Life without tRNA<sub>Arg</sub>–adenosine deaminase TadA: evolutionary consequences of decoding the four CGN codons as arginine in Mycoplasmas and other Mollicutes

Shin-ichi Yokobori<sup>1,*</sup>, Aya Kitamura<sup>2</sup>, Henri Grosjean<sup>3</sup> and Yoshitaka Bessho<sup>2,*</sup>

<sup>1</sup>Laboratory of Extremophiles, Department of Applied Life Sciences, School of Life Sciences, Tokyo University of Pharmacy and Life Sciences, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan, <sup>2</sup>RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan and <sup>3</sup>Centre de Génétique Moléculaire, UPR 3404, CNRS, Associée à l'Université Paris-Sud 11, FRC 3115, 1 Avenue de la Terrasse, 91190 Gif-sur-Yvette, France

Received February 29, 2012; Revised April 13, 2013; Accepted April 15, 2013

ABSTRACT

In most bacteria, two tRNAs decode the four arginine CGN codons. One tRNA harboring a wobble inosine (tRNA<sub>Arg</sub><sub>I</sub><sup>CG</sup>) reads the CGU, CGC and CGA codons, whereas a second tRNA harboring a wobble cytidine (tRNA<sub>Arg</sub><sub>C</sub><sup>CG</sup>) reads the remaining CGG codon. The reduced genomes of Mycoplasmas and other Mollicutes lack the gene encoding tRNA<sub>Arg</sub><sub>C</sub><sup>CG</sup>. This raises the question of how these organisms decode CGG codons. Examination of 36 Mollicute genomes for genes encoding tRNA<sub>Arg</sub> and the TadA enzyme, responsible for wobble inosine formation, suggested an evolutionary scenario where the <i>tadA</i> gene mutations first occurred. This allowed the temporary accumulation of non-deaminated tRNA<sub>Arg</sub><sub>A</sub><sup>CG</sup>, capable of reading all CGN codons. This hypothesis was verified in <i>Mycoplasma capricolum</i>, which contains a small fraction of tRNA<sub>Arg</sub><sub>A</sub><sup>CG</sup> with a non-deaminated wobble adenosine. Subsets of Mollicutes continued to evolve by losing both the mutated tRNA<sub>Arg</sub><sub>C</sub><sup>CG</sup> and <i>tadA</i>, and then acquired a new tRNA<sub>Arg</sub><sub>U</sub><sup>CG</sup>. This permitted further tRNA<sub>Arg</sub><sub>A</sub><sup>CG</sup> mutations with tRNA<sub>Arg</sub><sub>G</sub><sup>CG</sup> or its disappearance, leaving a single tRNA<sub>Arg</sub><sub>U</sub><sup>CG</sup> to decode the four CGN codons. The key point of our model is that the A-to-I deamination activity had to be controlled before the loss of the <i>tadA</i> gene, allowing the stepwise evolution of Mollicutes toward an alternative decoding strategy.

INTRODUCTION

The genetic code is composed of 16 families of decoding boxes, each including four codons with the same first two nucleotides. Depending on the amino acid, these synonymous codons are read by one, two or at most three isoacceptor tRNA species harboring distinct anticodons. Therefore, fewer than 61 isoacceptor species (usually between 22 to a maximum of 46) are used to decode the 61 sense codons in mRNAs. These cellular tRNA repertoires are primarily responsible for the efficiency and accuracy of mRNA translation. The tRNA repertoires vary greatly from one organism and organelle to another, with most of the variability being found in the type of nucleotide present at the first ‘so-called’ wobble position of the anticodon (position 34), which is often post-transcriptionally modified. By interacting with the third base of the codon, the wobble-34 plays an essential role in determining the preferred codons to be read by the mature and functional tRNA (1–5).

Transfer RNAs harboring an unmodified wobble adenosine-34 are rare; thus, they are not frequently used during translation. The reason is that during tRNA maturation, the encoded wobble A<sub>34</sub> in the anticodon of the precursor tRNAs is generally enzymatically deaminated to inosine (6-deaminated adenosine–hypoxanthine base) by specific tRNA:A<sub>34</sub> deaminases. This results in the tRNA containing the A<sub>34</sub>-containing tRNA was predicted to base pair with a C-ending codon in the Watson–Crick mode and with U- and A-ending codons in a slightly different ‘wobble’ conformation (6), whereas the binding with a G-ending codon was forbidden, as reviewed previously (2,7). However, among

*To whom correspondence should be addressed. Tel: +81 791 58 2891; Fax: +81 791 58 2892; Email: bessho@spring8.or.jp
Correspondence may also be addressed to Shin-ichi Yokobori. Tel: +81 42 676 5035; Fax: +81 42 676 7145; Email: yokobori@ls.toyaku.ac.jp

© The Author(s) 2013. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
the three codons read by \( I_{34}\)-containing tRNA, the A-ending codon was expected to be difficult to translate, and this proposal was verified with *Escherichia coli* tRNA\(^{A^*_R}_{34}\) using an *in vitro* translation system (8). Confirmation of this wobble hypothesis, with both bases in the anti-conformation as initially predicted by Francis Crick, was finally obtained from the crystal structure of the 30S ribosomal subunit, with the anticodon stem loop derived from *E. coli* tRNA\(^{A^*_R}_{34}\) bound to the CGA codon in an mRNA fragment (9). Therefore, once a cell has evolved and begun using \( I_{34}\)-containing tRNA, the fourth remaining codon ending with G, in the corresponding four synonymous codons of the family box, has to be read by a second tRNA isoacceptor harboring a \( C_{34}\)-containing anticodon (Figure 1). Although this is the usual decoding strategy observed in many living cells (10,11), a few remarkable exceptions exist.

For example, in fungi and animals, all cytoplasmic tRNAs harboring a wobble \( A_{34}\) and a purine-35 (R\(_{35}\)) in the middle of the anticodon, as well as \( A_{34}\) in the cytoplasmic tRNA\(^{A^*_R}_{34}\) bound to the CGA codon in an mRNA fragment (9). Therefore, once a cell has evolved and begun using \( I_{34}\)-containing tRNA, the fourth remaining codon ending with G, in the corresponding four synonymous codons of the family box, has to be read by a second tRNA isoacceptor harboring a \( C_{34}\)-containing anticodon (Figure 1). Although this is the usual decoding strategy observed in many living cells (10,11), a few remarkable exceptions exist.

