Article
Chinese Science Bulletin
Bioinformatics
April 2011 Vol.56 No.10: 987–995
doi: 10.1007/s11434-011-4410-6

Accurate localization and excision of genomic islands in four strains of *Pseudomonas aeruginosa* and *Pseudomonas fluorescens*

SONG Lei[1,2] & ZHANG XueHong[1*]

[1] *Key Laboratory of Microbial Metabolism, Ministry of Education, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China;*
[2] *College of Life and Environment Sciences, Shanghai Normal University, Shanghai 200234, China*

Received August 17, 2010; accepted January 17, 2011

Mobile genomic islands (GIs) can be excised from the chromosome, then form a circular intermediate and be reintegrated into the chromosome by the GI internal integrase. Some mobile GIs can also be transferred into a new receptor cell by transformation, conjugation, or transduction. The action sites of the integrase are usually flanked direct repeats (DRs) of the GIs. Accurate localization of the flanking sequences is a precondition for determining the mobility of the GI. Mobile GIs are generally associated with transfer RNAs (tRNAs). Based on the correlation between flanking sequences and tRNA sequences, the flanking sequences of 11 putative mobile GIs in *Pseudomonas aeruginosa* PAO1, *P. aeruginosa* PA14, *P. fluorescens* Pf-5 and *P. fluorescens* Pf0-1 were identified. Among the 11 GIs, Pf0-1GI-1 is responsible for benzoate degradation. PAO1GI-1, Pf5GI-2, Pf5GI-3, and Pf5GI-4 were confirmed experimentally to be excised from a chromosome to form a circular intermediate. The action sites of the integrases are these GIs direct repeats. Due to distinct DRs, cutting sites for the internal integrase of PAO1GI-1, Pf5GI-2, Pf5GI-3 and Pf5GI-4 were determined outside the T-loop of the $tRNA^{Gly}$ gene, outside the anticodon loop of the $tRNA^{Ser}$ gene and $tRNA^{Lys}$ gene, and at the asymmetric 3′-end of the $tRNA^{Leu}$ gene, respectively. PAO1GI-1 and other mobile GIs may be transferred into many different strains that belong to different phyla because of the clear flanking sequences. This study describes basic information about the action sites of the integrases, assesses the mobility of GIs, and can help design and transfer mobile GIs to candidate strains.

***Pseudomonas*, genomic islands, accurate localization, excision, integrase**

Citation: Song L, Zhang X H. Accurate localization and excision of genomic islands in four strains of *Pseudomonas aeruginosa* and *Pseudomonas fluorescens*. Chinese Sci Bull, 2011, 56: 987–995, doi: 10.1007/s11434-011-4410-6

Horizontal gene transfer (HGT) from other sources, such as other prokaryotes, viruses, and even eukaryotes, has an important role in the evolution of prokaryotic genomes [1]. Genomic islands (GIs) are the products of horizontal gene transfer and are present in certain strains, but are absent in most closely related variants in prokaryotic chromosomes [2]. Consequently, they are critically important materials to study the evolution of microbial genomes. Various GIs have now revealed abundant information regarding accessory functions, such as pathogenicity, fitness, antibiotic resistance, symbiosis, ion uptake, xenobiotic degradation, primary and secondary metabolic activities, secretion activities, and other special features of prokaryotic genomes [3]. Some GIs can also be excised to form circular intermediates and are defined as mobile GIs [2,4]. Internal integrases catalyze the GIs excision directed at the flanking direct repeats (DRs) of GIs. A number of these mobile GIs can in turn be reintegrated into a new receptor cell by transformation, conjugation or transduction [2]. Mobile GIs often have the following basic characteristics: they are relatively large segments of DNA, (ranging in size from 5–500 kb); they may be identified by abnormal GC content, codon usage, and dinucleotide bias; they are often associated with tRNA genes and flanked by DRs; and they harbor some mobile

*Corresponding author (email: xuehzhang@sjtu.edu.cn)

elements like integrases (*int*), insertion elements (IS), and transposons [2,5].

With the evolution and rearrangement of prokaryotic genomes, some GIs become immobile because of mutations and deletions in important mobility elements, such as integrases and the GIs flanking sequences. Determination of clear flanking sequences becomes important, and the presence of the integrase in GIs must be ensured when performing searches for mobile GIs.

Pseudomonas spp. is a ubiquitous Gram-negative, non-spore forming, rod-shaped, mostly aerobic flagellated genus of *Proteobacteria*, normally found in soil, water, and eukaryotic hosts (humans, animals, and plants). Because of their widespread occurrence in nature, Pseudomonads are able to produce a wide variety of useful secondary metabolites. The current study investigated two Pseudomonads, *Pseudomonas aeruginosa*, an opportunistic pathogen of humans, plants, and animals, and *P. fluorescens,* a plant symbiont. *P. aeruginosa* is the most significant cause of chronic nosocomial infections in patients and its pathogenicity is due to the ability to produce a wide range of virulence factors [6].

The complete sequenced genome of *P. aeruginosa* PAO1 was published in 2000 [7]. Genomes of *P. aeruginosa* PAO1 and PA14 have been studied, and are very similar, but PA14 is significantly more virulent than PAO1 and has a wider range of hosts [8]. Fluorescent Pseudomonads are species that are often non-pathogenic and usually lack virulence factors. Fluorescent Pseudomonads have been investigated for their potential use in the bioremediation of various organic compounds, for example, the biodegradation of pollutants and biocontrol of agricultural pathogens. Currently, complete genomic sequences are available for *P. fluorescens* Pf-5 and *P. fluorescens* Pf0-1 and these are the model strains of *P. fluorescens*. *P. fluorescens* Pf-5 produces a variety of secondary metabolites, including various antibiotics, as well as several toxic products that suppress pathogens in the rhizosphere [9]. Therefore, *P. fluorescens* Pf-5 is studied as an organism for biological control or the control of rhizosphere colonization. In contrast, *P. fluorescens* Pf0-1 was isolated from agricultural soil and has long served as a model microorganism for analyzing bacterial proliferation and persistence in soils [10].

