University of Technology, Sydney

DNA on the move: Investigation into two mobile genetic elements in *Vibrio* species

Rita Antoinette Rapa

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy: Science

2014

CERTIFICATE OF ORIGINAL AUTHORSHIP

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student:

Date:

Acknowledgements:

The undertaking of my PhD has been one of the most challenging but rewarding experiences so far. Although at times completing my doctorate has felt like a lonely and isolating journey, there are many people who have supported me along the way and I simply could not have gotten to the end without them.

Firstly, I would like to acknowledge the unwavering support and guidance of my two supervisors Dr Maurizio Labbate and Dr Hatch Stokes. I simply could not have achieved the academic fulfilment I have without you both. I thank you both for your mentorship, assistance with publications and for the many hours of proofreading this thesis. Maurizio, a massive thank you for your ongoing help in the lab, teaching me everything you know and for your friendship over the past 4 ½ years. I will dearly miss our laughs and inside jokes; I expect regular email updates! Hatch, thank you for your kindness and ongoing support with my PhD and career, even after your retirement. To my co-supervisor Professor Ian Charles, thank you for your academic guidance, financial support, philosophical conversations and the 'tea and buns.'

A huge thank you to past members of the Stokes Lab, particularly Elena Martinez, Natasha Vanegas and Piklu Roy Chowdhury, you have made my PhD a happy and exciting period of my life. To my fellow PhD students: Alex Gale, Raquel Alvarado, Ben Raymond, Sam Burns, Amelia Hynen, Andrew Liew, Erin Gloag, Shirin Ansari, Chris Allen, Jessica Tout and Lauren Messer. Thank you all for putting up with my tears, whinging and incessant snorts, accompanied by random outbursts of song and dance over the years. You have all touched my life in so many ways and I am going to miss seeing every one of you on a daily basis.

To Leigh Monahan, I cannot thank you enough for all your support, praise, reassurance and encouragement over the course of my PhD. You have helped me accomplish every task I have set out to do and I know this would not have been possible without you.

To my family: Mum, Papa, Loretta, David, AnneMarie and Nicko. You have all offered words of support throughout my academic life. Thank you for all you do and most importantly for trying to appear enthused when I email you my publications to read. Lastly, I would like to acknowledge the financial assistance of the Faculty of Science, University of Technology, Sydney in providing me with a scholarship for 3 ½ years.

Table of contents:

List of Figures:

List of Tables:

List of publications:

Rita A. Rapa, Christopher Allen, Maurizio Labbate. 2014. Expression of gene cassettes affect polysaccharide phenotypes. (Manuscript in preparation).

Rita A. Rapa, Atiqul Islam, Leigh G. Monahan, Ankur Mutreja, Nicholas Thomson, Ian G. Charles, H. W. Stokes, Maurizio Labbate. 2014. A genomic island integrated into *recA* of *Vibrio cholerae* contains a divergent *recA* and provides multi-pathway protection from DNA damage. *Environmental Microbiology*. (Appendix 3)

Rita A. Rapa and Maurizio Labbate. 2013. The function of integron-associated gene cassettes in *Vibrio* species: the tip of the iceberg. *Frontiers in Microbiology*. 4:385 (Appendix 5)

Rita A. Rapa, Ronald Shimmon, Steven P. Djordjevic, H.W. Stokes, Maurizio Labbate. 2013. Deletion of Integron-associated gene cassettes impact on the surface properties of *Vibrio rotiferianus* DAT722. *PLoS ONE*. 8:e58430 (Appendix 2)

Abbreviations:

LGT	lateral gene transfer
Indels	insertions and deletions
MGE(s)	mobile genetic element(s)
spp.	species (plural)
sp.	species (singular)
µ	micro; 10^{-6}
m	metre
cm	centimetre
%	percent
mL	millilitre
L	litre
DNA	deoxyribonucleic acid
Kb	kilobases
bp	base pairs
kbp	kilo base pair
IS	insertion sequence
IR	inverted repeat
VPI	*Vibrio* Pathogenicity Island
VSP	*Vibrio* Seveth Pandemic
DR	direct repeat
ICE	integrative conjugative element
ORF	open reading frame
OD	optical density

hr	hours(s)
s	second(s)
ms	millisecond(s)
wt	wild-type
2D-PAGE	two-dimensional polyacrylamide gel electrophoresis
1D-PAGE	one-dimensional polyacrylamide gel electrophoresis
LC-MS/MS	liquid chromatography tandem mass spectrometry
kDa	kilo-dalton
LPS	lipopolysaccharide
CPS	capsular polysaccharide
EPS	extracellular polysaccharide
NMR	nuclear magnetic resonance
OMP	outer membrane protein/porin
VSD	*Vibrio* species DAT722
IPTG	isopropyl β-D-1-thiogalactopyranoside
V	volts
PCR	polymerase chain reaction
M	molar
UV	ultra violet
GI	genomic island

Abstract:

Vibrios are a group of Gram negative rod-shaped bacteria that are ubiquitous in marine and estuarine environments. They exist as both free-living organisms and in association with a variety of hosts such as humans, coral, marine animals and plants. *Vibrio cholerae* is the most notorious of vibrios, being the causative agent of the devastating intestinal disease cholera in humans.

Lateral gene transfer (LGT), a process that allows DNA transfer between bacterial cells, has largely driven the rapid evolution in *V. cholerae* and other *Vibrio* species. In some strains of *Vibrio* species at least 20% of genomic content has arisen from LGT events. With respect to *V. cholerae*, the two most important virulence factors: cholera toxin encoded by the *ctxAB* genes and intestinal adhesion encoded on the vibrio pathogenicity island (VPI-1) have been acquired *via* mobile genetic elements transferred by LGT. Thus, these two virulence factors convert toxigenic *V. cholerae* into a paradigm for the importance of LGT, demonstrating how seemingly avirulent strains of *V. cholerae* become capable of causing epidemic/pandemic outbreaks (Uma *et al.*, 2003).

Mobile genetic elements include but are not exclusive to: transposons, integrons, conjugative elements and genomic islands. Research performed in this thesis is focussed on the study of the integron and a genomic island and how phenotypes they confer contribute to the adaptation of two *Vibrio* species: *V. rotiferianus* and *V. cholerae.*

Briefly, integrons are a two-component genetic recombination system present in the chromosome of almost all *Vibrio* species. The integron incorporates mobile genes termed gene cassettes into a reserved genetic site *via* site-specific recombination, named the integron/gene cassette system. The integron consists of three basic elements: an integrase gene (*intI*), an attachment site (*attI*) and a promoter (P_c). Gene cassettes generally contain a single open reading frame (ORF) and an IntI-identifiable recombination site called *attC*. Insertion (and excision) of gene cassettes is driven by an integrase-mediated recombination between *attI* and *attC*. Multiple insertion events lead to the accumulation of cassettes to form a cassette array. In vibrios, cassette arrays are uniquely large, sometimes containing hundreds of cassettes that make up a 1-3% of the entire genome. There is a consensus that

these gene cassettes add to the adaptive potential of vibrios and have likely been an important driver in the evolution of vibrios into their respective niches. How this is achieved has been difficult to understand given that 80% of gene cassettes are novel and consequently of unknown physiological function. Using a number of chemical, proteomic and molecular techniques this thesis has shown that gene cassette(s) present in the chromosome of a model *Vibrio* organism (*V. rotiferianus* DAT722) are altering bacterial surface properties. Changes to bacterial surface properties can be important in bacterial-host interactions importantly; biofilm formation, protozoan grazing and pathogen-host association. This thesis also examines how another mobile genetic element; a novel genomic island, aids in the repair of damaged DNA in *V. cholerae*, giving the organism an advantage in both the environment and in the disease causing state within humans. Our knowledge of how LGT has and continues to drive bacterial adaptation and evolution has only uncovered the tip of the iceberg.

Chapter 1: Introduction

1.1: Preface

Our environment is continually changing and every species is forced to adapt in order to survive. Without adaptation reproducing populations would face the prospect of extinction. Humans like most multicellular organisms have long generation times, limiting their ability to rapidly alter and manipulate their genotype and as a consequence more importantly their physical traits (phenotype). In contrast, prokaryotes, despite being unicellular, lacking in intricate organelles and being $1/10^{th}$ the size of some eukaryotic counterparts, possess the striking ability to promptly adapt within small evolutionary time-frames (Ochman *et al.*, 2000). Eukaryotic adaptation generally occurs over long evolutionary time scales by the manipulation, modification and mutation of existing DNA sequences. On the other hand prokaryotes, namely bacteria, have the ability to quickly adapt to their environment by multiple mechanisms such as mutations and gene rearrangements. One of the most significant mechanisms that has shaped microbial evolution however, is the **transfer** or 'sharing' of DNA sequences between both closely related species of bacteria and those from different species altogether.

This thesis will focus on the importance of DNA sequences transferring between bacteria; a mechanism termed lateral gene transfer (LGT). Namely, it will demonstrate the role LGT has had in the adaptation and evolution of aquatic bacteria belonging to the *Vibrio* genus, focussing on two important mobile genetic elements (MGEs) associated with two different *Vibrio* species. This thesis will explore how lateral gene transfer events can shape the physiology of the bacterial cell altering bacteria-bacteria interactions, survival, virulence potential and pathogenicity.

1.2: *Vibrio* species: An overview

Vibrios, belonging to the *Gammaproteobacteria* class, are gram negative, motile rods ranging between 1.4-2.6 μm in length (Thompson *et al.*, 2004b;Igbinosa and Okoh, 2008). Vibrios are autochthonous in aquatic and estuarine environments and are commonly found

to be free living. However, members of this genus also interact both symbiotically and pathogenically with a plethora of higher order organisms such as prawns, coral, fish, invertebrates, plant, marine mammals and humans (Grimes *et al.*, 2009).The diverse roles held by *Vibrio* spp. is a testament to their extraordinary ability to adapt and evolve rapidly. Lateral gene transfer (LGT), the ability of bacteria to share DNA sequences, is now known to be a major driving force in *Vibrio* evolution (discussed in detail in section 1.3; (Zaneveld *et al.*, 2008).

This introductory chapter is made up of three main components; the first will describe the diverse role of vibrios, the second defines the mechanisms of LGT and the crucial role LGT has played in the evolution and adaptation of bacteria (including vibrios) into their respective niches and thirdly it will discuss important mobile genetic elements transferred *via* LGT and their relevance to vibrios. Specifically, how two mobile genetic elements: the integron/gene cassette system and genomic islands have shaped bacterial evolution will be highlighted. Finally, the aims of this PhD thesis and the model organisms used in this study will be outlined.

1.2.1: The diverse roles of vibrios

The *Vibrio* genus consists of at least 84 described species and studies into these organisms have proven them to be highly adaptable with respect to their ability to occupy various and diverse niches (Hazen *et al.*, 2010). This section of chapter 1 will briefly describe the diverse roles of *Vibrio* spp. providing the reader with an overview of the niches vibrios occupy. There are numerous relationships bacteria uphold, including vibrios, with higher order organisms. These range from mutual relationships where both bacteria and host benefit from the association to pathogenic where an associated microorganism can cause harm and disease to its host. This section will mainly focus on two types of relationships that fall under the broad umbrella of symbiotic: where no harm is caused to the host as a consequence of the relationship, and pathogenic associations: where the relationship is detrimental to the host (Willey *et al.*, 2011).

1.2.1.1: Mutual symbiotic relationships

A symbiotic relationship is defined as the interaction between two dissimilar organisms. The general notion that encompasses symbiotic relationships is one where each organism lives in harmony together and consequently the association is not unfavourable to either organism. See Table 1.1 for a list of some symbiotic relationships vibrios hold in the aquatic environment.

The best documented symbiotic relationship involving vibrios is the dynamic one *V. fischeri* has with the Hawaiian squid *Euprymna scolopes*, commonly known as the bobtail squid. *E. scolopes* has developed a specialised light organ that houses and promotes the growth of a monospecific culture of *V. fischeri*. In this light organ the density of *V. fischeri* cells can reach beyond 10^{11} per cm^3, this high cell density triggers the bacterial cells to emit bioluminescence. The squid can control this light emission and thus exploit the bacteria for anti-predation purposes that mimic moonlight during the squid's nocturnal feeding practices (Visick and McFall-Ngai, 2000).

Another example involving a symbiosis with a *Vibrio* sp. is the sea sponge *Verongia* sp. which is known to display antifungal and cytotoxic properties within its aquatic habitat. These properties protect the sponge from infectious organisms. Ivanova *et al.* showed that 52% of cultured bacterial isolates from the *Verongia* sp. sponge constituted *Vibrio* spp., and that these vibrios produced secondary metabolites that displayed biologically active characteristics (Ivanova *et al.*, 2000). This suggests vibrios play a key role in the production of antifungal and cytotoxic metabolites that protect the sponge (Ivanova *et al.*, 2000).

Plankton is often associated with strains of *V. cholerae* (an important human pathogen described in section 1.2.1.2.1). Although this association with plankton is not detrimental to the plankton itself, Rivera *et al.* have shown that transportation of *V. cholerae*-plankton symbionts in ballast water can potentially source and spread this deadly human pathogen (Rivera *et al.*, 2013).

Table 1.1: Representative list of vibrios that hold symbiotic relationships with aquatic higher organisms

Host	***Vibrio* species**	**Reference**
Plankton	*V. cholerae*	(Rivera *et al.*, 2013)
Squid	*V. fischeri, V. logei*	(Visick and McFall-Ngai, 2000;Thompson *et al.*, 2004b)
Crustaceans	*V. cholerae, V. harveyi*	(Thompson *et al.*, 2004b)
Rotifers	*V. rotiferianus*	(Gomez-Gil *et al.*, 2003)
Abalone	*V. halioticoli*	(Hooper and Gordon, 2001)
Prawn larvae e.g. *Penacus chinensis*	*V. pacinii*	(Thompson *et al.*, 2004b)
Mollusc larvae e.g. *Nodipecten nodosus*	*V. pomeroyi*	(Thompson *et al.*, 2004b)

1.2.1.2: Pathogenic relationships

In contrast to symbiotic relationships many well-studied bacteria are noted for their pathogenic ability and vibrios are no exception. Numerous members of the *Vibrio* genera are capable of causing disease in humans and marine animals. Table 1.2 provides a representative list of known *Vibrio* pathogens, demonstrating the extent of different hosts vibrios can infect.

Table 1.2: Representative list of vibrios that infect human and marine animals

Host	*Vibrio* species Pathogen	Reference
Humans	*V. cholerae* (serogroup O1 and O139), *V. parahaemolyticus*, *V. vulnificus V. alginolyticus*, *V. damsela, V. fluvialis*, *V. furnissii*, *V. hollisae*, *V. metschnikovii*, and *V. mimicus*	(Uma *et al.*, 2003)
Corals e.g. *Pocillopora damicornis*	*V. shiloii*, *V. mediterranei* and *V. coralliilyticus*	(Thompson *et al.*, 2004b)
Fish e.g. *Labeo rohita, Catla catla, Cirrhinus mrigalam*	*V. anguillarum, V. salmonicida V. vulnificus, V. fluvialis*, *V. damsel, V. logei* and *V. splendidus*	(Morris and J, 2003;Mishra *et al.*, 2009)
Prawns e.g. *Penaeus monodon*	*V. harveyi, V. penaeicida*	(Thompson *et al.*, 2004b)
Oysters	*V. splendidus*	(Grimes *et al.*, 2009)

1.2.1.2.1: Human pathogens

V. cholerae, *V. vulnificus* and *V. parahaemolyticus* cause the majority of *Vibrio*-related human illnesses. *V. vulnificus* is an important causative agent of wound infections and septicaemia (Thompson *et al.*, 2004b). Septicaemia generally occurs in immuno-suppressed individuals (dampened immune responses). Infection with *V. vulnificus* can be caused by exposing lacerations to contaminated sea water or ingestion of tainted seafood. Infections initiated by *V. vulnificus* can advance to tissue necrosis and thus a secondary sepsis can potentially develop. Certain health conditions place patients at higher risk of acquiring pathogenic *V. vulnificus* including: haemochromatosis, liver disease, HIV and cancer (Thompson *et al.*, 2004b). However, *V. vulnificus* is a less common *Vibrio* disease with the CDC (Centers for Disease Control and Prevention) reporting approximately 900 *V. vulnificus* cases from US Gulf states between 1988 and 2006 (CDC, 2013).

V. parahaemolyticus causes an acute gastroenteritis that is characterised by a self-limiting diarrhoeal disease which is generally not severe. Approximately 4500 cases of *V. parahaemolyticus* are reported annually in the United States, although it is believed this

figure is under represented (CDC, 2013). It is thought that pathology is caused by the production of haemolysins including a thermostable direct haemolysin (TDH) and/or TDH-related haemolysin (TRH) (Thompson *et al.*, 2004b). Furthermore, it has been suggested that the haemolysin determinants have been acquired by the bacterium *via* LGT (discussed in section 1.3) (Nishibuchi and Kaper, 1995;Park *et al.*, 2000). *V. parahaemolyticus* infections are transmitted *via* the faecal-oral route and are commonly acquired by ingestion of undercooked/contaminated seafood, usually oysters. Furthermore, swimming in affected coastal areas can lead to eye and ear infections by this organism (Thompson *et al.*, 2004b).

By far the most notorious *Vibrio* spp. to afflict humans is *V. cholerae,* which is responsible for the disease cholera. Pathology is caused by enterotoxin producing strains of *V. cholerae* (serogroup O1 and O139; groups of bacteria containing a common antigen), triggering profuse and highly liquid diarrhoea. In addition some strains of *V. cholerae* referred to as non-O1/O139 are known to cause cholera-like symptoms (Kaper *et al.*, 1995). Diarrhoea associated with cholera, commonly referred to as 'rice water stool,' is infused with bacterial cells which can then contaminate food and water; *via* the faecal-oral route of transmission (Thompson *et al.*, 2004b). Since the first documented appearance in 1817, cholera has remained a significant global public health issue. Today, explosive outbreaks occur mainly in developing endemic nations, particularly African and South East Asian countries (WHO, 2006). In 2014, the World Health Organisation estimated 3-5 million cases of cholera and 100 000-120 000 deaths annually (WHO, 2014).

The history of cholera is encompassed by seven distinct pandemics where new strains of *V. cholerae* replaced pre-existing strains driving the rapid emergence of new pandemics. Not until a few years prior to the sixth pandemic was *V. cholerae* able to be cultured, thus confirming it as the causative agent (Figure 1.1 shows a timeline of pandemic cholera over the past since 1817). The sixth pandemic was caused by a biotype (group of organisms with the same genotype) of *V. cholerae* O1 called the Classical biotype. The Classical biotype was probably responsible for the preceding five (Scrascia *et al.*, 2009). The seventh pandemic stemmed from Indonesia in 1961 with the causative agent *V. cholerae* O1 of the El Tor biotype. From 1992 developing in India and later causing a Bangladesh epidemic of cholera; appeared a new serogroup, O139. The emergence of the O139 biotype sparked

questions as to whether this was the start of an eighth pandemic (Safa *et al.*, 2010). O1 El Tor, however, has re-established its prominence and is responsible for the vast majority of cholera infections today. Nevertheless in recent years new variants of *V. cholerae* have appeared. Interestingly, these variants display a combination of both phenotypic and genotypic traits from the two Classical and El Tor biotypes leading to them being collectively referred to as 'hybrid' or 'atypical El Tor strains' (Safa *et al.*, 2010). The ability of *V. cholerae* to change from the Classical to El Tor biotype and from the O1 and O139 serogroup in addition to the emergence of atypical El Tor strains in the early 1990's, is a testament to the adaptability of vibrios, namely *V. cholerae,* which is largely attributed to LGT (more detail given in section 1.3).

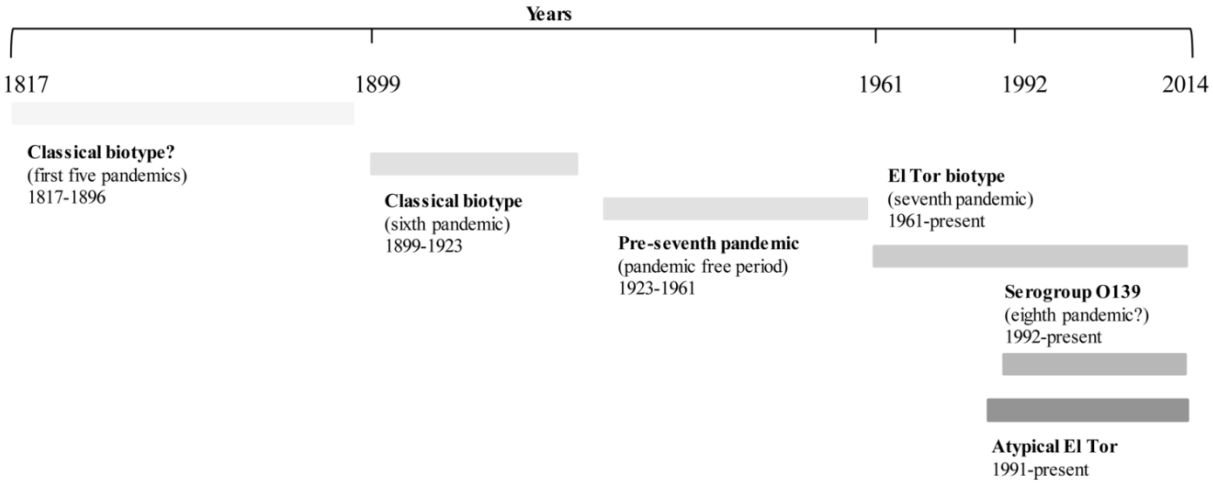


Figure 1.1: Timeline detailing the emergence of new cholera pandemics.
Outline of each pandemic documented over the past 200 years and the biotype/serogroup that was the causative agent of each. LGT has largely driven the emergence and progression of *V. cholerae* pathogens. Adapted from Safa *et al. Trends Microbiol,* 2009 (Safa *et al.*, 2010).

It is believed that non-O1/O139 strains of *V. cholerae* are gene reservoirs for O1/O139 toxigenic strains (Mukhopadhyay *et al.*, 2001;Faruque *et al.*, 2003;Faruque *et al.*, 2004;Pang *et al.*, 2007). In the aquatic environment strains of *V. cholerae* associate with a number or organisms including bacteriophage (discussed in detail in 1.4.3), protozoan grazers and a number of chitinous organisms where LGT events are enriched, potentially arming O1/O139 strains of *V. cholerae* with new genes. This may result in pathogenic *V. cholerae* transmission to humans that are now equipped with new virulence genes (see Figure 1.2 for a pictorial description of the lifestyles of *V. cholerae*).

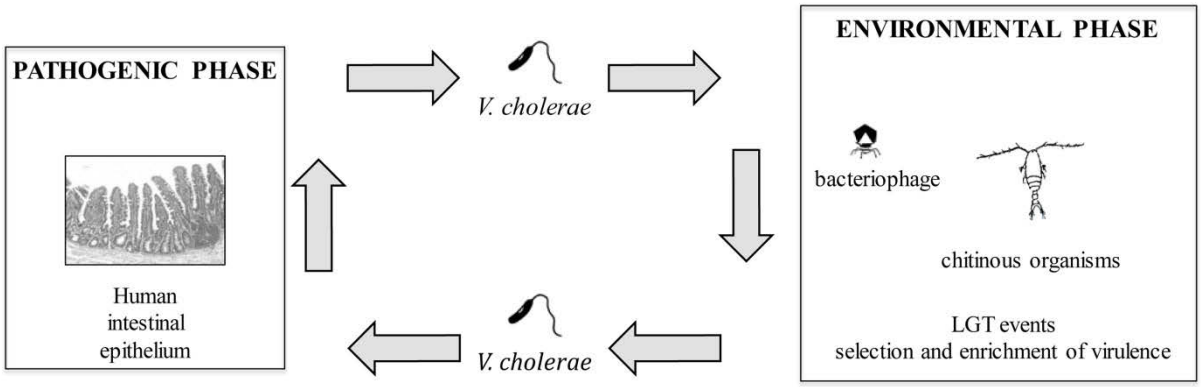


Figure 1.2: Pictorial description of the lifestyles of *V. cholerae.*
The two lifestyles of *V. cholerae* are described in each box; aquatic environmental and pathogenic in humans.

Other vibrios such as *V. alginolyticus, V. cincinnatiensis, V. fluvialis, V. furnisii, V. harveyi, V. metschnikovii, V. hollisae* and *V. mimicus* have been implicated in human disease (Thompson *et al.*, 2004b). Generally however, these organisms are considered less important in human disease than *V. cholerae, V. parahaemolyticus* and *V. vulnificus.* For example, *V. hollisae* has been isolated from a handful of patients presenting most commonly with gastroenteritis and rarely septicaemia upon consumption of contaminated seafood such as oysters (Carnahan *et al.*, 1994). It has been suggested that the reason for the infrequent isolation of this pathogen in association with human disease is due to its inability to grow on TCBS (thiosulfate-citrate-bile-salts-sucrose) agar; the medium used to clinically detect vibrios.

1.2.1.2.2: Marine animal pathogens

Vibrios are also pathogenic toward a number of marine animals. One of the most studied type of pathogenic association of vibrios with marine animals is with corals. With *Vibrio* spp. increasingly being implicated in coral bleaching, climate change is predicted to increase mass coral bleaching events (Luigi *et al.*, 2011). Coral reefs are unique and diverse ecosystems, made up of coral polyps (soft bodied invertebrates) in symbiosis with zooxanthellae; algae responsible for the vibrant colour of coral (Thompson *et al.*, 2004b). Coral bleaching is the loss of coral colour due to disturbances in the symbiotic relationship zooxanthellae have with coral. Coral bleaching is connected with an increase in seawater temperature and consequently is predicted to increase with climate change (Bourne *et al.*, 2009). Studies over the last 14 years have identified *Vibrio* spp. like *V. shiloii*, *V. mediterranei* and *V. coralliilyticus* to infect coral and cause coral bleaching (Kaper *et al.*, 1995). In *V. shiloii* and *V. coralliilyticus*, virulence factors for coral bleaching have been shown to be induced by increased temperature, linking increases in seawater temperature with *Vibrio* induced coral bleaching (Bourne *et al.*, 2009). The degradation of coral reef ecosystems by *Vibrio* infection has implications for decreased tourist influx to coral reef sanctuaries and biodiversity. For example, one of the seven natural wonders: The Great Barrier Reef situated off the Queensland coast of Australia is at threat due to coral associated diseases linked to the increase in sea water temperature. In August 2009, an article published in the Australian newspaper 'The Age,' stated that coral bleaching would result in the economic value of the reef tumbling by 73 per cent, from $51.4 billion to $13.7 billion. The loss in the Cairns region of Queensland would be more acute: 90% of an estimated economic value of $17.9 billion (Morton, 2009).

Vibrio anguillarum is the main causative agent of a deadly haemorrhagic septicaemic disease in fresh and salt water fish, molluscs and crustaceans (Frans *et al.*, 2011). This bacterium has a high morbidity and mortality rate and consequently is responsible for major economic losses worldwide in the aquaculture and larviculture industry. *V. anguillarum* causes disease in more than 50 fish species including salmon, sea bass, sea brim and cod. A number of environmental factors contribute to infection by this organism such as water quality, pollution, population density of fish and water temperature (Frans *et al.*, 2011).

Although the mechanism of pathogenesis has not been completely elucidated; genes involved in chemotaxis, motility, iron uptake and production of lipopolysaccharides with haemolytic activity have been implicated in disease (Frans *et al.*, 2011)

1.3: Lateral gene transfer: a major mechanism for bacterial evolution

The genomes of some bacterial species such as vibrios are in a constant state of flux and obtain a substantial amount of their genetic diversity by attaining DNA sequences from other cells including close or distantly related bacterial species, *via* a mechanism known as lateral gene transfer (LGT) (Ochman *et al.*, 2000). LGT is a two-step process that requires DNA to be 1) physically transferred between donor and recipient cells without the need for cell division and 2) the stable integration of DNA sequences into the bacterial genome so that they may be expressed and passed on during later cell division events. Cell division results in the vertical gene transfer of DNA to daughter cells following DNA replication and cell division. This results in all genetic information being replicated to produce a single identical clone (Lawrence, 1999). Undoubtedly, vertical transmission of genes is essential for the clonal inheritance of DNA sequences, however genetic diversity within closely related bacterial species is largely attributed to LGT.

There are three major mechanisms through which DNA is physically transferred into recipient cells: transformation, conjugation and transduction (Ochman *et al.*, 2000;Zaneveld *et al.*, 2008)(see Figure 1.3).

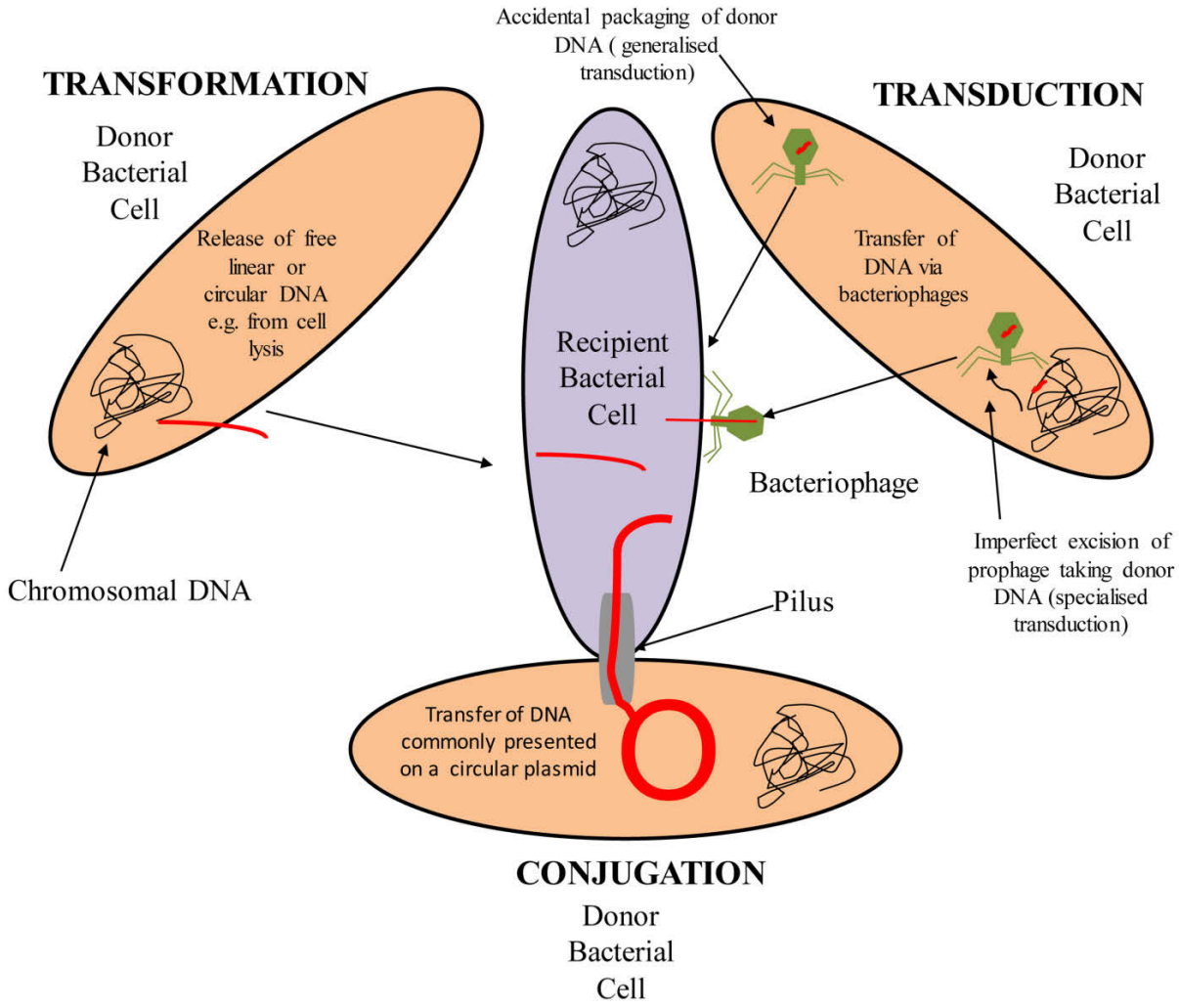


Figure 1.3: Mechanisms of LGT.
DNA transfer by transformation, conjugation and transduction are shown in this figure. Donor bacterial cells are shown in orange and the recipient bacterial cell is shown in purple. The DNA that is transferred *via* these three mechanisms is shown in red. The chromosomes of the recipient and donor bacterial cells are shown in black. The bacteriophage is represented in this figure in green. The pilus used for DNA transfer in conjugation is shown in grey.

1.3.1: Transformation

Free circular or linear DNA can be transferred into competent hosts *via* transformation (Frost *et al.*, 2005;Zaneveld *et al.*, 2008). Transformation occurs in naturally 'competent' bacteria. 'Competence' is a physiological state attained by bacteria through the regulation of expressed genes that encode functional proteins necessary for the uptake of DNA (Chen *et al.*, 2005). Transformation is responsible for a substantial amount of DNA transfer in aquatic environmental *Vibrio* spp. For example, in environmental strains of *V. cholerae* it was shown that DNA up to 44.9 kilobases (kb) could be transformed between bacterial cells (Miller *et al.*, 2007). Competence in multiple *Vibrio* spp. such as *V. cholerae, V. vulnificus* and *V. fischeri* is induced by chitin, a long chain polymer that makes up the exoskeletons of crustaceans. Many vibrios can colonise this exoskeleton and subsequently form biofilms (Meibom *et al.*, 2005;Blokesch and Schoolnik, 2007;Pollack-Berti *et al.*, 2010;Neiman *et al.*, 2011). This naturally induced competence known as natural transformation in the aquatic environment is an important driver in *V. cholerae* evolution. Perhaps the most important example that demonstrates how chitin-induced natural transformation contributes to the adaptability and evolution of strains of *V. cholerae* is the switch from the O1 to O139 serogroup. In a study the authors show chitin-induced natural transformation can mediate the serogroup conversion of a *V. cholerae* O1 pandemic strain *via* the acquisition of the O139-antigen encoding region by a **single** LGT event (Blokesch and Schoolnik, 2007). This highlights the potential LGT has to drive adaptation in bacteria facilitating the emergence of an evolved pathogen.

1.3.2: Conjugation

Conjugation is the transfer of DNA between two bacterial cells, commonly encoded on plasmids (discussed in section 1.4.1) *via* a specialised organ known as the conjugative pilus, (mating-pair formation apparatus (Mpf) shown in Figure 1.3). Other elements such as ICEs can be transferred *via* conjugation and are discussed further in section 1.4.5. Unlike transformation and transduction (section 1.3.3), direct cell to cell contact is required during conjugation (Thomas and Nielsen, 2005). Upon contact, the pilus retracts and brings cells close together and forms a pore between donor and recipient cells where mobilised DNA

can be transferred (Burrus *et al.*, 2002). Conjugative genes establish the 'mating pair channel' of proteins between the donor and recipient cells, allowing DNA to be transferred from one cell to another (Frost *et al.*, 2005). Conjugation is also one of the major mechanisms responsible for the rapid spread of antibiotic resistance genes in animal commensals.

1.3.3: Transduction

Transduction is the transfer of DNA sequences that is mediated by bacteriophage (phage; Figure 1.3) (Frost *et al.*, 2005). Bacteriophages are bacterial viruses that infect and use the cell to replicate. Bacteriophage can be lytic/virulent or lysogenic/temperate. Virulent/lytic bacteriophage, infect bacteria using the cell to replicate its genome and produce new phage that package the phage genome. Eventually, the bacteriophage overwhelms the host cell and causes it to lyse thereby releasing itself. Temperate/lysogenic phage integrate into the host genome (prophage), remaining dormant but replicating along with the bacterial genome. However, temperate phage can be activated and in turn become virulent (Zaneveld *et al.*, 2008). Figure 1.4 shows the lifecycles of temperate phage. Specialised and generalised transduction are phage mediated-LGT events that occur as a result of imperfect excision of lysogenic phage from the bacterial genome and accidental packaging of host DNA during phage assembly respectively. Specifically, specialised transduction occurs when the prophage integrates into the recipient cell at a particular attachment DNA site but incorrectly excises taking adjoining bacterial DNA with it. Generalised transduction is random packaging of donor DNA by the bacteriophage. In both instances, the donor DNA is transferred into the recipient cell by phage infection (Ochman *et al.*, 2000). With regard to aquatic bacteria such as vibrios, phage out number bacteria with ten viruses to one bacterial cell (Chibani-Chennoufi *et al.*, 2004). Consequently, this places a huge selection pressure on bacteria. As *Vibrio* spp. can be found as free living aquatic bacteria, it is not surprising that transduction is a common mechanism of LGT in vibrios (Jiang and Paul, 1998).

1.3.3.1: Lysogenic conversion

In some interactions between temperate/lysogenic phage and bacteria, lysogenic conversion can occur. Lysogenic conversion arises when the prophage itself that has integrated into the host bacterial genome alters the phenotype of the bacterial cell, sometimes enhancing fitness (Brussow *et al.*, 2004). For example, lysogenic conversion may enhance the bacterium's pathogenic capabilities. This is the case in *V. cholerae* where the CTX (cholera) toxin is carried by a temperate phage; consequently adding to the pathogenicity of *V. cholerae*. Without the CTX prophage, the cholera disease may not be as lethal or widespread. Lysogenic conversion of prophage have been crucial in shaping human pathogens other than *V. cholerae* by coding for other important toxins instigating human disease. For example, the *E. coli* Shiga and *C. botulinum* botulism toxins (Brussow *et al.*, 2004).

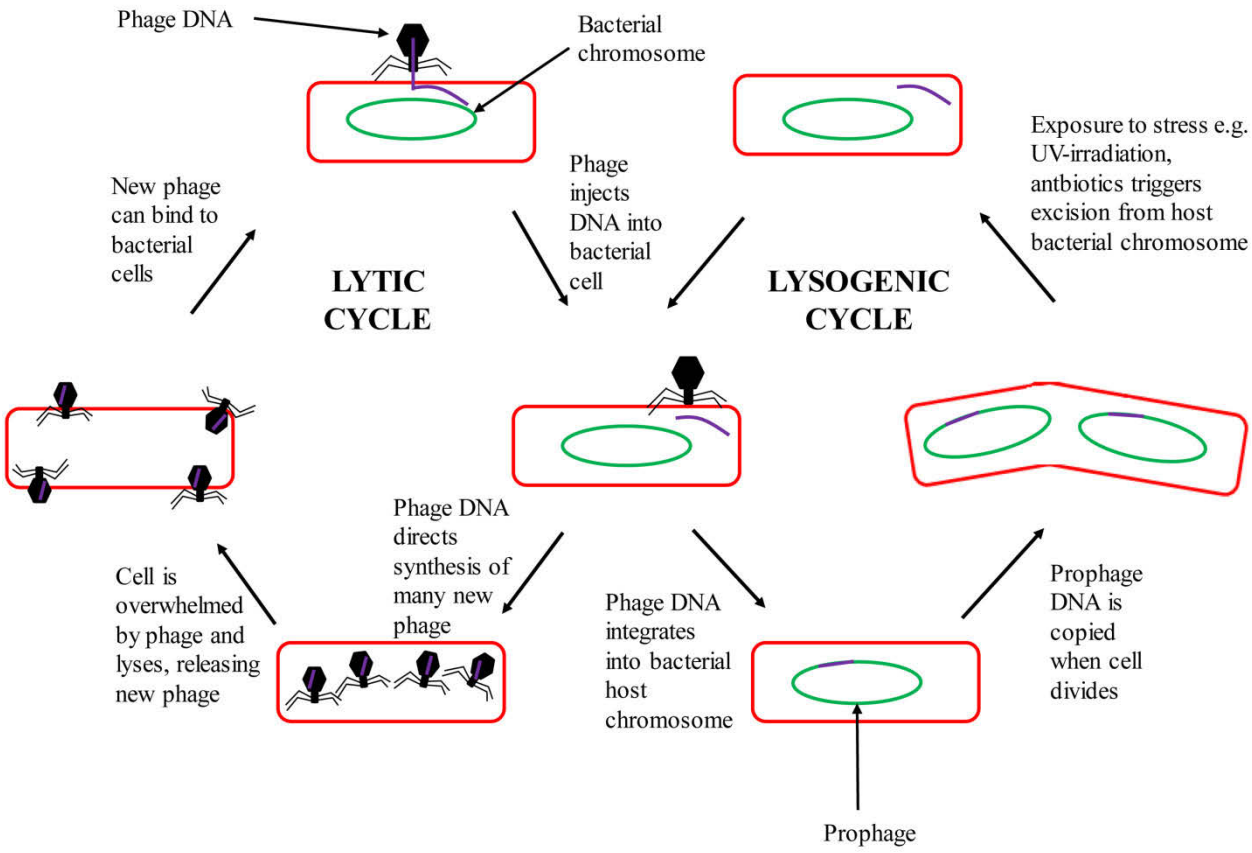


Figure 1.4: The lytic and lysogenic lifecycles of bacteriophage.
Virulent phage undergo only the lytic lifecycle. Temperate phage undergo two phases in their lifecycle: lytic and lysogenic. The lysogenic cycle allows the genome of the virus to replicate when the bacterial host genome is replicated *via* cell division. Certain environmental conditions such as UV-irradiation and exposure to chemical stressors such as antibiotics can cause a switch from the lysogenic to lytic cycle. In the lytic phase new phage are made and released when the bacterial host cell is overwhelmed by phage viruses.

1.3.4: Integration/maintenance of DNA sequences

The physical transfer of DNA between donor and recipient cells (either by transformation, conjugation or transduction) is the first step of LGT. Once DNA is physically transferred into the bacterial cell it must be integrated for it to be transcribed/translated, subsequently allowing phenotypes (observable physical or biochemical traits) to be displayed (Chen *et al.*, 2005;Thomas and Nielsen, 2005;Zaneveld *et al.*, 2008). If transferred/mobilised DNA is in the form of a plasmid, it can replicate autonomously. However, non-replicating DNA sequences must integrate into the bacterial genome or a plasmid to be maintained by the

cell *via* vertical transmission. Integration mechanisms include site-specific recombination, homologous recombination and transposition. Table 1.3 briefly summarises the modes of DNA transfer and integration (Thomas and Nielsen, 2005;Zaneveld *et al.*, 2008).

Site-specific recombination is the exchange of DNA strands between two DNA segments by enzymes that recognise *specific* sequences allowing recombination to occur. Site-specific recombinase enzymes cleave sequence specific DNA sequences, exchange the two DNA helices and re-join the two DNA strands (Ochman *et al.*, 2000). An important example of site-specific recombination in vibrios is the integron/gene cassette system discussed in section 1.6.

Homologous recombination is a process that results in the exchange of *homologous* DNA sequences (Frost *et al.*, 2005). Homologous DNA sequences are two sequences that are very highly similar or identical allowing bases to pair over an extended length of the two DNA sequences. This complex pathway involves the breaking and re-joining of paired DNA sequences carried out by two essential bacterial proteins, a single-stranded binding protein (SSB) and RecA (Orozco-Mosqueda *et al.*, 2009). Although homologous recombination requires sequences to be highly homologous, novel fragments of DNA can be recombined if flanked by homologous DNA sequence(s) such as insertion sequences.

Table 1.3: Mechanisms for DNA transfer and maintenance in the bacterial cell[1]

Mechanisms for Transfer of DNA into the Recipient Cell	Mechanisms for Maintenance of Laterally Transferred DNA
1. Conjugation	1. Autonomous replication (e.g.plasmids)
2. Transformation	2. Transposition (e.g. transposons, components of genomic islands)
3. Transduction	3. Site-specific recombination (e.g. integron/gene cassette system)
	4. Homologous recombination

[1]Table adapted from Stokes and Gillings, 2011, *FEMS Microbiol Rev* (Stokes, 2011).

1.4: Mobile genetic elements: a source of DNA

Mobile genetic elements (MGEs) are portions of DNA that are transferred *via* LGT. They generally encode enzymes and other proteins that mediate movement and integration of DNA within genomes or between bacterial cells. The collective term used to describe all MGEs in a genome is termed the mobilome (Frost *et al.*, 2005). Consequently, MGEs significantly contribute to the diversification of bacteria impacting on: genome plasticity and evolution, the dissemination of antibiotic resistance and virulence genes as well as enhance overall microbial fitness (Juhas *et al.*, 2009). This section will briefly describe common MGEs present in bacterial species and will then elaborate on MGEs specific to vibrios that have shaped their evolution into their respective niches.

1.4.1: Plasmids

Plasmids are usually circular DNA entities and, unlike most other MGEs, are usually self-replicating (Frost *et al.*, 2005). Plasmids are often vehicles for transmission of virulence and resistance genes. Plasmids usually exist as extrachromosomal replicons that are smaller than the bacterial chromosome, however some can also insert into the bacterial chromosome (Dobrindt *et al.*, 2004). Plasmids are most commonly covalently closed, circular double-stranded DNA molecules, however linear double-stranded plasmids have been found in some bacterial species (Hinnebusch and Tilly, 1993;Stewart *et al.*, 2005).

The general features that are common to plasmids are genes that encode for replication and transfer (*tra* genes). The *tra* genes establish a stable mating-pair formation apparatus and activate transfer from donor to recipient cell during conjugation (Frost *et al.*, 2005). In addition to replication, plasmids must also control their copy number and guarantee their inheritance during cell-division using a process known as partitioning and thus contain genes essential for this. Plasmids also contain variable accessory gene(s) that encode specific functions such as virulence that are often distinct from functions encoded on genes in the bacterial chromosome (Frost *et al.*, 2005). These are the genes that are important to bacterial evolution. Table 1.4 gives a representative list of some characterised plasmids present in vibrios.

Table 1.4: A short list of plasmids present in vibrios[1]

Host organism	Plasmid	Size (bp)	Phenotype	Number coding sequences
V. cholerae	pSIO1	4906	None known	3
V. alginolyticus	pVAE259	6075	None known	5
V. fischeri	pES213	5501	None known	9
V. nigripulchritudo	pSFn1	11 237	Pathogenicity	10
V. vulnificus	pC4602-1	56 628	Pathogenicity	69
V. vulnificus	pR99	68 446	Pathogenicity	71
V. anguillarum	pJM1	65 009	None known	57
V. parahaemolyticus	pSA19	4839	None known	4

[1]Table adapted from Hazen *et al.*, 2010 *FEMS Microbiol Rev*, (Hazen *et al.*, 2010).

1.4.2: Transposons

Transposons are genetic elements capable of moving from one DNA location to another and are commonly referred to as 'jumping genes' (Burrus *et al.*, 2002;Dobrindt *et al.*, 2004). In their basic form, transposons consist of two elements: a transposase gene, encoding the enzyme necessary for transposition and short inverted terminal DNA repeats (IR) flanking the transposase gene. Together, this arrangement is called an insertion

sequence (IS) (Salyers *et al.*, 1995). The transposase recognises the inverted repeats and excises the IS allowing it to relocate into a new DNA position. Transposons are diverse and there are numerous mechanisms for how they mobilise host DNA. Examples include IS elements that flank host DNA and mobilise as one unit (known as composite transposons, shown in Figure 1.5). For example Tn*5* is a composite transposon (cut and paste transposon) that has mobilised a kanamycin resistance gene. However, transposons have been shown to carry genes that are involved in metabolism such as Tn*951* which is involved in lactose fermentation (Michiels *et al.*, 1987). Other more complex transposons exist such as the replicative transposon (Brown and Evans, 1991). They have the basic structure of a composite transposon, however they carry an additional gene known as a resolvase, allowing replicative transposons to move by making a copy of themselves, unlike simple copy and paste composite transposons. One copy remains at the original insertion point and the other moves to a new site (Willey *et al.*, 2011).

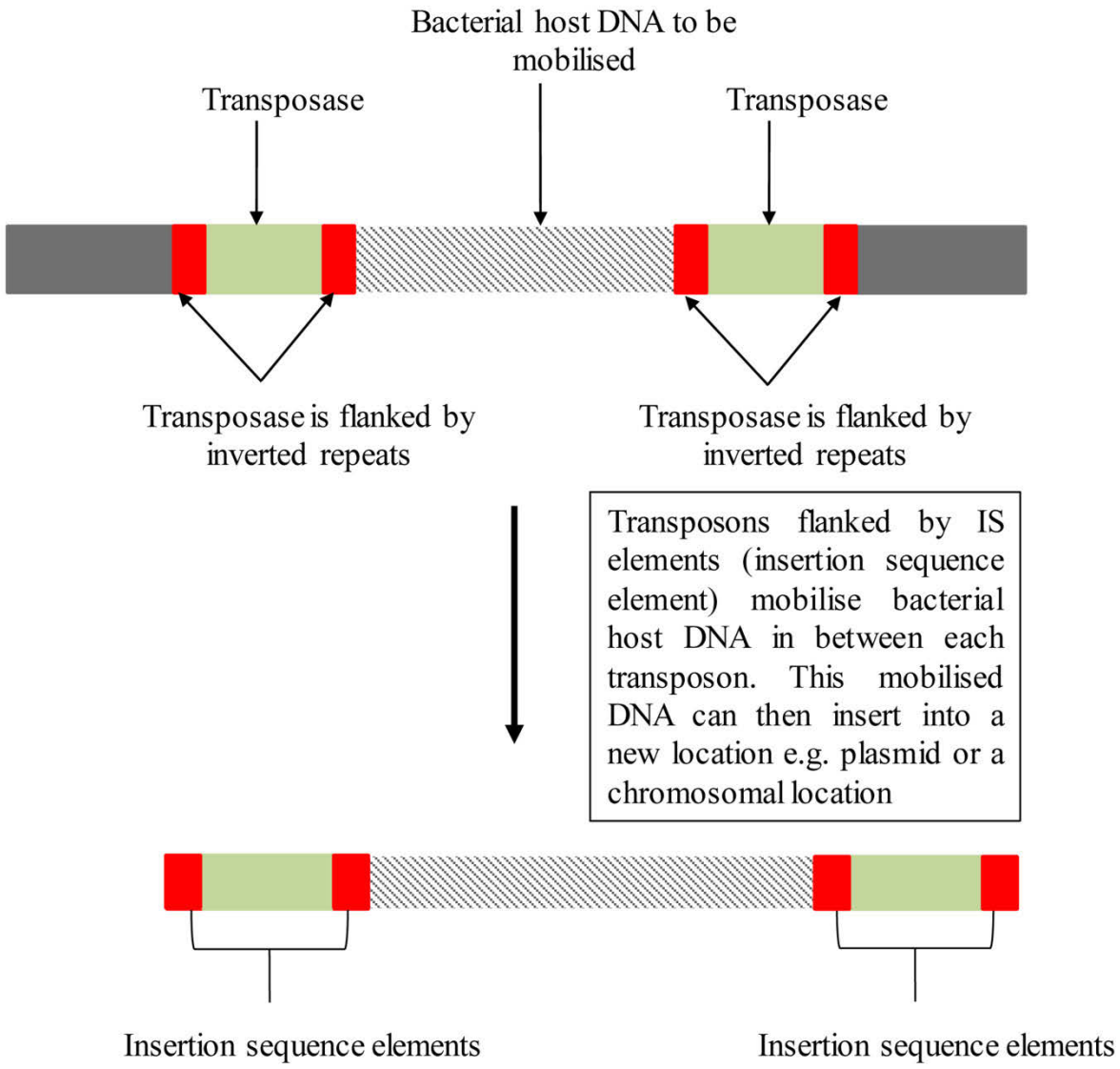


Figure 1.5: Composite transposon.
The inverted repeats (IRs) are shown here in red, the transposase gene is shown in light green and collectively they make the up the insertion sequences (IS). DNA surrounding the transposon is shown in dark grey and the DNA sequence that is mobilised is depicted by the diagonal stripes.

1.4.3: Bacteriophage

Bacteriophage or phage are viruses that infect bacteria and are also able to transfer host DNA sequences *via* transduction. The method of DNA transfer *via* the phage has been examined in the previous section 1.3.3. Bacteriophage are commonly exploited for their use

in molecular biology and genomics as they have the ability to transfer large portions (50-100kb) of a genome per transduction event. Interestingly, phage are the most abundant and most rapidly replicating organisms known with approximately 10^{25} infections per second (Frost *et al.*, 2005).

Following the characterisation of the important CTXϕ phage that encodes the cholera toxin genes *ctxAB* (Waldor and Mekalanos, 1996) significant research has been carried out on the diversity of phages specific to vibrios (vibriophages; Table 1.5 gives a list of different vibriophages). Interestingly, the CTXϕ phage has been shown to move from *V. cholerae* into strains of *V. mimicus,* leading to the emergence of toxigenic *V. mimicus* (Boyd *et al.*, 2000). This transfer of such important genes that are responsible for the fatal symptoms of the disease cholera into a completely different bacterial species, highlights the vital role LGT has had and continues to have in shaping the emergence of pathogens. Although the majority of vibriophages that have been described are relatively small (≤10 kb) a number of large phages have been characterised in *V. parahaemolyticus* (Chang *et al.*, 1998;Nasu *et al.*, 2000). One of these large phages is the 244.6 kb T4-like phage (a phage that infects *E. coli*) named KVP40. This phage encodes a number of proteins that may assist in host metabolism (Miller *et al.*, 2003).

As mentioned in section 1.3.3 bacteriophage outnumber bacteria in the aquatic environment placing enormous selection pressures on bacteria to develop mechanisms of evasion. It has been shown that bacteriophage specific to *V. cholerae* (cholera phages) are linked to cholera epidemics in Bangladesh and India. A distinctive epidemiological feature of cholera is its seasonal regularity in endemic areas, such as the Ganges Delta of Bangladesh and India (Faruque and Mekalanos, 2012). Studies over a 3 year period have shown phage that can infect and lyse *V. cholerae* play a significant role in kerbing epidemics (Faruque *et al.*, 2005a;Faruque *et al.*, 2005b;Jensen *et al.*, 2006;Agren *et al.*, 2008). These studies show that as the number of cholera presenting patients in Dhaka, Bangladesh increased as the number of lytic cholera phage decreased. Likewise, epidemics of cholera mostly ceased with large increases of cholera phage in the waters. This key correlation between the number of phage and the modulation of cholera epidemics in endemic regions, illustrates how selection pressures placed upon bacteria force adaptation, highlighting the importance

of LGT events within the environment. Furthermore, this example demonstrates the richness of DNA in the aquatic environment. The surge in abundance of DNA is an opportunity for surviving bacteria to harness the DNA from lysed cells, arming them with potential for new virulence mechanisms and increased fitness.

Table 1.5: Short representative list of vibriophages identified[1]

Host organism	**Phage**	**Size (bp)**	**Phenotype**	**Number of genes**
V. cholerae	CTXϕ	70 000	Pathogenicity	10
V. cholerae	K139	33 106	None known	44
V. parahaemolyticus	KVP4O	244 834	Metabolism	381
V. parahaemolyticus	Vf33	7965	None known	7
V. parahaemolyticus	VP93	43 931	None known	44
V. harveyi	VHML	43 198	None known	57
V. cholerae	fs1	6340	None known	15

[1]Table adapted from Hazen *et al.*, 2010 *FEMS Microbiol Rev*, (Hazen *et al.*, 2010)

1.4.4: Genomic islands

Genomic islands (GIs) are large (10-200 kb) discrete chromosomal segments of DNA differing between closely related bacterial strains to which mobility is attributed (Juhas *et al.*, 2009). GIs usually harbour a number of accessory genes that are beneficial to the bacterium in certain environments. Thus, it is understood that GIs contribute to the diversification and adaptation of microorganisms, having a significant impact on genome plasticity, adaptation and evolution (Juhas *et al.*, 2009). GIs contain DNA regions that have been previously transferred by other MGEs so generally have differing G + C content and are often inserted within host tRNA genes (Dobrindt *et al.*, 2004). As GIs are generally present in certain bacteria but are absent in other closely related strains of the same species, this makes them excellent markers for identifying recent LGT events and divergence. They are generally flanked by repeat units (short DNA sequences) similar to transposons and contain mobility genes such as integrases, recombinases and transposases that mediate their integration and excision from the host bacterial chromosome (Dobrindt *et al.*, 2004). Figure 1.6 shows the general features of genomic islands.

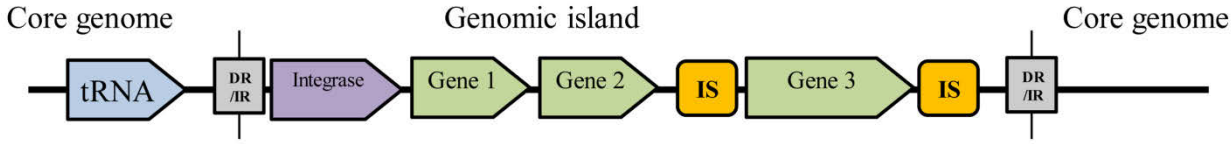


Figure 1.6: General features of genomic islands.
Genomic islands are generally large regions of DNA that encode for genes (shown in green) whose products add to the adaptive potential of their bacterial host. Genomic islands are often inserted adjacent to tRNA genes (blue) and are flanked by inverted (IR) or direct repeats (DR; grey). They generally contain genes that aid in mobilisation and integration such as an integrase gene (purple) and insertion sequences (IS; shown in yellow). Insertion sequences can translocate abutting DNA similar to transposons. For example, the two insertion sequences in this figure could transfer and integrate gene 3 into the core genome, or into other MGEs.

GIs frequently contain one or more genes that increase the fitness and adaptability of the host bacterium. Pathogenicity islands are a type of genomic island that generally contain genes that enhance virulence and consequently are commonly associated with pathogenic bacteria. GIs on the other hand have been reported to influence traits such as antibiotic resistance, symbiosis, fitness and general adaptation (Dobrindt *et al.*, 2004). Table 1.6 lists examples of the different types of bacteria that contain genomic islands.

Table 1.6: Examples of genomic islands and their encoded functions

Function	**Organism**	**Genomic location**	**Advantage**	**Reference**
Iron uptake	Faecal *E. coli*, *Salmonella enterica*, *Klebsiella* spp.	Chromosome	Increased adaptability	(Schubert *et al.*, 1998;Bach *et al.*, 2000;Oelschlaeger *et al.*, 2003)
Expression of adhesions	Faecal *E. coli*	Chromosome	Adhesion to gastrointestinal epithelium, colonisation	(Dobrindt *et al.*, 2003)
Sucrose uptake	*Salmonella senftenberg*	Chromosome	Increased metabolic versatility	(Hochhut *et al.*, 1997)
Resistance to methicillin	*Staphylococcus aureus*	Chromosome	Increased adaptability	(Ito *et al.*, 1999;Ito *et al.*, 2001)
Resistance to mercury and kanamycin	*Providencia rettgeri*	Chromosome	Increased adaptability	(Boltner *et al.*, 2002)
Degradation of phenolic compounds	*Pseudomonas putida*	Chromosome	Increased adaptability	(Ravatn *et al.*, 1998;van der Meer and Sentchilo, 2003)

1.4.4.1: Genomic islands and the evolution of pathogenic/pandemic V. cholerae

By far the most notable genomic islands that have driven virulence in pandemic-causing strains of *V. cholerae* are *Vibrio* Pathogenicity Islands I and II (VPI-I and VPI-II). Thus, as the name suggests these genomic islands harbour genes that encode important functions involved in human infection. Briefly, VPI-I is a 40 kb long island that carries genes that encode one of the two main virulence factors contributing to the disease state; TCP (previously described in section 1.3.6). TCP is a type IV pilus that acts as a colonisation factor. This type IV pilus aids in the adhesion of the bacterium to human intestinal epithelial cells (Schmidt and Hensel, 2004). TCP also acts as a receptor for the CTXϕ that carries the necessary cholera toxin genes. This toxin causes the diagnostic watery stool associated with cholera disease. As a result, the presence of VPI-I is a requirement for the lysogenic conversion of non-toxigenic *V. cholerae* strains by the CTXϕ. This lysogenic conversion aided by the presence of VPI-I in the chromosome of pandemic-causing strains, contributes to the appearance of new toxigenic strains (i.e. those that contain the cholera toxin) of *V. cholerae* in the environment (Schmidt and Hensel, 2004). It has also been shown that VPI-I can excise from the chromosome and form a closed circle intermediate, providing further support of the lateral transfer of this genomic island (Rajanna *et al.*, 2003).

VPI-II is also a pathogenicity island that was identified in toxigenic but not in non-toxigenic strains of O1 *V. cholerae*. VPI-II is a 57 kb region that encodes a neuraminidase (*nanH*) and amino sugar metabolism. The neuraminidase is required for *V. cholerae* growth on sialic acid as the sole carbon source. Sialic acid is present in large quantities in the human gut, and is thus thought to provide *V. cholerae* with an advantage when passing through the gut on its way to the intestines (Faruque and Mekalanos, 2012).

VSP I and II are genomic islands that are unique to seventh pandemic El Tor and O139 isolates (Dziejman *et al.*, 2002). VSP-I is a 16 kb region and VSP-II is a 7.5 kb region; which has homology to a genomic island in *V. vulnificus* (O'Shea *et al.*, 2004). The exact function of all the genes on VSP-I and II have not been elucidated, however the presence of these islands in pandemic-causing strains suggest they have a role in the evolutionary

fitness and spread of the seventh pandemic (Faruque and Mekalanos, 2012). Similar to VPI-II, both VSP-I and II have been shown to excise from the chromosome and form circular intermediates (Murphy and Boyd, 2008). The circularisation of these three pathogenicity islands is indicative of their transfer *via* LGT-mediated events. Further supporting the lateral transfer of VPI-II and VSP-II is the presence of an integrase that is likely used for insertion and excision of the circular intermediate.

1.4.5: Integrative conjugative elements: the tip of the ICEberg

Integrative conjugative elements (ICEs), like genomic islands are gene clusters that reside on the bacterial chromosome (Wozniak and Waldor, 2010). Similar to plasmids, ICEs are transferred *via* conjugation and like bacteriophage they can integrate into and use the host chromosome to replicate (Burrus and Waldor, 2004). They generally contain insertion sequences, transposons and recombinase genes and may promote the mobilisation of genomic islands (Burrus and Waldor, 2004). Originally the function of ICEs were thought to act as vectors for the transfer of antibiotic resistance genes, but are now known to encode diverse functions such as virulence factors, regulation of biofilm formation and metabolic processes (Hochhut *et al.*, 1997;Brassinga *et al.*, 2003;Wozniak and Waldor, 2010).

An example of an ICE that has shaped and driven adaption within vibrios is the SXT element. This ICE is ~100 kb in size and was first discovered in a 1992 *V. cholerae* O139 clinical isolate from India. This element and SXT-like ICEs harbour genes that confer multiple antibiotic resistance; including to sulfamethoxazole, trimethoprim, ciprofloxacin and streptomycin (Waldor *et al.*, 1996). Today, many strains of both the O1 and O139 serogroups have acquired SXT elements (Burrus *et al.*, 2006). Interestingly, a study showed that dissemination of SXT-encoded antibiotic resistance genes is regulated by the SOS-response (Beaber *et al.*, 2004). The SOS response is a rescue pathway induced by bacteria in response to DNA damage triggered by stress and drives the expression of genes necessary to ensure the survival of the cell (Aertsen and Michiels, 2007). The SOS-response is discussed further in the introduction of chapter 5. There are many environmental stimuli that trigger damage to DNA including UV-irradiation and antibiotics themselves. Hence, it is thought that the very treatment of cholera infections with

antibiotics stimulates the excision and subsequent dissemination of SXT elements, escalating the worldwide spread of these types of MGEs and multiple-antibiotic resistance (Hastings *et al.*, 2004;Kitaoka *et al.*, 2011).

1.5: LGT: a major mechanism for bacterial evolution

LGT is an extremely important process by which gene sequences, often encoding proteins of imperative function, can be transferred and spread between bacterial populations. The proportion of genes acquired by LGT that make up some bacterial genomes is significant, with estimates of up to 25% (Ochman *et al.*, 2000). In *Vibrio* spp. such as *V. cholerae* and *V. vulnificus* a large 20% of their pan genome (total collection of genes across all members of a species) can be composed of laterally acquired DNA (Deshpande *et al.*, 2011).

The most notable example of the impact LGT has had on bacterial adaptation and evolution is the expansion of antibiotic resistance. The efficacy of penicillin, which was discovered just in time to treat wounded soldiers suffering bacterial infections during World War II, was short lived. In an article for *The New York Times* in 1945 Sir Alexander Fleming warned that the widespread public use of penicillin would lead to 'mutant forms' of *Staphylococcus aureus* (Alanis, 2005). This proved true as only a few years later over 50% of *S. aureus* strains were no longer treatable with penicillin (Alanis, 2005). Since the discovery of the first antibiotic; penicillin, at least 17 different antibiotic classes (each with differing mechanisms of action against bacteria) have been developed. Regrettably, each class has at least one mode of action accumulated by bacteria that works against it. To make matters worse, multi-resistant bacteria have acquired multiple mechanisms of resistance against most antibiotic classes. Although bacteria can develop spontaneous chromosomal mutations and have other intrinsic mechanisms that allow them to evade antibiotics such as efflux pumps that actively expel antibiotics from the cell, the majority of antibiotic resistance is a consequence of LGT (Alanis, 2005). Thus, our understanding of LGT, its components and effects on microbial populations, is vital in the fight against a devastating post-antibiotic era that has seen an increase in infection related morbidity. For example, antibiotic resistance plasmids play a critical role in both multiple-antibiotic resistance and its subsequent dissemination. Resistance plasmids not only carry multiple

genes encoding evasion of antibiotics, but they are also vehicles for the transmission of other important MGEs such as genomic island, integrons (see section 1.6) and transposable elements (Bennett, 2008). This highlights how numerous MGEs can accumulate on a plasmid and consequently a single gene transfer event into one bacterial cell can initiate the spread of important phenotypes such as antibiotic resistance.

1.5.1: *V. cholerae*: a paradigm for the importance of LGT in adaptation and evolution

V. cholerae is by far the most devastating and comprehensively studied *Vibrio* spp. to cause pathology in humans (Uma *et al.*, 2003). As mentioned previously, in the aquatic environment numerous strains of non-pathogenic *V. cholerae* are present. Interestingly, when the genomes of these environmental *V. cholerae* strains are compared to pathogenic ones, the genes associated with virulence are often carried on mobile genetic elements (Figure 1.7) (Bik *et al.*, 1995). Therefore, LGT has been the major mechanism in the evolution of this pathogen making it a paradigm for studying how LGT events can transform a harmless strain into one capable of causing epidemics/pandemics (Uma *et al.*, 2003).

The ability of *V. cholerae* to cause pathology is largely due to the expression of two virulence factors; 1) the cholera enterotoxin (CT) and 2) an intestinal adhesin called toxin coregulated pilus (TCP) which is essential for attachment of the bacterium to the human intestinal wall. The *ctx*AB genes encoding CT are part of a genetic cluster known as the CTX element (Faruque *et al.*, 2003). The cholera toxin is encoded by a bacteriophage called the CTX phage (CTXϕ) and is thus a mobilisable element capable of being laterally transferred *via* phage mediated transduction events (Schmidt and Hensel, 2004). The gene cluster that encodes TCP is part of the *Vibrio* pathogenicity island (VPI). Consequently both pathogenicity islands and genomic islands are transferred by LGT (Schmidt and Hensel, 2004). TCP has likely been acquired by pathogenic strains of *V. cholerae* through a phage mediated LGT event. The exact type of phage involved in this transfer of genetic material is yet to be substantiated, but it is assumed to be a phage called VPI (*Vibrio* Pathogenicity Island) (Davis and Waldor, 2003). The TCP also acts as a receptor for the

CTXϕ, implying that the TCP gene cluster was acquired before CT (Waldor and Mekalanos, 1996).

The role lateral gene transfer has had on the emergence of new strains of cholera is best highlighted in the 1992 epidemic outbreak of cholera in Bangladesh, caused by a previously unreported serogroup, called O139, that had arisen from an O1 El Tor strain (Ramamurthy *et al.*, 1993). Strains of *V. cholerae* are diverse and encompass more than 200 O-antigen serogroups, with only O1 and O139 being pathogenic (Kotetishvili *et al.*, 2003). However, only serogroup O1 was responsible for epidemics and pandemics up until 1992, when a new serogroup O139 emerged causing an outbreak of cholera in Bangladesh (described previously in section 1.2.1.2) (Johnson *et al.*, 1994). Genetically, O139 is almost identical to O1 El Tor, indicating that O139 evolved from O1 El Tor. However, the region of difference is in the O-antigen genes that have a 22 kbp region deleted in O139 compared to O1 El Tor and this deletion is replaced by as 35 kbp fragment from an unknown source (Johnson *et al.*, 1994). This has most likely occurred from homologous recombination in the O-antigen region from DNA acquired *via* LGT (Johnson *et al.*, 1994;Bik *et al.*, 1995;Faruque and Mekalanos, 2003). The emergence a new epidemic *V. cholerae* strain (O139) is a classic example on the importance of LGT in the evolution of *Vibrio* species.

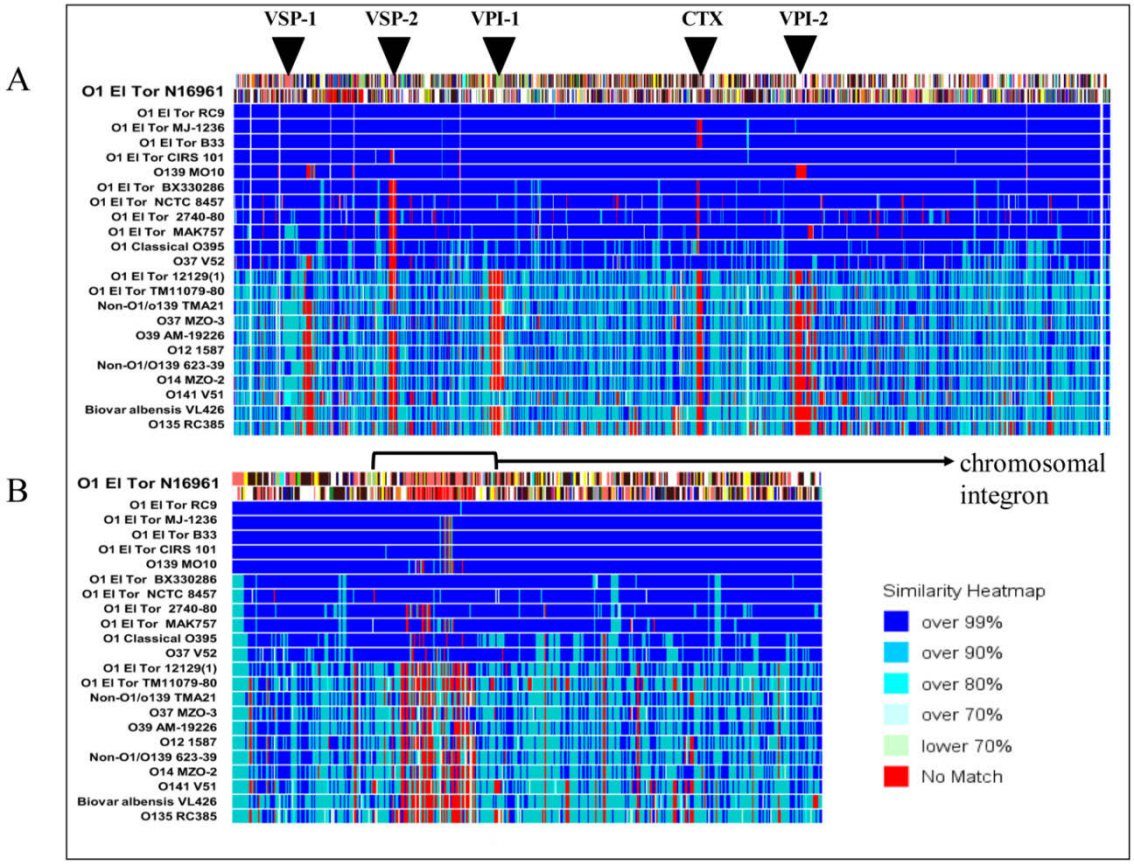


Figure 1.7: Importance of MGEs in the emergence of *V. cholerae* virulence.
This figure shows gene alignments of 23 *V. cholerae* strains highlighting gene conservation with *V. cholerae* O1 El Tor N16961 as the reference strain. (A) shows the large chromosome of *V. cholerae* and (B) the smaller chromosome (containing the integron/gene cassette system; described in 1.6 in detail). The top half of both (A) and (B) show pathogenic strains of *V. cholerae* of the O1 serogroup. The bottom half of both (A) and (B) show non-O1/O139 strains of *V. cholerae*. Red regions correspond to regions between all 23 genomes that are not similar to the reference strain N16961. These regions correspond largely to MGEs that have shaped the pathogenicity of *V. cholerae* and are indicated at the top on the figure by arrows. This figure has been adapted from Chun *et al., PNAS,* 2009 (Chun *et al.*, 2009).

1.6: The integron/gene cassette system

The integron/gene cassette system is an important MGE present in nearly all sequenced *Vibrio* genomes and this system facilitates rapid evolution in their bacterial host (Mazel *et al.*, 1998;Biskri *et al.*, 2005).

The integron/gene cassette system is a genetic platform that enables its bacterial host to rapidly acquire new genes as part of mobile elements called gene cassettes (shown in Figure 1.8) (Stokes and Hall, 1989;Mazel, 2006;Boucher *et al.*, 2007). The integron consists of three components: a gene (*intI*) that encodes an integrase (enzyme responsible for DNA recombination), an attachment site (*attI*) and a promoter region (P_c) (Hall *et al.*, 1991;Rowe-Magnus and Mazel, 2001). Gene cassettes are mobile genetic elements, commonly consisting of a single, promoter-less open reading frame (ORF; portion of an organisms gene that encodes a protein) and contain an *intI*-identifiable recombination site called *attC* (Rowe-Magnus *et al.*, 1999;Gillings *et al.*, 2008;Labbate *et al.*, 2009). The insertion (and excision) of gene cassettes is facilitated by an integrase-mediated recombination reaction between *attI* and *attC* where (Collis and Hall, 1992;Partridge *et al.*, 2000;Bouvier *et al.*, 2005;MacDonald *et al.*, 2006) cassettes can be accumulated forming a contiguous cassette array that are numbered sequentially (Rowe-Magnus *et al.*, 2003). Adjacent gene cassettes are inserted downstream of the promoter, P_c, which potentially transcribes the cassette-associated ORFs.

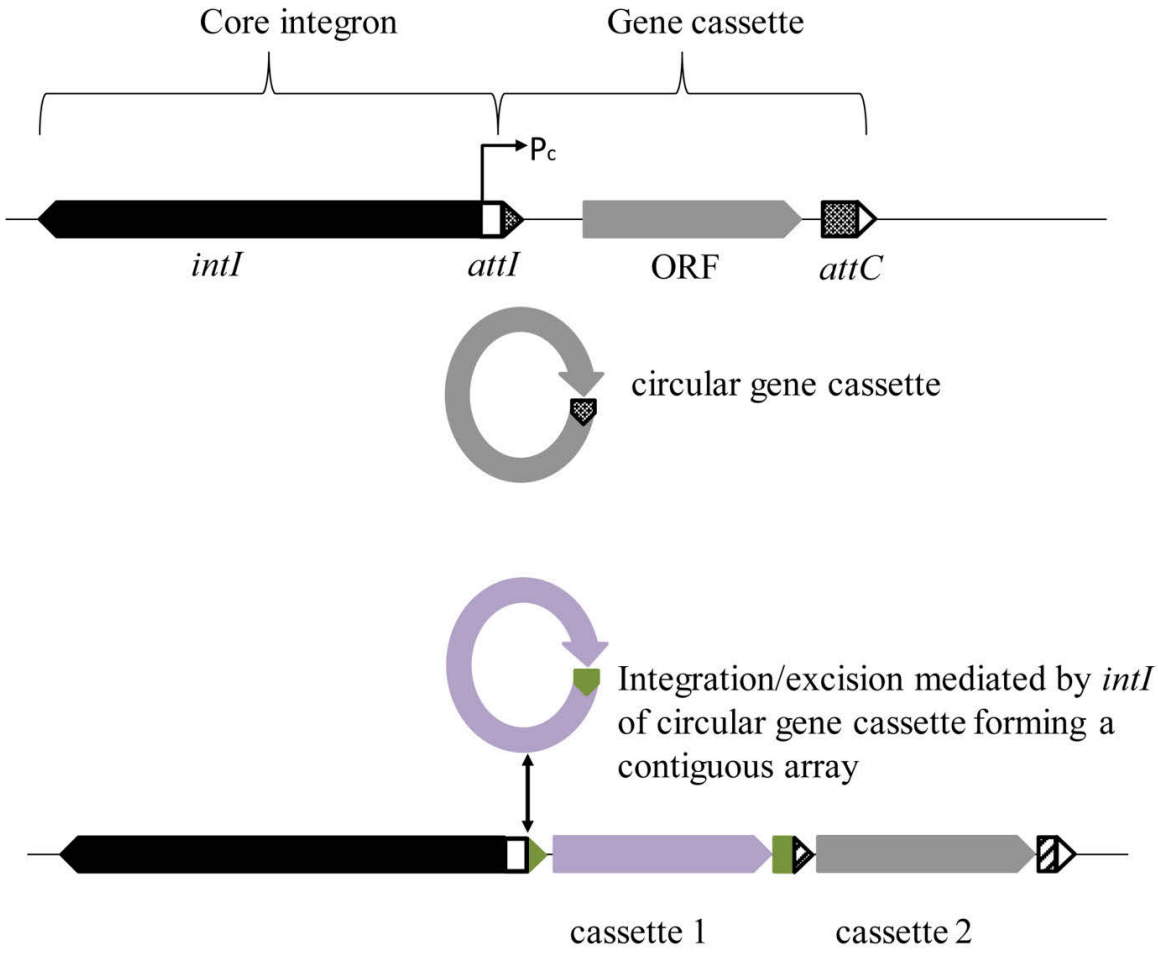


Figure 1.8: The structure of the integron/gene cassette system.
The integron assembly platform consists of the *intI* gene, *attI* site and P_c promoter. The gene cassette usually consists of a promoterless ORF and a recombination site *attC*. Freely, the gene cassette is circularised, and is recombined into the integron *via* the integrase site-specific recombination reaction between *attI* and *attC* (recombination can occur less frequently between *attC* and *attC*). This image shows a two-cassette array (normally numbered from the cassette closest to the *attI* site). Gene expression occurs *via* the P_c promoter.

Integrons were first described as a distinct genetic element in 1989 in association with mobile transposons and plasmids whose associated gene cassettes were identified to confer multiple antibiotic resistance in clinically-derived bacteria (Stokes and Hall, 1989). The first described integrons (since called class 1 integrons (Recchia and Hall, 1995)) are a common fixture in clinically important Gram negative pathogens and human commensals, commonly containing 1-6 gene cassettes. Integrons are a diverse family of elements and are catalogued into different classes based on nucleotide sequence of the integrase gene,

Currently, there are over 100 different integron classes, with most present on bacterial chromosomes and found in approximately 10% of sequenced bacterial genomes (Mazel, 2006;Boucher *et al.*, 2007). The identification of integrons distinct from class 1 integrons and located on immobile regions of the bacterial chromosome indicates that the integron has a much broader role in bacterial adaptation than simply conferring antibiotic resistance (Rowe-Magnus *et al.*, 2003;Mazel, 2006;Koenig *et al.*, 2009). This is supported by the presence of gene cassettes in chromosomal integrons coding for products unrelated to antibiotic resistance. Table 1.7 show a representative list of bacterial strains containing chromosomal integrons (Rowe-Magnus *et al.*, 2001;Mazel, 2006;Boucher *et al.*, 2007) It is evident from this table that gene cassette numbers vary and that members of the *Vibrio* genus generally carry large cassette arrays. Furthermore, the presence of antibiotic resistance genes in chromosomal integrons and the high homology of the clinical class 1 integron to integrons found in β-proteobacteria suggests that clinically important mobile class 1 integrons emerged from the mobilisation of a chromosomal integron from the β-proteobacteria into a transposon (Rowe-Magnus *et al.*, 2001;Mazel, 2006;Stokes *et al.*, 2006;Boucher *et al.*, 2007;Gillings *et al.*, 2008).

Table 1.7: A short list of bacterial species that contain chromosomal integrons

Bacterial group	**Bacterial strain**	**Number of gene cassettes**
Vibrionaceae	*Vibrio cholerae* N16961	179
	Vibrio vulnificus CMCP6	217
	Vibrio rotiferianus DAT722	116
Pseudoalteromonas	*Pseudoalteromonas tunicata* D2	7
	Pseudoalteromonas haloplanktis TAC125	5
Xanthomonadaceae	*Xanthomonas campestris* pv. campestris ATCC 33913	22
	Xanthomonas campestris pv. vesicatoria	3
Pseudomonadaceae	*Pseudomonas stutzeri* Q	>7
	Pseudomonas alcaligenes	32
Planctomycetes	*Rhodopirellula baltica* SH1	0
Spirochaetales	*Treponema denticola* ATCC35405	45

Although not as readily mobilisable as many class 1 integrons, chromosomal integrons have still been found to be present in bacteria separated by long phylogenetic distances (millions of years) (Mazel, 2006;Boucher *et al.*, 2007). When compared to mobile class 1 integrons, chromosomally located integrons are 'fixed' over evolutionary time periods in their bacterial host and as a result have co-evolved with their host and most likely influenced their evolution through LGT events and gene cassette rearrangements. For example, microbes inhabiting deep-sea hydrothermal vents have been found to contain integrons (Elsaied *et al.*, 2007). The discovery of integrons in bacteria from hostile and remote environments where bacterial populations are subject to numerous stresses including high water pressure, high temperatures and toxic gases, demonstrates two main points; 1) the integron/gene cassette system is an ancient genetic structure, 2) gene cassettes are clearly playing a much broader role in adaptation and evolution of bacteria than simply conferring antibiotic resistance (Rowe-Magnus *et al.*, 2003;Elsaied *et al.*, 2007).

1.6.1: A big black box in our understanding of gene cassette-associated function(s)

Cassette arrays in *Vibrio* species are generally quite large (sometimes >200 cassettes) and mostly consist of unique and novel genes with no identifiable function. A bioinformatics survey of gene cassettes from multiple genomes of sequenced *Vibrio* species found that 65% of cassette-associated proteins had no known homologues and that 13% had homologues of unknown function (Boucher *et al.*, 2007). The remainder showed a wide range of non-specific functions involved in metabolism, cellular processes and information storage. Similar statistics have been observed through PCR-amplification of gene cassettes from metagenomic DNA (Stokes *et al.*, 2001;Elsaied *et al.*, 2007;Koenig *et al.*, 2008;Koenig *et al.*, 2009). Putting aside this massive knowledge gap in cassette function, large cassette arrays provide an extra level of complexity. While some argue that P_c is the only driver of cassette transcription in large arrays (Kuadkitkan *et al.*;Guerin *et al.*, 2009;Cambray *et al.*, 2010), other studies have shown otherwise(Yildiz *et al.*, 2004;Michael and Labbate, 2010). A study of the 116-gene cassette array of *V. rotiferianus* DAT722 (used as a model organism in this thesis) showed that the majority of gene cassettes were transcribed and that numerous diverse promoters across the array were present that responded to different growth conditions (Michael and Labbate, 2010). The presence of diverse promoters along the entire cassette array means that integrated cassettes may be expressed/activated by multiple regulatory pathways. This gives the capacity for cassettes to re-arrange with different promoters potentially building operon-like structures that express complimentary cassette-associated proteins when required by the bacterium. Such an idea has been demonstrated in principal using artificial gene cassettes encoding for tryptophan biosynthesis (Bikard *et al.*, 2010). This level of intricacy is important when we consider that *Vibrio* spp. live in populations where gene cassettes might be considered a community resource not just a singular cell resource. For example, integrons might provide a mode for the community to break down and or extract energy from complex substrates without the entire pathway (and genetic burden) being owned by just one cell (Labbate *et al.*, 2009;Elsaied *et al.*, 2014).

Further to this a study in 2009 by Guerin *et al*. showed that the SOS-response induces up-regulation of *IntI* expression by 4.5 fold (Guerin *et al.*, 2009). The authors of this study have shown that the up-regulation of integrase expression by the SOS-response enables insertion of exogenous cassettes and/or shuffling of contained cassettes, consequently allowing expression of cassette products and potentially providing a selective advantage during times of stress. In conjunction with the recent finding that diverse cassette promoters respond to different conditions are present across large cassette arrays, the potential for specific gene cassettes to be matched to each other and to an appropriate promoter through cassette shuffling is intriguing. Therefore, the integron/gene cassette system is likely to be a dynamic tool for rapid evolution in unfamiliar or demanding environments.

Even with the limited understanding of gene cassette function(s), a number of studies have sampled the gene cassette metagenome from different environments and attempted to infer or determine how cassettes might influence adaptation and evolution (Elsaied *et al.*, 2007;Koenig *et al.*, 2008;Koenig *et al.*, 2009;Elsaied *et al.*, 2011;Koenig *et al.*, 2011;Elsaied *et al.*, 2014). Although correlations are observed such as homologues of genes in cassettes encoding potential pollution degrading enzymes from contaminated environments (Nemergut *et al.*, 2004;Koenig *et al.*, 2009) or environments showing a "gene cassette ecotype"(Koenig *et al.*, 2008), it is still the case that ~80% of the gene cassettes are of unknown function. In a study looking at gene cassettes from *Vibrio* spp. found in coral mucus, 12.5% of gene cassettes were implicated in biochemical processes identified in antibiotic resistance(Koenig *et al.*, 2011). The authors conclude that gene cassettes provide a competitive advantage by delivering protection from, or by synthesising, antimicrobials in the coral environment. This general inferred conclusion reflects the amount of research that has been done in the integron/resistance field. No other conclusions on the other cassette-assisted bacterial interactions present in the coral mucus could be made. This illustrates the knowledge gap in our understanding of how gene cassette-associated products contribute to general adaptation and evolution.

Of the many thousands of non-antibiotic resistance gene cassettes identified by genome or metagenome sequencing, only in less than 20 has the function of the encoded protein been

determined by functional characterisation (Table 1.8). In many instances, characterisation of these gene cassettes was selected based on some homology to a known protein such that a phenotype could be tested. This approach obviously does not really address the bulk of unknown and hypothetical gene cassette products. In other examples, their selection was based on the capacity of each cassette to be crystallised or identified as part of mutant library screen. In the instances where gene cassettes have been removed from their natural bacterial host and expressed in *E. coli* or where *in vitro* techniques have been used to study protein activity, caution must be taken in how their function is interpreted. Interactions of cassette proteins with host pathways may modify how these gene cassettes affect cell or community behaviour. This was observed in a study in *V. rotiferianus* DAT722 where deletion of a gene cassette encoding a putative topoisomerase I-like protein affected porin regulation. This phenotype could not have been predicted if characterised outside the host (Labbate *et al.*, 2011). This is also true of the bioinformatic studies described above where in the small proportion of gene cassette products that could be identified were often proteins such acetyltransferases, methylases or transcriptional regulators. Without knowing the primary substrate that is being modified by the acetyltransferase or methylase or the gene(s) controlled by the transcriptional regulator, the biological importance of the cassette(s) is still unclear.

It is clear that our ability to understand what cassette proteins "do" lags well behind our ability, *via* efficient high throughput DNA sequencing, to discover the associated cassettes. Since it has been estimated that gene cassettes comprise an enormous meta-gene pool with, as yet, no identifiable upper limit in terms of numbers (Michael *et al.*, 2004), continued sequencing as a cassette discovery tool is worthwhile. However, biological systems can only be understood by knowledge of how gene products function and interact with other cell components to make evolutionarily successful reproductive units. Merely knowing the size of a cell or community's gene pool is no help in this regard. Our research group has long argued that understanding the mobile gene pool in a functional sense is central to understanding prokaryotic biology and evolution given the fact that the mobile metagenome greatly outnumbers fixed genes in a prokaryotic community (Michael *et al.*, 2004;Gillings, 2005;Boucher *et al.*, 2007;Robinson *et al.*, 2008;Labbate *et al.*,

2009;Labbate *et al.*, 2011;Stokes, 2011;Labbate *et al.*, 2012). In addition to the intrinsic scientific importance of this metagenome, it is also clear that human activities are impacting on prokaryotic evolution, mostly to the detriment of the former and that the distinction between the 'environment' and 'clinical' settings are blurred (Holmes *et al.*, 2003;Gillings and Stokes, 2012;Gillings, 2013). This impact is such that it has been argued we are entering a new "Anthropocene" era (Gillings and Hagan-Lawson, 2014). Thus, at a number of levels, more systematic investigation of how mobile genes can impact bacterial evolution is highly desirable.

Table 1.8: Brief list of some non-antibiotic resistance functional ORFs in gene cassettes

Source of cassette	Function	Determination of function	Reference
Cassettes from *Vibrio* species			
Vibrio cholerae	Sulfate-binding protein	Complementation of *E. coli* mutation	(Rowe-Magnus and Mazel, 2001)
Vibrio cholerae OP4G	Transcriptional regulation	Crystal structure determination and drug binding assay	(Deshpande *et al.*, 2011)
Vibrio cholerae GP156	Heat stable enterotoxin	Active in suckling mouse assay when expressed in *E. coli*	(Ogawa, 1993)
Vibrio cholerae	mannose-fucose resistant hemagglutinin	Mutagenesis *in vivo* and testing in infant mouse model	(Franzon *et al.*, 1993;Barker, 1994)
Vibrio marinus	Psychrophilic lipase	Active when expressed in *E. coli* at 10 °C	(Rowe-Magnus and Mazel, 2001)
Vibrio vulnificus CMCP6	Cold shock	Complementation of cold shock phenotype in *E. coli*	(Rowe-Magnus, 2009)
Vibrio vulnificus CMCP98K	Secretion	Expression in *E. coli* mediates secretion of periplasmic proteins	(Kim *et al.*, 2003)
Vibrio rotiferianus DAT722 (cassette 21)	dNTP pyrophosphohydro-lase (iMazG)	Crystal structure determination. Expressed in *E. coli* and enzyme activity measured	(Robinson *et al.*, 2007)
Various large cassette arrays like this in *Vibrio* spp.	Toxin/antitoxin (TA) genes	Demonstration that presence of TA genes limits deletions in large cassette arrays	(Szekeres, 2007)
Vibrio vulnificus 1003	Capsular polysaccharide biosynthesis	Transposon mutagenesis *in vivo*	(Smith and Siebeling, 2003)

Vibrio rotiferianus DAT722 (cassette 11)	Porin regulation	Deletion of cassette *in vivo*	(Labbate *et al.*, 2011)
Vibrio rotiferianus DAT722 (multiple cassettes)	Surface polysaccharide modification	Deletion of cassettes *in vivo*	(Rapa *et al.*, 2013)
Cassettes from metagenomic DNA			
Soil metagenomic DNA	Potential transport protein	Crystal structure determination	(Robinson *et al.*, 2005)
Soil metagenomic DNA	ATPase activity	Expressed in *E. coli* and enzyme activity measured	(Nield *et al.*, 2001)
Soil metagenomic DNA	Methyltransferase activity	Expressed in *E. coli* and enzyme activity measured	(Nield *et al.*, 2001)

1.7: Project aims

The primary objective of the work presented in this thesis is to examine how LGT events can contribute to the adaptation and evolution of *Vibrio* spp. By examining how two specific MGEs transferred *via* LGT affect cell physiology and function in two different *Vibrio* spp. insight was gained into how the transfer of genetic information can contribute to the intricate functioning of the bacterium.

1.7.1: How does the integron/gene cassette system impact on cell physiology?

Results in chapters 3 and 4 aimed to elucidate the phenotype(s) conferred by gene cassette-associated products; apart from their involvement in antibiotic resistance. Given the enormous knowledge gap in the literature on the function of gene cassette-associate products with regard to broad adaptation, a novel approach to examining gene cassette-associated phenotypes was adopted in this study. This approach involved a subset of gene cassettes being deleted *in situ* in the large chromosomal integron cassette array of the model organism *Vibrio rotiferianus* DAT722. The data presented in chapter 3 has led to the discovery of novel gene cassette-associated phenotypes and has since been published in the peer reviewed journal *PLoS ONE*. By expressing specific gene cassettes to an environmental non-O1/O139 strain of *V. cholerae* designated S24 and inducing expression, specific phenotypes associated with an individual cassette from *V. rotiferianus* DAT722 is explored in chapter 4.

1.7.2: How does a novel genomic island impact on cell survival and overall fitness?

Strains of environmental *V. cholerae* are an important source of laterally acquired DNA for *V. cholerae* that is capable of causing large outbreaks of the disease cholera. The first aim of chapter 5 was to characterise a novel 32 kb genomic island (GI) found in environmental non-O1/O139 *V. cholerae* strain S24. Chapter 5 outlines the unique genes present on this GI, phylogeny studies on the associated *recA* gene carried by the GI and examines potential mechanisms of mobilisation/dissemination of the GI. The second aim of this chapter examines the phenotype(s) this novel GI confers in the non-indigenous *E. coli* host, using transposon mutagenesis of two specific genes on the genomic island. Specifically, the

second part of this chapter aims to elucidate if the GI can provide *E. coli* with increased fitness and survival when exposed to two forms of DNA damaging stress: UV-irradiation and DNA targeting antibiotics.

Chapter 2: Materials and Methods

2.1: Bacterial strains and growth conditions

Vibrio rotiferianus DAT722 and isogenic deletion mutants used throughout this thesis were grown in two media: Luria-Bertani supplemented with 2% (w/v) NaCl (LB20) and marine minimal salts medium supplemented with 0.2% glucose as a carbon source (2M + glucose). The individual constituents of these media are given in Tables 2.1 and 2.2. LB20 was sterilised by autoclaving. 2M + 0.2% glucose was prepared by combining six sterile component solutions, each prepared as outlined in Table 2.2. Solid media were made by adding 1.5% (w/v) agar to each medium and pouring approximately 20 mL of agar medium into 10 cm petri-dishes. *E. coli* and *V. cholerae* strains were routinely cultured in Luria-Bertani supplemented with 0.5% (w/v) NaCl (LB5) with the exception of *E. coli* strain WM3064 which is auxotrophic for DAP (diaminopimelic acid); in this case DAP was added to growth media at a final concentration of 0.3 mM.

The *V. rotiferianus* DAT722 strains were routinely cultured from frozen stocks (growth medium plus 15% glycerol) by streaking onto LB20 plates using aseptic technique and incubating at 28°C for at least 18 hr. *E. coli* and *V. cholerae* strains were also routinely cultured from frozen stocks by streaking onto LB5 plates using aseptic technique and incubating at 37°C for at least 18 hr. Starter cultures (referred to as overnight cultures) for subsequent experiments were created by inoculating a colony from streaked plates into a 50 mL Falcon tube containing 5 mL of medium and incubating overnight at 28°C (for *V. rotiferianus* strains) or 37°C (for *E. coli* and *V. cholerae* strains) for 16-20 hr. Kanamycin (Amresco) and chloramphenicol (Sigma-Aldrich) were routinely used when growing cultures in this work. The antibiotic kanamycin was used at a final concentration of 100 μg/mL and was made up in molecular biology water and filtered. The chloramphenicol antibiotic was used at a final concentration of 12.5 μg/mL and was made up using molecular grade ethanol (Sigma-Aldrich). All strains used in this study are given in Table 2.3.

Growth curves of all strains were conducted in 24 well microtitre plates (Nunc) containing 500 µL of medium per well. The inoculum was from overnight cultures grown in relevant media and then diluted to OD_{600} of 0.7 using the same medium used for growth. Growth curve cultures were inoculated at 1:100 and growth measurements recorded every 30 minutes using a microtitre plate reader (Synergy HT Bio-Tek) at OD_{600nm}, and Gen5 (Bio-Tek) software.

