A Comprehensive tRNA Genomic Survey Unravels the Evolutionary History of tRNA Arrays in Prokaryotes

Tam T.T. Tran, Hassiba Belahbib‡, Violaine Bonnefoy, and Emmanuel Talla*

Aix Marseille Université, CNRS, IGS, UMR 7256, IMM, France

‡Present address: Information Génomique et Structurale, UMR 7256, CNRS, Institut de Microbiologie de la Méditerranée, Aix Marseille Université, Marseille, France

*Corresponding author: E-mail: talla@imm.cnrs.fr.

Accepted: December 13, 2015

Abstract

Considering the importance of tRNAs in the translation machinery, scant attention has been paid to tRNA array units defined as genomic regions containing at least 20 tRNA genes with a minimal tRNA gene density of two tRNA genes per kilobase. Our analysis of *Acidithiobacillus ferrivorans* CF27 and *Acidithiobacillus ferrooxidans* ATCC 23270^T genomes showed that both display a tRNA array unit with syntenic conservation which mainly contributed to the tRNA gene redundancy in these two organisms. Our investigations into the occurrence and distribution of tRNA array units revealed that 1) this tRNA organization is limited to few phyla and mainly found in Gram-positive bacteria; and 2) the presence of tRNA arrays favors the redundancy of tRNA genes, in particular those encoding the core tRNA isoacceptors. Finally, comparative array organization revealed that tRNA arrays were acquired through horizontal gene transfer (from Firmicutes or unknown donor), before being subjected to tRNA rearrangements, deletions, and duplications. In Bacilli, the most parsimonious evolutionary history involved two common ancestors and the acquisition of their arrays arose late in evolution, in the genera branches. Functional roles of the array units in organism lifestyle, selective genetic advantage and translation efficiency, as well as the evolutionary advantages of organisms harboring them were proposed. Our study offers new insight into the structural organization and evolution of tRNA arrays in prokaryotic organisms.

Key words: tRNA array, core tRNA, tRNA gene redundancy, *Acidithiobacillus*.

Introduction

Transfer RNAs (tRNAs) play an essential role in the protein synthesis pathway, linking the genetic code with the corresponding amino acids (Widmann et al. 2010). Beyond their canonical role during protein biosynthesis, tRNAs are also implicated in other cellular functions such as the regulation of gene expression and cell death, cell wall formation, protein labelling for degradation, aminoacylation of cell membrane phospholipids, and antibiotic biosynthesis (for review, see Raina and Ibba [2014]).

Typically, tRNA is a short cloverleaf noncoding RNA of approximately 70–100 bases in length which folds into an L-shaped tertiary structure. tRNAs are highly conserved among Archaea, Bacteria, and Eukarya (Marck and Grosjean 2002; Sprinzl and Vassilenko 2005; Fujishima and Kanai 2014) suggesting that tRNA molecules had already evolved before the emergence of the last universal common ancestor. The tRNA multigene family usually comprises 20 amino acid accepting groups (also called amino acid isotypes), each containing one or more tRNA members that recognize different codons commonly referred as tRNA isoacceptors (Sprinzl and Vassilenko 2005; Goodenbour and Pan 2006). To function as substrate in protein biosynthesis, each tRNA molecule attaches to a single amino acid upon selectively base-pairing its three-base anticodon to a complementary three-base codon sequence of mRNA.

tRNA genes can be computationally predicted using several software (e.g., tRNAscan-SE, ARAGON), enabling the full identification of tRNAs (with their anticodon loop) within any genome (Lowe and Eddy 1997; Laslett and Canback 2004; Chan et al. 2011). tRNA databases have been constructed from draft and complete genomes as well as metagenomic sequences (Chan and Lowe 2009; Lee et al. 2009; Abe et al. 2011, 2014). As gene duplications and gene deletions continually occur (Bustamante et al. 2012) in a tRNA gene family, the number of tRNA genes is highly variable among different organisms (Marck and Grosjean 2002). However, ancestral set of tRNA genes within *Escherichia coli*

and related species or within all prokaryotic organisms have been proposed (Rocha 2004; Withers et al. 2006; Novoa et al. 2012). Although tRNA genes are present in multiple copies within the genome of most organisms, the number of gene copies for each tRNA isoacceptor (tRNAs with the same anticodon) varies widely from species to species (Marck and Grosjean 2002; Satapathy et al. 2010).

The tRNA genes are interspersed throughout genomic DNA, and contribute to its architecture and evolution. Frequently, they are positioned close to one another (or separated by few other genetic objects) on the same strand of the chromosome, constituting tRNA arrays. tRNA arrays have been reported in many organisms from eukaryotes to prokaryotes. In eukaryotic cells, tRNA arrays were mainly described for mitochondrial genomes (Jung et al. 2010; Friedrich et al. 2012; Smith 2015). Within the *Entamoeba* genus, a genome survey on tRNA array units revealed the same array unit organization within the five species, including the presence of short tandem repeats between adjacent tRNA genes (Tawari et al. 2008). In prokaryotes, arrays are generally transcribed as polycistronic transcripts from a common promoter, as shown in several prokaryotic strains including *Anabaena* sp. PCC7120, *Bacillus subtilis*, and *Staphylococcus aureus* (Green and Vold 1983, 1992, 1993; Puerto-Galan and Vioque 2012). In *B. subtilis*, tRNA genes are clustered in their genome, which are located nearby rRNA operons, suggesting that rRNA and tRNA genes represent a single transcriptional unit (Vold 1985). Similarly, array units of approximately 26 tRNA genes and pseudogenes were found within several cyanobacterial organisms including *Anabaena* sp. PCC7120, *Nostoc punctiforme* PCC73102, *Acaryochloris marina* MBIC11017, and *Oscillatoria* sp. PCC6506. These four arrays are clearly related and have a common origin, with the same order of the tRNA genes (Puerto-Galan and Vioque 2012). A tRNA array unit located in an integrative and conjugative element (ICE) has been reported in *Acidithiobacillus ferrooxidans* ATCC 23270^T. This array might mainly contribute to the translation of ICE encoded genes by dint of an enrichment of particular codons (Levican et al. 2009; Bustamante et al. 2012). Other functions of tRNA array units have been demonstrated such as to reduce the number of tRNA transcriptional units, and to coordinate the amount of tRNAs with translation rates (Rudner et al. 1993). For example, only eight transcriptional units over 62 known tRNA genes are seen in *B. subtilis* (Green and Vold 1992) whereas 79 tRNA genes in *E. coli* spread over 41 transcriptional units (Komine et al. 1990). This operonic organization of tRNA arrays was shown to be a significant factor that influences tRNA expression at high growth rates in *E. coli* (Ardell and Kirsebom 2005).

Despite the importance of tRNAs in protein translation, no large-scale comparative analysis has been performed on tRNA array units in prokaryotic organisms and little is known about their evolution. In order to understand how these structures evolved, investigations on the genomic structures of such tRNA array units as well as their emergence and evolutionary history are required. In this study, we first analyzed tRNA array units within *A. ferrooxidans* ATCC 23270^T and *Acidithiobacillus ferrivorans* CF27 genomes showing that both arrays displayed a tRNA gene synteny and mainly contributed to the tRNA gene redundancy observed in both organisms. Then, we undertook a large-scale genome survey of the tRNA array units within prokaryotic organisms, focusing on their occurrence, their distribution, and their genomic features. Comparative analysis allowed us to provide comprehensive insight into the origin and evolutionary history of tRNA array units in prokaryotes.

Materials and Methods

Genome Data Sets

The complete genomes of 2,765 prokaryotic (2,600 bacterial and 165 archaeal) organisms available in December 2013 were downloaded from National Center for Biotechnology Information (NCBI) ftp site (ftp:/ftp.ncbi.nih.gov/genomes/Bacteria/) and constituted the primary data source. Closely related organisms were not excluded as evolutionary differences (presence/absence of the tRNA gene array, loss/gain of a tRNA gene within an array) in species belonging to the same genus could be observed. The *A. ferrivorans* CF27 genome (accession number CCCS020000000) was retrieved from the MicroScope website (https://www.genoscope.cns.fr/agc/microscope/) (Vallenet et al. 2013) under the *AcidithioScope* project.

tRNA Gene Prediction and Identification of tRNA Array Units

The search for tRNA genes was carried out using the tRNAscan-SE 1.3.1 program (Lowe and Eddy 1997) with appropriate parameters for Bacteria and Archaea. Given that our analysis was mainly based on average tRNA abundances, any minor misannotations in tRNA genes using this prediction program were not statistically significant and thus should not have affected the final results of this work. It is noteworthy that selenocysteine tRNAs (TCA), suppressor tRNAs (CTA, TTA), and undetermined isoacceptor tRNAs represented less than 1% of the total tRNAs. In order to resolve some ambiguities, the Aragon tool (Laslett and Canback 2004) was also used to identify tRNA genes in the detailed analysis of the *Acidithiobacillus* genomes. Two tRNA genes were considered to be isoacceptors if they presented the same anticodon. The identified tRNA genes were categorized into 61 isoacceptor families and 20 amino acid (aa) isotypes. The tRNA isoacceptor was designated by its anticodon. In-house Perl scripts were then used to locate tRNA array units within the overall prokaryotic genomes. For that purpose, we first identified regions of 5, 10, 15, 20, 25 and 30 kb along the genome (incremental path of 1 kb) that contain at least 5, 10, 15, 20, 25, or 30 tRNA

genes, named array seeds (supplementary table S1, Supplementary Material online). Based on 1) a previous study (Bermudez-Santana et al. 2010) in eukaryotes in which two adjacent tRNA genes were considered as "clustered" if located less than 1 kb apart (in other words, a tRNA gene density higher than two tRNA genes per kb), and 2) that array seeds of 5 kb-size are all included within array seeds of 10 kb-size (see gray cells in supplementary table S1, Supplementary Material online), the 10 kb-length was chosen as the size for the tRNA array seeds with a minimal tRNA gene density set to 2, in order to focus on array units with high number of tRNAs. Consequently, tRNA array seeds were thus defined as a genomic region of 10 kb containing at least 20 tRNA genes. Second, the array seeds were extended in both directions with additional neighboring tRNA gene regions (while keeping their tRNA gene density higher than 2 tRNA genes per kb), yielding tRNA array units with new genomic boundaries (with tRNA gene limits). Therefore, smaller regions of the genome with high density of tRNA genes as well as larger regions with a high number of tRNA gene copies were accounted for, allowing the identification of all tRNA array units in prokaryotes. For simplicity, both tRNA array units and regions of an array are depicted by their encoded tRNAs using the standard single letter amino acid abbreviation and are enclosed in square brackets.

