Published online 8 September 2006 Nucleic Acids Research, 2006, Vol. 34, No. 16 4583–4592
doi:10.1093/nar/gkl597

Codon usage bias and tRNA over-expression in *Buchnera aphidicola* after aromatic amino acid nutritional stress on its host *Acyrthosiphon pisum*

Hubert Charles*, Federica Calevro, José Vinuelas, Jean-Michel Fayard and Yvan Rahbe

Laboratoire de Biologie Fonctionnelle Insectes et Interactions, UMR INRA/INSA de Lyon 203, Bâtiment Louis Pasteur, 69621 Villeurbanne Cedex, France

Received May 16, 2006; Revised July 12, 2006; Accepted August 1, 2006

ABSTRACT

Codon usage bias and relative abundances of tRNA isoacceptors were analysed in the obligate intracellular symbiotic bacterium, *Buchnera aphidicola* from the aphid *Acyrthosiphon pisum*, using a dedicated 35mer oligonucleotide microarray. *Buchnera* is archetypal of organisms living with minimal metabolic requirements and presents a reduced genome with high-evolutionary rate. Codon usage in *Buchnera* has been overcome by the high mutational bias towards AT bases. However, several lines of evidence for codon usage selection are given here. A significant correlation was found between tRNA relative abundances and codon composition of *Buchnera* genes. A significant codon usage bias was found for the choice of rare codons in *Buchnera*: C-ending codons are preferred in highly expressed genes, whereas G-ending codons are avoided. This bias is not explained by GC skew in the bacteria and might correspond to a selection for perfect matching between codon–anticodon pairs for some essential amino acids in *Buchnera* proteins. Nutritional stress applied to the aphid host induced a significant overexpression of most of the tRNA isoacceptors in bacteria. Although, molecular regulation of the tRNA operons in *Buchnera* was not investigated, a correlation between relative expression levels and organization in transcription unit was found in the genome of *Buchnera*.

INTRODUCTION

Associated with agricultural aphid pests, *Buchnera aphidicola* is one of the best-known symbiotic intracellular bacteria of insects. Aphids rely on *Buchnera* to support the rapid growth and reproductive potential that make them such versatile and notorious pests. *Buchnera* furnishes some vitamins and most essential amino acids that the aphid host cannot synthesize or find in sufficient quantities in plant phloem sap (1,2).

The *Buchnera* genome presents all the characteristics of a vertically transmitted intracellular bacterium (3): (i) a small size of 400–600 kb, depending on host species (4); (ii) a highly biased base composition towards A and T (5); and (iii) a high-evolutionary rate due to drastic bottlenecks and the absence of recombination, occurring in the population dynamics of the bacteria following transmission through host generations (6). During its intracellular evolution, *Buchnera* has conserved most of the genes encoding vital physiological functions for the symbiotic entity. As an example, enzymes of the essential amino acid biosynthesis pathways were conserved whereas most of the genes encoding proteins involved in non-essential amino acid biosynthesis (those the aphids would produce) are deleted. DNA replication and repair mechanisms have been partly deleted, as compared to its free-living relative *Escherichia coli*. The same observations can be made regarding the transcription machinery (i.e. lack of many transcriptional regulators, as well as recognizable terminator and promoter sequences). Such genes and structures have been lost during the genome shrinkage of *Buchnera*, probably early in the intracellular life style of the bacteria (7,8).

As generally in all bacteria, protein translation machinery is also highly conserved in *Buchnera*. Hence, Koonin (9), based on comparative genomics, determined a 'consensus' prokaryotic gene set composed of ∼60 proteins, primarily those involved in translation functions. Genome comparison between *Buchnera* and *E.coli*, using the KEGG database (http://www.genome.jp/kegg/), reveals that all of the 55 ribosomal proteins, the 12 translation factors and the 21 tRNA amino-acyl transferases found in *E.coli* are also present in *Buchnera*. It is noticeable that rRNA genes are present as a single copy in *Buchnera*, whereas seven copies are present in *E.coli*. This point is generally interpreted as a property of slow-growing organisms (10,11).