In contrast to cytoplasmic tRNA of eukaryotes, but similar to plant chloroplasts, inosine-34 in bacterial cytoplasmic mRNAs are translated into arginine. Only the chloroplast tRNA\(^{A^*_R}_{34}\) in *A. thaliana* (and probably in all land plants) contains a deaminated \( A_{34}\), and its formation is catalyzed by the nuclear encoded chloroplastic TadA, a deaminase that is similar to the bacterial ortholog (17,18).

In this report, we identified the tRNA\(^{A^*_R}_{34}\) set in the 36 fully sequenced genomes of Mollicutes currently available. This repertoire was then correlated with the presence or absence of a gene encoding a TadA deaminase in the Mollicute genome. This genomic analysis revealed that Mollicutes are evolving by setting up alternative, and probably more efficient, arginine decoding systems able to read all four \( G_{34}\) glycine codons, although the A-ending Gly-GGA codon was decoded with the lowest efficiency (21). Osawa and co-workers (22) experimentally proved in *vitro* that the naturally occurring \( A_{34}\)-containing tRNA\(^{A^*_R}_{34}\) from the bacterium *Mycoplasma capricolum* translates all four threonine ACN codons, and only the Thr-ACA codon showed greatly reduced efficiency. Notably, *M. capricolum* has evolved a second tRNA\(^{A^*_R}_{34}\) harboring an unmodified wobble \( U_{34}\) for reading the ACA codon without wobbling (23); therefore, it has naturally compensated for the difficulty of reading the Thr-ACA codon by \( A_{34}\)-containing tRNA\(^{A^*_R}_{34}\).

As for the mitochondria of the Tungus *Saccharomyces cerevisiae* and the nematode *Ascaris suum*, the *tadA* genes are missing in their nuclear genomes, and consequently, their encoded mitochondrial tRNA\(^{A^*_R}_{34}\) harbors an unmodified wobble \( A_{34}\) (24,25). As no other mitochondrial tRNA\(^{A^*_R}_{34}\) belonging to the same CGN arginine box exists, it was concluded that this unique tRNA\(^{A^*_R}_{34}\) must decode all four synonymous CGN codons. However, no experiments have been performed to verify this hypothesis.

*Escherichia coli* TadA and cytoplasmic *S. cerevisiae* Tad2/Tad3 are essential enzymes, and the deletions of the corresponding genes are lethal (13,19). Together, these experiments demonstrated that, at variance with the information reported in all textbooks, the essential inosine at the first anticodon position does not ‘extend’ the decoding capability of an \( A_{34}\)-containing tRNA. On the contrary, it ‘restricts’ the precursor tRNA harboring an unmodified wobble \( A_{34}\) to read only three of the four potential synonymous codons, excluding only the synonymous codon ending with G. This remaining synonymous G-ending codon of the same decoding box has to be decoded by a C\(_{34}\)-containing tRNA. However, as aforementioned, although \( I_{34}\)\(_{34}\)-A3 wobble pairing is possible (9), in practice it is inefficient (8), and cells usually limit the usage of codons involving \( I_{34}\) base pairing during translation (26–28).

In this report, we identified the tRNA\(^{A^*_R}_{34}\) set in the 36 fully sequenced genomes of Mollicutes currently available. This repertoire was then correlated with the presence or absence of a gene encoding a TadA deaminase in the Mollicute genome. This genomic analysis revealed that Mollicutes are evolving by setting up alternative, and probably more efficient, arginine decoding systems able to read all four \( G_{34}\) glycine codons, thus bypassing the requirement for the usually essential bacterial *tadA* gene.
MATERIALS AND METHODS

Data processing

All bacterial genomes analyzed were obtained from Genbank. They are listed in Supplementary Table S1. The genes encoding the TadA (tRNA-specific adenosine deaminase) and CDA (cytidine deaminase) protein sequences from the different Mycoplasmas analyzed were obtained from Genbank via BLASTP at NCBI, using TadA of Bacillus subtilis subsp. subtilis str. 168 (NP_387899.1) as the query sequence under the default conditions. The sequences of a few additional bacterial TadA proteins were obtained from published articles (Table 1). The tRNA<sup>Arg</sup> genes with the anticodons or TadA proteins were obtained from published articles (NP_387899.1) as the query sequence under the default conditions. The sequences of the different Mycoplasmas analyzed were deaminase) and CDA (cytidine deaminase) protein sequences from the different Mycoplasmas analyzed were obtained from Genbank. They are listed in Supplementary Table S1.

Data processing allowed us to assess the conserved amino acids and to distinguish the ones that are 'mechanistically' common under the default conditions. The TadA enzyme catalyzes the deamination of wobble A<sub>34</sub>-containing tRNA, whereas the CDA enzyme catalyzes the deamination of free cytidine to produce uridine. As the TadAs are apparently derived from an ancestral CDA (35), the comparison allowed us to assess the conserved amino acids and to distinguish the ones that are 'mechanistically' common to all members of the deaminase superfamily (CDA and TadA) from those that are specific to TadA, such as those composing the tRNA-binding motif.

cDNA analyses of M. capricolum and B. subtilis tRNA<sup>Arg</sup>

Bulk tRNA from B. subtilis strain 168 (wild-type) was obtained as described previously (36). Bulk tRNA from M. capricolum [American Type Culture Collection 27343 (kid)] at the late-log growth phase was obtained by the same procedure. Twenty micrograms of total tRNA from either M. capricolum or B. subtilis was treated with 4 U of Turbo DNase (Ambion), in the presence of 80 U of RNaseOUT (Invitrogen) for 30 min at 37°C. Following the suppliers’ protocols, the Turbo DNase was removed first, and then reverse transcription for first strand cDNA synthesis was performed, using 0.2 μg of total tRNA and 200 U of SuperScript III reverse transcriptase (Invitrogen). The primers for first strand cDNA synthesis of M. capricolum tRNA<sup>Arg</sup> and B. subtilis tRNA<sup>Arg</sup> were 5'-GGCCT-CA-CCCCAT-GTGA-3' (Mca-1st) and 5'-GGGAG-TGAA-CCTT-AAGG-GGGGAT-3' (Bsu-1st), respectively (black arrows in Figure 2A). In addition to the first strand cDNA synthesis primers, the following primers 5'-GCCCAGATCGAATGGATTGGAATCTT-3' (Mca-2nd) and 5'-GCCCG-TAGCT-CAATG-GATAGAGGCT-TGA-3' (Bsu-2nd) were used for further polymerase chain reaction (PCR) amplification of the cDNAs (gray arrows in Figure 2A).