In the present study, we determined clear flanking sequences of 11 GIs in *P. aeruginosa* PAO1, *P. aeruginosa* PA14, *P. fluorescens* Pf-5 and *P. fluorescens* Pf0-1. We show that 1 GI of *P. aeruginosa* PAO1 and 3 GIs of *P. fluorescens* Pf-5 can be excised from the chromosome to form a circular intermediate by the crossing-combining approach of primers [11]. Through the accurate localization of the GI flanking sequences, the action sites of the integrases of GIs can be predicted. Thus, excision of GIs from their chromosomal locations and the transference of GIs into appropriate receptors can be determined.

1 Materials and methods

1.1 Search strategy

Complete genomic sequences of 4 *Pseudomonas* strains were retrieved and downloaded from the *Pseudomonas* Genome Database V2 (http://www.pseudomonas.com/). For all four of these strains, the tRNA genes were retrieved from the Genomic tRNA Database (http://gtrnadb.ucsc.edu/) and tmRNA gene sequences were retrieved from the tmRNA Website (http://www.indiana.edu/~tmrna/). A BLASTN-based strategy was employed to determine the candidate GIs flanked by direct repeats that are related to the tRNA/tmRNA gene. Each tRNA or tmRNA gene from *P. aeruginosa* PAO1 was aligned with the complete sequence of *P. aeruginosa* PA14 using BLASTN algorithms (http://www.pseudomonas.com/blast.jsp) with default settings. Each hit and its downstream 13-bp sequences were searched as a query in the *P. aeruginosa* PA14 genome. If there were two or more identical hits (or its downstream 13-bp sequences) in this complete sequence and the genomic distance between these was 5–500 kb, the region between two identical hits (or its downstream 13-bp sequences) was defined as a candidate GI. A detailed search procedure for a candidate GI is described in Figure 1. Gene loci for putative integrases or transposases were retrieved from GenBank (http://www.ncbi.nlm.nih.gov). The candidate GIs were retained if integrases or transposases were present within the proximity of the DRs. Finally, GIs were identified by abnormal G+C content and dinucleotide frequency. Compositional bias of the dinucleotide frequency and G+C content were analyzed by the web-based application deltarho (http://deltarho.amc.uva.nl) [12]. This method calculates the genomic dissimilarity value, δ^* (the average dinucleotide relative abundance difference between the GI sequence and the corresponding genome sequence), where a high δ^* reveals a heterologous origin of the GI sequences. The δ^* value is affected by the length of the input sequence and the percentage of the genomic fragments with lower genomic dissimilarity than the input GI sequence was calculated.

1.2 Integrase analysis

The conserved domain database (CDD) is a resource for the functional annotation of proteins. The protein sequences of the integrases found in GIs were annotated by CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) [13]. CDD not only includes National Center for Biotechnology Information (NCBI)-curated domain models but also imports domain models from external sources, such as Pfam, SMART, TIGRFAM and NCBI Protein Clusters database.

1.3 Bacterial strain and culture conditions

Pseudomonas aeruginosa PAO1 and *Pseudomonas fluo-*

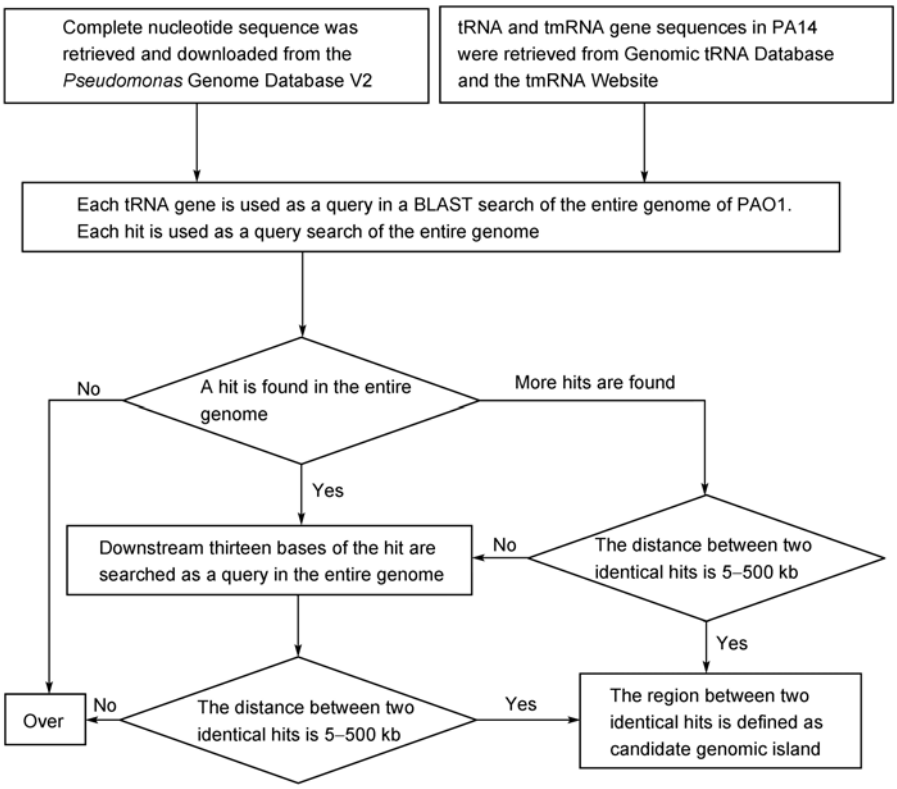

Figure 1 Process for searching for candidate genomic islands.

rescens Pf-5 strains were used in this study. *P. aeruginosa* PAO1 was grown at 37°C and *P. fluorescens* Pf-5 was grown at 28°C in Luria-Bertani (LB) medium. Solid medium contained 1.5% agar [14].

