Integrons: natural tools for bacterial genome evolution

Dean A Rowe-Magnus and Didier Mazel*

Integrons were first identified as the primary mechanism for antibiotic resistance gene capture and dissemination among Gram-negative bacteria. More recently, their role in genome evolution has been extended with the discovery of larger integron structures, the super-integrons, as genuine components of the genomes of many species throughout the γ-proteobacterial radiation. The functional platforms of these integrons appear to be sedentary, whereas their gene cassette contents are highly variable. Nevertheless, the gene cassettes for which an activity has been experimentally demonstrated encode proteins related to simple adaptive functions and their recruitment is seen as providing the bacterial host with a selective advantage. The widespread occurrence of the integron system among Gram-negative bacteria is discussed, with special focus on the super-integrons. Some of the adaptive functions encoded by these genes are also reviewed, and implications of integron-mediated genome evolution in the emergence of novel bacterial species are highlighted.

Addresses
Unité de Programmation Moléculaire et Toxicologie Génétique, Centre National de la Recherche Scientifique URA 1444, Département des Biotechnologies, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris, France.
URL: http://www.pasteur.fr/recherche/unites/pmtg/integ/
*Author for correspondence: Didier Mazel; e-mail: mazel@pasteur.fr

Current Opinion in Microbiology 2001, 4:565–569

1369-5274/01/$ – see front matter

Abbreviations

ORF	open reading frame
RI	resistant integron
SI	super-integron
VCR	*Vibrio cholerae* repeat

Introduction

The development of antibiotic resistance has led to the discovery of many natural mobile elements, including transposons and conjugative plasmids. Comparative sequence analysis of these elements ultimately led to the discovery of integrons — natural cloning and expression systems that incorporate open reading frames (ORFs) and convert them to functional genes (for reviews, see [1,2]). The functional integron platform consists of a gene (*intI*) that codes for an integrase of the tyrosine-recombinase family and a proximal primary recombination sequence called an *attI* site. The integrase mediates recombination between the *attI* site and a secondary target called an *attC* site (or 59-base element [59be]). The *attC* site is generally associated with a single ORF in a structure termed a gene cassette, and the gene cassette constitutes the mobile component of the system [3–6]. Insertion of the gene cassette at the *attI* site, which is located downstream of a resident promoter internal to the *intI* gene, drives expression of the encoded proteins. The integron platforms are defective for self-transposition, but they are often found associated with transposons and/or conjugative plasmids that can serve as vehicles for the intra- and interspecies transmission of genetic material. The Tn21 and Tn7 transposon families provide examples of this [7,8]. As such, they have been found in a variety of genetic contexts and among a large number of phylogenetically diverse Gram-negative and Gram-positive isolates.

Three classes of resistant integrons (RIs) have been defined on the basis of the divergence among their integrase genes, and each class appears to be able to acquire the same gene cassettes [9]. More than 70 different antibiotic resistance gene cassettes have been characterized in the three classes of RI [10] and most of their *attC* sites are unique. The lengths and sequences of the *attC* sites vary considerably (57–141 bp) and their similarities are primarily restricted to their boundaries, which correspond to the inverse core site (ICS), RYYYAAC, and the core site (CS), G↓TTRRRY (↓ denotes the recombination point) [5,11]. Studies by Mazel *et al.* [12] and Manning *et al.* [13] that examined the relationship between RI gene cassette arrays and the *Vibrio cholerae* repeat (VCR) cluster led to the discovery of the *Vibrio cholerae* super-integron (SI). This distinct type of integron is now known to be an integral component of many γ-proteobacterial genomes [14••].

In this review, we discuss the widespread occurrence of the integron system, with special focus on the SIs, among Gram-negative clinical and environmental bacteria. We also review some of the experimentally demonstrated adaptive functions encoded by these genetic reservoirs and highlight the implications of integron-mediated genome evolution in the emergence of novel bacterial species.