Comparison of the 3D structure of Staphylococcus TadA and the putative 3D structure of TadA from M. capricolum

A homology model of TadA from M. capricolum was created, based on its amino acid sequence and the crystal structure of TadA in complex with RNA from Staphylococcus aureus (PDB code: 2B3J) (37), using the SwissModel automatic modeling server from Exasy (http://swissmodel.expasy.org/). The hydrogen bonded contacts between TadA and tRNA were calculated by the LIGPLOT programs (38). Structure representations were prepared with the Pymol program (Schrödinger, LLC).

RESULTS

Decoding arginine codons in Mollicutes

Table 1 lists the frequencies of codon usage for each of the six arginine codons (4X CGN and 2X AGR, Figure 1), together with the corresponding usage of tRNA<sup>Arg</sup>
Table 1. Comparative usage of Arg codons, number of tRNA<sup>Arg</sup> genes and occurrence of the *tadA* gene in 10 bacterial and 36 parasitic Mollicute genomes

<table>
<thead>
<tr>
<th>Number</th>
<th>Species</th>
<th>Group</th>
<th>Number of Arg codons in ORFs</th>
<th>Anticodon and number of tRNA genes</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td><em>E. coli</em> K-12 substr. MG1655</td>
<td>Outer</td>
<td>28485 29996 4871 7432 2845 1651</td>
<td>4</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>1b</td>
<td>Nitrosononas europaea ATCC 19718</td>
<td>Outer</td>
<td>13425 14553 4584 10153 5082 3473</td>
<td>1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>1c</td>
<td>A. aerogenes V55</td>
<td>Outer</td>
<td>727 601 268 367 9229 12588</td>
<td>1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>1d</td>
<td>Streptomyces avermitilis MA 4680</td>
<td>Outer</td>
<td>19076 93823 7656 74019 2208 9827</td>
<td>1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>1e</td>
<td>Synechococcus elongatus PCC 6301</td>
<td>Outer</td>
<td>8173 24198 8010 787 24448 1135</td>
<td>1</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>1f</td>
<td>S. aureus subsp. aureus Mu50</td>
<td>Outer</td>
<td>10775 2603 3956 388 9321 1202</td>
<td>2</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>1g</td>
<td>Bacillus cereus ATCC 14579</td>
<td>Outer</td>
<td>20003 6523 7745 1911 13891 3604</td>
<td>4</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>1h</td>
<td>B. subtilis subsp. subtilis str. 168</td>
<td>Outer</td>
<td>9150 10889 4957 7839 13191 4704</td>
<td>4</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>1i</td>
<td>Listeria monocytogenes EGD e</td>
<td>Outer</td>
<td>10836 6301 5099 2578 5899 1102</td>
<td>2</td>
<td>1 1 1 1 1</td>
</tr>
<tr>
<td>1j</td>
<td>Oenococcus oeni PSU1</td>
<td>Outer</td>
<td>5934 2698 2965 2152 3951 1353</td>
<td>1</td>
<td>1 1 1 1 1</td>
</tr>
</tbody>
</table>

2. *A. laidlawii* PG-8A | IV | 4075 872 747 61 8639 670 | 1 1 1 |
3. *Aster yellows witches'-broom phytoplasma* AYWB | IV | 1183 710 305 42 2109 222 | 1 1 1 |
4. *Candidatus Phytoplasma australiense* | IV | 1332 804 505 77 2484 358 | 1 1 1 |
5. *Candidatus Phytoplasma multi* | IV | 1047 122 350 33 1972 168 | 1 1 1 |
6. *Onion yellow phytoplasma* OY-M | IV | 1455 843 379 58 2457 237 | 1 1 1 |

7. *Mesoplasma florum* L1 | I | 996 66 127 2 5444 190 | 1 1 1 |
8. *M. capricolum subsp. capricolum* ATCC 27343 | I | 904 100 153 6 6115 184 | 1 1 1 |
9. *Mycoplasma leachii* PG50 | I | 931 107 147 5 6154 175 | 1 1 1 |
10. *M. mycoides subsp. mycoides* SC str. PG1 | I | 1061 95 167 10 7324 272 | 1 1 1 |
11. *M. mycoides subsp. capri* LC str. 95010 | I | 1048 107 157 9 7275 252 | 1 1 1 |

12. *Mycoplasma agalactiae* | III | 1349 258 153 56 6250 711 | 1 1 1 |
13. *M. agalactiae* PG2 | III | 1186 255 163 57 5296 653 | 1 1 1 |
14. *Mycoplasma arthritidis* 15L8-1 | III | 1975 806 633 262 3233 327 | 1 1 1 |
15. *Mycoplasma bovis* PG45 | III | 1256 289 191 65 6129 758 | 1 1 1 |
16. *Mycoplasma conjunctivae* HRC/581 | III | 1786 718 728 175 3500 365 | 1 1 1 |
17. *M. crocodyli* MPI45 | III | 910 94 122 23 5564 337 | 1 1 1 1 |
18. *Mycoplasma hominis* ATCC 23114 | III | 899 181 136 46 3818 413 | 1 1 1 1 |
19. *Mycoplasma hypopneumoniae* 232 | III | 1485 938 1211 721 2858 745 | 1 1 1 |
20. *M. hypopneumoniae* 7448 | III | 1463 938 1210 672 2852 719 | 1 1 1 |
21. *M. hypopneumoniae* J | III | 1460 933 1196 665 2881 710 | 1 1 1 |
22. *Mycoplasma hyorhinis* HUB-1 | III | 913 125 340 41 4881 297 | 1 1 1 |
23. *M. mobile* 163K | III | 618 77 171 26 5441 411 | 1 1 1 |
24. *Mycoplasma synoviae* 53 | III | 986 136 96 60 4811 284 | 1 1 1 |
25. *M. fermentans* JER | III | 2030 258 314 63 5371 236 | 1 1 1 1 |
26. *M. fermentans* M64 | III | 2164 303 335 81 6439 315 | 1 1 1 1 |
27. *Mycoplasma penetrans* HF-2 | II | 467 15 52 26 8579 492 | 1 1 1 |
28. *Ureaplasma parvum* serovar 3 | II | 3098 447 946 122 1571 122 | 1 1 1 |
29. *Ureaplasma parvum* serovar 3 | II | 3087 450 946 122 1592 127 | 1 1 1 |
30. *Ureaplasma urealyticum* serovar 10 | II | 3671 369 1044 90 1652 77 | 1 1 1 |

