J. Microbiol. Biotechnol. (2016), 26(11), 1983–1992
http://dx.doi.org/10.4014/jmb.1605.05068

jmb

Identification and Characterization of Putative Integron-Like Elements of the Heavy-Metal-Hypertolerant Strains of *Pseudomonas* spp. S

Anna Ciok, Marcin Adamczuk, Dariusz Bartosik, and Lukasz Dziewit*

Department of Bacterial Genetics, Institute of Microbiology, Faculty of Biology, University of Warsaw, Warsaw 02-096, Poland

Received: May 23, 2016
Revised: July 14, 2016
Accepted: July 20, 2016

First published online
July 27, 2016

**Corresponding author*
Phone: +48-22-554-1406;
Fax: +48-22-554-1402;
E-mail: ldziewit@biol.uw.edu.pl

S upplementary data for this paper are available on-line only at http://jmb.or.kr.

pISSN 1017-7825, eISSN 1738-8872

Copyright© 2016 by
The Korean Society for Microbiology and Biotechnology

Pseudomonas strains isolated from the heavily contaminated Lubin copper mine and Zelazny Most post-flotation waste reservoir in Poland were screened for the presence of integrons. This analysis revealed that two strains carried homologous DNA regions composed of a gene encoding a DNA_BRE_C domain-containing tyrosine recombinase (with no significant sequence similarity to other integrases of integrons) plus a three-component array of putative integron gene cassettes. The predicted gene cassettes encode three putative polypeptides with homology to (i) transmembrane proteins, (ii) GCN5 family acetyltransferases, and (iii) hypothetical proteins of unknown function (homologous proteins are encoded by the gene cassettes of several class 1 integrons). Comparative sequence analyses identified three structural variants of these novel integron-like elements within the sequenced bacterial genomes. Analysis of their distribution revealed that they are found exclusively in strains of the genus *Pseudomonas*.

Keywords: Integron-like element, integron gene cassette, *Pseudomonas* spp., Lubin copper mine, Zelazny Most

Introduction

Integrons are widespread elements that play an important role in the adaptation of bacteria to changeable and stressful environmental conditions. These elements consist of a stable genetic platform known as the 5′-Conserved Sequence (5′-CS) and a variable array of incorporated gene cassettes. Some integrons (chromosomal superintegrons) may contain more than 200 such cassettes [7].

The 5′-CS consists of (i) an *intI* gene encoding an integrase (tyrosine recombinase), (ii) a recombination site *attI*, into which gene cassettes are inserted by an IntI integrase, and (iii) P_{intI} and P_c promoters driving the transcription of *intI* and the gene cassettes, respectively [23]. Many class 1 integrons contain an additional conserved region downstream of the cassette array (3′-CS), which confers resistance to quaternary ammonium compounds (*qacE*Δ*1*) and sulfonamides (*sul1*) [7].

When not integrated, the gene cassettes transiently exist as circular DNA entities, containing promoterless open reading frames (ORFs) and an *attC* recombination site. Incorporation of the cassettes into an integron (by site-specific recombination between *attC* and *attI* sites, mediated by the IntI integrases) places the ORFs adjacent to the external P_c promoter and converts them into functional genes [23, 29].

Integrons identified in clinical isolates of bacterial pathogens are of special interest owing to the important role that they play in the dissemination and evolution of multidrug resistance. Therefore, the best studied group of integron-associated gene cassettes encodes proteins involved in resistance to antibiotics, including beta-lactams, aminoglycosides, trimethoprim, chloramphenicol, streptothricin, erythromycin, rifampicin, lincomycin, fosfomycin, and quinolones, as well as quaternary ammonium compound antiseptics [9, 32, 37]. Much less is known about integron-mediated phenotypes of environmental bacteria, but recent metagenomic studies indicate that some strains may contain gene cassettes involved in the utilization of various anthropogenic pollutants; for example, polycyclic aromatic

hydrocarbons such as toluene, xylene, naphthalene, anthracene, carbazole, and biphenyl [24]. Nevertheless, the vast majority of gene cassettes found in nonpathogenic environmental bacteria carry genes of unknown function [5].

In this study, we have identified and characterized novel genetic elements, exhibiting characteristics of some integrons and occurring within the genomes of a pool of multiresistant heavy-metal-hypertolerant strains of *Pseudomonas* spp. inhabiting the Lubin copper mine and Zelazny Most post-flotation waste reservoir in Poland [13]. The Lubin underground mine is a source of polymetallic ore that is rich in copper (10 wt %) and silver (100 mg/kg). The Zelazny Most reservoir (the largest post-flotation pond in Europe) collects flotation tailings from the Lubin mine. The total volume of waste present in Zelazny Most is estimated to be 476 million m^3 with an annual deposition of flotation tailings of between 20 and 26 million tons. Both environments are highly contaminated with heavy metals (Cu, Pb, As, Ni, Co, Zn, and Cr) and various organic compounds, including polycyclic aromatic hydrocarbons such as anthracene, biphenyl, dibenzofurane, chrysene, fluoranthene, fluorene, naphthalene, phenanthrene, and pyrene [28, 39].

The genetic elements identified in this work may constitute a novel group of integron-like elements (ILEs). Therefore, we present here their structural and molecular characterization.

Table 1. Plasmids used and constructed in this study.

