0.17	1	0.00	0	0.00	0.00	0.00	0.00	NADH-quinone oxidoreductase subunit C/D
nuoG	0.17	1	0.00	0	0.00	0.00	0.00	0.00	NADH-quinone oxidoreductase subunit G
nuoI	0.17	1	0.00	0	0.00	0.00	0.00	0.00	NADH-quinone oxidoreductase subunit I
nuoM	0.17	1	0.00	0	0.00	0.00	0.00	0.00	NADH-quinone oxidoreductase subunit M
PA2570.1/PA2571	0.17	1	0.00	0	0.00	0.00	0.00	0.00	tRNA-Leu/two-component sensor
rpmH	0.17	1	0.00	0	0.00	0.00	0.00	0.00	50S ribosomal protein L34
rpoC	0.17	1	0.00	0	0.00	0.00	0.00	0.00	DNA-directed RNA polymerase subunit beta'
ccoN1	0.00	0	0.16	6	0.00	0.00	0.00	0.00	cbb3-type cytochrome C oxidase subunit I
PA0668.4	0.00	0	0.10	2	0.00	0.00	0.00	0.00	23S ribosomal RNA
PA1232	0.00	0	0.10	2	0.00	0.00	0.00	0.00	hypothetical protein
parS	0.00	0	0.10	2	0.00	0.00	0.00	0.00	two-component sensor ParS
sspA/PA4429	0.06	1	0.03	2	0.00	0.00	0.00	0.00	stringent starvation protein A/cytochrome C1
ccoN2	0.00	0	0.09	4	0.00	0.00	0.00	0.00	cbb3-type cytochrome C oxidase subunit I
dnaX	0.08	1	0.00	0	0.00	0.00	0.00	0.00	DNA polymerase III subunits gamma/tau
PA2019/PA2020	0.00	0	0.07	2	0.00	0.00	0.00	0.00	multidrug efflux lipoprotein/transcriptional regulator
nosD	0.00	0	0.05	1	0.00	0.00	0.00	0.00	copper-binding periplasmic protein
PA1016	0.00	0	0.05	1	0.00	0.00	0.00	0.00	thiolase
PA1059	0.00	0	0.05	1	0.00	0.00	0.00	0.00	monovalent cation/H+ antiporter subunit G
PA2264	0.00	0	0.05	1	0.00	0.00	0.00	0.00	hypothetical protein
PA4671/PA4672	0.00	0	0.05	1	0.00	0.00	0.00	0.00	50S ribosomal protein L25/general stress protein Ctc/ peptidyl-tRNA hydrolase
PA4690.5	0.00	0	0.05	1	0.00	0.00	0.00	0.00	16S ribosomal RNA
PA5471/PA5472	0.00	0	0.05	1	0.00	0.00	0.00	0.00	hypothetical protein/hypothetical protein
rne	0.00	0	0.05	1	0.00	0.00	0.00	0.00	ribonuclease E
rpmA/rplU	0.00	0	0.05	1	0.00	0.00	0.00	0.00	50S ribosomal protein L27/50S ribosomal protein L21
rpsL	0.00	0	0.05	1	0.00	0.00	0.00	0.00	30S ribosomal protein S12
tatC	0.00	0	0.05	1	0.00	0.00	0.00	0.00	transporter TatC
tufB	0.00	0	0.04	1	0.00	0.00	0.00	0.00	elongation factor Tu
PA0715	0.00	0	0.04	1	0.00	0.00	0.00	0.00	hypothetical protein
rpsA	0.00	0	0.04	1	0.00	0.00	0.00	0.00	30S ribosomal protein S1
PA4169	0.00	0	0.03	1	0.00	0.00	0.00	0.00	hypothetical protein
PA4869	0.00	0	0.03	1	0.33	0.00	0.00	0.00	hypothetical protein
phrD/PA0715	0.00	0	0.01	1	0.00	0.11	0.00	0.00	PhrD/hypothetical protein

Table 3.4: List of mutated chromosome genes in the ramping treatment at x1024 MIC, their average mutation frequency and the mutation count within each genotype. *(freq)* represents the cumulative frequencies of mutations found within each mutated plasmid gene, normalized by the number of samples and across each genotype while *(n)* represents the number of mutations present in this gene across all populations from this genotype. Cumulative mutation frequencies for these genes in 1/8 MIC and no antibiotic pooled control populations are indicated as reference.

3.4.3 x4 MIC time-point

In order to disentangle the importance of these different mechanisms in early evolution of aminoglycoside resistance, we went 'back in time' and investigated populations at the x4 MIC time-point. We obtained genomic sequences of 24 WTA3 and 22 Δ*intA3* whole populations selected randomly, of which 8 WTA3 and 3 Δ*intA3* populations went on to survive until the final time-point.

Cassette rearrangements

We already observe extensive rearrangements, with more than 70% of WTA3 populations containing rearranged arrays. At this earlier time-point, the frequency of mixed arrays populations is much higher, with more than 50% of populations containing more than one type of array (Figure 3.7 C). *aadB* duplications are again widespread, with the *aadB - blaVEB-1 - dfrA5 - aadB* array still being the most common rearrangement. We observe an additional cassette rearrangement where *aadB* is duplicated at the end of the array (*blaVEB-1 - dfrA5 - aadB - aadB*). While this rearrangement may be less beneficial than bringing the *aadB* cassette into first position, we can still expect additional cassette expression from the second *aadB* copy, especially as we predict a higher translation initiation rate of *aadB* when following an *aadB attC* site. The mixed *aadB - blaVEB-1 - dfrA5 - aadB/aadB* populations also point toward the insertion of *aadB* in first position followed by excision of the rest of the array as the potential mechanism to reach the *aadB* array. As expected, no rearrangements or duplications were visible in Δ*intA3* populations.

Plasmid mutations

We already observe extensive *blaVEB-1* mutations in every WTA3 populations and 75% of ΔintA3 populations (Table 3.5). We identified an extreme diversity of mutations with 21 different mutations (Supplementary Table 3.11), of which 12 are indels creating frame-shift, 5 are non-sense mutations and the rest are non-synonymous substitutions, showing that knocking-off *blaVEB-1* is heavily selected for. We only observed a mutation in the intergenic part of *aadB* in only

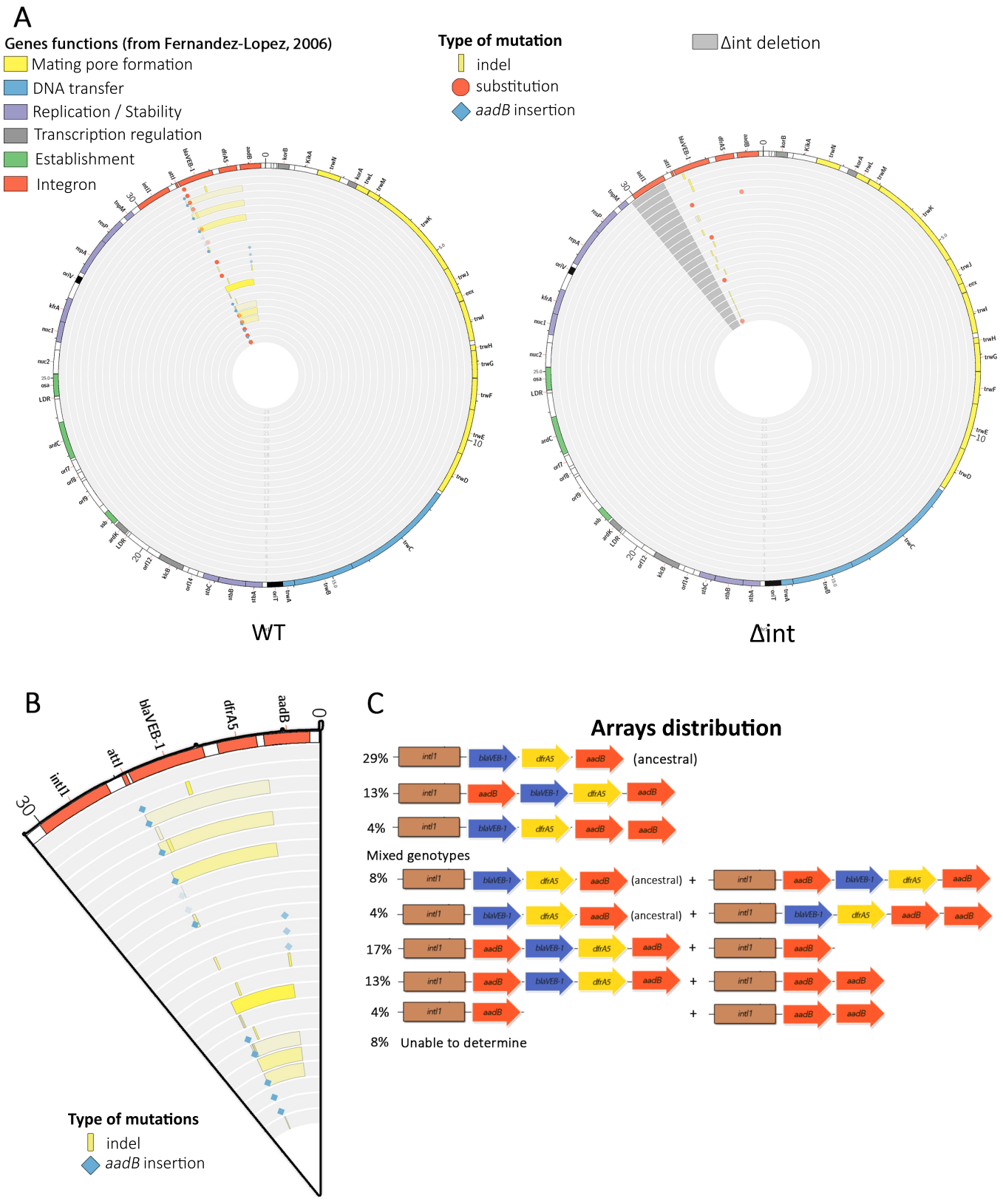


Figure 3.7: Plasmid mutations and rearrangements at x4 MIC. A) Representation of the full plasmid. Each circle represents a different population and its corresponding mutations/rearrangements. Plasmid gene functions are indicated by the colours of the outer circle. B) Snapshot of the integron recombinations of the WT populations. C) Rearranged arrays distribution within the WT populations.

one population, and no mutation in the backbone of the plasmid was identified. When excluding cassette rearrangements, the overall plasmid mutation frequency is similar between the two genotypes (3.13).

Gene name	Ramping (x4 MIC)			
	Delta (freq)	Delta (n)	WT (freq)	WT (n)
blaVEB-1 gene	0.74	17	1.00	31
dfrA5 gene	0.00	0	0.13	6
dfrA5/aadB gene	0.03	1	0.04	1
intI1/blaVEB-1 gene	0.05	1	0.00	0

Table 3.5: List of mutated plasmid genes and their average mutation frequency at x4 MIC. (freq) represents the cumulative frequencies of mutations found within each mutated plasmid gene, normalized by the number of samples, across each genotype. (n) represents the number of mutations present in this gene across populations from this genotype. (n) may be greater than the number of populations if a population contain multiple mutations.

Chromosomal mutations

The frequency of chromosomal mutations was much lower, with several populations keeping the ancestral chromosome sequence (Supplementary figure 3.8). Interestingly, we observed a significantly higher cumulative frequency ($t = -3.04$, $df = 136.32$, $p < 0.03$, Welch Two Sample t-test) of chromosomal mutations in theΔ*intA3* populations when compared with the WTA3 populations (Figure 3.13), due to a higher frequency of non-synonymous substitutions in the in theΔ*intA3* populations. This shows the increased occurrence of chromosomal selective sweeps in the absence of cassette rearrangements.

Surprisingly, there is reduced overlap between the mutated genes of the two time-points, with 15 out of 26 genes mutated at x4 MIC time point, but not at x1024 MIC. Among the genes in common, we observe again mutations in *PA2020, rluD* and *amgS* but at a reduced prevalence. Instead, we observe a slight increase in mutations in the ribosomal RNA and tRNA related genes. Interestingly, we observe again mutations in the *nuo* pathway exclusively in Δ*intA3* populations (Table 3.6).

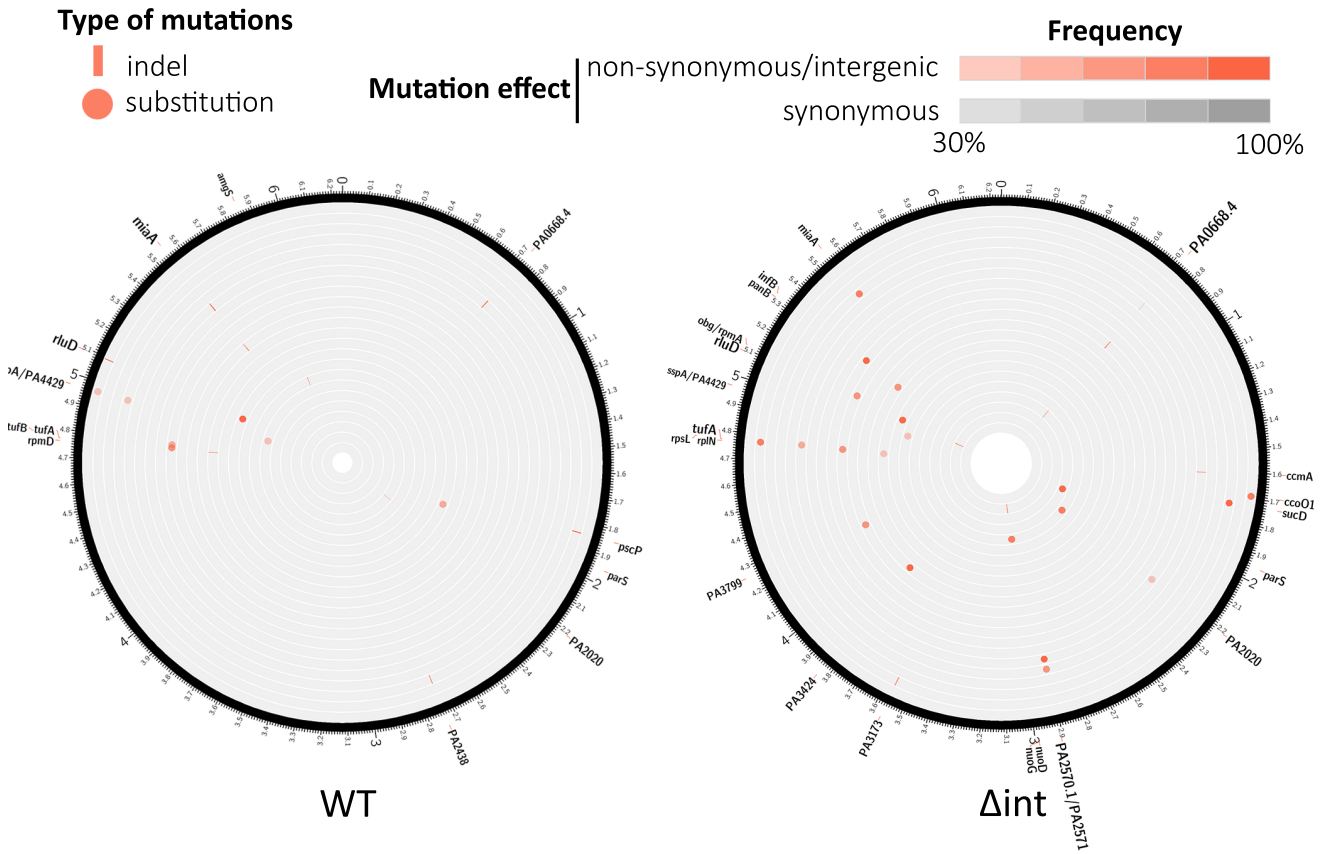


Figure 3.8: Chromosomal mutations at x4 MIC. Each circle represent a different population. Synonymous mutations are indicated in grey. The intensity of the markers represent the frequency of the mutation within the population

3.4.4 Duplications in pAMBL plasmids

We then investigated the potential for cassette duplications with another cassette, the metallo-b-lactamase bla_{VIM-1}.The cassettebla_{VIM-1} is found in two plasmids isolated in *P. aeruginosa* called pAMBL2 and pAMBL1. The plasmid AMBL2 contains three copies of the bla_{VIM-1} cassette and confers a greater resistance to meropenem in PAO1 than pAMBL1, which only possesses one. (San Millan *et al.*, 2015b). We therefore aimed to see if we could induce bla_{VIM-1} amplification in pAMBL1 using a similar experimental evolution protocol with meropenem (Supplementary material 3.7.5). We did not see a difference in evolvability between the different strains (**Figure** 3.14(B)), but were able to identify numerous potential duplications and rearrangements in pAMBL1 by PCR (**Figure** 3.14(C)), showing cassette duplication is not restricted to the *aadB* cassette but may be a widespread

Gene name	Ramping				Gene product
	Delta (freq)	Delta (n)	WT (freq)	WT (n)	
rluD	0.09	2	0.08	2	pseudouridine synthase
PA0668.4	0.11	3	0.04	1	23S ribosomal RNA
miaA	0.04	1	0.08	2	tRNA delta(2)-isopentenylpyrophosphate transferase
tufA	0.07	2	0.03	1	elongation factor Tu
PA2020	0.06	2	0.04	1	transcriptional regulator
PA2570.1/PA2571	0.08	2	0.00	0	tRNA-Leu/two-component sensor
parS	0.05	1	0.02	1	two-component sensor ParS
sspA/PA4429	0.02	1	0.04	3	stringent starvation protein A/cytochrome C1
infB	0.05	1	0.00	0	translation initiation factor IF-2
nuoG	0.05	1	0.00	0	NADH-quinone oxidoreductase subunit G
PA3424	0.05	1	0.00	0	hypothetical protein
sucD	0.05	1	0.00	0	succinyl-CoA ligase subunit alpha
ccoO1	0.04	1	0.00	0	cbb3-type cytochrome C oxidase subunit II
pscP	0.00	0	0.04	1	translocation protein in type III secretion
nuoD	0.04	1	0.00	0	NADH:-quinone oxidoreductase subunit C/D
PA3173	0.04	1	0.00	0	short-chain dehydrogenase
ccmA	0.03	1	0.00	0	cytochrome c biogenesis ATP-binding export protein CcmA
obg/rpmA	0.03	1	0.00	0	GTPase ObgE/50S ribosomal protein L27
PA3799	0.03	1	0.00	0	GTPase Der
panB	0.03	1	0.00	0	3-methyl-2-oxobutanoate hydroxymethyltransferase
PA2438	0.00	0	0.03	1	hypothetical protein
rpsL	0.03	1	0.00	0	30S ribosomal protein S12
rpmD	0.00	0	0.02	1	50S ribosomal protein L30
tufB	0.00	0	0.02	1	elongation factor Tu
amgS	0.00	0	0.02	1	protein AmgS
rplN	0.02	1	0.00	0	50S ribosomal protein L14

Table 3.6: List of mutated chromosome genes in the ramping treatment at x4 MIC, their average mutation frequency and the mutation count within each genotype. *(freq)* represent the cumulative frequencies of mutations found within each mutated plasmid gene, normalized by the number of samples and across each genotype while *(n)* represent the number of mutations present in this gene across all populations from this genotype.

consequence of integrase activity.

3.5 Discussion

Integrons are genetic shuffling devices allowing bacteria to excise and integrate gene cassettes within the integron array, modulating the expression of gene cassettes whose expression depends on their location from the start of the array. While the molecular mechanisms behind integron cassette shuffling are well-known (Cambray *et al.*, 2010b) and the selective advantage of the integrase in changing environments

has been shown theoretically (Engelstädter *et al.*, 2016), the evolutionary benefits of integrase-mediated recombination remain to be experimentally quantified. In this chapter we used the experimental system described previously to investigate the impact of integrase activity on the evolution of antibiotic resistance through the combination of experimental evolution and whole genome sequencing.

As shown previously, the expression of the *aadB* cassette is heavily influenced by its position within the array. We selected two arrays with *aadB* in second and last position where we expect the benefit of recombination to be maximal and quantified the benefits of integrase activity on survival rates by comparing strains with and without a functional integrase under increasing concentration of gentamicin.

We were able to identify a benefit from integrase activity: while all strains were able to evolve higher levels of gentamicin resistance, we observed a higher rate of survival in both strains containing the functional integrase, independently of the position (2nd or 3rd within the array) of the *aadB* cassette. This confirms that the effects of integrase activity can overcome the cost of integrase expression (Starikova *et al.*, 2012) and increase bacteria evolvability for antibiotic resistance. These results are in line with (Barraud and Ploy, 2015) which showed over-expression of the integrase speeds up the evolution of higher levels of chloramphenicol resistance, when compared with wild-type integrase expression or a deactivated integrase.

Using whole genome sequencing we identified extensive cassette rearrangements at both an early and final time-point. These rearrangements were exclusively in the populations containing a functional integrase and are expected to all be integrase-mediated. Nearly all these rearrangements brought the *aadB* cassette closer to the promoter and were sometimes linked with excision of the *blaVEB-1* and *dfrA5* cassettes. Analysis of the mixed populations shows insertion of the *aadB* cassette was often the first event sweeping through the population, and excision of the other cassettes followed after (Figure 3.9). The complexity of these rearrangements while the integrase is under control of LexA is interesting: some

arrays already present at the x4 MIC time-point are predicted to be the result of three successive recombination reactions. By comparison (Barraud and Ploy, 2015) saw a single rearrangement (excision of two middle cassettes) bringing *catB9* closer to the promoter under wild-type expression of the integrase after 5 days of transfers. Moreover all of their populations still contained the ancestral array, while rearrangements had already swept through more than half of our populations at x4 MIC. They had however minimal SOS induction, as chloramphenicol is a low SOS inducer in *E. coli* (Baharoglu and Mazel, 2011b) and they used a Δ*RecA E. coli*.

The prevalence of the insertion of *aadB* in first position in our experiment is quite striking: it is one of the main reshufflings in our experiment, while (Barraud and Ploy, 2015) never saw *catB9* insertion in first position when they over-expressed the integrase, and instead observed multiple cassettes deletions and duplications mainly due to cointegrate formation. Plasmid cointegrates (fusion of the plasmids) are the result of integrase-mediated intermolecular events occurring between two copies of a plasmid and their corresponding *attI* or *attC* sites (Escudero *et al.*, 2015), and can be a frequent product of integrase activity when the integrase is over-expressed (Aubert *et al.* (2012), Collis and Hall (1992b)). Cointegrates could produce some of the rearrangements we observed, such as the *blaVEB-1* - *dfrA5* - *aadB* - *aadB* array, and are difficult to identify with short-read sequencing. However the presence of only one type of array in most of the populations by the final time-point shows any co-integrates are likely to be resolved by the end of experiment, or do not generate array diversity. Moreover all rearrangements can be explained through involvement of a limited number of classical integron reactions involving the *aadB attC* site, which has been shown to be extremely recombinogenic (Aubert *et al.*, 2012). We therefore expect cointegrates to only play a minor role in our experiment, but only long-read sequencing or *in-vitro* measurement of the plasmid sizes would allow us to confirm their absence.

Interestingly, we never observed separation of the *blaVEB-1* and *dfrA5* cassettes, potentially due to the low recombination efficiency of the *blaVEB-1 attC* site, which joins the cassettes together. Excision of *blaVEB-1*/*dfrA5* seems to have always been

preceded by *aadB* insertion, which can be explained by the dynamics of the integrase reactions: excisions through $attI \times attC_{dfrA5}$ reactions are much less favorable than $attC_{aadB} \times attC_{aadB}$ or $attC_{aadB} \times attC_{dfrA5}$ reactions. Here we can see how an extremely mobile cassette, such as *aadB*, can enhance the mobility of other less recombinogenic cassettes.

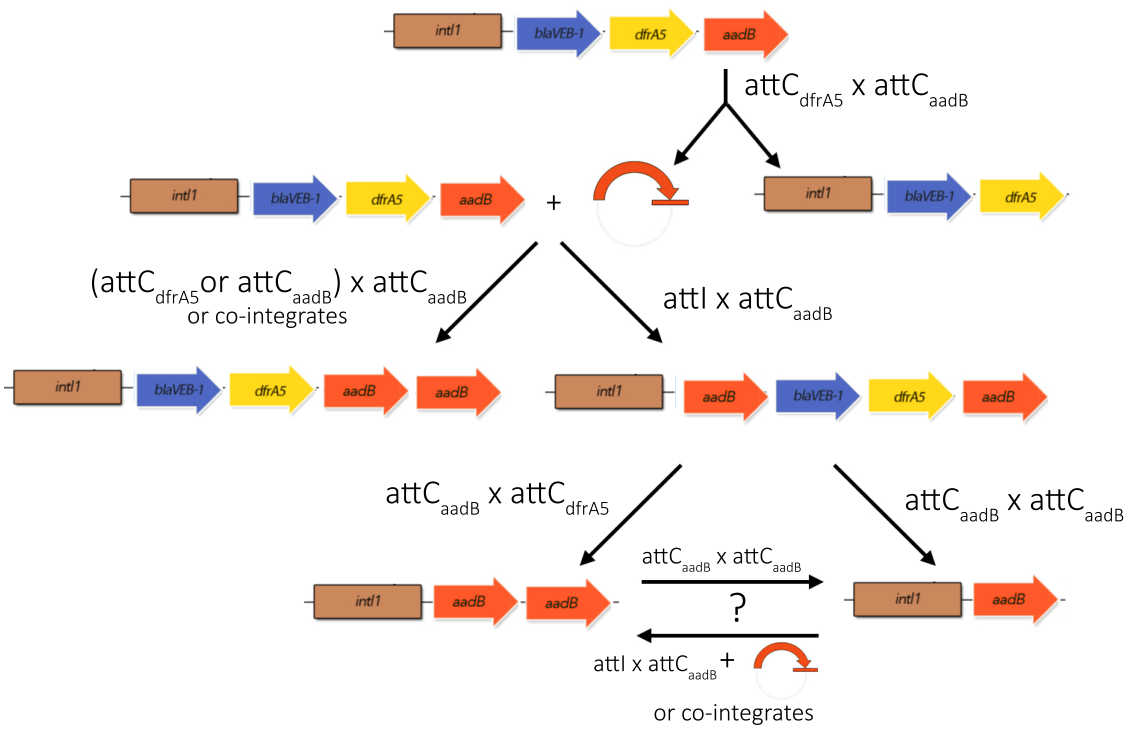


Figure 3.9: Representation of the proposed reactions leading to the various arrays observed.

Cassette duplications were a major consequence of integrase activity and the major pathway to evolve gentamicin resistance from an early time-point in the WTA3 populations. As mentioned previously, several integrase reactions can produce cassette duplications, through cointegrate formation or excision followed by reinsertion of a cassette in a different copy of the ancestral array. Surprisingly, we never observed the resulting excised array nor an array where the reinsertion happened within the same integron. The first absence is easy to understand: plasmids containing arrays without the *aadB* cassette create a cost without contributing to gentamicin resistance and should therefore be quickly lost during segregation. The

second absence is more puzzling: some final *aadB* arrays may be the result of a *blaVEB1/dfrA5* deletion from *aadB - blaVEB-1 - dfrA5* populations. Moreover *aadB* reinsertion in a different array could be selected for more strongly, as we saw in a previous chapter that a distant *aadB* does provide a small amount of gentamicin resistance, and *aadB - blaVEB-1 - dfrA5* arrays could lose to *aadB - blaVEB-1 - dfrA5 - aadB* if both were to arise within the same populations.

We showed cassette duplications were not limited to *aadB* but can also happen with the clinically relevant *blaVIM* cassette in the pAMBL1 plasmid. While we cannot strictly prove these duplications were mediated by integrase activity, similar *blaVIM* duplications were found in a plasmid called pAMBL2 isolated in an urology ward in Spain (San Millan *et al.*, 2015b) and provided higher levels of meropenem resistance. Duplicated cassettes have been found in large chromosomal integrons such as the *V. cholerae* super integron (Escudero *et al.*, 2015), but have rarely been described in mobile integrons (San Millan *et al.* (2015b), Stokes and Hall (1992)). We were able to identify only 50 arrays with duplicated cassettes, of which 5 contained *aadB*, among the class 1 integrons arrays listed by the INTEGRALL database (Moura *et al.*, 2009) (Supplementary Table 3.13), which represent less than 5% of the total of arrays containing two cassettes or more. This low number may be due to technological limitations, as duplications are difficult to resolve by short-read sequencing (Alkan *et al.*, 2011). However duplicated cassettes in mobile integrons may also be particularly prone to deletion in the absence of selection: when integrase expression is repressed, the integron cost comes largely from its cassette arrays and increases with the number of cassettes (Lacotte *et al.*, 2017). Moreover repetitive DNA sequences in bacteria are inherently unstable and may quickly collapse, even without integrase activity (Andersson and Hughes, 2009). Studying the stability of our duplicated arrays in the absence of selection would therefore be an interesting follow-up experiment.

Even if short-lived, cassette duplication may play an essential role in the evolutionary success of the integron. Alongside increased levels of gene expression, gene duplication offers potential for innovation and creation of novel functions from

existing genes (Taylor and Raes (2004),Toll-Riera *et al.* (2016)). Gene duplication may be especially likely in mobile integrons where integrons are present on several plasmid copies within a cell, but multiple copies of the integrons are not strictly necessary: as cassette excision is a semiconservative process (Escudero *et al.*, 2015) a copy of the un-excised array will always be present. Moreover, multifork replication can cause bacteria cells to be effectively polyploid, potentially letting several copies of a chromosomal integron coexist within a cell (Sun *et al.*, 2018).