The frequencies of arginine codons in protein-encoding ORFs in each genome were obtained from Genbank. The information about the presence or absence of a given tRNA<sup>Arg</sup> gene (the number corresponds to the number of genes encoding a tRNA with a given anticodon), as well as that about the *tadA* gene (always one when present), was obtained from the NCBI genome database, using BLASTN and BLASTP searches, respectively. The third base of the codon and the first wobble base of the anticodon are underlined. The accession numbers of the species, the subfamilies to which they belong and their hosts (in the cases of parasitic Mollicutes), their genome sizes, G+C% and references are provided in Supplementary Table S1. Species 2–6 correspond to Mollicutes of Group IV (Phytoplasmas), species 7–11 correspond to Mollicutes of Group I (Spiroplasmas), species 12–26 and 34 correspond to Mollicutes of Group III (Hominis) and finally species 27–33 and 35–37 correspond to Mollicutes of Group II (Pneumoniae). Descriptions of the different classes of Mollicutes are available (29–31). The CGG codon usages of Mollicute Groups IV (Phytoplasmas) and I (Spiroplasmas) are highlighted in bold letters.
isoacceptors, classified according to their anticodons (NGG and YCU) in 36 Mollicutes. This range of Mollicutes, all with reduced genome sizes (Supplementary Table S1), thoroughly covers the four major clades of the monophylegetic phylum of this group of bacteria, i.e. Group I for Spiroplasma (items 7–11), Group II for Pneumoniae (items 27–37, except for 34 belonging to Group III), Group III for Hominis (items 34+12–26) and Group IV for Phytoplasma and Acholeplasma (items 2–6). For comparison, the situations in a few selected bacterial genomes outside the Mollicute family (items 1a–1j) are also shown. The table includes information about the presence or absence of a gene encoding a homolog of *B. subtilis* TadA (accession No. NP_387899.1), as query sequence. The E-values of the candidate protein sequences in the BLASTP search are >1e-13 (10^{-13}). No other Mollicute proteins showed E-values >1e-09 (10^{-9}).

Inspection of Table 1 leads to the following conclusions:

(i) In contrast to most bacteria, no gene encoding a tRNAArg harboring the same anticodon is redundant. This trend fits with the gene economization strategy used by Mollicutes, with their small genome sizes. The only exception is for tRNAArgCGG in *Mycoplasma gallisepticum*, which is encoded by two genes differing by only a single base at position 25 in the D-stem (C25 or A25), thus creating a mismatch G10-A25 in one of the two tRNAs (Supplementary Figure S1, and indicated in the Group II- Pneumoniae of Supplementary Figure S2). *Mycoplasma crocodyli* also has two genes encoding tRNAArgCGG in its genome; however, these have exactly the same sequence (Supplementary Figure S1).

(ii) In contrast to most bacteria, none of the Mollicutes examined carries a gene encoding C34-containing tRNAArgCGG (row 13 in Table 1). This gene was obviously already lost in the genome of the common ancestor of Mollicutes. The lack of this gene is correlated with a drastic reduction, but not the complete elimination, of the CGG codons in mRNA (row 7 in Table 1), which are normally read by the missing tRNAArgCGG, especially in Spiroplasma (Group I, items 7–11) and Phytoplasma (Group IV, items 2–6, indicated in bold in Table 1). An analysis of the ORFs containing the few remaining Arg-CGG codons revealed that they are often used in genes encoding DNA and RNA modification enzymes, with only one codon in each gene, such as in Dam and DNA methylases, TruA, TruB, ThiI (indicated in bold in Supplementary Table S2) and even the tRNA-A34 deaminase TadA (indicated in bold in the same Supplementary Table S2). The presence of a problematic Arg-CGG codon at the beginning (second position) of the mRNA corresponding to the *tadA* gene of *Mycoplasma mycoides* (Spiroplasma) is notable, and it suggests that the level of TadA deaminase expression in this organism may depend on the ability of the remaining single tRNAArg of the Arg-CGN decoding box to read this rare CGG codon.

(iii) All Mollicutes belonging to Groups III (Hominis, items 12–26 and 34) and II (Pneumoniae, items 27–33 and 35–37) lack the *tadA* gene, whereas in all Mollicutes of Groups IV (Phytoplasma, items 2–6)
and I (Spiroplasma, items 7–11), the tadA gene is still present. The corollary is that A34, in the remaining single tRNAArgACG of the quartet decoding box, should normally be matured into I34 in all Groups I and IV Molllicutes, whereas in Groups II and III, the encoded wobble A34 will remain unmodified. Thus, the absence of the tRNA deaminase TadA in the Groups II and III Molllicutes obviously does not affect the viability of these cells, which have also adopted the strategy of preferring the arginine codon usage to mostly AGA of the duet decoding box (Table 1, compare the frequencies of codon usage in row 8 in Molllicutes—items 12–37, with those for bacteria—items 1a–1j). Groups I and IV of the Molllicutes (items 2–11) pose a more difficult problem because the cells have to read the four CGN codons with only a single I34-containing tRNAArgICG, which is normally unable to read CGG. Here, the dramatic reduction in CGG codon usage (indicated in bold in Table 1) and the preference for using the codon AGA of the duet decoding box instead is evident, especially in Spiroplasma (Group I, items 7–11). This AGA arginine codon will be read by the modified U*34-containing tRNAArgUCGU belonging to the duet decoding arginine box (see later in the text).

(iv) All Molllicutes of Group II (Pneumoniae), and Mycoplasma fermentans plus Mycoplasma pulmonis belonging to Group III-Hominis, have an additional tRNAArgACG harboring the anticodon UCG (row 12 in Table 1, items 25–37), thus alleviating the difficulty of reading both codons ending with A and G by A34- or I34-containing tRNAArg. Moreover, in most Pneumoniae with M. pulmonis (items 31–36), the A34-containing tRNAArgACG is replaced by the G34-containing tRNAArgICG. Together with the U34-containing tRNAArgUCGU, this allows all four CGN arginine codons to be easily read in contrast to the Hominis clade (items 12–24), with only a single A34-containing tRNAArgACG. Only Mycoplasma haemofelis (Pneumoniae, item 37) remains with a single tRNAArg harboring the UCG anticodon, with the wobble U34 probably kept unmodified to enable the reading of all four CGN codons by ‘superwobbling’ (four-way wobbling) (22,39,40).

