Published online 13 July 2009 *Nucleic Acids Research, 2009, Vol. 37, No. 17* ***e118***
doi:10.1093/nar/gkp561

Integrase-directed recovery of functional genes from genomic libraries

Dean A. Rowe-Magnus[1,2,*]

[1]Division of Clinical Integrative Biology, Sunnybrook Health Sciences Centre, 2075 Bayview Avenue, S1-26A, Toronto, Ontario M4N 3N5 and [2]Department of Laboratory Medicine & Pathobiology, Faculty of Medicine, University of Toronto, Toronto, Canada

Received March 9, 2009; Revised June 12, 2009; Accepted June 16, 2009

ABSTRACT

Large population sizes, rapid growth and 3.8 billion years of evolution firmly establish microorganisms as a major source of the planet's biological and genetic diversity. However, up to 99% of the microorganisms in a given environment cannot be cultured. Culture-independent methods that directly access the genetic potential of an environmental sample can unveil new proteins with diverse functions, but the sequencing of random DNA can generate enormous amounts of extraneous data. Integrons are recombination systems that accumulate open reading frames (gene cassettes), many of which code for functional proteins with enormous adaptive potential. Some integrons harbor hundreds of gene cassettes and evidence suggests that the gene cassette pool may be limitless in size. Accessing this genetic pool has been hampered since sequence-based techniques, such as hybridization or PCR, often recover only partial genes or a small subset of those present in the sample. Here, a three-plasmid genetic strategy for the sequence-independent recovery of gene cassettes from genomic libraries is described and its use by retrieving functional gene cassettes from the chromosomal integron of *Vibrio vulnificus* ATCC 27562 is demonstrated. By manipulating the natural activity of integrons, we can gain access to the caches of functional genes amassed by these structures.

INTRODUCTION

Integrons are natural gene cloning and expression systems that incorporate circularized open reading frames, called gene cassettes, and convert them to functional genes (1–3). The most notable integron gene cassettes code for antibiotic resistance determinants (4,5) and the stockpiling of these cassettes can produce what are often referred to as resistance integrons (RIs), which contain several antibiotic resistance genes in tandem. Five classes of RIs have been defined based on the homology of the integrase genes (6–12) and each is associated with a mobile DNA element. The paradigm for integron function is the class 1 integron (13). The integron platform consists of an integrase gene and an adjacent primary recombination site (*attI*). The integrase is a member of the tyrosine recombinase superfamily (14) and catalyses recombination between its cognate *attI* site and a secondary target called an *attC* site. The *attC* site is normally found associated with a single open reading frame (ORF), and the ORF–*attC* structure is termed a gene cassette (15). Integron gene cassettes are compact structures. They typically consist of the ORF, the *attC* site and little noncoding sequence (16). Consequently, most gene cassettes are promoterless. However, recombination between the *attI* and *attC* sites leads to insertion of the gene cassette downstream of a resident promoter, P_C, within the integron that drives expression of the encoded product. The strongest variant of P_C constitutes a perfect *Escherichia coli* consensus promoter and is six times more efficient than the de-repressed P_{Tac} promoter (17). The integron platforms are sedentary, but the common association of integrons with mobile DNA elements such as transposons and conjugative plasmids facilitates the transit of their amassed genes across phylogenetic boundaries, and augments the impact of integrons on bacterial evolution (2,18,19). The proficiency of this partnership is confirmed by the marked differences in codon usage among cassettes within the same RI, indicating that the captured genes are of diverse origins (20–23).

The ancestors of RIs are the chromosomal integrons (CIs) (24,25). The first was discovered in the *Vibrio cholerae* genome (26). Located on the smaller of the two circular chromosomes, this element spanned 126 kb and gathered at 179 cassettes into a single structure (24,27), dwarfing previously identified RIs. CIs have since been identified throughout the proteobacterial phylum through both systematic searches (25,28–30) and large-scale microbial genome sequencing efforts (27,31–33). CIs can span

*To whom correspondence should be addressed. Tel: +1 416 4806100; Fax: +1 416 4805737; Email: dean.rowe-magnus@sri.utoronto.ca

© 2009 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

hundreds of kilobases and harbour hundreds of gene cassettes. The CIs of the Vibrionaceae are among the largest identified to date, with those from the clinical *V. vulnificus* strains CMCP6 and YJ016 stretching 153 kb and 138 kb long, and containing 219 and 188 gene cassettes, respectively (34).

Just five *Vibrio* CIs contain a gene cassette reservoir that is equivalent in size to a small genome (35,36). Furthermore, PCR on environmental DNA samples with degenerate primers targeting the conserved regions of *attC* sites (37) has led to the recovery of novel integrase genes and hundreds of diverse gene cassettes (38–40). Importantly, accumulation analysis of the gene cassettes showed no signs of saturation (39), supporting the notion that the integron gene cassette pool is likely immense, and that the process of gene cassette genesis is constant and efficient. Although, the vast majority of the thousands of gene cassettes identified in CIs and environmental samples thus far have no counterparts in the database, or the sole homologues are unassigned ORFs, several have been shown to code for proteins that adopt characteristic three-dimensional structures and assemblies indicative of biological activity (41). The gene cassettes for which an activity has been experimentally demonstrated encoded proteins related to simple adaptive functions, including – metabolism (25,42), virulence (43,44) and antibiotic resistance (45–47). Others appear to be of viral, bacterial or eukaryotic origin, indicating their recruitment from all kingdoms of life (38–40,48). Thus, integron gene cassettes are an important source of new enzymes and novel proteins.

Difficulties in sampling or recognizing integron gene cassettes have limited our capacity to exploit this genetic resource. Although useful, a PCR-based approach for the sequence-independent retrieval of functional gene cassettes from environmental samples is limited by the diversity of the *attC* sites, most of which are unique in sequence and length (25,49). This will undoubtedly preclude the detection of many of the gene cassettes that may be present in a given sample. CI cassettes have been demonstrated to be substrates for recombination by the class 1 integrase of RIs (26,33,45,50,51). Here, the ability of the class 1 integron-integrase, IntI1, to recognize diverse *attC* target sites is exploited to develop an integron-based genetic selection system to recover functional gene cassettes from genomic libraries.

MATERIALS AND METHODS

Strains, plasmids and media

The *E. coli* strains, DH10B and EPI300, were from Invitrogen and EpiCentre, respectively. *Vibrio vulnificus* 75.4T (ATCC 27562), *Vibrio cholerae* N16961, *Xanthomonas campestris pv. campestris* 100069T (ATCC 33913) and *Shewanella oneidensis* 106686 (ATCC 700550; MR-1) were obtained from the Pasteur Institute (Paris, France). All strains were grown in LB media (Sigma). The class 1 integron-integrase gene, *int1*, was amplified with Platinum Pfx polymerase (Invitrogen) using primers intI1BspHI (5′-ggaattctcatgaaaaccgccactgcgcc) and intI1XbaI (5′-ctagtctagactacctctcactagtgaggg) and cloned into the NcoI and XbaI sites of pTRC99A. R388 and its In*3* integron have been previously described (5,52,53) and the pCC1FOS fosmid vector was from EpiCentre. Plasmid DNA purifications were performed using a Sigma Plasmid Miniprep and Midiprep kits. Antibiotics were used at the following concentration when indicated: ampicillin (Ap), 100 µg/ml; rifampicin (Rf), 100 µg/ml; chloramphenicol (Cm), 10 µg/ml; kanamycin (Km), 25 µg/ml; tetracycline (Tc), 15 µg/ml; IPTG 1 mM; L-arabinose (L-ara), 0.02%.

Overall strategy of the three-plasmid integron selection system

The overall strategy for the selection system is depicted in Figure 1. The system consists of three plasmids: p1 expresses the class 1 integron-integrase gene, *intI1*; p2 is a self-transmissible recipient plasmid that contains a class 1 integron; p3 is the substrate plasmid into which genomic DNA is cloned and some of the genomic DNA will contain integron gene cassette arrays. The genomic library is constructed in a specialized *E. coli* host strain (βDH10B) that allows for easy antibiotic-independent counter-selection by the simple omission of a specific nutrient from the media. p2 is transferred by conjugation to the *E. coli* strain containing the library. The transconjugants are pooled and transformed with p1 to complete assembly of the system. Integrase-mediated recombination between the *attI1* site of p2 and target *attC* sites on p3 plasmids harbouring integron gene cassettes fuses these plasmids together, and the fusions are selected following conjugation to a second *E. coli* strain and plating on the appropriate selective media that lacks the specific nutrient.

Constructing the βDH10B host strain

In *E. coli* strain β2155, the *dapA* gene is disrupted by a construct containing an erythromycin (*erm*) cassette and the *pir* gene (54). The disrupted gene was amplified by PCR with primers DAP1 (5′-atgttcacgggaagtattgtc) and DAP2 (5′-cagcaaaccggcatgcttaa) and purified. *E. coli* DH10B was transformed with pKOBEGA, which carries the λ red (*redαβγ*) recombination system under the control of the arabinose-inducible P_{BAD} promoter (55). This system can be used to generate gene knockouts in *E. coli* by allelic exchange. DH10B/pKOBEGA cells were made electrocompetent following arabinose induction and were transformed with the linear DAP1/DAP2 PCR product to create strain βDH10B. Transformants were selected on LB Erm plates containing diaminopimelic acid (DAP). Disruption of the *dapA* gene was conformed by PCR and the strains inability to grow on plates lacking DAP.