The super-integrons and their distribution

The integron discovered in chromosome 2 of *V. cholerae* has two characteristics that distinguish it from known RIs: the large number of cassettes that are gathered, and the high homology observed between the *attC* sites of these cassettes (the VCRs in the case of *V. cholerae*) [12]. These are the key features that define a SI. Both the *V. cholerae* SI and RIs possess specific and related integrases that are responsible for the insertion of ORFs into a unique chromosomal attachment site, leading to the formation of tandem arrays of genes. In the case of *V. cholerae*, the cluster of VCR-associated ORFs represents at least 216 unidentified genes in an array of 179 cassettes that starts from the *V. cholerae intIA* gene and occupies about 3% of the genome [15••]. The extent to which this system has impacted genome evolution has emerged with the discovery of SIs in several diverse proteobacterial genera. These include pathogens for man, other animals and plants, as well as non-pathogenic bacteria. SI structures have been

Table 1

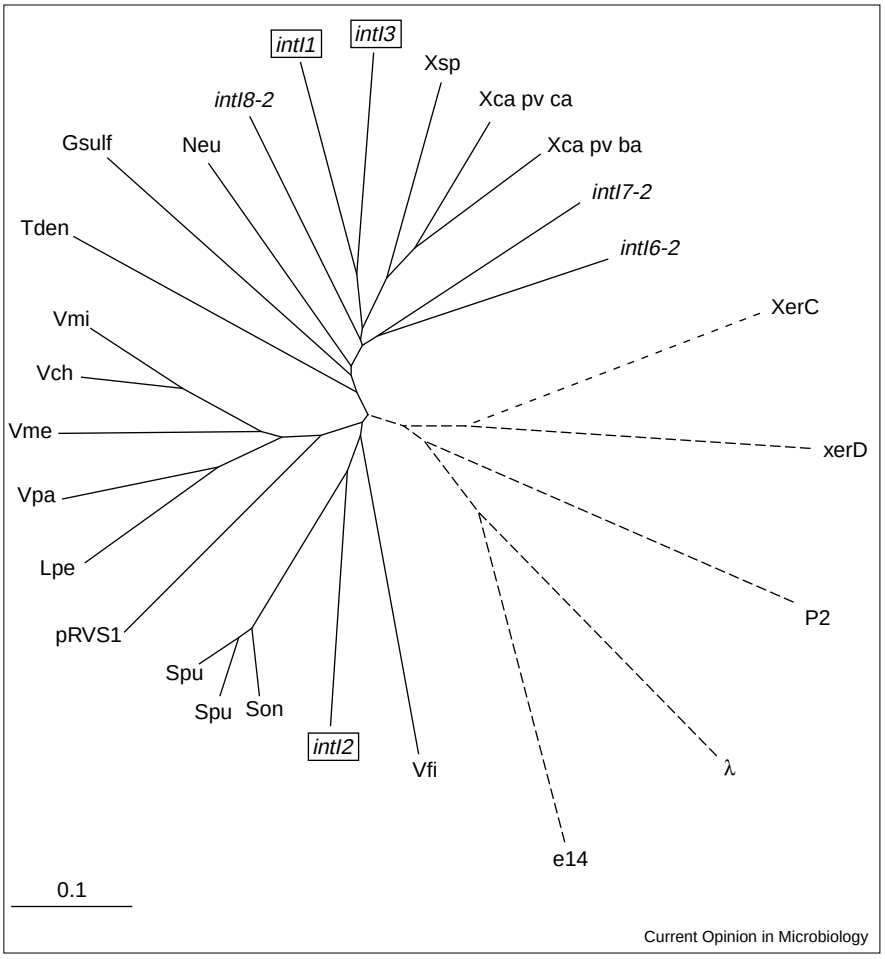
Bacterial species harbouring chromosomal super-integrons.

