Chronology and pattern of integration of tandem genomic islands associated with the tmRNA gene in *Escherichia coli* and *Salmonella enterica*

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

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

tmRNA, a combination of a tRNA-related fragment and a small mRNA fragment, was confirmed as the integration site of genomic islands (GIs). Using sequence alignment and comparative genomics, 68 GIs associated with tmRNA genes were identified among 13 genera of Enterobacteriaceae. Among them, 53 GIs were found in *Escherichia coli* and *Salmonella enterica*. Among these 53 GIs, tandem GIs were verified in eight *S. enterica* and two *E. coli* chromosomes. The downstream regions of the tmRNA genes in most of the *E. coli* and *S. enterica* chromosomes include one GI or tandem GIs region and a remnant variable region distal to the tmRNA. The chronology of integration of tandem GIs into the genome indicated that GIs farther from the tmRNA were incorporated into the genome earlier than those nearer from the tmRNA. The integrases of the tmRNA gene-associated GIs can be further categorized into three subtypes: HP1 integrases, PhiCTX integrases, and P4 integrases, which are the most predominant. The GIs were first integrated into the chromosome by the P4 integrase, subsequently by the PhiCTX integrase, and finally by the HP1 integrase. Thus, the tmRNA gene is an important site for investigating the genetics and evolution of tandem GIs.

genomic island, tmRNA gene, multiple integration, integration chronology

Genomic islands (GIs), first found in *Escherichia coli*, are a component of the flexible gene pools in prokaryotes [1,2]. GIs have various biological functions, such as pathogenicity, xenobiotic degradation, and antibiotic resistance [3]. The distinct structural characteristics of GIs have been described [4]. For example, the tRNA (or tmRNA) gene is usually utilized as the junction between the core chromosome and the GI, and a direct repeat of the 3′-end of the gene is formed at the other junction. GIs contain at least an integrase, a transposase, or a recombinase, which cleaves the direct repeats, integrates the GI into the corresponding genome, and deletes it from the genome.

The tRNA and tmRNA genes are frequently seen at integration sites of GIs [5–7], and the bias of the tmRNA gene (*ssrA*) for the integration site is presumed to be higher than that of the tRNA genes [8]. In 2004, 12 GIs associated with the tmRNA, which are widely distributed in *Proteobacteria*, *Firmicutes*, and *Actinobacteria*, were identified in 106 sequenced genomes using the Islander database at a frequency of 11% [6]. We also determined the genomic island associated with tmRNA gene in *Pseudomonas aeruginosa* UCBPP-PA14 [9]. The tmRNA includes part of the tRNA and a certain mRNA fragment. Its major functions are: (1) release of ribosomes bound to defective mRNA [10]; (2) encoding the protease for degradation of the incomplete protein produced by defective mRNA [10]; (3) regulation of

---

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

DNA-binding protein activity, such as cI protein [11]; and (4) support of phage growth [12]. The *Vibrio* pathogenicity island (VPI-1) can be excised from *Vibrio cholerae* with P4 integrase, which cleaves 5′-GCTGGGG-3′ (the 3′-end of the tmRNA gene) [13,14]. The CP4-57 prophage can also be excised from *E. coli* with P4 integrase, which cleaves the 3′-end (5′-CCGCCAGC-3′) of the tmRNA gene [15,16].

The tandem GIs that are associated with the tRNA^{Gly} gene have been experimentally confirmed in *Bordetella petrii* [17]. Tandem GIs are excellent materials for investigating the genetics and evolution of GIs. The downstream variable regions of the tmRNA genes, identified in the completely sequenced genomes of *S. enterica*, often contain two or more integrases; thus, tandem GIs may be found in these regions. The tmRNA sequences from Enterobacteriaceae were selected and the GIs flanked by tmRNA were determined by combining sequence alignment and comparative analysis. The tandem GIs were verified in *S. enterica* and *E. coli*, and the chronology of the integration of the tandem GIs into the genome was determined by analyzing the internal integrases and the characteristics of the boundary sequences. This study presents new perspectives for investigating the genetics and evolution of GIs.

1   Materials and methods

All of the Enterobacteriaceae tmRNA sequences were selected from the tmRNA website (http://www.indiana.edu/~tmrna/). Every tmRNA sequence was aligned with Enterobacteriaceae genomes in NCBI (http://www.ncbi.nlm.nih.gov/). If direct repeat sequences related to the tmRNA were found and the distances between them were within 500 kb, the regions between direct repeats were selected as a candidate GI. Considering that the direct repeat sequences found by the sequence algorithm were at least 13 bp, many GIs may be omitted, and many intraspecific sequenced genomes are present in *E. coli* and *S. enterica* of Enterobacteriaceae. Multiple comparison analysis was performed on the tmRNA site in all the sequenced intraspecies genomes by WebACT (http://www.webact.org/WebACT/generate) [18] to identify further candidate GIs. The GC content and dinucleotide bias of the candidate GIs were computed by δρ-WEB (http://deltarho.amc.uva.nl) [19]. This confirmed that candidate GIs have abnormal GC content and dinucleotide bias. The candidate GIs that contained at least an integrase, a transposase, or a recombinase were retained. The sequences of the integrase, transposase, or recombinase closest to the boundaries of the GIs were obtained from NCBI and annotated by CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) [20]. Newly confirmed GIs homologous to LT2GI$^{tmRNA-1}$ (Enterobacteria phage Fels-2) were subjected to phylogenetic analysis using the MEGA 4.02 software via the neighbor-joining method and the MEGA 5.04 software via the maximum likelihood method, using the default parameters for genetic distance calculation [21].