Fosmid library construction

Genomic DNA was prepared using the DNAzol genomic DNA isolation reagent (Invitrogen) as per the manufacturer's instructions. The DNA was then sheared by passage through a Hamilton syringe, end-repaired and ligated into Eco72I-digested pCC1FOS using the CopyControl Fosmid Library Production Kit (EpiCentre). The ligation was then transformed into the βDH10B strain and

colonies were selected on LB DAP Cm plates. Alternatively, sheared DNA was end-repaired and separated on a 1% low melting point PFGE gel. Gel slices containing fragments of 40 kb were excised and digested with gelase (InVitrogen) to recover the DNA. The DNA was concentrated by ethanol precipitation when necessary. The 40-kb DNA fragments were then ligated into Eco72I-digested pCC1FOS. MaxPlax lambda packaging extracts (Epicentre) were used to package the library and transduce βDH10B cells. Transductants were selected on LB DAP Cm plates at 37°C. The colonies were grown overnight at 37°C. To determine how many clones were required in order to obtain a given genome sequence with 99% probability, I used the equation $N = \ln(1-P)/\ln(1-I/GS)$, where N = number of clones, P = probability, I = insert size and GS = genome size (56). Six hundred clones of a 40-kb insert genomic library were assembled. Screening on X-gal indicated that >90% of the clones were recombinant.

Integrase-mediated recombination of fosmid clones containing gene cassettes

The R388ISceKT plasmid was conjugated from *E. coli* DH5α cells to the βDH10B library. βDH10B transconjugants were selected on LB DAP Cm Km plates at 37°C. These cells were then pooled and transformed with pTRC99A::*intI1* and transformants were selected on LB DAP Cm Km Ap plates at 37°C. The specific recovery of fosmid clones containing CI gene cassettes from the library was monitored as follows. The cells were pooled and incubated overnight in LB liquid media containing DAP Cm Km Ap and IPTG at 37°C with shaking. The next day, recombinant plasmids were isolated by conjugation to *E. coli* EPI300 cells and selection of transconjugants on LB Cm plates that lacked DAP. Higher yield of recombinant plasmids for PFGE, sequencing and activity screening was obtained by growing the transconjugants in media containing the induction solution that increases expression of *trfA* in the EPI300 strain.

Sequencing and protein analysis

Fusion plasmids were used in PCR reactions with primers intI1 and VVR1, VCR1, XCR1 or SPR1-3. The intI1 primer anneals within *intI1* and VVR1, VCR1 and XCR1 hybridize to the signature *attC* sites of *V. vulnificus*, *V. cholerae* and *X. campestris*, respectively (25,28). SPR1, SPR2 and SPR3 hybridize to three of the *attC* sites from the CI of *S. oneidensis* (27,52). PCR was conducted as follows: 94°C, 30 s; 60°C, 30 s; 72°C, 2 min, for 30 cycles with *Pfu* polymerase. Neither primer alone yielded a PCR product, nor did PCR with intI1 and the respective second primer on uninduced cultures. The amplified fragments were sequenced directly with the intI1 primer to confirm integration at the *attI* site. Sequencing was done at The Centre for Applied Genomics (TCAG, Hospital for Sick Children, Toronto). Homology searches were conducted using BLAST analysis (57), and protein parameters and domain predictions were performed with ProtParam (58) and the PROSITE research tool from the Expert Protein analysis system (59). Protein structure predictions were made using Phyre (60). The 35-kb CI region of *V. vulnificus* ATCC 27562 was sequenced at The Joint Genome Institute (JGI, University of California) and was deposited at GenBank (accession no. GQ292873).

RESULTS

The three-plasmid integron selection system

Based on the strategy outlined in Figure 1, a three-plasmid retrieval system was developed for the sequence-independent recovery of complete, functional open reading frames from CIs (Figure 2). The first plasmid, pTRC99A::*intI1*, expresses the integron-integrase gene, *intI1*, from the IPTG-inducible P_{TRC} promoter (61). The second, R388ISceKT, is the recipient plasmid. It is a Km- and Tc-resistant derivative of the 34-kb self-transmissible plasmid R388, which carries an *attI1* site within its class 1 integron, In*3* (52). In*3* contains two genes: *dfrB2*, which confers trimethoprim resistance, and *orfA*, of unknown function (62). However, *dfrB2* is the first gene cassette of In*3*. Recombination of an *attC* site with the *attI1* site would displace the promoterless *dfrB2* from the Pc promoter, potentially leading to trimethoprim sensitivity due to a lack of *dfrB2* expression. Furthermore, a defined media that lacks any thymine derivatives must be used when selecting with trimethoprim. The addition of the *aph* gene permits selection in complex media and avoids the potential problem of clone loss that can occur during growth of genomic libraries in defined media. The plasmid also includes a Tc resistance marker and I-SceI sites, which aid in the isolation of the insert. The third, pCC1FOS, is the substrate plasmid into which genomic DNA is cloned. pCC1FOS is an 8-kb fosmid that carries an F-factor origin of replication (*ori2*) and partitioning functions that maintain the fosmid in single copy. This circumvents the problem of under-representation of clones carrying genes that may be toxic when present in multiple copies. A second TrfA-dependent origin of replication, *oriV*, is also present on the fosmid. Genomic fragments of any size can be cloned into the unique Eco72 I (blunt end) site within *lacZα* and recombinants are selected by blue/white screening. If size selection is desired, the fosmid also contains cos sites for packaging into phage particles *in vitro*, ensuring that the library is comprised only of plasmids containing inserts of ∼40 kb under these conditions. The host strain for maintaining libraries is a derivative of the *E. coli* strain DH10B that has a disruption in *dapA*, a component of the diaminopimelic acid (DAP) biosynthetic pathway. DAP is incorporated into the bacterial cell wall peptidoglycan. *E. coli* mutants altered in the DAP pathway require exogenous DAP for growth. Use of this strain (βDH10B) as a donor in conjugation experiments allows for absolute counter-selection in an antibiotic-independent manner even in complex media such as LB, making it easy to isolate transconjugants. βDH10B has a high transformation efficiency for both large and small clones, readily accepts large DNA for the construction of large-insert genomic libraries, and is *endA1* and *recA1* to ensure high yields of DNA and greater stability of large cloned inserts.