Alongside cassette recombinations a striking result from our genomic analysis was the widespread mutations within the *blaVEB-1* gene, potentially stopping its expression. Trade-offs between resistance mechanisms, also called collateral sensitivity (Lázár *et al.*, 2013), have been shown to occur frequently during the evolution of antibiotic resistance and especially with aminoglycosides, but usually through modifications of the cell membrane potential. However *blaVEB-1* encodes for a β-lactamase which we do not expect to interact with other resistance mechanisms. An explanation could be that missfolded *blaVEB-1* proteins are particularly toxic: most of the mutations happened in the start of the gene, where the signal peptide for the periplasm is located. *blaVEB-1* missfolded proteins could potentially disturb the cell membrane. *blaVEB-1* could also simply be a very costly plasmid gene: unlike other plasmid genes (Fernández-Lópe *et al.*, 2014) integrons cassettes are constitutively expressed at a high level and, as shown in the previous chapter, *blaVEB-1* has a high translation rate. The translation of plasmid genes can be particularly costly due to discrepancies between codon usage by the plasmid gene and the host tRNA pool (San Millan and MacLean, 2017). However *blaVEB-1* is widely found in *P. aeruginosa* (Naas *et al.*, 2008). Determining the effect of these *blaVEB-1* mutations on bacteria fitness with and without gentamicin would give us an answer to this question.

Finally we identified several chromosomal mutations involved in gentamicin resistance and especially in three genes: *mexZ* (PA2020), a repressor of the efflux

pump *mexXY*, *amgS*, a regulator of the *mexAB-oprM* multidrug efflux operon and *rluD*, a pseudouridine synthase involved in ribosome assembly. Surprisingly we did not identify any signature of integrase activity within the chromosome. The integrase can mediate recombination between *attC*/*attI* sites and non-canonical sites in the genome (Francia *et al.* (1993),Recchia *et al.* (1994)), leading to insertion of integron cassettes outside of the integron (Recchia and Hall, 1995b) or potential deleterious deletions, which is why genomic instability was proposed as a potential cost of the integrase (Harms *et al.*, 2013). We observed no *aadB* insertion outside of the integron. To be expressed outside of the integron, *aadB* would be required to be inserted downstream of both a promoter and a Shine-Dalgarno sequence, making this unlikely, but not impossible, as reported in (Recchia and Hall, 1995b). Moreover, no extensive rearrangements were observed in the chromosome. (Harms *et al.*, 2013) showed *P. aeruginosa* has much less potential *attI* core site motifs (GTTRRRY) than expected, but still leaves more than 4500 potential sites. If anything, populations with a dysfunctional integrase had a higher rate of chromosomal mutations than the wild-types at the early time-point, potentially to compensate for the absence of cassette rearrangements. Our experiment is obviously biased against deleterious mutations, as we can only sequence the 'winners' of our experiment, but we could have expected mutations to hitch-hike within cells with beneficial plasmid rearrangements.

To conclude, we showed in this chapter how the integron integrase can increase bacteria evolvability and resistance to antibiotics through extensive cassette rearrangements and duplications. These results confirm the role of the integrase in the evolution of antibiotic resistance and give us a starting point for potential strategies to manage infections from integron-bearing bacteria. For example, inducing the excision of other resistance cassettes through integrase activity could be used as a strategy in combination with antibiotic cycling therapy to capitalize on the loss of resistance to these antibiotics (Beardmore and Peña-Miller, 2010). These results also provide another argument in favour of the development of drugs inhibiting

the SOS response, which could be used in combination with antibiotics to reduce the integron benefits (**?**)05Mo2016.

3.6 Acknowledgements

I would like to thank Natalia for her support with the ramping experiment, and Julio and Jess for their help with setting up this bioinformatic pipeline. We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of the sequencing data.

3.7 Supplementary materials

3.7.1 MIC in experimental conditions

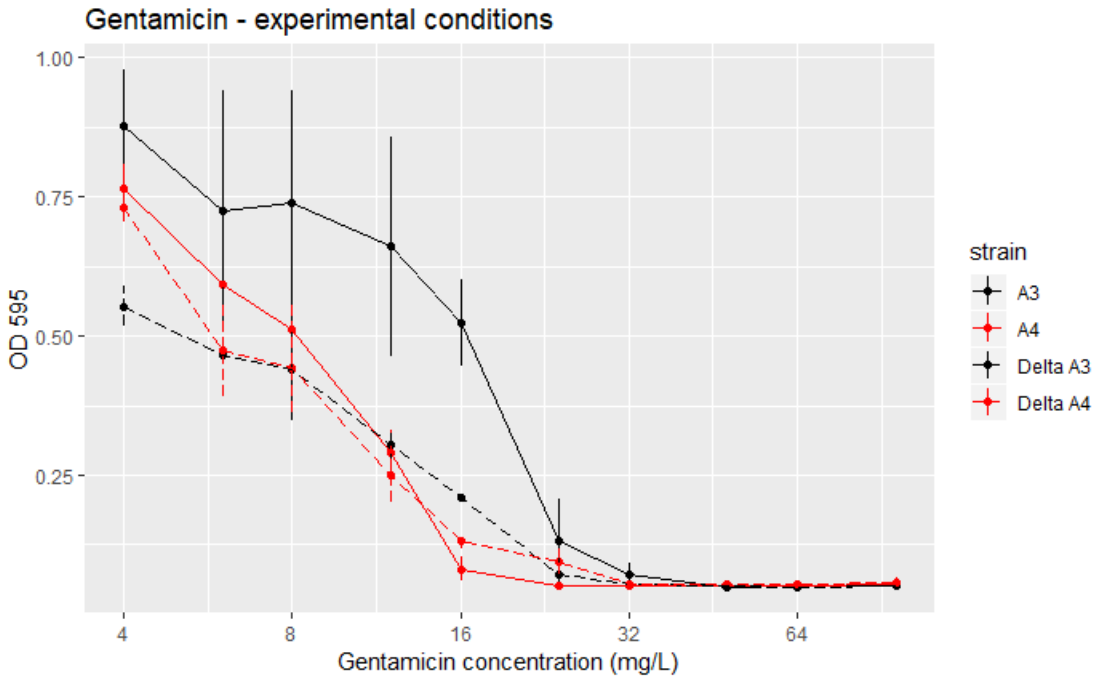


Figure 3.10: Growth of the PAO1:A3, PAO1:A4, PAO1:ΔA3 and PAO1:ΔA4 populations in increasing concentration of gentamicin using an inoculum similar to our ramping experiment. Δ strains are indicated by the dashed lines. The MIC was determined to be for all strains at 24 mg/L.

3.7.2 Recombinations identification by PCR

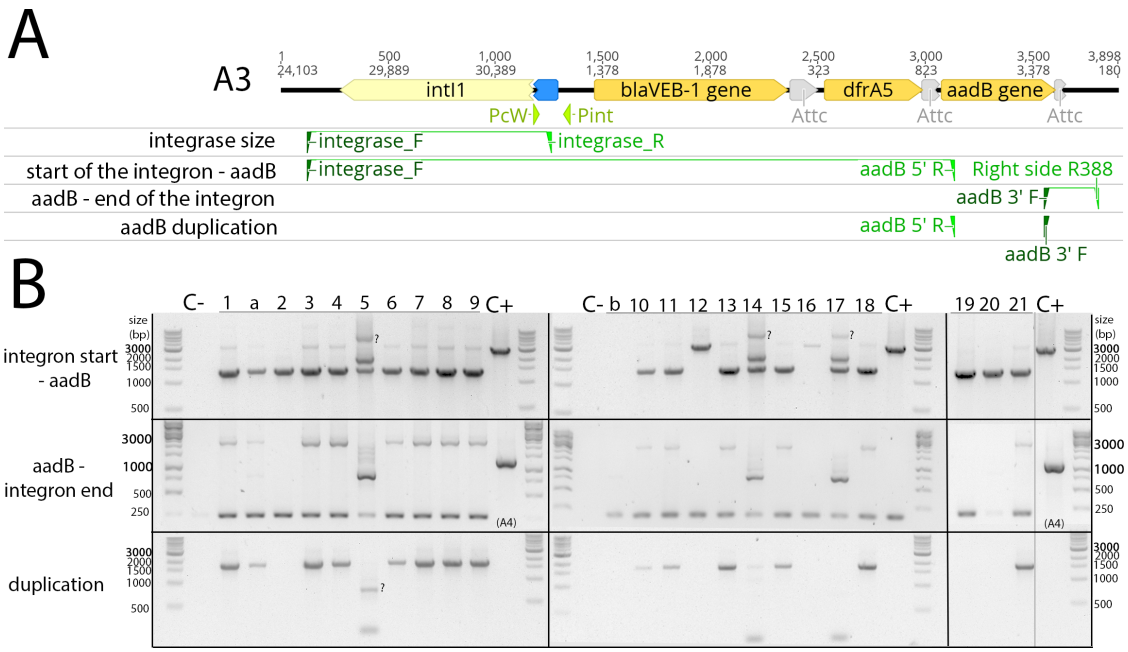


Figure 3.11: A. Primer binding sites and amplicons used in the screen for cassette rearrangements. B. Rearrangement PCR screen of the x1024 WTA3 populations. The population number identifier is indicated at the top. Samples with a letter were excluded from the analysis due to contamination. Bands which cannot be explained by the genomic data are indicated by a question mark. Non-relevant lanes were excluded from the right gel where indicated by the grey vertical lane and care was taken to keep the vertical alignment within the gel picture during the figure composition.

3.7.3 SNP & recombinations at x1024 MIC

Cassette rearrangements

	Junctions determined by breseq							Cassette copy number (CNOG pro)			Proposed array
	Ancestral junctions			Novel junctions							
Sample	intI1/blaVEB1	dfrA5/aadB	aadB/backbone	intI1/blaVEB -> dfrA5/aadB	aadB/backbone -> intI1/blaVEB1	aadB/backbone -> dfrA5/aadB		blaVEB1	dfrA5	aadB	
Control WTA3								1.03 (0.98 - 1.09)	1.20 (1.12 - 1.28)	1.02 (0.94 - 1.10)	ancestral array
Control DeltaA3								1.25 (1.19 - 1.32)	1.23 (1.15 - 1.31)	1.03 (0.96 - 1.10)	ancestral array
WT 1	0	1.28	0.99	1.16	1.11			1.04 (0.99 - 1.09)	1.21 (1.12 - 1.31)	2.19 (2.01 - 2.38)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 2	0			1				0.04 (0.00 - 0.13)	0.00 (0.00 - 0.00)	1.05 (0.94 - 1.14)	intI1 - aadB
WT 3	0	0.83	1.17	1.07	0.98			0.91 (0.87 - 0.96)	1.11 (1.05 - 1.17)	1.97 (1.84 - 2.11)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 4	0	1.27	0.87	1	1.09			1.10 (1.05 - 1.15)	1.16 (1.07 - 1.25)	2.05 (1.94 - 2.16)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 5	0	0	0.84	0.89		1.08		0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	2.00 (1.80 - 2.16)	intI1 - aadB - aadB
WT 6	0	1.17	0.81	1.08	1.11			1.03 (0.99 - 1.07)	1.17 (1.10 - 1.24)	2.01 (1.92 - 2.10)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 7	0	1.08	0.79	1.06	1.22			1.13 (1.07 - 1.19)	1.15 (1.07 - 1.23)	2.04 (1.92 - 2.16)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 8	0	1.27	0.84	1.1	1.22			1.22 (1.16 - 1.28)	1.17 (1.10 - 1.25)	2.17 (2.03 - 2.30)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 9	0	1.19	0.81	1.13	1.03			0.98 (0.94 - 1.02)	1.16 (1.10 - 1.22)	2.00 (1.88 - 2.12)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 10	0	1.24	0.86	1.01	1.2			1.22 (1.12 - 1.34)	1.22 (1.10 - 1.36)	2.04 (1.91 - 2.16)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 11	0	1.26	0.86	0.88	1.2			1.17 (1.06 - 1.27)	1.22 (1.12 - 1.32)	2.07 (1.95 - 2.20)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 12								1.22 (1.16 - 1.28)	1.22 (1.16 - 1.28)	1.47 (1.03 - 2.27)	ancestral array
WT 13	0	1.22	0.81	1.03	1.08			1.12 (1.07 - 1.16)	1.13 (1.06 - 1.21)	2.01 (1.89 - 2.13)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 14	0	0.35	1.02	1.17	0.33	0.76		0.55 (0.49 - 0.62)	0.36 (0.32 - 0.41)	2.33 (2.14 - 2.57)	intI1 - aadB - blaVEB - dfrA5 - aadB + intI1 - aadB - aaB
WT 15	0	1.19	0.71	1.3	1.11			1.24 (1.16 - 1.33)	1.23 (1.13 - 1.33)	2.06 (1.91 - 2.20)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 16								1.24 (1.17 - 1.32)	1.26 (1.14 - 1.39)	1.01 (0.93 - 1.09)	ancestral array
WT 17	0	0	0.91	1		0.91		0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	2.07 (1.86 - 2.24)	intI1 - aadB - aadB
WT 18	0	1.14	0.84	0.93	1.2			1.18 (1.11 - 1.24)	1.16 (1.06 - 1.27)	2.06 (1.92 - 2.19)	intI1 - aadB - blaVEB - dfrA5 - aadB
WT 19	0			1.13				0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	1.07 (0.94 - 1.18)	intI1 - aadB
WT 20	0			1.3				0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	1.35 (1.09 - 1.57)	intI1 - aadB
WT 21	0	1.16	0.86	1.04	1.01			0.98 (0.94 - 1.02)	1.07 (1.00 - 1.13)	1.97 (1.88 - 2.07)	intI1 - aadB - blaVEB - dfrA5 - aadB
Delta 1								1.21 (1.15 - 1.28)	1.18 (1.11 - 1.25)	1.02 (0.95 - 1.09)	ancestral array
Delta 2								1.22 (1.14 - 1.29)	1.28 (1.18 - 1.37)	1.05 (0.98 - 1.14)	ancestral array
Delta 3								1.17 (1.11 - 1.24)	1.20 (1.12 - 1.28)	0.99 (0.93 - 1.05)	ancestral array
Delta 4								1.19 (1.13 - 1.25)	1.22 (1.12 - 1.32)	1.00 (0.92 - 1.09)	ancestral array
Delta 5								1.23 (1.17 - 1.28)	1.25 (1.17 - 1.33)	1.02 (0.94 - 1.10)	ancestral array
Delta 6								1.26 (1.19 - 1.32)	1.27 (1.18 - 1.34)	1.03 (0.97 - 1.10)	ancestral array
WT 1/8								1.26 (1.19 - 1.33)	1.37 (1.26 - 1.47)	1.18 (1.10 - 1.26)	ancestral array
Delta 1/8								1.26 (1.20 - 1.32)	1.19 (1.12 - 1.26)	1.06 (0.98 - 1.12)	ancestral array
WT no AB								0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	no plasmid
Delta no AB								0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	no plasmid

Table 3.7: Results of the cassette rearrangements pipeline for the x1024 MIC populations. For each population and for each novel junction identified by breseq, coverage of the junction relative to the rest of the plasmid backbone is listed as well as the coverage of the ancestral sequence. Copy number of each cassette relative to the plasmid backbone is indicated, as determined by CNOGpro. The parentheses indicate the (0.95 - 1.05) confidence interval. The array deduced from the combination of these informations is indicated on the right.

Plasmid SNPs

Type	Position	Mutation	Gene Name	Effect	Ramping				1/8 MIC	
					Delta (avg)	Delta (n)	WT (avg)	WT (n)	Delta	WT
SNP	31242	G	blaVEB-1 gene	non synonymous	1.00	1	1.00	1	0.00	0.00
DEL	31273	1	blaVEB-1 gene	coding	1.00	1	1.00	1	0.00	0.11
INS	31211	T	blaVEB-1 gene	coding	1.00	2	0.60	5	0.31	0.08
SNP	31231	A	blaVEB-1 gene	nonsense	0.00	0	1.00	3	0.00	0.73
INS	31217	T	blaVEB-1 gene	coding	0.00	0	0.35	5	0.00	0.00
SNP	31172	T	blaVEB-1 gene	non synonymous	0.00	0	1.00	1	0.17	0.00
INS	31180	A	blaVEB-1 gene	coding	1.00	1	0.00	0	0.00	0.00
DEL	31186	1	blaVEB-1 gene	coding	0.00	0	1.00	1	0.00	0.00
SNP	31266	T	blaVEB-1 gene	nonsense	0.00	0	1.00	1	0.00	0.00
INS	31259	A	blaVEB-1 gene	coding	0.00	0	0.85	1	0.15	0.00
DEL	31185	1	blaVEB-1 gene	coding	0.00	0	0.69	1	0.00	0.00
DEL	31181	1	blaVEB-1 gene	coding	0.00	0	0.30	1	0.00	0.00
INS	31263	A	blaVEB-1 gene	coding	0.00	0	0.00	0	0.00	0.00
DEL	31283	1	blaVEB-1 gene	coding	0.00	0	0.00	0	0.00	0.06
SNP	31437	T	blaVEB-1 gene	nonsense	0.00	0	0.00	0	0.17	0.00
SNP	31848	T	blaVEB-1 gene	nonsense	0.00	0	0.00	0	0.18	0.00
DEL	32778	1	dfrA5/aadB gene	intergenic	1.00	2	0.00	0	0.00	0.00
DEL	32741	41	dfrA5/aadB gene	intergenic	0.00	0	1.00	1	0.00	0.00
SNP	32778	G	dfrA5/aadB gene	intergenic	1.00	1	0.00	0	0.00	0.00
SNP	32779	A	dfrA5/aadB gene	intergenic	1.00	1	0.00	0	0.00	0.00
SNP	32779	T	dfrA5/aadB gene	intergenic	0.00	0	0.00	0	0.00	0.00
DEL	1275	4	hypothetical protein CDS	coding	0.00	0	0.00	0	0.00	0.00
SNP	746	A	hypothetical protein CDS	non synonymous	0.00	0	0.00	0	0.00	0.12
SNP	746	T	hypothetical protein CDS	non synonymous	0.00	0	0.00	0	0.16	0.00
DEL	31139	62	dfrA5/aadB	intergenic	0.00	0	1.00	1	0.00	0.00
DEL	33362	8979	aadB/trwJ	coding	0.00	0	1.00	1	0.00	0.09
DEL	33362	8621	aadB/trwJ	coding	0.00	0	0.00	0	0.00	0.29

Table 3.8: List of plasmid SNPs and indels in the PAO1:A3 populations at x1024 MIC. For the ramping populations (avg) represents the average frequency of the mutation in the populations containing this variant. Mutations with an average of 1 swept completely through the population. (n) represents the number of populations in which the mutation is present. For the 1/8 MIC populations the number indicated represents the mutation frequency found in the equimolar pool of six separate populations. For reference a mutation which swept through one population but is absent in the other 5 populations will have a frequency of 0.17 in the pooled population.

Chromosomal mutations

Table 3.9 List of chromosomal SNPs and indels in the A3 populations at x1024 MIC. For the ramping populations (avg) represents the average frequency of the mutation in the populations containing this variant. For reference, mutations with an average frequency of 1 swept completely through the population. (n) represents the number of populations in which the mutation is present. For each variant found in the ramping populations the frequency of the corresponding mutant in the pooled 1/8MIC and no antibiotic controls is indicated as reference.

Type	Position	Mutation	Gene Name	Effect	Ramping				1/8 MIC		No Antibiotic		Gene product
					Delta (avg)	Delta (n)	WT (avg)	WT (n)	Delta	WT	Delta	WT	
DEL	5852365	12	amgS	coding	1.00	1	1.00	1	0	0	0	0	protein AmgS
SNP	4962241	G	sspA/PA4429	intergenic	0.38	1	0.33	2	0	0	0	0	stringent starvation protein A/cytochrome C1
SNP	5852197	C	amgS	non synonymous	0.00	0	1.00	3	0	0	0	0	protein AmgS
INS	5851829	CAGGCG	amgS	coding	0.00	0	1.00	1	0	0	0	0	protein AmgS
INS	5852255	CATCTG	amgS	coding	0.00	0	1.00	1	0	0	0	0	protein AmgS
DEL	5852414	3	amgS	coding	0.00	0	1.00	1	0	0	0	0	protein AmgS
DEL	5852404	9	amgS	coding	0.00	0	0.89	1	0	0	0	0	protein AmgS
SNP	5852051	A	amgS	non synonymous	0.00	0	0.91	1	0	0	0	0	protein AmgS
DEL	5852389	3	amgS	coding	0.00	0	0.87	1	0	0	0	0	protein AmgS
SNP	5852180	T	amgS	non synonymous	0.00	0	0.78	1	0	0	0	0	protein AmgS
DEL	2212767	11	PA2020	coding	1.00	1	0.00	0	0	0	0	0	transcriptional regulator
DEL	2212855	15	PA2020	coding	0.00	0	1.00	1	0	0	0	0	transcriptional regulator
DEL	2212859	70	PA2020	coding	0.00	0	1.00	1	0	0	0	0	transcriptional regulator
DEL	2212891	10	PA2020	coding	1.00	1	0.00	0	0	0	0	0	transcriptional regulator
DEL	2213041	11	PA2020	coding	1.00	1	0.00	0	0	0	0	0	transcriptional regulator
DEL	2213056	14	PA2020	coding	0.00	0	1.00	1	0	0	0	0	transcriptional regulator
DEL	2213133	6	PA2020	coding	0.00	0	1.00	1	0	0	0	0	transcriptional regulator
INS	2213158	C	PA2020	coding	0.00	0	1.00	1	0	0	0	0	transcriptional regulator
INS	2213159	TAT	PA2020	coding	0.00	0	1.00	1	0	0	0	0	transcriptional regulator
DEL	2213175	13	PA2020	coding	0.00	0	1.00	1	0	0	0	0	transcriptional regulator
DEL	2213272	6	PA2020	coding	0.00	0	1.00	1	0	0	0	0	transcriptional regulator
DEL	5090956	12	rluD	coding	0.00	0	1.00	1	0	0	0	0	pseudouridine synthase
SNP	5090980	G	rluD	non synonymous	0.00	0	1.00	1	0	0	0	0	pseudouridine synthase
INS	5091163	G	rluD	coding	1.00	1	0.00	0	0	0	0	0	pseudouridine synthase
DEL	5091317	11	rluD	coding	0.00	0	1.00	1	0	0	0	0	pseudouridine synthase
DEL	5091317	18	rluD	coding	0.00	0	1.00	1	0	0	0	0	pseudouridine synthase
DEL	5091355	1	rluD	coding	1.00	1	0.00	0	0	0	0	0	pseudouridine synthase
INS	5091355	GGGCCTCC	rluD	coding	1.00	1	0.00	0	0	0	0	0	pseudouridine synthase
DEL	5091370	1	rluD	coding	0.00	0	1.00	1	0	0	0	0	pseudouridine synthase
INS	5091640	GGACG	rluD	coding	0.00	0	1.00	1	0	0	0	0	pseudouridine synthase
DEL	5091653	11	rluD	coding	0.00	0	1.00	1	0	0	0	0	pseudouridine synthase
DEL	5091769	1	rluD	coding	0.00	0	1.00	1	0	0	0	0	pseudouridine synthase
SNP	1692578	C	ccoN1	synonymous	0.00	0	0.68	1	0	0	0	0	cbb3-type cytochrome C oxidase subunit I
SNP	1692584	C	ccoN1	synonymous	0.00	0	0.67	1	0	0	0	0	cbb3-type cytochrome C oxidase subunit I
SNP	1692612	C	ccoN1	non synonymous	0.00	0	0.59	1	0	0	0	0	cbb3-type cytochrome C oxidase subunit I
SNP	1692610	A	ccoN1	synonymous	0.00	0	0.58	1	0	0	0	0	cbb3-type cytochrome C oxidase subunit I
SNP	1692622	G	ccoN1	synonymous	0.00	0	0.47	1	0	0	0	0	cbb3-type cytochrome C oxidase subunit I
SNP	1692623	G	ccoN1	non synonymous	0.00	0	0.43	1	0	0	0	0	cbb3-type cytochrome C oxidase subunit I
INS	785686	TCCCCGCx9	phrD/PA0715	intergenic	0.00	0	0.31	1	0	0	0	0	PhrD/hypothetical protein
SNP	1334556	A	PA1232	non synonymous	0.00	0	1.00	1	0	0	0	0	hypothetical protein
SNP	1334556	G	PA1232	non synonymous	0.00	0	1.00	1	0	0	0	0	hypothetical protein
SNP	1951104	C	parS	non synonymous	0.00	0	1.00	1	0	0	0	0	two-component sensor ParS
SNP	1951335	G	parS	non synonymous	0.00	0	1.00	1	0	0	0	0	two-component sensor ParS
SNP	1696070	A	ccoN2	synonymous	0.00	0	0.54	1	0	0	0	0	cbb3-type cytochrome C oxidase subunit I
SNP	1696067	C	ccoN2	synonymous	0.00	0	0.52	1	0	0	0	0	cbb3-type cytochrome C oxidase subunit I
SNP	1696061	G	ccoN2	synonymous	0.00	0	0.45	1	0	0	0	0	cbb3-type cytochrome C oxidase subunit I
SNP	1696058	G	ccoN2	synonymous	0.00	0	0.42	1	0	0	0	0	cbb3-type cytochrome C oxidase subunit I
INS	2212665	G	PA2019/PA2020	intergenic	0.00	0	0.77	1	0	0	0	0	multidrug efflux lipoprotein/transcriptional regulator
INS	2212665	A	PA2019/PA2020	intergenic	0.00	0	0.77	1	0	0	0	0	multidrug efflux lipoprotein/transcriptional regulator
INS	1671363	CCGAGC	dnaX	coding	0.51	1	0.00	0	0	0	0	0	DNA polymerase III subunits gamma/tau
SNP	4771871	T	rpsL	non synonymous	0.00	0	1.00	1	0	0	0	0	30S ribosomal protein S12
DEL	1100709	12	PA1016	coding	0.00	0	1.00	1	0	0	0	0	thiolase
DEL	1147499	12	PA1059	coding	0.00	0	1.00	1	0	0	0	0	monovalent cation/H+ antiporter subunit G
SNP	2493497	C	PA2264	non synonymous	0.00	0	1.00	1	0	0	0	0	hypothetical protein
SNP	2906109	C	PA2570.1/PA2571	intergenic	1.00	1	0.00	0	0	0	0	0	tRNA-Leu/two-component sensor
DEL	2984581	131	nuoD	coding	1.00	1	0.00	0	0	0	0	0	NADH:-quinone oxidoreductase subunit C/D
DEL	2989872	2	nuoG	coding	1.00	1	0.00	0	0	0	0	0	NADH-quinone oxidoreductase subunit G
DEL	2991893	10	nuoI	coding	1.00	1	0.00	0	0	0	0	0	NADH-quinone oxidoreductase subunit I
DEL	2995699	1	nuoM	coding	1.00	1	0.00	0	0	0	0	0	NADH-quinone oxidoreductase subunit M
DEL	3334907	10	rne	coding	0.00	0	1.00	1	0	0	0	0	ribonuclease E
SNP	3799585	C	nosD	non synonymous	0.00	0	1.00	1	0	0	0	0	copper-binding periplasmic protein
SNP	4772491	A	rpoC	non synonymous	1.00	1	0.00	0	0	0	0	0	DNA-directed RNA polymerase subunit beta'
SNP	5116300	T	rpmA/rplU	intergenic	0.00	0	1.00	1	0	0	0	0	50S ribosomal protein L27/50S ribosomal protein L21
DEL	5240106	12	PA4671/PA4672	intergenic	0.00	0	1.00	1	0	0	0	0	50S ribosomal protein L25/general stress protein Ctc/ peptidyl-tRNA hydrolase
INS	5269237	C	PA4690.5	non coding	0.00	0	1.00	1	0	0	0	0	16S ribosomal RNA
DEL	5707756	12	tatC	coding	0.00	0	1.00	1	0	0	0	0	transporter TatC
DEL	6160737	69	PA5471/PA5472	intergenic	0.00	0	1.00	1	0	0	0	0	hypothetical protein/hypothetical protein
DEL	6257044	298	gidB	coding	1.00	1	0.00	0	0	0	0	0	gidB protein
DEL	6264325	1	rpmH	coding	1.00	1	0.00	0	0	0	0	0	50S ribosomal protein L34
SNP	3548495	T	rpsA	non synonymous	0.00	0	0.82	1	0	0	0	0	30S ribosomal protein S1
SNP	4784553	C	tufB	synonymous	0.00	0	0.94	1	0	0	0	0	elongation factor Tu
SNP	786915	T	PA0715	non synonymous	0.00	0	0.93	1	0	0	0	0	hypothetical protein
SNP	4666412	A	PA4169	non synonymous	0.00	0	0.53	1	0	0	0	0	hypothetical protein
SNP	5466832	T	PA4869	nonsense	0.00	0	0.53	1	0	0	0	0	hypothetical protein
DEL	726217	13	PA0668.4	coding	0.00	0	1.00	1	0	0	0	0	23S ribosomal RNA
DEL	2212868	292	PA2020	coding	0.00	0	0.93	1	0	0	0	0	transcriptional regulator
DEL	724412	14	PA0668.4	coding	0.00	0	1.00	1	0	0	0	0	23S ribosomal RNA

Table 3.9

Summary statistics

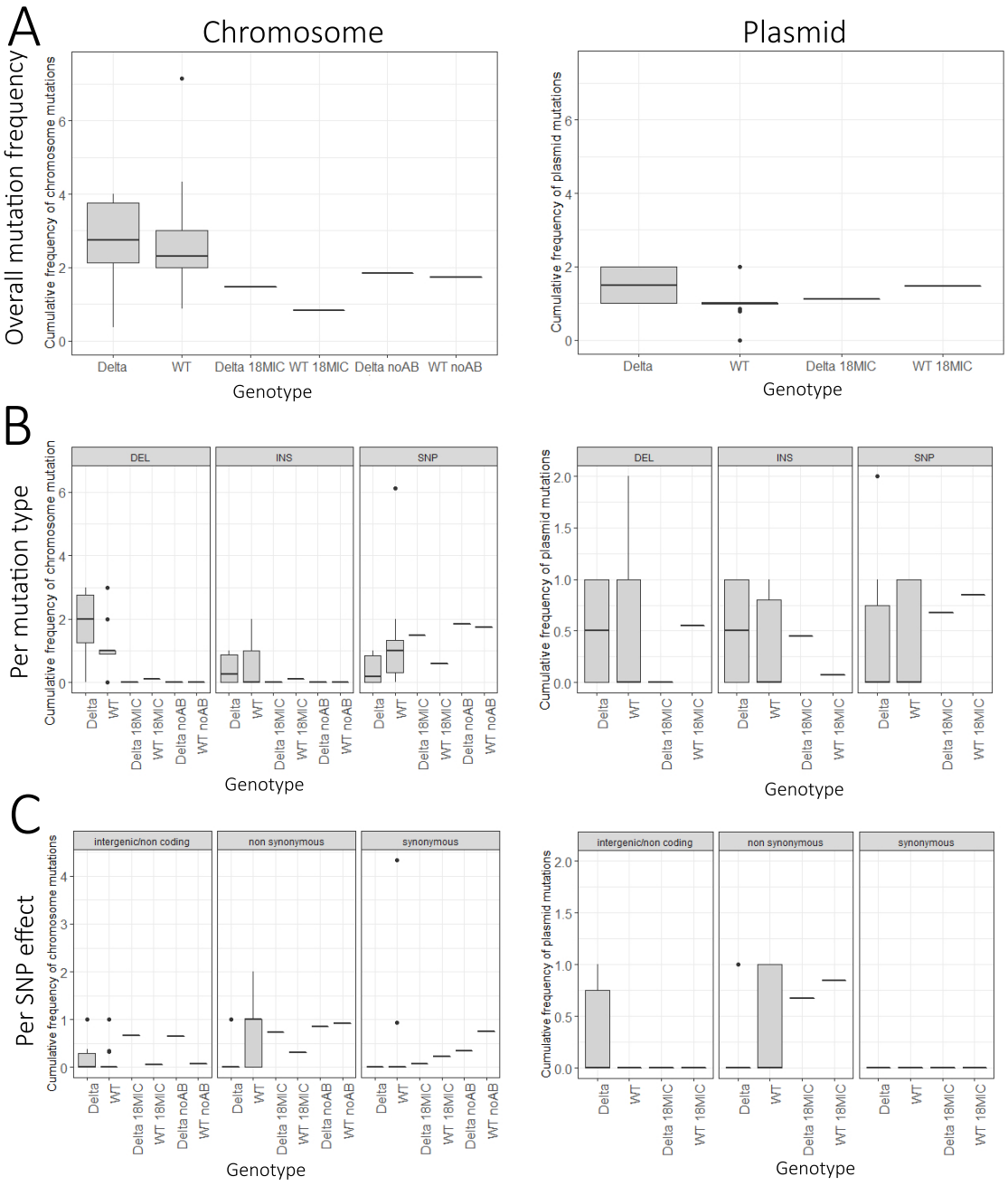


Figure 3.12: Box plots represent the average cumulative mutation frequencies for each population at the x1024 MIC time-point for A. all mutations B. per mutation type C. per mutation effect (SNP only). The lower and upper hinges correspond to the first and third quartiles while the middle line corresponds to the median. Cassette rearrangements (duplications and deletions) are not included.