1.4 PCR amplification and sequencing

An overnight bacterial culture was used as the DNA template for PCR. Amplification reactions were carried out with 2 μL of bacterial culture, 3 μL of 2.5 mmol/L deoxynucleoside triphosphate, 15 μL of 2×GC Buffer I (or 2×GC Buffer II), 100 pmol of primers, and 1 U of TaKaRa LA DNA Polymerase (TaKaRa Biotechnology, China). Distilled water was added to bring the final volume to 30 μL. The PCR conditions used for amplification were initial denaturation at 94°C for 10 min, followed by 45 cycles of 94°C for 30 s, 50–60°C (depending on the primer pair) for 30 s and 72°C for a time based on the size of the expected fragment (1 min/kb), finally 72°C for 10 min. The reaction products were subjected to electrophoresis in a 1.0%–1.5% agarose gel (according to length of products), stained with GoldView™, and visualized under UV light. The PCR fragment was then purified from the agarose gel using an AxyPrep™ DNA Gel Extraction Kit (Axygen, China). The nucleotide sequences for both DNA strands of the PCR products were determined using an automated DNA sequencer (Invitrogen Biotechnology, China).

2 Results

2.1 GIs associated with clear flanking sequences

Eleven genomic islands ranging in size from 5.8–115.6 kb were identified in 4 strains of *P. aeruginosa* and *P. fluorescens*. Eleven GIs, namely PAO1GI-1, PAO1GI-2, PA14GI-1, PA14GI-2, PA14GI-3, PA14GI-4, Pf5GI-1, Pf5GI-2, Pf5GI-3, Pf5GI-4, and Pf0-1GI-1, were recognized using a pre-determined search strategy (Table 1). The G+C contents of all 11 GIs were significantly lower than the overall chromosomal G+C contents. Seven GIs identified in 3 strains, namely *P. aeruginosa* PAO1, *P. fluorescens* Pf-5 and Pf0-1, and one from *P. aeruginosa* PA14, possessed a very high genomic dissimilarity value (δ^*) compared with the average value for corresponding genomes (fragment length is 20000 bp). The percentage of genomic fragments with lower genomic dissimilarity ranged from 79.1%–100% and revealed that all GIs were remarkably dissimilar to the core genome. Gene densities of PAO1GI-1, PAO1GI-2, PA14GI-4, Pf5GI-2 and Pf0-1GI-1 were higher than the gene densities of their complete genomes. In contrast, gene densities for PA14GI-1, PA14GI-2, PA14GI-3, Pf5GI-1, Pf5GI-3 and Pf5GI-4 were lower than the gene densities of their complete genomes (Table 1). Consequently, the 11 GIs identified were verified to be reliable GIs.

Table 1 Composition of GIs identified in *P. aeruginosa* PAO1, *P. aeruginosa* PA14, *P. fluorescens* Pf-5 and *P. fluorescens* Pf0-1

Island	Annotation bp position (Start and end of locus tag)	Insertion site (Locus tag)	*Int* or *tnp*	Size (bp)	GC%	Gene density (numbers/kb)	$\delta^{*a)}$ (×1000)	%genomic fragments with lower δ^*
Pseudomonas aeruginosa PAO1				6264404	66.6%	0.905	26.783[a]	
PAO1GI-1	785311-797747 (PA0715..PA0729.1)	$tRNA^{Gly}$ (PA0729.1)	PA0728	12437	56.39%	1.206	94.069	100%
PAO1GI-2	1060411-1069382 (PA0976.1-PA0987)	$tRNA^{Lys}$ (PA0976.1)	PA0978 PA0987	8972	53.53%	1.226	88.700	99.140%
Pseudomonas aeruginosa PA14				6537648	66.3%	0.914	27.910[a]	
PA14GI-1	290225-312990 (PA14_03250-PA14_03410)	$tRNA^{Arg}$ (PA14_03410)	PA14_03300	22766	58.55%	0.703	35.182	79.094%
PA14GI-2	4750058-4755861 (PA14_53570-PA14_53610)	tmRNA (4749731.. 4750083)	PA14_53570	5804	47.36%	0.861	87.185	96.98%
PA14GI-3	4860199-4895087 (PA14_54840-PA14_55090)	$tRNA^{Gly}$ (PA14_54840)	PA14_55060 PA14_55090	34889	61.47%	0.659	36.453	93.583%
PA14GI-4	5251438-5359447 (PA14_58910-PA14_60150)	$tRNA^{Lys}$ (PA14_60150)	PA14_60140	108010	59.73%	1.065	33.258	100%
Pseudomonas fluorescens Pf-5				7074893	63.3%	0.881	31.300[a]	
Pf5GI-1	2042156-2050551 (PFL_1841-PFL_1846)	$tRNA^{Ser}$ (PFL_1841)	PFL_1843	8396	46.38%	0.596	91.743	98.456%
Pf5GI-2	4338363-4395051 (PFL_3739-PFL_3794)	$tRNA^{Ser}$ (PFL_3739)	PFL_3794	56689	57.26%	0.988	56.187	98.387%
Pf5GI-3	5378464-5493562 (PFL_4658-PFL_4753)	$tRNA^{Other}$ (PFL_4658) $tRNA^{Lys}$ (PFL_4753)	PFL_4752	115099	56.36%	0.817	47.389	100%
Pf5GI-4	5728473-5745246 (PFL_4976-PFL_4984)	$tRNA^{Leu}$ (PFL_4976)	PFL_4977	16774	51.50%	0.477	72.404	97.862%
Pseudomonas fluorescens Pf0-1				6438405	60.5%	0.905	28.072[a]	
Pf0-1GI-1	3390830-3506476 (Pfl01_2950- Pfl01_r48)	$tRNA^{Leu}$ (Pfl01_r48)	Pfl01_3059 Pfl01_3064	115647	55.61%	0.977	49.430	100%

a) The superscript a denotes average δ^* value for the relative genome (fragment length is 20000 bp).