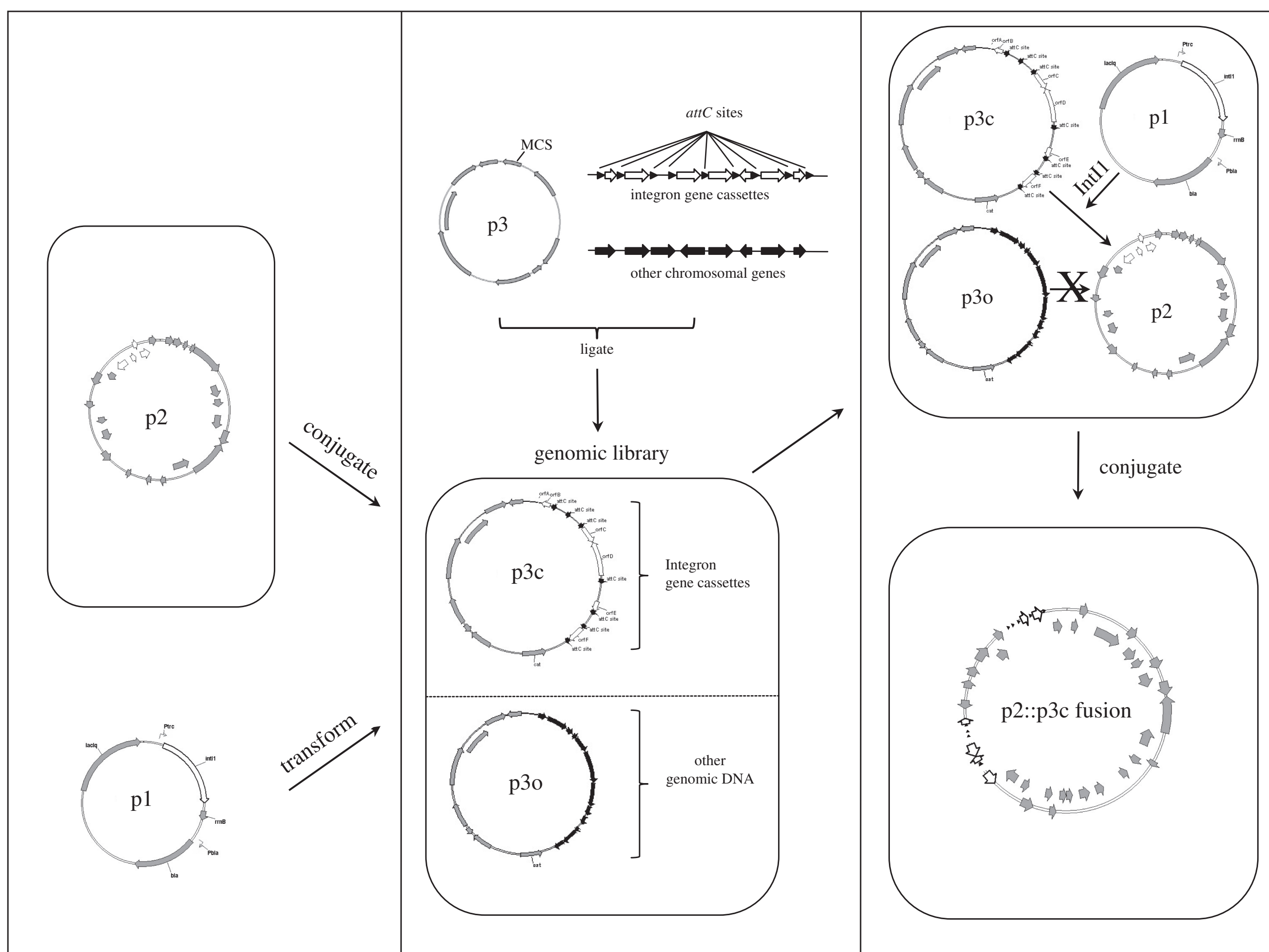

Figure 1. Strategy of the three-plasmid integron selection system. The system consists of three plasmids: p1 expresses the class 1 integron-integrase gene, *intI1* (white arrow); p2 is a self-transmissible recipient plasmid that contains a class 1 integron (white arrows); p3 is the substrate plasmid into which genomic DNA is cloned. A genomic library is constructed in p3 (middle panel), some of which will contain integron gene cassettes (p3c; white arrows, *orfs*; black arrowheads, *attC* sites). The rest will contain other chromosomal DNA (p3o; black arrows, *orfs*). A specialized *E. coli* strain that allows for easy antibiotic-independent counter-selection by the simple omission of a specific nutrient from the media is used as the host strain. p2 is transferred by conjugation (top left panel) to the *E. coli* strain containing the library. The transconjugants are pooled and transformed with p1 (bottom left panel) to complete assembly of the system (top right panel). Integrase-mediated recombination between the *attI1* site of p2 and target *attC* sites on p3c fuses these plasmids together, while p3o plasmids are not recombined. The fusions are selected following conjugation to a second *E. coli* strain and plating on the appropriate selective media that lacks the specific nutrient (bottom right panel).

The expression of *intI1* catalyzed recombination between the *attI1* site of In*3* on R388ISceKT and pCC1FOS fosmids from the library that contained inserts with at least one functional *attC* site, resulting in fusion of the two plasmids. Since the mobilizable R388ISceKT plasmid was Km^R, and the Cm^R pCC1FOS fosmid did not contain an origin of conjugative transfer, the recombinant plasmids were easily isolated by conjugation to a recipient *E. coli* strain (EPI300-T1^R) and selection on LB Km Cm plates that lacked DAP. EPI300-T1^R cells are *endA1* and *recA1*, and carry a copy-up allele of the *trfA* gene integrated into the chromosome under control of an inducible promoter. Induced expression of *trfA* allows for replication from *oriV* of pCC1FOS. The resulting 40–80-fold increase in the fosmid copy number provided for better fosmid yields for sequencing of the gene cassettes and aided activity-based screening of gene cassettes expressed from P_C of In*3*. If desired, the integrated fosmid of the recombinant plasmid could be recovered by digestion with I-SceI, self-ligation, transforming EPI300 cells and selecting clones on LB Cm plates. The fosmid harbouring the chromosomal fragment can replicate via the *oriV* of pCC1FOS.

Recovery of functional genes from a genomic library with the integron-based recombination system

The 153-kb CI of *V. vulnificus* strain CMCP6 represents 3% of its genome (34). Genomic DNA from strain CMCP6 was ligated into pCC1FOS and the library was either directly transformed into βDH10B or packaged into phage and transduced into the host strain. R388IsceKT and pTRC99A::*intI1* were then moved to these strains to complete assembly of the three-plasmid system. The expression of *intI1* was induced with IPTG and recombinant plasmids were retrieved by conjugation to EPI300-T1^R recipient cells. A representative gel of several random recombinant plasmids from the transformed and

Figure 2. The three-plasmid integron selection system. An example of the βDH10B host strain, which requires exogenous DAP for growth, carrying a pCC1FOS library clone containing gene cassettes is shown. The same strain also carries the pTRC99A::*intI1* and R388ISceKT plasmids. IPTG-induced expression from pTRC99A::*intI1* leads to integrase-mediated recombination between pCC1FOS plasmids containing *attC* sites and the *attI1* site of the class 1 integron on the conjugative R388 plasmid. The recombinant plasmid is specifically selected by conjugation to the EPI300-T1^R recipient and plating on media containing chloramphenicol but lacking DAP. The original class 1 integron of R388 is disrupted by the fusion. Small black arrows denote primers that can be used to sequence the gene cassettes of the recovered recombinant plasmids.

transduced libraries is shown in Figure 3. The wider size range for recombinant plasmids obtained with the transformed library relative to the more uniform plasmid size obtained from the transduced library was reflective of the size-selection imposed by packaging of the ligation reaction into phage particles prior to transduction. Restriction digests of randomly chosen recombinant plasmids suggested that they were indeed the products of recombination between the pCC1FOS library and R388 (data not shown). Specific recombination between the *attC* site of the substrate plasmid (pCC1FOS library) and the *attI1* site of recipient plasmid (R388ISceKT) was verified by PCR and sequencing.

One of the sequenced recombinant clones from the transformed pCC1FOS library contained a 5383-bp insert with 7 *attC* sites and corresponded to genes *vv12525*-*vv12533* of strain CMCP6 (Figure 4). One of the partial gene cassettes contained two small overlapping *orfs* (*vv12525* and *vv12526*) that coded for a toxin-antitoxin system that we previously characterized from *V. vulnificus* strain 27562 (63). Toxin/Antitoxin (TA) systems encode a stable toxin and a specific but unstable antitoxin (64–66). They enhance plasmid maintenance by preventing the proliferation of plasmid-free progeny because the continued *de novo* synthesis of the short-lived antitoxin is necessary to ensure cell survival. These toxins have proven useful in various molecular biology applications. The TA toxin CcdB (control of cell death), which poisons bacterial DNA gyrase causing degradation of the host chromosome and cell death (67–69), is a key component of the pSW2644T allelic replacement vector, where expression of *ccdB* is used as a counter-selection marker (70,71). Invitrogen's Zero Background^{TM} cloning methodology exploits the same toxin to eliminate high background in blunt-end cloning strategies. When an insert is ligated into the vector, the *ccdB* gene is disrupted, enabling only recombinant colonies to grow.

Another cassette (*vv12532*) coded for a protein with an Excalibur domain and a putative cold shock domain (CSD) (Figure 4). Excalibur domains bind extracellular calcium and contain the conserved DxDxDGxxCE motif, which is strikingly similar to the Ca^{2+}-binding loop of calmodulin-like EF-hand domains (72). CSD are commonly found in cold shock proteins (CSP) (73).

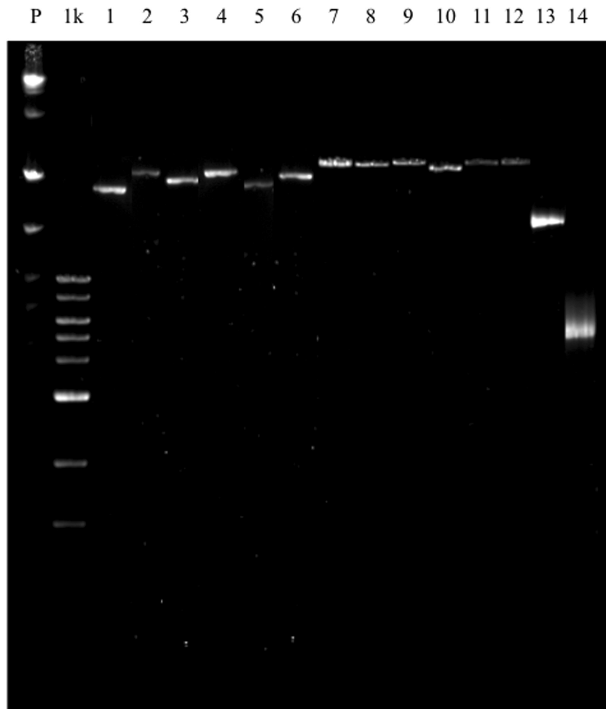

Figure 3. Pulse-field gel of the R388-pCC1FOS recombinant plasmids obtained from the transformed and transduced cosmid libraries. Plasmids were isolated and separated by PFGE. P, PFGE marker; 1 k, 1-kb maker; lanes 1–6, plasmids chosen at random from the transformed substrate library; lanes 7–12, plasmids chosen at random from the transduced substrate library; lane 13, R388; lane 14, the empty pCC1FOS cosmid.