Table 2.1: Media and solution constituents

Media/solution	Constituent(s)	Amount of constituent added per litre of medium/solution
Bacteriological media		
LB20 *Made up with distilled water and autoclaved*	Tryptone Yeast Extract NaCl	10 g 5 g 20 g
LB5 *Made up with distilled water and autoclaved*	Tryptone Yeast Extract NaCl	10 g 5 g 5 g
5X M9 salts *Used at 1X stock. Made up with distilled water and autoclaved. For growth, a carbon source was added to a final amount of 0.2% (w/v)*	$Na_2HPO_4.7H_2O$ KH_2PO_4 NaCl NH_4Cl	64 g 15 g 2.5 g 5 g
Phosphate buffered saline (PBS) *Buffered to pH 7.4 using NaOH, made up with distilled water and autoclaved.*	NaCl KCl Na_2HPO4 KH2PO4	8 g 0.2 g 1.44 g 0.24 g
2% NaCl *Made up with distilled water and autoclaved*	NaCl	20 g
Molecular biology reagents		
Molecular biology water *Filtered using a 0.2µM filter and autoclaved*	Distilled water	As required
XS buffer (for DNA extractions) *Filtered using a 0.2µM filter.*	C_2H_5OCSSK (Potassium ethyl xathanogenate) 4M $NH_4C_2H_3O_2$ (ammonium	10 g

Only 10 mL made up at any given time as the efficacy of this buffer decreases rapidly	acetate) 1M Tris-HCl pH 7.4 0.45 M EDTA 20% SDS	200 mL 100 mL 40 mL 50 mL
50X TAE (for DNA electrophoresis) *Made up with distilled water. No need for sterilisation. For use dilute using distilled water to 1X*	Tris Glacial acetic acid 0.5M EDTA	242 g 57.1 mL 100 mL
Solution 1* (for competent cells) *MES dissolved in 100 mL distilled water and pH adjusted to 6.2 with NaOH. Remaining constituents dissolved in 50 mL of distilled water and pH adjusted to 5.8 using acetic acid. Made up to a final volume of 200 mL with distilled water, filter sterilised and stored at 4°C*	10 mM MES (2-[N-morpholine]ethanesulfonic acid) 100 mM RbCl 10 mM $CaCl_2.2H_2O$ 50 mM $MnCl_2.4H2O$	0.39 g 2.42 g 0.294 g 1.98 g/ 200 mL
Solution 2* (competent cells) *Dissolved all constituents in distilled water and adjusted pH to 6.5 with KOH. Adjusted final volume to 200 mL and filter sterilised. Stored at 4°C*	10 mM MOPS (3-(N-morpholino) propanesulfonic acid) 75 mM $CaCl_2.2H_2O$ 10 mM RbCl 15% glycerol	0.419 g 2.2053 g 0.242 g 30 mL/ 200 mL

*amount per 200 mL not 1 L

Table 2.2: 2M + 0.2% glucose constituents

Components	Constituent chemical(s)	Amount of constituent chemical	Amount of constituent added per litre of medium
1. 0.55 x Nine Salt Solution (NSS) *Autoclaved*	NaCl Na_2SO_4 $NaHCO_3$ KCl KBr $MgCl_2.6H_2O$ $CaCl_2.2H_2O$ $SrCl_2.6H_2O$ H_3BO_3	8.8 g 0.735 g 0.04 g 0.125 g 0.02 g 0.935 g 0.205 g 0.004 g 0.004 g /920mL	920 mL
2. 1 M MOPS buffer *Adjusted to pH 7.4 using NaOH pellet and filter sterilised*	3-(N-Morpholino) propanesulfonic Acid (MOPS)	104.65 g/500 mL	40 mL
3. 0.4 M Tricine + 1mM $FeSO_4.7H_2O$ *Adjusted to pH 8.2 using 7M NaOH solution and filter sterilised*	Tricine $FeSO_4.7H_2O$	17.83 g/250 mL 0.07 g	10 mL
4. 953 mM NH_4Cl *Adjust to pH 7.8 using 7M NaOH solution and autoclaved*	NH_4Cl	12.73 g/250 mL	10 mL
5. 132 mM K_2HPO_4 *Autoclaved*	K_2HPO_4	5.75 g/250 mL	10 mL
6. 20% glucose *Filter sterilised*	Glucose	20 g/100 mL	10 mL

Table 2.3: List of bacterial strains and plasmids

Strain or plasmid	Relevant genotype[1]	Reference or source
V. rotiferianus		
DAT722	Wild-type	(Boucher *et al.*, 2006)
DAT722-Sm	DAT722; spontaneous Sm^R mutant	(Labbate *et al.*, 2011)
MD7	DAT722-Sm; Single recombination cross-over of pMAQ1081 into cassette 61, Km^R	(Labbate *et al.*, 2011)
SC-8B61	DAT722-Sm; Single recombination cross-over of pMAQ1081 into cassette 61, Km^R	This study
SC-8A91	DAT722-Sm; Single recombination cross-over of pMAQ1081 into cassette 93, Km^R	This study
d16-60	DAT722-Sm; Δcassettes 16-60, Sm^R, Km^R	(Labbate *et al.*, 2011)
d16-60a	DAT722-Sm; Δcassettes 16-60, Sm^R, Km^R	This study
d50-60	DAT722-Sm; Δcassettes 50-60, Sm^R, Km^R	This study
d72-92	DAT722-Sm; Δcassettes 72-92, Sm^R, Km^R	This study
d72-92a	DAT722-Sm; Δcassettes 72-92, Sm^R, Km^R	This study
V. cholerae		
S24	Wild-type (non-O1/O139)	This study
S22	Wild-type (non-O1/O139)	(Islam *et al.*, 2013)
S10	Wild-type (non-O1/O139)	(Islam *et al.*, 2013)
S11	Wild-type (non-O1/O139)	(Islam *et al.*, 2013)
S12	Wild-type (non-O1/O139)	(Islam *et al.*, 2013)
S16	Wild-type (non-O1/O139)	(Islam *et al.*, 2013)
S18	Wild-type (non-O1/O139)	(Islam *et al.*, 2013)
S22	Wild-type (non-O1/O139)	(Islam *et al.*, 2013)
S23	Wild-type (non-O1/O139)	(Islam *et al.*, 2013)
S25	Wild-type (non-O1/O139)	(Islam *et al.*, 2013)
E. coli		
XL-1-Blue	Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA*96 *relA1 lac [F'* proAB *lacI*^{q}*ZΔ*M15 *Tn*10 *(Tet*^{r}*)]*	Stratagene

RK600	ColE1 *oriV*; RP4*tra*$^{+}$ RP4 *oriT*; Cm^{R}; with helper plasmid in tri-parental matings	(Kessler *et al.*, 1992)
DH5αλpir	*endA1 glnV*44 *thi-1 recA1 relA1 gyrA96 deoR nupG* Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *hsdR17 λpir*	(Demarre *et al.*, 2005)
WM3064	Donor strain for conjugation: *thrB1*004 *pro thi rpsL hsdS lacZΔM15* RP4-1360 Δ(*araBAD)567* Δ*dapA*1341::[*erm pir*], Sm^{R}	(Saltikov and Newman, 2003)
EPI300^{TM}-T1^{R}	(F^{-} *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80d*lacZ*ΔM15 Δ*lacX*74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ^{-} *rps*L *nup*G *trf*A *ton*A *dhfr*], Sm^{R}, Tc^{R}	Epicentre Biotechnologies
TransforMax™ EC100™	*F^{-} mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ^{-} rpsL (Str^{R}) nupG*	Epicentre Biotechnologies
Plasmids/Fosmids		
pMAQ1080	pGEM-T Easy carrying a 1834-bp fragment. The fragment was created using fusion PCR and consists of, in order, a 448-bp of paralog group 1 sequence, a 964-bp fragment containing *aphA1* and a 410-bp paralog group 2 sequence abutted by *sal*I restriction sites.	(Labbate *et al.*, 2011)
pCVD442	Mobilisable *sacB* counter-selectable suicide vector, Ap^{R}	(Donnenberg and Kaper, 1991)
pGEM-T Easy	Cloning vector, Ap^{R}	Promega
pJAK16	Low copy IPTG-inducible expression vector, Cm^{R}	(Thomson *et al.*, 1999)
pJAK16::VSD31	pJAK16 containing the integron gene cassette VSD31	This study
pJAK16::VSD78	pJAK16 containing the integron gene cassette VSD78	This study
pSU-pBAD	Cloning vector, Cm^{R}	(Le Roux *et al.*, 2007)
pSU-pBAD::VSD31	pSU-pBAD containing the integron gene cassette VSD31	This study
pSU-pBAD::VSD78	pSU-pBAD containing the integron gene cassette VSD78	This study

pSU-pBAD::VSD54	pSU-pBAD containing the integron gene cassette VSD54	This study
pCC2FOS	Cloning vector, Cm^R	Epicentre
pCC2FOS::RME	pCC2FOS vector containing 32 kb insert from *V. cholerae* S24. The insert contains the *recA* GI and surrounding sequence, Cm^R	This study
pCC2FOS::RMEΔ$recA_{RME}$	pCC2FOS vector containing 32 kb insert from *V. cholerae* S24. The insert contains the *recA* GI and surrounding sequence and has *recA* on the GI insertionally inactivated by Tn5, Km^R, Cm^R	This study
pCC2FOS::RMEΔIS-2	pCC2FOS vector containing 32 kb insert from *V. cholerae* S24. The insert contains the *recA* GI and surrounding sequence and has insertion sequence (ISVuv4; RME012 see Figure 5.3) closest to IR_i end on the GI insertionally inactivated by Tn5, Km^R, Cm^R	This study
pCC2FOS::RMEΔ$umuC_{RME}$	pCC2FOS vector containing 32 kb insert from *V. cholerae* S24. The insert contains the *recA* GI and surrounding sequence and has *umuC* present on the GI insertionally inactivated by Tn5, Km^R, Cm^R	This study
$pOriVn_{700}$	Low copy mobilizable vector containing ori_{pB1067} (*Vibrio* specific) and *ori6K*, Sp^R	(Le Roux *et al.*, 2011)
$pOriVn_{700}$-$recA_{S22}$	$pOriVn_{700}$ with *recA* from *V. cholerae* S22 in between ori_{pB1067} and *ori6K*. The *recA* gene is reading toward ori_{pB1067}, Sp^R	This study
$pOriVn_{700}$-P_{lac}*gfp*	$pOriVn_{700}$ with P_{lac}*gfp* cloned in between ori_{pB1067} and *ori6K,* Sp^R	(Le Roux *et al.*, 2011)

[1] Sm^R, streptomycin resistance; Km^R, kanamycin resistance; Tc^R, tetracycline resistance; Sp^R, spectinomycin resistance; Cm^R, chloramphenicol resistance, Km^R ;kanamycin resistance; Ap^R ampicillin resistance.

2.2: List of primers

Table 2.4: List of primers

Primer	Sequence (5'-3')	Target	Source
pJAK16-F	GCTGTGCAGGTCGTAAATCA	Forward primer used for sequencing VSD31,VSD78 and VSD54 once cloned into pJAK16	This study
pJAK16-R	CTGGCAGTTCCCTACTCTCG	Reverse primer used for sequencing VSD31,VSD78 and VSD54 once cloned into pJAK16	This study
pSU-pBAD-R	CTGACGCTTTTATCGCAAC	Reverse primer used for sequencing VSD31,VSD78 and VSD54 once cloned into pSU-pBAD	(Le Roux *et al.*, 2007)
pBAD-out-2	CGAAAAAGGATGGATATACCG	Forward primer used for sequencing VSD31,VSD78 and VSD54 once cloned into pSU-pBAD	(Judson and Mekalanos, 2000)
EcoRI-pSU-pBAD-R*	TTTT**GAATTC**CGTTTCACTCCATCCCAAAAAAAC	Reverse primer for amplifying pSU-pBAD	This study
XbaI-pSU-pBAD-F*	CGAG**TCTAGA**CAGCGCTTTTCC	Forward primer for amplifying pSU-pBAD	This study
EcoRI-VSD78-F*	TTTT**GAATTC**ATGAAAGCACTAAGTAACAGAGAAA	VSD78 for cloning into pSU-pBAD	This study
NheI-VSD78-R*	TTTT**GCTAGC**TTAGGTGCGTGTTTCAAC	VSD78 for cloning into pSU-pBAD	This study
EcoRI-VSD54-F*	TTTT**GAATTC**ATGAAGATTGAGGTGCATGAA	VSD54 for cloning into pSU-pBAD	This study
Xbal-VSD54-R*	TTTT**TCTAGA**GTTATATTTTTGGCGGCTTACTG	VSD54 for cloning into pSU-pBAD	This study
EcoRI-VSD31-F*	TTTT**GAATTC**ATGGAGAACAAATTGAAAGATTA	VSD31 for cloning into pSU-pBAD	This study
NheI-VSD31-R*	TTTT**GCTAGC**CTTTGTTATACGAACATTTCATCG	VSD31 for cloning into pSU-pBAD	This study
Xbal-pSU-pBAD-F*	CGAG**TCTAGA**CAGCGCTTTTCC	Forward primer used for cloning VSD31,VSD78 and VSD54 into pSU-pBAD	This study
EcoRI-pSU-pBAD-R*	TTTT**GAATTC**CGTTTCACTCCATCCAAAAAAAC	Reverse primer used for cloning VSD31,VSD78 and VSD54 into pSU-pBAD	This study
BamHI-VSD31-F*	TTTT**GGATCC**TCGCAAAATTTAACATGGAGAA	VSD31 for cloning into pJAK16	This study
PstI-VSD31-R*	TTTT**CTGCAG**CTTTGTTATACGAACATTTCATCG	VSD31 for cloning into pJAK16	This study
BamHI-VSD78-F*	TTTT**GGATCC**CTTTGAGCAATCTTCAACAAG	VSD78 for cloning into pJAK16	This study
PstI-VSD78-R*	TTTT**CTGCAG**TTAGGTGCGTTTCAACC	VSD78 for cloning into pJAK16	This study

RME-R	GACGAGTCCAGCTCATGACA	integrase end of *recA* genomic island	This study
RME-F	GCTGCTAACGCTTTCTGCTT	*recA* end of *recA* genomic island	This study
S24-ctg675-F	CGGTTAGGAGGGGCTTTTAG	3' end of contig 675 and used for screening for Tn5 insertion into IS-2 of RME	This study
S24-ctg708-R	TATCGGCTGTGGTTGTTTGA	5' end of contig 675	This study
S24-ctg367-F	TAGCTAGAGCATTTGTCATA AGAAAAAGTAAG	3' end of contog 675	This study
S24-ctg367-R	ACTGGCAGCAGAAGAAGCA T	5' end contig 708	This study
S24-cinA-F	CAAGGTTGGCTCAAAGTG	*cinA* in *V. cholerae* S24	This study
S24-recX-R	GGCATCACTCAAATACCCTA	*recX* in *V. cholerae* S24	This study
S24-recA-F	CTGGAAATTTGTGATGCATT	*recA* in *V. cholerae* S24	This study
EcoRI-recA-F*	TTTT**GAATTC**TGGACGAGAA TAAACAGAAGG	*recA* in *V. cholerae* S22 & S24	This study
EcoRI-recA-R*	TTTT**GAATTC**AAACTCTTC TGGCACCGC	*recA* in *V. cholerae* S22 & S24	This study
EcoRI-Ori700-R*	TTTT**GAATTC**CGCGCTATCG CTTGTCG	ori_{pB1067} of pOriVn$_{700}$	This study
EcoRI-OriR6K-F*	TTTT**GAATTC**GTGTTCCTGT GTCACTCAAAATTG	*ori6k*	This study
Ori700-F	CCCTATTCCTCTTTAGTCCTG C	ori_{pB1067} of pOriVn$_{700}$	This study
Ori6K-R	TAACGCACTGAGAAGCCC	*ori6k*	This study
S24-phage-Int-F	GCCAAGATATGGCAGGAAA A	Integrase in *recA* genomic island	This study
S24-phage-Int-R	GGACGCTACCCAGTGAATGT	Integrase in *recA* genomic island	This study
recA-F	TGGACGAGAATAAACAGAA GGC	*recA*	(Boucher, 2006)
recA-R	CCGTTATAGCTGTACCAAGC GCCC	*recA*	(Boucher *et al.*, 2011)(3) (Boucher, 2006)
pCC2FOS-FP	GTACAACGACACCTAGAC	pCC2FOS sequencing primers (F)	Epicentre Biotechnologies
pCC2FOS-RP	CAGGAAACAGCCTAGGAA	pCC2FOS sequencing primers (R)	Epicentre Biotechnologies
recA-Tn5-F	CGCTCATAAGTCAGTAATGC TTCA	*recA* on genomic island. Used to screen for Tn*5* insertion.	This study
umuC-Tn5-F	GATGTATGGCTGAATCGACC A	*umuC* on genomic island. Used to screen for Tn*5* insertion.	This study
KAN-2 FP-1	ACCTACAACAAAGCTCTCAT CAACC	Forward primer inside Tn5 used to screen for Tn5 insertion.	Epicentre Biotechnologies
KAN-2 RP-1	GCAATGTAACATCAGAGATT TTGAG	Reverse primer inside Tn5 used to screen for Tn5 insertion.	Epicentre Biotechnologies

VSD14-F	AAAGCG GTTACATTCGGG	Forward primer used to screen for d16-60 and d16-60a deletion mutants	(Labbate *et al.*, 2011)
VSD47-F	CATTTTAAGTCGGCTCTTCC	Forward primer used to screen for the d50-60 deletion mutant	(Labbate *et al.*, 2011)
VSD70-F	GTGCGTGGATATCATAGAAC G	Forward primer used to screen for the d72-92 and d72-92a deletion mutants	This study
VSD94-R	CGCTCCTGCCAGCATATC	Reverse primer used to screen for the d72-92 and d72-92a deletion mutants	This study
VSD62-R	GTAGGTAATTTCGGC TTCTCG	Reverse primer used to screen for the d16-60, d16-60a and and d50-60 deletion mutants	(Labbate *et al.*, 2011)

*bold and underlined sequence shows the restriction enzyme cleavage site

2.3: DNA methods

2.3.1: PCR and agarose gel electrophoresis

Polymerase chain reaction (PCR) was performed to amplify DNA sequences using an Eppendorf Mastercycler (epgradientS) thermocycler. The following general protocol was performed: denaturation at 93°C, annealing of primers at specific temperatures according to primer design (primer sequences shown in Table 2.4), extension at 72°C for approximately 1 min per kb of product and a final 10 min cycle at 72°C. Samples were then held at 10°C until required. Denaturation, annealing and extension were repeated for 30 cycles. PCR reactions were performed using 2X PCR MangoMix (Bioline) containing DNA polymerase, dNTPs, and $MgCl_2$ and forward and reverse primers added to a final volume of 0.5 µM. PCR products were run in a 1% agarose (made with TAE see section 2.1) gel (Bioline) in gel tanks (BioRad) at 100V. Gels were stained using ethidium bromide (0.0005 mg/mL in distilled water; Progen Industries Limited) and imaged using a UV transilluminator (LKB Bromma).

2.3.2: Genomic DNA extraction

2.3.2.1: Crude extraction for PCR

Genomic DNA from a single colony grown on an agar plate was extracted for use as template DNA in PCR reactions. This was done by growing cells overnight with shaking from a single colony in relevant media (see section 2.1 for a list of media used in this thesis). 500 µL of this overnight culture was taken and spun down in a 1.5 mL eppendorf tube at 16, 000 *x g* for 5 min. Supernatant was removed carefully using a pipette and the pellet was resuspended in 100 µL of molecular biology water and heated at 80°C for 15 min in a dry heating block (Labnet International). Cells were spun down at 16, 000 *x g* for 5 min and supernatant containing genomic DNA transferred to a sterile 1.5 mL tube and stored on ice or at -20°C until required.

2.3.2.2: XS buffer method

Genomic DNA (gDNA) purified by this method was used for construction of transposon mutant libraries and any other protocol requiring the use of gDNA other than PCR reactions. The protocol was adapted from Tillet and Neilan, 2000 (Tillett and Neilan, 2000). Briefly, cells were grown overnight and 1 mL was transferred to a 1.5 mL tube and cells pelleted at 16,000 *x g* for 5 min. Supernatant was then removed and cell pelleted resuspended in 750 µL of XS buffer (see Table 2.1) and incubated at 70°C for 1 hr. Samples were then mixed by vortexing for 10 s and incubated on ice for a further 30 min. Samples were again centrifuged at 14,000 *x g* for 10 min and supernatant carefully removed, transferred to a new tube and 1 volume of 100% isopropanol (Sigma-Aldrich) added. Samples were then incubated at room temperature for 5 min and centrifuged at 14,000 *x g* for 10 min. The pellet was then washed with 70% ethanol (Sigma-Aldrich) and resuspended in 200 µL of molecular biology water (see section 2.1). If DNA was used for cloning, it was treated with 5 µL of a 10 mg/mL RNase A (Amresco) stock for 30 min at room temperature to remove any contaminating RNA. A phenol/chloroform extraction was then performed by mixing a 1:1 phenol/chloroform solution (Amresco) and an equal volume then added to the DNA sample. The sample was then inverted to mix and centrifuged at 16 000 *x g* for 1 min. The upper phase was removed and transferred to a new 1.5 mL tube, and the sample extracted twice more with an equal volume of chloroform. This removes any residual phenol which is highly inhibitive to downstream reactions. After the final extraction, $1/10^{th}$ volume of 3M sodium acetate (Sigma-Aldrich) was added and precipitated with ethanol using 2.5 volumes of molecular grade ethanol (Sigma-Aldrich). Samples were then centrifuged at 16 000 *x g* DNA for 1 min to collect DNA and the pellet washed with freshly made 70 % ethanol. This removes salt from the sample. Samples were finally resuspended in 50 µL of molecular biology water (more if necessary).

2.3.3: Plasmid extraction

Extraction of plasmids was performed using two kits. For low yield plasmid preparation the PureYield Plasmid Miniprep system (Promega) was used. For high yield plasmid

preparation the PureYield Plasmid Maxiprep system (Promega) was utilised. Both protocols were performed according to manufacturer instructions.

2.3.4: Fosmid library construction, transposon mutagenesis and screening of RME

To clone the *recA* genomic island (RME) from *V. cholerae* strain S24, genomic DNA was digested with *Nae*I according to manufacturer instructions (New England Biolabs) and a library constructed using the CopyControl Fosmid Library Production Kit (Epicentre). *Nae*I digestion of *V. cholerae* strain S24 genomic DNA creates a fragment of 38,913-bp containing the entire 32,787-bp RME. The library was screened for a fosmid clone containing the 38,913-bp *Nae*I fragment using primers targeting the phage integrase in the RME (see section 2.3.1). A positive clone designated pCC2FOS::RME was confirmed by sequencing the ends of the cloned insert using the pCC2FOS vector primers FP and RP (see section 2.3.1). To create the pCC2FOS no insert control, linearised and dephosphorylated pCC2FOS (Epicentre) was treated with T4 polynucleotide kinase (New England Biolabs) and circularised by ligation. A mutant library of pCC2FOS::RME was constructed using the EZ-Tn5 Kan-2 Insertion Kit (Epicentre biotechnologies) according to manufacturer instructions in *E. coli* EC100. EC100 mutants containing knockouts of individual genes ($recA_{RME,}$ $umuC_{RME}$ and IS-2) present on the genomic island were screened by PCR using primers reading out from EZ-Tn5 Kan-2 and a primer targeting the gene of interest (Table 2.4). Once the required gene knockout was identified by PCR screening, the pCC2FOS containing RME with the specific gene knockout induced to high copy (only a single copy of fosmid per cell prior to induction) as per manufacturer instructions, extracted (section 2.3.3) and transformed into *E. coli* EPI300 (section 2.3.8) for subsequent experiments. It is interesting to note that when inducing expression of pCC2FOS containing the genomic island discussed in chapter 5 to high copy, induction for only 2 hr was required as compared to the manufacturer instruction of 5 hr-overnight. Induction longer than 2 hr resulted in degraded fosmid when pCC2FOS::RME was extraction *via* plasmid preparation. Although the exact reason for this is unknown, it is presumed genes present on this genomic island were impacting on the cell's ability to grow normally and affecting fosmid quality.

2.3.5: DNA and whole genome sequencing

Sequencing of PCR products was performed at Macrogen (Seoul) and samples were prepared according to Macrogen procedures (http://www.macrogen.com/eng/). Whole genome sequencing of *V. cholerae* S24 was carried out at the Wellcome Trust Sanger Institute, United Kingdom. Whole genome sequencing of *V. rotiferianus* DAT722 was performed at the Ramaciotti Centre (University of New South Wales, Sydney) using 454 sequencing technology (Chowdhury *et al.*, 2011).

2.3.6: Extraction, purification and quantification of DNA samples from agarose gels

Once DNA products were run on a 1% agarose gel as described in 2.3.1, bands containing the product of interest were excised and DNA was extracted and purified using Wizard SV Gel and PCR Clean-Up System (Promega) according to manufacturer instructions. Once extracted and purified, DNA was quantified using the NanoDrop Spectrophotometer (NS-1000) and acquisition software ND-1000 (version 3.8.1).

2.3.7: Preparation of competent cells

E. coli cells (strains WM3064, DH5α and XL-1-Blue) were grown in 50 mL LB5 to an OD_{600nm} of 0.5 and kept on ice for 15 min. Cells were then centrifuged at 4,600 *x g* and the supernatant discarded, and cells resuspended in 16 mL of ice cold solution 1 (Table 2.1). Cells were then kept on ice for 15 min and centrifuged at 4,600 *x g*. The supernatant was removed and cells resuspended in 1.5 mL of solution 2 (Table 2.1) and kept on ice for 15 min. Cells were stored in 100 µL aliquots at -80°C until required for transformation.

2.3.8: Transformation of competent cells

Competent cells were removed from -80°C storage and thawed on ice. Up to 2 µL (~5 ng) of plasmid DNA was added and cells incubated on ice for 30 min. Cells were then heated to 42°C for 90 s and recovered in 500 µL of LB5 at 37°C for 1 hr. After incubation cells were plated out in LB5 + 12.5 µg/mL chloramphenicol (Sigma-Aldrich) and incubated overnight at 37°C. The following day a single colony was picked and streaked out for purity.

Overnight cultures were then grown at 37°C and glycerol stocks made in LB5 + 15% glycerol (see section 2.1).

2.3.8.1: Chitin induced transformation of non-O1/O139 V. cholerae strains with RME

Prior to attempting chitin induced transformation of 8 non-O1/O139 strains of *V. cholerae* with the *recA* genomic island, the genomic island first had to be tagged with an antibiotic resistance marker to enable selection. This was done by *in vitro* transposon mutagenesis (see section 2.3.4) to introduce the Tn5 transposon carrying a kanamycin resistance marker into the second insertion sequence (ISVuv4; RME012) of the genomic island (see Figure 5.3). By choosing ISVuv4 it ensured that the function of any genes carried by RME were not insertionally inactivated. The chitin transformation protocol was performed according to Marvig and Blokesch (Marvig and Blokesch, 2010). Briefly, 80 mg of chitin powder (Sigma-Aldrich) was sterilised by autoclaving and dried prior to transformation. The 8 *V. cholerae* strains were grown in M9 medium (Table 2.1) supplemented with 32 mM $MgSO_4$ and 5 mM $CaCl_2$ at 30°C to an OD_{600nm} of approximately 0.5. Cells were then centrifuged at 16,000 *x g* for 10 min washed and resuspended in M9 medium. The autoclaved chitin powder was inoculated with 500 µL of resuspended cells, mixed thoroughly and incubated at 30°C for 16-20 hr. After this incubation *Nae*I digested RME was added to the cells (now competent) at an amount of ~ 2.5 µg and cells incubated at 30°C for 24 hr. Cells were then detached from chitin *via* vortexing for 30 s and plated out on LB5 + 50 µg/mL kanamycin.

2.3.9: Conjugation of integron gene cassettes from *E. coli* into *Vibrio*

2.3.9.1: Tri-parental conjugation of DAT722 deletion mutants

E. coli XL-1-Blue does not contain the necessary *tra* genes (transfer) required for successful conjugation of DNA into another bacterial strain. Thus, a tri-parental conjugation was performed in order to conjugate the gene cassettes cloned into the pJAK16 vector into the relevant integron deletion mutants. This was done by utilising the helper *E. coli* strain RK600; which possesses the necessary transfer genes in a helper plasmid. This plasmid will then transfer to donor cells *via* conjugation, allowing successful transfer of the plasmid of interest (pJAK16) containing genes to be expressed, to recipient cells. Donor,

helper and recipient cells were grown overnight and 200 µL of recipient cells (DAT722 deletion mutants), 50 µL helper cells (*E. coli* RK600) and 50 µL of donor cells (*E. coli* XL-1-Blue with donor plasmid) were combined and added to 5 mL of LB20 + 1 mM $MgSO_4$. Cell suspension was centrifuged at 4,600 *x g,* supernatant removed and the cells resuspended in 50 µL of LB20 + 1 mM $MgSO_4$ and spotted onto a 0.2 µM filter (Millipore) placed on a plate containing LB20 agar. The filter was incubated on the plate for 7 hr at 28°C. After incubation, cells were removed from filter using a vortex and serially diluted to 10^{-3} and 100 µL plated onto LB20 medium containing 100 µg/mL kanamycin (Sigma-Aldrich) and 10 µg/mL chloramphenicol. The following day a single colony was picked and streaked out for purity. Overnight cultures were then grown at 28°C and glycerol stocks made in LB20 + 15% glycerol.

2.3.9.2: Bi-parental conjugation of DAT722 deletion mutants

Integron gene cassettes cloned into the pSU-pBAD vector and transformed into *E. coli* WM3064 were able to be conjugated into the DAT722 integron deletion mutants directly from the *E. coli* WM3064 donor strain as it has the required *tra* genes. 500 µL of an overnight culture of both donor and recipient strains were combined in a 1.5 mL tube and centrifuged at 16,000 *x g* for 5 min and the supernatant removed. The pellet was then resuspended in 50 µL of LB20 and spotted onto a 0.2 µm filter placed on a plate containing LB agar containing 0.3 mM DAP (diaminopimelic acid is required for growth of *E. coli* WM3064). The filter was incubated on the plate for 7 hr at 28°C. After incubation, cells were removed from filter by placing in a tube with diluent and detached by vigorous vortexing. After cells were removed from the filter they were serially diluted to 10^{-3} and 100 µL plated onto LB20 medium containing 100 µg/mL kanamycin (Sigma-Aldrich), 10 µg/mL chloramphenicol and 1% glucose (to repress the pBAD promoter). The following day a single colony was picked and streaked out for purity. Overnight cultures were then grown at 28°C and glycerol stocks made in LB20 + 15% glycerol.

2.3.10: Cloning of genes

2.3.10.1: Cloning of integron gene cassettes into pJAK16

Plasmid pJAK16 (Figure 2.1) was purified as described in section 2.3.3 and treated with two restriction enzymes *Pst*I and *Bam*HI according to manufacturer instructions (New England Biolabs) and gel purified as described in section 2.3.6. VSD31 (***V****ibrio* **s**pecies **D**AT722) was amplified using primer pairs BamHI-VSD31-F/PstI-VSD31-R (see Table 2.4) and VSD78 with BamHI-VSD78-F/PstI-VSD78-R (Table 2.4) and gel purified (section 2.3.6). Digested plasmid and gene cassettes were ligated together using the T4 DNA ligase (New England Biolabs). After cloning VSD31 and VSD78 into pJAK16 they were transformed into competent *E. coli* XL-1-Blue cells where they were then conjugated using tri-parental conjugation into d16-60 and d72-92 respectively. Sequencing of gene cassettes was then performed using primers pJAK16-F/pJAK16-R (Table 2.4).

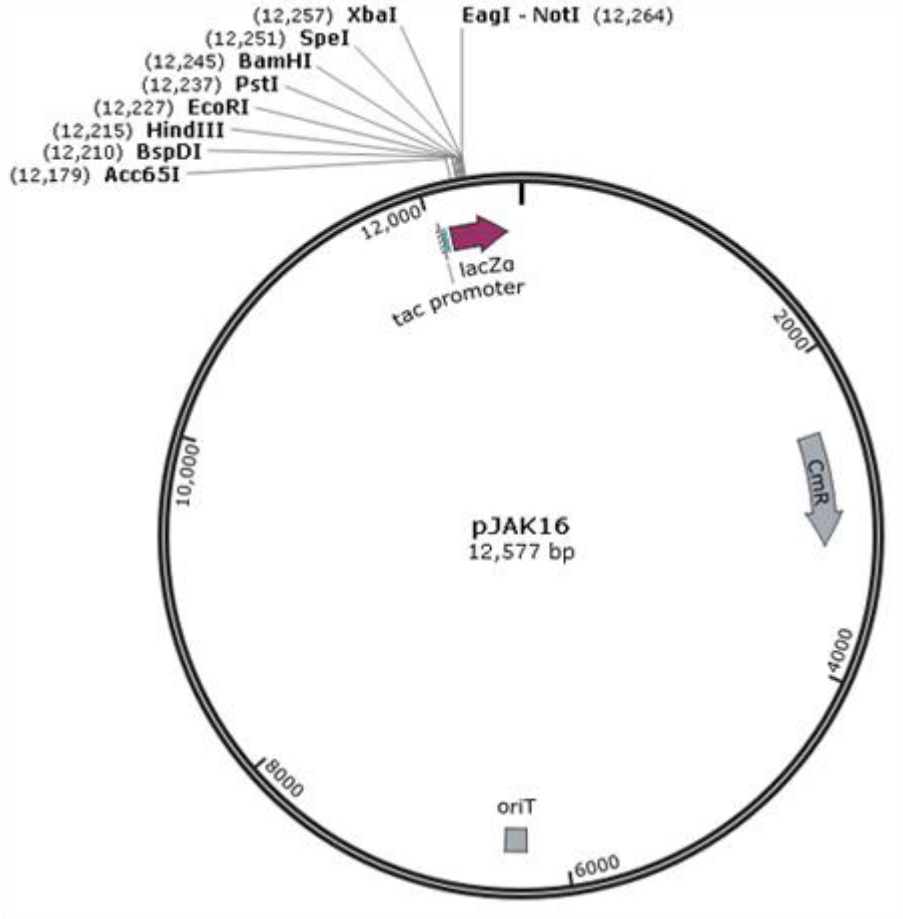


Figure 2.1: pJAK16 map.
Map of the pJAK16 cloning plasmid. This figure was created by importing the nucleotide sequence into SnapGene Viewer (snapgene.com) version 1.5.3.

2.3.10.2: Cloning of integron gene cassettes into pSU-pBAD

Plasmid pSU-pBAD (Figure 2.2) was amplified using primers EcoRI-pSU-pBAD-R/XbaI-pSU-pBAD-F and treated with two sets of restriction enzymes *Xba*I/*Eco*RI and *Eco*RI/*Nhe*I according to manufacturer instructions (New England Biolabs). VSD31 was amplified using primer pairs EcoRI-VSD31-F/NheI-VSD31-R (see Table 2.4), VSD78 with EcoRI-VSD78-F/NheI-VSD78-R (Table 2.4), VSD54 with EcoRI-VSD54-F/XbaI-VSD54-R (NheI restriction enzyme cuts in the VSD54 gene) (Table 2.4) and gel purified (section 2.3.6). Digested plasmid and gene cassettes were ligated using the T4 DNA ligase (New

England Biolabs). After cloning VSD31, VSD78 and VSD54 into pSU-pBAD they were transformed into competent *E. coli* WM3064 cells where they were then conjugated using bi-parental conjugation into d16-60, d72-92 and d60-50 respectively. These constructs were also used for conjugation into wt S24 as described in chapter 4. Sequencing of gene cassettes was then performed using primers pSU-pBAD-R and pBAD-Out-2 (Table 2.4).

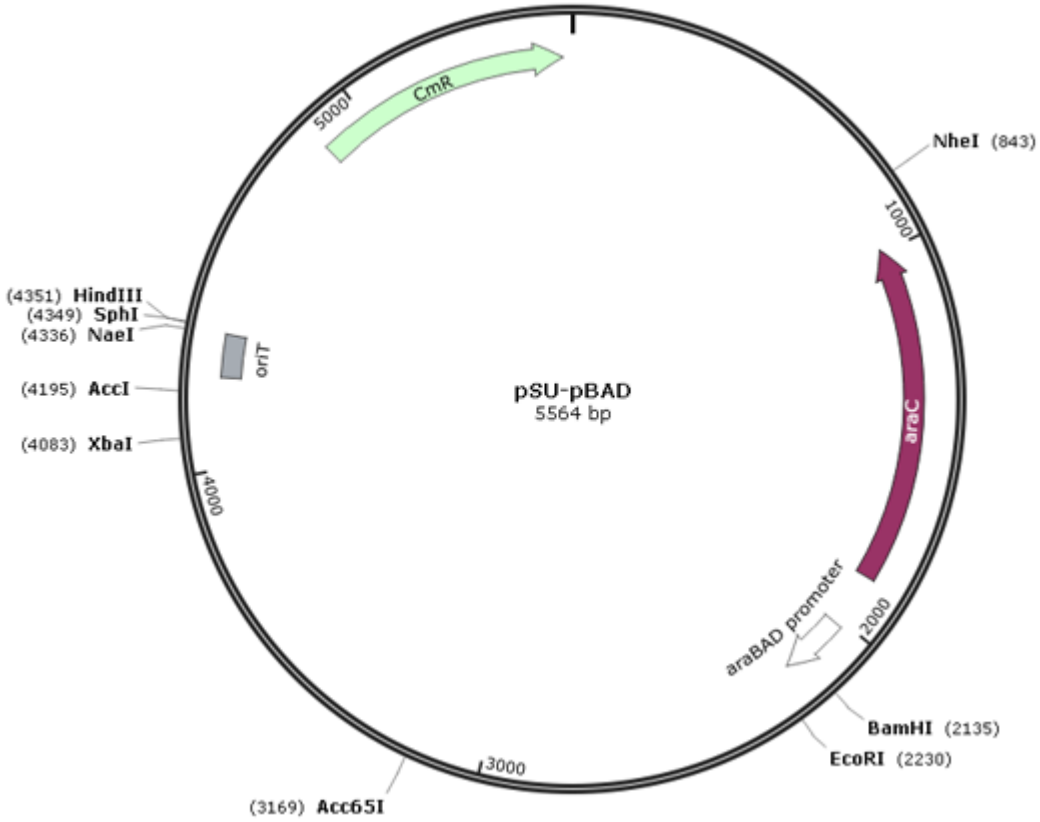


Figure 2.2: pSU-pBAD map.
Map of the pSU-pBAD cloning plasmid. This figure was created by importing the nucleotide sequence into SnapGene Viewer (snapgene.com) version 1.5.3.

2.3.10.3: Cloning of $recA_{S22}$ and gfp into $pOriVn_{700}$

Vector $pOriVn_{700}$ and *recA* from *V. cholerae* S22, a strain that is closely related to *V. cholerae* S24, were amplified using primer pairs EcoRI-Ori700-R/EcoRI-Ori6K-F and EcoRI-recA-F/EcoRI-recA-R respectively (Table 2.4). Since the primers contained engineered *Eco*RI sites, the resulting amplicons of $recA_{S22}$ and $pOriVn_{700}$ were purified, digested with *Eco*R1 and then ligated together using T4 DNA ligase (New England Biolabs). The ligation mix was then transformed into *E. coli* DH5αλpir to produce $pOriVn_{700}$-$recA_{S22}$. The construct was then extracted and transformed into the conjugation donor strain *E. coli* WM3064.

2.4: Construction of integron gene cassette deletion mutants in *V. rotiferianus* DAT722

The cassette array of DAT722 is fully sequenced and consists of 116 gene cassettes, although there are 94 different cassette types due to the presence of paralogous cassettes (Boucher, 2006;Chowdhury *et al.*, 2011). Construction of the deletion mutants (Table 2.1) was carried out as described previously (Labbate *et al.*, 2011). Briefly, pMAQ1081 containing a 1834 bp fragment inserted into the *sacB*-counter selectable suicide vector pCVD442 (Donnenberg and Kaper, 1991) was used to create deletions in the cassette array of *V. rotiferianus* DAT722. The fragment consisted of two sequences with homology to different paralogous cassettes across the array disrupted with a kanamycin resistance gene. Conjugation of this construct into *V. rotiferianus* DAT722 allowed for allele replacement and deletion of cassettes between these two sets of paralogous cassettes. Deletion mutants d16-60 and d50-60 were created by taking a merodiploid (designated MD7) consisting of pMAQ1081 recombined into cassette 61 and screening colonies counter selected on sucrose medium with primers targeting unique cassettes outside the expected deletions (Table 2.4). An identical approach was taken for creating d16-60a. An independently derived but identical merodiploid to MD7 (designated SC-8B61) was used. Deletion mutant d72-92 was isolated as a double-crossover and did not undergo sucrose counter selection. Deletion mutant d72-92a was created by taking a merodiploid (designated SC-8A91) consisting of pMAQ1081 recombined into cassette 93 and screening colonies counter selected on 10% sucrose medium as described above.

2.5: Protein methods

2.5.1: Preparation of bacterial cells for 2D-PAGE analysis

2.5.1.1: Preparation of bacterial cells grown in LB20

Cells were harvested from 100 mL of LB20 medium grown to mid-logarithmic phase (OD_{600} ~ 0.5) and stationary phase (OD_{600} ~ 2.5-3.0). Inoculation and growth of the culture was carried out as described in section 2.1. 100 mL of culture was harvested for cells in mid-logarithmic phase and 20 mL for cells grown to stationary phase. Cells were collected by centrifugation for 10 min at 4000 *x g* (Megafuge 2.0R; Heraeus instruments), the supernatant discarded, the cells washed once with 50 mL 2% NaCl and the cells collected by centrifugation using the above conditions. Cell pellets were frozen at -20ºC until required for processing. All samples were prepared in triplicate.

2.5.1.2: Preparation of bacterial cells grown in 2M + glucose

Cells were harvested from 100 mL of LB20 medium grown to mid-logarithmic phase (OD_{600} ~ 0.5) and stationary phase (OD_{600} ~ 2.0). Inoculation and growth of the culture was carried out as described in section 2.1. 100 mL of culture was harvested for cells in mid-logarithmic phase and 25 mL for cells grown to stationary phase. Cells were collected by centrifugation for 10 min at 4000 *x g* (Megafuge 2.0R Heraeus instruments) and the supernatant discarded. Initially, cells were washed once with 50 mL of 2% NaCl. However, it was determined that three 2% NaCl washes were necessary to remove a contaminating substance that prevented proper isoelectric focusing of proteins from the d16-60 mutant cells. Cell pellets were frozen at -20ºC until required for processing. All samples were prepared in triplicate.

2.5.2: Preparation of protein from bacterial cells for 2D-PAGE and secretome analysis

Frozen cell pellets prepared as described in 2.5.1 were resuspended in 1 mL of CUTTL solution (1% C7bz0, 2 M thiourea, 7 M urea, 40 mM Tris and 50 mM LiCl) and transferred into a 1.5 mL eppendorf tube. C7bz0 is a non-ionic detergent that improves protein solubility. Thiourea and urea are denaturing agents that break hydrogen bonds present in proteins. Tris is added to increase the pH to 8.3 for subsequent reduction and alkylation and LiCl is added to interfere with interactions between proteins and bacterial cell wall components thus releasing cell wall bound proteins.

Cells were disrupted by ultrasonic sonication using a VibraCell sonicator (Sonics and Materials). Specifically, this was done at tune 25 with each round of sonication running for two 20 s bursts with a cooling step (on ice) in between bursts, starting with an output of 40%. This was repeated a further six times, each time increasing output by 10% until 100% was reached. Sonication destroys the structural integrity of the bacterial cell, allowing whole cell proteins to be solubilised. Furthermore, sonication breaks up high molecular weight DNA that can disrupt the IEF (isoelectric focusing). After sonication, the samples were spun at 4838 *x g* to remove any unlysed cells and cellular debris and supernatant transferred into a clean 1.5 mL eppendorf tube. The proteins in the sample were then reduced and alkylated by adding tributylphosphine (TBP) and acrylamide to final concentrations of 5 mM and 20 mM respectively and incubating at room temperature for 1.5 hr. These reagents break all disulfide bonds formed between cysteine residues and modify the remaining free cysteine residues to prevent reformation of these disulfide bonds. After this incubation time, the alkylation reaction was quenched using 10 mM DTT.

2.5.2.1: Removal of DNA and cell wall material from whole cell protein samples

Contaminants such as high molecular weight DNA not disrupted by sonication, cell wall material and polysaccharides affect the IEF of proteins in 2D-PAGE and were removed using a 100 kDa cut off filter (Microcon Centrifugal Filter Devices). This is because the vast majority of proteins in bacterial samples are less than 100 kDa and will flow through the filter into the filtrate. Initially, 400 µL of sample was placed into the top chamber of the 100 kDa cut off filter and spun at 8000 *x g* for 30 min and the filtrate retained. However, it

was determined that this volume was an overload for the filter resulting in clogging of the filter and thus likely not allowing all the protein to pass through. Consequently, samples with diluted levels of protein were recovered and were of insufficient concentration for subsequent experiments. To address this, 250 µL of sample was filtered per filter with multiple cut off filters used to process enough sample.

2.5.2.2: Removal of salts and other small contaminants from whole cell protein samples

Filtered samples were run through Micro Bio-Spin 30 Column, 25 (catalog# 732-6221; Bio-Rad) to remove LiCl and other small contaminants such as salts. Micro Bio-Spin columns are size exclusion traps that retain small contaminants, allowing proteins to pass through unaltered. The use of the columns was carried out as described by the manufacturer. Briefly, three Micro Bio-Spin columns were used to process 225 µL of filtered sample (75 µL per column). The columns were placed into a 1.5 mL wash eppendorf tube and centrifuged at 1000 *x g* for 2 min to remove the buffer already present in the column. The filtrate was discarded from all three columns and 500 µL of 1% C7bz0, 2 M thiourea and 7 M urea was added. The column was then spun at 1000 *x g* for 1 min and the filtrate discarded. This was repeated twice more. 75 µL of filtered sample was added to each of the Micro Bio-Spin columns and spun at 1000 *x g* for 4 min. The samples were collected in fresh 1.5 mL eppendorf tubes. The columns were then discarded and the filtrate from each of the columns was pooled into one tube, producing a sample volume of approximately 250 µL.

2.5.2.3: Supernatant protein extraction and gel electrophoresis

Cells were grown in 2M + glucose minimal medium overnight at 28ºC. Cells were collected by centrifugation (4000 *x g*), the supernatant was collected and filtered through a 0.2 µm filter. To precipitate supernatant proteins, five volumes of acetone was added, mixed by gentle inversion and incubated at -20ºC for 30 min. Precipitated proteins were collected by centrifugation at 3000 *x g* for 3 mins and supernatant discarded. The protein pellet was dried at 37 ºC overnight, weighed and then resuspended to a concentration of 40 mg/mL in 2 M thiourea, 7 M urea and 1% C7bz0.

Prior to gel electrophoresis samples were run through a Micro-Biospin Column (2.5.2.2) according to the manufacturer's protocol to remove excess salt from the sample due to media. 400 µg of protein was loaded onto a 4-12% Bis/Tris precast polyacrylamide 1D gel (Bio-Rad) and run at 160 V for ~60 min. Gels were then fixed for 30 min in 10% acetic acid (v/v) and 40% methanol (v/v) prior to staining with Flamingo protein stain (Bio-Rad). The Ladder used was Bio-Rad Precision Plus Protein Unstained Standard (catalog # 161-0363).

2.5.3: Quantification of protein in samples

To ensure that differences in protein spot intensity in 2D-PAGE gels between the wt DAT722 and d16-60 gels was not due to differences in the amount of loaded protein, the protein concentration of each sample was determined by gel densitometry of loaded protein sample in a 1-dimension gel and comparing to a protein standard of known concentration.

2.5.3.1: Preparation of a protein standard for use in quantification

Since the dye used in staining proteins (Flamingo protein staining solution; Bio-Rad) can vary in the staining of specific proteins, it was decided that use of commercially available bovine serum albumin (BSA) as a known protein standard would not be accurate for comparing to wt DAT722 proteins. As a result, a protein standard derived from wt DAT722 extracted proteins was created to ensure the most accurate determination of the protein concentration in protein extracted samples.

1 L of mid-logarithmic phase wt DAT722 cells grown in LB20 medium at 28°C was harvested by centrifugation (10 min at 4000 *x g*). The supernatant was discarded, and the cells washed by re-suspension in 2% NaCl solution. Cells were again collected by centrifugation using the above conditions, the supernatant discarded and the cells resuspended in 17 mL CUTTL solution (1% C7bz0, 2 M thiourea, 7 M urea, 40 mM Tris and 50 mM LiCl). The resuspended pellet was then disrupted by ultrasonic sonication (100% output, tune 25) at 20 s intervals until the solution was clear and no longer viscous. Between bursts, the solution was allowed to cool on ice. Cell extracts were then spun down

at 4000 *x g* for 10 min to remove cellular debris and unbroken cells (Megafuge 2.0R; Heraeus instruments).

Proteins in the cell extracts were reduced and alkylated using tributylphosphine (TBP) and acrylamide as in section 2.5.2. Proteins were precipitated by adding 5 volumes of acetone, mixing by gentle inversion and incubating at room temperature for 30 min. Precipitated protein was then collected by centrifugation (3000 *x g* for 5 min; Megafuge 2.0R; Heraeus instrument) and the supernatant carefully discarded. Excess acetone was removed by placing tubes containing protein pellets upside down on absorbent paper. The pellet was then air dried overnight at 37ºC to remove remaining acetone by evaporation. The dried protein was weighed and resuspended in 7 M urea, 2 M thiourea and 1% C7bz0 (UTC7) to a final concentration of 20 mg/mL. Unused dried protein pellet was stored at -80ºC and the 20 mg/mL protein standard solution was aliquoted into 1.5 mL eppendorf tubes in 100 µL volumes and stored at -20ºC until required.

2.5.3.2: Quantification of total protein by densitometry

Quantification of protein concentration in all whole cell protein samples was conducted to determine the appropriate volume required to load onto IPG strips (pH gradient strip used for isoelectric focusing; section 2.5.4.1) for 2D-PAGE analysis (section 2.5.4.2).

A standard serial 2-fold dilution assay using 2x protein loading buffer (0.125 M Tris pH 8.8, 30% (v/v) glycerol, 0.4% (w/v) SDS, 0.012 mg/mL (w/v) bromophenol blue in ddH_2O) was carried out using the protein standard described in section 2.5.3.1 and the samples to be analysed by 2D-PAGE.

Once the protein standard and sample dilutions were prepared, 10 µL of each dilution was loaded into the individual wells of a Criterion XT 1D 26-well precast gel (4-12% Bis-Tris; Bio-Rad). 500 mL of 1x MES running buffer was prepared per gel by diluting a 50x stock in ddH_2O (BioRad) and poured into the top and bottom chambers of the criterion gel tank. Gels were run at 160 V until the protein from each well had entered the top of the gel (generally 3-4 min). The gel was removed and then fixed and stained with Flamingo protein staining solution for visualisation and densitometry.

2.5.3.3: Fixing, staining and imaging of protein gels

Gels were placed in 200 mL of fixing solution (10% acetic acid, 40% methanol, 50% ddH_2O) for 30 min. Fixing solution precipitates and fixes proteins into the gel ensuring they remain appropriately fixed in their position on the criterion gel. After 30 min the fixing solution was replaced with 100 mL of 1x Flamingo protein staining solution (Bio-Rad; diluted from a 10x stock with ddH_2O) and incubated for 1 hr. To prevent photo-bleaching of protein gels, gels were covered in aluminum foil during and after staining. Gels were scanned using the Pharos FX Plus Molecular Imager using software program Quantity One (Bio-Rad) at high intensity with a resolution of 50 μm. Because the total protein was run for only a short time in the gel (3-4 min), separation of individual proteins is not achieved and proteins appear as a single large band that can be subjected to densitometry for quantification (Figure 2.3). Some 1-Dimensional gels were stained for the presence of glycoproteins, and this was performed using Pro-Q Emerald 300 Glycoprotein stain kit (Molecular Probes) according to manufacturer's instructions.

2.5.3.4: Quantification of samples by densitometry

Using Quantity One software (Bio-Rad) the density of each band representative of the separate dilutions was analysed and a standard curve calculated using the known amounts of protein standard loaded into each well. A representative image of one such quantification is shown in Figure 2.3. The sample dilutions were interpolated on the standard curve and the protein concentration in each sample was determined in μg/μL.

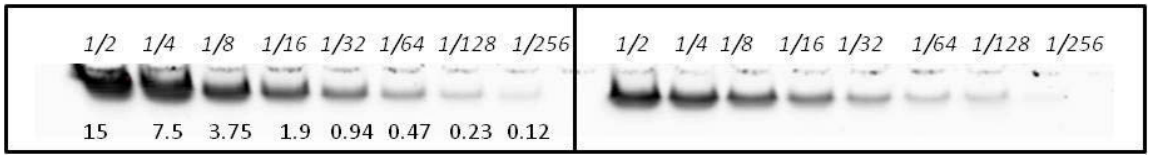


Figure 2.3: Quantification of protein samples by densitometry.
This image was captured using Pharos FX Molecular Imager (Bio-Rad). The left most and right most wells show the serial 2-fold dilution assay of the protein standard and whole cell protein extract of a representative wt DAT722 wild-type sample respectively. The dilution of the loaded sample is given above the well in both the protein standard and representative sample. The actual total protein loaded for the protein standard is given below in µg. The Quantity One software program (Bio-Rad) was used to construct a standard curve from the density of each band represented in the protein standard and used to calculate the concentration of the protein in the sample.

2.5.4: 2D-PAGE

2.5.4.1: Isoelectric focusing – the 1st dimension of separation

Isoelectric focusing (IEF) is the first dimension of protein separation. IEF separates proteins on the basis of the isoelectric point (pI), the pH at which the protein carries a net charge of zero.

2.5.4.1.1: Rehydration of the IPG strip

Quantified samples were centrifuged at 16,000 *x g* for 10 min (Quantum Scientific 1.15; Sigma centrifuge) to remove any insoluble material that may have still been present in the sample. The conductivity of the sample was determined using the Horiba Conductivity Meter B-173 to identify any salt contamination. Conductivity measurements higher than 100 µS/cm were rejected and re-run through Micro Bio-Spin Chromatography Columns (section 2.5.2.2).

After conductivity was measured, 2 µL of orange ProteomIQ IEF Tracking Dye was added to the sample to allow for easy identification of the direction of current flow during IEF. The volume of sample totalling 400 µg of protein was determined from the quantification protocol (section 2.5.3.2) and used to rehydrate a pH 4-7 Bio-Rad 11 cm IPG strip. It should be noted that alkaline proteins are particularly difficult to resolve using 2D-PAGE (Chevalier, 2010) and since more than 70% of proteins produced by *Vibrio* spp. fall into the

pI range of pH 4-7, IPG strips with a pH gradient of pH 4-7 were selected (Schwartz *et al.*, 2001).

Two 'in-gel' methods of rehydration were used in this study: active and passive. Active rehydration involves the partial rehydration of the IPG strip with a buffer; in this case UTC7, before the sample is loaded onto the strip with an applied voltage. This was done by pipetting 100 µL of buffer along the length of an IEF tray (Proteome Systems) and placing the IPG strip face down onto the buffer and incubating at room temperature for 30 min. After rehydration the IEF apparatus was set up as in section 2.5.4.1.2 and 150 µL of sample made up to the remaining volume required using the quantification protocol with UTC7. This combination of UTC7 and sample was pipetted underneath the IPG strip set up in the IEF tray in section 2.5.4.1.2.

Passive rehydration involves pipetting the sample evenly along the length of an IEF tray, adding the IPG strip gel side down onto the sample and allowing the protein sample to rehydrate the strip for approximately 4-6 hr at room temperature. Passive rehydration was used for rehydration of IPG strips for the 2D-PAGE experiments performed for the completion of this thesis.

2.5.4.1.2 IEF–apparatus set-up and loading the strip

IPG wicks (small pieces of blotting paper used to overlap the ends of the IPG strip during IEF) were saturated with ddH_2O and carefully blotted on paper towel to remove excess water. The IPG wicks were placed into a Bio-Rad 11 cm IEF tray, ensuring that the wicks covered the electrode wire in the bottom of the tray. The rehydrated IPG strip was then placed into the tray gel side down making sure there was contact between the IPG wick and IPG strip. The IEF tray was then covered with paraffin oil to ensure the IPG strip did not dry out. The lid was placed on the tray and the tray was then placed into the Protean IEF cell (Bio-Rad). The instrument was programmed to run the following three step program: Step 1: slow ramp to 3,000 V for 4 hr, Step 2: linear ramp to 10,000 V for 4 hr and Step 3: 10,000 V rapid ramp for 24 hr. The run was limited to 50 µA per IPG strip for the duration of the run. This program is used to limit the voltage at the beginning of the run when the current will be high. In this way the overall power (subsequently heat also) can be

controlled. It is important to reach the maximum voltage during the final step, to ensure the best possible resolution when run in the second dimension. Figure 2.4 shows the concept of protein separation in the first and second dimensions. The run was complete when the volt/hours reached 100kVh (usually after about 16 hr).

2.5.4.2: Polyacrylamide gel electrophoresis–the 2nd dimension of separation

Separation of proteins in the second dimension is based on their molecular size. Proteins are coated and linearised by SDS and since SDS is negatively charged the proteins will migrate toward the cathode separating in the polyacrylamide gel according to their size.

2.5.4.2.1: IPG strip equilibration

Once isoelectric focusing was complete the IPG strips were carefully removed with excess oil being drained away using absorbent paper. The IPG strips were then laid gel side up in a clean IEF tray. Each IPG strip was covered in 3 mL of equilibration solution (2% (w/v) SDS, 6 M urea, 250 mM Tris-HCl pH 8.5, 0.01% (w/v) bromophenol blue) for 20 min at room temperature. This equilibration solution coats all proteins with SDS, maintaining solubility and coating all proteins with a uniform negative charge. Additionally, the solution adjusts the pH of the IPG strip to the appropriate pH for the second dimension. IPG strips were incubated in equilibration solution for 20 min.

2.5.4.1.2: Polyacrylamide gel electrophoresis–apparatus set-up and loading the strip

Equilibrated IPG strips were carefully loaded on top of a Criterion XT Precast gel (Bio-Rad 4-12% Bis-Tris IPG + 1 well) for running the second dimension. Specifically, the criterion gels were placed into the criterion running tank, the integrated buffer chamber was filled with 1x MES running buffer (diluted from a 50x stock in ddH_2O) first, followed by filling of the outer chamber until the buffer reached the bottom of the integrated upper chamber. The IPG strip was carefully loaded on top of the criterion gel, with the gel side of the IPG strip facing inwards. Using a thin bladed spatula the IPG strip was gently pushed against the top of the gel, ensuring it was level and in full contact with the criterion gel. 3 µL of Bio-Rad Precision Plus Protein Unstained Standard (catalog # 161-0363) was added to the single well in the criterion gel for sizing the proteins. The lid of the criterion tank was then

attached to the power supply and placed on top of the tank. The criterion gels were then run at 160 V until the bromophenol blue dye front reached the bottom of the gel. This ensured the full length of the gel gradient was used.

Upon completion of the run, the IPG strip was discarded and the gel fixed and stained as described in 2.5.3. The gels were scanned into software program Quantity One (Bio-Rad version 4.6.1) using the Pharos FX Plus Molecular Imager (Bio-Rad) at high intensity with a resolution of 50 μm and saved as high resolution JPEG files for later analysis (section 2.5.5).

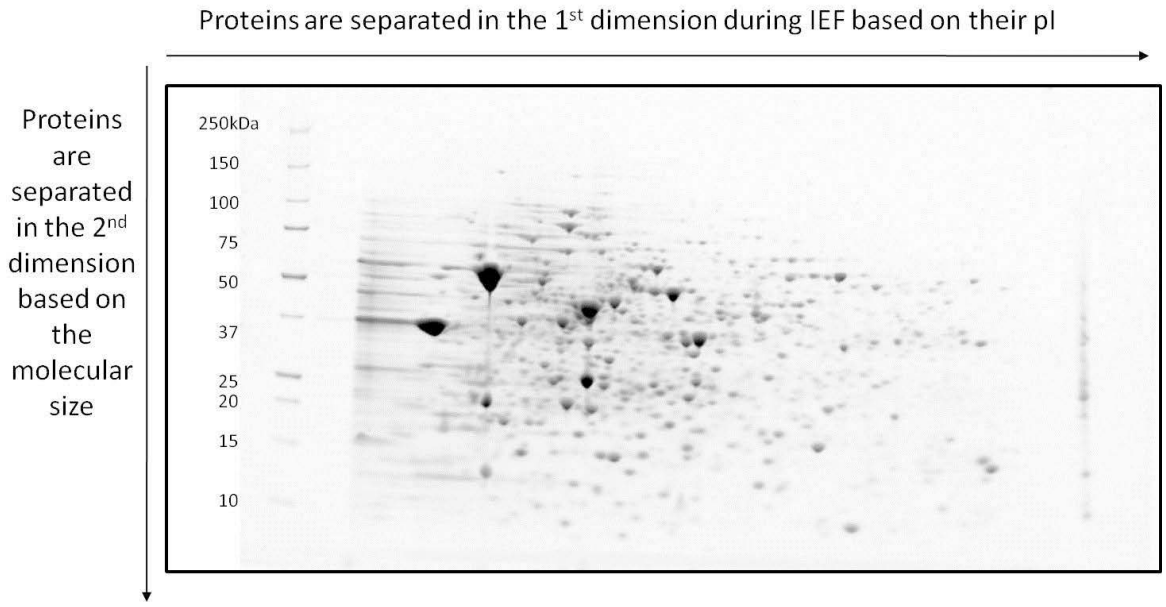


Figure 2.4: Representative image of a gel derived from 2D-PAGE of whole cell proteins from *Vibrio rotiferianus* DAT722.
Proteins are separated in the first dimension along the horizontal axis during IEF based on their pI (isoelectric point-pH at which their charge is neutral). On the vertical axis, proteins are separated in the second dimension during polyacrylamide gel electrophoresis based on their molecular size. Size molecular protein standard is shown on the left of the gel. Sizes of each protein in the protein standard are given on the image.

2.5.5: Identification of differentially expressed or shifted proteins between wt DAT722 and the d16-60 mutant

2.5.5.1: PDQuest

Images of triplicate 2D-PAGE gels of the wt DAT722 and d16-60 mutant (separate biological replicates described in section 2.5.1) were differentially compared for any changes in expression or shifts in proteins (indicative of post-translational modification) using software program PDQuest (version 8.0; Bio-Rad). PDQuest identifies at least 2-fold changes in protein spot intensity and/or spot migration in triplicate gels using an intersection of both a quantitative and statistical analysis (student t-test with 95% confidence interval). Spots that were identified to be both quantitatively expressed and statistically significant were cut out of the gels and their protein sequence identified using LC-MS/MS (Liquid Chromatography Tandem Mass Spectrometry).

2.5.5.2: Excision and trypsin digestion of protein spots

Proteins of interest were excised and subject to LC-MS/MS for identification. Since the Flamingo protein stain is a fluorescent dye and can only be detected at 300 nm, all gels were post-stained with Coomassie Blue (10% (w/v) ammonium sulfate, 20.6% methanol, 1.9% (w/v) phosphoric acid, 0.077% (w/v) coomassie blue G250 in ddH_2O) so proteins could be visualised and excised. After staining with Coomassie Blue gels were placed in destain (1% acetic acid made up with ddH_2O) and stored in 1% (v/v) acetic acid until required for protein spot excision.

Excised protein spots were destained by washing twice with 50% acetonitrile/50 mM NH_4HCO_3 pH 9 for 10 min with vortexing. The gel spots were dehydrated with 100% acetonitrile for 10 min and rehydrated in 20 µL of NH_4HCO_3 pH 9.0 containing 12.5 ng/µL of trypsin (Promega). After incubating at 4°C for 30 min, 25 µL of NH_4HCO_3 pH 9 was added and the samples incubated overnight at 37°C. The digest solution was transferred to a new tube and 30 µL of 50% acetonitrile, 2% formic acid added and incubated for 10 min in a sonicating water bath (PowerSonic420) at full power. The solution was removed and pooled with the digest solution and lyophilised in an Eppendorf concentrator 5201 at 30ºC to 15 µL and transferred to an autosampler vial. This 15 µL of concentrated protein sample was given to Dr. Matthew Padula (UTS Proteomics Technology Centre of Expertise) for LC-MS/MS analysis.

2.5.5.3. Liquid chromatography tandem mass spectrometry (LC-MS/MS)

LC-MS/MS was used to identify proteins of interest. Using an Eksigent AS-1 autosampler connected to a Tempo nanoLC system (Eksigent, USA), 10 µL of the sample was loaded at 20 µl/min with MS buffer A (2% Acetonitrile + 0.2% Formic Acid) onto a C8 trap column (Michrom, USA). After washing the trap for 3 min, the peptides were washed off the trap at 300 nL/min onto a IntegraFrit column (75 µm x 100 mm) packed with ProteoPep II C18 resin (New Objective, Woburn, MA). Peptides were eluted from the column and into the source of a QSTAR Elite hybrid quadrupole-time-of-flight mass spectrometer (Applied Biosystems/MDS Sciex) using the following program: 5-50% MS buffer B (98% Acetonitrile + 0.2% Formic Acid) over 15 min, 50-80% MS buffer B over 5 mins, 80% MS

buffer B for 2 min, 80-5% for 3 min. The eluting peptides were ionised with a 75 µm ID emitter tip that tapered to 15 µm (New Objective) at 2300 V. An Intelligent Data Acquisition (IDA) experiment was performed, with a mass range of 375-1500 Da continuously scanned for peptides of charge state 2+-5+ with an intensity of more than 30 counts/s. Selected peptides were fragmented and the product ion fragment masses measured over a mass range of 100-1500 Da. The mass of the precursor peptide was then excluded for 15s.

2.5.5.4. Identification of protein spots by comparing generated peptides to predicted proteins from the annotated genome sequence of Vibrio rotiferianus DAT722

The MS/MS data files produced by the QSTAR were searched using PEAKS Studio (Bioinformatics Solutions Inc.). PEAKS was used to directly search peptides matching the *Vibrio rotiferianus* DAT722 protein sequence FASTA output derived from the recent completion of the *Vibrio rotiferianus* DAT722 genome (Chowdhury *et al.*, 2011). The highest PEAKS score (percentage based on a p-value < 0.05) was taken as the closest peptide match.

The *Vibrio rotiferianus* DAT722 genome was sequenced at the Ramaciotti Centre (University of New South Wales, Sydney) using 454 sequencing technology (Chowdhury *et al.*, 2011). The genome was then run through the online program RAST (http://rast.nmpdr.org/) which is a fully-automated service for annotating bacterial and archaeal genomes (Aziz *et al.*, 2008). The FASTA protein sequence output was generated from RAST and used for identification of protein spots.

To predict where proteins may be localised within the cell (e.g. cytosolic) the amino acid sequences of relevant proteins was run through the online program PSORTb (version 3.0; http://www.psort.org/psortb/).

2.6: Polysaccharide methods

2.6.1: Extraction and purification of loosely attached surface polysaccharide

As described in chapter 1, extracellular (EPS) and capsular (CPS) surface polysaccharide structures are loosely attached to the cell surface. To extract these moieties a method by Enos-Berlage and McCarter was adapted (Enos-Berlage and McCarter, 2000). Briefly, a lawn of cells were grown on an agar plate of 2M + glucose for 20 hr and scraped from the plate using pipetting and resuspended in 5 mL of Phophate-Buffered Saline (PBS). Cells were then vortexed for 1 min, shaken at 200 rpm on a rotary shaker for 1.5 hr at 28°C and the vortex and shaking steps repeated. Centrifugation was performed at 10,000 *x g* for 15 min and the supernatant collected. RNase A (Sigma-Aldrich), DNase I (Sigma-Aldrich) and $MgCl_2$ (Sigma-Aldrich) were added to final concentrations of 50 µg/mL, 50 µg/mL and 10 mM, respectively, to remove contaminating RNA and DNA. This mixture was then incubated at 37°C for 8 hr in order for the RNA and DNA to be digested. After this step Proteinase K (Astral Scientific) was added to a final concentration of 200 µg/mL and incubated at 37°C for 17 hr. This step was performed to remove any contaminating protein. Samples were then extracted twice with equal volumes of phenol-chloroform (Amresco) and then EPS/CPS was precipitated by adding 2.5 volumes of molecular grade ethanol (Sigma-Aldrich). Samples were spun at 16,000 *x g* for 10 min to collect EPS/CPS and pellet washed with 70% ethanol (made up using ddH_2O). EPS/CPS pellets were then made up to a 1 mL volume with molecular biology water, stored overnight at -20°C and lypholised using a Dynavac freeze drier.

2.6.2: Extraction of whole cell polysaccharide

Whole cell polysaccharide from mutants grown in 2M + 0.2% glucose overnight was extracted using a proteinase K, phenol-water method adapted from (Apicella, 2008). For the ^1H NMR scans of extracts made from the integron deletion mutants in chapter 3, benzonase was used to remove contaminating DNA and RNA (Figure 3.13). For the ^1H NMR scans of extracts shown in Figure 3.14, DNase I and RNase A was used to remove contaminating DNA and RNA.

2.6.3: Gel electrophoresis of whole cell and loosely attached polysaccharides

After polysaccharide moieties had been lypholysed approximately 200 µg of polysaccharide was solubilised with molecular biology water and mixed with an equal volume of 0.1 M Tris-HCl buffer (pH 6.8; Amresco), 2% SDS (w/v), 20% sucrose (w/v), 1% 2-mercaptoethanol (Sigma-Aldrich) and 0.001% bromophenol blue. Samples were then heated for 5 min at 100°C, loaded onto a 7.5% Criterion TGX precast polyacrylamide gel (BioRad) and run at 160V for approximately 30 min. Gels were run in a BioRad gel tank using 1x Tris/Glycine buffer: 0.12 M tris, 0.96 M glycine and 0.5% SDS. Gels were then silver stained using the method described in section 2.6.4.

2.6.4: Silver staining of polysaccharide 1D gels

The silver staining method for polysaccharide polyacrylamide gels was taken from Kittelberger, 1993 (Kittelberger and Hilbink, 1993). The steps of this procedure in outlined in the Table 2.5 below.

Table 2.5: Silver staining protocol

Procedure	Reagent	Duration
Fixation/Oxidation		
Fixation	30% ethanol. 10% acetic acid	Overnight
Oxidation	0.7% periodic acid in fixative	10 min
Wash x 3	Water wash	30 min
Silver Staining		
Silver	0.1% silver nitrate	30 min
Wash x 1	Water wash	10 s
Develop	3% sodium carbonate, 0.02% formaldehyde	20 min
Stop	1% acetic acid	5 min
Wash x 3	Water wash	10 min
Reduction	Farmers reducer (0.3% sodium thiosulfate, 0.15% potassium ferricyanide, 0.05% sodium carbonate)	10-30 s
Wash x 3	Water wash	10 min

2.6.5: Congo red binding assays

2.6.5.1: Congo red colony morphology

Overnight cultures of bacteria were grown as stated in section 2.1 and diluted to 10^{-7} in the media they were grown in (either LB20 or 2M + glucose). 100 µL of this dilution was spread plated onto LB20 + 0.001% (w/v) congo red (Sigma-Aldrich) and 2M + glucose + 0.001% (w/v) congo red and incubated for 7 days at 28°C and imaged as described in section 2.7.3.

2.6.5.2: Congo red liquid binding assays

Congo red colony morphology assays described above were qualitative and the results were sometimes subjective. For a more definitive measure of congo red binding, a method was adapted from Colvin *et al.* (Colvin *et al.*, 2011) where the ability of cells to bind congo red in a liquid medium was examined. Briefly, overnight cultures were grown in the presence of 40 µg/mL congo red. Cultures were then centrifuged at 4,000 *x g* to pellet the cells, and the optical density at 495_{nm} of the supernatant (containing unbound congo red) was recorded. The lower the optical density reading the more the cells were able to bind congo red.

2.6.6: Nuclear magnetic resonance of whole cell polysaccharide

Equal amounts of purified whole cell polysaccharide (section 2.6.2) were exchanged three times using D_2O (deuerated water; Sigma-Aldrich) as a solvent. 1D water suppression ^1H NMR experiments were performed on purified whole cell polysaccharide resuspended in 600 µL of D_2O using an Agilent Technologies 500 MHz NMR instrument at 28ºC with the internal reference of the sodium salt of 3-(trimethylsilyl)-3,3,2,2-tetradeuteropropionic acid. The typical acquisition parameters utilized were spectral width 8012Hz, acquisition time 4.089 s, relaxation delay 1.5 s and line boarding frequency 0.5 Hz.

2.7: Microscopy

2.7.1. Preparation of cells for fluorescence microscopy

V. rotiferianus DAT722 overnight cultures were grown in 2M + 0.2% glucose. For detection of polysaccharides, 100 µL of 1 mg/mL concanvalin A conjugated to the FITC fluorophore made up with ddH_2O (Sigma-Aldrich) was added to 1 mL of an overnight culture. To examine these cells using microscopy a 1.5 × 1.6 cm rectangular well was created on the surface of a glass slide using a 65 µL Gene Frame (ABgene; slides were obtained from Livingstone). Agarose (type 1; Sigma-Aldrich) was dissolved at 2% (w/v) in ddH_2O and 65 µL of this solution was applied to the centre of the well. A coverslip was immediately placed on top and the agarose was allowed to solidify within the well for ~20 min, producing a flat 'agarose pad'. To mount cells for fluorescence microscopy, the coverslip was gently lifted away and 3 µL of the concanvalin A stained live cell culture was placed on the surface of the agarose pad. The coverslip was then reapplied and the cells were viewed immediately.

2.7.2. Phase-contrast and fluorescence microscopy

Live mounted cells were observed using a Zeiss Axioplan 2 fluorescence microscope equipped with a 100X Plan ApoChromat phase-contrast objective (numerical aperture 1.4; Zeiss) and an AxioCam MRm cooled CCD camera (Zeiss). For fluorescence excitation, light from a 100-watt high pressure mercury lamp was passed through the Filter Set 02 (Zeiss; 365 nm excitation filter, 420 nm LP barrier filter). Images were collected using AxioVision software, version 4.5 (Zeiss).

2.7.3: Microscopy of congo red stained colonies

After cells were grown on an agar plate for 7 days in the presence of congo red they were imaged using an Olympus SZX12 light microscope at 100X using a DP70 colour camera. Acquisition of images was performed using the DP manager and controller software (Olympus; version 3.1.1.267).

2.7.4: Inverted microscopy of crystal violet stained biofilms

Batch biofilm assays described in section 2.9 were imaged using an inverted Nikon DS-Fi2 fluorescence microscope under brightfield using a 40X objective. Images were acquired using the DS-L3 software (Nikon).

2.8: Flow cytometry

2.8.1: Calcofluor staining of surface polysaccharide

A 1 in 10 dilution (in 2% NaCl) of an overnight culture of wt *V. rotiferianus* DAT722 grown in 2M + glucose was stained with 0.05% calcofluor (Sigma-Aldrich) made up in water for 20 min at 28°C. Cells, washed and unwashed, were run through the flow cytometer. For cells that remained unwashed no further steps were required. For washing, cells were pelleted at 16,000 *x g* for 5 min and washed twice with 2% NaCl before being subjected to flow cytometry.

2.8.2: Concanvalin A staining of surface polysaccharide

Cells to be stained with concanvalin A (conjugated to the FITC fluorophore that emits at 520 nm) were grown as stated in 2.8.1 but stained with 100 µg/ml concanvalin A made up in 0.1 M sodium bicarbonate (pH 8.3) for 15 min at 28°C. Cells were then washed as stated in 2.8.1.

2.8.3: Performing flow cytometry

Following staining of polysaccharide on the surface of cells, as stated in section 2.8.1 and 2.8.2, cells were transferred to a BD Falcon round bottom test tube designed for flow cytometry (Becton, Dickinson and Company) and subjected to flow cytometry using the LSRII Cytometer BD and acquisition software BD FACSDiva (version 6.1.3; Becton, Dickinson and Company). Analysis of histograms was performed in Flowjo (version 7.2.5). The voltage of forward scatter (FSC), side scatter (SSC) and the DAPI and FITC filters are provided in Table 2.6.

Table 2.6: Flow cytometry parameters

Parameter	Voltage
Forward scatter (FSC)	758
Side scatter (SSC)	385
DAPI (461 nm used for calcofluor stained cells)	500
FITC (520 nm used for concanvalin A stained cells)	693

2.9: Biofilm assays

For each strain to be tested, overnight cultures were prepared in relevant media, diluted 1:100 and 500 µl added to the wells of a flat bottom, hydrophilic (tissue culture treated) and hydrophobic (non-tissue culture treated) plastic 24-well microtitre plate (Nunc ThermoFisher) in quadruplet. Plates were incubated for 24 hr at 28°C for *V. rotiferianus* DAT722 strains or 37°C for *V. cholerae* S24 strains with shaking at 200 rpm. Quadruplet control wells were also set up containing media only. Unattached planktonic cells were then carefully removed by pipetting and wells were washed twice with 2% NaCl for *V. rotiferianus* DAT722 strains or PBS for *V. cholerae* S24 strains. 500 µl of crystal violet (0.2% w/v; Sigma-Aldrich) was then added to wells and incubated for 15 min to stain the adhered cells. The crystal violet was removed and the wells washed three times with 2% NaCl/PBS to remove excess stain. To measure the biomass of adhered cells, the crystal violet was solubilised with 30% (v/v) acetic acid, transferred to a clean microtitre plate and optical density (600 nm) measured using a Synergy HT plate reader (Bio-Tek). This experiment was repeated on three separate occasions.

2.10: Bioinformatic analysis

Alignment of nucleotide and protein sequences was performed using the free online program ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The identification of putative functions of gene and superfamily domains was performed by the inputting both protein and nucleotide sequences into BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Both the *V. rotiferianus* DAT722 and *V. cholerae* S24 genome sequences were run through RAST (**R**apid **A**nnotation using **S**ubsystem **T**echnology), which is a fully automated online annotation tool (http://rast.nmpdr.org/) (Aziz *et al.*, 2008). To upload the genomic island sequence described in chapter 5 to the Genbank server, the Sequin software (NCBI) was utilised. All sequences of PCR products and the RME assemblage from the three separate contigs were analysed using the Geneious (version 6.1.4) software.

2.10.1: Phylogenetic tree construction

Phylogenetic analysis of $recA_{S24}$ and $recA_{RME}$ was done using bioinformatics program Geneious version 6.1.6 and FigTree version 1.4.0. Phylogenetic tree parameters were taken from (Thompson *et al.*, 2004a). Distance estimations were obtained using the Jukes and Cantor model and tree built using the Neighbor-Joining method. Bootstrap percentages were calculated after 100 simulations. The *Campylobacter jejuni* subsp. *jejuni* NCTC 11168 *recA* sequence was used as an outgroup.

2.11: Ultraviolet-light irradiation assays

UV stress experiments were adapted from (Lin and Wang, 2001). Strains were grown for 16-20 hr at 37°C with shaking at 230 rpm in 5 ml LB broth supplemented with appropriate antibiotic. Cells were centrifuged at 4,000 *x g*, corrected for differences in optical density at 600nm and resuspended in an equal volume of 1x M9 salts (Sambrook *et al.*, 1989) supplemented with $MgSO_4.7H_2O$ to a final concentration of 0.002 M. The entire cell suspension was placed in a clear bottom 10 cm plastic petri dish and subjected to 0.8 mJ/cm^2 UV-C for 0, 10, 20, 40 and 60 s using an Amersham Life Science Ultraviolet Crosslinker. After each time interval a 150 µL aliquot was removed and placed in a 1.5 mL eppendorf tube in the dark. The remaining liquid culture was thoroughly mixed using a pipette to avoid clumping of cells, after it was noted that enumeration for cell counts were inconsistent without this thorough mixing. After the final UV-C exposure time point, cells were diluted in M9 salts + $MgSO_4.7H_2O$ to 10^{-6} and enumerated by the drop plate method on LB5 agar. Plates were incubated in the dark to prevent photo-reactivation at 37°C overnight and colony forming units were calculated the following day.

2.12: Environmental stress assays

V. rotiferianus DAT722 and deletion mutants were subjected to environmental stress using the following conditions. All experiments were carried out in triplicate with data given in chapter 3 representative of the triplicate data.

2.12.1: Oxidative stress assays

1 mL of an overnight culture grown in LB20 was washed with 0.55x NSS and diluted 1:10 in 0.55x NSS. The diluted culture was exposed to 0.5 mM hydrogen peroxide (Sigma-Aldrich) with samples taken for enumeration at 0, 30 and 60 min post addition of H_2O_2.

2.12.2: Iron depletion stress

100 µL of an overnight LB20 culture was inoculated into 2M + glucose containing 0.1 mM of the iron chelating agent 2'2',-dipyridyl (DP) and incubated at 28°C with shaking for 12 days. Samples were taken daily for enumeration of viable cells.

2.12.3: Cold shock assays

Cells were grown to mid-logarithmic phase in 2M + glucose (OD_{600} ~ 0.3) and then placed at 4°C. Samples were taken daily for enumeration of viable cells.

2.13: Antibiotic assays

2.13.1: Minimum inhibitory concentration (MIC) assays

MICs of nalidixic acid, ciprofloxacin and bleomycin (Sigma-Aldrich) were determined by broth micro-dilution using standard methods (Clinical and Laboratory Standards Institute, 2003) except that LB5 broth was used as the growth medium instead of Mueller Hinton medium. Each MIC was performed in triplicate.

2.13.2: Mutation frequency assays

The mutation frequency experiment was designed using the guidelines described in (Pope *et al.*, 2008). Specifically, mutation frequencies were determined using LB5 supplemented with 50 µg/mL nalidixic acid and 100 µg/mL rifampicin. Ten replicate overnight cultures for each strain were grown in 5 mL LB5 (chloramphenicol was added for those strains carrying pCC2FOS and derivatives). Each overnight culture was then diluted to ~10^4 CFU/mL with fresh LB5 (no chloramphenicol added) and 5 mL for each replicate was transferred into a 15 mL tube and incubated for 16-20 hr at 37°C with shaking at 230 rpm. The following day, 200 µL from each tube was spread plated onto LB5 agar supplemented with the appropriate antibiotic (rifampicin or nalidixic acid) and incubated for 48 hr at 37°C, counting the number of visible colonies after 24 and 48 hr. This was repeated in triplicate. In order to calculate total colony counts, cells were enumerated on LB5 agar with no antibiotic. Note that these experiments were performed in a Class II Biosafety Hood to avoid contamination. Mutation frequencies were calculated as number of antibiotic resistant CFUs/total number of CFUs after 24 and 48 hr.

2.14: *recA* targeting experiments

Conjugations using pOriVn$_{700}$-*rec*A$_{S22}$ and pOriVn$_{700}$-P$_{lac}$*gfp* were performed by combining equal volumes of overnight cultures in LB of the donor and recipient strains. These mixtures were then centrifuged at 3,000 *x g* and cells resuspended in 50 µL of LB and spotted onto a 0.2 µm filter (Millipore) that had been placed on an LB5 agar plate containing 0.3 mM DAP. Donor and recipient cells were left to incubate for 4 hr at 37 °C and cells were then removed from the filter by vortexing. The resuspended cells were then

plated on LB5 + 125 μg/ml spectinomycin and incubated at 37 °C overnight. One colony per mating was picked and appropriate junction PCR was conducted using primers in plasmid backbone (Table 2.4; Ori700-F/Ori6K-R) and primers reading out from RME (Table 2.4; RME-F/RME-R).

Chapter 3: Investigating the role of the integron/gene cassette system on *Vibrio* physiology

3.1: Introduction

The majority of the results in this chapter have been published in the journal *PLoS ONE*, see appendix 2 for published manuscript (Rapa *et al.*, 2013). As described in chapter 1, integrons are genetic elements that integrate and express genes located in mobilisable elements termed gene cassettes. Unlike class 1 integrons that mostly include cassettes that possess antibiotic resistance genes and are commonly carried on plasmids, most integrons in environmental bacteria are in chromosomal locations. Approximately 10% of sequenced prokaryotic genomes harbour chromosomal integrons (Boucher *et al.*, 2007). In *Vibrio* species the integron/gene cassette system comprises ~1-3% of the entire genome and as such is a substantial source of laterally acquired DNA. Exactly how the integron influences the adaptation and evolution of *Vibrio* spp. remains largely unknown with ~80% of gene cassettes in vibrios having no known biochemical function (Boucher *et al.*, 2007). Given this, integrons are regarded as having a more general role in evolution than simply carrying and expressing antibiotic resistance genes (Boucher *et al.*, 2007;Cambray *et al.*, 2010).

Using the wild type model organism *Vibrio rotiferianus* DAT722 (referred to as wt DAT722 for the remainder of this thesis) this chapter aims to *ascertain how the chromosomal integron of DAT722 affects cellular physiology* thus providing clues on how it might affect adaptation and evolution. DAT722 was isolated from an aquaculture tank in Darwin, Northern Territory, Australia and is mildly pathogenic towards mud-crab larvae (Boucher *et al.*, 2006). The genome sequence has been published for this strain and it contains a large chromosomal integron cassette array of 116 gene cassettes (Chowdhury *et al.*, 2011). See appendix 1 for a table listing each cassette and its putative identification and putative conserved superfamily domains

A feature of chromosomal arrays is that large groups of contiguous cassettes can excise and consequently be deleted from the array at any given time (Labbate *et al.*, 2007). For example, *V. cholerae* O1 El Tor strains can be divided into separate phylogenetic clades

based on the presence of a large gene cassette deletion (Labbate *et al.*, 2007). However, how this has affected the evolution or pathogenicity of this pandemic strain is unknown (Labbate *et al.*, 2007). In an effort to understand how deletion of contiguous gene cassettes may affect *Vibrio* adaptation and evolution three different subsets of gene cassettes were deleted from wt DAT722 using homologous recombination (Labbate *et al.*, 2011 and section 2.4) (Labbate *et al.*, 2011). These three isogenic deletion mutants have been designated d16-60, d72-92 and d50-60 whereby 45, 21 and 11 cassettes respectively have been deleted from the array and replaced with a kanamycin resistance gene for selection. To aid understanding throughout this and other chapters, it is also worth noting that reference to, for example, deletion of cassettes 16-60 means that cassettes in the ordered DAT722 array from number 16 to 60 inclusive (as numbered from *attI*), are removed.

To address the aim of this chapter, the following specific objectives were examined:

1. Determine whether gene cassette deletions influence growth in DAT722
2. Determine whether gene cassette deletions affect environmental stress survival in DAT722
3. Determine whether the gene cassette deletions affect regulation and post-translational modifications of whole cell and secreted proteins
4. Determine whether the gene cassette deletions affect surface polysaccharide structures

3.2: Results

3.2.1: Do gene cassette deletions influence growth in *V. rotiferianus* DAT722?

Previous integron/gene cassette studies using *V. rotiferianus* strain DAT722 have recently shown that deletion of a different subset of gene cassettes to the three used in this thesis produced a mutant with a substantial loss of central metabolic functions affecting growth fitness (Labbate *et al.*, 2011). This study demonstrated that acquisition of gene cassettes (or potentially any mobile DNA) has the capacity to be rapidly integrated into fundamental cell networks in the bacterial cell (Labbate *et al.*, 2011). In order to determine if the deletion of gene cassettes had any impact on bacterial cell metabolism, growth curves were performed (see section 2.1). Growth curves were carried out in LB20 (a complete medium) and in 2M

salts (a marine minimal medium that mimics sea water) supplemented with various carbon sources including glucose, pyruvate, fumarate, succinate and aspartic acid.

When growth rates were compared on a logarithmic scale wt DAT722 and isogenic deletion mutants did not differ in all media types. However, minor reproducible differences were observed in optical density (OD) at specific time points in 2M salts + various carbons sources. Namely, wt DAT722 had at least 2-fold less optical density when grown in 2M + aspartic acid compared to all isogenic deletion mutants at 60 hr growth (Figure 3.1C). In 2M + succinate, d72-92 had approximately 3-fold less optical density at 12 hr growth when compared to wt DAT722 and other deletion mutants (Figure 3.1D). In 2M + fumarate d50-60 and d72-92 both had approximately a 3-fold increase in optical density when compared to wt DAT722 and the d16-60 deletion mutant after 24 hr of growth (Figure 3.1E). In 2M + pyruvate d16-60 and d50-60 had increased optical density readings at 7 and 35 hr of growth (Figure 3.1F). Growth in complete medium; LB20 revealed no differences when wt DAT722 and deletion mutants were compared at every time point (Figure 3.1A) or in 2M + glucose (Figure 3.1B). The observed fluctuations in optical density readings for these "no difference" comparisons are consistent with either proportional variability in the number of cells present in the sample or variability in the volume of individual cells.

Where differences were found data can be explained by one or more scenarios. 1) the deletion of gene cassettes has altered the metabolic balance of the cell resulting in minor variations in growth, 2) as growth differences were only evident in minimal media, the deletions have resulted in the removal of cassettes involved in stress adaptation upon nutrient deprivation or 3) the permeability of the cells has changed thereby varying the uptake of nutrients into the cell. While gene cassette deletions did not appear to drastically affect growth, the minor variations observed in cell densities may impact on environmental competiveness.

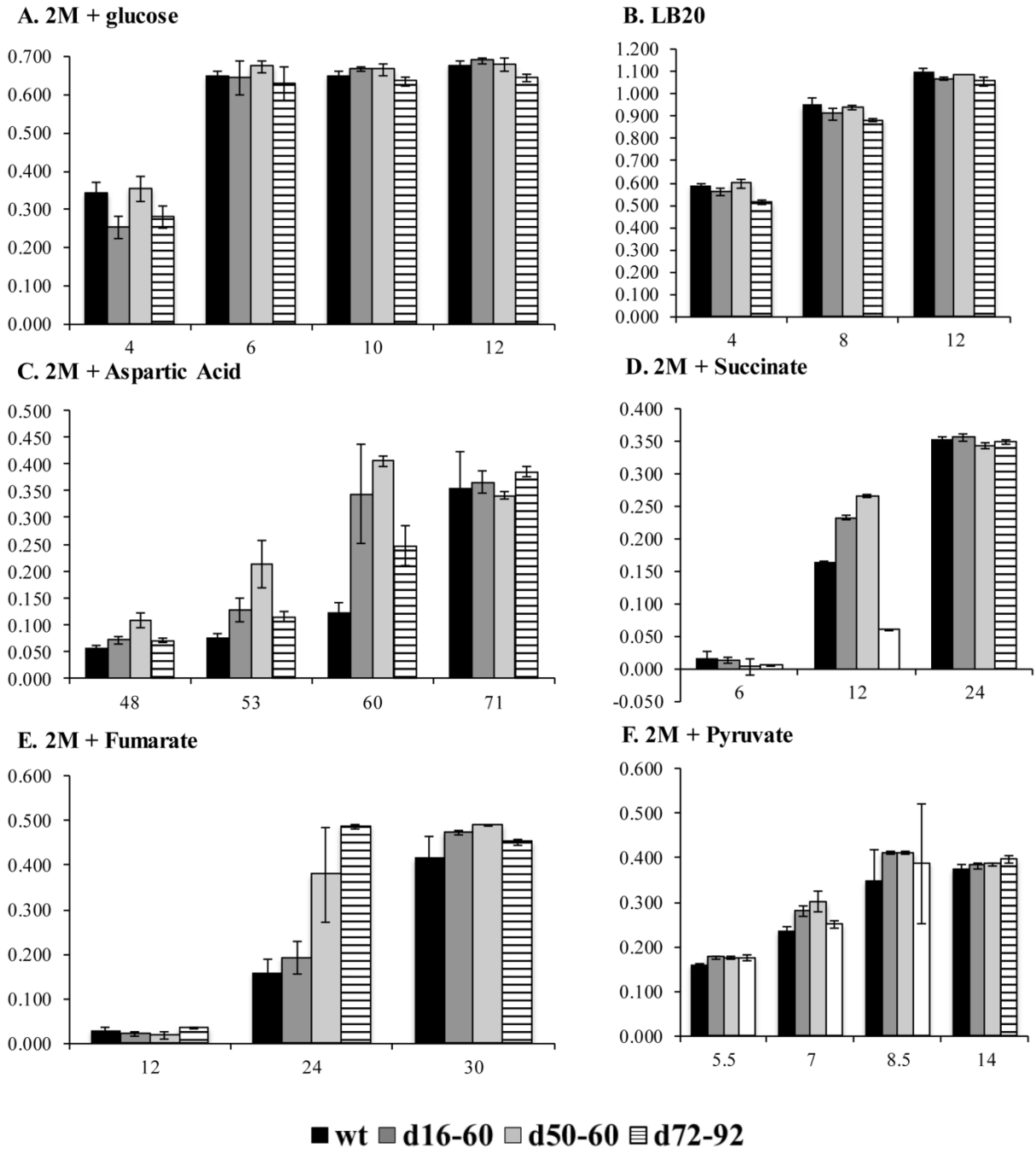
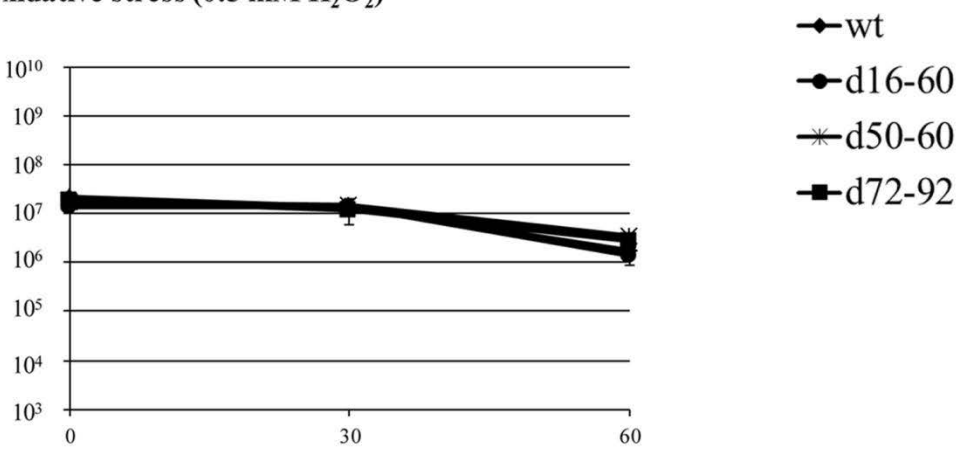


Figure 3.1: Growth of wt DAT722 and isogenic deletion mutants.
Growth of wt DAT722 and isogenic deletion mutants in LB20 (complete media) and 2M salts + various carbon sources (marine minimal media). Indicated on the horizontal axis are the times intervals (hr) of growth. Indicated on the vertical axis are optical density (OD) readings recorded at 600nm. Error bars show statistical significance of triplicate optical density readings.