Structural Organization of tRNA Array Units

The organization of tRNA between tRNA array units was compared as follows: 1) Each tRNA array unit was first converted into an amino acid (aa) sequence, in which each tRNA gene was turned into the corresponding aa isotype; 2) then, incremental patterns (incremental path of 1) of 20 amino acids (corresponding to the tRNA genes) were extracted from each aa sequence and an overall amino acid pattern database associated with tRNA array units was constituted; and 3) the *fuzzpro* program (http://emboss.sourceforge.net) was used for pattern matching analysis between each incremental pattern and the pattern database, allowing 0–19 maximal pattern mismatches. These represented all putative local alignments of 20 adjacent tRNA genes between tRNA array units. From each couple of local alignment matches, similarity between two incremental patterns was settled as the ratio of the number of amino acid matches over the total number of amino acids (here 20). Finally, the array structure similarity between two tRNA array units was defined as the best similarity observed between all of their incremental pattern match combinations. From the array structure similarities, a clustered heatmap graphical representation was performed using the CIMminer tool (http://discover.nci.nih.gov/cimminer/) with default parameters. Syntenic gene maps of tRNA array neighboring regions were obtained from the MicroScope website.

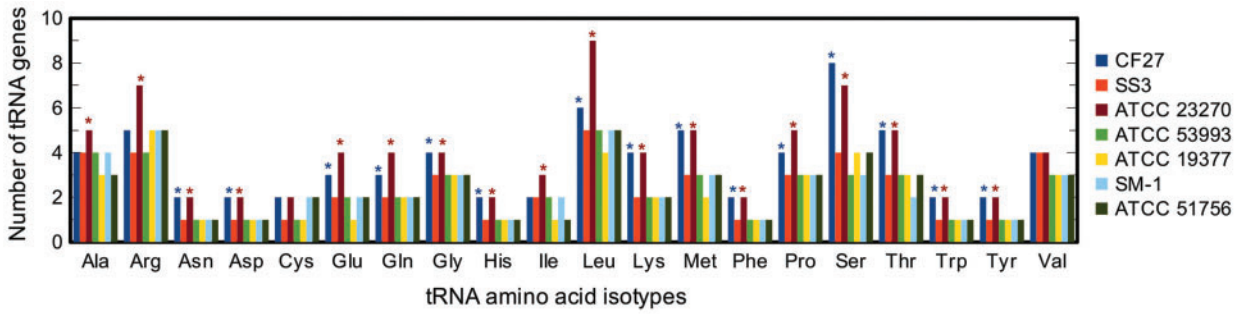
Results

Atypical tRNA Array Units Were Found in Two Gram-Negative Bacterial Species

Acidithiobacillus organisms are Gram-negative, acidophilic sulfur-oxidizing bacteria that catalyze the oxidative dissolution of sulfide minerals and play major roles in biomining and acid mine drainage. Compared with other *Acidithiobacillus* organisms, *A. ferrivorans* CF27 were recently found to harbor a high number of tRNA genes (71 tRNA genes) (Talla et al. 2014), as previously reported in *A. ferrooxidans* ATCC 23270^T (80 tRNA genes) (Levican et al. 2009). This observation could reflect the physiological importance of protein translation in these proteobacterial organisms. Therefore, the structural organization of the tRNA genes in *Acidithiobacillus* species was analyzed. An increased number of tRNA genes for each tRNA amino acid (tRNA-AA) isotype (fig. 1*A*) except for tRNA-Cys and tRNA-Val genes were found in at least one of the two strains. The additional (or extra) tRNA-AA isotypes (e.g., up to four in *A. ferrooxidans* ATCC 23270^T and *A. ferrivorans* CF27 for tRNA-Leu and tRNA-Ser, respectively) explain the higher number of tRNA genes in these two organisms.

Analysis of the corresponding tRNA isoacceptors (fig. 1*B*) showed that the extra tRNA-AA genes in CF27 or ATCC 23270^T were mainly related to specific tRNA isoacceptors. Indeed, two copies of tRNA-Arg^{ACG} and tRNA-Asn^{GUU} genes were found present in *A. ferrivorans* CF27 and *A. ferrooxidans* ATCC 23270^T against only one copy in other *Acidithiobacillus* strains. Notably, the genes of all extra tRNAs present in both *A. ferrivorans* CF27 and *A. ferrooxidans* ATCC 23270^T were found clustered in a specific region (tRNA array units) in each strain (fig. 1*C*). The tRNA array unit of *A. ferrivorans* CF27 encompassed a 6-kb region with 26 tRNA genes (36.6% of the total), whereas that in *A. ferrooxidans* ATCC 23270^T comprised 34 tRNA genes (42.5% of the total) within a 9-kb region. These arrays also harbored gene-encoding proteins of unknown function. The presence of tRNA array units in both species could be advantageous in terms of reducing the number of tRNA transcriptional units, thereby allowing coordinated transcription and processing of the precursors. The absence of such arrays in other *Acidithiobacillus* strains, combined with the location of the ATCC 23270^T tRNA array within an ICE (of about 300 kb) (Levican et al. 2009; Bustamante et al. 2012), suggests that such tRNA array units were independently acquired through horizontal gene transfer (HGT). Moreover, tRNA gene synteny between the two array units (fig. 1*C*) clearly suggests that they have a common ancestor (likely similar to the CF27 array unit) followed by gene rearrangements/insertions/deletions during evolution. Indeed, we found tRNA gene insertions/deletions in *A. ferrooxidans* ATCC 23270^T/*A. ferrivorans* CF27 (e.g., tRNA-Ser^{ACU}, tRNA-Pro^{CGG}, and tRNA-Leu^{CAA}), as well as tRNA gene translocation for both tRNA-codons tRNA-Gly^{UCC} and tRNA-Ala^{UGC} genes between the two arrays. It is noteworthy

A



B

tRNA isoacceptor types		CF27	SS3	ATCC 23270	ATCC 53993	ATCC 19377	SM-1	ATCC 51756
Ala	CGC	1	1	1	1	1	1	1
	UGC	2	2	**3**	2	1	2	1
	GGC	1	1	1	1	1	1	1
	AGC	0	0	0	0	0	0	0
Arg	CCU	1	1	1	1	2	1	1
	UCU	1	1	**2**	1	1	1	1
	CCG	1	1	1	1	1	2	2
	UCG	0	0	**1**	0	0	0	0
	GCG	0	0	0	0	0	0	0
	ACG	**2**	1	**2**	1	1	1	1
Asn	GUU	**2**	1	**2**	1	1	1	1
	AUU	0	0	0	0	0	0	0
Asp	GUC	**2**	1	**2**	1	1	1	1
	AUC	0	0	0	0	0	0	0
Cys	GCA	2	1	2	1	1	2	2
	ACA	0	0	0	0	0	0	0
Glu	CUC	**2**	1	**2**	1	0	1	1
	UUC	1	1	**2**	1	1	1	1
Gln	CUG	**2**	1	**2**	1	1	1	1
	UUG	1	1	**2**	1	1	1	1
Gly	CCC	1	1	1	1	1	1	1
	UCC	**2**	1	**2**	1	1	1	1
	GCC	1	1	1	1	1	1	1
	ACC	0	0	0	0	0	0	0
His	GUG	**2**	1	**2**	1	1	1	1
	AUG	0	0	0	0	0	0	0
Ile	UAU	0	0	0	0	0	0	0
	GAU	2	2	**3**	2	1	2	1
	AAU	0	0	0	0	0	0	0
Leu	CAG	**2**	1	**2**	1	1	1	1
	UAG	1	1	**2**	1	1	1	1
	GAG	1	1	1	1	1	1	1
	AAG	0	0	0	1	0	0	0
	CAA	1	1	**2**	1	0	1	1
	UAA	1	1	**2**	1	1	1	1
Lys	CUU	1	1	**2**	1	1	1	1
	UUU	**3**	1	**2**	1	1	1	1
Met	CAU	**5**	3	**5**	3	2	3	3
Phe	GAA	**2**	1	**2**	1	1	1	1
	AAA	0	0	0	0	0	0	0
Pro	CGG	1	1	**2**	1	1	1	1
	UGG	**2**	1	**2**	1	1	1	1
	GGG	1	1	1	1	1	1	1
	AGG	0	0	0	0	0	0	0
Ser	GCU	**2**	1	**2**	1	1	1	1
	ACU	**1**	0	0	0	0	0	0
	CGA	**2**	1	**2**	1	1	0	1
	UGA	1	0	1	0	1	1	0
	GGA	2	1	2	1	1	1	2
	AGA	0	0	0	0	0	0	0
Thr	CGU	**2**	1	**2**	1	1	0	1
	UGU	**2**	1	**2**	1	1	1	1
	GGU	1	1	1	1	1	1	1
	AGU	0	0	0	0	0	0	0
Trp	CCA	**2**	1	**2**	1	1	1	1
Tyr	GUA	**2**	1	**2**	1	1	1	1
	AUA	0	0	0	0	0	0	0
Val	CAC	1	2	1	1	1	1	1
	UAC	**2**	1	**2**	1	1	1	1
	GAC	1	1	1	1	1	1	1
	AAC	0	0	0	0	0	0	0
Total		71	47	80	46	43	47	47