The tRNA set of *Buchnera* is composed of 32 isoacceptors and is almost minimal, being only surpassed in compactness by 8 organisms of the TIGR CMR (http://www.tigr.org), all of them from small genome Mollicutes (*Mycoplasma*), and

*To whom correspondence should be addressed. Tel: +33 4 72 43 80 85; Fax: +33 4 72 43 85 34; Email: hubert.charles@insa-lyon.fr

© 2006 The Author(s).
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1. Codon usage and tRNA isoacceptors in the genome of *B.aphidicola*

	T				C				A				G			
	cod[a]	AA	RCSU[b]	N	cod	AA	RCSU	N	cod	AA	RCSU	N	cod	AA	RCSU	N
T	TTT	Phe	**1.83**	8662	TCT	Ser	**2.56**	5821	TAT	Tyr	**1.73**	5882	TGT	Cys	**1.64**	1874
	TTC		0.17	814	**TCC**		0.25	570	**TAC**		0.27	910	**TGC**[c]		0.36	409
	TTA	Leu	**3.87**	12 029	**TCA**		**1.49**	3388	TAA	STOP	2.27	433	TGA	STOP Trp	0.34	65
	TTG		0.62	1932	TCG		0.21	468	TAG		0.39	75	**TGG**		1	1701
C	CTT	Leu	0.73	2251	CCT	Pro	**1.89**	2678	CAT	His	**1.75**	3470	**CGT**	Arg	**2**	2384
	CTC		0.1	320	CCC		0.29	404	**CAC**		0.25	504	CGC		0.29	342
	CTA		0.56	1727	**CCA**		**1.52**	2144	**CAA**	Gln	**1.74**	5226	CGA	Arg	0.95	1128
	CTG		0.12	370	CCG		0.3	423	CAG		0.26	790	**CGG**		0.09	102
A	ATT	Ile	**1.65**	11 901	ACT	Thr	**1.82**	3900	AAT	Asn	**1.72**	11 599	AGT	Ser	**1.23**	2801
	ATC		0.25	1783	**ACC**		0.26	551	**AAC**		0.28	1864	**AGC**		0.26	596
	ATA	Ile	**1.11**	8013	**ACA**		**1.71**	3674	**AAA**	Lys	**1.84**	17 059	**AGA**	Arg	**2.5**	2962
	ATG	Met	1	4023	ACG		0.21	453	AAG		0.16	1514	AGG		0.18	211
G	GTT	Val	**1.8**	4102	GCT	Ala	**1.76**	3688	GAT	Asp	**1.76**	7153	GGT	Gly	**1.67**	4244
	GTC		0.28	640	**GCC**		0.25	524	**GAC**		0.24	966	**GGC**		0.3	782
	GTA		**1.61**	3665	**GCA**		**1.72**	3618	**GAA**	Glu	**1.82**	9417	**GGA**		**1.78**	4529
	GTG		0.32	726	GCG		0.27	572	GAG		0.18	942	GGG		0.25	643

[a]Perfect match codons are bold-underlined (codons with a corresponding perfect match anticodon-harbouring isoacceptor tRNA in *Buchnera*).
[b]RSCU, relative synonymous codon usage (ΣRSCU = *n*, the degeneracy level for each amino acid); boldfaced figures are prefered codons (RCSU > 1).
[c]BU330 described as pseudo-tRNA (GCA anticodon) in the automated annotation of TIGR-CMR.

the euryarcheota *Methanopyrus kandleri*. *Mycoplasma mobile* is the only bacterium with a putative remnant set of 28 tRNA genes in its genome (11), including just two tRNA-Met.