(v) The only tRNAArg present in all Molllicutes analyzed is tRNAArgUCGU of the duet decoding Arg-box (Figure 1 and Table 1), where U* stands for 5-carboxymethylaminomethyluridine (cmmUT), as demonstrated in M. capricolum tRNAArgUCGU (41). The modification of U34 in this tRNAArgUCGU is catalyzed by the multi-protein complex MnmE/MnmG present in almost all bacteria, including Molllicutes (42,43). Together with a second C34-containing tRNAArgCCU of the same duet decoding arginine box (only present in a few Molllicutes, Table 1), they translate the frequently used Arg codons AGA and AGG (AGR). From an evolutionary point of view, the existence of a second decoding box for arginine probably greatly facilitated the progressive shift in the decoding strategy within the other arginine decoding box.

In M. capricolum, the wobble A34 of a small fraction of tRNAArgACG is not deaminated

The nucleotide sequence of the naturally occurring tRNAArgICG of M. capricolum has been sequenced (41). However, no information was provided about the possibility that a small fraction of this tRNA population was not completely matured, especially at the wobble A34 position (Figure 2A). To clarify this point, we sequenced the anticodon region of cDNAArgICG, obtained after reverse transcription of the naturally occurring tRNAArgICG present in the bulk tRNA of M. capricolum (Figure 2A). As inosine behaves like G during transcription, we expected to obtain a G at the corresponding position in the cDNAArg. In contrast, if a fraction of the wobble A34 in the tRNA sample is not modified into I34, then some cDNAArgG clones will now carry A at position 34, and the proportion of ‘A’-clones over ‘G’-clones will provide information about the degree of A34 to I34 modification in the original M. capricolum tRNA sample. As shown in Figure 2B (upper part), among 86 cDNA clones analyzed, 5 clones (6%) have A at the anticodon first position, and the remaining 81 cDNA clones have G (94%). To confirm this result, several control experiments were performed. First, when the reverse-transcribed tRNA solution was used as the PCR template, only the cDNAs of M. capricolum tRNAArgICG were amplified (Figure 2C, lanes 1 and 3). Second, in the absence of reverse transcriptase, no cDNA products were PCR amplified (Figure 2C, lanes 2 and 4), confirming the absence of DNA contamination (even without DNase treatment). The results shown in Figure 2B were obtained using the cDNA shown in lane 1 of Figure 2C. The second series of control experiments involved performing the same analysis with bulk tRNA obtained from B. subtilis (Figure 2B and C). The tRNAArgICG sequence in this bacterium is similar to its M. capricolum homolog (Figure 2A) (15). The results from the analysis of 82 clones obtained from the cDNA (lane 6 in Figure 2C) indicated that, in contrast to the bulk tRNA from M. capricolum, no clone contained a cDNAArg with an A at the anticodon position 34, and only G34 was detected (100% - Figure 2B), corresponding to the fully matured I34 in the original sample of B. subtilis tRNAArgICG. These experiments demonstrated that in naturally occurring M. capricolum cells, a minor fraction of tRNAArg with unmodified wobble A34 (anticodon AGC) does exist and probably functions in translating all Arg-CGN codons (21,22).

The enzymatic deamination of A34 in tRNAArgACG in Molllicutes is probably not as efficient as in other bacteria

A small fraction of non-deaminated tRNAArgACG may also exist in other Molllicutes with genomes encoding tadA. This possibility could result from insufficient tadA gene expression and/or an abnormally inefficient (degenerate) deaminase. To examine this latter possibility, we compared the amino acid sequences of 10 TadA proteins
Figure 3. Amino acid sequence alignment of the genes encoding TadA. The TadA amino acid sequences from the species listed in Table 1 were retrieved from Genbank and aligned by Clustal X (34), under the default conditions. The amino acid numbers from \( E. \ coli \) are indicated above the alignment. The amino acid numbers from other species are indicated at the beginning and the end of the sections. The TadA-specific conserved amino acids are highlighted with a red or orange background. The conserved amino acids common among TadA and CDA are highlighted with a black or gray background. The conserved deaminase catalytic and zinc-binding sequences are highlighted in blue or light blue. Structurally and functionally important residues of TadA, inferred from the tertiary structures of the \( A. \ aeolicus \) and \( S. \ aureus \) TadAs (37,44), are indicated above the alignment. The terms 'nnb' and 'stack' mean non-bonded (hydrophobic) contacts and stacking interactions, respectively. The red boxes in Mollicutes (sequences 2–11) indicate the variations from other bacterial TadAs (sequences 1a–1j). Conserved amino acids involved in tRNA interactions, which are depicted by stick models in Figure 4, are indicated by arrows below the sequences.
encoded in the genomes of various bacteria (sequences 1a–1j in Figure 3), with those of 10 TadA proteins of the Mollicutes of Groups I (Spiroplasma) and IV (Phytoplasma), all encoding the \( \text{tadA} \) gene (sequences 2–11 in Figure 3). The list includes the well-characterized TadA from \( \text{E. coli} \) (sequence 1a) (19,45), \( \text{Aquifex aeolicus} \) (sequence 1c) (44) and \( \text{S. aureus} \) (sequence 1f) (37). The amino acids with identical locations in the sequences are highlighted with black or colored backgrounds, and the systematic sequence deviations among these invariant or semi-invariant amino acids are boxed. The correspondence of these remarkable amino acids within the architecture of the TadA enzyme (indicated with black and colored backgrounds), and of the nucleotide position in tRNA (indicated in black), is depicted at the top of the figure. This information was deduced from the crystal structure of \( \text{S. aureus} \) TadA in complex with a chemically synthesized anticodon stem loop (16mer) bearing nebularine-34 as a substrate, in place of inosine-34 (Figure 4A) (37). For clarity, all other important elements of the anticodon branch in contact with the deaminase are not shown, as they are similar in the tRNA\(^{\text{Arg}}\)\(^{\text{ACG}}\) of both \( \text{S. aureus} \) and \( \text{M. capricolum} \) (15).

