



Published in final edited form as:
. 2012 ; 7: 11–34.

Decoding in *Candidatus Riesia pediculicola*, close to a minimal tRNA modification set?

Valérie de Crécy-Lagard^{1,*}, Christian Marck², and Henri Grosjean³

¹Department of Microbiology and Cell Science, University of Florida, P.O. Box 110700, Gainesville, FL 32611-0700, USA

²Institut de Biologie et de Technologies de Saclay (iBiTec-S) Bât 144, CEA/Saclay, F-91191 Gif-sur-Yvette Cedex

³Centre de Génétique Moléculaire, UPR 3404, CNRS, Associée à l'Université Paris-Sud 11, FRC 3115, 91190 Gif-sur-Yvette, France

Abstract

A comparative genomic analysis of the recently sequenced human body louse unicellular endosymbiont *Candidatus Riesia pediculicola* with a reduced genome (582 Kb), revealed that it is the only known organism that might have lost all post-transcriptional base and ribose modifications of the tRNA body, retaining only modifications of the anticodon-stem-loop essential for mRNA decoding. Such a minimal tRNA modification set was not observed in other insect symbionts or in parasitic unicellular bacteria, such as *Mycoplasma genitalium* (580 Kb), that have also evolved by considerably reducing their genomes. This could be an example of a minimal tRNA modification set required for life, a question that has been at the center of the field for many years, especially for understanding the emergence and evolution of the genetic code.

Keywords

tRNA; maturation; translation; modified nucleosides; comparative genomics

INTRODUCTION

As adapters between the mRNA and the elongating peptide, tRNAs are the central decoding molecules in translation. Their overall efficiency in protein synthesis depends both on the sequence/structure of the whole set of the tRNA repertoire and on modified nucleotides that are formed during the tRNA maturation process. Depending on the organism considered, a single functional tRNA isoacceptor may contain from 2 to 17 modified nucleosides [2]. These post-transcriptional modifications are required to maintain tRNA structure, insure correct mRNA decoding, optimize translation accuracy and efficiency, and/or regulate tRNA turn-over or its cellular localization (reviewed in: [3] and [4]).

*Corresponding author: vcrecy@ufl.edu.

Note added in proofs: It was recently found that the *E. coli rlmN* gene encodes the missing m²A37 methyltransferase (Eugenia Armengod, personal communication). RlmN homologs are present in most insect symbiont genomes, including *C. R. pediculicola*. A37 is therefore most certainly methylated into m²A37 in a few *C. R. pediculicola* tRNAs, which fits with our general conclusion above.

ABBREVIATIONS

Full names for the different acronyms used to define a given modified base can be found in [1].

Several studies have attempted to define a minimal, possibly ancestral tRNA modification set. By comparing the modification profiles in all available sequenced tRNAs from different kingdoms (Bacteria, Eucarya and Archaea, a total of about 500 tRNA in 1998), it was predicted that eight, possibly nine, modifications were present in the putative last universal common ancestor (LUCA or Cenancestor) [5, 6]. These modifications are the Ψ residues at positions 13, 38, 39, 55, C_m at position 34 [5] or C_m at position 32 [6], Q at position 34, t⁶A and m¹G adjacent to the anticodon at position 37, and m¹A at position 58 in the highly conserved sequence of the so-called TΨ-loop. Another study that combined comparative genomics and essentiality data predicted that LUCA harbored only three modifications [7]: Q34, Ψ13, and Ψ39. Finally Church and colleagues proposed that six modifications (k²C34, xs²U34 derivatives, I34, m¹G37, t⁶A37, ms²i⁶A37) are required for the minimal bacterial translation set [8]. The discrepancies are due to inherent flaws in all the prediction methods used. Predictions based solely on gene essentiality can be misleading, as a dispensable tRNA modification can become essential if other modifications are missing [9, 10]. Moreover, ancient-primordial genes may have considerably diverged in different phyla of organisms, so that they are now unrecognizable by any sequence-relatedness algorithms [11]. Alternatively, distinct enzyme families can introduce the exact same modification (functional type of enzyme evolution [12, 13, 14]). These will also be missed with methods based on ortholog searches. Finally, several genes of unknown function predicted to be present in LUCA [7] have since turned out to be involved in tRNA modification [15–17] and had therefore been missed in previous searches.