3.7.4 SNPs & recombinations at x4 MIC

Arrays rearrangements

Table 3.10: List of plasmid SNPs and indels in the PAO1:A3 populations at x1024 MIC. For the ramping populations (avg) represents the average frequency of the mutation in the populations containing this variant. Mutations with an average of 1 swept completely through the population. (n) represents the number of populations in which the mutation is present. For the 1/8 MIC populations the number indicated represents the mutation frequency found in the equimolar pool of six separate populations. For reference a mutation which swept through one population but is absent in the other 5 populations will have a frequency of 0.17 in the pooled population.

	Ancestral junctions			Novel recombinations				Copy number (CNOGpro)			Proposed array
Sample	intI1/blaVEB1	dfrA5/aadB	aadB/backbone	intI1/blaVEB1 -> dfrA5/aadB	aadB/backbone -> intI1/blaVEB1	aadB/backbone -> dfrA5/aadB	dfrA5/aadB -> aadB/backbone	blaVEB1	dfrA5	aadB	
Delta 1								1.11 (0.93 - 1.43)	1.19 (1.12 - 1.26)	1.02 (0.94 - 1.09)	ancestral array
Delta 2								1.13 (0.94 - 1.48)	1.13 (1.07 - 1.19)	0.99 (0.92 - 1.06)	ancestral array
Delta 3								1.12 (1.05 - 1.18)	1.21 (1.14 - 1.28)	1.04 (0.94 - 1.14)	ancestral array
Delta 4								1.14 (1.10 - 1.19)	1.20 (1.13 - 1.27)	1.11 (1.04 - 1.18)	ancestral array
Delta 5								1.16 (1.11 - 1.21)	1.22 (1.14 - 1.29)	1.02 (0.96 - 1.08)	ancestral array
Delta 6								0.76 (0.70 - 0.82)	1.06 (0.99 - 1.14)	1.03 (0.97 - 1.09)	ancestral array
Delta 7								0.93 (0.88 - 0.97)	1.16 (1.10 - 1.23)	0.99 (0.93 - 1.04)	ancestral array
Delta 8								1.04 (1.01 - 1.08)	1.22 (1.14 - 1.29)	1.09 (1.02 - 1.16)	ancestral array
Delta 9								1.09 (1.05 - 1.13)	1.22 (1.14 - 1.31)	0.97 (0.91 - 1.04)	ancestral array
Delta 10								1.07 (1.03 - 1.12)	1.20 (1.13 - 1.27)	1.00 (0.94 - 1.05)	ancestral array
Delta 11								1.28 (1.21 - 1.34)	1.28 (1.19 - 1.39)	1.03 (0.97 - 1.10)	ancestral array
Delta 12								1.16 (1.11 - 1.20)	1.17 (1.12 - 1.23)	1.03 (0.96 - 1.10)	ancestral array
Delta 13								0.79 (0.74 - 0.84)	1.14 (1.05 - 1.25)	1.00 (0.94 - 1.07)	ancestral array
Delta 14								0.97 (0.81 - 1.30)	1.11 (1.05 - 1.16)	1.02 (0.96 - 1.08)	ancestral array
Delta 15								0.95 (0.90 - 1.00)	1.17 (1.12 - 1.24)	0.99 (0.91 - 1.07)	ancestral array
Delta 16								1.10 (1.06 - 1.15)	1.19 (1.13 - 1.24)	1.03 (0.96 - 1.10)	ancestral array
Delta 17								1.25 (1.19 - 1.31)	1.24 (1.15 - 1.33)	1.01 (0.93 - 1.08)	ancestral array
Delta 18								1.16 (1.11 - 1.22)	1.22 (1.15 - 1.30)	1.03 (0.97 - 1.09)	ancestral array
Delta 19								1.08 (1.03 - 1.13)	1.17 (1.11 - 1.24)	1.04 (0.99 - 1.10)	ancestral array
Delta 20								1.06 (1.02 - 1.10)	1.21 (1.13 - 1.29)	1.08 (1.02 - 1.13)	ancestral array
Delta 21								1.11 (1.07 - 1.16)	1.24 (1.17 - 1.32)	0.97 (0.91 - 1.05)	ancestral array
Delta 22								0.92 (0.86 - 0.97)	1.16 (1.10 - 1.22)	1.04 (0.97 - 1.11)	ancestral array
WT 1								1.06 (1.02 - 1.10)	1.27 (1.19 - 1.36)	1.06 (0.99 - 1.14)	ancestral array
WT 2	0	1.11	0.86	0.99	1.07	0.06		1.10 (1.05 - 1.15)	1.05 (0.98 - 1.12)	2.01 (1.90 - 2.11)	intI1 - aadB - dfrA5 - blaVEB - aadB
WT 3	0	0.6	0.89	1.02	0.6			0.77 (0.73 - 0.81)	0.66 (0.62 - 0.71)	1.64 (1.54 - 1.74)	intI1 - aadB - dfrA5 - blaVEB - aadB + intI1 - aadB
WT 4								1.10 (1.04 - 1.16)	1.19 (1.11 - 1.27)	0.97 (0.92 - 1.04)	ancestral array
WT 5	0.02	0.4	0.9	1.02	0.35	0.41		0.61 (0.54 - 0.67)	0.44 (0.40 - 0.49)	1.85 (1.71 - 2.03)	intI1 - aadB - dfrA5 - blaVEB - aadB + intI1 - aadB - aadB
WT 6								1.16 (1.10 - 1.21)	1.14 (1.07 - 1.21)	0.95 (0.88 - 1.03)	ancestral array
WT 7	0	0.23	0.9	1.1	0.2			0.44 (0.35 - 0.53)	0.25 (0.21 - 0.32)	1.28 (1.10 - 1.56)	intI1 - aadB - dfrA5 - blaVEB - aadB + intI1 - aadB
WT 8	0.88	1.22	0.88	0.25	0.25			1.00 (0.96 - 1.05)	1.14 (1.07 - 1.21)	1.27 (1.20 - 1.35)	ancestral array + intI1 - aadB - dfrA5 - blaVEB - aadB
WT 9	0.98	1.19	0.89	0.34	0.38			1.24 (1.18 - 1.30)	1.21 (1.13 - 1.30)	1.30 (1.24 - 1.36)	ancestral array + intI1 - aadB - dfrA5 - blaVEB - aadB
WT 10	0	1.16	0.82	1.01	1.14			1.17 (1.11 - 1.22)	1.13 (1.06 - 1.21)	2.04 (1.93 - 2.16)	int1 - aadB - blaVEB - dfrA- aadB
WT 11		1.24	0.84			0.9		1.07 (1.02 - 1.11)	1.19 (1.15 - 1.24)	1.97 (1.84 - 2.10)	int1- blaVEB - dfrA - aadB - addB
WT 12	0.19	1.13	0.86	0.15	0.19	0.78		1.11 (1.06 - 1.16)	1.15 (1.06 - 1.22)	1.97 (1.86 - 2.09)	ambiguous
WT 13		1.22	0.92			0.67		1.00 (0.96 - 1.05)	1.14 (1.09 - 1.19)	1.76 (1.64 - 1.86)	ancestral array + int1 - blaVEB - dfrA - aadB - aadB
WT 14								1.08 (1.04 - 1.13)	1.13 (1.07 - 1.19)	1.33 (0.97 - 1.99)	ancestal array
WT 15								1.31 (1.07 - 1.73)	1.23 (1.15 - 1.32)	1.06 (1.00 - 1.13)	ancestral array
WT 16		0	0.9			0.08		0.00 (0.00 - 0.00)	0.00 (0.00 - 0.00)	1.11 (1.00 - 1.19)	int1 - aadB + int1 - aadB - aadB
WT 17								1.20 (1.14 - 1.25)	1.13 (1.07 - 1.19)	1.01 (0.95 - 1.07)	ancestral array
WT 18	0.01	0.19	0.87	1.06	0.19	0.52	0.06	0.45 (0.36 - 0.54)	0.27 (0.23 - 0.34)	1.96 (1.69 - 2.40)	int1 - aadB - blaVEB - dfrA - aadB + int1 - aadB - aadB
WT 19	0	0.69	0.91	1.01	0.59	0.35		0.72 (0.69 - 0.76)	0.64 (0.60 - 0.67)	1.96 (1.84 - 2.08)	int1 - aadB - blaVEB - dfrA - aadB + int1 - aadB - aadB
WT 20	0	0.05	0.85	1.06	0.05			0.29 (0.17 - 0.42)	0.08 (0.05 - 0.12)	1.85 (0.92 - 3.66)	int1 - aadB - blaVEB - dfrA - aadB + int1 - aadB
WT 21	0.01	0.51	0.84	1.05	0.52		0.04	0.62 (0.56 - 0.68)	0.57 (0.52 - 0.62)	1.51 (1.38 - 1.67)	int1 - aadB - blaVEB - dfrA - aadB + int1 - aadB
WT 22	0	1.26	0.89	1.04	1.11	0.19		1.01 (0.97 - 1.05)	1.17 (1.10 - 1.25)	2.27 (2.15 - 2.39)	ambiguous
WT 23	0.03	1.18	0.82	0.96	1.08			1.02 (0.97 - 1.06)	1.12 (1.05 - 1.20)	1.99 (1.88 - 2.09)	int1 - aadB - blaVEB - dfrA- aadB
WT 24								1.03 (0.97 - 1.08)	1.19 (1.12 - 1.25)	1.01 (0.94 - 1.07)	ancestral array

Table 3.10

Plasmid mutations

Type	Position	Mutation	Gene Name	Effect	Ramping			
					Delta (avg)	Delta (n)	WT (avg)	WT (n)
INS	31440	A	blaVEB-1 gene	coding	1.00	3	1.00	1
SNP	31242	G	blaVEB-1 gene	non synonymous	1.00	1	1.00	1
DEL	31430	11	blaVEB-1 gene	coding	1.00	1	1.00	1
SNP	31231	A	blaVEB-1 gene	nonsense	0.83	2	1.00	6
DEL	31283	1	blaVEB-1 gene	coding	0.77	2	0.94	1
INS	31211	T	blaVEB-1 gene	coding	1.00	6	0.67	4
SNP	31266	T	blaVEB-1 gene	nonsense	0.00	0	1.00	1
SNP	31172	A	blaVEB-1 gene	non synonymous	0.00	0	1.00	1
SNP	31172	T	blaVEB-1 gene	non synonymous	0.00	0	1.00	1
SNP	31211	T	blaVEB-1 gene	non synonymous	0.00	0	1.00	1
DEL	31340	11	blaVEB-1 gene	coding	1.00	1	0.00	0
SNP	31437	T	blaVEB-1 gene	nonsense	1.00	1	0.00	0
INS	31778	A	blaVEB-1 gene	coding	0.00	0	1.00	1
SNP	31239	T	blaVEB-1 gene	nonsense	0.00	0	0.77	1
INS	31207	T	blaVEB-1 gene	coding	0.00	0	0.75	1
DEL	31185	1	blaVEB-1 gene	coding	0.00	0	0.64	1
DEL	31186	1	blaVEB-1 gene	coding	0.00	0	0.53	1
INS	31217	T	blaVEB-1 gene	coding	0.00	0	0.38	1
DEL	31181	1	blaVEB-1 gene	coding	0.00	0	0.35	1
INS	31167	ATATGAA	intI1/blaVEB-1 gene	intergenic	1.00	1	0.00	0
DEL	32741	41	dfrA5/aadB gene	intergenic	0.00	0	1.00	1
SNP	32779	A	dfrA5/aadB gene	intergenic	0.67	1	0.00	0

Table 3.11: List of plasmid SNPs and indels in the PAO1:A3 populations at x4 MIC. For the ramping populations (avg) represents the average frequency of the mutation in the populations containing this variant. Mutations with an average of 1 swept completely through the population. (n) represents the number of populations in which the mutation is present.

Chromosomal mutations

Type	Position	Mutation	Gene Name	Effect	Ramping				Gene product
					Delta (avg)	Delta (n)	WT (avg)	WT (n)	
SNP	4962241	G	sspA/PA4429	intergenic	0.36	1	0.33	3	stringent starvation protein A/cytochrome C1
SNP	5091142	C	rluD	non synonymous	1.00	1	0.00	0	pseudouridine synthase
INS	5091163	G	rluD	coding	1.00	1	0.00	0	pseudouridine synthase
SNP	5091581	A	rluD	nonsense	0.00	0	1.00	1	pseudouridine synthase
INS	5091111	A	rluD	coding	0.00	0	0.88	1	pseudouridine synthase
SNP	2906123	A	PA2570.1/PA2571	intergenic	1.00	1	0.00	0	tRNA-Leu/two-component sensor
SNP	2906109	C	PA2570.1/PA2571	intergenic	0.77	1	0.00	0	tRNA-Leu/two-component sensor
SNP	4768227	C	tufA	non synonymous	0.77	2	0.00	0	elongation factor Tu
SNP	4767839	A	tufA	non synonymous	0.00	0	0.70	1	elongation factor Tu
DEL	2989872	2	nuoG	coding	1.00	1	0.00	0	NADH-quinone oxidoreductase subunit G
INS	2213159	TAT	PA2020	coding	0.00	0	0.93	1	transcriptional regulator
SNP	2213182	A	PA2020	non synonymous	0.85	1	0.00	0	transcriptional regulator
SNP	2212809	T	PA2020	non synonymous	0.36	1	0.00	0	transcriptional regulator
SNP	1731788	T	sucD	non synonymous	1.00	1	0.00	0	succinyl-CoA ligase subunit alpha
INS	1844903	GTTGGC	pscP	coding	0.00	0	1.00	1	translocation protein in type III secretion
SNP	1951271	G	parS	non synonymous	1.00	1	0.00	0	two-component sensor ParS
SNP	3834101	T	PA3424	non synonymous	1.00	1	0.00	0	hypothetical protein
SNP	5328724	A	infB	non synonymous	1.00	1	0.00	0	translation initiation factor IF-2
SNP	1950755	A	parS	non synonymous	0.00	0	0.57	1	two-component sensor ParS
INS	5549417	C	miaA	coding	0.00	0	1.00	1	tRNA delta(2)-isopentenylpyrophosphate transferase
DEL	5549418	1	miaA	coding	0.00	0	1.00	1	tRNA delta(2)-isopentenylpyrophosphate transferase
SNP	5549083	A	miaA	nonsense	0.89	1	0.00	0	tRNA delta(2)-isopentenylpyrophosphate transferase
SNP	1691011	A	ccoO1	non synonymous	0.94	1	0.00	0	cbb3-type cytochrome C oxidase subunit II
SNP	5311881	C	panB	non synonymous	0.60	1	0.00	0	3-methyl-2-oxobutanoate hydroxymethyltransferase
SNP	2985030	A	nuoD	nonsense	0.89	1	0.00	0	NADH:-quinone oxidoreductase subunit C/D
DEL	3563610	18	PA3173	coding	0.87	1	0.00	0	short-chain dehydrogenase
DEL	1602835	11	ccmA	coding	0.76	1	0.00	0	cytochrome c biogenesis ATP-binding export protein CcmA
SNP	5115994	A	obg/rpmA	intergenic	0.68	1	0.00	0	GTPase ObgE/50S ribosomal protein L27
SNP	4258565	A	PA3799	non synonymous	0.67	1	0.00	0	GTPase Der
INS	2735109	CGTGAT	PA2438	coding	0.00	0	0.65	1	hypothetical protein
SNP	4771991	C	rpsL	non synonymous	0.58	1	0.00	0	30S ribosomal protein S12
SNP	4762509	G	rplN	non synonymous	0.37	1	0.00	0	50S ribosomal protein L14
DEL	4758963	6	rpmD	coding	0.00	0	0.58	1	50S ribosomal protein L30
SNP	4785418	A	tufB	non synonymous	0.00	0	0.54	1	elongation factor Tu
INS	5852255	CATCTG	amgS	coding	0.00	0	0.41	1	protein AmgS
DEL	724470	5	PA0668.4	coding	0.40	1	0.00	0	23S ribosomal RNA
DEL	726019	46	PA0668.4	coding	1.00	1	0.00	0	23S ribosomal RNA
DEL	725416	983	PA0668.4	coding	1.00	1	0.00	0	23S ribosomal RNA
DEL	726217	13	PA0668.4	coding	0.00	0	1.00	1	23S ribosomal RNA

Table 3.12: List of chromosomal SNPs and indels in the PAO1:A3 populations at x4 MIC. For the ramping populations (avg) represents the average frequency of the mutation in the populations containing this variant. For reference, mutations with an average of 1 swept completely through the population. (n) represents the number of populations in which the mutation is present.

Summary statistics

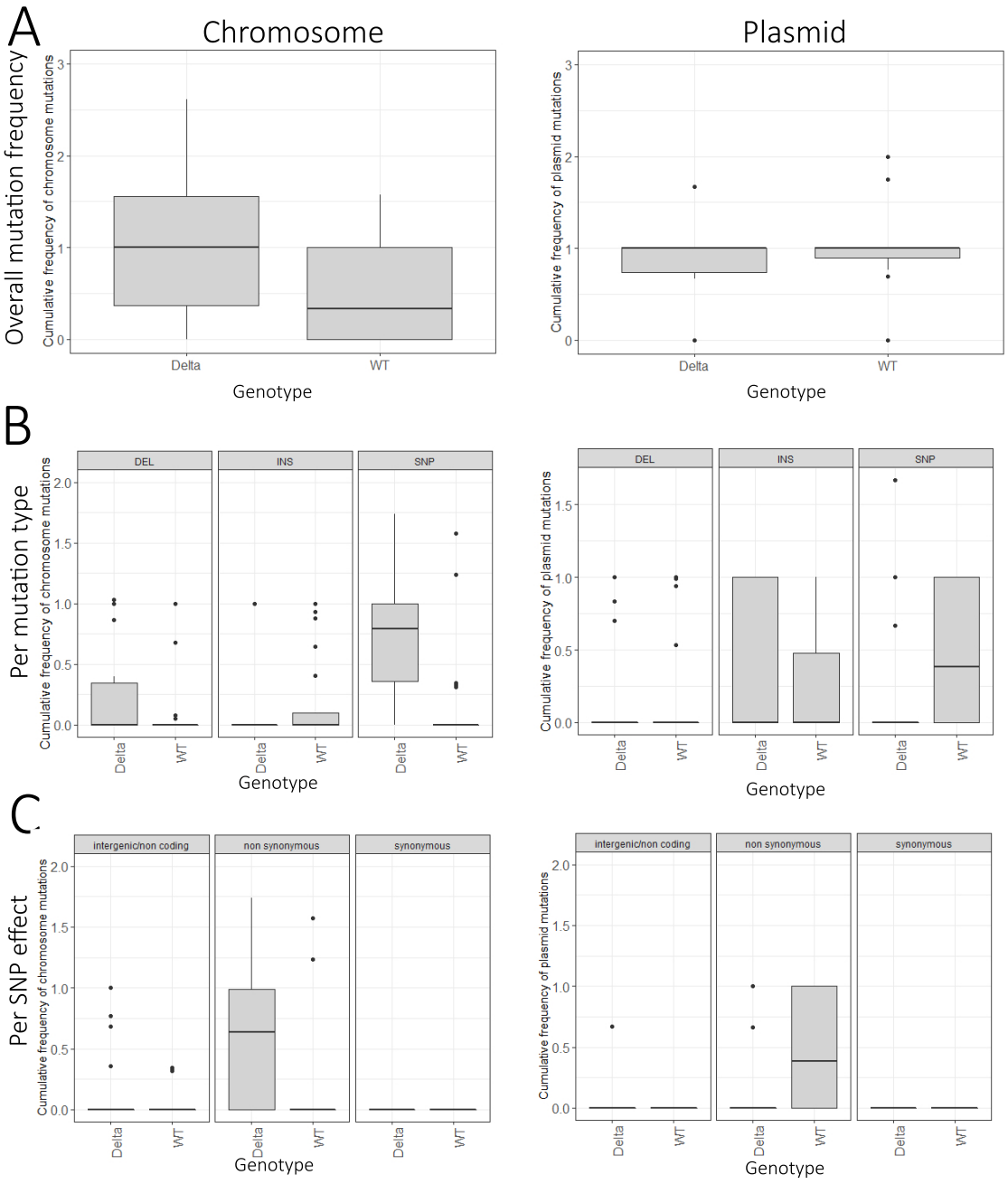


Figure 3.13: Box plots represent the average cumulative mutation frequencies for each population at the x4 MIC time-point for A. all mutations B. per mutation type C. per mutation effect (SNP only). The lower and upper hinges correspond to the first and third quartiles while the middle line corresponds to the median. Cassette rearrangements (duplications and deletions) are not included.

3.7.5 pAMBL experimental evolution

Experimental protocol

Thirty colonies of each strain (PAO1, PAO1:pAMBL1, PAO1:pAMBL2) were inoculated in 100 μL of MH2 broth and transferred every day in doubling concentrations of meropenem with a 1/10 000 dilution and frozen in 15% glycerol every other day. Population survival was monitored each day after reaching a concentration of x1 MIC by plating every well on a MH2 agar plate without antibiotic using a pin replicator. Extinction of a population was defined as the absence of a visible colony after 24h incubation at 37°C.

Surviving populations at x2 MIC were grown on a MH2 agar plate without antibiotic and used as substrate for PCR. PAO1:pAMBL1 populations were controlled for the absence of pAMBL2 DNA and primers were used to identify potential duplications of the *blaVIM* cassette by PCR (Appendices Table 5.2). PCR reactions were carried out using the GoTaq G2 DNA mastermix (Promega) and the following protocol: 30s at 95°C, 30s at 55°C, 3 minutes at 72°C for 30 cycles.

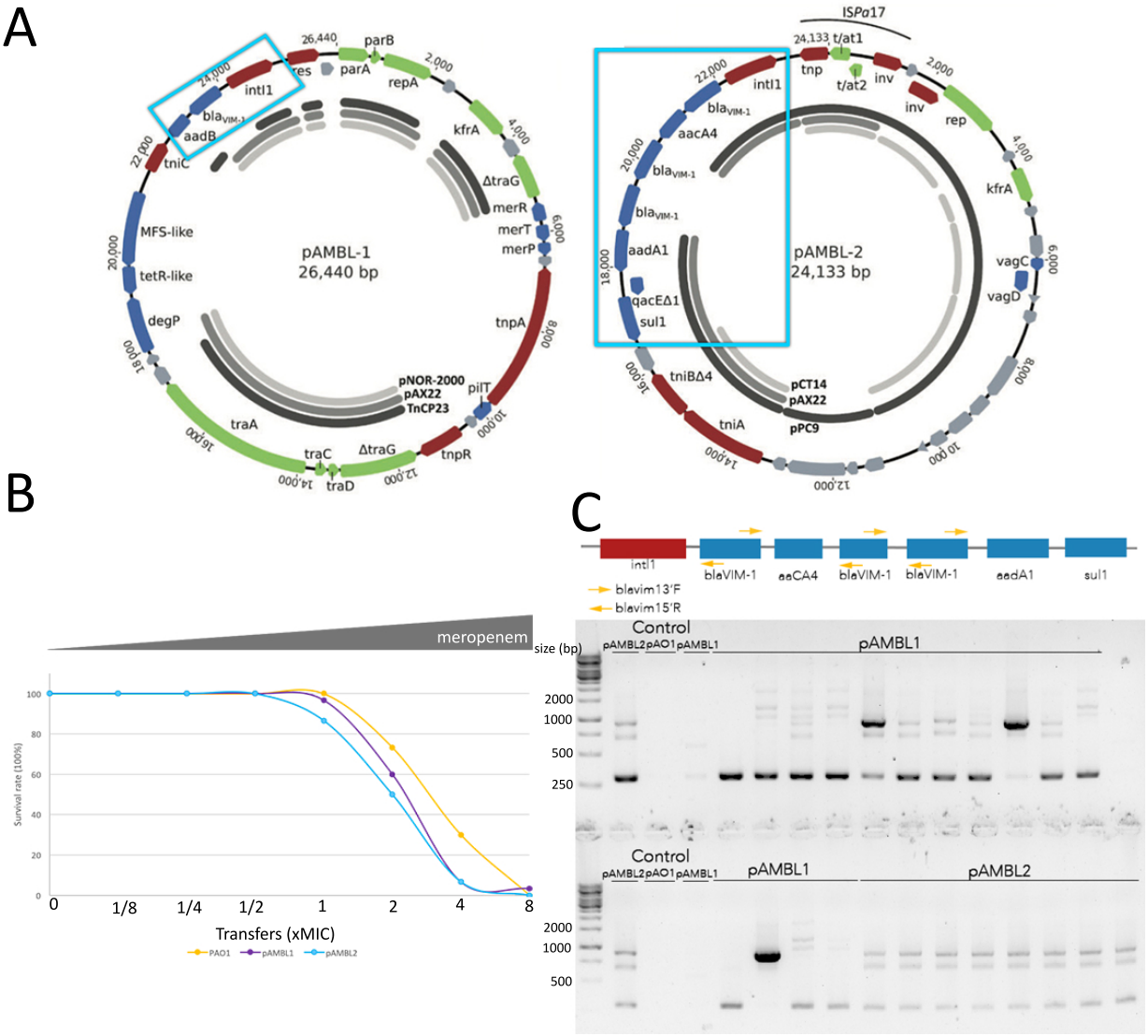


Figure 3.14: **(a)** Schematic representation of the plasmids pAMBL-1 (left) and pAMBL-2 (right). The integron is highlighted with the blue box. Figure modified from (San Millan *et al.*, 2015b). **(b)** Resulting survival curves of the ramping experiment with meropenem. **(c)** Detection of blaVIM duplication by PCR on populations from day 5 (x2 MIC transfer). Top : schematic representation of pAMBL-2 with the expected position of the PCR primers. The primers face the opposite direction from each other, providing an amplicon only in the presence of blaVIM duplications.

3.7.6 Cassette duplications in INTEGRALL

Duplicated cassette	Arrays count
aacA29b	1
aacA29e	1
aacA4	7
aacA4'-17	1
aacA4'-3	5
aacA4-5	2
aacA48	1
aacA7	3
aadA1a	1
aadA1b	5
aadA2	1
aadA7	1
aadB	5
blaIMP-1	1
blaOXA-2	3
blaVIM-1	1
blaVIM-4	1
fosL	1
gcu110	1
gcu114	1
gcu2b	1
gcuD	1
gcuP	4
qacG	1

Table 3.13: Cassettes found in duplicates within the arrays listed in the INTEGRALL database. 'Arrays count' represents the number of different arrays in which a duplicated cassette was observed.