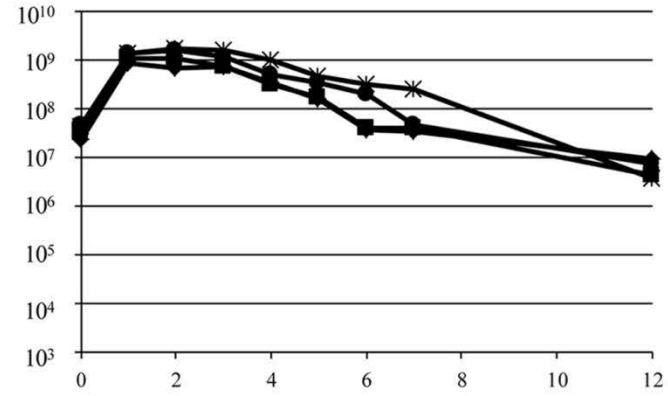
3.2.2: Gene cassette deletions do not affect environmental stress survival

Recent studies have shown that SOS-induced stress activates the integron-integrase; *intI* whereby gene cassette shuffling, integration and deletion events are induced (Guerin *et al.*, 2009). Thus, it has been suggested that during episodes of stress, gene cassette shuffling, integration and deletion may lead to events in some cells that allow for stress adaptation. Given that only minor variations were observed in growth rates between wt DAT722 and deletion mutants, environmental stress assays were performed under conditions that wt DAT722 might be expected to be exposed to in the natural environment (see section 2.12). These were oxidative (addition of H_2O_2 results in the production of reactive oxygen species), iron depletion (addition of 2,2-dipyridyl (chelates iron ions) and cold shock stresses (Figure 3.2). For each stress factor tested there was no major differences evident when wt DAT722 and deletion mutants were compared. This indicates that at least of the cassettes deleted and of the stress assays tested, gene cassettes seem to play no major role in adaptation to environmental stress.

A. Oxidative stress (0.5 mM H_2O_2)



B. Iron depletion stress (addition of 2,2-dipyridyl)



C. Cold shock stress (exposure to 4 °C)

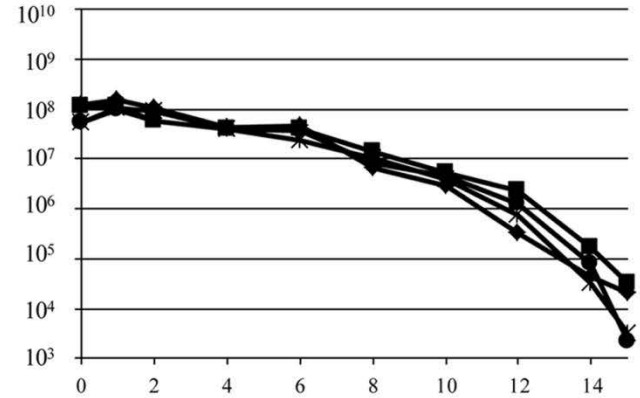


Figure 3.2: Environmental stress assays.
Environmental stress assays showing (A) oxidative, (B) iron depletion and (C) cold shock stresses. The horizontal axis represents time in hr for (A) and days (B) and (C). The vertical axis represents colony forming units. The figures shown here are representative of three independent experiments

3.2.3: Do gene cassette deletions affect regulation and post-translational modifications of whole cell and secreted proteins?

Deletion of gene cassettes 16-60 produced cells with a healthy growth phenotype and showed no impact in the tested stress assays when compared to the wt DAT722 (Figure 3.1 and 3.2). Thus, a whole cell proteomic approach (see section 2.5 for method) was adopted to observe whether deletion of gene cassettes affect protein regulation and post-translational modification. 2D-PAGE and LC-MS/MS were performed on whole cell protein extracts and 1D-PAGE on supernatant protein extracts, to compare whether deletion of gene cassettes had any influence on the up/down regulation and post-translational modification of proteins within the cell.

3.2.3.1: How does deletion of gene cassettes affect whole cell proteome?

Gene cassettes have been shown to have adverse effects on growth, indicating that mobile genes integrate into host cellular pathways (Labbate *et al.*, 2011). However, the deletion of gene cassettes in the mutants studied in this thesis did not have an observable detrimental effect on growth. Thus, in order to determine how a large deletion of gene cassettes, not adversely affecting growth would impact on whole cell protein regulation and modification, 2D-PAGE was performed. It was hypothesised that deleting a large number of gene cassettes (45 in total), as evident in d16-60 would result in the loss of some protein functions leading to changes in cellular pathways. Deletion of gene cassettes may affect cellular pathways *via* a number of mechanisms such as: 1) they may regulate other proteins through transcriptional activators or *via* RNA precursors, 2) they could influence other proteins thus altering their function or through post-translational modification(s), for example *via* phosphorylation, acetylation, glycosylation and 3) they may affect cellular structures for example, they may influence polysaccharide biosynthesis and/or the assembly of polysaccharide structures. Although a proteomic approach will not detect any change in

polysaccharide structures *per se*, proteins involved in the synthesis of these structures may be detected on a 2D-PAGE gel. Identifying these changes may provide insight into gene cassette function and the influence mobile DNA has on the adaptive potential of the cell.

Isogenic deletion mutant d16-60 (the largest array deletion) was subjected to 2D-PAGE and compared to wt DAT722 when grown in complete (LB20) and minimal (2M + glucose) media. Additionally, protein expression in stationary and mid-logarithmic phase was compared under both nutrient conditions. For cells grown in LB20 to stationary phase, two out of 325 resolved protein spots were shown to be differentially expressed between wt DAT722 and d16-60. In mid-logarithmic phase, four out of 357 resolved protein spots were differentially expressed; see Table 3.1. For cells grown in minimal media to stationary phase, three out of 360 resolved protein spots were differentially expressed and in mid-logarithmic phase, three out of 201 protein spots were differentially expressed; see Table 3.2. Thus, overall approximately 0.5-1% of protein spots identified between wt DAT722 and d16-60 were at least 2-fold differentially expressed with a maximum fold difference of 3.7 for protein spot 2CM (Table 3.2). Furthermore, no post-translational modification of proteins was observed such as: acetylation, glycosylation or phosphorylation. This proteomic approach revealed subtle differences in the whole cell proteome between wt DAT722 and d16-60, at least under the conditions tested. This indicates that although 45 gene cassettes have been removed from the array (~31kb) there has been minimal impact on the overall proteome. It is possible however, changes in lower abundance proteins not detected using 2D-PAGE as a methodology occurred.

Due to co-migration of proteins not all protein spots could be unambiguously identified. Out of the unambiguous protein spots that were differentially expressed, surface-associated proteins were identified (Tables 3.1 and 3.2). These include OmpA (outer membrane protein; spot 1CM), an OmpU-like protein (spot 2CM) and an unknown protein (spot 1CS). Since the gene that encodes this unknown protein is located in a region of the genome responsible for polysaccharide (polymeric carbohydrate molecules made up of monosaccharide units joined by glycosidic bonds) biosynthesis and PSORTb analysis (described in section 2.5.5.4) predicts its localisation to be extracellular, it is likely that the unknown protein is associated with polysaccharide(s) biosynthesis (Figure 3.3). Metabolic

proteins were also identified as being differentially expressed including proteins with homology to: nitrogen regulatory protein P-II (spot 2CS), alkyl hydroperoxide reductase subunit C-like protein (spot 2MM) and 6,7-dimethyl-8-ribityllumaxine synthase (spot 3MM). An uncharacterised protein (spot 2MS) was also identified as being differently expressed. Given that most of the gene cassettes in *V. rotiferinaus* DAT722 are transcribed it was surprising that no gene cassette-associated proteins were identified as missing in the d16-60 mutant (Michael and Labbate, 2010). Presumably this is due to the presence of high abundance proteins masking the detection of those present in lower abundance (Shaw and Riederer, 2003;Heppelmann *et al.*, 2007).

Table 3.1: Differentially expressed spots between deletion mutant d16-60 and wt *V. rotiferianus* DAT722 in LB20

Growth Phase	Differentially expressed spot	LC-MS/MS match(s)	Number of matched peptides	PEAKS score* (%)	Fold change in mutant	Accession number
Mid-logarithmic phase						
	1CM[1]	OmpA; VrotD_16305	8	99.0	-2.31	ZP_08911638
	2CM[1]	OmpU-like outer membrane protein	10	98.8	-3.68	ZP_08912594
	^3CM[1]	50S ribosomal protein L9	10	98.8	-2.61	ZP_08908422
		OmpA-like membrane protein; VrotD_08232	7	97.5		ZP_08910036
	^4CM[1]	ATP-dependent Clp protease proteolytic subunit	9	98.3	-2.90	ZP_08909655
		S-ribosylhomocysteinase/Autoinducer-2 production protein LuxS	5	97.9		
		Type VI secretion-related protein				ZP_08911325
		Shikimate kinase I	3	96.8		
			3	96.4		ZP_08909611
						ZP_08908430
Stationary phase						
	1CS[2]	Unknown protein (gene is surrounded by genes encoding O-antigen biosynthesis or export); VrotD_02720	15	99.0	+2.30	ZP_08908944
	2CS[2]	Nitrogen regulatory protein P-II	13	96.9	-2.75	ZP_08911087

* highest PEAKS score (percentage based on a p-value <0.05) was taken as the closet peptide match

^ denotes co-migrating protein spots

[1]CM denotes cells grown in **c**omplete medium (LB20) to **m**id-logarithmic phase

[2]CS denotes cells grown in **c**omplete medium (LB20) to **s**tationary phase

Table 3.2: Differentially expressed spots between deletion mutant d16-60 and wt *V. rotiferianus* DAT722 in 2M + glucose

Growth Phase	Differentially expressed spot	LC-MS/MS match(s)	Number of matched peptides	PEAKS score* (%)	Fold change in mutant	Accession number
Mid-logarithmic phase						
	^1MM[1]	Putative membrane protein; VrotD_04538	6	98.6	+2.73	ZP_08909304
		Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (Histidine biosynthesis)	4	95.1		ZP_08910780
	2MM[1]	Alkyl hydroperoxide reductase subunit C-like protein	3	98.3	+2.45	ZP_08909276
	3MM[1]	6,7-dimethyl-8-ribityllumazine synthase	15	99.0	-2.01	ZP_08909366
Stationary phase						
	^1MS[2]	Cysteine synthase A	12	95.4	+2.11	ZP_08909513
		Acetyl-coenzyme A carboxyl transferase alpha chain	7	92.0		ZP_08911090
		OmpT	17	90.2		ZP_08909742
	2MS[2]	Unknown protein; VrotD_21668	10	98.3	-2.00	ZP_08912704
	^3MS	Unknown protein; VrotD_07347	16	96.1	+2.55	ZP_08909861
		NAD-dependent glyceraldehyde-3-phosphate dehydrogenase	13	96.0		ZP_08910944
		Phenylalanyl-tRNA synthetase alpha chain	9	94.4		ZP_08910491
		Universal stress protein E	10	93.8		ZP_08910109

* highest PEAKS score (percentage based on a p-value <0.05) was taken as the closet peptide match
^ denotes co-migrating protein spots
[1] MM denotes cells grown in **m**inimal medium (2M + glucose) to **m**id-logarithmic phase
[2] MS denotes cells grown in **m**inimal medium (2M + glucose) to **s**tationary phase

	Size (nt)	Strand	Function	Accession number
1	1044	-	Undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase	ZP_08908938
2	249	+	Hypothetical protein	ZP_08908939
3	966	+	Glycosyl transferase, group 2 family protein	ZP_08908940
4	2163	+	Hypothetical protein	ZP_08908941
5	876	+	Putative glycosyl transferase	ZP_08908942
6	966	+	Alpha-L-Rha alpha-1,3-L-rhamnosyltransferase	ZP_08908943
7	1560	+	Unknown protein (spot 1CS)	ZP_08908944
8	1083	+	dTDP-glucose 4,6-dehydratase	ZP_08908945
9	891	+	dTDP-4-dehydrorhamnose reductase	ZP_08908946
10	885	+	Glucose-1-phosphate thymidylyltransferase	ZP_08908947
11	537	+	dTDP-4-dehydrorhamnose 3,5-epimerase	ZP_08908948
12	786	+	O-antigen export system permease protein RfbD	ZP_08908949
13	1662	+	General secretion pathway protein E / Type II secretory pathway, ATPase PulE	ZP_08908950

Figure 3.3: Genomic localisation of unknown protein.
This figure shows the genetic context for the gene (shaded in blue) encoding the differentially expressed unknown protein (spot 1CS) in *V. rotiferianus* DAT722.

3.2.3.2: An extracellular contaminating substance produced by d16-60 interferes with the IEF step in 2D-PAGE and gives insight into how gene cassettes affect DAT722 physiology

During 2D-PAGE analysis it was observed that whole cell protein extracts from stationary phase d16-60 cells grown in 2M + glucose consistently contained a substance that interfered with the isoelectric focusing (IEF) step in 2D-PAGE (Figure 3.4). This substance resulted in the streaking of protein spots and was eventually removed from the sample by washing the cells with 2% NaCl pre-protein extraction (see section 2.5.1.2). This represented a point of physiological difference between wt DAT722 and deletion mutant d16-60.

Multiple substances can interfere with the IEF step of 2D-PAGE including: DNA, cell wall material and polysaccharides (Shaw and Riederer, 2003). Methods were utilised to remove contaminating DNA and cell wall (see methods section 2.5). The contaminating substance found to interfere with IEF was extracellular as it could be easily removed by washing, raising the notion that it may be polysaccharide in nature. To support this hypothesis, a protein putatively involved in polysaccharide biosynthesis was also observed in 2D-PAGE analysis (spot 2CS). Although this substance is yet to be identified, it is extracellular, charged and is weakly associated with the cell surface (as washing readily removed it) and thus was hypothesised to be polysaccharide in nature.

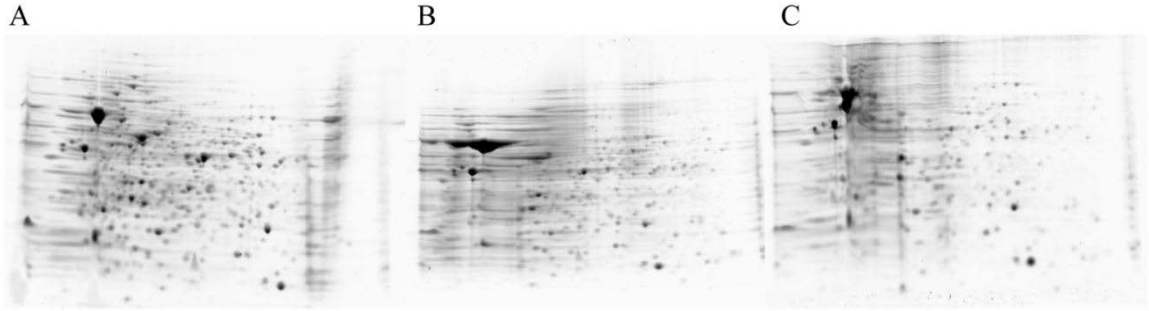


Figure 3.4: Contaminating substance affecting IEF of 2D-PAGE analysis.
(A), (B) and (C) show biological triplicate 2D-PAGE gels all affected by the contaminating substance present in whole cell protein extracts from d16-60 cells grown to stationary phase in 2M + glucose.

3.2.3.3: Does deletion of gene cassettes impact on the secretome?

In addition to 2D-PAGE, 1D-PAGE was performed on the supernatant of all deletion mutants grown to stationary phase in 2M + glucose to determine whether the deletion of gene cassettes is influencing secreted proteins (Figure 3.5). No major differences were found, however slight variations in the abundance of some proteins were observed.

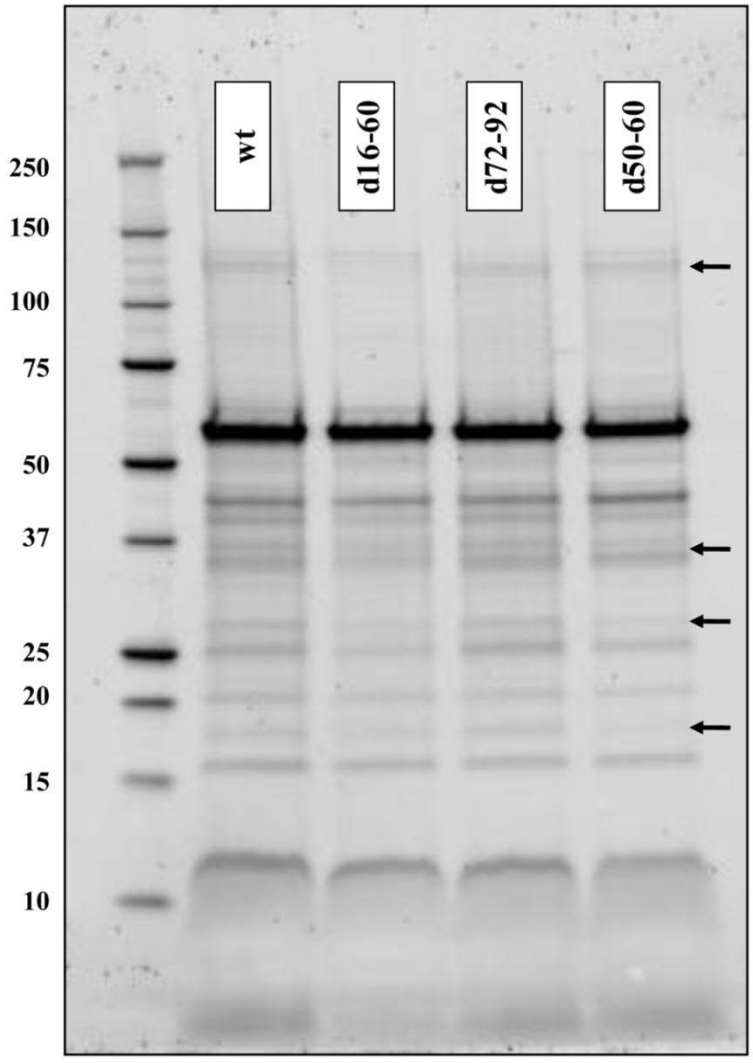


Figure 3.5: Gel electrophoresis of supernatant proteins.
1D-PAGE of precipitated supernatant proteins from wt DAT722 and deletion mutants grown in 2M + glucose to stationary phase. Protein standards are indicated left of the gel and are sized in kDa. Lanes are labelled with strain names and arrows indicate protein bands of differing abundance.

3.2.3.4: Are gene cassettes involved in modification of glycosylated proteins?

Bioinformatic analysis using RAST (**R**apid **A**nnotation **S**ubsystem **T**echnology; see section 2.10) revealed at least four genes that may encode functions associated with the glycosylation of proteins named glycoproteins (Table 3.3). Glycoproteins are proteins covalently linked to oligosaccharide side chains. Glycosylation of proteins occurs is a post-translational modification and as stated in section 3.2.3.1 was not a point of difference between wt DAT722 and d16-60 detected by 2D-PAGE analysis. To further substantiate the 2D-PAGE analysis performed and determine whether deletion of gene cassettes from wt DAT722 impacts on glycoprotein expression, protein extracts in 2M + glucose and LB20 were stained with the glycoprotein specific stain Pro-Q Emerald (Figure 3.6; see section 2.5.3.3). However, there was no change in glycoprotein profiles between wt DAT722 and deletion mutant d16-60 (Figure 3.6). These data correlate with the lack of post-translational modification of proteins identified during 2D-PAGE. It should be noted that this staining technique will only identify differences in the glycosylation of proteins and is not directly related to any other changes occurring to surface polysaccharide as a consequence of gene cassette deletions. Thus, the remainder of this chapter details experiments performed in order to identify whether gene cassettes are influencing the polysaccharide of *V. rotiferianus* DAT722.

Table 3.3: Putative genes in *V. rotiferianus* DAT722 involved in glycosylation of proteins

Functional category	Functional subcategory	RAST subsystem classification	Putative Role
Protein Metabolism	Protein processing and modification	N-linked Glycosylation in Bacteria	UDP-N-acetylglucosamine 4,6-dehydratase
Protein Metabolism	Protein processing and modification	N-linked Glycosylation in Bacteria	UDP-glucose 4-epimerase
Protein Metabolism	Protein processing and modification	N-linked Glycosylation in Bacteria	Lipid carrier : UDP-N-acetylgalactosaminyltransferase
Protein Metabolism	Protein processing and modification	N-linked Glycosylation in Bacteria	4-keto-6-deoxy-N-Acetyl-D-hexosaminyl-(Lipid carrier) aminotransferase

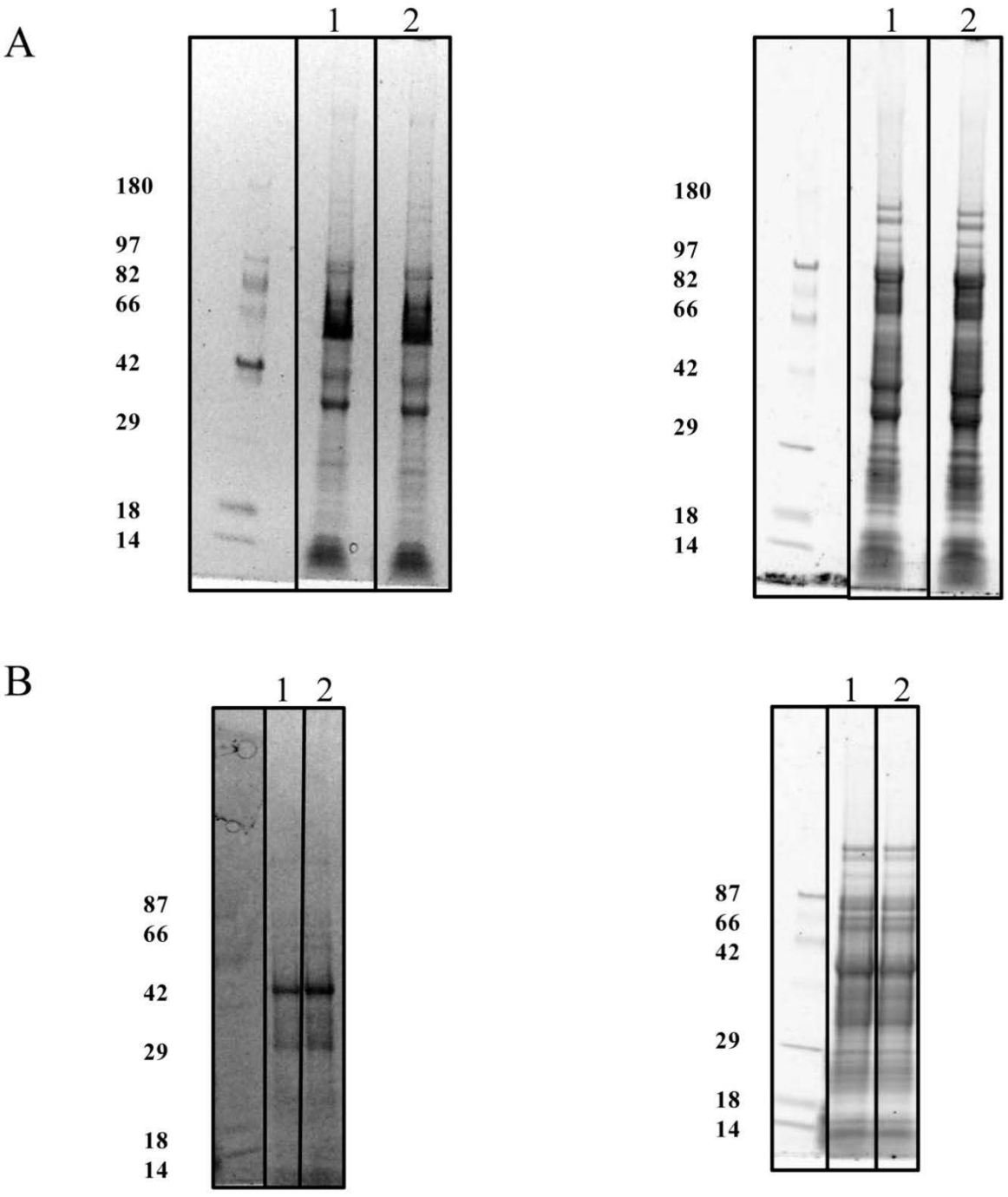


Figure 3.6: Glycoprotein gel electrophoresis.
Whole cell protein extracts of wt *V. rotiferianus* DAT722 (1) and d16-60 (2) cells grown in LB20 (A) and 2M + glucose (B) stained with Pro-Q Emerald gels (images on the left) for glycoproteins and SYPRO Ruby for proteins (images on the right). Protein/glycoprotein standards are labelled and given in bold as kDa.

3.2.4: Investigation into whether gene cassette deletions modify surface polysaccharide

Given that the contaminating substance produced by deletion mutant d16-60 was most likely be polysaccharide in nature, a number of assays were performed in order to confirm this.

Gram negative bacterial cells, including vibrios, have three major types of surface associated polysaccharide (carbohydrates made up of repeating mono- or di-saccharide units joined by a glycosidic bond): lipopolysaccharide (LPS), capsular polysaccharide (CPS) and extracellular polysaccharide (EPS) (Chen *et al.*, 2010) (see Figure 3.7).

LPS is made up of three components: lipid A, core polysaccharide and a variable O-antigen (Reeves *et al.*, 1996). The lipid A is composed of sugars and fatty acids, which anchor the LPS into the outer membrane. The core is composed of sugars and their derivatives and the O-antigen is a polysaccharide that protrudes from the cell surface, consisting of repeating oligosaccharide units that are composed of 3-6 sugars. The O-antigen is often a virulence factor in many bacteria, including *V. cholerae* (Reeves *et al.*, 1996). CPS or K-antigen is composed of high molecular weight polysaccharide and forms a dense, protective layer on the outside of bacterial cells (Reeves *et al.*, 1996;Whitfield, 2006). Often, the K- and O-antigens are encoded on separate regions of the bacterial chromosome, with ~80 K-antigen and ~ 170 O-antigen variations present in gram negative bacteria. It is no wonder that pathogens, through the use of antigenic variation (changing the K- and O-antigen present on their surface) can successfully evade their host/s immune response (Whitfield, 2006). Thus identifying mechanisms that alter CPS is important when studying pathogen evolution.

EPS forms a loose slime layer outside the cell that often makes up the matrix in a biofilm, or also forms a high molecular weight capsule, without CPS (Nyunt *et al.*, 1998). In *V. cholerae* when EPS is expressed in cells it is displayed as a rugose (wrinkled) colony phenotype. This rugose phenotype is associated with cell aggregation, pathogenicity, environmental stress protection and nutrient acquisition (Nyunt *et al.*, 1998).

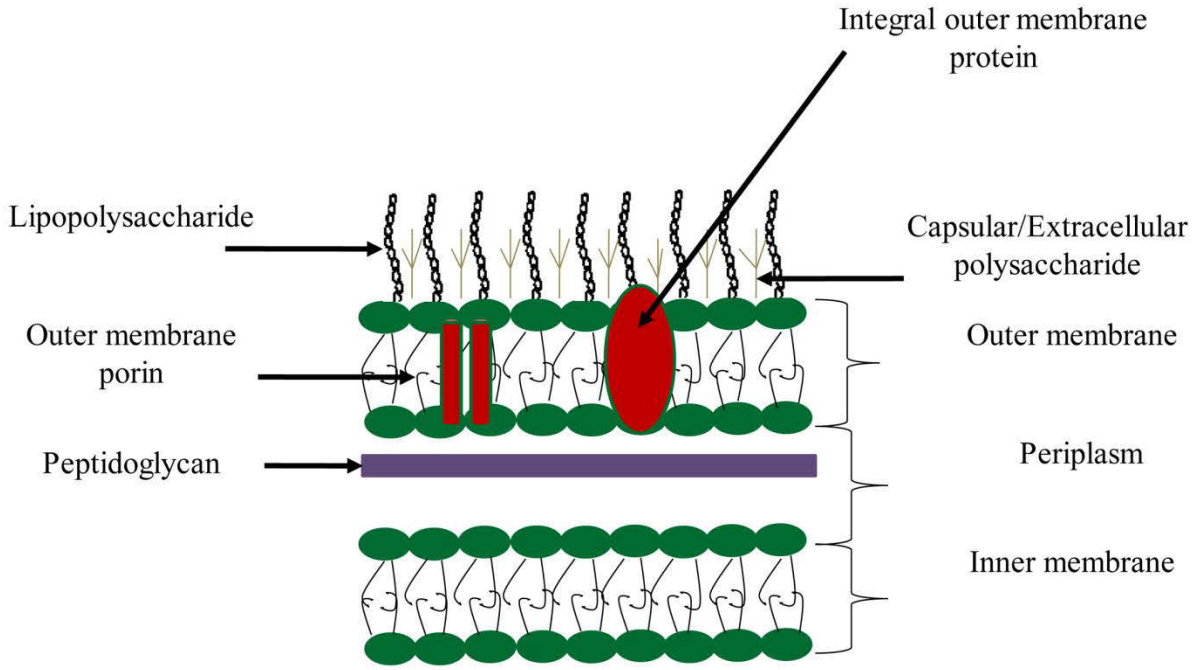


Figure 3.7: Overview of cell envelope.
An overview of specific structures present in the inner and outer membrane of gram negative bacterial cells. The cell envelope is divided here into three sections: inner membrane, periplasm and outer membrane. Polysaccharide structures LPS, CPS and EPS are indicated also with LPS imbedded in the outer membrane. Other structures such as outer membrane proteins/porins are also indicated in red.

3.2.4.1: Flow cytometry of cells stained with calcofluor and fluorescently labelled lectin

To determine whether there was a difference in the nature of extracellular polysaccharide(s) between wt DAT722 and deletion mutants, cells were grown to stationary phase in 2M + glucose (the medium that the contaminating substance was found in) and stained with the polysaccharide stain calcofluor and subjected to flow cytometery (see 2.8 for method). Calcofluor binds to beta-glucans which are polysaccharides of D-glucose monomers linked by beta-glycosidic bonds. Calcofluor emits in the DAPI range (461nm) and it was expected that there would be a difference in the way the deletion mutants bind calcofluor. However, using the required filter, wt DAT722 cells were auto-fluorescent within this range, thus any shift in fluorescence emission due to changes in the ability of the stain to bind polysaccharide would not be noticeable and interpretation would be ambiguous (Figure

3.8). This result is not entirely surprising as it has been shown that bacterial cells naturally fluoresce and is suggested as a typing/identification technique (Giana *et al.*, 2003).

Given the levels of auto-fluoresence of wt DAT722 cells using the calcofluor stain that emits in the DAPI range, flow cytometry was then performed using cells stained with the concanvalin A lectin conjugated to a FITC fluorophore (emits at 594nm). This lectin binds alpha mannose and glucose monomers, which are common sugars in the backbone of polysaccharides. Thus any alterations in these common sugars, due to gene cassette deletions, would be of importance. Unfortunately, flow cytometry revealed no differences between wt DAT722 and all deletion mutants (Figure 3.9). This could be explained by one of two reasons 1) *V. rotiferianus* DAT722 does not have any binding sites for this lectin or 2) the lectin was compromised. Thus, wt DAT722 was stained using this lectin and visualised using basic fluorescence microscopy (see section 2.7.2) to acertain whether the lectin had been compromised. Microscopy revealed the stain was able to bind to the surface of wt DAT722 cells (Figure 3.10). It was concluded that using flow cytometry to reveal differences in polysaccharide(s) between wt DAT7722 and deletion mutants was inconclusive as the method was not sensitive or robust enough to distinguish polysaccharide changes.

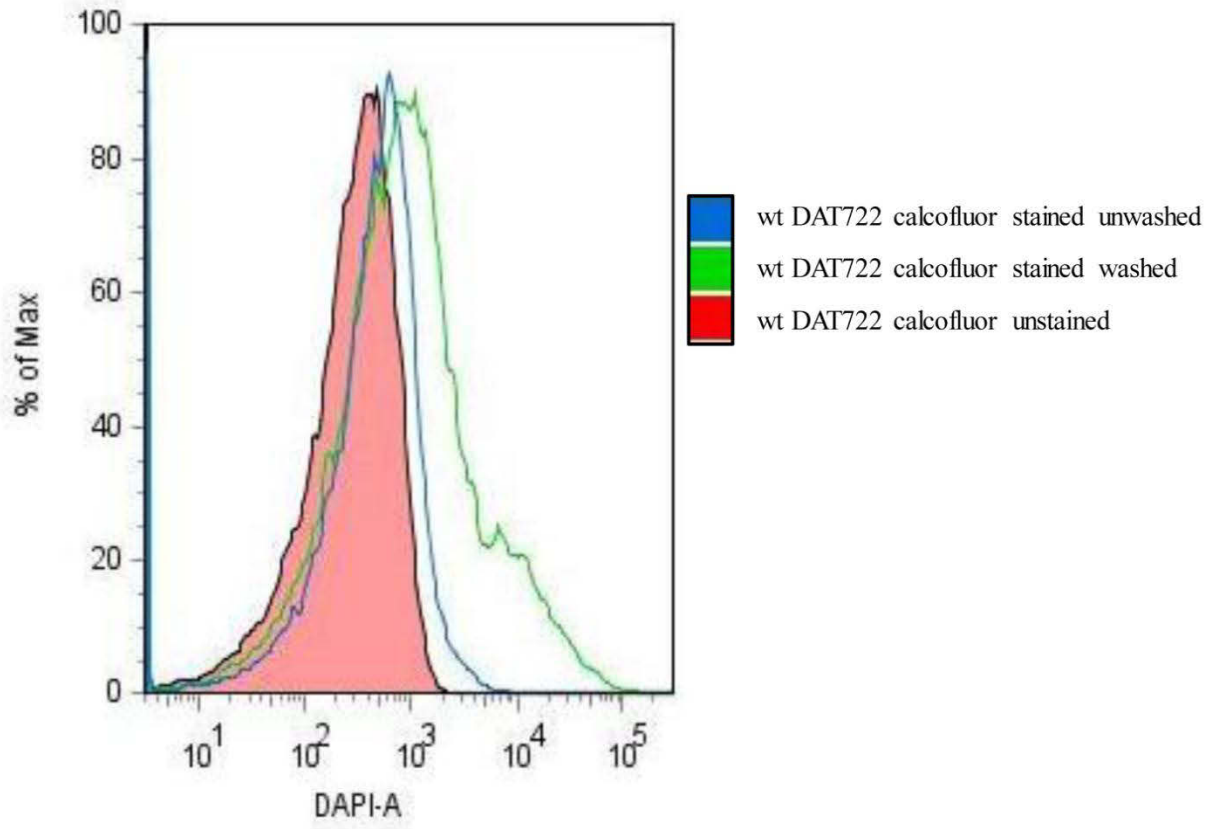


Figure 3.8: Auto-fluorescence in *V. rotiferianus* DAT722.
When wt DAT722 cells were stained with calcofluor to investigate changes to surface polysaccharide, an overlap of stained cells that had been washed post-calcofluor treatment (green) and cells that remained unwashed-calcofluor and unstained (red) cells was apparent. This resulted in the inability to detect changes due to changes in polysaccharide. It also was demonstrated that washing to remove excess calcofluor stain did not alter the shift in fluorescence intensity drastically. The horizontal axis indicates the relative emission using the DAPI filter and the vertical axis is the recorded number of bacterial cells.

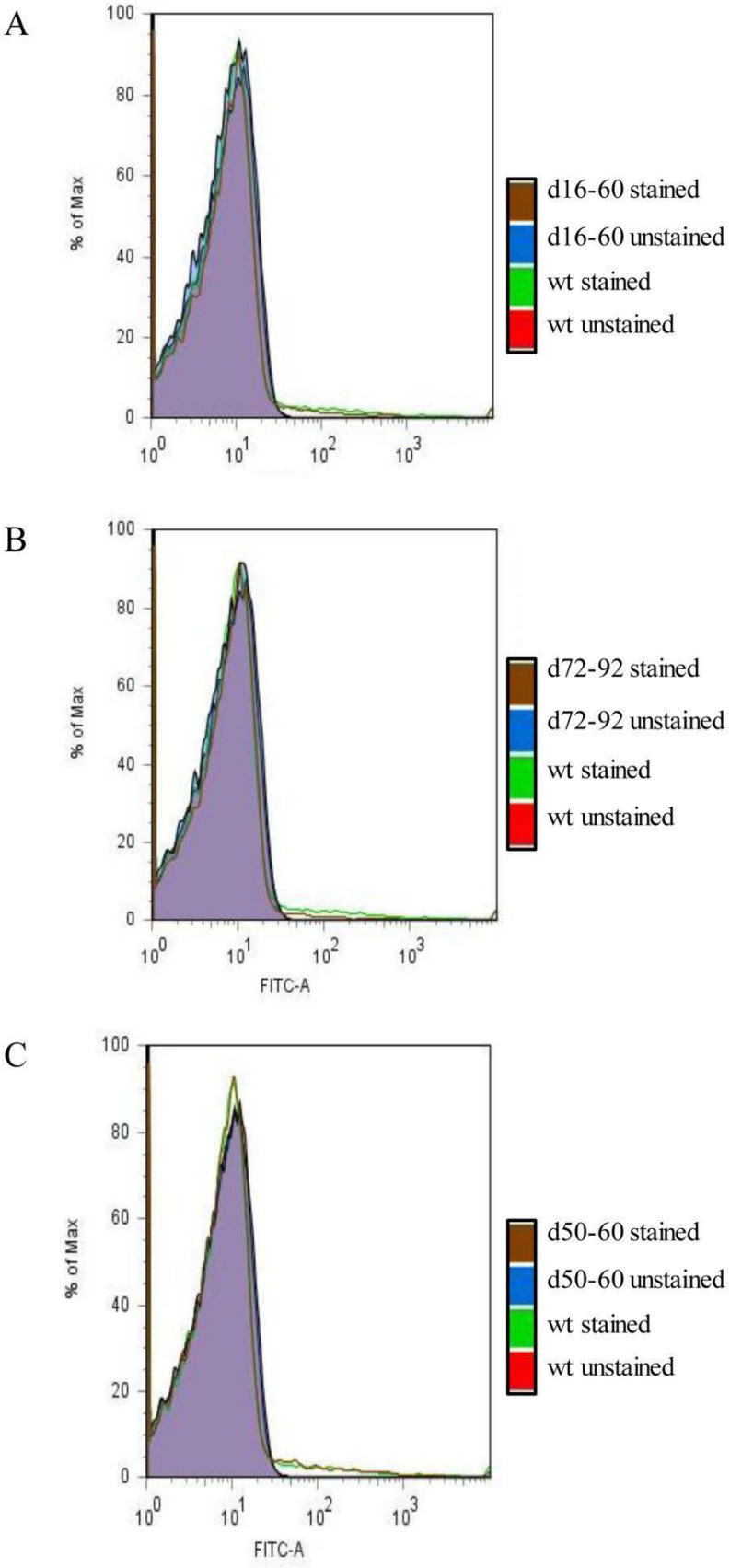

A
100
80
60
40
20
0
% of Max
10^0
10^1
10^2
10^3
d16-60 stained
d16-60 unstained
wt stained
wt unstained
B
100
80
60
40
20
0
% of Max
10^0
10^1
10^2
10^3
FITC-A
d72-92 stained
d72-92 unstained
wt stained
wt unstained
C
100
80
60
40
20
0
% of Max
10^0
10^1
10^2
10^3
FITC-A
d50-60 stained
d50-60 unstained
wt stained
wt unstained

Figure 3.9: Concanvalin A lectin binding of *V. rotiferianus* DAT722 and deletion mutant cells.
When d16-60 (A) was stained with concanvalin A and compared to wt DAT722 using flow cytometry, no shift in emission intensity was identified, this was similar for d72-92 (B) and d50-60 (C) when stained with this lectin. The horizontal axis indicates the relative emission using the FITC filter and the vertical axis is the recorded number of bacterial cells.

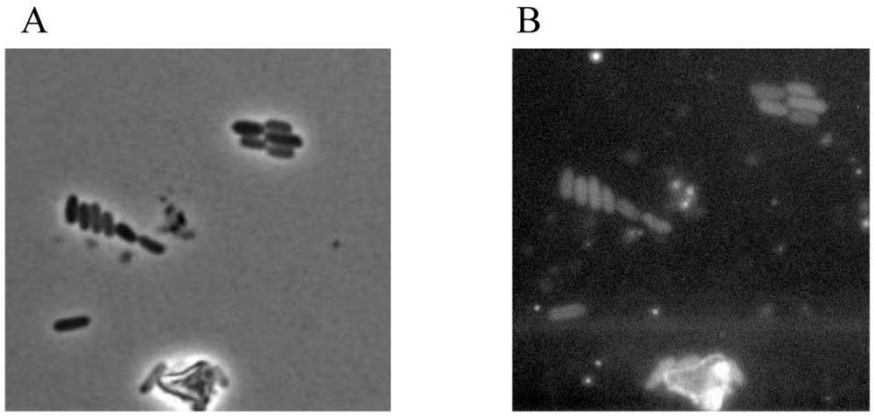


Figure 3.10: Fluorescence microscopy of wt DAT722 cells stained with concanvalin A.
Image (A) shows the phase contrast representation of wt DAT722 cells stained with concanvalin A conjugated to the FITC fluorophore at 100X. Image (B) shows the identical cells when imaged using the DAPI filter at 100X. It can be seen that the cells are able to be stained by the concanvalin A lectin. The fluorescence image was taken with an exposure time of 1000 ms.

3.2.4.2: Gel electrophoresis of extracellular and lipopolysaccharides reveals potential minor changes in polysaccharide structures between wt DAT722 and deletion mutant d16-60

In order to elucidate whether the deletion of gene cassettes 16-60 in *V. rotiferianus* DAT722 was impacting on LPS and/or EPS/CPS, both were extracted (see section 2.6.1 and 2.6.2 for methods) from cells grown in 2M + glucose (media in which contaminating substance in IEF was detected) and subjected to gel electrophoresis and silver staining (Figure 3.11). LPS studies were performed on whole cell polysaccharide extractions and EPS/CPS analysis was carried out using cells that had been shaken to remove this loosely attached polysaccharide. Minor subtle differences were identified in two out of three replicates performed and Figure 3.11 shows representative gels. From a combined EPS and CPS extraction it was noted that when wt DAT722 and d16-60 extracts were compared, wt

DAT722 EPS/CPS had two extra bands present at ~75 and ~37 kDa (Figure 3.11) that were absent from d16-60. Likewise, when total polysaccharide was subjected to gel electrophoresis the lipid A core region of the wt DAT722 extract (seen as a smear between <10 and 15 kDa; Figure 3.10) appeared larger. Also noted were a few extra bands in wt DAT722 total polysaccharide within the range of 37-50 kDa (Figure 3.11) that were again absent from d16-60. Figure 3.11 also indicates that there are at least two polysaccharide structures produced by *V. rotiferianus* DAT722. However, this data alone although providing evidence for a gene cassette(s) role in polysaccharide biosynthesis, is ambiguous and minor variations may be a consequence of loading discrepancies between samples.

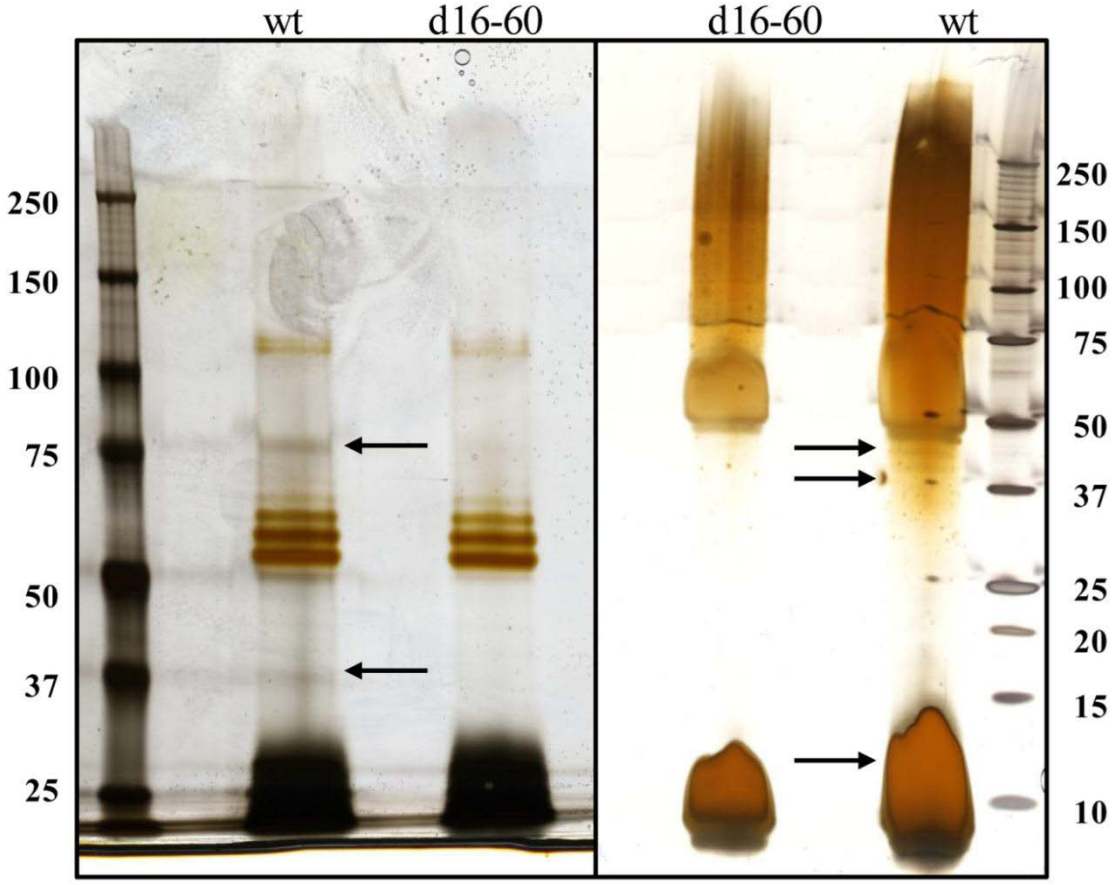


Figure 3.11: Gel electrophoresis of total surface polysaccharides.
EPS/CPS (loosely attached polysaccharide; left hand image) was extracted from wt DAT722 and d160-60 cells grown in 2M + glucose. Arrows indicate two bands at ~ 75 and 37 kDa that are present in wt DAT722 EPS/CPS extracts but not d16-60. Total polysaccharide (including LPS) was also extracted from wt DAT722 and d16-60 cells grown in 2M + glucose (right hand image). Arrows indicate additional bands present in the wt DAT722 extract between 37-50 kDa. An arrow also draws attention to the lipid A core region of the lane (<10-15 kDa). This region appears to be larger in the wt DAT722 extract than the d16-60 extract. Protein standards were used in these gels and are indicated on either side of each image.

3.2.4.3: Congo red colony staining

To substantiate the gel electrophoresis findings, congo red colony staining was performed on cells grown on LB20 and 2M + glucose agar for 7 days at 28°C (see section 2.6.5.1 for method). Colony wrinkling has previously been shown to be associated with EPS/CPS in vibrios (Chen *et al.*, 2010). Interestingly, no differences could be observed between wt DAT722 and all deletion mutants when grown on 2M + glucose supplemented with congo red (Figure 3.12B). This was surprising, especially since it was in this medium that d16-60 was producing the contaminating substance. However, substantial variations in colony wrinkling were observed when cells were grown on LB20 (Figure 3.12A). On LB20, wt DAT722 colonies showed repeated characteristic wrinkling architecture, whereas the deletion mutants produced differing levels of wrinkling. This indicated a change in the amount or structure of produced polysaccharide in these mutants. The lack of colony wrinkling when cells were grown on 2M + glucose medium indicates that under these nutrient conditions, *V. rotiferianus* DAT722 produces a different polysaccharide(s) that does not result in colony wrinkling. The production of different polysaccharide(s) under varying growth conditions has been reported in other bacteria (Poutrel *et al.*, 1995;Joseph and Wright, 2004).

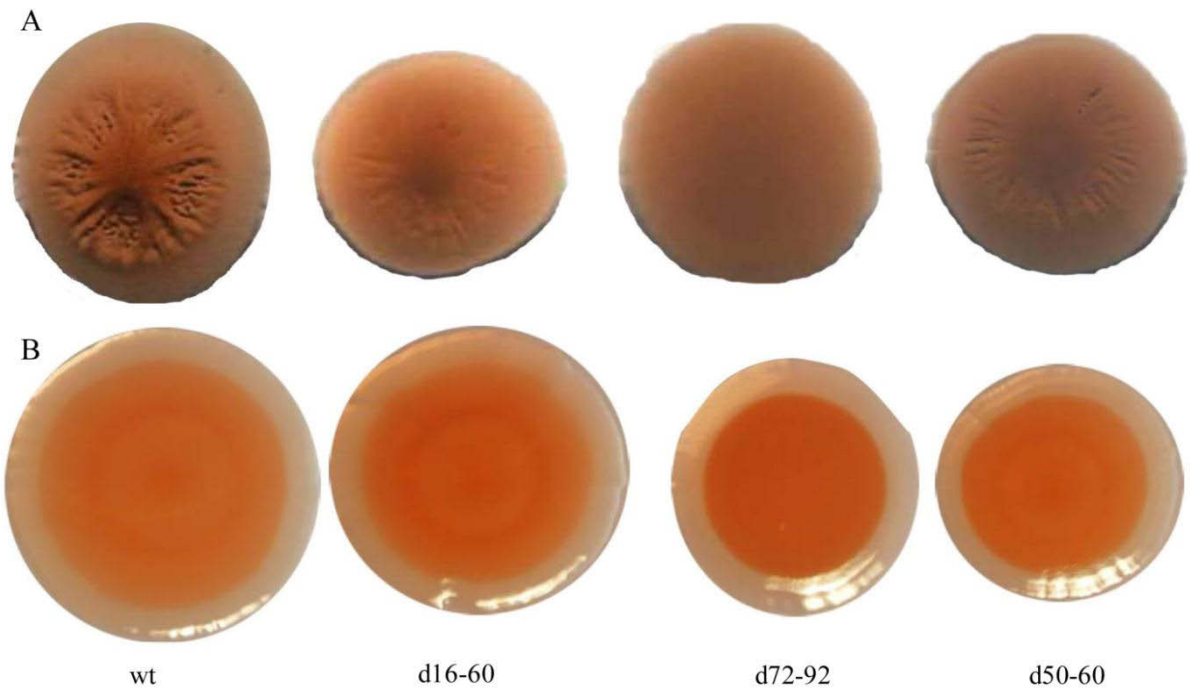


Figure 3.12: Congo red colony morphology.
Colonies of *V. rotiferianus* DAT722 and deletion mutants grown on agar plates supplemented with 0.001% congo red. Colonies grown on LB20 and 2M + glucose supplemented with congo red are shown in panels (A) and (B) respectively. All colonies were imaged after 7 days growth at 28°C.

3.2.4.4: Proton nuclear magnetic resonance spectroscopy of whole cell polysaccharides

Previous data presented in section 3.2.4 suggested that deletion of gene cassettes impact on polysaccharide structures. However, the experiments performed were not conclusive. To confirm that deletion of gene cassettes had modified cell surface polysaccharides, total polysaccharide from wt DAT722, d16-60 and d72-92 was extracted from cells grown in 2M + glucose medium (see section 2.6.6 for method) and subjected to ^{1}H NMR analysis. These two mutants were selected for ^{1}H NMR as they represented unique gene cassette deletions (d50-60 has 11 cassettes deleted that are in common with d16-60). To ensure changes were a result of the deletions and not any secondary mutation(s), total whole cell polysaccharide was also extracted and purified from identical but independently derived mutants: d16-60a and d72-92b (described in section 2.4).

Whole cell polysaccharide extractions include: LPS and EPS/CPS structures, so any changes are directly related to these structures. However, which polysaccharide moiety is altered cannot be identified.

The ^{1}H NMR spectra of wt DAT722, d16-60 and d72-92 were initially compared (Figure 3.13). The spectra showed dissimilarity between the wt DAT722, d16-60 and d72-92, as well as variation between the two deletion mutants themselves. The differences noted were in the chemical shift region of approximately 3.0-4.5 ppm (Figure 3.13). Fresh extractions of the wt DAT722, d16-60a and d72-92a produced ^{1}H NMR spectra similar to the first batch (Figure 3.13). However, due to a slight modification in the extraction methodology (see section 2.6.2) differences in purity were observed. Nevertheless, changes were again observed in the chemical shift region between 3.0-4.5ppm (Figure 3.14). Therefore, these changes to whole cell polysaccharide are confirmed to be a direct result of the gene cassette deletions.

The 3.0-4.5 ppm section is identified as the carbohydrate ring proton region of the diagnostic NMR spectrum. Alterations in the chemical shift between wt DAT722 and deletion mutants indicate that there are changes in the functional groups attached to carbohydrate molecules. A peak labelled HDO (deuterated water) is present at 4.75 ppm (Figure 3.13 and 3.14). Differences in the width of this peak indicated a saturated water signal as a consequence of incomplete exchange of all hydrogen atoms present in water with deuterium (see section 2.6.6 for method). This does not affect the results presented. The anomeric region (4.5-6.0 ppm) labelled in Figure 3.13 shows little change in the chemical shift of peaks between wt DAT722 and mutants indicating that there are no alterations in the type of sugar(s) (i.e glucose, sucrose, fructose) present along the backbone of the polysaccharide structure(s). Overall, it is not surprising that there are distinct differences between d16-60 and d72-92 mutants as these have different gene cassette deletions..

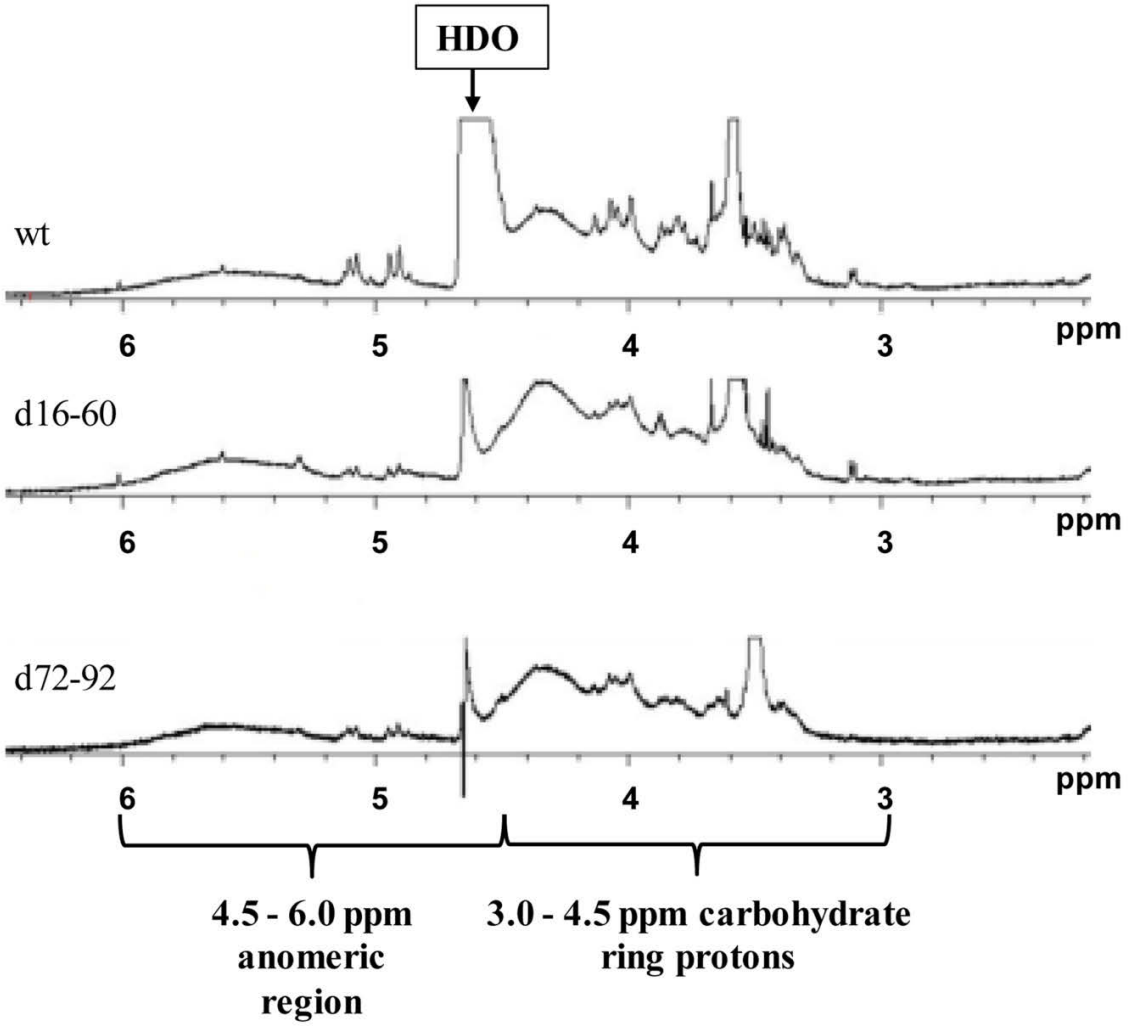


Figure 3.13: ^{1}H NMR spectra.
^{1}H NMR spectra of wt *V. rotiferianus* DAT722, d16-60 and d72-92 whole cell polysaccharide.

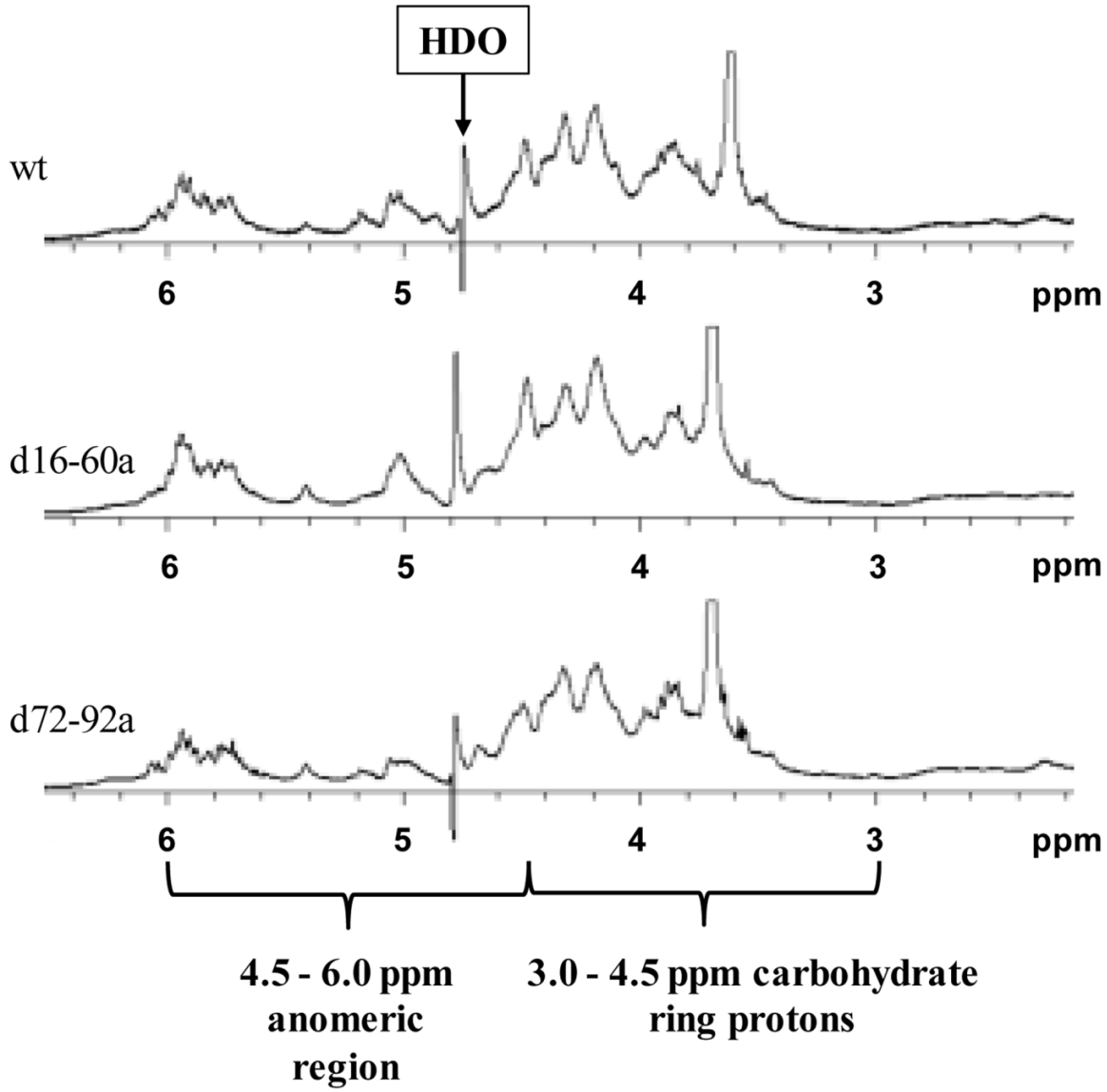


Figure 3.14: ^{1}H NMR replicate spectra.
^{1}H NMR spectra of wt *V. rotiferianus* DAT722, d16-60a and d72-92a whole cell polysaccharide.

3.2.5: A gene cassette deletion impacts on biofilm formation

^{1}H NMR confirmed that deletion of gene cassettes indeed impacts on whole cell polysaccharide. As alterations to bacterial surface polysaccharide can affect how the bacterial cells interact with the environment, simple batch biofilm assays (see section 2.7.4 and 2.9 for methods) were performed to identify whether deletion of cassettes influence biofilm formation. Deletion mutant d72-92 showed statistically significant higher biomass when compared to wt DAT722, d16-60 and d50-60 when stained with 0.2% crystal violet (Figure 3.15 and 3.16). It is likely that the changes to polysaccharide as a result of deleting cassettes 72-92 have altered the cell's ability to adhere or form a biofilm. Biofilm assays were performed using a hydrophilic plastic surface. However, different affects may be observed in the deletion mutants if biofilm experiments were conducted on other surfaces such as metal, chitin or eukaryotic tissues.

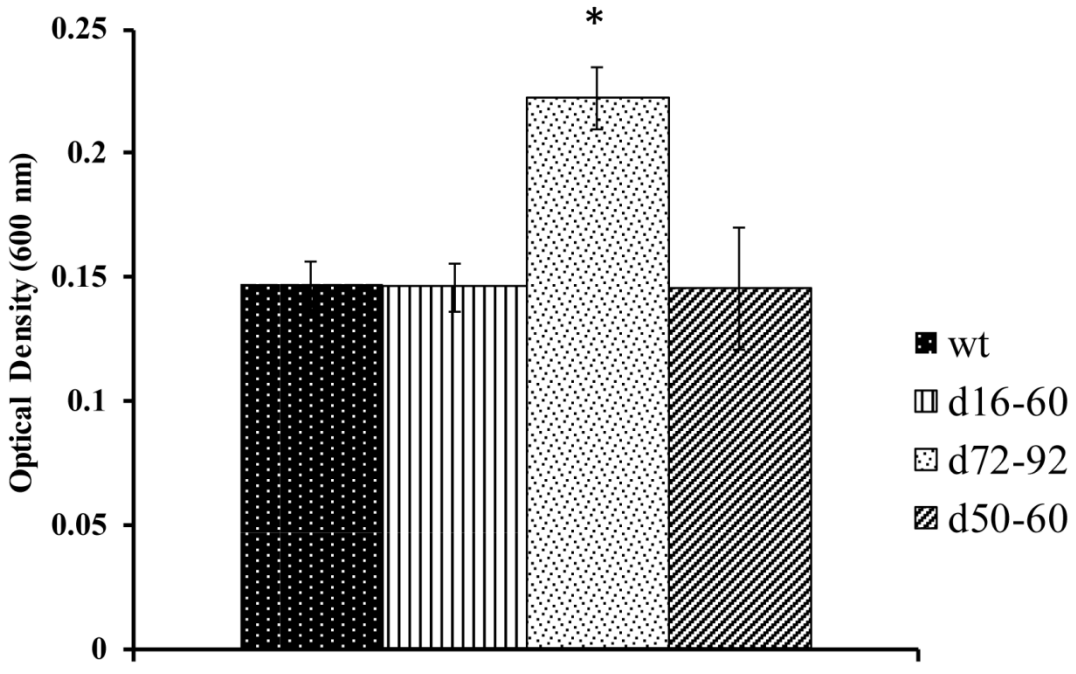


Figure 3.15: Graph showing optical density of crystal violet stained cells.
Bar graph showing significantly higher adhesion of d72-92 on a hydrophilic surface indicated by the asterisk.

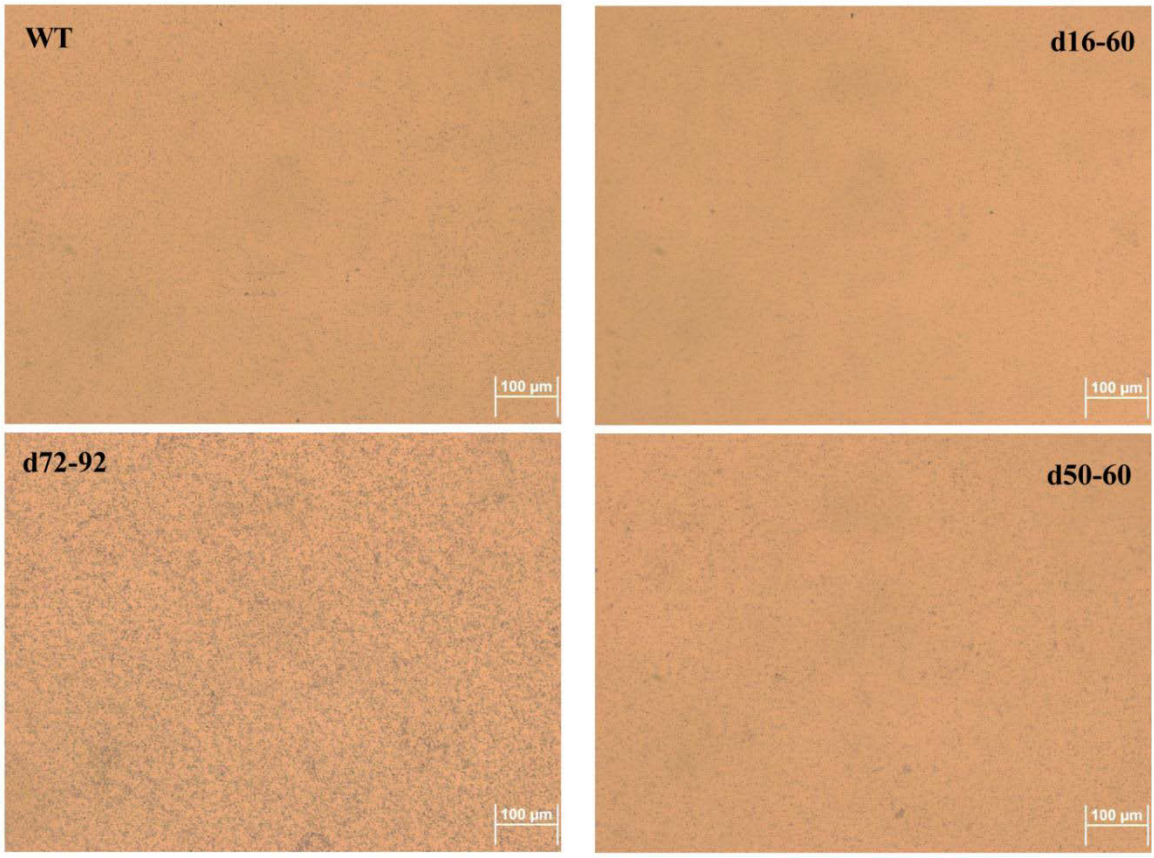


Figure 3.16: Microscopy of crystal violet stained cells.
Inverted phase-contrast microscopy of crystal violet stained cells adhered to a hydrophilic surface. Images taken using a 40X objective.

3.3: Discussion

As stated in chapter 1, the integron/gene cassette system is an ancient structure and the mobile gene cassette metagenome represents a vast reservoir of novel genes (Michael *et al.*, 2004;Gillings, 2005;Boucher *et al.*, 2007;Koenig *et al.*, 2008;Cambray *et al.*, 2010). While the majority of the predicted gene products have no known function, environmental studies strongly imply they are adaptive and where examined, structural and other studies support this concept (Nield, 2004;Robinson *et al.*, 2005;Robinson *et al.*, 2008). This chapter aimed at identifying how the chromosomal integron of DAT722 affects cellular physiology.

3.3.1: Gene cassette-associated products influence whole cell polysaccharide

During analyses, evidence emerged that the gene cassette deletions in wt DAT722 altered host polysaccharide. This was shown through congo red colony staining (Figure 3.12) and ^{1}H NMR analysis of whole cell purified polysaccharide (Figures 3.13 and 3.14). Congo red colony staining revealed differences in the wrinkling of colonies grown on media supplemented with congo red. This phenotype is known to be associated with changes in polysaccharide and thus provides evidence that gene cassettes are influencing polysaccharide structure(s) (Chen *et al.*, 2010;Colvin *et al.*, 2011). Gel electrophoresis of both loosely attached polysaccharide (EPS/CPS) and total polysaccharide (which includes LPS) gave further insight into differences between wt DAT722 and the deletion mutants polysaccharide moieties. Gel electrophoresis data showing differences between the banding profiles of wt DAT722 and d16-60 of loosely attached and whole cell polysaccharide provided further evidence that the contaminating substance during 2D-PAGE analysis was indeed polysaccharide in nature. To confirm that deletion of gene cassettes were altering polysaccharide structures, preliminary whole polysaccharide ^{1}H NMR analyses were performed. Changes between the wt DAT722 and both the d16-60 and d72-92 mutants, as well as dissimilarities between the two mutants themselves was identified in the carbohydrate ring proton region. This indicates that gene cassette-associated products are most likely influencing functional groups linked to the sugar component of polysaccharide structure(s). For example, gene cassette products could be adding/removing functional groups such as NH_2 or CH_3 groups. The biological ramifications for such changes are

substantial since they can affect processes such as bacterial-host interactions or virulence. For example, a recent study determined that the polysaccharide component of LPS plays a critical role in the colonisation of the light organ of squid species *Euprymna scolopes,* hence altering the composition of LPS components has the potential to impact on this colonisation (Post *et al.*, 2012). CPS is also widely known to be important in virulence including pandemic strains of *V. parahaemolyticus*, thus altering CPS moieties has the capacity to increase or decrease pathogen virulence (Chen *et al.*, 2010). Changes to surface polysaccharide are also likely to affect resistance to bacteriophage and biofilm formation (Labrie *et al.*, 2010). The data in this chapter showed that as a consequence of deleting 21 cassettes to produce mutant d72-92, biofilm formation to a plastic hydrophilic surface was significantly increased when compared to wt DAT722, d16-60 and d50-60 (Figures 3.15 and 3.16). This increased biofilm formation is likely due to the alterations observed in surface polysaccharide, particularly since surface polysaccharides are known to influence biofilm formation (Fong *et al.*, 2010).

At this stage it not known how gene cassette products are altering host polysaccharide as further chemical characterisation of host polysaccharide is required. However, it is intriguing to hypothesise that gene cassette products modify or decorate host polysaccharide through addition/removal of functional groups or sugars. A prior study had identified a gene cassette encoding an uncharacterised gene as important for CPS biosynthesis in a strain of *V. vulnificus* (Smith and Siebeling, 2003). In the *V. rotiferianus* DAT722 array, there are some gene cassettes that suggest a role for polysaccharide modification or biosynthesis (Appendix 1). For example, cassette 31 contains a gene that encodes a putative phosphorylated carbohydrate phosphatase protein. Briefly, a phosphatase is an enzyme which removes a phosphate group from its substrate. Further to this, cassette 78 encodes a putative maltose O-acetyltransferase. In addition to the putative maltose O-acetyltransferase, there are numerous acetytransferases present along the entire cassette array. Notably, five acetyltransferases span the region between cassettes 16-60 and four are present between cassettes 72-92.

Overall, there is significant biological importance to surface-associated polysaccharide and its modification. This is evident in the literature and includes biofilm formation, bacterial

cell co-aggregation, bacteriophage resistance, evasion of immune cells and resistance to antimicrobial peptides (Pier *et al.*, 2001;Scholl *et al.*, 2005;Westman *et al.*, 2008;Vu *et al.*, 2009;Lee *et al.*, 2013). Figure 3.17 demonstrates a proposed model detailing a mechanism derived from the studies arising from this thesis for how gene cassettes are influencing surface properties of the bacterial cell. In this model it is suggested that deletions, insertion and rearrangements of gene cassettes within the array trigger alterations to surface polysaccharide which in turn influence a number of phenotypes.

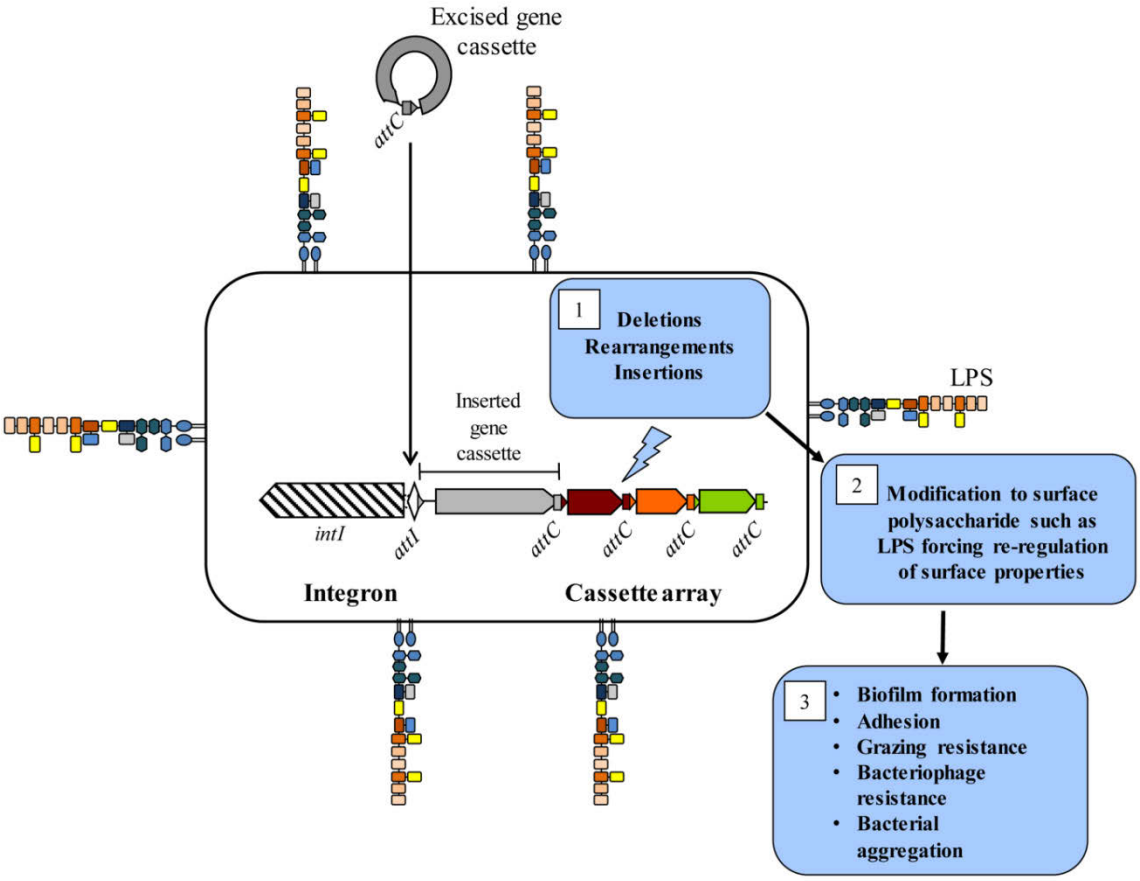


Figure 3.17: Proposed mechanism for the production of surface polysaccharide diversity through deletions, rearrangements and insertions in the cassette array. Deletion (as in the case of d16-60, d72-92 and d50-60), rearrangements and insertions in the cassette array (1) results in modification of cell surface polysaccharide (2) that may alter processes involved in biofilm formation, adhesion to different surfaces, grazing resistance, bacteriophage resistance and bacterial aggregation (3).

3.3.2: Further insight into the effect of a gene cassette deletion on cellular physiology from the 2D-PAGE analysis

Given the size of the gene cassette deletion in d16-60 (~1/3 of the array), only 0.5-1% of the total proteome was differentially expressed with fold changes at approximately the 2-fold range (maximum change 3.6 fold). These changes were consistent across two different media (complete and minimal) and two different growth phases (mid-logarithmic and stationary). Since 2D-PAGE detects only the most abundant proteins in the cell, it can be determined that the deletion has not affected the major metabolic pathways of the cells. However, it is possible that there is a higher degree of change to lower abundance proteins not detected using this methodology. Nevertheless, in conjunction with the secreteome analysis, these data indicate that deletion of cassettes 16-60 has not adversely impacted on the major cellular pathways of the cell in contrast to the previous study using mutant d8-60 (Labbate *et al.*, 2011). Furthermore, d16-60 and all other deletion mutants described in this thesis were not disadvantaged in environmental stress assays including oxidative stress, iron stress and cold stress. These data combined indicate that the majority of gene cassettes are maintained independently of host cell networks and are most likely not involved in stress adaptation.

Despite the limitations of 2D-PAGE in solubilising membrane associated proteins, various outer membrane proteins/porins were identified as being differentially expressed with statistical significance. This suggests that the products encoded by the cassette deletion between genes 16-60 affect the physiology of the outer membrane. Based on data as mentioned in section 3.3.1, the deletions appear to have affected surface polysaccharide with no major role in environmental stress survival and minor changes to overall protein expression as observed by 2D-PAGE and secretome analysis. In fact, some of the differentially expressed proteins observed in d16-60 are explained by changes to surface polysaccharide. Since polysaccharide changes would most likely alter the permeability of the cell and when considering the interconnected nature of cell envelope structures, re-regulation of general porins (e.g. OmpU; spot 2CM in Table 2) and outer membrane proteins (OmpA; spot 1CM in Table 2) probably occurred to compensate (Figure 3.16).

This change in permeability and subsequent re-regulation of porins may explain the minor variations in growth for all the deletion mutants.

For example, OmpA (significantly down-regulated in d16-60 cells grown to mid-logarithmic phase in complete media) is a major outer membrane protein in gram negative bacteria. Due to its presence in such high copy numbers, OmpA is a multifaceted protein that can function as a porin, biofilm formation, a receptor for several bacteriophage and is a target for antibiotics (Smith *et al.*, 2007). OmpA primarily functions as a porin and consequently acts as a permeability barrier for the bacterial cell, allowing nutrients and wastes to move in and out of the cell. A decrease in OmpA may result in less nutrients being able to enter the cell and waste to exit. This might offer an explanation for the minor variations observed in growth on differing carbon sources (Figure 3.1).

Similar to OmpA, OmpU and OmpT (co-migrated with cysteine synthase or acetyl coenzyme A) are major outer membrane porins found in members of the *Vibrio* genus and have been found to be general diffusion porins, commonly transporting hydrophilic solutes (Simonet *et al.*, 2003). OmpU and OmpT porins are both regulated by the ToxR transcriptional regulator in *V. cholerae* and *V. vulnificus*. Hence, it is possible that gene cassette products are 1) altering OmpU and OmpT expression levels, by regulating the ToxR homologue in DAT722 or most likely 2) the ToxR DAT722 homologue is responding to a modification in cell permeability due to alterations in surface polysaccharide.

Apart from outer membrane structures, some proteins of general metabolic function were differentially expressed between wt DAT722 and the d16-60 mutant such as: Nitrogen regulatory protein P-II (2CS) which was significantly down regulated in the d16-60 mutant. Nitrogen regulatory protein P-II is known to have many complex roles in nitrogen metabolism including, but not exclusively, uptake of nitrogenous compounds into the bacterial cell and their further catabolism (Arcondeguy *et al.*, 2001). Nitrogen is an essential element in bacteria and is necessary for the production of amino acid, nucleotides and amino sugars; which are all essential building blocks of bacterial cell structures. A down-regulation in nitrogen regulatory protein P-II may result in less uptake of nitrogenous

material into the cell and decrease catabolism of nitrogen resulting in an obstruction in the biosynthesis of essential nitrogenous building blocks or secretion of excess nitrogen wastes. It should be noted however, that this down-regulation of spot 2CS did not alter the growth rate of d16-60 in comparison to the wt DAT722, which means that either nitrogen regulatory protein P-II is not important for cell metabolism, or it was not down-regulated to the extent where physiological changes in growth could be observed in the complete and minimal media tested. However, this may not be in the case if cells are grown in a nitrogen deficient environment. Interestingly, deletion of the gene encoding nitrogen regulatory protein P-II in gram negative *E. coli* cells is not detrimental, indicating that it is not an essential regulatory protein (Arcondeguy *et al.*, 2001), which could provide explanation for why the down-regulation in wt DAT722 did not affect growth.

Spot 1CS (an unknown protein) was shown to be up-regulated in d16-60 cells grown in complete medium to stationary phase. The gene that encodes the unknown protein was found to be surrounded by genes that encode proteins involved in extracellular polysaccharide biosynthesis in wt DAT722 (Figure 3.3). The identification of the unknown protein further highlights that gene cassette products are likely affecting polysaccharide structures (Whitfield, 2006).

3.4: Future directions and conclusions

Using a number of techniques including molecular, protein, chemical and biological methods, this chapter has demonstrated that by deleting subsets of gene cassettes physiological changes occur which are localised to the bacterial cell surface. Namely, changes to surface polysaccharide and outer membrane structures such as porins and proteins were identified. It was confirmed using ^{1}H NMR that gene cassette-associated products are most likely influencing functional groups attached to the backbone of surface polysaccharides when cassettes 16-60 and 72-92 are removed respectively from wt DAT722. This chapter also revealed that when wt DAT722 and isogenic deletion mutants were allowed to form biofilms on a hydrophilic surface, d72-92 had significantly increased levels of biomass. This suggests a biological link between alteration to surface

polysaccharide and biofilm formation in this deletion mutant. This concept will be further explored in chapter 4 by expressing specific gene cassette associated proteins *in trans*.

Importantly, the data presented in this chapter has answered important questions as well as advanced the integron biology field. Gene cassette-associated products are highly novel and at this stage of research an attempt to identify cassette related phenotypes is challenging and relies on random physiological traits to test. By determining that surface polysaccharide is a point of change in these deletion mutants, future research can be aimed at determining how gene cassettes are indeed modifying surface polysaccharide moieties. Interestingly, indel (insertion and deletions) events are common in the cassette array of vibrios, indicating that they have evolutionary and adaptive potential (Boucher *et al.*, 2007). For the first time the results presented in this thesis have shown how this can affect *Vibrio* physiology by showing that cassette deletions largely affected surface polysaccharide. Thus, future experiments should be aimed at elucidating the structure of surface polysaccharides; namely EPS/CPS and LPS. A number of chemical techniques can be used to do this. For example, capillary electrophoresis could be performed in order to identify the carbohydrate composition of the polysaccharide structures present in *V. rotiferianus* DAT722 (Kabir, 1982). From preliminary H^1 NMR analysis carried out on whole cell surface polysaccharide (EPS, CPS and LPS) it was deduced that gene cassette-associated products were manipulating functional groups attached to the carbohydrate backbone of the polysaccharide. Further chemical analysis of functional groups such as amine, phosphate, methyl and acyl groups would also clarify the specific changes to surface polysaccharide as a consequence of gene cassette deletions. To further enhance our understanding of the surface topography of wt DAT722 and how deletion of gene cassettes are effecting surface properties of the cell, lectin-binding assays would be beneficial. Lectins are carbohydrate binding proteins or glycoproteins that are specific to certain sugar moieties (e.g. concanvalin A) (Gildemeister *et al.*, 1994). Thus, by investigating the different binding affinities of various lectins; the structure and also binding capacity/ability of wt DAT722 surface polysaccharides will be revealed. This can then be compared to the lectin binding capacity of the gene cassette deletion mutants to decipher specific difference within the polysaccharide structure.

Some protein spots in section 3.2.3.1 were identified as co-migrating when 2D-PAGE analysis was performed. Co-migration results when proteins have a similar isoelectric point (charge) and molecular size, appearing as one spot on a polyacrylamide gel. Co-migration of proteins was a clear limitation of the 2D-PAGE studies performed in this thesis. Co-migration may be tackled using a variety of methods such as separating samples on a narrow pH gradient (4-7 used in this study see section 2.5.4.1.1) in order to resolve closely distributed spots. Another method is to utilise SRM (selection reaction monitoring). SRM targets a predetermined set of peptides and gives a quantitative differential analysis between two samples (Lange *et al.*, 2008). Thus, SRM is a technique that can quantitatively measure the differences in protein concentrations between the wt DAT722 and deletion mutant samples, potentially eliminating co-migrating proteins that are not truly up/down regulated as a direct consequence of deleting gene cassettes.

Chapter 4: Expression of foreign integron-associated gene cassettes: the identification of a novel gene cassette phenotype

4.1: Introduction

As described in chapter 3, proteomic, chemical and biological assays showed that gene cassette-associated products influence cell surface structures, namely polysaccharide in *V. rotiferianus* DAT722. It was shown using NMR that gene cassettes are decorating surface polysaccharide by altering functional groups attached to polysaccharide.

In order to elucidate how individual gene cassettes may impact on phenotypes associated with surface polysaccharide, three cassettes were chosen to complement the gene cassette deletion mutants; d16-60, d72-92 and d650-60 of *V. rotiferianus* DAT722. These gene cassettes were selected based on their putative bioinformatic identification as potentially being involved in carbohydrate modification. Thus this chapter aims to *identify phenotypes associated with single gene cassette expression.*

To address the aim of this chapter, the following specific objectives were examined:

1. Clone three individual gene cassettes and induce expression in a *Vibrio* host
2. Examine how expression of gene cassettes affect growth in a *Vibrio* host
3. Examine how expression of gene cassettes in a *Vibrio* host affect biofilm formation on two different surfaces
4. Examine how expression of gene cassettes affect congo red binding in a *Vibrio* host

4.2: Results

4.2.1: Bioinformatic analysis of three gene cassettes

Prior to cloning and expression, the entire 116 gene cassette array was examined (Appendix 1) using BLAST (section 2.10 for method) to identify any cassette associated genes putatively involved in polysaccharide biosynthesis. Three cassettes; 31, 54 and 78 were chosen for cloning (see section 2.3.10 for method) and expression into deletion mutants d16-60, d50-60 and d72-72 respectively. These cassettes are named VSD31, VSD54 and VSD78 (**V**ibrio **S**pecies **D**AT722) for the remainder of this thesis. The cassette 31 amino acid sequence was identified as being a putative phosphorylated carbohydrate phosphatase (Figure 4.1).

Cassette 31: Phosphorylated carbohydrate phosphatase

A

MENKLKDYEVYLFDMDGTLVNSEPLKGKALALACADYGAQVDHNIYKDVMGESWQVVTGHFFT
HAGISPDLGEFNRYFRAHYELMLNDELELNAGAKAYIEHLKKAGKKCGVVSSAATWMVENILT
SLQLETAFDLVITQEHVTKHKPDPEAYNLALANLDVSPEHTIVFEDSTAGISAGKSSGCDVIA
VRHEFNGKNDLSGALRSISVYDEMFV

B

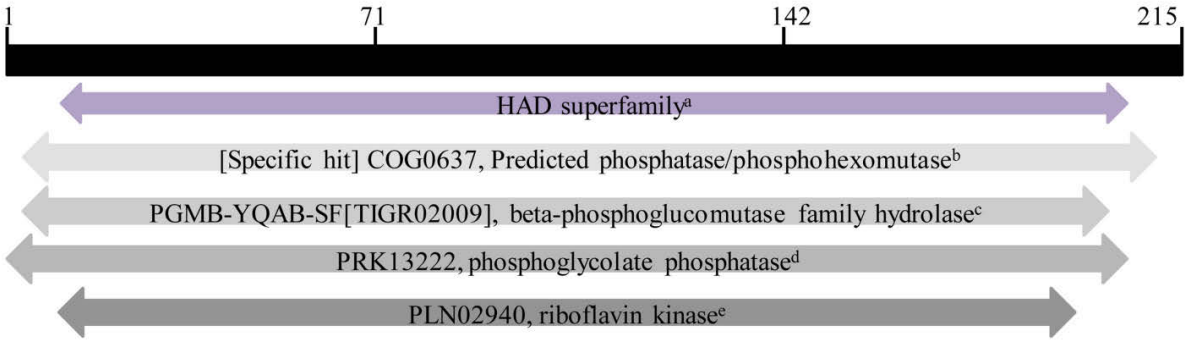


Figure 4.1: Putative identification for protein encoded by VSD31.
Panel (A) shows the 165 amino acid sequence of the protein encoded by cassette 31 (that is the protein sequence of the cassette in the 31st position away from *attI*). Panel (B) highlights relative domains of the protein and their functions. Shown in purple is the superfamily domain of the protein and in grey the multi-domains of the protein. [a]Haloacid dehalogenase-like hydrolases. The haloacid dehalogenase-like (HAD) superfamily includes L-2-haloacid dehalogenase, epoxide hydrolase, phosphoserine phosphatase, phosphomannomutase, phosphoglycolate phosphatase, P-type ATPase, and many others, all of which use a nucleophilic aspartate in their phosphoryl transfer reaction. All members possess a highly conserved alpha/beta core domain, and many also possess a small cap domain, the fold and function of which is variable, [b]Predicted phosphatase/phosphohexomutase, [c]Beta-phosphoglucomutase family hydrolase, [d]Phosphoglycolate phosphatase, [e]Riboflavin kinase. Date searched: 2nd April, 2014.

The gene in cassette 54 (VSD54) was identified as a putative acetyltransferase (Figure 4.2). The *V. rotiferianus* DAT722 gene cassette array has ten putatively identified genes with acetyltransferase activity. This identification is broad: acetyltransferases, as the name implies, are enzymes that transfer acetyl groups. However, exactly where and to what motif an acetyl group is being transferred is unknown in this case. In an attempt to determine the role the acetyltransferases play within the cell, VSD54 was cloned and complemented into deletion mutant d50-60. The protein encoded by VSD54 has a predicted amino acid length of 165. Figure 4.2 shows the protein domains of VSD54.

Cassette 54: Acetyltransferase

A

MKIEVHEISPDEKELIQNLLQFYEYEFSIYEEDDVDENGDFEIADVDEY
FEIREYTPLLVRVSGQPAGFVIVNSDPSAENGRYLIEEFFIMKRFQEKG
VGREVAGQVFDMFGQDWVVRVISENLKGQSFWEKVISSYTSGRFLTEVF
NDESWEGPIYTFSKPPKI

B

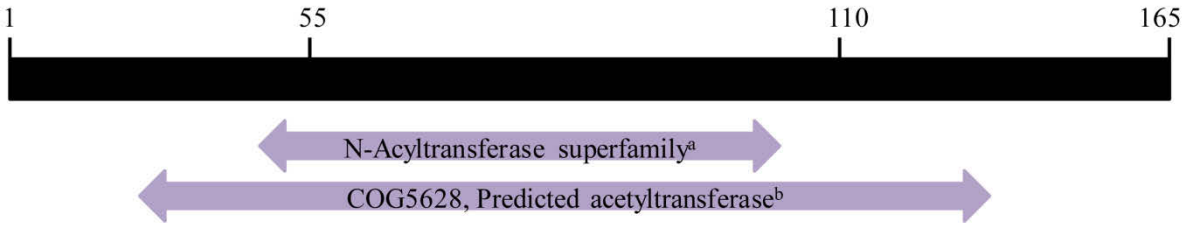


Figure 4.2: Putative identification for protein encoded by VSD54.
Panel (A) shows the amino acid sequence of cassette 54 (that is the cassette in the 54th position away from *attI)*. Panel (B) highlights relative domains of the protein and their functions. Shown in purple are the superfamily domains of the protein. [a]N-Acyltransferase superfamily: Various enzymes that characteristically catalyse the transfer of an acyl group to a substrate, [b]Predicted acetyltransferase superfamily. Date searched: 2nd April, 2014.

Lastly, the gene in cassette 78 (VSD78) was identified to be a putative maltose O-acetyltransferase (Figure 4.3). This gene was cloned and complemented into deletion mutant d72-92 and the protein encoded by this gene has a predicted amino acid length of 195.

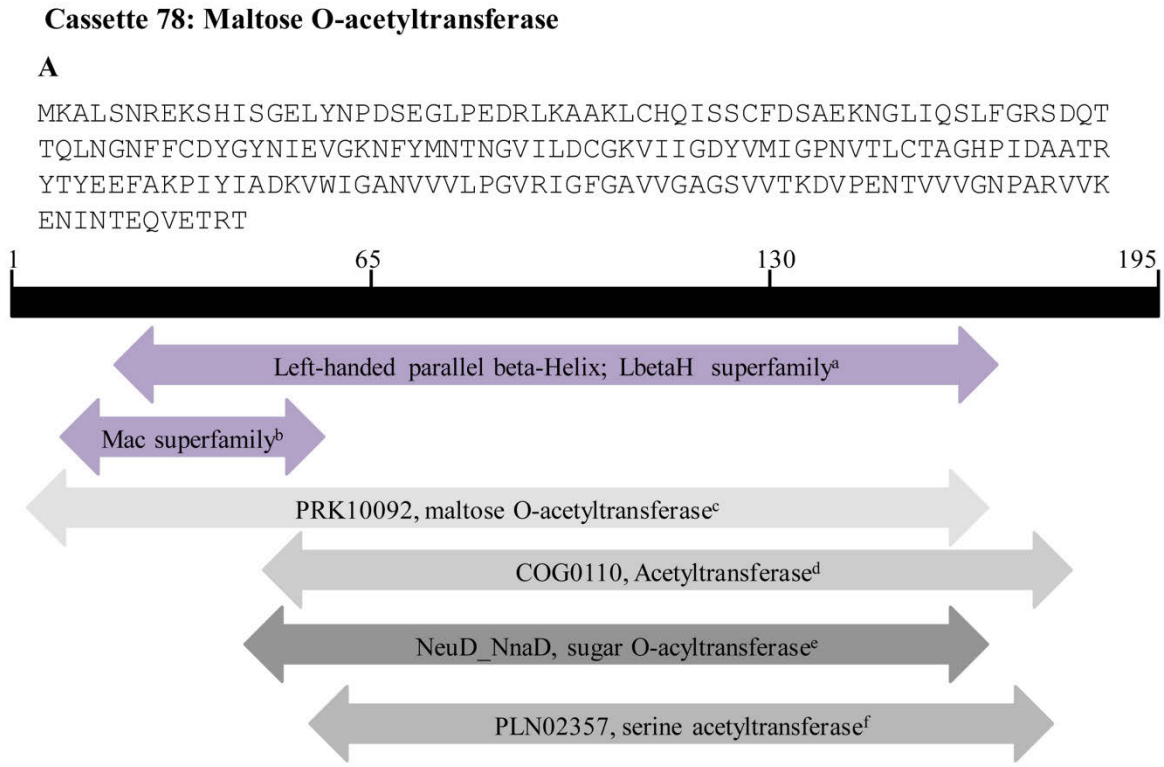


Figure 4.3: Putative identification for protein encoded by VSD78.
Panel (A) shows the amino acid sequence of cassette 78 (that is the cassette in the 78th position away from *attI)*. Panel (B) highlights relative domains of the protein and their functions. Shown in purple are the superfamily domains of the protein and in grey the multi-domains of the protein. [a]Left-handed parallel beta-Helix (LbetaH or LbH) superfamily domain: Proteins containing hexapeptide repeats are often enzymes showing acyltransferase activity, however, some subfamilies in this hierarchy also show activities related to ion transport or translation initiation, [b]Maltose acetyltransferase superfamily; this domain family is found in bacteria, archaea and eukaryotes, and is approximately 50 amino acids in length. Mac uses acetyl-CoA as acetyl donor to acetylated cytoplasmic maltose, [c]PRK10092, maltose O-acetyltransferase, [d]COG0110, Acetyltransferase (isoleucine patch superfamily), [e]IGR03570, sugar O-acyltransferase, sialic acid O-acetyltransferase NeuD family; this family of proteins includes the characterised NeuD sialic acid O-acetyltransferase enzymes from *E. coli* and *Streptococcus agalactiae* (group B strep). The [f]PLN02357, serine acetyltransferase. Date searched: 2nd April, 2014.