Strain	Radiation	Characteristics
Vibrionaceae and close relatives	γ-proteobacteria	
Vibrio cholerae		Etiological agent of cholera in humans
V. mimicus		Certain serogroups are enterotoxic human pathogens
V. metschnikovii		Certain serogroups are enterotoxic human pathogens
V. parahaemolyticus		Certain strains can cause seafood-borne gastroenteritis in humans
V. harveyi		Pathogen of black tiger prawns
V. hollisae		Certain strains can cause gastroenteritis in humans
V. anguillarium		Marine fish pathogen, causative agent of vibriosis (terminal haemorrhagic septicaemia)
V. salmonicida		Marine fish pathogen
V. fischeri		Non-pathogenic, luminescent bacterium
Listonella anguillarium		Marine fish pathogen
L. pelagia		Marine bacterium that produces tetrodotoxin
Alteromonas macleodii		Marine bacterium
Photobacterium phosphoreum		Luminescent bacterium
Moritella marina		Psychrophilic marine bacterium
Shewanella	γ-proteobacteria	Diverse genus of bacteria that are well known for their ability to utilize dissimilar compounds as electron acceptors and that figure prominently in fish spoilage and hull oxidation in the marine industry
Shewanella oneidensis		
S. putrefaciens		
Xanthomonads	γ-proteobacteria	Gram-negative phytopathogenic species responsible for disease in virtually all major taxa of plant life
Xanthamonas campestris pv. *campestris*		
X. campestris pv. *badrii*		
X. species 102397		
X. sp. 102336		
X. sp. 102338		
X. sp. 105155		
X. oryzae		
Pseudomonads	γ-proteobacteria	Include opportunistic pathogens
P. pseudoalcaligenes		
P. alcaligenes		
P. mendocina		
P. stutzeri		
Pseudomonas sp. NEB 376		
Acidithiobacillus ferroodoxans	γ-proteobacteria	Iron-acidizing bacteria
Nitrosomonas europaea	β-proteobacteria	Soil bacterium
Geobacter sulfurreducens	δ-proteobacteria	Fe(III)-reducing bacterium
Treponema denticola	Spirochaetales	Cause of periodontal disease

identified among the Vibrionaceae and their close relatives, the *Shewanella*, the Xanthomonads [14••,16•] and the Pseudomonads ([14••]; R Vaisvila, R Morgan, E Raleigh, unpublished data) (Table 1). They share the same general characteristics (a large size and a high homology between their endogenous cassette *attC* sites) and clearly pre-date the antibiotic era, as they are present in isolates from the last century [12]. Integron-integrase-like genes have also been identified in the genomes of other proteobacteria, including *Acidithiobacillus ferroodoxans*, *Nitrosomonas europaea*, *Geobacter sulfurreducens* and *Treponema denticola* (Table 1), but they have not been further characterized [14••,17•].

Using PCR primers directed against conserved regions of the integron-integrase genes and *attC* sites, Nield *et al.* [17•] were able to identify three new classes of integrons from four markedly different environmental DNA samples that had no known previous exposure to antibiotics. The protocol they used to retrieve the integron-integrase loci allowed recovery of the majority of the *intI* gene, the

Figure 1



Phylogenetic relationship of the integron *intI* genes among the proteobacteria. Unrooted dendrogram based on known *intI* gene sequences. The integrases from the three classes of RI are boxed. Organism abbreviations for the SIs are as follows: Vch, *Vibrio cholerae*; Vmi, *V. mimicus*; Vme, *V. metschnikovii*; Vpa, *V. parahaemolyticus*; Vfi, *V. fischeri*; Lpe, *Listonella pelagia*; Son, *Shewanella oneidensis*; Spu, *S. putrefaciens*; Xca pv ca, *Xanthomonas campestris* pathovar *campestris*; Xca pv ba, *Xanthomonas campestris* pathovar *badrii*; Xsp, *X. species*; Neu, *Nitrosomonas europaea*. The sources of *intI6-2*, *intI7-2*, *intI8-2* and the *intI* genes of plasmid pRVS1 (GenBank accession number AJ277063) are unknown. Classic representatives of the tyrosine recombinase family (dashed lines) are the integrases of phage λ, P2, e14 (an *E. coli* lambdoid phage) and the XerC and XerD recombinases of *E. coli*. The scale bar represents 10% divergence at the nucleotide level.

sequence covering the *attI* site and the first cassette up to the ICS of its associated *attC* site. Unfortunately, such a short sequence does not permit determination of the source of these integrons, be it the endogenous SI of a soil bacterium or an integron located in a mobile structure. However, the fact that integrons are widespread among bacterial populations either as components of mobile DNA elements or the chromosome, and that they are not confined to pathogenic or multidrug-resistant bacteria, support the hypothesis developed by the discovery of SIs, that is, that integrons function as a general gene-capture system in bacterial adaptation.