The idea of defining an “absolute minimal” set when talking about tRNA modifications might be inherently flawed and probably elusive. First, parallel and convergent solutions are deployed by different organisms both for modifications involved in decoding (discussed below) and in maintaining tRNA structural integrity. For example, ribothymidine, (m⁵U54) that is critical for tRNA stability in bacteria [18, 19], is replaced by m¹Ψ or Um in many Archaea [2, 20, 21]. Likewise, different modified uridines are used at the wobble position of tRNA to fulfill decoding requirements in different organisms [22]. Second, one cannot separate nucleoside modifications from the sequence context of a given tRNA repertoire as there is a clear co-evolution between the two sets. For example, the *tis* gene responsible for the k²C (lysidine) modification at the wobble position 34 was lost in *Mycoplasma mobile*. This loss occurred with a concomitant change of the sequence of the minor tRNA^{Ile} that decodes AUA codons from a CAU to a UAU anticodon [23, 24], a cellular strategy that has been experimentally verified in *B. subtilis* [25]. Third, the G+C content at the third codon position conditions the use of modified bases at the wobble position of tRNA [26]. Lastly, the requirements for modifications are going to be extremely dependent on environmental and physiological factors and will hence vary from one organism to another, for example, halophiles are predicted to require less modifications (see discussion of [24]). It is therefore not a minimal tRNA modification set but a minimal set of organism specific functional constraints that needs to be defined. An efficient and biologically relevant method to tentatively identify these minimal sets of essential tRNA modification enzymes, possibly the most reluctant to be lost during cellular evolution, is to analyze organisms with reduced genomes, such as parasitic or symbiotic intracellular and extracellular bacteria.

tRNA modification sets in Mollicutes

Mollicutes are parasitic, small unicellular bacteria normally living within eukaryotic cells. They originated from gram-positive bacteria (phylum: Firmicutes) by considerably reducing their genomes [27]. The Mollicute with the smallest genome identified so far is *Mycoplasma genitalium* (580 kb encoding 480 predicted ORFs) [28]. When cultivated in extremely rich medium, several of these Mollicutes can grow as free-living organisms, albeit very poorly and thus are considered to have minimal genomes [29]. In agreement with gene

economization strategies, all Mollicutes display a minimalist, non-redundant set of tRNAs (from 28 to 35 with distinct anticodons), that is sufficient to decode all sense codons corresponding to 20 canonical amino acids [24]. In this same study, we analyzed the presence or absence of genes coding for corresponding enzymes and predicted the tRNA modifications sets in 15 Mollicutes covering the four major clades (*Spiroplasma*, *Pneumonia*, *Hominis* and *Phytoplasma*). The genes were identified by homology with model systems such as *Escherichia coli* and *Bacillus subtilis*, and further validated from the knowledge of the modified nucleosides in the full set of 29 sequenced tRNAs of *Mycoplasma capricolum* [24]. The main conclusion was that only a few modification enzymes, all acting on nucleotides of the anticodon loop in tRNA (m¹G37, t⁶A37 and cmnm⁵U34), seemed resistant to gene loss. However, all the Mollicutes analyzed retained additional genes coding for enzymes inserting modifications in the tRNA body. For example, TruB catalyzing the Ψ55 insertion and TrmB catalyzing the methylation of G47 (m⁷G47) are found in the majority of Mollicutes, and therefore resistant to loss. Inspection of 20 additional complete genome sequences of Mollicutes, made available since this study, does not fundamentally change the initial conclusion (S. Yokobori, H. Grosjean and S. Bessho, personal communication).

tRNA repertoires in insect bacterial symbionts

In the present work, we performed a similar computational analysis of 14 genomes of bacterial symbionts and endosymbionts of insects, covering *Wolbachia* (3 strains which infect arthropod species and some nematodes), *Buchnera* (6 strains, which infect aphids), Candidatus *Blochmannia* (2 strains, which infect bacteriocytes and ant ovaries), *Baumannia cicadellinicola* (infecting bacteriomes of sharpshooter leafhoppers), *Wigglesworthia glossinidia* (infecting the gut of the tsetse fly) and Candidatus *Riesia pediculicola* (infecting human body louse). All of these species are derived from gram-negative Proteobacteria, mainly gamma-proteobacteria and related to *E. coli*, with the exception of *Wolbachia* (an alpha-proteobacteria). Unlike most bacteria and Mollicutes, members of this group cannot live as free-living organisms and form an obligate relationship (intimate symbioses) with their eucaryal hosts. These symbionts are predominantly vertically transmitted along with their host, and thus extend the heritable genetic variation of the host cells [30–33].

The genome sizes of the set of organisms analyzed (Supplemental Table 1) varied from 416 kb with 371 predicted CoDing Sequences (CDSs) (*Buchnera aphidicola* str. CC) to 1,483 kb with 1586 CDSs (*Wolbachia pipientis quinquefasciatus* Mel) (numbers of CDS taken from the Rast server [34]). 557 CDSs have been predicted in *C. R. pediculicola*, but around 80 of these are very small (between 19 and 70 aa) with no homology to any known proteins. These types of small proteins are not found in the other insect symbiont genomes analyzed and might be overpredictions.