Among the important invariant amino acids to be considered in the A-to-I deaminase TadA, some are also common within the C-to-U deaminase CDA (35), including the AE motif of the deaminase catalytic center, and PCxxC of the zinc-binding motif (Figure 3). In addition, the TadA proteins from Mollicutes (sequences 2–11) share several other identity elements in common with some selected bacterial TadA proteins (red or orange background), i.e. the EVPV and TLE motifs of the TadA-structural core, and

**Figure 4.** (A) Homology model of \( \text{M. capricolum} \) TadA, superposed on \( \text{S. aureus} \) TadA complexed with tRNA\(^{\text{Arg}}\)\(^{\text{ACG}}\). Both TadA proteins are represented by ribbon models, colored green for \( \text{M. capricolum} \) and gray for \( \text{S. aureus} \). The \( \text{S. aureus} \) tRNA is depicted by a stick model. Conserved amino acids involved in tRNA interactions, which are indicated by arrows in Figure 3, are shown in stick models. The amino acids specific to Mycoplasma, indicated in the red boxes in Figure 3, are circled. (B) Sequences of the anticodon branches of the tRNA\(^{\text{Arg}}\)\(^{\text{ACG}}\) from \( \text{S. aureus} \) and \( \text{M. capricolum} \) (15).
several amino acids at conserved positions, such as His57, Lys111 and Phe149 (E. coli numbers), which is precisely the region in contact with the tRNA anticodon loop (37). More interesting are the systematic sequence deviations and the absence of certain amino acids (gaps, indicated by dashes) in the TadA sequences of Mollicutes (sequences 2–11, positions in red boxes), as compared with the TadA sequences of other bacteria.

To better visualize the implications of these different amino acids within the active site architecture of the deaminase, the sequence of TadA from *M. capricolum* (item 8 in Figure 3) was superposed on the 3D architecture of TadA from *S. aureus* (item 1f in Figure 3) in complex with a 16nt mini substrate. As shown in Figure 4A, it is now clear that Asn71 and Lys95 in *M. capricolum* (indicated in green and encircled in red) replaced Arg70 and Arg94 in *S. aureus* (indicated in blue). Therefore, the ribose phosphate backbone of nucleotides G37 and G36 in the anticodon loop, which H-bond with these amino acids in the case of the *S. aureus* TadA–RNA complex, may not be well fixed, or exist in a slightly different configuration in the case of the putative complex of the same RNA with *M. capricolum* TadA. Moreover, in the vicinity of the essential zinc motif and nebularine-34, and thus within the catalytic center of the deaminase, Ser105 (indicated in green and encircled in red) in *M. capricolum* replaces the important Asp104 in *S. aureus* (indicated in blue), which normally H-bonds with the ribose of U at position 33, adjacent to nucleoside 34 of the anticodon loop. The absence of an interaction with the ribose of U at position 33, adjacent to nucleoside 34 of the anticodon loop, and the absence of H-bonding because of the amino acid replacements at positions 70/71 and 94/95 discussed earlier in the text, may affect the dynamics (flexibility/adaptability) of the entire anticodon branch within the active site of the deaminase. Consequently, this may limit the accessibility of the amine target of the wobble A34 for deamination, which is catalyzed by the neighboring zinc atom (in the brown background) around His53/54.

A global inspection of the 3D architecture of *S. aureus* TadA in complex with its RNA mini substrate (37) revealed that the A31-U39 base pair at the beginning of the anticodon stem does not interact with any amino acids of the deaminase. Only the C37-A38 pair interacts with Lys106 and Asn123 (Supplementary Figure S3). However, Lys106 (Lys107 in *M. capricolum*) is conserved in all TadA proteins examined (Figure 3), whereas Asn123 (Asn122 in *M. capricolum*) is replaced with different amino acids among the various Mollicutes; therefore, it may not be important for the catalytic function of the deaminase. It is likely that only the mutations in the *tadA* gene corresponding to the catalytic core of the deaminase, as discussed earlier in the text, contribute to the modulation of the A34-deamination efficiency and ultimately play a role in decoding all four arginine CGN codons.

**DISCUSSION**

During protein synthesis, tRNAs bearing the complementary anticodons read mRNA codons. However, because different types of relaxed base pairing are allowed between the often modified ‘wobble’ base at position 34 of the anticodon and the last nucleotide of the codon, some tRNA species can read two, three or even four synonymous codons. Therefore, the number of isoacceptor tRNAs with distinct anticodons needed to read all synonymous codons of a given amino acid is usually lower than the number of codons specifying that particular amino acid in the genetic code. Various organisms apply different rules to adapt their tRNA sets, attesting to the existence of distinct cellular strategies for reading the almost universal genetic code (4). Here, we focused on reading the quartet arginine codons in the quickly evolving Mollicutes with reduced genomes (0.6–1.5 Mb, Supplementary Table S1).

**Reading arginine codons in *M. capricolum***

In *M. capricolum*, only two kinds of tRNA^{Arg} exist for reading the six arginine codons (four in the quartet and two in the duet family boxes). One tRNA contains an anticodon with a wobble inosine (tRNA^{Arg}_{CGG}) and the other contains an anticodon with a modified wobble uridine (cmnm^5U_{34}, tRNA^{Arg}_{U*UCCU}) (41). Because of the wobble inosine-34, tRNA^{Arg}_{ACG} was expected to read only the three arginine codons ending with U, C or A of the quartet family box (8,9,16). Paradoxically, a tRNA^{Arg} harboring the anticodon CCG, needed to read the remaining fourth arginine codon CGG, as found in the majority of other bacteria (Table 1, items 1a–1j), was absent (41). Here, we demonstrated that a small fraction of the cellular A_{34} containing tRNA^{Arg}_{ACG} precursor is not enzymatically deaminated in *M. capricolum*. The key point of our report is the correlation with a few characteristic amino acid variants that exist within the active sites of the TadA’s of *M. capricolum* and other Mollicutes, as compared with other well characterized bacterial TadA’s considered as references. We hypothesize that these point mutations are needed for reducing the enzymatic activity of the tRNA:A_{34} deamination (degenerate TadA*), allowing the accumulation of a small but sufficient amount of the non-deaminated A_{34}-containing tRNA^{Arg}_{ACG}, which is competent for reading all four arginine codons of the quartet CGN decoding box (Step 1 in Figure 5A). To use a term that was first applied in the case of unmodified U_{34}-containing tRNAs, this decoding strategy would correspond to a sort of ‘superwobbling’, facilitating the translation of synonymous codons with a reduced set of tRNAs (40). Therefore, the useless C_{34}-containing tRNA^{Arg}_{CCG} can be lost (Step 2 in Figure 5A). This process was probably facilitated by limiting the usage of the problematic CGG codon (Step 2). Indeed, among 1163 CGN codons, only 6 such rare CGG codons, each in different mRNAs, were detected in the ORFs of *M. capricolum*.