Integrons are ancient evolutionary apparatuses

All characterized integron-integrases clearly group together and form a specific clade within the tyrosine recombinase family (Figure 1). Furthermore, it has been noticed that all contained a specific stretch of 16 amino acids [17•] located between conserved patches II and III of the tyrosine recombinase family [18]. The role of this integron-integrase-specific sequence is unknown. The integron platform is undoubtedly ancient, as shown by the species-specific clustering of the respective SI integrase genes in a pattern that adheres to the line of descent among the bacterial species in which they are found (Figure 1) [14••]. Thus, the establishment of SIs likely pre-dates speciation within the respective genera, indicating that integrons are ancient structures that have been involved in the evolution of bacterial genomes for hundreds of millions of years. It is, however, possible that transfer of either a part or all of a SI occurred (from an Hfr-type strain, for example) during such a long period of evolution. This might be the origin of the discrepency observed in the SI-integrase and 16S rRNA gene trees for the *Vibrio fischeri* branching points among the other bacterial species [14••]. However, the phylogenetic analysis performed on the *rplT* genes from the same species, which encode the conserved ribosomal L20 protein, produced a branching order identical to the one found for the *intI* dendogram (see Figure 1). This supports the idea that the *V. fischeri* SI integrase gene has co-existed with the rest of the *V fischeri* genome.

Cassette functions

The SIs identified to date are collectively equal in size to a small genome, suggesting that the process of cassette genesis is constant and efficient. The majority of the cassettes examined thus far appeared to be unique to the

host species. Furthermore, most of their encoded genes have no counterparts in the database or the sole homologues are unassigned ORFs of viral, bacterial or eukaryotic origins, indicating their recruitment from all kingdoms of life [15••,19]. A precise inventory of the functions encoded by the cassettes remains to be established. However, a preliminary study indicates that many of the SI cassettes encode adaptive functions, in a broad sense, beyond pathogenicity and antibiotic resistance. In *V. cholerae*, three pathogencity genes (the heat-stable toxin gene *sto* [20], the mannose-fucose-resistant haemagglutinin gene *mrhA* and the gene *mrhB*, which is in the same operon as *mrhA*) [21] as well as a lipoprotein gene have been found to be cassette-encoded [22]. We have determined the metabolic function of three SI cassettes: a sulfate-binding protein in a *V. cholerae* SI cassette, a psychrophilic lipase in a *Moritella marina* SI cassette, and a restriction enzyme (*Xba* I) and its cognate methylase in a *Xanthomonas campestris* pathovar *badrii* SI cassette [14••]. Genes with homology to DNA methylases, immunity proteins, restriction endonucleases, dNTP triphophohydrolases, periplasmic sulphate-binding proteins, lipases and 8-oxoguanine triphosphatases (MutT), among others, have been found [14••,15••,16•,19]. Although a known antibiotic resistance gene cassette has not yet been identified within a SI, several potential progenitor cassettes with significant homology to aminoglycoside, phosphinotricin, fosfomycin, streptothricin and chloramphenicol resistance genes are present. The determination of the metabolic activities of several SI cassettes, whose activities are not related to antibiotic resistance or virulence, confirms that integrons operate as a general gene capture system in bacterial adaptation [14••,22]. If each bacterial species harboring a SI has its own cassette pool, the resource in terms of gene cassette availability will be immense and the functions of the encoded genes have fantastic potential from both genetic and biotechnological standpoints.

Intraspecific cassette content variations

The activity of integron cassettes offers a fast track to bacterial innovation. The sizes of SIs and the ancient and dynamic nature of the system is a reminder that the cassettes that currently occupy SIs represent only a fraction of those that may have participated in the evolution of the host, as the cassettes will presumably be subject to episodic selection. The more than 165 different O serotypes of *V. cholerae* are represented by species of ecological, geographical and temporal diversity. Thus, comparison of SI organization from recent and earlier isolates as well as between recent isolates from different geographical locations and ecological niches may yield valuable information. Clark *et al.* [16•] examined the global SI organization of 65 different *V. cholerae* O serotypes by PCR and Southern hybridization. Extensive restriction polymorphism was observed even among closely related isolates, suggesting a plasticity in these structures and in their microevolution through integrase-mediated gene acquisition, gene loss [23] and gene cassette rearrangement events [6,24,25]. With regard to cassette rearrangements, an important question would be the following: can cassettes be mobilized in clusters? The SI organization of two *V. cholerae* strains suggests that they can. The cassettes in positions 1–4 of the SI of *V. cholerae* strain 569B were found to be in the same order as in the SI of *V. cholerae* strain N16961 [15••], but they occupied positions 79–82 (a displacement of more than 40 Kb) [19]. This is also observed in other cassettes [16•]. These observations may represent a true group mobilization event or simply temporal differences in cassette acquisition.