4

Experimental evolution with *blaVEB-1*: costly integrase and plasmid rearrangements

4.1 Abstract

Integrons can act as genetic shuffling device and modulate the integron cassette expression levels by re-ordering the array cassette order. As we saw in the previous chapter, integrase activity can increase bacteria evolvability against antibiotics through extensive cassette rearrangements and duplications. However, this benefit may vary depending on the cassette characteristics, such as its mobility or its expression gradient across the array. In this chapter, we quantify the adaptive benefits of shuffling another cassette, *blaVEB-1*, which is characterized by a weak expression gradient and a low recombinogenic potential, against increasing concentrations of piperacillin. We observe a deleterious effect of integrase activity on evolvability combined with reduced cassette arrangements, but identify extensive rearrangements in the plasmid conjugation machinery, including integrase-specific inversions. This demonstrates the benefits of integrase expression may highly depend on the array's cassette mobility, as well as the potential for integrase activity to have a wider impact on the rest of the genome through non-canonical recombinations.

4.2 Introduction

With more than 130 different cassettes identified in class 1 integrons, integron cassettes provide resistance against nearly all antibiotic families (Escudero *et al.*, 2015). These cassettes can greatly differ in their characteristics. Cassettes in mobile integrons present a variety of *attC* sites with almost no sequence identity but a conserved palindromic structure. Depending on their structure these cassettes have differing degree of mobilisation potential (Larouche and Roy, 2011), with recombination frequencies varying up to 50-fold between *attC* sites (Aubert *et al.*, 2012). Based on the genes they encode, different integron cassettes impose different fitness cost on their host, ranging from negligible to a 17% reduction in host growth (Lacotte *et al.*, 2017). Some cassettes are characterized by the absence of an efficient translation initiation sequence, making their expression reliant on their proximity to an open reading frame overlapping the integration site *attI* (Hanau-Berçot *et al.*, 2002). Mathematical models show (Engelstädter *et al.*, 2016) these characteristics (cassette mobility, cost, expression gradient along the array) can alter the benefits of cassette shuffling which has been predicted to increase bacteria adaptability to changing environments.

We confirmed experimentally in Chapter 3 that integrase activity could increase evolvability for higher levels of gentamicin resistance by reshuffling and duplicating the *aadB* cassette. Given the diversity of cassette characteristics we aim to investigate in this chapter if this result holds true for a different cassette, *blaVEB-1*, and different antibiotic, piperacillin. *blaVEB-1* encodes for an extended spectrum β-lactamase and was isolated for the first time in *E. coli* (Poirel *et al.*, 1999) but is now widespread worldwide in Gram-negative bacteria (Aubert *et al.*, 2012). In comparison to *aadB*, *blaVEB-1* has been shown to possess a much less mobilizable *attC* site (Aubert *et al.*, 2012) and the resistance levels it provides are largely unaffected by cassette position (Chapter 2).

Using the synthetic arrays introduced previously we investigated the benefit of cassette shuffling in an array where *blaVEB-1* is in a position distant from the promoter. While bringing the *blaVEB-1* cassette closer to the start of the array

should only bring a small increase in piperacillin resistance, the integrase still has the potential to increase *blaVEB-1* expression through cassette duplication. Given *blaVEB-1* expression is not strongly impacted by position, the effect of cassette amplification may even be stronger than in the previous chapter, as all *blaVEB-1* copies would be similarly expressed.

Combining experimental evolution and whole genome sequencing we studied the evolution of *P. aeruginosa* populations containing an array with *blaVEB-1* in third position against increasing concentration of piperacillin. Surprisingly we found populations containing a functional integrase had a lower survival rate than population where the integrase was inactivated, with extremely scarce cassette rearrangements. Instead, we observed multiple mutations and rearrangements in the plasmid conjugation machinery as well as novel recombinations due to off-target effects of the integrase. These results confirm the major effect of cassette characteristics on the benefits of cassette shuffling and reinforce the idea that the impact of integrase activity is not limited to integron cassettes but may have wider impact on the rest of the genome.

4.3 Materials & Methods

Strains and growth conditions

The strains and plasmids used in this chapter are listed in Table 4.1. *P. aeruginosa* strains were grown in LB Miller broth or agar supplemented with 100 mg/L of ceftazidime while *E. coli* strains were cultivated with MH broth or agar supplemented with 100 mg/L of carbenicillin.

Experimental evolution

Minimum inhibitory concentration in experimental conditions

Minimum inhibitory concentrations for piperacillin were determined for all strains in conditions matching the experimental evolution set-up: overnight cultures were

Strain/Plasmid	Description	Source
Strain		
PAO1	*P. aeruginosa* strain PAO1	
MG1655	*E. coli* strain MG1655	
Plasmids		
WTA4	Custom integron array *intI1 - dfrA5 - aadB - blaVEB1* on R388 plasmid backbone	Previous chapters
ΔintA4	Array WTA4 with a 948bp deletion of the integrase intI1	Previous chapters
WTA5	Custom integron array *intI1 - aadB - blaVEB1 - dfrA5* on R388 plasmid backbone.	Previous chapters
ΔintA5	Array WTA5 with a 948bp deletion of the integrase intI1	Previous chapters

Table 4.1: Strains used in this chapter

diluted 1/10 000 and supplemented with doubling concentrations of piperacillin. MICs were determined after 20h of incubation. This process was repeated twice. In these conditions the MIC were 64 mg/L of piperacillin for PAO1:WTA4 and PAO1:ΔintA4 and 8 mg/L for MG1655:WTA5 and MG1655:ΔintA5.

Ramping experiment

Similarly to the previous experimental evolution experiment, 90 populations of each strain (PAO1:WTA4, PAO1:ΔintA4, MG1655:WTA5 and MG1655:ΔintA5) were initially inoculated from individual colonies grown on selective agar in either MH2 or MH at a piperacillin concentration of 1/8 MIC. The outer wells of every plate were kept as negative controls and strain distribution among the plates was kept balanced to limit plate effects. Plates were kept at 37°C with shaking at 225 rpm. Every day the populations were diluted 1/10 000 in a doubling concentration of MIC until reaching x8 MIC. As controls, 15 populations of each strain were transferred either without antibiotic or at a constant piperacillin concentration of 1/8 MIC. The OD_{600} of each population was measured each day and a population was considered extinct when its OD fell below 0.1 after 20h of incubation. All populations were frozen in 15% glycerol every two days.

Piperacillin resistance maintenance

As a proxy for plasmid persistence, maintenance of the piperacillin resistance phenotype in the absence of selection pressure in the *P. aeruginosa* populations was measured each day. The *P. aeruginosa* control populations grown without antibiotic were transferred in a piperacillin concentration corresponding to 1/8 MIC with a similar 1/10 000 dilution factor and incubated for 20h before measuring survival rates through OD measurements.

PCR analysis

Final cassette rearrangements

At the x2 and x4 MIC timepoint for the *P. aeruginosa* genotypes and at the final x8 MIC transfer for the *E. coli* strains, surviving populations were plated using a pin replicator on LB agar supplemented with piperacillin at a concentration corresponding to x1 MIC, x2 MIC and x4 MIC respectively and incubated for 48h at 37°C. For the x2 and x8 MICs time-points 12 populations of each genotype were picked randomly, boiled for 10 minutes at 95°C in distilled water. This was used as a substrate for four different PCRs: amplification of the integrase size (as control for cross-contamination), localisation of the *blaVEB-1* cassette relative to the start of the integrase, localisation of the *blaVEB-1* cassette relative to the end of the *dfrA5* cassette and identification of *blaVEB-1* cassettes duplications. Primer binding sites are represented in Figure 4.6.3, while primer sequences are indicated in Appendice Table 5.3. PCRs were performed using the GoTaq G2 DNA mastermix (Promega) for 30 cycles with 30s at 95°C denaturation, 30s at 55°C annealing and 3 minutes at 72°C elongation. The same amplifications were also carried out on all the *P. aeruginosa* surviving populations at x4 MIC.

Plasmid backbone deletions and inversions

Surviving PAO1:WTA4 and PAO1:ΔA4 populations at x1 MIC were screened for inversions between *resP* and *trwJ* and deletions between *blaVEB-1* and *stbA*. Inversions were detected using a pair of primers specific to the integrase and *trwG*

genes and binding on the same strand, therefore only producing an amplicon if the intervening sequence is reversed. Deletions were targeted using primers specific to *blaVEB-1* and *stbA*: given *blaVEB-1* and *stbA* are separated by 17 kbp on the ancestral plasmid, amplification is only possible in the presence of deletion bringing the primers closer together. Populations were plated from frozen stock using a pin replicator on LB agar supplemented with 1/2 MIC concentration of piperacillin and incubated for 48h at 37°C. Whole populations were swabbed, boiled and used as template for PCR amplification with the GoTaq G2 DNA mastermix (Promega) and the following protocol: 30s at 95°C denaturation, 30s at 55°C annealing and 1 minute at 72°C elongation. In total, 82 PAO1:ΔA4 and 66 PAO1:WTA4 populations were screened, with genomic DNA from the x2 MIC populations with deletions and inversions used as positive control.

Bioinformatic analysis

DNA extraction

PAO1:WTA4 and PAO1:ΔintA4 populations from the x2 MIC time-point were re-grown from frozen stock in LB Miller media supplemented with a piperacillin concentration of x1 MIC (ramping populations), x1/8 MIC (1/8 MIC controls), no antibiotic (no antibiotic controls) or 100 mg/L of ceftazidime (ancestral PAO1:WTA4 and PAO1:ΔintA4 populations). DNA extraction was performed using the DNeasy Blood & Tissue Kit (Qiagen) through the QiaCube extraction platform. Samples were treated with RNAse before extraction and DNA concentration was measured using the Quantifluor dsDNA system (Promega). For each control treatment (1/8 MIC and no antibiotic, for each genotype) the DNA of six populations was pooled in equimolar ratio and sequenced as one sample.

Next Generation Sequencing

Library preparation and sequencing was performed by the Oxford Genomics Centre at the Wellcome Centre for Human Genetics. Sequencing was performed in two batches. Twelve samples of each genotype as well as the ancestral populations (PAO1,

PAO1:WTA4 and PAO1:ΔintA4) and the pooled controls were first sequenced on the NextSeq 500 sequencing system (Illumina). Five PAO1:WTA4 populations as well as the ancestral strains were contaminated by phage during the DNA extraction process and were re-extracted and re-sequenced alongside the remaining PAO1:ΔintA4 populations (14 populations) on the HiSeq4000 platform.

Bioinformatic pipeline

Similarly to the previous chapter, PCR duplicates and optical artifacts were removed using MarkDuplicates (Picard (Institute, 2018)) while low quality bases and adaptors were trimmed using Trimmomatic v.039 (Bolger *et al.*, 2014) and overall quality control was performed using FastQC (Andrews *et al.*, 2012) and multiQC (Ewels *et al.*, 2016). Phage contamination in the first batch of samples was determined at this stage.

Variant calling and rearrangement identifications were performed using the *breseq* pipeline in polymorphism mode. Variants present in the un-evolved ancestor populations were filtered out irrespective of frequency. Due to the lower quality of the samples processed by NextSeq 500, which led to an over-identification of low frequency variants, a filtering process harsher than previous chapter's was carried out: any variant present in the controls evolved without antibiotic (samples sequenced with the NextSeq 500) or at a frequency of less than 30% was removed from the ramping populations dataset. The evidence for any variant above this threshold and present in several NextSeq 500 populations was manually examined and the variant removed if present in a region of low read quality. For the pooled 1/8 MIC populations a lower threshold of 5% (which corresponds to a variant present in 30% of one out of 6 populations) was applied and manual examination was carried out to exclude low quality variants. Additional evidence for plasmid rearrangements identified by *breseq* (large scale deletions, inversions, chromosomal insertion) was obtained using CNOGpro (Brynildsrud, 2018) (gene copy number) and plasmidSPAdes (Antipov *et al.*, 2016).

Statistical analysis

Statistical analysis was carried out using R (version 3.6.1) and RStudio (Version 1.2.5). Survival analysis and log-rank test were performed using the *survminer* (survminer) and *survival* (Therneau, 2015) packages.

4.4 Results

4.4.1 Integrase activity can decrease survivability

We selected an array with the *blaVEB-1* cassette in last position (PAO1:WTA4) and its corresponding Δint mutant (PAO1:ΔintA4) in *P. aeruginosa*. As with gentamicin (Figure 3.7.1) we observed no difference in piperacillin resistance between the PAO1:WTA4 and PAO1:ΔintA4 mutants (Figure 4.6).

We used an experimental evolution approach similar to the previous chapter's to determine the impact of integrase activity on piperacillin resistance evolution. We followed 90 populations of each strain under doubling concentrations of piperacillin, starting from 1/8 MIC (ramping treatment), while 15 populations of each strain were transferred at either a constant piperacillin concentration of 1/8 MIC or without antibiotic (control treatments) and we measured survival rates by OD_{600} reading (Figure 4.1 A). We continued the experiment until complete extinction of all ramping populations at x8 MIC.

We observed population extinctions from the x1 MIC transfer for both strains, showing not all populations were able to evolve higher levels of piperacillin resistance (Figure 4.1 B). We observed a better survival rate of the PAO1:ΔintA4 populations in comparison to the PAO1:WTA4 strain, with PAO1:ΔintA4 populations surviving until x4 MIC. While minor, the difference in survival rates was statistically significant (log-rank test: Chisq= 12.4 on 1 degree of freedom, p= 4e-04).

We also tracked the persistence of the resistance phenotype in the 15 populations transferred without antibiotic as a proxy for plasmid maintenance (Supplementary Figure 4.7). We observed rapid disappearance of the piperacillin resistance phenotype, with absence of resistance in every population after 5 transfers, potentially

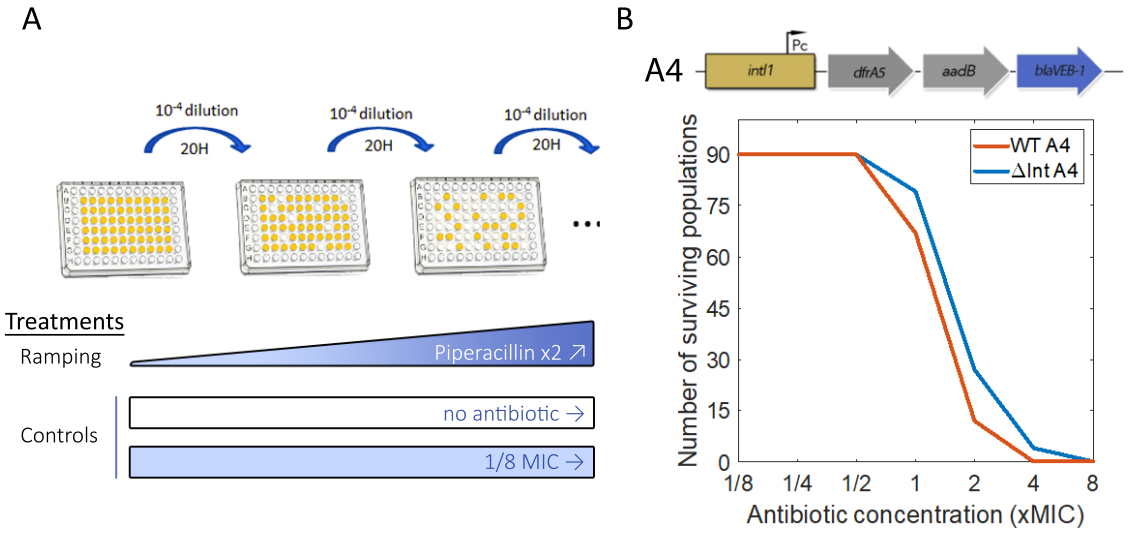


Figure 4.1: A. Schematic representation of the experimental evolution protocol. **B**. Top: representation of the WTA4 array, with the β-lactamase *blaVEB-1* cassette highlighted in color. Bottom: Survival curves of the PAO1:WTA4 and PAO1:ΔintA4 populations during the ramping experiment. Population survivals were estimated through $OD_{6}00$ measurement

due to plasmid loss. We observed a faster loss of resistance in the populations with a functional integrase than in the mutants, which could be due to the cost of the integrase.

4.4.2 Extensive rearrangements of the plasmid backbone & mutations in the conjugation machinery

We first investigated the prevalence of cassette shuffling in the surviving populations at x2 MIC by PCR. Only one rearrangement bringing the *blaVEB-1* cassette closer to the start of the array was identified and no *blaVEB-1* duplications were detected. We then used next-generation sequencing to confirm this low level of rearrangement, search for any visible effect of integrase activity and determine the mechanisms involved in the evolution of piperacillin resistance.

We first focused on the integron-bearing plasmid (Figure 4.2). We confirmed the loss of the plasmid in the controls without antibiotic and the presence of only one type of cassette rearrangement in the ramping populations: the excision of the *aadB* cassette in one population, bringing *blaVEB-1* closer to the start of the

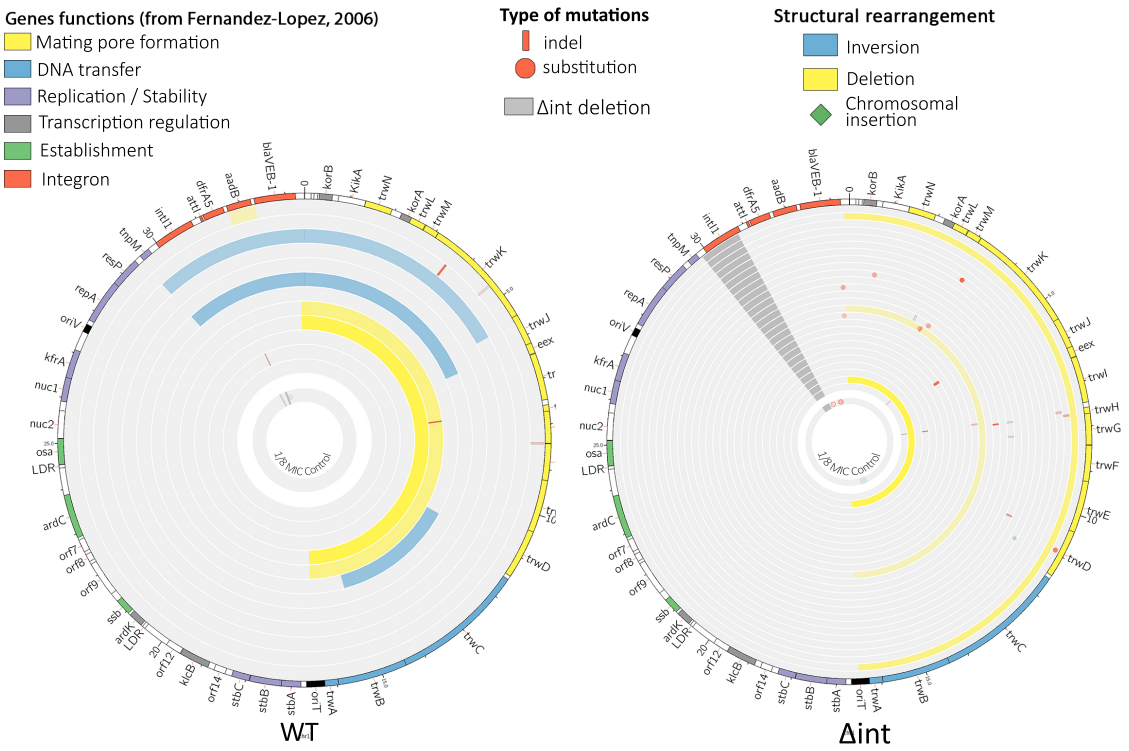


Figure 4.2: Representation of the plasmid mutations and rearrangements in the surviving PAO1:WTA4 (left - 12 populations) and PAO1:ΔintA4 (right - 26 populations) populations at x2 MIC, mapped to the R388 reference sequence. Each circle represents a separate population, with the inner circle representing the variants present in an equimolar pool of six 1/8 MIC control populations. Large scale inversions and deletions are represented in blue and yellow respectively, while indels and single nucleotide substitutions are represented in red. The color intensity represents the frequency of the corresponding mutation. The dark grey area in the PAO1:ΔintA4 populations represents the location of the *IntI1* deletion. The function of each R388 gene as described in (Fernández-López *et al.*, 2006) is indicated by a specific color in the outer circle.

array (Supplementary Table 4.5). We identified a limited number of mutations and rearrangements within the integron itself (Supplementary Table 4.4). In the ramping samples, we observed the presence of one non-synonymous SNP within *blaVEB-1* in two ΔintA4 populations, as well as a 20bp deletion in the intergenic region between the *attI* site and the start of *dfrA5* in one WTA4 population. The 1/8 MIC control pooled populations contained several mutations targeting the other two cassettes: we identified several mutations targeting the *dfrA5* cassette, all nonsense or frame-shifting mutations, as well as the insertion within the *aadB* cassette of ISPa11 (Partridge and Hall, 2003) initially located on the chromosome. We also observed a 131bp deletion within the integrase in the 1/8 MIC WTA4 control, compatible with

the hypothesis of a deleterious integrase, even at low antibiotic concentration.

While we observed very few cassette shuffling events, we identified extensive mutations and recombinations in the rest of the plasmid backbone and especially in genes involved with conjugation. Alongside the complete deletion of the DNA transfer replication and mating pore formation modules (Frumerie *et al.*, 2010) in five populations, we identified several mutations disabling the *trwK*, *trwH*, *trwD*, *trwG* and *trwM* genes (Table 4.2), with 12 out of 17 mutations being either non-sense or frame-shifting mutations (Supplementary Table 4.4). These five genes are all part of a Type 4 Secretion System, encoded by the operon *trwDEFGHIJKLMN* (Fernández-López *et al.*, 2006) whose genes *trwD,E,F,G,I,J,K,L* and *H* have been shown to be essential for R388 conjugation (de Paz *et al.*, 2005). We also identified four inversions in this region of the plasmid, mainly in the WTA4 populations: two inversions from *resP* to *trwJ*, one from *trwD* to *trwB*, and a smaller 773 bp inversion within *trwC*. We can expect these inversions to disrupt the genes at the border of the rearrangements but also to impact the expression of other downstreams genes, as the *trwL-D* genes are separated in two transcriptional units, *korA-trwI* and *trwH-trwD*. Finally we identified the insertion within the *trwD* gene of one Δint population of a 1.2kbp chromosomal sequence containing the genes *PA0979* and *PA0978*, which encode for an integrase (Song and Zhang, 2011). Several of these *trw* variants are present at a high frequency and swept through the populations (Supplementary Table 4.4).

We then examined the mutations located in the chromosome (Figure 4.3 and Table 4.3) and observed only a limited number of variants reached a frequency over 30% (Supplementary Table 4.6). The region of the genome with the most mutations was the *PA1766-PA1768* operon. These genes are predicted to be involved in post-translation modification of cell wall proteins and have been linked with polymyxin resistance in *P. aeruginosa* (Gutu *et al.*, 2013). Moreover, a *PA1767* mutant from a PA14 genomic transposon mutant library was shown to be x4 more resistant to piperacillin than its ancestor (Dötsch *et al.*, 2009), but nothing more is known

Gene name	Ramping				1/8 MIC	
	Delta (freq)	Delta (n)	WT (freq)	WT (n)	Delta	WT
trwH	0.13	10	0.25	3	0.00	0.00
trwD	0.13	6	0.25	3	0.00	0.00
trwJ	0.03	3	0.30	4	0.00	0.00
trwG	0.08	5	0.21	3	0.00	0.00
trwB	0.03	3	0.25	3	0.06	0.00
trwM	0.07	5	0.17	2	0.00	0.00
trwL	0.04	4	0.17	2	0.00	0.00
eex	0.03	3	0.17	2	0.00	0.00
korA	0.03	3	0.17	2	0.00	0.00
korB	0.03	3	0.17	2	0.00	0.00
trwA	0.03	3	0.17	2	0.00	0.00
trwC	0.03	3	0.17	2	0.00	0.00
trwE	0.03	3	0.17	2	0.00	0.00
trwF	0.03	3	0.17	2	0.00	0.00
trwI	0.03	3	0.17	2	0.00	0.00
trwN	0.03	3	0.17	2	0.00	0.00
resP /hypothetical protein	0.00	0	0.14	2	0.00	0.00
intI1/dfrA5	0.00	0	0.06	1	0.00	0.00
hypothetical protein	0.02	1	0.00	0	0.00	0.00
blaVEB-1	0.02	2	0.00	0	0.00	0.00
aadB	0.00	0	0.00	0	0.00	0.07
aadB gene/blaVEB-1	0.00	0	0.00	0	0.33	0.00
dfrA5	0.00	0	0.00	0	0.32	0.37
intI1	0.00	0	0.00	0	0.00	0.14

Table 4.2: List of plasmid genes mutated by indels and substitutions in the ramping populations and their mutation average frequency. Numbers represent the cumulative frequencies of mutations found within each mutated plasmid gene, normalized by the number of samples, across each genotype. Cumulative mutation frequency for these genes in the 1/8 MIC pooled populations is indicated as reference. The inversions are noted solely as impacting the genes where the junctions are located, but we also expect the expression of the genes from the *trw* operon to be affected due to its organisation in two transcriptional units.

of the molecular mechanisms linking this operon to β-lactam resistance. In the same transposon screen, the deactivation of *galU*, a pyrophosphorylase required for lipopolysaccharide synthesis (Dean and Goldberg, 2002), and *PA1195* were shown to induce higher levels of piperacillin resistance. We found a mutation in another gene also involved in lipopolysaccharide formation *waaL* (Abeyrathne *et al.*, 2005), which was not previously linked with β-lactam resistance. Other genes have been shown to be involved with antibiotic resistance: both *clpA* (an

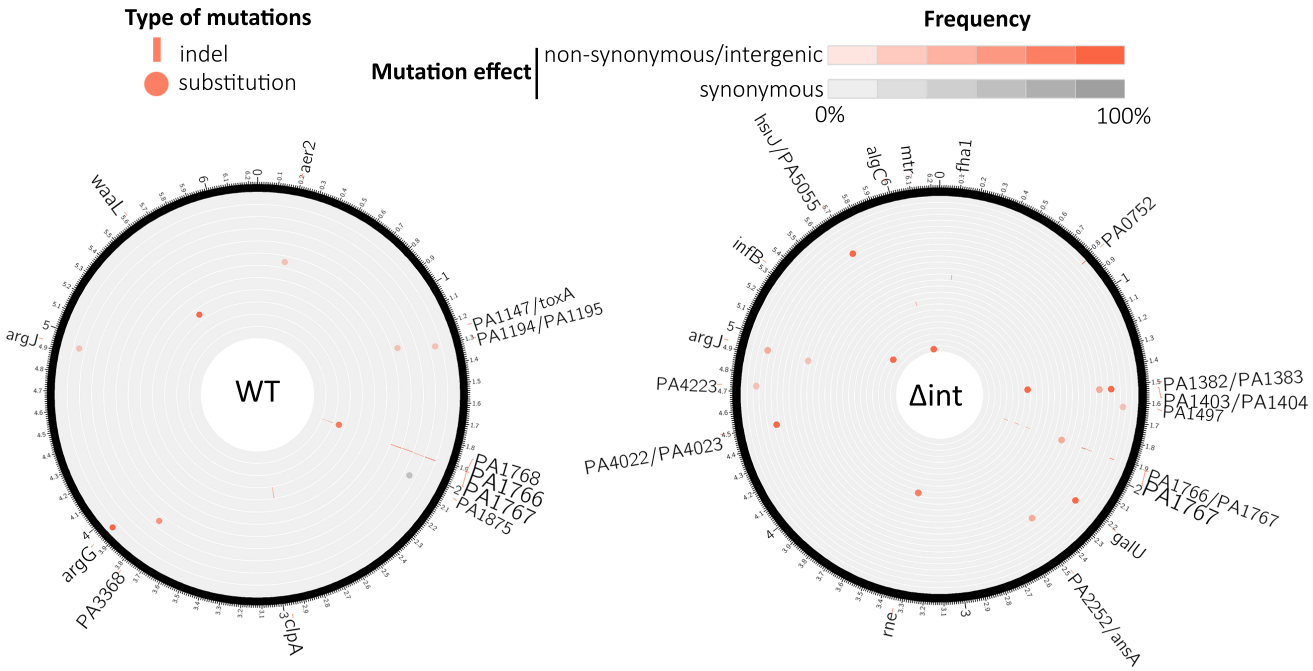


Figure 4.3: Representation of the chromosome mutations in the surviving PAO1:WTA4 (left - 12 populations) and PAO1:ΔintA4 (right - 26 populations) populations at x2 MIC mapped to the R388 reference sequence. Each circle represents a separate population. The type (indel, substitution) of each mutation is represented by the shape of the marker (line, circle), while the marker colour represents the effect of the mutation (non-synonymous/intergenic vs synonymous) and its colour intensity represents its frequency. The size of the gene labels on the outer ring represents the overall cumulative frequency of mutations present in this gene across all populations from this genotype.

intracellular protease) and *infB* (a translation initiation factor) have been linked with ceftazidime resistance (Sanz-García *et al.*, 2018), while *PA1875* is part of the ABC transporter, an efflux pump (Zhang and Mah, 2008). *argG* and *argJ* are involved in the arginine metabolic pathway (Lu *et al.*, 2004) and the same PA14 transposon mutant library screen showed *argJ* to be linked with cephalosporin but not piperacillin resistance. We were not able to identify any link with antibiotic resistance for the remaining 16 out of 26 mutated genes.