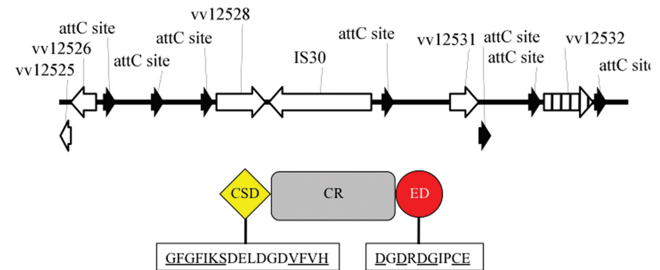

Figure 4. Schematic of the insert from a R388-pCC1FOS recombinant plasmid obtained from the transformed cosmid library of *V. vulnificus* strain CMCP6. (Top) The 5383-bp insert contained seven *attC* sites (black arrows) and corresponded to genes *vv12525-vv12533* of the CI of strain CMCP6. One of the partial gene cassettes contained two small overlapping *orfs* (*vv12525* and *vv12526*) that coded for a toxin–antitoxin system that we previously characterized from *V. vulnificus* strain 27562 (28,63). Another cassette, *vv12532* (striped), coded for a putative cold shock protein (CSP). (Bottom) The domain structure of VV12532. Cold shock domains (CSDs) contain 11 residues (underlined) that are important for CSP function. Excalibur domains (ED) bind extracellular calcium and contain the conserved DxDxDGxxCE motif (underlined). CR, central region.

CSP typically acts as transcription anti-terminators to prevent the formation of secondary structures in RNA molecules, thus facilitating translation at low temperature. Eleven residues that are important for CSP function are proposed to form two RNA-binding motifs, RNP-1 and RNP-2, on the surface of the CSP (74–79). An alignment between VV12532 and the major CSP of *E. coli*, *Thermotoga maritima* and *Bacillus caldolyticus* suggested that the RNP-1 and RNP-2 motifs were present in VV12532 (Figure 5). The VV12532 CSD was predicted to adopt an oligomer binding (OB)-fold similar to that of the *E. coli*, *T. maritime* and *B. caldolyticus* CSP, and the surface location of the RNP-1 and RNP-2 motifs was also conserved (Figure 5).

These observations suggested that VV12532 might function as a CSP. At least four separate *csp* genes must be deleted in *E. coli* to obtain a cold-sensitive phenotype due to overlapping activity of the CSP family of proteins (80). In *E. coli*, complementation by any one allele can revert the cold-sensitive phenotype of the quadruple deletion *ΔcspABGE* strain (81). The genomes of strain YJ016 and CMCP6 contain 4 and 5 CSP homologues, respectively. If a similar situation to that in *E. coli* occurs in *V. vulnificus*, targeted disruption of the putative *vv12532 csp* gene cassette alone would likely have little or no effect on the adaptation of *V. vulnificus* to cold temperature. Therefore, I opted to test the VV12532 CSP homologue by expressing it in the *ΔcspABGE E. coli* cold-sensitive mutant. The *vv12532* gene was cloned into the arabinose-inducible pBAD24 expression vector (82) and transformed into the *E. coli cspABEG* quadruple mutant. The cold-sensitive phenotype of the quadruple-deletion strain was suppressed in cells that expressed *vv12532* but not in cells that carried the empty vector (Figure 6), suggesting that VV12532 was a CSP. The protein was thus designated CspA^{vvu}. To determine if CspA^{vvu} acted as a transcription anti-terminator, pBAD24::*cspA*^{vvu} was transformed into the *E. coli* strain RL211 (83,84). RL211 contains a *cat* gene that is preceded by the ρ-independent *trpL* terminator. This strain is sensitive to chloramphenicol but becomes chloramphenicol resistant when transcription termination at *trpL* is reduced and thus provides a measure of antitermination activity. RL211 cells carrying pBAD24::*cspA*^{vvu} grew on plates containing chloramphenicol under inducing conditions, whereas the parental strain carrying pBAD24 did not (Figure 6). Hence, CspA^{vvu} was a CSP that appeared to act as a transcription antiterminator.

There were only 27 bp between the end of the 5′ *attC* site and the start ATG of *cspA*^{vvu}. A translation initiation region (TIR) was apparent but the sequence did not include a discernable promoter. Hence, I sought to directly screen for the activity of CspA^{vvu} *in vivo* by relocating gene cassettes from the *V. vulnificus* CMCP6 CI to the *attI1* site of the class 1 integon on R388ISceKT. This would place any incorporated CI cassettes downstream of the P_C promoter to ensure their expression. R388ISceKT and pTRC99A::*intI1* were introduced into *V. vulnificus* CMCP6 by conjugation. Following induction of *intI1* gene expression, R388 was transferred to the *E. coli ΔcspABGE* cold-sensitive mutant and transconjugants were selected on LB plates containing Tc and Km at 15°C. Sequence analysis of several transconjugants revealed that a CI cassette, corresponding to *cspA*^{vvu}, had been precisely inserted into In*3*.


```
              *     ::* :   :*:***   :      ***** : :   :*::     *: * * :          *
EcolCspA     MT-GIVKWFNADKGFGFITPDDGSKDVFVHFSAIQNDGYKSLDEGQKVSFTIESGAKGPAAGNVTSL     66
Bcald1c9oa   MQRGKVKWFNNEKGYGFIEVEGGS-DVFVHFTAIQGEGFKTLEEGQEVSFEIVQGNRGPQAANVVKL     66
TmarCspA     MR-GKVKWFDSKKGYGFITKDEGG-DVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVV     65
VvulVV12532  MKGKILRWVD-ERGFGFIKSDELDGDVFVHISKFP-QGYRRPQVGDHVEFHLANNQPKLSAASARL-     64
```

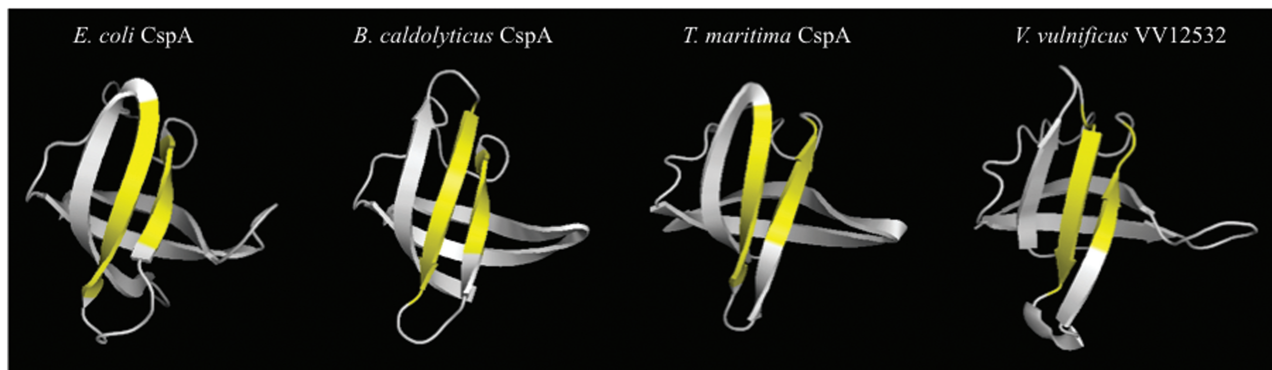

Figure 5. Comparison of the CSD of VV12532 to the major CSP of *E. coli*, *Thermotoga maritima* and *Bacillus caldolyticus*. (Top) Alignment between the CSD of VV12532 and the major CSP of the indicated species. Protein lengths are indicated on the right. Asterisks and colons mark identical and similar amino acids, respectively. The 11 residues that form the two RNA-binding motifs, RNP-1 and RNP-2, and that are important for CSP function are boxed. Bottom, the known structures of the major CSP of *E. coli*, *T. maritima* and *B. caldolyticus* are shown. The two RNA-binding motifs, RNP-1 and RNP-2 (yellow regions), are on the surface of the CSP. A proposed structure for the CSD of VV12532 is also shown.

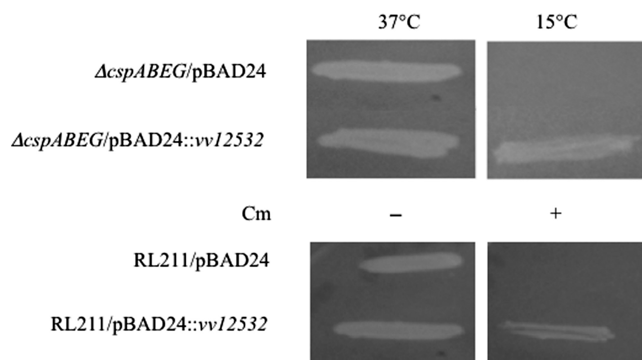

Figure 6. CspA^{vvu} is a CSP that causes transcription antitermination. Top, the *E. coli* cspABEG quadruple mutant with pBAD24::vv12532 or the empty vector were streaked on plates containing L-ara and were incubated at the indicated temperatures. Bottom, *E. coli* RL211 cells containing a *cat* gene cassette positioned downstream of the *trpL* terminator were transformed with pBAD24::*cspA*^{vvu} and streaked on plates containing 25 μg/ml chloramphenicol and arabinose. Plates were incubated overnight at 37°C.

