# Structure Analysis of Genomic Islands Flanked by 5' and 3' End of the tRNA$^{Gly\text{-}CCC}$ Gene in *Delftia Acidovorans* SPH-1

Lingwei Su
College of Life and Environmental Sciences
Shanghai Normal University
Shanghai, China

Qingze Zha
College of Life and Environmental Sciences
Shanghai Normal University
Shanghai, China

Xuan Peng
College of Life and Environmental Sciences
Shanghai Normal University
Shanghai, China

Lei Song*
College of Life and Environmental Sciences
Shanghai Normal University
Shanghai, China
*songlei@shnu.edu.cn

**Abstract—*Delftia acidovorans* SPH-1 is a beta-Proteobacterium that degrades a number of organic compounds. Genomic islands play the important role in horizontal gene transfer. Two prophage-related genomic islands (GIs), SPHGI$^{Gly\text{-}CCC}$-1 and SPHGI$^{Gly\text{-}CCC}$-2, were determined and are integrated into 5' and 3' end of the tRNA$^{Gly\text{-}CCC}$ gene (Daci_R0056) that was aligned with *D. acidovorans* SPH-1 chromosome. The genomic islands that have similar flanking direct repeats and homologous integrases with SPHGI$^{Gly\text{-}CCC}$-1 and SPHGI$^{Gly\text{-}CCC}$-2 verified in *Delftia* sp. Cs1-4, *Bordetella petrii* strain DSM 12804, *Pseudomonas stutzeri* DSM 4166, and several *Pseudomonas aeruginosa* strains, respectively. The cutting sites of the integrases in SPHGI$^{Gly\text{-}CCC}$-1 and SPHGI$^{Gly\text{-}CCC}$-2 that are annotated as DNA breaking-rejoining enzyme and P4 integrase are located inside DHU loop and downstream region of T Ψ C loop, respectively. SPHGI$^{Gly\text{-}CCC}$-1, SPHGI$^{Gly\text{-}CCC}$-2, and their structure similar GIs should be transfer into many beta-Proteobacterium chromosomes that have highly similar tRNA$^{Gly\text{-}CCC}$ gene with *D. acidovorans* SPH-1, and these strains that capture the GIs will obtain new character.**

***Keywords-genomic islands (GIs); Delftia acidovorans SPH-1; tRNA$^{Gly\text{-}ccc}$ gene; integrase; structure similarity***

## I. INTRODUCTION

*Delftia acidovorans* SPH-1 is a strain of the microbial consortia that contain three representative organisms, *Parvibaculum lavamentivorans*, *Comamonas testosterone*, and *D. acidovorans*, and can completely degrade commercial linear alkylbenzenesulfonate (LAS) [1]. Meanwhile, taurine moiety can be utilized and produce cholate via TauXY, Xsc, a sulfite exporter and sulfite dehydrogenase in *D. acidovorans* SPH-1 [2, 3, 4]. Genomic islands are products of horizontal gene transfer (HGT) that plays an important role for the evolution of prokaryote chromosomes. DAGI-1 and DAGI-2 were determined as the metal resistance genomic islands (GIs) in *D. acidovorans* SPH-1 through comparative genomics analysis with *Cupriavidus metallidurans* CH34, and their integration sites are not the tRNA$^{Gly}$ genes [5]. Accidentally, two integrases (Daci_4138 and Daci_4139) that exist nearby left and right flanking sequence of tRNA$^{Gly\text{-}CCC}$ gene (Daci_R0056) were found in *D. acidovorans* SPH-1, respectively. The single copy tRNA$^{Gly\text{-}CCC}$ gene sequence was aligned with *D. acidovorans* SPH-1 chromosome, homologous sequences of 5' and 3' end of the tRNA$^{Gly\text{-}CCC}$ gene were located in its upstream and downstream. Two GIs that simultaneously exist nearby 5' and 3' end of the tRNA$^{Gly\text{-}CCC}$ gene were determined in *D. acidovorans* SPH-1. Then, the integrase genes and flanking direct repeats (DRs) were aligned with all sequenced chromosomes using Tblastn and Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The structure similar GIs were verified in some *Delftia*, *Bordetella*, and *Pseudomonas* strains. The action sites of the integrases in two GIs are localized in DHU stem-loop and downstream region of TΨC loop through analysis of flanking DRs in structure similar GIs.