Plasmid	Characteristics	Reference or source
pBluescript KSII	Ap^r; *oriV* ColE1; *lacZ*α; cloning vector	[1]
pBS-MFEI	Ap^r; Cm^r; Km^r; pBluescript KSII derivative carrying 17.0-kb MfeI restriction fragment of LM13::pDS-GCLM13KM genomic DNA cloned in EcoRV site	This study
pDS-SACI	Cm^r; pDS132 derivative carrying 8.2 kb SacI restriction fragment of LM13::pDS-GCLM13KM genomic DNA	This study
pBS-SACIIA	Ap^r; Km^r, pBluescript KSII derivative carrying 11.2 kb SacII restriction fragment of LM7::pDS-GCLM7KM genomic DNA cloned in EcoRV site	This study
pBS-SACIIB	Ap^r; pBluescript KSII derivative carrying 4.9 kb SacII restriction fragment of LM7::pDS-GCLM7KM genomic DNA cloned in EcoRV site	This study
pCAB132	Km^r; pCM132 derivative with deletion of EcoRI-BglII fragment; promoterless *lacZ* reporter gene	[3]
pCM132	Km^r; *oriV* RK2; *oriT* RK2; promoterless *lacZ* reporter gene	[26]
pCM132-1A	Km^r; pCM132 derivative carrying putative P_c promoter for the LM13 gene cassettes	This study
pCM132-1B	Km^r; pCM132 derivative carrying putative P_{intI} promoter for the LM13 integrase gene	This study
pDIY-KM	Km^r; Ap^r; *oriV* ColE1; source of kanamycin resistance cassette	[11]
pDS132	Cm^r; *oriV* R6K; *oriT* RK2; *sacB*; suicide vector in *Pseudomonas* spp.	[33]
pDS-GCLM13	Cm^r; pDS132 derivative carrying gene cassette (*orf_05*) of LM13 strain (derived from pGEM-GCLM13) cloned between PstI and SphI sites	This study
pDS-GCLM13KM	Cm^r; Km^r; pDS-GCLM13 derivative carrying kanamycin resistance cassette inserted in Eco31I site of the gene cassette (*orf_05*) of LM13 strain	This study
pDS-GCLM7	Cm^r; pDS132 derivative carrying gene cassette (*orf_05*) of LM7 strain (derived from pGEM-GCLM7) cloned between PstI and SphI sites	This study
pDS-GCLM7KM	Cm^r; Km^r; pDS-GCLM7 derivative carrying kanamycin resistance cassette inserted in Eco31I site of the gene cassette (*orf_05*) of LM7 strain	This study
pET28b(+)	Km^r; *ori* pBR322; *ori* f1; expression vector	Novagen
pET-lexA	Km^r; pET28b(+) derivative carrying *lexA* gene of *P. aeruginosa* PAO1161 cloned in-frame	This study
pGEM-GCLM13	Ap^r; pGEM-T Easy derivative carrying gene cassette (*orf_05*) of LM13 strain (PCR amplified with primers HS286 and HS287)	This study
pGEM-GCLM7	Ap^r; pGEM-T Easy derivative carrying gene cassette (*orf_05*) of LM7 strain (PCR amplified with primers HS286 and HS287)	This study
pGEM-T Easy	Ap^r; *oriV* ColE1; *lacZ*α; cloning vector	Promega

Materials and Methods

Bacterial Strains, Plasmids, and Culture Conditions

The following bacterial strains were used in this study: *Escherichia coli* BL21 (λDE3), TG1 [38], and DH5α λpir [6]; *Pseudomonas aeruginosa* PAO1161 and PAO1161R [3]; *Pseudomonas* spp. ZM1, ZM2, LM5, LM6, LM8, LM11, LM12, LM13, LM14, LM15, and LM25; *Pseudomonas mendocina* LM7 and *Pseudomonas aeruginosa* LM10 [27, 42]; and *Pseudomonas* sp. LM13R and LM7R [12]. All strains were grown in Luria-Bertani (LB) medium [38] at 37°C (*E. coli* strains and *P. aeruginosa* PAO1161) or 30°C (other *Pseudomonas* strains). Where necessary, the medium was supplemented with antibiotics: ampicillin (100 µg/ml), chloramphenicol (20 µg/ml), kanamycin (50 µg/ml for *E. coli* strains or 500 µg/ml for strain PAO1161), and rifampicin (50 µg/ml). The plasmids used and constructed in this study are listed in Table 1.

DNA Manipulations, PCR Amplification, and Introduction of Plasmid DNA into Bacterial Cells

Plasmid DNA isolation and common DNA manipulation techniques were performed as described by Sambrook and Russell [38]. PCR amplification was performed with the primers listed in Table 2. *E. coli* and *P. aeruginosa* strains were transformed using the methods of Kushner [25] and Irani and Rowe [21], respectively. Triparental mating was performed as described previously [4].

Table 2. Oligonucleotide primers used in this study.

Primer	Sequence (5′ → 3′)[a]	Reference
HS286	TCSGCTKGARCGAMTTGTTAGVC	[40]
HS287	GCSGCTKANCTCVRRCGTTAGSC	[40]
Int1f	AAAACCGCCACTGCGCCGTTA	[15]
Int1r	GAAGACGGCTGCACTGAACG	[15]
Int3f	ACTTTCAGCACATGCG	[15]
Int3r	TCTGTGGACCCACAAAC	[15]
IntM2-D	CAACGGAGTCATGCAGATG	[41]
IntM2-U	GTGCAACGCATTTTGCAGG	[41]
LlexAPAO	gcCATATGCAGAAGCTGACGCCC	This study
LsosLM13	AGGGCTGATGGCTGTAGGTA	This study
LsosPAO	AGCGGCAGATGACGCACCAG	This study
L1PLM13	gcgaattcCTTGCTGTCGATGGCTAA	This study
L2PLM13	taggatccCTTGCTGTCGATGGCTAA	This study
RlexAPAO	taGAATTCTCAGCGCCGGATCACGC	This study
RsosLM13	TCACAGCATGTCGCCAAAGC	This study
RsosPAO	GAAAGCCGTGGTCTTCCAGGC	This study
R1PLM13	taggatccACCGCACAGATAACGACT	This study
R2PLM13	gcgaattcACCGCACAGATAACGACT	This study

[a]Nucleotides not complementary to the target sequences are shown in lowercase letters; restriction sites are underlined.

Identification and Cloning of DNA Regions Adjacent to GCLM7 and GCLM13

The identified gene cassettes (GCLM7 and GCLM13) were initially cloned into vector pDS132 to create the plasmids pDS-GCLM7 and pDS-GCLM13. Then, a kanamycin resistance gene cassette was inserted into the Eco31I site within GCLM7 and GCLM13 to produce plasmids pDS-GCLM7KM and pDS-GCLM13KM, respectively. The plasmids were then introduced via triparental mating into their respective strains LM7R and LM13R. Since these pDS-derivative plasmids were unable to replicate in *Pseudomonas* spp., the obtained transconjugants contained co-integrates generated by homologous recombination between the introduced plasmids and the chromosomes of the recipient strains.