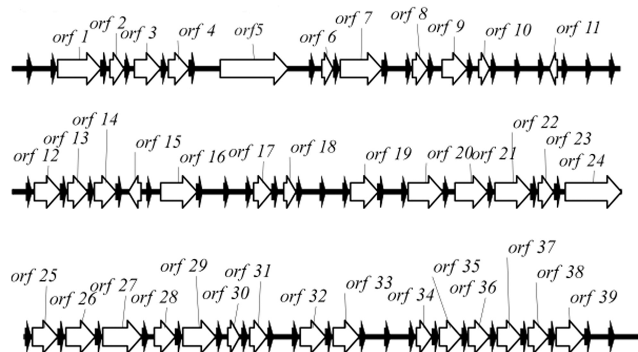

Figure 7. Schematic of a 35-kb region of the *V. vulnificus* ATCC 27562 CI. The insert of p2190 was sequenced and annotated. Gene cassettes with discernable *orfs* (white arrows) are numbered. The *attC* sites are shown as black arrows.

Comparative genomic analysis of a 35-kb contig containing CI gene cassettes from *V. vulnificus* strain ATCC 27562

Genomic comparisons of the CI cassette contents of the two sequenced *V. vulnificus* strains CMCP6 and YJ016 revealed that they share only 30–39% of their gene cassettes (34,85). The genome of strain 27562 has not been sequenced but we previously demonstrated that it contained a CI (28). To recover fosmids carrying large inserts from the CI, genomic DNA from strain 27562 was ligated into pCC1FOS, the library was packaged into phage and then transduced into the βDH10B host strain. R388IsceKT and pTRC99A::*intI1* were then moved to this strain to complete assembly of the system. Recombination was induced with the addition of IPTG and recombinant plasmids were isolated by conjugation to strain EPI300. The 35-kb insert of one of the recovered clones (p2190) was sequenced and annotated (Figure 7), and the *attC* sites and gene cassette boundaries were extracted with the XXR program we developed (28). Of the 55 complete and two partial gene cassettes that we identified, 39 (71%) contained a discernable *orf* (Table 1), 23 (59%) had homologues in the CIs of other members of the *Vibrionaceae* and only 16 (29%) had homologues in either the CMCP6 or YJ016 CIs. Furthermore, only about 12% of the CI cassettes from

Table 1. Homologues of the *orfs* from the 27562 CI

orf	Bacterial species	*E*-value	Function	Accession no.
1	NH		Hypothetical protein	
2	NH		Hypothetical protein	
3	*Shewanella sp.* MR7	$8e^{-30}$	preprotein translocase subunit SecA	YP_736195.1
4	NH		Hypothetical protein	
5	*Legionella pneumophila*	$3e^{-152}$	Reverse transcriptase	YP_001251128.1
6	*V. vulnificus*	$5e^{-44}$	Hypothetical protein	NP_934729.1
7	NH		Hypothetical protein	
8	*V. vulnificus*	$4e^{-55}$	Hypothetical protein	NP_761344.1
9	NH		Hypothetical protein	
10	*Aeromonas salmonicida*	$6e^{-31}$	Hypothetical protein	YP_001143449.1
11	*Shewanella* sp. MR4	$2e^{-17}$	Hypothetical protein	YP_734249.1
12	*S. lioihica* PV4	$7e^{-5}$	Hypothetical protein	YP_001095272.1
13	*V. vulnificus*	$2e^{-80}$	Hypothetical protein	NP_934596.1
14	*Aliivibrio salmonicida*	$4e^{-18}$	Acetyltransferase	YP_002262874
15	*V. vulnificus*	$1e^{-41}$	Hypothetical protein	NP_761251.1
16	*V. fischeri*	$3e^{-108}$	Esterase/lipase	YP_002156256.1
17	*V. vulnificus*	$5e^{-73}$	Hypothetical protein	NP_761320.1
18	*S. lioihica* PV4	$9e^{-19}$	Hypothetical protein	YP_001094978.1
19	NH		Hypothetical protein	
20	*V. parahaemolyticus*	$3e^{-151}$	CBS domain protein	NP_798204.1
21	*V. vulnificus*	$5e^{-131}$	Hypothetical protein	NP_761333.1
22	*V. vulnificus*	$2e^{-165}$	Hypothetical protein	NP_761335.1
23	*V. parahaemolyticus*	$1e^{-60}$	Hypothetical protein	NP_798185.1
24	*Oenococcus oeni*	$5e^{-59}$	Permease of the major facilitator superfamily	YP_809940.1
25	NH		Hypothetical protein	
26	NH		Hypothetical protein	
27	*V. vulnificus*	$2e^{-142}$	Hypothetical protein	NP_761333.1
28	*A. hydrophila*	$3e^{-10}$	Hypothetical protein	
29	NH		Hypothetical protein	
30	*V. vulnificus*	$5e^{-27}$	Acetyltransferase	NP_761288.1
31	*Photobacterium profundum*	$2e^{-17}$	Hypothetical protein	YP_132162.1
32	NH		Hypothetical protein	
33	NH		Hypothetical protein	
34	*V. vulnificus*	$4e^{-47}$	Hypothetical protein	
35	NH		Hypothetical protein	
36	*V. vulnificus*	$9e^{-71}$	Lactoglutathione lyase	
37	*V. vulnificus*	$1e^{-86}$	Hypothetical protein	NP_934598.1
38	*V. vulnificus*	$7e^{-81}$	4-oxalocrotonate tautomerase	NP_761363.1
39	*V. vulnificus*	$3e^{-109}$	Adenylate kinase	NP_934722.1

the *V. vulnificus* CMCP6, YJ016 and 27562 strains had homologues in the CI of *V. cholerae* N16961. These results support the notion that CIs contribute to the extensive heterogeneity of the *Vibrionaceae* and that the pool of gene cassettes available for exchange among bacteria is likely immense.

Several of the CI cassettes showed homology to enzymes involved in energy metabolism and detoxification (Table 1). These included a 4-oxalocrotonate tautomerase that enables bacteria to use various aromatic hydrocarbons as their sole energy source (86), a carbohydrate transport protein of the major facilitator superfamily, an adenylate kinase that interconverts ADP to AMP and ATP for driving energy requiring processes and a lactoylglutathione lyase detoxification enzyme that converts toxic methylglyoxal (MG) to lactic acid (87).

To demonstrate that the system could recover integron gene cassettes from different sources within the same sample, genomic DNA from *V. cholerae*, *X. campestris pv campestris* and *S. oneidensis* was isolated, mixed in equal quantities and cloned into pCC1FOS. While each of these species is known to harbour a CI (27,30,88), they are of varying size. The CIs of *V. cholerae*, *X. campestris* and *S. oneidensis* harbour 216, 23 and three gene cassettes, respectively (27,32,89), and represent 3, 0.3 and 0.1% of the genomes. The three-plasmid system was assembled and integrase-mediated recombination was induced. Recovered fusion plasmids were used in PCR reactions with primers intI1 and VVR1, VCR1, XCR1 or SPR1-3, and the products were sequenced. Products corresponding to integron gene cassettes from the CIs of each species were identified, suggesting that the system can indeed recover low abundance integron gene cassettes from different sources within the same sample.

Sequencing of integron gene cassette arrays with the integron-based recombination system

The system can also be applied to the sequencing of integron gene cassette arrays. The p2190 plasmid was used as the sole substrate plasmid in the integron-based recombination assay with R388ISceKT. Recombination was induced by the addition of IPTG and recombinant plasmids were isolated by conjugation. The recombinant

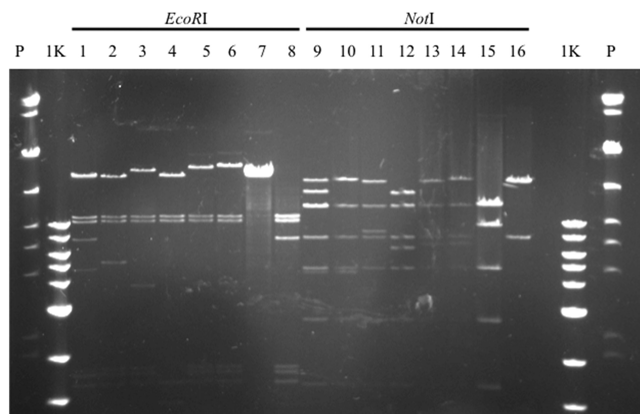

Figure 8. Pulse-field gel of R388-p2190 fusion plasmids. Fusion plasmids that were obtained by integrase-mediated recombination between the integron platform of R388 and the gene cassettes of p2190 were digested with NotI or EcoRI and separated by PFGE. Lanes are as follows: P, PFGE markers; 1 K, 1-Kb markers; 1 and 9, clone 1; 2 and 10, clone 2; 3 and 11, clone 3; 4 and 12, clone 4; 5 and 13, clone 5; 6 and 14, clone 6; 7 and 15, R388; 8 and 16, p2190.

plasmids from six randomly chosen transconjugants were digested with either EcoRI or NotI and separated by PFGE (Figure 8). Although each of the plasmids maintained the same overall size, each exhibited a different banding pattern, suggesting that multiple *attC* sites of p2190 could serve as substrates for recombination with the *attI1* site on R388ISceKT. This can prove useful since a series of plasmids with differing restriction patterns can be sequenced using two sets of unique primers that anneal within the 5′ conserved region and the dfrb2 gene of the class 1 integron, and within the *cat* gene and *cos* site of pCC1FOS, to obtain the sequence of the gene cassette array (Figure 2).