## II. MATERIALS AND METHODS

All chromosomes, tRNA$^{Gly}$ gene, and the integrase protein sequences were extracted from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/). The tRNA$^{Gly}$ genes (Daci_R0004, Daci_R00036, Daci_R00037, Daci_R00038, Daci_R00040, and Daci_R00056) were aligned with the *D. acidovorans* SPH-1 chromosome sequence through Blastn from NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The region of GIs will be determined if the distance between every tRNA$^{Gly}$ gene sequence and its homologous sequence is 5kb-500kb and at least one integrase exists nearby the tRNA$^{Gly}$ gene or its homologous sequence. The GIs will be accurately localized when the flanking DRs are further determined using Blastn. The integrases of the GIs was aligned with the all sequenced chromosomes through Tblastn from NCBI. The tRNA$^{Gly}$ gene that is the insert site of the GI

 

was aligned with the chromosome that contains homologous integrase of the GI. The structure similar GIs will be verified if the flanking DRs that are relative to $tRNA^{Gly}$ gene were found and the homologous integrase exists between them. The second structure of all flanking $tRNA^{Gly}$ genes were predicted by RNA structure 5.7 [6] in structure similar GIs. The phylogenetic relationship between the integrases was analyzed by MEGA 6 in structure similar GIs [7]. The flanking DRs of structure similar GIs were aligned through Clustal W2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The integrase proteins were annotated by CDD (Conserved Domain Database) (http://www.ncbi.nlm.nih.gov/cdd/). The GC% of every GI and the correlative chromosome were calculated as the characteristics marker of HGT.

## III. RESULTS AND DISCUSSION

### A. *Two genomic islands were integrated into 5' and 3' end of $tRNA^{Gly\text{-}CCC}$ gene (Daci_R0056)*

Six $tRNA^{Gly}$ genes that their anticodons are GCC, TCC and CCC were annotated in *D. acidovorans* SPH-1. Three $tRNA^{Gly}$ genes (Daci_R00036, Daci_R00037, and Daci_R00038) that form a cluster (from 2912398 to 2912751) and an adjacent $tRNA^{Gly}$ gene (Daci_R00040) (from 2914166 to 2914241) are same nucleotide sequence, and their anticodon is GCC. Only $tRNA^{Gly\text{-}CCC}$ gene (Daci_R0056) (Fig.1a) was become the integration site of GIs using this gene sequence alignment with the *D. acidovorans* SPH-1 chromosome. Interesting, two GIs, namely $SPHGI^{Gly\text{-}CCC}$-1 and $SPHGI^{Gly\text{-}CCC}$-2, were simultaneously inserted into 5' and 3' end of the $tRNA^{Gly\text{-}CCC}$ gene, and their DRs are located in symmetrical DHU loop and downstream region of TΨC loop (See TABLE 1 and Fig. 1a). Only one integrase exists in every genomic island. The integrase of $SPHGI^{Gly\text{-}CCC}$-1 is annotated as DNA breaking-rejoining enzyme, and the integrase of $SPHGI^{Gly\text{-}CCC}$-2 is annotated as P4 integrase. The GC% of $SPHGI^{Gly\text{-}CCC}$-1 and $SPHGI^{Gly\text{-}CCC}$-2 are obviously lower than of the *D. acidovorans* SPH-1 chromosome.