**Reading arginine codons in other Mollicutes (Spiroplasma and Acholeplasma/Phytoplasma)**

Combining our comparative genome analysis with information about the evolutionary origin of Mollicutes (29,30) revealed that the decoding strategy for *M. capricolum* is still in use in all Mollicutes of Groups I (Spiroplasmas, items 7–11) and IV (Acholeplasmas/Phytoplasmas, items
Figure 5. Hypothetical scenario for the evolution of the CGN decoding system for arginine in Mollicutes. (A) Schematic view of the five sequential events leading from a ‘classical bacterial’ arginine decoding strategy involving two tRNA^Arg, one with a wobble inosine-34 and the other with a wobble C34, to another Arg decoding strategy involving only one tRNA^Arg with an unmodified wobble U34. In M. capricolum, this latter situation exists in many other quartet decoding boxes (Leu, Val, Ser, Pro, Ala and Gly), as well as in most mitochondria of eukarya. (B) The same events as in A, but depicted within the Mollicute evolutionary framework. Because of the degenerated TadA*, partial A-to-I deamination occurs at the first anticodon position of tRNA^Arg^CG (Step 1), generating a situation where a mixture of both deaminated (in black) and non-deaminated tRNA^Arg^CG (in red) molecules co-exist in the cell. In addition to the three synonymous arginine codons normally decoded by tRNA^Arg^CG, tRNA^Arg^ACG also decodes the CGG codon, but probably inefficiently (see text). The gene encoding tRNA^Arg^CG could then be lost (Step 2), along with the gene encoding tad* (Step 3). Further reorganization of the tRNA repertoire could occur by gaining an extra U34-containing tRNA^Arg^UCG (Step 4). The original A34-containing tRNA^Arg^ACG can undergo a mutation in its anticodon to generate a G34-containing tRNA^Arg^CG (Step 5a), or simply be lost (Step 5b). The species of Mollicutes in which these different events occurred are indicated by numbers, corresponding to the organisms listed in Table 1. The phylogenetic relationships among the different Mollicutes were adapted from the literature (29–31).
2–6), as shown in Table 1 and the green background in Figure 5B. Obviously, the two events (Steps 1 and 2 described earlier in the text) occurred early in evolution, almost at the root of the monophyletic Mollicute tree. These Mollicutes currently have the same original set of two genes: one gene encoding an A$_{34}$-containing tRNA$^{\text{Arg}_{\text{AGG}}}_{\text{ACG}}$ for reading a minimum number of CGN codons, and a second one harboring a U$^*_{\text{CGG}}$ (tRNA$^{\text{Arg}_{\text{UUCU}}}_{\text{UGG}}$) for reading the other most frequently used arginine codons AGA and AGG; only the original TadA is now the mutant Tad*A.

**Further stepwise evolution of the decoding strategy in Hominis and Pneumoniae**

To become less dependent on the activity of the hypothetical degenerate TadA*, a subset of the newly evolved Mollicutes lost the degenerated tadA* gene (Step 3). This new evolutionary event occurred before the divergence into Groups III (Hominis) and II (Pneumoniae), items 12–37—all indicated with a yellow background in Figure 5A and B. Interestingly, the usage of the earlier problematic and rare CGG codon in these newly evolved Mollicutes became more frequent again, confirming that a Mollicute lacking the tadA gene and encoding an unmodified wobble A$_{34}$-containing tRNA$^{\text{Arg}_{\text{AGG}}}_{\text{ACG}}$ (items 12–24 in Figure 5) is perfectly viable because of its ability to read all four Arg-CGN codons.

In a subset (items 25–37) of Groups II and III (Hominis/ Pneumoniae), the reading of the four synonymous arginine CGN codons was probably improved by gaining a new U$_{34}$-containing tRNA$^{\text{Arg}_{\text{UGC}}}_{\text{UGG}}$ (Step 4 in Figure 5A and B). This U$_{34}$-containing tRNA$^{\text{Arg}_{\text{UGC}}}_{\text{UGG}}$ could have originated in diverse manners. It may have arisen from the duplication of the gene encoding A$_{34}$-containing tRNA$^{\text{Arg}_{\text{AGG}}}_{\text{ACG}}$, followed by a few mutations, including the wobble A$_{34}$-to-U$_{34}$. It may also have resulted from duplication and subsequent recruitment/mutation of a gene encoding a tRNA possibly from the other duet Arg-AGR coding box, or belonging to another amino acid coding box. The mutations in the tRNA$^{\text{Arg}_{\text{AGG}}}_{\text{ACG}}$ substrate itself may modulate the efficiency of A$_{34}$-deamination and ultimately play a role in decoding all four arginine CGN codons (Supplementary Figure S2). Unfortunately, a phylogenetic analysis of all of the tRNA genes retrieved from the 36 Mollicutes examined did not allow us to confidently determine which one of these two alternatives prevailed because of the low-bootstrap values in constructing such phylogenetic trees with relatively short tRNAs, including many conserved and semi-conserved nucleotides and invariant regions under strong selective pressure (46,47).

Among the few species (items 25–30) of Groups II and III (Hominis/Pneumoniae), the four arginine CGN codons are read by a tRNA$^{\text{Arg}}_{\text{ACG}}$ pair, one with a non-deaminated wobble A$_{34}$ and the other with a wobble U$_{34}$ (Figure 5, yellow background; U$_{34}$ is probably not modified, see later in the text). This decoding strategy is also the one used presently for reading the four CGN codons as arginine in a few other non-Mollicute bacteria, such as *Clostridium perfringens*, *Chlamydia trachomatis*, *Geobacter metallireducens* and *Haloplasma contractile*, the four CUN codons as leucine in *Lactobacillus lactis*, and as mentioned in the ‘Introduction’ section, for reading the four ACN codons as threonine in *M. capricolum* (11,41).

Other species of Group II-Pneumoniae (items #31-36, including *M. pulmonis*) continued to evolve by using a slightly different decoding strategy (Step 5a). In these species, the CGN codons are now read by another type of tRNA$^{\text{Arg}}_{\text{UGC}}$ set, one with a wobble U$_{34}$ and the other one with G$_{34}$ (Figure 5, yellow background). Because of the close sequence homology between the new G$_{34}$-containing tRNA$^{\text{Arg}}_{\text{UGC}}$ and the A$_{34}$-containing tRNA$^{\text{Arg}}_{\text{ACG}}$ in the other Pneumoniae (data not shown), this new G$_{34}$-containing tRNA$^{\text{Arg}}_{\text{UGC}}$ is believed to have arisen via a simple A$_{34}$-to-G$_{34}$ mutation and additional base mutations within the rest of the tRNA$^{\text{Arg}_{\text{UGC}}}_{\text{UGG}}$ structure. This last decoding strategy is most frequently used in bacteria for decoding the sense codons of quartet synonymous codon boxes, at least in bacteria with moderate or low G+C content in their ORFs, as in *Borrelia burgdorferi*, *Campylobacter jejuni*, *Helicobacter pylori*, *Treponema pallidum*, *Thermotoga maritima* and a few others (11).