DISCUSSION

Fundamentally, the integron platform is a simple structure. There is an integrase (*intI*) and an adjacent, cognate recombination site (*attI*). The substrates for this system are the compact gene cassettes, many of which code for functional proteins or proteins that can adopt a structure of known biological significance. Further interest in the nature of the gene cassette pool has emerged with the demonstration that metabolic (25,42), virulence (43,44) and antibiotic resistance (45–47) genes are structured as gene cassettes and that the gene cassette pool may be limitless in size (1,90). Accessing this genetic pool has been hampered since sequence-based techniques such as hybridization with probes based on known genes or PCR with primers directed at conserved sequence motifs, often lead to the recovery of only partial genes or a small subset of those present in the sample. A previous method for the sequence-independent recovery of integron gene cassettes from environmental samples relied on cassette detection by PCR using degenerate primers that targeted conserved sequences within the *attC* sites (37). Although useful, this approach is limited by the diversity of the *attC* sites, which will undoubtedly preclude the detection of many of the gene cassettes that may be present in a given sample. Here, a genetic system to extract fosmids harboring integron gene cassettes from a bulk genomic pool is described. The system is comprised of three plasmids. The first expresses the class 1 integron-integrase under the control of an IPTG-inducible promoter. The second is a self-transmissible recipient plasmid that carries an *attI1* site for the integration of gene cassettes via recombination with their *attC* sites. The third is the donor plasmid that contains the genomic DNA library. IntI1-catalysed recombination between the *attI* site of recipient plasmid and target *attC* sites on the inserts of donor plasmids results in the fusion of the two plasmids. The fusion plasmid can then be isolated by conjugation into a recipient host and recovered for sequencing or activity screening of the gene cassettes.

The system has several advantages over existing methods for the detection of gene cassettes. Since integron-integrases are able to recognize diverse *attC* sites as targets for recombination (39,51,91,92), a recovery system based on the recombination activity of the integron integrase improves the likelihood of recovering gene cassettes from samples relative to detection by PCR or hybridization. Primers targeting the *cos* site, *dfrB2*, *cat* and *intI1* genes can be used to sequence the insert of the fusion plasmid. The I-SceI sites flanking the Km and Tc resistance markers of the fusion plasmid (Cm^R, Km^R, Tc^R) can be used to isolate clones lacking the R388 backbone (Cm^R Km^S, Tc^S) by replica plating on media containing the appropriate antibiotics. Obtaining such clones may be desirable for sequencing purposes or for activity-based screening. Finally, the ability to directly implement activity-based screening is a significant advantage. The activity-based screening of gene banks allows for the recovery of novel enzymes without any prior knowledge of the sequence and relies only on the ingenuity of the screening method. Chitinases, amylases, DNAses, esterase/lipases, proteases, oxygenases, dehydrogenases and antimicrobial activities have been identified via activity screens of genomic libraries (93,94). Many of these enzymes are used as key active components in detergents, various food/feed processes, pulp and paper manufacturing, molecular biology applications and medical treatment protocols. Activity screening requires gene expression and proper folding of the resulting protein in a heterologous host. In *E. coli*, the minimal set of requirements for gene expression includes the presence of a promoter for transcription and a translation initiation region (TIR) comprised of a Shine–Dalgarno sequence, the initiation codon, and an adequate spacer between them. In such systems, three modes of gene expression can be anticipated: (i) independent gene expression with both the promoter and the TIR provided by the insert; (ii) expression as a transcriptional fusion with only the TIR located on the insert; and (iii) expression as a translational fusion depending on both the promoter and the TIR of the vector. Genes of the last category are virtually inaccessible by shotgun cloning because of the low frequency of functional constructs. Class 1 integrons naturally incorporate many of the attributes necessary for the proper expression of exogenous genes. The Pc promoter drives expression of promoterless gene cassettes integrated at the *attI* site. Often, a TIR

can be recognized at the 5′ boundary of the gene cassette. Even cassettes that have no plausible TIR can be translated. A short open reading frame (ORF-11) that is separated by an octameric spacer from a consensus SD tetramer overlaps the *attI* site of class 1 integrons. Hanau-Berçot *et al.* (95) demonstrated that translation of the TIR-deficient aminoglycoside 6′-N-acetyltransferase (*aac(6′)-Ib7*) gene cassette was dependent on the translation of ORF-11 but not on the amino acid sequence of the corresponding peptide. These results indicated that ORF-11 and its TIR could substantially enhance expression at the translational level of TIR-deficient gene cassettes. Hence, the integron/gene cassette system can support all three modes of gene expression. Even potential translational attenuation signals, the *attC* sites themselves (96), are present. Using integron-based activity screening, we previously identified a novel *cat* gene in the CI of *V. cholera* (45) and here I demonstrate the activity of a CSP encoded by a *V. vulnificus* CI gene cassette.

Greater than 99% of the microorganisms present in many natural environments are unculturable. Hence their genetic wealth has eluded exploitation for medical, biotechnological or basic research purposes. Cultivation-independent approaches, in which DNA is directly isolated from environmental samples and used to construct complex genomic libraries, permits analysis of the genetically and metabolically rich microbial communities in their entirety (97). This obviates the need for cultivation and allows the direct cloning of the collective genomes of all organisms present in a habitat. The quest to identify functional genes in the vast amounts of genetic data contained in these metagenomic libraries presents a variety of challenges, not the least of which is filtering out the often large mounts of undesired DNA that codes for informational (transcription, translation and related processes) and conserved operational (housekeeping) genes (98,99). This can be overcome, in part, by focusing on genes that are structured as cassettes in integrons, since they tend to code for proteins of unknown function or with tremendous adaptive potential (1,100–104), including structural (105), metabolic (25,42,106), resistance (40,107), toxin (45,46,108) and virulence proteins (44,109). Thus, integron gene cassettes may be a tremendous source of new enzymes and novel proteins from diverse origins. By manipulating the natural activity of integrons, it should be possible to gain access to many of the functional gene cassettes contained within a metagenomic sample.

ACKNOWLEDGEMENTS

I would like to thank Dr Jan-Fang Cheng (JGI) for sequencing of fosmid clones, and Erica Susky and Bella Kuchuk for technical assistance.

FUNDING

The Natural Sciences and Engineering Research Council of Canada (293281) and the Canadian Institutes of Health Research (MOP-89794). Funding for open access charge: CIHR.

Conflict of interest statement. None declared.