2   Results and discussion

2.1   Multiple integration at the tmRNA site and the presence of tandem GIs in *E. coli* and *S. enterica*

The tmRNA is an integration hotspot for GIs in Enterobacteriaceae. By combining sequence alignment and comparative genomics, 68 GIs associated with the tmRNA gene were found in Enterobacteriaceae (Table S1). About 100 tmRNA sites are not integrated with GIs in Enterobacteriaceae. The GIs flanked by the tmRNA gene were detected in 13 genera, whereas the sequenced Enterobacteriaceae genomes in the NCBI database were distributed over 22 genera. Among the GIs, 53 GIs belong to *E. coli* and *S. enterica* chromosomes. The other 15 GIs were distributed in *Escherichia*, *Cronobacter*, *Citrobacter*, *Dickeya*, *Enterobacter*, *Klebsiella*, *Pantoea*, *Edwardsiella*, *Shigella*, *Photorhabdus*, *Yersinia*, and *Sodalis*. Nine tmRNA sites are not integrated with GIs in from among the chromosomes of 32 *E. coli* strains, and three tmRNA sites are not integrated with GIs in the chromosomes of 22 *S. enterica* strains (Tables S2 and S3).

The GIs flanked by tmRNA genes were detected in the chromosomes of 23 of 32 (71.88%) *E. coli* strains; the nine exceptions being *E. coli* B str. REL606, *E. coli* HS, *E. coli* O111:H- str. 11128, *E. coli* O103:H2 str. 12009, *E. coli* O127:H6 str. E2348/69, *E. coli* SE11, *E. coli* SMS-3-5, *E. coli* SE15, and *E. coli* 536. The GI or the variable region was not integrated into the tmRNA site in *E. coli* SE15 and *E. coli* 536. A fragment of the variable region was found at this site in the other seven completely sequenced genomes (Table S2). GIs with tmRNA were integrated into the chromosome of 19 of 22 (86.36%) *S. enterica* strains, the exceptions being *S. enterica* subsp. *enterica* serovars Typhi str. CT18, Gallinarum str. 287/91, and Enteritidis str. P125109. However, a fragment of the variable region was found in these strains (Table S3).

The tmRNA is a multiple integration hotspot for GIs in Enterobacteriaceae. Comparative genomic analysis was performed to determine the variable region at the tmRNA sites in *E. coli* and *S. enterica* (Tables S2 and S3). Based on the results, variable regions were widely found in *E. coli* and *S. enterica*, except for *E. coli* SE15 and *E. coli* 536. Seventeen variable regions had a GI region and a remnant variable region in the chromosomes of 32 *E. coli* strains (Table S2). Eighteen variable regions had a GI region and a remnant variable region in the chromosomes of 22 *S. enterica* strains (Table S3). The GI region consisted of one GI or tandem GIs (Tables 1 and S1). The GC content and dinucleotide bias of the remnant variable region indicated that this region was obtained by horizontal gene transfer (Tables S2 and S3). These remnant variable regions, located downstream of the GI region (tmRNA was considered as an up-

stream area), had been integrated into the genomes for a long time. Therefore, these remnant variable regions are probably the historical remains of GIs that had lost their basic GI structures after deletion and rearrangement of chromosomes, particularly the loss or pseudogenization of integrases and mutation or deletion of direct repeat sequences at the restriction site of the integrases. The remnant variable region is estimated to have been integrated into the genome earlier than the detected GI. Thus, these genomes had undergone two or more integrations at the tmRNA site. Some tRNA sequences were confirmed to be products of horizontal gene transfer, because the tRNA$^{\text{Met}}$ gene was found in the remnant regions of some *E. coli* strains, and two tRNA$^{\text{Arg}}$ genes and one tRNA$^{\text{Ile}}$ gene were found in 11368GI$^{\text{tmRNA}}$.

The downstream variable region of the tmRNA gene in *E. coli* S88 consisted of S88GI$^{\text{tmRNA-1}}$ and S88GI$^{\text{tmRNA-2}}$. S88GI$^{\text{tmRNA-1}}$ is highly homologous with ED1aGI$^{\text{tmRNA}}$, whereas S88GI$^{\text{tmRNA-2}}$ is highly homologous with APECO1GI$^{\text{tmRNA}}$ (Figure 1). ED1aGI$^{\text{tmRNA}}$ and APECO1GI$^{\text{tmRNA}}$ are integrated into the tmRNA gene of *E. coli* 536. Therefore, S88GI$^{\text{tmRNA-1}}$ and S88GI$^{\text{tmRNA-2}}$ were confirmed as tandem GIs. Among the 53 GIs in *E. coli* and *S. enterica*, tandem GIs were detected in *S. enterica* subsp. *enterica* serovar Agona str. SL483, *S. enterica* subsp. *enterica* serovar Newport str. SL254, *S. enterica* subsp. *enterica* serovar Typhimurium LT2, *S. enterica* subsp. *enterica* serovar Paratyphi A str. AKU_12601, *S. enterica* subsp. *enterica* serovar Paratyphi A str. ATCC 9150, *S. enterica* subsp. *enterica* serovar Typhimurium SL1344, *S. enterica* subsp. *enterica* serovar Virchow str. SL491, *S. enterica* subsp. *enterica* serovar Paratyphi B str. SPB7, *E. coli* S88, and *E. coli* O55:H7 str. CB9615 (Table 1 and Figure 2). The tandem GIs were found in eight of the 22 genomes (36.4%) in *S. enterica* (Table 1 and Figure 2), and two of the 32 genomes (6.25%) in *E. coli* (Table 1 and Figure 2). Thus, tandem GIs are more frequently seen in *S. enterica* than in *E. coli*.