### B. *The structure similar GIs with $SPHGI^{Gly\text{-}CCC}$-1 and $SPHGI^{Gly\text{-}CCC}$-2*

Using protein sequence similar alignment with all sequenced chromosomes, the integrase (DelCs14_2650) is high similar (≥98%) with Daci_4138 in $SPHGI^{Gly\text{-}CCC}$-1. $Cs1\text{-}4GI^{Gly\text{-}CCC}$ that contains the DelCs14_2650 was accurately localized in *Delftia* sp. Cs1-4 through sequence alignment of DRs of $SPHGI^{Gly\text{-}CCC}$-1. Similarly, $Bp12804GI^{Gly\text{-}CCC}$, $Ps4166GI^{Gly\text{-}CCC}$, $PaNCGM2.S1GI^{Gly\text{-}CCC}$, $PaNCGM257GI^{Gly\text{-}CCC}$, $PaVRFPA04GI^{Gly\text{-}CCC}$-1 and $PaVRFPA04GI^{Gly\text{-}CCC}$-2 that own one sequence similar integrase (Daci_4139) (≥98%) and DRs with $SPHGI^{Gly\text{-}CCC}$-2 were verified in *B. petrii* strain DSM 12804, *P. stutzeri* DSM 4166 and some *P. aeruginosa* strains (TABLE 1). The integrases of $SPHGI^{Gly\text{-}CCC}$-1, $SPHGI^{Gly\text{-}CCC}$-2 and their structure similar GIs were analyzed by MEGA 6. The result showed that the integrases were divided into two types (Fig. 2). One group is annotated as DNA breaking-rejoining enzyme and belongs to two structure similar GIs associated with 5' end of $tRNA^{Gly\text{-}CCC}$ gene, another is annotated as P4 integrase and belongs to seven structure similar GIs flanked by 3' end of $tRNA^{Gly\text{-}CCC}$ gene. The cutting and binding sites of the DNA breaking-rejoining enzyme were predicted in symmetrical DHU stem-loop (5'GTTCAATGGCAGGAC3') of DRs that were aligned by ClustalW2 in $SPHGI^{Gly\text{-}CCC}$-1 and $Cs1\text{-}4GI^{Gly\text{-}CCC}$. So, the cutting site of the integrases is 5'AATGGCA3', and the binding site of the integrases is 5'GTTC3'(5'GGAC3'). The cutting site of the P4 integrase was predicted in downstream region of TΨC loop (5'GATTCCCTTCGCCCGCTCCA3') of DRs that were aligned by ClustalW2 in $SPHGI^{Gly\text{-}CCC}$-2 and its structure similar GIs. Interesting, two copies $tRNA^{Gly\text{-}CCC}$ genes were become integration site of GIs, namely $PaVRFPA04GI^{Gly\text{-}CCC}$-1 and $PaVRFPA04GI^{Gly\text{-}CCC}$-2 in *P. aeruginosa* VRFPA04. $PaVRFPA04GI^{Gly\text{-}CCC}$-1 contains $PaVRFPA04GI^{Gly\text{-}CCC}$-2, and their partial sequences are high similar (26kb). Thus, $PaVRFPA04GI^{Gly\text{-}CCC}$-1 and $PaVRFPA04GI^{Gly\text{-}CCC}$-2 that form the GIs group are named as the nest GIs. The gene contents between $Cs1\text{-}4GI^{Gly\text{-}CCC}$ and $SPHGI^{Gly\text{-}CCC}$-1 are almost completely different. Both *Delftia* sp. Cs1-4 and *D. acidovorans* SPH-1 will obtain new characteristics if transformation between them is operated each other. $SPHGI^{Gly\text{-}CCC}$-1 contains a number of hypothetical proteins and phage-related proteins, a pathogenesis-like transcriptional factor (Daci_4092) and a phage transcriptional regulator AlpA (Daci_4137) that is highly similar (≥94%) with the protein (DelCs14_2651) of $Cs1\text{-}4GI^{Gly\text{-}CCC}$. AlpA can active the P4 integrase in *Escherichia coli* [8], and it will be verify whether the AlpA protein (Daci_4137) also can regulate DNA breaking-rejoining enzyme type integrase (Daci_4138). $Cs1\text{-}4GI^{Gly\text{-}CCC}$ also contains a pathogenesis-like transcriptional factor (DelCs14_2724) that is different with Daci_4092 and is also phage-related GI.