REFERENCES

1. Rowe-Magnus,D.A. and Mazel,D. (2001) Integrons: natural tools for bacterial genome evolution. *Curr. Opin. Microbiol.*, **4**, 565–569.
2. Mazel,D. (2006) Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.*, **4**, 608–620.
3. Rowe-Magnus,D.A. and Mazel,D. (1999) Resistance gene capture. *Curr. Opin. Microbiol.*, **2**, 483–488.
4. Rowe-Magnus,D.A. and Mazel,D. (2002) The role of integrons in antibiotic resistance gene capture. *Int. J. Med. Microbiol.*, **292**, 115–125.
5. Mazel,D. and Davies,J. (1999) Antibiotic resistance in microbes. *Cell Mol. Life Sci.*, **56**, 742–754.
6. Martinez,E. and de la Cruz,F. (1988) Transposon Tn21 encodes a RecA-independent site-specific integration system. *Mol. Gen. Genet.*, **211**, 320–325.
7. Martinez,E. and de la Cruz,F. (1990) Genetic elements involved in Tn*21* site-specific integration, a novel mechanism for the dissemination of antibiotic resistance genes. *EMBO J.*, **9**, 1275–1281.
8. Hansson,K., Sundstrom,L., Pelletier,A. and Roy,P.H. (2002) IntI2 integron integrase in Tn7. *J. Bacteriol.*, **184**, 1712–1721.
9. Sundstrom,L., Roy,P.H. and Skold,O. (1991) Site-specific insertion of three structural gene cassettes in transposon Tn7. *J. Bacteriol.*, **173**, 3025–3028.
10. Arakawa,Y., Murakami,M., Suzuki,K., Ito,H., Wacharotayankun,R., Ohsuka,S., Kato,N. and Ohta,M. (1995) A novel integron-like element carrying the metallo-beta-lactamase gene *blaIMP*. *Antimicrob. Agents Chemother.*, **39**, 1612–1615.
11. Hochhut,B., Lotfi,Y., Mazel,D., Faruque,S.M., Woodgate,R. and Waldor,M.K. (2001) Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins. *Antimicrob. Agents Chemother.*, **45**, 2991–3000.
12. Sorum,H., Dommarsnes,K., Sandersen,K., Sundstrom,L., Gullberg,M. and Solberg,A. (2001) Genbank accession no. AJ277063.
13. Liebert,C.A., Hall,R.M. and Summers,A.O. (1999) Transposon Tn21, flagship of the floating genome. *Microbiol. Mol. Biol. Rev.*, **63**, 507–522.
14. Nunes-Duby,S.E., Kwon,H.J., Tirumalai,R.S., Ellenberger,T. and Landy,A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.*, **26**, 391–406.
15. Hall,R.M. (1997) Mobile gene cassettes and integrons: moving antibiotic resistance genes in gram-negative bacteria. *Ciba Found. Symp.*, **207**, 192–202.
16. Recchia,G.D. and Hall,R.M. (1995) Gene cassettes: a new class of mobile element. *Microbiology*, **141**, 3015–3027.
17. Levesque,C., Brassard,S., Lapointe,J. and Roy,P.H. (1994) Diversity and relative strength of tandem promoters for the antibiotic-resistance genes of several integrons. *Gene*, **142**, 49–54.
18. Rowe-Magnus,A.D., Davies,J. and Mazel,D. (2002) Impact of integrons and transposons on the evolution of resistance and virulence. *Curr. Top. Microbiol. Immunol.*, **264**, 167–188.
19. Rowe-Magnus,D.A., Davies,J. and Mazel,D. (2002) In Hacker,J. and Kaper,J. (eds), *Pathogenicity Islands and the Evolution of Pathogenic Microbes*, Vol. 264/2, pp. 167–188.
20. Davies,J.E. (1997) Origins, acquisition and dissemination of antibiotic resistance determinants. *Ciba Found. Symp.*, **207**, 15–27; discussion 27–35.
21. Davies,J. (1994) Inactivation of antibiotics and the dissemination of resistance genes. *Science*, **264**, 375–382.
22. Peters,E.D., Leverstein-Van Hall,M.A., Box,A.T., Verhoef,J. and Fluit,A.C. (2001) Novel gene cassettes and integrons. *Antimicrob. Agents Chemother.*, **45**, 2961–2964.
23. Recchia,G.D. and Hall,R.M. (1997) Origins of the mobile gene cassettes found in integrons. *Trends Microbiol.*, **5**, 389–394.
24. Rowe-Magnus,D.A., Guerout,A.-M. and Mazel,D. (1999) Super-Integrons. *Res. Microbiol.*, **150**, 641–651.
25. 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. USA*, **98**, 652–657.
26. 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.
27. Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L. *et al.* (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature*, **406**, 477–483.
28. 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.
29. Vaisvila,R., Morgan,R.D., Posfai,J. and Raleigh,E.A. (2001) Discovery and distribution of super-integrons among Pseudomonads. *Mol. Microbiol.*, **42**, 587–601.
30. Gillings,M.R., Holley,M.P., Stokes,H.W. and Holmes,A.J. (2005) Integrons in Xanthomonas: a source of species genome diversity. *Proc. Natl Acad. Sci. USA*, **102**, 4419–4424.
31. Coleman,N., Tetu,S., Wilson,N. and Holmes,A. (2004) An unusual integron in Treponema denticola. *Microbiology*, **150**, 3524–3526.
32. Heidelberg,J.F., Paulsen,I.T., Nelson,K.E., Gaidos,E.J., Nelson,W.C., Read,T.D., Eisen,J.A., Seshadri,R., Ward,N., Methe,B. *et al.* (2002) Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis. *Nat. Biotech.*, **20**, 1118–1123.
33. Holmes,A.J., Holley,M.P., Mahon,A., Nield,B., Gillings,M. and Stokes,H.W. (2003) Recombination activity of a distinctive integron-gene cassette system associated with Pseudomonas stutzeri populations in soil. *J. Bacteriol.*, **185**, 918–928.
34. Chen,C.Y., Wu,K.M., Chang,Y.C., Chang,C.H., Tsai,H.C., Liao,T.L., Liu,Y.M., Chen,H.J., Shen,A.B., Li,J.C. *et al.* (2003) Comparative genome analysis of Vibrio vulnificus, a marine pathogen. *Genome Res.*, **13**, 2577–2587.
35. Fraser,C.M., Gocayne,J.D., White,O., Adams,M.D., Clayton,R.A., Fleischmann,R.D., Bult,C.J., Kerlavage,A.R., Sutton,G. and Kelley,J.M. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science*, **270**, 397–403.
36. Himmelreich,R., Hilbert,H., Plagens,H., Pirkl,E., Li,B.C. and Herrmann,R. (1996) Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae. *Nucleic Acids Res.*, **24**, 4420–4449.
37. 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.
38. Nield,B.S., Holmes,A.J., Gillings,M.R., Recchia,G.D., Mabbutt,B.C., Nevalainen,K.M. and Stokes,H.W. (2001) Recovery of new integron classes from environmental DNA. *FEMS Microbiol. Lett.*, **195**, 59–65.
39. Collis,C.M., Recchia,G.D., Kim,M.J., Stokes,H.W. and Hall,R.M. (2001) Efficiency of recombination reactions catalyzed by class 1 integron integrase IntI1. *J. Bacteriol.*, **183**, 2535–2542.
40. 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.
41. Robinson,A., Wu,P.S., Harrop,S.J., Schaeffer,P.M., Dosztanyi,Z., Gillings,M.R., Holmes,A.J., Helena Nevalainen,K.M., Stokes,H.W., Otting,G. *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.
42. Nield,B.S., Willows,R.D., Torda,A.E., Gillings,M.R., Holmes,A.J., Nevalainen,K.M., Stokes,H.W. and Mabbutt,B.C. (2004) New enzymes from environmental cassette arrays: functional attributes of a phosphotransferase and an RNA-methyltransferase. *Protein Sci.*, **13**, 1651–1659.
43. 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.
44. 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.
45. Rowe-Magnus,D.A., Guerout,A.M. and Mazel,D. (2002) Bacterial resistance evolution by recruitment of super-integron gene cassettes. *Mol. Microbiol.*, **43**, 1657–1669.
46. Melano,R., Petroni,A., Garutti,A., Saka,H.A., Mange,L., Pasteran,F., Rapoport,M., Rossi,A. and Galas,M. (2002) New carbenicillin-hydrolyzing beta-lactamase (CARB-7) from *Vibrio cholerae* non-O1, non-O139 strains encoded by the VCR region of the *V. cholerae* genome. *Antimicrob. Agents Chemother.*, **46**, 2162–2168.
47. Petroni,A., Melano,R.G., Saka,H.A., Garutti,A., Mange,L., Pasteran,F., Rapoport,M., Miranda,M., Faccone,D., Rossi,A. *et al.* (2004) CARB-9, a carbenicillinase encoded in the VCR region of Vibrio cholerae non-O1, non-O139 belongs to a family of cassette-encoded beta-lactamases. *Antimicrob. Agents Chemother.*, **48**, 4042–4046.
48. Tennstedt,T., Szczepanowski,R., Braun,S., Pühler,A. and Schlüter,A. (2003) Occurrence of integron-associated resistance gene cassettes located on antibiotic resistance plasmids isolated from a wastewater treatment plant. *FEMS Microbiol. Ecol.*, **45**, 239–252.
49. Stokes,H.W., O'Gorman,D.B., Recchia,G.D., Parsekhian,M. and Hall,R.M. (1997) Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol. Microbiol.*, **26**, 731–745.
50. Drouin,F., Melancon,J. and Roy,P.H. (2002) The IntI-like tyrosine recombinase of *Shewanella oneidensis* is active as an integron integrase. *J. Bacteriol.*, **184**, 1811–1815.
51. 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.
52. Avila,P. and de la Cruz,F. (1988) Physical and genetic map of the IncW plasmid R388. *Plasmid*, **20**, 155–157.
53. Hall,R.M., Brown,H.J., Brookes,D.E. and Stokes,H.W. (1994) Integrons found in different locations have identical 5′ ends but variable 3′ ends. *J. Bacteriol.*, **176**, 6286–6294.
54. 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.
55. Chaveroche,M.K., Ghigo,J.M. and d'Enfert,C. (2000) A rapid method for efficient gene replacement in the filamentous fungus Aspergillus nidulans. *Nucleic Acids Res.*, **28**, E97.
56. Clarke,L., Hitzeman,R. and Carbon,J. (1979) Selection of specific clones from colony banks by screening with radioactive antibody. *Methods Enzymol.*, **68**, 436–442.
57. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
58. Gasteiger,E., Hoogland,C., Gattiker,A., Duvaud,S., Wilkins,M.R., Appel,R.D. and Bairoch,A. (2005) In Walker,J.M. (ed.), *The Proteomics Protocols Handbook, Humana Press*. Humana Press, NJ, pp. 571–607.
59. Hulo,N., Bairoch,A., Bulliard,V., Cerutti,L., De Castro,E., Langendijk-Genevaux,P.S., Pagni,M. and Sigrist,C.J.A. (2006) The PROSITE database. *Nucleic Acids Res.*, **34**, D227–D230.
60. Kelley,L.A. and Sternberg,M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.*, **4**, 363–371.
61. Amann,E., Ochs,B. and Abel,K.J. (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli. *Gene*, **69**, 301–315.
62. Huovinen,P. (1987) Trimethoprim resistance. *Antimicrob. Agents Chemother.*, **31**, 1451–1456.
63. Szekeres,S., Dauti,M., Wilde,C.D.M. and Rowe-Magnus,D.A. (2007) Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Mol. Microbiol.*, **43**, 1657–1669.
64. Engelberg-Kulka,H., Amitai,S., Kolodkin-Gal,I. and Hazan,R. (2006) Bacterial programmed cell death and multicellular behavior in bacteria. *PLoS Genet.*, **2**, e135.
65. Hayes,F. (2003) Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science*, **301**, 1496–1499.