C

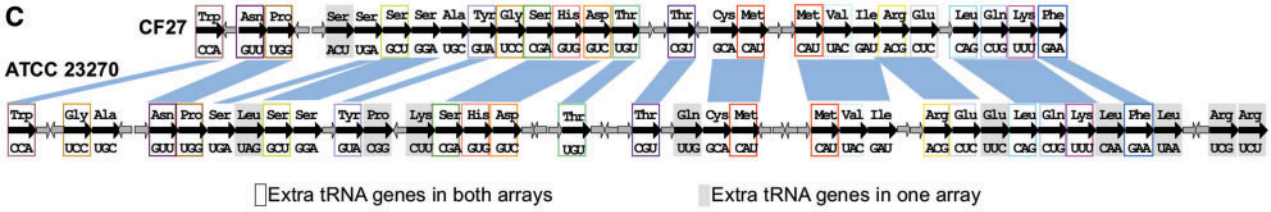


FIG. 1.—Occurrence of the tRNA amino acid isotypes and tRNA isoacceptors in *Acidithiobacillus* genomes. In (*A*), significant differences between *A. ferrivorans* CF27 or *A. ferrooxidans* ATCC 23270^T and other *Acidithiobacillus* species are shown by an asterisk. In (*B*), colored backgrounds highlight tRNA isoacceptor genes with a high number of copies in both CF27 and ATCC 23270, compared with other organisms. For CF27 or ATCC 23270, the number of tRNA gene copies is shown in bold if higher than others. (*C*) Gene organization of CF27 and ATCC 23270 tRNA array units. The tRNA genes and protein-encoding genes are represented in black and gray arrows, respectively. Extra tRNA isoacceptors in both CF27 and ATCC 23270 (compared with other *Acidithiobacillus*) are shown in colored boxes. The same color codes are used in (*B*) and (*C*) for extra tRNA genes in both arrays. Extra tRNA genes present in only one array are shaded with gray background. Syntenic regions between CF27 and ATCC 23270 arrays are illustrated by blue bands. The strain names are: CF27, *A. ferrivorans* CF27; SS3, *A. ferrivorans* SS3; ATCC 23270, *A. ferrooxidans* ATCC 23270^T; ATCC 53993, *A. ferrooxidans* 53993; ATCC 19377, *A. thiooxidans* ATCC 19377^T; SM-1, *A. caldus* SM-1; and ATCC 51756, *A. caldus* ATCC 51756^T.

that neighboring regions of the two array units showed no gene synteny between the two species. As such tRNA array units have only been described in a few organisms, these findings in the two acidithiobacilli species then prompted investigations into the organization and evolution of tRNA array units in all prokaryotic organisms.

Occurrence and Distribution of tRNA Array Units

An in silico analysis of tRNA genes within complete bacterial and archaeal genomes (see Materials and Methods; supplementary table S2, Supplementary Material online) was performed. In this way, 33 (e.g., tRNA-Ala^{GGC} and tRNA-Val^{GAC}) and 27 (e.g., tRNA-Arg^{ACG} and tRNA-Leu^{CAA}) core tRNA isoacceptors were defined in Archaea and Bacteria, respectively (supplementary information, Supplementary Material online). Twenty-five tRNA genes (e.g., tRNA-Met^{CAU} and tRNA-Ser^{GGA}) were found present in both archaeal and bacterial core sets and can therefore be considered as the minimal set of tRNA genes (present in at least 90% of the organisms) found in the ancestor of Archaea and Bacteria. The number of tRNA genes and the tRNA gene diversity among this large number of organisms were strongly correlated to the rRNA and GC content, respectively, as previously reported (Kanaya et al. 1999; Lee et al. 2009; Satapathy et al. 2010). This result clearly supports the notion of tRNA and rRNA gene numbers concomitantly increasing to favor efficient protein synthesis. A detailed analysis of these data is available in the supplementary information, Supplementary Material online.

A large-scale in silico identification and analysis of tRNA array units (supplementary table S3, Supplementary Material online) on all complete genomes of prokaryotic organisms was then performed to investigate the propensity for formation of tRNA array units, as well as how these arose in prokaryotic groups during evolution. A tRNA array unit was defined as a genomic region containing at least 20 tRNA genes with a minimal tRNA gene density of two tRNA genes per kilobase. This strategy led to the identification of 383 tRNA gene arrays in 349 organisms (13% of all organisms) that were mainly located within the chromosome (98.4% of array units). The remaining tRNA array units (six in total) were located in plasmids within six bacterial strains: In *Arthrobacter chlorophenolicus* A6, *Nostoc* PCC7120, *Nostoc* PCC7524, and *Methylibium petroleiphilum* PM1, they were located in plasmids but not in chromosome; in *Paenibacillus polymyxa* SC2 and *Bacillus thuringiensis* Bt407, the tRNA arrays were located in both plasmids and chromosomes. Four tRNA arrays that had previously been found in cyanobacterial strains (Puerto-Galan and Vioque 2012), as well as the two array units in *A. ferrivorans* CF27 and *A. ferrooxidans* ATCC 23270^{T} (see above) were identified, thus validating the tRNA array detection procedure. Almost all of the tRNA arrays were present in Bacteria, but in a limited number of phyla including Firmicutes, Cyanobacteria, Actinobacteria and in the γ-proteobacteria class (fig. 2 and supplementary table S4, Supplementary Material online). About 322 Firmicutes organisms (53.7% of all species in this phylum) harbored nearly 93% of all tRNA arrays, most of which were in Bacilli (64.4% of all Bacilli organisms) and Clostridia (26.8% of all Clostridia organisms) taxonomic classes, with, respectively, 311 and 45 tRNA arrays. In these bacteria, the presence of tRNA array units therefore represents one of the major distinguishing features of their tRNA genomic organization. On the contrary, the tRNA arrays were less represented in Cyanobacteria (13.3%), γ-Proteobacteria (2%), and Actinobacteria (1%), suggesting that they may have been acquired through HGT (from Firmicutes or from an unknown donor) during evolution. Among the γ-Proteobacteria shown to have a high number of tRNA genes, only few were found to harbor tRNA arrays (fig. 2 and supplementary table S4, Supplementary Material online), therefore arguing against this feature being linked to a high number of tRNA genes

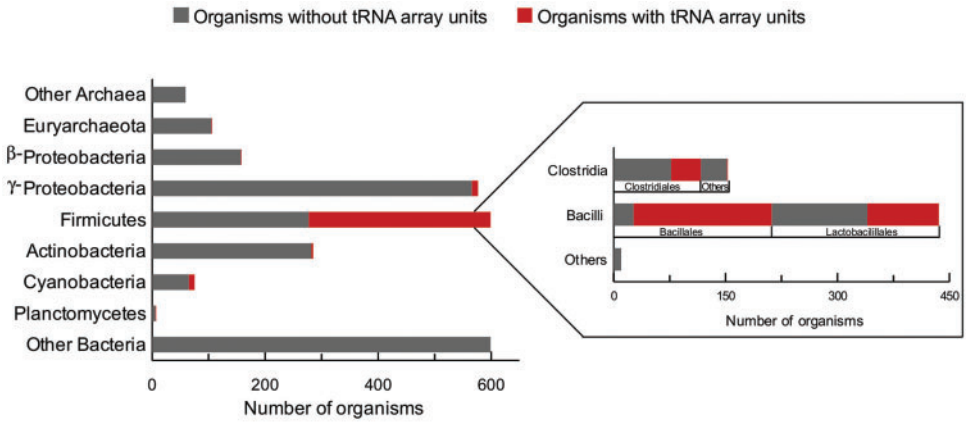


Fig. 2.—Distribution of organisms with (red bar) or without (gray bar) tRNA array units across prokaryotic phyla and specifically in Firmicutes.

within the organism (see supplementary information, Supplementary Material online). Surprisingly, only one archaeal organism (i.e., the rumen methanogen *Methanobrevibacter ruminantium*) was found to possess a tRNA array unit of 21 tRNA genes for a total of 60 tRNA genes. Concomitantly, the higher number of tRNA genes in this organism, compared with other *Methanobacteriales* species (~36-48 tRNA genes), indicated that extra-tRNA genes were probably acquired through HGT after the divergence of the *Methanobacteriales* order, and may be linked to the specific lifestyle of this organism. Unfortunately, no distinctive physiological features were found between members of the *Methanobrevibacter* genus that might have explained a lifestyle adaption for *M. ruminantium*. Among the organisms with tRNA arrays, 316 (90% of the total) were found to have only one tRNA array in their genome, 32 to possess two and one to harbor three tRNA arrays (*B. thuringiensis* Bt407) within their genome. Detailed analysis of genomic features associated with tRNA array regions is available in the supplementary information, Supplementary Material online.