Flanking sequences are important markers of the action sites of integrases or transposases that can insert or delete a "mobile" GI. For this reason, it is necessary to define or characterize the flanking regions in the study of these GIs. The tRNA sequence of *P. aeruginosa* PA14 was used as a query for a BLAST search against the entire genome of *P. aeruginosa* PAO1. The hits of every tRNA and the downstream 13-bp sequences of every hit were searched for identical sequences within the genome, respectively. To this end, we were able to take into account the comparative genome results and inferred that the GI exists in certain strains but is absent in closely related strains. Flanking DRs were defined as the start and end sites of the 11 GIs (Table 1). Therefore, the 11 identified GIs are not only reliable but also have clear DRs as flanking sequences.

2.2 Pf0-1GI-1 is a xenobiotic-degradation GI

P. fluorescens Pf0-1 harbors a 115.6-kb GI, designated Pf0-1GI-1, which is integrated into the $tRNA^{Leu}$, and contains 113 coding sequences (CDSs). Among these, 69 genes were predicted as special functional genes, 42 genes were classified as hypothetical genes, and two genes were identified as pseudogenes. Pf0-1GI-1 contains a cluster of 10 genes, from Pfl01_2960 through Pfl01_2969 (*BenABCDEKR* and *catABC* gene clusters) that are involved in benzoate degradation (http://www.genome.jp/kegg/pathway.html). This gene cluster encodes products that are involved in the transport of benzoate into the cells and produce 2-oxo-2,3-dihydrofuran-5-acetate through 5 chemical reactions: BenE and BenK are transport proteins; BenABC catalyzes the first reaction of degradation; BenR is the *BenABC* operon transcriptional activator; and BenD produces catechol. The catABC catalyzes the catechol and finally produces 2-oxo-2,3-dihydrofuran-5-acetate through 3 reactions. Because of the xenobiotic characteristics of benzoate, Pf0-1GI-1 is defined as a xenobiotic-degradation GI.

2.3 Detection of the circular excision of GIs

Mobile GIs can be excised from the chromosome and be reintegrated into the chromosome using internal integrases, which have dual roles as an endonuclease and ligase (Figure 2(a)). In the current study, we selected a crossing-combining approach of primers to determine whether the GIs were mobile [11]. A set of PCR primers were designed to determine whether PAO1GI-1, Pf5GI-1, Pf5GI-2, Pf5GI-3, and Pf5GI-4 were present in the chromosome and whether they can be excised from the chromosome (Table 2). Primer pairs 1F+1R and 2F+2R were used to amplify the upstream and downstream junction sequences between the chromosome and the GI, respectively; primers 1R+2F were used to detect the presence of a circular GI, and primer pair 1F+2R