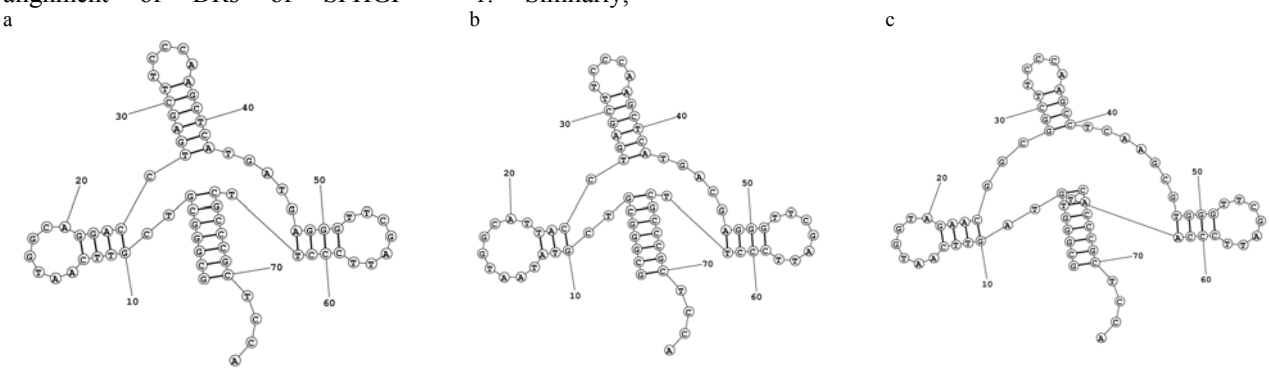


Figure 1. The second structure of $tRNA^{Gly\text{-}CCC}$ genes in *Delftia* (a), *Bordetella* (b), and *Pseudomonas* (c).

TABLE 1. The characteristics of GIs flanked by tRNA$^{Gly\text{-}CCC}$ genes in *D. acidovorans* SPH-1 and some strains