The Presence of tRNA Array Units Favors the Redundancy of Core tRNA Isoacceptors

Next, the distribution of tRNA genes within prokaryotes with or without tRNA array units according to tRNA amino acid isotype and corresponding tRNA isoacceptor was explored (supplementary fig. S1, Supplementary Material online). All of the 20 isotypes were found located within tRNA arrays with the most represented being tRNA-Met (11.3%), tRNA-Leu (9.7%) and tRNA-Gly (9.8%) genes, and the rarest being tRNA-Cys, tRNA-Trp and tRNA-Tyr, each representing just 2% (supplementary fig. S1*A*, Supplementary Material online). It is noteworthy that selenocysteine tRNA (TCA) and possible suppressor tRNAs (CTA, TTA) were absent from the tRNA array units. For each tRNA isoacceptor, two tRNA gene indexes were defined as the total number of tRNA genes over the total number of prokaryotes either with or without tRNA array units. By comparing these two indexes, we assessed the propensity of an isoacceptor to be located in organisms with or without arrays. As shown in supplementary fig. S1*B*, Supplementary Material online, 29 tRNA isoacceptor genes (e.g., tRNA-Glu^{UUC} and tRNA-Val^{UAC}) had higher indexes for organisms with arrays suggesting a favored location of these tRNA genes within array units. Moreover, no differences were observed between organisms with and those without array units with regards to the codon usage (data not shown). The frequency of each tRNA isoacceptor in the array units or nonarrayed regions within array organisms highlighted (supplementary fig. S1*C*, Supplementary Material online): 1) A highly variable proportion of tRNA isoacceptor genes within the array units compared with nonarrayed regions, 2) 16 tRNA isoacceptor genes that were more represented in tRNA array units than in nonarrayed regions (e.g., tRNA-Leu^{UAG} and tRNA-Met^{CAU}), and 3) that on the contrary, some tRNA genes were underrepresented in the tRNA array units (e.g., tRNA-Arg^{CCU} and tRNA-Ser^{GGA}). Except for the tRNA-Arg^{UCG} gene, all tRNA genes displaying a high frequency in array units were included in the set of high tRNA indexes and interestingly, 14 of them were found to belong to the minimal set of tRNA isoacceptor genes (see supplementary information, Supplementary Material online). Altogether, these results suggest that the presence of tRNA array units may contribute toward tRNA gene redundancy in prokaryotic organisms.

The tRNA gene redundancy was then computed for each prokaryote and found to range from 1 (organisms with a unique tRNA isoacceptor gene in at least 2 copies) (e.g., *Pyrobaculum calidifontis* JCM 11548 and *Aciduliprofundum boonei* T469) to 43 (organisms in which 43 tRNA isoacceptors were present in at least two copies) (e.g., *Treponema pallidum* SS14 and *Candidatus cloacamonas acidaminovorans*) (supplementary fig. S2*A*, Supplementary Material online). Interestingly, the strong correlation (*r* value of 0.895) observed between the number of tRNA genes and the tRNA gene redundancy (supplementary fig. S2*B*, Supplementary Material online) clearly suggests that the redundancy is one of the factors associated with the increasing number of tRNA genes in many organisms such as *A. ferrivorans* CF27 (tRNA gene redundancy of 23) and *A. ferrooxidans* ATCC 23270^T (tRNA gene redundancy of 31). This then leads to the following question: Is the presence of tRNA arrays linked to the number of tRNA genes within an organism, to the tRNA gene redundancy, or to the tRNA gene diversity? To answer this, several distribution analyses were performed on organisms bearing 0, 1, or 2 tRNA array units. As shown in supplementary figure S3*A*, Supplementary Material online, the number of tRNA genes clearly increased with the number of tRNA array units per organism. Indeed, the median values associated with genomes containing 0, 1 and 2 tRNA arrays were 48, 67 and 104, respectively, suggesting a high correlation between the number of tRNA genes and the presence of tRNA arrays within the genome. Consequently, the tRNA gene redundancy also correlated with the presence of tRNA arrays within the genome, as the median tRNA gene redundancy increased from 5 to 20, and 22 for organisms with 0, 1 and 2 tRNA arrays, respectively (supplementary fig. S3*B*, Supplementary Material online). On the contrary however, the tRNA gene diversity slightly decreased with the presence of tRNA array units (supplementary fig. S3*C*, Supplementary Material online).

Next, the number of tRNA genes, the tRNA gene diversity, as well as the tRNA gene redundancy at the codon level were compared between prokaryotes with or without tRNA arrays. The tRNA gene diversity index was calculated as the proportion of arrayed or nonarrayed organisms harboring a particular isoacceptor. The results (fig. 3*A*) show that the majority of isoacceptors (e.g., tRNA-Ala^{UGC} and tRNA-Phe^{GAA}) had a similar tRNA gene diversity index between the two sets of

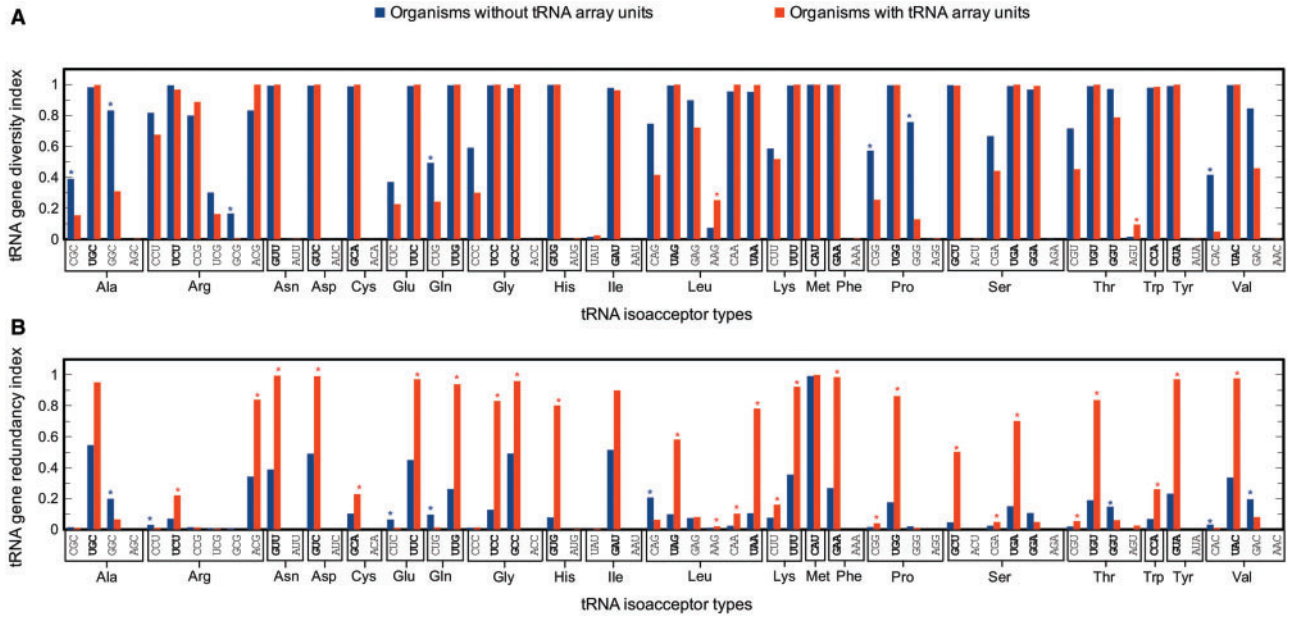
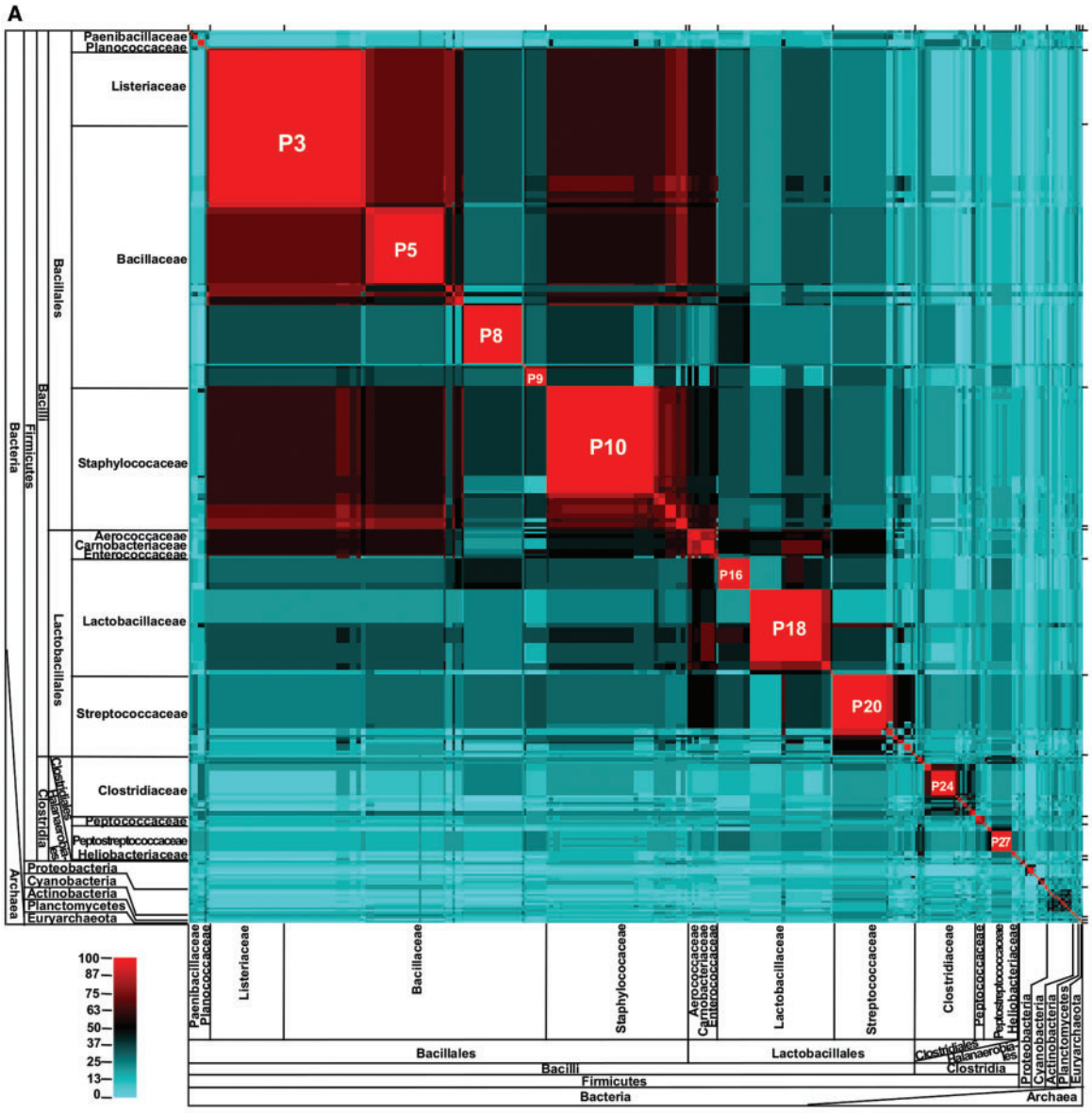