In the next step, total DNA isolated from the co-integrates was digested with different restriction enzymes (ApaI, MfeI, MluI, NdeI, SacI, or SacII). The restriction fragment mixtures were self-ligated (plasmid rescue technique) or ligated with linearized vector pBluescript KSII and introduced via chemical transformation into *E. coli* strains DH5α λpir (strain enabling replication of the pDS-derivative plasmids) and TG1. The selected transformants contained the following plasmids: (i) pBS-SACIIA and pBS-SACIIB (strain LM7), and (ii) pBS-MFEI and pDS-SACI (strain LM13). These plasmids were used as templates for sequencing.

DNA Sequencing

Nucleotide sequences of the ILEs and PCR-amplified DNA fragments were determined using a dye terminator sequencing kit and an automated sequencer (ABI 377 Perkin Elmer) in the Laboratory of DNA Sequencing and Oligonucleotide Synthesis (oligo.pl firm) at the Institute of Biochemistry and Biophysics, Polish Academy of Sciences. The plasmid templates for sequencing were constructed by cloning the DNA fragments into the vectors pBluescript KSII, pDS132, or pGEM T-Easy (Promega, USA). Primer walking was used to obtain the complete nucleotide sequences of the cloned fragments.

Bioinformatic Analyses

Nucleotide sequences were analyzed using Artemis software [8]. Similarity searches were performed using the BLAST programs [2] provided by the NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Promoters were predicted using the BDGP Neural Network Promoter Prediction tool [35]. The CLANS tool was used for graphical sequence clustering [16]. For the identification of transmembrane domains, the PSORTb 3.0 localization prediction tool was used [47]. Phylogenetic analyses were performed using the MEGA6 program [43], applying the neighbor-joining algorithm with Kimura corrected distances and 1,000 bootstrap replicates.

β-Galactosidase Assay

Promoter activity was examined by assaying β-galactosidase activity in *P. aeruginosa* PAO1161 strains carrying transcriptional fusions with a *lacZ* reporter gene (in promoter probe vector

pCM132), as described previously [44].

DNA Blotting and Hybridization

The distribution of ILEs in *Pseudomonas* spp. genomes was analyzed by DNA blotting and hybridization. DNA dot blots were produced using a Bio-Dot apparatus (Bio-Rad) according to the manufacturer's instructions. Southern blotting was carried out as described by Sambrook and Russell [38]. Molecular probes specific for *orf_int*, *orf_gc1*, *orf_gc2*, and *orf_gc3* of LM13 were prepared by PCR amplification of selected DNA fragments (primer pairs listed in Table 2). These fragments were then gel-purified and labeled with digoxigenin (Roche). DNA-DNA hybridization and visualization of bound digoxigenin-labeled probes were carried out as recommended by the supplier (Roche).

Cloning of the *Pseudomonas aeruginosa lexA* Gene, and Expression and Purification of the Recombinant Protein

The *lexA* gene of *P. aeruginosa* PAO1161 was amplified by PCR (primers LlexA and RlexA; Table 2) and cloned into vector pET28b(+). The resulting plasmid pET-lexA was introduced into *E. coli* BL21 (λDE3). This strain was cultured, and expression of the His-tagged recombinant LexA protein was induced in log-phase cells by the addition of IPTG to a final concentration of 0.1 mM. After a 4-h induction at 37°C, the cells were collected by centrifugation and processed to purify the LexA-6His protein. The pelleted cells were resuspended in buffer A (50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 20 mM imidazole, 50 μM PMSF, 10 mM 2-mercaptoethanol, 0.1% Tween 20, and 10% glycerol) and lysed by sonication. Cell debris was removed by centrifugation, and the supernatant containing the recombinant protein was mixed with 50% Ni-NTA agarose (Qiagen) and incubated at 4°C for 30 min with gentle agitation to permit binding. The resin was then settled in a chromatographic column and washed three times with buffer A to remove nonspecifically bound proteins. The LexA-6His protein was eluted in successive 1 ml portions of buffer A containing an increasing concentration of imidazole (50, 100, 200, 250, and 500 mM).

Electrophoretic Mobility Shift Assay

The DNA binding activity of the LexA-6His protein was examined using the electrophoretic mobility shift assay (EMSA). The DNA fragments used for the assay were PCR amplicons containing the LexA-like box of the *Pseudomonas* sp. LM13 integron-like element or the LexA-box of PAO1161 (245 bp region upstream of the *lexA* gene, positive control) (Table 2). The total volume of binding reaction contained 10 ng of DNA fragment, 10 ng of competitor DNA, purified LexA-6His protein (5, 10, 20, 50, 100, or 200 ng), and 4 μl of (5×) EMSA buffer (100 mM HEPES-NaOH (pH 8.0), 100 mM Tris-HCl (pH 8.0), 500 mM KCl, 10 mM EDTA, 10 mM DTT, and 50% glycerol). Following a 30-min incubation at 30°C, the reactions were loaded on a non-denaturing polyacrylamide gel and separated by electrophoresis at 100 V. DNA fragments were visualized by staining with ethidium bromide.

Nucleotide Sequence Accession Numbers

The nucleotide sequences of the ILEs and the adjacent DNA regions of strains LM7 and LM13 have been annotated and deposited in GenBank (NCBI) with the respective accession numbers KU366689 and KU366690.

Results and Discussion

Identification of Integron Gene Cassettes

Thirteen *Pseudomonas* strains isolated from the Lubin copper mine and Zelazny Most post-flotation waste reservoir were examined for the presence of integrons. Genomic DNA isolated from each strain was screened by PCR with primers targeting integrase genes of class 1, 2, and 3 integrons (primer pairs Int1f/Int1r, IntM2-D/IntM2-U, and Int3f/Int3r, respectively) or integron gene cassettes (degenerate primers HS286 and HS287, complementary to the *attC* sites) (Table 2).

This analysis failed to detect any *intI* genes, but 10 products were obtained with use of primer pair HS286 and HS287 targeting putative integron gene cassettes. These products (ranging in size from 0.35 to 2.0 kb) were cloned into vector pGEM-T Easy and sequenced. In silico analysis of the obtained nucleotide sequences revealed that only two possessed characteristics typical for integron gene cassettes. The other DNA fragments resulted from amplification of random genomic regions, mostly representing internal parts of chromosomal genes (*e.g.*, encoding FMN-dependent NADH-azoreductase, ATP-dependent DNA helicase RecQ, and hypothetical proteins), which is not in line with the definition of integron gene cassettes (*i.e.*, DNA fragment containing promoterless open reading frame). This suggests that those were nonspecific PCR amplicons.