When analysing the codon distribution of the tRNA genes present in the *Buchnera* genomes (Table 1), it seems clear that the occurrence of genes is almost optimized to fulfil *a minima* the standard wobble rule (encoded anticodons starting with G or U, i.e. codons ending with C or A). The charging of proline and isoleucine through a single tRNA implies, for these two amino acids (and arginine), that a part of the tRNA population is modified with non-standard bases to ensure a proper recognition of the three non-perfect match codons. For arginine and proline, it is noteworthy that the encoded anticodons do not follow the general rule, and that the tRNAs matching the CGC and CCC codons were avoided, in contrast to all the other C-terminated codons, which have their matching tRNAs. This situation is not unique among small genomes, as many other bacteria displayed very similar sets of tRNA-Arg, including *Wolbachia pipientis*, *Wiggelsworthia glossinidia brevipalpis* and *Blochmania floridanus*.

The frequency of alternative synonymous codons varies among bacterial species and often also among genes from a single genome. Factors influencing codon usage in bacteria are the GC skew (i.e. differential mutation rate between lagging and leading strand) (12), horizontal gene transfer between bacteria exhibiting different codon usage bias (13) and selection for translation efficiency. The latter hypothesis was formulated first in *E.coli*. Ikemura (14,15) showed that tRNA abundances are correlated with codon usage: highly expressed isoacceptors are those corresponding to the more frequently used codons (rule 1). This correlation has also been found in a few other bacteria [for a review see (11)]. In contrast, selection for codon usage is often inefficient in bacteria with extreme GC content (16–18). In the case of *Buchnera*, several authors have claimed the absence of selection for codon usage (19,20).

As each isoacceptor recognizes several codons, additional rules are needed to fully explain codon choice in bacteria [listed in (21): rule 2: modified uridines such as thiolated uridine and 5-carboxymethyluridine at the wobble position produce a preference for A over G at the codon third position; rule 3: an inosine at the anticodon wobble position produces a preference for T or C over A at the third position; rule 4: in two-codon sets of the (A/T)-(A/T)-pyrimidine type, C rather than T at the third position promotes an optimal interaction strength between codon and anticodon (the 'good choice' codon rule of Grosjean *et al*. (22) and Grantham *et al*. (23)].

Regulation of rRNA and tRNA expression is complex in bacteria. In *E.coli*, rRNA and tRNA operons are negatively regulated by the free nontranslating ribosomes (24). However, the regulation mechanism is unclear and probably occurs during the four following steps: transcription, processing of precursors, degradation of precursors and degradation of mature tRNA (25). Recently, it has been shown that tRNA levels corresponding to major isoacceptors increase while those corresponding to minor isoacceptors decrease when the bacterial growth rate increases (26,27). Dittmar *et al*. (25) also demonstrated, in *Bacillus subtilis*, that the relative abundances of tRNA were higher when the bacteria were placed in nutritional conditions producing high growth rates.

In *Buchnera*, the 32 tRNA genes are organized into 18 groups of contiguous sequences (dashed lines in the Table 2). Two of them include the rRNA genes *rrs* (BU243) and *rrf*-*rrl* (BU490-BU491), and the associated tRNAs are highly expressed. As a comparison, *E.coli* K-12 has 88 tRNA genes organized in 40 transcription units, of which 7 are associated with rRNA and are highly expressed. Global transcriptional regulators are known to be involved in the regulation of tRNA expression (28). Among them, the histone-like proteins FIS and H-NS are preponderant. Such regulators are conserved in the genome of *Buchnera* but their role in tRNA regulation has not yet been demonstrated.

This work addresses the following two questions: (i) Is codon usage only driven by mutational bias in *Buchnera*, as is generally claimed? and (ii) To what extent is the regulation of tRNA expression functional in the degenerate genome of *Buchnera*? tRNA relative abundances were measured with

Table 2. Chromosome location and isoacceptor tRNA expression in *B.aphidicola* from *A.pisum* reared on YF^0 and control medium (AP3)