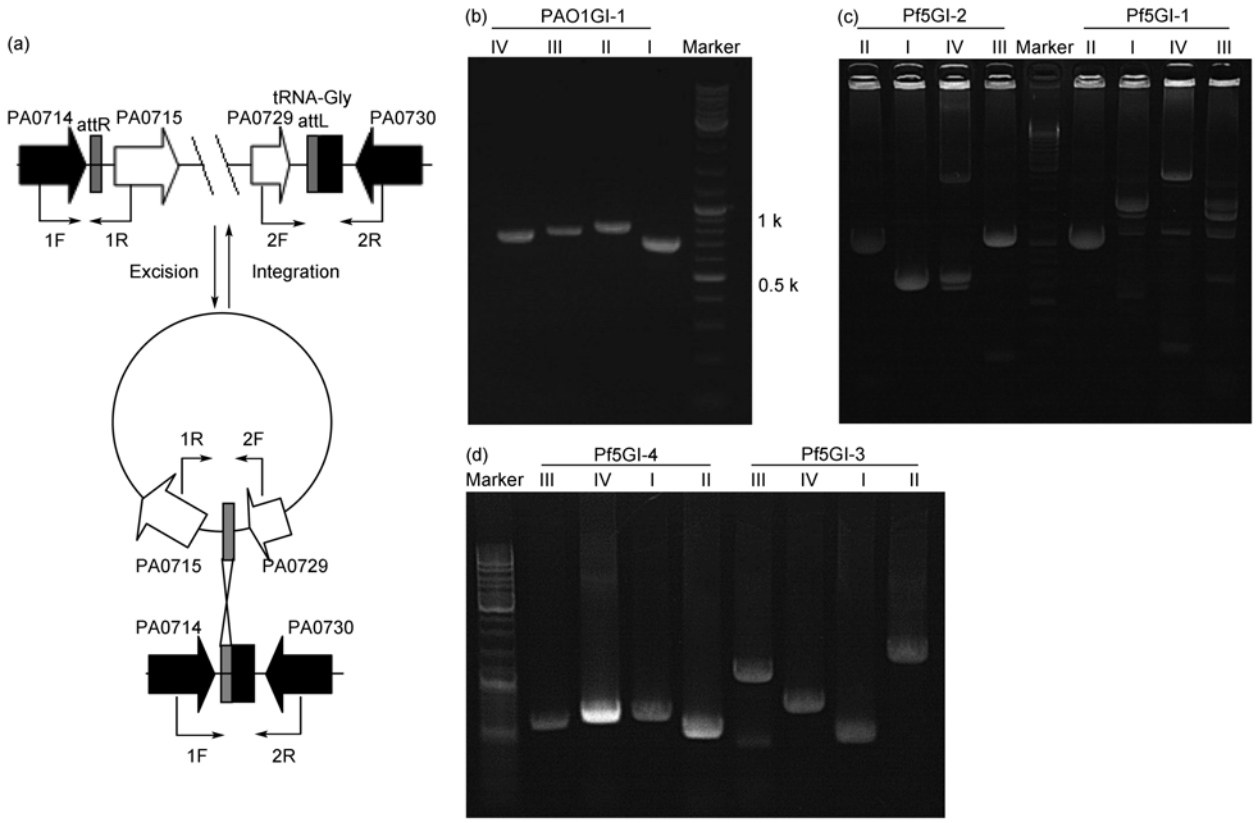

Figure 2 Detection of integrated and circular PAO1GI-1, Pf5GI-1, Pf5GI-2, Pf5GI-3, Pf5GI-4. (a) Schematic representation of a partial PAO1GI-1 region indicating the locations of primer pairs used to detect integrated and circularized PAO1GI-1. Black arrows represent core region genes. Boxes show the *att* sites. Arrowheads indicate the primers used for detection of integrated and circular PAO1GI-1 in *P. aeruginosa* PAO1. In (b), (c) and (d), I, II, III, and IV are PCR amplification products with primer pairs 1F+1R, 2F +2R, 1F+2R and 1R+2F, respectively.

was used to confirm the GI excision (Figure 2(a)). The 4 PCR products were found to be consistent with predicted sizes and were sequenced (Table 2 and Figure 2(b)–(d)). The sequences of the 4 PCR products demonstrated that PAO1GI-1, Pf5GI-2, Pf5GI-3 and Pf5GI-4 were in a circular form and reintegrated into the chromosome using the flanking DRs of the GIs. The 4 PCR products of Pf5GI-1 were present, but the limited III and IV products were not sequenced. The mobility of the 4 GIs was confirmed by the presence of clear flanking sequences.

2.4 Integrase analysis

Analysis of the internal integrases from the 11 GIs showed that a number of the internal integrases could be categorized as tyrosine-type integrases, while the others could be categorized as the Rve family of integrases (Table 3). The tyrosine recombinase/integrase family is one of two families of all site-specific recombinases. It utilizes a catalytic tyrosine to break and rejoin single strands in pairs to form a Holliday junction intermediate [15]. DRs are generally required to identify recombination partners. The integrase proteins from human immunodeficiency virus 1 (HIV-1) and avian sarcoma virus (ASV) have been the most extensively studied Rve family integrases. The integrases of retroviruses and the transposases of many transposons in both prokaryotes and eukaryotes show similar catalytic mechanisms in their transposition reactions [16]. Haren et al. [16] provided strong evidence that the DDE triad (aspartic acid, aspartic acid and glutamic acid) lies at the center of the catalytic site of the integrase in HIV-1.

Williams [17] suggested that integration sites can occur at the anticodon-loop, the T-loop, and the asymmetric 3′-end of tRNA genes. We determined that PAO1GI-1 can be excised from the genome and form circular elements that can be reintegrated into the core genome of *P. aeruginosa* PAO1 through DRs (5′-AGGGTTCGATTCCCTTCGCCC-GCTCCA-3′). Consequently, the repeat sequences at the 3′-end of $tRNA^{Gly}$ not only includes recognition loci but also the excision loci of PA0728. In this study, we infer that AGGG and CCCT are inverted sequences; therefore, as the binding locations of this integrase, the excision loci of this integrase are located outside of the T-loop (TTCGATT) of the $tRNA^{Gly}$ gene. The AGGGTTCGATTCCCT sequence was determined to be the essential sequence for the excision and reintegration of PAO1GI-1 within flanking DRs. The binding and cutting locations of the tyrosine family integrases of other GIs are predicted (Table 3). The cutting loci

Table 2 PCR primers used to detect the PAO1GI-1, Pf5GI-1, Pf5GI-2, Pf5GI-3 and Pf5GI-4 deletion and the circular forms of mobile elements

Name	Sequence (5′→3′)	Product size (bp)	PCR annealing temperature (°C)
PAO1GI-1 1F	AGTTATTGGTCGTGGTTGCTTC	708 (1F-1R)	53
PAO1GI-1 1R	CTACGGAGTAAGACGCCTAAAA	815 (2F-2R)	53
PAO1GI-1 2F	GGTGTTCCTGCCTATGTGGT	771 (1F-2R)	53
PAO1GI-1 2R	AACCCAGCGAAATCAACAAA	752 (1R-2F)	53
Pf5GI-1 1F	GATAGTCCGCCGACAGCA	1703(1F-1R)	52
Pf5GI-1 1R	ATATTAAACTCACGCAAACCTG	1075 (2F-2R)	52
Pf5GI-1 2F	GCAGTATCCTGGCTGACAAA	1678 (1F-2R)	52
Pf5GI-1 2R	CCTCCAACAACCGACCAGTA	1100 (1R-2F)	52
Pf5GI-2 1F	GAGTCACGGAGTTTGTTGGT	670 (1F-1R)	52
Pf5GI-2 1R	GTGGTGTCGGGTGGAATAG	1082 (2F-2R)	52
Pf5GI-2 2F	GGAAGGCAGGCGAGTTAC	1069 (1F-2R)	52
Pf5GI-2 2R	AGCATTGTCTCGCTCACTTAT	683 (1R-2F)	52
Pf5GI-3 1F	ACTGCGAAGTTACCGACCCT	464 (1F-1R)	52
Pf5GI-3 1R	AGTCCTATGATATGCCGTGAAA	1528 (2F-2R)	52
Pf5GI-3 2F	AGCCAAAGCCTGAACCTCT	1202 (1F-2R)	52
Pf5GI-3 2R	TGAACAAGCACAGAACCCATC	790 (1R-2F)	52
Pf5GI-4 1F	CGGAAAGTTTAGCGGATTA	692 (1F-1R)	52
Pf5GI-4 1R	GCCTTTAGCCAAGCGTCA	538 (2F-2R)	52
Pf5GI-4 2F	GCGGAAACCCAGTAGCGT	600 (1F-2R)	52
Pf5GI-4 2R	AGTGGCAACACCCTGACG	630 (1R-2F)	52