4.2.2: Cloning of gene cassettes into an IPTG-inducible expression vector results in amino acid substitution within gene cassette-encoded proteins

Initially VSD31 and VSD78 were cloned into the IPTG-inducible vector pJAK16, transformed into the *E. coli* strain XL-1-Blue and then conjugated into *V. rotiferianus* DAT722 deletion mutants d16-60 and d72-92 respectively using tri-parental conjugation (see section 2.3.9.1 for method). After conjugation into each of the deletion mutants d16-60 and d72-92, gene cassettes were amplified *via* PCR and sequenced (section 2.3.1 for method). Interestingly, a single point mutation was identified resulting in an amino acid change in the protein(s) encoded by the cloned gene cassettes (Figures 4.4 and 4.5). These single point mutations were discovered following conjugation into the DAT722 deletion strains. It is likely that these point mutations in VSD31 and VSD78 occurred in *E. coli* XL-1-Blue due to residual expression of the gene cassettes resulting in nucleotide changes that are more favourable for growth in the *E. coli* XL-1-Blue strain. It is thus less likely that the mutations occurred after conjugation into the DAT722 deletion mutants as these genes were derived from wt DAT722, therefore are more adapted to this host in comparison to *E. coli.*

Interestingly, the point mutations that occurred in the nucleotide sequences of both VSD31 and VSD78 resulted in a change to an arginine amino acid. Arginine is a positively charged amino acid and curiously in the VSD31 mutation there was a switch from serine (neutral and polar) and for VSD78 from a cysteine (also neutral and polar). Thus, the point mutation that occurred in both VSD31 and VSD78 resulted in the change of a neutral amino acid to one that was positively charged. Each mutation therefore, has the potential to alter the secondary folding of the protein. Further to this, removing a cysteine from the translated amino acid sequence of VSD78 may impact on disulfide bonding. Disulfide bonding plays an integral role in the stabilisation of the protein folding process and consequently is of interest in studies related to structural and functional properties of specific proteins (Ceroni *et al.*, 2006).

A >cloned VSD31
MENKLKDYEVYLFDMDGTLVNSEPLKGKALALACADYGAQVDHNIYKDVMGESWQVVTGH
FFTHAGISPDLGEFNRYFRAHYELMLNDELELNAGAKAYIEHLKKAGKKCGVVSSAATWM
VENILTSLQLETAFDLVITQEHVTKHKPDPEAYNLALANLDVSPEHTIVFEDSTAGISAG
KSRGCDVIAVRHEFNGKNDLSGALRSISVYDEMFV

B >VSD31
MENKLKDYEVYLFDMDGTLVNSEPLKGKALALACADYGAQVDHNIYKDVMGESWQVVTGH
FFTHAGISPDLGEFNRYFRAHYELMLNDELELNAGAKAYIEHLKKAGKKCGVVSSAATWM
VENILTSLQLETAFDLVITQEHVTKHKPDPEAYNLALANLDVSPEHTIVFEDSTAGISAG
KSSGCDVIAVRHEFNGKNDLSGALRSISVYDEMFV

Figure 4.4: Translated protein sequence of cassette 31 shows an amino acid substitution
Panel (A) shows the amino acid sequence encoded by gene cassette 31 after it had been conjugated from *E. coli* XL-1-Blue into deletion mutant d16-60. Highlighted in yellow at position 182 is the amino acid arginine (R). Panel B shows the amino acid sequence encoded by the wt cassette 31. Highlighted in yellow is the amino acid serine (S). This amino acid substitution at position 182 results in the substitution of a serine to an arginine.

A >cloned VSD78

MKALSNREKSHISGELYNPDSEGLPEDRLKAAKLRHQISSCFDSAEKNGLIQSLFGRSDQ
TTQLNGNFFCDYGYNIEVGKNFYMNTNGVILDCGKVIIGDYVMIGPNVTLCTAGHPIDAA
TRYTYEEFAKPIYIADKVWIGANVVVLPGVRIGFGAVVGAGSVVTKDVPENTVVVGNPAG
VVKENINTEQVETRT

B >VSD78

MKALSNREKSHISGELYNPDSEGLPEDRLKAAKLCHQISSCFDSAEKNGLIQSLFGRSDQ
TTQLNGNFFCDYGYNIEVGKNFYMNTNGVILDCGKVIIGDYVMIGPNVTLCTAGHPIDAA
TRYTYEEFAKPIYIADKVWIGANVVVLPGVRIGFGAVVGAGSVVTKDVPENTVVVGNPAR
VVKENINTEQVETRT

Figure 4.5: Translated protein sequence of cassette 78 shows an amino acid substitution
Panel (A) shows the amino acid sequence encoded by gene cassette 78 after it had been conjugated from *E. coli* XL-1-Blue into deletion mutant d72-92. Highlighted in yellow at position 35 is the amino acid arginine (R). Panel B shows the amino acid sequence encoded by the wt cassette 78. Highlighted in yellow is the amino acid cysteine (C). This aminio acid substitution at position 35 results in the substitution of a cysteine to an arginine.

4.2.3: Does cloning gene cassettes downstream of an arabinose-inducible promoter avoid point mutations?

The presence of the *tac* promoter (*lac* operon) on pJAK16 means that cloned genes are expressed in the presence of IPTG. It has been known for many years that the *lac* operon is subject to an 'all or none' expression response (Khlebnikov *et al.*, 2000). This 'all or none' concept arises because the transporter for the inducer (e.g. IPTG) is under control of the inducer itself. When the threshold level of the inducer accumulates inside the cell, expression of the operon is driven (Khlebnikov *et al.*, 2000). Unfortunately, expression systems are 'leaky' meaning that there can be some residual expression even when there is no inducer present provided that the repressor itself is not present to turn off expression. As the P_{BAD} expression system has successfully been used to express toxic genes in *E. coli* and various other *Vibrio* spp. such as *V. cholerae* (Le Roux *et al.*, 2007;Abuaita and Withey, 2009;Houot *et al.*, 2010;Miyata *et al.*, 2013) this expression system was adopted in order to clone genes without a point mutation.

The genes contained in gene cassettes VSD31, VSD54 and VSD78 were cloned in front of an arabinose inducible promoter into the pSU-pBAD vector and transformed into *E. coli* WM3064 using 1% glucose to repress the P_{BAD} promoter (see section 2.3.10.2 for method). Thes contsructs were then conjugated into deletion mutants d16-60, d50-60 and d72-92 respectively. When the gene cassettes were amplified *via* PCR and sequenced, the amino acid sequence of the protein encoded by the gene cassette was identitical to wt, unlike the results obtained in section 4.2.2. Figures 4.6, 4.7 and 4.8 show the amino acid sequence encoded by the cloned VSD31, VSD54 and VSD78 genes aligned with the amino acid sequence from wt DAT722

```
cloned          MENKLKDYEVYLFDMDGTLVNSEPLKGKALALACADYGAQVDHNIYKDVMGESWQVVTGH 60
VSD31           MENKLKDYEVYLFDMDGTLVNSEPLKGKALALACADYGAQVDHNIYKDVMGESWQVVTGH 60
                ************************************************************

cloned          FFTHAGISPDLGEFNRYFRAHYELMLNDELELNAGAKAYIEHLKKAGKKCGVVSSAATWM 120
VSD31           FFTHAGISPDLGEFNRYFRAHYELMLNDELELNAGAKAYIEHLKKAGKKCGVVSSAATWM 120
                ************************************************************

cloned          VENILTSLQLETAFDLVITQEHVTKHKPDPEAYNLALANLDVSPEHTIVFEDSTAGISAG 180
VSD31           VENILTSLQLETAFDLVITQEHVTKHKPDPEAYNLALANLDVSPEHTIVFEDSTAGISAG 180
                ************************************************************

cloned          KSSGCDVIAVRHEFNGKNDLSGALRSISVYDEMFV 215
VSD31           KSSGCDVIAVRHEFNGKNDLSGALRSISVYDEMFV 215
                *************************************
```

Figure 4.6: Protein sequence encoded by cloned cassette 31 without any amino acid substitution.
From this alignment of the protein sequence encoded by the cloned VSD31 into pSU-pBAD (upper amino acid sequence) and the protein sequence encoded by wt VSD31 (lower amino acid sequence) it is evident that there are no point mutations in the DNA sequence that will lead to a change in amino acid sequence.

```
cloned          MKIEVHEISPDEKELIQNLLQFYEYEFSIYEEDDVDENGDFEIADVDEYFEIREYTPLLV 60
VSD54           MKIEVHEISPDEKELIQNLLQFYEYEFSIYEEDDVDENGDFEIADVDEYFEIREYTPLLV 60
                ************************************************************

cloned          RVSGQPAGFVIVNSDPSAENGRYLIEEFFIMKRFQEKGVGREVAGQVFDMFGQDWVVRVI 120
VSD54           RVSGQPAGFVIVNSDPSAENGRYLIEEFFIMKRFQEKGVGREVAGQVFDMFGQDWVVRVI 120
                ************************************************************

cloned          SENLKGQSFWEKVISSYTSGRFLTEVFNDESWEGPIYTFSKPPKI 165
VSD54           SENLKGQSFWEKVISSYTSGRFLTEVFNDESWEGPIYTFSKPPKI 165
                *********************************************
```

Figure 4.7: Protein sequence encoded by cloned cassette 54 without any amino acid susbstitution.

From this alignment of the protein sequence encoded by the cloned VSD54 into pSU-pBAD (upper amino acid sequence) and the protein sequence encoded by wt VSD54 (lower amino acid sequence) it is evident that there are no point mutations in the DNA sequence that will lead to a change in amino acid sequence.

```
cloned          MKALSNREKSHISGELYNPDSEGLPEDRLKAAKLCHQISSCFDSAEKNGLIQSLFGRSDQ 60
VSD78           MKALSNREKSHISGELYNPDSEGLPEDRLKAAKLCHQISSCFDSAEKNGLIQSLFGRSDQ 60
                ************************************************************

cloned          TTQLNGNFFCDYGYNIEVGKNFYMNTNGVILDCGKVIIGDYVMIGPNVTLCTAGHPIDAA 120
VSD78           TTQLNGNFFCDYGYNIEVGKNFYMNTNGVILDCGKVIIGDYVMIGPNVTLCTAGHPIDAA 120
                ************************************************************

cloned          TRYTYEEFAKPIYIADKVWIGANVVVLPGVRIGFGAVVGAGSVVTKDVPENTVVVGNPAR 180
VSD78           TRYTYEEFAKPIYIADKVWIGANVVVLPGVRIGFGAVVGAGSVVTKDVPENTVVVGNPAR 180
                ************************************************************

cloned          VVKENINTEQVETRT 195
VSD78           VVKENINTEQVETRT 195
                ***************
```

Figure 4.8: Protein sequence encoded by cloned cassette 78 without any amino acid substitution.
From this alignment of the protein sequence encoded by the cloned VSD78 into pSU-pBAD (upper amino acid sequence) and the protein sequence encoded by wt VSD78 (lower amino acid sequence) it is evident that there are no point mutations in the DNA sequence that will lead to a change in amino acid sequence.

4.2.4: Does expression of gene cassettes affect growth?

Growth curves were carried out in complete (LB20) and minimal media (2M + glucose) using complemented *V. rotiferianus* DAT722 deletion mutants now designated d16-60/pSU-pBAD::VSD31, d50-60/pSU-pBAD::VSD54 and d72-92/pSU-pBAD::VSD78. Growth curves were performed in order to determine whether expressed genes in these cassettes impact on growth. Intriguingly, even though cloning was successful and there were no point mutations present in any of the three gene cassettes, the vector (without any cloned gene) altered the growth rate of cells with (Figure 4.9A) and without the addition of arabinose to drive expression (Figure 4.9B) in LB20 in d50-60. The effect on growth due to the vector without (Figure 4.10A) and with (Figure 4.10B) the addition of arabinose was also apparent when cells were grown in 2M + glucose in d72-92 and d50-60.

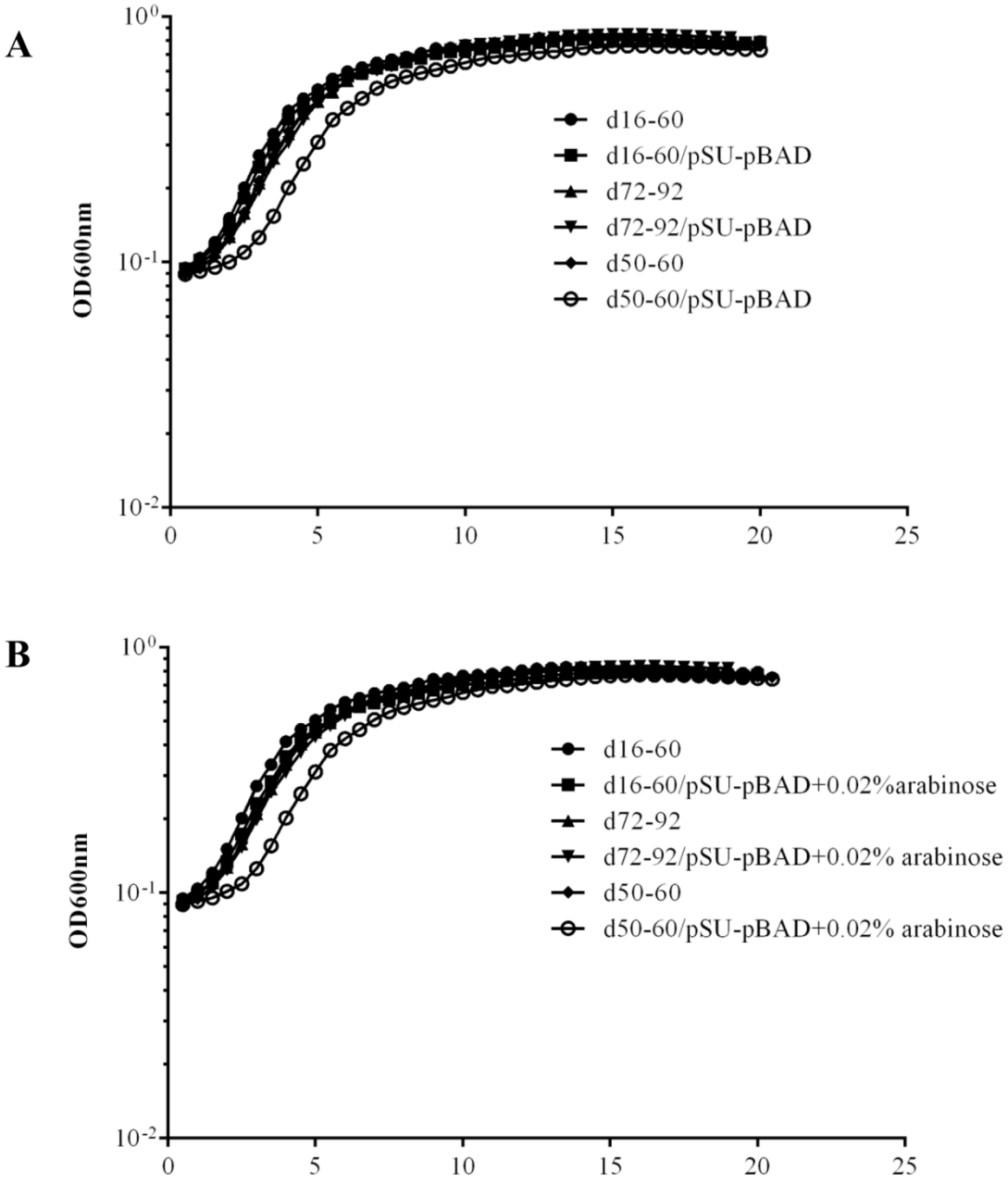


Figure 4.9: Growth curve of deletion mutants with pSU-pBAD in LB20.
Panel (A) shows growth curves of deletion mutants and deletion mutants with pSU-pBAD only as a control without the addition of arabinose. (B) shows the growth of deletion mutants and deletion mutants with pSU-pBAD only as a control with the addition of arabinose. The y-axis shows optical density readings at 600 nm absorbance and the x-axis is given in time (hr).

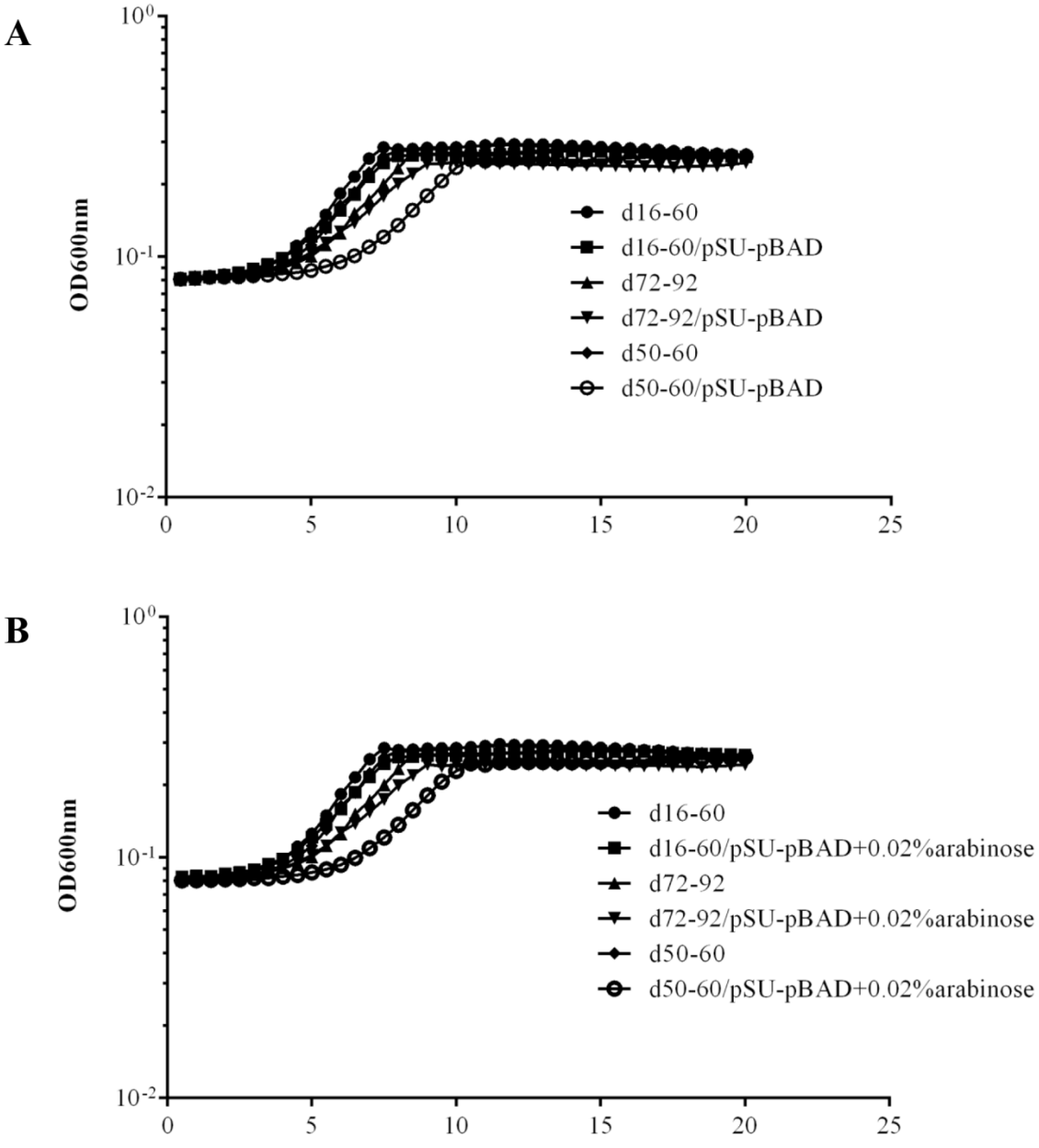


Figure 4.10: Growth curve of deletion mutants with pSU-pBAD in 2M+glucose. Panel (A) shows growth curves of deletion mutants and deletion mutants with pSU-pBAD only as a control without the addition of arabinose. (B) shows the growth of deletion mutants and deletion mutants with pSU-pBAD only as a control with the addition of arabinose. The y-axis shows optical density readings at 600 nm absorbance and the x-axis is given in time (hr).

4.2.5: Expression of *V. rotiferianus* DAT722 gene cassettes in *V. cholerae* S24

4.2.5.1: Do non-native gene cassettes affect growth of V. cholerae S24?

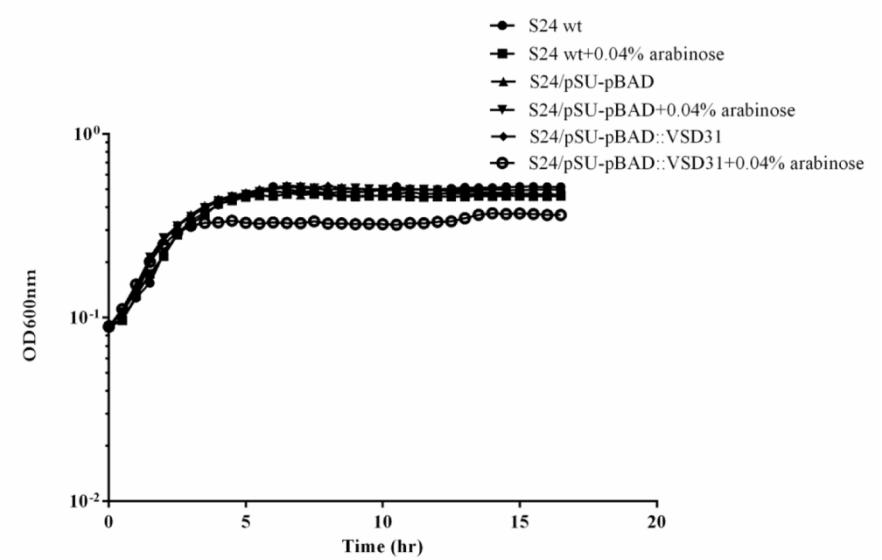
As described in section 4.2.4 the pSU-pBAD vector appears to impact on the growth of d50-60 when grown in LB20 and 2M + glucose. The P_{BAD} expression system has been used to successfully clone and express genes in various in *V. cholerae* strains (Le Roux *et al.*, 2007;Abuaita and Withey, 2009;Houot *et al.*, 2010;Miyata *et al.*, 2013). Thus, each cloned cassette VSD31, VSD78 and VSD54 carried on pSU-pBAD was conjugated (see section 2.3.9.1 for method) into an environmental non-O1/O139 strain of *V. cholerae* strain S24 (see chapter 5 for a detailed description of this strain) not containing gene homologues of VSD31, VSD78 or VSD54. These strains are designated *V. cholerae* S24/pSU-pBAD::VSD31, *V. cholerae* S24/pSU-pBAD::VSD78, *V. cholerae* S24/pSU-pBAD::VSD54 and wild type *V. cholerae* S24 will be referred to as wt S24 for the remainder of this thesis.

Under all media conditions tested (Figures 4.11 and 4.12) the pSU-pBAD vector did not impact on the growth on *V. cholerae* S24 in contrast to what was observed for some of the *V. rotiferianus* DAT722 deletion mutants (section 4.2.4).

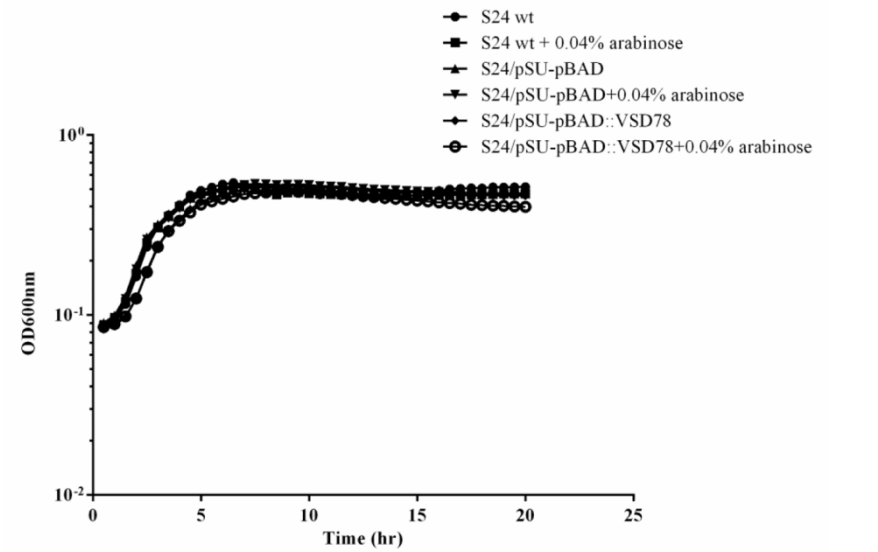
Figure 4.11A also shows that expression of VSD31 appeared to impact the growth of wt S24 during entry into stationary phase. Upon closer examination however, it was evident that the decrease in OD was actually due to the aggregation of cells expressing VSD31. Hence, the decrease in optical density was a consequence of cell clumping and not the direct result of a change in growth rate. Interestingly, aggregation of cells only occurred upon growth with shaking at 200 rpm and never when grown statically (data now shown). This result was always apparent when expressing VSD31 in wt S24 cells that were shaken on a rotary shaker at 200 rpm. Cells did not aggregate when VSD78 and VSD54 were expressed in wt S24 cells grown in LB5.

When VSD31, VSD78 and VSD54 are expressed in wt S24 cells grown in minimal salts medium 2M + glucose as the sole carbon source growth was unaffected. Furthermore, no cell aggregation occurred as was observed in wt S24 cells expressing VSD31 (Figure 4.12).

A


S24 wt
S24 wt+0.04% arabinose
S24/pSU-pBAD
S24/pSU-pBAD+0.04% arabinose
S24/pSU-pBAD::VSD31
S24/pSU-pBAD::VSD31+0.04% arabinose
OD600nm
Time (hr)

B


S24 wt
S24 wt + 0.04% arabinose
S24/pSU-pBAD
S24/pSU-pBAD+0.04% arabinose
S24/pSU-pBAD::VSD78
S24/pSU-pBAD::VSD78+0.04% arabinose
OD600nm
Time (hr)

C

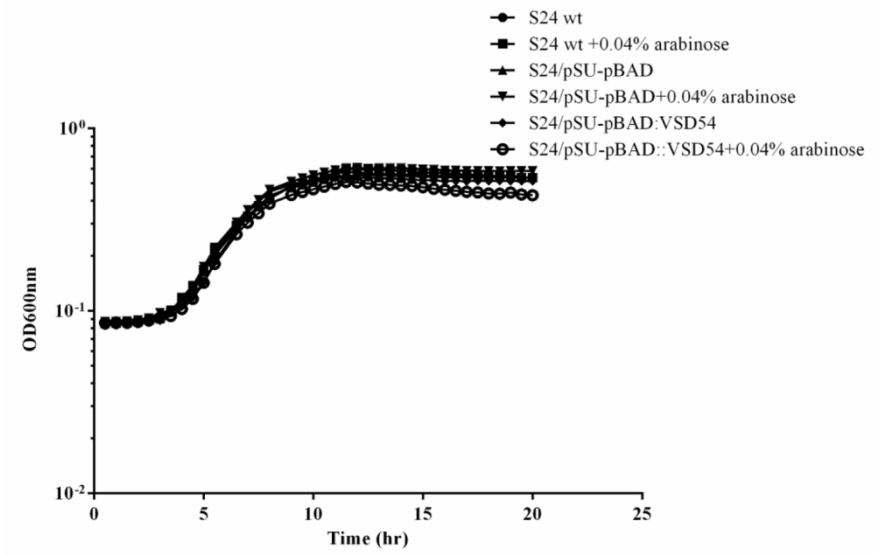

S24 wt
S24 wt +0.04% arabinose
S24/pSU-pBAD
S24/pSU-pBAD+0.04% arabinose
S24/pSU-pBAD:VSD54
S24/pSU-pBAD::VSD54+0.04% arabinose
OD600nm
Time (hr)

Figure 4.11: Effect of VSD31, VSD78 and VSD54 on growth of *V. cholerae* S24 in LB5. Panel (A) shows the effect of expressing VSD31 in wt S24. (B) shows the effect of expressing VSD78 in wt S24 and (C) shows the effect of VSD54 expressed in wt S24. It can be seen that expression of VSD31 with 0.04% arabinose impedes the growth of wt S24 cells. However, both VSD78 and VSD54 do not affect growth of wt S24.

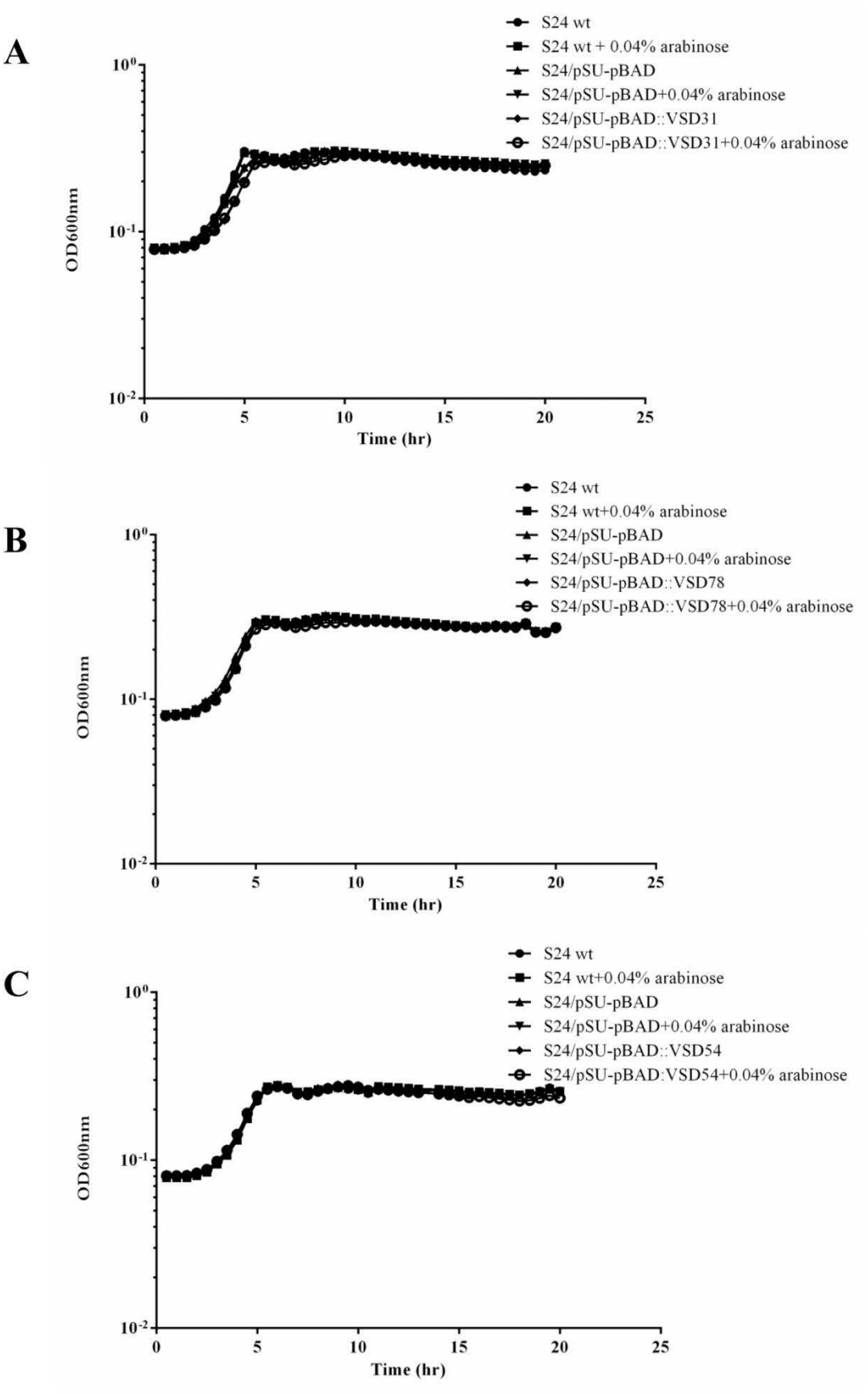

A
S24 wt
S24 wt + 0.04% arabinose
S24/pSU-pBAD
S24/pSU-pBAD+0.04% arabinose
S24/pSU-pBAD::VSD31
S24/pSU-pBAD::VSD31+0.04% arabinose
OD600nm
Time (hr)
B
S24 wt
S24 wt+0.04% arabinose
S24/pSU-pBAD
S24/pSU-pBAD+0.04% arabinose
S24/pSU-pBAD::VSD78
S24/pSU-pBAD::VSD78+0.04% arabinose
OD600nm
Time (hr)
C
S24 wt
S24 wt+0.04% arabinose
S24/pSU-pBAD
S24/pSU-pBAD+0.04% arabinose
S24/pSU-pBAD::VSD54
S24/pSU-pBAD:VSD54+0.04% arabinose
OD600nm
Time (hr)

Figure 4.12: Effect of VSD31, VSD78 and VSD54 on growth of *V. cholerae* S24 in 2M + glucose.
Panel (A) shows the effect of expressing VSD31 in wt S24. (B) shows the effect of expressing VSD78 in wt S24 and (C) shows the effect of VSD54 expressed in wt S24. It can be seen that in 2M+glucose VSD31, VSD78 and VSD54 have no impact on the growth of wt S24.

4.2.5.2: Expression of cassette 31 causes cell aggregation in V. cholerae S24

The pSU-pBAD vector does not adversely impact on cell growth of *V. cholerae* S24 as it did in the *V. rotiferianus* DAT722 deletion mutants. Also, in section 4.2.5.1 it was apparent that in LB5 when VSD31 was expressed by the addition of 0.04% arabinose cell aggregation was observed. The aggregation of these cells was further examined using phase contrast microscopy (see section 2.7.2 for method).

Figure 4.13B shows images of cell aggregation at 100X of wt S24 cells when VSD31 expression is induced using 0.04% arabinose. This figure shows a representative image of cell clumping. When a single field of view is examined single cells are difficult to identify, with clumped cells occupying the vast majority of the view. In contrast when arabinose is removed from LB5 media and cells containing pSU-pBAD::VSD31 are grown there is no aggregation, as seen in Figure 4.13A. To account for any clumping caused by the addition of arabinose itself, images were captured when wt S24 cells were grown in the presence of 0.04% arabinose. Figure 4.13C shows a representative image of wt S24 cells when grown in LB5 + 0.04% arabinose with no cell aggregation present. Likewise, when 0.04% arabinose is added to wt S24 cells containing pSU-pBAD only in LB5 no cell aggregation is observed (Figure 4.13D), ruling out any aggregation affects as a consequence of the vector alone.

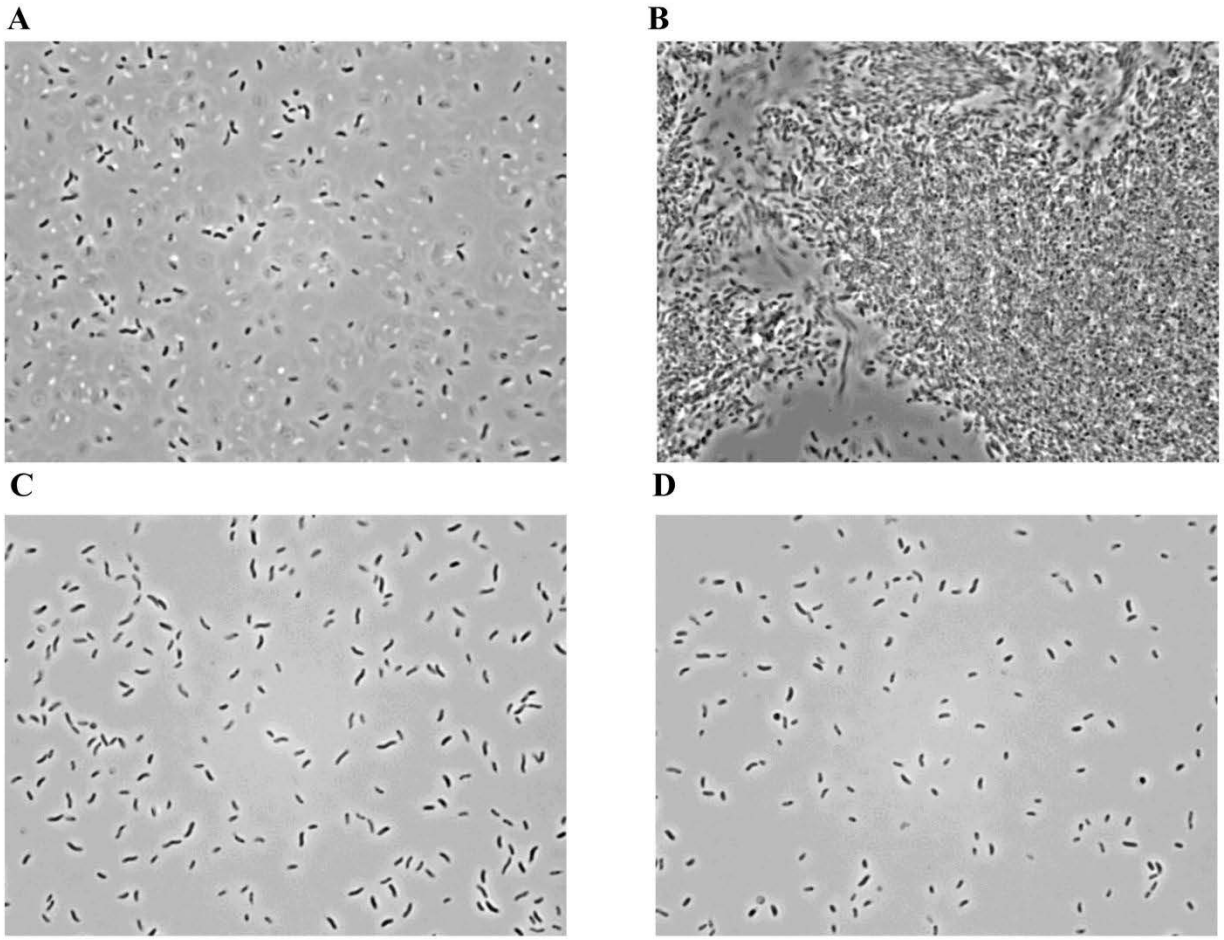


Figure 4.13: Phase contrast microscopy of cell aggregation due to expression of VSD31.

Cell aggregation was examined under phase contrast microscopy using a Zeiss Axioplan 2 fluorescence microscope. Panel (A) shows wt S24/pSU-pBAD::VSD31 cells without addition of arabinose, (B) wt S24/pSU-pBAD::VSD31 with the addition of 0.04% arabinose, (C) wt S24 with the addition of 0.04% arabinose and (D) wt S24/pSU-pBAD with the addition of 0.04% arabinose. Images were examined at 100X.

4.2.5.3: Expression of gene cassettes affects congo red binding capacity of V. cholerae S24 cells

Chapter 3 results showed that cassette associated genes are modifying surface polysaccharide. Congo red staining was also performed in chapter 3 as it is a common dye used to identify bacterial phenotypes associated with polysaccharide structures such as LPS and EPS/CPS. Congo red binding and liquid culture aggregation of bacterial cells are both two important phenotypes that have been shown to be associated with changes to polysaccharide (Colvin *et al.*, 2011). As it has already been demonstrated in this chapter that expression of VSD31 in wt S24 cells causes them to aggregate (a phenotype often associated with surface polysaccharide), thus congo red liquid culture binding assays were carried out to confirm whether expression of the cloned gene cassettes affected congo red binding. Figure 4.14 shows that when pSU-pBAD::VSD31 was expressed with the addition of 0.04% arabinose, cells are able to bind more congo red which is indicated by a decrease in optical density at 495nm (see section 2.6.5.2 for method). Interestingly, this decrease in optical density was also observed when pSU-pBAD::VSD54 was expressed indicating alterations to polysaccharide due to expression of this cassette unrelated to bacterial aggregation (not evident when cells express VSD54). This decrease in optical density is a result of the cell's ability to bind more congo red dye.

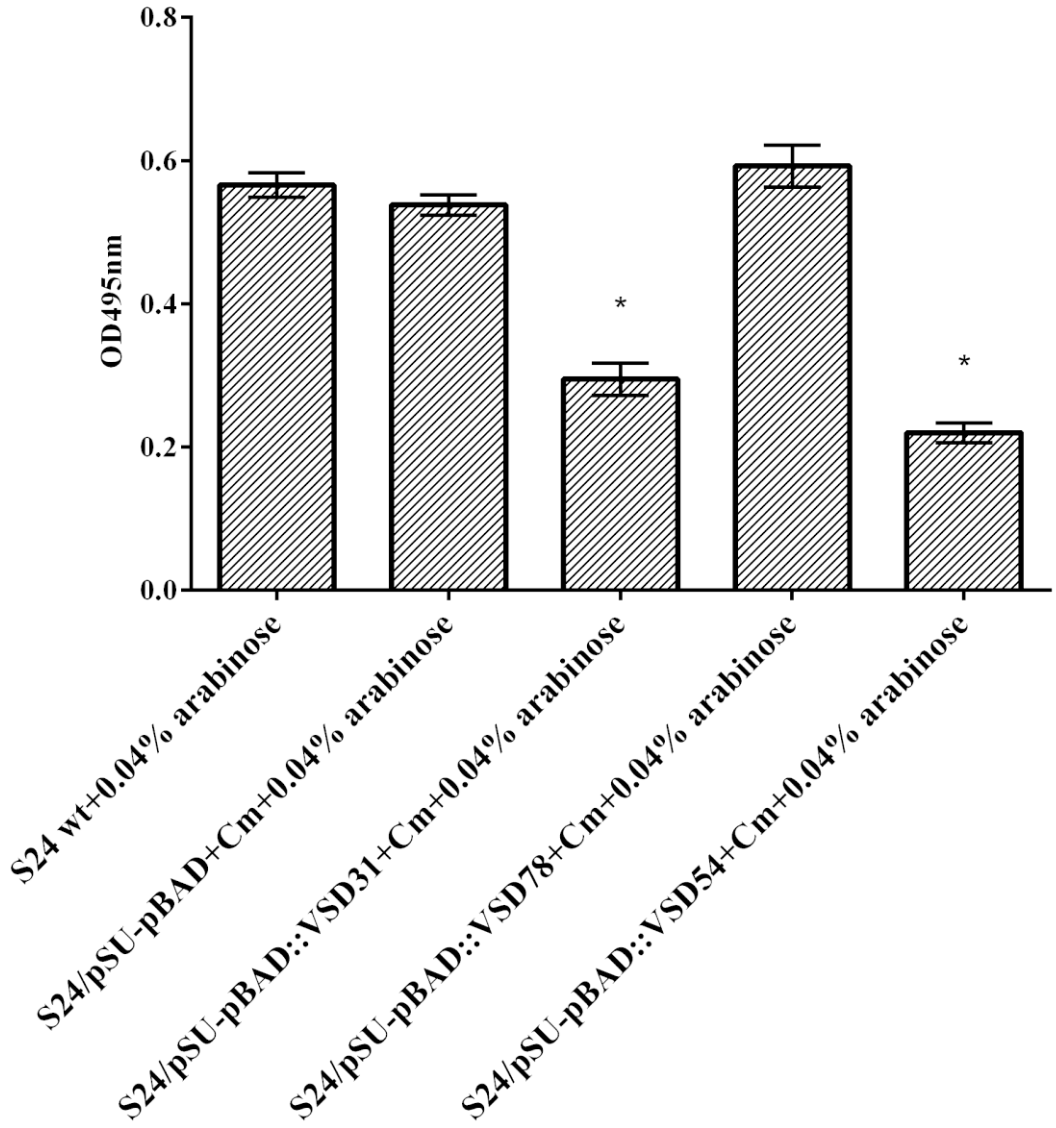


Figure 4.14: Congo red liquid culture binding.
Cells grown in LB5 broth with the addition of 40 μg/mL of congo red in the presence of 0.04% arabinose. S24/pSU-pBAD::VSD31 and S24/pSU-pBAD::VSD54 have a decreased absorbance at 495nm relative to wt S24 and S24/pSU-pBAD. Asterisks indicate congo red binding of statistically significant difference. Error bars are given as standard error of the mean (SEM) and readings were carried out in triplicate.

4.2.5.4: Bacterial cell aggregation alters biofilm formation on hydrophilic and hydrophobic surfaces

Bacterial polysaccharide is known to play an integral role in biofilm formation (Joseph and Wright, 2004;Colvin *et al.*, 2011). Having observed aggregation as a result of cassette 31 expression and increased congo red binding as a result of expression of genes in both cassettes 31 and 54. Thus, it was of interest to observe if expression of all gene cassettes altered biofilm formation using simple batch biofilm assays (see section 2.9 for method).

Figure 4.15 illustrates that when VSD31 is expressed in wt S24 cells grown on a hydrophilic surface, there is a significant increase in crystal violet staining compared to uninduced cells containing VSD31. This result similar when VSD31 expressed in wt S24 biofilms are grown on hydrophobic surface (Figure 4.16). However, on a hydrophobic surface, *V. cholerae* S24/pSU-pBAD::VSD31 cells form less biomass overall when compared to biofilms grown on a hydrophilic surface in LB5 + 0.04% arabinose.

When *V. cholerae* S24/pSU-pBAD::VSD54 cells are grown in the presence of 0.04% arabinose on a hydrophilic surface (Figure 4.15) it was also noted that there was a minor difference in biofilm formation when compared to uninduced *V. cholerae* S24/pSU-pBAD::VSD54 cells. These data correlate with the congo red results obtained in section 4.2.5.3 where both *V. cholerae* S24/pSU-pBAD::VSD31 and *V. cholerae* S24/pSU-pBAD::VSD54 cells induced with 0.04% arabinose increase cell binding of congo red.

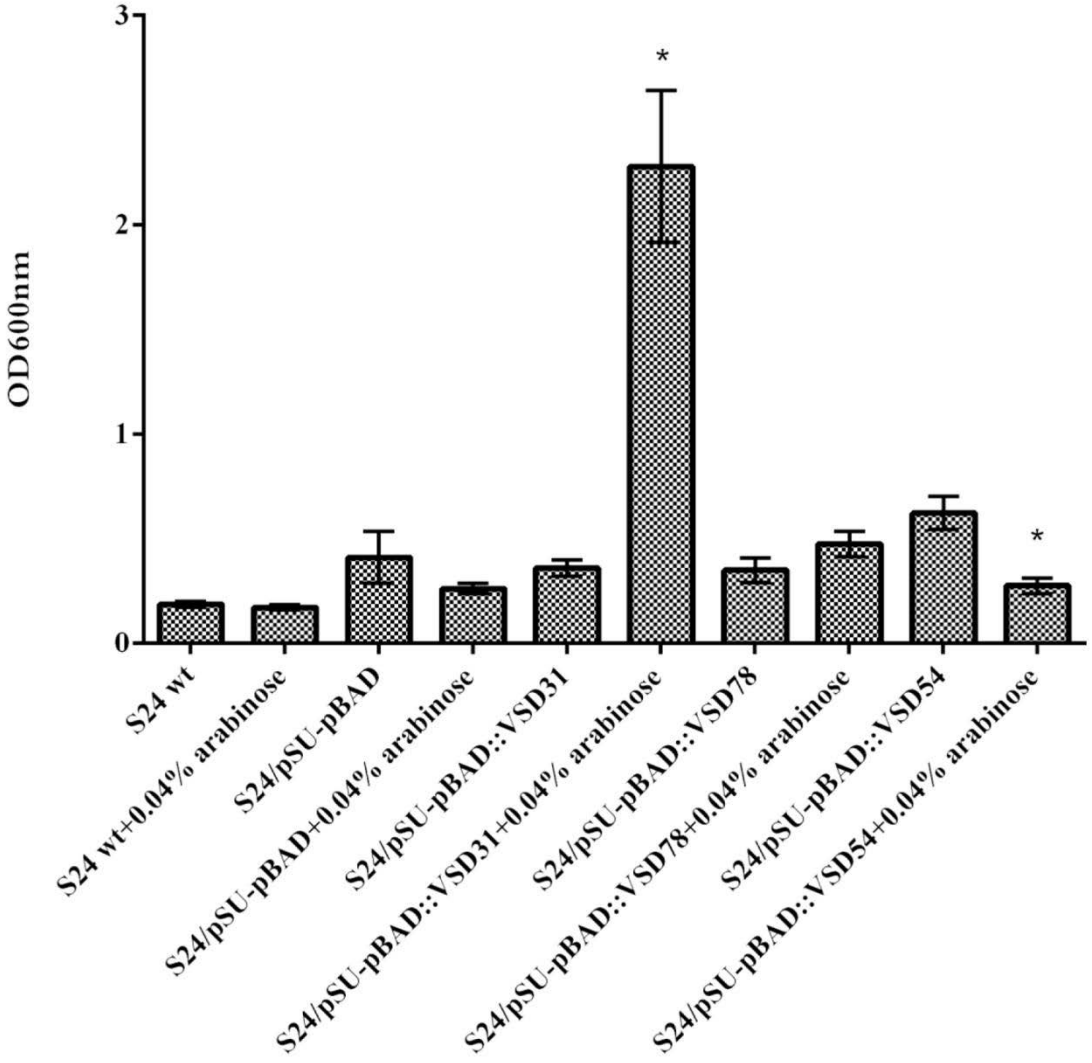


Figure 4.15: Biofilms grown on a hydrophilic surface.
Biofilms grown in LB5 with and without addition of 0.04% arabinose to induce expression of gene cassettes carried on pSU-pBAD on a hydrophilic surface. Cells were grown for 24 hr in a 24-well microtitre plate and optical density was measured at 600nm post crystal violet staining of attached biomass. Experiment was performed in triplicate and standard error of the mean (SEM) expressed as error bars. Statistically different biomass measurements are indicated by an asterisk.

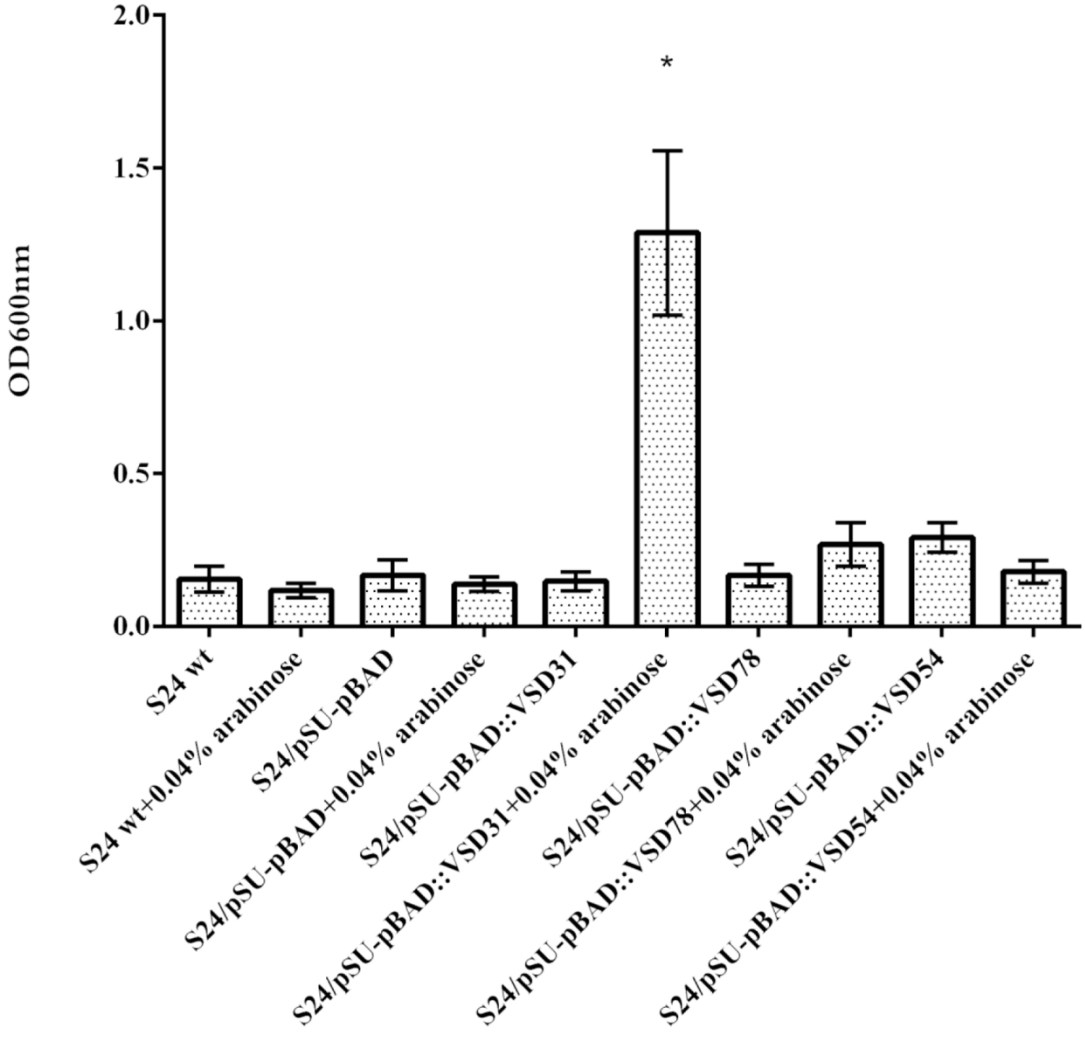


Figure 4.16: Biofilms grown on a hydrophobic surface.
Biofilms grown in LB5 with and without addition of 0.04% arabinose to induce expression of gene cassettes carried on pSU-pBAD on a hydrophobic surface. Cells were grown for 24 hr in a 24-well microtitre plate and optical density was measured at 600nm post crystal violet staining of attached biomass. Experiment was performed in triplicate and standard error of the mean (SEM) expressed as error bars. Statistically different biomass measurements are indicated by an asterisk.

4.3: Discussion

As stated in chapter 3, approximately 80% of gene cassette-associated products encode proteins of no identifiable biochemical function. Those that have been attributed a function on a whole are mainly classified into broad categories e.g. an acetyltransferase. This chapter aimed to identify phenotypes associated with the expression of three single gene cassettes from the *V. rotiferianus* DAT722 integron gene cassette array. Three specific gene cassettes were chosen to examine what impact they would have on vibrio cell physiology. These gene cassettes were VSD31, VSD78 and VSD54 encoding a putative phosphorylated carbohydrate phosphatase, acetyltransferase and maltose O-acetyltransferase protein respectively.

4.3.1: Plasmid expression systems in vibrios

It was noted from Figures 4.9 and 4.10 that the pSU-pBAD vector appears to affect the growth of deletion mutants d72-92 and d50-60 in LB20 and 2M + glucose. The effects seen on growth due to the vector with or without the addition of arabinose can be explained by two possible scenarios: 1) the *araC* gene product that is responsible for regulation of the P_{BAD} operon may be affecting chromosomal genes in *V. rotiferianus* DAT722 and 2) the presence of the pSU-pBAD vector itself could be altering growth of cells even in the absence of arabinose.

As the P_{BAD} expression system has successfully been used to clone and express toxic genes in *E. coli* and various *Vibrio* spp. such as *V. cholerae* (Le Roux *et al.*, 2007;Abuaita and Withey, 2009;Houot *et al.*, 2010;Miyata *et al.*, 2013) pSU-pBAD::VSD31, pSU-pBAD::VSD78 and pSU-pBAD::VSD54 were conjugated into the environmental *V. cholerae* strain S24. In this strain the pSU-pBAD vector did not have a negative impact on growth. Similarly, when VSD31, VSD78 and VSD54 were expressed in wt S24 cells they did not impact on growth of cells.

4.3.2: New insight into how gene cassette-associated products influence bacterial cell physiology and virulence

Upon observation of the growth medium (LB5) containing wt S24 cells expressing VSD31, it was apparent that the decrease in optical density was an artefact of cell aggregation (see Figure 4.13). Cassette 31 (VSD31) was identified using BLAST to be a putative phosphorylated carbohydrate phosphatase. Phosphatases are a group of enzymes that remove phosphate groups from a substrate; in this case a carbohydrate/polysaccharide molecule. Phosphate ions are negatively charged and thus removing a phosphate group from the polysaccharide molecule may alter the overall charge of the polysaccharide structure. This alteration in the charge of the polysaccharide could account for cell aggregation and the increase in biofilm formation when VSD31 is expressed. Alternatively, modification of sugars on the polysaccharide by the phosphatase may be facilitating aggregation *via* a lectin-sugar interaction. Further to this in *V. cholerae* there is a phosphotransferase cascade system that transfers phosphates to carbohydrate/polysaccharide molecules. Thus, cassette 31 may be regulating expression of genes that result in polysaccharide changes. Interestingly, the cascade system in *V. cholerae* has been shown to regulate biofilm formation on abiotic surfaces (Houot *et al.*, 2010). Acetylation of the O-antigen (part of LPS) is also known to be important for biofilm formation in *Pseudomonas aeruginosa* (Nivens *et al.*, 2001). When VSD54 was expressed in wt S24, cells were are able to bind more congo red and formed less biofilm on a hydrophilic surface. Although VSD54 was identified to be a putative acetyltransferase it is not known what substrate the acetyl functional group is being transferred to. The increased congo red binding and decreased biofilm formation data both suggest that this acetyltransferase gene cassette is acetylating polysaccharide moieties. These phenotypes observed (changes in biofilm formation, cell aggregation and increased congo red binding) with expression of single gene cassettes are all associated with the transfer of functional groups to and from substrates, which correlates with the NMR data presented in chapter 3. NMR studies suggested that gene cassettes are decorating surface polysaccharide *via* the addition/removal of functional groups.

4.3.2.1: How might gene cassette-induced cell aggregation influence cholera outbreaks?

The spread of cholera during epidemics/pandemics is generally *via* water contaminated by human faecal matter containing toxigenic (*ctxA*) *V. cholerae* (Bari *et al.*, 2013). Many factors that enhance the waterborne spread of cholera epidemics are unclear, and detection of toxigenic *V. cholerae* can be challenging (Faruque *et al.*, 2006). In regions where cholera is endemic, such as Bangladesh, viable *V. cholerae* is readily detected throughout the course of the epidemic. However, it is the inter-epidemic period where isolation of *V. cholerae* becomes problematic. This is thought to be due to predation by lytic bacteriophage as mentioned in chapter 1 (Faruque *et al.*, 2005a). It is also known that during this inter-epidemic period, *V. cholerae* can remain in a state know as a conditionally viable environmental cell (CVEC) or more commonly known as viable but nonculturable (VBNC) (Faruque *et al.*, 2006;Bari *et al.*, 2013). Bacteria in the CVEC state are dormant and evade isolation *via* routine microbiological culturing methods. An interesting study showed that cells of pathogenic *V. cholerae* in the CVEC state that avoid typical laboratory cultivation techniques actually exist as surface aggregates of 'dormant' cells (Faruque *et al.*, 2006). Even more remarkable is that the authors demonstrated that these cells in CVEC were resuscitated as complete virulent bacteria by inoculating water into rabbit intestines (Faruque *et al.*, 2006). Interestingly, the authors also demonstrate that both aggregated and planktonic forms of pathogenic *V. cholerae* were present in patients presenting with cholera and that these aggregate-forms of *V. cholerae* show enhanced infectivity. Subsequently, it was suggested that these aggregates then disperse in the human intestine yielding a high infectious dose, accounting for the enhanced infectivity (Faruque *et al.*, 2006). In this chapter, results show that VSD31 (from *V. rotiferianus* DAT722) when expressed in the environmental *V. cholerae* strain S24 allowed large aggregates to form. Although wt S24 is not a strain of *V. cholerae* capable of causing outbreaks of cholera, this chapter highlights how a single mobile gene acquired by LGT might affect the virulence of pathogenic *V. cholerae*.

4.4: Future directions and conclusions

In this chapter it was demonstrated that VSD31 from *V. rotiferianus* DAT722 when cloned and expressed on pSU-pBAD in *V. cholerae* S24 significantly modified the phenotypic behaviour of this bacterium. Specifically, by mediating the induction of cell aggregation/clumping, affecting biofilm formation and increasing congo red binding. These data suggest that the physiological changes are related to modifications to polysaccharide (also given that VSD31 is a putative phosphorylated carbohydrate phosphatase). These two phenotypes: aggregation and biofilm formation have not been previously reported to be linked to expression of a gene cassette-associated product.

The induction of cell aggregation and increased biofilm formation on hydrophilic and hydrophobic surfaces due to the expression of a single gene cassette has potential to increase infectivity of pathogenic strains. Hence, studies of VSD31 in a clinically important toxigenic strain of *V. cholerae* that is capable of causing outbreaks of the disease cholera, and observing the aggregation phenotype, might elucidate the importance of this gene cassette in an infection setting. In addition to this, the LPS structure of *V. cholerae* has been chemically resolved. It would be interesting to perform NMR on the toxigenic strain containing the gene cassettes in order to localise where modifications to the polysaccharide chemical structure are occurring. Further to this, performing batch biofilm assays using strains with expressed VSD31, VSD78 and VSD54 on surfaces involved in propagating *V. cholerae* infections such as chitin surfaces and intestinal epithelial cells may reveal more information about the role these cassettes play in *Vibrio* physiology

Chapter 5: Characterisation and function of a genomic island inserted into the chromosome of non-O1/O139 environmental *Vibrio cholerae* strain S24

5.1: Introduction

The results presented in chapter 5 have been published in the journal of *Environmental Microbiology* (see Appendix 3; (Rapa *et al.*, 2014)). As described in chapter 1, mobile genetic elements (MGE) have been pivotal in the adaptation and evolution of vibrios. Chapters 3 and 4 detailed how one such MGE, the integron/gene cassette system, can influence adaptation in vibrios by cassette encoded proteins leading to modifications to surface polysaccharide and the subsequent impact of these changes on associated bacterial phenotypes such as biofilm formation and cell aggregation. In this chapter of this thesis I will focus on a novel genomic island (GI) and characterise its importance in adaptation and evolution of *V. cholerae*. GIs are defined as large chromosomal regions that have features suggestive of recent LGT (Boyd *et al.*, 2009). Many GIs have the capacity to excise and form circular intermediates and often target tRNA loci for their integration. In *V. cholerae,* the causative agent of cholera, GIs have been implicated in causing human disease and enhancing environmental survival. For example, the replacement of the O1 Classical biotype by the O1 El Tor biotype in the 1960s is suggested to be due to the acquisition of VSP-1 and VSP-2 (described in chapter 1) that have most likely enhanced epidemic spread (Faruque *et al.*, 2003). Moreover, VPI-1 (*Vibrio* Pathogenicity Island) codes for important virulence genes such as: the toxin-coregulated pilus (TCP), which is an essential colonisation factor for the *V. cholerae* bacterium, as well as the accessory colonisation factor (ACF) and virulence regulators ToxT and TcpPH (Murphy and Boyd, 2008). Non-O1/O139 *V. cholerae* strains are considered to be a major source of laterally acquired DNA for cholera-causing strains (Meibom *et al.*, 2005) and thus a deeper understanding of the diverse genetic elements present in *V. cholerae* strains is important for understanding and alleviating the emergence of new pathogenic strains of *V. cholerae.* Thus the overall aim of this chapter is to *characterise a novel genomic island* present in a non-O1/O139 environmental strain of *V. cholerae* designated strain S24. This strain was isolated from the

Georges River in Sydney, Australia (Islam *et al.*, 2013) as part of a study of *V. cholerae* strains indigenous to a cholera non-endemic region. Aspects of this study, not involving S24, were published in 2013 (Islam *et al.*, 2013).

In bacteria, errors in DNA can occur as part of normal DNA replication or damage can be induced by external stimuli, including but not exclusive to: UV-radiation, antibiotics and the host immune system (Janion, 2008). During DNA repair, the SOS response first induces genes involved in error-free DNA repair including base excision repair (BER), nucleotide excision repair (NER) and recombinational DNA repair (Rattray and Strathern, 2003;Janion, 2008). Mismatch repair (MMR) is an example of a more stringent DNA repair; mediated by at least nine different proteins (Polosina and Cupples, 2010;Lenhart *et al.*, 2012). However, if DNA damage is extensive the mutagenic phase of the SOS response is triggered, known as SOS mutagenesis (Goodman, 2002). This phase is mediated by DNA polymerases that replicate past template lesions in a process called translesion DNA synthesis (TLS) that is inherently error-prone (Goodman, 2002). In particular, DNA polymerase V, encoded by the *umuDC* operon is largely responsible for the ~100-fold increase in DNA damage-induced chromosomal mutations in *Escherichia coli* (Patel *et al.*, 2010).

Briefly, the *recA* gene is an essential housekeeping gene and is known to have four main functions in *E. coli*: 1) it catalyses DNA strand exchange between ssDNA (damaged DNA) and dsDNA; this is known as recombinational DNA repair (involved in homologous recombination), 2) it induces the SOS response by autocatalytic cleavage of the LexA repressor 3) it activates UmuD *via* cleavage to its active form UmuD' and thus 4) it induces the SOS mutagenesis pathway by the activation of DNA polymerase V (Schlacher *et al.*, 2006;Jiang *et al.*, 2009). UmuD' undergoes a conformational change and forms a complex with UmuC giving UmuD'$_2$C (Patel *et al.*, 2010). When maximally expressed there are only about 60 UmuD'$_2$C molecules present in the cell (Sommer *et al.*, 1998) implying that this operon system is induced by the cell when stress levels are extreme.

The novel genomic island isolated from wt S24 was identified to house a number of genes putatively involved in DNA repair such as *recA* and *umuDC* and will be further summaried in section 5.2.1 of this chapter.

To address the aim of this chapter, the following specific objectives were examined:

1. Use bioinformatics to characterise genes present on the genomic island
2. Identify whether the genomic island excises from the S24 chromosome
3. Identify if the genomic island preferentially inserts into the *recA* gene
4. Perform phylogenetic analysis on the *recA* gene carried on the genomic island to decipher genetic relationships
5. Determine whether the genomic island enhances protection against DNA damaging agents including UV-radiation and antibiotics in *E. coli*
6. Use transposon mutagenesis to test the functionality of two genes present on the genomic island: *recA* and *umuC* in *E. coli*

5.2: Results

5.2.1: Identification of a novel genomic island in *V. cholerae* S24 containing *recA* and other DNA repair genes

S24 is an environmental, non-O1/O139 *V. cholerae* strain isolated from the Georges River in Sydney, Australia with other *V. cholerae* strains from a previous study (Islam *et al.*, 2013). During a preliminary phylogenetic analysis of *V. cholerae* S24 (performed by Atiqul Islam), which involved amplification of the housekeeping genes *adk*, *gyrB*, *mdh* and *recA,* the host *recA,* here designated $recA_{S24}$, was found to generate a predominant product of ~1.5 kb and a less dominant product of ~1.1 kb, instead of the expected ~850 bp. To confirm this intriguing result, the experiment was repeated for this thesis and is shown in Figure 5.1.

After confirming the size of the $recA_{S24}$ amplicon, the sequence of this region was further interrogated by generating a draft genome sequence for S24 (to be released at a later date; see section 2.3.5). Analysis of the S24 genome revealed that recA_{S24} had been disrupted and was present on two separate contigs (contiguous, overlapping DNA sequence reads resulting from the reassembly of the small DNA fragments generated during sequencing). The two contigs were pieced together by PCR and joined to an intervening third contig to produce a contig of 262,869-bp (Figure 5.2). Analysis of the region disrupting recA_{S24} (indigenous *recA* gene in S24) identified a putative GI of 32,787 bp. Consistent with being a GI, the GC content is 41.3% in comparison to the rest of the genome at 47.2%. On either side of the genomic island, 9 bp inverted repeats were identified designated IR_R (for ***r****ecA* end) and IR_i (for **i**ntegrase end) (Figure 5.3). Bioinformatic analysis of the GI identified 23 ORFs (Figure 5.3) including a complete copy of *recA,* designated $recA_{RME}$, at the IR_R end and a phage integrase at the IR_i end. Because of the presence of *recA*, the element has been designated: *recA* mobile element (RME).

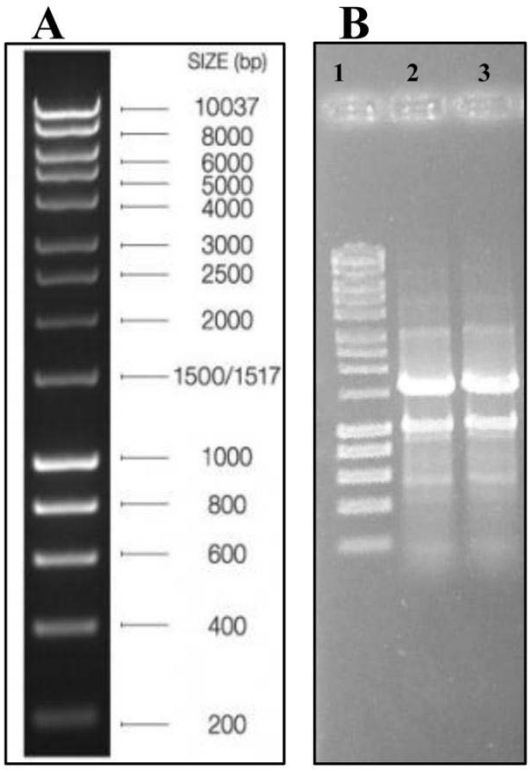


Figure 5.1: Gel electrophoresis of product generated from *recA*$_{S24}$ PCR.
Image (A) shows the standard ladder (Hyperladder 1; Bioline). (B) shows the two products (lane 2 and 3; replicates) generated from PCR using primers to amplify the *recA*$_{S24}$ gene (recA-F/recA-R). There is a predominant product at ~1500 bp and a less dominant product at ~1100 bp. Lane 1 contains 5 µL of the standard ladder as shown in (A).

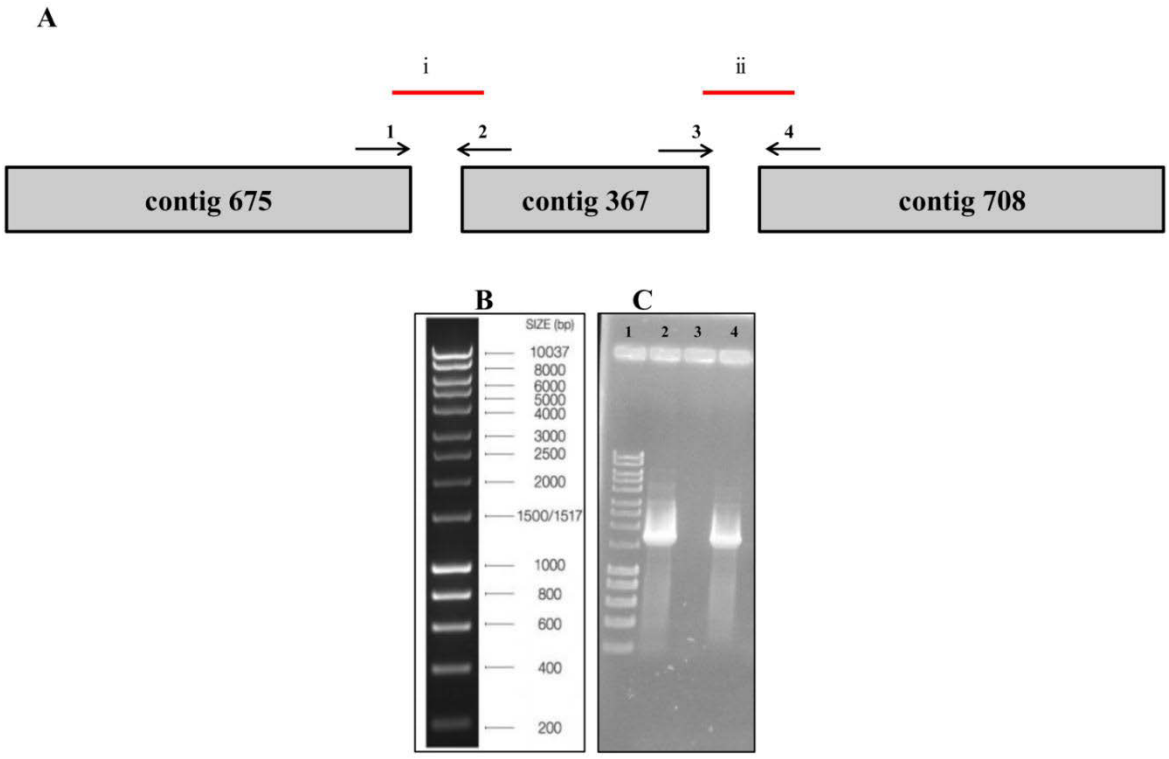
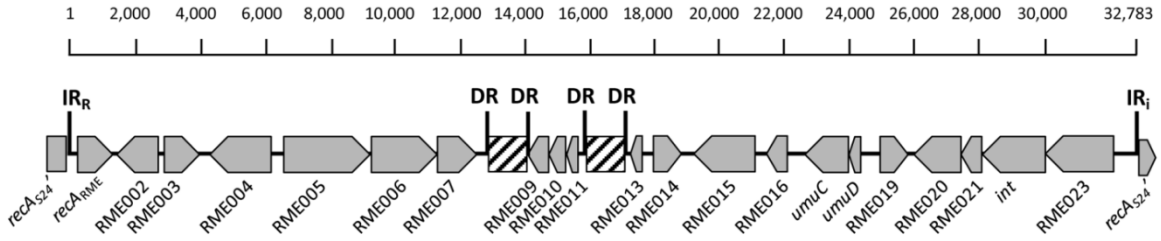


Figure 5.2: Pictorial representation of how contig gaps containing the genomic island were closed and gel electrophoresis of PCR products generated.
RME needed to be pieced together using PCR. It was present on three separate contigs as shown in this figure (A). Arrows indicate primers and primer pair 1 and 2 (ctg675-F and ctg367-R) gave product (i) and primer pair 3 and 4 (ctg367-F and ctg708-R) gave product (ii). These were sequenced and aligned to the genome of S24 and assembled using Geneious software. Lane 2 (C) shows the product generated from primers (ctg675-F and ctg367-R) designed to close the gap between contigs 675 and 367. Lane 4 (C) shows the product generated from primers (ctg367-F and ctg708-R) designed to close the gap between contigs 708 and 367. Lanes 1 standard ladder (Hyperladder 1; Bioline) provided in (B).

Besides *recA*$_{RME}$, a number of genes encoding putative DNA-binding proteins were present in RME including *umuDC* encoding the error-prone DNA polymerase V. Also present is a gene putatively encoding a MutL-like protein (Figure 5.3), a component of the mismatch DNA repair (MMR) system (Polosina and Cupples, 2010). This protein showed 86% identity to MutL from *Pseduoalteromonas piscicida* (WP_010377309.1). MMR corrects for mismatched and unmatched bases that occurs during normal DNA replication. A number of other genes on RME that are homologous to genes involved in DNA processes/repair include: a ParB-like nuclease (91% identity to *Vibrio alginolyticus* 12G01;

WP_005381205.1), a redox sensitive transcriptional activator with a SoxR-domain (96% identity to *Vibrio* sp. 712i1; WP_017634100.1), a type II restriction enzyme containing a methylase subunit (77% identity to *Vibrio splendidus*; WP_017082665.1) and a helicase (90% identity to *Vibrio brasiliensis* LMG 20546; WP_006880978.1).

Two insertion sequence (ISVvu4) elements were identified at positions 12,877 – 14,083 and 15,897 – 17,103 (striped boxes in Figure 5.3) of RME. In both instances, 7-bp direct repeats (DR) were evident abutting the ISVvu4 elements indicating insertion by transposition. The DR for each ISVvu4 element is different, indicating independent insertion events. *In silico* removal of the ISVvu4 elements from the sequence did not restore any ORFs indicating that their insertion had not led to gene disruption. As expected, the promoter regions of both $recA_{RME}$ and the *umuDC* operon have the characteristic LexA binding sequence of CTGT-$(AT)_4$-ACAG indicating control by the SOS response (Saenz-de-Miera *et al.*;Wertman and Mount, 1985). Present on RME also are genes putatively involved in mobilisation/integration such as: a phage integrase (RME022) and a site-specific recombinase XerS (RME019). Remaining are a number of hypothetical genes that have no known putative identifications (Figure 5.3). The RME sequence has been annotated and is uploaded into GenBank (accession number KJ123688) and has been provided as an appendix to this thesis (Appendix 4).



Locus tag	start	stop	orientation	Putative protein identification
RME001	204	1271	+	Recombinase protein RecA
RME002	1373	2593	-	Par-B like nuclease
RME003	2783	3916	+	Hypothetical
RME004	6225	4204	-	MutL mismatch repair domain-containing protein
RME005	6547	9219	+	Type II restriction enzyme, methylase subunit
RME006	9247	11268	+	Hypothetical; DEAD/DEAH box helicase
RME007	11284	12531	+	Hypothetical; YeeC-like nuclease domain-containing protein
RME008	13036	14025	+	Transposase of ISVvu4
RME009	14731	14114	-	Hypothetical
RME010	15192	14734	-	Hypothetical
RME011	15609	15199	-	Hypothetical; DnaJ domain-containing protein
RME012	16056	17045	+	Transposase of ISVvu4
RME013	17645	17211	-	Redox-sensitive transcriptional activator, SoxR
RME014	17963	18844	+	Permease of the drug/metabolite transporter (DMT) superfamily
RME015	21105	19225	-	Hypothetical; chromosome segregation ATPase-containing protein
RME016	22080	21598	-	Hypothetical
RME017	23939	22683	-	Error-prone, lesion bypass DNA polymerase V (UmuC)
RME018	24199	23939	-	Error-prone, lesion bypass DNA polymerase V (UmuD)
RME019	24984	25739	+	Multi domain XerS site-specific tyrosine recombinase XerS
RME020	27461	26001	-	Hypothetical
RME021	28181	27465	-	Hypothetical
RME022	30198	28165	-	Phage integrase family domain protein
RME023	32131	30185	-	Hypothetical

Figure 5.3: Genes carried by the *recA* mobile element and genetic context within *V. cholerae* S24.

Genetic structure and gene content of the *recA* genomic island. The RME contains 9-bp inverted repeats at each end (IR_R and IR_i) and 23 ORFs inclusive of the transpose genes from the ISVuv4 elements (striped boxes). The ISVuv4 elements are abutted by 7-bp direct repeats (DR) indicating insertion by transposition. RME contains multiple genes associated with DNA repair including a full copy of *recA* (RME001), the *umuDC* operon (RME017 and RME018) encoding the two subunits of DNA polymerase V and a gene encoding a protein with a MutL mismatch repair domain (RME004).

5.2.2: Phylogenetic analysis of *recA* on genomic island proves it divergent from $recA_{S24}$

As RME essentially inactivates the indigenous S24 *recA;* $recA_{S24}$, it was of interest to establish how closely related (if at all) the *recA* carried on the RME; $recA_{RME}$ was to *recA* genes resident in *V. cholerae.* Phylogenetic analysis (see section 2.10.1 for method) of *recA* sequences from the *Vibrionaceae* family determined that $recA_{S24}$ groups with the *V. cholerae* clade (consistent with the known phylogeny of S24 based on the analysis of other chromosomal genes) whereas $recA_{RME}$ is not closely related to *V. cholerae*. It does however group with *recA* genes of the *Vibrio* genus and thus is likely derived from a non-*cholerae* member of this genus (Figure 5.4). It is clear $recA_{RME}$ has been derived by LGT. An important point to be emphasised here is that housekeeping genes are not normally subject to LGT, and while this particular mobilisation event is obvious, this may not always be the case. This has implications for the use of housekeeping genes in phylogenetic and taxonomic analyses given that *recA* is commonly used for these purposes. This problem has been a focus of debate in the microbial evolution literature, for example studies conducted by Bapteste *et al.* and Creevey *et al.* (Bapteste *et al.*, 2004;Creevey *et al.*, 2004).

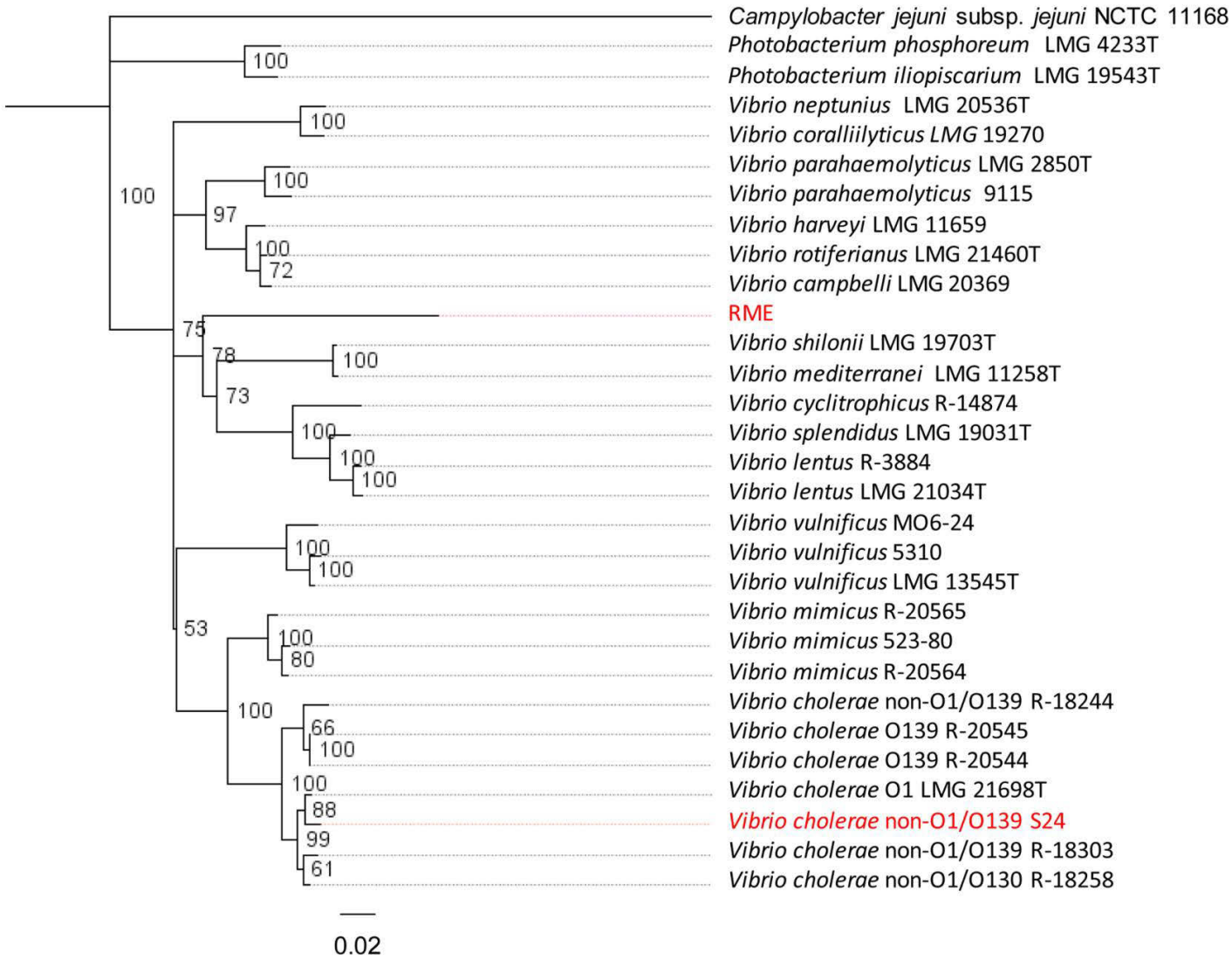

Campylobacter jejuni subsp. jejuni NCTC 11168
Photobacterium phosphoreum LMG 4233T
Photobacterium iliopiscarium LMG 19543T
Vibrio neptunius LMG 20536T
Vibrio coralliilyticus LMG 19270
Vibrio parahaemolyticus LMG 2850T
Vibrio parahaemolyticus 9115
Vibrio harveyi LMG 11659
Vibrio rotiferianus LMG 21460T
Vibrio campbelli LMG 20369
RME
Vibrio shilonii LMG 19703T
Vibrio mediterranei LMG 11258T
Vibrio cyclitrophicus R-14874
Vibrio splendidus LMG 19031T
Vibrio lentus R-3884
Vibrio lentus LMG 21034T
Vibrio vulnificus MO6-24
Vibrio vulnificus 5310
Vibrio vulnificus LMG 13545T
Vibrio mimicus R-20565
Vibrio mimicus 523-80
Vibrio mimicus R-20564
Vibrio cholerae non-O1/O139 R-18244
Vibrio cholerae O139 R-20545
Vibrio cholerae O139 R-20544
Vibrio cholerae O1 LMG 21698T
Vibrio cholerae non-O1/O139 S24
Vibrio cholerae non-O1/O139 R-18303
Vibrio cholerae non-O1/O130 R-18258
100
100
100
100
97
100
72
75
78
100
73
100
100
100
100
100
53
100
80
100
66
100
100
88
99
61
0.02

Figure 5.4: Phylogenetic tree of 30 *recA* nucleotide sequences from the *Vibrio* genus. Phylogenetic analysis of $recA_{S24}$ and $recA_{RME}$ (RME highlighted in red). $recA_{S24}$ (also highlighted in red) groups with *V. cholerae* strains whereas $recA_{RME}$ groups with *recA* from other *Vibrio* species indicating that $recA_{RME}$ was mobilised from another member of the *Vibrio* genus. *Campylobacter jejuni* has been used as an outgroup.

5.2.3: The *recA* genomic island circularises and is excised from the S24 genome

Many GIs are known to excise from their location in the chromosome (Boyd *et al.*, 2009). In order to determine whether RME circularises and excises from the S24 chromosome, inverse PCR was performed using primers reading out from the IR_R and IR_i ends (RME-F/RME-R). A product of ~560 bp was amplified and sequenced (Figure 5.5A). Sequencing analysis of the interrupted $recA_{S24}$ indicated that exact excision at IR_R and IR_i would leave behind 4-bp causing $recA_{S24}$ to be out of frame and therefore non-functional (Figure 5.6A). Analysis of the sequence showed that excision occurs in one of two possible ways (Figure 5.6B): 1) Excision occurs at 2-bp on either side of the IR_R and IR_i ends (black arrows in Figure 5.6A) and/or 2) Excision occurs exactly at the IR_i end and at 4-bp before the IR_R end (grey arrows in Figure 5.6A). Either way, excision was predicted to restore an uninterrupted and therefore functional copy of $recA_{S24}$ in the chromosome (Figure 5.6C). Amplification of $recA_{S24}$ using primer set S24-cinA-F/S24-recX-R was performed and DNA extracted from the band at the predicted size of ~1600 bp (Figure 5.5A). This length of ~1600 bp is consistent with RME having excised from the S24 genome. A nested PCR (Figure 5.5B) was then performed (EcoRI-recA-F/EcoRI-recA-R) to obtain sufficient product for sequencing of an intact "empty" insertion site (Figure 5.5B). Inverse PCR was also conducted in *E. coli* containing a fosmid clone (pCC2FOS::RME) that bears a complete intact RME (see section 2.3.4 for methods). Interestingly, excision of RME was undetectable by PCR analysis in *E. coli* even when subjected to UV-irradiation implying that factors within *V. cholerae* S24 are required for or enhance excision.

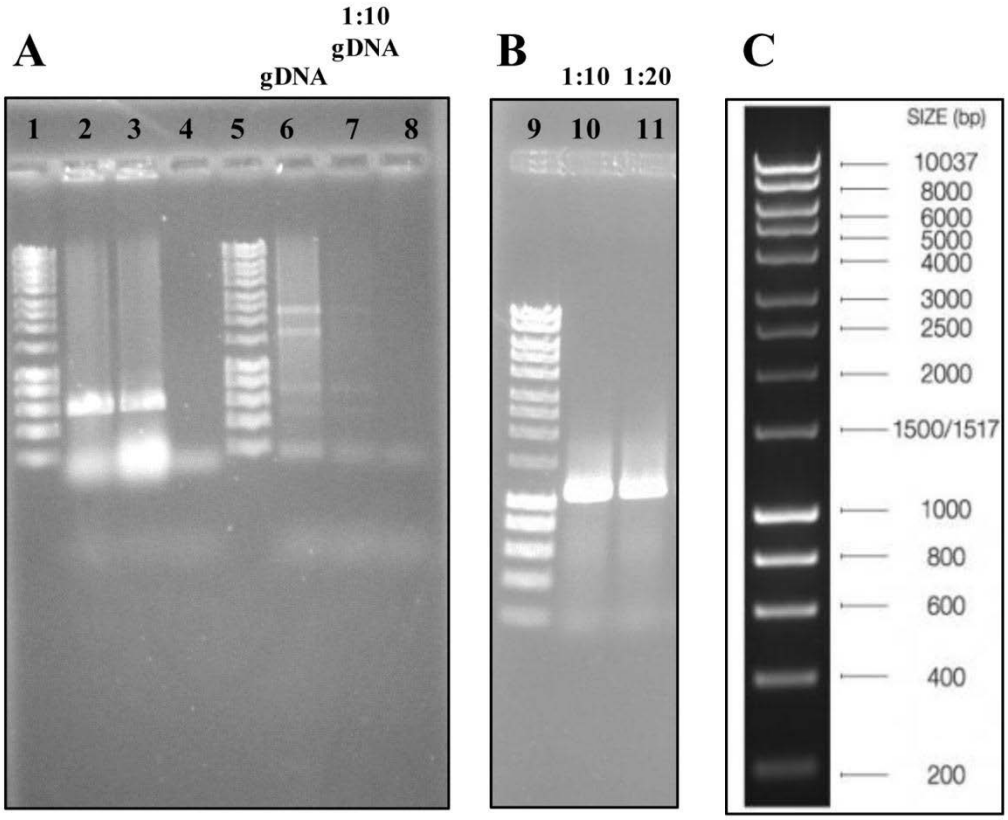


Figure 5.5: RME can circularise from the S24 chromosome and leaves behind an intact copy of *recA*$_{S24}$.

PCR gel profiles showing inverse PCR where RME circularises (A) and gives the expected product size of ~600 bp in both lanes 2 and 3. The template DNA used for amplification in lanes 2 and 3 was derived from a colony lysate. The negative control is given in lane 4 where water is used as the template (RME-F/RME-R). Lane 6 shows the products obtained when the empty insertion site of RME is amplified using S24-cinA-F and S24-recX-R. The template DNA used for this PCR reaction is undiluted S24 genomic DNA. Lane 7 shows the product of a 1:10 dilution of S24 genomic DNA. Lane 8 the negative control (water used as template DNA also). Shown in (B) is product amplified using primers EcoRI-recA-F/EcoRI-recA-R from extracted DNA from band at ~1600 bp in lane 6. Lane 10 is a 1:10 dilution of extracted PCR product and Lane 11 is a 1:20 dilution. Lanes 1,5 and 9 contain 5 µL of Hyperladder 1 (C).

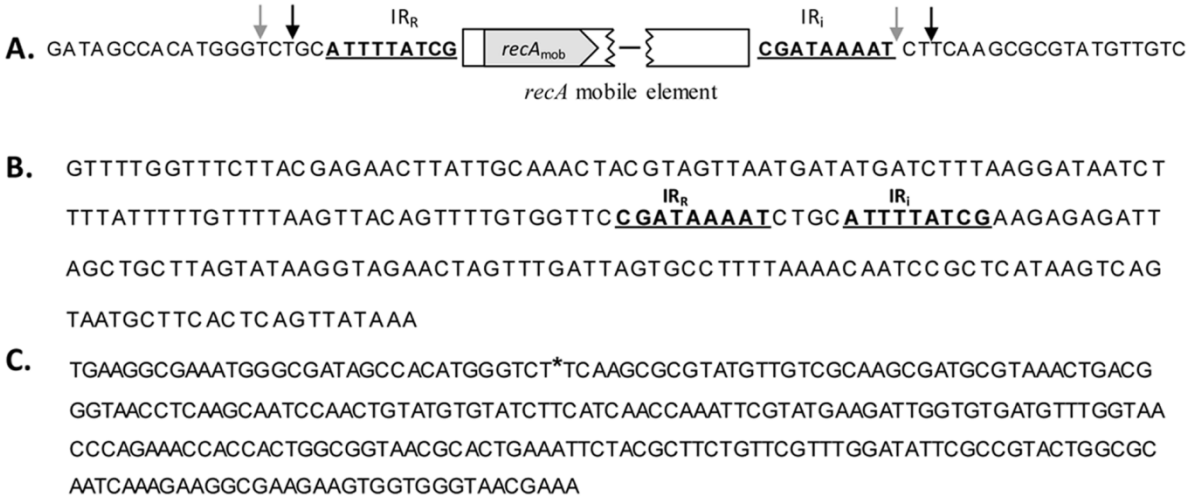


Figure 5.6: RME insertion into and excision from the S24 host genome.
Sequence abutting the insertion point of the *recA* genomic island (A). The black and grey arrows demarcate the possible excision points for the RME. Sequence of the product derived from inverse PCR of the excised RME (B). The sequence shows that excision does not precisely occur at IR_R and IR_i and either occurs by two possible methods (see text for more details). Sequence of the "empty" $recA_{S24}$ insertion site (C) shows that excision leaves behind a complete copy of $recA_{S24}$. The asterix marks the point of RME insertion.

5.2.4: The *recA* genomic island targets the *recA* gene

To determine whether the RME was capable of translocating into a new location, a vector ($pOriVn_{700}$-$recA_{S22}$) containing the *recA* gene from a closely related strain of *V. cholerae* S24, strain S22 (Islam *et al.*, 2013) was introduced into *V. cholerae* S24 by conjugation (see section 2.14 for method). A control vector substituting $recA_{S22}$ with *gfp* ($pOriVn_{700}$-P_{lac}*gfp*) was also introduced into *V. cholerae* S24. Primers targeting the vector backbone and the ends of the RME were used in a PCR reaction to determine whether the RME had moved into either $pOriVn_{700}$-$recA_{S22}$ or $pOriVn_{700}$-P_{lac}*gfp* (Figure 5.7A). In four independent conjugation experiments of $pOriVn_{700}$-P_{lac}*gfp* into *V. cholerae* S24, a product was never detected. However, products were detected for insertion of the RME into $recA_{S22}$. Across the clones containing insertion events in $recA_{S22}$ both orientations were represented. That is, some clones had the RME inserted in the orientation seen in S24 and others had the RME in the opposite orientation. Examples of both orientations were found across three of four independent conjugation assays (Figure 5.7B).

Sequence analysis of the amplified products from the *V. cholerae* S24/pOriVn$_{700}$-*recA*$_{S22}$ transconjugants indicated movement of the RME into *recA*$_{S22}$. It should be noted that homologous recombination between *recA*$_{S22}$ in the vector and *recA*$_{S24}$ in *V. cholerae* S24 genome could result in merodiploids that generate the same sized PCR products as those for RME inserted in the orientation found in *V. cholerae* S24 (Figure 5.8). However, in the immediate 2-bp of the IR$_R$ end for three of the transconjugants (Figure 5.7B; ia, iiia and iva), there is a G to T substitution and at the immediate 3-bp of the IRi end, two transconjugants (Figure 5.7B; iiia and iva) showed a T to G substitution. Since the *recA*$_{S22}$ sequence is identical to *recA*$_{S24}$ around the insertion point, homologous recombination should result in identical sequences immediately surrounding the RME. Since this is not the case, homologous recombination is unlikely the result for these derived sequences. Furthermore, homologous recombination cannot explain insertion of the RME in the inverse orientation indicating insertion *via* a non-homologous process. Since insertions were never detected in pOriVn$_{700}$-P$_{lac}$*gfp*. These data suggest that the RME is targeting *recA* or at the very least has a preference for integration into *recA*.