Figure 1 (right part) shows that all 14 symbionts analyzed harbor genes coding for a full set of tRNAs able to read all sense codons for the 20 canonical amino acids, indicating that no tRNAs from the host are needed. Like Mollicutes and at variance with bacteria with large genomes, these uncultivable symbionts display a quasi-non-redundant set of tRNAs, with each isoacceptor having a distinct anticodon (compare columns #1 through #14 with column #15 for *E. coli*). The total number of tRNAs varies from 31 for *Buchnera aphidicola* str Cc (#8) to 40 for *C. Blochmannia pennsylvanicus* (#10). These correspond to tRNA repertoires typically found in Bacteria and not in Eucarya and Archaea [22, 35]. For example, tRNA genes containing the wobble T34 and G34 are almost always present, while tRNA genes containing C34 are often absent (blue background in Figure 1). In both of the quartet boxes corresponding to Pro and Ala (boxed in red in Figure 1), only one tRNA gene harboring a wobble T34 is present. For the isoleucine triplet decoding box and the arginine quartet decoding box, the T34-containing tRNA genes are systematically replaced by a C34-

containing tRNA^{Ile} and an A34-containing tRNA^{Arg}, respectively (indicated with yellow and green background in Figure 1). tRNA usage is usually correlated with codon usage, which in turn controls the efficiency of decoding [36]. By comparing the relative codon usage in each of the decoding box, it appears that G34-containing tRNAs more frequently read codons ending with the wobble U3 while U34-containing tRNAs mainly read codons ending with A3 (Watson-Crick base pairing). When the C34-containing tRNA isoacceptor is absent, U34 also reads codons ending with the wobble G3 (compare information about codon usage on the left part of Figure 1 with the identity of the wobble base in the tRNA, under the column symbol 'AC' for anticodon). This trend reflects the low average G+C content in ORFs of insect symbionts analyzed (from 23 to 35% compared to 52% in *E. coli*; Supplemental Table 1), particularly at the third position of codons (data not shown), and reflects the type of modified nucleotide present at the wobble position of tRNA. Non-redundancy of tRNA isoacceptor may affect cellular tRNA abundance, and hence the growth rate of the symbiont [37]. Finally, in contrast to *Mycoplasma* [24], UGA is a genuine stop codon in these insect symbiotic organisms, correlating with the presence of Release Factor 1 and Release Factor 2 (see the "tRNA modification *E. coli*" subsystem available on the Public SEED, <http://pubseed.theseed.org/SubsysEditor.cgi>).

tRNA modifications sets in insect bacterial symbionts

Genes coding for tRNA modification enzymes in the 14 genomes analyzed were identified by BLAST analysis against the genes found in *E. coli* (see [1] and Figure 2 legend). In *E. coli*, all but four of the fully matured isoacceptor tRNAs have been sequenced, and genes coding for most of the corresponding tRNA modification enzymes have been identified. Surprisingly, the recently sequenced human louse endosymbiont *C. R. pediculicola* [38] appears to have lost all modifications of the tRNA body, retaining only a few modifications of the anticodon loop and proximal stem (Figure 2): Ψ at position 38 and 39; I, k²C, xs²U derivatives and xo⁵U at position 34; and m¹G, t⁶A, i⁶A and ms²i⁶A at position 37. As it is technically impossible to extract enough tRNA from the human louse symbiont to analyze the modifications profiles, we cannot rule out that additional or unknown modifications are present in this organism. For example the acp³U47 modification gene has not yet been identified in any organism, and could be present in *C. R. pediculicola* (Figure 2).

An identical analysis was performed on the remaining 13 symbionts (#1 to #13, Figure 3). In some genomes, additional modifications were predicted to be present: s⁴U8, s⁴U9, D17, 20, 20a, Q34, m⁷G46 and Ψ55. However, all symbionts analyzed except *C. R. pediculicola* contain at least one modification outside the anticodon-stem loop (Figure 3).

Decoding strategy of synonymous codons in Candidatus *R. pediculicola*

Analysis of the sequences of both of the louse endosymbiont *C. R. pediculicola* and its host reveals that no eukaryotic genes, including putative tRNA modification enzymes, have been transferred to the insect bacterial genome, and that the genome reduction in *C. R. pediculicola* has not been associated with gene transfer to the host [38]. In the 14 proteobacterial symbionts analyzed, we are confident that the only genes coding for tRNAs and tRNA modification enzymes are those reported in Figures 1, 2 and 3 (except, see Figure 2 legend, for enzymes catalyzing acp³U47 and m²A37, for which the corresponding genes in *E. coli* are yet to be identified).

Beside the lack of some tRNA isoacceptors in the insect symbionts (discussed above and Figure 1), both nucleotide identities and post-transcriptional modifications are very similar when comparing tRNA isoacceptors from *E. coli* and *C. R. pediculicola*, attesting closely related and typical bacterial decoding strategies [22]. The differences between the two organisms are indicated in red in Figure 4. The main difference is the additional

modification of several *E. coli* bases. These are 2' *O*-methylation of C32 (Cm) in tRNA^{Ser} (*UGA) and tRNA^{Trp} (CCA) or U32 (Um) in tRNAs specific for Pro, His and Gln, as well as C34 (Cm) or U34 (Um) in *E. coli* tRNA^{Leu} (*UmAA) and tRNA^{Leu} (CmAA). Many complex modifications are predicted to be missing in *C. R. pediculicola* tRNAs: the acetyl group in elongator tRNA^{Met} (ac⁴C34), the *N*6-methylations of A37 or t⁶A37 in tRNA^{Thr} (GGU/UGU), Q34 or GluQ34 in tRNA specific for Tyr, His, Asn and Asp, and the thio-group on C32 in several tRNA for Arg and Ser. The mnm⁵U34 modification found in some *E. coli* Gln, Lys, Glu, Arg and Gly tRNAs, is predicted to be cmnm⁵U34 in the corresponding *C. R. pediculicola* tRNAs, because of the presence of MnmE and MnmG and the absence of MnmC. This last enzyme normally catalyzes the stepwise decarboxylation of the 'cmnm' group attached to C5 of U34, followed by methylation of the resulting 'nm' group into the final product mnm⁵U34 [39]. An alternative nmnm⁵U34 biosynthetic pathway using ammonium instead of glycine as a cofactor has been demonstrated in *E. coli* MnmE/MnmG mutants *in vitro* [40]. Therefore, it is possible that such an alternative ammonium mediated biosynthetic pathway leading to the final nm⁵U34 derivative is used in the insect symbiont (thus by-passing the formation of cmnm⁵U34). The cmo⁵U34 modification is found in *E. coli* tRNAs belonging to quartet decoding boxes for Leu (anticodon *UAG), Val (*UAC), Ser (*UGA), Pro (*UGG), Thr (*UGU) and Ala (*UGC). Its synthesis requires at least three enzymes, only two of which are known (CmoA-*yecO* and CmoB-*yecP* [41, 42]). The methylester of cmo⁵U at position 34 (mcmo⁵U not mentioned in any of the cases in Figure 4) is reported to be base labile and thus only cmo⁵U is usually detected during most analyses of modified nucleosides. In *E. coli*, tRNA^{Ser} and tRNA^{Ala}, but not tRNA^{Val} were reported to be substrates for the *E. coli* and *Salmonella* CmoA methyltransferase [3]. Remarkably, genes coding for CmoA-CmoB are found in *C. R. pediculicola* but absent in all the other 13 insect symbionts analyzed (Figure 3) as well as in *Mycoplasma* [24, 43]. This suggests that the cmo⁵U34 modification is dispensable, and *cmoA/cmoB* could be the next set of modification gene lost by *C. R. pediculicola*. Alternatively, the maintenance of cmo⁵U in several *C. R. pediculicola* tRNAs could result from subtle decoding constraints specific to that organism. Several studies exploring the function of (m)cmo⁵U derivatives versus ho⁵U modified U34 [3, 41, 42] concluded that the (m)cmo group added to the C5 atom of the wobble U base enhances the ability of the tRNA to pair with all four codons, a property that was also demonstrated for a non-modified wobble U-base [44]. These observations again suggest that U34 modification of tRNA belonging to quartet decoding boxes can be dispensable, however only in certain extended anticodon contexts [45].

The situation is different in the cases of bacterial tRNAs belonging to the split/duet decoding boxes, such as tRNAs specific for Phe/Leu, His/Gln, Asn/Lys, Asp/Gly and Ser/Arg that depend strictly on the identity of modified nucleotides at wobble U base. *E. coli*, all of the insect symbionts analyzed in this work, and all of the Mollicutes analyzed earlier rely on xm⁵U and xm⁵s²U derivatives to allow accurate and efficient discrimination of the duet codons ending with a pyrimidine U or C. The other duet codons of the same decoding box ending with a purine A or G being efficiently read by a G34 or modified G34-containing tRNA (reviewed in: [22, 46]).

Other important, modified nucleotides conserved in *C. R. pediculicola*, and possibly essential in all bacteria, are the pseudouridines at positions 38, 39 and/or 40 and the modified purine at position 37 found in tRNAs harboring an anticodon ending with A36, G36 or U36, modified into (ms²)i⁶A37, m¹G37 or m²A37 and (m⁶)t⁶A37, respectively (Figure 4). Removal of these modifications has been shown to have a detrimental effect on efficiency and accuracy of decoding (reviewed in [47, 48]). However, one cannot generalize the essentiality of these modifications to all tRNA sets, as the naturally occurring *E. coli* tRNA^{Ser}(GGA) and/or tRNAs harboring an anticodon ending with A37 in most *Mycoplasmas* lack i⁶A37 or ms²-i⁶A37 derivatives [24, 49]. It is clear from our analysis that

the genes responsible for the insertion of Ψ 38-40, m^1G37 , t^6A37 , $cmnm^5U34$ and s^2U34 remain resistant to loss. This suggests that these genes emerged early during cellular evolution, and, once fixed in the genome, became essential for the cell.

DISCUSSION

Both the insect symbionts and Mollicutes analyzed in our work are derived from bacteria with larger genomes (gram-negative Proteobacteria and gram-positive Firmicutes, respectively). During their evolutionary adaptation to their specific eukaryotic host cell, these organisms have massively lost genes, including genes coding for many isoacceptor tRNA and tRNA modification enzymes. With their minimal genomes, and unlike more specialized organelles, they are generally considered to correspond to the simplest living, autonomous organisms. We purposely did not include in our analysis genomes of insect symbionts with extremely reduced genomes (below 300 kb), such as Candidatus *Carsonella ruddii*, the endosymbiont of the psyllid *Pachypsylla venusta* (genome size of 160 kb with 183 CDSs [50]) and the very recently sequenced Candidatus *Tremblaya princeps* str PCVAL of the citrus mealybug *Planococcus citri* (genome size of 138kb, about 110 CDSs [51]). Both *C. Carsonella ruddii* and *C. Tremblaya princeps* have lost several enzymes required for self-replication, several ribosomal RNA, and many aminoacyl-tRNA synthetases. *C. Tremblaya princeps* has even lost most of its tRNA genes. These organisms must therefore rely on host proteins and tRNAs. They resemble organelles (mitochondria and plastids) [32, 52, 53], and cannot be used in our analysis as we cannot predict the presence of modifications from the presence of the corresponding genes in the endosymbiont.

The finding that *C. R. pediculicola* has lost all modifications of the tRNA body suggests that the structural and recognition roles of modifications outside the anticodon region (reviewed in [3] and [4]) are dispensable in the context of intracellular organisms with slow growth rates and probably with limited sets of nucleases genes and whertRNA degradation might be less of an issue [9]. Indeed, one can expect that protein synthesis might not be as accurate in Mollicutes and insect symbionts as in more sophisticated free-living bacteria. However, since these organisms are not in constant competition with other bacteria, they can certainly survive with a less efficient translation system. The positions of these parasites on the bacterial phylogenetic tree suggest that these are fast evolving bacteria with elevated mutation rates ([29] and several chapters of [54]). Proteins generated by an inaccurate translation system might provide an advantage to the parasite to evolve faster than other bacteria producing a more homogeneous proteome (discussed in [55–57]) and could be an advantage for fast adaptation to the host.

The conservation of genes coding for modification enzymes acting at the wobble position as well as the proximal anticodon bases (position 37–40), at least in organisms having a relatively low G+C content (below 35%, like Mollicutes and most insect symbionts), definitively pointed out the importance of these modifications for maintaining minimalist accuracy and efficacy in reading the genetic code based on 61 sense codons for 20 amino acids. Analyzing genomes of organisms having progressively reduced the size of their genomes allows for identification of the genes more resistant to loss. Hence, from an evolutionary perspective, Mollicutes and insect symbionts constitute excellent biological specimens to identify strategies developed during evolution for reading the genetic code with a minimal set of tRNAs and modification enzymes, a situation that could correspond to what might have occurred at an early stage of life, when the genetic code was just emerging [24, 58].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health (R01 GM70641) to VdC. HG hold a position of Emeritus Scientist at the Center of Molecular Genetic of the CNRS in Gif-sur-Yvette (France) in the laboratory of Dominique Fourmy. We thank David Reed and Bret Boyd for introducing us to the field of louse symbionts, Basma El Yacoubi for numerous and insightful discussions and Patrick C. Thiaville and Jennifer J. Thiaville for editing the manuscript.

References

1. Czerwoniec A, Dunin-Horkawicz S, Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, Grosjean H, Rother K. *Nucleic Acids Res.* 2009; 37:D118. [PubMed: 18854352]
2. Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J. *Nucleic Acids Res.* 2009; 37:D159.
3. Björk, GR.; Hagervall, TG. *Escherichia coli* and *Salmonella*. In: Böck, A.; Curtis, R.; Kaper, JB.; Neidhardt, FC.; Nyström, T.; Squires, CL., editors. *Cellular and Molecular Biology*. ASM. Press; Washington DC: 2005. <http://www.ecosal.org> Module 4.6.2
4. Phizicky EM, Hopper AK. *Genes & Dev.* 2010; 24:1832. [PubMed: 20810645]
5. Cermakian, N.; Cedergren, R. Modification and Editing of RNA. Grosjean, H.; Benne, R., editors. ASM Press; Washington DC: 1998. p. 535
6. Björk GR. *Chemica Scripta.* 1986; 26B:91.
7. Ouzounis CA, Kunin V, Darzentas N, Goldovsky L. *Res Microbiol.* 2006; 157:57. [PubMed: 16431085]
8. Forster AC, Church GM. *Mol Syst Biol.* 2006;2.
9. Alexandrov A, Chernyakov I, Gu W, Hiley SL, Hughes TR, Grayhack EJ, Phizicky EM. *Mol Cell.* 2006; 21:87. [PubMed: 16387656]
10. Grosshans H, Lecointe F, Grosjean H, Hurt E, Simos G. *J Biol Chem.* 2001; 276:46333. [PubMed: 11571299]
11. Gerlt JA, Babbitt PC, Jacobson MP, Almo SC. *J Biol Chem.* 2012; 287:29. [PubMed: 22069326]
12. Christian T, Evilia C, Williams S, Hou YM. *J Mol Biol.* 2004; 339:707. [PubMed: 15165845]
13. Urbonavicius J, Skouloubris S, Myllykallio H, Grosjean H. *Nucleic Acids Res.* 2005; 33:3955. [PubMed: 16027442]
14. Galperin MY, Koonin EV. *J Biol Chem.* 2012; 287:21. [PubMed: 22069324]
15. El Yacoubi B, Lyons B, Cruz Y, Reddy R, Nordin B, Agnelli F, Williamson JR, Schimmel P, Swairjo MA, de Crécy-Lagard V. *Nucleic Acids Res.* 2009; 37:2894. [PubMed: 19287007]
16. El Yacoubi B, Hatin I, Deutsch C, Kahveci T, Rousset JP, Iwata-Reuyl D, Murzin AG, de Crécy-Lagard V. *EMBO J.* 2011; 30:882. [PubMed: 21285948]
17. Srinivasan M, Mehta P, Yu Y, Prugar E, Koonin EV, Karzai AW, Sternglanz R. *EMBO J.* 2011; 30:873. [PubMed: 21183954]
18. Davanloo P, Sprinzl M, Watanabe K, Albani M, Kersten H. *Nucleic Acids Res.* 1979; 6:1571. [PubMed: 377228]
19. Sengupta R, Vainauskas S, Yarian C, Sochacka E, Malkiewicz A, Guenther RH, Koshlap KM, Agris PF. *Nucleic Acids Res.* 2000; 28:1374. [PubMed: 10684932]
20. Chatterjee AK, Blaby I, Thiaville PC, Majumder M, Grosjean H, Yuan YA, Gupta R, de Crécy-Lagard V. *RNA.* 2012; 18:421. [PubMed: 22274953]
21. Wurm JP, Griese M, Bahr U, Held M, Heckel A, Karas A, Soppa J, Wohnert J. *RNA.* 2012; 18:412. [PubMed: 22274954]
22. Grosjean H, de Crécy-Lagard V, Marck C. *FEBS Lett.* 2010; 584:252. [PubMed: 19931533]
23. Silva FJ, Belda E, Talens SE. *Nucleic Acids Res.* 2006; 34:6015. [PubMed: 17071718]

24. de Crécy-Lagard V, Marck C, Brochier-Armanet C, Grosjean H. *IUBMB Life*. 2007; 59:634. [PubMed: 17852564]
25. Fabret C, Dervyn E, Dalmais B, Guillot A, Marck C, Grosjean H, Noirot P. *Mol Microbiol*. 2011; 80:1062. [PubMed: 21435031]
26. van der Gulik P, Hoff W. *J Mol Evol*. 2011; 73:59. [PubMed: 22076654]
27. Maniloff, J. *Molecular Biology and Pathogenicity of Mycoplasmas*. Razin, S.; Hermann, R., editors. Kluwer Academic/Plenum Publisher; New York: 2000. p. 31
28. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman JL, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS. *Science*. 1995; 270:397. [PubMed: 7569993]
29. Sirand-Pugnet P, Citti C, Barré A, Blanchard A. *Res Microbiol*. 2007; 158:754. [PubMed: 18023150]
30. Moya A, Pereto J, Gil R, Latorre A. *Nat Rev Genet*. 2008; 9:218. [PubMed: 18268509]
31. John PM. *Curr Opin Microbiol*. 2010; 13:73. [PubMed: 20044299]
32. McCutcheon JP, Moran NA. *Nat Rev Micro*. 2012; 10:13.
33. Douglas AE. *Cell Host Microbe*. 2011; 10:359. [PubMed: 22018236]
34. Aziz R, Bartels D, Best A, DeJongh M, Disz T, Edwards R, Formsma K, Gerdes S, Glass E, Kubal M, Meyer F, Olsen G, Olson R, Osterman A, Overbeek R, McNeil L, Paarmann D, Paczian T, Parrello B, Pusch G, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. *BMC Genomics*. 2008; 9:75. [PubMed: 18261238]
35. Marck C, Grosjean H. *RNA*. 2002; 8:1189. [PubMed: 12403461]
36. Higgs PG, Ran W. *Mol Biol Evol*. 2008; 25:2279. [PubMed: 18687657]
37. Dong H, Nilsson L, Kurland CG. *J Mol Biol*. 1996; 260:649. [PubMed: 8709146]
38. Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubío JMC, Ribeiro JMC, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Kraus E, Mittapalli O, Margam VM, Li HM, Meyer JM, Johnson RM, Romero-Severson J, VanZee JP, Alvarez-Ponce D, Vieira FG, Aguadé M, Guirao-Rico S, Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang N, Dasch GA, Struchiner CJ, Madey G, Hannick LI, Bidwell S, Joardar V, Caler E, Shao R, Barker SC, Cameron S, Bruggner RV, Regier A, Johnson J, Viswanathan L, Utterback R, Sutton GG, Lawson D, Waterhouse RM, Venter JC, Strausberg RL, Berenbaum MR, Collins FH, Zdobnov EM, Pittendrigh BR. *Proc Natl Acad Sci USA*. 2010; 107:12168. [PubMed: 20566863]
39. Bujnicki JM, Oudjama Y, Roovers M, Owczarek S, Caillet J, Droogmans L. *RNA*. 2004; 10:1236. [PubMed: 15247431]
40. Moukadiri I, Prado S, Piera J, Velázquez-Campoy A, Björk GR, Armengod ME. *Nucleic Acids Res*. 2009; 37:7177. [PubMed: 19767610]
41. Näsvall SJ, Chen P, Björk GR. *RNA*. 2007; 13:2151. [PubMed: 17942742]
42. Nasvall SJ, Chen P, Björk GR. *RNA*. 2004; 10:1662. [PubMed: 15383682]
43. Samuelsson T, Guindy YS, Lustig F, Boren T, Lagerkvist U. *Proc Natl Acad Sci USA*. 1987; 84:3166. [PubMed: 3554232]
44. Takai K, Okumura S, Hosono K, Yokoyama S, Takaku H. *FEBS Lett*. 1999; 447:1. [PubMed: 10218569]
45. Ledoux S, Olejniczak M, Uhlenbeck OC. *Nat Struct Mol Biol*. 2009; 16:359. [PubMed: 19305403]
46. Takai K, Yokoyama S. *Nucleic Acids Res*. 2003; 31:6383. [PubMed: 14602896]
47. Agris PF. *Nucleic Acids Res*. 2004; 32:223. [PubMed: 14715921]
48. Atkins JF, Björk GR. *Microbiol Mol Biol Rev*. 2009; 73:178. [PubMed: 19258537]
49. Grosjean H, Nicoghosian K, Haumont E, Söll D, Cedergren R. *Nucleic Acids Res*. 1985; 13:5697. [PubMed: 3898020]
50. Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M. *Science*. 2006; 314:267. [PubMed: 17038615]

51. López-Madrigal S, Latorre A, Porcar M, Moya A, Gil R. *J Bacteriol.* 2011; 193:5587. [PubMed: 21914892]
52. Tamames J, Gil R, Latorre A, Pereto J, Silva F, Moya A. *BMC Evolutionary Biology.* 2007; 7:181. [PubMed: 17908294]
53. Douglas AE, Raven JA. *Phil Trans R Soc Lond B Biol Sci.* 2003; 358:5. [PubMed: 12594915]
54. Blanchard, A.; Browning, GF. *Mycoplasmas: Molecular Biology, Pathogenicity and Strategies for control.* Horizon Scientific Press; Norwich UK: 2005.
55. Drummond AD, Wilke CO. *Nat Rev Genet.* 2009; 10:715. [PubMed: 19763154]
56. Li L, Boniecki MT, Jaffe JD, Imai BS, Yau PM, Luthey-Schulten ZA, Martinis SA. *Proc Natl Acad Sci USA.* 2011; 108:9378. [PubMed: 21606343]
57. Meyerovich M, Mamou G, Ben-Yehuda S. *Proc Natl Acad Sci USA.* 2010; 107:11543. [PubMed: 20534550]
58. Novozhilov A, Koonin E. *Biology Direct.* 2009; 4:44. [PubMed: 19925661]
59. Lowe TM, Eddy SR. *Nucl Acids Res.* 1997; 25:955. [PubMed: 9023104]
60. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. *Nucleic Acids Res.* 1997; 25:3389. [PubMed: 9254694]
61. Benítez-Páez A, Villarroya M, Douthwaite S, Gabaldón T, Armengod ME. *RNA.* 2010; 16:2131. [PubMed: 20855540]
62. de Crécy-Lagard, V. *Practical Bioinformatics.* Bujnicki, J., editor. Springer-Verlag; Berlin Heidelberg: 2004. p. 169
63. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Ruckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. *Nucleic Acids Res Symp Series.* 2005; 33:5691.

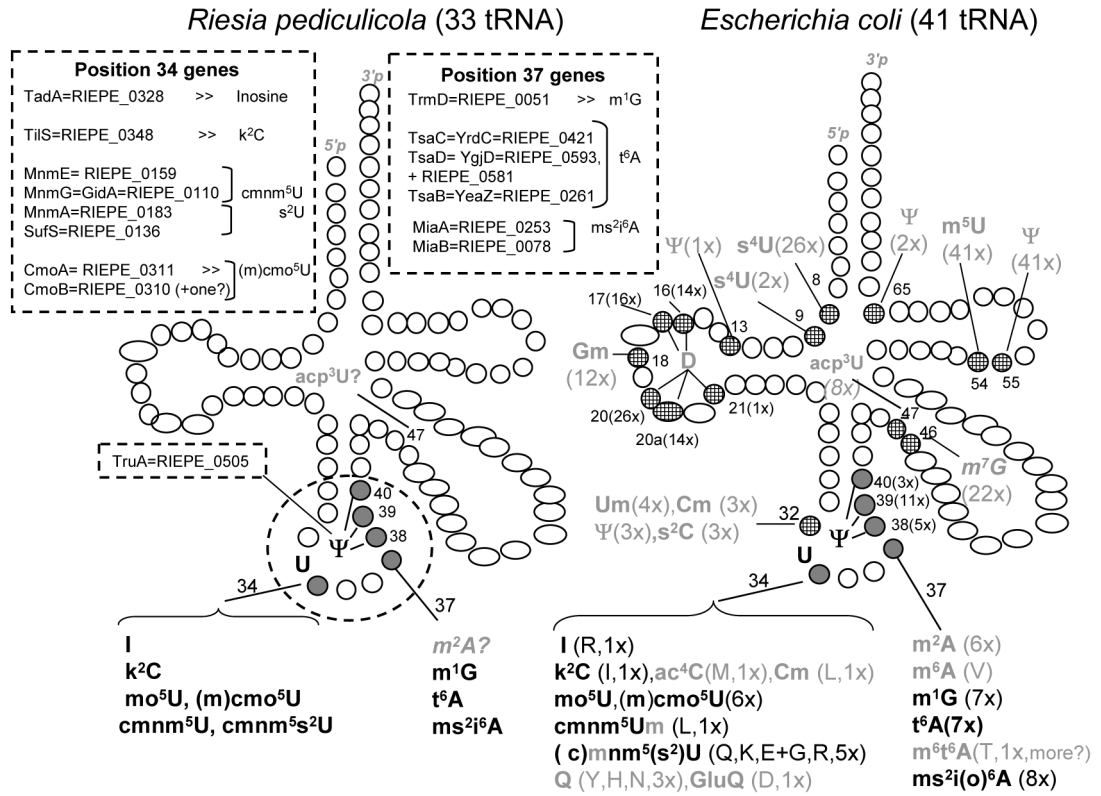


Figure 2. Prediction of the tRNA modifications present in *C. R. pediculicola* and comparison with *E. coli*.

The analysis of the modification genes present in the genome of *C. R. pediculicola* was performed using the SEED database. We constructed a subsystem containing all known *E. coli* tRNA modifications genes (see “tRNA modification *E. coli*” subsystem available on the Public SEED, <http://pubseed.theseed.org/SubsysEditor.cgi>) and extended it to *C. R. pediculicola*. A manual search of the genome (NC_014109, NC_013962) using BlastP and tBlastN [60] with the *E. coli* proteins from the “tRNA modification *E. coli*” subsystem as input was performed. The gene list used is also found in Table 1 of [61] with the addition of the gene encoding TsaA involved in m⁶t⁶A formation (T. Suzuki and V. de Crécy-lagard, personal communication) and TsaD/YgjD involved in t⁶A formation [16]. In *E. coli*, IscS and TusABCDE are required for thiol transfer [3], but no TusACDE homologs were found in *C. R. pediculicola* and SufS is the only IscS homolog in this organism. The m²A³⁷ methylase encoding gene has not been identified in any organism. The same is true for the acp³U⁴⁷ gene, hence the question marks. We previously predicted *yfiF* encodes the missing methylase [62], but this has not been experimentally validated. No *yfiF* homolog or no other methylase of unknown function could be identified in the *C. R. pediculicola* genome, making the presence of m²A³⁷ in this organism unlikely. Finally, to make sure no other genes had been missed, all known tRNA modification genes from *B. subtilis* and *S. cerevisiae* were queried in *C. R. pediculicola* (using the subsystems “tRNA modification Bacteria”, “tRNA modification yeast cytoplasmic” and “tRNA modification yeast mitochondrial” [63]). The genes present in *C. R. pediculicola* are listed in the dashed boxes, with prediction of the resulting modification. Assuming that gene products in *C. R. pediculicola* exhibit the same specificity as the *E. coli* homologs, one can predict which modifications are found in the 33 tRNAs of the symbiont. They are all localized in the anticodon loop and proximal stem (indicated by numbered grey circles, the whole cluster of

modified nucleotides being encircled by dashed line). Only acp^3U , normally present at position 47, cannot be excluded because the gene coding for the corresponding enzyme is unknown. For the same reason, it is not certain if $\text{m}^2\text{A}37$ is present. For comparison, the same tRNA cloverleaf is shown with all the modified nucleotides identified so far by sequencing the 37 fully mature *E. coli* tRNA, as indicated in Figure 1 (only 4 isoacceptor tRNA remain to be sequenced, see Figure 4). The modified nucleotides common to both bacteria are indicated in black, while the ones found only in *E. coli* are indicated in grey. In brackets, the number of isoacceptor tRNAs containing a given modification is indicated. When this number is low, the identity of the modified tRNA is also indicated using the one letter code for amino acid. Open circles correspond to positions in *E. coli* tRNAs where no modification has been found. This compilation was adapted from previously published data [2, 3]. Full names for the different acronyms used to define a given modified base can be found in the MODOMICS database [1].