2.2   Chronology of the integration of tandem GIs

The chronology of the integration of the tandem GIs indicates that the GI farthest from the integration site was the first to be integrated into the chromosome and the GI nearest to the integration site was the last. The variable region associated with the tmRNA gene in 16 *E. coli* and 11 *S. enterica* genomes consisted of one GI and one remnant variable region. The GI was integrated into the chromosome later than the remnant variable region, as indicated by the defective structure of the latter (Tables S2 and S3). A GI integrated earlier into the tmRNA gene often has defective flanking direct repeats or a pseudo integrase. Therefore, SL483GI$^{\text{tmRNA-1}}$ was integrated into the chromosome later than SL483GI$^{\text{tmRNA-2}}$, and that LT2GI$^{\text{tmRNA-1}}$ was inserted into the chromosome later than LT2GI$^{\text{tmRNA-2}}$ as revealed by the

truncated or the pseudo integrase in SL483GI$^{\text{tmRNA-2}}$ and LT2GI$^{\text{tmRNA-2}}$ (Figure 2). The variable region associated with the tmRNA site in *S. enterica* subsp. *Enterica* serovar Typhimurium SL1344 consisted of tandem GIs (SL1344GI$^{\text{tmRNA-1}}$, SL1344GI$^{\text{tmRNA-2}}$, and SL1344GI$^{\text{tmRNA-3}}$) and a fragment of the remnant variable region. Sequence alignment indicated that the remnant variable fragment (2927877–2938549) associated with the tmRNA site in *S. enterica* subsp. *Enterica* serovar Typhimurium SL1344 is highly homologous to the remnant variable fragment (2905309–2915973) associated with the tmRNA site in *S. enterica* subsp. *Enterica* serovar Typhimurium str. LT2. SL1344GI$^{\text{tmRNA-3}}$ is highly homologous to LT2GI$^{\text{tmRNA-2}}$, SL1344GI$^{\text{tmRNA-2}}$ is homologous to the GI (4456335–4466068) associated with tRNA$^{\text{Leu}}$ in *S. enterica* subsp. *Enterica* serovar Paratyphi A str. AKU_12601, and SL1344GI$^{\text{tmRNA-1}}$ is homologous to LT2GI$^{\text{tmRNA-1}}$ (Figure 3). The remnant variable fragment has a pseudo bacteriophage integrase (SL1344_2747). No flanking direct repeats were found as action sites for an integrase. The remnant variable fragment farthest from the tmRNA gene was likely the first to be integrated into the tmRNA gene of *S. enterica* subsp. *Enterica* serovar Typhimurium SL1344 by the integrase (SL1344_2747). Over time, this region has lost mobility because of the pseudogenization of integrases and the rearrangement or deletion of the flanking sequences. Soon afterwards, SL1344GI$^{\text{tmRNA-3}}$ was integrated into the tmRNA gene and lost its mobility because of the pseudogenization of integrases (SL1344_2723) and probable point mutation of the flanking sequences (5′-ACGCCAGC-3′) (Table 1, Figures 2 and 3). SL1344-GI$^{\text{tmRNA-2}}$ has a P4 integrase and complete direct repeats (5′-CCGCCAGC-3′) (Table 1, Figures 2 and 3). SL1344-GI$^{\text{tmRNA-1}}$ has an HP1 integrase, and the binding site of this integrase is 5′-AGGAATTT-3′ in the flanking direct repeats [22] (Table 1, Figures 2 and 3). P4 integrases were present in the earlier genomic islands integrated into tmRNA gene, as revealed by the predominance of P4 integrases in 68 genomic islands with tmRNA gene. This was the third time that SL1344GI$^{\text{tmRNA-2}}$ had been integrated into the tmRNA gene. It should retain its mobility because of the presence of functional P4 integrase and the complete restriction site (5′-CCGCCAGC-3′) in the flanking direct repeats. The last integration of SL1344GI$^{\text{tmRNA-1}}$ into the tmRNA gene should retain its mobility because of the presence of a functional HP1 integrase and the complete binding site (5′-AGGAATTT-3′) in the flanking direct repeats. Therefore, the GI or the remnant variable region that integrated earliest into the genome was farthest from the integration site.