| GI | Strain | Range | Insertion site | DR(UP/DOWN) | Size (bp) | integrase | CDD | Genomic island GC% (genome GC%) |
|---|---|---|---|---|---|---|---|---|
| SPHGI$^{Gly\text{-}CCC}$-1 | *Delftia acidovorans* SPH-1 | 4481370-4531427 | Daci_R0056 4531404..4531477 | AAGCGGGCGTCGTTCAATGGCAGGAC/AAGCGGGCGTCGTTCAATGGCAGGAC | 50.058 | Daci_4138 4529912..4531123 | DNA breaking-rejoining enzyme | 62.55% (66.48%) |
| SPHGI$^{Gly\text{-}CCC}$-2 | | 4531458-4614944 | Daci_R0056 4531404..4531477 | GATTCCCTTCGCCCGCTCCA/GATTCCCTTCGCCCGCTCCA | 83.487 | Daci_4139 4531718..4533646 | P4 integrase | 60.70% (66.48%) |
| Cs1-4GI$^{Gly\text{-}CCC}$ | *Delftia* sp. Cs1-4 | 3023199-3084775 | DelCs14_R0050 complement(3023149..3023222) | GTCCTGCCATTGAACGACGCCCGCTT/GTCCTGCCATTGAACGACGCCCGCTT | 61.577 | DelCs14_2650 3023429..3024712 | DNA breaking-rejoining enzyme | 63.78% (66.7%) |
| Bp12804GI$^{Gly\text{-}CCC}$ | *Bordetella petrii* strain DSM 12804 | 4417743-4576856 | Bpettrna_44 complement(4576839..4576912) | TGGAGCGGGTGATGGGAA/TGGAGCGGGTGATGGGAA | 159.114 | Bpet4316 complement(4574670..4576598) | P4 integrase | 61.15% (65.5%) |
| Ps4166GI$^{Gly\text{-}CCC}$ | *Pseudomonas stutzeri* DSM 4166 | 1309004-1377715 | PSTAA_1291 complement(1377696..1377769) | TGGAGCGGGCGAAGGGAATC/TGGAGCGGGCGAAGGGAATC | 68.712 | PSTAA_1290 complement(1375527..1377455) | P4 integrase | 63.03% (64.0%) |
| PaNCGM2.S1GI$^{Gly\text{-}CCC}$ | *Pseudomonas aeruginosa* NCGM2.S1 | 1513064-1597104 | NCGM2_1453 complement(1597085..1597158) | TGGAGCGGGCGAAGGGAATC/TGGAGCGGGCGAAGGGAATC | 84.041 | NCGM2_1452 complement(1594915..1596843) | P4 integrase | 60.85% (66.1%) |
| PaNCGM257GI$^{Gly\text{-}CCC}$ | *Pseudomonas aeruginosa* DNA, complete genome, strain: NCGM257 | 523363-5330002 | PA257_5183 5233579..5233652 | GATTCCCTTCGCCCGCTCCA/GATTCCCTTCACCCGCTCCA | 96.370 | PA257_5184 5233894..5235822 | P4 integrase | 69.40% (65.9%) |
| PaVRFPA04GI$^{Gly\text{-}CCC}$-1 | *Pseudomonas aeruginosa* VRFPA04 | 1639727-1812416 | P797_08570 complement(1812397-1812470) | TGGAGCGGGCGAAGGGAATC/TGGAGCGGGCGAAGGGAATC | 172.690 | P797_08565 complement(1810228..1812156) | P4 integrase | 63.08% (66.5%) |
| PaVRFPA04GI$^{Gly\text{-}CCC}$-2 | | 1657845-1745231 | P797_08195 complement(1745212-1745285) | TGGAGCGGGCGAAGGGAATC/TGGAGCGGGCGAAGGGAATC | 87.387 | P797_08190 complement(1743042..1744970) | P4 integrase | 60.83% (66.5%) |

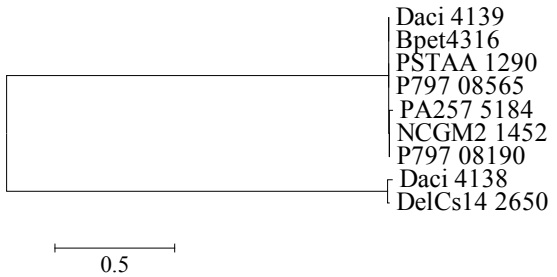


Figure 2. The genetics analysis of the integrases in structure similar GIs flanked by tRNA$^{Gly\text{-}CCC}$ genes by MEGA 6.

The gene contents of SPHGI$^{Gly\text{-}CCC}$-2 is partially similar with its structure similar GIs (about 50kb). SPHGI$^{Gly\text{-}CCC}$-2 and its structure similar GIs also contain a highly similar phage transcriptional regulator AlpA protein. This enzyme should be the teanscriptional active factor of the P4 integrase in SPHGI$^{Gly\text{-}CCC}$-2 and its structure similar GIs. SPHGI$^{Gly\text{-}CCC}$-2 is high similar with some sequence of Bp12804GI$^{Gly\text{-}CCC}$ that also contains two extra regions (about 12kb and 62kb, respectively) and of PaNCGM2.S1GI$^{Gly\text{-}CCC}$ and PaVRFPA04GI$^{Gly\text{-}CCC}$-2 that