These data show that RME is capable of mobilisation and preferentially targets a specific site within *recA*. By carrying its own functional copy of *recA,* the GI does not affect any of the vital cell pathways associated with disruption of this gene during integration. Furthermore, specific targeting of *recA* may be necessary to ensure successful maintenance and dissemination of the GI. Since RecA does not function as a monomer but polymerizes to form a filament structure (Yu *et al.*, 2004) disruption of the indigenous *recA* prevents a situation where two divergent RecA proteins might negatively interact resulting in reduced cell fitness.

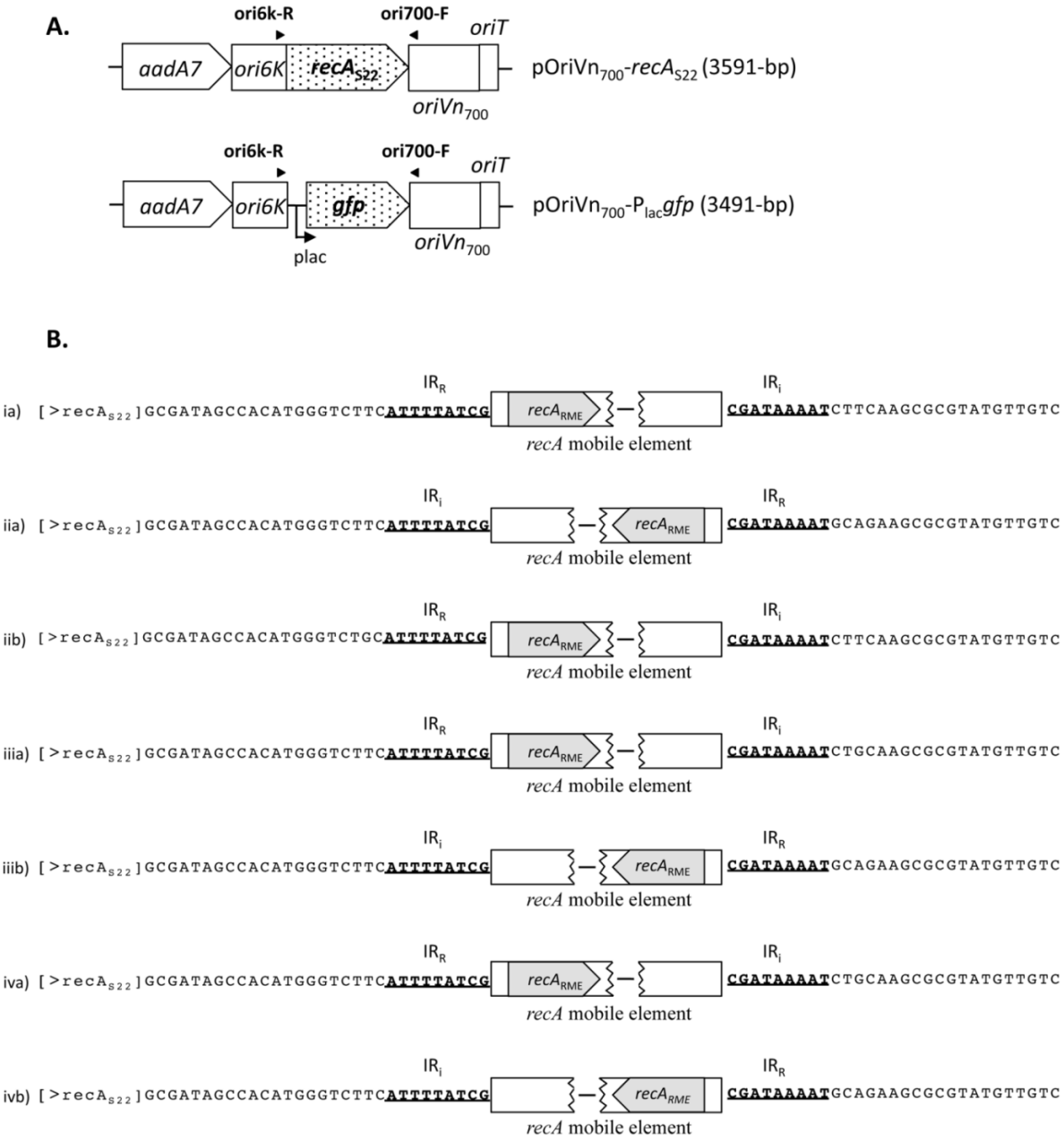
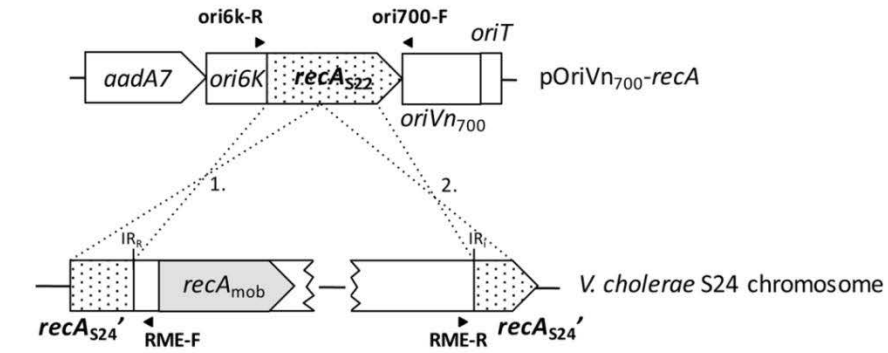
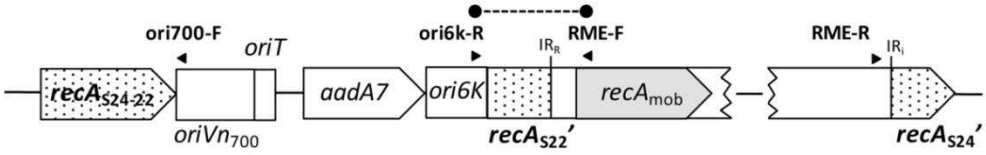


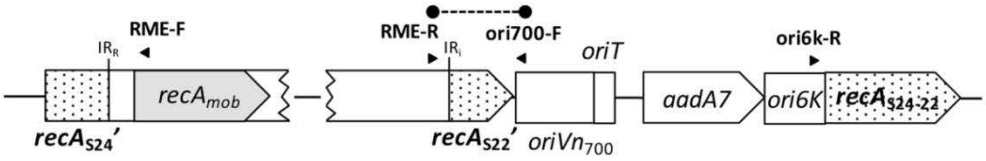
Figure 5.7: The *recA* genomic island targets *recA*.
Movement of the *recA* genomic island into a replicating vector. Genetic structure of the vectors pOriVn$_{700}$-*recA*$_{S22}$ and pOriVn$_{700}$-P$_{lac}$*gfp* and the placement of primers ori6k-R and ori700-F are shown (A). Sequence of products derived using vector specific and RME specific primers from PCR of *V. cholerae* S24 transconjugates. Each transconjugant is denoted by i), ii), iii) and iv). The sequence indicates specific insertion into the same site of *recA*$_{S22}$.



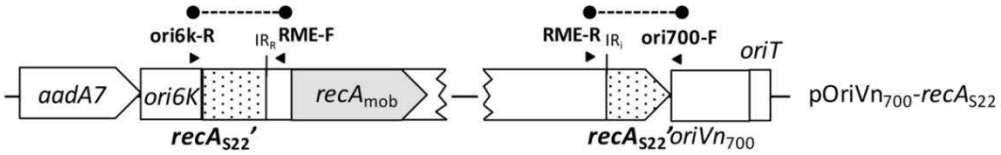
cross-over 1



cross-over 2



Insertion 1



Insertion 2

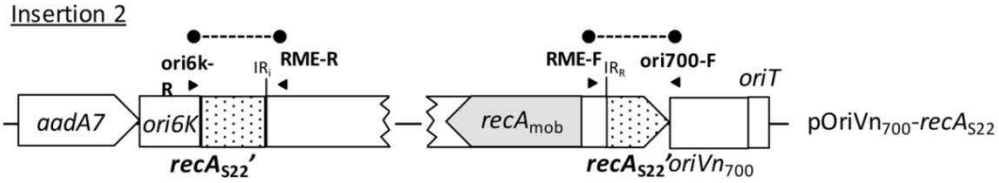


Figure 5.8: Possible insertion events of RME into *recA* from *V. cholerae* S22.
Possible insertion and homologous recombination events of the *recA* genomic island with pOriVn$_{700}$-*recA*$_{S22}$. Production of merodiploids due to homologous recombination between *recA*$_{S22}$ and *recA*$_{S24}$ could result in two genetic structures (cross-over 1 and cross-over 2). In both instances, PCR products could be generated using the vector specific and RME specific primers used to detect insertion of the RME into *recA*$_{S22}$ (broken lines) in the orientation found in the *V. cholerae* S24 genome (insertion 1). Insertion of the RME in the inverse orientation (insertion 2) would generate products using inverse primer pairs. The inverse insertion cannot be explained by homologous recombination and indicates an integration event.

5.2.5: The *recA* genomic island provides protection against ultraviolet irradiation in *E. coli*

Section 5.2.1 describes a number of genes involved in DNA pathways, including DNA repair are present on RME such as *recA*. To investigate the role *recA*$_{RME}$ has in protecting the cell from DNA damage caused by UV-radiation and exposure to antibiotics, natural transformation of the genomic island into eight environmental *V. cholerae* strains, designated S10, S11, S12, S16, S18, S22, S23 and S25 was attempted (see Table 2.3; (Islam *et al.*, 2013)). It has been shown that when grown on chitin, *V. cholerae* becomes naturally competent and can uptake linear fragments of DNA (Meibom *et al.*, 2005). This experiment was performed in order to test the ability of RME to protect against DNA damage.

Firstly, the entire RME was cloned into a fosmid (fosmid vector pCC2FOS) and transformed into the *recA*$^-$ *E. coli* strain EPI300 (see section 2.3.4 for method). Fosmid vectors are particularly valuable when cloning large portions of DNA, as the technology exploits the ability of a phage to preferentially package only large sections of DNA (~40 kb; RME is ~32 kb in size) and incorporate this DNA carried by the fosmid vector into the *E. coli* strain EPI300. The other advantage of employing a fosmid to transfer RME, is that the fosmid vector (containing RME) does not over-replicate in the cell and thus there is only one copy of the vector and RME insert per cell. This means that any resulting phenotype due to the presence of RME is not a consequence of overexpression of genes present on the genomic island. Using this method the construction of a fosmid containing

RME in *E. coli* EPI300 was successful, and the resulting strain was designated EPI300/pCC2FOS::RME. A strain containing pCC2FOS only was also constructed to be a control in subsequent assays. Using transposon mutagenesis (section 2.3.4 for method), the ISVvu4 insertion sequence at 16,056-17,045 bp (RME012) in Figure 5.3 (for simplicity referred to here as IS-2) was insertionally tagged with the Tn5 transposon carrying a kanamycin resistance gene resulting in the strain designated EPI300/pCC2FOS::RMEΔIS-2. IS-2 was chosen as a target for transposon mutagenesis for two reasons: 1) to introduce a selection marker for chitin transformation assays and 2) as ISVvu4 does not carry any apparent functional ORFs, insertionally inactivating this region of RME would not disrupt the function of the GI. Restriction enzyme sites on either side of RME were exploited to linearise the genomic island from its location on pCC2FOS, as only linear DNA has been shown to be successfully transferred in chitin transformation experiments (Meibom *et al.*, 2005). Chitin transformation was performed to move RME across (see section 2.3.8.1 for method) into eight environmental non-O1/O139 *V. cholerae* strains co-isolated with S24 (Islam *et al.*, 2013). Unfortunately, all attempts to transform RME into another environmental *V. cholerae* strain *via* this method were unsuccessful. In an attempt to artificially mediate natural transformation of linearised RME *via* chitin, a vector with the major regulator gene of chitin competency; *tfox*, under the control of an arabinose-inducible promoter was conjugated into two out of the eight non-O1/O139 *V. cholerae* strains (construct made previously in the laboratory by Dr Maurizio Labbate) (Pollack-Berti *et al.*, 2010). Chitin experiments were performed again after inducing expression of *tfox* with addition of 0.2% arabinose. Unfortunately, transformation of RME into *V. cholerae* was again unsuccessful.

Given the inability to transform RME into another strain of *V. cholerae* to determine its role in DNA repair, UV-irradiation assays were performed in the *recA*$^-$ *E. coli* strain EPI300 (see section 2.11 for method). Initially, growth assays were performed to confirm that the presence of RME did not affect the growth of EPI300 (Figure 5.9). No effect on growth of EPI300 due to pCC2FOS or RME was observed and thus UV-irradiation assays were performed. UV-light is a known DNA damaging agent, Figure 5.10A shows that the entire RME enhanced bacterial cell survival when exposed to 0.8 mJ/cm^2 of UV-C. It can be seen

that EPI300 and vector only controls are completely killed at 20s exposure (Figure 5.10A and B), however EPI300 containing the RME survive up until 60s of exposure. Also, RME provides the cell with 1000-fold increased survivability at 10s and 10-fold higher survivability after 20s exposure. Figure 5.10A shows that when $recA_{RME}$ is insertionally inactivated (strain designated EPI300/pCC2FOS::Δ$recA_{RME}$; using transposon mutagenesis method section 2.3.4), the level of cell survivability decreases to that of cells containing no RME. This demonstrates that, at least under the conditions tested, $recA_{RME}$ is functional and is the gene mainly responsible for the enhanced protection from UV-irradiation provided by the presence of the RME.

To confirm pCC2FOS alone was not providing the cell with any improved protection against UV-C, EPI300 alone was compared to EPI300 containing cloning vector pCC2FOS only (Figure 5.10B). This showed no difference in protection levels against UV-C radiation.

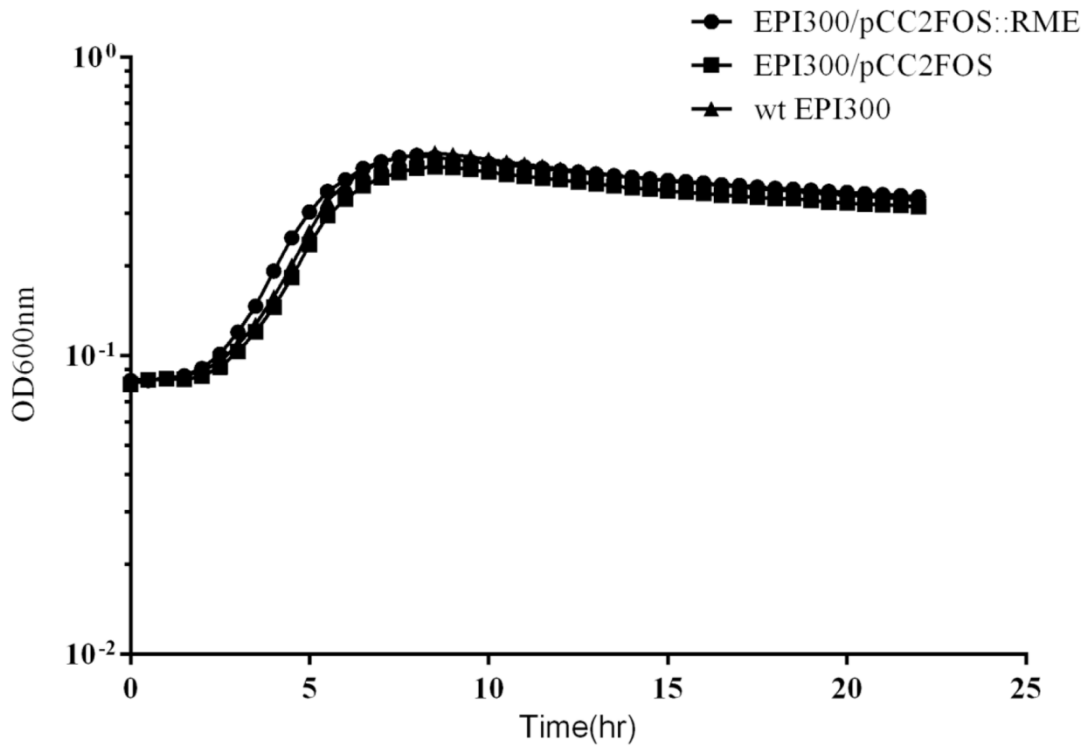


Figure 5.9: Growth of fosmid constructs in EPI300 in LB5 medium.
This figure shows growth curve data of wt EPI300, EPI300/pCC2FOS (vector only) and EPI300/pCC2FOS::RME (vector with genomic island) in LB5. No difference in growth was observed due to the presence of the genomic island or fosmid vector pCC2FOS. The y-axis shows optical density readings at 600 nm absorbance and the x-axis is given in time (hr).

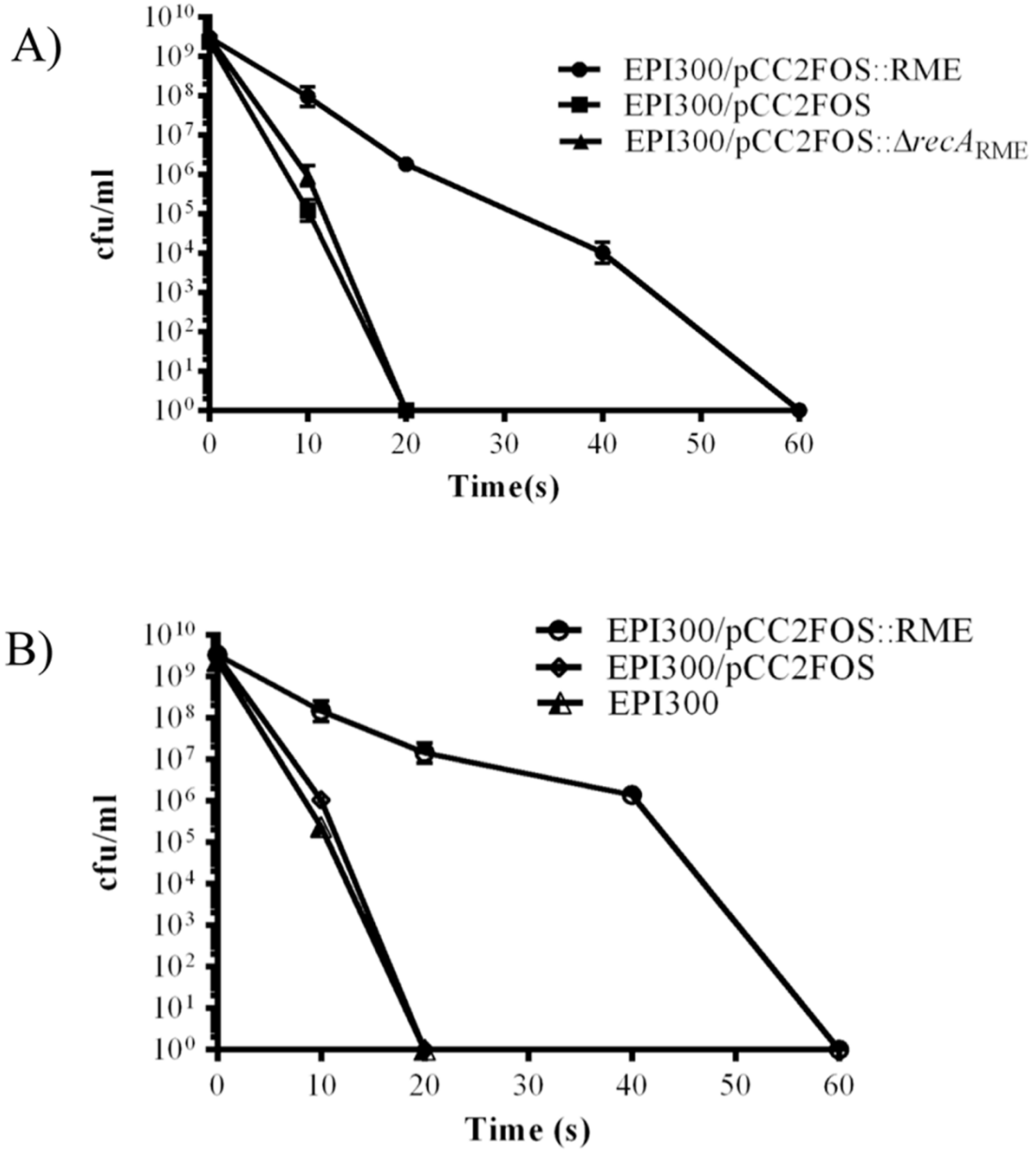


Figure 5.10: Ultraviolet-irradiation assays of *E. coli* containing *recA* genomic island. Survival of *E. coli* carrying the *recA* genomic island and control strains when exposed to UV-C stress. Time points are given at 0, 10, 20, 40 and 60 s. UV-C exposure was set to 0.8 mJ/cm^2. Error bars indicate the standard error of the mean (SEM) of three replicates.

5.2.6: The *recA* genomic island provides *E. coli* with increased protection against antibiotics

Damage to DNA is induced in the bacterial cell by many mechanisms, one such method being ultraviolet light as shown section 5.2.5. However, in the clinical environment bacteria encounter a variety of stressors that induce DNA damage other than UV-radiation. Repair of damaged DNA is a necessity for bacterial cell survival. It has been shown in 5.2.1 that RME has multiple genes involved in DNA repair. To test whether RME provided enhanced protection against three DNA targeting antibiotics: nalidixic acid, ciprofloxacin and bleomycin, MICs were performed in EPI300 (see section 2.13.1 for method). Interestingly, the MIC of nalidixic acid (1st generation quinolone) did not vary between any of the tested strains, including the RME (data not shown). However, when the strains given in Table 5.1 were tested using ciprofloxacin, a 2nd generation quinolone (i.e. fluoroquinolone) a greater MIC was observed for strains containing the GI. Likewise, when cells containing the GI were tested using bleomycin, a 2X higher MIC was recorded.

In the case of ciprofloxacin it is apparent that $recA_{RME}$ is the main driver in providing protection. Consistent with this, when $recA_{RME}$ is insertionally inactivated from the GI the MIC of ciprofloxacin drops to that of wt EPI300. However, in the case of bleomycin, the MIC is the same as wt EPI300 when either $recA_{RME}$ or $umuC_{RME}$ are insertionally inactivated. This shows both *recA* and *umuC* are required for protection against bleomycin. Based on evidence in the literature (detailed in the discussion 5.3.1), it is likely that RecA is required for activation of *umuC*.

Table 5.1: Minimal inhibitory concentration (MICs*)

Strain	Ciprofloxacin	Bleomycin
EPI300	0.015625	8
EPI300/pCC2FOS	0.015625	8
EPI300/pCC2FOS::RME	0.0625	16
EPI300/pCC2FOS::RMEΔ*umuC*$_{RME}$	0.0625	8
EPI300/::RMEΔ pCC2FOS*recA*$_{RME}$	0.015625	8

*MIC given as μg/mL

5.2.7: The *recA* genomic island increases spontaneous mutation frequency in *E. coli* when grown on media containing antibiotics

Apart from DNA repair, *recA* and DNA polymerase V (encoded by *umuDC*) are known to increase spontaneous mutation frequencies resulting in the emergence of antibiotic resistance mutants (Thi *et al.*, 2011). The induction of DNA polymerase V initiates the SOS mutagenesis pathway, which acts as the cell's last attempt to repair damaged DNA. Spontaneous mutation within the bacterial chromosome in *V. cholerae* is well documented to cause resistance to a variety of antibiotics (Goss *et al.*, 1965;Gellert *et al.*, 1977;Sugino *et al.*, 1977;Allen *et al.*, 1979;Kitaoka *et al.*, 2011). Thus, mutation frequencies using two antibiotics: rifampicin which acts on RNA synthesis and nalidixic acid which targets DNA replication by inhibiting the A and B subunit of DNA gyrase (Komp *et al.*, 2003), were performed (see section 2.13.2 for method).

Table 5.2 shows mutation frequencies resulting from growth on LB + 100 μg/ml rifampicin after 24 and 48 hr. It can be seen from the mutation frequencies over the three replicates (30 tubes, 10 sets of tubes per replicate) that EPI300 containing the intact GI when compared to control strains (wt EPI300 and EPI300/pCC2FOS) showed no increase in spontaneous mutant growth. Likewise, when *umuC* was insertionally inactivated in RME (strain designated EPI300/pCC2FOSΔ*umuC*$_{RME}$) there was no change in mutation frequency when compared to control *E. coli* EPI300 strains. This result was not surprising as rifampicin targets the β-subunit of RNA polymerase (Kitaoka *et al.*, 2011).

Following these results mutation frequencies were then performed using LB5 supplemented with 50 μg/ml nalidixic acid. Nalidixic acid is a 1st generation quinolone that is known to target DNA replication by inhibiting the A subunit of DNA gyrase, resulting in generation of the relaxed DNA form; interrupting the DNA replication fork. Table 5.3 shows the mutation frequencies generated after growth on nalidixic acid after 24 and 48 hr. Over the replicates it is evident that wt EPI300 and vector only control consistently do not produce any spontaneous mutants. However, *E. coli* EPI300/pCC2FOS::RME has an increased mutation frequency after both 24 and 48 hr incubation on nalidixic acid. *E. coli* EPI300/pCC2FOS::RMEΔ*umuC*$_{RME}$ generally has the same mutation frequency as *E. coli* EPI300/pCC2FOS::RME. However, replicate 3 showed that this strain produced no mutations after 24 and 48 hr. It can be concluded from these experiments that *E. coli* EPI300 with the complete RME displays elevated mutation frequency when exposed to the DNA damaging antibiotic; nalidixic acid.

Table 5.2: Rifampicin[a] mutation frequencies

Strain	Mutation frequency 24h	Mutation frequency 48h
EPI300	$1.6x10^{-8}$ – $8.7x10^{-8}$	$3.7x10^{-8}$ – $1.2x10^{-7}$
EPI300/pCC2FOS	$1.6x10^{-8}$ - $1.1x10^{-7}$	$4.0x10^{-8}$ - $1.2x10^{-7}$
EPI300/pCC2FOS::RME	$1.9x10^{-8}$ – $3.3x10^{-8}$	$4.0x10^{-8}$ – $9.7x10^{-8}$
EPI300/pCC2FOS::RMEΔ*umuC*$_{RME}$	$1.9x10^{-8}$ – $4.1x10^{-8}$	$4.5x10^{-8}$ - 9.5 $x10^{-8}$

Values indicate the range for three replicates
[a]concentration of rifampicin = 100 μg/ml

Table 5.3: Nalidixic acid[b] mutation frequencies

Replicate 1

Strain	Mutation frequency 24h	Mutation frequency 48h
EPI300	$<1.1 \times 10^{-11}$	$<1.1 \times 10^{-11}$
EPI300/pCC2FOS	$<1.7 \times 10^{-11}$	$<1.7 \times 10^{-11}$
EPI300/pCC2FOS::RME	1.4×10^{-8} (4)	1.5×10^{-8} (4)
EPI300/pCC2FOS::RMEΔ$umuC_{RME}$	1.6×10^{-9} (3)	2.1×10^{-9} (4)

Replicate 2

Strain	Mutation frequency 24h	Mutation frequency 48h
EPI300	$<1.1 \times 10^{-11}$	$<1.1 \times 10^{-11}$
EPI300/pCC2FOS	$<1.7 \times 10^{-11}$	$<1.7 \times 10^{-11}$
EPI300/pCC2FOS::RME	8.3×10^{-10} (1)	1.4×10^{-9} (3)
EPI300/pCC2FOS::RMEΔ$umuC_{RME}$	2.8×10^{-9} (6)	4.4×10^{-9} (7)

Replicate 3

Strain	Mutation frequency 24h	Mutation frequency 48h
EPI300	$<1.1 \times 10^{-11}$	$<1.1 \times 10^{-11}$
EPI300/pCC2FOS	$<1.3 \times 10^{-11}$	$<1.3 \times 10^{-11}$
EPI300/pCC2FOS::RME	4.8×10^{-10} (1)	1.7×10^{-9} (4)
EPI300/pCC2FOS::RMEΔ$umuC_{RME}$	$<9.1 \times 10^{-12}$	$<9.1 \times 10^{-12}$

"<" indicates that zero colonies appeared in all 10 sets

Numbers in brackets indicates the number of replicates in which one or more colonies appeared.

[b]concentration of nalidixic acid = 50 µg/ml

5.3: Discussion

Mobile genetic elements such as genomic islands have been pivotal in the evolution of the cholera-producing pathogen, *Vibrio cholerae* (Faruque and Mekalanos, 2003). Although a pathogen, *V. cholerae* is also an environmental microorganism and it is in the aquatic environment where *V. cholerae* acquires novel genetic material resulting in the emergence of new strains (e.g. recent emergence of the hybrid strains; discussed in chapter 1). Non-O1/O139 strains are likely to be a significant source of novel DNA for pathogenic strains thus, a better understanding for the diverse genetic elements present in the *V. cholerae* species is important in understanding the emergence of new epidemic strains. The results observed in the chapter reveal a novel genomic island that is integrated into the genome of the environmental non-O1/O139 *V. cholerae* strain S24. This genomic island is novel as it a) carries a plethora of genes putatively involved in multiple DNA repair pathways and DNA processes, b) is the first genomic island associated with the transport of the housekeeping gene *recA* and c) preferentially integrates into *recA*.

5.3.1: The RME genomic island carries genes involved in DNA repair that are functional in *E. coli* and protect against induced DNA damage

The results presented in this chapter revealed a novel mobile genetic element that is integrated into the genome of the environmental non-O1/O139 *V. cholerae* strain S24. Perhaps the most intriguing feature of this mobile genetic element is that it has inserted into the *V. cholerae* S24 host *recA* gene while carrying a copy of *recA* thereby not disrupting vital cell pathways associated with this gene. Interestingly, a 2011 study showed in *C. difficile* a 4.2 kb insert that disrupts the gene encoding the thymidylate synthetase enzyme. This insert provides its own functional copy of the enzyme that was shown to be more functionally active (Knetsch *et al.*, 2011). It is intriguing to suggest that *recA* carried by RME may also be more functionally active with regards to DNA repair mechanisms.

The preferential targeting of *recA* by RME (Figure 5.7) has consequences for the use of housekeeping genes, such as *recA*, when they are used in phylogenetic analyses. Housekeeping genes such as *recA*, *gyrB*, and *rpoB* are commonly used to group bacterial species together, as they are 1) highly conserved and 2) ubiquitous among all bacteria

(Marechal *et al.*, 2000). Further to this, a study carried out in 2004 also shows that *recA* can be used as an alternative phylogenetic marker in the *Vibrionaceae* family (Thompson *et al.*, 2004a). In this study we showed that RME carries a copy of *recA* ($recA_{RME}$). Phylogenetic analysis revealed that $recA_{RME}$ is phylogenetically distant from representative *recA* genes derived from a number of *V. cholerae* strains.mThis raises an array of misclassification issues when using this gene as a phylogenetic marker if the potential for LGT is ignored.

As stated previously (section 5.1), the expression of error-prone polymerases is induced when DNA damage within the bacterial cell cannot be repaired by more stringent mechanisms. The SOS induction of error-prone polymerases like polymerase V is considered a final response where although induced mutation(s) may kill a number of cells, a random mutation at some point may be favourable, resulting in progeny survivability. An alternative view for the function of error-prone polymerases is that they are used to generate genetic diversity for survival in environments to which the host is maladapted. To support this hypothesis, transcription of error-prone polymerases has been observed in the absence of SOS-induced DNA damage (Yeiser *et al.*, 2002). Furthermore, error-prone polymerase mutants are less competitive than their parents during starvation (McKenzie *et al.*, 2000;Yeiser *et al.*, 2002;Tark *et al.*, 2005) and some antibiotics (e.g. quinolones like nalidixic acid shown in this chapter) induce the SOS mutagenic response, increasing the frequency of resistant mutants (Piddock and Wise, 1987;Ysern *et al.*, 1990). Most interestingly, UmuD'$_2$C and *umuDC*-like operons are commonly associated with mobile genetic elements (Permina *et al.*, 2002;Tark *et al.*, 2005;Hare *et al.*, 2012). The presence of DNA polymerase V on MGEs allows for the transfer of these genes that provide the cell with enhanced evolutionary fitness during times of severe stress. However, *recA* or *recA* homologues have never been reported in association with MGEs. This chapter showed that *recA* not only is mobile but is also found on a GI that contains other putative DNA repair genes such as *umuDC* and other DNA repair genes such as a *mutL*-like gene. UV-irradiation experiments described in this chapter show the complete mobile element provides enhanced protection against this type of induced DNA damage. When $recA_{RME}$ was inactivated, there was no measurable protection conferred by the RME.

Cells with a Δ*recA* background are known to be extremely sensitive to UV-induced DNA damage such as the EPI300 strain used in assays for this chapter (Aranda *et al.*, 2011). Hence a lack of any detectable protection from UV-irradiation due to the genes present on RME such as: *mutL*-like ORF (RME004), the two nuclease-like genes (RME002 and RME007) and the methylase-like gene (RME005); may be due to the absence of any cells surviving UV-irradiation long enough for expression of these genes to be initiated. Further to this point there may be other host *E. coli* repair pathways where activation upon induced DNA damage supersedes those on RME. Thus, performing similar experiments in a *Vibrio cholerae* background would be an important and logical next step. To address this it was attempted to transform RME into 8 non-O1/O139 *V. cholerae* strains using chitin-induced transformation after tagging RME with a kanamycin resistance gene (Marvig and Blokesch, 2010). These experiments however, were unsuccessful. As it is well known that *recA* is required for the activation of $UmuD_2C$ its removal could provide another possible explanation for the lack of protection provided by pCC2FOS::RMEΔ$recA_{RME}$ in UV-irradiation experiements (Patel *et al.*, 2010). Thus, testing the $umuC_{RME}$ deletion mutant in UV-irradiation assays will be cruicial to further elucidating its role in DNA repair.

Likewise, *E. coli* cells containing the entire mobile element showed an increased mutation frequency when grown in the presence of nalidixic acid, but not rifampicin. However, when *umuC* was inactivated from RME no difference was observed in mutation frequency when compared to wt. This implies that the protection provided by RME is most likely a $recA_{RME}$ mediated response ($recA_{RME}$ activates *E. coli* host *umuDC*); showing for the first time that a *recA* gene transferred on a mobile genetic element is functionally active.

MIC experiments carried out using both ciprofloxacin and bleomycin showed that the complete mobile element provided the cell with 4 fold and 2 fold increased protection levels respectively. However when $umuC_{RME}$ and $recA_{RME}$ were independently inactivated from the genomic island, the MIC, when challenged with bleomycin, decreases to that of wt cells. This clearly shows that both these genes are active players in driving the identified protection provided by RME. Conversely, ciprofloxacin MICs showed that when $umuC_{RME}$ and $recA_{RME}$ were independently inactivated from RME, only the strain with $recA_{RME}$ inactivated from RME had an MIC identical to wt EPI300. This result showed that when

challenged with ciprofloxacin the *recA* gene (and not *umuC*) on the genomic island is a key player in providing protection from this fluoroquinolone. The lack of protection driven by *umuC* when cells were exposed to ciprofloxacin (not bleomycin) may be explained by their differing modes of action. Bleomycin acts by inducing breaks in dsDNA (D'Andrea and Haseltine, 1978) and ciprofloxacin targets DNA gyrase preventing the decatenation of replicating DNA (Kitaoka *et al.*, 2011). Thus, the abundance of ssDNA in the cell due to the addition of bleomycin may give DNA polymerase V an enhanced opportunity to bind and synthesise nucleotides past the many lesions.

5.3.2: Are there implications of antibiotic protection due to RME?

With antibiotic resistance becoming an increasing public health issue worldwide, it is interesting to speculate that in cells that already contain genes directly responsible for resistance to most antibiotics, RME may further enhance the survival of such strains. Thus, potentially pushing the MIC of these strains beyond what can be reached at the site of infection.

Antibiotic resistance in *V. cholerae* has become widespread since the late 1970s, with the isolation of **m**ultiply **a**ntibiotic-**r**esistance ***V. cholerae***; given the acronym MARV (Sack *et al.*, 2001). Since then, the majority of resistance profiles have been attributed to mobile genetic elements, although chromosomal mutations conferring resistance have been reported (Sack *et al.*, 2001;Ehara *et al.*, 2004;Jabeen *et al.*, 2008;Chander *et al.*, 2009;Kim *et al.*, 2010;Kitaoka *et al.*, 2011). However, many strains described as having antibiotic resistance profiles during pandemic outbreaks have no identified mechanism responsible for the resistance (Kitaoka *et al.*, 2011). Interestingly, in 2001 (Garg *et al.*, 2001) a study showed that *V. cholerae* O1 of the El Tor biotype isolated from cholera patients in Calcutta, had an increased resistance to ciprofloxacin, a fluoroquinolone used to treat cholera infections (and used in this chapter) (Garg *et al.*, 2001). More recently, ciprofloxacin was unable to be used as a treatment during the pandemic outbreak in Haiti (2010-2013) due to ciprofloxacin resistance (CDC, 2011). Prior to these studies, ciprofloxacin resistance had only been reported in non-O1/O139 *V. cholerae* strains (Mukhopadhyay *et al.*, 1996). This suggests that resistance to ciprofloxacin may have originated from a non-O1/O139 strain

providing evidence for the importance of non-O1/O139 strains in driving the evolution of environmental *V. cholerae* strains into one capable of causing pandemics/epidemics. Since then, a 2010 study (Kim *et al.*, 2010) identified specific genes present on the ICE (integrative conjugative element) SXT, responsible for ciprofloxacin resistance in *V. cholerae*. This element was shown to be mobilisable and responsible for the spread of multiple antibiotic resistance in *V. cholerae*, as well as demonstrating a similar function (antibiotic resistance) in *E. coli*. The genomic island studied in this chapter provides protection against ciprofloxacin, reflected by the 4-fold increase in MIC. Moreover, it is mobilisable and thus has the potential to disseminate throughout the aquatic environment from non-O1/O139 cholera strains to pandemic ones and although speculative also within the human gut to normal flora during cholera infection; potentially reducing ciprofloxacin susceptibility in a range of bacterial species. This transfer could have significant impact of the treatment of non-cholera diarrheoal infections as ciprofloxacin is used to treat a wide-range of bacterial infections transmitted *via* the faecal-oral route. This chapter also demonstrates an increase in bleomycin MIC by 2-fold in strains containing RME. Although, bleomycin is not routinely used as a treatment of cholera infections, a genome study by Feng *et al*. (Feng *et al.*, 2008) using O1 pandemic clones of *V. cholerae* showed the presence of a bleomycin resistance protein. Bleomycin is commonly used as an anticancer drug and acts by creating breaks in single-stranded and double-stranded DNA in both mammalian and bacterial DNA (D'Andrea and Haseltine, 1978). Identified resistance proteins in clinical isolates have shown them to be in association with mobile elements, ultimately acting to sequester bleomycin cleavage of DNA (Mori *et al.*, 2008). In conjunction with the results presented in this chapter the suggestion must be raised as to the suitability of bleomycin as a future antibiotic treament for patients with cholera.

Intriguingly, a putative redox sensitive transcriptional activator, SoxR (RME013) is present on RME. SoxR is an activator of a complex regulon in response to superoxide-generating stress (Amabile-Cuevas and Demple, 1991). A study in 2007 shows that all antibiotics (irrespective of their mode of action i.e. targeting cell wall, proteins or DNA) induce free radical production which subsequently leads to the triggering of the SOS response pathway (which is mediated by RecA) and eventually cell death. This study further suggests that

targeting the SOS response; that is the bacterium's ability to repair damaged DNA, may be a new target for anti-bacterial drugs (Kohanski *et al.*, 2007). The presence of a putative *soxR* gene on RME and proven ability of the genomic island in *E. coli* to enhance cell survival against DNA induced damage, suggests RME may provide an adaptive advantage in bacteria upon the induction of free radicals.

5.4: Future directions and conclusions

This study is the first to show that a functional copy of the housekeeping gene *recA* can be mobilised, highlighting the ramifications for accurate construction of phylogenetic relationships. The genomic island described is novel in the sense that it contains many genes on a single mobile genetic element involved in different DNA repair pathways and pathways involving numerous DNA processes. The results of experiments performed in this chapter show this genomic island protects against antibiotic induced damage as well as UV-irradiation. Furthermore, this element is mobilised by means of circularisation, suggesting RME may be capable of disseminating amongst bacteria. It is intriguing to speculate that acquisition of RME could provide a selective advantage against DNA targeting antibiotics in a clinical setting.

Although attempts to transfer RME into another *V. cholerae* strain to observe phenotypic changes as a result of this genomic island in *Vibrio* spp. were unsuccessful, the absence of this data does not detract from the novelty of the identified genomic island (RME) in chapter 5. However, it would be prudent to study the effects of this genomic island in an isogenic mutant where RME has been deleted perhaps by homologous recombination where the genomic island is replaced by a marker of some description such as a kanamycin resistance gene (similar to the isogenic integron deletion mutants studied in chapters 3 and 4).

A handful of genes coding for proteins of unknown function were identified as being present on RME. In order to elucidate the function(s) of these unknown genes it would be useful to employ the transposon mutagenesis method used in this thesis to insertionally inactive $recA_{RME}$ and $umuC_{RME}$ to subsequently inactivate the unknown(s). Once this was

done subjecting these strains to UV-irradiation and DNA targeting antibiotic assays would aid in substantiating their role in DNA repair.

One of the interesting features of RME is its ability to target *recA*. A measure of recombination frequency *via* conduction assays would support the targeting data presented in this chapter. Keeping with this theme as mentioned the RecA protein performs a variety of roles within the cell and is a crucial player in homologous recombination. Therefore, it would be of great interest to compare the recombinational efficiency of $recA_{RME}$ in order to deduce how effective it is compared to *recA* genes in other vibrios. This is interesting not only with regard to the recombination of genetic material within bacteria in an environmental setting, but $recA_{RME}$ could be exploited for use as a molecular tool during cloning procedures to increase recombination efficiencies.

Chapter 6: General Discussion

As discussed in chapter 1 the genomes of many bacterial species including members of the *Vibrio* genus, are in a constant state of change and flux; mainly due to lateral gene transfer events (Ochman *et al.*, 2000). The genomics era has revealed that a remarkable 1.6-32.6% of every bacterial genome shows signatures of recent LGT (Koonin *et al.*, 2001). However, LGT particularly with regards to the integron/gene cassette system, has largely been studied in relation to its direct impact on human behaviour; with antibiotic resistance determinants carried by MGEs being a primary focus.

The impact LGT has had on bacterial adaptation and evolution in the broader context can no longer be segregated into their role in 'environmental' and 'clinical' settings. The results presented in this thesis highlight the marriage of these milieus using two examples of MGEs in two different environmental *Vibrio* species: *V. rotiferianus* DAT722 and *V. cholerae* S24.

6.1: The function of integron-associated gene cassettes in *Vibrio* species: the tip of the iceberg revealed

As highlighted in chapter 1, the integron is a genetic element that incorporates genes termed gene cassettes into a reserved genetic site *via* site-specific recombination (Stokes and Hall, 1989). The integron/gene cassette system is best known for its role in antibiotic resistance with one type of integron, the class 1 integron, being a major contributor of antibiotic resistance in Gram negative pathogens and commensals. Integrons however, are ancient structures with over 100 classes (including the class 1) present in bacteria from the wider environment. While the class 1 integron is only one example of an integron being mobilised into the clinical environment, it is by far the most successful. Unlike class 1 integrons which are largely found on plasmids, other integron classes are located on the chromosome of bacteria and carry a repertoire of gene cassettes, indicating roles(s) involved in phenotypes other than antibiotic resistance. However, there is very limited knowledge on what these alternative roles are. This is particularly relevant to *Vibrio* species where gene cassettes make up approximately 1-3% of the entire genome (Labbate *et al.*,

2009). The strong emphasis of research on the class 1 integron has resulted in a limited understanding by the wider research community on the role integrons play in the broader environments. Furthermore, there is a significant lack of knowledge about how gene cassettes in *Vibrio* species drive adaptation and evolution. *V. rotiferianus* DAT722 is the only organism where extensive physiological analysis has been conducted on isogenic mutants with gene cassettes deleted. This data (chapter 3) has revealed new insights into how gene cassettes might affect adaptation and evolution. In a previous study, a cassette could not be deleted without a compensatory mutation (Labbate *et al.*, 2011). The resulting mutants exhibited abnormal regulation of their porins and impaired growth in minimal salts medium. The gene cassette in question was the 11^{th} cassette from *attI* and appears to be strain specific as it lacks close relatives elsewhere. The cassette 11 protein contains two domains, one with weak homology to nucleases and the other a C4-zinc finger domain commonly found in topoisomerase 1 proteins. These domains indicate a DNA binding/processing protein that is likely to have a regulatory role potentially through controlling the coiling state of gene promoters. Irrespective of the exact role, this study is important in demonstrating that cellular networks can rapidly integrate an incoming mobile gene cassette such that it becomes advantageous for survival and adaptation. It also shows that benefit need not necessarily come from acquisition of a novel functional gene(s) but through modification of existing host cellular networks (Labbate *et al.*, 2012).

In a follow up to this study (results presented in chapter 3), the impact of deletions on the cassette array of *V. rotiferianus* DAT722 was addressed (Rapa *et al.*, 2013). Indels have been regularly observed in *V. cholerae* arrays and are likely in all large arrays. However, their impact on bacterial physiology has not been addressed (Labbate *et al.*, 2007;Szekeres, 2007;Boucher *et al.*, 2011). In chapter 3, three mutants with deletions of gene cassettes 16-60 (mutant d16-60), 50-60 (mutant d50-60) and 72-92 (mutant d72-92) in the array were subjected to physiological growth, stress, proteomic and chemistry-based techniques to determine the effect of contiguous cassette deletions on vibrio physiology. The total deleted cassettes encompassed 58% of the DAT722 array. Surprisingly, growth and stress assays of these mutants showed little change compared to the wild-type. Furthermore, two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) analysis of d16-60 in different

media and growth stages showed only 0.5-1% change in the proteome. This indicates that unlike deletion of cassette 11, the majority of cassettes are not integrated into host pathways and do not affect the major metabolic pathways of the cell, at least under the conditions observed.

Importantly, analyses in chapter 3 did identify changes to host surface polysaccharide in mutants with Proton Nuclear Magnetic Resonance on whole cell polysaccharide from d16-60 and d72-92. These results indicated that gene cassette-associated products decorate host cell polysaccharide *via* the addition or removal of functional groups. Consistent with this result, the d72-92 mutant had modified biofilm-forming capabilities in a simple batch biofilm assay (Rapa *et al.*, 2013). This is a significant result as it focuses future researchers who are addressing gene cassette function in vibrios to surface polysaccharide. This thesis proposes that at least a subset of cassettes are involved in modifying host surface polysaccharide and that deletion (and most likely rearrangements and acquisition) of cassettes may be a mechanism for creating surface property diversity. There is significant biological importance to surface-associated polysaccharide and its modification as evidenced in the literature. This includes biofilm formation (Lee *et al.*, 2013), bacterial cell co-aggregation (Vu *et al.*, 2009), bacteriophage resistance (Scholl *et al.*, 2005), evasion of human immune cells (Pier *et al.*, 2001) as well as resistance to antimicrobial peptides (Westman *et al.*, 2008).

Following on from the results presented in chapter 3, chapter 4 explored the phenotype(s) associated with the expression of the gene carried by cassette 31 (VSD31) from DAT722. The phenotypes observed due to the expression of this gene cassette were shown in the environmental strain of *V. cholerae* S24. Importantly, this chapter confirms three polysaccharide linked phenotypes in *V. cholerae*; biofilm formation, enhanced congo red binding and cell aggregation. These phenotypes are attributed to the expression of VSD31 in S24. Cassette 31 is not indigenous to the S24 strain, thus highlighting the central role a **single** LGT event (i.e. the transfer of VSD31) has in shaping the ability of a strain to form biofilms and increasing the tendency of cells to aggregate/clump (as in the case of S24/pSU-pBAD::VSD31 cells; chapter 4). As brought to the forefront in chapter 4, the ability of toxigenic strains of *V. cholerae* to form aggregates during inter-epidemic periods

of cholera outbreaks is a major driving force in not only the survival and persistence of these strains in the environment, but aggregation actually enhances subsequent virulence in the human host upon activation from their dormant state (Faruque *et al.*, 2006). The importance of aggregation and the findings presented in chapter 4 highlight the interconnected relationship between the 'environmental' and 'clinical' settings, such that a single mobile gene cassette from an environmental *Vibrio* species has the potential capacity to enhance the virulence of the cholera-producing pathogen.

From the research performed in this thesis (chapters 3 and 4) using the integron/gene cassette system of *V. rotiferianus* DAT722, new insight into how gene cassettes affect cellular physiology reveals **novel** roles for genes carried by gene cassettes. In this thesis it was shown that at least a subset of gene cassettes are involved in host surface polysaccharide modification, impacting on adhesion/biofilm formation and bacterial aggregation.

6.2: A genomic island in an environmental strain of *V. cholerae* highlights the importance of the 'environment' in driving adaptation of bacteria

Chapter 5 reports a **novel** genomic island found in an environmental, non-toxigenic strain of *V. cholerae* isolated in Sydney, Australia. This genomic island was unique for three main reasons: 1) it preferentially inserts into the housekeeping gene *recA*, 2) it carries a copy of *recA* so not to disrupt any vital cell pathways involving this essential gene and 3) it houses a number of genes identified to be putatively involved in DNA repair pathways and other DNA processes such as restriction modification. Further to this, chapter 5 demonstrated that this genomic island (RME) protects *E. coli* from two different DNA damaging stimuli: UV-irradiation and DNA targeting antibiotics. These two stimuli emulate two different DNA stressors that members of the *Vibrionaceae* are exposed to in the clinic (DNA damage induced by antibiotics used to treat infections in humans) and the environment (UV-irradiation from the sun in the marine setting). This highlights the importance of the two lifestyles of *V. cholerae* exists within: the pathogenic and environmental (Figure 1.2).

This 'inter-linked' relationship between the environmental and clinical milieus has recently been highlighted in US and Australian mainstream media. Triclosan is an antimicrobial

ingredient (known as a biocide) added to many consumer products such as toothpaste and household soap to reduce bacterial contamination (FDA, 2010). Triclosan acts against the cell *via* multiple non-specific targets at sub-lethal concentrations (Yazdankhah *et al.*, 2006). Studies over the past 10 years have reported bacterial resistance mechanisms to triclosan in a range of micro-organisms (McMurry *et al.*, 1998;Chuanchuen *et al.*, 2001;Fan *et al.*, 2002) bringing to the forefront the over-zealous use of household products containing triclosan. Further to this, it has been suggested that exposure of bacteria to sub-lethal concentrations of this biocide presents an increased risk for developing and promoting cross- or co-resistance to other clinically relevant antibiotics (Chuanchuen *et al.*, 2001;Yazdankhah *et al.*, 2006). Thus, household waste, which ultimately ends up in the environment, represents a 'soup' of resistance determinants potentially transferable to human pathogens (Barbosa and Levy, 2000). So much so, the state of Minnesota in the USA has banned household use of triclosan. This example highlights the importance the environment has in enriching for LGT events, forcing bacteria to adapt.

When linking the current triclosan debate to the studies performed in chapter 5 of this thesis, it can be emphasised how important the environment can be in enriching for genes involved in virulence and pathogenicity-associated phenotypes amongst bacteria. Although RME is a genomic island present in a non-pathogenic strain of *V. cholerae*, results in chapter 5 show it excises from the host genome, and therefore has potential to disseminate throughout the environment where it may be integrated into a pathogenic strain of *V. cholerae,* or even a different bacterial genus/species. Functionality of RME in a different genus of bacteria aside from *Vibrio* is not difficult to propose, as RME, or at the very least $recA_{RME}$ and $umuC_{RME}$ were shown to be functional within *E. coli* (a member of a completely different family of bacteria).

The concept of co-resistance and cross-resistance to antimicrobials (as mentioned in the above paragraph) is an interesting point to make with regards to RME. Co-resistance involves the transfer of multiple genes into the same bacterium and/or the acquisition of mutations in different genetic loci that affect bacterial resistance to a range of antimicrobials (Canton and Ruiz-Garbajosa, 2011). Cross-resistance on the other hand, is the tolerance/resistance to an antimicrobial agent due to the bacterium's exposure to a

different antimicrobial that acts *via* a similar mechanism (e.g. a class of antibiotics that target DNA; quinolones and fluoroquinolones). For example, exposure to nalidixic acid (a quinolone) may reduce the susceptibility of some bacteria to other quinolones (or fluoroquinolones), which has been reported for this class of antimicrobial agent (Sanders, 2001). Studies performed in chapter 5 show that the presence of RME increases the tolerance of its host to the DNA targeting antibiotics ciprofloxacin and bleomycin. As ciprofloxacin is commonly used to treat not only patients presenting with cholera infections but other gastrointestinal infections it is interesting to suggest that the acquisition of RME by a bacterium may reduce its susceptibility to other DNA targeting antimicrobial therapies *via* co-and/or cross-resistance mechanisms. Further to this, when exposed to nalidixic acid (a quinolone targeting DNA gyrase), an increase in spontaneous mutations was recorded for *E. coli* cells containing RME. It is again intriguing to hypothesise that one mutation somewhere in the chromosome may give rise to co-resistance against other antimicrobials.

RME was identified in an environmental strain of *V. cholerae,* where this strain (S24) most likely expresses gene(s) present on this genomic island to enhance its environmental survival. It is likely, however, that in the event RME is transferred into a strain capable of causing cholera, RME will enhance virulence. This study draws the environment and clinical settings closer together, highlighting how identifying how bacteria survive in each setting relevant and applicable to both situations.

6.3: Overall conclusions

By adopting a novel approach to study gene cassette function in their native host by creating isogenic gene cassette deletion mutants, chapter 3 revealed gene cassette-associated product(s) are modifying polysaccharide functional groups associated with the surface of *V. rotiferianus* DAT722. Chapter 4 then attempted to complement three individual gene cassettes (whose associated genes were identified to be putatively involved in polysaccharide biogenesis) into their respective isogenic DAT722 deletion mutants. Following the unsuccessful expression of these genes in DAT722 it was discovered that one cassette (VSD31) when cloned and expressed in the environmental strain of *V. cholerae* S24, biofilm and aggregation phenotypes were significantly affected. These

important bacterial phenotypes have not previously been reported to be associated with genes carried by gene cassettes. Chapter 5 goes on to characterise both the genes and phenotypes associated with a ~32 kb in length genomic island (RME) found in S24. This genomic island not only is unique in its ability to target the housekeeping gene *recA* (and carries its own *recA* gene as a substitute) but it also contained a plethora of genes identified to be putatively involved in DNA repair pathways such as the SOS response and mutagenesis pathway. Furthermore, two genes present on this genomic island; *recA* and *umuC* were found to be functional. Consequently, RME was shown to protect *E. coli* against DNA damage induced by both UV-irradiation and DNA targeting antibiotics.

Although the studies performed in this thesis highlight the diversity two MGEs (the integron and a novel genomic island) bring to members of the *Vibrio* genus, this study only reveals the tip of the iceberg. In the past, little progress has been made in the precise mechanisms gene cassette products contribute to adaptation and evolution within *Vibrio* species. Similarly, like gene cassettes focus on the function of other MGEs such as genomic islands has largely been studied with regards to their clinical significance. However, if we are to learn more about the roles MGEs play in shaping bacterial evolution, researcher focus needs to shift to studying the impact LGT had in a broader sense. Subsequently, placing laterally acquired DNA into context within the clinical setting.

References:

Abuaita, B.H., and Withey, J.H. **(2009)**. Bicarbonate induces *Vibrio cholerae* virulence gene expression by enhancing ToxT activity. *Infect Immun* 77**,** 4111-4120.

Aertsen, A., and Michiels, C.W. **(2007)**. The high-pressure shock response in *Escherichia coli*: a short survey. *High Pressure Res* 27**,** 121-124.

Agren, D., Schnell, R., Oehlmann, W., Singh, M., and Schneider, G. **(2008)**. Cysteine Synthase (CysM) of *Mycobacterium tuberculosis* Is an O-Phosphoserine Sulfhydrylase. *J Biol Chem* 283**,** 31567-31574.

Alanis, A.J. **(2005)**. Resistance to antibiotics: are we in the post-antibiotic era? *Arch Med Res* 36**,** 697-705.

Allen, J.G., Atherton, F.R., Hall, M.J., Hassall, C.H., Holmes, S.W., Lambert, R.W., Nisbet, L.J., and Ringrose, P.S. **(1979)**. Phosphonopeptides as antibacterial agents: alaphosphin and related phosphonopeptides. *Antimicrob Agents Chemother* 15**,** 684-695.

Amabile-Cuevas, C.F., and Demple, B. **(1991)**. Molecular characterization of the *soxRS* genes of *Escherichia coli*: two genes control a superoxide stress regulon. *Nucleic Acids Res* 19**,** 4479-4484.

Apicella, M.A. **(2008)**. Isolation and characterization of lipopolysaccharides. *Methods in Molecular Biology (Clifton, N.J.)* 431**,** 3-13.

Aranda, J., Bardina, C., Beceiro, A., Rumbo, S., Cabral, M.P., Barbe, J., and Bou, G. **(2011)**. *Acinetobacter baumannii* RecA protein in repair of DNA damage, antimicrobial resistance, general stress response, and virulence. *J Bacteriol* 193**,** 3740-3747.

Arcondeguy, T., Jack, R., and Merrick, M. **(2001)**. P-II signal transduction proteins, pivotal players in microbial nitrogen control. *Microbiol Mol Biol Rev* 65**,** 80.

Aziz, R.K., Bartels, D., Best, A.A., Dejongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., Mcneil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. **(2008)**. The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* 9**,** 15.

Bach, S., De Almeida, A., and Carniel, E. **(2000)**. The *Yersinia* high-pathogenicity island is present in different members of the family *Enterobacteriaceae*. *FEMS Microbiol Lett* 183**,** 289-294.

Bapteste, E., Boucher, Y., Leigh, J., and Doolittle, W.F. **(2004)**. Phylogenetic reconstruction and lateral gene transfer. *Trends Microbiol* 12**,** 406-411.

Barbosa, T.M., and Levy, S.B. **(2000)**. The impact of antibiotic use on resistance development and persistence. *Drug Resist Updat* 3**,** 303-311.

Bari, S.M., Roky, M.K., Mohiuddin, M., Kamruzzaman, M., Mekalanos, J.J., and Faruque, S.M. **(2013)**. Quorum-sensing autoinducers resuscitate dormant *Vibrio cholerae* in environmental water samples. *Proc Natl Acad Sci U S A* 110**,** 9926-9931.

Barker, A., Clark, C.A. **(1994)**. Identification of VCR, a repeated sequence associated with a locus encoding a hemagglutinin in *Vibrio cholerae* O1. *J Bacteriol* 176**,** 5450-5458.

Beaber, J.W., Hochhut, B., and Waldor, M.K. **(2004)**. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 427**,** 72-74.

Bennett, P.M. **(2008)**. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *Br J Pharmacol* 153 Suppl 1**,** S347-357.

Bik, E.M., Bunschoten, A.E., Gouw, R.D., and Mooi, F.R. **(1995)**. Genesis of the novel epidemic *Vibrio cholerae* O139 strain- evidence for horizontal transfer of genes involved in polysaccharide synthesis. *EMBO J* 14**,** 209-216.

Bikard, D., Julie-Galau, S., Cambray, G., and Mazel, D. **(2010)**. The synthetic integron: an in vivo genetic shuffling device. *Nucleic Acids Res* 38.

Biskri, L., Bouvier, M., Guerout, A.M., Boisnard, S., and Mazel, D. **(2005)**. Comparative study of class 1 integron and *Vibrio cholerae* superintegron integrase activities. *J Bacteriol* 187**,** 1740-1750.

Blokesch, M., and Schoolnik, G.K. **(2007)**. Serogroup conversion of *Vibrio cholerae* in aquatic reservoirs. *PLoS Pathog* 3**,** e81.

Boltner, D., Macmahon, C., Pembroke, J.T., Strike, P., and Osborn, A.M. **(2002)**. R391: a conjugative integrating mosaic comprised of phage, plasmid, and transposon elements. *J Bacteriol* 184**,** 5158-5169.

Boucher, Y., Cordero, O.X., Takemura, A., Hunt, D.E., Schliep, K., Bapteste, E., Lopez, P., Tarr, C.L., and Polz, M.F. **(2011)**. Local mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio cholerae* populations. *mBio* 2.

Boucher, Y., Labbate, M., Koenig, J.E., and Stokes, H.W. **(2007)**. Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol* 15**,** 301-309.

Boucher, Y., Nesbo, C., Joss, M., Robinson, A., Mabbutt, B., Gillings, M., Doolittle, W.F., and Stokes, H.W. **(2006)**. Recovery and evolutionary analysis of complete integron gene cassette arrays from *Vibrio*. *BMC Evol Biol* 6**,** 3.

Boucher, Y., Stokes.W.H. **(2006)**. "The roles of lateral gene transfer and vertical descent in vibrio evolution," in *The Biology of Vibrios,* ed. B.A. Fabiano Lopes Thompson, J. G. Swings.).

Bourne, D.G., Garren, M., Work, T.M., Rosenberg, E., Smith, G.W., and Harvell, C.D. **(2009)**. Microbial disease and the coral holobiont. *Trends Microbiol* 17**,** 554-562.

Bouvier, M., Demarre, G., and Mazel, D. **(2005)**. Integron cassette insertion: a recombination process involving a folded single strand substrate. *EMBO J* 24**,** 4356-4367.

Boyd, E.F., Almagro-Moreno, S., and Parent, M.A. **(2009)**. Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends Microbiol* 17**,** 47-53.

Boyd, E.F., Moyer, K.E., Shi, L., and Waldor, M.K. **(2000)**. Infectious CTXPhi and the *Vibrio* pathogenicity island prophage in *Vibrio mimicus*: evidence for recent horizontal transfer between *V. mimicus* and *V. cholerae*. *Infect Immun* 68**,** 1507-1513.

Brassinga, A.K., Hiltz, M.F., Sisson, G.R., Morash, M.G., Hill, N., Garduno, E., Edelstein, P.H., Garduno, R.A., and Hoffman, P.S. **(2003)**. A 65-kilobase pathogenicity island is unique to Philadelphia-1 strains of *Legionella pneumophila*. *J Bacteriol* 185**,** 4630-4637.

Brown, N.L., and Evans, L.R. **(1991)**. Transposition in prokaryotes: transposon Tn501. *Res Microbiol* 142**,** 689-700.

Brussow, H., Canchaya, C., and Hardt, W.D. **(2004)**. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* 68**,** 560-602.

Burrus, V., Marrero, J., and Waldor, M.K. **(2006)**. The current ICE age: biology and evolution of SXT-related integrating conjugative elements. *Plasmid* 55**,** 173-183.

Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. **(2002)**. Conjugative transposons: the tip of the iceberg. *Mol Microbiol* 46**,** 601-610.

Burrus, V., and Waldor, M.K. **(2004)**. Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiology* 155**,** 376-386.

Cambray, G., Guerout, A.M., and Mazel, D. **(2010)**. Integrons. *Annu Rev Genet* 44**,** 141-166.

Canton, R., and Ruiz-Garbajosa, P. **(2011)**. Co-resistance: an opportunity for the bacteria and resistance genes. *Curr Opin Pharmacol* 11**,** 477-485.

Carnahan, A.M., Harding, J., Watsky, D., and Hansman, S. **(1994)**. Identification of *Vibrio hollisae* associated with severe gastroenteritis after consumption of raw oysters. *J Clin Microbiol* 32**,** 1805-1806.

Centers for Disease Control and Prevention **(2011)**. Haiti cholera training manual:a full course for healthcare providers.

Centers for Disease Control and Prevention **(2013)**. http://www.cdc.gov/vibrio/vibriov.html.

Ceroni, A., Passerini, A., Vullo, A., and Frasconi, P. **(2006)**. DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Res* 34**,** W177-W181.

Chander, J., Kaistha, N., Gupta, V., Mehta, M., Singla, N., Deep, A., and Sarkar, B.L. **(2009)**. Epidemiology & antibiograms of *Vibrio cholerae* isolates from a tertiary care hospital in Chandigarh, North India. *Indian J Med Res* 129**,** 613-617.

Chang, B., Taniguchi, H., Miyamoto, H., and Yoshida, S. **(1998)**. Filamentous bacteriophages of *Vibrio parahaemolyticus* as a possible clue to genetic transmission. *J Bacteriol* 180**,** 5094-5101.

Chen, I., Christie, P.J., and Dubnau, D. **(2005)**. The ins and outs of DNA transfer in bacteria. *Science* 310**,** 1456-1460.

Chen, Y., Dai, J., Morris, J.G., Jr., and Johnson, J.A. **(2010)**. Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio parahaemolyticus* O3:K6. *BMC Microbiol* 10.

Chevalier, F. **(2010)**. Highlights on the capacities of "Gel-based" proteomics. *Proteome Science* 8**,** 23.

Chibani-Chennoufi, S., Bruttin, A., Dillmann, M.-L., and Brussow, H. **(2004)**. Phage-host interaction: an ecological perspective. *J. Bacteriol.* 186**,** 3677-3686.

Chowdhury, P.R., Boucher, Y., Hassan, K.A., Paulsen, I.T., Stokes, H.W., and Labbate, M. **(2011)**. Genome sequence of *Vibrio rotiferianus* strain DAT722. *J Bacteriol* 193**,** 3381-3382.

Chuanchuen, R., Beinlich, K., Hoang, T.T., Becher, A., Karkhoff-Schweizer, R.R., and Schweizer, H.P. **(2001)**. Cross-resistance between triclosan and antibiotics in *Pseudomonas aeruginosa* is mediated by multidrug efflux pumps: exposure of a susceptible mutant strain to triclosan selects *nfxB* mutants overexpressing MexCD-OprJ. *Antimicrob Agents Chemother* 45**,** 428-432.

Chun, J., Grim, C.J., Hasan, N.A., Lee, J.H., Choi, S.Y., Haley, B.J., Taviani, E., Jeon, Y.S., Kim, D.W., Brettin, T.S., Bruce, D.C., Challacombe, J.F., Detter, J.C., Han, C.S., Munk, A.C., Chertkov, O., Meincke, L., Saunders, E., Walters, R.A., Huq, A., Nair, G.B., and Colwell, R.R. **(2009)**. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc Natl Acad Sci U S A* 106**,** 15442-15447.

Clinical and Laboratory Standards Institute **(2003)**. *Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically.*

Collis, C.M., and Hall, R.M. **(1992)**. Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. *J Bacteriol* 174**,** 1574-1585.

Colvin, K.M., Gordon, V.D., Murakami, K., Borlee, B.R., Wozniak, D.J., Wong, G.C.L., and Parsek, M.R. **(2011)**. The Pel polysaccharide can serve a structural and protective role in the biofilm matrix of *Pseudomonas aeruginosa*. *PLoS Pathog* 7**,** e1001264.

Creevey, C.J., Fitzpatrick, D.A., Philip, G.K., Kinsella, R.J., O'connell, M.J., Pentony, M.M., Travers, S.A., Wilkinson, M., and Mcinerney, J.O. **(2004)**. Does a tree-like phylogeny only exist at the tips in the prokaryotes? *Proc Biol Sci* 271**,** 2551-2558.

D'andrea, A.D., and Haseltine, W.A. **(1978)**. Sequence specific cleavage of DNA by the antitumor antibiotics neocarzinostatin and bleomycin. *Proc Natl Acad Sci U S A* 75**,** 3608-3612.

Davis, B.M., and Waldor, M.K. **(2003)**. Filamentous phages linked to virulence of *Vibrio cholerae*. *Current Opinion in Microbiology* 6**,** 35-42.

Demarre, G., Guerout, A.M., Matsumoto-Mashimo, C., Rowe-Magnus, D.A., Marliere, P., and Mazel, D. **(2005)**. A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalpha) conjugative machineries and their cognate *Escherichia coli* host strains. *Res Microbiol* 156**,** 245-255.

Deshpande, C.N., Harrop, S.J., Boucher, Y., Hassan, K.A., Di Leo, R., Xu, X., Cui, H., Savchenko, A., Chang, C., Labbate, M., Paulsen, I.T., Stokes, H.W., Curmi, P.M.G., and Mabbutt, B.C. **(2011)**. Crystal structure of an integron gene cassette-associated protein from *Vibrio cholerae* identifies a cationic drug-binding module. *PLoS ONE* 6.

Dobrindt, U., Agerer, F., Michaelis, K., Janka, A., Buchrieser, C., Samuelson, M., Svanborg, C., Gottschalk, G., Karch, H., and Hacker, J. **(2003)**. Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *J Bacteriol* 185**,** 1831-1840.

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. **(2004)**. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2**,** 414-424.

Donnenberg, M.S., and Kaper, J.B. **(1991)**. Construction of an *eae* deletion mutant of enteropathogenic *Escherichia coli* by using a positive-selection suicide vector. *Infect. Immun.* 59**,** 4310-4317.

Dziejman, M., Balon, E., Boyd, D., Fraser, C.M., Heidelberg, J.F., and Mekalanos, J.J. **(2002)**. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. *Proc Natl Acad Sci U S A* 99**,** 1556-1561.

Ehara, M., Nguyen, B.M., Nguyen, D.T., Toma, C., Higa, N., and Iwanaga, M. **(2004)**. Drug susceptibility and its genetic basis in epidemic *Vibrio cholerae* O1 in Vietnam. *Epidemiol Infect* 132**,** 595-600.

Elsaied, H., Stokes, H.W., Kitamura, K., Kurusu, Y., Kamagata, Y., and Maruyama, A. **(2011)**. Marine integrons containing novel integrase genes, attachment sites, attI, and associated gene cassettes in polluted sediments from Suez and Tokyo Bays. *ISME J* 5**,** 1162-1177.

Elsaied, H., Stokes, H.W., Nakamura, T., Kitamura, K., Fuse, H., and Maruyama, A. **(2007)**. Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents. *Environ Microbiol* 9**,** 2298-2312.

Elsaied, H., Stokes, H.W., Yoshioka, H., Mitani, Y., and Maruyama, A. **(2014)**. Novel integrons and gene cassettes from a Cascadian submarine gas-hydrate-bearing core. *FEMS Microbiol Ecol* 87**,** 343-356.

Enos-Berlage, J.L., and Mccarter, L.L. **(2000)**. Relation of capsular polysaccharide production and colonial cell organization to colony morphology in *Vibrio parahaemolyticus*. *J Bacteriol* 182**,** 5513-5520.

Fan, F., Yan, K., Wallis, N.G., Reed, S., Moore, T.D., Rittenhouse, S.F., Dewolf, W.E., Jr., Huang, J., Mcdevitt, D., Miller, W.H., Seefeld, M.A., Newlander, K.A., Jakas, D.R., Head, M.S., and Payne, D.J. **(2002)**. Defining and combating the mechanisms of triclosan resistance in clinical isolates of *Staphylococcus aureus*. *Antimicrob Agents Chemother* 46**,** 3343-3347.

Faruque, S.M., Biswas, K., Udden, S.M., Ahmad, Q.S., Sack, D.A., Nair, G.B., and Mekalanos, J.J. **(2006)**. Transmissibility of cholera: in vivo-formed biofilms and their relationship to infectivity and persistence in the environment. *Proc Natl Acad Sci U S A* 103**,** 6350-6355.

Faruque, S.M., Chowdhury, N., Kamruzzaman, M., Dziejman, M., Rahman, M.H., Sack, D.A., Nair, G.B., and Mekalanos, J.J. **(2004)**. Genetic diversity and virulence potential of environmental *Vibrio cholerae* population in a cholera-endemic area. *Proc Natl Acad Sci U S A* 101**,** 2123-2128.

Faruque, S.M., Islam, M.J., Ahmad, Q.S., Faruque, A.S., Sack, D.A., Nair, G.B., and Mekalanos, J.J. **(2005a)**. Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. *Proc Natl Acad Sci U S A* 102**,** 6119-6124.

Faruque, S.M., Kamruzzaman, M., Meraj, I.M., Chowdhury, N., Nair, G.B., Sack, R.B., Colwell, R.R., and Sack, D.A. **(2003)**. Pathogenic potential of environmental *Vibrio cholerae* strains carrying genetic variants of the toxin-coregulated pilus pathogenicity island. *Infect Immun* 71**,** 1020-1025.

Faruque, S.M., and Mekalanos, J.J. **(2003)**. Pathogenicity islands and phages in *Vibrio cholerae* evolution. *Trends Microbiol* 11**,** 505-510.

Faruque, S.M., and Mekalanos, J.J. **(2012)**. Phage-bacterial interactions in the evolution of toxigenic *Vibrio cholerae*. *Virulence* 3**,** 556-565.

Faruque, S.M., Naser, I.B., Islam, M.J., Faruque, A.S., Ghosh, A.N., Nair, G.B., Sack, D.A., and Mekalanos, J.J. **(2005b)**. Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. *Proc Natl Acad Sci U S A* 102**,** 1702-1707.

Feng, L., Reeves, P.R., Lan, R., Ren, Y., Gao, C., Zhou, Z., Cheng, J., Wang, W., Wang, J., Qian, W., Li, D., and Wang, L. **(2008)**. A recalibrated molecular clock and independent origins for the cholera pandemic clones. *PLoS One* 3**,** e4053.

Fong, J.C.N., Syed, K.A., Klose, K.E., and Yildiz, F.H. **(2010)**. Role of *Vibrio* polysaccharide (*vps*) genes in VPS production, biofilm formation and *Vibrio cholerae* pathogenesis. *Microbiology* 156**,** 2757-2769.

Food and Drug Administration. **2010**. Triclosan: what consumers should know. [Accessed 30 May, 2014].

Frans, I., Michiels, C.W., Bossier, P., Willems, K.A., Lievens, B., and Rediers, H. **(2011)**. *Vibrio anguillarum* as a fish pathogen: virulence factors, diagnosis and prevention. *J Fish Dis* 34**,** 643-661.

Franzon, V., Barker, A., and Manning, P. **(1993)**. Nucleotide sequence encoding the mannose-fucose-resistant hemagglutinin of *Vibrio cholerae* O1 and contstruction of a mutant. *Infect Immun* 61**,** 3032-3037.

Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. **(2005)**. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 3**,** 722-732.

Garg, P., Sinha, S., Chakraborty, R., Bhattacharya, S.K., Nair, G.B., Ramamurthy, T., and Takeda, Y. **(2001)**. Emergence of fluoroquinolone-resistant strains of *Vibrio cholerae* O1 biotype El Tor among hospitalized patients with cholera in Calcutta, India. *Antimicrob Agents Chemother* 45**,** 1605-1606.

Gellert, M., Mizuuchi, K., O'dea, M.H., Itoh, T., and Tomizawa, J.I. **(1977)**. Nalidixic acid resistance: a second genetic character involved in DNA gyrase activity. *Proc Natl Acad Sci U S A* 74**,** 4772-4776.

Giana, H., Silveira, L., Jr., Zângaro, R., and Pacheco, M. **(2003)**. Rapid Identification of Bacterial Species by Fluorescence Spectroscopy and Classification Through Principal Components Analysis. *Journal of Fluorescence* 13**,** 489-493.

Gildemeister, O.S., Zhu, B.C., and Laine, R.A. **(1994)**. Chitovibrin: a chitin-binding lectin from *Vibrio parahemolyticus*. *Glycoconj J* 11**,** 518-526.

Gillings, M., Boucher, Y., Labbate, M., Holmes, A., Krishnan, S., Holley, M., and Stokes, H.W. **(2008)**. The evolution of class 1 integrons and the rise of antibiotic resistance. *J. Bacteriol.* 190**,** 5095-5100.

Gillings, M., and Hagan-Lawson, E. **(2014)**. The cost of living in the Anthropocene. *Earth Perspectives* 1**,** 2.

Gillings, M., Holley Marita, Stokes H.W., Holmes Andrew **(2005)**. Integrons in *Xanthomonas*: A source of species genome diversity. *Proc Natl Acad Sci U S A* 102**,** 4419-4424.

Gillings, M.R. **(2013)**. Evolutionary consequences of antibiotic use for the resistome, mobilome and microbial pangenome. *Front Microbiol* 4**,** 4.

Gillings, M.R., and Stokes, H.W. **(2012)**. Are humans increasing bacterial evolvability? *Trends Ecol Evol* 27**,** 346-352.

Gomez-Gil, B., Thompson, F.L., Thompson, C.C., and Swings, J. **(2003)**. *Vibrio rotiferianus* sp nov., isolated from cultures of the rotifer *Brachionus plicatilis*. *Int J Syst Evol Micr* 53**,** 239-243.

Goodman, M.F. **(2002)**. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* 71**,** 17-50.

Goss, W.A., Deitz, W.H., and Cook, T.M. **(1965)**. Mechanism of action of nalidixic acid on *Escherichia coli*. Inhibition of deoxyribonucleic acid synthesis *J Bacteriol* 89**,** 1068-1074.

Grimes, D., Johnson, C., Dillon, K., Flowers, A., Noriea, N., and Berutti, T. **(2009)**. What genomic sequence information has revealed about *Vibrio* ecology in the ocean-A review. *Microb Ecol* 58**,** 447-460.

Guerin, E., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Da Re, S., Gonzalez-Zorn, B., Barbe, J., Ploy, M.C., and Mazel, D. **(2009)**. The SOS response controls integron recombination. *Science* 324**,** 1034-1034.

Hall, R.M., Brookes, D.E., and Stokes, H.W. **(1991)**. Site-specific insertion of genes into integrons: role of the 59-base element and determination of the recombination cross-over point. *Mol Microbiol* 5**,** 1941-1959.

Hare, J.M., Bradley, J.A., Lin, C.L., and Elam, T.J. **(2012)**. Diverse responses to UV light exposure in *Acinetobacter* include the capacity for DNA damage-induced mutagenesis in the opportunistic pathogens *Acinetobacter baumannii* and *Acinetobacter ursingii*. *Microbiology* 158**,** 601-611.

Hastings, P.J., Rosenberg, S.M., and Slack, A. **(2004)**. Antibiotic-induced lateral transfer of antibiotic resistance. *Trends Microbiol* 12**,** 401-404.

Hazen, T.H., Pan, L., Gu, J.D., and Sobecky, P.A. **(2010)**. The contribution of mobile genetic elements to the evolution and ecology of Vibrios. *FEMS Microbiol Ecol* 74**,** 485-499.

Heppelmann, C.J., Benson, L.M., and Bergen, H.R. **(2007)**. A simple method to remove contaminating salt from IPG strips prior to IEF. *Electrophoresis* 28**,** 3988-3991.

Hinnebusch, J., and Tilly, K. **(1993)**. Linear plasmids and chromosomes in bacteria. *Mol Microbiol* 10**,** 917-922.

Hochhut, B., Jahreis, K., Lengeler, J.W., and Schmid, K. **(1997)**. CTnscr94, a conjugative transposon found in enterobacteria. *J Bacteriol* 179**,** 2097-2102.

Holmes, A.J., Gillings, M.R., Nield, B.S., Mabbutt, B.C., Nevalainen, K.M.H., and Stokes, H.W. **(2003)**. The gene cassette metagenome is a basic resource for bacterial genome evolution. *Environ Microbiol* 5**,** 383-394.

Hooper, L.V., and Gordon, J.I. **(2001)**. Commensal host-bacterial relationships in the gut. *Science* 292**,** 1115-1118.

Houot, L., Chang, S., Absalon, C., and Watnick, P.I. **(2010)**. *Vibrio cholerae* phosphoenolpyruvate phosphotransferase system control of carbohydrate transport, biofilm formation and colonization of the germfree mouse intestine. *Infect Immun* 78**,** 1482-1494.

Igbinosa, E.O., and Okoh, A.I. **(2008)**. Emerging *Vibrio* species: an unending threat to public health in developing countries. *Res Microbiol* 159**,** 495-506.

Islam, A., Labbate, M., Djordjevic, S.P., Alam, M., Darling, A., Melvold, J., Holmes, A.J., Johura, F.T., Cravioto, A., Charles, I.G., and Stokes, H.W. **(2013)**. Indigenous *Vibrio cholerae* strains from a non-endemic region are pathogenic. *Open Biol* 3**,** 120181.

Ito, T., Katayama, Y., Asada, K., Mori, N., Tsutsumimoto, K., Tiensasitorn, C., and Hiramatsu, K. **(2001)**. Structural comparison of three types of staphylococcal cassette chromosome *mec* integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* 45**,** 1323-1336.

Ito, T., Katayama, Y., and Hiramatsu, K. **(1999)**. Cloning and nucleotide sequence determination of the entire *mec* DNA of pre-methicillin-resistant *Staphylococcus aureus* N315. *Antimicrob Agents Chemother* 43**,** 1449-1458.

Ivanova, E., Kuznetsova, T., and Mikhailov, V. **(2000)**. Antifungal compounds produced by bacterial populations of the Sea Sponge *Verongia* sp. *Russ J of Mar Biol* 26**,** 118-119.

Jabeen, K., Zafar, A., and Hasan, R. **(2008)**. Increased isolation of *Vibrio cholerae* O1 serotype Inaba over serotype Ogawa in Pakistan. *East Mediterr Health J* 14**,** 564-570.

Janion, C. **(2008)**. Inducible SOS response system of DNA repiar and mutagenesis in *Escherichia coli*. *Int J Biol Sci* 4**,** 338-344.

Jensen, M.A., Faruque, S.M., Mekalanos, J.J., and Levin, B.R. **(2006)**. Modeling the role of bacteriophage in the control of cholera outbreaks. *Proc Natl Acad Sci U S A* 103**,** 4652-4657.

Jiang, Q., Karata, K., Woodgate, R., Cox, M.M., and Goodman, M.F. **(2009)**. The active form of DNA polymerase V is UmuD'(2)C-RecA-ATP. *Nature* 460**,** 359-363.

Jiang, S.C., and Paul, J.H. **(1998)**. Gene Transfer by Transduction in the Marine Environment. *Appl. Environ. Microbiol.* 64**,** 2780-2787.

Johnson, J.A., Salles, C.A., Panigrahi, P., Albert, M.J., Wright, A.C., Johnson, R.J., and Morris, J.G. **(1994)**. *Vibrio cholerae* O139 synonym Bengal is closely related to *Vibrio cholerae* El Tor by has important differences. *Infect Immun* 62**,** 2108-2110.

Joseph, L.A., and Wright, A.C. **(2004)**. Expression of *Vibrio vulnificus* capsular polysaccharide inhibits biofilm formation. *J Bacteriol* 186**,** 889-893.

Judson, N., and Mekalanos, J.J. **(2000)**. TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. *Nat Biotechnol* 18**,** 740-745.

Juhas, M., Van Der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., and Crook, D.W. **(2009)**. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* 33**,** 376-393.

Kabir, S. **(1982)**. Characterization of the lipopolysaccharide from *Vibrio cholerae* 395 (Ogawa). *Infect Immun* 38**,** 1263-1272.

Kaper, J.B., Morris, J.G.J., and Levine, M.M. **(1995)**. Cholera [published erratum appears in Clin Microbiol Rev 1995 Apr;8(2):316]. *Clin Microbiol Rev* 8**,** 48-86.

Kessler, B., De Lorenzo, V., and Timmis, K.N. **(1992)**. A general system to integrate *lacZ* fusions into the chromosomes of gram-negative eubacteria: regulation of the Pm promoter of the TOL plasmid studied with all controlling elements in monocopy. *Mol Gen Genet* 233**,** 293-301.

Khlebnikov, A., Risa, O., Skaug, T., Carrier, T.A., and Keasling, J.D. **(2000)**. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J Bacteriol* 182**,** 7029-7034.

Kim, H.B., Wang, M., Ahmed, S., Park, C.H., Larocque, R.C., Faruque, A.S., Salam, M.A., Khan, W.A., Qadri, F., Calderwood, S.B., Jacoby, G.A., and Hooper, D.C. **(2010)**. Transferable quinolone resistance in *Vibrio cholerae*. *Antimicrob Agents Chemother* 54**,** 799-803.

Kim, Y.R., Lee, S.E., Kim, C.M., Kim, S.Y., Shin, E.K., Shin, D.H., Chung, S.S., Choy, H.E., Progulske-Fox, A., Hillman, J.D., Handfield, M., and Rhee, J.H. **(2003)**.

Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. *Infect Immun* 71**,** 5461-5471.

Kitaoka, M., Miyata, S.T., Unterweger, D., and Pukatzki, S. **(2011)**. Antibiotic resistance mechanisms of *Vibrio cholerae*. *J Med Microbiol* 60**,** 397-407.

Kittelberger, R., and Hilbink, F. **(1993)**. Sensitive silver-staining detection of bacterial lipopolysaccharides in polyacrylamide gels. *J Biochem Bioph Meth* 26**,** 81-86.

Knetsch, C.W., Hensgens, M.P., Harmanus, C., Van Der Bijl, M.W., Savelkoul, P.H., Kuijper, E.J., Corver, J., and Van Leeuwen, H.C. **(2011)**. Genetic markers for *Clostridium difficile* lineages linked to hypervirulence. *Microbiology* 157**,** 3113-3123.

Koenig, J.E., Boucher, Y., Charlebois, R.L., Nesbo, C., Zhaxybayeva, O., Bapteste, E., Spencer, M., Joss, M.J., Stokes, H.W., and Doolittle, W.F. **(2008)**. Integron-associated gene cassettes in Halifax Harbour: assessment of a mobile gene pool in marine sediments. *Environ Microbiol* 10**,** 1024-1038.

Koenig, J.E., Bourne, D.G., Curtis, B., Dlutek, M., Stokes, H.W., Doolittle, W.F., and Boucher, Y. **(2011)**. Coral-mucus-associated *Vibrio* integrons in the Great Barrier Reef: genomic hotspots for environmental adaptation. *ISME J* 5**,** 962-972.

Koenig, J.E., Sharp, C., Dlutek, M., Curtis, B., Joss, M., Boucher, Y., and Doolittle, W.F. **(2009)**. Integron gene cassettes and degradation of compounds associated with industrial waste: the case of the Sydney tar ponds. *PLoS ONE* 4**,** e5276.

Kohanski, M.A., Dwyer, D.J., Hayete, B., Lawrence, C.A., and Collins, J.J. **(2007)**. A common mechanism of cellular death induced by bactericidal antibiotics. *Cell* 130**,** 797-810.

Komp, L.P., Karlsson, A., and Hughes, D. **(2003)**. Mutation rate and evolution of fluoroquinolone resistance in *Escherichia coli i*solates from patients with urinary tract infections. *Antimicrob Agents Chemother* 47**,** 3222-3232.

Koonin, E.V., Makarova, K.S., and Aravind, L. **(2001)**. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* 55**,** 709-742.

Kotetishvili, M., Stine, O.C., Chen, Y., Kreger, A., Sulakvelidze, A., Sozhamannan, S., and Morris, J.G., Jr. **(2003)**. Multilocus sequence typing has better discriminatory ability for typing *Vibrio cholerae* than does pulsed-field gel electrophoresis and provides a measure of phylogenetic relatedness. *J. Clin. Microbiol.* 41**,** 2191-2196.

Kuadkitkan, A., Wikan, N., Fongsaran, C., and Smith, D.R. Identification and characterization of prohibitin as a receptor protein mediating DENV-2 entry into insect cells. *Virology* 406**,** 149-161.

Labbate, M., Boucher, Y., Joss, M.J., Michael, C.A., Gillings, M.R., and Stokes, H.W. **(2007)**. Use of chromosomal integron arrays as a phylogenetic typing system for *Vibrio cholerae* pandemic strains. *Microbiology* 153**,** 1488-1498.

Labbate, M., Boucher, Y., Luu, I., Roy Chowdhury, P., and Stokes, H.W. **(2012)**. Integron associated mobile genes: Just a collection of plug in apps or essential components of cell network hardware? *Mob Genet Elem* In Press.

Labbate, M., Boucher, Y., Roy Chowdhury, P., and Stokes, H. **(2011)**. Integration of a laterally acquired gene into a cell network important for growth in a strain of *Vibrio rotiferianus*. *BMC Microbiol* 11**,** 253.

Labbate, M., Case, R.J., and Stokes, H.W. **(2009)**. The integron/gene cassette system: an active player in bacterial adaptation. *Methods Mol Biol* 532**,** 103-125.

Labrie, S.J., Samson, J.E., and Moineau, S. **(2010)**. Bacteriophage resistance mechanisms. *Nat Rev Microbiol* 8**,** 317-327.

Lange, V., Picotti, P., Domon, B., and Aebersold, R. **(2008)**. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* 4.

Lawrence, J.G. **(1999)**. Gene transfer, speciation, and the evolution of bacterial genomes. *Curr Opin Microbiol* 2**,** 519-523.

Le Roux, F., Binesse, J., Saulnier, D., and Mazel, D. **(2007)**. Construction of a *Vibrio splendidus* mutant lacking the metalloprotease gene vsm by use of a novel counterselectable suicide vector. *Appl Environ Microbiol* 73**,** 777-784.

Le Roux, F., Davis, B.M., and Waldor, M.K. **(2011)**. Conserved small RNAs govern replication and incompatibility of a diverse new plasmid family from marine bacteria. *Nucleic Acids Res* 39**,** 1004-1013.

Lee, K.J., Kim, J.A., Hwang, W., Park, S.J., and Lee, K.H. **(2013)**. Role of capsular polysaccharide (CPS) in biofilm formation and regulation of CPS production by quorum-sensing in *Vibrio vulnificus*. *Mol Microbiol* 90**,** 841-857.

Lenhart, J.S., Schroeder, J.W., Walsh, B.W., and Simmons, L.A. **(2012)**. DNA repair and genome maintenance in *Bacillus subtilis*. *Microbiol Mol Biol Rev* 76**,** 530-564.

Lin, K., and Wang, A. **(2001)**. UV mutagenesis in *Escherichia coli* K-12: Cell survival and mutation frequency of the chromosomal genes *lacZ*, *rpoB*, *ompF*, and *ampA*. *J Exp Microbiol Immun* 1**,** 32-46.

Luigi, V., Brettar, I., Pezzati, E., Reid, P.C., Colwell, R.R., Hofle, M.G., and Pruzzo, C. **(2011)**. Long-term effects of ocean warming on the prokaryotic community: evidence from the vibrios. *ISME J* 11.

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

Marechal, J., Clement, B., Nalin, R., Gandon, C., Orso, S., Cvejic, J.H., Bruneteau, M., Berry, A., and Normand, P. **(2000)**. A *recA* gene phylogenetic analysis confirms the close proximity of *Frankia* to *Acidothermus*. *Int J Syst Evol Microbiol* 50 Pt 2**,** 781-785.

Marvig, R.L., and Blokesch, M. **(2010)**. Natural transformation of *Vibrio cholerae* as a tool - optimizing the procedure. *BMC Microbiol* 10.

Mazel, D. **(2006)**. Integrons: agents of bacterial evolution. *Nature Reviews Microbiology* 4**,** 608-620.

Mazel, D., Dychinco, B., Webb, V.A., and Davies, J. **(1998)**. A distinctive class of integron in the *Vibrio cholerae* genome. *Science* 280**,** 605-608.

Mckenzie, G.J., Harris, R.S., Lee, P.L., and Rosenberg, S.M. **(2000)**. The SOS response regulates adaptive mutation. *Proc Natl Acad Sci U S A* 97**,** 6646-6651.

McMurry, L.M., Oethinger, M., and Levy, S.B. **(1998)**. Overexpression of *marA*, *soxS*, or *acrAB* produces resistance to triclosan in laboratory and clinical strains of *Escherichia coli*. *FEMS Microbiol Lett* 166**,** 305-309.

Meibom, K.L., Blokesch, M., Dolganov, N.A., Wu, C.-Y., and Schoolnik, G.K. **(2005)**. Chitin induces natural competence in *Vibrio cholerae*. *Science* 310**,** 1824-1827.

Michael, C., and Labbate, M. **(2010)**. Gene cassette transcription in a large integron-associated array. *BMC Genet* 11**,** 82.

Michael, C.A., Gillings, M.R., Holmes, A.J., Hughes, L., Andrew, N.R., Holley, M.P., and Stokes, H.W. **(2004)**. Mobile gene cassettes: A fundamental resource for bacterial evolution. *Am Nat* 164**,** 1-12.

Michiels, T., Cornelis, G., Ellis, K., and Grinsted, J. **(1987)**. Tn2501, a component of the lactose transposon Tn951, is an example of a new category of class II transposable elements. *J. Bacteriol.* 169**,** 624-631.

Miller, E.S., Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Durkin, A.S., Ciecko, A., Feldblyum, T.V., White, O., Paulsen, I.T., Nierman, W.C., Lee, J., Szczypinski, B., and Fraser, C.M. **(2003)**. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J Bacteriol* 185**,** 5220-5233.

Miller, M.C., Keymer, D.P., Avelar, A., Boehm, A.B., and Schoolnik, G.K. **(2007)**. Detection and transformation of genome segments that differ within a coastal population of *Vibrio cholerae* strains. *Appl Environ Microbiol* 73**,** 3695-3704.

Mishra, P., Samanta, M., Maiti, N.K., and Sarangi, N. **(2009)**. Characterization of extracellular cytotoxic protein of *Vibrio* spp. isolated from freshwater carps and prawns. *Indian J Fish* 56**,** 307-311.

Miyata, S.T., Unterweger, D., Rudko, S.P., and Pukatzki, S. **(2013)**. Dual Expression Profile of Type VI Secretion System Immunity Genes Protects Pandemic *Vibrio cholerae*. *PLoS Pathog* 9**,** e1003752.

Mori, T., Mizuta, S., Suenaga, H., and Miyazaki, K. **(2008)**. Metagenomic screening for bleomycin resistance genes. *Appl Environ Microbiol* 74**,** 6803-6805.

Morris, J., and J, G. **(2003)**. Cholera and other types of vibriosis: A story of human pandemics and oysters on the half shell. *Clin Infect Dis* 37**,** 272.

Morton, A. **(2009)**. $38bn warning on Great Barrier Reef bleaching. *The Age* August 10, 2009.

Mukhopadhyay, A.K., Chakraborty, S., Takeda, Y., Nair, G.B., and Berg, D.E. **(2001)**. Characterization of VPI pathogenicity island and CTXphi prophage in environmental strains of *Vibrio cholerae*. *J Bacteriol* 183**,** 4737-4746.

Mukhopadhyay, A.K., Garg, S., Mitra, R., Basu, A., Rajendran, K., Dutta, D., Bhattacharya, S.K., Shimada, T., Takeda, T., Takeda, Y., and Nair, G.B. **(1996)**. Temporal shifts in traits of *Vibrio cholerae* strains isolated from hospitalized patients in Calcutta: a 3-year (1993 to 1995) analysis. *J Clin Microbiol* 34**,** 2537-2543.

Murphy, R.A., and Boyd, E.F. **(2008)**. Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates. *J Bacteriol* 190**,** 636-647.

Nasu, H., Iida, T., Sugahara, T., Yamaichi, Y., Park, K.S., Yokoyama, K., Makino, K., Shinagawa, H., and Honda, T. **(2000)**. A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains. *J Clin Microbiol* 38**,** 2156-2161.

Neiman, J., Guo, Y., and Rowe-Magnus, D.A. **(2011)**. Chitin-induced carbotype conversion in *Vibrio vulnificus*. *Infect Immun* 79**,** 3195-3203.

Nemergut, D.R., Martin, A.P., and Schmidt, S.K. **(2004)**. Integron diversity in heavy-metal-contaminated mine tailings and inferences about integron evolution. *Appl Environ Microbiol* 70**,** 1160-1168.

Nield, B.S., Holmes, A.J., Gillings, M.R., Recchia, G.D., Mabbutt, B.C., Nevalainen, K.M.H., and Stokes, H.W. **(2001)**. Recovery of new integron classes from environmental DNA. *FEMS Microbiol Lett* 195**,** 59-65.