Labels[a]	nameS	g-A	M	FC	Up/down
BU017	tRNA-Phe-GAA	12.53	−0.175	0.885	0
BU041	tRNA-Thr-GGU	10.84	−0.030	0.979	0
BU042	tRNA-Gly-GCC	12.61	0.646	1.565	+
BU043	tRNA-Tyr-GUA	15.40	0.244	1.184	0
BU044	tRNA-Thr-UGU	16.09	0.406	1.325	+*
BU068	tRNA-Lys-UUU	11.76	0.154	1.113	0
BU069	tRNA-Val-GAC	13.44	0.231	1.173	0
BU071	tRNA-Ala-GGC	14.25	−0.028	0.980	0
BU111	tRNA-Val-UAC	11.85	0.176	1.130	0
BU244	tRNA-Ile-GAU	28.08	0.490	1.405	+
BU245	tRNA-Ala-UGC	15.54	0.026	1.018	0
BU249	tRNA-Asp-GUC	12.77	0.013	1.009	0
BU329	tRNA-Leu-GAG	15.95	0.851	1.804	+**
BU330	tRNA-Cys-PseudoGCA	15.31	0.258	1.200	+*
BU331	tRNA-Ser-GCU	10.72	0.934	1.911	+
BU379	tRNA-Leu-UAA	20.41	−0.276	0.825	−
BU405	tRNA-Ser-GGA	16.43	0.373	1.295	+
BU406	tRNA-Arg-ACG	12.77	0.115	1.083	0
BU412	tRNA-Gln-UUG	15.27	0.297	1.228	+
BU413	tRNA-Leu-UAG	15.19	0.510	1.424	+*
BU414	tRNA-Met-CAU	10.99	0.072	1.051	0
BU457	tRNA-Met-CAU	11.29	0.193	1.143	0
BU485	tRNA-Arg-CCG	10.63	0.404	1.324	+
BU492	tRNA-Glu-UUC	20.28	−0.456	0.728	−
BU540	tRNA-Ser-UGA	21.73	0.967	1.955	+
BU557	tRNA-Asn-GUU	25.57	−0.016	0.988	0
BU558	tRNA-Met-CAU	10.11	0.417	1.335	+
BU575	tRNA-Gly-UCC	14.26	0.271	1.207	+
BU593	tRNA-Pro-UGG	19.59	0.923	1.896	+
BU594	tRNA-His-GUG	8.78	0.179	1.132	0
BU595	tRNA-Arg-UCU	10.00	0.226	1.169	0*
BU601	tRNA-Trp-CCA	10.18	0.246	1.186	0

Expression levels (g-A) are given as g-normalized A values with $A = \log 2(\sqrt{YF^\circ \times AP3})$, Logratio are $M = \log 2(YF^\circ/AP3)$ and fold changes (FC) are 2^M. + and − are up- and down-regulated genes with fold changes greater than 1.2 or smaller than 0.83. Genes marked with an asterisk show M value statistically not equal to 0 (**P-value < 0.05, *P-value < 0.1). Dashed lines in the table separate non-contiguous genes on the chromosome.
[a]The three numerical digits represent chromosome locations (gene numbering from the origin of replication).

an oligonucleotide microarray dedicated to the *Buchnera* genome when the aphid *Acyrthosiphon pisum* diet is depleted of tyrosine and phenylalanine, two important amino acids for which *Buchnera* furnishes the phenolic nucleus to its host when supply is low. An analysis of the different rules for assigning optimal codons in *Buchnera* is presented.

MATERIALS AND METHODS

Aphids and artificial diets

A long-established parthenogenetic clone (LL01) of *A.pisum* (Harris) was maintained at 21°C with a 16 h light photoperiod. Alate viviparous adults, reared at low density on *Vicia faba* seedlings for 2 days, were allowed to lay progeny on young plants. Neonate aphids (aged 0–12 h) were used for all the experiments. All experiments were initiated by transferring 2-day-old aphids from plants to artificial diets. Two diet formulations were used in this study, differing only in their amino acid composition. Diet AP_3 was nutritionally optimized, and based on the total amino acid profiles of whole aphid tissues, as described by Febvay *et al*. (29). The YF^0 diet shared the omission of both phenylalanine and tyrosine from the AP_3 complete diet, and was designed to stimulate the biosynthetic activities of *Buchnera* for the important classes of essential amino acids harbouring the phenolic nucleus. For the microarray experiments, *Buchnera* were purified from the aphids as described by Charles and Ishikawa (30).