2.3   Analyses of the integrases in the GIs

The integrases in the GIs related to the tmRNA gene were mainly categorized into three types: HP1 integrase, PhiCTX integrase, and P4 integrase (Table S4). Williams [23] estimated

Table 1 tmRNA gene-associated tandem genomic islands detected in *Escherichia coli* and *Salmonella enterica*[a]

Island	Strain	Coordinates	tmRNA	DR(attL/attR)	Int	Size (bp)	Integrase type annotated by CDD(PDB)	GC% of genomic island (genome GC%)	δ^* (×1000)	Genomic fragments with lower δ^*(%)
SL483GI^tmRNA-1	*Salmonella enterica* subsp. *enterica* serovar Agona str. SL483	2787019–2820998	2786701..2787063	ATGTAGGAATTTCGGACGCGGGTTCAACTCCCGC CAGCTCCACCAAA/ATGTAGGAATTTCGGACGCGG GGTTCAACTCCCGCCAGCTCCACCAAA	SeAg_B2882 2819809..2820834	33980	HP1 integrase	51.08 (52.1)	65.518	100
SL483GI^tmRNA-2	Agona str. SL483	2820982–2824289	No	CCGCCAGC/ACGCCAGC	eAg_B2883 2821164..2821745	3308	DNA breaking-rejoining enzyme	38.24 (52.1)	87.655	91.793
SL254GI^tmRNA-1	*Salmonella enterica* subsp. *enterica* serovar Newport str. SL254	2846418–2878144	2846103..2846465	AGGATGTAGGAATTTCGGACGCGGGTTCAACTCC CGCCAGCTCCACCAAAT/AGGATGTAGGAATTTCG GACGCGGGTTCAACTCCCGCCAGCTCCACCAAAT	SNSL254_A2951 2876970..2877980	31727	DNA breaking-rejoining enzyme (HP1 integrase)	50.52 (52.2)	67.188	100
SL254GI^tmRNA-2	Newport str. SL254	2878119–2888742	No	TTCAACTCCCGCCAGCT/TTCAGACTCCGCCAGCT	SNSL254_A2952 (2878326..2879519)	10624	P4 integrase	47.94 (52.2)	60.835	90.969
LT2GI^tmRNA-1	*Salmonella enterica* subsp. *enterica* serovar Typhimurium LT2	2844265–2878004	2843947..2844309	ATGTAGGAATTTCGGACGCGGGTTCAACTCCCGC CAGCTCCACCAAA/ATGTAGGAATTTCGGACGCGG GGTTCAACTCCCGCCAGCTCCACCAAA	STM2739 2876815..2877840	33740	HP1 integrase	52.49 (52.2)	67.606	100
LT2GI^tmRNA-2	Typhimurium LT2	2877988–2905308	No	CCGCCAGC/ACGCCAGC	STM2740 (2878169..2879233)	27321	P4 integrase	47.66 (52.2)	55.235	94.35
12601GI^tmRNA-1	*Salmonella enterica* subsp. *enterica* serovar Paratyphi A str. AKU_12601	2657604–2691233	2657286.. 2657648	ATGTAGGAATTTCGGACGCGGGTTCAACTCCCGC CAGCTCCACCAAATAAAACAAATGGGTTACGTGAA AACGTAGCCCCTTTTTCTT/ATGTAGGAATTTCGG ACGCGGGTTCAACTCCCGCCAGCTCCACCAAATA AAACAAGGGGTTACGTGAAAACGTAGTCCCTTTT TGTT	Sspa2423 2690004..2691029	33630	HP1 integrase	51.60 (52.2)	63.441	99.265
12601GI^tmRNA-2		2691147–2723458	No	ATGTAGGAATTTCGGACGCGGGTTCAACTCCCGC CAGCTCCACCA/ATGTAGGAATTTCGGACGCGGGGT TCAACTCCCGCCAGCTCCACCA	Sspa2447 2721297..2722310	32312	DNA breaking-rejoining enzyme (HP1 integrase)	49.93 (52.2)	64.371	98.582
9150GI^tmRNA-1	*Salmonella enterica* subsp. *enterica* serovar Paratyphi A str. ATCC 9150	2662068–2695697	2661750.. 2662112	ATGTAGGAATTTCGGACGCGGGTTCAACTCCCGC CAGCTCCACCAAATAAAACAAATGGGTTACGTGAA AACGTAGCCCCTTTTTCTT/ATGTAGGAATTTCGG ACGCGGGTTCAACTCCCGCCAGCTCCACCAAATA AAACAAGGGGTTACGTGAAAACGTAGTCCCTTTT TGTT	Spa2600 2694468..2695493	33630	HP1 integrase	51.60 (52.2)	63.426	99.265
9150GI^tmRNA-2		2695611–2727922	No	ATGTAGGAATTTCGGACGCGGGTTCAACTCCCGC CAGCTCCACCA/ATGTAGGAATTTCGGACGCGGGGT TCAACTCCCGCCAGCTCCACCA	2725761.. 2726774	32312	DNA breaking-rejoining enzyme (HP1 integrase)	49.93 (52.2)	64.509	99.291
SL1344GI^tmRNA-1	*Salmonella enterica* subsp. *enterica* serovar Typhimurium SL1344	2855454–2889891	2855136.. 2855498	ATGTAGGAATTTCGGACGCGGGTTCAACTCCCGC CAGCTCCACCAAAT/ATGTAGGAATTTCGGACGCG GGTTCAACTCCCGCCAGCTCCACCAAAT	SL1344_2709 2887497..2888522	34438	HP1 integrase	51.39 (52.2)	206.422	94.444
SL1344GI^tmRNA-2	Typhimurium SL1344	2889874–2900569	No	CCGCCAGCTCCACCAAA/CCGCCAGCGCCCACCAAA	SL1344_2712 2890073..2891269	10696	P4 integrase	47.57 (52.2)	225.454	93.305
SL1344GI^tmRNA-3		2900554–2927876	No	CCGCCAGC/ACGCCAGC	SL1344_2723 2900734..2901979	27323	P4 integrase	47.67 (52.2)	226.938	92.857