Table 3 Analysis of flanking direct repeats and integrases in eleven identified genomic islands [a)]

GIs	*Int* or *tnp*	Direct repeats (*attL*/*attR*) (5′→3′)	CDD	Integrase type
PAO1GI-1	PA0728	AGGGTTCGATTCCCTTCGCCCGCTCCA	Phage HP1 integrase	Tyrosine recombinase /integrase family
PAO1GI-2	PA0978	CGAATCCTACACGACCCACCAT	Rve	Rve family
	PA0987		Rve	Rve family
PA14GI-1	PA14_03300	GGGTCGTGGGTTCGAATCCCGCCGGATGCGCCATA	Rve	Rve family
PA14GI-2	PA14_53570	GGGT(G)TCAAATCCCCCCGGCTCCACCAAACG	DNA breaking-rejoining enzyme	Tyrosine recombinase /integrase family
PA14GI-3	PA14_55090	TTCCCTTCGCCCGCTCCA	Rve	Rve family
	PA14_55060		Rve	Rve family
PA14GI-4	PA14_60140	GAG(C)CAGTTGG(T)CTTTTAACCAATTG GTTCGTAGGTTCGAATCCTACACGACCCACCAT	DNA breaking-rejoining enzyme	Tyrosine recombinase /integrase family
Pf5GI-1	PFL_1843	TTCGAATCTCTCCTTCACCGCCACATTCTACAAC(A)C ACAAACCCCTGACTTTCCTAAAGAAAGTCGGGGGTTT GTGGTTTTTGGCGTCTGAAAAATGGCC	DNA breaking-rejoining enzyme	Tyrosine recombinase /integrase family or similar
Pf5GI-2	PFL_3794	CGGTCTTGAAAACCGCCGACTGTAACAGGTCCTAGAG TTCGAATCTCTACGCCTCCGCCAAATCTCAACGAA(G) AAAGCCCTGATTTTTCAGGGCTTTTTTGTGTCTG	DNA breaking-rejoining enzyme	Tyrosine recombinase /integrase family
Pf5GI-3	PFL_4752	CAG(A)TTGG(T)CTTTTAACCAATTGGTCGTAGGTTCG AATCCCACACGACCCACCATATT	DNA breaking-rejoining enzyme	Tyrosine recombinase /integrase family
Pf5GI-4	PFL_4977	AGTCTCCCTCGGGGCACCA	Bacteriophage P4 integrase	Tyrosine recombinase /integrase family
Pf0-1GI-1	Pfl01_3064	GGTTCGATTCCGGCTTCGGGCACCATCT	Bacteriophage P4 integrase	Tyrosine recombinase /integrase family
	Pfl01_3059		Rve	Rve family

a) Letters annotated by Lines (---) represent reverted sequences that are binding locations of the integrases. Letters annotated by Lines (—) represent possible cutting location of the integrases. Letters inside brackets represents letters of *attR*.

of integrases in PAO1GI-1, PA14GI-2, and Pf0-1GI-1 are located outside of the T-loop of the $tRNA^{Gly}$ gene, the tRNA part of the tmRNA gene, and $tRNA^{Leu}$ gene. The cutting loci of one integrase in Pf5GI-1 are outside of the T-loop of the $tRNA^{Ser}$ gene, but the binding loci of this integrase are incomplete. Therefore, it is difficult to excise Pf5GI-1 from the chromosome, which is consistent with the excision experimental result (Figure 2(c)). The cutting loci of inte-

grases in PA14GI-4, Pf5GI-2, and Pf5GI-3 are located outside the anticodon-loop of the $tRNA^{Lys}$ gene, $tRNA^{Ser}$ gene, and $tRNA^{Lys}$ gene, and the binding loci are also present. Integrase sites of PAO1GI-2, PA14GI-3, and Pf5GI-4 are at the asymmetric 3′-end of the $tRNA^{Lys}$ gene, the $tRNA^{Gly}$ gene, and the $tRNA^{Leu}$ gene. Through analysis of the integrase action sites, the important sites within flanking DRs were determined. The sequence of binding and cutting sites used by integrase is necessary to delete the GI from the chromosome. The integrase catalyzes the identical sequence of the receiving chromosome and integrates the GI into the chromosome. Therefore, the transformation of some useful functional GIs into appropriate receiving strains is possible.

3 Discussion

Eleven genomic islands were identified in *P. aeruginosa* PAO1, *P. aeruginosa* PA14, *P. fluorescens* Pf-5, and *P. fluorescens* Pf0-1 that had clear flanking DRs. The genomic location of six GIs in *P. aeruginosa* was identical to those in our previous report [18], but their mobility was not determined. A number of studies have also described the regions of GIs in these four strains [14,19–23]. These data have been presented in comparison with the 11 GIs of the present study (Table 4). The Islander program can determine the GIs that possess tRNA/tmRNA-type flanking sequence and tyrosine-type integrases [19]. However, we obtained more mobile GIs than ones identified by Islander. GIs found by IslandViewer did not contain flanking direct sequences compared with our results [20]. Mathee et al. [21] identified more GI regions in *P. aeruginosa* PAO1 and PA14 than the current study, but did not accurately localize these GIs. Mavrodi et al. [22] obtained GIs with incomplete flanking DRs in *P. fluorescens* Pf-5, in particular, Pf0-1GI-1 was divided into 2 parts [23]. Because the 11 GIs identified in this study have clear flanking DRs, the endpoints of the 11 GIs in four strains could be used to determine the binding and cutting sites of integrases or transposases in the GIs (Table 3). Thus, because accurate flanking sequences assist the design of effective primers, the mobility of GIs can be determined.