have almost same sequence and also contain about 3kb extra region. Bp12804GI$^{Gly\text{-}CCC}$ that contains many multi-antibiotics resistance genes has been named as GI6 [9]. The two GIs, namely PAO1GI-1 and PA14GI-6, associated with tRNA$^{Gly\text{-}CCC}$ gene have been determined in *P. aeruginosa* PAO1 and UCBPP-PA14 [10]. The integrase (PA0728) of PAO1GI-1 is annotated as the HP1 integrase that is different with the integrase of GIs that were found in this research, and its flanking DRs are symmetrical TΨC stem-loop of the tRNA$^{Gly\text{-}CCC}$ gene (5'AGGGTTCGATTCCCT3'). So, the cutting site of the integrase is 5'TTCGATT3', and its binding sites are 5'AGGG3' (5'CCCT3'). PAO1GI-1 has been deleted and reintegrated into the corresponding chromosome, but PA14GI-6 can't be deleted and reintegrated [11]. The integrases of PA14GI-6 are different with the integrase of SPHGI$^{Gly}$-2 and its flanking DRs are shorter than of SPHGI$^{Gly}$-2, so it is probable that the tRNA$^{Gly\text{-}CCC}$ gene has various action sites of the integrases, not only symmetrical anticodon stem-loop, symmetrical TΨC loop, and asymmetrical 3' end of the tRNA gene[12], but also asymmetrical anticodon stem-loop, asymmetrical TΨC loop, symmetrical and asymmetrical DHU stem-loops, asymmetrical 5' end of the tRNA gene. In addition, the cutting sites of the integrases in GIs flanked by asymmetrical 3' end of the tRNA$^{Gly\text{-}CCC}$ gene maybe different because the lengths of the GIs' flanking DRs are different. SPHGI$^{Gly\text{-}CCC}$-1, Cs1-4GI$^{Gly\text{-}CCC}$, SPHGI$^{Gly\text{-}CCC}$-2, Bp12804GI$^{Gly\text{-}CCC}$, Ps4166GI$^{Gly\text{-}CCC}$, and PaNCGM2.S1GI$^{Gly\text{-}CCC}$ were determined through Islander [13], but the flanking DRs that were accurately localized in this research should be more reliable and are slightly different with the results of Islander.

### C. *Structure characteristics of tRNA$^{Gly\text{-}CCC}$ genes in GIs*

Genomic tRNA Database (http://gtrnadb.ucsc.edu/) provides the tRNA messages of 629 Bacteria. The anticodons of tRNA$^{Gly}$ genes are GCC, TCC, and CCC in Bacteria. 250 tRNA$^{Gly\text{-}CCC}$ genes, 610 tRNA$^{Gly\text{-}TCC}$ genes, and 923 tRNA$^{Gly\text{-}GCC}$ genes exist in 629 Bacteria in Genomic tRNA Database. Structure similar GIs that were integrated into tRNA$^{Gly\text{-}CCC}$ genes belong to *Delftia*, *Bordetella*, and *Pseudomonas* that have almost single-copy and intragenus same sequence of tRNA$^{Gly\text{-}CCC}$ gene. The tRNA$^{Gly\text{-}CCC}$ genes that are located in all similar structure GIs are highly conservative in DHU stem-loop, TΨC stem-loop, and anticoden stem-loop (See Fig. 1a-c). The anticoden stem-loops of tRNA$^{Gly\text{-}CCC}$ genes in *Delftia*, and *Pseudomonas* are the more conservative than in *Bordetella*, but the DHU stem-loops of tRNA$^{Gly\text{-}CCC}$ genes in *Delftia*, and *Bordetella* are the more conservative than in *Pseudomonas*. All stains belong to Betaproteobacteria in this research. Now about 3000 chromosomes are sequenced in Bacteria. So, analysis action sites of the integrases that are located in GIs flanked by tRNA$^{Gly\text{-}TCC}$ gene will be recognized though the tRNA$^{Gly\text{-}TCC}$ gene alignment with the correlative chromosome.