Fig. 3.—The tRNA gene diversity index (*A*) and tRNA gene redundancy index (*B*) in organisms with and without tRNA array units. For each tRNA isoacceptor, the tRNA gene diversity index (i.e., the proportion of organisms harboring the isoacceptor) in arrayed and nonarrayed organisms was computed. Ratios between the tRNA gene diversity index with array and that without (or the opposite) were calculated. Also for each tRNA isoacceptor, the tRNA gene redundancy index (i.e., the proportion of organisms harboring a redundancy for the isoacceptor) in arrayed and nonarrayed organisms was computed. Ratios between the tRNA gene redundancy index with array units and that without (or the opposite) were calculated. Ratios ≥2 are shown by a blue or red star. Only the tRNA isoacceptors with more than ten genes (across all organisms) were considered relevant for this analysis. Core tRNA isoacceptor genes (as defined in supplementary information, Supplementary Material online) are in bold characters.

organisms. In seven specific cases (e.g., tRNA-Ala^{CGC} and tRNA-Val^{CAC}), the tRNA gene diversity index was significantly higher (more than 2-fold) in nonarrayed organisms whereas the opposite was true for two tRNA genes (tRNA-Leu^{AAG} and tRNA-Thr^{AGU}). Similarly, the tRNA gene redundancy index was assessed as the proportion of arrayed or nonarrayed organisms harboring a redundancy for a particular isoacceptor (fig. 3*B*). Comparison between arrayed and nonarrayed organisms revealed a significantly higher tRNA gene redundancy index in arrayed organisms for 27 tRNA isoacceptors (over 59 represented tRNA types in arrayed organisms), therefore suggesting that tRNA array units favor tRNA gene redundancy but not diversity in prokaryotic organisms. In addition, 82% of these tRNA isoacceptors belonged to the core set of essential tRNA genes (supplementary information, Supplementary Material online), suggesting that in order to increase the number of essential tRNA genes, one strategy adopted by prokaryotes has been to favor the presence of array units in their genomes. One advantage to species with tRNA gene redundancy (and consequently high tRNA gene copy number) is that loss of one tRNA gene may have relatively little impact on the overall tRNA population in the cell for that particular tRNA functional isoacceptor. As a high redundancy of tRNA genes indeed facilitates tRNA anticodon mutation as previously suggested (Rogers and Griffiths-Jones 2014), organisms with tRNA array units would be subjected to anticodon shifts.

The tRNA Gene Array Organization Reveals Two Distinct tRNA Array Ancestors in Bacilli Class

The genomic arrangement of tRNA array units was next investigated in a wide range of prokaryotes, to identify common patterns and taxon-specific peculiarities in gene evolutionary events (rearrangements, deletions, insertions, and duplications). As shown in the results (fig. 4), arrays from distinct taxonomic phyla and classes displayed no similarities. However, the tRNA array units from different taxonomic families, genera, and species displayed over 50% structural similarity; this was the case for Listeriaceae array units, which shared almost 70% similarity with those of Staphylococcaceae and Bacillaceae families. The tRNA array units from *A. ferrivorans* CF27 and *A. ferrooxidans* ATCC 23270^{T} shared 75% structural similarity. Such similarities between arrays indicate that they had a common ancestor that was followed by tRNA gene rearrangement events during evolution. Finally, 314 tRNA array units (~82% of the total) belonging to 286 organisms could be clustered into 29 groups of different core tRNA array patterns (here the shared structural similarity was 100%, red diagonal blocks in fig. 4*A*) (supplementary table S5, Supplementary Material



B

Core tRNA array patterns

P1	EYNVDFKLGCRPATSSVTSL	P11	VTKLGLRPAMMSSMDFTYWH	P21	ENAVDKLTGLRPMMSMFGIS
P2	ANMSEVMDFTYHQKLGCRPG	P12	VTKLGLRPAMMSMDSMDFTY	P22	EAVDKLTGLRRPMMSMFGIS
P3	VTKLGLRPAMMSMDFHGINS	P13	VTKLGLRPAMMSMDFTYWHQ	P23	HQKLGGRGKHQKLGGRKPGR
P4	VTKLGLRPAMMSMDFTHGIN	P14	VKLTGLRPAMMSMDGISESM	P24	RHQKLGGRKHQKLGGRKPGR
P5	VTHLGLRPAMMSMDFTKGIN	P15	VKLTGLRPAMMSMDGIESMD	P25	QKKLGGRPGRHQKLGGRAIN
P6	VTKLGRPAMMSMDFHGINSE	P16	IAVKTGLRPMMSMDFGISEM	P26	GRTLSNMGEVDYMQKGRTLS
P7	VTKLGLRPAMSMDFHGINSE	P17	VKVKTGLRPMMSMDFGISEM	P27	GVDTYLRQSFMMPHKCNLME
P8	VYQKLGLRPASSMDFTWINE	P18	RPMMMDFGISESMDFYWHQC	P28	ANEVDTYGRQKSSPWPIFMA
P9	MDFTKGLRPAMMSTHGINSE	P19	VKLTGLRPMMMFGISESMDF	P29	EEEEEEEEEEAVVVVVVVVV
P10	VTKLGLRPAMMSDSMDFTYW	P20	DKLTGLRPAVDKLTGLRPMM		

Fig. 4.—(*A*) Comparative analysis of tRNA array units. The array structure similarity between two tRNA array units was computed as described in Materials and Methods. A heatmap graphical representation that displays the array structure similarities between all array tRNA units (383 × 383 scale matrix) is shown, with their associated taxonomic levels (phylum, class, order, and family). The color scale represents the percentage of array structure similarity between two tRNA array units. The main tRNA array patterns (P3, P5, P8, P9, P10, P16, P18, P20, P24, and P27) are highlighted. (*B*) Core tRNA array patterns associated with tRNA array units. From the comparative analysis, strong array structure similarities (100% identity, red blocks in *A*) between tRNA array units allowed core tRNA array patterns to be defined (P1–P29) (see also supplementary table S5, Supplementary Material online). The tRNA gene content of each pattern is shown using the single-letter amino acid code.

online). The tRNA array units in the Bacilli class could be clustered into four main core patterns P3, P5, P10 and P18 containing 65, 30, 45 and 31 array units, respectively. Except for P3 and P14, respectively, shared by Listeriaceae/Bacillaceae and Carnobacteriaceae/Enterococcaceae, each core pattern appeared specific to a particular family or genus. More surprisingly, although P5 and P8 belong to two distinct tRNA array units from the same species, they shared less than 30% structural similarity. Similar observations were made with the three tRNA array units present in *B. thuringiensis* Bt407. Indeed, the tRNA array unit located on the plasmid displayed no structural similarity with either of the two other arrays. In addition, each of the chromosomal arrays displayed either a P5 or a P8 pattern. The high diversity of the core array patterns combined with the distinct number of tRNA arrays in each core pattern (fig. 4*B* and supplementary table S5, Supplementary Material online) may suggest their distinct origins and probably multiple evolutionary histories.

Furthermore, multiple alignments of the full length core tRNA patterns revealed that these patterns (representing ~82% of all the tRNA arrays) could be clustered into seven distinct groups (supplementary table S6, Supplementary Material online). Within each group, differences in the core array patterns were due to a small number of insertion/deletions or duplication events. This was the case for Group III, with the duplication of tRNA[PGRHQKLGGR] set (in square brackets are given the tRNA-AA isotypes using the standard single letter amino acid abbreviation) and the translocation of tRNA[PGR] or insertions of tRNA-Gly and tRNA-Lys in some tRNA arrays, and Group V, with the duplication of the tRNA[NLMEGVDTYLRQSFMMPHKC] set and a loss of the tRNA-Pro in one of the duplicated copies. Interestingly, multiple duplications of tRNA-Glu and tRNA-Val resulted in a P29 array pattern (Group VII) in *Schewanella* strains, with more than 12 and 8 copies of the same tRNA gene, respectively.

A detailed and comparative analysis of core patterns from Group II (P3–P22, Bacilli class) mainly revealed (fig. 5): 1) Numerous tRNA arrays that had undergone various evolutionary events—deletions (48 in total, e.g., tRNA-Ala^{UGC} in P7), duplications (40 in total, e.g., tRNA-Leu^{CAA} in P19), translocations (8 in total, e.g., tRNA-Lys^{UUU} in P5), and insertions (38 in total, e.g., tRNA-His^{GUG})—indicating that tRNA arrays were likely subjected to tRNA rearrangements which then led to unstable genomic regions; 2) similarities between some structures (B and D in fig. 5) that could represent the common ancestral tRNA array region; 3) a probable insertion of the tRNA[SMDGISE] set occurring in P14, P15, P18, and P19 (C in fig. 5), that may have resulted from duplication in P14 and P15 (C and E1 in fig. 5); 4) a similar duplication event also occurring in P20 with tRNA[AVDKLTGLRP] genes (A in fig. 5); and 5) that the last tRNA region (E in fig. 5) harbored two distinct conserved structures E1 and E2 shared by P3–P7, P9, P14–P17, P20–P22 (fig. 5, left) and P8, P10, P11–P13, P18–P19 (fig. 5, right), respectively. The relative number of duplications in the patterns highlights tRNA gene redundancy as the likely driving factor responsible for the evolutionary history of array units. Altogether, these findings support the existence of at least two distinct "common ancestor" array sets (tRNA[AVTKLTGLRPAMMSMDFGINSE] (AC1) and tRNA[AVTKLTGLRPAMMSMDFYWHQCGGL] (AC2); probable translocation event for the T in gray, see fig. 5) for the Bacilli class, except for the Paenibacillaceae family. Interestingly, all 26 distinct tRNA isoacceptors of AC1 and AC2 belonged to the minimal tRNA gene set. The location of AC1 in 198 organisms (representing, respectively, 70% of the tRNA arrayed Bacilli organisms and 56.7% of all organisms with tRNA arrays), combined with the P18 and P19 array patterns constituting a mix of the AC1 and AC2 ancestors, does however allow the rationale of a unique common ancestor likely similar to AC1 to be reasonably inferred.