(To be continued on the next page)

(Continued)

Island	Strain	Coordinates	tmRNA	DR(*attL/attR*)	*Int*	Integrase type annotated by CDD(PDB)	Size (bp)	GC% of genomic island (genome GC%)	δ^* (×1000)	Genomic fragments with lower δ^*(%)
SL491GI^tmRNA-1	*Salmonella enterica* subsp. *enterica* serovar Virchow str. SL491	gcontig 1112707573 884 2274696–2307156	No	GCTGGCGGGGAGTTGAACCCGCGTCCGAAATTCCT ACAT/GCTGGCGGGGAGTTGAACCCGCGTCCGAAA TTCCTACAT	(2275837..2276850)	DNA breaking-rejoining enzyme (HP1 integrase)	32461	50.07 (52.2)	65.847	97.872
SL491GI^tmRNA-2		2307070–2342654	2342610-2342972	AACAAAAAGGGGGCTACGTTTTCACGTAACCCCTT GTTTTTATTTGGTGGAGCTGGCGGGAGTTGAACCC GGGTCCGAAATTCCTACAT/AACAAAAAGGGGCT ACGTTTTCACGTAACCCCTTGTTTTTATTTGGTTGGA GCTGGCGGGGAGTTGAACCCGCGTCCGAAATTCCT ACAT	(2307274..2308299)	HP1 integrase	35585	51.95 (52.2)	71.350	100
SPB7GI^tmRNA-1	*Salmonella enterica* subsp. *enterica* serovar Paratyphi B str. SPB7	2858026–2868501	2857678.. 2858039	CCGCCAGCTCCACCAAA/CCGCCAGCCCCACCAAA	SPAB_03412 (2858224..2859417)	P4 integrase	10476	48.78 (52.1)	62.457	90.713
SPB7GI^tmRNA-2		2868476– 2876283	No	TATTCAGACTCCGCCAGCCCA/TATTCAGATTCCG CCAGCTCA	SPAB_03429 (2868695..2869891)	P4 integrase	7808	42.17 (52.1)	77.723	92.444
S88GI^tmRNA-1	*Escherichia coli* S88	2799240– 2832381	2797922.. 2798284	ATGTAGGAATTTCGGACGCGCGGGTTCAACTCCCGC CAGCTCCACCAA/ATGTAGGAATTTCGGACGCGCGG GTTCAACTCCCGCCAGCTCCACCAA	ECS88_2851 2829972..2830991	DNA break-ing-rejoining en-zyme (HP1 inte-grase)	34142	50.31 (50.7)	40.315	89.116
S88GI^tmRNA-2		2832346– 2875497	No	TTCGGACGCGCGGGTTCAACTCCCGCCAGC/TTTGGA CTCGGGGTTCAACTCCCGCCAGC	ECS88_2917 (2872287..2873459)	PhiCTX integrase	43152	49.40 (50.7)	45.501	93.966
CB9615GI^tmRNA-1	*Escherichia coli* O55:H7 str. CB9615	3258465.. 3313225	3258138.. 3258500	TTTCGGACGCGCGGGTTCAACTCCCGCCAGCTCCAC CACTTT/TTTCGGACACGGGTTCAACTCCCGCCAG CTCCACCACTTT	G2583_3276 3310496..3311680	PhiCTX integrase	54761	50.98 (50.5)	57.167	98.98
CB9615GI^tmRNA-2		3313207– 3337896	No	CCGCCAGCTCCACCA/CTGCCAGCCCACCA	G2583_3278 (3313387..3314616)	P4 integrase	24690	36.00 (50.5)	121.065	100