The two predicted gene cassettes, designated GCLM7 (LM7 strain) and GCLM13 (LM13), were nearly identical (99% nucleotide sequence identity). They contained one ORF, encoding a putative protein displaying a high level of amino acid sequence similarity to several related hypothetical proteins of unknown function with the COG4337 domain. Interestingly, homologous proteins are encoded by gene cassettes of class 1 integrons present in three bacterial strains (*Thauera* sp. strain B4 [17] and uncultured strains OricaB11 and OricaD1 [18]). Based on these findings, we hypothesized that strains LM7 and LM13 may contain integrons with embedded gene cassettes GCLM7 and GCLM13, respectively.

Identification and Sequence Analysis of Putative Integron-Like Elements

To investigate the genetic context of the GCLM7 and GCLM13 cassettes in the LM7 and LM13 genomes, we used a combination of homologous recombination, plasmid rescue, and sequencing, as described in Materials and Methods. This strategy resulted in chromosomal DNA sequences encompassing these cassettes: 8,574 bp from strain LM7 (contains GCLM7) and 9,184 bp from strain LM13 (contains GCLM13).

Detailed inspection of these nucleotide sequences identified several putative genes, which are described in Table S1. Both gene cassettes (GCLM7 and GCLM13) are located within highly conserved (99% nucleotide sequence identity), three-component gene arrays (*orf_gc1-3*), which exhibit a lower GC content than adjacent sequences (Fig. 1). These gene arrays are placed in close proximity to genes encoding homologous integrases (*orf_int*), which share 92% amino acid sequence identity. Within the LM13 genome, the putative integrase gene is located upstream of the gene array and is transcribed in the opposite direction, which is a feature of typical integrons. In contrast, the *orf_int* of LM7 is placed downstream of the *orf_gc1-3* gene cluster (Fig. 1).

Interestingly, additional complete or truncated genes encoding related integrases were also found within the sequenced DNA fragments: *orf8* of LM13, encoding an integrase sharing 95% amino acid sequence identity with the LM13 Orf_int, and truncated *orf9* of LM7, encoding a 58-amino-acid polypeptide sharing 85% amino acid sequence identity with the *N*-terminal region of the LM7 Orf_int.

In silico analysis of the LM7 and LM13 Orf_int amino acid sequences revealed the presence of a conserved C-terminal catalytic domain (DNA_BRE_C), which is found in several DNA breaking-rejoining enzymes, including type IB topoisomerases and tyrosine recombinases. Moreover, we identified a conserved "signature" amino acid tetrad, R-H-R-Y (positions 57-154-157-189 in the Orf_int of LM7 and LM13), which is characteristic of tyrosine recombinases, being responsible for their catalytic activity [31].

A typical feature of proteins of the tyrosine recombinase family is their ability to catalyze site-specific recombination between substrates sharing very limited DNA homology. This protein family gathers several subgroups of distinct recombinases, including the integrases of bacteriophages, integrative and conjugative elements (ICEs), and integrons [34]. To enable proper classification of the predicted LM7 and LM13 integrases, we performed sequence clustering using CLANS software [16]. This analysis distinguished seven major subgroups of tyrosine recombinases: (i) XerC/XerD-like recombinases, (ii) integrases of lambda-like phages, (iii) integrases of other phages, (iv) integrases of ICEs, (v) integrases of class 2 integrons, (vi) integrases of class 1 and 3 integrons, and (vii) integrases containing the DNA_BRE_C domain (including putative proteins identified in this study) (Fig. 2). BLASTp analysis revealed that DNA_BRE_C domain-containing integrases share only limited amino acid sequence identity (25–35%) with the integrases of known integrons (Fig. 2), whereas all integrases of class 1, 2, and 3 integrons share amino acid sequence identity of at least 41% [30].

Comparative analysis revealed synteny and sequence

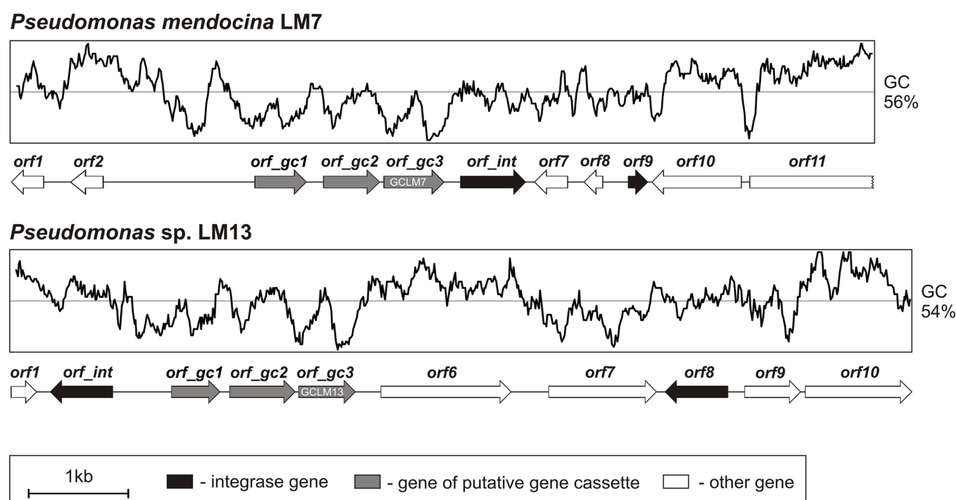

Fig. 1. Genetic organization of integron-like elements and adjacent DNA regions of *P. mendocina* LM7 and *Pseudomonas* sp. LM13. Arrows indicate the transcriptional orientation of the genes. The plots show the GC content of the sequences; the average GC content is indicated.