Finally, one Mycoplasma in Group II, *M. haemofelis* (item 37 in Table 1), lost the ancient A$_{34}$-(or G$_{34}$)-containing tRNA$^{\text{Arg}_{\text{AGG}}}_{\text{ACG}}$ (Step 5b); thus, it has only one U$_{34}$-containing tRNA$^{\text{Arg}_{\text{UGC}}}_{\text{UGG}}$ for reading the four synonymous Arg-CGN codons. This situation corresponds to the minimal set of tRNA$^{\text{Arg}}_{\text{UGC}}$ that a Mollicute can use to continue decoding all CGN codons as arginine, with no need for the enzyme TadA and probably with better efficiency than that with a single A$_{34}$-containing tRNA$^{\text{Arg}_{\text{AGG}}}_{\text{ACG}}$. This decoding strategy was also used in other quartet decoding boxes corresponding to Leu, Val, Ser, Pro, Ala and Gly in *M. capricolum*, *M. mycoides* and the mitochondria of *S. cerevisiae* and mammals (24,39,41); reviewed in (5,48). The sequences of the corresponding tRNAs revealed the presence of an unmodified U$_{34}$ in their anticodons (15).

**Analogy to a similar situation in the chloroplasts of higher plants**

Gene knockout experiments in the plastids of the moss *Physcomitrella patens* demonstrated the dispensability of the C$_{34}$-containing tRNA$^{\text{Arg}_{\text{CCC}}}_{\text{CCG}}$, whereas the chloroplastic A$_{34}$-containing tRNA$^{\text{Arg}_{\text{AGG}}}_{\text{ACG}}$ and the chloroplastic TadA enzyme are encoded in the plastid and nuclear genomes, respectively (49). This situation corresponds to that of the Groups I (Spiroplasmas) and IV (Acholeplasmas/Phytoplasmas) Mollicutes (Table 1), which also lack C$_{34}$-tRNA$^{\text{Arg}_{\text{CCC}}}_{\text{CCG}}$ (see earlier in the text). On the other hand, the chloroplasts of *A. thaliana* lack C$_{34}$-tRNA$^{\text{Arg}_{\text{CCC}}}_{\text{CCG}}$, and only two kinds of tRNA$^{\text{Arg}}_{\text{ACG}}$ are encoded on the plastid genome: one with the anticodon ACG and the other one with the anticodon UCU. In this species, the inhibition of the chloroplastic tadA gene expression by RNAi (not the cytoplasmic Tad2/Tad3) allows plant survival, and only the chloroplast translation and photosynthesis activities were hindered (17,18). This situation corresponds to the one described earlier in the text for the Mollicutes of Hominis Group III. By analogy with our results in the case of *M. capricolum*,

Nucleic Acids Research, 2013, Vol. 41, No. 13 6541
we anticipate that in the chloroplasts of wild-type *A. thaliana*, and probably in other plant plastids, a fraction of the chloroplastic A$_{34}$-containing tRNA$_{\text{Arg}_{\text{ACG}}}$ also remains naturally unmodified, allowing super-wobbling for decoding all CGN codons, including the rare Arg-CGG (16).

**Evolutionary scenario of the Mollicute decoding process**

The scenario proposed in Figure 5 illustrates the evolvability of the decoding process. However, changing the decoding strategy during cellular evolution depends on a series of sequentially ordered events, such as point mutations in modification enzymes (probably also in the tRNA), gene loss, gene duplication and possibly the recruitment of a gene encoding a tRNA from another decoding box. The driving forces of this evolutionary process are almost certainly the efficacy and accuracy of translation. The authors have proposed, to explain the elimination of the essential deaminase TadA in Mollicutes, also an application to the rRNA–lysidine synthase TilS, responsible for the k$^2$C modification at the wobble position 34 of tRNA$^{\text{ile}_{\text{CAU}}}_{\text{ACU}}$. Indeed, although it is encoded in the genomes of 35 Mollicutes, the *tilS* gene is notably absent in *Mycoplasma mobile*, with a concomitant change in the sequence of the minor tRNA$^{\text{ile}}_{\text{ACU}}$ that decodes AU$^3$C codons, from a CAU to a UAU anticodon (50, 51). A similar cellular strategy has been experimentally verified in the case of *B. subtilis*, after the deletion of its essential *tilS* (52).

Finally, the idea of first reducing the activity of an enzyme (here, TadA or TilS) by point mutations, before its complete loss later in evolution, is reminiscent of recent work describing the progressive degeneration of aminoacyl-tRNA synthetases in *M. mobile* and other closely related Mycoplasmas of Group III-Hominis (53, 54). In these cases, the degenerated aminoacyl-tRNA synthetases, while still performing the normal aminoacylation function, occasionally misacylate the cognate tRNA with a non-cognate amino acid. This allows the generation of a small number of cellular proteins with an incorrect amino acid substitution (statistical mutations). It was proposed that such misacylation reactions, if they are not too frequent, would provide an advantage to the Mycoplasma, which are indeed evolving faster than other extant bacteria by producing a more homogeneous proteome (55).

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online: Supplementary Tables 1–2 and Supplementary Figures 1–3.

**ACKNOWLEDGEMENTS**

The authors thank the Eminent Professors Syozo Osawa (Nagoya University) and Kimitsuna Watanabe (University of Tokyo) for valuable discussions about evolving genetic codes. The authors thank Dr Kazuyuki Takai (Ehime University) for discussions about the A$_{34}$ decoding system. The authors also thank Pascal Sirand-Pugnet (INRA, Bordeaux, France) for advice on the phylogeny of Mollicutes, and Juan Alfonzo (Ohio State University, Columbus, USA) and Valérie de Crécy-Lagard (University of Florida, Gainesville, USA) for advice and manuscript editing. The authors thank Dr Chisato Ushida and Dr Akira Wada for bacterial sample preparation.

**FUNDING**


Conflict of interest statement. None declared.

**REFERENCES**