a) The double underlined bases are the binding site of HP1 integrase. The dotted line bases denote the binding sites of PhiCTX integrase. The underlined bases denote the cleavage site of the integrase.
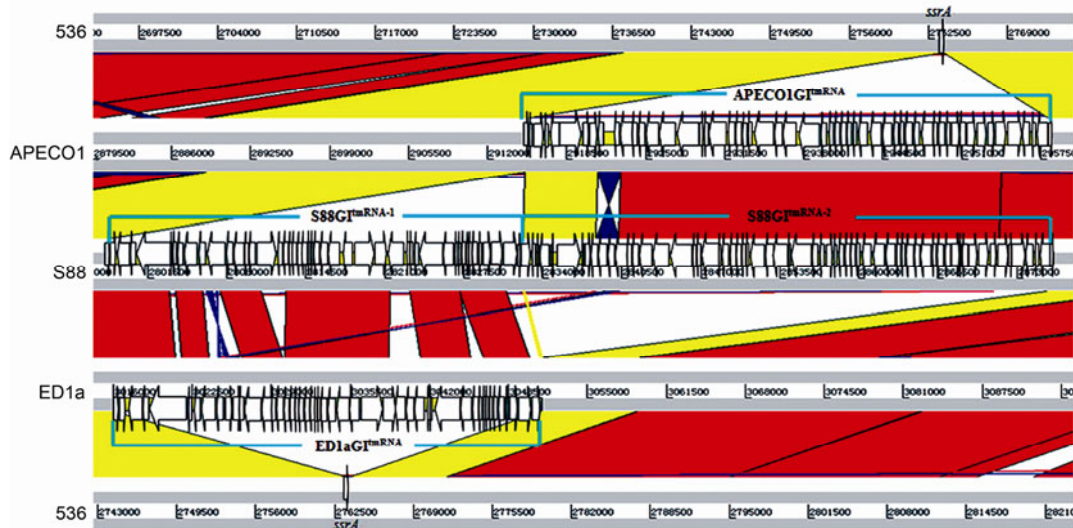
Figure 1   Identification of the tandem genomic islands flanked by tmRNA in *Escherichia coli* S88 using WebACT. The red or blue shading shows the homologous regions between the chromosomes. The yellow lines indicate the insertion sites of the genomic island. The genes in the genomic island are represented by arrows.
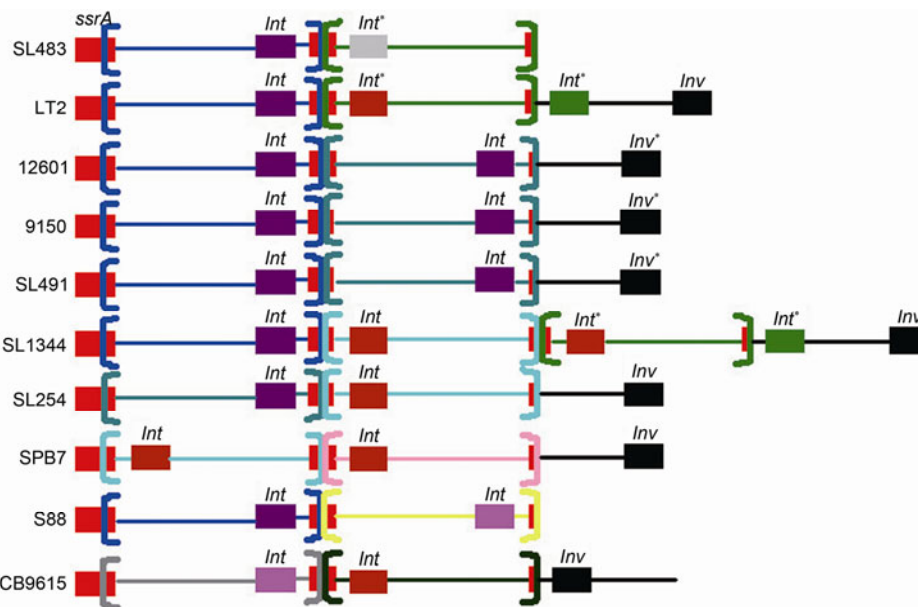

Figure 2   Schematic diagram (not to scale) of variable regions in genomes containing tandem genomic islands associated with the tmRNA gene. Symbols: red, *ssrA* or its fragment; brackets, proposed *att* sites; *Int*, integrase; dark purple represents the HP1 integrase, light purple represents the PhiCTX integrase, orange represents P4 integrase, gray represents the DNA breaking-rejoining enzyme; black, *Inv*, DNA invertase; the thin line within the brackets denotes the genomic islands region. Brackets and lines of the same color show that these regions are homologous to each other; the thin black line is used to represent the remnant variable region; asterisk, pseudogene or truncated gene.