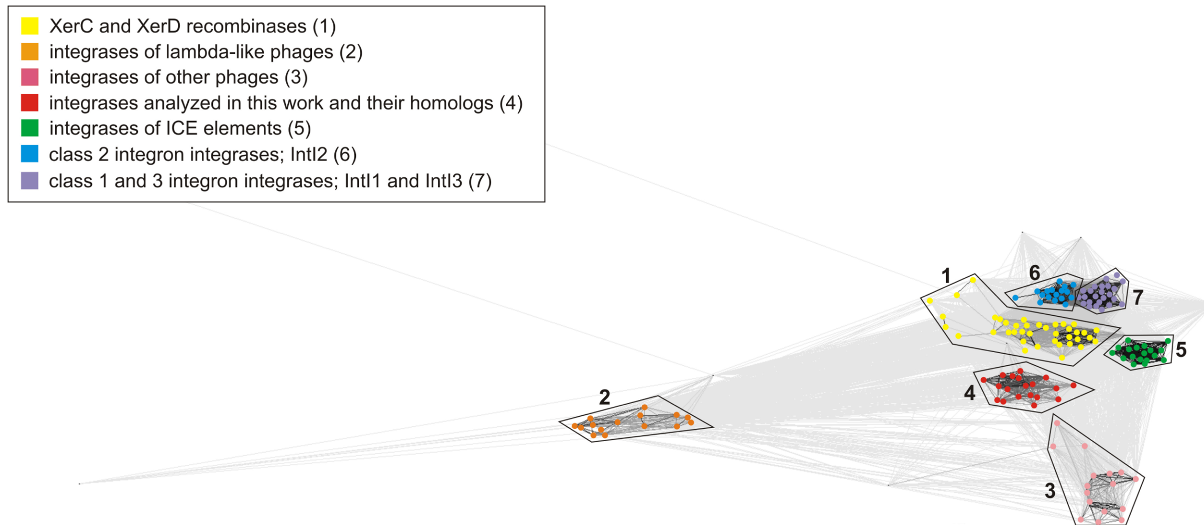

Fig. 2. CLANS clustering of 165 tyrosine recombinases.

Seven subgroups of tyrosine recombinases, indicated by different colors (as denoted in the key), were compared in this analysis. The number of sequences included in each subgroup is indicated in parentheses. Each cluster is numbered and surrounded by a solid border.

conservation of the predicted three-gene cassette arrays of strains LM7 and LM13. Besides the aforementioned GCLM7 and GCLM13 cassettes (which possess *orf_gc3*), they also contain genes *orf_gc1*, encoding proteins that share 99% amino acid sequence identity, and *orf_gc2*, encoding identical proteins. The putative Orf_gc1 proteins of LM7 and LM13 contain four predicted transmembrane helical domains and show similarity to several hypothetical membrane proteins of various gram-negative bacteria. The *orf_gc2* genes encode proteins with amino acid sequence similarity to GCN5 family acetyltransferases (E.C. 2.3.1.-). GCN5-related *N*-acetyltransferases (GNATs) are members of a large superfamily of *N*-acyltransferases that mostly catalyze the transfer of an acyl group to a substrate. These enzymes are responsible for a variety of functions; for example, several GNATs, such as aminoglycoside *N*-acetyltransferases, confer resistance to antibiotics [46].

A characteristic feature of the integron-associated gene cassettes is the presence of *attC* sites, which vary in size from 57 to 141 bp. These DNA regions may be identified by their location and the presence of imperfect palindromic sequences at their outer ends: RYYYAAC (R, purine; Y, pyrimidine) and GTTRRRY (where the point of recombination is between the G and T bases) [29, 40]. The *attC* sites consist of two potential core sites, named L′–L′′ and R′′–R′, separated by a central region [29].

In silico sequence analysis of the intergenic regions of the predicted GCLM7 and GCLM13 cassettes revealed that only the *orf_gc3* genes are associated with typical *attC* sites (Fig. 3), whereas a conserved motif (GTTRRRY) is present in the case of the *orf_gc1* and *orf_gc2* genes. The predicted *orf_gc3 attC* site has the potential to create a secondary structure with the extra-helical bases (G and T) (Fig. 3), which is required for DNA strand recognition by the integrase [22]. It is noteworthy that this putative *attC* site shares 94% nucleotide sequence identity with the recombination site identified downstream of the *orf186* gene cassette (99% sequence identity with *orf_gc3*) embedded within the class 1 integron of *Thauera* sp. strain B4 [17].

Based on these findings, we speculate that the identified gene clusters might constitute arrays of integron gene cassettes, which together with the accompanying integrase genes may form a novel group of ILEs.

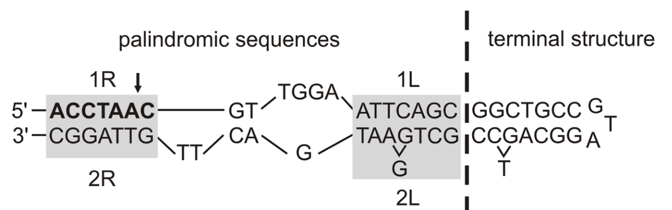

Fig. 3. Potential secondary structure of the bottom strand of the putative *attC* recombination sites identified downstream of the *orf_gc3* genes in strains LM7 and LM13.

The inverted repeats L′, L′′, R′, R′′ (putative integrase binding sites) are shaded gray. The arrow indicates the putative recombination crossover site. The highly conserved motif RYYYAAC is shown in bold font.

Identification of Promoters within the LM13 Integron-Like Element

Detailed inspection of the nucleotide sequence of the LM13 ILE using the BDGP promoter prediction tool [35] led to the identification of a putative P_c promoter [TTGGCA(N)$_{17}$CATTAT, coordinates 1027-1055], displaying sequence similarity to one of the proposed P_c variants of class 1 integrons, P_cIn42 [TTGGCA(N)$_{17}$TAAACT] [23]. However, no putative promoter of the LM13 ILE integrase gene (P_{int}) could be identified by in silico analysis.

To confirm the activity of the predicted P_c promoter and to test for the presence of P_{int}, an 810 bp region of the LM13 ILE (containing a 165 bp fragment of the initial part of *orf_int* and the complete *orf_int – orf_gc1* intergenic region; Fig. 1) was amplified by PCR and inserted into the broad-host-range promoter probe vector pCM132 (functional in *Pseudomonas* spp.) to generate transcriptional fusions with a promoterless *lacZ* reporter gene. The DNA fragment was cloned into the vector in both orientations: the "A" orientation to test the putative P_c promoter (pCM132-1A), and the "B" orientation to identify the presence of a putative (inversely oriented) P_{intI} promoter (pCM132-1B). The two resulting plasmids were introduced into *P. aeruginosa* PAO1161, and β-galactosidase assays were performed to assess promoter activity.