66. Engelberg-Kulka,H. and Glaser,G. (1999) Addiction modules and programmed cell death and antideath in bacterial cultures. *Annu. Rev. Microbiol.*, **53**, 43–70.
67. Bernard,P. and Couturier,M. (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J. Mol. Biol.*, **226**, 735–745.
68. Bernard,P., Kezdy,K.E., Van Melderen,L., Steyaert,J., Wyns,L., Pato,M.L., Higgins,P.N. and Couturier,M. (1993) The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase. *J. Mol. Biol.*, **234**, 534–541.
69. Critchlow,S.E., O'Dea,M.H., Howells,A.J., Couturier,M., Gellert,M. and Maxwell,A. (1997) The interaction of the F plasmid killer protein, CcdB, with DNA gyrase: induction of DNA cleavage and blocking of transcription. *J. Mol. Biol.*, **273**, 826–839.
70. Bernard,P., Gabant,P., Bahassi,E.M. and Couturier,M. (1994) Positive-selection vectors using the F plasmid ccdB killer gene. *Gene*, **148**, 71–74.
71. 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.
72. Rigden,D.J., Jedrzejas,M.J. and Galperin,M.Y. (2003) An extracellular calcium-binding domain in bacteria with a distant relationship to EF-hands. *FEMS Microbiol. Lett.*, **221**, 103–110.
73. Ermolenko,D.N. and Makhatadze,G.I. (2002) Bacterial cold-shock proteins. *Cell Mol. Life Sci.*, **59**, 1902–1913.
74. Feng,W., Tejero,R., Zimmerman,D.E., Inouye,M. and Montelione,G.T. (1998) Solution NMR structure and backbone dynamics of the major cold-shock protein (CspA) from Escherichia coli: evidence for conformational dynamics in the single-stranded RNA-binding site. *Biochemistry*, **37**, 10881–10896.
75. Newkirk,K., Feng,W., Jiang,W., Tejero,R., Emerson,S.D., Inouye,M. and Montelione,G.T. (1994) Solution NMR structure of the major cold shock protein (CspA) from Escherichia coli: identification of a binding epitope for DNA. *Proc. Natl Acad. Sci. USA*, **91**, 5114–5118.
76. Reid,K.L., Rodriguez,H.M., Hillier,B.J. and Gregoret,L.M. (1998) Stability and folding properties of a model beta-sheet protein, Escherichia coli CspA. *Protein Sci.*, **7**, 470–479.
77. Schindelin,H., Jiang,W., Inouye,M. and Heinemann,U. (1994) Crystal structure of CspA, the major cold shock protein of Escherichia coli. *Proc. Natl Acad. Sci. USA*, **91**, 5119–5123.
78. Mueller,U., Perl,D., Schmid,F.X. and Heinemann,U. (2000) Thermal stability and atomic-resolution crystal structure of the Bacillus caldolyticus cold shock protein. *J. Mol. Biol.*, **297**, 975–988.
79. Kremer,W., Schuler,B., Harrieder,S., Geyer,M., Gronwald,W., Welker,C., Jaenicke,R. and Kalbitzer,H.R. (2001) Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium Thermotoga maritima. *Eur. J. Biochem.*, **268**, 2527–2539.
80. Xia,B., Ke,H. and Inouye,M. (2001) Acquirement of cold sensitivity by quadruple deletion of the cspA family and its suppression by PNPase S1 domain in Escherichia coli. *Mol. Microbiol.*, **40**, 179–188.
81. Phadtare,S., Inouye,M. and Severinov,K. (2002) The nucleic acid melting activity of Escherichia coli CspE is critical for transcription antitermination and cold acclimation of cells. *J. Biol. Chem.*, **277**, 7239–7245.
82. Guzman,L.M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.*, **177**, 4121–4130.
83. Landick,R., Stewart,J. and Lee,D.N. (1990) Amino acid changes in conserved regions of the beta-subunit of Escherichia coli RNA polymerase alter transcription pausing and termination. *Genes Dev.*, **4**, 1623–1636.
84. Bae,W., Xia,B., Inouye,M. and Severinov,K. (2000) Escherichia coli CspA-family RNA chaperones are transcription antiterminators. *Proc. Natl Acad. Sci. USA*, **97**, 7784–7789.
85. Boucher,Y., Nesbo,C.L., Joss,M.J., Robinson,A., Mabbutt,B.C., Gillings,M.R., 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.
86. Whitman,C.P. (2002) The 4-oxalocrotonate tautomerase family of enzymes: how nature makes new enzymes using a beta-alpha-beta structural motif. *Arch. Biochem. Biophys.*, **402**, 1–13.
87. Ferguson,G.P., Totemeyer,S., MacLean,M.J. and Booth,I.R. (1998) Methylglyoxal production in bacteria: suicide or survival? *Arch. Microbiol.*, **170**, 209–218.
88. Romine,M.F., Carlson,T.S., Norbeck,A.D., McCue,L.A. and Lipton,M.S. (2008) Identification of mobile elements and pseudogenes in the Shewanella oneidensis MR-1 genome. *Appl. Environ. Microbiol.*, **74**, 3257–3265.
89. da Silva,A.C., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida,N.F., Alves,L.M. *et al.* (2002) Comparison of the genomes of two Xanthomonas pathogens with differing host specificities. *Nature*, **417**, 459–463.
90. 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.
91. Collis,C.M., Grammaticopoulos,G., Briton,J., Stokes,H.W. and Hall,R.M. (1993) Site-specific insertion of gene cassettes into integrons. *Mol. Microbiol.*, **9**, 41–52.
92. 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.
93. Hardeman,F. and Sjoling,S. (2007) Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiol. Ecol.*, **59**, 524–534.
94. Rondon,M.R., August,P.R., Bettermann,A.D., Brady,S.F., Grossman,T.H., Liles,M.R., Loiacono,K.A., Lynch,B.A., MacNeil,I.A., Minor,C. *et al.* (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.*, **66**, 2541–2547.
95. Hanau-Berçot,B., Podglajen,I., Casin,I. and Collatz,E. (2002) An intrinsic control element for translational initiation in class 1 integrons. *Mol. Microbiol.*, **44**, 119–130.
96. Jacquier,H., Zaoui,C., Sanson-le Pors,M.J., Mazel,D. and Berçot,B. (1995) Translation regulation of integrons gene cassette expression by the attC sites. *Mol. Microbiol.*, **72**, 1475–1486.
97. Schmeisser,C., Stockigt,C., Raasch,C., Wingender,J., Timmis,K.N., Wenderoth,D.F., Flemming,H.C., Liesegang,H., Schmitz,R.A., Jaeger,K.E. *et al.* (2003) Metagenome survey of biofilms in drinking-water networks. *Appl. Environ. Microbiol.*, **69**, 7298–7309.
98. Charlebois,R.L. and Doolittle,W.F. (2004) Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome Res.*, **14**, 2469–2477.
99. Fitz-Gibbon,S.T. and House,C.H. (1999) Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res.*, **27**, 4218–4222.
100. 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.
101. Michael,C.A. (2004) Mobile gene cassettes: a fundamental resource for bacterial evolution. *Am. Nat.*, **164**, 1–12.
102. Rowe-Magnus,D.A., Zouine,M. and Mazel,D. (2006) In Thompson,F.L., Austin,B. and Swings,J. (eds), *The Biology of Vibrios*. American Society fro Microbiology, Washington.
103. Robinson,A., Guilfoyle,A.P., Sureshan,V., Howell,M., Harrop,S.J., Boucher,Y., Stokes,H.W., Curmi,P.M. and Mabbutt,B.C. (2008) Structural genomics of the bacterial mobile metagenome: an overview. *Methods Mol. Biol.*, **426**, 589–595.
104. Robinson,A., Wu,P.S., Harrop,S.J., Schaeffer,P.M., Dosztanyi,Z., Gillings,M.R., Holmes,A.J., Nevalainen,K.M., Stokes,H.W., Otting,G. *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.

105. Barker,A. and Manning,P.A. (1997) VlpA of *Vibrio cholerae* O1: the first bacterial member of the alpha 2-microglobulin lipocalin superfamily. *Microbiology*, **143**, 1805–1813.
106. 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.
107. Sobecky,P.A. and Coombs,J.M. (2009) Horizontal gene transfer in metal and radionuclide contaminated soils. *Methods Mol. Biol.*, **532**, 455–472.
108. Fonseca,E.L., Dos Santos Freitas,F., Vieira,V.V. and Vicente,A.C. (2008) New qnr gene cassettes associated with superintegron repeats in Vibrio cholerae O1. *Emerging Infect. Dis.*, **14**, 1129–1131.
109. Franzon,V.L., Barker,A. and Manning,P.A. (1993) Nucleotide sequence encoding the mannose-fucose-resistant hemagglutinin of Vibrio cholerae O1 and construction of a mutant. *Infect. Immun.*, **61**, 3032–3037.