that the sites of action of the integrases in the GIs associated with tRNA gene or tmRNA should be two symmetrical sites, namely, the anticodon and the TψC stem-loop, and the 3′ asymmetric end of tRNA. The boundary direct repeats and integrases of the 68 genomic islands were subjected to further analysis, which revealed that the integrases in 17 GIs are HP1 integrases and their binding site may be 5′-AGG-AATTT-3′ [22]. The effective site for this integrase is localized at the upstream region of the TψC ring in a tmRNA gene, because of the absence of the anticodon loop. The integrases of 10 GIs are PhiCTX integrases and their effective site is 5′-CGGGTTCAACTCCCG-3′. Among them, the binding site was CGGG(CCCG) and the cleavage site was outside the TψC ring (TTCAACT) of the tmRNA sequence. The integrases in the remaining GIs are mainly P4 integrases and their effective sites are often 5′-CCGCCAGC-3′
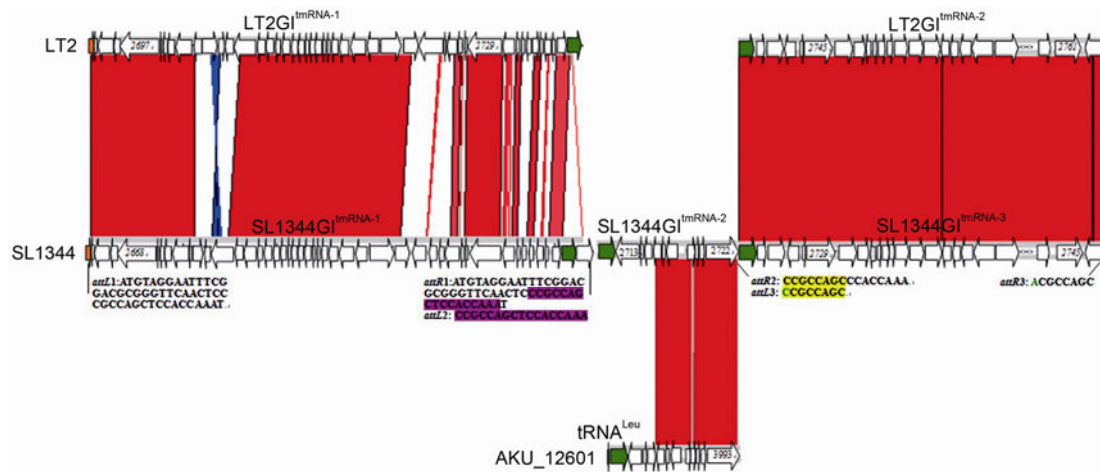
Figure 3 Verifying the chronology of integration of tandem genomic islands in *Salmonella enterica* subsp. *enterica* serovar Typhimurium SL1344. The red or blue shading shows the homologous region between the chromosomes. The arrows represent the genes in the GI. The orange arrows indicate the tmRNA gene (*ssrA*); the green arrows indicate the integrase. The bases with the purple background are those common to *attR1* and *attL2*; the bases with the yellow background are those common to *attR2* and *attL3*.
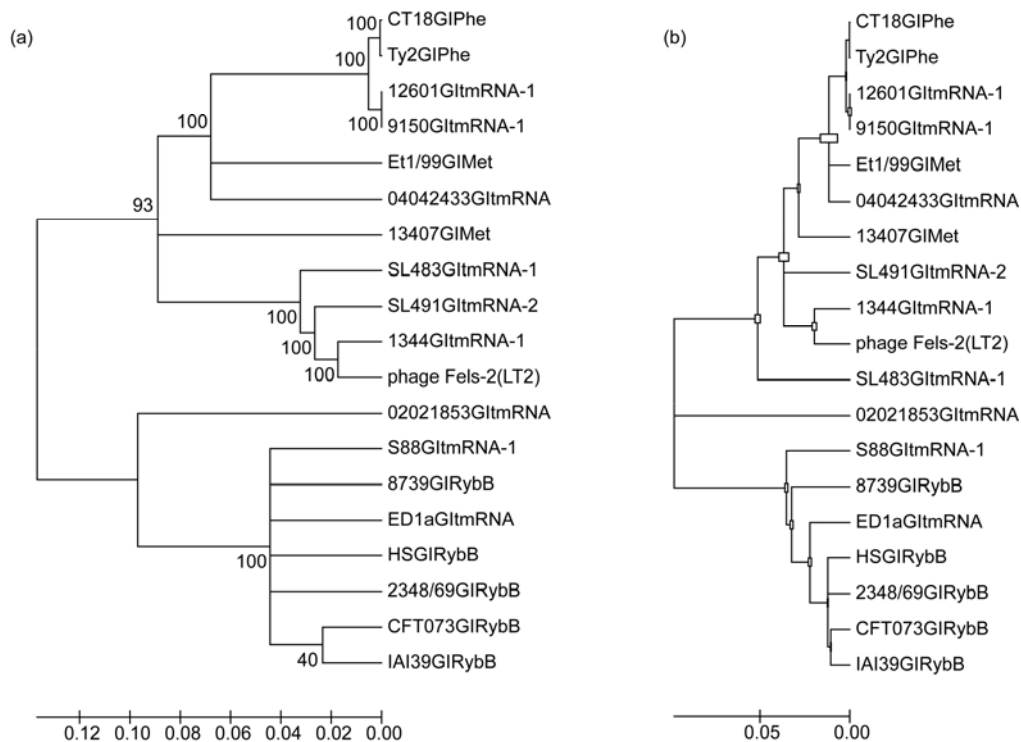

Figure 4 The panoramic phylogenetic tree constructed by MEGA4 using the neighbor-joining method and by MEGA5 using the maximum likelihood method for the genomic islands homologous to Enterobacteria phage Fels-2. The numbers indicate the bootstrap percentage values (500 replicates). The scale below represents the inferred evolutionary distance. (a) Neighbor-joining method; (b) maximum likelihood method.

in the downstream region of the TψC ring of the tmRNA sequence (Table S4). Through deletion experiments on the CP4-57 prophage, Wang et al. [16] found that the integrase of this phage is a P4 integrase and the cleavage site for this integrase is outside of 5′-CCGCCAGC-3′. Therefore, the reliability of the flanking direct repeats of GIs associated with P4 integrases can be confirmed. Figure 2 shows that

the integrases of the GI nearest to the integration site (the tmRNA gene) in the tandem GIs of *S. enterica* and *E. coli* is usually an HP1 integrase, then next nearest involve a PhiCTX integrase, and the farthest by a P4 integrase. Therefore, the sites of action of integrases in tandem GIs are initially likely to be the asymmetric 3′-end of the tRNA, second the TψC ring, and finally the anticodon loop. That is,

the effective sites of integrases in tandem GIs move from the 3′-end to the 5′-end of the tRNA or the tmRNA gene.