Significantly increased β-galactosidase activity was observed in PAO1161 containing pCM132-1A (309.4 ± 3.6 Miller units) compared with the strain containing control plasmid pCAB132 (82.5 ± 3.4 Miller units; PAO1161 basal β-galactosidase activity). The strain containing pCM132-1B showed no increase in β-galactosidase activity (85.2 ± 2.1 Miller units), which suggested that the cloned fragment does not carry the P_{int} promoter. These results indicated the presence of a P_c promoter that most probably drives the

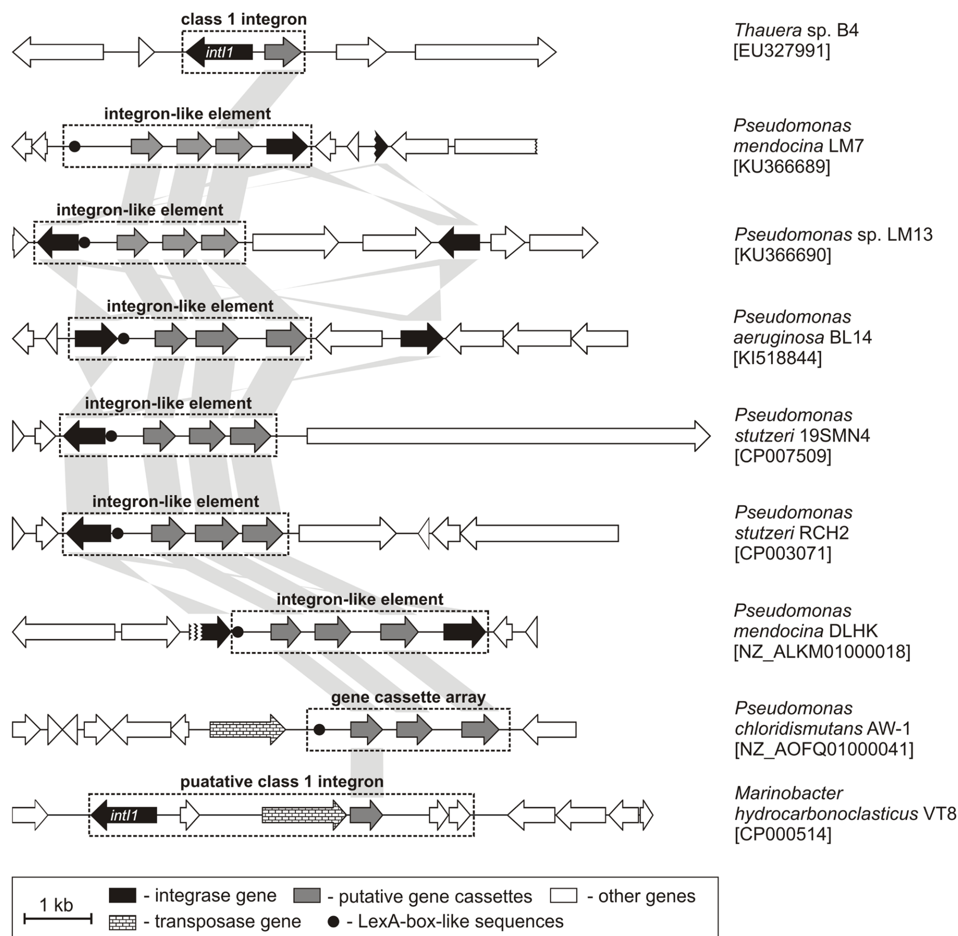

Fig. 4. Comparative analysis of the integron-like elements identified in *P. mendocina* LM7 and *Pseudomonas* sp. LM13.

Arrows indicate the transcriptional orientation of the genes. The gray-shaded areas connect homologous regions within identified integron-like elements and class 1 integrons.

expression of the ILE gene cassettes.

Identification and Analysis of a Putative LexA Box within the LM13 Integron-Like Element

In many integrons, the transcription of the *intI* gene is controlled by the bacterial stress (SOS) response. Under non-inducing conditions, the global transcriptional regulator LexA (repressor of the SOS regulon) binds to a specific LexA box overlapping P_{intI} and prevents transcription of the integrase gene [19].

Using the sequence $CTGT(N)_8ACAG$ as a query (consensus LexA-box sequence for Gammaproteobacteria) [14], we searched for putative LexA binding sites within the aforementioned 810 bp DNA fragment of the LM13 ILE. The partially matching sequence C<u>TTG</u>$(N)_7$ACAG (position 1026-1040) was identified, overlapping the predicted P_c promoter (nucleotides in common with the -35 P_c hexamer underlined). Interestingly, identical LexA-box-like sequences were also found within several other related ILEs identified in the genomic sequences of other *Pseudomonas* strains. In each case, this conserved box is located upstream of *orf_gc1*, regardless of the presence or orientation of the integrase gene (Fig. 4).

To verify whether these DNA regions constitute a target site for the LexA protein, EMSA was performed. Since the genomic sequence of strain LM13 is unknown, the *lexA* gene of another *Pseudomonas* strain was used for LexA protein purification. The amino acid sequence of the LexA repressor from *P. aeruginosa* PAO1161 (Acc. No. NP_251697) is 100% identical to the LexA protein from *Pseudomonas aeruginosa* BL14 (Acc. No. ERV48649), which carries a homologous ILE like strain LM13 (Fig. 4). Moreover, within the BL14 ILE, an identical (like in LM13) LexA box was identified.

The EMSA analysis revealed that LexA-6His did not interact with the LM13 LexA-box-like sequence (Fig. S1). Nevertheless, the high conservation of this sequence and its location within the predicted P_c promoters strongly suggest its importance in the expression of the gene cassettes of a group of related ILEs. It may be speculated that these sequences constitute a target site for an as yet unidentified cellular regulator.

Distribution of Putative Integron-Like Elements in *Pseudomonas* spp. Genomes

DNA blotting and hybridization analysis were performed to examine the distribution of the identified ILEs in the genomes of *Pseudomonas* spp. originating from the Lubin mine and Zelazny Most post-flotation reservoir. These analyses revealed that sequences related to the *orf_int* (integrase gene) of LM13 are present in six (out of 13) analyzed strains (ZM1, LM5, LM7, LM12, LM13, and LM15), whereas the *orf_gc1-3* gene array of LM13 is present in four strains (LM5, LM7, LM12, and LM13) (data not shown).