Deciphering the Evolutionary History of tRNA Array Units

tRNA genes are ancient molecules that can be traced back to the putative RNA World before the separation of the three domains of life (Sun and Caetano-Anolles 2008; Bermudez-Santana et al. 2010) and their origin and evolution represents an exciting topic (Di Giulio 2006, 2012; Fujishima and Kanai 2014). To shed light on the origin and evolutionary history of tRNA arrays, a tree representing the occurrence and distribution of tRNA arrays over the taxonomic phyla, classes, orders, families and genera of Bacteria and Archaea, as well as the evolutionary events that occurred was set up. As tRNA array units are mainly found in Firmicutes, one can reasonably assume that they probably originated in this phylum through an HGT event (from an unknown donor) before their propagation in other Firmicutes and in other phyla through vertical inheritance or HGT events. This hypothesis is supported by the presence of syntenic gene locations among main species groups within a genus (e.g., *Bacillus* strains with P3, P5–P7 array patterns from the AC1 ancestor). Such features were not found in organisms outside of the genus such as *Geobacillus* and *Listeria* species, which were found to harbor arrays with P3–P4 and P9 patterns (supplementary fig. S4*A*, Supplementary Material online).

Four possible evolutionary histories of the tRNA arrays in Bacilli class were then hypothesized based on a unique common ancestor (AC1) or two common ancestors (AC1 and AC2) with late or early acquisitions. Evolutionary histories involving either two ancestors with early acquisition (supplementary fig. S5*A*, Supplementary Material online), or one unique ancestor with late acquisition (supplementary fig. S5*B*, Supplementary Material online) seemed less parsimonious, as these implied 1) a huge number of AC1 or AC2 losses in the same downstream family clades (see supplementary fig. S5*A*, Supplementary Material online), or 2) multiple and independent AC1 acquisition events followed by massive loss of tRNA[GINSE] units and acquisitions of tRNA[YWHQCGGL]

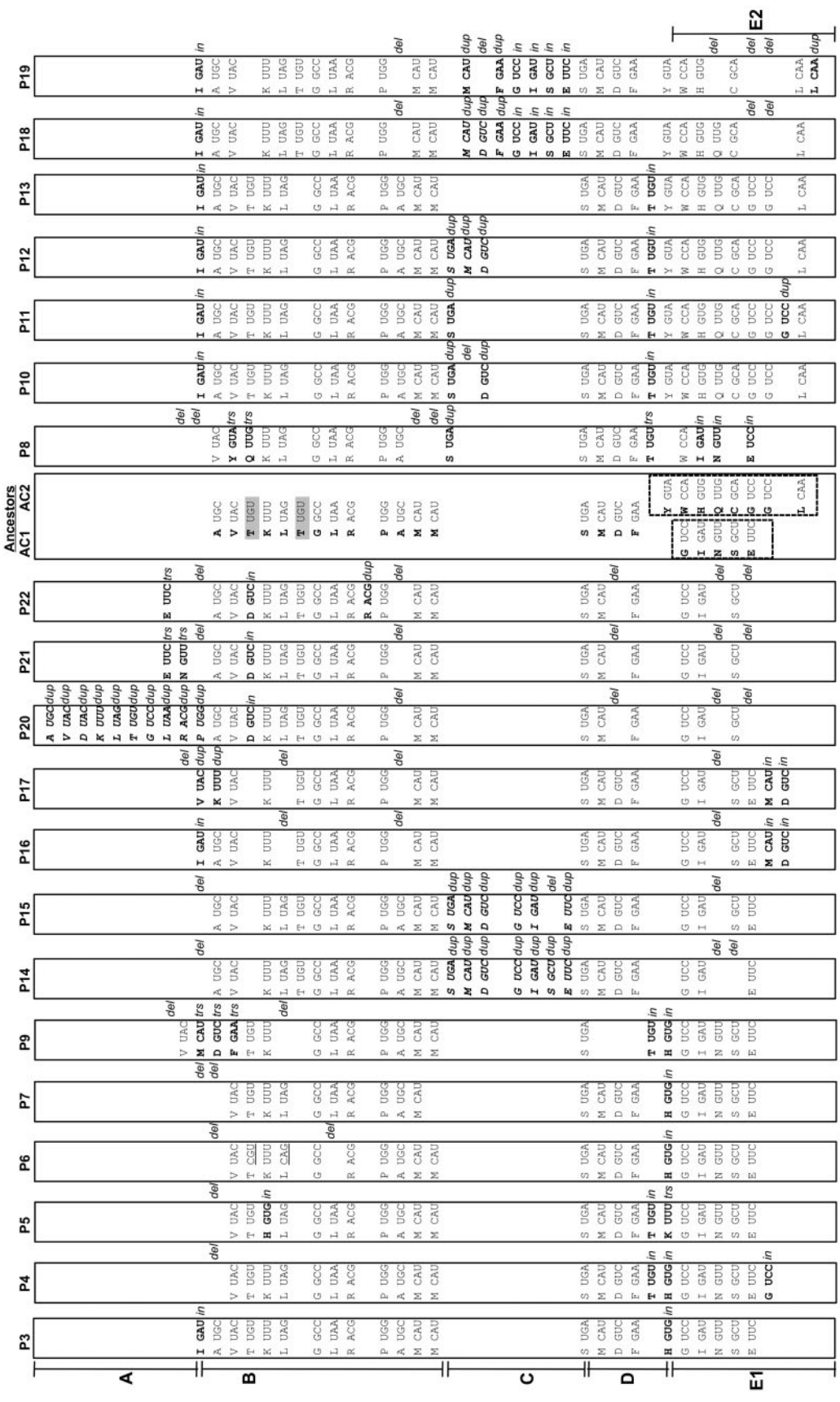


Fig. 5.—The tRNA gene evolutionary events along the tRNA array units in Bacilli class. The tRNA isoacceptors are given with their codons and the corresponding amino acids. Codon changes are underlined. Chromosomal mutations including translocation (*trs*), insertion (*in*), deletion (*del*), and duplication (*dup*) are in bold. Blocks of tRNA gene duplications are shown in italics. Small unit conserved structures (A, B, C, D, E1, and E2) between tRNA array patterns allow two probable common ancestors (AC1 and AC2) to be defined. Within the ancestors, the tRNA-Thr^{UGU} gene (in gray background) was subjected to translocation.

units in several Bacilli family clades, respectively (supplementary fig. S5*B*, Supplementary Material online).

One of the most parsimonious evolutionary histories involving two common ancestors (AC1 and AC2) suggests that acquisition of these arrays arose late in evolution, probably in genera branches and at the species level (fig. 6*A*). Indeed, evidence of AC1 acquisition was found in *Bacillus* and *Geobacillus* genera (for Bacillaceae family) and in some species (e.g., *Streptococcus* sp.) of the Lactobacillales order, whereas that of AC2 was specific to the Staphylococcaceae family branch and numerous *Lactobacillus* species. In particular, these acquisitions in *Enterococcus* species (related to P14 and P15 patterns) or in *Lactobacillus* species (related to P18 and P19 patterns) were followed by duplication or insertion of the tRNA[SMDGISE] gene set, respectively (as described above). Interestingly, some species, including *Bacillus anthracis*, *B. thuringiensis*, and *Bacillus cereus*, harbored two copies of tRNA arrays (from P5 and P8 patterns) that originated independently of the two common ancestors, with acquisition of AC1 occurring after the emergence of *Bacillus* and *Geobacillus* genera, followed by a specific acquisition of AC2 in the three species groups (fig. 6*A*). This evolutionary hypothesis is supported by comparative gene synteny around the array unit insertions (supplementary fig. S4*A*, Supplementary Material online), which mainly shows that for AC1, insertion positions of tRNA array units from *Bacillus* species (harboring P3, P5–P7 patterns) are the same but differ from those in *Geobacillus* and *Listeria* organisms (harboring P3–P4 and P9 patterns). This observation also indicates that specific and independent acquisitions of AC1 occurred in species of *Bacillus*, *Geobacillus*, and *Listeria* with different insertion times (after the speciation of *Bacillus* and *Geobacillus* for both genera; probably in the family branch for *Listeria*). The same observations can be made for AC2 (supplementary fig. S4*B* with *Bacillus* strains harboring P10–P13 patterns, Supplementary Material online) which has been independently and specifically acquired in *B. anthracis*, *B. thuringiensis*, *B. cereus*, and Staphylococcaceae branches.