In the current study, Pf5GI-2, Pf5GI-3 and Pf5GI-4 were determined to be excised from and reintegrated into the corresponding chromosome. Webb et al. [14] extracted the replicative-form (RF) DNA of Pf4 (PAO1GI-1) from Pf4-infected cells, and used primers Pf4F and Pf4R to confirm recirculation of prophage Pf4 by the repeat sequence (3′-TGGAGCGGGCGAAGGGAATCGAACCCT-5′). This result was similar to our study, however, we selected a more precise crossing-combining approach of primers [11], and also determined that PAO1GI-1 (prophage Pf4) not only exists in the chromosome, but also can be excised from the chromosome to form a circular intermediate. Mathee et al. [21] also found a similar region (RGP5) with PAO1GI-1 (Table 4) that only determined the excision of RGP5 from the genome. Further, the circular intermediate was not detected by Mathee et al. [21] because GIs with unclear boundaries were identified and crossing-combining primers were nonspecific resulting in the inability to produce all four amplification products, including the circular intermediates. PA14GI-4, namely PAPI-1, was also determined similarly [11]. We also designed primers to determine the mobility of PAOGI-2 and Pf0-1GI-1, but our study revealed that these 2 GIs were not deleted from the chromosomes proving that Rve family proteins may not catalyze the excision of GIs.

Table 4 Flanking sequence comparison of genomic islands with other prediction methods

Island	Annotation bp position (Start and end of locus tag)	Islander [19]	IslandViewer [20] (By at least one method)	Mathee et al. [21]	Mavrodi et al. [22]	Silby et al. [23]
PAO1GI-1	785311–797747 (PA0715..PA0729.1)	Yes	783012–792990	RGP5 (PA0715–PA0729)	–	–
PAO1GI-2	1060411–1069382 (PA0976.1..PA0987)	No	1060513–1080012	RGP7 (PA0977–PA0987)	–	–
PA14GI-1	290225–312990 (PA14_03250..PA14_03410)	No	288290–297901	RGP2 (PA14_03163–PA14_03400)	–	–
PA14GI-2	4750058–4755861 (PA14_53570..PA14_53610)	Yes	4748023–4760960	RGP6 (PA14_53570–PA14_53670)	–	–
PA14GI-3	4860199–4895087 (PA14_54840..PA14_55090)	No	4860218–4872744 4873469–4891796	RGP5 (PA14_54850–PA14_55090)	–	–
PA14GI-4	5251438–5359447 (PA14_58910..PA14_60150)	Yes	5301173–5308773 5347270–5355678	RGP41 (PA14_58910–PA14_60140)	–	–
Pf5GI-1	2042156–2050551	Yes	No	–	2042157–2050549	–
Pf5GI-2	4338363–4395051	Yes	4373058–4394818	–	4338335–4395005	–
Pf5GI-3	5378464–5493562	Yes	5491834–5506798	–	5378468–5493586	–
Pf5GI-4	5728473–5745246	Yes	5728657–5747337	–	5728474–5745256	–
Pf0-1GI-1	3390830–3506476	Yes	3389143–3399593 3487378–3502629	–		3389342..3474979 3489458..3506828

The GI, PAPI-1 (PA14GI-4), can also be integrated into the chromosome of *P. aeruginosa* PAO1. Action sites of the integrase are PA0976.1, which is homologous to the flanking DRs of PAPI-1 [11]. Using the BLASTN algorithm, we discovered that the action sites of the integrase of PAO1GI-1, namely PA0729.1, is identical to Avin_60400 (belonging to *Azotobacter vinelandii* DJ), PLES_46152 (belonging to *P. aeruginosa* LESB58), PSPA7_4794 (belonging to *P. aeruginosa* PA7), PST_1260 (belonging to *P. stutzeri* A1501), and PA14_54840 (belonging to *P. aeruginosa* UCBPP-PA14) that all belong to the *Pseudomonadaceae*. We suggest that PAO1GI-1 (prophage Pf4) could transfer into the other 5 *Pseudomonadaceae* strains, especially *Azotobacter vinelandii* DJ. We also discovered that the genomic sequences from 87 completely sequenced strains include the DRs similar to PAO1GI-1, namely, 3′-TGGAGCGGGCGAAGGGAATCGAACCCT-5′. There- fore, these strains can be recommended as candidate receptors of PAO1GI-1 (prophage Pf4). These 86 strains belong to the *Proteobacteria*, whereas *Veillonella parvula* DSM 2008 belongs to *Firmicutes*. Therefore, PAO1GI-1 (prophage Pf4) may transfer between the *Proteobacteria* and *Firmicutes*. By BLASTN analysis, we discovered that 62 completely sequenced strains belonging to the *Gammaproteobacteria* can be recommended as candidate receiving strains of Pf5GI-3, and that 8 completely sequenced strains belonging to *Pseudomonas* can be recommended as candidate receiving strains of Pf5GI-4. However, because the DRs of Pf5GI-2 are unique, Pf5GI-2 is not likely to be transferred into other completely sequenced chromosomes.