BLAST searches were also performed to identify related DNA sequences in the GenBank database. Four homologous elements, composed of an integrase gene and putative three-gene cassettes, were found in different *Pseudomonas* spp. (*P. aeruginosa* BL14, *P. stutzeri* 19SMN4, *P. stutzeri* RCH2, and *P. mendocina* DLHK). Interestingly, one strain (*P. chloridismutans* AW-1) carried only the conserved array of gene cassettes, and no adjacent integrase gene was identified (Fig. 4). The integrases of the identified ILEs (as well as related putative proteins encoded by the predicted gene cassettes) share at least 85% amino acid sequence identity.

It is worth mentioning that genes homologous to *orf_gc1* and *orf_gc3* are also present in the genomes of *Thauera* sp. B4 and *Marinobacter hydrocarbonoclasticus* VT8, respectively, where they are located in the vicinity of a class 1 integron integrase gene (Fig. 4).

Comparative genomic analysis identified three structural variants of the newly characterized ILEs: (i) with the integrase gene located downstream and in the same orientation as the predicted gene array (in *P. mendocina* LM7 and DLHK), (ii) with the integrase gene located upstream and in the same orientation as the predicted gene array (in *P. aeruginosa* BL14), and (iii) with the integrase gene located upstream and in the opposite orientation to the predicted gene array (in *Pseudomonas* sp. LM13, and *P. stutzeri* 19SMN4 and RCH2). Notably, all of the putative ILEs identified in this study were found in strains of *Pseudomonas* spp., which suggests that those ILEs may be genus specific.

In conclusion, it is worth emphasizing that, to our best knowledge, only three integrons of *Pseudomonas* spp. isolated from non-nosocomial habitats have been characterized so far [10, 20, 45]. These contain gene cassettes with unknown functions and are not associated with resistance phenotypes. Moreover, a group of ILEs (not related with elements identified in this work) present in several environmental pseudomonads has recently been described, whose members contain a XerD-like recombinase gene and a variable region encoding some adaptive traits [36].

In this study, we have characterized a novel group of putative ILEs, which show no sequence similarity to previously described elements. These ILEs share some

characteristics with typical integrons: they encode a putative tyrosine recombinase, contain *attC* sites (or at least the conserved motif GTTRRRY) within the predicted gene cassettes, and possess a strong promoter for expression of the cassette array. However, unlike integrons, they lack a recognizable *attI* recombination site, and their tyrosine recombinases show no homology to integron integrases identified previously. All of the identified putative ILEs contain a conserved array of gene cassettes, which suggests that they are somehow tightly linked with the elements. In future studies, it will be necessary to determine whether these ILEs are able to mobilize and incorporate genetic cassettes: two features of typical integrons. If they possess such abilities, they may constitute a novel group of integrons; otherwise, they may be recognized as an orphan integron gene cassette array located within a potentially mobile DNA region or recombination hot spot where other genetic elements are integrating.

Acknowledgments

We thank R. Matlakowska and A. Sklodowska for providing *Pseudomonas* strains ZM1, ZM2, LM5, LM6, LM7, LM8, LM10, LM11, LM12, LM13, LM14, LM15, and LM25, and A.A. Bartosik and G. Jagura-Burdzy for providing strains PAO1161 and PAO1161R and plasmid pCAB132. This work was supported by the National Science Center, Poland (Grant No. N N303 579238).

References

1. Alting-Mees MA, Short JM. 1989. pBluescript II: gene mapping vectors. *Nucleic Acids Res.* **17:** 9494.
2. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389-3402.
3. Bartosik AA, Glabski K, Jecz P, Mikulska S, Fogtman A, Koblowska M, Jagura-Burdzy G. 2014. Transcriptional profiling of ParA and ParB mutants in actively dividing cells of an opportunistic human pathogen *Pseudomonas aeruginosa*. *PLoS One* **9:** e87276.
4. Bartosik D, Szymanik M, Wysocka E. 2001. Identification of the partitioning site within the *repABC*-type replicon of the composite *Paracoccus versutus* plasmid pTAV1. *J. Bacteriol.* **183:** 6234-6243.
5. Boucher Y, Labbate M, Koenig JE, Stokes HW. 2007. Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol.* **15:** 301-309.
6. Bullock WO, Fernandez JM, Short JM. 1987. XL1-Blue – a high-efficiency plasmid transforming *recA Escherichia coli* strain with β-galactosidase selection. *BioTechniques* **5:** 376-378.
7. Cambray G, Guerout AM, Mazel D. 2010. Integrons. *Annu. Rev. Genet.* **44:** 141-166.
8. Carver T, Berriman M, Tivey A, Patel C, Bohme U, Barrell BG, *et al.* 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **24:** 2672-2676.
9. Cho HH, Kwon GC, Kim S, Koo SH. 2015. Distribution of *Pseudomonas*-derived cephalosporinase and metallo-β-lactamases in carbapenem-resistant *Pseudomonas aeruginosa* isolates from Korea. *J. Microbiol. Biotechnol.* **25:** 1154-1162.
10. Coleman NV, Holmes AJ. 2005. The native *Pseudomonas stutzeri* strain Q chromosomal integron can capture and express cassette-associated genes. *Microbiology* **151:** 1853-1864.
11. Dziewit L, Adamczuk M, Szuplewska M, Bartosik D. 2011. DIY series of genetic cassettes useful in construction of versatile vectors specific for Alphaproteobacteria. *J. Microbiol. Methods* **86:** 166-174.
12. Dziewit L, Pyzik A, Matlakowska R, Baj J, Szuplewska M, Bartosik D. 2013. Characterization of *Halomonas* sp. ZM3 isolated from the Zelazny most post-flotation waste reservoir, with a special focus on its mobile DNA. *BMC Microbiol.* **13:** 59.
13. Dziewit L, Pyzik A, Szuplewska M, Matlakowska R, Mielnicki S, Wibberg D, *et al.* 2015. Diversity and role of plasmids in adaptation of bacteria inhabiting the Lubin copper mine in Poland, an environment rich in heavy metals. *Front. Microbiol.* **6:** 152.
14. Erill I, Escribano M, Campoy S, Barbe J. 2003. In silico analysis reveals substantial variability in the gene contents of the gamma proteobacteria LexA-regulon. *Bioinformatics* **19:** 2225-2236.
15. Fonseca EL, Vieira VV, Cipriano R, Vicente AC. 2005. Class 1 integrons in *Pseudomonas aeruginosa* isolates from clinical settings in Amazon region, Brazil. *FEMS Immunol. Med. Microbiol.* **44:** 303-309.
16. Frickey T, Lupas A. 2004. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20:** 3702-3704.
17. Gillings M, Boucher Y, Labbate M, Holmes A, Krishnan S, Holley M, Stokes HW. 2008. The evolution of class 1 integrons and the rise of antibiotic resistance. *J. Bacteriol.* **190:** 5095-5100.
18. Gillings MR, Xuejun D, Hardwick SA, Holley MP, Stokes HW. 2009. Gene cassettes encoding resistance to quaternary ammonium compounds: a role in the origin of clinical class 1 integrons? *ISME J.* **3:** 209-215.
19. Guerin E, Cambray G, Sanchez-Alberola N, Campoy S, Erill I, Da Re S, *et al.* 2009. The SOS response controls integron recombination. *Science* **324:** 1034.
20. Holmes AJ, Holley MP, Mahon A, Nield B, Gillings M, Stokes HW. 2003. Recombination activity of a distinctive integron-gene cassette system associated with *Pseudomonas*