No significant differences in overall mutation frequencies, mutation type or mutation effect were found at the chromosomal level between the WTA4 and ΔintA4 populations (Supplementary Figure 4.9). At the gene level, mutations in the *PA1766-PA1768* operon were more likely to be through large deletion in the WTA4 than in the ΔintA4 populations but a larger sample size would be required

Treatment	Ramping				1/8 MIC		
Gene name	Delta (freq)	Delta (n)	WT (freq)	WT (n)	Delta	WT	Gene product
PA1767	0.157	6	0.122	3	0	0	hypothetical protein
PA1766	0	0	0.135	2	0	0	hypothetical protein
argG	0	0	0.083	1	0	0	argininosuccinate synthase
waaL	0	0	0.083	1	0	0	O-antigen ligase WaaL
PA1768	0	0	0.078	1	0	0	hypothetical protein
clpA	0	0	0.072	1	0	0	ATP-binding protease component ClpA
PA3368	0	0	0.064	1	0	0	acetyltransferase
argJ	0.031	2	0.027	1	0	0	acetyltransferase/N-acetylglutamate synthase
PA1403/PA1404	0.056	2	0	0	0	0	transcriptional regulator/hypothetical protein
galU	0.038	1	0	0	0	0	uridylyltransferase
hslU/PA5055	0.038	1	0	0	0	0	protease ATP-binding subunit HslU/hypothetical protein
infB	0.038	1	0	0	0	0	translation initiation factor IF-2
mtr	0.038	1	0	0	0	0	tryptophan permease
PA0752	0.038	1	0	0	0	0	hypothetical protein
PA1382/PA1383	0.038	1	0	0	0	0	type II secretion system protein/hypothetical protein
PA4022/PA4023	0.038	1	0	0	0	0	aldehyde dehydrogenase/transporter
PA1875	0	0	0.034	1	0	0	hypothetical protein
rne	0.032	1	0	0	0	0	ribonuclease E
PA1147/toxA	0	0	0.030	1	0	0.287	amino acid permease/exotoxin A
PA1766/PA1767	0.029	1	0	0	0	0	PA1766/PA1767
aer2	0	0	0.029	1	0	0	aerotaxis transducer Aer2
PA1194/PA1195	0	0	0.027	1	0	0	amino acid permease/hypothetical protein
fha1	0.023	1	0	0	0	0	Fha domain-containing protein
algC	0.022	1	0	0	0	0	phosphomannomutase
PA2252/ansA	0.021	1	0	0	0	0	AGCS symporter/L-asparaginase I
PA4223	0.012	1	0	0	0	0	ABC transporter ATP-binding protein
PA1497	0.012	1	0	0	0	0	transporter

Table 4.3: List of mutated chromosome genes in the ramping treatment, their mutation average frequency and the mutation count within each genotype. *(freq)* represents the cumulative frequencies of mutations found within each mutated plasmid gene, normalized by the number of samples and across each genotype while *(n)* represents the number of mutations present in this gene across all populations from this genotype. Cumulative mutation frequency for these genes in the 1/8 MIC pooled populations is indicated as reference.

to confirm this effect (Supplementary Table 4.6).

4.4.3 Recombinations and potential off-target effects of the integrase

Junction analysis

Since plasmid backbone rearrangements, and especially inversions, were found more frequently in WTA4 than in ΔintA4 populations, we investigated a potential implication of the integrase in these rearrangements. To do so, we searched the recombination junctions sites for integrase secondary sites (Figure 4.4). The recognition characteristics of *IntI1* require little sequence identity, with only a three base long conserved motif, 5'-GTT-3', where cleavage usually happens between the

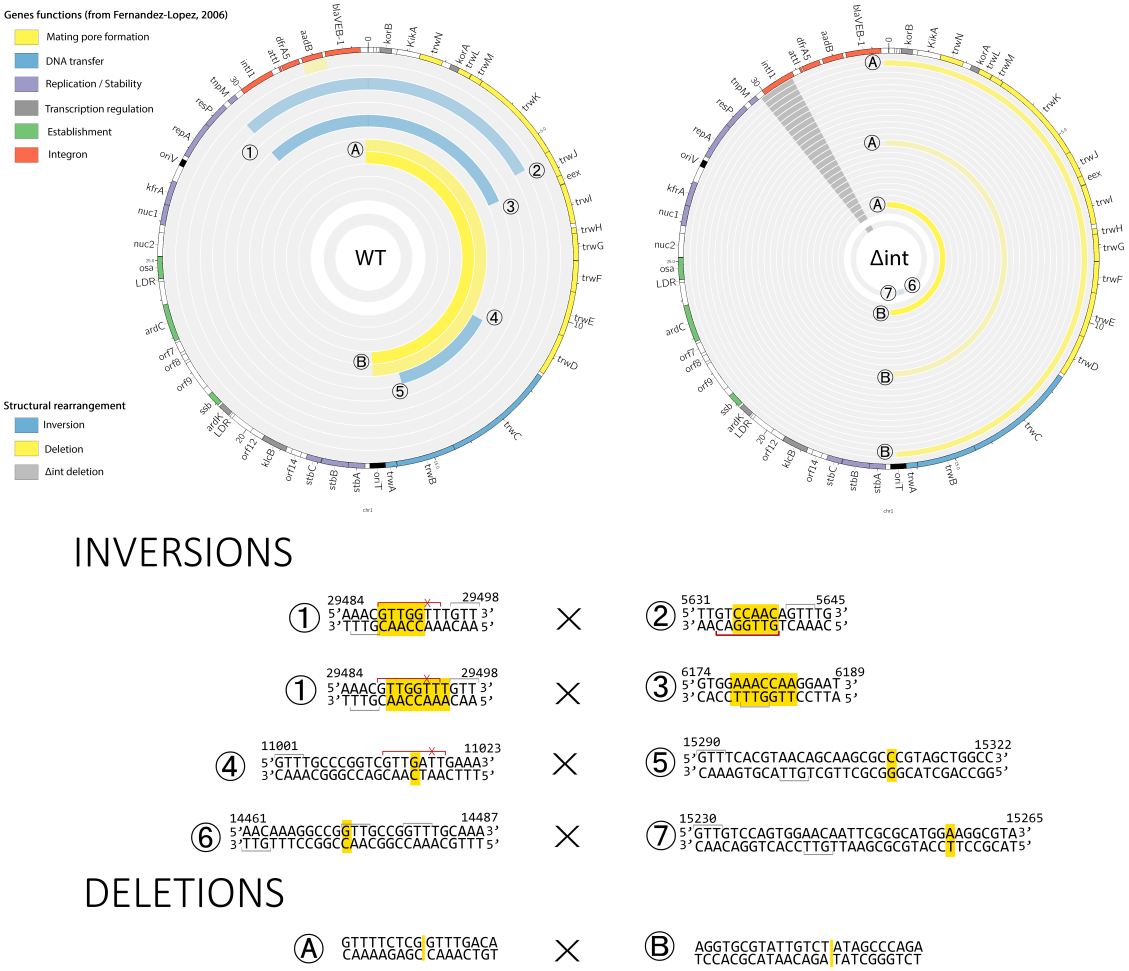


Figure 4.4: Plasmid rearrangement junction sites. Top: Representation of the plasmid rearrangement. Each junction site is indicated by a number (inversion) or a letter (deletion). Bottom: junction sequences for each rearrangement. Junction site is indicated in yellow. When the crossover point is unclear due to sequence homology between each junction the entire homology is highlighted. Motifs close to the GTTRRRY integron secondary sequence are indicated by a red line, with red crosses indicating the potential mismatch. Simple 5'-GTT-3' triplets are indicated by a grey line. The *nic* site of the origin of transfer *oriT* is indicated by a black arrow.

base A and C of the complementary strand (Escudero *et al.*, 2015). In the case of *attC* sites the conserved region can be extended to form a 7bp GTTRRRY motif.

The inversions between *resP* and *trwJ* match well with integrase activity: while we cannot determine the precise cross-over point due to sequence homology we can find a 5'-GTT-3' triplet at each site, which can often be extended to form a 7bp sequence similar to the GTTRRRY motif. Moreover in both these inversions the 5'-GTT-3' triplets of each recombination site are in inverted orientation from

each other, with one 5'-GTT-3' triplet located on the top strand and one triplet located on the bottom strand. It has been shown that the integrase can create sequence inversions through intramolecular $attI \times attI$ reactions when the sites are in inverted orientation (Escudero *et al.*, 2016). In this case the Holliday junction formed during the first step of the recombination process is resolved through second strand exchange instead of replication, as is the case with the canonical $attI \times attC$ or $attC \times attC$ reactions. While we cannot rule out other mechanisms due to the homology between the sequences, this integrase pathway could be the origin of these inversions, which seems specific to WTA4 populations.

The involvement of the integrase in the $trwD \times trwB$ inversion is less clear. We identified a motif close to GTTRRRY at the *trwD* end, but the cleavage point is after the 5'-GTT-3' triplet. It has been shown that the constraints on the second nucleotide of the triplet are fairly lax, and that cleaving can happen between the G and the A of a 5'-GAT-3', as observed here, especially in the case of reaction with secondary sites (Francia *et al.*, 1993). However, we were not able to identify any 5'-GTT-3' triplet in the vicinity of the *trwB* junction, casting doubt on the implication of integrase activity in this inversion. As a control, we also investigated the small inversion located within the 1/8 MIC Δint pooled sample: we identified one 5'-GTT-3' triplet on one side, but not on the other.

Finally we looked into the extended $blaVEB\text{-}1 \times oriT$ deletions. These deletions are found in both genotypes but are marginally more frequent in the WTA4 populations and all share the same junction sites. Similarly to the deletions identified in the previous chapter, one border is located within the *attC* site of a cassette. In this case however the junction does not happen at the recombination site itself but in the variable terminal structure (VTS) of the *attC* site. No motif specific to the integrase was easily recognizable at the *oriT* end, which casts into doubt the involvement of the integrase. Interestingly, we found that the junction within the origin of transfer *oriT* happens precisely at the R388 *nic* site (Llosa *et al.*, 1995). During R388 conjugation the *TrwC* relaxase binds to *oriT* and initiates the

nicking of the DNA at the *nic* site to create single stranded DNA which is then exported to the recipient cell and recircularized (De La Cruz *et al.*, 2010). Moreover it has been shown that *TrwC* can mediate site-specific recombination between two *oriT* sites even in the absence of conjugation (Llosa *et al.*, 1994) which leads to loss of the intervening DNA (Draper *et al.*, 2005). It is unclear if the *TrwC* enzyme can bind to the sequence at the *blaVEB-1 attC* junction site: recognition of the *nic* site is supposed to be highly specific (Lucas *et al.*, 2010) and we did not identify any obvious sequence homology between the two junctions sites. Recombinations between the two sites may however be facilitated by the structure of the *attC* site: the limiting step for *oriT* recombination is the generation of a single strand at *oriT* which facilitates *TrwC* nicking (César *et al.*, 2006) while the VTS region of *blaVEB-1 attC* is characterized by a complex secondary structure (Aubert *et al.*, 2012) which may render the junction site available as single stranded DNA.

Rearrangements at x1 MIC

In order to study the involvement of the integrase in these recombinations with a bigger sample size we screened 66 WTA4 and 82 ΔintA4 surviving populations at x1 MIC for the presence of $resp \times trwJ$ inversions and *blaVEB-1* $\times$ *oriT* deletions by PCR. The $resp \times trwJ$ inversions were sparse, with only one additional WTA4 population identified containing an inversion (Figure 4.5). No similar inversions were detected in any of the 82 ΔintA4 populations, confirming the hypothesis these inversions are integrase mediated. We detected more *blaVEB-1* $\times$ *oriT* deletions in both genotypes, with a higher frequency within the WTA4 populations, but this difference was not significant (Pearson's Chi-squared test, X-squared = 2, df = 1, p-value = 0.1573).

4.4.4 Experimental evolution with *E. coli*

We observed a much lower frequency of rearrangement in this experiment than in the previous chapter, which may be due to a low induction level of the integrase. As seen in chapter 2, piperacillin does not induce the SOS response or over-expression

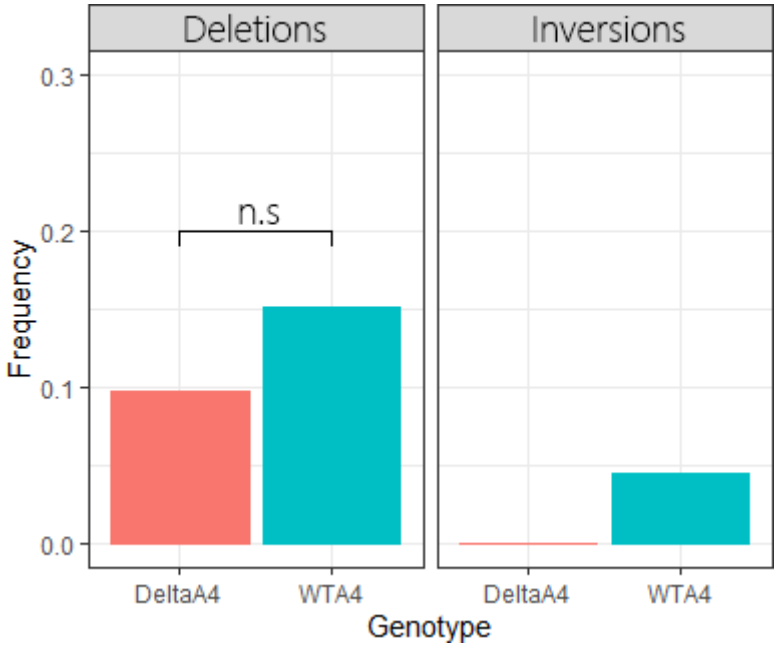


Figure 4.5: Frequency of populations containing *resp* × *trwJ* inversions and *blaVEB-1* × *oriT* deletions at the x1 MIC point for each genotype.

of the integrase in *P. aeruginosa*. This is not the case in *E. coli* where a dedicated pathway links cell wall stress to the SOS response (Miller *et al.*, 2004b). In order to investigate if higher levels of integrase expression may generate more cassette rearrangements and an increase in evolvability, we carried out a similar experimental evolution set-up using the *E. coli* MG1655 strain as genetic background. We used the array with the lowest piperacillin resistance in *E. coli*, the array A5. As seen in Chapter 2, this array has *blaVEB-1* in second and not in last position, since the main factor influencing *blaVEB-1* expression in *E. coli* is the presence of the *dfrA5* cassette in front of it, rather than its absolute position within the array. We expect MG1655:WTA5 to be able to increase its piperacillin resistance levels either by inserting *dfrA5* in front of *blaVEB-1* or by bringing *blaVEB-1* closer to the promoter.

Using again a ramping treatment doubling concentration of piperacillin each day, we measured the survival rates of the MG1655:WTA5 and MG1655:ΔintA5 populations (Supplementary Figure 4.10). Both genotypes evolved higher levels of resistance relative to the *P. aeruginosa* strains, with populations surviving up to x8 MIC, and we observed no differences in evolvability between the two genotypes.

We then screened the surviving populations at x8 MIC for the presence of cassette rearrangements by PCR: we determined the distance between *blaVEB-1*, *aadB* and the start of the array as well as any *blaVEB-1* duplications or rearrangement bringing *dfrA5* in front of *blaVEB-1* (Supplementaty Figure 4.11). We observed no duplications or rearrangement bringing *blaVEB-1* closer to the start of the array. We only observed a faint band in one population in the screen for *dfrA5/blaVEB-1* rearrangements, but this rearrangement was not confirmed when looking at the other amplifications targeting the position of the *blaVEB-1* and *aadB* cassettes.

4.5 Discussion

Integron cassettes display a diversity of fitness costs, recombination rates and expression gradients along the array, all of which may impact the benefits of integrase activity. We therefore aimed in this chapter to quantify the impact of cassette shuffling for another cassette, the β-lactamase encoding *blaVEB-1*. We used a similar experimental evolution approach to quantify difference in piperacillin resistance evolution of strains with and without a functional integrase, both in *P. aeruginosa* and *E. coli*.

Unlike the previous experiment with *aadB*, we observed a lower survival rate of the *P. aeruginosa* populations with a functional integrase. This shows integrase activity can reduce bacteria antibiotic resistance evolvability, which is expected if the fitness costs of integrase activity exceed its adaptive benefits. Integrase costs have already been reported: in the absence of the SOS response the integrase has been shown to impose a 7% to 11 % fitness reduction in *Acinetobacter baylyi*, leading to the emergence of inactivated integrase mutants (Starikova *et al.*, 2012), similar to the numerous *intI* pseudogenes found across phyla (Nemergut *et al.*, 2008). This cost has been hypothesized to originate from non-canonical recombinations between secondary sites present in the genome, which may lead to the deletion of essential genes (Starikova *et al.*, 2012). The integrase cost can be alleviated through *LexA* regulation (Guerin *et al.*, 2009)(Lacotte *et al.*, 2017), but solely in the absence of stressors. Given we saw piperacillin does not induce the SOS response

at sub-MIC level in *P. aeruginosa* (Chapter 2), the cost we observed may therefore come from the basal level of expression of the integrase or from an SOS response induction only visible at higher concentrations of antibiotics, which would need to be confirmed through competition experiments.

Interestingly this integrase cost was not seen in *E. coli* when performing a similar ramping experiment: MG1655 WTA5 and ΔintA5 had the same survival rate. This result is surprising, as we saw previously that piperacillin was better at inducing the SOS response in *E. coli* and this species has a dedicated pathway linking cell wall stress to the SOS response (Miller *et al.*, 2004a). We observed in the previous chapter similar ratio of cassette to integrase transcription in both species but a difference in overall plasmid transcription levels cannot be ruled out, and our *P. aeruginosa* strains may have overall higher integrase levels than in *E. coli*. An alternative explanation is a species-specific cost of the integrase, which we can expect to be linked to the number of integrase secondary sites within the genome. However given *P. aeruginosa* chromosome was shown to possess a distinctively lower frequency of GTTRRRY secondary sites than the *E. coli* genome, the exact mechanisms and interactions behind the integrase fitness cost remain to be determined.

We observed very limited levels of cassette shuffling both in the *P. aeruginosa* and *E. coli* strains, with the partial deletion of the *aadB* cassette in only one population but no shuffling of the *blaVEB-1* cassette. This may be due to the low recombinant potential of the *blaVEB-1 attC* site which is 50 times less efficient for recombination than the *aadB attC* site due to integrase binding sites diverging from the *attC* consensus sequence (Aubert *et al.*, 2012). Duplication events, which require excision followed by reintegration of the cassette, may therefore be too unlikely and limited to highly mobile cassettes. The rearrangement we observed with *blaVEB-1* is similar to the results observed in a previous study with *catB9* (Barraud and Ploy, 2015). In the absence of integrase overexpression, *catB9* was brought closer to the promoter only through excision of the proximal cassettes. Like *blaVEB-1*'s, the

catB9 attC site is characterized by an extensive variable terminal structure which may hinder proper *attC* folding (Barraud and Ploy, 2015). In comparison to our previous chapter, cassette rearrangement may also be limited by a reduced integrase induction, as we observed a stronger SOS induction by gentamicin in *P. aeruginosa* than by piperacillin in both *E. coli* and *P. aeruginosa*.

Instead of cassette rearrangements we observed numerous mutations and recombinations disabling the plasmid conjugation machinery, and especially its Type 4 Secretion System (T4SS) encoded by the *trwDEFGHIJKLMN* operon. Given the high frequency of these mutations in our populations, a question comes to mind: are we observing solely the deactivation of costly mechanisms or is the T4SS additionally increasing piperacillin susceptibility? The T4SS may impose an important fitness cost in *P. aeruginosa*. As we saw, the R388 plasmid is lost in less than 5 daily transfers in the absence of selection, which may be due to the cost of the conjugation machinery. The T4SS is the only secretion system not present in the *P. aeruginosa* genome (Filloux, 2011) and *P. aeruginosa* has been shown to react to T4SS pili by counter-attacking the donor cell through its Type 6 Secretion System (T6SS) (Ho *et al.*, 2013). The presence of the T4SS in *P. aeruginosa* populations may therefore interact with its other warfare mechanisms. To our knowledge, no-one has ever reported interactions between β-lactam antibiotics and the T4SS. However, as β-lactam antibiotics target the bacteria cell wall by binding with penicillin-binding proteins and interrupting cell wall synthesis, their effect may interact with the presence of the T4SS inserted in the bacteria membrane (Wallden *et al.*, 2010). Another argument towards this hypothesis is the reduced number of T4SS mutations observed in the previous chapter with gentamicin, even at the x1024 MIC time-point, which points toward a piperacillin-specific impact on T4SS cost.

Measuring the fitness of the mutants with a deleted T4SS in the presence and absence of piperacillin would bring us closer to answering this question. It would also be useful to observe if the deletion of the conjugation machinery improves the stability of the R388 plasmid in *P. aeruginosa*. As *E. coli* cells containing R388

have been shown to be more piliated than *P. aeruginosa* cells (David and Cohen, 1976), it would also be interesting to see if mutations in the T4SS are more or less frequent in the *E. coli* populations than in the *P. aeurginosa* populations evolved with piperacillin.

In our experiment the inactivation of the conjugation machinery happened through a wide diversity of mechanisms: SNPs, insertion of transposable elements and, notably, large scale inversions and deletions of the plasmid backbone, which were more frequent in the populations with a functional integrase.

By looking at the sequences of the junction sites of the $resp \times trwJ$ inversions we identified patterns in line with integrase activity. For both inversions the cross-over happened within sequences containing 5'-GTT-3' triplets in inverted orientation from each other. These inversions may therefore be the product of a reaction similar to the $attI \times attI$ recombination when the sites are in inverted orientation, which has been shown to result in the inversion of the sequence between the two *attI* sites through second strand exchange (Escudero *et al.*, 2016). Other mechanisms, notably recombination between homologous sequences, can also generate inversions (Periwal and Scaria, 2015). However we confirmed through a screen of the surviving populations at the x1 MIC time-point that these two inversions were specific, even if rare, to the populations with a functional integrase. If confirmed, we provide here some of the first evidence of the usefulness of the $attI \times attI$ reactions, whose biological relevance and adaptive meaning was so far unclear (Escudero *et al.*, 2016).

We observed repeated parallel deletions between the *attC* site of *blaVEB-1* and the plasmid origin of transfer *oriT*. As the deletions were present in both genotypes, an integrase-specific mechanism is excluded, but the exact junctions sites of these deletions were puzzling: one site is located within the variable terminal sequence of the *blaVEB-1 attC* site while the other side happens at the *nic* site within *oriT*, which corresponds to the cutting site of the *TrwC* relaxase (Llosa *et al.*, 1995). *TrwC* has been shown to mediate site-specific recombination between two *oriT* sites outside of conjugation (Llosa *et al.*, 1994) and reactions between *oriT* and

an *oriT*-like sequence could be a potential mechanism leading to the *blaVEB-1* × *oriT* deletion. More work is required to determine if *TrwC* can effectively bind to the *attC* site of *blaVEB-1* since *TrwC* is theoretically highly specific to *oriT*. (Lucas *et al.*, 2010). These results hint toward the existence of interactions between the integron cassette *attC* sites and other recombinases, which has never been reported before. Such interactions may have wide consequences on the spread of integrons by facilitating the jump of integrons and integrons cassettes to new genomic backgrounds. It also enhances the ability of the integron to act as a hot spot for genomic diversity, even in the absence of integrase activity. Given the location of the recombination site within the VTS of the *blaVEB-1 attC* site, more work is required to see if this interaction is specific to this cassette or if it is a widespread phenomenon.

We observed a relatively low level of chromosomal mutations mostly located in genes without an obvious link with antibiotic resistance. We identified several mutations in the *PA1766-PA1768* operon, which deactivation had been previously identified in a transposon screen as increasing piperacillin resistance (Dötsch *et al.*, 2009). These genes are predicted to be involved in modifications of cell wall proteins but little is know about their precise function (Dötsch *et al.*, 2009).

In conclusion, we showed the benefits of integrase activity are highly dependent on its cassette cargo. In absence of beneficial cassette reshuffling due to a cassette with low mobilisation potential, integrase activity can decrease the survivability of bacteria populations. While integrase-mediated recombination between secondary sites outside of the integron can provide novel pathways for adaptation, these recombinations were not enough to compensate for the cost of integrase activity. While SOS regulation helps to reduce the cost of integron carriage in the absence of stressors, the integrase remains a double-edged sword for survival against killing agents. This is similar to other bacteria adaptation mechanisms, such as SOS-induced mutagenesis or competence, which create similar trade-offs. The outcome

of the trade-off between integrase cost and adaptability benefits will depend on a variety of factors, such as its cassette cargo (mobile or not, costly or not), the nature of the stressors (inducing the SOS response or not) or the effect of the integrase on the host genomic background, which is potentially species-specific. Studying the components of this trade-off is essential if we want to understand the dynamics and selection pressures controlling the distribution of integron arrays and their resistance gene cargo across different bacterial species and environments.

4.6 Supplementary materials

4.6.1 MIC in experimental conditions

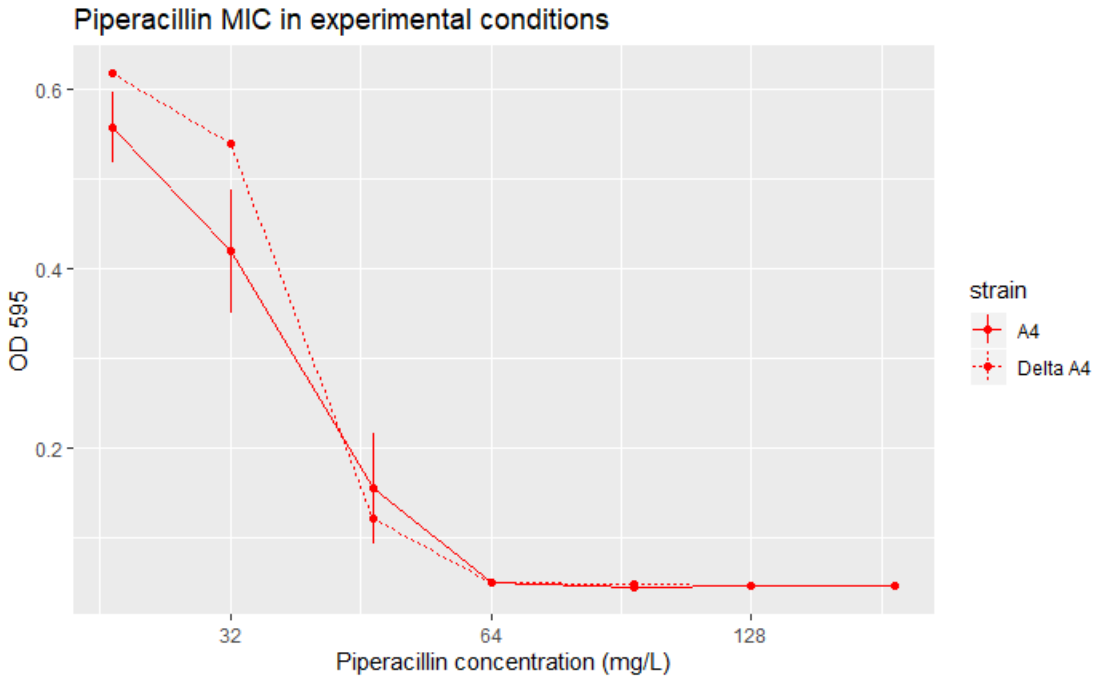


Figure 4.6: Growth of the PAO1:A4 and PAO1:ΔA4 populations in increasing concentration of piperacillin using an inoculum similar to our ramping experiment. Δint strains are indicated by the dashed lines. The MIC was determined for both strains at 64 mg/L.

4.6.2 Loss of the plasmid in the absence of antibiotics

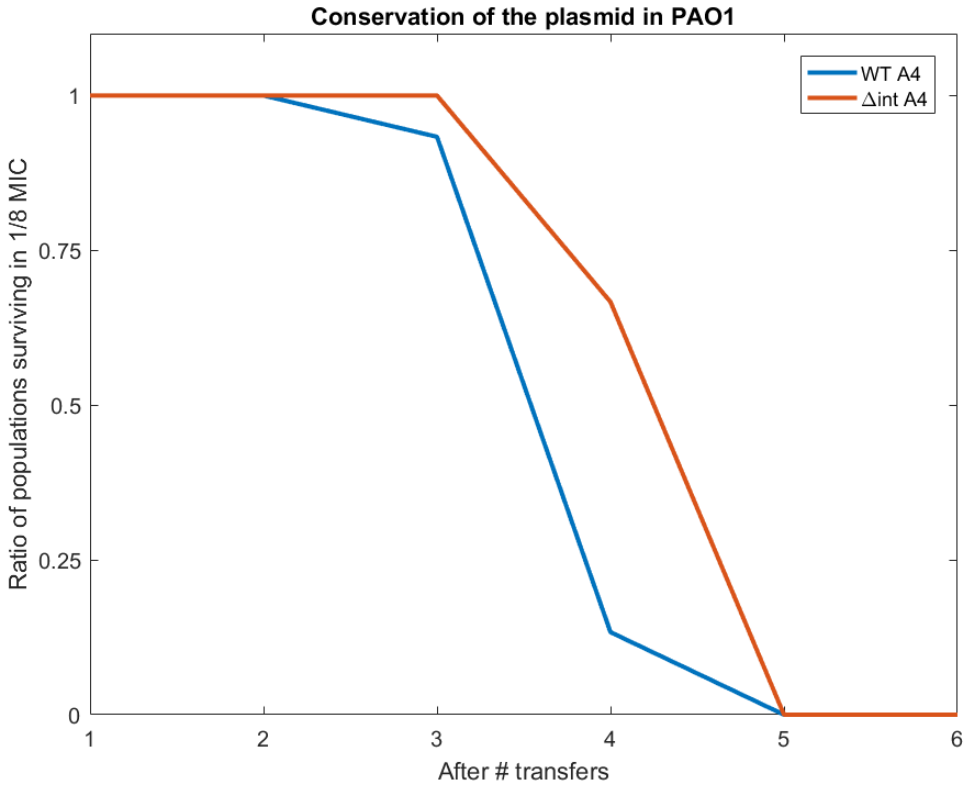


Figure 4.7: Reversal of the resistance phenotype as a proxy for plasmid loss across the ramping experiment. Each day the populations passaged without antibiotic were tested for the presence of the piperacillin resistance phenotype by transferring them in media supplemented with 1/8 MIC followed by 24h incubation. Survival rates were measured through OD_{600}.