Nield, B.S., Willows, R.D., Torda, A.E., Gillings, M.R., Holmes, A.J., Nevalainen, K.M., H., Stokes, H.W., Mabbutt, B.C. **(2004)**. New enzymes from environmental cassette arrays: functional attributes of a phosphotransferase and an RNA-methyltransferase. *Protein Sci* 13**,** 1651-1659.

Nishibuchi, M., and Kaper, J.B. **(1995)**. Thermostable direct hemolysin gene of *Vibrio parahaemolyticus*: a virulence gene acquired by a marine bacterium. *Infect Immun* 63**,** 2093-2099.

Nivens, D.E., Ohman, D.E., Williams, J., and Franklin, M.J. **(2001)**. Role of alginate and its O acetylation in formation of *Pseudomonas aeruginosa* microcolonies and biofilms. *J Bacteriol* 183**,** 1047-1057.

Nyunt, W.S., Mizunoe, Y., Takade, A., Kawabata, S.-I., and Yoshinda, H.-I. **(1998)**. *Vibrio cholerae* O1 strain TSI-4 produces the exopolysaccharide materials that determine colony morphology, stress resistance and biofilm formation. *Appl Environ Microbiol* 64**,** 3648-3655.

O'shea, Y.A., Finnan, S., Reen, F.J., Morrissey, J.P., O'gara, F., and Boyd, E.F. **(2004)**. The *Vibrio* seventh pandemic island-II is a 26.9 kb genomic island present in *Vibrio cholerae* El Tor and O139 serogroup isolates that shows homology to a 43.4 kb genomic island in *V. vulnificus*. *Microbiology* 150**,** 4053-4063.

Ochman, H., Lawrence, J.G., and Groisman, E.A. **(2000)**. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405**,** 299.

Oelschlaeger, T.A., Zhang, D., Schubert, S., Carniel, E., Rabsch, W., Karch, H., and Hacker, J. **(2003)**. The high-pathogenicity island is absent in human pathogens of *Salmonella enterica* subspecies I but present in isolates of subspecies III and VI. *J Bacteriol* 185**,** 1107-1111.

Ogawa, A., Takeda,T. **(1993)**. The gene encoding the heat-stable enterotoxin of *Vibrio cholerae* is flanked by 123-bp direct repeats. *Microbiol Immun* 37**,** 607-616.

Orozco-Mosqueda, M.D.C., Altamirano-Hernandez, J., Farias-Rodriguez, R., Valencia-Cantero, E., and Santoyo, G. **(2009)**. Homologous recombination and dynamics of rhizobial genomes. *Res Microbiol* 160**,** 733-741.

Pang, B., Yan, M., Cui, Z., Ye, X., Diao, B., Ren, Y., Gao, S., Zhang, L., and Kan, B. **(2007)**. Genetic diversity of toxigenic and nontoxigenic *Vibrio cholerae* serogroups O1 and O139 revealed by array-based comparative genomic hybridization. *J Bacteriol* 189**,** 4837-4849.

Park, K.S., Iida, T., Yamaichi, Y., Oyagi, T., Yamamoto, K., and Honda, T. **(2000)**. Genetic characterization of DNA region containing the *trh* and *ure* genes of *Vibrio parahaemolyticus*. *Infect Immun* 68**,** 5742-5748.

Partridge, S.R., Recchia, G.D., Scaramuzzi, C., Collis, C.M., Stokes, H.W., and Hall, R.M. **(2000)**. Definition of the attI1 site of class 1 integrons. *Microbiology* 146 (Pt 11)**,** 2855-2864.

Patel, M., Jiang, Q., Woodgate, R., Cox, M.M., and Goodman, M.F. **(2010)**. A new model for SOS-induced mutagenesis: how RecA protein activates DNA polymerase V. *Crit Rev Biochem Mol Biol* 45**,** 171-184.

Permina, E.A., Mironov, A.A., and Gelfand, M.S. **(2002)**. Damage-repair error-prone polymerases of eubacteria: association with mobile genome elements. *Gene* 293**,** 133-140.

Piddock, L.J., and Wise, R. **(1987)**. Induction of the SOS response in *Escherichia coli* by 4-quinolone antimicrobial agents. *FEMS Microb Lett* 41**,** 289-294.

Pier, G.B., Coleman, F., Grout, M., Franklin, M., and Ohman, D.E. **(2001)**. Role of alginate O acetylation in resistance of mucoid *Pseudomonas aeruginosa* to opsonic phagocytosis. *Infect Immun* 69**,** 1895-1901.

Pollack-Berti, A., Wollenberg, M.S., and Ruby, E.G. **(2010)**. Natural transformation of *Vibrio fischeri* requires *tfoX* and *tfoY*. *Environ Microbiol* 12**,** 2302-2311.

Polosina, Y.Y., and Cupples, C.G. **(2010)**. Wot the 'L-Does MutL do? *Mutat Res* 705**,** 228-238.

Pope, C.F., O'sullivan, D.M., McHugh, T.D., and Gillespie, S.H. **(2008)**. A practical guide to measuring mutation rates in antibiotic resistance. *Antimicrob Agents Chemother* 52**,** 1209-1214.

Post, D.M.B., Yu, L., Krasity, B.C., Choudhury, B., Mandel, M.J., Brennan, C.A., Ruby, E.G., Mcfall-Ngai, M.J., Gibson, B.W., and Apicella, M.A. **(2012)**. The O-antigen and core carbohydrate of *Vibrio fischeri* lipopolysaccharide: Composition and analysis of their role in *Euprymna scolopes* light organ colonization. *J Biol Chem*.

Poutrel, B., Gilbert, F.B., and Lebrun, M. **(1995)**. Effects of culture conditions on production of Type-5 capsular polysaccharide by human and bovine *Staphylococcus aureus* strains. *Clin Diagn Lab Immunol* 2**,** 166-171.

Rajanna, C., Wang, J., Zhang, D., Xu, Z., Ali, A., Hou, Y.M., and Karaolis, D.K. **(2003)**. The vibrio pathogenicity island of epidemic *Vibrio cholerae* forms precise extrachromosomal circular excision products. *J Bacteriol* 185**,** 6893-6901.

Ramamurthy, T., Garg, S., Sharma, R., Bhattacharya, S.K., Nair, G.B., Shimada, T., Takeda, T., Karasawa, T., Kurazano, H., Pal, A., and Takeda, Y. **(1993)**. Emergence of novel strain of Vibrio cholerae with epidemic potential in Southern and Eastern India. *Lancet* 341**,** 703-704.

Rapa, R.A., Islam, A., Monahan, L.G., Mutreja, A., Thomson, N., Charles, I.G., Stokes, H.W., and Labbate, M. **(2014)**. A genomic island integrated into *recA* of *Vibrio cholerae* contains a divergent *recA* and provides multi-pathway protection from DNA damage. *Environ Microbiol*. doi:10.1111/1462-2920.12512

Rapa, R.A., Shimmon, R., Djordjevic, S.P., Stokes, H.W., and Labbate, M. **(2013)**. Deletion of integron-associated gene cassettes impact on the surface properties of *Vibrio rotiferianus* DAT722. *PLoS ONE* 8**,** e58430.

Rattray, A.J., and Strathern, J.N. **(2003)**. Error-prone DNA polymerases: when making a mistake is the only way to get ahead. *Annu Rev Genet* 37**,** 31-66.

Ravatn, R., Studer, S., Springael, D., Zehnder, A.J., and Van Der Meer, J.R. **(1998)**. Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. Strain B13. *J Bacteriol* 180**,** 4360-4369.

Recchia, G.D., and Hall, R.M. **(1995)**. Gene cassettes: a new class of mobile element. *Microbiology* 141 (Pt 12)**,** 3015-3027.

Reeves, P.R., Hobbs, M., Valvano, M.A., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D., Raetz, C.R.H., and Rick, P.D. **(1996)**. Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol* 4**,** 495-503.

Rivera, I.N., Souza, K.M., Souza, C.P., and Lopes, R.M. **(2013)**. Free-living and plankton-associated vibrios: assessment in ballast water, harbor areas, and coastal ecosystems in Brazil. *Front Microbiol* 3.

Robinson, A., Guilfoyle, A.P., Harrop, S.J., Boucher, Y., Stokes, H.W., Curmi, P.M.G., and Mabbutt, B.C. **(2007)**. A putative house-cleaning enzyme encoded within an integron array: 1.8 Å crystal structure defines a new MazG subtype. *Mol Microbiol* 66**,** 610-621.

Robinson, A., Guilfoyle, A.P., Sureshan, V., Howell, M., Harrop, S.J., Boucher, Y., Stokes, H.W., Curmi, P.M.G., and Mabbutt, B.C. **(2008)**. Structural genomics of the bacterial mobile metagenome: an overview. *Method Mol Biol* 426**,** 589-595.

Robinson, A., Wu, P.S.C., Harrop, S.J., Schaeffer, P.M., Dosztányi, Z., Gillings, M.R., Holmes, A.J., Helena Nevalainen, K.M., Stokes, H.W., Otting, G., Dixon, N.E., Curmi, P.M.G., and Mabbutt, B.C. **(2005)**. Integron-associated mobile gene cassettes code for folded proteins: the structure of Bal32a, a new member of the adaptable [alpha]+[beta] barrel family. *J Mol Biol* 346**,** 1229-1241.

Rowe-Magnus, D.A. **(2009)**. Integrase-directed recovery of functional genes from genomic libraries. *Nucleic Acids Res* 37**,** e118.

Rowe-Magnus, D.A., Guérout, A.-M., and Mazel, D. **(1999)**. Super-integrons. *Res Microbiol* 150**,** 641-651.

Rowe-Magnus, D.A., Guerout, A.-M., Ploncard, P., Dychinco, B., Davies, J., and Mazel, D. **(2001)**. The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc Natl Acad Sci U S A* 98**,** 652-657.

Rowe-Magnus, D.A., Guerout, A.M., Biskri, L., Bouige, P., and Mazel, D. **(2003)**. Comparative analysis of superintegrons: engineering extensive genetic diversity in the Vibrionaceae. *Genome Res* 13**,** 428-442.

Rowe-Magnus, D.A., and Mazel, D. **(2001)**. Integrons: natural tools for bacterial genome evolution. *Curr Opin Microbiol* 4**,** 565-569.

Sack, D.A., Lyke, C., Mclaughlin, C., and Suwanvanichkij, V. **(2001)**. "Antimicrobial resistance in shigellosis, cholera and campylobacteriosis", (ed.) World Health Organization.).

Saenz-De-Miera, B., Thrasher, J.F., Chaloupka, F.J., Waters, H.R., Hernandez-Avila, M., and Fong, G.T. Self-reported price of cigarettes, consumption and compensatory behaviours in a cohort of Mexican smokers before and after a cigarette tax increase. *Tob Control*.

Safa, A., Nair, G.B., and Kong, R.Y. **(2010)**. Evolution of new variants of *Vibrio cholerae* O1. *Trends Microbiol* 18**,** 46-54.

Saltikov, C.W., and Newman, D.K. **(2003)**. Genetic identification of a respiratory arsenate reductase. *Proc Natl Acad Sci USA* 100**,** 10983-10988.

Salyers, A.A., Shoemaker, N.B., Stevens, A.M., and Li, L.Y. **(1995)**. Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol. Rev*. 59**,** 579-590.

Sambrook, J., Fritsch, E.F., and Maniatis, T. **(1989)**. *Molecular cloning: A laboratory manual.*

Sanders, C.C. **(2001)**. Mechanisms responsible for cross-resistance and dichotomous resistance among the quinolones. *Clin Infect Dis* 32**,** S1-S8.

Schlacher, K., Pham, P., Cox, M.M., and Goodman, M.F. **(2006)**. Roles of DNA polymerase V and RecA protein in SOS damage-induced mutation. *Chem Rev* 106**,** 406-419.

Schmidt, H., and Hensel, M. **(2004)**. Pathogenicity islands in bacterial pathogenesis. *Clin. Microbiol. Rev*. 17**,** 14-56.

Scholl, D., Adhya, S., and Merril, C. **(2005)**. *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. *Appl Environ Microbiol* 71**,** 4872-4874.

Schubert, S., Rakin, A., Karch, H., Carniel, E., and Heesemann, J. **(1998)**. Prevalence of the "high-pathogenicity island" of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infect Immun* 66**,** 480-485.

Schwartz, R., Ting, C.S., and King, J. **(2001)**. Whole proteome pI Values correlate with subcellular localizations of proteins for organisms within the three domains of life. *Genome Res* 11**,** 703-709.

Scrascia, M., Pugliese, N., Maimone, F., Mohamud, K.A., Ali, I.A., Grimont, P.a.D., and Pazzani, C. **(2009)**. Cholera in Ethiopia in the 1990s: Epidemiologic patterns, clonal analysis, and antimicrobial resistance. *Intl J Med Microbiol* 299**,** 367-372.

Shaw, M.M., and Riederer, B.M. **(2003)**. Sample preparation for two-dimensional gel electrophoresis. *Proteomics* 3**,** 1408-1417.

Simonet, V.C., Basle, A., Klose, K.E., and Delcour, A.H. **(2003)**. The *Vibrio cholerae* Porins OmpU and OmpT have distinct channel properties. *J Biol Chem* 278**,** 17539-17545.

Smith, A.B., and Siebeling, R.J. **(2003)**. Identification of genetic loci required for capsular expression in *Vibrio vulnificus*. *Infect Immun* 71**,** 1091-1097.

Smith, S.G., Mahon, V., Lambert, M.A., and Fagan, R.P. **(2007)**. A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiol Lett* 273**,** 1-11.

Sommer, S., Boudsocq, F., Devoret, R., and Bailone, A. **(1998)**. Specific RecA amino acid changes affect RecA-UmuD'C interaction. *Mol Microbiol* 28**,** 281-291.

Stewart, P.E., Byram, R., Grimm, D., Tilly, K., and Rosa, P.A. **(2005)**. The plasmids of *Borrelia burgdorferi*: essential genetic elements of a pathogen. *Plasmid* 53**,** 1-13.

Stokes, H.W., Gillings, M.R. **(2011)**. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram negative pathogens. *FEMS Microbiol Rev*.

Stokes, H.W., and Hall, R.M. **(1989)**. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions-integrons. *Mol Microbiol* 3**,** 1669-1683.

Stokes, H.W., Holmes, A.J., Nield, B.S., Holley, M.P., Nevalainen, K.M., Mabbutt, B.C., and Gillings, M.R. **(2001)**. Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Appl Environ Microbiol* 67**,** 5240-5246.

Stokes, H.W., Nesbo, C.L., Holley, M., Bahl, M.I., Gillings, M.R., and Boucher, Y. **(2006)**. Class 1 integrons potentially predating the association with tn402-like transposition genes are present in a sediment microbial community. *J Bacteriol* 188**,** 5722-5730.

Sugino, A., Peebles, C.L., Kreuzer, K.N., and Cozzarelli, N.R. **(1977)**. Mechanism of action of nalidixic acid: purification of *Escherichia coli nalA* gene product and its

relationship to DNA gyrase and a novel nicking-closing enzyme. *Proc Natl Acad Sci U S A* 74**,** 4767-4771.

Szekeres, S., Dauti, M., Wilde, C., Mazel, D., Rowe-Magnus, D.A. **(2007)**. Chromosomal toxin-antitoxin loci can diminish large- scale genome reductions in the absence of selection. *Mol Microbiol* 63**,** 1588-1605.

Tark, M., Tover, A., Tarassova, K., Tegova, R., Kivi, G., Horak, R., and Kivisaar, M. **(2005)**. A DNA polymerase V homologue encoded by TOL plasmid pWW0 confers evolutionary fitness on *Pseudomonas putida* under conditions of environmental stress. *J Bacteriol* 187**,** 5203-5213.

Thi, T.D., Lopez, E., Rodriguez-Rojas, A., Rodriguez-Beltran, J., Couce, A., Guelfo, J.R., Castaneda-Garcia, A., and Blazquez, J. **(2011)**. Effect of *recA* inactivation on mutagenesis of *Escherichia coli* exposed to sublethal concentrations of antimicrobials. *J Antimicrob Chemother* 66**,** 531-538.

Thomas, C.M., and Nielsen, K.M. **(2005)**. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* 3**,** 711-721.

Thompson, C.C., Thompson, F.L., Vandemeulebroecke, K., Hoste, B., Dawyndt, P., and Swings, J. **(2004a)**. Use of *recA* as an alternative phylogenetic marker in the family *Vibrionaceae*. *Int J Sys Evol Micr* 54**,** 919-924.

Thompson, L.F., Iida, T., and Swings, J. **(2004b)**. Biodiversity of Vibrios. *Microbiol Mol Biol Rev* 68**,** 403-431.

Thomson, V.J., Bhattacharjee, M.K., Fine, D.H., Derbyshire, K.M., and Figurski, D.H. **(1999)**. Direct selection of IS903 transposon insertions by use of a broad-host-range vector: isolation of catalase-deficient mutants of *Actinobacillus actinomycetemcomitans*. *J Bacteriol* 181**,** 7298-7307.

Tillett, D., and Neilan, B.A. **(2000)**. Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. *J Phycol* 36**,** 251-258.

Uma, G., Chandrasekaran, M., Takeda, Y., and Nair, G.B. **(2003)**. Recent advances in cholera genetics. *Curr Sci* 85**,** 1538-1545.

Van Der Meer, J.R., and Sentchilo, V. **(2003)**. Genomic islands and the evolution of catabolic pathways in bacteria. *Curr Opin Biotechnol* 14**,** 248-254.

Visick, K.L., and Mcfall-Ngai, M.J. **(2000)**. An Exclusive Contract: Specificity in the *Vibrio fischeri-Euprymna scolopes* Partnership. *J Bacteriol* 182**,** 1779-1787.

Vu, B., Chen, M., Crawford, R.J., and Ivanova, E.P. **(2009)**. Bacterial Extracellular Polysaccharides Involved in Biofilm Formation. *Molecules* 14**,** 2535-2554.

Waldor, M.K., and Mekalanos, J.J. **(1996)**. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272**,** 1910-1914.

Waldor, M.K., Tschape, H., and Mekalanos, J.J. **(1996)**. A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. *J Bacteriol* 178**,** 4157-4165.

Wertman, K.F., and Mount, D.W. **(1985)**. Nucleotide sequence binding specificity of the LexA repressor of *Escherichia coli* K-12. *J Bacteriol* 163**,** 376-384.

Westman, E.L., Preston, A., Field, R.A., and Lam, J.S. **(2008)**. Biosynthesis of a rare di-N-acetylated sugar in the lipopolysaccharides of both *Pseudomonas aeruginosa* and *Bordetella pertussis* occurs via an identical scheme despite different gene clusters. *J Bacteriol* 190**,** 6060-6069.

Whitfield, C. **(2006)**. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Ann Rev Biochem***,** 39-68.
Willey, J.M., Sherwood, L.M., and Woolverton, C.J. **(2011)**. *Microbiology*. McGraw-Hill. Edition 8
World Health Organization **(2006)**. Weekly epidemiological record. 297-308.
World Health Organization **(2014)**. Factsheet http://www.who.int/mediacentre/factsheets/fs107/en/.
Wozniak, R.A., and Waldor, M.K. **(2010)**. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* 8**,** 552-563.
Yazdankhah, S.P., Scheie, A.A., Hoiby, E.A., Lunestad, B.T., Heir, E., Fotland, T.O., Naterstad, K., and Kruse, H. **(2006)**. Triclosan and antimicrobial resistance in bacteria: an overview. *Microb Drug Resist* 12**,** 83-90.
Yeiser, B., Pepper, E.D., Goodman, M.F., and Finkel, S.E. **(2002)**. SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. *Proc Natl Acad Sci U S A* 99**,** 8737-8741.
Yildiz, F.H., Liu, X.S., Heydorn, A., and Schoolnik, G.K. **(2004)**. Molecular analysis of rugosity in a *Vibrio cholerae* O1 El Tor phase variant. *Mol Microbiol* 53**,** 497-515.
Ysern, P., Clerch, B., Castaňo, M., Gilbert, I., Barbé, J., and Llagostera, M. **(1990)**. Induction of SOS genes in *Escherichia coli* and mutagenesis in *Salmonella typhimurium* by fluoroquinolones. *Mutagenesis* 5**,** 63-66.
Yu, X., Vanloock, M.S., Yang, S., Reese, J.T., and Egelman, E.H. **(2004)**. What is the structure of the RecA-DNA filament? *Curr Protein Pept Sci* 5**,** 73-79.
Zaneveld, J.R., Nemergut, D.R., and Knight, R. **(2008)**. Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. *Microbiology* 154**,** 1-15.

Appendix:

Appendix 1: List of 116 *V. rotiferianus* DAT722 gene cassettes with putative identification and putative conserved superfamily domains of proteins encoded by genes contained within the gene cassettes.

Cassette	Putative identification	Number of amino acids	Putative conserved superfamily domains
1	Hypothetical protein	200	None
2	Phage-related protein	277	None
3	Hypothetical protein	234	None
4	FRG domain protein	239	FRG domain[a]
5	ThiJ/PfpI family protein	202	GAT_1 domain[2]
6	Hypothetical protein	133	None
7	Hypothetical protein	93	None
8	Hypothetical protein	187	None
9	Hypothetical protein	153	None
10	Hypothetical protein	150	None
11	DNA topoisomerase I	237	Mrr_cat domain[3]; zf-C4_Topoisom domain[4]
12	Hypothetical protein	188	None
13	Hypothetical protein	190	None
14	Hypothetical protein	79	None
15	Hypothetical protein	106	None
16	Hypothetical protein	190	None
17	Hypothetical protein	258	None
18	Hypothetical protein	208	None
19	Hypothetical protein	226	DUF4145 domain[5]
20	Antibiotic biosynthesis monooxygenase	114	ABM domain[6]
21	MazG nucleotide phosphohydroloase	94	NTP-PPase domain[7]

22	No significant coding region		
23	Hypothetical protein	119	None
24	Acetyltransferase	255	NAT_SF domain[8]
25	Hypothetical protein	282	None
26	No significant coding region		
27	Hypothetical protein	126	None
28	Hypothetical protein	214	None
29	Putative restriction endonuclease	289	HNHc domain[9]
30	Putative PAAR-containing motif	176	None
31	Phosphorylated carbohydrate phosphatase	215	HAD-like domain[10]
32	Putative GNAT family acetyltransferase	149	NAT_SF domain[8]
33	Hypothetical protein	246	None
34	No significant coding region		
35	Hypothetical protein	209	None
36	No significant coding region		
37	No significant coding region		
38	Hypothetical protein	139	None
39	GNAT family acetyltransferase	154	NAT_SF domain[8]
40	No significant coding region		
41	Hypothetical protein	159	None
42	Hypothetical protein	209	None
43	Hypothetical protein	218	None
44	Hypothetical protein	106	DUF2834[11]
45	Hypothetical protein	223	None
46	No significant coding region		
47	Putative ribonuclease inhibitor	93	Barstar_like domain[12]

48	No significant coding region		
49	No significant coding region		
50	No significant coding region		
51	Hypothetical protein	194	None
52	Hypothetical protein	105	None
53	No significant coding region		
54	Acetyltransferase	165	NAT_SF domain[8]
55	No significant coding region		
56	Hypothetical protein	110	None
57	Acetyltransferase	151	NAT_SF domain[8]
58	Aminoglycoside phosphotransferase	274	PKc_like domain[13]
59	DNA topology modulation protein	165	P-loop_NTPase domain[14]
60	No significant coding region		
61	No significant coding region		
62	Hypothetical protein	116	None
63	No significant coding region		
64	Retinol acyltransferase domain protein	161	LRAT domain [15]
65	hypothetical acetyltransferase	142	NAT_SF domain[8]
66	No significant coding region		
67	Hypothetical protein	241	None
68	Hypothetical protein	78	None
69	Histone acetyltransferase HPA2	156	NAT_SF domain[8]
70	Hypothetical protein	118	None
71	Hypothetical protein	116	None
72A	Type VI secretion system-associated	159	DUF4285 domain[16]
72B	Hypothetical protein	121	None

73	No significant coding region		
74	haemagglutinin associated protein	224	AdoMet-MTases[17]
75	Hypothetical protein	318	TIR_2 domain[18]
76	Hypothetical protein	97	None
77	Hypothetical protein	91	None
78	Maltose O-acetyltransferase	195	LbetaH domain[19]; Mac domain[20]
79	Hypothetical protein	246	None
80	Hypothetical protein	124	None
81	Hypothetical protein	102	None
82	No significant coding region		
83	No significant coding region		
84	Acetyltransferase	157	NAT_SF domain[8]
85	Hypothetical protein	154	GIY-YIG nuclease domain[21]
86	Hypothetical protein	157	None
87	No significant coding region		
88	Acetyltransferase	173	NAT_SF domain[8]
89	Hypothetical protein	102	None
90	No significant coding region		
91	Acetyltransferase	173	NAT_SF domain[8]
92	Hypothetical protein	102	None
93	No significant coding region		
94	Hypothetical protein	292	None
95	Hypothetical	238	None
96	No significant coding region		
97	Cold shock protein	155	S1_like domain[22];Excalibur calcium-binding domain[23]
98	Hypothetical protein	210	None

99	Hypothetical protein	318	None
100	No significant coding region		
101	Hypothetical protein	289	none
102	Toxin-antitoxin plasmid stability protein	107	Plasmid stabilisation system domain[24]
103	Acetyltransferase	143	NAT_SF domain[8]
104	Hypothetical protein	152	Transglut_core domain[25]
105	No significant coding region		
106	Hypothetical protein	127	None
107	Glyoxalase/bleomycin resistance protein/dioxygenase protein	117	Glo_EDI_BRP_like domain[26]
108	Hypothetical protein	244	None
109	Restriction endonuclease-like protein	264	HNHc domain[9]
110	Putative phage-related membrane protein	155	None
111	Hypothetical protein	100	None
112	Putative cytoplasmic protein	172	DUF2778 domain[27]
113	Hypothetical	122	None
114	Cytoplasmic protein	161	DUF4285 domain[15]
115	No significant coding region		
116	Hypothetical protein	149	None

[a] This presumed domain contains a conserved N-terminal (F/Y)RG motif. It is functionally uncharacterised.
[2] Type 1 glutamine amidotransferase (GATase1)-like domain.
[3] Prokaryotic family found in type II restriction enzymes containing the hallmark (D/E)-(D/E)XK active site. Presence of catalytic residues implicates this region in the enzymatic cleavage of DNA.
[4] Topoisomerase DNA binding C4 zinc finger.
[5] This domain is found in a variety of restriction endonuclease enzymes. It is functionally uncharacterised.
[6] This domain is found in monooxygenases involved in the biosynthesis of several antibiotics by Streptomyces species.
[7] This superfamily contains enzymes that hydrolyze the alpha-beta phosphodiester bond of all canonical NTPs into monophosphate derivatives and pyrophosphate (PPi).
[8] N-Acyltransferase superfamily: Various enzymes that characteristically catalyse the transfer of an acyl group to a substrate.
[9]HNH endonuclease signature which is found in viral, prokaryotic, and eukaryotic proteins.

[10] The haloacid dehalogenase (HAD) superfamily includes carbon and phosphorus hydrolases. These proteins catalyse nucleophilic substitution reactions at phosphorus or carbon centres, using a conserved Asp carboxylate in covalent catalysis.
[11] Protein of unknown function (DUF2834) ;This is a bacterial family of uncharacterised proteins.
[12]Barstar is an intracellular inhibitor of barnase, an extracellular ribonuclease of Bacillus amyloliquefaciens. Barstar binds tightly to the barnase active site and sterically blocks it, thus inhibiting its potentially lethal RNase activity inside the cell.
[13]Protein Kinases, catalytic domain ;The protein kinase superfamily is mainly composed of the catalytic domains of serine/threonine-specific and tyrosine-specific protein kinases.
[14]P-loop containing Nucleoside Triphosphate Hydrolases ;Members of the P-loop NTPase domain superfamily are characterized by a conserved nucleotide phosphate-binding motif.
[15]Lecithin retinol acyltransferase ;The full-length members of this family are representatives of a novel class II tumour-suppressor family.
[16] This family of proteins is functionally uncharacterised.
[17]S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I; AdoMet-MTases are enzymes that use S-adenosyl-L-methionine (SAM or AdoMet) as a substrate for methyltransfer, creating the product S-adenosyl-L-homocysteine (AdoHcy).
[18]This is a family of bacterial Toll-like receptors.
[19]Left-handed parallel beta-Helix (LbetaH)domain: The alignment contains 5 turns, each containing three imperfect tandem repeats of a hexapeptide repeat motif .Proteins containing hexapeptide repeats are often enzymes showing acyltransferase activity, however, some subfamilies in this hierarchy also show activities related to ion transport or translation initiation.
[20]Maltose acetyltransferase ;This domain family is found in bacteria, archaea and eukaryotes, and is approximately 50 amino acids in length. Mac uses acetyl-CoA as acetyl donor to acetylated cytoplasmic maltose.
[21]The GIY-YIG nuclease domain superfamily includes a large and diverse group of proteins involved in many cellular processes, such as class I homing GIY-YIG family endonucleases, prokaryotic nucleotide excision repair proteins UvrC and Cho. All of these members contain a conserved GIY-YIG nuclease domain that may serve as a scaffold for the coordination of a divalent metal ion required for catalysis of the phosphodiester bond cleavage.
[22]Ribosomal protein S1-like RNA-binding domain. Found in a wide variety of RNA-associated proteins. Originally identified in S1 ribosomal protein. This superfamily also contains the Cold Shock Domain (CSD), which is a homolog of the S1 domain. Both domains are members of the Oligonucleotide/oligosaccharide Binding (OB) fold.
[23]Extracellular Ca2+-dependent nuclease YokF from Bacillus subtilis and several other surface-exposed proteins from diverse bacteria are encoded in the genomes in two paralogous forms that differ by a ~45 amino acid fragment, which comprises a novel conserved domain.
[24] Plasmid stabilisation system protein ;Members of this family are involved in plasmid stabilisation. The exact molecular function of this protein is not known. This family also encompasses RelE/ParE . RelE/StbE family ;Plasmids may be maintained stably in bacterial populations through the action of addiction modules, in which a toxin and antidote are encoded in a cassette on the plasmid. In any daughter cell that lacks the plasmid, the toxin persists and is lethal after the antidote protein is depleted. Toxin/antitoxin pairs are also found on main chromosomes, and likely represent selfish DNA.
[25]This family includes animal transglutaminases and other bacterial proteins of unknown function.
[26]This domain superfamily is found in a variety of structurally related metalloproteins, including the type I extradiol dioxygenases, glyoxalase I and a group of antibiotic resistance proteins.
[27]Protein of unknown function (DUF2778) ;This is a bacterial family of uncharacterised proteins.

Appendix 2

OPEN ACCESS Freely available online

PLOS | ONE

Deletion of Integron-Associated Gene Cassettes Impact on the Surface Properties of *Vibrio rotiferianus* DAT722

Rita A. Rapa[1], Ronald Shimmon[2], Steven P. Djordjevic[1], H. W. Stokes[1], Maurizio Labbate[1]*

1 The ithree Institute, University of Technology, Sydney, Australia, **2** Chemical Technology and Forensic Science, University of Technology, Sydney, Australia

Abstract

Background: The integron is a genetic recombination system that catalyses the acquisition of genes on mobilisable elements called gene cassettes. In *Vibrio* species, multiple acquired gene cassettes form a cassette array that can comprise 1–3% of the bacterial genome. Since 75% of these gene cassettes contain genes encoding proteins of uncharacterised function, how the integron has driven adaptation and evolution in *Vibrio* species remains largely unknown. A feature of cassette arrays is the presence of large indels. Using *Vibrio rotiferianus* DAT722 as a model organism, the aim of this study was to determine how large cassette deletions affect vibrio physiology with a view to improving understanding into how cassette arrays influence bacterial host adaptation and evolution.

Methodology/Principal Findings: Biological assays and proteomic techniques were utilised to determine how artificially engineered deletions in the cassette array of *V. rotiferianus* DAT722 affected cell physiology. Multiple phenotypes were identified including changes to growth and expression of outer membrane porins/proteins and metabolic proteins. Furthermore, the deletions altered cell surface polysaccharide with Proton Nuclear Magnetic Resonance on whole cell polysaccharide identifying changes in the carbohydrate ring proton region indicating that gene cassette products may decorate host cell polysaccharide via the addition or removal of functional groups.

Conclusions/Significance: From this study, it was concluded that deletion of gene cassettes had a subtle effect on bacterial metabolism but altered host surface polysaccharide. Deletion (and most likely rearrangement and acquisition) of gene cassettes may provide the bacterium with a mechanism to alter its surface properties, thus impacting on phenotypes such as biofilm formation. Biofilm formation was shown to be altered in one of the deletion mutants used in this study. Reworking surface properties may provide an advantage to the bacterium's interactions with organisms such as bacteriophage, protozoan grazers or crustaceans.

Citation: Rapa RA, Shimmon R, Djordjevic SP, Stokes HW, Labbate M (2013) Deletion of Integron-Associated Gene Cassettes Impact on the Surface Properties of *Vibrio rotiferianus* DAT722. PLoS ONE 8(3): e58430. doi:10.1371/journal.pone.0058430

Editor: Laurence Van Melderen, Universite Libre de Bruxelles, Belgium

Received April 17, 2012; **Accepted** February 6, 2013; **Published** March 6, 2013

Funding: The research in this paper was funded by: 1. UTS Early Career Research Grant, 2. ithree Institute internal grant, 3. UTS Faculty of Science Postgraduate Scholarship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: maurizio.labbate@uts.edu.au

Introduction

Integrons are genetic elements that include site-specific recombination functions. They integrate and express genes present on mobilisable elements called gene cassettes. The integron consists of three components, a gene (*intI*) encoding an integrase, an attachment site (*attI*) where gene cassettes insert and a promoter (P_c) adjacent to *attI* that drives transcription of inserted gene cassettes [1,2]. Numerous classes of integrons have been identified and these classes are defined by the sequence that encodes the integrase [3]. First discovered in clinical settings, class 1 integrons are commonly found on resistance plasmids within pathogens and commensals. They carry small cassette arrays of, commonly, 1–6 gene cassettes and are a major contributor to the problem of antibiotic resistance [4]. In the natural environment, the integron is present in chromosomal locations with approximately 10% of sequenced genomes containing chromosomal integrons [5]. In these organisms, cassette arrays can vary substantially in size (0 – >200 cassettes) and rarely carry known antibiotic resistance gene cassettes. Given this, integrons are regarded as having a more general role in evolution than simply carrying and expressing antibiotic resistance genes [1,5].

Vibrio species are free-living marine bacteria that carry out diverse roles and occupy a wide range of niches in association with higher organisms. They can be found in symbiotic or pathogenic relationships with a wide variety of marine hosts such as prawns, coral, fish, invertebrates, plants and marine mammals [6]. One of the major drivers of the diversification of *Vibrio* species is lateral gene transfer (LGT) [7]. *Vibrio* species carry particularly large cassette arrays with the integron and associated cassettes making up 1–3% of the entire bacterial genome and as such is a substantial source of laterally acquired DNA in vibrios. How the integron influences the evolution of *Vibrio* species remains largely unknown although recent studies have provided new insight into the biology of integrons. The SOS response induces the integron-integrase resulting in enhanced rates of acquisition, deletion and movement of gene cassettes across the array [8]. This suggests that the bacterial host uses the integron as a mechanism for adaptation to stressful environments. Furthermore, in *V. rotiferianus* DAT722, the majority of the 116-gene cassettes in the array are transcribed,

with different cassettes transcribed in response to differing growth conditions [9]. These data indicate the presence of diverse promoters other than P_c within the array and demonstrates that almost all gene cassettes are able to add to the adaptive potential of the cell.

Approximately 75% of gene cassettes in *Vibrio* species encode proteins that are of unknown function [5] although a handful have been characterised and show to produce functional proteins [2,10,11,12,13,14]. In a recent study, mutants with deletions in the integron cassette array of *V. rotiferianus* DAT722 were created to assist in identifying phenotypes for uncharacterised gene cassettes. In this study, it was shown that deletion of a specific group of cassettes substantially impacted growth and porin regulation demonstrating that apart from providing accessory functions, gene cassette products can integrate into complex regulatory pathways [15]. A feature of cassette arrays is that large groups of contiguous cassettes can be deleted from an array at a given time [16]. For example, a large 38 gene cassette deletion was found in some strains of pathogenic *V. cholerae* O1 El Tor strains but not in others with the deletion being the only known difference between these strains [16]. How this impacts the bacterial cell and its survivability in the environment is unknown.

In an effort to understand how deletion of contiguous gene cassettes might affect vibrio adaptation and evolution, we used physiological growth, stress assays, proteomic and chemistry-based techniques to characterise how engineered deletions of gene cassettes in *V. rotiferianus* DAT722 affects vibrio physiology. We show that deletion of gene cassettes affects surface structures of the bacterial cell, specifically, properties of bacterial polysaccharide. We hypothesise that acquisition, movement or deletion of some gene cassettes within the array might provide the host organism with a mechanism for altering surface properties. In this study, we show that deletion of gene cassettes can alter biofilm formation and hypothesise that modifying surface properties may also have implications for how vibrios interact with bacteriophage, protozoan grazers and higher marine organisms.

Results

In this study, we compared the physiological effects of deleting multiple genes cassettes from the cassette array of *V. rotiferianus* DAT722. Construction and description of the mutants is described in the Materials and Methods (strains used in this study are shown in Table 1). Deletion mutants d16-60, d50-60 and d72-92 have had gene cassettes 16–60, 50–60 and 72–92 deleted from the 116 gene cassette array respectively. A table providing details on the deleted cassettes is included as supplementary material (Table S1).

Does Deletion of Gene Cassettes Affect Growth?

Comparison growth curves (Figure 1) of the parent and isogenic deletion mutants were conducted to determine whether the deletion of gene cassettes affected growth. Growth curves were conducted in LB20 and in 2M salts+various carbon sources including glucose, fumarate, succinate, aspartic acid and pyruvate. On a logarithmic scale, growth of the wild-type (wt) and deletion mutants did not reveal any obvious changes in lag phase or growth rate (data not shown). However, minor reproducible changes were observed in optical density when growth of the deletion mutants on 2M+various carbon sources was compared to the wild-type (wt) parent. Specifically, the wt had at least 2-fold less cells when grown in 2M+aspartic acid compared to d16–60, d50–60 and d72–92 at 60 hours growth (Figure 1C). In 2M+succinate, d72–92 had approximately 3-fold less cells at 12 hours growth when compared to the wt and other deletion mutants (Figure 1D). In 2M+fumarate d50–60 and d72–92 grew faster than the wt and d16–60 mutant with the d72–92 having approximately 3-fold more cells than the wt at 24 hours (Figure 1E). In 2M+pyruvate d16–60 and d50–60 have higher optical density readings at after 7 and 8.5 hours (Figure 1F).

These data indicate one or more of multiple possibilities. The deletions have altered the metabolic balance within the cell resulting in minor variation in growth, the deletions have removed cassettes involved in stress upon nutrient deprivation or that permeability of the cells has been changed altering the uptake of the carbon source or other nutrients. While the deletions did not appear to have affected final cell densities, the minor variations in growth could affect competitiveness in the environment.

Gene Cassettes do not Affect Environmental Stress Survival

Recent studies have shown that SOS-inducing stress activates the integron-integrase gene and subsequent gene cassette shuffling [8]. As a result, it has been suggested that in times of stress, gene cassette shuffling may be a mechanism for surviving environmental stress. To determine whether this is the case, the wt parent and deletion mutants were subject to stress conditions that the organism might encounter in the natural environment including oxidative, iron depletion and cold shock stresses (Figure 2). For each tested stress, no major difference was observed between the deletion mutants and wt indicating that at least for the deleted gene cassettes, there is no specific role in environmental stress.

How does Deletion of Gene Cassettes Affect Whole Cell Protein Regulation and Secretion of Proteins?

Gene cassettes can have adverse effects on growth indicating that mobile genes integrate into host cellular networks [15]. Thus, we were interested in how a large gene cassette deletion, not adversely affecting growth rates would affect host whole cell physiology. It was hypothesised that deletion of cassettes, thereby resulting in the loss of some proteins, would result in changes to cell networks. Identifying these changes may provide further insight into gene cassette function and the influence mobile genes have on the adaptive potential of the cell.

Mutant d16–60 (the largest array deletion) was subjected to 2D-PAGE and compared to the wt when grown under complete (LB20) and minimal growth conditions (2M+glucose). Furthermore, protein expression at stationary and mid-logarithmic phases was compared under both nutrient conditions. For cells grown in LB20 to stationary phase, two out of 325 protein spots were shown to be differentially expressed between wt and d16–60, in mid-logarithmic phase, four out of 357 were differentially expressed (Table 2). For cells grown in minimal media to stationary phase, three out of 360 protein spots were differentially expressed and in mid-logarithmic phase, three out of 201 were differentially expressed (Table 3). Thus, approximately 0.5–1% of protein spots identified between wt and d16–60 were at least 2-fold differentially expressed with a maximum fold difference of 3.7 for spot 2CM (Table 2). This analysis showed quite subtle differences in the whole cell proteome between wt and d16–60 in the conditions tested, indicating only a minor effect of the 46 cassette deletion (~31 kb) in the expression of detectable high abundance proteins in the proteome. It is possible that changes to lower abundance proteins not detected using this methodology is occurring.

Due to co-migration of proteins not all protein spots could be unambiguously identified. Out of the unambiguous protein spots that were differentially expressed, surface-associated proteins were identified (Tables 2 and 3), including OmpA (spot 1CM), an

Table 1. List of strains and plasmids.

Strain or plasmid	Relevant genotype[1]	Reference
Vibrio rotiferianus DAT722	wild-type	[43]
DAT722-Sm	DAT722; spontaneous Sm^R mutant	[15]
MD7	DAT722-Sm; Single recombination cross-over of pMAQ1081 into cassette 61, Km^R	[15]
SC-8B61	DAT722-Sm; Single recombination cross-over of pMAQ1081 into cassette 61, Km^R	This study
SC-8A91	DAT722-Sm; Single recombination cross-over of pMAQ1081 into cassette 93, Km^R	This study
d16-60	DAT722-Sm; Δcassettes 16-60, Sm^R, Km^R	[15]
d16–60a	DAT722-Sm; Δcassettes 16–60, Sm^R, Km^R	This study
d50–60	DAT722-Sm; Δcassettes 50-60, Sm^R, Km^R	This study
d72–92	DAT722-Sm; Δcassettes 72–92, Sm^R, Km^R	This study
d72–92a	DAT722-Sm; Δcassettes 72–92, Sm^R, Km^R	This study

[1]Sm^R, streptomycin resistance; Km^R, kanamycin resistance.

doi:10.1371/journal.pone.0058430.t001

OmpU-like protein (spot 2CM) and an unknown protein (spot 1CS). Since the gene that encodes this unknown protein is located in a region of the genome responsible for polysaccharide biosynthesis and PSORTb analysis shows its localisation to be extracellular [17], the unknown protein is likely be involved in polysaccharide synthesis (Table S2). Metabolic proteins were also differentially expressed including proteins with homology to nitrogen regulatory protein P-II (spot 2CS), alkyl hydroperoxide reductase subunit C-like protein (spot 2MM) and 6,7-dimethyl-8-ribityllumazine synthase (spot 3MM). An uncharacterised protein (spot 2MS) was also identified as differentially expressed. No gene cassette proteins were identified as missing in the d16-60 mutant even though most gene cassettes in *V. rotiferianus* DAT722 are transcribed [9]. Presumably this is due to high abundance proteins masking the presence of lower abundance proteins [18,19].

In addition to 2D-PAGE, we compared the secreted proteins of wt with all deletion mutants from cells grown in 2M+glucose and found no major differences in terms of presence or absence of protein bands. However, slight variations in the abundance of some proteins were observed and are labelled in Figure 3.

Congo Red Staining of Bacterial Colonies Show Differences in All Deletion Mutants

During 2D-PAGE analysis, it was observed that protein extracted from stationary phase d16–60 cells grown in 2M+glucose consistently contained a substance that interfered with the isoelectric focusing (IEF) step in 2D-PAGE (Figure S1). This substance was removed from the sample by washing cells with 2% NaCl pre-protein extraction and represents a point of physiological difference between wt and the d16–60 mutant. Although this substance is yet to be identified, it is extracellular and weakly associated with the cell surface.

Multiple contaminating substances can interfere with the IEF step including DNA, cell wall material and polysaccharides [19]. Since the substance specific to d16–60 is extracellular in nature and a protein putatively involved in polysaccharide biosynthesis was identified by the 2D-PAGE analysis (spot 2CS), we hypothesised the contaminating substance was polysaccharide in nature.

To confirm this, we used congo red staining to determine whether deleting cassettes affected colony wrinkling in all deletion mutants. Colony wrinkling has previously been shown to be associated with the extracellular/capsular polysaccharide in vibrios [20]. Interestingly, no differences could be observed between the wt and deletion mutants on 2M+glucose medium especially since it was in this medium that d16–60 was producing the contaminating substance (Figure 4). However, substantial changes in colony wrinkling were observed when cells were grown on LB20. On LB20, wt colonies showed repeated characteristic wrinkling architecture whereas the deletion mutants showed differing levels of wrinkling indicating a change in the amount or structure of produced polysaccharide in these mutants. The lack of wrinkling of colonies when grown on 2M+glucose medium indicates that *V. rotiferianus* DAT722 produces different polysaccharide(s) when exposed to different growth conditions that does not bind congo red. This production of different polysaccharides under different growth conditions has been reported in other bacteria [21,22].

Proton NMR Spectroscopy of Whole Cell Polysaccharides

To confirm that deletion of gene cassettes had modified cell surface polysaccharides, total polysaccharide from wt, d16–60 and d72–92 was extracted from cells grown in 2M+glucose medium using a hot phenol extraction method (Material and Methods) and subjected to preliminary ^{1}H NMR analysis. These two mutants were selected for ^{1}H NMR due to the different cassettes deleted. To ensure any changes were a result of the deletions and not any secondary mutation(s), total whole cell polysaccharide was also extracted and purified from identical but independently derived mutants: d16–60a and d72–92a.

Whole cell polysaccharide extractions include, lipopolysaccharide (LPS; made up of the O-antigen, core polysaccharide and lipid A core) and capsular/extracellular (C/EPS) polysaccharide structures, so any changes are related directly to these structures. However which polysaccharide moiety being altered cannot be identified.

We initially compared the ^{1}H NMR scans of the wt with d16–60 and d72–92 (Figure 5). The scans showed dissimilarity between the wt and both d16–60 and d72–92, as well as dissimilarity between the two deletion mutants themselves in the chemical shift region of approximately 3.0–4.5 ppm (Figure 5). Fresh extractions of the wt, d16–60a and d72–92a produced ^{1}H NMR scans showed similar scans to the first batch however, due to a small modification in the extraction methodology (see Materials and Methods) differences in purity were observed. Nevertheless, changes were again observed in the chemical shift region of approximately 3.0–4.5 ppm (Figure S2). Therefore, these changes

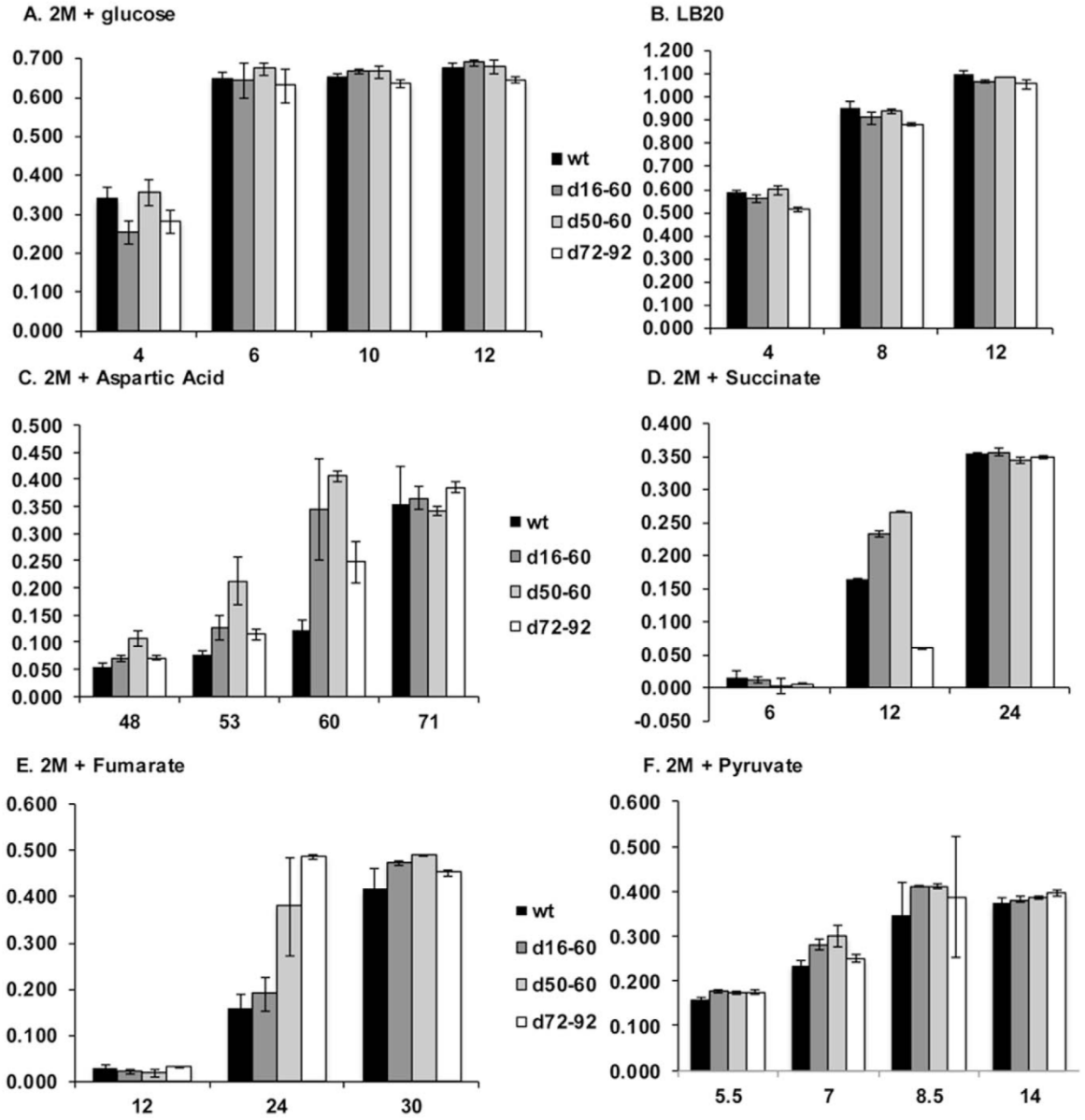
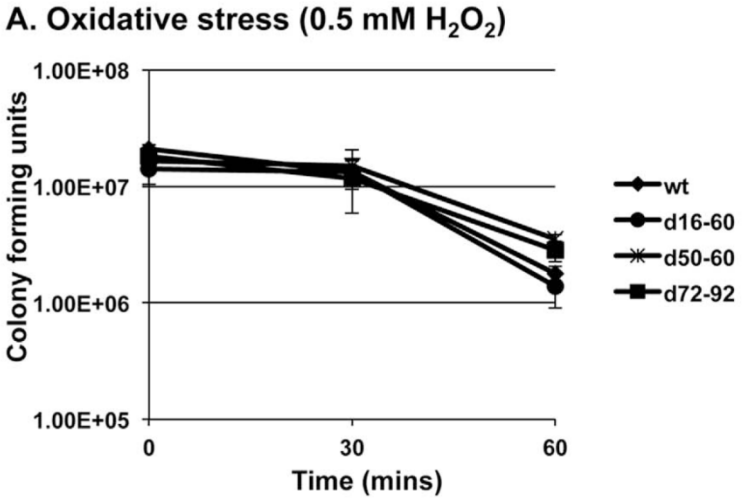


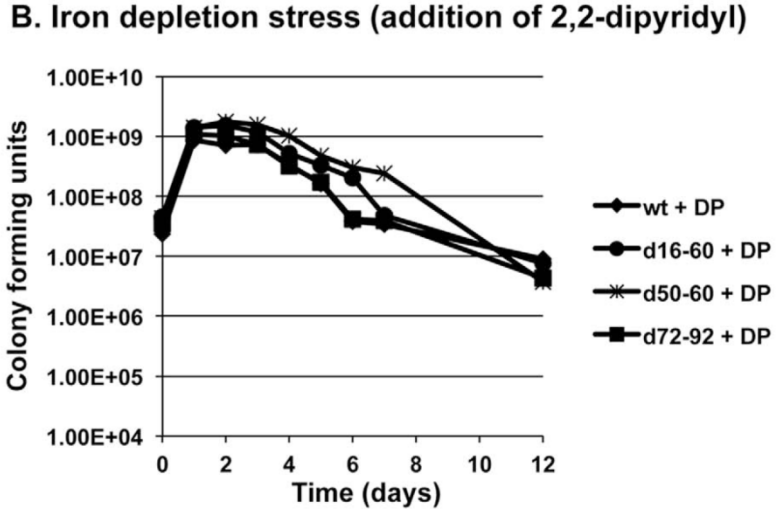
Figure 1. Growth curves. Growth of wt and deletion mutants in LB20 and 2M salts+differing carbon sources. Indicated on the horizontal axis are time intervals (hrs) where differences in growth between wt and deletion mutants were identified. Error bars show statistical significance of triplicate optical density readings.
doi:10.1371/journal.pone.0058430.g001

to the whole cell polysaccharide are confirmed to be a direct result of the gene cassette deletions.

The 3.0–4.5 ppm region is identified as the carbohydrate ring proton region of the diagnostic NMR spectrum. Alterations in the chemical shift between wt and deletion mutants indicate that there are changes in the functional groups attached to carbohydrate molecules. A peak labelled HDO (deuterated water) is present at 4.75 ppm (Figure 5 and Figure S2). Differences in the width of this peak indicated a saturated water signal as a consequence of incomplete exchange of all hydrogen atoms present in water with deuterium (see Materials and Methods). This does not affect results presented. The anomeric region labelled in Figure 5 shows little change in the chemical shift of peaks between wt and mutants indicating that there are no alterations in the type of sugar (i.e. glucose, sucrose, fructose) present along the backbone of the polysaccharide structure(s).

Overall, it is not surprising that there are distinct differences between the d16–60 and d72–92 mutants as these contain


A. Oxidative stress (0.5 mM H_2O_2)
Colony forming units
1.00E+08
1.00E+07
1.00E+06
1.00E+05
0
30
60
Time (mins)
wt
d16-60
d50-60
d72-92


B. Iron depletion stress (addition of 2,2-dipyridyl)
Colony forming units
1.00E+10
1.00E+09
1.00E+08
1.00E+07
1.00E+06
1.00E+05
1.00E+04
0
2
4
6
8
10
12
Time (days)
wt + DP
d16-60 + DP
d50-60 + DP
d72-92 + DP

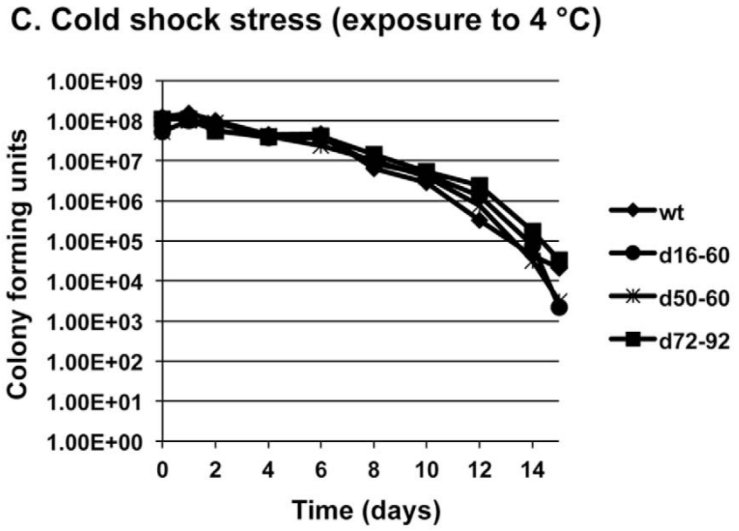

C. Cold shock stress (exposure to 4 °C)
Colony forming units
1.00E+09
1.00E+08
1.00E+07
1.00E+06
1.00E+05
1.00E+04
1.00E+03
1.00E+02
1.00E+01
1.00E+00
0
2
4
6
8
10
12
14
Time (days)
wt
d16-60
d50-60
d72-92

Figure 2. Environmental stress assays. Environmental stress assays showing oxidative (A), iron depletion (B) and cold shock stress (C). The figures shown here are representative of three independent experiments.
doi:10.1371/journal.pone.0058430.g002

different gene cassette deletions across the array. To elucidate exactly how the deletions have altered the polysaccharide(s) from *V. rotiferianus* DAT722, further comprehensive chemical characterisation of the polysaccharides is required. Our preliminary analyses has shown that there are at least two polysaccharide structures are produced by *V. rotiferianus* DAT722 (data not shown) and these would need to be separated, purified and the structure determined for each polysaccharide.

A Gene Cassette Deletion Affects Biofilm Formation

As alterations to bacterial surface polysaccharide can affect how the bacterial cell interacts with the environment, studies were performed to identify whether the deletion of cassettes impacted on biofilm formation. Deletion mutant d72–92 showed statistically significant higher biomass when compared to the wt, d16–60 and d50–60 when stained with crystal violet (Figure 6). It is likely that the changes in polysaccharide as a result of the 72–92 deletion, has altered the organism's ability to adhere or form a biofilm. Biofilm assays were carried out on a hydrophilic plastic surface however, different affects might be observed in the deletion mutants if adhesion or biofilm experiments were conducted on other surfaces such as metal, chitin or eukaryotic tissues.

Discussion

The integron/gene cassette system was first identified as a consequence of its contribution to the acquisition by bacteria of antibiotic resistance genes [23]. In clinical isolates, class 1 integrons have accumulated diverse tandem arrays of gene cassettes, most of which encode antibiotic resistance functions [24] driven by the strong selection imposed by the broad use of antibiotics by humans over the last 70 years. However, integrons and gene cassettes are ancient structures and the mobile gene cassette metagenome represents a vast reservoir of novel genes [1,5,25,26,27]. While the majority of the predicted gene products have no known function, environmental surveys strongly imply they are adaptive and where examined, structural and other studies support this notion [28,29,30].

Using *Vibrio rotiferianus* strain DAT722 as a model, we have recently shown that deletion of cassettes 8–60 produced mutants with a substantial loss of growth fitness [15]. Furthermore, this large deletion could not be supported without a suppressor mutation. This study demonstrated that recently acquired gene cassettes (or any mobile DNA) have the capacity to be rapidly integrated into pivotal cell networks [15]. This is in contrast to the generally held view that gene cassettes only impart accessory roles (e.g. antibiotic degradation), an issue we recently explored [31]. Following on from this study, it was of interest to determine how gene cassettes affect bacterial physiology in the absence of any detrimental growth defect. In contrast to d8–60, deletion of cassettes 16–60 (d16–60) had a healthy growth phenotype [15] and thus we utilised 2D-PAGE to determine how the large 46 cassette deletion altered protein physiology. Given the size of the deletion (~1/3 of the array) only 0.5–1% of the total proteome was differentially expressed in d16–60 with fold changes at approximately the 2-fold range (maximum change 3.6 fold). These changes were consistent across two different media (complete and minimal) and two different growth phases (mid-logarithmic and stationary). Since 2D-PAGE detects only the most abundant proteins in the cell, it can be stated that the deletion has not affected the major metabolic pathways of the cell. However, it is possible that there is a higher degree of change to lower abundance proteins not detected here. Nevertheless, in conjunction with the secreteome analysis, these data indicate that deletion of cassettes 16–60 has not adversely impacted on the major cellular pathways of the cell in contrast to our previous study with mutant d8–60 [15]. Furthermore, d16–60 and all other deletion mutants described in this study were not disadvantaged in environmental stress assays including oxidative stress, iron stress and cold stress. These data combined indicate that the majority of gene cassettes are maintained independently of host cell networks and are most likely not involved in stress adaptation.

During our analyses, evidence emerged that the gene cassette deletions alter host polysaccharide. This was confirmed with congo red staining and ^{1}H NMR analysis of whole cell purified polysaccharide. Multiple polysaccharide structures can exist on the Gram negative cell surface including lipopolysaccharide (LPS), a polysaccharide covalently linked to a lipid (lipid A core) that is embedded in the membrane and capsular/extracellular polysaccharide (C/EPS), a polysaccharide closely associated with the cell surface. From our preliminary ^{1}H NMR analyses, changes between the wt and both d16–60 and d72–92 mutants, as well as dissimilarities between the two mutants themselves was identified in the carbohydrate ring proton region. This indicates that gene cassette-associated products are most likely influencing functional groups linked to the sugar component of polysaccharide structure(s). For example, gene cassette products could be adding/removing functional groups such as NH_2 or CH_3 groups. The biological ramifications for such changes are substantial since they can affect processes such as bacterial-host interactions or virulence. For example, a recent study determined that the polysaccharide component of LPS plays a critical role in the colonisation of the light organ of squid species *Euprymna scolopes* [32]. CPS is also widely known to be important in virulence including a pandemic strain of *V. parahaemolyticus* [20]. Changes to surface polysaccharide are also likely to affect resistance to bacteriophage and biofilm formation [33]. We showed that as a consequence of deleting 22 cassettes to produce mutant d72–92, biofilm formation to a plastic hydrophilic surface was significantly increased when compared to wt, d16–60 and d50–60. This increased biofilm formation is likely due to the alterations observed in surface polysaccharide, especially since surface polysaccharides are known to influence biofilm formation [34].

Based on these data, the deletions appear to have affected surface polysaccharide with no major role in environmental stress survival and minor changes to overall protein expression as observed by 2D-PAGE and secretome analysis. In fact, some of the differentially expressed proteins observed in d16–60 could be explained by changes to surface polysaccharide. These changes would most likely alter the permeability of the cell and when considering the interconnected nature of cell envelope structures, re-regulation of general porins (e.g. OmpU; spot 2CM in Table 2) and outer membrane proteins (OmpA; spot 1CM in Table 2) probably occurred to compensate. Consistent with this, outer membrane protein extractions also showed higher expression of an OmpU-like protein in d72–92 (data not shown). This change in permeability and subsequent re-regulation of porins may explain the minor variations in growth for all the deletion mutants. At this stage we cannot know how gene cassette products are altering host polysaccharide as further chemical characterisation of host

Table 2. Differentially expressed spots between deletion mutant d16–60 mutant and wild-type *Vibrio rotiferianus* DAT722 in LB20.

Growth Phase	Differentially expressed spot	LC-MS/MS match(s)	Number of matched peptides	PEAKS score* (%)	Fold change in mutant	Accession number
Mid-logarithmic phase						
	1CM	OmpA; VrotD_16305	8	99.0	−2.31	ZP_08911638
	2CM	OmpU-like outer membrane protein	10	98.8	−3.68	ZP_08912594
	^3CM	50S ribosomal protein L9	10	98.8	−2.61	ZP_08908422
		OmpA-like membrane protein; VrotD_08232	7	97.5		ZP_08910036
	^4CM	ATP-dependent Clp protease proteolytic subunit	9	98.3	−2.90	ZP_08909655
		S-ribosylhomocysteinase/Autoinducer-2 production protein LuxS	5	97.9		ZP_08911325
		Type VI secretion-related protein	3	96.8		ZP_08909611
		Shikimate kinase I	3	96.4		ZP_08908430
Stationary phase						
	1CS	Unknown protein (gene is surrounded by genes encoding O-antigen biosynthesis or export); VrotD_02720	15	99.0	+2.30	ZP_08908944
	2CS	Nitrogen regulatory protein P-II	13	96.9	−2.75	ZP_08911087

*highest PEAKS score (percentage based on a p-value <0.05) was taken as the closest peptide match

^denotes co-migrating protein spots.

doi:10.1371/journal.pone.0058430.t002

Table 3. Differentially expressed spots between deletion mutant d16–60 mutant and wild-type *Vibrio rotiferianus* DAT722 in 2M+glucose.

Growth Phase	Differentially expressed spot	LC-MS/MS match(s)	Number of matched peptides	PEAKS score* (%)	Fold change in mutant	Accession number
Mid-logarithmic phase						
	^1MM	Putative membrane protein; VrotD_04538	6	98.6	+2.73	ZP_08909304
		Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (Histidine biosynthesis)	4	95.1		ZP_08910780
	2MM	Alkyl hydroperoxide reductase subunit C-like protein	3	98.3	+2.45	ZP_08909276
	3MM	6,7-dimethyl-8-ribityllumazine synthase	15	99.0	−2.01	ZP_08909366
Stationary phase						
	^1MS	Cysteine synthase A	12	95.4	+2.11	ZP_08909513
		Acetyl-coenzyme A carboxyl transferase alpha chain	7	92.0		ZP_08911090
		OmpT	17	90.2		ZP_08909742
	2MS	Unknown protein; VrotD_21668	10	98.3	−2.00	ZP_08912704
	^3MS	Unknown protein; VrotD_07347	16	96.1	+2.55	ZP_08909861
		NAD-dependent glyceraldehyde-3-phosphate dehydrogenase	13	96.0		ZP_08910944
		Phenylalanyl-tRNA synthetase alpha chain	9	94.4		ZP_08910491
		Universal stress protein E	10	93.8		ZP_08910109

*highest PEAKS score (percentage based on a p-value <0.05) was taken as the closest peptide match.
^denotes co-migrating protein spots.
doi:10.1371/journal.pone.0058430.t003

polysaccharide is required, however, it is intriguing to hypothesise that gene cassette products modify or decorate host polysaccharide through addition of functional groups or sugars. A prior study had identified a gene cassette encoding an uncharacterised gene as important for CPS biosynthesis in a strain of *V. vulnificus* [35]. In the *V. rotiferianus* DAT722 array, there are some gene cassettes that suggest a role for polysaccharide modification or biosynthesis (Table S1). Cassette 31 contains a gene that encodes a putative β-phosphoglucomutase protein, a phosphotransferase that transfers a phosphate group to glucose and cassette 78 which encodes a putative O-acetyltransferase. Furthermore, there are numerous acetytransferases in the cassette array, four each span the deletions in d16–60 and d72–92. Research in our laboratory is underway to express such genes *in trans* to see whether they change the ^{1}H NMR spectra and level of biofilm formation in the deletion mutants.

This study has answered important questions as well as advanced the integron biology field. Firstly, gene cassette products are highly novel and at this stage of research an attempt to identify phenotypes is difficult and relies on randomly selecting a phenotype to test. By determining that surface polysaccharide is a target of change in these deletion mutants, future research can be aimed at determining how gene cassettes modify surface polysaccharide. Secondly, this study conclusively demonstrated that gene cassettes do not need to be expressed from P_c to have an physiological impact and corroborates our previous study that identified numerous promoters in the *V. rotiferianus* DAT722 cassette array [9]. Thirdly, indel events are common in the cassette array of vibrios [5] indicating they have an important evolutionary and adaptive role. For the first time, we have shown how this affects *Vibrio* physiology by showing that a large deletion appears to largely affect surface polysaccharide. Future experiments are aimed at determining the biological ramifications of such changes by testing the deletion mutants in more biological assays such as biofilm formation to various substrata and bacteriophage assays.

Conclusions

From this study we conclude that deletion of subsets of gene cassettes along the 116 cassette long array impacts on surface cell polysaccharide structures. How gene cassettes are altering polysaccharide structures requires further chemical analysis, however, any structural change to bacterial host polysaccharide is likely to impact on how the cell interacts with its environment and with other organisms within its environment.

Materials and Methods

Bacterial Strains and Growth Conditions

Bacterial strains and plasmids used in this study are listed in Table 1. *Vibrio* strains were routinely grown on Luria-Bertani medium supplemented with 2% NaCl (LB20) at 28°C. *Escherichia coli* strains were routinely grown on Luria-Bertani medium. Growth curves of all *Vibrio* strains were conducted in 24 well microtitre plates containing 1 ml of medium per well. The inoculum was from overnight cultures grown in LB20 and then diluted to OD_{600} of 0.7 using 2% NaCl. Growth curve cultures were inoculated at 1:100 and growth measured using a microtitre plate reader (Synergy HT Bio-Tek) at OD_{590}, and Gen5 (Bio-Tek) software. In experiments comparing growth of the wt and deletion mutants with different carbon sources, a marine minimal salts medium (2M) which mimics a seawater environment [36] was used supplemented with a carbon source (glucose at 11.1 mM and

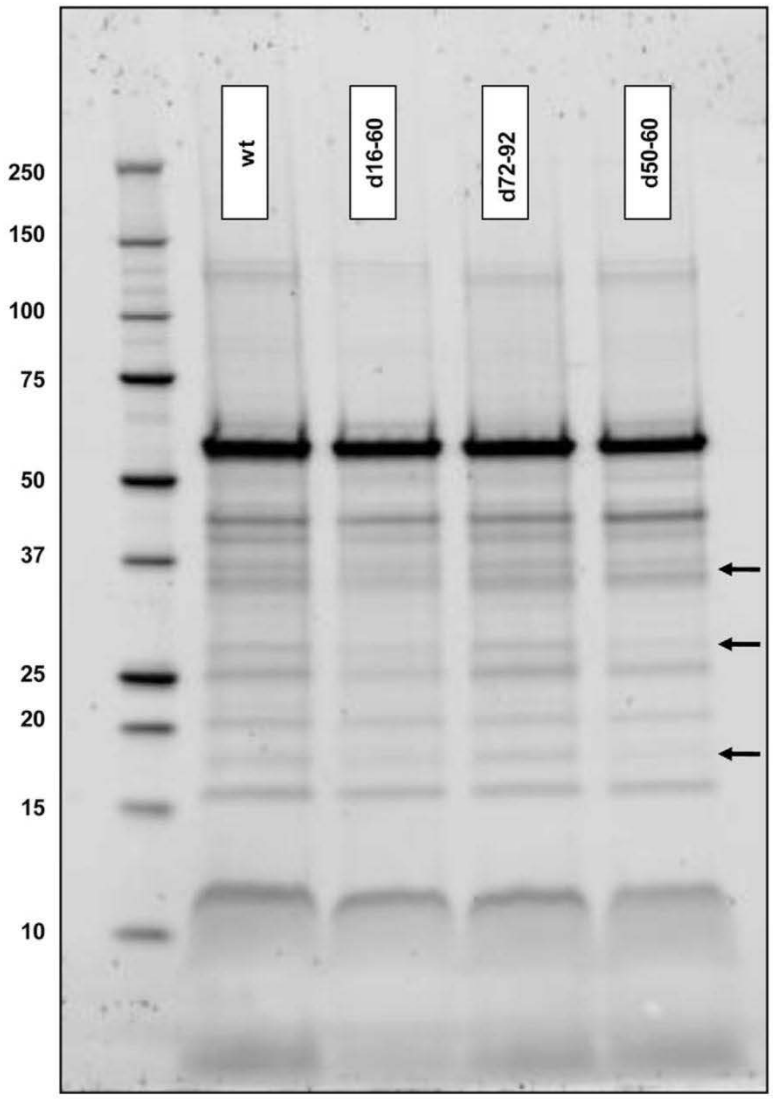


Figure 3. Gel electrophoresis of supernatant proteins. Gel electrophoresis of precipitated supernatant proteins from *V. rotiferianus* DAT722 (wt) and deletion mutants grown in 2M+glucose. Protein standards indicated left of the gel have sizes in kDa. Lanes are labeled with strain names and arrows indicate protein bands of differing abundance.
doi:10.1371/journal.pone.0058430.g003

aspartic acid, succinate and fumarate at 20 mM respectively). Kanamycin was used at 100 μg/ml.

DAT722 Cassette Analysis and Strain Construction

The cassette array of DAT722 is fully sequenced and consists of 116 gene cassettes although there are 94 different cassette types due to the presence of paralogous cassettes [37,38]. Construction of the deletion mutants (Table 1) is as described previously [15]. Briefly, pMAQ1081 containing a 1834 bp fragment inserted into the *sacB*-counter selectable suicide vector pCVD442 [39] was used to create deletions in the cassette array of *V. rotiferianus* DAT722. The fragment consisted of two sequences with homology to different paralogous cassettes across the array disrupted with a kanamycin resistance gene. Conjugation of this construct into *V. rotiferianus* DAT722 allowed for allele replacement and deletion of cassettes between these two sets of paralogous cassettes. Deletion mutants d16–60 and d50–60 were created by taking a merodiploid (designated MD7) consisting of pMAQ1081 recombined into cassette 61 and screening colonies counter selected on sucrose medium with primers targeting unique cassettes outside the expected deletions. An identical approach was taken for creating d16–60a. An independently derived but identical merodiploid to MD7 (designated SC-8B61) was used. Deletion mutant d72–92 was isolated as a double-crossover and did not undergo sucrose counter selection. Deletion mutant d72–92a was created by taking a merodiploid (designated SC-8A91) consisting of pMAQ1081

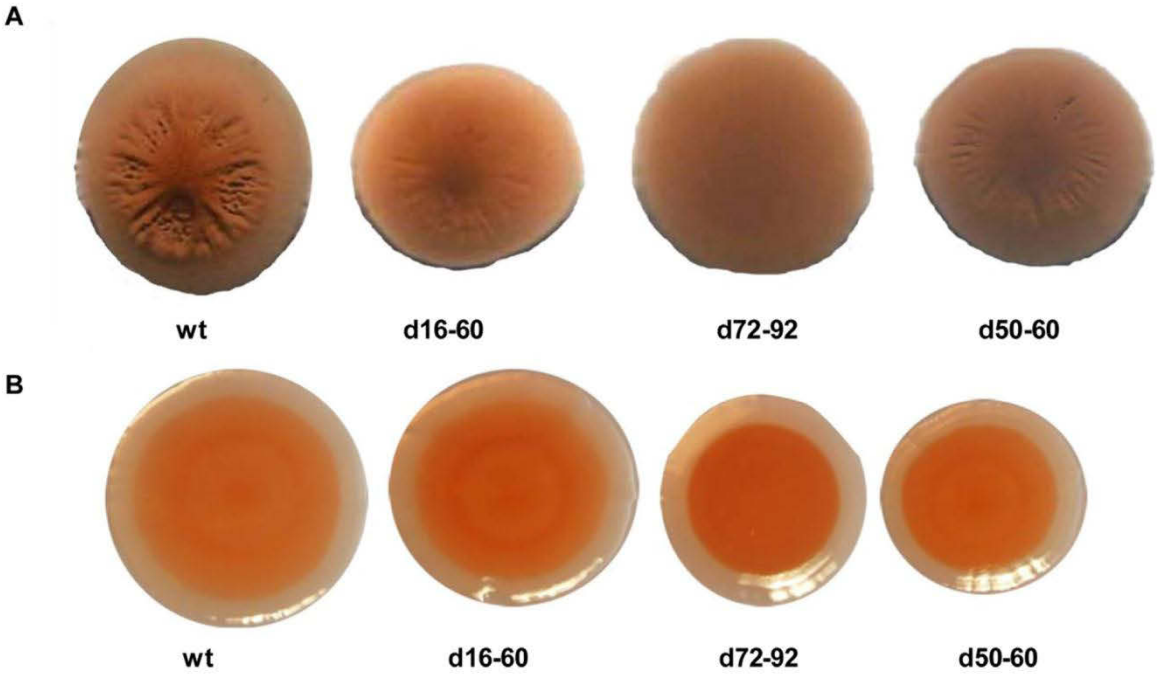


Figure 4. Congo red colony morphology. Colonies of *V. rotiferianus* DAT722 and deletion mutants grown on agar plates supplemented with 0.001% congo red. Colonies grown on LB20 and 2M +0.2% glucose supplemented with congo red are shown in panels A and B respectively. All colonies were imaged after 7 days growth at 28°C.
doi:10.1371/journal.pone.0058430.g004

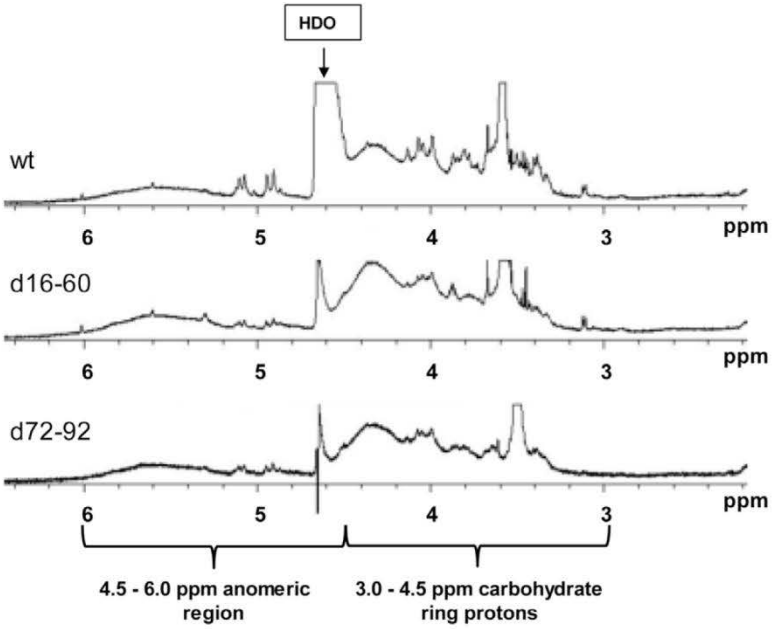


Figure 5. Water suppression H^1 NMR spectra. H^1 NMR spectra of wt, d16–60 and d72–92 whole cell polysaccharide.
doi:10.1371/journal.pone.0058430.g005

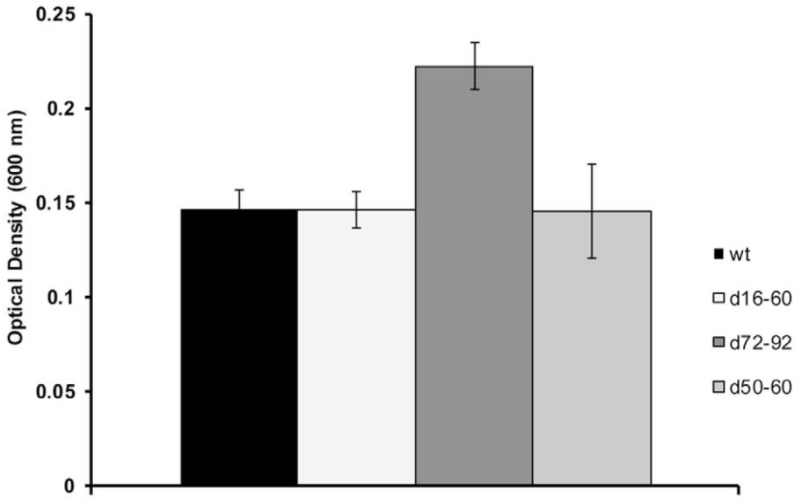


Figure 6. Graph showing optical density of crystal violet stained cells. Bar graphs showing significantly higher adhesion of d72–92 indicated by the asterisk.
doi:10.1371/journal.pone.0058430.g006

recombined into cassette 93 and screening colonies counter selected on sucrose medium as described above.

Stress Assays

V. rotiferianus DAT722 and deletion mutant were stressed with the following conditions as follows. All experiments were carried out in triplicate with data given in figures representative of the triplicate data.

Oxidative stress. 1 mL of an overnight culture grown in LB20 was washed with 0.55×NSS and diluted 1:10 in 0.55×NSS. The diluted culture was exposed to 0.5 mM hydrogen pyroxide with samples taken for enumeration at 0, 30 and 60 minutes post addition of H_2O_2.

Iron depletion stress. 100 µl of an overnight LB20 culture was inoculated into 2M+glucose containing 0.1 mM of the iron chelating agent 2′2′,-dipyridyl (DP) and incubated at 28°C with shaking for 12 days. Samples were taken daily for enumeration of viable cells.

Cold shock. Cells were grown to mid-logarithmic phase in 2M+glucose (OD_{600} ~ 0.3) and then placed at 4°C. Samples were taken daily for enumeration of viable cells.

Differential Display Analysis of 2D-PAGE Gels

Overnight cultures of wt and mutant strains grown in LB20 and 2M+glucose were resuspended in 1 ml solution containing 1% C7bz0, 2M thiourea, 7M urea, 40 mM Tris and 50 mM LiCl. Resuspended samples were processed according to [40] and run on a 2D-PAGE gel in triplicate. Triplicate wt and mutant 2D-PAGE gels were analysed for post-translational modifications and up/down-regulations using software program PDQuest (Bio-Rad ver 8.0). Differentially displayed protein spots were cut out, trypsin digested and run through LC-MS/MS for identification at the UTS Protein and Proteomics Core Facility. LC-MS/MS data was run through PEAKS Studio software (Bioinformatics Solutions) in order to compare LC-MS/MS identified peptides to a protein output file acquired from the RAST annotated *V. rotiferianus* DAT722 genome [41].

Supernatant Protein Extraction and Gel Electrophoresis

Cells were grown in 2M +0.2% glucose minimal media for 17 hrs at 28°C. Cells were collected by centrifugation (4000×*g*) and the supernatant was collected and filtered through a 0.2 µM filter. To precipitate supernatant proteins, five volumes of acetone was added, mixed by gentle inversion and incubated at −20°C for 30 mins. Precipitated proteins were collected by centrifugation at 3000×*g* for 3 mins and supernatant discarded and the pellet dried at 37°C overnight prior and total protein to weighing. The protein pellet was then resuspended to a concentration of 40 mg/ml in 2M thiourea, 7M urea and 1% C7bz0.

Prior to gel electrophoresis samples were run through a Micro-Biospin Column (Bio-Rad) according to their protocol to remove excess salt from the sample due to media. 400 µg of protein was loaded onto a 4–12% Bis/Tris precast polyacrylamide 1D gel (Bio-Rad) and run at 160V for ~60 mins. Gels were then fixed for 30 mins in 10% acetic acid (v/v) and 40% methanol (v/v) prior to staining with Flamingo protein stain (Bio-Rad). Ladder used was Bio-Rad Precision Plus Protein Unstained Standard (catalog # 161-0363).

Congo Red Staining of Bacterial Colonies

Wild-type *Vibrio rotiferianus* DAT722 and isogenic deletion mutants were plated out for single isolated colonies on LB20 and 2M +0.2% glucose plates containing 0.001% congo red (Sigma) and left for 7 days at 28°C. After 7 days colonies were observed on an Olympus SZX12 stereomicroscope at a magnification of 125× using a Colourview colour camera and images were collected using Image Analysis software (Olympus).

Extraction of Whole Cell Polysaccharide

Whole cell polysaccharide of mutants grown in 2M +0.2% glucose overnight was extracted using a proteinase K, phenol-water method adapted from Apicella, 2008 [42]. For the ^{1}H NMR scans of extracts shown in Figure 5, benzonase was used to remove contaminating DNA and RNA. For the ^{1}H NMR scans of extracts shown in Figure S2, DNase I and RNase A was used to remove contaminating DNA and RNA.

NMR Spectroscopy of Whole Cell Polysaccharide

Equal amounts of purified polysaccharide were exchanged three times using D_2O as a solvent. 1D water suppression ^1H NMR experiments were performed on purified whole cell polysaccharide resuspended in 600 µL of D_2O using an Agilent Technologies 500 MHz NMR instrument at 28°C with the internal reference of the sodium salt of 3-(trimethylsilyl)-3,3,2,2-tetradeuteropropionic acid. The typical acquisition parameters utilized were spectral width 8012 Hz, acquisition time 4.089s, relaxation delay 1.5s and line boarding frequency 0.5 Hz.

Biofilm/adhesion Assays

Overnight cultures of wt, d16–60, d72–92 and d50–60 grown in 2M+glucose were diluted 1:100 and 500 µl added to the well on a flat bottom, hydrophilic, plastic 24-well microtitre plate (Nunc ThermoFisher) in quadruplet and incubated for 24 hrs at 28°C with shaking at 200 rpm. Quadruplet control wells were also set up containing media only. Unattached planktonic cells were then carefully removed by pipetting and wells were washed twice with 2% NaCl. 500 µl of crystal violet (0.2% w/v) was then added to wells and incubated for 15 mins to stain the adhered cells. The crystal violet was removed and the wells washed three times with 2% NaCl to remove excess stain. To measure the biomass of adhered cells, the crystal violet was solubilised with 30% (v/v) acetic acid, transferred to a clean microtitre plate and optical density (600 nm) measured using a Synergy HT plate reader (Bio-Tek). This experiment was repeated on three separate occasions.

Supporting Information

Figure S1 Isoelectric focusing contamination. Triplicate 2D-PAGE gels of protein extracted from d16-60 cells grown in 2M+glucose to stationary phase. Horizontal streaking indicates disruption of the IEF stage due to a contaminating substance.
(TIF)

Figure S2 Water suppression NMR spectra replicate. H^1 NMR spectra of wt, d16-60a and d72-92a whole cell polysaccharide.
(TIF)

Table S1 DAT722 ordered cassette array. List of 116 *V. rotiferianus* DAT722 gene cassettes with putative identification and putative conserved superfamily domains of proteins encoded by genes contained within the gene cassettes.
(DOCX)

Table S2 Genome region encoding unknown protein found in 2D-PAGE. *V. rotiferianus* DAT722 genome region containing gene encoding unknown protein identified in spot 1CS. The gene encoding this protein is the genomic region responsible for polysaccharide biosynthesis. Sequence found at accession # NZ_AFAJ01000014.
(TIF)

Acknowledgments

We thank Dr. Matthew Padula for assistance with the 2D-PAGE and LC-MS/MS experiments.

Author Contributions

Conceived and designed the experiments: ML HS RR SD. Performed the experiments: RR ML. Analyzed the data: ML RR HS RS SD. Contributed reagents/materials/analysis tools: ML HS. Wrote the paper: RR ML HS.

References

1. Cambray G, Guerout AM, Mazel D (2010) Integrons. In: Campbell A, Lichten M, Schupbach G, editors. Annu Rev Genet: Annual Reviews. 141–166.
2. Labbate M, Case RJ, Stokes HW (2009) The integron/gene cassette system: an active player in bacterial adaptation. Methods Mol Biol 532: 103–125.
3. Nield BS, Holmes AJ, Gillings MR, Recchia GD, Mabbutt BC, et al. (2001) Recovery of new integron classes from environmental DNA. FEMS Microbiol Lett 195: 59–65.
4. Partridge SR (2011) Analysis of antibiotic resistance regions in Gram-negative bacteria. FEMS Microbiol Rev 35: 820–855.
5. Boucher Y, Labbate M, Koenig JE, Stokes HW (2007) Integrons: mobilizable platforms that promote genetic diversity in bacteria. Trends Microbiol 15: 301–309.
6. Thompson LF, Iida T, Swings J (2004) Biodiversity of Vibrios. Microbiol Mol Biol Rev 68: 403–431.
7. Hazen TH, Pan L, Gu JD, Sobecky PA (2010) The contribution of mobile genetic elements to the evolution and ecology of Vibrios. FEMS Microbiol Ecol 74: 485–499.
8. Guerin E, Cambray G, Sanchez-Alberola N, Campoy S, Erill I, et al. (2009) The SOS response controls integron recombination. Science 324: 1034–1034.
9. Michael C, Labbate M (2010) Gene cassette transcription in a large integron-associated array. BMC Genet 11: 82.
10. Barker A, Clark C.A. (1994) Identification of VCR, a repeated sequence associated with a locus encoding a hemagglutinin in *Vibrio cholerae* O1. J Bacteriol 176: 5450–5458.
11. Barker A, Manning P.A (1997) VlpA of *Vibrio cholerae* O1: the first bacterial member of the alpha 2-microglobulin lipocalin superfamily. Microbiology 143: 1805–1813.
12. Deshpande CN, Harrop SJ, Boucher Y, Hassan KA, Di Leo R, et al. (2011) Crystal structure of an integron gene cassette-associated protein from *Vibrio cholerae* identifies a cationic drug-binding module. PLoS ONE 6.
13. Franzon VL, Barker A, Manning P (1993) Nucleotide sequence encoding the mannose-fucose-resistant hemagglutinin of *Vibrio cholerae* O1 and contstruction of a mutant. Infect Immun 61: 3032–3037.
14. Ogawa A, Takeda,T. (1993) The gene encoding the heat-stable enterotoxin of *Vibrio cholerae* is flanked by 123-bp direct repeats. Microbiol and Immunol 37: 607–616.
15. Labbate M, Boucher Y, Roy Chowdhury P, Stokes H (2011) Integration of a laterally acquired gene into a cell network important for growth in a strain of *Vibrio rotiferianus*. BMC Microbiol 11: 253.
16. Labbate M, Boucher Y, Joss MJ, Michael CA, Gillings MR, et al. (2007) Use of chromosomal integron arrays as a phylogenetic typing system for *Vibrio cholerae* pandemic strains. Microbiology 153: 1488–1498.
17. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, et al. (2010) PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26: 1608–1615.
18. Heppelmann CJ, Benson LM, Bergen HR (2007) A simple method to remove contaminating salt from IPG strips prior to IEF. Electrophoresis 28: 3988–3991.
19. Shaw MM, Riederer BM (2003) Sample preparation for two-dimensional gel electrophoresis. Proteomics 3: 1408–1417.
20. Chen Y, Dai J, Morris JG, Jr., Johnson JA (2010) Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio parahaemolyticus* O3: K6. BMC Microbiol 10.
21. Joseph LA, Wright AC (2004) Expression of *Vibrio vulnificus* capsular polysaccharide inhibits biofilm formation. J Bacteriol 186: 889–893.
22. Poutrel B, Gilbert FB, Lebrun M (1995) Effects of culture conditions on production of Type-5 capsular polysaccharide by human and bovine *Staphylococcus aureus* strains. Clin Diagn Lab Immun 2: 166–171.
23. Hall RM (1997) Mobile gene cassettes and integrons: Moving antibiotic resistance genes in Gram-negative bacteria. In: Chadwick DJGJ, editor. Antibiotic Resistance: Origins, Evolution, Selection and Spread. 192–202.
24. Partridge SR, Tsafnat G, Coiera E, Iredell JR (2009) Gene cassettes and cassette arrays in mobile resistance integrons. FEMS Microbiol Rev 33: 757–784.
25. Gillings M, Holley Marita, Stokes H.W., Holmes Andrew (2005) Integrons in *Xanthomonas*: A source of species genome diversity. PNAS 102: 4419–4424.
26. Koenig JE, Boucher Y, Charlebois RL, Nesbo C, Zhaxybayeva O, et al. (2008) Integron-associated gene cassettes in Halifax Harbour: assessment of a mobile gene pool in marine sediments. Environ Microbiol 10: 1024–1038.
27. Michael CA, Gillings MR, Holmes AJ, Hughes L, Andrew NR, et al. (2004) Mobile gene cassettes: A fundamental resource for bacterial evolution. Am Nat 164: 1–12.
28. Nield BS, Willows R.D., Torda A.E., Gillings M.R., Holmes A.J., Nevalainen K.M., H Stokes, H.W Mabbutt, B.C. (2004) New enzymes from environmental cassette arrays: Functional attributes of a phosphotransferase and an RNA-methyltransferase. Protein Sci 13: 1651–1659.
29. Robinson A, Guilfoyle AP, Sureshan V, Howell M, Harrop SJ, et al. (2008) Structural genomics of the bacterial mobile metagenome: an overview. Methods Mol Biol (Clifton, NJ) 426: 589–595.

30. Robinson A, Wu PSC, Harrop SJ, Schaeffer PM, Dosztányi Z, et al. (2005) Integron-associated mobile gene cassettes code for folded proteins: the structure of Bal32a, a new member of the adaptable [alpha]+[beta] barrel family. J Mol Biol 346: 1229–1241.
31. Labbate M, Boucher Y, Luu I, Roy Chowdhury P, Stokes HW (2012) Integron associated mobile genes: Just a collection of plug in apps or essential components of cell network hardware? Mob Genet Elem In Press.
32. Post DMB, Yu L, Krasity BC, Choudhury B, Mandel MJ, et al. (2012) The O-antigen and core carbohydrate of Vibrio fischeri lipopolysaccharide: Composition and analysis of their role in Euprymna scolopes light organ colonization. J Biol Chem.
33. Labrie SJ, Samson JE, Moineau S (2010) Bacteriophage resistance mechanisms. Nature Rev Microbiol 8: 317–327.
34. Fong JCN, Syed KA, Klose KE, Yildiz FH (2010) Role of *Vibrio* polysaccharide (*vps*) genes in VPS production, biofilm formation and *Vibrio cholerae* pathogenesis. Microbiology 156: 2757–2769.
35. Smith AB, Siebeling RJ (2003) Identification of genetic loci required for capsular expression in *Vibrio vulnificus*. Infect Immun 71: 1091–1097.
36. Paludan-Müller C, Weichart D, McDougald D, Kjelleberg S (1996) Analysis of starvation conditions that allow for prolonged culturability of *Vibrio vulnificus* at low temperature. Microbiology 142: 1675–1684.
37. Boucher Y, Stokes.W.H. (2006) The roles of lateral gene transfer and vertical descent in *Vibrio* evolution. In: Fabiano Lopes Thompson BA, J. G Swings, editor. The Biology of Vibrios.
38. Chowdhury PR, Boucher Y, Hassan KA, Paulsen IT, Stokes HW, et al. (2011) Genome Sequence of Vibrio rotiferianus Strain DAT722. J Bacteriol 193: 3381–3382.
39. Donnenberg MS, Kaper JB (1991) Construction of an *eae* deletion mutant of enteropathogenic *Escherichia coli* by using a positive-selection suicide vector. Infect Immun 59: 4310–4317.
40. Jobbins SE, Hill CJ, D'Souza-Basseal JM, Padula MP, Herbert BR, et al. (2010) Immunoproteomic Approach to Elucidating the Pathogenesis of Cryptococcosis Caused by *Cryptococcus gattii*. J Proteome Res 9: 3832–3841.
41. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, et al. (2008) The RAST server: Rapid annotations using subsystems technology. BMC Genomics 9: 15.
42. Apicella MA (2008) Isolation and characterization of lipopolysaccharides. Methods Mol Biol (Clifton, NJ) 431: 3–13.
43. Boucher Y, Nesbo C, Joss M, Robinson A, Mabbutt B, et al. (2006) Recovery and evolutionary analysis of complete integron gene cassette arrays from *Vibrio*. BMC Evol Biol 6: 3.

Appendix 3

environmental microbiology

Environmental Microbiology (2014) doi:10.1111/1462-2920.12512

A genomic island integrated into *recA* of *Vibrio cholerae* contains a divergent *recA* and provides multi-pathway protection from DNA damage

Rita A. Rapa,[1,2] Atiqul Islam,[1] Leigh G. Monahan,[1] Ankur Mutreja,[3] Nicholas Thomson,[3] Ian G. Charles,[1] Harold W. Stokes[1] and Maurizio Labbate[1,2]*

[1]*ithree Institute,* [2]*Department of Medical and Molecular Biosciences, University of Technology, PO Box 123 Broadway, Sydney, NSW 2007, Australia.*

[3]*Wellcome Trust Sanger Institute, Cambridge, UK.*

Summary

Lateral gene transfer (LGT) has been crucial in the evolution of the cholera pathogen, *Vibrio cholerae*. The two major virulence factors are present on two different mobile genetic elements, a bacteriophage containing the cholera toxin genes and a genomic island (GI) containing the intestinal adhesin genes. Non-toxigenic *V. cholerae* in the aquatic environment are a major source of novel DNA that allows the pathogen to morph via LGT. In this study, we report a novel GI from a non-toxigenic *V. cholerae* strain containing multiple genes involved in DNA repair including the recombination repair gene *recA* that is 23% divergent from the indigenous *recA* and genes involved in the translesion synthesis pathway. This is the first report of a GI containing the critical gene *recA* and the first report of a GI that targets insertion into a specific site within *recA*. We show that possession of the island in *Escherichia coli* is protective against DNA damage induced by UV-irradiation and DNA targeting antibiotics. This study highlights the importance of genetic elements such as GIs in the evolution of *V. cholerae* and emphasizes the importance of environmental strains as a source of novel DNA that can influence the pathogenicity of toxigenic strains.