Our study revealed that one GI closely related to Pf0-1GI-1 exists in the chromosome of *P. fluorescens* SBW25. This island is integrated into the $tRNA^{Asp}$ gene (PFLUt47), spans 140065 bp (from 2050486–2190550), has a G+C content of 56.22%, which is much lower than the 60.5% GC content of the SBW25 complete genome, and has a high genomic dissimilarity value (δ^*=48.371). An integrase (PFLU1878) as an important mobile element in the GI is proximate to the upstream flanking region of the GI, and was identified as bacteriophage P4 integrase by CDD. The GI in *P. fluorescens* SBW25 is homologous to Pf-01GI-1 about 32.4 kb. The homologous region included the chromosome partitioning protein, SOJ, the anti-restriction protein, the type-IV secretion system protein, a plasmid protein of unknown function, DNA topoisomerase III, and a number of hypothetical proteins.

4 Conclusion

The current study accurately localized 11 potentially mobile genomic islands in four strains of *Pseudomonas* by comparing results from algorithms used to identify GIs, such as, IslandViewer, Islander, and several related references. Of the GIs identified in this study, PAO1GI-1, Pf5GI-2, Pf5GI-3, and Pf5GI-4 were determined to be mobile regions. We also predicted the binding and cutting sites of the tyrosine-type integrases in flanking DRs. The binding and cutting sites of these integrases are essential for the deletion of the GI from the chromosome. We speculate that PAO1GI-1 can be transferred between *Proteobacteria* and *Firmicutes*, and Pf5GI-3 can be transferred between some strains of *Gammaproteobacteria*. Clear flanking DRs and determining the mobility of the GIs will help researchers transfer mobile GIs into appropriate candidate receptor strains of bacteria.

We thank Dr. Dieter Haas for providing the strain P. aeruginosa PAO1, and Dr. Ou Hongyu for discussion about this strategy. This work was supported by the National Natural Science Foundation of China (30821005 and 30870075), the National Basic Research Program of China (2009CB118906), and the Shanghai Leading Academic Discipline Project (B203).

1 de la Cruz F, Davies J. Horizontal gene transfer and the origin of species: Lessons from bacteria. Trends Microbiol, 2000, 8: 128–133
2 Juhas M, van der Meer J R, Gaillard M, et al. Genomic islands: Tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev, 2009, 33: 376–393
3 Dobrindt U, Hochhut B, Hentschel U, et al. Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol, 2004, 2: 414–424
4 Glöckner G, Albert-Weissenberger C, Weinmann E, et al. Identification and characterization of a new conjugation/type IVA secretion system (*trb/tra*) of *Legionella pneumophila* Corby localized on two mobile genomic islands. Int J Med Microbiol, 2008, 298: 411–428
5 Hacker J, Blum-Oehler G, Mühldorfer I, et al. Pathogenicity islands of virulent bacteria: Structure, function and impact on microbial evolution. Mol Microbiol, 1997, 23: 1089–1097
6 Lyczak J B, Cannon C L, Pier G B. Establishment of *Pseudomonas aeruginosa* infection: Lessons from a versatile opportunist. Microbes Infect, 2000, 2: 1051–1060
7 Stover C K, Pham X Q, Erwin A L, et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature, 2000, 406: 959–964
8 Lee D G, Urbach J M, Wu G, et al. Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. Genome Biol, 2006, 7: R90
9 Paulsen I T, Press C M, Ravel J, et al. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. Nat Biotechnol, 2005, 23: 873–878
10 Kim W, Levy S B. Increased fitness of *Pseudomonas fluorescens* Pf0-1 leucine auxotrophs in soil. Appl Environ Microbiol, 2008, 74: 3644–3651
11 Qiu X, Gurkar A U, Lory S. Interstrain transfer of the large pathogenicity island (PAPI-1) of *Pseudomonas aeruginosa*. Proc Natl Acad Sci USA, 2006, 103: 19830–19835
12 van Passel M W, Luyf A C, van Kampen A H, et al. δρ-Web, an online tool to assess composition similarity of individual nucleic acid sequences. Bioinformatics, 2005, 21: 3053–3055
13 Marchler-Bauer A, Anderson J B, Chitsaz F, et al. CDD: Specific functional annotation with the Conserved Domain Database. Nucleic Acids Res, 2009, 37: D205–D210
14 Webb J S, Lau M, Kjelleberg S. Bacteriophage and phenotypic variation in *Pseudomonas aeruginosa* biofilm development. J Bacteriol, 2004, 186: 8066–8073
15 Grindley N D, Whiteson K L, Rice P A. Mechanisms of site-specific recombination. Annu Rev Biochem, 2006, 75: 567–605
16 Haren L, Ton-Hoang B, Chandler M. Integrating DNA: Transposases

and retroviral integrases. Annu Rev Microbiol, 1999, 53: 245–281
17 Williams K P. Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: Sublocation preference of integrase subfamilies. Nucleic Acids Res, 2002, 30: 866–875
18 Song L, Zhang X H. Innovation for ascertaining genomic islands in PAO1 and PA14 of *Pseudomonas aeruginosa*. Chinese Sci Bull, 2009, 54: 3991–3999
19 Mantri Y, Williams K P. Islander: A database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities. Nucleic Acids Res, 2004, 32: D55–D58
20 Langille M G, Brinkman F S. IslandViewer: An integrated interface for computational identification and visualization of genomic islands. Bioinformatics, 2009, 25: 664–665
21 Mathee K, Narasimhan G, Valdes C, et al. Dynamics of *Pseudomonas aeruginosa* genome evolution. Proc Natl Acad Sci USA, 2008, 105: 3100–3105
22 Mavrodi D V, Loper J E, Paulsen I T, et al. Mobile genetic elements in the genome of the beneficial rhizobacterium *Pseudomonas fluorescens* Pf-5. BMC Microbiol, 2009, 9: 8
23 Silby M W, Cerdeño-Tárraga A M, Vernikos G S, et al. Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. Genome Biol, 2009, 10: R51