4.6.3 PCR primers binding site

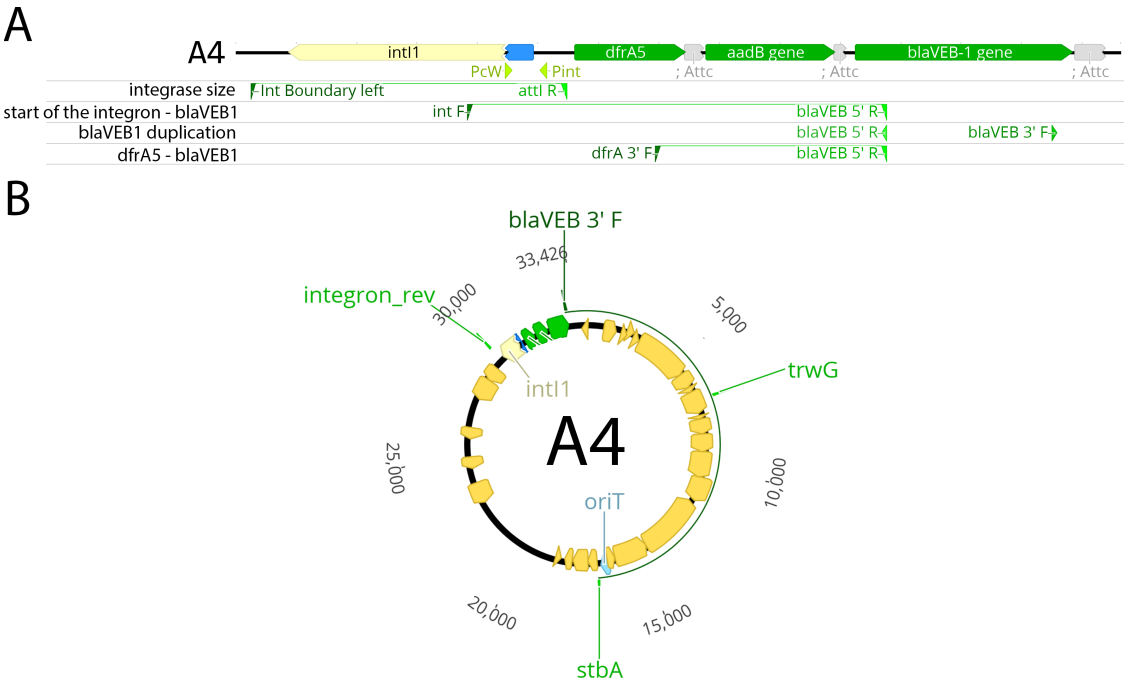


Figure 4.8: Binding sites and amplicons of the primers used in the PCR screen for A. cassettes rearrangements and B. plasmid rearrangements in the PAO1 populations.

4.6.4 Genomic variants

Summary statistics

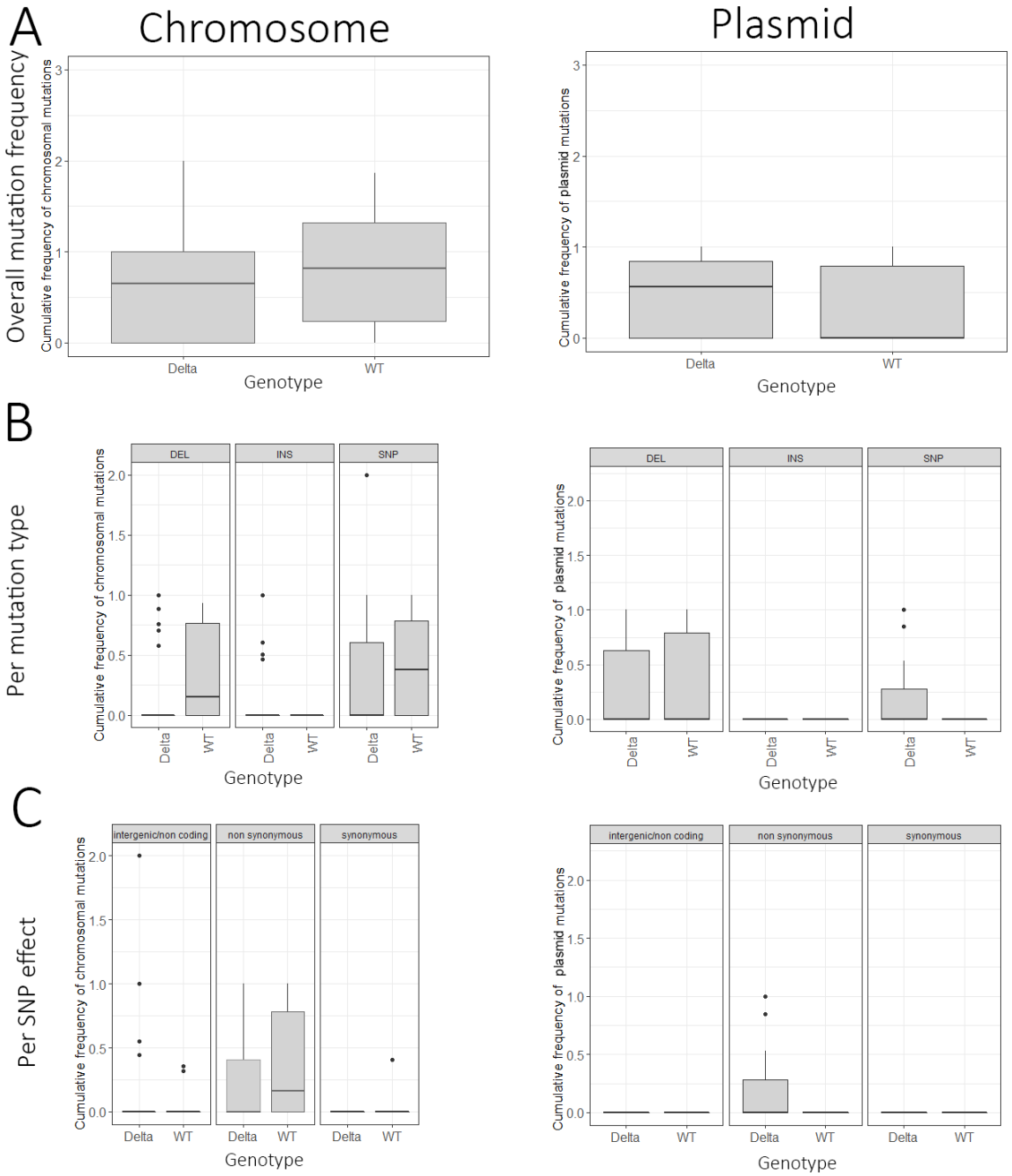


Figure 4.9: Box plots represent the cumulative mutation frequencies for each population of A. all mutations B. per mutation type C. per mutation effect (SNP only). The lower and upper hinges correspond to the first and third quartiles while the middle line corresponds to the median.

Plasmid variants

Type	Position	Mutation	Gene name	Mutation effect	Ramping				1/8 MIC	
					Delta (avg)	Delta (n)	WT (avg.)	WT (n)	Delta	WT
DEL	7620	1	trwH	coding	0.62	4	1.00	1	0.00	0.00
DEL	7702	30	trwH	coding	0.64	1	0.00	0	0.00	0.00
DEL	7702	19	trwH	coding	0.36	1	0.00	0	0.00	0.00
SNP	3235	A	trwK	nonsense	1.00	1	0.00	0	0.00	0.00
DEL	3614	1	trwK	coding	0.00	0	1.00	1	0.00	0.00
DEL	5148	126	trwK	coding	1.00	1	0.00	0	0.00	0.00
SNP	3181	A	trwK	nonsense	0.46	1	0.00	0	0.00	0.00
DEL	4659	1	trwK	coding	0.00	0	0.34	1	0.00	0.00
DEL	4170	12	trwK	coding	0.30	1	0.00	0	0.00	0.00
DEL	10645	11	trwD	coding	0.95	1	0.00	0	0.00	0.00
SNP	10927	T	trwD	nonsense	0.85	1	0.00	0	0.00	0.00
SNP	33208	G	blaVEB-1	non synonymous	0.40	2	0.00	0	0.00	0.00
DEL	7702	45	trwH /trwG	coding	1.00	1	0.00	0	0.00	0.00
SNP	2960	A	trwM	nonsense	0.51	1	0.00	0	0.00	0.00
DEL	2998	1	trwM	coding	0.43	1	0.00	0	0.00	0.00
SNP	793	A	hypothetical protein	nonsense	0.53	1	0.00	0	0.00	0.00
DEL	8422	12	trwG	coding	0.00	0	0.52	1	0.00	0.00
DEL	2556	73	trwL	coding	0.36	1	0.00	0	0.00	0.00
DEL	8184	1	trwG	coding	0.33	1	0.00	0	0.00	0.00
DEL	31136	20	intI1/dfrA5	intergenic	0.00	0	0.76	1	0.00	0.00
SNP	32317	T	aadB gene/blaVEB-1	intergenic	0.00	0	0.00	0	0.19	0.00
DEL	30751	131	intI1	coding	0.00	0	0.00	0	0.00	0.14
SNP	32289	T	aadB gene/blaVEB-1	intergenic	0.00	0	0.00	0	0.14	0.00
DEL	31363	13	dfrA5	coding	0.00	0	0.00	0	0.00	0.13
DEL	31495	1	dfrA5	coding	0.00	0	0.00	0	0.00	0.12
DEL	31494	1	dfrA5	coding	0.00	0	0.00	0	0.00	0.12
SNP	31272	A	dfrA5	nonsense	0.00	0	0.00	0	0.11	0.00

Table 4.4: List of plasmid SNPs and indels in the *P. aeruginosa* populations at x1 MIC. For the ramping populations (avg) represents the average frequency of the mutation in the populations containing this variant. Mutations with an average of 1 swept completely through the population. (n) represents the number of populations in which the mutation is present. For the 1/8 MIC populations the number indicated represents the mutation frequency found in the equimolar pool of six separate populations. For reference a mutation which swept through one population but is absent in the other 5 populations will have a frequency of 0.17 in the pooled population.

Plasmid new junctions

Rearrangement	Position	Position type	Ramping				1/8 MIC		Gene
			Delta (avg)	Delta (n)	WT (avg)	WT (n)	Delta	WT	
Deletion	33313	Left side	0.80	3	0.99	2	0.00	0.00	blaVEB-1 attC
	16493	Right side							oriT
Deletion	31705	Left side	0	0	0.57	1	0.00	0.00	dfrA5 attC
	32308	Right side							aadB attC
Inversion	29493	Left side	0.00	0	0.63	1	0.00	0.00	ResP/hypothetical protein
	5635	Right side							TrwJ
Inversion	29496	Left side	0.00	0	0.99	1	0.00	0.00	ResP/hypothetical protein
	6178	Right side							TrwJ
Inversion	14480	Left side	0	0	0	0	0.06	0.00	trwB
	15253	Right side							TrwB
Insertion	11163	Insertion site (plasmid)	0.78	1	0.00	0	0.00	0.00	TrwD
	1061199	Insert left side (chr)							PA0976.1/PA0978
	1062434	Insert right side (chr)							PA0979/PA0981
Insertion	31820	Insertion site (plasmid)	0.00	0	0.00	0	0.00	0.07	aadB
	499829	Insert left side (chr)							PA0444/PA0445
	501199	Insert right side (chr)							PA0445/PA0446

Table 4.5: List of plasmid rearrangements and their junctions in the *P. aeruginosa* populations at x1 MIC. Here (avg) represents the average frequency of the mutation in the populations containing this variant and (n) represents the number of populations in which the mutation is present for this genotype. For the 1/8 MIC populations the number indicated represents the mutation frequency found in the equimolar pool of six separate populations. The gene in which each junction is located is indicated on the right.

Chromosomal mutations

Type	Position	Mutation	Gene Name	Effect	Ramping				1/8 MIC		Gene product
					Delta (avg)	Delta (n)	WT (avg)	WT (n)	Delta	WT	
INS	821291	GGTGCCGATCAG	PA0752	coding	1.00	1	0.00	0	0	0	hypothetical protein
SNP	1501107	G	PA1382/PA1383	intergenic	1.00	1	0.00	0	0	0	type II secretion system protein/hypothetical protein
SNP	1526403	G	PA1403/PA1404	intergenic	1.00	1	0.00	0	0	0	transcriptional regulator/hypothetical protein
DEL	1911359	1	PA1767	coding	1.00	1	0.00	0	0	0	hypothetical protein
SNP	2215142	A	galU	non synonymous	1.00	1	0.00	0	0	0	uridylyltransferase
SNP	3945048	G	argG	non synonymous	0.00	0	1.00	1	0	0	argininosuccinate synthase
SNP	4503813	G	PA4022/PA4023	intergenic	1.00	1	0.00	0	0	0	aldehyde dehydrogenase/transporter
SNP	5328724	A	infB	non synonymous	1.00	1	0.00	0	0	0	translation initiation factor IF-2
SNP	5616586	C	waaL	non synonymous	0.00	0	1.00	1	0	0	O-antigen ligase WaaL
SNP	5694496	C	hslU/PA5055	intergenic	1.00	1	0.00	0	0	0	protease subunit HslU/hypothetical protein
SNP	6115211	T	mtr	non synonymous	1.00	1	0.00	0	0	0	tryptophan permease
DEL	1912260	10	PA1768	coding	0.00	0	0.93	1	0	0	hypothetical protein
DEL	1911772	11	PA1767	coding	0.89	1	0.00	0	0	0	hypothetical protein
DEL	2962464	1	clpA	coding	0.00	0	0.86	1	0	0	ATP-binding protease component ClpA
SNP	1910320	C	PA1766	non synonymous	0.00	0	0.83	1	0	0	hypothetical protein
SNP	3334558	C	rne	non synonymous	0.82	1	0.00	0	0	0	ribonuclease E
DEL	1910216	192	PA1766	coding	0.00	0	0.79	1	0	0	hypothetical protein
SNP	3778776	T	PA3368	non synonymous	0.00	0	0.77	1	0	0	acetyltransferase
DEL	1910642	345	PA1766/PA1767	TBD	0.76	1	0.00	0	0	0	PA1766/PA1767
DEL	1911619	151	PA1767	coding	0.00	0	0.76	1	0	0	hypothetical protein
SNP	4933454	C	argJ	non synonymous	0.39	1	0.32	1	0	0	acetyltransferase/N-acetylglutamate synthase
DEL	1911252	6	PA1767	coding	0.70	1	0.00	0	0	0	hypothetical protein
INS	98901	TGGCTG	fha1	coding	0.61	1	0.00	0	0	0	Fha domain-containing protein
DEL	5993105	6	algC	coding	0.58	1	0.00	0	0	0	phosphomannomutase
SNP	2480780	G	PA2252/ansA	intergenic	0.55	1	0.00	0	0	0	AGCS /L-asparaginase I
SNP	1911203	A	PA1767	nonsense	0.53	1	0.00	0	0	0	hypothetical protein
INS	1911755	G	PA1767	coding	0.49	2	0.00	0	0	0	hypothetical protein
SNP	1526418	G	PA1403/PA1404	intergenic	0.45	1	0.00	0	0	0	transcriptional regulator/hypothetical protein
SNP	4933463	G	argJ	non synonymous	0.41	1	0.00	0	0	0	acetyltransferase/N-acetylglutamate synthase
SNP	2044672	T	PA1875	synonymous	0.00	0	0.41	1	0	0	hypothetical protein
DEL	1911445	156	PA1767	coding	0.00	0	0.40	1	0	0	hypothetical protein
SNP	1240553	G	PA1147/toxA	intergenic	0.00	0	0.36	1	0	0.28	amino acid permease/exotoxin A
SNP	201563	G	aer2	non synonymous	0.00	0	0.35	1	0	0	aerotaxis transducer Aer2
SNP	1296272	C	PA1194/PA1195	intergenic	0.00	0	0.32	1	0	0	amino acid permease/hypothetical protein
SNP	4729434	C	PA4223	non synonymous	0.32	1	0.00	0	0	0	ABC transporter ATP-binding protein
DEL	1911734	12	PA1767	coding	0.00	0	0.31	1	0	0	hypothetical protein
SNP	1625028	C	PA1497	non synonymous	0.30	1	0.00	0	0	0	transporter

Table 4.6: List of chromosomal SNPs and indels in the *P. aeruginosa* populations at x1MIC. For the ramping populations (avg) represents the average frequency of the mutation in the populations containing this variant. For reference, mutations with an average frequency of 1 swept completely through the population. (n) represents the number of populations in which the mutation is present. For each variant found in the ramping populations the frequency of the corresponding mutant in the pooled 1/8MIC controls is indicated as reference.

4.6.5 Experimental evolution with MG1655:A5

Ramping experiment

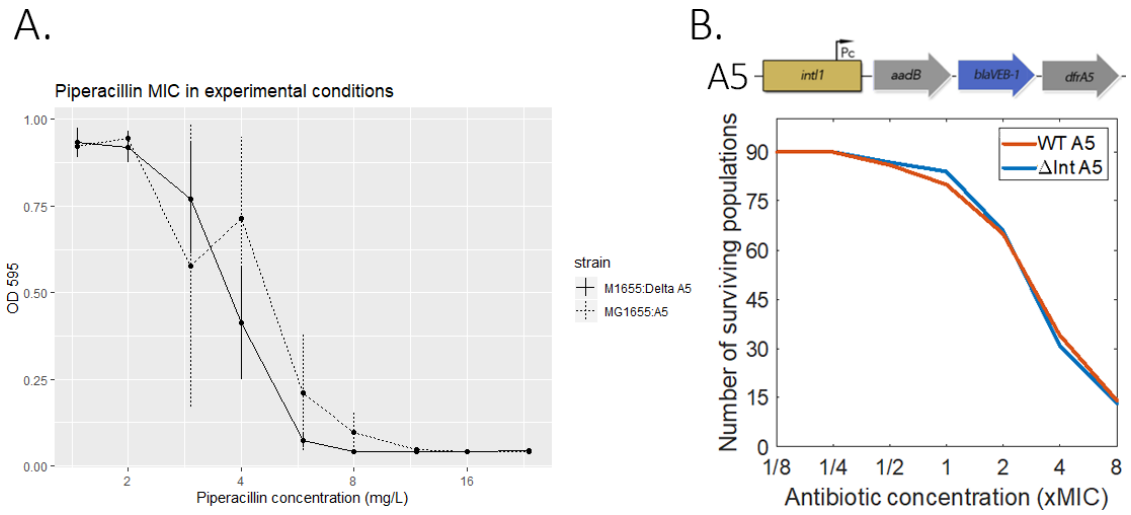


Figure 4.10: A. Growth of the MG1655:WTA5 and MG1655:ΔintA5 populations in increasing concentration of piperacillin in conditions similar to the ramping experiment.The Δint strain is indicated by the dashed line. The MIC was determined for both strains at 8 mg/L. B. Top: representation of the A5 array with *blaVEB-1* highlighted in color. Bottom: Survival curves of the MG1655:WTA5 and MG16555:ΔintA5 populations.

PCR screening

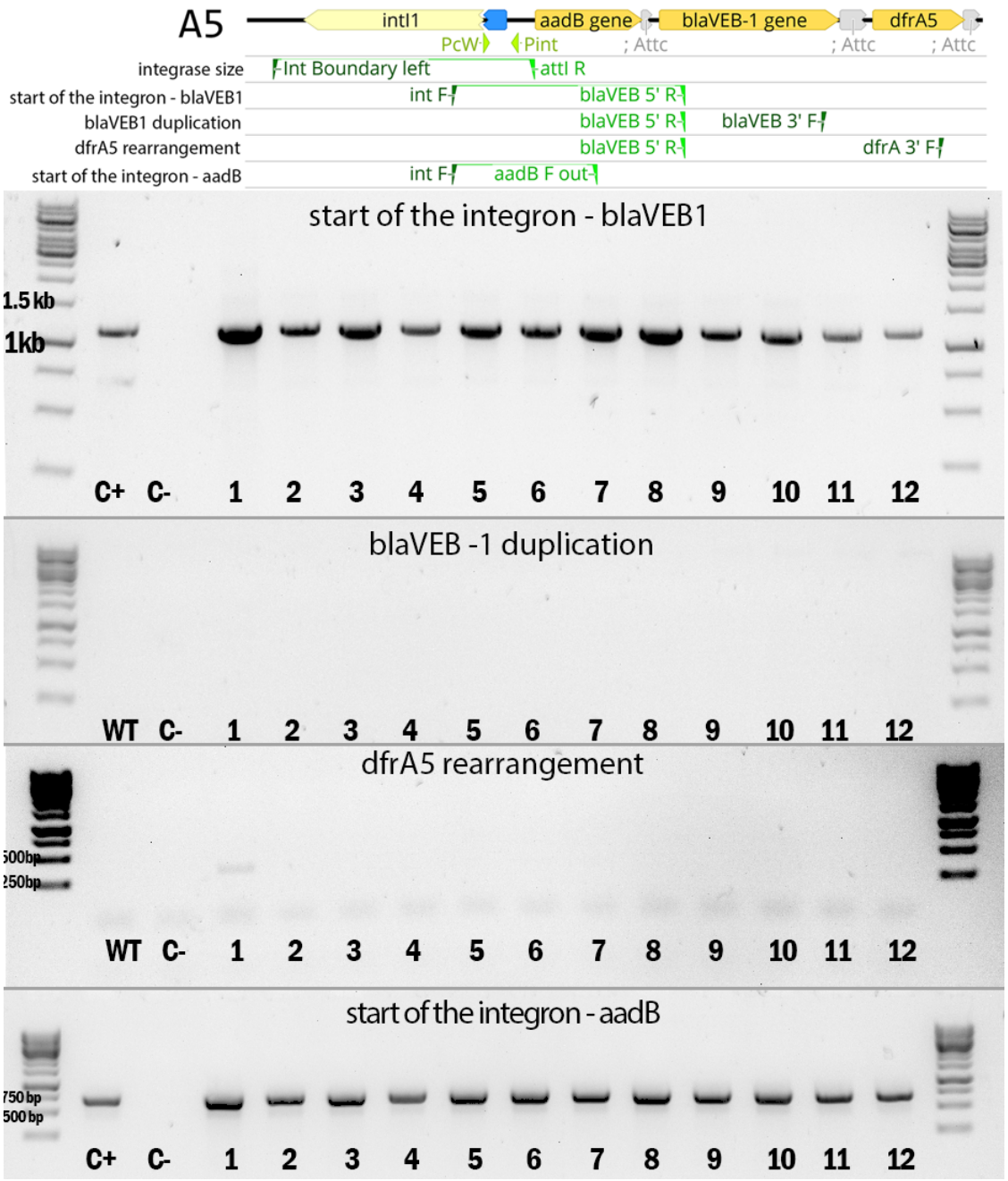


Figure 4.11: Top: Binding sites and amplicons of the PCR primers used in the *E. coli* cassette recombination screen/ Bottom: Amplicons from the PCR screen for each primer combination. Each lane numbered 1 to 12 corresponds to a separate MG1655:WTA5 population which survived up to x8 MIC. The C+/WT control is amplified from DNA from the ancestral MG1655:WTA5 strain.

5 Conclusion

5.1 Summary

Antibiotics are one of the cornerstones of modern medicine, but their use is threatened by the worldwide rise of antibiotic resistance. Solving this crisis, which is reinforced by the economic and scientific difficulties of bringing new antibiotics to the market, requires improving our management of currently available antibiotics to limit the emergence and spread of resistance. To achieve this goal, an in-depth understanding of the mechanisms and forces shaping the evolution of antibiotic resistance in bacteria is necessary, of which the integron is a prime example.

Widespread among multidrug resistant bacteria, integrons are genetic platforms allowing bacteria to acquire, excise and reorder gene cassettes, which often encode for antibiotic resistance genes. The integron can improve bacteria evolvability for antibiotic resistance in two ways: first through facilitating the acquisition of novel genes in combination with Horizontal Gene Transfer and second through the modulation of resistance levels by reshuffling the gene cassette order. As a cassette's expression depends on its position within the array, integrons have been hypothesised to provide a versatile adaptation mechanism. Costly resistance cassettes can be kept at low cost at the end of the array but are ready to be brought

forward when required for enhanced expression. I focused on this 'adaptation on demand' hypothesis and aimed to quantify the adaptability benefits given by cassette shuffling through the use of custom integron arrays and the combination of experimental evolution and whole genome sequencing.

First, I investigated the effect of cassette position on antibiotic resistance in a three cassette system at the phenotypic, transcriptional and translational level in two different clinically relevant species, *E. coli* and *P. aeruginosa*. I showed the effect of cassette position extensively varied depending on the antibiotic, the cassette and the host background, and was generally weak in a three-cassette array. I found an overall transcription gradient across positions for all cassettes, which was then reinforced or reduced by cassette-specific effects at the translational level. In particular, I showed that in *P. aeruginosa* the levels of gentamicin resistance provided by the *aadB* cassette are heavily impacted by its distance from the start of the array, due to its weak Shine-Dalgarno sequence, while the *blaVEB-1* cassette showed a reduced effect of position on piperacillin resistance. We can therefore expect varying adaptability benefits of cassette shuffling between integron arrays depending on the array cassette composition and the antibiotic considered. Moreover, I identified diverging patterns of cassette and integrase expression depending on the host background, showing that the evolutionary trajectories of mobile integrons may vary as they move across species.

Second, I quantified the adaptation benefit provided by the integrase when a distant cassette can be brought closer to the promoter, in the case of two different cassettes and two antibiotics. I applied a combination of experimental evolution and whole genome sequencing to the custom integron three-cassette system introduced previously and investigated the differences in both evolvability and genomic evolution between strains with and without a functional integrase.

In the case of the *aadB* cassette and the antibiotic gentamicin, I observe an increase in adaptability associated with extensive cassette rearrangements and

duplications in strains with a functional integrase, which confirms integrase activity can increase evolvability. I demonstrate for the first time that complex cassette rearrangements resulting from multiple recombination reactions can happen at a high frequency in the absence of integrase over-expression. I also highlight the prevalence of cassette duplications as a result of integrase activity, both in our model integron system and in a clinical plasmid. While cassette duplication has only been detected for a subset of all integron cassettes, it may have profound consequences on the evolution of antibiotic resistance genes since gene amplifications can facilitate the evolution of novel genes (Bergthorsson *et al.*, 2007). Cassette duplication also provides another pathway for cassette shuffling to modulate expression levels that does not require a cassette expression gradient; it therefore extends the potential benefits of integrase activity.

I then investigated the impact of cassette shuffling for another cassette, *blaVEB-1*, and the antibiotic piperacillin. In contrast to the previous experiment, I observed a deleterious effect of integrase activity on adaptation with limited cassette rearrangement. This result confirms the limited recombinogenic potential of the *blaVEB-1* cassette (Aubert *et al.*, 2012) and shows an evolutionary cost of integrase expression in the absence of accessible beneficial cassette rearrangement. Interestingly this deleterious effect was not visible in *E. coli*, pointing toward a species-specific integrase cost. I also identified recombinations between integrase secondary sites similar to *attI* x *attI* recombinations, showing for the first time a beneficial application of the evolutionarily elusive *attI* x *attI* reactions. Alongside the deletions originating from reactions between an *attC* site and a secondary site identified in the previous experiment, these inversions show that the consequences of integrase activity are not limited to integron cassettes but may have a wider impact on plasmid and genome evolution.

In conclusion, I found that the adaptive power of mobile integron shuffling is highly dependent on its cassette cargo. In the presence of mobile cassettes the integron can modulate cassette expression through reshuffling and duplication -

effectively providing 'adaptation on demand' against antibiotics. In their absence the integrase may also act on other regions on the genome, but with overall deleterious consequences.

5.2 Future directions

We saw in this work how integrase activity could mediate extensive cassette rearrangements, but how frequent is cassette shuffling outside the lab? We already know of one example (Hocquet *et al.*, 2012), where the excision of a *gcuF1* cassette led to enhanced ceftazidime resistance within a patient. The PcW integrase variant, which is linked with high integrase but low cassette expression, has been found to be the most frequent among class 1 integrons (Jové *et al.*, 2010) and points toward the selection of integrons with shuffling potential. Evolutionary studies in more realistic settings, such as in biofilms or infection models, can provide us with a better idea of the real costs and benefits of the integrase. It has already been shown than integrase expression and cassette excision is enhanced in biofilms in comparison to planktonic cultures (Strugeon *et al.*, 2016), while the SOS response is involved in the colonization of the mammalian gut (Samuels *et al.*, 2019), leading us to expect an impact of the integrase in both these models. Experimental evolution with biofilms model populations is well described (Steenackers *et al.*, 2016), frequently used to study antibiotic resistance evolution (Santos-Lopez *et al.*, 2019), and could easily be applied to create a similar study. We can expect infections models to be more challenging, due to the difficulty and ethical concers of creating long-term bacterial infection in animals models (Pletzer *et al.*, 2017). Small difference in evolvability would also be more difficult to identify, due to the small sample size coming with this type of studies.

Another approach would be to follow the dynamics of integrons present within patients microbiome through longitudinal studies. This would help us bridging the gap between our extensive knowledge of integron biology coming from lab experimental and epidemiological sampling studies, which both have their limitations. The former can only give us an idea of the *potential* benefits of integron activity in

the clinic, while the latter only provides a snapshot of the integron's prevalence in the environment, without knowledge of its dynamics.