Introduction

Vibrio cholerae is a common inhabitant of marine and estuarine waters and is the causative agent of the diarrheal disease cholera. Although there are over 200 O-antigen serogroups among *V. cholerae* strains, only two, O1 and O139, are known to cause pandemics of cholera disease (Kaper *et al.*, 1995). Lateral gene transfer (LGT) has largely contributed to the emergence of new pandemic strains of cholera (Faruque and Mekalanos, 2003; Keymer and Boehm, 2011). The appearance of the O139 serogroup and the so-called hybrid strains in the early 1990s are prime examples (Ramamurthy *et al.*, 2003; Safa *et al.*, 2009). Mobile genetic elements (MGEs) have been pivotal in the evolution of *V. cholerae* including diverse elements, such as the genomic islands (GIs) VPI-1, VPI-2, VSP-1, VSP-2, an integrative conjugative element, and the bacteriophage CTX (Faruque and Mekalanos, 2003; Grim *et al.*, 2010). GIs are defined as large chromosomal regions that have features suggestive of recent LGT (Boyd *et al.*, 2008). They have the capacity to excise and form circular intermediates and often target tRNA loci for their integration. In *V. cholerae*, GIs have been implicated in causing human disease and in environmental survival. For example, the replacement of the O1 classical biotype by the O1 El Tor biotype in the 1960s is suggested to be due to the acquisition of VSP-1 and VSP-2 that have probably enhanced epidemic spread (Faruque and Mekalanos, 2003). VPI-2 is a 57.3 kb island integrated at *tRNA-Ser* and encodes a neuraminidase important for converting higher-order sialogangliosides to GM1 gangliosides, the receptor for cholera toxin (Galen *et al.*, 1992). Moreover, VPI-1 encodes for the toxin-coregulated pilus (TCP), an essential intestinal colonization factor, as well as the accessory colonization factor (ACF), and virulence regulators ToxT and TcpPH (Everiss *et al.*, 1994; Murphy and Boyd, 2008). Non-O1/O139 *V. cholerae* strains are considered to be the major source of laterally acquired DNA for O1/O139 strains (Meibom *et al.*, 2005) thus, a better understanding of the diverse genetic elements present in the *V. cholerae* species is important for predicting and mitigating the emergence of new pandemic strains.

In bacteria, errors in DNA can occur as part of normal DNA replication or can be induced by external stimuli (e.g. UV irradiation) (Janion, 2008). There are several

Received 31 January, 2014; accepted 13 May, 2014. *For correspondence. E-mail maurizio.labbate@uts.edu.au; Tel. (61-2) 9514 4064; Fax (61-2) 9514 8206.

genetic systems involved in error-free DNA repair including base excision repair (BER), nucleotide excision repair (NER), recombinational DNA repair and mismatch repair (MMR) (Rattray and Strathern, 2003; Janion, 2008; Polosina and Cupples, 2010; Lenhart *et al*., 2012). However, if DNA damage is extensive the mutagenic phase of the SOS response is triggered (Goodman, 2002). This response is mediated by DNA polymerases that replicate past template lesions in a process called translesion DNA synthesis (TLS) that is inherently error-prone (Goodman, 2002). For example, DNA polymerase V, encoded by the *umuDC* operon (Patel *et al*., 2010). The SOS induction of error-prone polymerases is considered a final response where although induced mutation(s) may be deleterious to the host cell, this is balanced against the need for rapid DNA repair (Goodman, 2002). An alternative view for the function of error-prone polymerases is that they act to generate genetic diversity that may have a role in environments where the host is maladapted by providing a bank of pre-existing genetic diversity within that population, some of which may confer a positive selective advantage. To support this second view, transcription of error-prone polymerases has been observed in the absence of SOS inducing DNA damage (Yeiser *et al*., 2002). Furthermore, error-prone polymerase mutants are less competitive than the parent cells during starvation (McKenzie *et al*., 2000; Yeiser *et al*., 2002; Tark *et al*., 2005), and some antibiotics (e.g. quinolones) induce the SOS mutagenic response increasing the frequency of resistant mutants (Piddock and Wise, 1987; Ysern *et al*., 1990).

In this study we report a novel GI inserted into *recA* of *V. cholerae* non-O1/O139 strain S24 isolated from an estuarine river in Sydney, Australia. This strain lacks the major virulence factors: cholera toxin and the toxin-coregulated pilus, thus is not capable of causing cholera. The GI carries (i) a *recA* gene phylogenetically distant from the disrupted host *recA*, designated $recA_{RME}$; (ii) a *umuDC* operon, designated $umuDC_{RME}$, encoding DNA polymerase V; and (iii) genes encoding hypothetical proteins, proteins with DNA processing domains including a MutL domain involved in MMR, and proteins involved in site-specific recombination. The GI can excise as a closed circle and preferentially inserts into a specific site within *recA*. We also show that $recA_{RME}$ is functional and provides protection from UV irradiation, a common source of DNA damage encountered in the shallow waters of marine and estuarine environments. Furthermore, the GI provides protection from the antibiotics bleomycin and ciprofloxacin. Acquisition of this GI by O1/O139 toxigenic *V. cholerae* would not only enhance survival of this pathogen in the natural environment but may also provide enhanced protection from DNA targeting antibiotics such as ciprofloxacin.

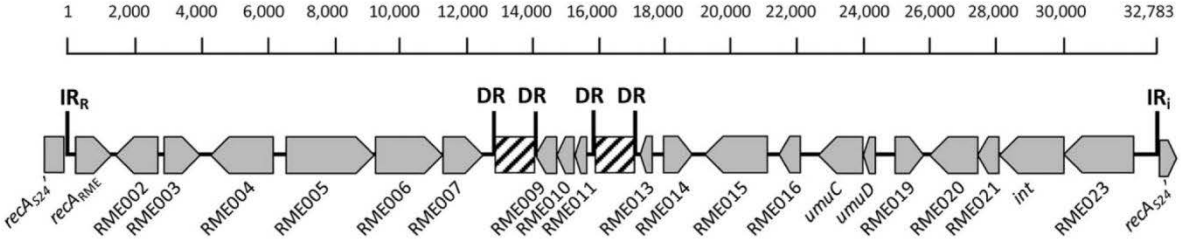
Results and discussion

Identification of a novel genomic island in V. cholerae *S24 containing* recA

S24 is an environmental, non-O1/O139 *V. cholerae* strain isolated from Georges River in Sydney, Australia, as described in a previous study (Islam *et al*., 2013). It was noted during multilocus sequence analysis of housekeeping genes *adk*, *gyrB*, *mdh* and *recA* using primers designed to amplify the *V. cholerae* S24 *recA,* (designated here $recA_{S24}$), that a product of ~ 1.5 kb was identified instead of the expected ~ 850 bp. When the *V. cholerae* S24 draft genome sequence (to be released at a later date) was interrogated, it was noted that $recA_{S24}$ was present on two separate contigs. PCR, using primers designed to sequence within these contigs, was used to close this region of the genome (described in *Experimental procedures*) resulting in a final contig of 262,869 bp. Within this contig and disrupting $recA_{S24}$ at 494 bp into the 1065 bp gene was a GI of 32,787 bp we have designated *recA* mobile element (RME). Consistent with RME being a mobile genetic element, the GC content is 41.3% compared with the genome average of 47.2%, it encodes mobility functions (see below) and is bordered by 9 bp inverted repeats, designated IR_R (for *recA* end) and IR_i (for integrase end) (Fig. 1). Bioinformatic analysis of the GI identified 23 coding sequences (CDSs) (Fig. 1) including a complete copy of *recA,* designated $recA_{RME}$ at the IR_R end and a phage integrase at the IR_i end. To our knowledge, this is the first mobile genetic element associated with the lateral movement of the critical gene, *recA*.

A number of other genes similar to those known to be involved in DNA processing are also found in RME including *umuDC* encoding the error-prone DNA polymerase V and a gene encoding a protein with a partial domain found in MutL (COG0323; Fig. 1), a component of the MMR pathway (Polosina and Cupples, 2010). A number other genes on RME, homologous to those involved in DNA processes include those encoding a ParB-like nuclease (91% identity to *Vibrio alginolyticus* 12G01; WP_005381205.1), a redox sensitive transcriptional activator with a SoxR-domain (96% identity to *Vibrio* sp. 712i1; WP_017634100.1), a type II restriction enzyme containing a methylase subunit (77% identity to *Vibrio splendidus*; WP_017082665.1) and a helicase (90% identity to *Vibrio brasiliensis* LMG 20546; WP_006880978.1).

Two insertion sequence (ISVvu4) elements were identified at positions 12,877 – 14,083 and 15,897 – 17,103 (striped boxes in Fig. 1) of RME. In both instances, 7 bp direct repeats (DR) were evident bordering the ISVvu4 elements indicating insertion by transposition. The DR for each ISVvu4 element is different, indicating independent



Locus tag	start	stop	orientation	Putative protein identification
RME001	204	1271	+	Recombinase protein RecA
RME002	1373	2593	-	Par-B like nuclease
RME003	2783	3916	+	Hypothetical
RME004	6225	4204	-	MutL mismatch repair domain-containing protein
RME005	6547	9219	+	Type II restriction enzyme, methylase subunit
RME006	9247	11268	+	Hypothetical; DEAD/DEAH box helicase
RME007	11284	12531	+	Hypothetical; YeeC-like nuclease domain-containing protein
RME008	13036	14025	+	Transposase of ISVvu4
RME009	14731	14114	-	Hypothetical
RME010	15192	14734	-	Hypothetical
RME011	15609	15199	-	Hypothetical; DnaJ domain-containing protein
RME012	16056	17045	+	Transposase of ISVvu4
RME013	17645	17211	-	Redox-sensitive transcriptional activator, SoxR
RME014	17963	18844	+	Permease of the drug/metabolite transporter (DMT) superfamily
RME015	21105	19225	-	Hypothetical; chromosome segregation ATPase-containing protein
RME016	22080	21598	-	Hypothetical
RME017	23939	22683	-	Error-prone, lesion bypass DNA polymerase V (UmuC)
RME018	24199	23939	-	Error-prone, lesion bypass DNA polymerase V (UmuD)
RME019	24984	25739	+	Multi domain XerS site-specific tyrosine recombinase XerS
RME020	27461	26001	-	Hypothetical
RME021	28181	27465	-	Hypothetical
RME022	30198	28165	-	Phage integrase family domain protein
RME023	32131	30185	-	Hypothetical

Fig. 1. Genetic structure and gene content of the *recA* genomic island. The RME contains 9 bp inverted repeats at each end (IR_R and IR_i) and 23 ORFs inclusive of the transposase genes from the ISVuv4 elements (striped boxes). The ISVuv4 elements are abutted by 7 bp direct repeats (DR) indicating insertion by transposition. RME contains multiple genes in DNA repair including a full copy of *recA* (RME001), the *umuDC* operon (RME017 and RME018) encoding the two subunits of DNA polymerase V and a gene encoding a protein with a MutL mismatch repair domain (RME004).

insertion events. *In silico* removal of the ISVvu4 elements from the sequence did not restore any CDSs indicating that their insertion had not led to gene disruption. As expected, the promoter regions of both $recA_{RME}$ and the $umuDC_{RME}$ operon have the characteristic LexA binding sequence of CTGT-$(AT)_4$-ACAG indicating control by the SOS response (Wertman and Mount, 1985; Sanchez-Alberola *et al.*, 2012). Present on RME are also genes putatively involved in mobilization/integration such as a phage integrase (RME022) and a site-specific recombinase XerS (RME019) (Fig. 1).

Phylogenetic analysis of *recA* sequences from the *Vibrionaceae* determined that $recA_{S24}$ is characteristic of *recA* genes found within the *V. cholerae* clade, whereas $recA_{RME}$ is not. It does, however, group with other more distantly related *recA* genes found in other members of the *Vibrio* genus (Fig. 2). This is consistent with $recA_{RME}$ having been acquired by LGT. *recA* is an excellent phylogenetic marker for resolving relationships within the *Vibrionaceae* family (Stine *et al.*, 2000; Thompson *et al.*, 2004). Although the acquisition of a divergent *recA* in *V. cholerae* S24 is easily evident, this data reminds us that LGT of critical housekeeping genes like *recA* can and does occur. Less evident would be LGT of *recA* between closely related strains within the *V. cholerae* species confounding phylogenetic trees

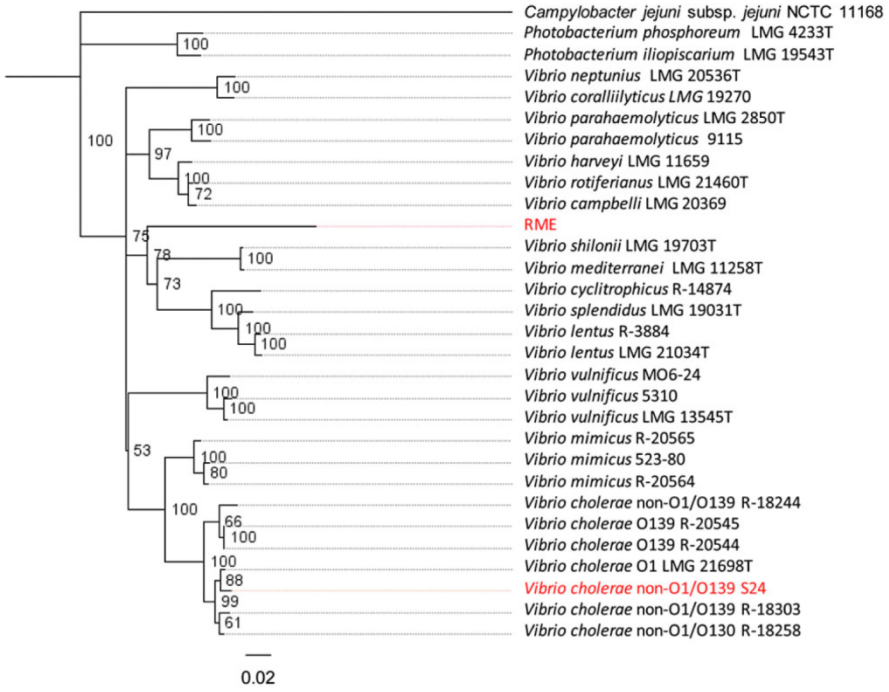


Fig. 2. Phylogenetic analysis of *recA*$_{S24}$ and *recA*$_{RME}$ (RME highlighted in red) *recA*$_{S24}$ (also highlighted in red) groups with *V. cholerae* strains whereas, *recA*$_{RME}$ groups with *recA* from other *Vibrio* species indicating that *recA*$_{RME}$ was mobilized from another member of the *Vibrio* genus.

using a single marker (Bapteste *et al.*, 2004; Creevey *et al.*, 2004).

The recA *mobile element excises as a closed circle and targets a specific site in* recA

Many GIs are known to excise from their location in the chromosome (Boyd *et al.*, 2008). Analysis of RME suggested that it integrated into *recA*$_{S24}$ using site-specific recombination. In site-specific recombination, a DNA recombinase recognizes specific sequences (usually inverted sequences) allowing for DNA breakage and joining reactions that result in integration or excision of the element (Hallet and Sherratt, 1997). Exact excision of RME at IR$_R$ and IR$_i$ was predicted to leave behind a 4 bp scar introducing a frame shift in *recA*$_{S24}$ (Fig. 3A). To determine whether RME excision would leave behind an excision scar, an inverse PCR was conducted using primers reading out from the IR$_R$ and IR$_i$ ends (primers RME-F/RME-R in Table 1). A product of ~ 560 bp was amplified (see gel image in Fig. 3B) and sequenced. Analysis of the sequence showed that excision occurred in one of two possible ways (Fig. 3B): (1) Excision occurred at 2 bp on either side of the IR$_R$ and IR$_i$ ends (black arrows in Fig. 3A) and/or (2) precise excision occurred at the end of IR$_i$ and at 4 bp before the end of IR$_R$ (grey arrows in Fig. 3A). Either way, excision was predicted to restore an uninterrupted and therefore functional copy of *recA*$_{S24}$ in the chromosome, consistent with site-specific recombination. This was confirmed by amplification of an intact 'empty' insertion site using primers S24-cinA-F/S24-recX-R and excising the predicted ~ 1.6 kb fragment (marked with a diamond in Fig. 3C). A nested PCR was then performed on the purified excised fragment using primers EcoRI-recA-F/EcoRI-recA-R (see gel image in Fig. 3C) and the product sequenced (Fig. 3C).

To determine whether the RME was capable of translocating from the genome of *V. cholerae* S24 into a new location, a vector (pOriVn$_{700}$-*recA*$_{S22}$; see Fig. 4A) containing the *recA* gene from a closely related strain of *V. cholerae* S24, strain S22 (Islam *et al.*, 2013), was introduced into *V. cholerae* S24 by conjugation. Here, the RME is expected to excise from the genome of *V. cholerae* S24 and insert into *recA*$_{S22}$ present on pOriVn$_{700}$-*recA*$_{S22}$. A control vector substituting *recA*$_{S22}$ with *gfp* (pOriVn$_{700}$-P$_{lac}$*gfp*; see Fig. 4A) was also introduced into *V. cholerae* S24 by conjugation as a control. Primers (ori6k-R and ori700-F; see Fig. 4A) targeting the vector backbone and the ends of the RME were used in a PCR reaction to determine whether the RME had mobilized into either pOriVn$_{700}$-*recA*$_{S22}$ or pOriVn$_{700}$-P$_{lac}$*gfp*. In

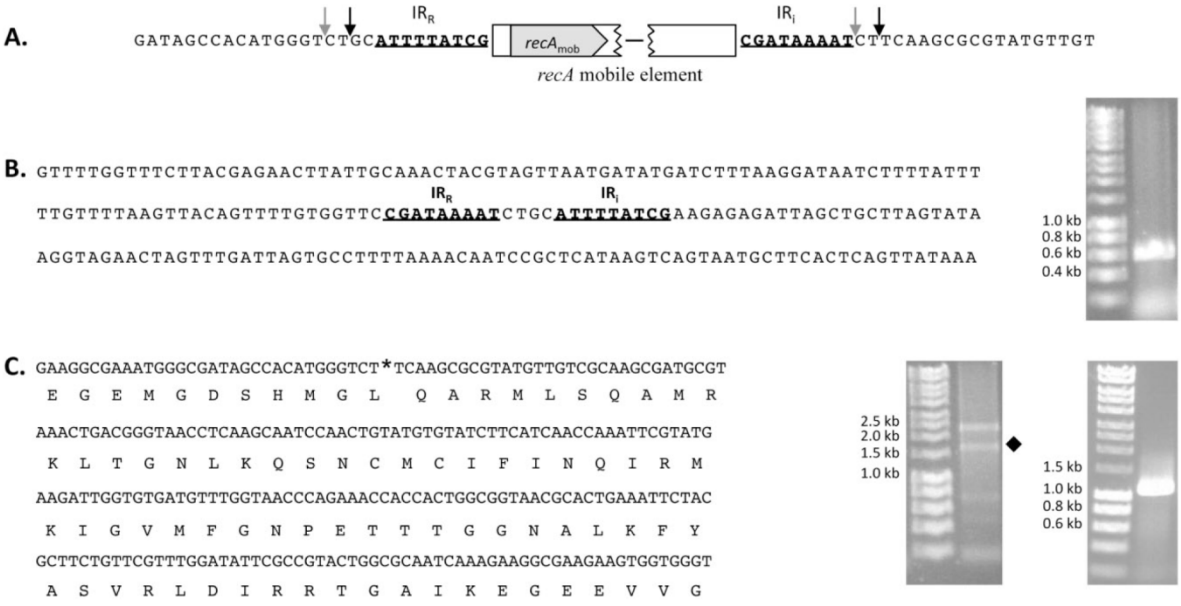


Fig. 3. A. Sequence abutting insertion of the *recA* genomic island in *V. cholerae* S24. The black and grey arrows demarcate the possible excision points for the RME.
B. Sequence and gel image of the product derived from inverse PCR of the excised RME. The sequence shows that excision does not precisely occur at IR_R and IR_i and either occurs by two possible methods shown in Fig. 3A (see text for more details).
C. Sequence of the 'empty' $recA_{S24}$ insertion site and translated peptide sequence shows excision restores an uninterrupted $recA_{S24}$. The asterisk marks the point of RME insertion. Amplification of the 'empty' $recA_{S24}$ site gave a faint product (marked by diamond in left gel image). This was excised, purified and a nested PCR (right gel image) was conducted to generate sufficient product for sequencing.

four independent experiments where pOriVn$_{700}$-P$_{lac}$*gfp* was successfully introduced into *V. cholerae* S24 by conjugation, a product was never detected in the transconjugates (see representative gel in Fig. 4B). However, when pOriVn$_{700}$-$recA_{S22}$ was successfully introduced into *V. cholerae* S24, products were amplified (see representative gel in Fig. 4B) demonstrating insertion of the RME in the equivalent DNA site of $recA_{S22}$ and in both orientations, with respect to $recA_{S22}$. Sequence of the PCR products are shown in Fig. 4C, demonstrating successful insertion of RME into the equivalent $recA_{S24}$ insertion site into $recA_{S22}$ in pOriVn$_{700}$-$recA_{S22}$.

It should be noted that homologous recombination between $recA_{S22}$ in pOriVn$_{700}$-$recA_{S22}$ and $recA_{S24}$ in the *V. cholerae* S24 genome could result in merodiploids that generate the same amplicons as those for RME inserted in the orientation (relative to $recA_{S24}$) found in *V. cholerae* S24 (see Supporting Information Fig. S1 on expected merodiploids). However, in the immediate 2 bp of the IR_R end for three of the transconjugates (Fig. 4C; ia, iiia and iva), there is a G to T substitution and at the immediate 3 bp of the IRi end, two transconjugants (Fig. 4C; iiia and iva) showed a T to G substitution. Since the $recA_{S22}$ sequence is identical to $recA_{S24}$ around the insertion point, homologous recombination should result in identical sequences immediately surrounding the RME. Furthermore, the RME was also found in both orientations, with respect to $recA_{S22}$ (Fig. 4C). Consequently, homologous recombination is unable to explain these results.

These data show that RME is capable of mobilization and preferentially targets a specific site within *recA*. By carrying its own functional copy of *recA,* the GI does not affect any of the vital cell pathways associated with disruption of this gene during integration. Furthermore, specific targeting of *recA* may be necessary to ensure successful maintenance and dissemination of the GI. Since RecA does not function as a monomer but polymerizes to form a filament structure (Yu *et al.*, 2004), disruption of the indigenous *recA* prevents a situation where two divergent RecA proteins might negatively interact resulting in reduced cell fitness.

The recA *mobile element provides* E. coli *protection from UV irradiation*

The presence of multiple genes involved in DNA repair prompted us to look at whether the GI could protect against a common DNA-damaging process faced by *V. cholerae* – UV irradiation. To investigate if $recA_{RME}$ has a role in protecting the cell from DNA damage, the RME was cloned into a fosmid and used to transform *recA*$^-$ *E. coli* strain EPI300. The resultant transformant was

Table 1. Primers used in this study.

Primer	Sequence (5′-3′)	Target	Source
RME-R	GACGAGTCCAGCTCATGACA	integrase end of *recA* genomic island	This study
RME-F	GCTGCTAACGCTTTCTGCTT	*recA* end of *recA* genomic island	This study
S24-ctg675-F	CGGTTAGGAGGGGCTTTTAG	3′ end of contig 675	This study
S24-ctg708-R	TATCGGCTGTGGTTGTTTGA	5′ end of contig 675	This study
S24-ctg367-F	TAGCTAGAGCATTTGTCATAAGAAAAAGTAAG	3′ end of contog 675	This study
S24-ctg367-R	ACTGGCAGCAGAAGAAGCAT	5′ end contig 708	This study
S24-cinA-F	CAAGGTTGGCTCAAAGTG	*cinA* in *V. cholerae* S24	This study
S24-recX-R	GGCATCACTCAAATACCCTA	*recX* in *V. cholerae* S24	This study
S24-recA-F	CTGGAAATTTGTGATGCATT	*recA* in *V. cholerae* S24	This study
EcoRI-recA-F[a]	TTTT**GAATTC**TGGACGAGAATAAACAGAAGG	*recA* in *V. cholerae* S22 & S24	This study
EcoRI-recA-R[a]	TTTT**GAATTC**AAACTCTTCTGGCACCGC	*recA* in *V. cholerae* S22 & S24	This study
EcoRI-Ori700-R[a]	TTTT**GAATTC**CGCGCTATCGCTTGTCG	ori_{pB1067} of pOriVn$_{700}$	This study
EcoRI-OriR6K-F[a]	TTTT**GAATTC**GTGTTCCTGTGTCACTCAAAATTG	*ori6k*	This study
Ori700-F	CCCTATTCCTCTTTAGTCCTGC	ori_{pB1067} of pOriVn$_{700}$	This study
Ori6K-R	TAACGCACTGAGAAGCCC	*ori6k*	This study
S24-phage-Int-F	GCCAAGATATGGCAGGAAAA	Integrase in *recA* genomic island	This study
S24-phage-Int-R	GGACGCTACCCAGTGAATGT	Integrase in *recA* genomic island	This study
recA-F	TGGACGAGAATAAACAGAAGGC	*recA*	(Boucher *et al.*, 2011)
recA-R	CCGTTATAGCTGTACCAAGCGCCC	*recA*	(Boucher *et al.*, 2011)
pCC2FOS-FP	GTACAACGACACCTAGAC	pCC2FOS sequencing primers (F)	Epicentre Biotechnologies
pCC2FOS-RP	CAGGAAACAGCCTAGGAA	pCC2FOS sequencing primers (R)	Epicentre Biotechnologies
recA-Tn5-F	CGCTCATAAGTCAGTAATGCTTCA	*recA* on genomic island. Used to screen for Tn*5* insertion.	This study
umuC-Tn5-F	GATGTATGGCTGAATCGACCA	*umuC* on genomic island. Used to screen for Tn*5* insertion.	This study
KAN-2 FP-1	ACCTACAACAAAGCTCTCATCAACC	Forward primer inside Tn5 used to screen for Tn5 insertion.	Epicentre Biotechnologies
KAN-2 RP-1	GCAATGTAACATCAGAGATTTTGAG	Reverse primer inside Tn5 used to screen for Tn5 insertion.	Epicentre Biotechnologies

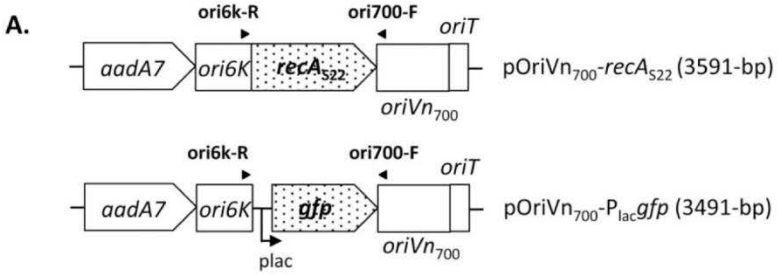
a. Bold and underlined sequence shows the *Eco*RI restriction site.

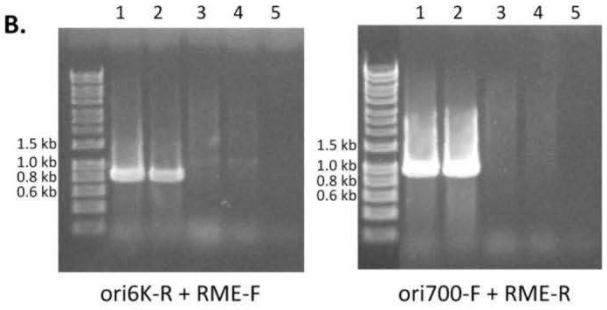
subjected to UV-C irradiation. Fig. 5A shows that the presence of the RME element conferred enhanced bacterial cell survival when exposed to 0.8 mJ cm^{-2} of UV-C. From Fig. 5A and B it can be seen that EPI300 and EPI300 transformed by vector only controls are completely killed by exposure to 20 s 0.8 mJ cm^{-2} of UV-C. However, EPI300 transformants containing the RME survive for up to 60 s of UV-C exposure and show a 100-fold increase in survival at 10 s and up to 10,000,000-fold higher survival after 20 s UV-C exposure. Fig. 5A shows that when $recA_{RME}$ is insertionally inactivated, the level of cell survivability decreases to a level comparable with the vector-only control (Fig. 5B). This demonstrates that $recA_{RME}$ is functional and is the gene mainly responsible for the protection provided by the presence of the RME. An interesting future question would be whether $recA_{RME}$ is more efficient in DNA repair than the host *recA* (i.e. $recA_{S24}$). There is precedent for such an idea, in a strain of *Clostridium difficile,* a 4.2 kb insert disrupts a gene encoding a thymidylate synthetase (involved in DNA synthesis and repair) but contains a more functionally active version of the disrupted gene (Knetsch *et al.*, 2011).

The recA *mobile element provides* E. coli *with increased protection against antibiotics*

Since RME has multiple genes involved in DNA repair, we tested whether RME provided enhanced protection against three DNA-targeting antibiotics: nalidixic acid, ciprofloxacin and bleomycin. Minimum inhibitory concentrations (MICs) were determined using nalidixic acid, ciprofloxacin and bleomycin (Table 2). Interestingly, the MIC of the first-generation quinolone, nalidixic acid, did

Fig. 4. Translocation of the *recA* genomic island from the genome of *V. cholerae* S24 into a replicating vector containing $recA_{S22}$.
A. Genetic structure of the replicating vectors pOriVn$_{700}$-$recA_{S22}$ and pOriVn$_{700}$-P$_{lac}$*gfp* and the placement of primers ori6k-R and ori700-F used for amplifying the boundaries of the translocated genomic island are shown.
B. Representative gel showing amplification using vector specific and RME specific primers from colonies derived from conjugation of pOriVn$_{700}$-$recA_{S22}$ (lanes 1 and 2) and pOriVn$_{700}$-P$_{lac}$*gfp* (lanes 3 and 4) into *V. cholerae* S24. Lane 5 shows negative dH$_2$O control.
C. Sequence of products derived using vector-specific and RME-specific primers from PCR of *V. cholerae* S24 transconjugates from four independent conjugations. Each transconjugant is denoted by i, ii, iii and iv. In most instances (iib, iiib and ivb), the same transconjugate showed insertion of RME in both orientations relative to $recA_{S22}$. The sequences indicate specific insertion of RME into the same site of $recA_{S22}$ (the equivalent insertion site in $recA_{S24}$).





Lane	PCR template
1	Colony from conjugation 1 of pOriVn$_{700}$-*recA*$_{S22}$ into *V. cholerae* S24
2	Colony from conjugation 1 of pOriVn$_{700}$-*recA*$_{S22}$ into *V. cholerae* S24
3	Colony from conjugation 1 of pOriVn$_{700}$-P$_{lac}$*gfp* into *V. cholerae* S24
4	Colony from conjugation 1 of pOriVn$_{700}$-P$_{lac}$*gfp* into *V. cholerae* S24
5	dH$_2$0 negative control

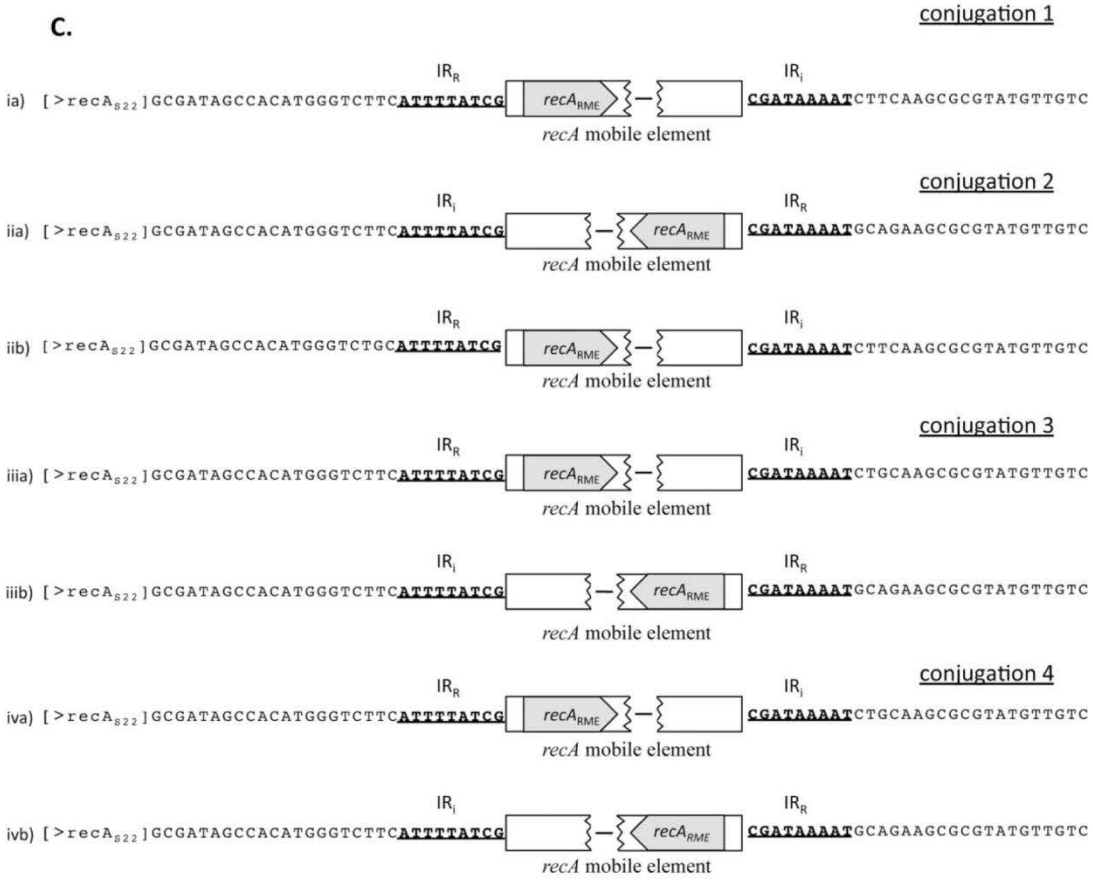


Table 2. Minimal inhibitory concentration (MICs[a]).

Strain	Ciprofloxacin	Bleomycin
EPI300	0.015625	8
EPI300/pCC2FOS	0.015625	8
EPI300/pCC2FOS::RME	0.0625	16
EPI300/pCC2FOS::RMEΔ$umuC_{RME}$	0.0625	8
EPI300/pCC2FOS::RMEΔ $recA_{RME}$	0.015625	8

a. MIC given as μg ml^{-1}.

not vary between any of the tested strains, including the RME (data not shown). However, when the strains given in Table 2 were tested using ciprofloxacin, a second-generation quinolone, a fourfold increase in MIC was observed for strains containing the RME. In the case of ciprofloxacin, it is apparent that $recA_{RME}$ is responsible for the increased resistance. When $recA_{RME}$ is insertionaslly inactivated from the genomic island, the MIC drops to a level equivalent to that seen for the *E. coli* strain EPI300. The importance of *recA* in protection against ciprofloxacin and other antibiotics has previously been documented [e.g. in *Acinetobacter baumannii* (Aranda *et al.*, 2011)]. RME also provided protection from bleomycin in *E. coli* EPI300. However, in contrast to ciprofloxacin, when $recA_{RME}$ and *umuC* from the RME (designated $umuC_{RME}$) are both inactivated independently, the MIC against bleomycin is the same as *E. coli* EPI300, indicating that $RecA_{RME}$ activation of the DNA polymerase V subunit UmuD to UmuD' encoded on the GI is responsible for protections against bleomycin (Patel *et al.*, 2010). Since $umuDC_{RME}$ was able to provide protection from bleomycin in a genetic background that already contains *umuDC*, it is hypothesized that in the *V. cholerae* species, where *umuDC* is only sporadically found, this element may provide increased protection from DNA damage compared with what was observed here in *E. coli* EPI300.

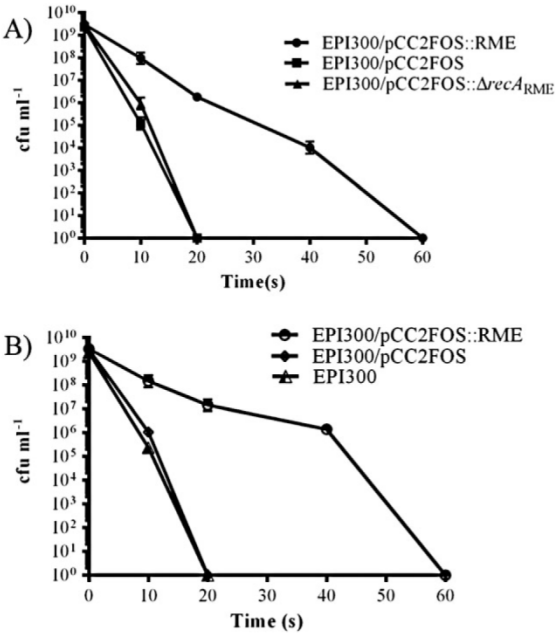


Fig. 5. Survival of *E. coli* carrying the *recA* genomic island and control strains when exposed to UV-C stress. Time points are given at 0, 10, 20, 40 and 60 s. UV-C exposure was set to 0.8 mJ cm^{-2}.

Apart from DNA repair, *recA* and DNA polymerase V (encoded by *umuDC*) are known to increase spontaneous mutation frequencies resulting in the emergence of antibiotic resistance mutants (Thi *et al.*, 2011). Spontaneous mutation in *V. cholerae* is well documented to cause resistance to a variety of antibiotics (Goss *et al.*, 1965; Gellert *et al.*, 1977; Sugino *et al.*, 1977; Allen *et al.*, 1979; Kitaoka *et al.*, 2011). Here we chose to examine the mutation frequency of two antibiotics, rifampicin which acts on protein synthesis and nalidixic acid which targets DNA replication by inhibiting the A subunit of DNA gyrase. Mutation frequencies after 24 and 48 h showed no differences between *E. coli* containing RME and the controls on 100 μg ml^{-1} rifampicin (data not shown). This may be because of rifampicin acting on protein synthesis and therefore not inducing the SOS response which induces transcription of *umuDC*. However, when the experiments were repeated with 50 μg ml^{-1} nalidixic acid, *E. coli* EPI300 and the vector-only control consistently did not produce any spontaneous mutants, while *E. coli* EPI300/pCC2FOS::RME showed an increased mutation frequency after both 24 and 48 h incubation in the presence of nalidixic acid (Table 3). *E. coli* EPI300/pCC2FOS::RMEΔ$umuC_{RME}$ generally had the same mutation frequency as the complete RME. However, experiment 3 (Table 3) showed that this strain produced no mutants after 24 and 48 h. It can be concluded from these experiments that *E. coli* EPI300 with the complete RME provides an adaptive advantage by increasing the mutation rate resulting in subsequent resistance to nalidixic acid, but this effect could not be wholly attributed to the activity of $umuC_{RME}$. One possible explanation is the activation of the indigenous *E. coli* UmuD by RecA provided by the GI. However, it cannot be excluded that other genes on the RME are elevating the spontaneous mutation rate. Nevertheless, *umuDC*-like operons are commonly associated with mobile genetic elements (Permina *et al.*, 2002; Tark *et al.*, 2005; Hare *et al.*, 2012) and do provide a general adaptive advantage to hosts that house them (Yeiser *et al.*, 2002; Tark *et al.*, 2005). Although we failed in our attempts to transfer RME into seven non-O1/O139 *V. cholerae* strains from Sydney using chitin

transformation, this element is likely to do the same in a *V. cholerae* genetic background.

To conclude, this study reports a novel GI in *V. cholerae* that contains genes involved in multiple DNA repair pathways, including the critical housekeeping gene *recA* and genes encoding DNA polymerase V which in this study, we show to be functional. The presence of other DNA processing genes may provide *V. cholerae* with alternative DNA repair pathways. Since this element can excise from its chromosomal location, it has the potential to mobilize into other strains, such as cholera toxin-producing O1/O139 pandemic strains. Such mobilization could have implications for increased environmental survival or resistance to certain antibiotics.

Experimental procedures

Bacterial strains, plasmids and growth conditions

All strains and plasmids used are shown in Table 4. *V. cholerae* strain S24 was collected from Georges River in the greater Sydney (Australia) urban area as previously described (Islam *et al.*, 2013). All *E. coli* and *V. cholerae* strains were routinely grown on Luria–Bertani (LB) broth at 37°C under aerobic conditions. For *E. coli* WM3064, diaminopimelic acid (DAP) was added to a final concentration of 0.3 mM. Spectinomycin was used for *E. coli* and *V. cholerae* at 50 μg ml^{-1} and 125 μg ml^{-1} respectively. Chloramphenicol was used at 12.5 μg ml^{-1}.

Whole genome sequencing, PCR, DNA extraction and sequencing methods

DNA was extracted using the Wizard genomic DNA purification kit (Promega). Plasmid and PCR/gel extractions were done using PureYield Plasmid Miniprep and Wizard SV Gel and PCR clean-up systems respectively (Promega). Purified DNA from *V. cholerae* S24 was sequenced at the Wellcome Trust Sanger Institute using Illumina-based technology.

All primers used in this study are shown in Table 1. Standard PCR was performed using the PCR master mix (Promega) containing 25 units ml^{-1} of *Taq* DNA polymerase, 800 μM dNTPs and 1.5 mM $MgCl_2$. Primers were used at a final concentration of 0.5 μM each. All PCRs were performed with 30 cycles of denaturation at 94°C for 30 s, the appropriate annealing temperature for 30 s and an extension of 72°C (1 min kb^{-1}) and sequencing performed at Macrogen. From whole genome sequencing (Wellcome Trust Sanger Institute) it became evident that the host *recA* had been disrupted and was present on two separate contigs. The two contigs (contigs 675 and 708) were pieced together by PCR and joined to an intervening third contig to produce a contig of 262,869 bp (contig 367) using primers described in Table 1. The accession number for RME is KJ123688.

Cloning of RME and transposon mutagenesis of recA *genomic island*

To clone the *recA* genomic island (RME) from *V. cholerae* strain S24, genomic DNA was digested with *Nae*I and a

Table 3. Nalidixic acid[a] mutation frequencies.

Experiment 1		
Strain	Mutation frequency 24 h	Mutation frequency 48 h
EPI300	$< 1.1 \times 10^{-11}$	$< 1.1 \times 10^{-11}$
EPI300/pCC2FOS	$< 1.7 \times 10^{-11}$	$< 1.7 \times 10^{-11}$
EPI300/pCC2FOS::RME	1.4×10^{-8} (4)	1.5×10^{-8} (4)
EPI300/pCC2FOS::RMEΔ*umuC*$_{RME}$	1.6×10^{-9} (3)	2.1×10^{-9} (4)
Experiment 2		
Strain	Mutation frequency 24 h	Mutation frequency 48 h
EPI300	$< 1.1 \times 10^{-11}$	$< 1.1 \times 10^{-11}$
EPI300/pCC2FOS	$< 1.7 \times 10^{-11}$	$< 1.7 \times 10^{-11}$
EPI300/pCC2FOS::RME	8.3×10^{-10} (1)	1.4×10^{-9} (3)
EPI300/pCC2FOS::RMEΔ*umuC*$_{RME}$	2.8×10^{-9} (6)	4.4×10^{-9} (7)
Experiment 3		
Strain	Mutation frequency 24 h	Mutation frequency 48 h
EPI300	$< 1.1 \times 10^{-11}$	$< 1.1 \times 10^{-11}$
EPI300/pCC2FOS	$< 1.3 \times 10^{-11}$	$< 1.3 \times 10^{-11}$
EPI300/pCC2FOS::RME	4.8×10^{-10} (1)	1.7×10^{-9} (4)
EPI300/pCC2FOS::RMEΔ*umuC*$_{RME}$	$< 9.1 \times 10^{-12}$	$< 9.1 \times 10^{-12}$

a. Concentration of nalidixic acid = 50 μg ml^{-1}.
'<' indicates that zero colonies appeared in all 10 replicates (see *Experimental procedures*).
Numbers in brackets indicates the number of replicates in which one or more colonies appeared.

Table 4. List of strains and plasmids.

Strain or plasmid	Relevant genotype[a]	Reference or source
V. cholerae		
S24	Wild-type (non-O1/O139)	This study
S22	Wild-type (non-O1/O139)	(Islam *et al.*, 2013)
E. coli		
DH5αλpir	*endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG* Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *hsdR17 λpir*	(Demarre *et al.*, 2005)
WM3064	Donor strain for conjugation: *thrB1*004 *pro thi rpsL hsdS lacZΔM15* RP4-1360 Δ(*araBAD)567* Δ*dapA*1341::[*erm pir*], Sm^R	(Saltikov and Newman, 2003)
EPI300™-$T1^R$	[F^- mcrA Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80d*lacZ*ΔM15 Δ*lacX*74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ^- *rps*L *nup*G *trf*A *ton*A *dhfr*], Sm^R, Tp^R	Epicentre Biotechnologies
Plasmids/fosmids		
pCC2FOS	Cloning vector, Cm^R	
pCC2FOS-RME	pCC2FOS vector containing 32 kb insert from *V. cholerae* S24. The insert contains the *recA* GI and surrounding sequence, Cm^R	This study
pCC2FOS::RMEΔ$recA_{RME}$	pCC2FOS vector containing 32 kb insert from *V. cholerae* S24. The insert contains the *recA* GI and surrounding sequence and has *recA* on the GI insertionally inactivated by Tn5, Km^R, Cm^R	This study
pCC2FOS::RMEΔ$umuC_{RME}$	pCC2FOS vector containing 32 kb insert from *V. cholerae* S24. The insert contains the *recA* GI and surrounding sequence and has *umuC* present on the GI insertionally inactivated by Tn5, Km^R, Cm^R	This study
$pOriVn_{700}$	Low copy mobilizable vector containing ori_{pB1067} (vibrio specific) and *ori6K*, Sp^R	(Le Roux *et al.*, 2011)
$pOriVn_{700}$-$recA_{S22}$	$pOriVn_{700}$ with *recA* from *V. cholerae* S22 in between ori_{pB1067} and *ori6K.* The *recA* gene is reading toward ori_{pB1067}, Sp^R	This study
$pOriVn_{700}$-P_{lac}*gfp*	$pOriVn_{700}$ with P_{lac}*gfp* cloned in between ori_{pB1067} and *ori6K,* Sp^R	(Le Roux *et al.*, 2011)

a. Tc^R, tetracycline resistance; Sm^R, streptomycin resistance; Sp^R, spectinomycin resistance; Cm^R, chloramphenicol resistance, Km^R, kanamycin resistance.

library constructed using the CopyControl Fosmid Library Production Kit (Epicentre). *Nae*I digestion of *V. cholerae* strain S24 genomic DNA creates a fragment of 38, 913 bp containing the entire 32, 787 bp RME. The library was screened for a fosmid clone containing the 38, 913-bp *Nae*I fragment using primers targeting the phage integrase in the RME (Table 1). A positive clone designated pCC2FOS-RME was confirmed by sequencing the ends of the cloned insert using the pCC2FOS vector primers FP and RP (Table 1). To create the pCC2FOS no insert control, linearized and dephosphorylated pCC2FOS (Epicentre) was treated with T4 polynucleotide kinase and circularized by ligation. A mutant library of pCC2FOS-RME was constructed using the EZ-Tn5 Kan-2 Insertion Kit (Epicentre Biotechnologies) according to manufacturer instructions. Mutants containing knockouts of individual genes present on the genomic island were screened by PCR using primers reading out from EZ-Tn5 Kan-2 and a primer targeting the gene of interest (Table 1).

Phylogenetic analysis

Phylogenetic analysis of $recA_{S24}$ and $recA_{RME}$ was done using bioinformatics program Geneious version 6.1.6 and FigTree version 1.4.0. Phylogenetic tree parameters were taken from (Thompson *et al.*, 2004). Distance estimations were obtained using the Jukes and Cantor model and tree built using the neighbour-joining method. Bootstrap percentages were calculated after 100 simulations. The *Campylobacter jejuni* subsp. *jejuni* NCTC 11168 *recA* sequence was used as an outgroup.

recA *targeting experiments*

Vector $pOriVn_{700}$ and *recA* from a strain of *V. cholerae* S22 that is closely related to *V. cholerae* S24 were amplified using primer pairs EcoRI-Ori700-R/EcoRI-Ori6K-F and EcoRI-recA-F/EcoRI-recA-R respectively (Table 1). Since the primers contained engineered *Eco*RI sites, the resulting amplicons of $recA_{S22}$ and $pOriVn_{700}$ were purified, digested with *Eco*R1 and then ligated together using T4 DNA ligase (Fermentas). The ligation mix was then transformed into *E. coli* DH5αλpir to produce $pOriVn_{700}$-$recA_{S22}$. The construct was then extracted and transformed into the conjugation donor strain *E. coli* WM3064.

Conjugations using $pOriVn_{700}$-$recA_{S22}$ and $pOriVn_{700}$-P_{lac}*gfp* were performed by combining equal volumes of overnight cultures in LB from both donor and recipient strains. These were then centrifuged at 3000 × *g* and re-suspended in 50 μL of LB and spotted onto a 0.2 μM filter (Millipore) that had been placed on an LB agar plate containing 0.3 mM DAP. Donor and recipient cells were left to incubate for 4 h at 37°C and cells were then removed from the filter by vortexing. The re-suspended cells were then plated on LB + 125 μg ml^{-1} spectinomycin and incubated at 37°C overnight. One colony per mating was picked and appropriate junction PCR was conducted using primers in plasmid backbone (Table 1; Ori700-F/Ori6K-R) and primers reading out from RME (Table 1; RME-F/RME-R).

UV stress experiments

UV stress experiments were adapted from Lin and Wang (2001). Strains were grown for 16–20 h at 37°C with shaking

at 230 r.p.m. in 5 ml LB broth supplemented with appropriate antibiotic. Cells were centrifuged at 4000 × *g*, corrected for differences in optical density (OD) at 600 nm and re-suspended in equal volumes of M9 salts (Sambrook *et al*., 1989) supplemented with $MgSO_4.7H_2O$ to a final concentration of 0.002 M. The entire cell suspension was placed in a clear bottom 10 cm plastic petri dish and subjected to 0.8 mJ cm^{-2} UV-C for 0, 10, 20, 40 and 60 s using an Amersham Life Science Ultraviolet Crosslinker. After each time interval, 150 ul aliquot was removed and placed in a 1.5 ml Eppendorf tube in the dark. The remaining liquid culture was thoroughly re-suspended using a pipette to avoid clumping of cells. After the final UV-C exposure time point, cells were diluted in M9 salts+ $MgSO_4.7H_2O$ to 10^{-6} and enumerated by the drop plate method on LB agar. Plates were incubated in the dark to prevent photoreactivation at 37°C overnight and colony-forming units (CFUs) were calculated the following day.

Minimum inhibitory concentration experiments and antibiotic mutation frequency experiments

MICs of nalidixic acid, ciprofloxacin and bleomycin were determined by broth microdilution using standard methods (Clinical and Laboratory Standards Institute, 2003) except that LB broth was used as the growth medium instead of Mueller Hinton. Each MIC was performed in triplicate. The mutation frequency experiment was designed using the guidelines described in (Pope *et al*., 2008). Specifically, mutation frequencies were determined using LB supplemented with 50 μg ml^{-1} nalidixic acid and 100 μg ml^{-1} rifampicin. Ten replicate overnight cultures for each strain were grown in 5 ml LB (chloramphenicol was added for those strains carrying pCC2FOS and derivatives). Each overnight culture was then diluted to ~ 10^4 CFU ml^{-1} with fresh LB5 (no chloramphenicol added) and 5 ml for each replicate was transferred into a 15 ml tube and incubated for 16–20 h at 37°C with shaking at 230 r.p.m. The following day, 200 μl from each tube was spread plated onto LB5 agar supplemented with the appropriate antibiotic (rifampicin or nalidixic acid) and incubated for 24 h and then 48 h at 37°C when colonies were counted. This was repeated in triplicate. In order to calculate total colony counts, cells were enumerated on LB5 agar with no antibiotic. Note that these experiments were performed in a Class II Biosafety Hood to avoid any contamination. Mutation frequencies were calculated as number of antibiotic-resistant CFUs/total number of CFUs after 24 and 48 h.

Acknowledgements

We would like to acknowledge Associate Professor Aaron Darling for his help with the phylogenetic analysis. This work was supported by the ithree institute, UTS. Rita Rapa is a recipient of a UTS postgraduate scholarship. The sequencing of *V. cholerae* S24 was funded under Wellcome Trust grant 098051.

References

Allen, J.G., Atherton, F.R., Hall, M.J., Hassall, C.H., Holmes, S.W., Lambert, R.W., *et al*. (1979) Phosphonopeptides as antibacterial agents: alaphosphin and related phosphonopeptides. *Antimicrob Agents Chemother* **15:** 684–695.

Aranda, J., Bardina, C., Beceiro, A., Rumbo, S., Cabral, M.P., Barbe, J., and Bou, G. (2011) *Acinetobacter baumannii* RecA protein in repair of DNA damage, antimicrobial resistance, general stress response, and virulence. *J Bacteriol* **193:** 3740–3747.

Bapteste, E., Boucher, Y., Leigh, J., and Doolittle, W.F. (2004) Phylogenetic reconstruction and lateral gene transfer. *Trends Microbiol* **12:** 406–411.

Boucher, Y., Cordero, O.X., Takemura, A., Hunt, D.E., Schliep, K., Bapteste, E., *et al*. (2011) Local mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio cholerae* populations. *MBio* **2:** e00335-10.

Boyd, E.F., Almagro-Moreno, S., and Parent, M.A. (2008) Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends Microbiol* **17:** 47–53.

Clinical and Laboratory Standards Institute. (2003) *Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically*.

Creevey, C.J., Fitzpatrick, D.A., Philip, G.K., Kinsella, R.J., O'Connell, M.J., Pentony, M.M., *et al*. (2004) Does a tree-like phylogeny only exist at the tips in the prokaryotes? *Proc Biol Sci* **271:** 2551–2558.

Demarre, G., Guérout, A., Matsumoto-Mashimo, C., Rowe-Magnus, D.A., Marliére, P., and Mazel, D. (2005) A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPα) conjugative machineries and their cognate *Escherichia coli* host strains. *Res Microbiol* **156:** 245–255.

Everiss, K.D., Hughes, K.J., and Peterson, K.M. (1994) The accessory colonization factor and toxin-coregulated pilus gene clusters are physically linked on the *Vibrio cholerae* 0395 chromosome. *DNA Seq* **5:** 51–55.

Faruque, S.M., and Mekalanos, J.J. (2003) Pathogenicity islands and phages in *Vibrio cholerae* evolution. *Trends Microbiol* **11:** 505–510.

Galen, J.E., Ketley, J.M., Fasano, A., Richardson, S.H., Wasserman, S.S., and Kaper, J.B. (1992) Role of *Vibrio cholerae* neuraminidase in the function of cholera toxin. *Infect Immun* **60:** 406–415.

Gellert, M., Mizuuchi, K., O'Dea, M.H., Itoh, T., and Tomizawa, J.I. (1977) Nalidixic acid resistance: a second genetic character involved in DNA gyrase activity. *Proc Natl Acad Sci USA* **74:** 4772–4776.

Goodman, M.F. (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* **71:** 17–50.

Goss, W.A., Deitz, W.H., and Cook, T.M. (1965) Mechanism of action of nalidixic acid on *Escherichia coli*. Inhibition of deoxyribonucleic acid synthesis. *J Bacteriol* **89:** 1068–1074.

Grim, C.J., Hasan, N.A., Taviani, E., Haley, B., Chun, J., Brettin, T.S., *et al*. (2010) Genome sequence of hybrid *Vibrio cholerae* O1 MJ-1236, B-33, and CIRS101 and comparative genomics with *V. cholerae*. *J Bacteriol* **192:** 3524–3533.

Hallet, B., and Sherratt, D.J. (1997) Transposition and site-specific recombination: adapting DNA cut-and-paste

mechanisms to a variety of genetic rearrangements. *FEMS Microbiol Rev* **21:** 157–178.

Hare, J.M., Bradley, J.A., Lin, C., and Elam, T.J. (2012) Diverse responses to UV light exposure in *Acinetobacter* include the capacity for DNA damage-induced mutagenesis in the opportunistic pathogens *Acinetobacter baumannii* and *Acinetobacter ursingii*. *Microbiology* **158:** 601–611.

Islam, A., Labbate, M., Djordjevic, S.P., Alam, M., Darling, A., Melvold, J., *et al.* (2013) Indigenous *Vibrio cholerae* strains from a non-endemic region are pathogenic. *Open Biol* **3:** 120181.

Janion, C. (2008) Inducible SOS response system of DNA repair and mutagenesis in *Escherichia coli*. *Int J Biol Sci* **4:** 338–344.

Kaper, J., Morris, J., Jr, and Levine, M. (1995) Cholera. *Clin Microbiol Rev* **8:** 48–86.

Keymer, D.P., and Boehm, A.B. (2011) Recombination shapes the structure of an environmental *Vibrio cholerae* population. *Appl Environ Microbiol* **77:** 537–544.

Kitaoka, M., Miyata, S.T., Unterweger, D., and Pukatzki, S. (2011) Antibiotic resistance mechanisms of *Vibrio cholerae*. *J Med Microbiol* **60:** 397–407.

Knetsch, C.W., Hensgens, M.P., Harmanus, C., van der Bijl, M.W., Savelkoul, P.H., Kuijper, E.J., *et al.* (2011) Genetic markers for *Clostridium difficile* lineages linked to hypervirulence. *Microbiology* **157:** 3113–3123.

Le Roux, F., Davis, B.M., and Waldor, M.K. (2011) Conserved small RNAs govern replication and incompatability of a diverse new plasmid family from marine bacteria. *Nucleic Acids Res* **39:** 1004–1013.

Lenhart, J.S., Schroeder, J.W., Walsh, B.W., and Simmons, L.A. (2012) DNA repair and genome maintenance in *Bacillus subtilis*. *Microbiol Mol Biol Rev* **76:** 530–564.

Lin, K., and Wang, A. (2001) UV mutagenesis in *Escherichia coli* K-12: cell survival and mutation frequency of the chromosomal genes *lacZ*, *rpoB*, *ompF*, and *ampA*. *J Exp Microbiol Immunol* **1:** 32–46.

McKenzie, G.J., Harris, R.S., Lee, P.L., and Rosenberg, S.M. (2000) The SOS response regulates adaptive mutation. *Proc Natl Acad Sci USA* **97:** 6646–6651.

Meibom, K.L., Blokesch, M., Dolganov, N.A., Wu, C.-Y., and Schoolnik, G.K. (2005) Chitin induces natural competence in *Vibrio cholerae*. *Science* **310:** 1824–1827.

Murphy, R.A., and Boyd, E.F. (2008) Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates. *J Bacteriol* **190:** 636–647.

Patel, M., Jiang, Q., Woodgate, R., Cox, M.M., and Goodman, M.F. (2010) A new model for SOS-induced mutagenesis: how RecA protein activates DNA polymerase V. *Crit Rev Biochem Mol Biol* **45:** 171–184.

Permina, E.A., Mironov, A.A., and Gelfand, M.S. (2002) Damage-repair error-prone polymerases of eubacteria: association with mobile genome elements. *Gene* **293:** 133–140.

Piddock, L.J., and Wise, R. (1987) Induction of the SOS response in *Escherichia coli* by 4-quinolone antimicrobial agents. *FEMS Microbiol Lett* **41:** 289–294.

Polosina, Y.Y., and Cupples, C.G. (2010) Wot the 'L-Does MutL do? *Mutat Res* **705:** 228–238.

Pope, C.F., O'Sullivan, D.M., McHugh, T.D., and Gillespie, S.H. (2008) A practical guide to measuring mutation rates in antibiotic resistance. *Antimicrob Agents Chemother* **52:** 1209–1214.

Ramamurthy, T., Yamasaki, S., Takeda, Y., and Nair, G.B. (2003) *Vibrio cholerae* O139 Bengal: odyssey of a fortuitous variant. *Microbes Infect* **5:** 329–344.

Rattray, A.J., and Strathern, J.N. (2003) Error-prone DNA polymerases: when making a mistake is the only way to get ahead. *Annu Rev Genet* **37:** 31–66.

Safa, A., Bhuiya, A.N., Murphy, D., Bates, J., Nusrin, S., Kong, R.Y.C., *et al.* (2009) Multilocus genetic analysis reveals that the Australian strains of *Vibrio cholerae* O1 are similar to the pre-seventh pandemic strains of the El Tor biotype. *J Med Microbiol* **58:** 105–111.

Saltikov, C.W., and Newman, D.K. (2003) Genetic identification of a respiratory arsenate reductase. *Proc Natl Acad Sci USA* **100:** 10983–10988.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular cloning: A laboratory manual*.

Sanchez-Alberola, N., Campoy, S., Barbé, J., and Erill, I. (2012) Analysis of the SOS response of *Vibrio* and other bacteria with multiple chromosomes. *BMC Genomics* **13:** 58.

Stine, O.C., Sozhamannan, S., Gou, Q., Zheng, S., Morris, J.G., Jr, and Johnson, J.A. (2000) Phylogeny of *Vibrio cholerae* based on *recA* sequence. *Infect Immun* **68:** 7180–7185.

Sugino, A., Peebles, C.L., Kreuzer, K.N., and Cozzarelli, N.R. (1977) Mechanism of action of nalidixic acid: purification of *Escherichia coli nalA* gene product and its relationship to DNA gyrase and a novel nicking-closing enzyme. *Proc Natl Acad Sci USA* **74:** 4767–4771.

Tark, M., Tover, A., Tarassova, K., Tegova, R., Kivi, G., Hõrak, R., and Kivisaar, M. (2005) A DNA polymerase V homologue encoded by TOL plasmid pWW0 confers evolutionary fitness on *Pseudomonas putida* under conditions of environmental stress. *J Bacteriol* **187:** 5203–5213.

Thi, T.D., Lopez, E., Rodriguez-Rojas, A., Rodriguez-Beltran, J., Couce, A., Guelfo, J.R., *et al.* (2011) Effect of *recA* inactivation on mutagenesis of *Escherichia coli* exposed to sublethal concentrations of antimicrobials. *J Antimicrob Chemother* **66:** 531–538.

Thompson, C.C., Thompson, F.L., Vandemeulebroecke, K., Hoste, B., Dawyndt, P., and Swings, J. (2004) Use of recA as an alternative phylogenetic marker in the family *Vibrionaceae*. *Int J Syst Evol Microbiol* **54:** 919–924.

Wertman, K.F., and Mount, D.W. (1985) Nucleotide sequence binding specificity of the LexA repressor of *Escherichia coli* K-12. *J Bacteriol* **163:** 376–384.

Yeiser, B., Pepper, E.D., Goodman, M.F., and Finkel, S.E. (2002) SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. *Proc Natl Acad Sci USA* **99:** 8737–8741.

Ysern, P., Clerch, B., Castaňo, M., Gilbert, I., Barbé, J., and Llagostera, M. (1990) Induction of SOS genes in *Escherichia coli* and mutagenesis in *Salmonella typhimurium* by fluoroquinolones. *Mutagenesis* **5:** 63–66.

Yu, X., VanLoock, M.S., Yang, S., Reese, J.T., and Egelman, E.H. (2004) What is the structure of the RecA-DNA filament? *Curr Protein Pept Sci* **5:** 73–79.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Possible insertion and homologous recombination events of the *recA* genomic island with pOriVn$_{700}$-*recA*$_{S22}$. Production of merodiploids because of homologous recombination between *recA*$_{S22}$ and *recA*$_{S24}$ could result in two genetic structures (crossover 1 and crossover 2). In both instances, PCR products could be generated using the vector-specific and RME-specific primers used to detect insertion of the RME into *recA*$_{S22}$ (broken lines) in the orientation found in the *V. cholerae* S24 genome (insertion 1). Insertion of the RME in the inverse orientation (insertion 2) would generate products using inverse primer pairs. The inverse insertion cannot be explained by homologous recombination and indicates an integration event.

Appendix S1. Genbank file of the *recA* mobile element (RME); Accession Number KJ123688.

Appendix 4

```
LOCUS       RecA_mobile_element_Copy32783 bp   DNA     linear       20-DEC-2013
DEFINITION  Vibrio cholerae strain S24.
ACCESSION
VERSION
KEYWORDS    .
SOURCE      Vibrio cholerae
  ORGANISM  Vibrio cholerae
            Unclassified.
REFERENCE   1  (bases 1 to 32783)
  AUTHORS   Rapa,R.A.
  TITLE     An integrated mobile genetic element in Vibrio cholerae provides
            multi-pathway protection from DNA damage
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 32783)
  AUTHORS   Rapa,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-2013) The ithree institute, University of
            Technology, Sydney, Sydney, NSW 2007, Australia
COMMENT     ##Assembly-Data-START##
            Assembly Method       :: NA v. NA
            Sequencing Technology :: Sanger dideoxy sequencing; Illumina
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..32783
                     /organism="Vibrio cholerae"
                     /mol_type="genomic DNA"
                     /strain="S24"
     mobile_element  1..32783
                     /mobile_element_type="other:Genomic Island: RecA mobile
                     element"
     repeat_region   1..9
                     /note="Inverted repeart for RecA mobile element"
                     /rpt_type=inverted
     protein_bind    98..113
                     /bound_moiety="Site for binding of LexA"
                     /function="LexA box"
     gene            204..1271
                     /gene="RME001"
                     /allele="recA"
     CDS             204..1271
                     /gene="RME001"
                     /allele="recA"
                     /note="recombinase A protein"
                     /codon_start=1
                     /transl_table=11
                     /product="RecA"
                     /translation="MDENKQKALAAALGQIEKQFGKGSIMKLGDNRAMDVETISTGSL
                     SLDIALGAGGLPMGRIVEIFGPESSGKTTLTLEVIASAQKQGKTCAFIDAEHALDPIY
                     AQKLGVNIDELLVSQPDTGEQALEICDALARSGAVDVIVVDSVAALTPKAEIEGEMGD
                     SHMGLQARLLSQSMRKMTGNLKASNCMCIFINQIRMKIGVMFGNPETTTGGNALKFYA
                     SVRLDIRRTGAIKEGDEVVGNETRIKVVKNKIAAPFKEANTQILYGKGFNRYGELIDL
                     GVKHKLVEKSGAWYSYQGDKIGQGKANACKFIEEHSHISQELEAKLREVLLSPVKAEG
                     SAVTDHEVFIDSDEATGHEEF"
     gene            complement(1373..2593)
                     /gene="RME002"
                     /allele="ParB-like nuclease"
     CDS             complement(1373..2593)
                     /gene="RME002"
                     /allele="ParB-like nuclease"
                     /codon_start=1
                     /transl_table=11
                     /product="ParB-like nuclease"
                     /translation="MSSKNENEHENNEIEDSIVQGLMDKYQKSPSEYLASSKQLIESS
                     SLRLDRKVMDEIFKSNSQYYPQLKIINLETPIRHEVVHISPTVARDMLMFSSRGKINT
                     DLSNRRVNRSTVKKFTEDMNEGRWCLTGEPIIIGHDGEILDGHTRLEAASKSNYGFIA
                     IIIWGISDALAFAHTDTGNIRSRANVLEMAGVSVDARVLSQVAMLSKSFEMTANPYAF
                     RGTQGTSFQPAEILDYVQGKDELAWSVNLVASLAREHKQQIQAPQATLAFAHYLINEK
                     LKTYEGDEIPINPDMYISNIISGLGQTSIESIEYQVRDYLESIRKESTSYALLCRLSC
                     IFKGWNMYFNIPVVGKKVAIRRVANYVKNADGERVPAKAAGNIREAFTIPLSPPGKTP
                     IRLKKQPAMKKVIR"
     gene            2783..3916
                     /gene="RME003"
                     /allele="Hypothetical"
     CDS             2783..3916
                     /gene="RME003"
                     /allele="Hypothetical"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="Hypothetical protein"
                     /translation="MHKFGKVKQLTSYVKVIDLFNLFSLFDIPSELEEAFQSIHSRKN
                     QKRIELAKTSIEAGLDNLVEVPPPCITFVVAKVLNHKSLSRSIVELEYDPLDTMIVDG
                     VITLFAMMQISGFSHPFEKKRISKDLTQKNCRVRQELASYPVQVNLLFSPTEPLSQKA
                     CITLYKKYSQTESNIYAPLIESLNAELPLNTYVKEIADDIALDSFGGMKTSSVRLSVK
                     DPYVTTEATMIRLVLGAIGGADYQDKNKVDLFGSGPFSAEHIDKIKPYICIFMEAWLD
                     SVKVQLLSNKNGFHYSTTLWQSLGLVIHNLYLEKKNLEEFAKAGAVLGRLDYSKSAKH
                     WGRCGALELDVLGQSYKNVTGGGRALRIALTNYLLGYLSEQEK"
gene                 complement(4204..6225)
                     /gene="RME004"
                     /allele="DNA mismatch repair;MutL"
CDS                  complement(4204..6225)
                     /gene="RME004"
                     /allele="DNA mismatch repair;MutL"
                     /note="Proetin with DNA mismatch repair MutL domain"
                     /codon_start=1
                     /transl_table=11
                     /product="MutL-like protein"
                     /translation="MAILQERTTGITDGSIKKRFGSIPAWKMLSEYIWNGLDAGASNV
                     NVTININELGGVESIEIHDNGEGIDFHNLDKNFDNWDDSSKKQVTLKGSQGRGRYSFH
                     KYAATATWYTRRNNENARIVIDSSCIKTYKFQPLDEIEQIKDILNNGSGTSVFLSGIT
                     SEKSQKIPPIESIIDSLSNEFGWKLIVCSEINISVNGTLLTPPQHKLFQDTIEIDENQ
                     FSSNIIQWLNKPSGEKSYNYFRDLKDELQHRTLTGFNYKNDFFISGYIQSEWFNDFQA
                     TSSLNNDFFVDNEKTDGSKVLDKLLKELRKKTDEIYQKFLRERAAQLVDEFESKGYFP
                     VYKWEQDADRLVRIEHTKKLVTSICVADPSAFNGLKAKQTKIIIALLDRLSTSSENDS
                     LLEILENILDLDKEHLDEFASQISRSKLDHIISTIGHLQKCDLVIQKMKYLFKEHAKD
                     VLETPDLQGIIEANTWLFGSQYTTIGAEEDDFSKTAKKLRDSDLEILDGDKISSSDLI
                     EGATIEGAKGQVDLFLARKMITVDHSTHDEFIKCTIIEIKRPSVALNKKHLAQAERYA
                     EVLDKHPAFNDERMRFDIVLVGTKISQNDTQINRRLKDLAGKGGAGLVSGADSRIRVY
                     VKSWSTIFNDFEVTHRSLLQKLKIQRDELEYKSKTTLVDELQVPLTEEL"
gene                 6547..9219
                     /gene="RME005"
                     /allele="Type II restriction enzyme"
CDS                  6547..9219
                     /gene="RME005"
                     /allele="Type II restriction enzyme"
                     /note="methylase subunit"
                     /codon_start=1
                     /transl_table=11
                     /product="Type II restriction enzyme"
                     /translation="MAVNQAKIFEQLEKLVEAPDPSEFIYGFLTAFNFPKATLTQIRQ
                     GGNRNVAKIKGHVGLKNKLYYFPAEGDQDIDTALDDVLADQMIAQNKIRYVLTTDFER
                     FLAWDTATNERLDIDFEVLHRNYSFFLPLVGLEKAILNSEKPADVKAAVKMGKLFDLI
                     RVHNHLNTPEDIHALNVFLTRLLFCLFAEDTGIFPKKSQFTSAIKSMTSEDGSDLDQF
                     LSDLFVVLNSPEGSDPRNRLPQHLIDFPYVNGGLFESDESIPKLGKKPRRILIDCGLE
                     DWSAINPDIFGSMFQAVIDVDQRARLGQHYTSYSNIMKVIQPLFLDPLRLELEKQRKS
                     ANGLKRLLVRLGKIKVFDPACGSGNFLIIAYKELRLLEIEVIQALMKIDQGFFISNIH
                     LDQFYGIEIDDFACEIARLSLWLAEHQVNKQWEEHIGPPQPALPLKTTGRVVSANSLD
                     IDWHTVCPNSGNEEIYVIGNPPFKGHRGRSDNQRSEMVRVFKGFKSIGLLDYVACWFW
                     KGAEYIKGTNAELALVSTNSISQGEQVGTLWPSIFNLGVSIHLAYPTFTWANNAKDKA
                     AVHVVIIGLTAKSKKGLLFERVEEDWHSRVVDSISPYLLEGGKACVLAAKAPLNSSVP
                     PLLFGNMPNDGGYLLLNRSERDELIAKEPEAKKWIKKVLGADEFLNSKERWCLWLKDA
                     TKSDLASMPLVLQRIEKVAEKRRKSSDKGTQKLADRPHQFRDLNNPSKYILMPSVSSA
                     RRRYVPVGVFDESVISTNLNYIIPNGTMYEFAILSSLIHNDWMRLVAGRMKSDYRYSA
                     SVVYNTFPWITPSDAQRKELEVLAEDILFAREDYPGKTLAELYNPDSMPLGLLSAHQK
                     LDTAVDRLYRKKPFNDTSERLACLLSRYEDMVTK"
gene                 9247..11268
                     /gene="RME006"
                     /allele="Hypothetical"
CDS                  9247..11268
                     /gene="RME006"
                     /allele="Hypothetical"
                     /note="Protein with helicase and phosphorothioation system
                     restriction enzyme domains"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical Protein"
                     /translation="MKNNILQVSYNQTGQSASHNDMGMREMQARAFAERASQYLLIKA
                     PPASGKSRALMFLGLDKLIHQSVRKLIVAVPEVSIGGSFKDTDLMKFGFFANWTVKPE
                     NNLCISGSESGKTNAFKRFMESDDQILVCTHSTLRSVFDKLLPTDFDNCLVAIDEFHH
                     VSADENSRLGSLIDVLMKQSSAHIIAMTGSYFRGDTVPILLPEDEAKFTKVTYTYYEQ
                     LNGYKHLKTLGIGYHFYQGRYVNALPSVLDPTRKTIIHIPNVNSVESTKDKYTEVDSI
                     LEVLGDVMMQDPETGIYQVKCAKTGKTLLVANLVDDSDMRPKVQAYLRNIESAEQMDI
                     IIALGMAKEGFDWPYCEHVLTIGYRSSMTEIVQIIGRATRDSEGKSHAQFTNLIAQPD
                     AQDDDVKVSVNNMLKAITTSLLMEQILAPSIQFKPRSQWNGEDLPANTLIVDDTTTSV
                     SQKVLDILNRGKDEILSALLSKESVVKGAIADTMPAEVINEIELPSVIQTLYPDLEEH
```

```
                EIEQIRTGVLHSLYIGQKGGLIDGKDLPEDAVIEDGGASYGNGHKSKDNGESGTSNQF
                IKMGDKFINIESLDIDLIDSVNPFHGAYEILSKSVDAAVLKTIQQTVRASQANVSEEE
                AVMLWPRIKAFKVEHQREPSLNSSDPIEQRYADVLAYIRRMKQQRMAQQ"
gene            11284..12531
                /gene="RME007"
                /allele="Hypothetical"
CDS             11284..12531
                /gene="RME007"
                /allele="Hypothetical"
                /note="YeeC-like nuclease domain-containing protein"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical Protein"
                /translation="MIRLRSSKATEKFDLDDIFSEQDDLGLLDVAPLRAKIPTGNILA
                NQFEEISNFYEQNGRVPRSDAQSFYEKRLARRLNAFKSNPEQCDVLSEYDRYSLLDSD
                LAAQENVQRDKVEDKSLPTELDKSELVTSLDDIFDDDDDGLLEFDAPHLFTKTHVPAE
                KKSQPNEIARRQPCAEFHRFSPIFETIQQEIRSGSASLERFRHELQMRVGDVFILNGL
                IGYVHSAGERLEGYSSYNARLHLIFENGTEMHMLFQSLTHGLVRDERGCKIIREGESL
                EPDDTPVPAGLVYVLATKSTDSALAPYKANLYKVGFTDGKVEERIKYAEKDKTFLEAP
                VRVVMTTECYNIDAHKLETLIHGFLGHRRLNVTLKGHAGQYSPREWFYVPLNTVLEVI
                GYILDGTISQYRMDNTTGKVVKK"
repeat_region   12870..12876
                /note="Direct repeat for insertion of ISVuv4"
                /rpt_type=direct
                /rpt_unit_seq="ccttaag"
mobile_element  12877..14083
                /mobile_element_type="insertion sequence:ISVvu4"
gene            13036..14025
                /gene="RME008"
                /allele="Transposase for ISVvu4"
CDS             13036..14025
                /gene="RME008"
                /allele="Transposase for ISVvu4"
                /codon_start=1
                /transl_table=11
                /product="Transposase"
                /translation="MQKNSVMYLEPVRLWGYQEILSTVXQELVETGGIDALINRSRRA
                PNLKNRVDSETEQAVIKYAIDFPAHGQVRTSNELRKLGVFISPSGVRSIWLRNDLENF
                KKRLIALEKQVAENGIILTDEQVAALERKKHDDEACGEIETAHPGYLGSQDTFYVGNL
                KGVGRIYQQTFVDTYSKVAFAKLYTTKTPITAADILNDKVLPFFEAHELPMLRILTDR
                GTEYCGRVEQHDYQLYLAINDIDHTKTKAMSPQTNGICERFHKTILNEFYQVTFRKKL
                YGSIEELQKDLDEWMDYYNNHRTHQGKMCCGRTPIETLEDGKSIWAEKNLAQI"
repeat_region   14084..14090
                /note="Direct repeat for insertion of ISVuv4"
                /rpt_type=direct
                /rpt_unit_seq="ccttaag"
gene            complement(14114..14731)
                /gene="RME009"
                /allele="Hypothetical"
CDS             complement(14114..14731)
                /gene="RME009"
                /allele="Hypothetical"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical Protein"
                /translation="MPMFKYHLDTTKKLSVNGQGFSVLQVYTDTKVTNSQLFINVNDL
                VENSPLTRGEVNEHVANASEEQVIIDQEQTLIRVSSALKLNDPKLRDVDPNVRSQAQQ
                FEQVIDKINMMPKLNEERAIASETVKTKSTKAKQDYKNQRVIQGLGNVCEKTNQPIPQ
                GDNLHIHHDPREADFPELAAEEASLSAIGSTVHSEGHKNDNNPFN"
gene            complement(14734..15192)
                /gene="RME010"
                /allele="Hypothetical"
CDS             complement(14734..15192)
                /gene="RME010"
                /allele="Hypothetical"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical Protein"
                /translation="MFMSDTNLPIELKPLSELIDVKPIEISPDLDEKLTENNQVLVSK
                SIMKIDHQTKTPTPFFSVDSLVSCIGTDRKPFRELMADAADGEVIKINNKYLIRSDLT
                KQFLQERSEQPRSCGERARIEATRSIVNEVGKLDYEQVIALLNNKVQGDE"
gene            complement(15199..15609)
                /gene="RME011"
                /allele="Hypothetical"
CDS             complement(15199..15609)
                /gene="RME011"
                /allele="Hypothetical"
```

```
                /note="DnaJ domain-containing protein"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical Protein"
                /translation="MTSLYVVVFALLLTVVQLLFLLRKYKKKIQELQSTYVESSTTAE
                EADLQVNLVRTSTDDMAYFKSENDRILFLLLEVDGKRRNQLLGITSEMYEDEDAAKKW
                YKSLSNKVHPDKNDDPRAAEAFDKLKQLYNKVTY"
repeat_region   15890..15896
                /note="Direct repeat for insertion of ISVuv4"
                /rpt_type=direct
                /rpt_unit_seq="cctaaat"
mobile_element  15897..17103
                /mobile_element_type="insertion sequence:ISVvu4"
gene            16056..17045
                /gene="RME012"
                /allele="Transposase for ISVvu4"
CDS             16056..17045
                /gene="RME012"
                /allele="Transposase for ISVvu4"
                /codon_start=1
                /transl_table=11
                /product="Transposase"
                /translation="MQKNSVMYLEPVRLWGYQEILSTVIQELVETGGIDALINRSRRA
                PNLKNRVDSETEQAVIKYAIDFPAHGQVRTSNELRKLGVFISPSGVRSIWLRNDLENF
                KKRLIALEKQVAENGIILTDEQVAALERKKHDDEACGEIETAHPGYLGSQDTFYVGNL
                KGVGRIYQQTFVDTYSKVAFAKLYTTKTPITAADILNDKVLPFFEAHELPMLRILTDR
                GTEYCGRVEQHDYQLYLAINDIDHTKTKAMSPQTNGICERFHKTILNEFYQVTFRKKL
                YGSIEELQKDLDEWMDYYNNHRTHQGKMCCGRTPIETLEDGKSIWAEKNLAQI"
repeat_region   17104..17110
                /note="Direct repeat for insertion of ISVuv4"
                /rpt_type=direct
                /rpt_unit_seq="cctaaat"
gene            complement(17211..17645)
                /gene="RME013"
                /allele="Transcriptional activator SoxR"
CDS             complement(17211..17645)
                /gene="RME013"
                /allele="Transcriptional activator SoxR"
                /note="Redox-sensitive transcriptional activator SoxR"
                /codon_start=1
                /transl_table=11
                /product="Transcriptional activator SoxR"
                /translation="MIRNDYLTVGQLSERSGVATSALRFYETKQLITSFRTNSNQRRY
                LPSMLRRVALIQVAQSVGFTLEEIRYELSTLPMNKTATKQDWDRVAKKWQQDLDKKMV
                KIQSLKDNLSGCIGCGCLSMKKCHLFNPEDGLYEQGAQRTIR"
gene            17963..18844
                /gene="RME014"
                /allele="Drug/metabolite transporter (DMT) superfamily"
CDS             17963..18844
                /gene="RME014"
                /allele="Drug/metabolite transporter (DMT) superfamily"
                /note="Permease of the drug/metabolite transporter (DMT)
                superfamily"
                /codon_start=1
                /transl_table=11
                /product="Permease of the drug/metabolite transporter
                superfamily"
                /translation="MLKNYFTLIAIGLIWGSQFIFQETALEAFTPVWIGTFRAVLGAT
                TLFIICRIMRICSSSKQWKLYTVIGFLEATVPFVLVPFAQKELSSSVTAILMGTLPFY
                ALLLAPLFIRNTKITKGNLFSITIGFFGLVVLFYPDLISSSQRIELVSVLAVLLAAVC
                FAIALLLLNRVQREHPLIVARNVLTMASIQLIIVALITTTRMTYKAPSPSIFIAVIYL
                GVMCAGVVYYLYMTSVKNAGAVFTSMTNYLVPAVGVLIAALVTNESIQTTTWLALGII
                LSALLLNQMVEKYEQKT"
gene            complement(19225..21105)
                /gene="RME015"
                /allele="Hypothetical"
CDS             complement(19225..21105)
                /gene="RME015"
                /allele="Hypothetical"
                /note="chromosome segregation ATPase-containing protein"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /translation="MTQYKPFLLVKRLVITKGKSIAYDEKFTAGVNIIRGVNSSGKST
                ISDFIFYGLGGDLTKLKNEAKQCSFVFVEASLSGKVFTLKREIAEGGRKGMDIFSGDY
                ESAYVASVTDWIRYPYNANTKESFYQALFKELGMPYSKSDDKNSITMHQLLRMLYVDQ
                MTSPDRLFKFDKFDSPNKRQAIGELLIGLSDFELYEKRVRLQSLKLALENRIKEIKTI
```

```
                HSFLGGTIKSVSEINGEIEEKRKEIDALELEVESFSSPSEDGCTEDEVLKNLIVEVQE
                ARGKYTASQQEIAKTSFEINDSQMFIESLSRRVKALKETQDTINALSDVAFNHCPACQ
                TEVQARPTGCSLCGATKPESENNIDPTFKVRKEIEFQIAESILLIEKKQIRLDEQKVE
                SNQLETKLGELERDLEIIRKPQRAVNIQLRQKLSEIGGLRNEIRTLNNSKKEFAKLYG
                LYDERDTLQTDFNTLNDEITRLTQRMENELKAKKRKLSEATLSILKADAGHEEIFVDG
                SKVEFDFAEDRVTIDDRALFSASSMVYLKNAFRLAMLKCSCEDSSYLYPRFLLMDNIE
                DKGMEEERSQLFQREIVKLSNSLDVEHQIIFTTSMIDSDLNHSEYCVGDFYNMHNKTL
                KV"
gene            complement(21598..22080)
                /gene="RME016"
                /allele="Hypothetical"
CDS             complement(21598..22080)
                /gene="RME016"
                /allele="Hypothetical"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /translation="MYRQDHRLRELVEEHEKEIQKDEQYKEFSSQLNNFLERKVEGKL
                RDLGQKLTDGNRDFLLDYAMDVKERVSKKIMQFSHYKSAQELYTYILTNIRTTFLHEI
                SPRIKSGDFKVYQINDLVEEKIIQPVLESVHGCSLNIDKDELYGLLYILTGNCYVEWD
                "
gene            complement(22683..23939)
                /gene="RME017"
                /allele="Error prone, lesion bypass DNA polymerase V
                (UmuC)"
CDS             complement(22683..23939)
                /gene="RME017"
                /allele="Error prone, lesion bypass DNA polymerase V
                (UmuC)"
                /note="Error prone, lesion bypass DNA polymerase V (UmuC)"
                /codon_start=1
                /transl_table=11
                /product="UmuC"
                /translation="MFALVDANSFYCSAEQVFRPDWRGKPMIVLSNNDGCIVAANRQA
                KEVGIEKFKPYFQVKALCEQKGVIALSSNYELYADLSAKMMQVIGRFAPEQHIYSIDE
                SFLSFEHSFPAIRSLKEHGMKLRRAVWRECRLPVCVGFGQTLTLAKVANHAAKKIEGY
                NGVCVLDSERERKAVLGQLHVSEVWGIGRKLTHRLEMMGINTALKLANYPPALIRKEF
                NVEVERTVRELNGQKCKGWDAARADKKQIFSTRSAGQRITDLESLQQALCKHANIASF
                KARKQKSLCRVMLCFANSSPFDSPPVARRAVHRFAYPTADVTLITQAASRLAEQLFQQ
                DVRFYKIGVGLIDLVDGQHEQPDLFNLTPNNPKLMNVYDTLNNRFGNDAIFLAAQGIT
                QKWAMRREMLSPQYTTRWQDLPKIKC"
gene            complement(23939..24199)
                /gene="RME018"
                /allele="Error-prone repair protein (UmuD)"
CDS             complement(23939..24199)
                /gene="RME018"
                /allele="Error-prone repair protein (UmuD)"
                /note="Error-prone repair protein UmuD"
                /codon_start=1
                /transl_table=11
                /product="UmuD"
                /translation="MQDVGIFDKDLLIVDRSLDVQDFDIIVANLNGEFICKQIDLNRR
                LLLSANERYQPVPIHEFDQFSLEGIVTRSIRCHRVSPLLRTR"
protein_bind    24359..24374
                /bound_moiety="Site for binding of LexA"
                /function="LexA box"
gene            24984..25739
                /gene="RME019"
                /allele="Hypothetical"
CDS             24984..25739
                /gene="RME019"
                /allele="Hypothetical"
                /note="Multi domain XerS site-specific tyrosine
                recombinase XerS"
                /codon_start=1
                /transl_table=11
                /product="Hypothetical protein"
                /translation="MLWSKMPSKWIAEGILRESFSGGRSVSDDIAALKLFIVLCFLAK
                QVNRSSGSSVDDVLEIRATYDQLTDMCSLSRSLVSRGLKKLHITGLLSSAGVRTKTYT
                FIQGVYRGWCKLPKRALVKNEEEIPAMKAFLNRYSHERDALKCFLYILASRSNSRTYV
                DLSRGTIAKKTGVSLDAIDGAIGFLQSTSLISKVEDKGFLANSIRRDLSERLHRYWVI
                GSSSLNYKTYSVESEDSIFVREPSLYFMENNRR"
gene            complement(26001..27461)
                /gene="RME020"
                /allele="Hypothetical"
CDS             complement(26001..27461)
                /gene="RME020"
```

```
                     /allele="Hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /translation="MKLDSLKIDVKDCLPSDESDPNISWESSLFLHGNGEVNWFATDF
                     LMSDAIPNRSKKTAKGMIRYFLEYLECYENWKYNDIQGRPFPIGLITDSHLYDYVQYI
                     EDDIGLNRNAIANRVRMALRFLEYVQKYYHLSYTLIAIANTDGEYFTKGLVNAERKIS
                     PYGKRYLHHDCIPHCESYGSRSPITDTAIESLYDDLDILEAEGDLYRFEFFSTLISLL
                     EATGIRVSEAANIDTHTIEVLRAQVNASLSGKAIGLDEIISLNKLTINTQSLQAAQAI
                     YRKSALGSANDQLIWIKIKTTKGKNKDKFRIIPISFTTAQYLIRFYDDYIVNELDRIS
                     KGLAKVNRAKFGKLFVHPSSHLPMSGIMISRLFYDVFSRKFKSKHKRSPHLFRHRFIT
                     LLVLQQLKALKTNIGGTQLAILILNRIKGLTGHASIKAMLHYVELAEAELYEDEDESE
                     VFDRVTRDHLVAELGAEHVAEIEAGLRLKKAKQAFI"
     gene            complement(27465..28181)
                     /gene="RME021"
                     /allele="Hypothetical"
     CDS             complement(27465..28181)
                     /gene="RME021"
                     /allele="Hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /translation="MKMNLKRELSPNHNGPKGRRDTLEQVKLRVILLNDFIENGVPDG
                     FTAHISLKALLAYSDSGKIQSRSYPAIYNKKKILVKEIDPSFTGSASSVVDCKDYLLL
                     KIEELRSKIETTSTPSEQVTEAENECNNDDDPKVKTKGELRSALKEQAELIESLAREI
                     LRQRSANNTLISLIKERDKYASRTLKSYYEDHQTELCKVRTVIKPALRETISNLNKLS
                     HEFDDIFAESEENIVSIFGK"
     gene            complement(28165..30198)
                     /gene="RME022"
                     /allele="Phage integrase"
     CDS             complement(28165..30198)
                     /gene="RME022"
                     /allele="Phage integrase"
                     /note="Protein containing phage integrase family domain
                     protein"
                     /codon_start=1
                     /transl_table=11
                     /product="Phage integrase"
                     /translation="MTLININNVKLREGADWSYKSDDTSKWHFAWLHGNFSDSVWHID
                     NGSGKSGSDLPYISFNDYIPTTGDTLDEDKYLIVTNTIRKVIALNRTGHLSTETKNSS
                     LLTRFSSMNSNAIALKILALFLIKEYGESAVSEEGFNLLTTTDIRKFHLEIATGRADK
                     ATGLTDIILEKISTMDRESAIQLFDINVESKNNELPVGEFLQRIGMQNTSISDFTMNL
                     ISDRLVERFPDLSLTITGNNNGKKSEFPNSCPKLFTGKENNISSSTFDALTKGPLLLK
                     RYSNYLPELENYCNPIAKVSKSFKKNHIRSKGRTPNIPTKTALHYLNEAIRIVVVYGE
                     SIVATKINCELQLNKIHKNNPSYLRSHIFEGRYPVNVAIPKNKFTQDFKVTRYNELQK
                     GSTPKKLRENVTVLFGYRILLAATYILTHTFCIKRTTEILELKESNLELGFWGGYELF
                     FGIRKAAPTENSILVTGRPIPHVLFEGISYLAEANEHYYCENEDPFLFPEKYTSAAEG
                     LPPKNSKMSRFAMSNILKDFGDFIQVPTEMVNGIESRFYLSRTHVLRRFAARAFYALT
                     DISDFPAISWLMGHRSTQETWRYLLEDVSNEELSEEEAQGVLDAIYKKGVDTSQVENV
                     IKTELNIKFNNQSDELAKEFIKEQLLSGSKIYSYTDESGKVIVWMEVGNEDES"
     gene            complement(30185..32131)
                     /gene="RME023"
                     /allele="Hypothetical"
     CDS             complement(30185..32131)
                     /gene="RME023"
                     /allele="Hypothetical"
                     /codon_start=1
                     /transl_table=11
                     /product="Hypothetical protein"
                     /translation="MNIKEHTINGMHCQKIMAHQYAVNMSISITKILSGKSDSYKRRT
                     TRMLNEFFTFCYKNNLHMKTGADVSLCYSAWQEFYFSGAVKKQPHTLLSDSRSLWSAL
                     LRLCKRSVNGLPKTNHITPPRRLNTLVFSDTTNSQAKVLGGFQLSTNTTGESLTLILE
                     TDVDAFLGNLISTMKKHRDVIYKVSKSYLIDACDRFTFGQEAMEKIDASVFENNPTLL
                     HPSLLNGKGSRLSLFNPDLPNGECGKTNLLAYIGHCKNGLVDRHFLGGNNHLYRFTSN
                     QYELREHFGLSTLSAVACCNIIITESGINVDSLRKLALSAQGSINKIFEPSQHGFRVS
                     YHKARAKEHLKRNLRHATDTPFIEKAFNYIIEATSYHRSLASGNDANRLFLYQTATAT
                     HKVMPMSDMAFKSGFVRLLVEAKKILKNTPNWCEGVTPECIDEVLMAAPTAKKLRATE
                     GVIRWYESGGNPAVAAKYLGNSEAVSIRNYLPQELQLAVYNQRVRRFQHVLISSATDG
                     QKYQIKALKLKDETELTEYLSRLDERIPHWRSVIETISDNNRNSSTASDIKITLDICP
                     ENIALMKACHEIAMQRIDSANHTEDSISELSFVYQGLTGYLNTHPDRRLMRTLQKGNL
                     LYEKKTGSLLDLKIIKQEVQYDSN"
     repeat_region   32775..32783
                     /note="Inverted repeart for RecA mobile element"
                     /rpt_type=inverted
                     /rpt_unit_seq="cgataaaat"
BASE COUNT     9474 a   6453 c   6605 g  10250 t      1 others
ORIGIN
```

```
   1 attttatcga agagagatta gctgcttagt ataaggtaga actagtttga ttagtgcctt
  61 ttaaaacaat ccgctcataa gtcagtaatg cttcactcag ttataaataa ttgaatgtat
 121 attcatttta gatttgaatt agaaagacta aattgaatat gcattacgtg cataactaaa
 181 tagaattcag agggtatata aagatggatg agaacaagca gaaagcgtta gcagcagcgc
 241 tcggccaaat tgaaaaacaa ttcggtaaag gctctatcat gaaattgggt gacaaccgag
 301 caatggatgt tgaaacaatt tctacaggtt cactttcttt agatattgct ctcggtgcgg
 361 gtggtttacc tatgggtcgc atagtagaga tctttggccc tgagtcatca ggtaaaacaa
 421 cgcttactct tgaggtaatt gcttcggcac agaagcaagg taagacttgt gctttcattg
 481 atgcagagca tgcgctagac ccgatttatg ctcaaaaact aggtgttaat attgatgaat
 541 tacttgtttc tcaacctgat acaggtgagc aagctcttga gatctgtgat gcgcttgctc
 601 gttcaggagc tgttgatgta atcgttgttg actctgtagc agcattgaca ccaaaagctg
 661 aaattgaagg tgaaatgggc gattcacata tgggcctcca ggccaggctt cttagccagt
 721 cgatgcgtaa aatgactggt aatttgaagg cttcgaattg catgtgtatc tttattaacc
 781 agattcgtat gaagataggt gttatgtttg gcaacccaga aacaactact ggtggtaatg
 841 ctctaaaatt ctacgcttct gttcgtctgg atattcgtcg tactggtgcg atcaaagaag
 901 gtgatgaggt cgtgggtaac gaaacgcgta tcaaagttgt caagaacaag attgcagcac
 961 catttaaaga ggcaaacacg caaatccttt atggcaaggg cttcaaccgc tatggtgagc
1021 taattgacct tggcgttaaa cacaaactgg tcgaaaagtc tggcgcttgg tacagctatc
1081 aaggtgataa gatcggtcag ggtaaagcaa atgcctgcaa gtttattgaa gagcatagtc
1141 atatttctca agagttagaa gctaaactgc gtgaggtact attatctcct gtcaaggctg
1201 aaggttcagc ggtcaccgat catgaagtgt tcatcgactc tgatgaggca actggacacg
1261 aagagtttta atctttagtt tgacacctaa aggtaaattc tacaaaaagc gccctgcata
1321 atgggcgctt ttacatttca tatattcctt aacttattgt ttagggctgt ttttatcgaa
1381 tcaccttctt catagcaggt tgctttttta acctaatagg cgtctttccg ggggggggaaa
1441 gaggtattgt gaaagcttct cttatatttc ctgctgcttt tgctggaact cgttcaccat
1501 cagcattttt cacatagttc gctacacggc gaattgctac ctttttgcct acaaccggaa
1561 tattaaaata catgttccat cctttgaaaa tgcatgataa acggcatagg agtgcataag
1621 atgttgattc tttccttatc gattcaagat agtctcgaac ctgatattca atactctcta
1681 ttgacgtttg ccctaaccct gagatgatat tagatatgta catatcagga ttaattggaa
1741 tttcatcgcc ttcgtaggtt tttagttttt catttatgag ataatgtgca aatgccaaag
1801 ttgcttgagg tgcttgaatc tgttgcttat gttctctagc taaagaagca accaaattaa
1861 ctgaccacgc taactcatct ttaccttgta cgtaatcgag tatttcagca ggttggaatg
1921 aagtgccttg agtacctcta aatgcatacg gattggctgt catctcaaaa gattttgaaa
1981 gcattgcaac ctgagacagt actctagcat ctacgctaac acctgccatt tcaaggacgt
2041 tagctctgga gcgaatatta cctgtatctg tgtgagcaaa tgctagagca tcagagatcc
2101 cccatataat gatcgcaata aatccataat ttgattttga tgcggcctct agtcttgtat
2161 gaccatccag tatttcacca tcatgaccga taatgatcgg ttctccagta aggcaccacc
2221 gtccttcatt catgtcttcc gtgaactttt tgactgtact tcgattaaca cgacgattgc
2281 ttaaatcggt attaattttt cctctacttg aaaacatgag catgtctctg gcgacggttg
2341 gtgaaatatg cacaacctca tgtcgaatag gagtcctcta atttatgatc ttgagttgag
2401 ggtaatactg ggagtttgat ttaaagattt catccattac cttcctatct agtctgagac
2461 tggatgattc aattaactgc tttgagctgg ctaggtattc actgggagat ttttgatatt
2521 tatccatcaa cccctgaaca attgaatcct ctatctcatt gttttcatgt tcattttcat
2581 ttttgctact catcatttac ctcaaatctt tcaggttcat agatcaaaaa ccttccaatt
2641 ccacaaagtt agtgtttaat tgttgatcgg acggggcgtc cgatcaatat ttattttaat
2701 ctcaatttag aatgcaaact caacatagtg aggggttcat tcattgaaaa agacatcagt
2761 gaggctggaa ggcttatgct caatgcacaa gtttggtaaa gttaaacaac taacttcata
2821 cgttaaagtc attgatttat ttaacttatt ctcgctattt gacataccaa gtgaacttga
2881 agaagcattt caatcgatac acagccgaaa gaatcaaaaa aggatcgagc tagcaaaaac
2941 ttctattgaa gcgggcttag ataacttagt tgaggtgcct ccgccatgta tcacatttgt
3001 tgtggcgaaa gtcttaaatc ataagagttt aagtagaagt atcgtagaac tagaatacga
3061 tccactcgat actatgattg tggatggtgt gattacgttg tttgccatga tgcaaattag
3121 tggttttagc catccatttg agaaaaagcg tatatctaag gatttaactc agaaaaactg
3181 tcgagttagg caagaactcg caagttatcc cgtacaagtt aacttgttgt tctcacctac
3241 agaaccattg tcacaaaagg catgtatcac tctttataaa aaatacagcc aaactgaaag
3301 caacatttat gcacctttga ttgaaagtct aaatgcagag ttgcctctga acacttatgt
3361 aaaagaaatt gcagatgata tcgctcttga ttctttcggg ggaatgaaaa cgtcgtcagt
3421 taggctgtcc gtaaaagacc catatgtgac tactgaggca acaatgattc gtctagttct
3481 tggtgcaatc ggtggtgcag attaccaaga taaaaacaaa gtcgatttgt ttggttctgg
3541 gccatttagt gccgagcata tcgataaaat aaagccttat atttgtatat ttatggaagc
3601 gtggcttgac tcagtaaagg ttcaactgct ttcaaataaa aatggctttc attactcaac
3661 gacgctttgg caatcactag ggctagtcat acataacctt taccttgaga aaaaaactt
3721 agaagagttt gctaaggctg gggctgtttt aggtcgtctg gattactcca aatcagctaa
3781 gcactgggga aggtgcggtg cgctggaact tgatgtgttg ggacaaagct ataaaaacgt
3841 tacaggaggt ggtagagctt tgcgtattgc cttaactaac taccttcttg gatacttatc
3901 agagcaagaa aagtaagaat gtttctattg tccactcatg gtcaatgatg gtcaatggtg
3961 aacaattgct taatttaaaa gaggaaaggt atgagcttta aggcagttaa atgacaacct
4021 aacatttagt taatgtttga attgtcgagg attttattct tgaagtacaa gtaagagctg
4081 tctctaggtt tgttctgtaa agtatttgcc caccccccc aaaaaaaaag gcaagctatg
4141 gggtgggcta tgtagttgag ttgaagccgc gcctgtgaac aaacaacctc aacactagtg
4201 aaatcacaat tcttcagtta gaggaacttg taattcatcg acaagagtgg ttttggattt
4261 atactccagc tcatctcttt gtatttttag cttctgcaac aaagatctgt gggtaacctc
4321 aaaatcattg aatatcgtag accaactttt gacgtataca cgtattcttg agtcagcccc
4381 agaaactaga cctgcaccac ctttacctgc aaggtctttt aagcgcctat tgatttgggt
4441 gtcattctga gagatttttg taccaactag aacaatatcg aatcgcattc tttcatcgtt
4501 aaaagctgga tgcttatcta aaacttcagc atagcgctca gcttgagcca aatgtttctt
4561 atttagagcg acacttggcc tctttatctc tatgattgta cattttatga actcatcatg
```