2.4 Analyses of Enterobacteria phage Fels-2 and its homologous GIs

LT2GI$^{tmRNA-1}$, defined as Enterobacteria phage Fels-2 [24], is homologous to certain GIs in many completely sequenced chromosomes using BLASTN alignment. These GIs have similar phage-related structural proteins. Nine GIs associated with the tmRNA gene have already been verified (Table S1), and nine remnant GIs were also detected (Table S5). Eighteen homologous GIs were distributed in *E. coli*, *S. enterica*, *Erwinia tasmaniensis*, and *Enterobacter cloacae*. The tmRNA, *RybB* gene, tRNA^{Met}, and tRNA^{Phe} were identified as the integration sites of these GIs. These 19 GIs were subjected to evolutionary molecular genetics analysis using MEGA 4.02 via the neighbor-joining method and MEGA 5.04 via the maximum likelihood method (Figure 4(a) and (b)). Both analyses showed that the GIs formed two categories. Enterobacteria phage Fels-2 is more similar to the GIs found in *S. enterica*. The boundaries of these GIs are mainly the tmRNA gene, and two of these GIs, namely, CT18GI^{Phe} and Ty2GI^{Phe}, are integrated into the *S. enterica* tRNA^{Phe} site. Another two GIs, namely, 13047GI^{Met} and Et1/99GI^{Met}, are inserted into the tRNA^{Met} site in *E. tasmaniensis* and *E. cloacae* chromosomes. The homology of the Enterobacteria phage Fels-2 is relatively far from the GIs discovered in *E. coli*, and the boundaries of these GIs are tmRNA and *RybB* genes. Further investigations are still required to explain whether convergence occurred in this genomic sequence after the GIs were integrated into the genome.

3 Conclusion

A total of 68 GIs were found integrated into the tmRNA site of 13 Enterobacteriaceae genera, mainly in *E. coli* and *S. enterica*. The presence of tandem GIs at this site was determined in *E. coli* and *S. enterica*. The GIs farthest from tmRNA may have been integrated earliest into the genome, whereas those nearer to the tmRNA may have been integrated into the genome later. Furthermore, the integrases in the tandem GIs, from the nearby sites to the distant sites, were identified as HP1, PhiCTX, and P4 integrases, respectively. Enterobacteria phage Fels-2 and its homologous GIs were distributed in 19 chromosomes in Enterobacteriaceae, mainly in *E. coli* and *S. enterica*, and mainly are integrated into tmRNA or the *RybB* gene.

1 Hacker J, Bender L, Ott M, et al. Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extraintestinal *Escherichia coli* isolates. Microb Athog, 1990, 8: 213–225

2 Hacker J, Carniel E. Ecological fitness, genomic islands and bacterial pathogenicity. EMBO Rep, 2001, 2: 376–381

3 Dobrindt U, Hochhut B, Hentschel U, et al. Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol, 2004, 2: 414–424

4 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

5 Boyd E F, Almagro-Moreno S, Parent M A. Genomic islands are dynamic, ancient integrative elements in bacterial evolution. Trends Microbiol, 2009, 17: 47–53

6 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

7 Ou H Y, He X, Harrison E M, et al. MobilomeFINDER: Web-based tools for in silico and experimental discovery of bacterial genomic islands. Nucleic Acids Res, 2007, 35: W97–W104

8 Williams K P. Traffic at the tmRNA gene. J Bacteriol, 2003, 185: 1059–1070

9 Song L, Zhang X H. Innovation for ascertaining genomic islands in PAO1 and PA14 of *Pseudomonas aeruginosa*. Chinese Sci Bull, 2009, 54: 3991–3999

10 Muto A, Sato M, Tadaki T, et al. Structure and function of 10Sa RNA: *Trans*-translation system. Biochimie, 1996, 78: 985–991

11 Retallack D M, Friedman D I. A role for a small stable RNA in modulating the activity of DNA-binding proteins. Cell, 1995, 83: 227–235

12 Withey J, Friedman D. Analysis of the role of trans-translation in the requirement of tmRNA for lambdaimmP22 growth in *Escherichia coli*. J Bacteriol, 1999, 181: 2148–2157

13 Karaolis D K, Johnson J A, Bailey C C, et al. A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. Proc Natl Acad Sci USA, 1998, 95: 3134–3139

14 Rajanna C, Wang J, Zhang D, et al. The vibrio pathogenicity island of epidemic *Vibrio cholerae* forms precise extrachromosomal circular excision products. J Bacteriol, 2003, 185: 6893–6901

15 Kirby J E, Trempy J E, Gottesman S. Excision of a P4-like cryptic prophage leads to Alp protease expression in *Escherichia coli*. J Bacteriol, 1994, 176: 2068–2081

16 Wang X, Kim Y, Wood T K. Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms. ISME J, 2009, 3: 1164–1179

17 Lechner M, Schmitt K, Bauer S, et al. Genomic island excisions in *Bordetella petrii*. BMC Microbiol, 2009, 9: 141

18 Abbott J C, Aanensen D M, Bentley S D. WebACT: An online genome comparison suite. Methods Mol Biol, 2007, 395: 57–74

19 van Passel M W, Luyf A C, van Kampen A H, et al. Deltarho-web, an online tool to assess composition similarity of individual nucleic acid sequences. Bioinformatics, 2005, 21: 3053–3055

20 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

21 Tamura K, Dudley J, Nei M, et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol, 2007, 24: 1596–1599

22 Esposito D, Thrower J S, Scocca J J. Protein and DNA requirements of the bacteriophage HP1 recombination system: A model for intasome formation. Nucleic Acids Res, 2001, 29: 3955–3964

23 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

24 Pelludat C, Mirold S, Hardt W D. The SopEPhi phage integrates into the *ssrA* gene of *Salmonella enterica* serovar Typhimurium A36 and is closely related to the Fels-2 prophage. J Bacteriol, 2003, 185: 5182–5191