Alongside giving us hindsight on the prevalence of cassette shuffling, longitudinal studies would also provide us with precious information on the rates of cassette acquisition and integron dispersion occurring through horizontal gene transfer within patients. Indeed, the combination of the integrase enzyme and horizontal gene transfer has great adaptive potential, since the integrase can allow the acquisition of resistance genes at a lower cost and a greater frequency than other genomic mechanisms. For example, the integrase can easily facilitate the creation of multidrug resistant plasmids by promoting the movement of cassettes from chromosomal integrons to integrons located on plasmids (Rowe-Magnus *et al.*, 2002) or the transfer of cassettes from mobile integron to mobile integron. In bacteria containing several resistance plasmids of varying fitness costs, this may allow the stockpiling of resistance cassettes on the less costly plasmid. This can reduce the fitness cost of multidrug resistance, as the cost of an additional cassette is less than the cost of an entire additional plasmid (Lacotte *et al.*, 2017). In combination with transformation, the integrase allows the insertion of novel genes without the need for homologous sequences, facilitating the acquisition of genes from foreign species (Domingues *et al.*, 2012a). Such studies would also allow us to investigate the constraints around cassette number per array in class 1 integrons: unlike the 'super-integron' present in the *Vibrio* chromosome, the number of cassettes in the majority of mobile integrons arrays remains relatively low (three cassettes or less in more than 70% of mobile integrons (Zhang *et al.*, 2018)). This could either be due to a limited rate of cassette acquisition or due to strong constraints on the number of cassettes, leading to frequent cassette turn-over.

Most of integron evolutionary studies have focused so far on selection at the integrase level (Engelstädter *et al.* (2016),Starikova *et al.* (2012)). However, cassettes, as mobile genetic elements with varying characteristics, will also be subject to their

own level of selection. Cassette mobility, for example, is an easily mutable trait through changes in the *attC* site (Loot *et al.*, 2017) which we can expect to be the target of adaptation pressure. High cassette mobility could help the spread of a cassette to new host and new environment but also increase the risk of cassette loss in stressful environments if the cassette does not provide a selective advantage. Highly mobile cassettes may mitigate this risk by more frequently occupying the first position of the array, where they should be less subject to excision, due to the preference of the *attI* x *attC* reaction toward insertion rather than excision, unlike the *attC* x *attC* reaction. The benefits of a high mobility may also depend on the rest of the cassette characteristics, such as its fitness cost or the effect of position on its expression. So far, no obvious correlation has been observed between the recombinogenic activity of the *attC* site and the fitness cost of the cassette (Lacotte *et al.*, 2017), but more in-depth investigation of the cassette characteristics landscape would allow us to identify more complex relationships. For this, the recent development of genomic tools to identify novel cassette and integrons from genomic data (Cury *et al.* (2016), Zhang *et al.* (2018)) is especially useful, as we now have access to new sources of integron cassette sequences.

Interestingly the existence of different transmission dynamics between the static integron platform and mobile cassettes could lead to potential genomic conflict between the two (Werren, 2011): integron cassettes may possess a certain degree of selfishness, enhancing their own transmission or prevalence to the detriment of the rest of the integron. Similarly to other mobile genetic elements such as transposons or insertion sequences (Werren, 2011) we can imagine 'selfish' integrons cassettes providing little to no benefits for its bacteria host and the integron platform but with high mobility and excision rate, and therefore able to jump and spread efficiently141 between integrons. Interactions between cassettes may also strongly impact the cassette selection dynamics. We saw in this work an enhanced expression of *blaVEB-1* when located after *dfrA5* in *E. coli*. A recent study has even shown the existence of ORF-less cassettes containing promoters which enhance the expression of neighbouring cassettes (Tansirichaiya *et al.*, 2019), cassettes which must therefore

rely on co-selection with other cassettes to ensure their spread. As integron arrays often contain multiple resistance cassettes, they are also a prime spot for collateral sensitivities or synergies between the different resistance mechanisms encoded by the cassettes, which may lead to cassettes co-occurring more or less frequently together.

The study of cassette dynamics could therefore strongly benefit from further theoretical models of the integron. So far, the sole modelling study on the integron modelled cassettes as interchangeable cargo only characterized by the phenotype they provided (Engelstädter *et al.*, 2016). Expanding this work by considering cassettes with different excision and expression patterns would provide us with valuable insights on the forces governing cassette prevalence and dispersion.

Finally, understanding the integron and the cassette's evolutionary dynamics is crucial if we want to understand, and potentially restrict, the integron's contribution to AMR.

Given integrons often contain multiple resistance genes which can be co-selected, the presence of integrons severely threatens the reversibility of resistance once the use of an antibiotic is stopped. This has already been seen with the antibiotic trimethoprim: only a marginal reduction in trimethoprim resistance was observed in *E. coli* isolated from patients' urinatery tracts after an 85% reduction in trimethoprim prescriptions in Sweden (Sundqvist *et al.*, 2010). The main resistance mechanism was identified as *dfr* genes located on integrons, whose prevalence remained constant during the two years of the study. Similarly, the integron-associated *sul1* gene is partly responsible for the persistence of sulphonamide resistance in the UK despite a 10-year ban (Enne *et al.*, 2001).

The extensive prevalence of the mobile integron in the environment - to the point it has been described as an 'invasive species' (Gillings, 2017) - makes the integron a high risk vector for the spread of novel resistance genes, given their potential as hot spots for gene exchange between species and environments. The mechanisms leading to the creation of new integron cassettes remain to be determined (Escudero *et al.*,

2015), and so the risk of novel resistance determinant emerging as integron cassettes is unknown, but new resistance cassettes against newly discovered antibiotics could easily spread across species to human bacterial pathogens, speeding up the rise of resistance. The same integron arrays have already been observed in multiple bacteria species (Domingues *et al.*, 2012b). In the lab, movement of integrons between species has been observed, with cassettes going from *V. cholerae* to *E. coli* (Rowe-Magnus *et al.*, 2002) or from multiple Gram-negative species to *Acinetobacter baylyi* (Domingues *et al.*, 2012b).

How can we counter integrons? Limiting the prevalence of mobile integrons in the environment will only be achievable through reducing the overall selection pressure for resistance as the prevalence of the class 1 integron integrase has been shown to be closely linked to anthropogenic pollution (Gillings *et al.*, 2015). Wastewater treatment plans, for example, can play a key role in the spread of integrons in effluents (An *et al.*, 2018).

In clinical settings a promising approach can be found in the development of adjuvant drugs which do not kill bacteria but instead limit the evolution of resistance. Examples are *LexA* inhibitors (Mo *et al.*, 2018), which would stop the induction of the SOS response, or conjugation inhibitors (Buckner *et al.* (2018), Getino *et al.* (2016)), which limit the exchange of cassettes between plasmids. Finally CRISPR-Cas nucleases have been proposed as a potential strategy to specifically target resistance genes (Bikard *et al.*, 2014) and could also be used to target the integron's integrase.

Starting as ancient bacterial structures, mobile integrons became mechanisms uniquely suited to the challenges posed to bacteria by the human use of antibiotics. Until we change our constant need for antibiotics in both modern medicine and agriculture, mobile integrons will remain challenging components of bacteria's adaptive arsenal.

List of Primers

Chapter 2

Primer name	Primer sequence	Description/Function	Primer origin	Efficiency
Creation of Pseudomonas strains				
Ecoli E F	CGCCGTCTGGTGATGTAAGT	E. coli specific primers	Rice (2009)	
Ecoli E R	CAACGGAAGGCAGCGGAGTA		Rice (2009)	
PA431 F	CTGGGTCGAAAGGTGGTTGTTATC	P. aeruginosa specific primers	Choi *et al.* (2013)	
PA431 R	GCGGCTGGTGCGGCTGAGTC		Choi *et al.* (2013)	
Left side R388	AGCCAAGCCGAGAAGAAAGC	Plasmid specific primers	Jose Antonio Escudero	
atti R	AACTTTGTTTTAGGGCGACTGCC		Jose Antonio Escudero	
Plasmid copy number determination				
orf9 F	GGCGTGACGTTTCAAAAGCT	Plasmid target	This work	0.997
orf9 R	CTTGAGCAACAACAACGCGA		This work	
orf9 probe	GTAGGTGTAGCCGTTGCCTT		This work	
rpoD F	GGGCGAAGAAGGAAATGGTC	P. aeruginosa chromosome target	Savli *et al.* (2003)	0.923
rpoD R	CAGGTGGCGTAGGTGGAGAA		Savli *et al.* (2003)	
rpoD probe	ATCCAGGAAGGCAACATCGG		This work	
qssrA F	ACGGGGATCAAGAGAGGTCAAAC	E. coli reference target	Moura de Sousa *et al.* (2017)	0.962
qssrA R	CGGACGGACACGCCACTAAC		Moura de Sousa *et al.* (2017)	
Cassette transcription				
acpP F	ACTCGGCGTGAAGGAAGAAG	Chromosomal reference genes (P. aeruginosa)	Lenz *et al.* (2008)	1.037
acpP R	CGACGGTGTCAAGGGAGT		Lenz *et al.* (2008)	
atpA F	TACGCCAGATCACGCCCGGA		Qi *et al.* (2014)	0.839
atpA R	GGCGGCGTCTACGGTATGGC		Qi *et al.* (2014)	
qrrsA F	ACGGGGATCAAGAGAGGTCAAAC	Chromosomal reference genes (E. coli)	Moura de Sousa *et al.* (2017)	0.936
qrrsA R	CGGACGGACACGCCACTAAC		Moura de Sousa *et al.* (2017)	
qssrA F	as above			
qssrA R	as above			
qaadB F	TAGCGGAGATCGGGGATGAG	aadB transcription	This work	0.990
qaadB R	CCTCCCAGCTGTTACAACGG		This work	
qblaVEB F	CGACTTCCATTTCCCGATGC	blaVEB-1 transcription	This work	0.973
qblaVEB R	TCGGACTCCACGTTTTAGGC		This work	
qdfrA F	CAACCAGTGGCTTTTGGTGG	dfrA5 transcription	This work	0.990
qdfrA R	ACATGGCCTCTTCGATCGAC		This work	
Integrase expression				
INTI1-RT1	TTCATCCGTTTCCACGGTGTGCG	Integrase transcription	Hocquet *et al.* (2012)	0.878
INTI1-RT2	GAGACCGAAACCTTGCGCTCGTT		Hocquet *et al.* (2012)	
LEXA-RT1	CTCCTTCATCAAGCGCTGCCTGG	LexA transcription	Hocquet *et al.* (2012)	0.7958
LEXA-RT2	GTTCGGCGACTTGAAGCCGAGTT		Hocquet *et al.* (2012)	
qdfrA F	as above	Plasmid reference gene		
qdfrA R	as above			
acpP F	as above	Chromosomal reference gene (P. aeruginosa)		
acpP R	as above			
qssrA F	as above	Chromosomal reference gene (E. coli)		
qssrA R	as above			

Table 5.1

Chapter 3

Primer name	Primer sequence	Description/Function	Primer origin
Identification of cassette rearrangements in PAO1:A3			
integrase F	CTATAGCAGCGGAGGGGTTG	Integrase size	This work
integrase R	CGTTCGGTCAAGGTTCTGGA		This work
aadB 3' F	GGAAAAGGTTGAGGTCTTGC	outward facing aadB primers	Jose Antonio Escudero
aadB 5' R	CAGATTTCGCTCATCTGCC		Jose Antonio Escudero
Right side R388	CTGCAAGTGATCTTGAAGCC	End of the integron	Jose Antonio Escudero
pAMBL rearrangement analysis			
blaVIM1 3'F	GAAGCAGAGGTCGTCATTCC	outward facing blaVIM primers	This work
blaVIM1 5'R	GAATGGGCTAACGGACTTGC		This work
tniA F	GCAAGGCTCAAAAACTGCGA	pAMBL2 specific primers	This work
tniA R	CCGATGATCCGTTCCACGAT		This work

Table 5.2

Chapter 4

Primer name	Primer sequence	Description/Function	Primer origin
Identification of cassette rearrangements			
int Boundary left	TATAGCAGCGGAGGGGTTGG	Integrase size	Jose Antonio Escudero
attI R	as above		
int F	CCGTAGAAGAACAGCAAGGC	Start of the array	Jose Antonio Escudero
blaVEB 5'R	CCTATTCTGGCATTTTTTGCC	outward facing blaVEB-1 primers	Jose Antonio Escudero
blaVEB 3'F	CAGAGTCCAAAGAAACTTCGG		Jose Antonio Escudero
dfrA 3'F	GATTGATATTGAGCCGGAAGG	outward facing dfrA5 primer	Jose Antonio Escudero
aadB F out	GGAGCCTCCGCGATTTCATA	outward facing aadB primer	This work
Identification of plasmid backbone inversion and deletions			
integron rev	TTGCCTGATGGATCCAACCC	Inversion identification	This work
trwG	AGCAGAGCAAGGGCAATCTT		This work
stbA	ATCTTTCAGGGCGTCCGTTT	Deletion identification	This work
blaVEB 3'F	as above		

Table 5.3

List of Strains

Name	Description	Origin
Strains		
PAO1	P. aeruginosa strain PAO1	
MG1655	E. coli strain MG1655	
Plasmids		
WTA1	Custom integron array drfA5 - blaVEB1 - aadB on R388 plasmid backbone	Jose Antonio Escudero
WTA2	Custom integron array blaVEB1 - aadB - dfrA5 on R388 plasmid backbone	Jose Antonio Escudero
WTA3	Custom integron array blaVEB1 - dfrA5 - aadB on R388 plasmid backbone	Jose Antonio Escudero
WTA4	Custom integron array dfrA5 - aadB - blaVEB1 on R388 plasmid backbone	Jose Antonio Escudero
WTA5	Custom integron array aadB - blaVEB1 - dfrA5 on R388 plasmid backbone	Jose Antonio Escudero
WTA6	Custom integron array aadB - dfrA5 - blaVEB1 on R388 plasmid backbone	Jose Antonio Escudero
ΔintA3	Array WTA3 with a 948bp deletion of the integrase IntI1	Jose Antonio Escudero
ΔintA4	Array WTA4 with a 948bp deletion of the integrase IntI1	Jose Antonio Escudero
ΔintA5	Array WTA5 with a 948bp deletion of the integrase IntI1	Jose Antonio Escudero
pAMBL1	Clinical plasmid containing a blaVIM-aadB array	San Millan *et al.* (2015b)
pAMBL2	Clinical plasmid containing a blaVIM - aacA4 - blaVIM - blaVIM - aadA1 array	San Millan *et al.* (2015b)

Table 5.4

Bibliography

Abeyrathne PD, Daniels C, Poon KK, Matewish MJ, and Lam JS (2005) Functional characterization of WaaL, a ligase associated with linking O-antigen polysaccharide to the core of Pseudomonas aeruginosa lipopolysaccharide. *Journal of Bacteriology* 187(9):3002–3012.

Alkan C, Sajjadian S, and Eichler EE (2011) Limitations of next-generation genome sequence assembly. *Nature methods* 8(1):61–5.

Aminov RI (2010) A brief history of the antibiotic era: Lessons learned and challenges for the future. *Frontiers in Microbiology* 1(DEC).

Amyes SG, Towner KJ, *et al.* (1990) Trimethoprim resistance; Epidemiology and molecular aspects. *Journal of Medical Microbiology* 31(1):1–19.

An XL, Chen QL, *et al.* (2018) Impact of wastewater treatment on the prevalence of integrons and the genetic diversity of integron gene cassettes. *Applied and Environmental Microbiology* 84(9).

Andersson DI and Hughes D (2009) Gene Amplification and Adaptive Evolution in Bacteria. *Annual Review of Genetics* 43(1):167–195.

Andersson DI and Levin BR (1999) The biological cost of antibiotic resistance.

Andrews S, Krueger F, *et al.* (2012) Fastqc. Babraham Institute.

Antipov D, Hartwick N, *et al.* (2016) plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 32(22):btw493.

Aubert D, Naas T, and Nordmann P (2012) Integrase-mediated recombination of the veb1 gene cassette encoding an extended-spectrum β-lactamase. *PloS one* 7(12):e51602.

Avila P and de la Cruz F (1988) Physical and genetic map of the IncW plasmid R388. *Plasmid* 20(2):155–157.

Baharoglu Z, Bikard D, and Mazel D (2010) Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLoS genetics* 6(10):e1001165.

Baharoglu Z, Krin E, and Mazel D (2012) Connecting environment and genome plasticity in the characterization of transformation-induced SOS regulation and carbon catabolite control of the Vibrio cholerae integron integrase. *Journal of bacteriology* 194(7):1659–67.

Baharoglu Z, Krin E, Mazel D, Kulasekara H, and Bertrand X (2013) RpoS Plays a Central Role in the SOS Induction by Sub-Lethal Aminoglycoside Concentrations in Vibrio cholerae. *PLoS Genetics* 9(4):e1003421.

Baharoglu Z and Mazel D (2011a) Vibrio cholerae triggers SOS and mutagenesis in response to a wide range of antibiotics: a route towards multiresistance. *Antimicrobial agents and chemotherapy* 55(5):2438–41.

Baharoglu Z and Mazel D (2011b) Vibrio cholerae triggers SOS and mutagenesis in response to a wide range of antibiotics: a route towards multiresistance. *Antimicrobial agents and chemotherapy* 55(5):2438–41.

Baharoglu Z and Mazel D (2014a) Influence of very short patch mismatch repair on SOS inducing lesions after aminoglycoside treatment in Escherichia coli. *Research in Microbiology* 165(6):476–480.

Baharoglu Z and Mazel D (2014b) SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiology Reviews* 38(6):1126–1145.

Barber M (1947) Staphylococcal Infection due to Penicillin-resistant Strains. *British Medical Journal* 2(4534):863.

Barraud O and Ploy MC (2015) Diversity of Class 1 Integron Gene Cassette Rearrangements Selected under Antibiotic Pressure. *Journal of bacteriology* 197(13):2171–8.

Barrick JE, Colburn G, *et al.* (2014) Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics* 15(1):1039.

Baym M, Lieberman TD, *et al.* (2016) Spatiotemporal microbial evolution on antibiotic landscapes. *Science* 353(6304):1147–1151.

Beardmore RE and Peña-Miller R (2010) Rotating antibiotics selects optimally against antibiotic resistance, in theory. *Mathematical Biosciences and Engineering* 7(3):527–552.

Becker B and Cooper MA (2013) Aminoglycoside antibiotics in the 21st century. *ACS Chemical Biology* 8(1):105–115.

Bergthorsson U, Andersson DI, and Roth JR (2007) Ohno's dilemma: Evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences of the United States of America* 104(43):17004–17009.

Bikard D, Euler CW, *et al.* (2014) Exploiting CRISPR-cas nucleases to produce sequence-specific antimicrobials. *Nature Biotechnology* 32(11):1146–1150.

Bissonnette L, Champetier S, Buisson JP, and Roy PH (1991) Characterization of the nonenzymatic chloramphenicol resistance (cmlA) gene of the In4 integron of Tn1696: similarity of the product to transmembrane transport proteins. *Journal of bacteriology* 173(14):4493–502.

Blahna MT, Zalewski CA, *et al.* (2006) The role of horizontal gene transfer in the spread of trimethoprim–sulfamethoxazole resistance among uropathogenic Escherichia coli in Europe and Canada. *Journal of Antimicrobial Chemotherapy* 57(4):666–672.

Blount ZD (2015) The unexhausted potential of E. coli.

Bolger AM, Lohse M, and Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 30(15):2114–20.

Boucher H, Talbot G, *et al.* (2009) Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America. *Clinical Infectious Diseases* 48(1):1–12.

Bouvier M, Demarre G, and Mazel D (2005) Integron cassette insertion: a recombination process involving a folded single strand substrate. *The EMBO journal* 24(24):4356–67.

Bouvier M, Ducos-Galand M, *et al.* (2009) Structural Features of Single-Stranded Integron Cassette attC Sites and Their Role in Strand Selection. *PLoS Genetics* 5(9):e1000632.

Breidenstein EB, de la Fuente-Núñez C, and Hancock RE (2011) Pseudomonas aeruginosa: all roads lead to resistance. *Trends in Microbiology* 19(8):419–426.

Brent R (1982) Regulation and autoregulation by lexA protein. *Biochimie* 64(8-9):565–569.

Brogden RN, Carmine AA, Heel RC, Speight TM, and Avery GS (1982) Trimethoprim: A Review of its Antibacterial Activity, Pharmacokinetics and Therapeutic Use in Urinary Tract Infections. *Drugs* 23(6):405–430.

Brolund A, Sundqvist M, Kahlmeter G, and Grape M (2010) Molecular characterisation of trimethoprim resistance in Escherichia coli and Klebsiella pneumoniae during a two year intervention on trimethoprim use. *PloS one* 5(2):e9233.

Brook I (1989) Inoculum Effect. *Reviews of Infectious Diseases* 11(3):361–368.

Bruger EL and Marx CJ (2018) A decade of genome sequencing has revolutionized studies of experimental evolution.

Brynildsrud O (2018) Read Depth Analysis to Identify CNV in Bacteria Using CNOGpro. 73–81. Humana Press, New York, NY.

Buckling A, Craig MacLean R, Brockhurst MA, and Colegrave N (2009) The Beagle in a bottle.

Buckner MM, Ciusa ML, and Piddock LJ (2018) Strategies to combat antimicrobial resistance: Anti-plasmid and plasmid curing.

Bush K and Bradford PA (2016a) β-lactams and β-lactamase inhibitors: An overview. *Cold Spring Harbor Perspectives in Medicine* 6(8).

Bush K and Bradford PA (2016b) β-lactams and β-lactamase inhibitors: An overview. *Cold Spring Harbor Perspectives in Medicine* 6(8):a025247.

Cambray G, Guerout AM, and Mazel D (2010a) Integrons. *http://dx.doi.org/10.1146/annurev-genet-102209-163504* .

Cambray G, Guerout AM, and Mazel D (2010b) Integrons. *Annual Review of Genetics* 44(1):141–166.

Card KJ, LaBar T, Gomez JB, and Lenski RE (2019) Historical contingency in the evolution of antibiotic resistance after decades of relaxed selection. *PLOS Biology* 17(10):e3000397.

César CE, Machón C, de la Cruz F, and Llosa M (2006) A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences. *Molecular Microbiology* 62(4):984–996.

Chevalier S, Bouffartigues E, *et al.* (2017) Structure, function and regulation of Pseudomonas aeruginosa porins. *FEMS Microbiology Reviews* 41(5):698–722.

Choi HJ, Kim MH, *et al.* (2013) Improved PCR for identification of Pseudomonas aeruginosa. *Applied Microbiology and Biotechnology* 97(8):3643–3651.

CIDRAP (2019) Antibiotic developer melinta files for bankruptcy.

Cirz RT, O'Neill BM, Hammond JA, Head SR, and Romesberg FE (2006) Defining the Pseudomonas aeruginosa SOS response and its role in the global response to the antibiotic ciprofloxacin. *Journal of bacteriology* 188(20):7101–10.

Clift C (2019) Review of Progress on Antimicrobial Resistance: background and analysis (October).

CLSI (2017) *Performance Standards for Antimicrobial Susceptibility Testing.* 27th edition.

Collis CM and Hall RM (1992a) Gene cassettes from the insert region of integrons are excised as covalently closed circles. *Molecular microbiology* 6(19):2875–85.

Collis CM and Hall RM (1992b) Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. *Journal of bacteriology* 174(5):1574–85.

Collis CM and Hall RM (1995) Expression of antibiotic resistance genes in the integrated cassettes of integrons. *Antimicrobial agents and chemotherapy* 39(1):155–62.

Conly JM and Johnston BL (2005) Where are all the new antibiotics? The new antibiotic paradox.

Cox G, Stogios PJ, Savchenko A, and Wright GD (2015) Structural and molecular basis for resistance to aminoglycoside antibiotics by the adenylyltransferase ANT(2)-Ia. *mBio* 6(1).

Coyne S, Guigon G, Courvalin P, and Périchon B (2010) Screening and quantification of the expression of antibiotic resistance genes in Acinetobacter baumannii with a microarray. *Antimicrobial agents and chemotherapy* 54(1):333–40.

Cury J, Jové T, Touchon M, Néron B, and Rocha EP (2016) Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic acids research* 44(10):4539–50.

Cystic Fibrosis Foundation (2013) Cystic Fibrosis Foundation Patient Registry: 2012 Annual Data Report. Technical report, Bethesda, MD.

D 'Costa VM, King CE, *et al.* (2011) Antibiotic resistance is ancient. *Nature* 477.

da Fonseca ÉL and Vicente ACP (2012) Functional characterization of a Cassette-specific promoter in the class 1 integron-associated qnrVC1 gene. *Antimicrobial agents and chemotherapy* 56(6):3392–4.

David B and Cohen D (1976) Basic Characterization of W-pili (1973).

Davies J and Davies D (2010) Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews* 74(3):417–433.

de Kraker MEA, Stewardson AJ, and Harbarth S (2016) Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? *PLOS Medicine* 13(11):e1002184.

De La Cruz F, Frost LS, Meyer RJ, and Zechner EL (2010) Conjugative DNA metabolism in Gram-negative bacteria.

de Paz HD, Sangari FJ, *et al.* (2005) Functional interactions between type IV secretion systems involved in DNA transfer and virulence. *Microbiology* 151(11):3505–3516.

Dean CR and Goldberg JB (2002) Pseudomonas aeruginosa galU is required for a complete lipopolysaccharide core and repairs a secondary mutation in a PA103 (serogroup O11) wbpM mutant . *FEMS Microbiology Letters* 210(2):277–283.

Deatherage DE and Barrick JE (2014) Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods in molecular biology (Clifton, N.J.)* 1151:165–88.

Dettman JR, Rodrigue N, and Kassen R (2015) Genome-Wide Patterns of Recombination in the Opportunistic Human Pathogen Pseudomonas aeruginosa. *Genome Biology and Evolution* 7(1):18–34.

Domingues S, Harms K, *et al.* (2012a) Natural Transformation Facilitates Transfer of Transposons, Integrons and Gene Cassettes between Bacterial Species. *PLoS Pathogens* 8(8):e1002837.

Domingues S, Harms K, *et al.* (2012b) Natural transformation facilitates transfer of transposons, integrons and gene cassettes between bacterial species. *PLoS pathogens* 8(8):e1002837.

Dörr T, Lewis K, and Vulić M (2009) SOS response induces persistence to fluoroquinolones in Escherichia coli. *PLoS Genetics* 5(12).

Dötsch A, Becker T, *et al.* (2009) Genomewide identification of genetic determinants of antimicrobial drug resistance in Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy* 53(6):2522–2531.

Draper O, César CE, Machón C, De La Cruz F, and Llosa M (2005) Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. *Proceedings of the National Academy of Sciences of the United States of America* 102(45):16385–16390.

Dutta D, Shatalin K, Epshtein V, Gottesman ME, and Nudler E (2011) Linking RNA polymerase backtracking to genome instability in E. coli. *Cell* 146(4):533–543.

Dwyer DJ, Collins JJ, and Walker GC (2015) Unraveling the Physiological Complexities of Antibiotic Lethality. *Annual Review of Pharmacology and Toxicology* 55(1):313–332.

ECDC/EMEA (2009) *A call to narrow the gap between multidrug-resistant bacteria in the EU and the development of new antibacterial agents - Joint Technical Report.*

Engelstädter J, Harms K, and Johnsen PJ (2016) The evolutionary dynamics of integrons in changing environments. *The ISME journal* .

Enne VI, Livermore DM, Stephens P, and Hall LM (2001) Persistence of sulphonamide resistance in Escherichia coli in the UK despite national prescribing restriction. *Lancet* 357(9265):1325–1328.

Escudero JA, Loot C, Nivina A, and Mazel D (2015) The Integron: Adaptation On Demand. *Microbiology spectrum* 3(2):MDNA3–0019–2014.

Escudero JA, Loot C, *et al.* (2016) Unmasking the ancestral activity of integron integrases reveals a smooth evolutionary transition during functional innovation. *Nature Communications* 7:10937.

Espah Borujeni A, Cetnar D, *et al.* (2017) Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Research* 45(9):5437–5448.

Ewels P, Magnusson M, Lundin S, and Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32(19):3047–3048.

Fernández-Lópe R, del Campo I, Revilla C, Cuevas A, and de la Cruz F (2014) Negative Feedback and Transcriptional Overshooting in a Regulatory Network for Horizontal Gene Transfer. *PLoS Genetics* 10(2).

Fernández-López R, Pilar Garcillán-Barcia M, *et al.* (2006) Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiology Reviews* 30(6):942–966.

Fetar H, Gilmour C, *et al.* (2011) mexEF-oprN multidrug efflux operon of Pseudomonas aeruginosa: Regulation by the MexT activator in response to nitrosative stress and chloramphenicol. *Antimicrobial Agents and Chemotherapy* 55(2):508–514.

Filloux A (2011) Protein secretion systems in Pseudomonas aeruginosa: An essay on diversity, evolution, and function. *Frontiers in Microbiology* 2(JULY).

Fleming A (1964) Sir Alexander Fleming - Nobel Lecture: Penicillin.

Fonseca ÉL, Vieira VV, Cipriano R, and Vicente AC (2005) Class 1 integrons in Pseudomonas aeruginosa isolates from clinical settings in Amazon region, Brazil. *FEMS Immunology and Medical Microbiology* 44(3):303–309.

Francia MV, de la Cruz F, and Lobo JM (1993) Secondary sites for integration mediated by the Tn21 integrase. *Molecular Microbiology* 10(4):823–828.

Francia MV, Zabala JC, De La Cruz F, and García Lobo JM (1999) The IntI1 integron integrase preferentially binds single-stranded DNA of the attC site. *Journal of Bacteriology* 181(21):6844–6849.

Frieden T (2013) Antibiotic resistance threats in the United States. *Centers for Disease Control and Prevention* 114.