The final most parsimonious evolutionary history (fig. 6*B*) hypothesizes the presence of a unique common ancestor (here AC1), and is based on a single and early acquisition of AC1 ancestor in the Bacilli branch, followed by vertical inheritance into Bacillales and Lactobacillales orders and subsequent spreading over the genera and species. Consequently, the tRNA array units would share high structural similarity, as described above with patterns P3–P22. The low number (96 of 224) of Lactobacillales organisms containing tRNA array units (compared with that observed in the Bacillales branch) clearly indicates a massive loss of the tRNA array units in the Lactobacillales. Based on this scenario, the first main evolutionary highlight is that the tRNA arrays in *B. anthracis*, *B. thuringiensis*, and *B. cereus* which harbored two copies of tRNA arrays might have resulted from the duplication of the AC1 ancestor within the branch: The ones incorporating the P5 pattern remaining nearly identical to the common ancestor AC1, whereas the second copy relating to the P8 pattern were subjected to deletions (e.g., tRNA-Gly and tRNA-Ser) and insertion of tRNA-Trp within the tRNA[SMDGISE] gene set during evolution. A recent study suggested that rather than a single duplication preceding the *Bacillus* ancestor, two inverted duplications around the origin of replication independently occurred in each group [*B. cereus*, *B. anthracis*] and [*B. licheniformis*, *B. subtilis*] that was followed by losses of the same regions in each of the two monophyletic groups (Holloway et al. 2013). The second evolutionary event based on this scenario would have been the deletion of the tRNA[GINSE] set and acquisition of the tRNA[YWHQCGGL] unit in the tRNA arrays of the *Macrococcus* and *Staphylococcus* species (with P10, P11, P12, and P13 patterns), whereas the formation of P14 and P15 pattern array units derived from specific duplication of the tRNA[SMDGISE] gene set in Carnobacteriaceae and Enterococcaceae family branches (after loss of the tRNA-Asn^{GUU} gene). Similar evolutionary history can be drawn for some clusters of *Lactobacillus* species (with P18 and P19 patterns) for which the early acquisition of the AC1 ancestor would have been followed by duplication of the tRNA[SMDGINSE] gene set, followed by a later recent deletion of the tRNA[GINSE] set within the downstream duplicated copy and finally a recent insertion of the tRNA[YWHQCGGL] set. The appearance of tRNA array units (relative to P1 and P2 patterns) in *Paenibacillus* species may be the result of a complete rearrangement of the tRNA genes from the AC1 ancestor copy with multiple tRNA gene losses, or from independent and specific insertion events from a P1/P2-like ancestor. Based on the data currently available, it remains impossible to distinguish the most likely explanation. We consider that both evolutionary hypotheses in Bacilli are equally parsimonious as they involve almost the same number of major evolutionary events (20 and 18 for the first and the second, respectively).

The possible evolutionary histories of tRNA arrays in the Clostridia taxonomic class are as follows (see supplementary fig. S5*C*, Supplementary Material online): 1) For the Clostridiaceae order, the AC3 ancestor may resemble tRNA[PGRHQKLGGR] and was probably recently acquired at the genus level in 25 strains (over 46 organisms), prior to duplication of AC3 ancestor and minor rearrangement events (translocation of tRNA[PGR] set, insertion of tRNA-Gly and tRNA-Lys) in some arrays; 2) similarly, specific and recent acquisitions of tRNA[EVDYMQKGRTLSNMG] (AC6 ancestor) and [NLMEGVDTYLRQSFMMPHQC] (AC7 ancestor) occurred in four *Peptococcaceae* strains (over 19) and ten *Peptostreptococcaceae* strains (over 11), respectively, that were followed by their duplications; and 3) independent and specific insertion events from AC5 and AC8 ancestors recently occurred in *Heliobacterium* and *Halobacteroides* genera, respectively, as demonstrated by the lack of structural similarity

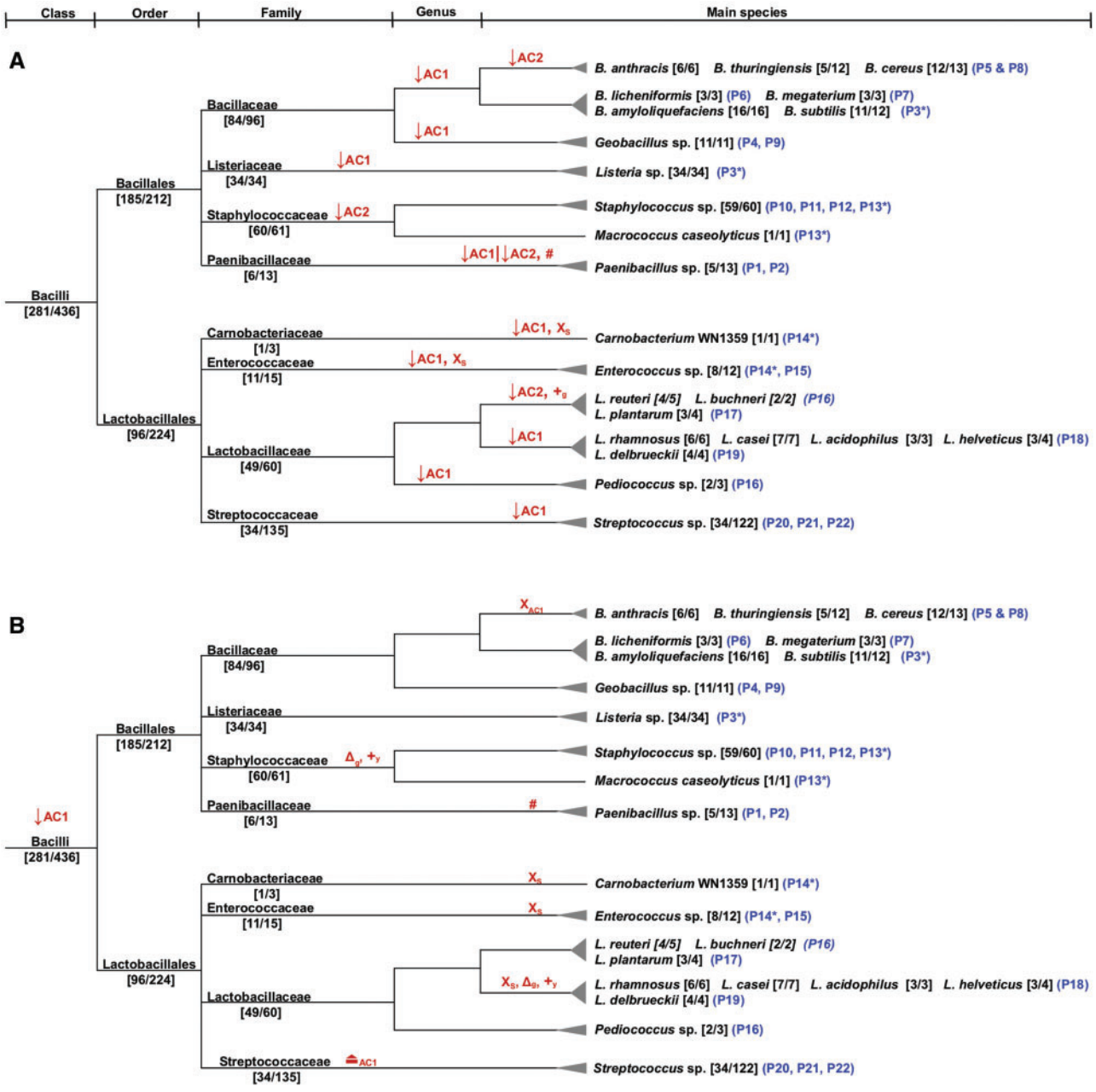


Fig. 6.—Evolutionary histories of tRNA array units in Bacilli class. Taxonomic subdivisions were taken from the NCBI taxonomic browser. Only Bacilli clades harboring tRNA array units with P patterns are shown. The two most parsimonious scenarios ([*A*] two Bacilli common ancestors [AC1 and AC2] with late acquisition; [B] unique common ancestor [AC1] with early acquisition) are shown. The numbers in brackets represent the number of organisms harboring tRNA array units and all the organisms of the clade. Core tRNA array patterns are shown in parentheses. ↓AC, acquisition of the tRNA array units; X_{AC}, duplication of ancestor tRNA array unit; X_S, duplication of the tRNA[SMDGISE] gene set; Δ_g, deletion of the tRNA[GINSE] gene set; $+_g$, insertion of the tRNA[SMDGISE] gene set; $+_\gamma$, insertion of the tRNA[YWHQCGGL] gene set; ⏏$_{AC}$, massive loss of tRNA array units in the clade; #, multiple events of tRNA gene losses and insertions; and *, tRNA array patterns shared by two genera.

between these tRNA arrays and others. As tRNA array units located in other phyla (Actinobacteria, Cyanobacteria, Euryarchaeota, Planctomycetes, and Proteobacteria) were found present in a low number of organisms (see supplementary table S3, Supplementary Material online) and that these array units shared a low percentage of structural similarity with others, they may have originated from independent insertions from various ancestors. Altogether, our findings indicate that tRNA array units have been subjected to complex evolutionary histories over time.

Discussion

In this study, tRNA array units within prokaryotic organisms have been identified and characterized. These genomic structures appeared to be a general feature of Firmicutes genomes, but were highly polymorphic and largely confined to the Bacilli class. The degree of similarity in tRNA array organization reflected phylogenetic relatedness, with high structural similarity between members of the same genus and species but low structural similarities with organisms outside of the genus. As this type of organization is widespread in low GC Gram-positive bacteria, it may be important during spore germination and outgrowth, when rapid activation of the translation machinery is required (Green and Vold 1993; Rudner et al. 1993). However, as tRNA array units were found in nonsporulating prokaryotes, such clustered tRNA gene organization does not appear to be an exclusive requirement for sporulation. In *E. coli*, tRNA arrays significantly influence global tRNA expression at high growth rates (Ardell and Kirsebom 2005). By molecular differential analysis at the population level, tRNA arrays located at the 3′-end of rRNA operons were found to be specific to virulent subgroups of *Streptococcus agalactiae* strains (Rolland et al. 2002). This raises the possibility of a link between these tRNA genes and the virulence of the bacterium. No relationship was found between the presence of the array units and the organism's lifestyle (aerobic or anaerobic, optimal growth temperature, pathogenicity or environmental conditions) (data not shown). Even when *A. ferrivorans* CF27/*A. ferrooxidans* ATCC 23270^T were compared with other *Acidithiobacillus* organisms, no specific physiological features were able to explain the acquisition and maintenance of tRNA array units.