stutzeri populations in soil. *J. Bacteriol.* **185:** 918-928.
21. Irani VR, Rowe JJ. 1997. Enhancement of transformation in *Pseudomonas aeruginosa* PAO1 by Mg^{2+} and heat. *BioTechniques* **22:** 54-56.
22. Johansson C, Kamali-Moghaddam M, Sundstrom L. 2004. Integron integrase binds to bulged hairpin DNA. *Nucleic Acids Res.* **32:** 4033-4043.
23. Jove T, Da Re S, Denis F, Mazel D, Ploy MC. 2010. Inverse correlation between promoter strength and excision activity in class 1 integrons. *PLoS Genet.* **6:** e1000793.
24. Koenig JE, Sharp C, Dlutek M, Curtis B, Joss M, Boucher Y, Doolittle WF. 2009. Integron gene cassettes and degradation of compounds associated with industrial waste: the case of the Sydney tar ponds. *PLoS One* **4:** e5276.
25. Kushner SR. 1978. An improved method for transformation of *E. coli* with ColE1 derived plasmids, pp. 17-23. *In* Boyer HB, Nicosia S (eds.). *Genetic Engineering*. Elsevier/North-Holland, Amsterdam.
26. Marx CJ, Lidstrom ME. 2001. Development of improved versatile broad-host-range vectors for use in methylotrophs and other gram-negative bacteria. *Microbiology* **147:** 2065-2075.
27. Matlakowska R, Sklodowska A. 2009. The culturable bacteria isolated from organic-rich black shale potentially useful in biometallurgical procedures. *J. Appl. Microbiol.* **107:** 858-866.
28. Matlakowska R, Sklodowska A. 2011. Biodegradation of Kupferschiefer black shale organic matter (Fore-Sudetic Monocline, Poland) by indigenous microorganisms. *Chemosphere* **83:** 1255-1261.
29. Mazel D. 2006. Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.* **4:** 608-620.
30. Nield BS, Holmes AJ, Gillings MR, Recchia GD, Mabbutt BC, Nevalainen KM, Stokes HW. 2001. Recovery of new integron classes from environmental DNA. *FEMS Microbiol. Lett.* **195:** 59-65.
31. Nunes-Duby SE, Kwon HJ, Tirumalai RS, Ellenberger T, Landy A. 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* **26:** 391-406.
32. Partridge SR, Tsafnat G, Coiera E, Iredell JR. 2009. Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS Microbiol. Rev.* **33:** 757-784.
33. Philippe N, Alcaraz JP, Coursange E, Geiselmann J, Schneider D. 2004. Improvement of pCVD442, a suicide plasmid for gene allele exchange in bacteria. *Plasmid* **51:** 246-255.
34. Rajeev L, Malanowska K, Gardner JF. 2009. Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. *Microbiol. Mol. Biol. Rev.* **73:** 300-309.
35. Reese MG. 2001. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput. Chem.* **26:** 51-56.
36. Rhodes G, Bosma H, Studholme D, Arnold DL, Jackson RW, Pickup RW. 2014. The *rulB* gene of plasmid pWW0 is a hotspot for the site-specific insertion of integron-like elements found in the chromosomes of environmental *Pseudomonas fluorescens* group bacteria. *Environ. Microbiol.* **16:** 2374-2388.
37. Rowe-Magnus DA, Mazel D. 2002. The role of integrons in antibiotic resistance gene capture. *Int. J. Med. Microbiol.* **292:** 115-125.
38. Sambrook J, Russell DW. 2001. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
39. Sklodowska A, Matlakowska R, Bal K. 2005. Extracellular polymer produced in the presence of copper minerals during bioleaching. *Geomicrobiol. J.* **22:** 1-9.
40. Stokes HW, Holmes AJ, Nield BS, Holley MP, Nevalainen KM, Mabbutt BC, Gillings MR. 2001. Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Appl. Environ. Microbiol.* **67:** 5240-5246.
41. Su J, Shi L, Yang L, Xiao Z, Li X, Yamasaki S. 2006. Analysis of integrons in clinical isolates of *Escherichia coli* in China during the last six years. *FEMS Microbiol. Lett.* **254:** 75-80.
42. Szuplewska M, Ludwiczak M, Lyzwa K, Czarnecki J, Bartosik D. 2014. Mobility and generation of mosaic non-autonomous transposons by Tn*3*-derived inverted-repeat miniature elements (TIMEs). *PLoS One* **9:** e105010.
43. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30:** 2725-2729.
44. Thibodeau SA, Fang R, Joung JK. 2004. High-throughput beta-galactosidase assay for bacterial cell-based reporter systems. *BioTechniques* **36:** 410-415.
45. Vaisvila R, Morgan RD, Posfai J, Raleigh EA. 2001. Discovery and distribution of super-integrons among pseudomonads. *Mol. Microbiol.* **42:** 587-601.
46. Vetting MW, de Carvalho LPS, Yu M, Hegde SS, Magnet S, Roderick SL, Blanchard JS. 2005. Structure and functions of the GNAT superfamily of acetyltransferases. *Arch. Biochem. Biophys.* **433:** 212-226.
47. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, *et al.* 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26:** 1608-1615.