Frimmersdorf E, Horatzek S, Pelnikevich A, Wiehlmann L, and Schomburg D (2010) How Pseudomonas aeruginosa adapts to various environments: a metabolomic approach. *Environmental Microbiology* 12(6):1734–1747.

Frost LS, Leplae R, Summers AO, and Toussaint A (2005) Mobile genetic elements: The agents of open source evolution.

Frumerie C, Ducos-Galand M, Gopaul DN, and Mazel D (2010) The relaxed requirements of the integron cleavage site allow predictable changes in integron target specificity. *Nucleic acids research* 38(2):559–69.

Gassama A, Aïdara-Kane A, Chainier D, Denis F, and Ploy MC (2004) Integron-Associated Antibiotic Resistance in Enteroaggregative and Enteroinvasive Escherichia coli. *Microbial Drug Resistance* 10(1):27–30.

Getino M and de la Cruz F (2014) Natural and Artificial Strategies To Control the Conjugative Transmission of Plasmids. *Microbiology Spectrum* 2(3):1–16.

Getino M, Fernández-López R, *et al.* (2016) Tanzawaic Acids, a Chemically Novel Set of Bacterial Conjugation Inhibitors. *PLOS ONE* 11(1):e0148098.

Gillings M, Boucher Y, *et al.* (2008) The evolution of class 1 integrons and the rise of antibiotic resistance. *Journal of bacteriology* 190(14):5095–100.

Gillings MR (2014) Integrons: past, present, and future. *Microbiology and molecular biology reviews : MMBR* 78(2):257–77.

Gillings MR (2017) Class 1 integrons as invasive species. *Current Opinion in Microbiology* 38:10–15.

Gillings MR, Gaze WH, *et al.* (2015) Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *The ISME Journal* 9(6):1269–1279.

Gillings MR, Holley MP, Stokes HW, and Holmes AJ (2005) Integrons in Xanthomonas: A source of species genome diversity. *Proceedings of the National Academy of Sciences of the United States of America* 102(12):4419–4424.

Gonzalez-Padilla M, Torre-Cisneros J, *et al.* (2015) Gentamicin therapy for sepsis due to carbapenem-resistant and colistin-resistant Klebsiella pneumoniae. *Journal of Antimicrobial Chemotherapy* 70(3):905–913.

Gotoh H, Kasaraneni N, Devineni N, Dallo SF, and Weitao T (2010) SOS involvement in stress-inducible biofilm formation. *Biofouling* 26(5):603–611.

Gravel A, Fournier B, and Roy PH (1998) DNA complexes obtained with the integron integrase IntI1 at the attI1 site. *Nucleic acids research* 26(19):4347–55.

Group WHOA, Surveillance I, and Resistance A () *WHO | WHO list of Critically Important Antimicrobials (CIA).*

Guerin E, Cambray G, *et al.* (2009) The SOS response controls integron recombination. *Science (New York, N.Y.)* 324(5930):1034.

Guérin E, Jové T, Tabesse A, Mazel D, and Ploy MC (2011) High-level gene cassette transcription prevents integrase expression in class 1 integrons. *Journal of bacteriology* 193(20):5675–82.

Gutu AD, Sgambati N, *et al.* (2013) Polymyxin resistance of Pseudomonas aeruginosa phoQ mutants is dependent on additional two-component regulatory systems. *Antimicrobial Agents and Chemotherapy* 57(5):2204–2215.

Hall BG and Barlow M (2004) Evolution of the serine beta-lactamases: past, present and future. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy* 7(2):111–23.

Hall RM, Brookes DE, and Stokes HW (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):1941–59.

Hanau-Berçot B, Podglajen I, Casin I, and Collatz E (2002) An intrinsic control element for translational initiation in class 1 integrons. *Molecular Microbiology* 44(1):119–130.

Händel N, Schuurmans JM, Brul S, and Ter Kuilea BH (2013) Compensation of the metabolic costs of antibiotic resistance by physiological adaptation in escherichia coli. *Antimicrobial Agents and Chemotherapy* 57(8):3752–3762.

Harms K, Starikova I, and Johnsen PJ (2013) Costly Class-1 integrons and the domestication of the the functional integrase. *Mobile genetic elements* 3(2):e24774.

Higgins S, Heeb S, *et al.* (2018) Differential Regulation of the Phenazine Biosynthetic Operons by Quorum Sensing in Pseudomonas aeruginosa PAO1-N. *Frontiers in cellular and infection microbiology* 8:252.

Hishida T, Han YW, *et al.* (2004) Role of the Esicherichia coli RecQ DNA helicase in SOS signaling and genome stabiliziation at stalled replication forks. *Genes and Development* 18(15):1886–1897.

Ho BT, Basler M, and Mekalanos JJ (2013) Type 6 secretion system-mediated immunity to type 4 secretion system-mediated gene transfer. *Science* 342(6155):250–253.

Hocquet D, Llanes C, Thouverez M, and et Al (2012) Evidence for Induction of Integron-Based Antibiotic Resistance by the SOS Response in a Clinical Setting. *PLoS Pathogens* 8(6):e1002778.

Holmes AJ, Holley MP, *et al.* (2003) Recombination activity of a distinctive integron-gene cassette system associated with Pseudomonas stutzeri populations in soil. *Journal of Bacteriology* 185(3):918–928.

Hover BM, Kim SH, *et al.* (2018) Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nature Microbiology* 3(4):415–422.

Hughes VM and Datta N (1983) Conjugative plasmids in bacteria of the 'pre-antibiotic' era. *Nature* 302(5910):725–6.

Imai Y, Meyer KJ, *et al.* (2019) A new antibiotic selectively kills Gram-negative pathogens. *Nature* .

Institute B (2018) Picard toolkit.

Ioannidou E, Siempos II, and Falagas ME (2007) Administration of antimicrobials via the respiratory tract for the treatment of patients with nosocomial pneumonia: A meta-analysis. *Journal of Antimicrobial Chemotherapy* 60(6):1216–1226.

Jackson J, Chen C, and Buising K (2013) Aminoglycosides: How should we use them in the 21st century? *Current Opinion in Infectious Diseases* 26(6):516–525.

Jacquier H, Zaoui C, Sanson-Le Pors MJ, Mazel D, and Berçot B (2009) Translation regulation of integrons gene cassette expression by the attC sites. *Molecular Microbiology* 72(6):1475–1486.

Jena NR (2012) DNA damage by reactive species: Mechanisms, mutation and repair. *Journal of Biosciences* 37(3):503–517.

Johansson C, Kamali-Moghaddam M, and Sundström L (2004) Integron integrase binds to bulged hairpin DNA. *Nucleic acids research* 32(13):4033–43.

Jové T, Da Re S, Denis F, Mazel D, and Ploy MC (2010) Inverse correlation between promoter strength and excision activity in class 1 integrons. *PLoS genetics* 6(1):e1000793.

Kato T, Sato Y, Iyobe S, and Mitsuhashi S (1982) Plasmid-mediated gentamicin resistance of Pseudomonas aeruginosa and its lack of expression in Escherichia coli. *Antimicrobial Agents and Chemotherapy* 22(3):358–363.

Koenig JE, Bourne DG, *et al.* (2011) Coral-mucus-associated Vibrio integrons in the Great Barrier Reef: Genomic hotspots for environmental adaptation. *ISME Journal* 5(6):962–972.

Kohanski MA, Dwyer DJ, and Collins JJ (2010) How antibiotics kill bacteria: From targets to networks. *Nature Reviews Microbiology* 8(6):423–435.

Krause KM, Serio AW, Kane TR, and Connolly LE (2016) Aminoglycosides: An overview. *Cold Spring Harbor Perspectives in Medicine* 6(6).

Krzywinski M, Schein J, *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome research* 19(9):1639–45.

Lacotte Y, Ploy MC, and Raherison S (2017) Class 1 integrons are low-cost structures in Escherichia coli. *The ISME Journal* 11(7):1535–1544.

Larouche A and Roy PH (2011) Effect of attC structure on cassette excision by integron integrases. *Mobile DNA* 2(1):3.

Lau CHF, Fraud S, Jones M, Peterson SN, and Poole K (2013) Mutational activation of the AmgRS two-component system in aminoglycoside-resistant Pseudomonas aeruginosa. *Antimicrobial agents and chemotherapy* 57(5):2243–51.

Lau CHF, Krahn T, Gilmour C, Mullen E, and Poole K (2015) AmgRS-mediated envelope stress-inducible expression of the mexXY multidrug efflux operon of Pseudomonas aeruginosa. *MicrobiologyOpen* 4(1):121–35.

Lázár V, Pal Singh G, *et al.* (2013) Bacterial evolution of antibiotic hypersensitivity. *Molecular Systems Biology* 9(1):700.

Lenski RE and Travisano M (1994) Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences of the United States of America* 91(15):6808–6814.

Lenz AP, Williamson KS, Pitts B, Stewart PS, and Franklin MJ (2008) Localized gene expression in Pseudomonas aeruginosa biofilms. *Applied and Environmental Microbiology* 74(14):4463–4471.

Levesque RC (2006) In Vivo Functional Genomics of Pseudomonas: PCR-Based Signature-Tagged Mutagenesis. *Pseudomonas*, 99–120. Springer US.

Liebert CA, Hall RM, and Summers AO (1999) Transposon Tn21, flagship of the floating genome. *Microbiology and molecular biology reviews : MMBR* 63(3):507–22.

Lim JY, Yoon JW, and Hovde CJ (2010) A brief overview of Escherichia coli O157:H7 and its plasmid O157.

Lin J, Nishino K, *et al.* (2015) Mechanisms of antibiotic resistance. *Frontiers in Microbiology* 6(FEB):1–24.

Ling LL, Schneider T, *et al.* (2015) A new antibiotic kills pathogens without detectable resistance. *Nature* 517(7535):455–459.

Lister PD, Wolter DJ, and Hanson ND (2009) Antibacterial-resistant Pseudomonas aeruginosa: Clinical impact and complex regulation of chromosomally encoded resistance mechanisms.

Liu PC, Lee YT, Wang CY, and Yang YT (2016) Design and Use of a Low Cost, Automated Morbidostat for Adaptive Evolution of Bacteria Under Antibiotic Drug Selection. *Journal of Visualized Experiments* (115):e54426–e54426.

Llosa M, Bolland S, and De la Cruz F (1994) Conjugation-independent, site-specific recombination at the oriT of the IncW plasmid R388 mediated by TrwC. *Journal of Bacteriology* 176(11):3210–3217.

Llosa M, Grandoso G, and de la Cruz F (1995) Nicking Activity of TrwC Directed Against the Origin of Transfer of the IncW Plasmid R388. *Journal of Molecular Biology* 246(1):54–62.

Lodge J, Williams R, Bell A, Chan B, and Busby S (1990a) Comparison of promoter activities in Escherichia coli and Pseudomonas aeruginosa: use of a new broad-host-range promoter-probe plasmid. *FEMS Microbiology Letters* 67(1-2):221–226.

Lodge JM, Minchin SD, Piddock LJ, and Busby JW (1990b) Cloning, sequencing and analysis of the structural gene and regulatory region of the Pseudomonas aeruginosa chromosomal ampC β-lactamase. *Biochemical Journal* 272(3):627–631.

Loot C, Ducos-Galand M, Escudero JA, Bouvier M, and Mazel D (2012) Replicative resolution of integron cassette insertion. *Nucleic acids research* 40(17):8361–70.

Loot C, Nivina A, *et al.* (2017) Differences in Integron Cassette Excision Dynamics Shape a Trade-Off between Evolvability and Genetic Capacitance. *mBio* 8(2):e02296–16.

Lu CD, Yang Z, and Li W (2004) Transcriptome analysis of the ArgR regulon in Pseudomonas aeruginosa. *Journal of Bacteriology* 186(12):3855–3861.

Lucas M, González-Pérez B, *et al.* (2010) Relaxase DNA binding and cleavage are two distinguishable steps in conjugative DNA processing that involve different sequence elements of the nic site. *Journal of Biological Chemistry* 285(12):8918–8926.

Lucassen R, Rehberg L, Heyden M, and Bockmuhl D (2019) Strong correlation of total phenotypic resistance of samples from household environments and the prevalence of class 1 integrons suggests for the use of the relative prevalence of intI1 as a screening tool for multi-resistance. *PLoS ONE* 14(6):1–13.

Ma L, Li AD, Yin XL, and Zhang T (2017) The Prevalence of Integrons as the Carrier of Antibiotic Resistance Genes in Natural and Man-Made Environments. *Environmental Science & Technology* 51(10):5721–5728.

MacDonald D, Demarre G, Bouvier M, Mazel D, and Gopaul DN (2006) Structural basis for broad DNA-specificity in integron recombination. *Nature* 440(7088):1157–1162.

Magiorakos AP, Suetens C, *et al.* (2013) The rise of carbapenem resistance in Europe: Just the tip of the iceberg.

Mathee K, Narasimhan G, *et al.* (2008) Dynamics of Pseudomonas aeruginosa genome evolution. *Proceedings of the National Academy of Sciences of the United States of America* 105(8):3100–3105.

Mazel D (2006) Integrons: Agents of bacterial evolution. *Nature Reviews Microbiology* 4(8):608–620.

McKenzie GJ, Harris RS, Lee PL, and Rosenberg SM (2000) The SOS response regulates adaptive mutation. *Proceedings of the National Academy of Sciences of the United States of America* 97(12):6646–6651.

Michael CA and Labbate M (2010) Gene cassette transcription in a large integron-associated array. *BMC genetics* 11:82.

Miller C, Thomsen LE, *et al.* (2004a) SOS Response Induction by ß-Lactams and Bacterial Defense Against Antibiotic Lethality. *Science* 305(5690).

Miller C, Thomsen LE, *et al.* (2004b) SOS Response Induction by -Lactams and Bacterial Defense Against Antibiotic Lethality. *Science* 305(5690):1629–1631.

Miller GH, Sabatelli FJ, *et al.* (1997) The Most Frequent Aminoglycoside Resistance Mechanisms–Changes with Time and Geographic Area: A Reflection of Aminoglycoside Usage Patterns? *Clinical Infectious Diseases* 24(Supplement 1):S46–S62.

Mo CY, Culyba MJ, *et al.* (2018) Inhibitors of LexA Autoproteolysis and the Bacterial SOS Response Discovered by an Academic-Industry Partnership. *ACS Infectious Diseases* 4(3):349–359.

Mo CY, Manning SA, *et al.* (2016) Systematically Altering Bacterial SOS Activity under Stress Reveals Therapeutic Strategies for Potentiating Antibiotics. *mSphere* 1(4).

Moradali MF, Ghods S, and Rehm BH (2017) Pseudomonas aeruginosa lifestyle: A paradigm for adaptation, survival, and persistence.

Moura A, Soares M, *et al.* (2009) INTEGRALL: A database and search engine for integrons, integrases and gene cassettes. *Bioinformatics* 25(8):1096–1098.

Moura de Sousa J, Balbontín R, *et al.* (2017) Multidrug-resistant bacteria compensate for the epistasis between resistances. *PLOS Biology* 15(4):e2001741.

Naas T, Mikami Y, Imai T, Poirel L, and Nordmann P (2001) Characterization of In53, a Class 1 Plasmid- and Composite Transposon-Located Integron of Escherichia coli Which Carries an Unusual Array of Gene Cassettes. *Journal of Bacteriology* 183(1):235–249.

Naas T, Poirel L, and Nordmann P (2008) Minor extended-spectrum β-lactamases. *Clinical Microbiology and Infection* 14(SUPPL. 1):42–52.

Nakaya R, Nakamura A, and Murata Y (1960) Resistance transfer agents in Shigella. *Biochemical and Biophysical Research Communications* 3(6):654–659.

Nemergut DR, Robeson MS, *et al.* (2008) Insights and inferences about integron evolution from genomic data. *BMC genomics* 9:261.

Nivina A, Escudero JA, Vit C, Mazel D, and 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.

Okii M, Iyobe S, and Mitsuhashi S (1983) Mapping of the gene specifying aminoglycoside 3'-phosphotransferase II on the Pseudomonas aeruginosa chromosome. *Journal of bacteriology* 155(2):643–9.

O'Neill J (2014) Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations The Review on Antimicrobial Resistance Chaired .

Pankey GA and Sabath LD (2004) Clinical Relevance of Bacteriostatic versus Bactericidal Mechanisms of Action in the Treatment of Gram-Positive Bacterial Infections. *Clinical Infectious Diseases* 38(6):864–870.

Papagiannitsis CC, Tzouvelekis LS, Tzelepi E, and Miriagou V (2017) attI1-Located Small Open Reading Frames ORF-17 and ORF-11 in a Class 1 Integron Affect Expression of a Gene Cassette Possessing a Canonical Shine-Dalgarno Sequence. *Antimicrobial agents and chemotherapy* 61(3).

Partridge SR and Hall RM (2003) The IS1111 Family Members IS4321 and IS5075 Have Subterminal Inverted Repeats and Target the Terminal Inverted Repeats of Tn21 Family Transposons. *Journal of Bacteriology* 185(21):6371–6384.

Partridge SR, Tsafnat G, Coiera E, and Iredell JR (2009) Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiology Reviews* 33(4).

Periwal V and Scaria V (2015) Insights into structural variations and genome rearrangements in prokaryotic genomes. *Bioinformatics* 31(1):1–9.

Pletzer D, Mansour SC, Wuerth K, Rahanjam N, and Hancock RE (2017) New mouse model for chronic infections by gram-negative bacteria enabling the study of anti-infective efficacy and host-microbe interactions. *mBio* 8(1).

Ploy MC, Courvalin P, and Lambert T (1998) Characterization of In40 of Enterobacter aerogenes BM2688, a class 1 integron with two new gene cassettes, cmlA2 and qacF. *Antimicrobial Agents and Chemotherapy* 42(10):2557–2563.

Poirel L, Madec JY, *et al.* (2018) Antimicrobial Resistance in Escherichia coli. *Microbiology Spectrum* 6(4).

Poirel L, Naas T, *et al.* (1999) Molecular and biochemical characterization of VEB-1, a novel class A extended-spectrum beta-lactamase encoded by an Escherichia coli integron gene. *Antimicrobial agents and chemotherapy* 43(3):573–81.

Poole K (2012) Stress responses as determinants of antimicrobial resistance in Gram-negative bacteria. *Trends in Microbiology* 20(5):227–234.

Providenti MA, O'Brien JM, Ewing RJ, Paterson ES, and Smith ML (2006) The copy-number of plasmids and other genetic elements can be determined by SYBR-Green-based quantitative real-time PCR. *Journal of Microbiological Methods* 65(3):476–487.

Qi Q, Preston GM, and Maclean RC (2014) Linking system-wide impacts of RNA polymerase mutations to the fitness cost of rifampin resistance in Pseudomonas aeruginosa. *mBio* 5(6):e01562.

Quinones M, Kimsey HH, and Waldor MK (2005) LexA cleavage is required for CTX prophage induction. *Molecular Cell* 17(2):291–300.

Ramirez MS and Tolmasky ME (2010) Aminoglycoside modifying enzymes. *Drug Resistance Updates* 13(6):151–171.

Rao AN, Barlow M, *et al.* (2006) Class 1 integrons in resistant Escherichia coli and Klebsiella spp., US hospitals. *Emerging Infectious Diseases* 12(6):1011–1014.

Recchia GD and Hall RM (1995a) Gene cassettes: A new class of mobile element. *Microbiology* 141(12):3015–3027.

Recchia GD and Hall RM (1995b) Plasmid evolution by acquisition of mobile gene cassettes: plasmid pIE723 contains the aadB gene cassette precisely inserted at a secondary site in the IncQ plasmid RSF1010. *Molecular Microbiology* 15(1):179–187.

Recchia GD, Stokes HW, and Hall RM (1994) Characterisation of specific and secondary recombination sites recognised by the integron DNA integrase. *Nucleic acids research* 22(11):2071–8.

Rice L (2008) Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE. *The Journal of Infectious Diseases* 197(8):1079–1081.

Rice W (2009) Design and evaluation of PCR primers which differentiate Escherichia coli O157:H7 and related serotypes. *Journal of Applied Microbiology* 106(1):149–160.

Rowe-Magnus DA, Guerout AM, Biskri L, Bouige P, and Mazel D (2003) Comparative analysis of superintegrons: Engineering extensive genetic diversity in the vibrionaceae. *Genome Research* 13(3):428–442.

Rowe-Magnus DA, Guerout AM, and Mazel D (2002) Bacterial resistance evolution by recruitment of super-integron gene cassettes. *Molecular Microbiology* 43(6):1657–1669.

Rowe-Magnus DA, Guerout AM, *et al.* (2001) The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proceedings of the National Academy of Sciences of the United States of America* 98(2):652–7.

Sakakibara Y and Chow CS (2017) Pseudouridine modifications influence binding of aminoglycosides to helix 69 of bacterial ribosomes. *Org. Biomol. Chem.* 15(40):8535–8543.

Salis HM (2011) The Ribosome Binding Site Calculator. *Methods in Enzymology* 498:19–42.

Salis HM, Mirsky EA, and Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology* 27(10):946–950.

Samuels AN, Roggiani M, Zhu J, Goulian M, and Kohli RM (2019) The SOS response mediates sustained colonization of the mammalian gut. *Infection and Immunity* 87(2).

San Millan A and MacLean RC (2017) Fitness Costs of Plasmids: a Limit to Plasmid Transmission. *Microbiology Spectrum* 5(5).

San Millan A, Santos-Lopez A, *et al.* (2015a) Small-plasmid-mediated antibiotic resistance is enhanced by increases in plasmid copy number and bacterial fitness. *Antimicrobial agents and chemotherapy* 59(6):3335–41.

San Millan A, Toll-Riera M, *et al.* (2015b) Sequencing of plasmids pAMBL1 and pAMBL2 from Pseudomonas aeruginosa reveals a blaVIM-1 amplification causing high-level carbapenem resistance. *Journal of Antimicrobial Chemotherapy* 70(11):3000–3003.

Santos-Lopez A, Marshall CW, Scribner MR, Snyder DJ, and Cooper VS (2019) Evolutionary pathways to antibiotic resistance are dependent upon environmental structure and bacterial lifestyle. *eLife* 8.

Sanz-García F, Hernando-Amado S, and Martínez JL (2018) Mutation-driven evolution of pseudomonas aeruginosa in the presence of either ceftazidime or ceftazidime-avibactam. *Antimicrobial Agents and Chemotherapy* 62(10).

Savli H, Karadenizli A, *et al.* (2003) Expression stability of six housekeeping genes: A proposal for resistance gene quantification studies of Pseudomonas aeruginosa by real-time quantitative RT-PCR. *Journal of Medical Microbiology* 52(5):403–408.

Schmidt F (1984) MGG The role of insertions, deletions, and substitutions in the evolution of R6 related plasmids encoding aminoglycoside transferase ANT-(2). *Mol Gen Genet* 194:248–259.

Schoonover LL, Occhipinti DJ, Rodvold KA, and Danziger LH (1995) Piperacillin/tazobactam: A new beta-lactam/beta-lactamase inhibitor combination. *Annals of Pharmacotherapy* 29(5):501–514.

Schwartz RS (2004) Paul Ehrlich's Magic Bullets. *New England Journal of Medicine* 350(11):1079–1080.

Sengupta S, Chattopadhyay MK, and Grossart HP (2013) The multifaceted roles of antibiotics and antibiotic resistance in nature. *Frontiers in microbiology* 4:47.

Solovyev V and Salamov A (2011) *Automatic Annotation of Microbial Genomes and Metagenomic Sequences*. Nova Science Publishers.

Song L and Zhang XH (2011) Accurate localization and excision of genomic islands in four strains of Pseudomonas aeruginosa and Pseudomonas fluorescens. *Chinese Science Bulletin* 56(10):987–995.

Spanjaard RA and van Duin J (1989) Translational Reinitiation in the presence and absence of a Shine and Dalgarno sequence 17(14):5501–5507.

Starikova I, Harms K, *et al.* (2012) A trade-off between the fitness cost of functional integrases and long-term stability of integrons. *PLoS pathogens* 8(11):e1003043.

Steenackers HP, Parijs I, Foster KR, and Vanderleyden J (2016) Experimental evolution in biofilm populations. *FEMS Microbiology Reviews* 40(3):373–397.

Stewart PS (2002) Mechanisms of antibiotic resistance in bacterial biofilms. *International Journal of Medical Microbiology* 292(2):107–113.

Stokes H and Hall RM (1992) The integron In1 in plasmid R46 includes two copies of the oxa2 gene cassette. *Plasmid* 28(3):225–234.

Stokes HW and Hall RM (1989) A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Molecular microbiology* 3(12):1669–83.

Stover CK, Pham XQ, *et al.* (2000) Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature* 406(6799):959–964.

Strugeon E, Tilloy V, Ploy MC, and Da Re S (2016) The Stringent Response Promotes Antibiotic Resistance Dissemination by Regulating Integron Integrase Expression in Biofilms. *mBio* 7(4):e00868–16.

Sun L, Alexander HK, *et al.* (2018) Effective polyploidy causes phenotypic delay and influences bacterial evolvability. *PLOS Biology* 16(2):e2004644.

Sundqvist M, Geli P, *et al.* (2010) Little evidence for reversibility of trimethoprim resistance after a drastic reduction in trimethoprim use. *Journal of Antimicrobial Chemotherapy* 65(2):350–360.

Sundström L, Rådström P, Swedberg G, and Sköld O (1988) Site-specific recombination promotes linkage between trimethoprim- and sulfonamide resistance genes. Sequence characterization of dhfrV and sulI and a recombination active locus of Tn 21. *MGG Molecular & General Genetics* 213(2-3):191–201.

survminer () Survminer package.

Tansirichaiya S, Mullany P, and Roberts AP (2019) Promoter activity of ORF-less gene cassettes isolated from the oral metagenome. *Scientific reports* 9(1):8388.

Taylor JS and Raes J (2004) Duplication and Divergence: The Evolution of New Genes and Old Ideas. *Annual Review of Genetics* 38(1):615–643.

Therneau TM (2015) *A Package for Survival Analysis in S*. Version 2.38.

Toll-Riera M, San Millan A, Wagner A, and MacLean RC (2016) The genomic basis of evolutionary innovation in pseudomonas aeruginosa. *PLOS Genetics* 12(5):1–21.

Torres-Barceló C, Kojadinovic M, *et al.* (2015) The SOS response increases bacterial fitness, but not evolvability, under a sublethal dose of antibiotic. *Proceedings. Biological sciences / The Royal Society* 282(1816):20150885.

Valot B, Guyeux C, *et al.* (2015) What it takes to be a Pseudomonas aeruginosa? The core genome of the opportunistic pathogen updated. *PLoS ONE* 10(5).

Van Acker H and Coenye T (2017) The Role of Reactive Oxygen Species in Antibiotic-Mediated Killing of Bacteria.

Van Boeckel TP, Pires J, *et al.* (2019) Global trends in antimicrobial resistance in animals in low- And middle-income countries. *Science* 365(6459).

Van Boxtel RA and Van De Klundert JA (1998) Expression of the Pseudomonas aeruginosa gentamicin resistance gene aacC3 in Escherichia coli. *Antimicrobial Agents and Chemotherapy* 42(12):3173–3178.

Vogne C, Aires JR, Bailly C, Hocquet D, and Plésiat P (2004) Role of the multidrug efflux system MexXY in the emergence of moderate resistance to aminoglycosides among Pseudomonas aeruginosa isolates from patients with cystic fibrosis. *Antimicrobial agents and chemotherapy* 48(5):1676–80.

Von Wintersdorff CJ, Penders J, *et al.* (2016) Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Frontiers in Microbiology* 7(FEB):173.

Wallden K, Rivera-Calzada A, and Waksman G (2010) Type IV secretion systems: Versatility and diversity in function.

Wang L, Pulk A, *et al.* (2012) Allosteric control of the ribosome by small-molecule antibiotics. *Nature Structural & Molecular Biology* 19(9):957–963.

Weinstein RA, Gaynes R, and Edwards JR (2005) Overview of Nosocomial Infections Caused by Gram-Negative Bacilli. *Clinical Infectious Diseases* 41(6):848–854.

Werren JH (2011) Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proceedings of the National Academy of Sciences of the United States of America* 108(SUPPL. 2):10863–10870.

Wick RR, Schultz MB, Zobel J, and Holt KE (2015) Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31(20):3350–3352.

Wilson KS and von Hippel PH (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proceedings of the National Academy of Sciences of the United States of America* 92(19):8793–7.

Wimberly H, Shee C, *et al.* (2013) R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. *Nature Communications* 4.

Wistrand-Yuen E, Knopp M, *et al.* (2018) Evolution of high-level resistance during low-level antibiotic exposure. *Nature Communications* 9(1):1599.

Xu Z, Li N, *et al.* (2009) Occurrence and characteristics of class 1 and 2 integrons in pseudomonas aeruginosa isolates from patients in southern china. *Journal of Clinical Microbiology* 47(1):230–234.

Yang Q, Li M, *et al.* (2017) Balancing mcr-1 expression and bacterial survival is a delicate equilibrium between essential cellular defence mechanisms. *Nature Communications* 8(1).

Yoshimura F and Nikaido H (1982) Permeability of Pseudomonas aeruginosa outer membrane to hydrophilic solutes. *Journal of Bacteriology* 152(2):636–642.

Ysern P, Clerch B, *et al.* (1990) Induction of SOS genes in Escherichia coli and mutagenesis in Salmonella typhimurium by fluoroquinolones. *Mutagenesis* 5(1):63–66.

Zhang AN, Li LG, *et al.* (2018) Conserved phylogenetic distribution and limited antibiotic resistance of class 1 integrons revealed by assessing the bacterial genome and plasmid collection. *Microbiome* 6(1):130.

Zhang L and Mah TF (2008) Involvement of a novel efflux system in biofilm-specific resistance to antibiotics. *Journal of Bacteriology* 190(13):4447–4452.