For most tRNA array units (98% of the total), the tRNA genes are transcribed in the same orientation, suggesting that transcription may occur as one (or in a limited number of) operon(s). Hence, this leaves us with an important question: Do these array unit organization convey a selective genetic advantage to their host? One advantage of tRNA gene clustering is that it reduces the number of tRNA transcriptional units, suggesting that the tRNA content in organisms with tRNA array units could be maintained by a simple gene dosage effect. Such adaptations may favor a more efficient translational regulation thus allowing a rapid and coordinated response to environmental stresses. Indeed, it was recently shown that elevation in tRNA levels accelerated translation and protected *E. coli* against oxidative stress caused by hydrogen peroxide and the antibiotic ciprofloxacin (Zhong et al. 2015). On the contrary, one disadvantage of having a highly clustered tRNA gene organization is the difficulty finely tuning the transcription of each tRNA gene under different growth conditions. One last compelling theory is that controlling the expression of individual and clustered tRNA species enables different levels of translational control, which may be useful for specific gene products or under certain growth conditions.

Another important question that remains is: Does the presence of tRNA arrays enhance the efficiency, fidelity or specificity of the translation? For that purpose, the structural organization (same order in amino acids) of the arrays was analyzed as a protein motif or pattern within proteomes of organisms harboring array units compared with nonarrayed organisms. However, there is no difference between the two sets of organisms suggesting that the structural organization of the arrays is not directly linked to the translation efficiency (data not shown). One unique feature of genetic code redundancy allows choosing between alternative codons for the same amino acid. As the efficiency of translation mainly depends on the concentrations of each corresponding tRNA (Fernandez-Vazquez et al. 2013), the increase in redundancy of core tRNA gene sets within organisms harboring array units would be expected to favor translation efficiency.

Comparative array organization revealed that tRNA array units were acquired through HGT (from Firmicutes or unknown donor), before being subjected to tRNA rearrangements, deletions, insertions, and duplications. The evolutionary history of tRNA array units in Bacilli class shows that they may have originated from at least two common ancestors, from unknown donors. More surprisingly, the low number of tRNA arrays in phyla other than Firmicutes and the low structural similarities among each array unit clearly suggest a reduced mobility of these array units across the taxonomic clades and consequently their evolutionary specificity to one clade. Interestingly, the evolutionary history of tRNAs favored the tRNA gene redundancy instead of tRNA gene diversity, in particular the core tRNA isoacceptors. Therefore, the evolutionary bias for tRNA gene redundancy of the core tRNA isoacceptors within the array units remains in agreement with the wobble rules. This observation leads to an open question: Is the wobble effect stands as a consequence of the tRNA gene redundancy or the opposite? As the wobble effect can explain the degeneracy of the genetic code, this will suggest a possible link between the tRNA gene redundancy and the degeneracy of the code, which in turn will stimulate further research and developments in the RNA field.

Supplementary Material

Supplementary information, figures S1–S5, and tables S1–S6 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by Aix-Marseille Université (AMU), Centre National de la Recherche Scientifique (CNRS), and the Vietnamese Ministry of Education and Training (PhD fellowship to T.T.T.T.). The authors thank Patricia Chan (Lowe Lab, University of California Santa Cruz) for her kind advice on the tRNA gene detection procedure as well as to Sebastien

Tempel for his critical reading of the manuscript and to Sophie Mangenot, Zoe Rouy and Valerie Barbe for preliminary annotations of the *Acidithiobacillus ferrivorans* CF27 strain.

Literature Cited

Abe T, et al. 2011. tRNADB-CE 2011: tRNA gene database curated manually by experts. Nucleic Acids Res. 39:D210–D213.
Abe T, et al. 2014. tRNADB-CE: tRNA gene database well-timed in the era of big sequence data. Front Genet. 5:114.
Ardell DH, Kirsebom LA. 2005. The genomic pattern of tDNA operon expression in *E. coli*. PLoS Comput Biol. 1:e12.
Bermudez-Santana C, et al. 2010. Genomic organization of eukaryotic tRNAs. BMC Genomics 11:270.
Bustamante P, et al. 2012. ICE Afe 1, an actively excising genetic element from the biomining bacterium *Acidithiobacillus ferrooxidans*. J Mol Microbiol Biotechnol. 22:399–407.
Chan PP, Cozen AE, Lowe TM. 2011. Discovery of permuted and recently split transfer RNAs in Archaea. Genome Biol. 12:R38.
Chan PP, Lowe TM. 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res. 37:D93–D97.
Di Giulio M. 2006. The non-monophyletic origin of the tRNA molecule and the origin of genes only after the evolutionary stage of the last universal common ancestor (LUCA). J Theor Biol. 240:343–352.
Di Giulio M. 2012. The origin of the tRNA molecule: independent data favor a specific model of its evolution. Biochimie 94:1464–1466.
Fernandez-Vazquez J, et al. 2013. Modification of tRNA(Lys) UUU by elongator is essential for efficient translation of stress mRNAs. PLoS Genet. 9:e1003647.
Friedrich A, Jung PP, Hou J, Neuveglise C, Schacherer J. 2012. Comparative mitochondrial genomics within and among yeast species of the *Lachancea genus*. PLoS One 7:e47834.
Fujishima K, Kanai A. 2014. tRNA gene diversity in the three domains of life. Front Genet. 5:142.
Goodenbour JM, Pan T. 2006. Diversity of tRNA genes in eukaryotes. Nucleic Acids Res. 34:6137–6146.
Green CJ, Vold BS. 1983. Sequence analysis of a cluster of twenty-one tRNA genes in *Bacillus subtilis*. Nucleic Acids Res. 11:5763–5774.
Green CJ, Vold BS. 1992. A cluster of nine tRNA genes between ribosomal gene operons in *Bacillus subtilis*. J Bacteriol. 174:3147–3151.
Green CJ, Vold BS. 1993. *Staphylococcus aureus* has clustered tRNA genes. J Bacteriol. 175:5091–5096.
Holloway P, Swenson K, Ardell D, El-Mabrouk N. 2013. Ancestral genome organization: an alignment approach. J Comput Biol. 20:280–295.
Jung PP, et al. 2010. Complete mitochondrial genome sequence of the yeast *Pichia farinosa* and comparative analysis of closely related species. Curr Genet. 56:507–515.
Kanaya S, Yamada Y, Kudo Y, Ikemura T. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene 238:143–155.
Komine Y, Adachi T, Inokuchi H, Ozeki H. 1990. Genomic organization and physical mapping of the transfer RNA genes in *Escherichia coli* K12. J Mol Biol. 212:579–598.
Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32:11–16.
Lee ZM, Bussema C 3rd, Schmidt TM. 2009. rrnDB: documenting the number of rRNA and tRNA genes in Bacteria and Archaea. Nucleic Acids Res.37:D489–D493.
Levican G, et al. 2009. A 300 kpb genome segment, including a complete set of tRNA genes, is dispensable for *Acidithiobacillus ferrooxidans*. Adv Mat Res. 71–73:187–190.
Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.
Marck C, Grosjean H. 2002. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. RNA 8:1189–1232.
Novoa EM, Pavon-Eternod M, Pan T, Ribas de Pouplana L. 2012. A role for tRNA modifications in genome structure and codon usage. Cell 149:202–213.
Puerto-Galan L, Vioque A. 2012. Expression and processing of an unusual tRNA gene cluster in the cyanobacterium *Anabaena* sp. PCC 7120. FEMS Microbiol Lett. 337:10–17.
Raina M, Ibba M. 2014. tRNAs as regulators of biological processes. Front Genet. 5:171.
Rocha EP. 2004. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res. 14:2279–2286.
Rogers HH, Griffiths-Jones S. 2014. tRNA anticodon shifts in eukaryotic genomes. RNA 20:269–281.
Rolland K, Mereghetti L, Watt S, Chatellier S, Quentin R. 2002. tRNA gene clusters at the 3′ end of rRNA operons are specific to virulent subgroups of *Streptococcus agalactiae* strains, as demonstrated by molecular differential analysis at the population level. Microbiology 148:1493–1499.
Rudner R, et al. 1993. Two tRNA gene clusters associated with rRNA operons *rrnD* and *rrnE* in *Bacillus subtilis*. J Bacteriol. 175:503–509.
Satapathy SS, Dutta M, Ray SK. 2010. Variable correlation of genome GC% with transfer RNA number as well as with transfer RNA diversity among bacterial groups: alpha-proteobacteria and tenericutes exhibit strong positive correlation. Microbiol Res. 165:232–242.
Smith DR. 2015. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? Brief Funct Genomic. p1–8.
Sprinzl M, Vassilenko KS. 2005. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 33:D139–D140.
Sun FJ, Caetano-Anolles G. 2008. Evolutionary patterns in the sequence and structure of transfer RNA: early origins of archaea and viruses. PLoS Comput Biol. 4:e1000018.
Talla E, et al. 2014. Insights into the pathways of iron- and sulfur-oxidation, and biofilm formation from the chemolithotrophic acidophile *Acidithiobacillus ferrivorans* CF27. Res Microbiol. 165:753–760.
Tawari B, et al. 2008. Patterns of evolution in the unique tRNA gene arrays of the genus *Entamoeba*. Mol Biol Evol. 25:187–198.
Vallenet D, et al. 2013. MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. Nucleic Acids Res. 41:D636–D647.
Vold BS. 1985. Structure and organization of genes for transfer ribonucleic acid in *Bacillus subtilis*. Microbiol Rev. 49:71–80.
Widmann J, Harris JK, Lozupone C, Wolfson A, Knight R. 2010. Stable tRNA-based phylogenies using only 76 nucleotides. RNA 16: 1469–1477.
Withers M, Wernisch L, dos Reis M. 2006. Archaeology and evolution of transfer RNA genes in the *Escherichia coli* genome. RNA 12:933–942.
Zhong J, et al. 2015. Transfer RNAs mediate the rapid adaptation of *Escherichia coli* to oxidative stress. PLoS Genet. 11:e1005302.

Associate editor: Emmanuelle Lerat