4621 ggtcgagtga tcaaccgtaa tcatcttacg agcaaggaat agatcaacct gccccttgc
4681 cccctctatg gttgcccctt caatcagatc gctggagcta atcttgtccc cgtctagtat
4741 ttctaaatct gaatctctca gttttttgc tgttttagaa aaatcatctt cctcagctcc
4801 tatagttgta tactgagagc caaataacca tgtgttagct tcaataatac cttgtaaatc
4861 aggagtttcg agaacatctt tagcatgctc tttgaataga tatttcattt tttgtatgac
4921 taagtcacac ttctgaaggt gtccaatagt gctaattatg tgatcaagtt tagatctact
4981 gatctgagag gcaaattcgt ctaagtgctc cttgtctaag tctaaaatat tttccaatat
5041 ttccagtaag ctatcatttt cactagacgt tgagagtcta tccagcaaag cgatgataat
5101 tttagtttgt tttgccttta ggccattgaa agcggatggg tctgcaacgc agatgctggt
5161 gactaacttt ttagtatgct caattcttac caacctgtct gcatcttgtt cccatttata
5221 tactggaaag tatccttttg actcgaactc gtctacaagt tgagcggctc tttctctcag
5281 aaacttctga taaatttcat cagtcttttt gcgcaactct tttaatagtt tatctagaac
5341 tttagaccca tcggtttttt cattatcaac gaaaaaatcg ttattgagag atgaggtggc
5401 ttgaaaatca ttaaaccatt cagactggat gtaaccgctt atgaagaaat cattttgta
5461 gttaaaacca gtaagagttc ggtgttgcag ttcatctttt aagtccctaa agtagttgta
5521 agatttttcc cctgatggtt tgtttagcca ttgaataata tttgatgaaa actgattttc
5581 atcaatctct atagtatctt gaaatagctt gtgttgaggg ggtgttagta gagtaccgtt
5641 gactgaaatg ttgatttcgg agcaaacaat tagcttccag ccaaattcat ttgacagaga
5701 atctatgata ctttcaattg gtggtatttt ttgagatttt tcactagtga tgcctgaaag
5761 gaaaactgat gttccactac cattattaag aatgtctttt atttgttcaa tttcgtctaa
5821 tggctgaaat ttgtaagttt tgatacagct tgagtcaatt acaatacgcg cattttcatt
5881 atttcttctg gtataccaag tcgctgtggc ggcatacttg tgaaacgaat aacggcctct
5941 gccctgacta cctttgagcg tcacttgttt ttttgaagaa tcgtcccaat tgtcaaagtt
6001 tttatctagg ttatgaaagt ctataccctc tccattgtca tggatttcaa ttgactcaac
6061 gccaccgagt tcattgatgt taatagttac attgacattt gaagcgcctg catccaaacc
6121 attccagatg tattcactca acattttcca agcaggaatt gaaccaaatc ttttttttat
6181 cgaaccatct gtaatgccag ttgttctttc ttgaagtatc gccactgact aacccttttct
6241 tcagtacatt ccacacaaat catcttaacc taagtcacgg gaattattaa ttataattga
6301 tgaacctttg ctgtccgact aggccatgta ccatctatta cagcacctgt taagcacata
6361 tcacggaagc atcaaagatc cttccacaaa aggcatggtt acgacagtgg aaaattttct
6421 agcagtagca atccccttg tttctgtgtg gtagatattt gatatgaatc tcaatgatta
6481 aagtgtgttt tatttgtata atttcccgca tttatggcgc gattcgcaaa ccaatttagg
6541 aacaccatgg ccgtcaacca agcaaaaatc ttcgaacaac tagagaaact tgttgaagct
6601 cctgacccta gtgagttcat ctatggattt cttacagcat tcaatttccc taaagcaacg
6661 cttactcaaa tacgtcaagg tggcaaccgt aacgttgcca agattaaagg tcatgttggg
6721 ctaaagaata agctgtatta cttcccagct gaaggcgatc aagatatcga tactgcttta
6781 gatgatgtcc ttgcagacca aatgatcgcc cagaataaga ttcgctacgt cctgaccact
6841 gactttgaac gtttcttggc ttgggatact gccactaatg aacggctaga tattgacttt
6901 gaagtcttac accggaatta cagtttcttt ttaccactgg tagggttaga gaaggcaatc
6961 cttaacagtg aaaagccagc ggatgtaaaa gcggcagtca agatgggtaa gctgtttgac
7021 cttattcgcg ttcataatca tttgaacacc cctgaagata ttcatgcact gaacgttttt
7081 ctgactcgcc tgctgttctg cttatttgca gaagatactg gtatttttcc aaagaaaagc
7141 caatttacgt ctgcgattaa gagcatgacc tcggaagatg gtagtgattt agatcagttc
7201 ctttctgatt tgtttgtggt tcttaatagc cctgaaggaa gtgacccacg taatcgtttg
7261 ccgcagcacc tgattgattt cccatacgtt aacggtggct tatttgaaag tgatgagtcc
7321 attccgaagt tagggaaaaa accccgccgt attctgattg actgtggact ggaggattgg
7381 agcgctatta accctgatat cttcggaagt atgttccaag ctgttattga tgtggatcag
7441 agagcacgac tagggcagca ctatacctct tacagcaata tcatgaaggt gattcaaccg
7501 ttatttctag acccccttacg acttgaatta gaaaagcagc gtaaaagtgc taacgggtta
7561 aaacgccttc ttgttcgatt aggtaaaatt aaagtttttg accctgcttg tggttctggc
7621 aactttctga tcattgctta taaagagtta cgtttgcttg agattgaggt tattcaggca
7681 ctgatgaaaa tcgaccaagg cttctttatc agcaatatac atctggatca gttttatggc
7741 atcgagatcg atgactttgc ttgtgagatc gcacgtcttt ctctttggtt ggctgagcat
7801 caagtcaaca aacaatggga agaacacatc ggcccccccac agccagcact acctcttaaa
7861 acaacaggta gagttgttag cgctaatagt ttggatattg attggcatac agtttgccct
7921 aactctggga atgaagagat ctatgtgatt ggtaaccctc cgtttaaagg gcatagagga
7981 agaagtgata atcagcgttc tgaaatggta agagtattta aaggctttaa atctataggt
8041 ttattagatt atgtggcctg ttggttttgg aaaggtgctg agtatattaa gggtacgaat
8101 gctgaattag cgcttgtgtc gactaattca atatctcaag gagaacaagt tggtactctg
8161 tggccaagta tttttaattt gggagtttcg attcatttag cttatccaac tttcacttgg
8221 gcaaacaatg ctaaagataa agctgcggtg catgtagtaa ttataggatt aaccgcaaaa
8281 agcaagaaag ggctattatt tgaacgagtt gaagaagatt ggcattcaag agtcgttgac
8341 agtattagcc catacctatt ggaaggaggg aaagcttgtg tgttggctgc aaaggctcct
8401 ctaaatagtt cagttccacc cctgctattt gggaatatgc ctaatgatgg tggatactta
8461 ctgctaaatc gaagtgaacg agatgagtta atagcaaagg aaccagaagc taaaaaatgg
8521 attaaaaaag tactaggtgc agatgaattt ctaaattcaa aagaacgctg gtgcttatgg
8581 ctaaaagatg caacaaaatc tgatttggca tcaatgcccc ttgtactaca aaggattgaa
8641 aaagtagctg aaaaaagaag aaaaagctca gataaaggaa cccaaaagct agcagataga
8701 cctcatcaat ttcgagactt aaataatcca agtaaatata tacttatgcc tagcgtttct
8761 tctgcccgac gaagatatgt gcctgttgga gtgtttgatg aaagtgtcat ttcgactaac
8821 ctgaattaca taattcctaa tggaacaatg tatgagtttg cgattctgtc ttcgctgatc
8881 cataacgact ggatgagact ggtcgccgga agaatgaaga gtgactatag gtattccgca
8941 tcagttgtct acaacacttt cccttggatt acacctagtg atgcacaaag aaaagagctt
9001 gaagttttag ctgaagatat tctatttgcg agagaagatt accctggaaa gactcttgct
9061 gagttatata acccagattc tatgccatta ggcctattaa gtgcgcatca gaagttagat
9121 acggctgttg atagactgta tcgtaaaaaa ccttttaacg atacttctga gcgcttagcg
9181 tgccttctat ctcgttatga agatatggta acaaaataaa aactaaagtt ggcaaaggtt

```
 9241 tagataatga aaaacaatat cttacaggtt agttataacc aaacaggtca aagtgcttca
 9301 cacaacgaca tgggaatgcg agaaatgcag gctcgtgcat tcgctgaacg tgctagccaa
 9361 tatttactca tcaaagctcc tccagcatct ggtaaatcac gcgcactgat gtttttggga
 9421 ttggataagc ttattcatca gtcggttcgt aaactgatcg tcgctgtacc tgaagtgtcc
 9481 ataggtggct cttttaaaga tactgactta atgaagttcg gtttcttcgc taattggaca
 9541 gtaaaaccag aaaacaatct gtgtatcagt ggaagtgaga gtggtaagac taatgctttt
 9601 aagcgcttta tggagagtga tgatcaaatt ttagtttgca ctcactcaac acttcgaagc
 9661 gtctttgata agttgttacc tactgacttc gataattgcc ttgttgctat cgatgagttt
 9721 caccatgttt ctgctgatga aaatagccga ctgggtagtc ttattgatgt tttaatgaag
 9781 cagtcatcag ctcatattat cgcgatgact ggctcttact tccgaggtga tacggttccc
 9841 atattgctcc ctgaggatga agcgaagttt actaaggtga catataccta ctacgagcag
 9901 ttaaatggct ataagcacct aaaaactttg ggtattggtt accattttta tcaaggccgc
 9961 tatgtcaatg cgttaccatc tgtacttgat ccgacaagga aaaccattat ccatatccca
10021 aacgtaaatt ctgttgagtc taccaaggat aaatacaccg aagtcgactc tattcttgag
10081 gtgctcgggg atgtaatgat gcaagaccca gagactggta tttatcaggt gaaatgtgca
10141 aaaacaggta agacgttgtt ggttgcaaac ttagttgatg acagtgatat gcggcctaaa
10201 gttcaggcat acctgcgtaa tatagagtca gcagaacaaa tggacatcat cattgcattg
10261 gggatggcaa aagaaggttt tgactggcct tactgcgagc atgttctgac tattggatat
10321 cgtagttcta tgacggagat cgttcagatt attgggcgtg caactcgtga tagtgagggt
10381 aaaagccacg ctcagtttac gaaccttatc gctcaacctg atgcacagga tgatgacgta
10441 aaggtttcgg ttaacaacat gctgaaagct attacaacct cgctcttaat ggagcaaata
10501 ctagcgccga gcattcaatt taaacctcgc tctcagtgga atggtgagga tttaccagcg
10561 aataccctaa tcgttgatga caccaccaca tctgtttcgc aaaaagttct ggatatcctt
10621 aatcggggca aggacgagat tctttcagca ttattatcta aagagtcggt tgttaaaggg
10681 gcgattgccg atacgatgcc agcagaggtc attaatgaga ttgaactacc gtctgtgatt
10741 caaacgttat atccagatct agaagaacac gaaattgaac agataagaac gggtgtgttg
10801 cacagtcttt atattggtca gaaggggggc ttaattgatg gtaaagacct tcctgaagat
10861 gcggtgattg aagatggtgg tgcctcctat ggtaatggtc ataagtctaa agataatggg
10921 gaatcaggga cgagtaatca gtttataaaa atgggtgata agttcattaa tattgagagt
10981 ttagatatag acttgatcga ctcagttaac ccttttcatg gtgcttacga aatcctatct
11041 aagtccgtcg atgctgccgt gttgaaaact attcagcaaa cagttcgagc tagtcaggca
11101 aatgtttctg aagaggaggc tgttatgctt tggcctagaa tcaaagcctt caaagtcgaa
11161 caccaacgtg agccttcact aaactcctca gatcctatag agcagcgtta tgctgacgtt
11221 cttgcttata ttcgtaggat gaaacagcaa cgcatggcgc agcaataaga gggttatcac
11281 caaatgatcc gattacgctc ttctaaggct acagagaagt tcgacttaga tgatattttt
11341 tcagaacagg atgatttggg gctactcgat gtagctccct tgagggcaaa gattcccact
11401 gggaatatat tagccaatca atttgaagaa atttctaatt tctacgaaca aaatggtcga
11461 gtcccacgta gtgatgctca atcattttat gaaaagcgcc tagcgcgccg tctcaatgcg
11521 tttaagtcta atcctgaaca atgtgacgta ttatctgaat acgaccgtta tagcttgcta
11581 gatagcgatt tagccgcgca agagaatgta caaagagata aggtcgaaga taagtcttta
11641 cctaccgagc ttgataagtc tgaactcgta acttctctag atgatatctt tgacgatgac
11701 gacgatggcc ttttagagtt tgatgctcct cacctgttca ccaagaccca tgtgcctgct
11761 gagaaaaaat ctcagccaaa tgagattgcc aggaggcaac cttgtgctga gttccaccgt
11821 ttttccccta tattcgaaac gatccaacag gaaattagat caggctctgc atcactagag
11881 agatttcgtc atgaacttca gatgcgtgtt ggggatgtct ttatccttaa cggcctgatt
11941 ggatatgtcc attcagcagg cgaacggtta gaggggtata gtagctataa tgccagattg
12001 cacctgatct ttgagaatgg tacagaaatg catatgttgt tccagtcgtt gacgcatggt
12061 ttagttcgcg atgaacgcgg gtgtaagatt attcgtgaag gtgagtctct tgagcctgat
12121 gataccctg ttccggcggg actcgtttat gttctagcta ctaagagtac agatagtgca
12181 ttagcacctt ataaggcaaa cctttataag gtcggcttta cagatggtaa ggttgaagaa
12241 cgtattaaat atgccgagaa ggataaaacg ttcttagagg ctcctgttcg tgttgttatg
12301 accaccgagt gttacaacat cgatgctcat aaacttgaaa cattgataca tggattctta
12361 gggcatcgta ggttaaacgt cactctcaaa gggcatgcag gacaatatag tccaagagag
12421 tggttttatg tgccgttgaa tactgtttta gaggttattg gctacatctt agatggaacg
12481 atatcccaat acaggatgga taacacgacg ggtaaagttg taaaaaaata agctgcttta
12541 agcctttgtg tagttgaaag cattatcgtg gttatgttct tagcataacc cgttttgatg
12601 acgtatgtcg gttaggaggg gcttttagaa agagcagaaa gtccccaagg ccgctaggag
12661 gcgttattgt ctgtagtaac aactcatact ctcacatccg ctgttttgga cacgattaga
12721 tactgtcagc ctgagggtat atcaactcgc aattggatat ttcctagttg aaccattgct
12781 aatgtagtaa agcctaaatt aagtaaggta tggcaaatta tccgttagtc tagtgtcaag
12841 agtcacaaga gaacatcaac aaatgaatgc cttaagtgta ttagttccga ctagatctga
12901 cacttcaatt tgaaaagaag agagttaccc actttgatta aaggtgctta atcactaatc
12961 aaagacaaaa agaggtaact ctcatgcttc atactagcaa tccaattatc aaacacaaag
13021 cgggccttct caatcttgca gaagaactcg gtaatgtatc tagagcctgt aaggttatgg
13081 gggtatcaag agatactttc taccgttatn caagagttgg ttgagacggg gggtattgat
13141 gctctgatta accgtagccg aagagcaccg aatttgaaga accgtgttga tagtgaaact
13201 gagcaagccg ttatcaaata cgccatcgac ttccctgctc atggacaagt tagaacgagt
13261 aatgaattac gtaaattggg agtgtttatc tctccaagtg gcgtacgctc aatctggctt
13321 cgcaatgacc tagagaactt caagaaacgt cttattgcac tggagaaaca ggtcgcagag
13381 aacggtatta tcctaacgga cgagcaagtt gcggctcttg agcgtaagaa gcacgatgat
13441 gaggcttgtg gtgagataga aacagcgcat ccaggttatc tcggctctca agacacattc
13501 tatgttggca acttgaaagg tgttgggcgc atctatcaac aaaccttcgt tgatacctac
13561 agtaaagtcg cctttgccaa gctctacaca acgaaaacac caatcaccgc agcggatata
13621 ttgaatgata aggttctacc gttctttgag gcgcatgaac tgccaatgct gcgaatcttg
13681 actgaccgag gcactgaata ctgtggtcgt gttgagcagc acgattacca gctctaccta
13741 gccattaatg atatcgacca cacgaaaact aaagcgatgt caccacaaac aaatggtatc
13801 tgcgagcgct tccacaagac catattgaat gagttctacc aagtgacatt cagaaagaaa
```

13861 ctgtatggtt ctatcgaaga gttgcagaaa gatctggacg aatggatgga ctactacaac
13921 aatcaccgta ctcatcaagg aaaaatgtgc tgtggcagaa cgccgataga aacattagag
13981 gatgggaaat caatctgggc tgaaaagaat ctagcccaga tataatctga caggcaccaa
14041 gtcgaaaagt gggtaactgt cagatcaggt ctgaattagt acaccttaag tagtcaattt
14101 ttcagatttc attctagtta aaggggttat tatcgttttt gtgcccctca ctgtgaacgg
14161 ttgaaccaat ggcagataaa gatgcttctt ctgctgccag ttctggaaag tccgcttctc
14221 tgggatcgtg gtgaatatga aggttatcgc cctgaggtat gggctgattt gtttttcac
14281 acacgttacc taaaccctgt ataactcttt gattcttata atcttgcttt gccttggtag
14341 actttgtttt caccgtctct gatgcaatgg cccgctcctc attaagctta ggcatcatat
14401 tgattttatc aataacttgc tcaaattgct gagcttgaga acgcacgttg gggtcaacat
14461 ctctcaactt agggtcattc agtttcaaag cactagaaac acgaattaag gtttgctctt
14521 gatcaatgat cacttgctct tcagatgcat ttgcgacatg ctcgttcact tcaccacgag
14581 tcaatggaga gttttcaacg agatcgttaa cattgataaa caattgacta tttgtaacct
14641 tagtatctgt atacacttgt aatacactaa atccttgtcc attcacggaa agtttttttg
14701 tggtatctaa gtgatattta aacatgggca tttttattca tcaccttgaa ctttgttgtt
14761 taacagcgca attacctgtt cataatctaa cttacctacc tcattcacta tacttcgtgt
14821 tgcctcaatc cgcgctcgct ctccgcaaga acgtggttgc tccgagcgct cttgcaaaaa
14881 ttgtttggta agatcgctgc gaataaggta tttattattt atttttatta cctcaccatc
14941 tgccgcatct gccattaact ctctgaatgg ctttctatca gttccaatgc atgacactaa
15001 tgaatcaaca ctaaaaaatg gagtaggtgt tttggtctga tgatcgatct tcataatagа
15061 cttacttact aatacttgat tgttttcggt aagtttttca tccaaatcgg gggagatttc
15121 gatgggtttc acgtcaatca attccgatag tggcttaagc tctatgggta aatttgtgtc
15181 actcatgaac attacccact aatatgtaac tttgttatag agttgcttta atttatcgaa
15241 agcctccgcg gcccttggat catcattttt gtctggatga actttatttg acagtgactt
15301 ataccatttt ttagctgcat cctcgtcctc atacatctct gatgtgatac caagcaattg
15361 gtttcgacgc ttaccatcaa cctctaacaa caaaaacaaa atgcgatcgt tttccgactt
15421 gaagtacgcc atatcatcag ttgatgtcct gacaagatta acctgtagat cagcctcttc
15481 agcggtcgta ctagattcaa cataggtgga ttgtaattct tggatttttt tcttgtattt
15541 tctcaacaag aaaagaagtt gcacaactgt taataggagt gcaaaaacaa caacatacag
15601 tgatgtcata tatatttcct tggtagtctg aaaagtatag gaatgtttta taccacatca
15661 cctcaggctt gcataagtga gacagttagg aaaggcatat cacattgatg aattcgtgtg
15721 gcggatgaga aaatgtggcc acttagctag agcatttgtc ataagaaaaa gtaagtcaaa
15781 gaccaaactg aatgctatgc acgaaagcaa cctaacaaac aatgggaat gaacgtgcca
15841 gaacggatcg gcctgacacg agtagggaca aatccttgtc agtgactttc ctaaattgta
15901 ttagttccga ctagatctga cacttcaatt tgaaaagaag agagttaccc actttgatta
15961 aaggtgctta atcactaatc aaagacaaaa agaggtaact ctcatgcttc atactagcaa
16021 tccaattatc aaacacaaag cgggccttct caatcttgca gaagaactcg gtaatgtatc
16081 tagagcctgt aaggttatgg gggtatcaag agatactttc taccgttatc caagagttgg
16141 ttgagacggg gggtattgat gctctgatta accgtagccg aagagcaccg aatttgaaga
16201 accgtgttga tagtgaaact gagcaagccg ttatcaaata cgccatcgac ttccctgctc
16261 atggacaagt tagaacgagt aatgaattac gtaaattggg agtgtttatc tctccaagtg
16321 gcgtacgctc aatctggctt cgcaatgacc tagagaactt caagaaacgt cttattgcac
16381 tggagaaaca ggtcgcagag aacggtatta tcctaacgga cgagcaagtt gcggctcttg
16441 agcgtaagaa gcacgatgat gaggcttgtg gtgagataga aacagcgcat ccaggttatc
16501 tcggctctca agacacattc tatgttggca acttgaaagg tgttgggcgc atctatcaac
16561 aaaccttcgt tgatacctac agtaaagtcg cctttgccaa gctctacaca acgaaaacac
16621 caatcaccgc agcggatata ttgaatgata aggttctacc gttctttgag gcgcatgaac
16681 tgccaatgct gcgaatcttg actgaccgag gcactgaata ctgtggtcgt gttgagcagc
16741 acgattacca gctctaccta gccattaatg atatcgacca cacgaaaact aaagcgatgt
16801 caccacaaac aaatggtatc tgcgagcgct tccacaagac catattgaat gagttctacc
16861 aagtgacatt cagaaagaaa ctgtatggtt ctatcgaaga gttgcagaaa gatctggacg
16921 aatggatgga ctactacaac aatcaccgta ctcatcaagg aaaaatgtgc tgtggcagaa
16981 cgccgataga aacattagag gatgggaaat caatctgggc tgaaaagaat ctagcccaga
17041 tataatctga caggcaccaa gtcgaaaagt gggtaactgt cagatcaggt ctgaattagt
17101 acacctaaat tagcatttaa ctgattacct cagacattgt aaactctaac tgcctcgccg
17161 aattcactta tccactacag actgtgttga atatagaatt gccgaaaaca tcatctaata
17221 gttcgttgcg ctccttgctc atataaccca tcctctgggt taaaaagatg acacttcttc
17281 atactcaaac aaccacagcc gatacaccca gagaggttat cttttagtga ctggatcttt
17341 accatttttt tgtcaagatc ttgctgccac ttcttcgcaa ctctatccca atcttgttta
17401 gtagctgttt tgttcatagg taaggtactt aactcatacc ttatttcttc tagggtaaac
17461 cctactgact gtgccacttg gataagtgcg accctacgta acatagaagg caagtaacga
17521 cgttgattac tattggttcg aaaagatgtg atcaactgtt tggtctcata aaaacgtaaa
17581 gccgatgtcg cgacaccact gcgttctgat aactgcccaa cggtaaggta atcatttcta
17641 atcataatta ttctaacttt tgtcaattta tcgctttact taaagttaac tttaacttttt
17701 atactccttt caaatataag actcaacgca taaaaggata tgatcatgcc gaaatccccc
17761 cccataagtg aagagatgcc aaattgcatg tcggcagata cgcagtcaat aagcaactta
17821 attcctgtac ctaaactagc gctggatagc ataacaattg atgttgatgc cgttacttttt
17881 aaatcggctt agccagtacg ccataatctg tttcaacttt tcatacagct caacctgttg
17941 agctgcgttt ggattcttta ctatgcttaa aaactatttc actttgatcg caattgggct
18001 gatttgggga tcacaattta tttttcaaga aacagcacta gaagcattca ctccagtatg
18061 gataggcaca tttcgtgcgg tactcggtgc cacaaccctg tttattattt gtcgaataat
18121 gcgaatttgc agtagtagta aacagtggaa actctacact gtaatcggtt ttcttgaagc
18181 aactgttcct ttcgttctcg tgccttttgc tcaaaaagag ttgagtagtt cagtgactgc
18241 aattttaatg ggaacattac cctttttacgc acttcttctc gcgccattat ttattaggaa
18301 tacaaagatc accaaaggaa acctgttcag tataaccatc ggtttttttg gtttagttgt
18361 acttttctac cctgacctaa tttcatcatc tcaaaggatt gagcttgtca gcgtgctagc
18421 cgtcttactt gccgcagtat gctttgctat tgctttgcta ttgctaaacc gtgttcaaag

18481 agagcatcct ctgatagtcg caagaaatgt tctgaccatg gcgagcattc aactgatcat
18541 tgtcgctctc attacaacga cgagaatgac atataaagca ccatcgccga gcatttttat
18601 tgctgtaatc tatcttggtg tcatgtgtgc gggagttgtc tattacttat atatgactag
18661 cgtcaagaat gcaggagctg tttttacgtc aatgactaac tacctcgtac cagctgtagg
18721 tgtattaatc gctgcgctag tcacgaatga aagcatacag actacaactt ggctggcact
18781 tggtattatt ttatcagctc ttttattgaa tcaaatggtc gaaaagtatg aacaaaaaac
18841 ttagcgggaa tgttcaggct atagttgtaa ctttttagta aaacaaaacg ctcggtgtac
18901 ttcatcgggc gtgtgtactt taaagcctta tctaatctta cttgattata ggttttatta
18961 cccatcagtc taacttgctt caaatgtctt aggttcaaat agtaaaactg aaagcgttac
19021 acttcaaaag agatacctat tagttttgat catgaattat aacgcatccc gaaacaaact
19081 ttggggcaaa aactagttag ggtgtttaag taaaatagcg ctgtaaggat taggagggat
19141 aatgaaaaca taccgttgca gggcttaaaa acaacaaaga gacaatttaa aaattatctc
19201 tttgtaagtt gggctacaag agagtcatac ttttaaagtc ttattatgca tattgtagaa
19261 atcacctaca caatattcac tgtgattaag gtcgctatca atcatagatg tagtgaaaat
19321 gatttggtgc tctacatcta gagagttgga tagtttcacg atctcacgct ggaaaagttg
19381 gcttcgttct tcttccatcc ccttatcttc aatattgtcc ataagtaaga agcgagggta
19441 caaataagaa gagtcttcac atgagcactt aagcatagcc aatctaaatg cgttctttag
19501 ataaaccatc gaactagcag aaaacaaggc tctatcatca atggtcactc gatcttctgc
19561 aaagtcgaac tcgactttgc taccatcaac gaatatttct tcatgtccag cgtcagcttt
19621 aagaatactt aacgttgctt ctgatagctt tctcttcttc gctttaagct cattttccat
19681 tcgctgcgtc aatctagtga tttcatcatt cagcgtattg aagtctgtct gcaacgtatc
19741 tcgttcatca tataatccat agagtttagc gaactctttc ttactattgt tcagtgttct
19801 gatttcattg cgtaaaccac ctatttcact gagcttttgc cttagctgta tattcaccgc
19861 tctttgaggt ttgcggatga tctcaaggtc tcgttccaac tcgcctaact tagtttcaag
19921 ctggttcgat tcaaccttct gctcatcaag ccttatttgt tttttctcta tcagaagaat
19981 tgactcggca atctggaact ctatttcttt tctcacttta aatgtagggt cgatattatt
20041 ctcagattca ggttttgttg caccacacaa agaacaacct gtaggtcttg cttggacctc
20101 tgtttggcaa gctgggcagt gattaaatgc aacgtcagat aaagcgttga tagtatcttg
20161 ggtttcttta agagccttaa ctctacggct taaagactca ataaacatct gagagtcatt
20221 gatttcaaaa gacgttttag ctatttcttg ttggcttgct gtgtactttc ctcttgcctc
20281 ctgaacctct acaatcaaat ttttaagtac ttcatcttct gtacaaccat cttccgatgg
20341 gctagaaaac gattcaacct ctagctccag cgcatcaatc tctttacgct tttcttcaat
20401 ctcaccattt atttcactaa cagattttat cgtaccaccc aaaaatgagt gtatggtttt
20461 aatctcttta attcgatttt ccaacgctag tttcaagctc tgtaatctga cccgtttttc
20521 atagagttca aaatctgaca agccgatcaa tagctccccg atagcttgac gcttgtttgg
20581 tgaatcgaac ttatcaaact tgaatagtcg atctggtgac gtcatctgat caacatagag
20641 cattctaagt aactgatgca ttgtaatact gttcttgtca tccgactttg aatacggcat
20701 accaagctct ttgaacaagg cttgataaaa actttctttt gtattggcgt tataagggta
20761 acgaatccaa tcagtaacag atgcaacgta tgctgattca tagtctccgc taaatatatc
20821 catccctttc cttccaccct ccgcgatttc ccgcttagg gtgaatacct tgccgcttag
20881 agaagcttcg acaaaaacga agctacactg cttcgcttca ttcttaagct tagttaaatc
20941 accgccaagg ccgtagaata taaagtctga gatcgtagac ttaccagaac tgtttacgcc
21001 tctgataata ttcacaccag cagtgaactt ttcatcatag gcaatacttt tgcctttggt
21061 aatgactagt cttttgacta ataagaaagg tttatactgc gtcatatctg tattccatta
21121 agcctgttct tttcttaagt ccactttcac caacaatctc aatcgcattt agggtttgga
21181 atagcttgca aaaaagctcc gtagactcat tagagttatc gttaatcagc ttttgaatag
21241 acgggctata gaaccctct gagatccgca agttaccgtt ttcttcctca aaaataccct
21301 ttgcttttag tatttgaagt gcctgtatat ggtagtcgcc catctctgaa aatagtcgtt
21361 tcttatctgg tagaacttcg tatggttgat ttattgagcc aaagctacgc ttcaactcag
21421 cgattccttt agttcttgga aactcaatat cttttacaag gtgtggaaat aggtaaaaca
21481 aatcgtaaat acgcagtctt accgagtcaa aatcctctaa atcttgcata cttgctatgc
21541 tgataaacct aaacatacag tggtaaagat cgtttcttgg atgataaata atcacttta
21601 atcccactct acatagcagt tacctgtaag tatatacagt agtccgtaca gctcatcttt
21661 gtcaatattg aggctgcacc catgaacact ttccaataca ggttggatga ttttttcttc
21721 gactaaatca ttgatttggt agactttaaa gtcgccagat ttaatgcgtg gtgaaatctc
21781 gtggagaaac gtggttctga tattagtgag gatatatgtg tagagttcct gagccgattt
21841 gtaatgcgag aactgcatta tcttcttcga tacacgctct ttgacatcca tagcgtaatc
21901 caagagaaaa tctctgttac catcggtgag cttttgccct agatctcgaa gcttcccttc
21961 gacttttcgt tccagaaagt tattgagctg actagaaaac tctttgtact gttcatcttt
22021 ctgaatttct ttttcatgct cttcaacaag ctctcgcagc ctatgatctt gacgatacat
22081 aactggagct gcgtgattat atgttatgtg ggttgagtga tcatgaacat ctcgtccagc
22141 gacactacgt ccagcattaa cgttgtcgaa agaggcttta tttgaatcca gtaccgacac
22201 ttatgaccct tttttgtccc taccaacaac atcgcctcca gcgcgaatat tgctcattgt
22261 tacattgttg gaagatgctt ttgacctaaa gaagaaataa cctataactg cgacaatagc
22321 cgctacgatt aggctaacaa cgtaaacacc taagccagag aacaaaacat cgaagttttc
22381 gcagacccag tttatagctt catccaccct aacatcctcg tggtttttcc ctaaacttta
22441 caagtattta attttattca caaatcgaca agccaaaaat atctacttac ttctcaatat
22501 atgcaatcgg ttaggtcact atgtagatgt atggctgaat cgaccaagct tctttgaaca
22561 ttgcactgat tagaaaaatt acttcaaatt gtacacacta gatatgatgg attcgcagtc
22621 taaatcgaaa tggggcaaaa ccaagctagt tctgtattag atagagcata attcacacta
22681 tctcagcatt tgatttttgg taaatcttgc caacgggtgg tgtattgagg cgacaacatt
22741 tcacgccgca ttgcccattt ttgagttatg ccttgagcag caagaaagat agcatcattc
22801 ccaaatcgat tgtttaacgt gtcatacaca ttcattagct taggattatt cggggtcagg
22861 ttgaacagat cgggctgctc atgctgaccg tcaactaagt caatcagccc aacaccaatt
22921 ttataaaatc tcacatcctg ttggaataac tgctctgcca gtcgagaggc ggcttgagtt
22981 atcagagtca catcagcagt aggatacgca aaacgatgaa ccgctcggcg tgccacagga
23041 ggcgagtcaa atggtgaact gttagcgaaa caaagcatca ctcgacacag tgatttttgt

23101 tttcgagcct taaaggatgc aatatttgca tgcttgcaca acgcttgctg caacgactct
23161 aggtcagtaa ttctctgccc tgcactgcgg gtagaaaata tttgcttttt atccgctcta
23221 gccgcatccc agcccttaca tttttgtcca ttgagttcac gtactgtacg ctccacttct
23281 acgttaaact cttttctaat cagtgcagga gggtaattgg caagcttgag cgctgtatta
23341 atacccatca tctctaatcg atgggttagt ttcctgccta tgccccacac ctctgagaca
23401 tgcagttgcc ctaaaaccgc cttacgctct cgttcgctat ctaagacaca tacgccattg
23461 tatccttcga ttttttcgc agcatgattg gccacttttg ctagcgtcaa tgtttgccca
23521 aatccgacac atacgggtaa tcgacattca cgccaaacag cgcgtcgtaa tttcattcca
23581 tgctctttga gagaacgaat ggctggaaag ctatgctcaa atgaaaggaa agactcatcg
23641 atactgtata tgtgttgttc aggtgcgaag cgaccaataa cctgcatcat tttggcagat
23701 aaatccgcat acaactcata gttcgatgac agagcaatca caccttctg ttcacaaagt
23761 gctttaactt gaaagtatgg tttgaacttt tctatgccaa cttctttagc ttgtcgattt
23821 gctgcgacaa tgcaaccgtc gttgttggac aaaacgatca tcggctttcc acgccaatcg
23881 gggcgaaaaa cttgctcagc gctacagtag aaagaattag catcaaccaa agcaaacatt
23941 atctcgttct caacaatgga ctcactcggt ggcaacgaat tgagcgagtt acaatccctt
24001 cgagggaaaa ttgatcaaat tcgtggatag gaacgggctg gtatcgctca ttagcagaga
24061 ggagtaatcg acgattgagg tcgatttgtt tacaaataaa ttccccattg aggtttgcca
24121 ctatgatgtc gaaatcttgc acatctaatg agcgatctac aataagtaaa tctttgtcga
24181 agatacccac atcttgcata gaatctccac atgcgtatcc cagaaaggtg gcacaaggat
24241 gctcaataag aagctggtct agatctaaag cgagttgagt atactcttca gcaggactct
24301 cgaagcctgt gataccagca tgagcaaaga tagggatgac tttcatacaa aacatatgct
24361 gtatattgat acagtatatt gtatttcaga ttaggcattt tgcaacagca ctgttgagat
24421 aggatttctg agcttaatat catcaagaga actattgatt gacaactgga ctcaatatga
24481 gttgtctcaa tgggatgatg ctcgtacatt cttggataat tccgagttac agtatttgct
24541 aaattagtag ttagcggacg aattgaggta tggcaatacc ttctctacgc agaatggtct
24601 accttgaggc gcgtgtctac agatatggtg gtgttatact tctccatgaa ctaatagggt
24661 attgccatgt actaatggga gggtttttat ttttaaatac ctgtttttta tcagttgttt
24721 aaggtttatt tctcttttct attttctaga aagattatgc gaagacatat ttaaagacag
24781 caatgcttca tgtaaaaagt actcatgttc tacgtatttg aaggttttca ggaggtagtg
24841 atgttagagc tgtttttaag tagtttggag agggtcgcta aaatacaacc tcagataact
24901 atcctcattc tagtgttagc cttattaggt agtgacttga tctcgttcat caatgacttt
24961 atggcggtaa tgaaatgact ttaatgctct ggtctaaaat gccctcaaaa tggattgcag
25021 agggaatact aagagagtcc tttagtggtg gtaggagtgt tagtgatgat atcgctgctt
25081 tgaagttgtt cattgtactg tgttttcttg ccaagcaagt taaccgttcg agtggtagct
25141 cggtggatga tgttttggaa atcagggcaa catatgatca gcttacagat atgtgttctc
25201 tgagccgttc tttggtgtct agaggcttaa agaagttaca tataacaggg ttgttatcat
25261 ccgctggtgt tcgaacaaaa acatatactt ttattcaagg agtctaccga ggatggtgca
25321 agcttccaaa aagagcgctt gttaaaaatg aggaagaaat cccagcgatg aaagcttttc
25381 tcaataggta ttcccacgaa cgtgatgctc taaaatgctt tttgtacata cttgctagtc
25441 gaagcaatag tagaacttac gtcgatctta gtcggggaac tatagccaaa aagactggcg
25501 taagtctgga tgctattgat ggcgctatcg gatttctgca aagcactagt ttgatttcta
25561 aagtagagga caaaggcttc ttggctaact ctattcgtcg tgatttgtca gaacgactgc
25621 ataggtactg ggtgattggt agcagcagtt tgaactacaa gacatattcg gttgaaagtg
25681 aagactctat ttttgttagg gaaccatcgt tgtactttat ggagaacaac aggagataat
25741 gcatatttca ttctcacgta attcaataat ttatctatta attcaatatg ttaattgacg
25801 tattgagagg ttggactgtt cagtgataca taatttctag atgtgctgaa ttcagataac
25861 ttttggtaag gtaaaatcgt attactagaa tgtgcatgag tatttgctct aagtgaaaat
25921 ggggcagaac cgatttaggg ggacttgtca gatctatcca tatcggggaa acgaaattga
25981 cagattgtat tcagaccaat tcaaataaac gcctgtttag ctttttcag cctcaatcca
26041 gcttcaatct ctgcaacatg ttcagctcct aattcggcaa ctaaatggtc acgagttact
26101 ctatcaaaaa cctcactttc atcctcatct tcataaagtt cagcttccgc gagttcgaca
26161 tagtgaagca ttgctttgat actcgcatgc ccagttaagc ctttaattct gtttaggatt
26221 aggattgcga gttgagtccc gccgatgttg gtttttagag ccttaagttg ttgtaataca
26281 agtagcgtaa taaaccgatg cctaaacagg tgggggctgc gtttatgttt gctcttaaac
26341 ttccgactga atacatcata gaataagcgc gaaatcatta ttccactcat cggcaagtga
26401 gagcttgggt gcacgaataa cttcccgaat tttgcacggt taactttagc aagccctttg
26461 cttatgcgat caagttcatt aactatatag tcatcataaa accgaatgag gtattgtgct
26521 gttgtaaatg aaatcggaat aatcctaaat ttatctttgt tcttgccttt ggtggtttta
26581 atttttatcc aaataagctg gtcgttagcg ctgcctaaag ctgatttacg gtaaatagct
26641 tgagctgcct gtaagctctg tgtattgatg gtaagcttat ttagactaat aatttcatcc
26701 agtcctatag ccttaccact caaagatgca tttacctgcg cacgtaagac ttcaattgta
26761 tgagtgtcaa tgtttgcggc ttcacttact cgaatgcctg ttgcctcaag caaactaatt
26821 agagtagaaa agaactcaaa acgatatagg tcaccttccg cttccaagat atctaggtca
26881 tcatatagac tttcgattgc agtatctgtt attggacttc tactaccata gctctcacag
26941 tggggtatgc agtcgtgatg caagtaacgt ttcccataag gactgatttt tctttctgca
27001 ttaaccaacc cctttgtaaa gtattcacca tcagtattag ctatagctat taacgtatag
27061 ctcaaatggt agtatttttg tacatattca agaaaccgta gtgccatcct aactctgtta
27121 gctatagcat ttcgatttag gccaatatcg tcttcgatat attggacgta gtcatataaa
27181 tgagagtcgg taataaggcc aatagggaaa gggcgtcctt gaatatcgtt atatttccaa
27241 ttttcataac attctaggta ttcaagaaaa taccgaatca tacctttggc agttttcttt
27301 gaacggttcg gtattgcatc tgacattaaa aagtctgtcg caaaccaatt aacctcacca
27361 ttaccatgca aaaacaaact tgattcccaa ctgatgttag gatcactttc atctgagggc
27421 agacagtcct taacatctat ttttaggcta tctaatttca tatgttattt gccaaatatt
27481 gacacaatat tttcttcact ttcagcgaaa atatcatcaa attcatgaga gagtttgttt
27541 aagtttgaaa ttgtctccct tagagcgggt ttaataaccg ttctaacttt acatagttcg
27601 gtctggtgat cctcgtagta tgatttcaac gtccgtgaag cgtatttgtc acgttctttt
27661 attaaagata ttaaagtgtt gttagcactt ctttgccgca atatctccct agctaacgat

27721 tctataagtt ctgcttgctc tttaagggca ctccgtagtt cacctttagt tttaactttg
27781 ggatcgtcgt cattattgca ttcattctct gcttcagtaa cttgctcgga tggagtgctt
27841 gtagtttcaa tcttgcttct taattcttct atttttagga gtaggtagtc cttacagtca
27901 accactgatg aagcagagcc agtgaaggat ggatctattt cttttactag gatttttttc
27961 ttattgtaaa tggcaggata gctacgtgat tgtattttgc cactatcaga gtatgctaaa
28021 agtgctttta aactgatatg ggcggtaaat ccatctggca ctccattctc aataaaatca
28081 ttaagaagga ttacacgaag tttaacctgt tcaagcgtat ccctccgccc tttagggccg
28141 ttgtggtttg gcgataattc ccttttaaga ttcatcttca ttacccacct ccatccaaac
28201 aattaccttt cctgattcat ctgtgtagct atatatcttc gagccagata gaagttgctc
28261 ttttataaat tcttttgcca actcatctga ctgattattg aatttaatgt ttagctcagt
28321 tttgattaca ttttccactt gtgatgtatc aacacctttt ttgtaaattg catctaatac
28381 accctgtgct tcttcttctg atagctcttc gttacttacg tcctctagca agtaacgcca
28441 agtttcctga gtgctgcggt gccccatgag ccaagatatg gcaggaaaat cacttatatc
28501 agtaagcgca tagaaggcac gagcagcgaa tcgtcttaaa acatgggtac gagataaata
28561 aaagcgtgat tcaataccat tcaccatttc tgttggaact tggataaagt cgccaaaatc
28621 tttcagtatg ttactcattg caaagcgact cattttactg ttttttgggg gcagaccttc
28681 cgctgctgag gtgtatttt caggaaacaa aaaggggtct tcgttttcac aatagtaatg
28741 ctcattcgct tcagccagat atgagatccc ttcaaataaa acatgaggta taggtcgtcc
28801 agtaacaaga atggaattct ctgtaggcgc agcctttcgg ataccaaaaa acaactcata
28861 gccaccccaa aatccgagtt ccaaattaga ttctttgagt tcaagtatct ctgttgttcg
28921 cttaatgcaa aaggtatgtg ttagaatata tgttgctgca agcaatatcc tgtatccgaa
28981 taaaacagtg acattttcac gtagtttctt tggtgtagat cctttctgta actcattata
29041 tctagttacc ttgaaatctt gagtaaactt gtttttaggg atggctacat tcactgggta
29101 gcgtccttca aaaatatgag aacggagata acttggattg ttcttgtgga ttttatttag
29161 ctgtaattca cagtttatct ttgttgcaac aatactttct ccgtaaacaa ccacaatccg
29221 aatagcttca ttcaaataat gtagagctgt tttggtagga atattaggag tgcgaccctt
29281 tgatctgatg tgatttttt taaatgattt actaaccttt gcaataggat tgcaataatt
29341 ctccagttca ggtaaatagt ttgaatatct ctttaaaagc aatggtcctt tagtcagagc
29401 atcaaaagta ctcgatgaaa tattattttc ttttcctgta aataactttg ggcatgaatt
29461 aggaaactca gatttttttac cattattatt gccagtgatt gtcagtgata aatcaggaaa
29521 tcgttccact aatctgtctg aaattaaatt cattgtaaag tctgatatac ttgtgttctg
29581 cataccaatc ctttgtaaaa attcacctac tggtaattcg ttgtttttgg attctacatt
29641 tatgtcaaat agttggattg ctgattctct atccatagtc gatatttttt ctaaaataat
29701 atcagtaagc cctgtagctt tgtctgccct ccctgtggca atttccaagt gaaattttct
29761 tatatcggtg gtagtcagaa ggttgaatcc ttcttcggat acagcacttt caccgtactc
29821 tttaattaag aatagagcta atattttgag cgcgatggca ttgctattca tacttgaaaa
29881 gcgtgtcagc aatgacgaat ttttggtctc tgttgataaa tgtcccgttc tatttagagc
29941 aataactttc cttatagtgt tagttactat caggtattta tcctcatcaa gagtatcacc
30001 tgttgtcgga atgtaatcat taaacgaaat atatgggagg tcggaaccac tttttcctga
30061 accattatca atatgccata ctgagtctga aaaattgcca tgaagccaag caaagtgcca
30121 ttttgaggtg tcatctgatt tataagacca gtcagcccct tcacgaagct ttacgttatt
30181 gatattaatt agagtcatac tgaacctctt gctttattat ttttagatca agcaagcttc
30241 cagtttttt ctcatatagt aggttaccct tctgcaaagt tcgcattagc ctcctatctg
30301 gatgtgtgtt caagtatccc gtcaacccct gatacacgaa tgatagttct gaaatgctgt
30361 cttctgtatg attggcagaa tctatgcgct gcattgctat ctcatgacaa gccttcatta
30421 aggctatatt ttcaggacag atgtctagag taatttaat atcactggct gtactggaat
30481 tcctattgtt atcagatata gtctctatca cgcttctcca gtgaggaatg cgctcatcta
30541 gccgagataa atactccgtt agctcagttt cgtctttcag cttcagagcc ttaatttggt
30601 atttttgtcc atcagtagcg ctgctgatca atacatgctg gaaccgcctt actcgctggt
30661 tataaaccgc caactgtaac tcttggggta agtaatttct tatggacaca gcttcagagt
30721 tcccaagata tttagctgca acagcagggt ttccaccact ttcataccaa cgaatgacac
30781 cctcagtagc acggagcttc ttggctgttg gagctgccat tagcacttca tcaatacatt
30841 ctggcgttac tccttcacac caattagggg tgtttttaag gatctttttt gcttcgacaa
30901 gtaggcgaac gaaaccactc ttaaaagcca tatcactcat tggcattact ttatgagttg
30961 cagtagctgt ctgatacaaa aataatctat tggcatcatt gcccgatgcg agagagcggt
31021 gatatgaggt tgcctcaatg atgtagttaa aagccttttc aatgaacgga gtatccgttg
31081 catgcctcaa attgcgcttt aaatgctctt tagctcttgc tttatggtag ctgactcgga
31141 atccgtgttg agatggttcg aaaatcttgt taatcgagcc ttgagctgac aatgctaatt
31201 tacggagact atcaacgtta atacccgatt cggtgatgat tatgttgcag caagcgacgg
31261 ccgataacgt ggacaaacca aaatgctctc ttaactcata ttgattactt gtgaatctgt
31321 aaagatggtt gttaccgccc aaaaaatgac gatcaactag tccatttta caatgtccaa
31381 tgtatgctag taagttggtt ttcccacatt caccgttagg taaatctgga ttaaaaggc
31441 tgagtcgaga gccttttccg tttaaaaggg agggatgaag tagagtgggg ttattttcaa
31501 acacactagc atctattttt tccattgcct cttgcccaaa ggtaaaccta tcgcaagcat
31561 caatcagata gcttttagat actttataga tcacgtcacg atgcttttc atcgtggata
31621 tcaaattacc tagaaaggca tctacatccg tttcaagtat taaggtgagg ctttcccccg
31681 tagtgttagt tgataactga aaacctccga ggactttggc ttgggagttt gtagtatcag
31741 agaaaactaa agtatttaaa cgtcttgggg gagtgatgtg gttggttttt ggtaggccat
31801 taactgaacg tttacaaagt ctcaacaaag cactccaaag gcttcgggaa tcactcagca
31861 aggtgtgtgg ttgtttcttg acagcgccgc taaaatagaa ctcttgccat gcagagtagc
31921 ataacgatac atcagctccc gtcttcatat gtaggttgtt cttataacag aatgtgaaga
31981 attcgtttaa catcctcgtc gtacgtcttt tgtaagagtc tgatttacca gagagaattt
32041 tagtgatgga tatgctcata tttacagcat actgatgtgc catgattttc tggcagtgca
32101 taccgtttat tgtatgctct tttatgttca taaacacttc caccgtataa tgtcattgct
32161 aatactttga aagatattag cactaaggag tgttgtgtaa agcattgttt tttatactta
32221 tttttatgtt agtgcatgag gttaaaagaa cgatagtgga agattttgtt caatatattg
32281 agatgctgaa gttagatgag gacaggctcg atttatttat tataaccaat gtcagtatgg

```
    32341 atgggggaa gtgtgagttt atcctaacta tagattcacc agtgcctaag ctgtcgagta
    32401 agtggttaat ttcatgtgag gattgtctga agctacatat agatagtacg acgagtccag
    32461 ctcatgacat tgccataaaa tacggcattt tagttattgg aacttcgtat gtaacaggag
    32521 cctatttcaa agccgtacga ctaaacgttt aagttttat tcggaagtgt agcattttta
    32581 tttggttctt catttgtggt gacgtgcgtc gatttctact tgcctacaat aaatcctacc
    32641 ttataggtgt ttcatttctt accttataat tgtgttttgg tttcttacga gaacttattg
    32701 caaactacgt agttaatgat atgatcttta aggataatct tttatttttg ttttaagtta
    32761 cagttttgtg gttccgataa aat
//
```

Appendix 5

frontiers in
MICROBIOLOGY

MINI REVIEW ARTICLE
published: 09 December 2013
doi: 10.3389/fmicb.2013.00385

The function of integron-associated gene cassettes in *Vibrio* species: the tip of the iceberg

***Rita A. Rapa*[1,2] *and Maurizio Labbate*[1,2] ***

[1] *ithree Institute, University of Technology, Sydney, NSW, Australia*
[2] *Department of Medical and Molecular Biosciences, University of Technology, Sydney, NSW, Australia*

Edited by:
Daniela Ceccarelli, University of Maryland, USA

Reviewed by:
Mauro M. Colombo, Biotechnology Center, E. Mondlane University, Mozambique
Genevieve Garriss, Institut Pasteur, France

***Correspondence:**
Maurizio Labbate, ithree Institute, University of Technology, Sydney, NSW, Australia; Department of Medical and Molecular Biosciences, University of Technology, PO Box 123, Broadway 2007, Sydney, NSW, Australia
e-mail: maurizio.labbate@uts.edu.au

The integron is a genetic element that incorporates mobile genes termed gene cassettes into a reserved genetic site via site-specific recombination. It is best known for its role in antibiotic resistance with one type of integron, the class 1 integron, a major player in the dissemination of antibiotic resistance genes across Gram negative pathogens and commensals. However, integrons are ancient structures with over 100 classes (including class 1) present in bacteria from the broader environment. While, the class 1 integron is only one example of an integron being mobilized into the clinical environment, it is by far the most successful. Unlike clinical class 1 integrons which are largely found on plasmids, other integron classes are found on the chromosomes of bacteria and carry diverse gene cassettes indicating a non-antibiotic resistance role(s). However, there is very limited knowledge on what these alternative roles are. This is particularly relevant to *Vibrio* species where gene cassettes make up approximately 1–3% of their entire genome. In this review, we discuss how emphasis on class 1 integron research has resulted in a limited understanding by the wider research community on the role of integrons in the broader environment. This has the capacity to be counterproductive in solving or improving the antibiotic resistance problem into the future. Furthermore, there is still a significant lack of knowledge on how gene cassettes in *Vibrio* species drive adaptation and evolution. From research in *Vibrio rotiferianus* DAT722, new insight into how gene cassettes affect cellular physiology offers new alternative roles for the gene cassette resource. At least a subset of gene cassettes are involved in host surface polysaccharide modification suggesting that gene cassettes may be important in processes such as bacteriophage resistance, adhesion/biofilm formation, protection from grazers and bacterial aggregation.

Keywords: integron, gene cassette, *Vibrio*, mobile DNA, mobile genetic elements, mobile genes, lateral gene transfer

INTRODUCTION

Members of the *Vibrio* genus are ubiquitous in marine environments and show a wide range of niche specialization (Thompson et al., 2004). The capability of vibrios to occupy diverse niches is a testament to their ability to adapt and evolve. An important driver of this in vibrios is lateral gene transfer (LGT). LGT is the mechanism of DNA transfer from one bacterial cell to another without the requirement for cell division. It is followed by subsequent incorporation of the DNA into the recipients' genome such that DNA can be stably inherited, a process assisted by mechanisms such as homologous recombination or via a range of mobile genetic elements (MGEs) such as transposons and genomic islands (Stokes and Gillings, 2011). This mini review will focus on one important MGE called the integron, an element commonly known for its role in antibiotic resistance. The focus on the integron and its role in antibiotic resistance has driven a lack of understanding (and perhaps lack of interest) for the role this element plays in the broader environment. In contrast, we argue that understanding integron contribution to the antibiotic resistance problem requires an understanding of the role of integrons in their broad evolutionary context. Since integrons are present in almost all *Vibrio* species and comprise a significant proportion of their genome, they are excellent candidates for studying alternative roles of integrons outside of the clinical environment. Using recent work from *Vibrio rotiferianus* DAT722, we discuss possible environmental roles for this MGE.

WHAT ARE INTEGRONS?

An integron is a site-specific recombination system capable of integrating and expressing open reading frames (ORFs) contained in modular structures called gene cassettes (**Figure 1**; Mazel, 2006; Labbate et al., 2009). The integron is defined by three components, an integrase gene (*intI*) that encodes a site-specific recombinase, an attachment site (*attI*), and a promoter (P_c). The mobile units that insert into integrons are gene cassettes. Gene cassettes commonly consist of a single promoterless ORF and an IntI-identifiable recombination site called *attC*. The integration of gene cassettes is facilitated by an integrase-mediated recombination reaction between *attI* × *attC* and less commonly *attC* × *attC*. Multiple insertion events produce a contiguous cassette array with cassettes downstream of the P_c promoter being co-transcribed. Induction of *intI* can also cause excision and rearrangement of a gene cassette(s) into a different position.

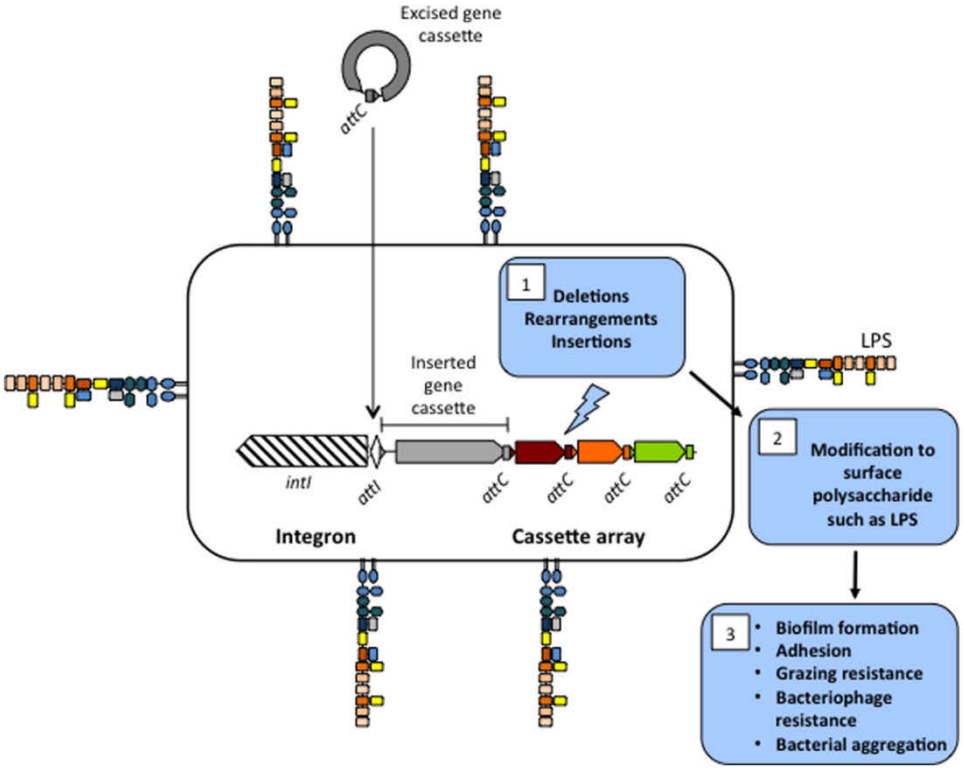


FIGURE 1 | Proposed mechanism for the production of surface polysaccharide diversity through deletions, rearrangements, and insertions in the cassette array of *Vibrio* species. The integron consists of an integrase gene (*intI*) and a primary recombination site called *attI*. Gene cassettes consists of a secondary recombination site called *attC* and are circularized in their excised format. Recombination of multiple cassettes into *attI* (or *attC*) results in a contiguous cassette array that in *Vibrio* species can make up 1–3% of their genome. Deletion, rearrangements, and insertions in the cassette array (1) results in modification of cell surface polysaccharide (2) that may affect processes such as biofilm formation, adhesion to different surfaces, grazing resistance, bacteriophage resistance, and bacterial aggregation (3).

Integrons are a diverse family of elements and are catalogued into classes based on the nucleotide sequence of the integrase gene. Currently, there are over 100 different integron classes, *most present on bacterial chromosomes* and found in approximately 10% of sequenced bacterial genomes (Boucher et al., 2007). The class 1 integron was the first described integron (Stokes and Hall, 1989), found linked to antibiotic resistance genes in resistance plasmids from Gram negative pathogens. This is because class 1 integrons in clinical contexts largely contain antibiotic resistance gene cassettes with approximately 130 so far described (Partridge et al., 2008). The accumulation of multiple resistance gene cassettes (up to about six) has associated these elements with multi-drug resistance (Leverstein-van Hall et al., 2002, 2003). Unlike clinical class 1 integrons which carry an identical integrase gene sequence, diverse class 1 integrons are also found in the chromosomes of environmental *Betaproteobacteria*, containing divergent integrase sequences and functionally diverse gene cassettes. This indicates that *Betaproteobacteria* were the original source of the clinical class 1 integron and that its initial capture by a transposon disseminated it across diverse Gram negative pathogens and human commensals (Gillings et al., 2008).

Although the class 1 integron is by far the most abundant integron in clinical contexts, others have been described (approximately five). An environmental source for all these clinical integrons strongly suggests that integrons have a much broader role in adaptation than conferring resistance to antibiotics in clinical environments (Rowe-Magnus et al., 2001; Boucher et al., 2007; Gillings et al., 2008). Phylogeny shows integrons to be ancient structures (Rowe-Magnus et al., 2001; Boucher et al., 2007) therefore, the gene cassette pool has been contributing to adaptation and evolution of bacteria for several hundred million years and not just in the last 70 years during the antibiotic revolution. This point is sometimes not well understood by researchers studying clinically derived integrons.

RESEARCHING THE BROADER ROLE OF INTEGRONS IN A FIELD FOCUSED ON THEIR CONTRIBUTION TO THE DISSEMINATION OF ANTIBIOTIC RESISTANCE

Due to the ongoing issue of bacterial antibiotic resistance, research is still heavily focused on the role of class 1 integrons. A PubMed search with the term "integron" retrieves in excess of 2200 publications. A search with the following terms "integron and (antibiotic or resistance or class 1)" retrieves 1847 publications showing that 83% focus on antibiotic resistance and/or class 1 integrons. Antibiotic resistance is a significant issue and we are certainly not suggesting that the emphasis on the role of integrons in this area is not justified however, we believe that this is impacting adversely on our understanding of these elements including in relation to the antibiotic resistance problem. Firstly, the focus on antibiotic resistance is skewing understanding for the general role of integrons in the wider research community. Given the hundreds of integron classes that exist, much of our knowledge is based on a single class (i.e., the class 1 integron). This overshadows the likely important role that integrons play in the broader environment and sometimes results in an erroneous dogma that knowledge of class 1 integrons can be extrapolated to all integron classes. The overshadowing of research in integrons outside the clinical setting is easily observed via a cursory examination of the literature over the last 3 years. Although we have known for nearly a decade that there are over 100 classes of integrons (most from non-clinical bacteria including in *Vibrio* species), publications still claim the existence of "4" (Madiyarov et al., 2010; Rezaee et al., 2012), "5" (Li et al., 2013), "10" (Salimizand et al., 2013), or "several" (Peymani et al., 2012) classes of integrons based on their knowledge of clinically derived integrons.

We and other authors have also experienced erroneous dogma in comments provided by expert reviewers for submitted manuscripts, particularly with regard to whether integrons described from natural environments correspond with what is known about "integrons" (mostly class 1). An excellent example as to why research from class 1 integrons cannot be extrapolated to all integron classes is shown in the recombination reaction rate of the class 1 integron and that from *Vibrio cholerae.* The *V. cholerae* integron has a 2600-fold higher rate of recombination in a *V. cholerae* background compared to an *Escherichia coli* background indicating the involvement of host factors (Biskri et al., 2005). In contrast, the class 1 integron shows no difference in rates of recombination in both backgrounds. The class 1 integron's capacity to operate in different backgrounds is the likely reason for why this particular integron has been successful in its mobilization across different bacteria. This trait and possibly a greater capacity to integrate cassettes with diverse *attC* sites (Biskri et al., 2005) are likely to make this integron an exception rather the rule.

Secondly, it has been over 12 years since the discovery that integrons are diverse and found in different environments (Mazel et al., 1998; Nield et al., 2001). Knowledge on the function of integrons with regard to site-specific recombination, transcription of gene cassettes, and regulation of the integron-integrase has significantly advanced and has been excellently reviewed elsewhere (Cambray et al., 2010; Roy Chowdhury et al., 2011). However, little progress has been made in addressing the precise ways in which gene cassette products contribute to the adaptation and evolution of bacteria outside of antibiotic resistance. Based on our knowledge of integrons in antibiotic resistance, we are aware of the power of this system in providing rapid adaptation under strong selection pressure(s). In approximately 70 years, the integron has assisted in making antibiotic treatment problematic and most likely obsolete in the next 10 years (World Health Organisation, 2013). Class 1 integrons are now a common fixture on plasmids from commensal bacteria and Gram negative pathogens. Re-entry of commensal and Gram negative pathogens into the broader environment through routes such as wastewater treatment ensures that access to the environmental gene cassette metagenome will be easy and rapid. Thus, a lack of understanding or distribution of misinformation regarding this greater resource, particularly in the antibiotic/clinical field, has the potential to be counterproductive in the quest to solve or improve the antibiotic resistance problem into the future.

In addressing the knowledge gap for the broader adaptive role of integrons, *Vibrio* species make excellent candidates. As already stated the cassette arrays of vibrios tend to be quite large comprising up to 3% of their genome and consisting of diverse and unique gene cassettes. To date, the largest cassette array is in *Vibrio vulnificus* CMCP6 consisting of 219 cassettes and totaling approximately 150 kb (Kim et al., 2003). In order to highlight the necessity for research into the role integrons play in bacterial adaptation and evolution and to focus attention on the lack of understanding that exists about the function of this element in bacteria generally, we will review and discuss the phenotypic functions of cassettes in *Vibrio* species in the context of what has been discussed above.

A BIG BLACK BOX IN OUR UNDERSTANDING OF GENE CASSETTE FUNCTION

Cassette arrays in *Vibrio* species are large and mostly consist of unique and novel genes with no identifiable function. In 2007, a bioinformatics survey of gene cassettes from multiple genome sequenced *Vibrio* species found that 65% of cassette proteins had no known homologs and that 13% had homologs of unknown function (Boucher et al., 2007). The remainder showed a wide range of non-specific functions in metabolism, cellular processes, and information storage. Similar statistics have been observed through PCR amplification of gene cassettes from metagenomic DNA (Elsaied et al., 2007; Koenig et al., 2008, 2009). Putting aside this massive knowledge gap in cassette function for the moment, large cassettes arrays provide an extra level of complexity. While some have argued that P_c is the only driver of cassette transcription in large arrays (Guerin et al., 2009; Cambray et al., 2010), other studies have shown otherwise (Yildiz et al., 2004; Michael and Labbate, 2010). A study of the 116-gene cassette array of *V. rotiferianus* DAT722 showed that the majority of gene cassettes were transcribed and that numerous diverse promoters across the array were present that responded to different growth conditions (Michael and Labbate, 2010). The presence of these diverse promoters provides integrated cassettes with multiple regulatory options. This gives the capacity for cassettes to re-arrange with different promoters potentially building operon-like structures that

express complimentary cassette proteins. Such an idea has been demonstrated in principal using artificial gene cassettes containing genes for tryptophan biosynthesis (Bikard et al., 2010). This complexity can be elevated when we consider that *Vibrio* species live in populations where gene cassettes might be considered a community resource not just a singular cell resource. For example, integrons might provide a way for the community to break down and/or extract energy from complex substrates without the entire pathway (and genetic burden) being owned by just one cell. Amusingly, the complexity of integrons has been used as proof for God/intelligent design (Hunter, 2010).

Even with the limited understanding of gene cassette function, a number of studies have sampled the gene cassette metagenome from different environments and attempted to determine how cassettes might influence adaptation and evolution (Elsaied et al., 2007, 2011, 2013; Koenig et al., 2008, 2009, 2011). Although correlations are observed such as homologs of genes in cassettes encoding pollution degrading enzymes from contaminated environments (Nemergut et al., 2004; Koenig et al., 2009) or environments showing a "gene cassette ecotype" (Koenig et al., 2008), it is still the case that ∼80% of the gene cassettes are of unknown function. In a study looking at gene cassettes from *Vibrio* species found in coral mucus, 12.5% of gene cassettes were implicated in biochemical processes also associated with antibiotic resistance (Koenig et al., 2011). The authors argued that gene cassettes provide a competitive advantage by delivering protection from, or by synthesizing, antimicrobials in the coral environment. While sound, the fact that this conclusion could be drawn clearly reflects the amount of research that has been done in the integron/resistance field. No other conclusions on the other cassette-assisted bacterial interactions present in the coral mucus could be made. So we are still left with a gaping hole in our understanding of how gene cassettes contribute to adaptation and/or evolution in this environment.

A handful of gene cassette products have been functionally characterized and these have been summarized in **Table 1**. In many instances, characterization of these gene cassettes was selected based on some homology to a known protein such that a phenotype could be tested which does not really address the bulk of unknown and hypothetical gene cassette products. In other instances, they were selected based on their capacity to be crystallized or were identified as part of mutant library or other screens. In the instances where gene cassettes have been removed from their natural bacterial host and expressed in *E. coli* or where *in vitro* techniques have been used to study protein activity, caution must be taken in how their function is interpreted. Interactions of cassette proteins with host pathways may modify how these gene cassettes affect cell or community behavior. This was observed in a study in *V. rotiferianus* DAT722 where deletion of a gene cassette encoding a putative topoisomerase I-like protein affected porin regulation. This phenotype could not have been predicted if characterized outside the host (Labbate et al., 2011). This is also true of the bioinformatic studies described above where in the small proportion of gene cassette products that could be identified were often proteins such acetyltransferases, methylases, or transcriptional regulators. Without knowing the primary substrate that is being modified by the acetyltransferase or methylase or the gene(s) controlled by the transcriptional regulator, the biological importance of the cassette(s) is still unclear. Therefore, an approach where gene cassettes are deleted or expressed in their natural host is arguably the best way to identify their physiological role.

NEW INSIGHT INTO GENE CASSETTE FUNCTION IN THE VIBRIOS

Vibrio rotiferianus DAT722 is the only microorganism where extensive physiological analysis has been conducted on isogenic mutants with gene cassettes deleted. This has revealed new insights into how gene cassettes affect adaptation and evolution. In one study, a cassette could not be deleted without a compensatory mutation (Labbate et al., 2011). The resulting mutants had abnormal regulation of their porins and impaired growth in minimal medium. The gene cassette in question was the 11th cassette from *attI* and appears to be strain specific by lacking close relatives elsewhere. The cassette 11 protein contains two domains, one with weak homology to nucleases and the other a C4-zinc finger domain commonly found in topoisomerase 1 proteins. These domains indicate a DNA binding/processing protein that is likely to have a regulatory role potentially through controlling the coiling state of gene promoters. Irrespective of the exact role, this study is important in demonstrating that cellular networks can rapidly integrate a mobile gene cassette such that it becomes advantageous for survival. It also shows that benefit need not necessarily come from acquisition of a novel functional gene(s) but through modification of existing host cellular networks (Labbate et al., 2012).

In a follow up to this study, the impact of deletions on the cassette array of *V. rotiferianus* DAT722 was addressed (Rapa et al., 2013). Indels are regularly observed in *V. cholerae* arrays and are likely in all large arrays however, their impact on bacterial physiology were unknown (Labbate et al., 2007; Szekeres et al., 2007; Yan et al., 2011). Three deletion mutants were subjected to physiological growth, stress, proteomic, and chemistry-based techniques to determine the effect of cassette deletions on vibrio physiology. The total deleted cassettes encompassed 58% of the array. Surprisingly, growth and stress assays of these mutants showed little change compared to the wild-type. Furthermore, proteomic analysis of one deletion mutant in different media and growth stages showed only 0.5–1% change in the proteome. This indicates that unlike deletion of cassette 11, the majority of cassettes are not integrated into host pathways and do not affect the major metabolic pathways of the cell, at least in the conditions observed.

Importantly, analyses did identify changes to host surface polysaccharide in the deletion mutants with proton nuclear magnetic resonance on whole cell polysaccharide indicating that gene cassette products decorate host cell polysaccharide via the addition or removal of functional groups. Consistent with this result, one mutant had modified biofilm-forming capabilities in a simple batch biofilm assay (Rapa et al., 2013). This is a significant result as it focuses future researchers who are addressing gene cassette function in vibrios to surface polysaccharide. We propose that at least a subset of cassettes are involved in modifying host surface polysaccharide and that deletion (and most likely rearrangements and acquisition) of cassettes is a mechanism for creating surface property diversity. There is significant biological

Table 1 | Non-antibiotic resistance and experimentally confirmed functional ORFs in gene cassettes.

Source of cassette	Function	Determination of function	Reference
Cassettes from *Vibrio* species			
Vibrio cholerae	Sulfate-binding protein	Complementation of *E. coli* mutation	Rowe-Magnus et al. (2001)
Vibrio cholerae OP4G	Transcriptional regulation	Crystal structure determination and drug binding assay	Deshpande et al. (2011)
Vibrio cholerae GP156	Heat stable enterotoxin	Active in suckling mouse assay when expressed in *E. coli*	Ogawa and Takeda (1993)
Vibrio cholerae	Mannose–fucose resistant hemagglutinin	Mutagenesis *in vivo* and testing in infant mouse model	Franzon et al. (1993), Barker et al. (1994)
Vibrio marinus	Psychrophilic lipase	Active when expressed in *E. coli* at 10°C	Rowe-Magnus et al. (2001)
Vibrio vulnificus CMCP6	Cold shock	Complementation of cold shock phenotype in *E. coli*	Rowe-Magnus (2009)
Vibrio vulnificus CMCP98K	Secretion	Expression in *E. coli* mediates secretion of periplasmic proteins	Kim et al. (2003)
Vibrio rotiferianus DAT722 (cassette 21)	dNTP-pyrophosphohydrolase (iMazG)	Crystal structure determination. Expressed in *E. coli* and enzyme activity measured	Robinson et al. (2007)
Various large cassette arrays like this in *Vibrio* spp.	Toxin/antitoxin (TA) genes	Demonstration that presence of TA genes limits deletions in large cassette arrays	Szekeres et al. (2007)
Vibrio vulnificus 1003	Capsular polysaccharide biosynthesis	Transposon mutagenesis *in vivo*	Smith and Siebeling (2003)
Vibrio rotiferianus DAT722 (cassette 11)	Porin regulation	Deletion of cassette *in vivo*	Labbate et al. (2011)
Vibrio rotiferianus DAT722 (multiple cassettes)	Surface polysaccharide modification	Deletion of cassettes *in vivo*	Rapa et al., 2013
Cassettes from metagenomic DNA			
Soil metagenomic DNA	Potential transport protein	Crystal structure determination	Robinson et al. (2005)
Soil metagenomic DNA	ATPase activity	Expressed in *E. coli* and enzyme activity measured	Nield et al. (2004)
Soil metagenomic DNA	Methyltransferase activity	Expressed in *E. coli* and enzyme activity measured	Nield et al. (2004)

importance to surface-associated polysaccharide and its modification as evidenced in the literature. This includes biofilm formation (Lee et al., 2013), bacterial cell co-aggregation (Vu et al., 2009), bacteriophage resistance (Scholl et al., 2005), evasion of immune cells (Pier et al., 2001) as well as resistance to antimicrobial peptides (Westman et al., 2008; **Figure 1**).

CONCLUSION

In the last 12 years, little progress has been made in the precise ways that cassette gene products contribute to adaptation and evolution of *Vibrio* species. One reason is the emphasis that is placed on studying integrons from clinical contexts. Another is that characterization of unknown genes is difficult and thus not considered a fruitful endeavor by researchers, especially in the current competitive research environment. However, if we are to learn more about the broader role of integrons, some of our focus needs to shift to identifying functions for gene cassettes. This will not only improve our understanding of this important genetic resource in a broader sense but give improved context for these elements clinically.

ACKNOWLEDGMENTS

We thank Dr. H. W. Stokes, who has since retired, for helpful comments and guidance in the preparation of this document. We also like to acknowledge his support as a mentor and for his contributions to the integron field throughout his career.

AUTHOR CONTRIBUTIONS

Rita A. Rapa and Maurizio Labbate both contributed to the writing of this manuscript.

REFERENCES

Barker, A., Clark, C. A., and Manning, P. A. (1994). Identification of VCR, a repeated sequence associated with a locus encoding a hemagglutinin in *Vibrio cholerae* O1. *J. Bacteriol.* 176, 5450–5458.

Bikard, D., Julié-Galau, S., Cambray, G., and Mazel, D. (2010). The synthetic integron: an in vivo genetic shuffling device. *Nucleic Acids Res.* 38, e153. doi: 10.1093/nar/gkq511

Biskri, L., Bouvier, M., Guérout, A.-M., Boisnard, S., and Mazel, D. (2005). Comparative study of class 1 integron and *Vibrio cholerae* superintegron integrase activities. *J. Bacteriol.* 187, 1740–1750. doi: 10.1128/JB.187.5.1740-1750.2005

Boucher, Y., Labbate, M., Koenig, J. E., and Stokes, H. W. (2007). Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol.* 15, 301–309. doi: 10.1016/j.tim.2007.05.004

Cambray, G., Guerout, A. M., and Mazel, D. (2010). "Integrons," in *Annu. Rev. Genet.* eds A. Campbell, M. Lichten, and G. Schupbach (Palo Alto: Annual Reviews), 141–166.

Deshpande, C. N., Harrop, S. J., Boucher, Y., Hassan, K. A., Di Leo, R., Xu, X., et al. (2011). Crystal structure of an integron gene cassette-associated protein from

Vibrio cholerae identifies a cationic drug-binding module. *PLoS ONE* 6:e16934. doi:10.1371/journal.pone.0016934

Elsaied, H., Stokes, H. W., Kitamura, K., Kurusu, K., Kamagata, Y., and Maruyama, A. (2011). Marine integrons containing novel integrase genes, attachment sites, attI, and associated gene cassettes in polluted sediments from Suez and Tokyo Bays. *ISME J.* 5, 1162–1167. doi:10.1038/ismej.2010.208

Elsaied, H., Stokes, H. W., Nakamura, T., Kitamura, K., Fuse, H., and Maruyama, A. (2007). Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents. *Environ. Microbiol.* 9, 2298–2312. doi: 10.1111/j.1462-2920.2007.01344.x

Elsaied, H., Stokes, H. W., Yoshioka, H., Mitani, Y., and Maruyama, A. (2013). Novel integrons and gene cassettes from a Cascadian submarine gas-hydrate-bearing core. *FEMS Microbiol. Ecol.* doi:10.1111/1574-6941.12227 [Epub ahead of print].

Franzon, V., L, Barker, A., and Manning, P. (1993). Nucleotide sequence encoding the mannose–fucose-resistant hemagglutinin of *Vibrio cholerae* O1 and construction of a mutant. *Infect. Immun.* 61, 3032–3037.

Gillings, M., Boucher, Y., Labbate, M., Holmes, A., Krishnan, S., Holley, M., et al. (2008). The evolution of class 1 integrons and the rise of antibiotic resistance. *J. Bacteriol.* 190, 5095–5100. doi: 10.1128/JB.00152-08

Guerin, E., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Da Re, S., et al. (2009). The SOS response controls integron recombination. *Science* 324, 1034. doi: 10.1126/science.1172914

Hunter, C. (2010). Integrons: evolution creates itself. Available at: http://darwins-god.blogspot.com.au/2010/02/integrons-evolution-creates-itself.html [accessed October 12, 2013].

Kim, Y. R., Lee, S. E., Kim, C. M., Kim, S. Y., Shin, E. K., Shin, D. H., et al. (2003). Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. *Infect. Immun.* 71, 5461–5471. doi: 10.1128/IAI.71.10.5461-5471.2003

Koenig, J. E., Boucher, Y., Charlebois, R. L., Nesbo, C., Zhaxybayeva, O., Bapteste, E., et al. (2008). Integron-associated gene cassettes in Halifax Harbour: assessment of a mobile gene pool in marine sediments. *Environ. Microbiol.* 10, 1024–1038. doi: 10.1111/j.1462-2920.2007.01524.x

Koenig, J. E., Bourne, D. G., Curtis, B., Dlutek, M., Stokes, H. W., Doolittle, W. F., et al. (2011). Coral-mucus-associated *Vibrio integrons* in the Great Barrier Reef: genomic hotspots for environment adaptation. *ISME J.* 5, 962–967. doi:10.1038/ismej.2010.193

Koenig, J. E., Sharp, C., Dlutek, M., Curtis, B., Joss, M., Boucher, Y., et al. (2009). Integron gene cassettes and degradation of compounds associated with industrial waste: the case of the Sydney tar ponds. *PLoS ONE* 4:e5276. doi:10.1371/journal.pone.0005276

Labbate, M., Boucher, Y., Chowdhury, P. R., and Stokes, H. W. (2011). Integration of a laterally acquired gene into a cell network important for growth in a strain of *Vibrio rotiferianus*. *BMC Microbiol.* 11:253. doi:10.1186/1471-2180-11-253

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

Labbate, M., Boucher, Y., Luu, I., Roy Chowdhury, P., and Stokes, H. W. (2012). Integron associated mobile genes: just a collection of plug in apps or essential components of cell network hardware? *Mob. Genet. Elem.* 2, 13–18. doi: 10.4161/mge.19510

Labbate, M., Case, R. J., and Stokes, H. W. (2009). The integron/gene cassette system: an active player in bacterial adaptation. *Methods Mol. Biol.* 532, 103–125. doi: 10.1007/978-1-60327-853-9_6

Lee, K. J., Kim, J. A., Hwang, W., Park, S. J., and Lee, K. H. (2013). Role of capsular polysaccharide (CPS) in biofilm formation and regulation of CPS production by quorum-sensing in *Vibrio vulnificus*. *Mol. Microbiol.* 90, 841–857. doi: 10.1111/mmi.12401

Leverstein-van Hall, M. A., Blok, H. E. M., Donders, R. T., Paauw, A., Fluit, A. C., and Verhoef, J. (2003). Multidrug resistance among *Enterobacteriaceae* is strongly associated with the presence of integrons and is independant of species or isolate origin. *J. Infect. Dis.* 187, 251–259. doi: 10.1086/345880

Leverstein-van Hall, M. A., Paauw, A., Box, A. T. A., Blok, H. E. M., Verhoef, J., and Fluit, A. C. (2002). Presence of integron-associated resistance in the community is widespread and contributes to multidrug resistance in the hospital. *J. Clin. Microbiol.* 40, 3038–3040. doi: 10.1128/JCM.40.8.3038-3040.2002

Li, B., Hu, Y., Wang, Q., Yi, Y., Woo, P. C., Jing, H., et al. (2013). Structural diversity of class 1 integrons and their associated gene cassettes in *Klebsiella pneumoniae* isolates from a hospital in China. *PLoS ONE* 8:e75805. doi:10.1371/journal.pone.0075805

Madiyarov, R. S., Bektemirov, A. M., Ibadova, G. A., Abdukhalilova, G. K., Khodiev, A. V., Bodhidatta, L., et al. (2010). Antimicrobial resistance patterns and prevalence of class 1 and 2 integrons in *Shigella flexneri* and *Shigella sonnei* isolated in Uzbekistan. *Gut Pathog.* 2, 18. doi: 10.1186/1757-4749-2-18

Mazel, D. (2006). Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.* 4, 608–620. doi: 10.1038/nrmicro1462

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

Michael, C. A., and Labbate, M. (2010). Gene cassette transcription in a large integron-associated array. *BMC Genet.* 11:82. doi:10.1186/1471-2156-11-82

Nemergut, D. R., Martin, A. P., and Schmidt, S. K. (2004). Integron diversity in heavy-metal-contaminated mine tailings and inferences about integron evolution. *Appl. Environ. Microbiol.* 70, 1160–1168. doi: 10.1128/AEM.70.2.1160-1168.2004

Nield, B. S., Holmes, A. J., Gillings, M. R., Recchia, G. D., Mabbutt, B. C., Nevalainen, K. M. H., et al. (2001). Recovery of new integron classes from environmental DNA. *FEMS Microbiol. Lett.* 195, 59–65. doi: 10.1111/j.1574-6968.2001.tb10498.x

Nield, B. S., Willows, R. D., Torda, A. E., Gillings, M. R., Holmes, A. J., Nevalainen, K. M. H., et al. (2004). New enzymes from environmental cassette arrays: functional attributes of a phosphotransferase and an RNA-methyltransferase. *Protein Sci.* 13, 1651–1659. doi: 10.1110/ps.04638704

Ogawa, A., and Takeda, T. (1993). The gene encoding the heat-stable enterotoxin of *Vibrio cholerae* is flanked by 123-base pair direct repeats. *Microbiol. Immunol.* 37, 607–616. doi: 10.1111/j.1348-0421.1993.tb01683.x

Partridge, S. R., Tsafnat, G., Coiera, E., and Iredell, J. R. (2008). Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol. Rev.* 33, 757–784. doi: 10.1111/j.1574-6976.2009.00175.x

Peymani, A., Farajnia, S., Nahaei, M. R., Sohrabi, N., Abbasi, L., Ansarin, K., et al. (2012). Prevalence of class 1 integron among multidrug-resistant *Acinetobacter baumannii* in Tabriz, Northwest of Iran. *Pol. J. Microbiol.* 61, 57–60.

Pier, G. B., Coleman, F., Grout, M., Franklin, M., and Ohman, D. E. (2001). Role of alginate O acetylation in resistance of mucoid *Pseudomonas aeruginosa* to opsonic phagocytosis. *Infect. Immun.* 69, 1895–1901. doi: 10.1128/IAI.69.3.1895-1901.2001

Rapa, R. A., Shimmon, R., Djordjevic, S. P., Stokes, H. W., and Labbate, M. (2013). Deletion of integron-associated gene cassettes impact on the surface properties of *Vibrio rotiferianus* DAT722. *PLoS ONE* 8:e58430. doi:10.1371/journal.pone.0058430

Rezaee, M. A., Langarizadeh, N., and Aghazadeh, M. (2012). First report of class 1 and class 2 integrons in multidrug-resistant *Klebsiella pneumoniae* isolates from Northwest Iran. *Jpn J. Infect. Dis.* 65, 256–259. doi: 10.7883/yoken.65.256

Robinson, A., Guilfoyle, A. P., Harrop, S. J., Boucher, Y., Stokes, H. W., Curmi, P. M., et al. (2007). A putative house-cleaning enzyme encoded within an integron array: 1.8. A crystal structure defines a new MazG subtype. *Mol. Microbiol.* 66, 610–621. doi: 10.1111/j.1365-2958.2007.05932.x

Robinson, A., Wu, P. S., Harrop, S. J., Schaeffer, P. M., Dosztanyi, Z., Gillings, M. R., et al. (2005). Integron-associated mobile gene cassettes code for folded proteins: the structure of Bal32a, a new member of the adaptable alpha + beta barrel family. *J. Mol. Biol.* 346, 1229–1241. doi: 10.1016/j.jmb.2004.12.035

Rowe-Magnus, D. A. (2009). Integrase-directed recovery of functional genes from genomic libraries. *Nucleic Acids Res.* 37, e118. doi: 10.1093/nar/gkp561

Rowe-Magnus, D. A., Guerout, A. M., Ploncard, P., Dychinco, B., Davies, J., and Mazel, D. (2001). The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc. Natl. Acad. Sci. U.S.A.* 98, 652–657. doi: 10.1073/pnas.98.2.652

Roy Chowdhury, P., Stokes, H. W., and Labbate, M. (2011). "Integrons: antibiotic resistance evolution and beyond," in *Bacterial Integrative Mobile Genetic Elements*, eds A. P. Roberts and P. Mullany (Austin, TX: Landes Biosciences), 53–69.

Salimizand, H., Shahcheraghi, F., Kalantar, E., and Badmasti, F. (2013). Molecular characterisation of class 1 integrons and gene cassetts in multidrug (MDR) *Klebsiella* spp. isolated from hospitalised and outpatients in Iran, 2009. *Iran J. Microbiol.* 5, 48–55.

Scholl, D., Adhya, S., and Merril, C. (2005). *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. *Appl. Environ. Microbiol.* 71, 4872–4874. doi: 10.1128/AEM.71.8.4872-4874.2005

Smith, A. B., and Siebeling, R. J. (2003). Identification of genetic loci required for capsular expression in *Vibrio vulnificus*. *Infect. Immun.* 71, 1091–1097. doi: 10.1128/IAI.71.3.1091-1097.2003

Stokes, H. W., and Gillings, M. R. (2011). Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiol. Rev.* 35, 790–819. doi: 10.1111/j.1574-6976.2011.00273.x

Stokes, H. W., and Hall, R. M. (1989). A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol. Microbiol.* 3, 1669–1683. doi: 10.1111/j.1365-2958.1989.tb00153.x

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

Thompson, F. L., Iida, T., and Swings, J. (2004). Biodiversity of vibrios. *Microbiol. Mol. Biol. Rev.* 68, 403–431. doi: 10.1128/MMBR.68.3.403-431.2004

Vu, B., Chen, M., Crawford, R. J., and Ivanova, E. P. (2009). Bacterial extracellular polysaccharides involved in biofilm formation. *Molecules* 14, 2535–2554. doi: 10.3390/molecules14072535

Westman, E. L., Preston, A., Field, R. A., and Lam, J. S. (2008). Biosynthesis of a rare di-*N*-acetylated sugar in the lipopolysaccharides of both *Pseudomonas aeruginosa* and *Bordetella pertussis* occurs via an identical scheme despite different gene clusters. *J. Bacteriol.* 190, 6060–6069. doi: 10.1128/JB.00579-08

World Health Organisation (WHO) (2013). *The Evolving Threat of Antimicrobial Resistance: Options for Action*. Geneva: WHO.

Yan, G., Bo, P., Hai Yin, W., Hai Jian, Z., Zhi Gang, C., and Biao, K. (2011). Structural variation of the superintegron in the toxigenic *Vibrio cholerae* O1 El Tor. *Biomed. Environ. Sci.* 24, 579–592. doi:10.3967/0895-3988.2011.06.001.

Yildiz, F. H., Liu, X. S., Heydorn, A., and Schoolnik, G. K. (2004). Molecular analysis of rugosity in a *Vibrio cholerae* O1 El Tor phase variant. *Mol. Microbiol.* 53, 497–515. doi: 10.1111/j.1365-2958.2004.04154.x

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

Received: 04 September 2013; accepted: 25 November 2013; published online: 09 December 2013.

Citation: Rapa RA and Labbate M (2013) The function of integron-associated gene cassettes in Vibrio species: the tip of the iceberg. Front. Microbiol. ***4****:385. doi: 10.3389/fmicb.2013.00385*

This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology.