

# Altered tRNA characteristics and 3' maturation in bacterial symbionts with reduced genomes

Allison K. Hansen\* and Nancy A. Moran

Department of Ecology and Evolutionary Biology, West Campus, Yale University, PO Box 27388 West Haven, CT 06516-7388, USA

Received March 18, 2012; Revised May 1, 2012; Accepted May 5, 2012

## ABSTRACT

**Translational efficiency is controlled by tRNAs and other genome-encoded mechanisms. In organelles, translational processes are dramatically altered because of genome shrinkage and horizontal acquisition of gene products. The influence of genome reduction on translation in endosymbionts is largely unknown. Here, we investigate whether divergent lineages of *Buchnera aphidicola*, the reduced-genome bacterial endosymbiont of aphids, possess altered translational features compared with their free-living relative, *Escherichia coli*. Our RNAseq data support the hypothesis that translation is less optimal in *Buchnera* than in *E. coli*. We observed a specific, convergent, pattern of tRNA loss in *Buchnera* and other endosymbionts that have undergone genome shrinkage. Furthermore, many modified nucleoside pathways that are important for *E. coli* translation are lost in *Buchnera*. Additionally, *Buchnera*'s A+T compositional bias has resulted in reduced tRNA thermostability, and may have altered aminoacyl-tRNA synthetase recognition sites. *Buchnera* tRNA genes are shorter than those of *E. coli*, as the majority no longer has a genome-encoded 3' CCA; however, all the expressed, shortened tRNAs undergo 3' CCA maturation. Moreover, expression of tRNA isoacceptors was not correlated with the usage of corresponding codons. Overall, our data suggest that endosymbiont genome evolution alters tRNA characteristics that are known to influence translational efficiency in their free-living relative.**

## INTRODUCTION

In the final step of protein synthesis, mRNA sequences must be accurately and efficiently translated into amino

acid proteins. Reliable and efficient translation depends critically on tRNA, which must exhibit specificity in aminoacylation, and correct pairing of the anticodon with its codon on the mRNA. The robust nature of the genetic code and numerous genome-encoded mechanisms promote translational accuracy (1,2), thus preventing deleterious events such as the reassignment of codons that can alter the function of thousands of genes. Nevertheless, tRNAs and the genetic code sometimes do change, especially in genomes undergoing size reduction as exemplified by mitochondria and plastids (1,3,4). These organelle genomes, which are derived from genomes of symbiotic bacteria (5,6), exhibit the most extreme cases of architectural alterations such as an increase in molecular evolutionary rate, inability to recombine, and massive gene loss that sometimes leads to tRNA loss and changes in the genetic code (7). Organelles encode a limited set of proteins and rely on other co-occurring genomes for enzymes and tRNAs (3,8,9).

The reduced genomes of some bacterial endosymbionts exhibit similar but less extreme alterations in genome sequence compared with organelles (10). However, unlike organelles, most endosymbionts are still autonomous in the sense that they possess their own core genetic machinery (11–13), including the conventional bacterial structure of tRNAs (14,15). Most endosymbionts retain the universal genetic code, but exceptions do exist among the tiniest genomes, in which UGA is sometimes recoded from Stop to Trp (16,17). In contrast, organelle tRNAs and their translational machinery are highly divergent from those of most bacteria (1,3,8,18). The question still remains as to how endosymbiont tRNAs and translational mechanisms differ from those of ancestral free-living genomes that are not reduced. Overall, we hypothesize that the process of genome shrinkage in endosymbionts results in a reduction of translational efficiency and integrity resembling a transitional stage between free-living ancestors and organelles.

Present day genomic features of bacterial endosymbionts result from their ancient transition from a free-living lifestyle to an obligate intracellular association (10).

\*To whom correspondence should be addressed. Tel: +1 203 737 3106; Fax: +1 203 737 3109; Email: allison.hansen@yale.edu

Many bacteria that replicate strictly in host intracellular environments possess reduced genomes with sequences that are A+T biased relative to those of their free-living ancestors (10,19,20). One such bacterium demonstrating these genomic shifts is *Buchnera aphidicola*, an obligate unculturable endosymbiont of aphids (21). *Buchnera* has coevolved with its aphid hosts for 200-250 million years (21), during which its genome shrank to only 416–652 kbp depending on the lineage (22–27). Based on previous gene expression and genomic studies in *Buchnera*, genome reduction and accelerated sequence evolution has resulted in changes that are hypothesized to lower the efficiency and accuracy of transcription and translation (28–31) as compared with the free-living relatives. We predict that *Buchnera* will also exhibit less optimal tRNA features. Presently, transcribed tRNAs and associated transcriptional mechanisms, which are key components of efficient and accurate translation, have not been extensively examined in *Buchnera* or any other bacterial endosymbiont.

Comprehensive characterization of transcribed endosymbiont tRNAs has previously been difficult largely because of the inability to isolate unculturable symbiont tRNAs free of host contamination. However, analysis of tRNAs beyond the level of DNA-encoded genes can reveal the nature of tRNA maturation, including the diversity of posttranscriptional processing that may occur. Taking advantage of new methodologies in high-throughput RNA sequencing (directional RNAseq), and the availability of several divergent *Buchnera* genomes (23,25,27), we investigated how genome reduction and A+T richness affect tRNA evolution in this model endosymbiont. This comparative framework provides us with an understanding of the conservation of tRNA sequences that influence specificity in aminoacylation and secondary structure as well as conservation of nucleoside modification pathways that influence anticodon-codon base pairing (1,2,32). From these data, we were able to address how *Buchnera* tRNAs and associated transcriptional fidelity mechanisms are altered relative to those of free-living relatives, exemplified by *Escherichia coli*. Additionally, because numerous reduced endosymbiont genomes have recently been sequenced (10), we investigated whether a pattern of tRNA loss was present among reduced endosymbiont genomes.

## MATERIALS AND METHODS

### Sample preparation

Four aphid species, *Acyrtosiphon pisum* (strains LSR1 and 5A), *Acyrtosiphon kondoi* (strain Ak), *Schizaphis graminum* (strain Sg) and *Uroleucon ambrosiae* (strain UA002, referred to as Ua), were reared in the same growth chamber at 20°C. *A. pisum* was reared on seedlings of *Vicia faba*, *A. kondoi* on *Medicago sativa*, *U. ambrosiae* on *Tithonia mexicana* and *S. graminum* on *Hordeum vulgare*.

For each aphid strain, *B. aphidicola* cells were filtered from 3 g of mixed age aphids. Filtration was done according to the study by Moran *et al.* (33), with modifications as

follows. First, modified buffer A (34) was used instead of PBS. Also, after the 1000 rpm centrifugation step, the pellet was resuspended and used for subsequent filtration steps instead of the supernatant. After the last centrifugation step, supernatant and the protein layer were discarded and the pellet was immediately immersed with Ambion TRI Reagent Solution. For RNA extraction, a similar protocol was used as in Hansen and Moran (34) except that, after step 5, Qiagen's miRNAeasy protocol under appendix A from Qiagen's miRNAeasy Mini Handbook was used to enrich for miRNA (i.e. RNA <200 bp). RNA was DNAase treated, and quality and quantity was checked as in Hansen and Moran (34). All filtration and extraction materials were treated with RNase AWAY (Molecular BioProducts, Inc, CA, USA), and all solutions were RNase free.

### RNA sequencing, read processing, mapping, expression and identification

The Yale Keck sequencing center carried out library preparation and sequencing of *Buchnera* tRNA for all five aphid strains. Briefly, for tRNA library preparation, the Illumina mRNA directional sequencing protocol was followed starting at the phosphatase treatment step. RNA <200 bp was directionally sequenced one lane per sample with Illumina 35 bp reads. The CLC Genomic Workbench Aarhus, Denmark was used for read processing and mapping. For all reads, small RNA adapters and reads with ambiguous nucleotides were trimmed from reads. Trimmed reads were then mapped to corresponding *Buchnera* genomes (Table 1; 23,25,27) with CLC Genomic Workbench short read local alignment mapping using the default settings for short reads. All *Buchnera* taxa used in this study possess similar genome sizes (Ap-5A = 642 122 bp; Ak = 641 794 bp; Ua = 615 380 bp; Sg = 641 454 bp). tRNA reads that mapped sense and anti-sense relative to the tRNA gene were converted into Reads Per Kilobase of exon model per million mapped reads (RPKM). Coverage per base pair was calculated using custom perl scripts and Microsoft Excel and was viewed in Artemis 13.0 (35) to visualize sense and anti-sense tRNA coverage. For each *Buchnera* strain, tRNA genes were annotated using genome annotations in NCBI, tRNAscan-SE 1.21 (14,15) and Artemis 13.0 (35) to verify whether 3' CCA was encoded in the genome. tRNA CCA 3' maturation occurs in all organisms and is essential for charging tRNAs with amino acids. To identify CCA 3' maturation, the last 3' 20 bp of annotated tRNA's were retrieved from all high quality raw reads. Reads that perfectly matched the last 20 bp were binned into the following three categories: (i) reads match the 3' tRNA end and no more nucleotides are processed, (ii) reads match the 3' tRNA end plus additional non-CCA nucleotides are transcribed and (iii) reads match the 3' tRNA end plus CCA is added by maturation. To analyse A+T richness in *Buchnera* and *E. coli* CDS and tRNA genes the program EMBOSS (36) was used. To calculate codon usage of 50 highly expressed *Buchnera* genes (37), E-cai (38) was used.

**Table 1.** Total reads mapping per sample from Illumina directional RNAseq

<i>Buchnera</i> RNA samples	Mapped to tRNA		Reference genome
	(sense)	(antisense)	
Ap-5A	5 704 865	73 652	NC_011833
Ap-LSR1	7 545 480	59 133	NC_011833
Ak	4 387 690	27 322	PRJNA65479
Ua	13 031 381	20 276	PRJNA65481
Sg	8 097 570	61 010	NC_004061

After consensus, RNAseq reads corresponding to tRNA genes were mapped and assembled, tRNA species were identified with tRNAscan-SE 1.21, with *E. coli* homology Blast searches (39), and with verification of the presence of signature identity elements relative to *E. coli* (32).

### Survey of tRNA complements in small genomes

The last comprehensive survey of tRNA genes from bacteria was conducted in 2002 and only included the endosymbiont genome of *Buchnera* strain APS (40). Because several smaller endosymbiont genomes have been sequenced since 2002, we surveyed several more genomes that varied drastically in genome size and phylogenetic placement. The tRNAscan-SE Genomic tRNA database (41) was used to characterize the presence of tRNA gene isoacceptors (i.e. a tRNA species that binds to one or more codons for a particular amino acid residue) in 16 genomes.

### High throughput detection of modified nucleoside bases in tRNAs

During library preparation, some modified bases cause the reverse transcriptase to either fall off at the modified position, and/or to incorporate a 'mismatch' relative to the reference genome sequence (42,43). To detect modified bases and potential posttranscriptional processing, we screened for mismatches in tRNA reads relative to the reference tRNA gene similar to Iida *et al.* (42) and Findeiß *et al.* (43). After mapping, only the sense tRNA reads in CLC (using the same mapping parameters as discussed earlier), we ran CLC single-nucleotide polymorphism (SNP) analyses to detect mismatches. Threshold criteria for counting a mismatch were established by identifying conserved mismatches in both Ap-5A and AP-LSR1 (two different strains from the same aphid species). These two strains shared 38 mismatches for which the mismatch rate was more than 1% per base (i.e. above Illumina's expected error rate per base) and the alternative variant count was at least eight reads. This mismatch criterion was then used to detect mismatches in other strains for a total of four divergent *Buchnera* taxa (Ap, Ak, Ua and Sg).

Predicted tRNA-modified bases and their pathways for each *Buchnera* tRNA were obtained from *E. coli* homologs using both <http://modomics.genesilico.pl/pathways/> (44) and <http://www.ecocyc.org/> (45). Divergent *Buchnera* genomes (23,25,27) were searched

for modification pathway enzymes using *E. coli* homologs using Blastp (39).

### tRNA secondary structure

Infernal (46) was used to generate tRNA sequence and secondary structure alignments among *Buchnera* strains and *E. coli*. The covariance model, RF00005 cm, was used, which accounts for tRNA secondary structure constraints. Using Infernal output, 4sale (47) was used to compute pairwise compensatory substitution tables from stems for all tRNAs among *Buchnera* strains and *E. coli*. Stability of tRNA secondary structure was measured as Delta G ( $\Delta G$ ), the change in Gibbs Free Energy (in units of kcal/mole). Thus, the more negative  $\Delta G$  is, the more thermodynamically stable the tRNA secondary structure.  $\Delta G$  was computed for tRNAs of each strain individually using RNAalifold (48,49) with constraints on tRNA constraint folding generated by tRNAscan-SE 1.21 (14,15).

All raw sense and anti-sense tRNA data were submitted to NCBI Genbank under SRA submission: SRA049863.3, under Bioproject #: (i) PRJNA82811, (ii) PRJNA82809, (iii) PRJNA82797, (iv) PRJNA82793, (v) PRJNA82789. All paired sample-t test (percent guanine-cytosine (%GC), tRNA length), correlation (pairwise RPKM comparisons) and regression (codon usage and tRNA expression) statistics were carried out using IBM SPSS Statistics. 2010 for Mac, standard version 19.0. New York, USA.

## RESULTS

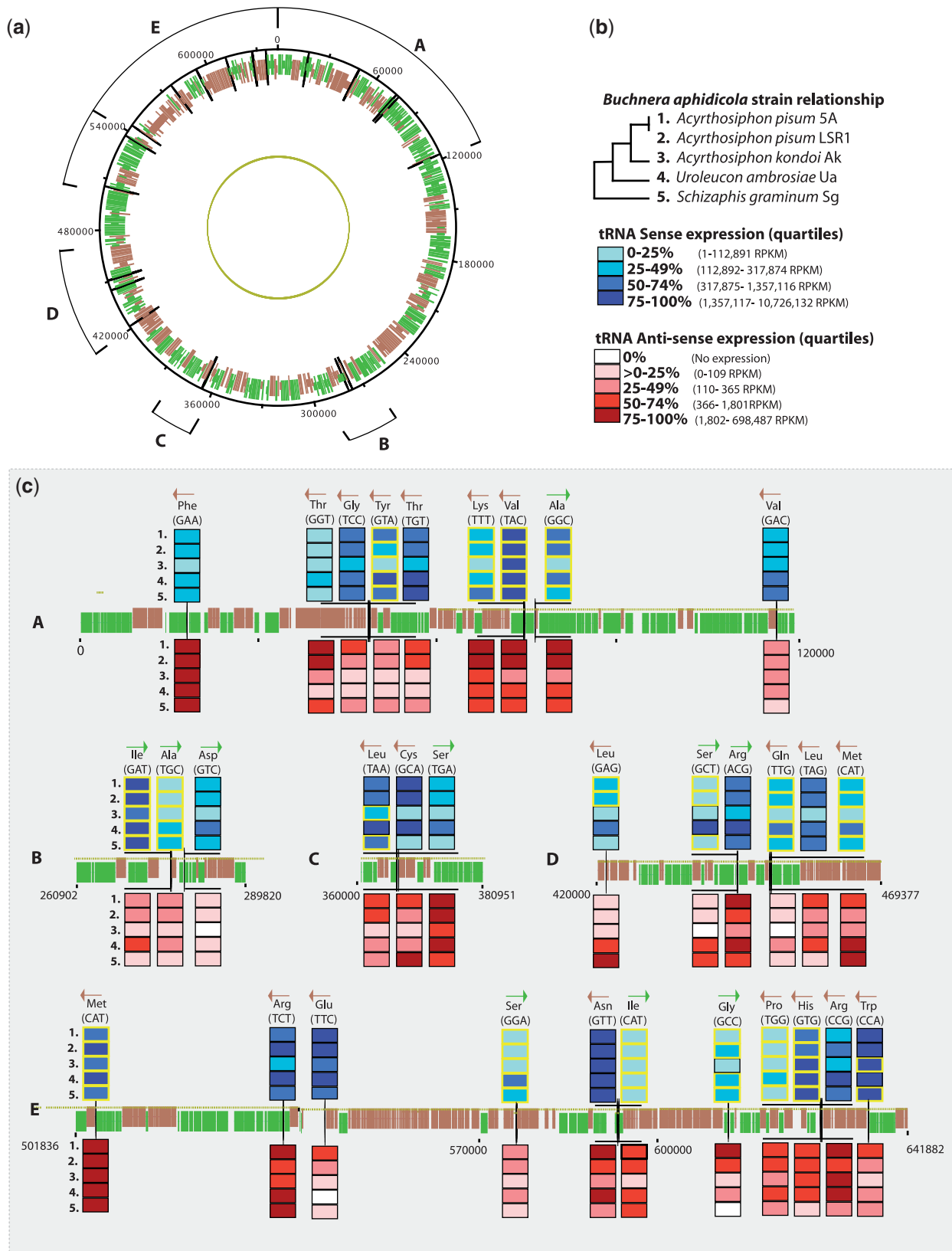
### All *Buchnera* tRNAs are transcribed

For all *Buchnera* genomes, tRNA genes occur in the same genomic positions (Figure 1). Based on tScan and blastn detection of homology with *E. coli* the same 32 tRNA genes and 29 anticodon types are conserved across *Buchnera* taxa (Figure 1).

As expected, directional RNAseq reads map primarily in the sense direction of tRNA genes, with antisense reads averaging less than 1% of the sense reads (Table 1, Figure 1). All *Buchnera* tRNA genes are expressed in the sense direction, but some lack antisense expression, depending on strain, and sense expression is always higher than antisense expression (except for Phe GAA in Ak and Sg) (Figure 1). tRNA sense expression is positively correlated across divergent *Buchnera* taxa (Table 2). The level of antisense expression is highly correlated across all *Buchnera* taxa, but the correlation is less for *Buchnera*-Sg, the most divergent taxon (Table 2). Transcriptional start sites and coverage curves for antisense RNAs varied widely across *Buchnera* taxa. Nevertheless, conserved 5' transcriptional start sites and coverage curves were identified for several antisense RNAs that occurred on or near tRNA genes for all five *Buchnera* taxa (Supplementary Table S1).

### Conservation of tRNA identity elements in *Buchnera*

Recognition of tRNAs by tRNA synthetases is essential to the fidelity of translation. Aminoacyl-tRNA synthetases



**Figure 1.** Sense and antisense expression of 32 tRNA genes from five *Buchnera* taxa based on RNAseq data. (a) *Buchnera* 5A's chromosome. Green and brown regions indicate forward and reverse coding regions respectively. Letters A-E specify regions where tRNA genes are coded, and correspond to (c). (b) phylogenetic distance of *Buchnera* taxa; number labels (1-5) correspond to (c). tRNA expression is presented as Reads Per Kilobase of exon model per Million mapped reads (RPKM). Blue and red boxes represent sense and antisense RNA expression, respectively, at different magnitudes. Yellow outlined boxes indicate tRNAs with genome encoded 3' CCA. tRNA gene direction is indicated by arrows above tRNA isoacceptor names.

(aaRS) must recognize multiple tRNA isoacceptors (i.e. different tRNA species that bind to alternative codons for the same amino acid residue) but discriminate against others. This recognition is dependent on tRNA identity elements, consisting of evolutionarily conserved bases at specific positions of tRNAs (Giege' *et al.* 1998). Based on RNAseq data from all taxa, unmodified identity elements for each tRNA are identical to those in *E. coli* except for base substitutions in Cys<sup>GCA</sup> (G15 to U15; A13 to G13), Ser<sup>GGA</sup> (G73 to A73), Ser<sup>GCT</sup> (variable loop 1 bp shorter, except in Sg) and Ala<sup>GGC</sup> (G20 to U20, 5A and Ua only; G20 to C20, Sg and Ak only). Based on blastp analyses, all 20 cognate aaRS are encoded within each *Buchnera* genome.

In contrast to *E. coli* and most other organisms with nonreduced genomes, *Buchnera* does not encode multiple tRNA genes with matching anticodons, except for three tRNA genes encoding the anticodon CAU. Two of these genes encode either an initiation or elongation Met tRNA based on tRNA identity elements and homology (Table 3). The other tRNA gene encoding a CAU anticodon possesses homology and identity elements corresponding to the Ile<sup>LAU</sup> anticodon (Table 3).

**Table 2.** Spearman rho correlations of tRNA sense expression (grey, below) and tRNA anti-sense expression (white, above) between pairs of *Buchnera* taxa\*

	Ap-5a	Ap-LSR1	Ak	Ua	Sg
Ap-5a	1.000	<b>0.984</b>	<b>0.82</b>	<b>0.501</b>	0.249
Ap-LSR1	<b>0.956</b>	1.000	<b>0.814</b>	<b>0.489</b>	0.252
Ak	<b>0.691</b>	<b>0.693</b>	1.000	<b>0.664</b>	0.312
Ua	<b>0.673</b>	<b>0.624</b>	<b>0.775</b>	1.000	<b>0.527</b>
Sg	<b>0.534</b>	<b>0.532</b>	<b>0.816</b>	<b>0.599</b>	1.000

\*Bold = significant ( $P < 0.01$ ); unbold = nonsignificant ( $P > 0.05$ ).  $N = 32$ .

### Selective loss of tRNA isoacceptors from *Buchnera* genomes

Numerous tRNA isoacceptors are present in *E. coli* but missing from all *Buchnera* strains. Many *Buchnera* tRNA isoacceptors that belong to 4-codon family boxes and to two-codon families (5'-NNR codon type) have been lost from *Buchnera* genomes (Table 3). 5'-CNN anticodons were preferentially lost in family boxes corresponding to Leu, Gly, Ser, Thr and Pro. Only one family box, corresponding to Pro, lost both 5'-CNN and 5'-GNN anticodons. For two-codon families, a 5' CNN anticodon was lost from Gln (and Leu and Arg for 6-codon families), relative to *E. coli* (Table 3). Based on Watson and Crick base-pairing and revised wobble rules (50,51), all tRNA isoacceptors encoded and expressed in *Buchnera* can base pair with the 61 possible codons (Table 3), which are all still encoded in *Buchnera*'s protein-coding genes at variable frequencies.

The pattern of tRNA gene isoacceptor loss was examined in 16 bacterial taxa representing a wide range of genome sizes and phylogenetic associations, including some with extremely reduced genomes (Figure 2). Reduced genomes show common patterns of retention of particular anticodons. For family box codons, 5'-CNN anticodons followed by 5'-GNN anticodons are consistently eliminated from the small genomes. For 5'-NNR two-box codons, 5'-CNN anticodons are eliminated. In the most reduced genomes, only 5'-UNN anticodons remain for both family box and two-box codons. Unmodified 5'-U anticodons can wobble and pair with all four base combinations for family box codons (50). Therefore, for 5'-NNR two-box codons, the 5'-U of anticodons must be modified to prevent mistranslation of neighboring two-box (NNY) codons (1, 2) (e.g. an unmodified 5' U in a Gln 5'-UUG anticodon can mispair with His codons 5'-CAU and CAC [Table 3]).

**Table 3.** *Buchnera* tRNA isoacceptors retained and expressed in *Buchnera* genomes (bolded)

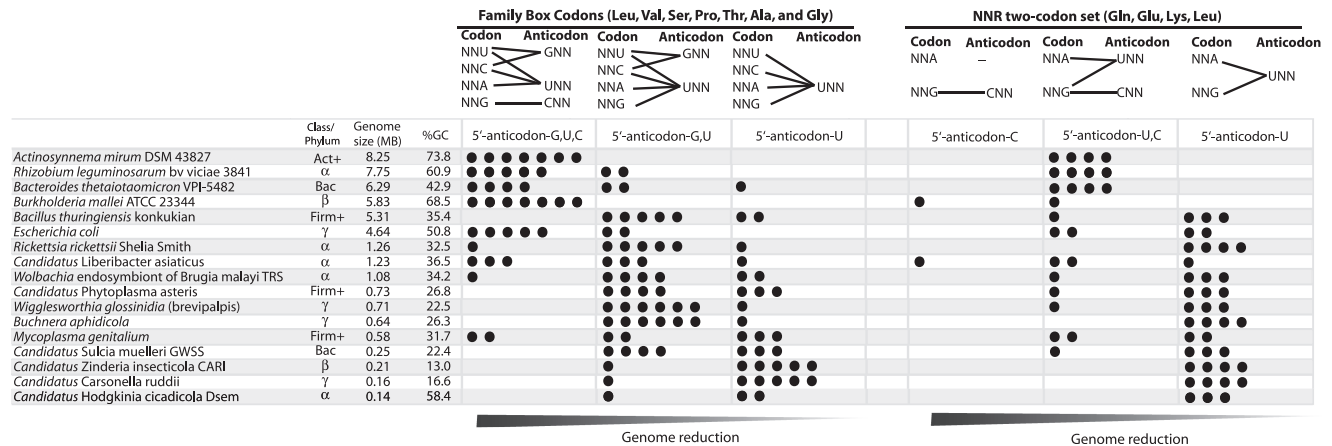
	Codon	AA	Anti-codon	Codon	AA	Anti-codon	Codon	AA	Anti-codon	Codon	AA	Anti-codon	
U	UUU	Phe		UCU	Ser		UAU	Tyr		UGU	Cys		U
	UUC	Phe	<b>GAA</b>	UCC	Ser	<b>GGA</b>	UAC	Tyr	<b>GUA</b>	UGC	Cys	<b>GCA</b>	C
	UUA	Leu	<b>UAA</b>	UCA	Ser	<b>UGA</b>	UAA	Stop		UGA	Stop		A
	UUG	Leu	<b>CAA<sup>c</sup></b>	UCG	Ser	<b>CGA<sup>c</sup></b>	UAG	Stop		UGG	Trp	<b>CCA</b>	G
C	CUU	Leu		CCU	Pro		CAU	His		<b>CGU</b>	Arg	<b>ICG</b>	U
	CUC	Leu	<b>GAG</b>	CCC	Pro	<b>GGG<sup>c</sup></b>	CAC	His	<b>GUG</b>	<b>CGC</b>	Arg		C
	CUA	Leu	<b>UAG</b>	CCA	Pro	<b>UGG</b>	CAA	Gln	<b>UUG</b>	<b>CGA</b>	Arg		A
	CUG	Leu	<b>CAG<sup>c</sup></b>	CCG	Pro	<b>CGG<sup>c</sup></b>	CAG	Gln	<b>CUG<sup>c</sup></b>	<b>CGG</b>	Arg	<b>CCG</b>	G
A	AUU	Ile		ACU	Thr		AAU	Asn		AGU	Ser		U
	AUC	Ile	<b>GAU</b>	ACC	Thr	<b>GGU</b>	AAC	Asn	<b>GUU</b>	AGC	Ser	<b>GCU</b>	C
	AUA	Ile	<b>LAU<sup>a</sup></b>	ACA	Thr	<b>UGU</b>	AAA	Lys	<b>UUU</b>	AGA	Arg	<b>UCU</b>	A
	AUG	Met	<b>CAU<sup>b</sup></b>	ACG	Thr	<b>CGU<sup>c</sup></b>	AAG	Lys		AGG	Arg	<b>CCU<sup>c</sup></b>	G
G	GUU	Val		GCU	Ala		GAU	Asp		<b>GGU</b>	Gly		U
	GUC	Val	<b>GAC</b>	GCC	Ala	<b>GGC</b>	GAC	Asp	<b>GUC</b>	<b>GGC</b>	Gly	<b>GCC</b>	C
	GUA	Val	<b>UAC</b>	GCA	Ala	<b>UGC</b>	GAA	Glu	<b>UUC</b>	<b>GGA</b>	Gly	<b>UCC</b>	A
	GUG	Val		GCG	Ala		GAG	Glu		<b>GGG</b>	Gly	<b>CCC<sup>c</sup></b>	G
	U		C		A		G						

Dark boxes represent family boxes of 4- and 6-fold degenerate codon families. Light grey boxes represent 5'-NNR(A/G) 2 codon boxes of 2- and 6-fold degenerate codon families.

<sup>a</sup>tRNA is homologous to *E. coli* Ile LAU; L modification enzyme encoded in *Buchnera* genomes.

<sup>b</sup>Represents two Met isoacceptors in *Buchnera* (initiation and elongation tRNAs).

<sup>c</sup>*Buchnera* lost this isoacceptor relative to its presence in *E. coli*.



**Figure 2.** Pattern of isoacceptor tRNA gene loss with decreasing genome size for a variety of free-living and symbiotic/pathogenic bacteria from multiple phyla and classes. Black dots represent the presence of a single tRNA isoacceptor. All bacteria have fully sequenced genomes, and their tRNA presence is based on tScan predictions. Note that some bacteria do not possess all seven family box codons and/or all four NNR two-codon sets. a = Alphaproteobacteria, b = Betaproteobacteria, g = Gammaproteobacteria, Bac = Bacteroidetes, Act = Actinobacteria, Firm = Firmicutes, += Gram-positive bacterium.

### *Buchnera* tRNAs are modified and RNA edited

Based on *E. coli* tRNA homologs, 26 different types of nucleoside modifications are predicted to occur in *Buchnera* tRNAs (Table 4, Supplementary Table S2, Supplementary Dataset 1). Nine of these modifications are important for the efficiency and fidelity of protein synthesis and occur in N34 tRNA positions (wobble) of *E. coli* (Table 4). We expect five of these N34 modifications to be retained to code for all cognate codon pairs and prevent mistranslation of other amino acids (e.g.  $\text{mnm}^5\text{u}$ ,  $\text{mnm}^5\text{s}^2\text{U}$ ,  $\text{cmnm}^5\text{Um}$ , I and  $\text{K}^2\text{C}$ ). An inosine (I) modification is important in *E. coli* because 5'-A from anticodon  $\text{Arg}^{\text{ACG}}$  is modified into I, which can wobble and pair with Arg codons CGA, CGU and CGC (55). Lysidine ( $\text{K}^2\text{C}$ ) is an important modification in *E. coli* because 5'-C from anticodon  $\text{Ile}^{\text{CAU}}$  is modified into  $\text{K}^2\text{C}$  (L), which pairs with Ile codon AUA (instead of the Met codon AUG) (59). Other expected N34 modifications ( $\text{mnm}^5\text{u}$ ,  $\text{mnm}^5\text{s}^2\text{U}$  and  $\text{cmnm}^5\text{Um}$ ) are important for modifying anticodon 5' U for NNR two codon boxes, thus preventing mistranslation (1,2). Based on *Buchnera* genome annotations, entire pathways are only present for expected wobble bases I,  $\text{K}^2\text{C}$  and  $\text{cmnm}^5\text{um}$  (Table 4); however, some pathways are only missing the last enzyme in a pathway (e.g.  $\text{mnm}^5\text{u}$  and  $\text{mnm}^5\text{s}^2\text{U}$ ), and/or are still unknown in *E. coli*.

High throughput mismatch evidence (see 'Materials and Methods' section) shared by multiple taxa supports the presence of a modified nucleoside at 5'-A from anticodon  $\text{Arg}^{\text{ACG}}$  in all *Buchnera* strains. These data support the presence of an inosine modification in all taxa. For example, we found a high frequency of anticodon 5'-ACG transcribed as 5'-GCG, where the frequency of 5'-G/A at this wobble base position was, Ap-5A: 61/39%; Ap-LSR1 = 70/30%; Ak = 69/31%; Ua = 27/73%; and Sg = 72/28%. Presence of transcripts containing a 5'-G for the  $\text{Arg}^{\text{ACG}}$  anticodon is strong indirect evidence for an inosine modification. For example,

during the reverse transcription process, the modified nucleoside inosine base pairs with C residues, and therefore 'G' is found in the consensus cDNA sequence instead of 'A' (60). Conserved high throughput mismatch evidence for Ap-5A, Ap-LSR1 and Ak also supports the presence of a modified base at N34 for  $\text{Lys}^{\text{TTT}}$ , suggesting that  $\text{mnm}^5\text{s}^2\text{U}$  is present in these strains even though the *E. coli* version of the pathway appears incomplete in *Buchnera*. Error evidence was not detected for other expected modified wobble positions relative to *E. coli*, even though full pathways are retained in the genome (Table 4).

Other tRNA modifications that are very important for the fidelity of protein synthesis are N37 modifications. N37 modifications are known to stabilize weak A:U and U:A base pairing between N36 of the anticodon and N1 of the codon (1,2,51). Based on *in vitro* experiments, N37 modifications are known to increase the interaction of the codon with the anticodon, preventing miscoding of amino acids and frameshifts (52,53,61–63). In turn, to maintain efficient translation, we expect these modifications to be retained. Based on modifications for the homologous tRNAs in *E. coli*, seven important N37 modified nucleosides are predicted in *Buchnera*. Among *Buchnera* genomes, four N37 nucleosides pathways are retained, two are missing, and one has an unknown pathway in *E. coli* (Table 4). High throughput mismatch evidence supports the presence of a modified base at N37 for  $\text{Phe}^{\text{GAA}}$ ,  $\text{Pro}^{\text{TGG}}$ ,  $\text{Leu}^{\text{GAG}}$  and  $\text{Leu}^{\text{TAG}}$  and thus suggests that  $\text{ms2i6A}$ ,  $\text{m1G}$ ,  $\text{xG}$  and  $\text{xG}$ , respectively, are present in all taxa. However, no mismatch was detected in Sg for  $\text{Leu}^{\text{GAG}}$ . The tRNA modifications at positions other than N34 and N37 that are supported by mismatch evidence are shown in Supplementary Table S2, and Supplementary Dataset 1. Mismatch evidence was also found at positions at which *E. coli* does not process modified nucleosides, suggesting the presence of new modified nucleoside sites and/or RNA editing of mature tRNAs (Supplementary Dataset 1). Collectively, all

**Table 4.** Mechanisms maintaining translational fidelity and efficiency

Modified base (Original base)	tRNA isoacceptors with modification in <i>E. coli</i>	tRNA base modified	Pathway enzymes present in <i>Buchnera</i> <sup>a</sup>	Rnaseq evidence at expected position <sup>b</sup>	General role (reference number)
m <sup>2</sup> A	2-methyladenosine (A)	Arg <sup>ACG</sup> <sub>UUC</sub> , Asp <sup>GUC</sup> <sub>UUG</sub> , Glu <sup>UUC</sup> <sub>UUG</sub> , His <sup>GUG</sup> <sub>UUG</sub>	N37	No	Prevents anticodon-codon mispairing for 5'-NNC/G anticodons (2)
ms <sup>2</sup> 1 <sup>6</sup> A	2-methylthio-N <sup>6</sup> -isopentenyladenosine (A)	Cys <sup>GCA</sup> <sub>UUA</sub> , Leu <sup>UAA</sup> <sub>UAA</sub> , Phe <sup>GAA</sup> <sub>UAA</sub> , Ser <sup>UGA</sup> <sub>CCA</sub> , Trp <sup>UAG</sup> <sub>Tyr</sub>	N37	Yes	Prevents anticodon-codon mispairing for 5'-NNA anticodons, Prevents miscoding of Phe in <i>E. coli</i> , Prevents frameshifting (2, 52, 53)
t <sup>6</sup> A	N <sup>6</sup> -threonylcarbamoyladenosine (A)	Arg <sup>UCU</sup> <sub>UUU</sub> , Asn <sup>GUU</sup> <sub>UUC</sub> , Ile <sup>GAU</sup> <sub>UUC</sub> , Lys <sup>UUU</sup> <sub>GCU</sub> , Met <sup>CAU</sup> <sub>UUC</sub> , Ile <sup>CAU</sup> <sub>UUC</sub> , Ser <sup>GCU</sup> <sub>UUC</sub>	N37	No	Prevents anticodon-codon mispairing for 5'-NNU anticodons (2)
m <sup>1</sup> G	1-methylguanosine (G)	Pro <sup>UGG</sup> <sub>GAG</sub> , Leu <sup>UAG</sup> <sub>UAC</sub>	N37	Yes	Prevents frameshifting (3,4,5)
xG	Unknown (G)	Leu <sup>UAG</sup> <sub>UAC</sub>	Unknown	Yes	Prevents anticodon-codon mispairing for 5'-NNG anticodons (2)
m <sup>6</sup> A	N <sup>6</sup> -methyladenosine (A)	Val <sup>UAC</sup> <sub>GGU</sub>	.....	No	
m <sup>6</sup> 1 <sup>6</sup> A	N <sup>6</sup> -methyl-N <sup>6</sup> -threonylcarbamoyladenosine (A)	Thr <sup>GGU</sup> <sub>GGU</sub>	<b>RimN</b> , <b>QueA</b> <sup>d</sup>	No	Increases the efficiency of Thr <sup>GCU</sup> to read the codon ACC (54)
I	Inosine (A)	Arg <sup>ACG</sup> <sub>UAC</sub>	<b>TadA</b>	Yes	Wobble to decode three Arg codons (55)
cmo <sup>5</sup> u	Uridine 5-oxyacetic acid (U)	Val <sup>UAC</sup> <sub>UAC</sub> , Thr <sup>UGT</sup> <sub>UAC</sub> , Pro <sup>UGG</sup> <sub>UAC</sub> , Ser <sup>UGA</sup> <sub>UAC</sub> , Ala <sup>UGC</sup> <sub>UAC</sub>	Unknown	No	Increases efficiency of 5' U wobble for family box codons (2)
m <sup>mm</sup> 5u	5-methylaminomethyluridine (U)	Arg <sup>UCU</sup> <sub>UCU</sub>	<b>MmmE</b> , <b>MmmG</b> <sup>e</sup>	No	Recognizes Arg codon AGA and much less efficiently codon AGG <i>in vivo</i> , constricting wobble from Ser AGU and AGC (2,56)
QtRNA	Queuosine (G)	Tyr <sup>GUA</sup> <sub>GUG</sub> , Asn <sup>GUU</sup> <sub>GUG</sub> , Asp <sup>GUC</sup> <sub>GUG</sub> , His <sup>GUG</sup> <sub>GUG</sub>	.....	No	Recognition of 5' NAU and NAC codons (57), Prevents frameshifting for Tyr and His codons (53)
xU	Unknown(U)	Gly <sup>UCC</sup> <sub>UCC</sub>	Unknown	No	Unknown modification
m <sup>mm</sup> 5s <sup>2</sup> U	5-methylaminomethyl-2-thiouridine (U)	Gln <sup>UUG</sup> <sub>UUG</sub> , Gln <sup>UUC</sup> <sub>UUC</sub> , Lys <sup>UUU</sup> <sub>UUU</sub>	<b>MmmA</b> , <b>MmmE</b> , <b>MmmG</b> , <b>IscS</b> , <b>GidA</b> <sup>e</sup>	Yes	Prevents frameshifting in Lys codons (53), Constricts 5' U wobble for NNR codons (2)
cm <sup>mm</sup> 5Um	5-carboxymethylaminomethyl-2-O-methyluridine (U)	Leu <sup>UAA</sup> <sub>UAA</sub>	<b>MmmE</b> , <b>MmmG</b>	No	Recognizes Leu codons UUA and UUG but constricts wobble from Phe codons UUU and UUC (2)
ac <sup>4</sup> c	N <sup>4</sup> -acetylcytidine (C)	elongator Met <sup>CAU</sup> <sub>CAU</sub>	.....	No	Prevent misreading of Ile AUA codons (58)
k <sup>2</sup> C	Lysidine (C)	Ile <sup>CAU</sup> <sub>CAU</sub>	<b>tifs</b>	No	Negative determinate for aminoacylation by MetRS (59)

Conservation of *Buchnera* taxa wobble -N34 and N37 tRNA modified base pathways and detection of modified base products in 32 tRNAs.

<sup>a</sup>*Buchnera* strains screened for homologous *E. coli* pathway enzymes are: Ap (5a and LSR1), Ak, Ua, and Sg.

<sup>b</sup>Directional RNAseq detection of modified bases among four divergent *Buchnera* taxa- Ap (5a and LSR1), Ak, Ua and Sg (see 'Materials and methods' section).

<sup>c</sup>Present in all *Buchnera* except pseudogene present in Sg.

<sup>d</sup>Present in Ap, Sg, but absent in Ua, and a pseudogene in Ak.

<sup>e</sup>Entire pathway present in all *Buchnera* taxa except for MmmC.

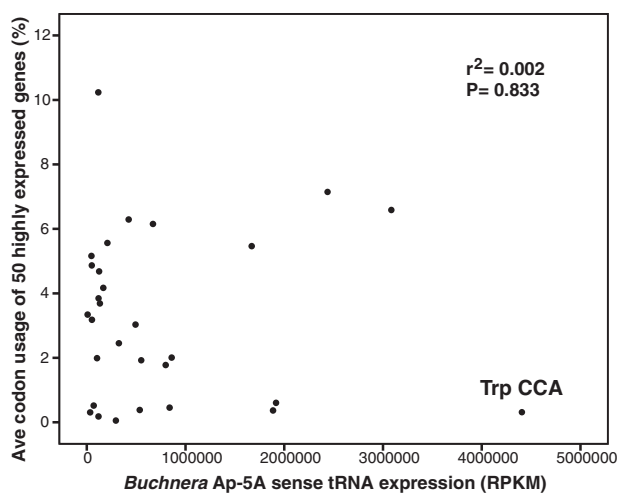
Dotted lines, *Buchnera* taxa where one or more enzymes are absent in the pathway.

Underlined tRNA isoacceptors indicate which tRNA shows RNAseq evidence of a modified base at either N34 or N37 for two or more divergent *Buchnera* taxa.

mismatch frequencies (with the exception of Arg<sup>ACG</sup>) were dominated by the reference sequence base at a frequency of ~90-99% relative to mismatches for all taxa. Mismatches were primarily not changes to a single nucleotide base, but were composed of three different bases other than the reference base.

### No relationship between codon frequencies and tRNA expression

In many species, tRNA abundances are positively correlated with codon usage for highly expressed genes (64,65). Anticodons of highly expressed tRNAs correspond to codons that are used frequently in these genes, thus improving the efficiency of translation (64,65). Based on Watson and Crick and revised wobble base-pairing rules (50,51), each *Buchnera* isoacceptor was paired with its corresponding codon pair. Met CAU, the only duplicate anticodon coding for the same codon, was excluded from analysis. The relationship between percent average codon usage of highly expressed genes and corresponding tRNA isoacceptor expression was examined for each *Buchnera* strain. No significant relationship was found between average codon usage of 50 highly expressed genes in Ap-5A (on leading and lagging strands) and cognate tRNA isoacceptor sense expression (Figure 3). No significant relationship was found on examining the relationship between highly expressed *Buchnera* genes (four chaperones and 54 ribosomal proteins) and cognate tRNA isoacceptor sense expression for all taxa (Supplementary Figure S1A and S1B). Examination of codon usage and tRNA expression scatterplots reveals that most tRNA isoacceptors, regardless of codon usage, are expressed at similar levels (e.g. for Ap-5a in RPKM the 75 percentile = 843 950, median = 309 742 and max = 4 407 138; Figure 3). Trp<sup>CCA</sup> is the highest expressed isoacceptor in all taxa (except Ua), even



**Figure 3.** The relationship between percent average codon usage and corresponding tRNA isoacceptor expression. Average codon usage calculations are based on 50 highly expressed genes in *Buchnera*-Ap (based on Charles *et al.* 2006).

though the corresponding codon occurs at low frequency (Figure 3 and Supplementary Figure S1).

### *Buchnera* tRNAs maintain secondary structure with compensatory base substitutions

As expected, *Buchnera* CDS are significantly more A+T rich relative to CDS of *E. coli* [Figure 4 (c)]. Within each *Buchnera* genome, tRNA genes are 2.2-fold more G+C rich relative to CDS, indicating that selection conserves higher %G+C in tRNA genes. Nevertheless, *Buchnera* tRNA genes are significantly more A+T rich than homologs in *E. coli* [Figure 4 (c)].

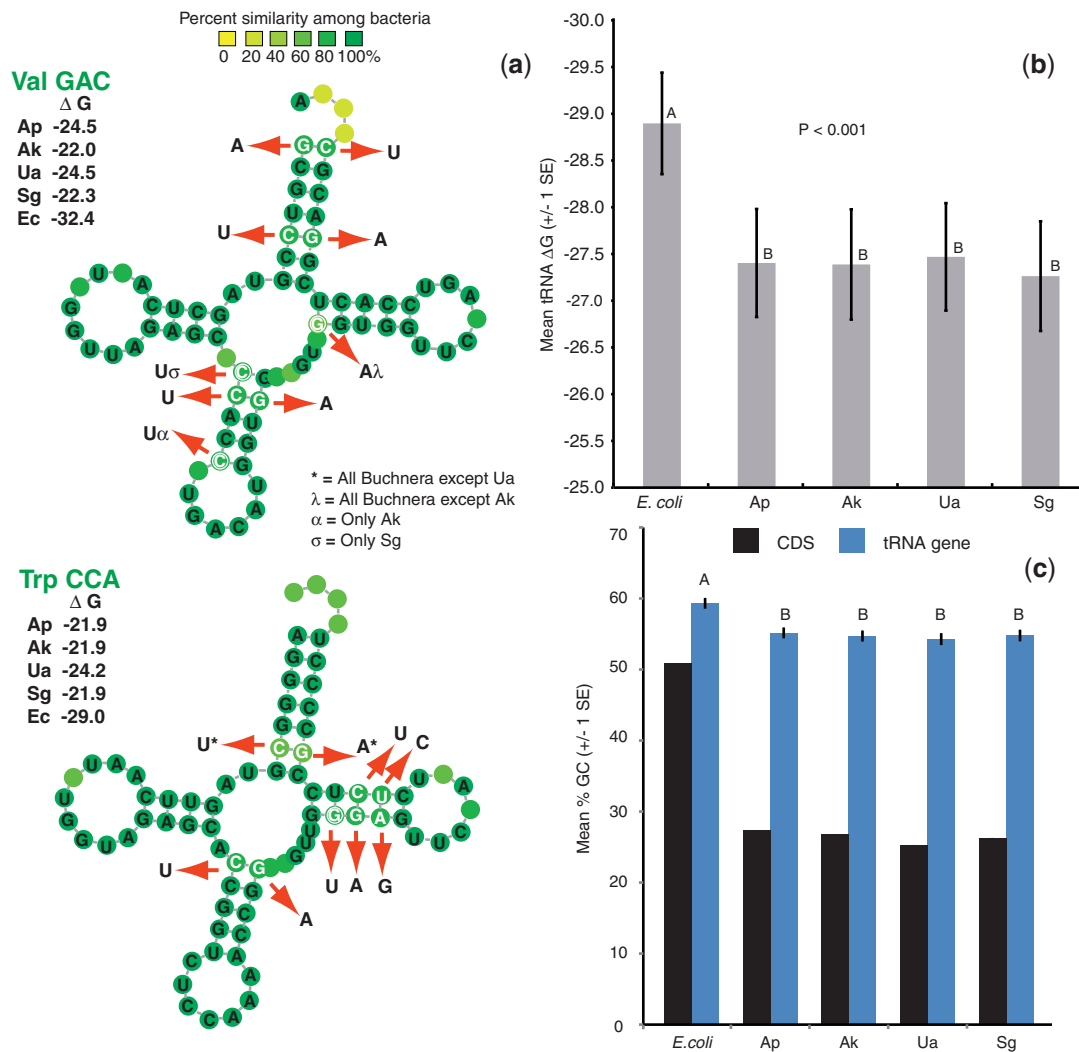
Stability of tRNA secondary structure can decrease with a reduction in %GC, especially in stem structures. Because *Buchnera* tRNAs are more A+T rich than those of *E. coli* [Figure 4 (c)], we measured the stability of *Buchnera* tRNA secondary structure.  $\Delta G$  was significantly more negative in *E. coli* tRNAs relative to homologs in *Buchnera* for all strains, indicating that *Buchnera* tRNAs have reduced stability *in vitro* [Figure 4 (b)]. Whether they have reduced stability *in vivo*, where stabilizing proteins may play a role, remains to be tested. Two tRNAs with the weakest secondary structure in all *Buchnera* relative to *E. coli* were Val<sup>GAG</sup> and Trp<sup>CCA</sup>; both tRNAs possess numerous compensatory and single base substitutions in the stem regions [Figure 4 (a)].

*Buchnera* tRNAs are more A+T biased and display weaker secondary structure than those of *E. coli* (Figure 4). However, a high frequency of compensatory base substitutions are expected in the stem regions as a mechanism for maintaining functionality of these essential molecules. Relative to *E. coli*, a total of 37–42 compensatory base substitutions were found in *Buchnera* tRNA stem regions (Table 5). Many of these compensatory substitutions were C/G to T/A directional changes (Table 5).

### *Buchnera* tRNA gene shrinkage and compensatory 3' maturation

Genome reduction primarily reflects loss of coding genes, as reduction in gene length is minor (<1%, 37), and gene packing is similar for bacterial genomes of different sizes (66). However, *Buchnera* tRNA genes are often shorter in length than their homologs in *E. coli* [Figure 5 (a)]. The difference in length is typically 3 bp and mostly reflects the loss of encoded 3' CCA in the *Buchnera* tRNA genes. At the 3' end of tRNAs, CCA is required for amino acid activation, and must either be encoded in the tRNA gene or added during tRNA maturation by the CCA-adding enzyme. Although *E. coli* and other close relatives of *Buchnera* such as *Vibrio* and *Pseudomonas* spp. all encode 3' CCA in all tRNA genes except that for selenocysteine, only half of *Buchnera* tRNA genes encode 3' CCA [14-17 depending on strain, Figure 5 (b)]. The remaining *Buchnera* tRNA genes have lost the 3' encoded CCA. Our analysis of directional RNAseq reads indicates that the mature transcript of these genes possesses a CCA at the 3' end [Figure 5 (b)], implying CCA-addition.





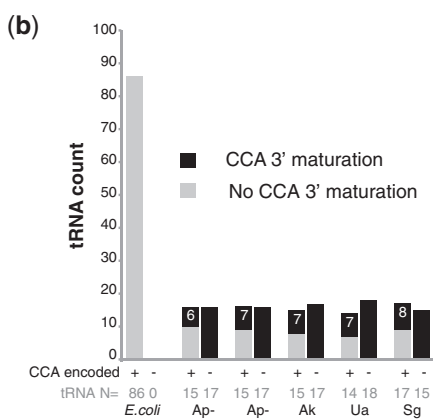
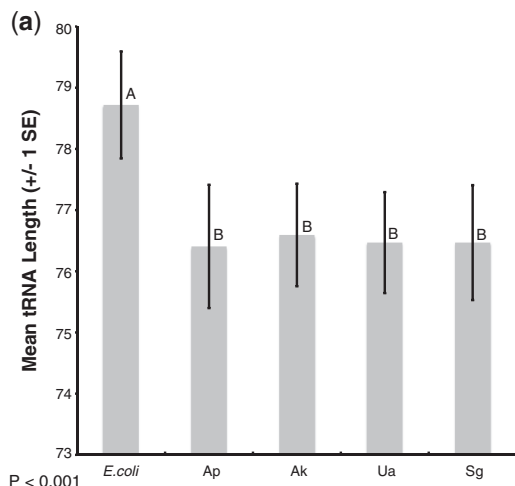
**Figure 4.** Reduction in stability of secondary structure and %GC in tRNAs of *Buchnera* relative to *E. coli*. *Buchnera* strains = (Ap-5A, Ak, Ua, Sg) (a) Two tRNAs in *Buchnera* with the lowest secondary structure stability relative to *E. coli* based on  $\Delta G$ . Red arrows indicate *Buchnera* base substitutions relative to *E. coli* in stem structures. White bases indicate compensatory substitutions in *Buchnera* and circled white bases indicate single base substitutions relative to *E. coli*. Black bases indicate conserved bases among all bacteria strains. (b) Mean tRNA  $\Delta G$  and (c) mean % GC for all 32 *Buchnera* tRNAs per strain and their corresponding tRNA homologs in *E. coli*.

**Table 5.** Total number of pairwise compensatory base changes (upper right) and percent of compensatory base changes with C/G to T/A (lower left) for 32 homologous tRNAs of *Buchnera* strains and *E. coli*

	Ap-5A	Ak	Ua	Sg	<i>E. coli</i>
Ap-5A	—	0	3	2	37
Ak	—	—	3	3	41
Ua	67%	67%	—	3	42
Sg	0%	33%	100%	—	41
<i>E. coli</i>	65%	66%	67%	63%	—

Some *Buchnera* tRNA genes with 3' CCA encoded also displayed CCA 3' maturation [Figure 5 (b)], resulting in double or triple CCA at the 3' end of tRNAs. Recently, it was shown that tRNAs with dual 3' CCA are targeted for degradation (67). More specifically, if a tRNA has 5'

Gs on bp 1 and 2, and its acceptor stem is structurally unstable, then the CCA-adding enzyme marks unstable tRNAs by adding dual 3' CCAs, targeting it for degradation by RnaseR (67). Such degradation also seems possible in *Buchnera* strains, which encode both the CCA-adding enzyme and RnaseR. Thus, we examined all *Buchnera* tRNAs with dual and triple 3' CCA maturation. First, we noted that all *E. coli* tRNAs with a 5' G at the 1st and 2nd base position encode dual or triple CCA on the 3' end of the tRNA gene [Figure 5 (c)]. Based on tRNAscan-SE 1.21, the penultimate CCA is always incorporated into the 3' acceptor stem, exposing a single 3' CCA for activation. Most *Buchnera* tRNAs that display dual or triple 3' CCA maturation still retain a 5' G at the 1st and 2nd bases and are homologs to dual or triple 3' CCA encoded *E. coli* tRNAs [Figure 5 (c)]. Three strain-specific tRNAs with dual 3' CCA maturation do not have *E. coli* homologs with dual CCAs encoded. These *Buchnera* tRNAs also do not encode 5' Gs at the



(c) CCA = genome encoded CCA at the 3' end  
 CCA = evidence of CCA maturation at 3' end

Encodes a G at the 1st and 2nd bases on 5' end      \* Second to last CCA not incorporated in 3' stem  
 Does not encode a G at the 1st and 2nd bases on 5' end      - - - Homolog not present

	<i>E.coli</i>	Ap-5A	Ap-LSR1	Ak	Ua	Sg
Leu TAA	CCACCA	CCACCA	CCACCA	CCACCA *	CCACCA	CCACCA
Ala GGC	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA
Ala TGC	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA
Gln TTG			CCACCA		CCACCA	
His GTG			CCACCA		CCACCA	CCACCA
Lys TTT	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA
Met CAT	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA
Ile CAT(y)	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA
Ser GGA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA
Tyr GGA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA
Tyr GTA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA
Val TAC	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA	CCACCA
Ile CAT(x)	CCACCA	- - -	- - -	- - -	- - -	- - -

**Figure 5.** Evidence of CCA 3' tRNA maturation in *Buchnera* (strains Ap-5A, Ap-LSR1, Ak, Ua, Sg). (a) Mean tRNA length for all 32 *Buchnera* tRNAs per strain and their corresponding tRNA homologs in *E. coli* (b) Presence and absence of 3' tRNA maturation in 86 *E. coli* tRNAs and all 32 *Buchnera* tRNAs for each strain. Grey numbers in CCA encoded black bars indicate the number of tRNA genes that both encode CCA and process an additional CCA during 3' tRNA maturation based on directional RNAseq (c) Pattern of dual or triple CCA 3' transcription in tRNAs among *E. coli* and *Buchnera* strains. tRNA isoacceptors in the left column correspond to the grey numbers in the black bars in panel ii for *Buchnera* strains. In both *E. coli* and *Buchnera*, the second to last CCA is always incorporated in the 3' stem except for one tRNA in Ak (asterisk).

1st and 2nd base. All *Buchnera* with dual or triple 3' CCA maturation incorporate the 2nd to last CCA into the 3' acceptor stem as in *E. coli*, except for one case, tRNA Leu<sup>TAA</sup> in Ak [Figure 5 (c)].

## DISCUSSION

The efficiency and fidelity of translation is reinforced by many mechanisms encoded in genomes. In reduced genomes, mutation rates are typically high, and selection becomes less effective in maintaining translational mechanisms. In this study, we found that bacterial endosymbiont lineages (*Buchnera*) that experience relaxed selection display less optimal tRNA characteristics relative to those of their free-living relative *E. coli*. Gene loss and A+T mutational bias in *Buchnera* have led to the loss of tRNA isoacceptors and loss of modified base pathways, the reduction of tRNA gene length, and the accumulation of base substitutions and indels (insertions/ deletions) in tRNA sequences that weaken tRNA secondary structure and possibly aminoacyl-tRNA synthetase recognition. These tRNA characteristics are conserved across four *Buchnera* lineages spanning 70 million years of divergence and may result in reduced translational efficiency and fidelity relative to their ancestors. However, we did detect compensatory base substitutions in *Buchnera* tRNAs, which are expected to maintain secondary structure of tRNA stem regions. Additionally, RNAseq reads reveal novel 3' maturation processes that compensate for tRNA gene length reduction.

Divergent *Buchnera* taxa in this study encode and express the same 32 tRNA genes composed of 32 different isoacceptor types (Figure 1). In turn, no duplication of tRNA gene isoacceptors was found. Based on a survey of 50 eukaryotic, eubacterial, and archaeal genomes, low tRNA gene redundancy (i.e. only one or two gene copies of a particular isoacceptor) was only found in all archaeans and several bacterial genomes, and was approximately correlated with genome size (40). In *Buchnera*, because of modified wobble rules (50,51), all mature tRNAs expressed can theoretically base pair with the 61 possible codons (Table 3, Figure 1), which are all still encoded in *Buchnera* CDS. One special *Buchnera* isoacceptor that has been identified previously in *Buchnera*-Ap (taxa type strain APS) is tRNA Ile<sup>CAU</sup> (40), where 5'-C is modified into lysidine by the enzyme TilS in *E. coli* (55), which all *Buchnera* strains still encode. This special Ile<sup>CAU</sup> isoacceptor codes for Ile instead of Met due to a wobble modification, and is ubiquitous in Eubacteria and Archaea (40).

During genome reduction, *Buchnera* has preferentially lost 5'-CNN, and to a lesser extent, 5'-GNN anticodons in family boxes and 5' CNN anticodons from two-codon NNR families (Table 3). This pattern of tRNA isoacceptor loss is common for many bacteria with reduced genomes (Figure 2), and is most likely related to gene deletion processes. Selective loss of these specific isoacceptors in family boxes and NNR two-codon families in Eubacteria was observed in previous studies (1,40,68,69) but was related to A+T sequence bias not deletion processes (1,70). We hypothesize that genome reduction, which is correlated with A+T bias, is the most likely explanation for this pattern of tRNA isoacceptor loss. First, the potential for wobble in codon-anticodon basepairing implies that some tRNA isoacceptors are not essential for pairing with

corresponding codons (e.g. 5'-CNN, 5'-GNN anticodons) and can be eliminated through mutation and deletion. Second, due to wobble rules, 5'-GNN anticodons followed by 5'-UNN anticodons are the most promiscuous isoacceptors when pairing with cognate codons; thus, it is not surprising that 5'UNN is always retained in family box and two-box NNR codons in the most reduced genomes. In turn, 5'-UNN anticodons are probably retained because of their ability to recognize alternative codons rather than because of the high frequency of cognate codons in A+T rich CDS. Typically in bacteria and eukaryotes 5'-CNN and 5'-GNN anticodons of family boxes and 5'-CNN anticodons from two-codon families along with 5' U anticodon modifications extending wobble are maintained by selection, because they increase the efficiency of translation (1,71). We predict that the loss of tRNA isoacceptors in *Buchnera* as well as other endosymbionts potentially results in less efficient translation.

Numerous unmodified nucleotides at specific nucleotide positions on tRNA isoacceptors are conserved phylogenetically and are known to play crucial roles in defining tRNA specificity for aminoacylation (32,72). These conserved nucleotides are called identity elements and are required for proper recognition by the cognate aaRS in addition to playing roles as deterrents to false recognition (32). Our results reveal that most *Buchnera* tRNAs have maintained identity elements homologous to those in *E. coli*, with the exceptions of Cys<sup>GCA</sup>, Ser<sup>GGA</sup>, Ser<sup>GCT</sup> and Ala<sup>GGC</sup>. In *E. coli* tRNA<sup>cys</sup>, the identity elements G15•G48 form an unusual tertiary base pair called a Levitt pair (73). Additionally, the *E. coli* identity elements A13•A22 are important in determining the structure of G15•G48 (74). Collectively, these *E. coli* identity elements are required for CysRS recognition due to their role in RNA tertiary structure (73). In all, *Buchnera* taxa, tRNA<sup>cys</sup> G15•G48 has mutated to U15•G48 and A13•A22 has mutated to G13•A22. Hou *et al.* (73) found that when G15•G48 is mutated to U15•G48, its backbone configuration is similar to the wild type tRNA<sup>cys</sup>; however, only partial aminoacylation (46.2%) occurs relative to the wild type. How both types of changes in identity element together affect tertiary structure is unknown.

In *Buchnera* tRNA Ala<sup>GGC</sup>, the identity element G20 is mutated to U20 in strains 5A and Ua and to C20 in Ak and Sg. In *E. coli* tRNA Ala<sup>GGC</sup>, these same base changes were shown to result in 6× and 50× reductions in alanine charging activity, respectively, relative to native tRNA Ala<sup>GGC</sup> (75). *Buchnera* Ala<sup>GGC</sup> does not possess this mutation. Potentially, if this mutation is deleterious in Ala<sup>GGC</sup> recognition, Ala<sup>GGC</sup> can wobble to all four alternative codons for the family box codon family for alanine. Interestingly, the smallest sequenced genome of *Buchnera*, for the host *Cinara cedri*, retains the same tRNA isoacceptors and aaRSs as other *Buchnera* taxa examined in this study; however, the Ala<sup>GGC</sup> tRNA gene has been lost, resulting in a total of only 31 tRNA genes.

In *Buchnera* tRNA Ser<sup>GGA</sup>, the identity element G73 (the discriminator base) has mutated to A73. Generally

a mutation in the discriminator base is known to result in the loss of cognate aminoacyl-tRNA synthetase recognition; however, Shimizu *et al.* (76) demonstrated that any four bases substituted in the discriminator base of *E. coli* Ser tRNA resulted in the same level of aminoacylation. Nevertheless, G73 in Ser tRNA is phylogenetically conserved (72) and has been shown to play minor roles in SerRS discrimination (77,78). Additionally, in *E. coli* Ser tRNA, the variable region plays a very important role as an identity element (77,79). In all *Buchnera* taxa, except Sg, the variable region length of the Ser<sup>GCT</sup> isoacceptor is 1 bp shorter than the *E. coli* Ser<sup>GCT</sup> isoacceptor. In summary, it is unknown how all these mutated identity elements affect *Buchnera* translation, but the same mutations in *E. coli* are known to significantly reduce the efficiency of aminoacylation.

In addition to requiring specificity in aminoacylation, reliable and efficient translation requires the anticodon to correctly pair with its codon. Modified nucleosides of tRNAs are essential mechanisms reinforcing translational fidelity and efficiency, especially at the wobble (N34) and 3' position immediately adjacent to the anticodon (P37), (1,2,51). Based on *E. coli* tRNA homologs, we expect 16 different types of modified bases to be present in the remaining 32 *Buchnera* tRNAs, for both N34 and N37 positions. In *E. coli*, 13 of these modified base pathways are known and *Buchnera* encodes complete pathways for six of these (Table 4). All *Buchnera* taxa have lost enzymes responsible for encoding N37 modified bases m<sup>2</sup>A and m<sup>6</sup>A, which are important in stabilizing 5'-NNC/G anticodons (2) (Table 4). Enzymes that synthesize the N37 modification m<sup>6</sup>t<sup>6</sup>A are conserved in only half of *Buchnera* taxa; this enzyme is known to slightly increase the efficiency of base pairing of the anticodon Thr<sup>GGU</sup> to the codon ACC in *E. coli* (54). All N37 modified base pathways important for preventing frameshifts and stabilizing A:U and U:A at the wobble position of the anticodon and the first position of the codon were retained in all *Buchnera* taxa (Table 4). These mechanisms may be essential for the fidelity of translation, especially for A+T rich genomes.

Modified nucleosides at the wobble base position (N34) of the anticodon are important for encoding the right amino acid, extending or restricting wobble, increasing the efficiency of base pairing and preventing frameshifts (2,53,55,56,58,59). *Buchnera* taxa all encode the enzyme TilS that is essential for the synthesis of the modified base lysidine, and is important for encoding the amino acid Ile instead of Met (59). All *Buchnera* taxa also encode the core enzymes MmmE and MnmG that are important for synthesizing the modified bases mnm<sup>5</sup>u, mnm<sup>5</sup>s<sup>2</sup>U, and cmnm<sup>5</sup>Um, which restrict 5'U wobble in NNR two-box codons, including Arg and Leucine (Table 4). All of these pathways are complete except for MmmC, which is involved in the last step for both modified bases, mnm<sup>5</sup>u and mnm<sup>5</sup>s<sup>2</sup>U. However, RNAseq mismatch evidence supports the presence of a modified base at the expected position of mnm<sup>5</sup>s<sup>2</sup>U (Table 4). Interestingly, the genes encoding MmmA, MmmE, MnmG, and IscS or SufS, but not MnmC are retained in several tiny endosymbiont genomes (10).

Conservation of these enzymes in reduced genomes indicates that these enzymes or derivatives are important for the production of the modified bases  $\text{mnm}^5\text{u}$ ,  $\text{cmnm}^5\text{Um}$ , and especially  $\text{mnm}^5\text{s}^2\text{U}$ , which is essential for preventing frameshifts and restricting wobble in NNR two codon boxes (Glu, Lys and Gln), thereby preventing the miscoding of amino acids. For incomplete pathways producing modified bases  $\text{mnm}^5\text{u}$  and  $\text{mnm}^5\text{s}^2\text{U}$ , either a derivative may be synthesized and/or the insect host may import MnmC. For example, the pea aphid, *A. pisum* expresses its *mnmC* homolog (XP\_003245837) in both its body and in the specialized aphid cells (bacteriocytes) that contain *Buchnera* cells (34).

Another key enzyme that is retained in *Buchnera* is TadA, which is responsible for synthesizing inosine in *E. coli* (55). This wobble modification is present on  $\text{Arg}^{\text{ACG}}$  in many bacteria and can wobble to three alternative codons of Arg (2,59). Rnaseq mismatch evidence highly supports this modification, as inosine is recognized as G during the reverse transcription process (60), and therefore we were able to measure a high frequency of modified  $\text{Arg}^{\text{ACG}}$  transcripts from all *Buchnera* taxa. Unfortunately, other modified bases do not appear to be recognized as specific bases and in turn incorporate different frequencies of any of the four bases during reverse transcription of modified transcripts (42,43). Collectively, Rnaseq evidence supported the presence of five modified bases, four in which the pathways are known and present (or near present for  $\text{mnm}^5\text{s}^2\text{U}$ ) and one in which the pathway is unknown (Table 4). If *Buchnera* tRNAs can be isolated without host contamination, modified base presence and identity can be confirmed.

In many bacterial species, tRNA abundances are positively correlated with codon usage for highly expressed genes, thus increasing translational efficiency (64,65). In addition to analysing specific tRNA characteristics that influence the accuracy and efficiency of translation, we examined whether codon usage correlates with tRNA expression. We found that tRNA sense expression is highly correlated across *Buchnera* taxa (Table 2), and many tRNA isoacceptors are expressed at similar levels within taxa (Figure 3, Supplementary Figure S1). A previous microarray study suggested that tRNA expression and codon usage of 50 highly expressed genes in *Buchnera*-Ap were positively correlated (37), but the relationship was weak and expression of sense and antisense tRNAs were not distinguished, possibly confounding results. Our directional RNaseq data show no relationship between tRNA expression and codon usage, for the same set of highly expressed genes in *Buchnera*-Ap under similar conditions (Figure 3). Furthermore, no relationship was detectable in three other *Buchnera* taxa (Supplementary Figure S1). Collectively, these results suggest that selection is not maintaining codon bias for highly expressed proteins. Interestingly,  $\text{Trp}^{\text{CCA}}$ , is the highest expressed isoacceptor in all *Buchnera* taxa (except Ua) and has very low codon usage. In all, *Buchnera* examined, isoacceptor  $\text{Trp}^{\text{CCA}}$  displays one of the lowest secondary structures relative to *E. coli*'s homolog; potentially  $\text{Trp}^{\text{CCA}}$  is highly expressed to compensate for low aminoacylation efficiency related to

numerous base substitutions that weaken its secondary structure [Figure 4(a)].

In this study, we found that *Buchnera* tRNAs have maintained high %GC relative to its CDS; however, its tRNAs are more A+T rich and less stable relative to homologs in *E. coli* (Figure 4). These results are consistent with previous findings (28) showing that 16S rRNAs of *Buchnera* and other endosymbiont species are more A+T rich and less stable than those of free-living relatives. Similarly, mitochondrial tRNAs from animals are more A+T rich and less stable than nuclear tRNAs (80). Collectively, these results suggest that the accumulation of deleterious mutations can lead to less stable secondary structures of essential RNAs involved in translation. Some selection for stabilization is also evident as numerous compensatory base substitutions have been fixed in the stem regions of both rRNAs (28) and tRNAs (Figure 5). Alternatively, *E. coli* tRNAs may possess higher %GC because its optimal growth temperature is higher than that of *Buchnera* (81), thus favoring higher %GC for increased thermal stability.

During genome reduction, 72–78% of *Buchnera* tRNA genes among all taxa have deleted 3 bp, due to the loss of 3' encoded CCA [Figure 5(a)]. Nevertheless, we found that all mature *Buchnera* tRNAs process 3' CCA, and therefore they all have potential for amino acid activation [Figure 5(b)]. In all *Buchnera* taxa, six to eight mature tRNAs process dual or triple 3' CCA [Figure 5(b)]. These characteristics, in addition to 5' G at the 1st and 2nd position and instability of the acceptor stem, result in tRNA degradation (67). Interestingly, these tRNAs in *Buchnera* and *E. coli* transcribe 5' G at the first and second base position and process dual or triple 3' CCA [Figure 5 (c)]. In these mature tRNAs in both *E. coli* and *Buchnera*, the second to last 3' CCA is always incorporated into the 3' acceptor stem. Potentially, the retention of encoded 5' G at N1 and N2 and the conservation of dual and triple 3' CCA maturation in these tRNAs [Figure 5 (c)] are essential to maintain the correct secondary structure and to police unstable tRNAs via the tRNA degradation pathway.

In conclusion, our observations of altered tRNA characteristics are consistent with the hypothesis that translational fidelity is lower in *Buchnera* compared with free-living relatives as represented by *E. coli*. First, *Buchnera* genome reduction has resulted in the loss of specific tRNA isoacceptors and modified nucleoside pathways that may reduce translational efficiency and fidelity. Second, *Buchnera*'s A+T mutational bias and reduced selection has resulted in the reduction of tRNA stability *in vitro* and specific tRNA base substitutions that may alter the efficiency of aaRS recognition. Moreover, reduced translational efficiency was supported by the lack of relationship between codon usage of highly expressed genes and cognate tRNA isoacceptor expression. Nevertheless, purifying selection appears to be strong enough in *Buchnera* genomes to maintain high %GC of tRNA genes relative to CDS. Also, CCA 3' maturation of shortened tRNA genes, and numerous compensatory base substitutions in tRNA stems help maintain tRNA secondary structure and function. Consequently, we predict that

the translational efficiency and fidelity evident in *Buchnera* are in an intermediate state between free-living bacteria and organelles.

## ACCESSION NUMBERS

All raw sense and anti-sense tRNA data were submitted to NCBI Genbank under SRA Submission: SRA049863.3, under Bioproject numbers: (i) PRJNA82811, (ii) PRJNA82809, (iii) PRJNA82797, (iv) PRJNA82793 and (v) PRJNA82789.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2, Supplementary Figure 1 and Supplementary Dataset 1.

## ACKNOWLEDGEMENTS

The authors thank Kim Hammond for rearing aphids and Dieter Söll, Jiqiang Ling, Patrick O'Donoghue and Markus Englert for helpful discussions and feedback on tRNA data. Also, they also thank Yogeshwar Kelkar, Rahul Raghavan and Patrick Degnan for helpful comments on the manuscript and thank four anonymous reviewers for their helpful comments and suggestions.

## FUNDING

Funding for open access charge: US Department of Agriculture [2011-67012-30707 to A.H.].

*Conflict of interest statement.* None declared.

## REFERENCES

- Osawa, S., Jukes, T.H., Watanabe, K. and Muto, A. (1992) Recent evidence for evolution of the genetic code. *Microbiol. Rev.*, **56**, 229–264.
- Yokoyama, S. and Nishimura, S. (1995) Modified nucleosides and codon recognition. In: Söll, D. and RajBhandary, U.L. (eds), *tRNA, Structure, Biosynthesis and function*. American Society of Microbiology, Washington, DC, USA, pp. 207–223.
- Martin, N.C. (1995) Organellar tRNAs: Biosynthesis and Function. In: Söll, D. and RajBhandary, U.L. (eds), *tRNA, Structure, Biosynthesis and function*. American Society of Microbiology, Washington, D.C., pp. 127–140.
- Watanabe, K. and Osawa, S. (1995) *tRNA Sequences and Variation in the Genetic Code*. American Society of Microbiology, Washington, DC, USA.
- Gray, M.W., Burger, G. and Lang, B.F. (1999) Mitochondrial evolution. *Science*, **283**, 1476–1481.
- Timmis, J.N., Ayliffe, M.A., Huang, C.Y. and Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.*, **5**, 123–135.
- Lynch, M. (2007) *The Origins of Genome Architecture*. Sinauer Associates, Inc., Sunderland, MA.
- Schneider, A. and Maréchal-Drouard, L. (2000) Mitochondrial tRNA import: are there distinct mechanisms? *Trends Cell Biol.*, **10**, 509–513.
- Mercer, T.R., Neph, S., Dinger, M.E., Crawford, J., Smith, M.A., Shearwood, A.M., Haugen, E., Bracken, C.P., Rackham, O., Stamatoyannopoulos, J.A. *et al.* (2011) The human mitochondrial transcriptome. *Cell*, **146**, 645–658.
- McCutcheon, J.P. and Moran, N.A. (2012) Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.*, **10**, 13–26.
- Theissen, U. and Martin, W. (2006) The difference between organelles and endosymbionts. *Curr. Biol.*, **16**, R1016–R1017.
- Keeling, P.J. and Archibald, J.M. (2008) Organelle evolution: what's in a name? *Curr. Biol.*, **18**, R345–R347.
- McCutcheon, J.P. (2010) The bacterial essence of tiny symbiont genomes. *Curr. Opin. Microbiol.*, **13**, 73–78.
- Schattner, P., Brooks, A.N. and Lowe, T.M. (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.*, **33**, W686–W689.
- Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
- McCutcheon, J.P., McDonald, B.R. and Moran, N.A. (2009) Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl Acad. Sci. USA*, **106**, 15394–15399.
- McCutcheon, J.P. and Moran, N.A. (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 million years of evolution. *Genome Biol. Evol.*, **2**, 708–718.
- Helm, M., Brulé, H., Friede, D., Giegé, R., Pütz, D. and Florentz, C. (2000) Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA*, **6**, 1356–1379.
- Moran, N.A. (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl Acad. Sci. USA*, **93**, 2873–2878.
- Mira, A. and Moran, N.A. (2002) Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb. Ecol.*, **44**, 137–143.
- Baumann, P., Baumann, L., Lai, C.-Y., Rouhbakhsh, D., Moran, N.A. and Clark, M.A. (1995) Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids. *Annu. Rev. Microbiol.*, **49**, 55–94.
- Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. and Ishikawa, H. (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, **407**, 81–86.
- Tamas, I., Klasson, L., Canback, B., Naslund, A.K., Eriksson, A.S., Wernegreen, J.J., Sandstrom, J.P., Moran, N.A. and Andersson, S.G. (2002) 50 million years of genomic stasis in endosymbiotic bacteria. *Science*, **296**, 2376–2379.
- Pérez-Brocá, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J.M., Silva, F.J., Moya, A. and Latorre, A. (2006) A small microbial genome: The end of a long symbiotic relationship? *Science*, **314**, 312–313.
- Moran, N.A., McLaughlin, H.J. and Sorek, R. (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science*, **323**, 379–382.
- Lamelas, A., Gosalbes, M.J., Manzano-Marin, A., Peretó, J., Moya, A. and Latorre, A. (2011) *Serratia symbiotica* from the aphid *Cinara cedri*: A missing link from facultative to obligate insect endosymbiont. *PLoS Genet.*, **7**, e1002357.
- Degnan, P.H., Ochman, H. and Moran, N.A. (2011) Sequence conservation and functional constraint on intergenic spacers in reduced genomes of the obligate symbiont *Buchnera*. *PLoS Genet.*, **7**, e1002252.
- Lambert, J.D. and Moran, N.A. (1998) Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. *Proc. Natl Acad. Sci. USA*, **95**, 4458–4462.
- Wilcox, J.L., Dunbar, H.E., Wolfinger, R.D. and Moran, N.A. (2003) Consequences of reductive evolution for gene expression in an obligate endosymbiont. *Mol. Microbiol.*, **48**, 1491–1500.
- Moran, N.A., Dunbar, H.E. and Wilcox, J.L. (2005) Regulation of transcription in a reduced bacterial genome: nutrient-provisioning genes of the obligate symbiont *Buchnera aphidicola*. *J. Bacteriol.*, **187**, 4229–4237.
- Tamas, I., Wernegreen, J.J., Nystedt, B., Kauppinen, S.N., Darby, A.C., Gomez-Valero, L., Lundin, D., Poole, A.M. and Andersson, S.G. (2008) Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. *Proc. Natl Acad. Sci. USA*, **105**, 14934–14939.

32. Giege, R., Sissler, M. and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.*, **26**, 5017–5035.
33. Moran, N.A., Degnan, P.H., Santos, S.R., Dunbar, H.E. and Ochman, H. (2005) The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proc. Natl Acad. Sci. USA*, **102**, 16919–16926.
34. Hansen, A.K. and Moran, N.A. (2011) Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc. Natl Acad. Sci. USA*, **108**, 2849–2854.
35. Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A. and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics*, **16**, 944–945.
36. Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
37. Charles, H., Calevro, F., Vinuelas, J., Fayard, J.M. and Rahbe, Y. (2006) Codon usage bias and tRNA over-expression in *Buchnera aphidicola* after aromatic amino acid nutritional stress on its host *Acyrtosiphon pisum*. *Nucleic Acids Res.*, **34**, 4583–4592.
38. Puigbo, P., Bravo, I.G. and Garcia-Vallve, S. (2008) E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics*, **9**, 65.
39. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
40. Marck, C. and 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.
41. Chan, P.P. and Lowe, T.M. (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.*, **37**, D93–D97.
42. Iida, K., Jin, H. and Zhu, J.-K. (2009) Bioinformatics analysis suggests base modifications of tRNAs and miRNAs in *Arabidopsis thaliana*. *BMC Genomics*, **10**, 155.
43. Findeiß, S., Langenberger, D., Stadler, P. and Hoffmann, S. (2011) Traces of post-transcriptional RNA modifications in deep sequencing data. *Biol. Chem.*, **392**, 305–313.
44. Dunin-Horkawicz, S., Czerwoniec, A., Gajda, M.J., Feder, M., Grosjean, H. and Bujnicki, J.M. (2006) MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res.*, **34**, D145–D149.
45. Keseler, I.M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muniz-Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T. et al. (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res.*, **39**, D583–D590.
46. Nawrocki, E.P., Kolbe, D.L. and Eddy, S.R. (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics*, **25**, 1335–1337.
47. Seibel, P., Muller, T., Dandekar, T., Schultz, J. and Wolf, M. (2006) 4SALE—A tool for synchronous RNA sequence and secondary structure alignment and editing. *BMC Bioinformatics*, **7**, 498.
48. Hofacker, I.L., Fekete, M. and Stadler, P.F. (2002) Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, **319**, 1059–1066.
49. Bernhart, S., Hofacker, I., Will, S., Gruber, A. and Stadler, P. (2008) RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics*, **9**, 474.
50. Crick, F.H. (1966) Codon—anticodon pairing: the wobble hypothesis. *J. Mol. Biol.*, **19**, 548–555.
51. Nishimura, S. (1997) *Transfer RNA: structure, properties and recognition*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
52. Wilson, R.K. and Roe, B.A. (1989) Presence of the hypermodified nucleotide N6-(delta 2-isopentenyl)-2-methylthioadenosine prevents codon misreading by *Escherichia coli* phenylalanyl-transfer RNA. *Proc. Natl Acad. Sci. USA*, **86**, 409–413.
53. Urbonavičius, J., Qian, Q., Durand, J.M., Hagervall, T.G. and Björk, G.R. (2001) Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.*, **20**, 4863–4873.
54. Qian, Q., Curran, J.F. and Björk, G.R. (1998) The methyl group of the N 6-methyl-N 6-threonylcarbamoyladenine in tRNA of *Escherichia coli* modestly improves the efficiency of the tRNA. *J. Bacteriol.*, **180**, 1808–1813.
55. Wolf, J., Gerber, A.P. and Keller, W. (2002) tadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *EMBO J.*, **21**, 3841–3851.
56. Spanjaard, R.A., Chen, K., Walker, J.R. and van Duin, J. (1990) Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: assigning a codon to argU tRNA and T4 tRNA<sup>Arg</sup>. *Nucleic Acids Res.*, **18**, 5031–5036.
57. Harada, F. and Nishimura, S. (1972) Possible anticodon sequences of tRNA<sup>His</sup>, tRNA<sup>Asn</sup>, and tRNA<sup>Asp</sup> from *Escherichia coli* B. Universal presence of nucleoside Q in the first position of the anticodon of three transfer ribonucleic acids. *Biochemistry*, **11**, 301–308.
58. Stern, L. and Schulman, L.H. (1978) The role of the minor base N4-acetylcytidine in the function of the *Escherichia coli* noninitiator methionine transfer RNA. *J. Biol. Chem.*, **253**, 6132–6139.
59. Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, S., Miyazawa, T. and Yokoyama, S. (1988) Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature*, **336**, 179–181.
60. Wittig, B. and Wittig, S. (1978) Reverse transcription of tRNA. *Nucleic Acids Res.*, **5**, 1165–1178.
61. Weissenbach, J. and Grosjean, H. (1981) Effect of threonylcarbamoyl modification (t6A) in yeast tRNA<sup>ArgIII</sup> on codon-anticodon and anticodon-anticodon interactions. *Eur. J. Biochem.*, **116**, 207–213.
62. Björk, G.R., Wikström, P.M. and Byström, A.S. (1989) Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. *Science*, **244**, 986–989.
63. Hagervall, T.G., Tuohy, T.M.F., Atkins, J.F. and Björk, G.R. (1993) Deficiency of 1-methylguanosine in tRNA from *Salmonella typhimurium* induces frameshifting by quadruplet translocation. *J. Mol. Biol.*, **232**, 756–765.
64. Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.*, **151**, 389–409.
65. Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, **2**, 13–34.
66. Mira, A., Ochman, H. and Moran, N.A. (2001) Deletional bias and the evolution of bacterial genomes. *Trends Genet.*, **17**, 589–596.
67. Wilusz, J.E., Whipple, J.M., Phizicky, E.M. and Sharp, P.A. (2011) tRNAs marked with CCACCA are targeted for degradation. *Science*, **334**, 817–821.
68. Osawa, S. and Jukes, T.H. (1988) Evolution of the genetic code as affected by anticodon content. *Trends Genet.*, **4**, 191–198.
69. Kano, A., Andachi, Y., Ohama, T. and Osawa, S. (1991) Novel anticodon composition of transfer RNAs in *Micrococcus luteus*, a bacterium with a high genomic G+C content: correlation with codon usage. *J. Mol. Biol.*, **221**, 387–401.
70. Andachi, Y., Yamao, F., Muto, A. and Osawa, S. (1989) Codon recognition patterns as deduced from sequences of the complete set of transfer RNA species in *Mycoplasma capricolum*: resemblance to mitochondria. *J. Mol. Biol.*, **209**, 37–54.
71. Rogalski, M., Karcher, D. and Bock, R. (2008) Superwobbling facilitates translation with reduced tRNA sets. *Nat. Struct. Mol. Biol.*, **15**, 192–198.
72. Sprinzl, M., Hartmann, T., Weber, J., Blank, J. and Zeidler, R. (1989) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **17**, R1–R172.
73. Hou, Y.M., Westhof, E. and Giegé, R. (1993) An unusual RNA tertiary interaction has a role for the specific aminoacylation of a transfer RNA. *Proc. Natl Acad. Sci. USA*, **90**, 6776–6780.
74. Hamann, C.S. and Hou, Y.-M. (1997) An RNA structural determinant for tRNA recognition. *Biochemistry*, **36**, 7967–7972.
75. Tamura, K., Asahara, H., Himeno, H., Hasegawa, T. and Shimizu, M. (1991) Identity elements of *Escherichia coli* tRNA(Ala). *J. Mol. Recognit.*, **4**, 129–132.

76. Shimizu, M., Asahara, H., Tamura, K., Hasegawa, T. and Himeno, H. (1992) The role of anticodon bases and the discriminator nucleotide in the recognition of some *E. coli* tRNAs by their aminoacyl-tRNA synthetases. *J. Mol. Evol.*, **35**, 436–443.
77. Himeno, H., Hasegawa, T., Ueda, T., Watanabe, K. and Shimizu, M. (1990) Conversion of aminoacylation specificity from tRNA<sup>Tyr</sup> to tRNA<sup>Ser</sup> in vitro. *Nucleic Acids Res.*, **18**, 6815–6819.
78. Rogers, M.J. and Söll, D. (1988) Discrimination between glutamyl-tRNA synthetase and seryl-tRNA synthetase involves nucleotides in the acceptor helix of tRNA. *Proc. Natl Acad. Sci. USA*, **85**, 6627–6631.
79. Sampson, J.R. and Saks, M.E. (1993) Contributions of discrete tRNA<sup>Ser</sup> domains to aminoacylation by *E. coli* seryl-tRNA synthetase: a kinetic analysis using model RNA substrates. *Nucleic Acids Res.*, **21**, 4467–4475.
80. Lynch, M. (1996) Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. *Mol. Biol. Evol.*, **13**, 209–220.
81. Dunbar, H., Wilson, A., Ferguson, N. and Moran, N. (2007) Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biol.*, **5**, e96.