There is evidence that not all repeated sequences are equally functional. Some are known to contain mutations or deletions within the CS that could render them non-functional (DA Rowe-Magnus, A-M Guerout, D Mazel, unpublished data). Therefore, their movement would have to be co-ordinated with those of other cassettes. Collis and Hall [26] demonstrated that integron gene cassettes are excised as covalently closed circles, and observed differences in the resulting recombination products. Some cassettes could be mobilized as individual units, whereas others were only excised in tandem with another cassette. It is not known if this cassette hitch-hiking is designed to ensure simultaneous transmission of genes, or is just a matter of coincidence.

Conclusions

The integron system is remarkably versatile in its ability to recognize highly variable target recombination sequences and its apparently limitless capacity to exchange and stockpile cassettes. Such flexibility permits rapid adaptation to the unpredictable flux of environmental niches by allowing bacteria to scavenge foreign genes that may ultimately endow increased fitness to the host. Likewise, genes that fail to provide a meaningful function may be readily eliminated. It is also quite likely that many of the cassettes that presently occupy the SI are not expressed, but they may nevertheless persist in the absence of selective pressure and provide a genetic basis for the evolution and subsequent retention of novel functions. In addition to the plethora of antibiotic resistance genes, two virulence genes of *V. cholerae* are also structured as gene cassettes, underscoring the potential of this system to participate in the establishment of pathogenicity islands. It is conceivable that any ORF can be structured as a gene cassette and it is vital to decipher the mechanism governing cassette genesis. According to the guanine + cytosine (GC) content and codon usage differences observed in the ORFs found in the cassettes, they must have many different origins, whereas their associated recombination elements are highly homologous (e.g. 74% nucleotide identity for the VCRs). This last characteristic suggests that the VCRs were added to the ORFs inside the *Vibrio* cell. Therefore, it is very likely that the capture process occurs *in vivo*, but the nature of this process remains unknown.

As such, integron-driven gene capture is likely to be an important factor in the more general process of horizontal gene transfer in the evolution of bacterial genomes. It

appears that multiresistant integrons have evolved from SIs through entrapment of *intI* genes and their cognate *attI* sites into highly mobile structures like transposons. The combination of this mobility and the selection pressure exerted by antibiotic use may have driven the specific capture of resistance cassettes from the many different kinds of SI cassette pools through multiple lateral transfers.

Acknowledgements

We are very grateful to Maurice Hofnung for his constant support and interest, and to Anne-Marie Guerout and Latefa Biskri for their helpful discussions. DRM is a European Molecular Biology Organization (EMBO) Post-doctoral fellow. This work was supported by the Institut Pasteur, the Centre National be la Recherche Scientifique and the Programme de Recherche Fondamentale en Microbiologie et Maladies Infectieuses et Parasitaires of the Ministère de l'Education Nationale, de la Recherche et de la Technologie (MERNT).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Hall RM, Collis CM: Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol Microbiol* 1995, 15:593-600.

2. Rowe-Magnus DA, Mazel D: Resistance gene capture. *Curr Opin Microbiol* 1999, 2:483-488.

3. Hall RM, Stokes HW: Integrons: novel DNA elements which capture genes by site-specific recombination. *Genetica* 1993, 90:115-132.

4. Recchia GD, Hall RM: Gene cassettes: a new class of mobile element. *Microbiology* 1995, 141:3015-3027.

5. Stokes HW, O'Gorman DB, Recchia GD, Parsekhian M, Hall RM: Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol Microbiol* 1997, 26:731-745.

6. Sundstrom L: The potential of integrons and connected programmed rearrangements for mediating horizontal gene transfer. *APMIS Supplementum* 1998, 84:37-42.

7. Liebert CA, Hall RM, Summers AO: Transposon Tn21, flagship of the floating genome. *Microbiol Mol Biol Rev* 1999, 63:507-22.

8. Sundstrom L, Roy PH, Skold O: Site-specific insertion of three structural gene cassettes in transposon Tn7. *J Bacteriol* 1991, 173:3025-3028.

9. Hall RM, Collis CM: Antibiotic resistance in gram-negative bacteria: the role of gene cassettes and integrons. *Drug Res Updates* 1998, 1:109-119.