## IV. CONCLUSIONS

SPHGI$^{Gly\text{-}CCC}$-1, SPHGI$^{Gly\text{-}CCC}$-2, and their structure similar GIs flanked by the tRNA$^{Gly\text{-}CCC}$ gene were determined in *D. acidovorans* SPH-1, *Delftia* sp. Cs1-4, *B. petrii* strain DSM 12804, *P. stutzeri* DSM 4166, and several *P. aeruginosa* strains. The nest GIs (PaVRFPA04GI$^{Gly\text{-}CCC}$-1 and PaVRFPA04GI$^{Gly\text{-}CCC}$-2) were found in *P. aeruginosa* VRFPA04. The action site of the integrases in SPHGI$^{Gly\text{-}CCC}$-1 and its structure similar GI was predicted as the DHU stem-loop, and the action site of the integrases in SPHGI$^{Gly\text{-}CCC}$-2 and its structure similar GIs was predicted in downstream region of TΨC loop.


## ACKNOWLEDGMENTS

This work was financially supported by National Program on Key Basic Research Project (973 Program) sub-topics (2012CB721005) and Shanghai Normal University Project (No. SK201335).


## REFERENCES


[1] Schleheck D, Knepper TP, Fischer K, et al. Mineralization of individual congeners of linear alkylbenzenesulfonate by defined pairs of heterotrophic bacteria. Appl Environ Microbiol, 2004, 70(7):4053–4063.

[2] Rösch V, Denger K, Schleheck D, et al. (2008) Different bacterial strategies to degrade taurocholate. Arch Microbiol. 190(1):11–18.

[3] Denger K, Smits THM, Cook AM Genome-enabled analysis of the utilization of taurine as sole source of carbon or nitrogen by *Rhodobacter sphaeroides* 2.4.1. Microbiology, 2006, 152:3167–3174.

[4] Denger K, Weinitschke S, Smits THM, et al. Bacterial sulfite dehydrogenases in organotrophic metabolism: separation and identification in *Cupriavidus necator* H16 and in *Delftia acidovorans* SPH-1. Microbiology, 2008, 154:256–263.

[5] Van Houdt R, Monchy S, Leys N, et al. (2009) New mobile genetic elements in *Cupriavidus metallidurans* CH34, their possible roles and occurrence in other bacteria. Antonie Van Leeuwenhoek. 96(2):205-226.

[6] Mathews DH. (2014) Using the RNAstructure Software Package to Predict Conserved RNA Structures. Curr Protoc Bioinformatics. 46:12.4.1-12.4.22.

[7] Tamura K, Stecher G, Peterson D, et al. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution. 30: 2725-2729.

[8] Kirby JE, Trempy JE, Gottesman S. (1994) Excision of a P4-like cryptic prophage leads to Alp protease expression in *Escherichia coli*. J Bacteriol. 176(7):2068-2081.

[9] Gross R, Guzman CA, Sebaihia M, et al. (2008) The missing link: *Bordetella petrii* is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic Bordetellae. BMC Genomics. 9: 449.

[10] Song L, and Zhang XH. (2009) Innovation for ascertaining genomic islands in PAO1 and PA14 of *Pseudomonas aeruginnosa*. Chinese Science Bulletin. 54:3991-3999.

[11] Song L, and Zhang XH. (2011) Accurate localization and excision of genomic islands in four strains of *Pseudomonas aeruginosa* and *Pseudomonas fluorescens*. Chinese Science Bulletin. 2011, 56:987-995.

[12] Williams KP. (2002) Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. Nucleic Acids Res. 30(4):866-875.

[13] Hudson CM, Lau BY, Williams KP. (2015) Islander: a database of precisely mapped genomic islands in tRNA and tmR NA genes. Nucleic Acids Res. 43(Database issue):D48-53.