10. Mazel D, Davies J: Antibiotic resistance in microbes. *Cell Mol Life Sci* 1999; 56:742-754.

11. Collis CM, Kim MJ, Stokes HW, Hall RM: Binding of the purified integron DNA integrase IntI1 to integron- and cassette-associated recombination sites. *Mol Microbiol* 1998, 29:477-490.

12. Mazel D, Dychinco B, Webb VA, Davies J: A distinctive class of integron In the ***Vibrio cholerae*** genome. *Science* 1998, 280:605-608.

13. Barker A, Clark CA, Manning PA: Identification of VCR, a repeated sequence associated with a locus encoding a hemagglutinin in ***Vibrio cholerae*** O1. *J Bacteriol* 1994, 176:5450-5458.

14. •• Rowe-Magnus DA, Guerout AM, Ploncard P, Dychinco B, Davies J, Mazel D: The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc Natl Acad Sci USA* 2001, 98:652-657.
The authors show that chromosomal super-integrons are not restricted to the Vibrionaceae. Equivalent superstructures are identified in various genera throughout the proteobacterial radiation and are suggested to be the source of contemporary mobile integrons.

15. •• Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ *et al.*: DNA sequence of both chromosomes of the cholera pathogen ***Vibrio cholerae***. *Nature* 2000, 406:477-483.
The true size of the *V. cholerae* super-integron is discovered upon completion of the genome sequence. Clustered in one region of chromosome II, this structure spans 126 Kb and harbours 214 ORFs of mainly unassigned function in 179 cassettes, dwarfing the largest described RIs.

16. • Clark CA, Purins L, Kaewrakon P, Focareta T, Manning PA: The ***Vibrio cholerae*** O1 chromosomal integron. *Microbiology* 2000, 146:2605-2612.
In this paper, a global comparison of the SIs from three *V. cholerae* strains reveals that even closely related isolates exhibit extensive restriction polymorphism within their respective SI structures, suggesting that SIs are dynamic structures. This study is one of the first to address the flexible nature of these structures and the microevolution occurring within SIs.

17. • Nield BS, Holmes AJ, Gillings MR, Recchia GD, Mabbutt BC, Nevalainen KM, Stokes HW: Recovery of new integron classes from environmental DNA. *FEMS Microbiol Lett* 2001, 195:59-65.
This is the first isolation of integrons from soil samples not known to have had previous exposure to antibiotics, and underscores the prevalence of integrons among environmental as well as clinical bacterial populations.

18. Nunes-Duby SE, Kwon HJ, Tirumalai RS, Ellenberger T, Landy A: Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res* 1998, 26:391-406.

19. Rowe-Magnus DA, Guerout AM, Mazel D: Super-integrons. *Res Microbiol* 1999, 150:641-651.

20. Ogawa A, Takeda T: The gene encoding the heat-stable enterotoxin of ***Vibrio cholerae*** is flanked by 123-base pair direct repeats. *Microbiol Immunol* 1993, 37:607-616.

21. van Dongen WMAM, van Vlerken MMA, De Graaf FK: Nucleotide sequence of a DNA fragment encoding a ***Vibrio cholerae*** haemagglutinin. *Mol Gen (Life Sci Adv)* 1987, 6:85-91.

22. Barker A, Manning PA: VlpA of Vibrio cholerae O1: the first bacterial member of the alpha 2-microglobulin lipocalin superfamily. *Microbiology* 1997, 143:1805-1813.

23. Takeda T, Peina Y, Ogawa A, Dohi S, Abe H, Nair GB, Pal SC: Detection of heat-stable enterotoxin in a cholera toxin gene-positive strain of ***Vibrio cholerae*** O1. *FEMS Microbiol Lett* 1991, 64:23-27.

24. Collis CM, Hall RM: Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. *J Bacteriol* 1992, 174:1574-1585.

25. Collis CM, Recchia GD, Kim MJ, Stokes HW, Hall RM: Efficiency of recombination reactions catalyzed by class 1 integron integrase IntI1. *J Bacteriol* 2001, 183:2535-2542.

26. Collis CM, Hall RM: Gene cassettes from the insert region of integrons are excised as covalently closed circles. *Mol Microbiol* 1992, 6:2875-2885.