Microarray experiments

Protocols for RNA isolation and labelling, microarray manufacturing (including probe design), hybridization reaction and washing, microarray scanning and statistical analyses of data are available on the SITRANS Database under the project name 'faromat' (31). Details of RNA preparation are provided in Calevro *et al*. (32). tRNA relative and g-normalized expressions (see further) are given in Table 2. Microarray data are also available in the Array Express database (accession no. E-TABM-83). The complete physiological and metabolic analysis has been submitted elsewhere.

The procedure is described briefly below. Oligonucleotide probe sequences (35 bases) were defined using the software ROSO (33). RNA was purified using the Trizol method and indirectly labelled by incorporating aminoallyl-dUTP into reverse transcript cDNA. Cy3 and Cy5 fluorescent dyes were then coupled with the targets in a dye swap experimental design including 2 × 3 microarrays (Cy3-YF^0/Cy5-AP3 and Cy3-AP3/Cy5-YF^0).

QMT aldheyde slides (Interchim, Montluçon, France) were used and hybridizations were performed manually and automatically using a Ventana Discovery automated station. Manual hybridization took place under a glass cover slip (24 × 60 mm) in a humidified slide chamber (Proteigene, Saint Marcel, France) and incubated at 50°C overnight for 16 h. The slides were washed twice in 2× SSC/0.1% SDS buffer for 10 min at room temperature and four times in 0.1× SSC for 1 min at room temperature, rinsed in distilled water and dried by centrifugation. In the case of automatic hybridization, target solution was inserted under an oil droplet (mineral oil LCS) at 45°C for 8 h followed by several washes of variable stringencies in a Ventana Medical Systems hybridization apparatus (Ventana Inc., Tucson, AZ).

Scanning was performed with a GeneTAC LSIV scanner (Genomic Solutions, Huntingdon, UK). Signal intensity values for each spot (pixel median) were recorded and quality analysis was performed with the GenePix 4.0 image analysis software (Axon Instruments, Foster City, CA).

Normalization and statistical analysis of microarray data were performed with the Bioconductor libraries (http://www.bioconductor.org/) of the R software (http://www.rproject.org). Dye and slide normalizations were performed on an invariant set of genes that was determined a posteriori following the non-parametric methods of Tseng *et al*. (34). A printTip Loess normalization was then applied using the R library 'marray' developed by Dudoit *et al*. (35). Statistical tests applied on microarray data are modified *t*-test (36) and ANOVA analysis using the R library 'maanova' (37). Non-parametric tests (i.e. simulated Pearson's Chi-squared, median and Wilcoxon tests) were performed with the R software.

Microarray technique usually does not allow absolute expression measurements because the hybridization reaction between the probes and their targets is highly variable. However, in bacteria, the use of genomic DNA (g-normalization) offers the facility to calibrate absolute gene expressions (38). This calibration was used here to estimate the expressions of the 32 tRNA isoacceptors of *Buchnera*, all sharing similar thermodynamic properties (i.e. length and GC content). Genomic DNA was classically purified with phenol/chloroform/isoamyl alcohol (25:24:1) protocol, followed by an RNase A treatement (Sigma-Aldrich, St Quentin Fallavier, France). The labelling was performed using the Nick Translation kit (Amersham) with 5 μg genomic DNA and incorporating the dUTP-Cy3 fluorochrome (Amersham). Non-incorporated fluorescent dyes were eliminated by purification on AutoseqTM G-50 columns (Amersham). Obtained targets have been independently hybridized on four additional slides to estimate the relative hybridization rate of each *Buchnera* gene probe. As no gene repeat occurs in *Buchnera*, genomic DNA fluorescence signals (gF_g) should be homogeneous over the 617 *Buchnera* genes. Assuming this homogeneity, we calculated a calibration coefficient for each gene $K_g = 617 * gF_g / \sum_1^{617} gF_g$. g-Normalized tRNA expressions were then expressed as percentage values within the group of tRNA genes.

Expression datasets and codon usage indices

Highly and poorly expressed gene sets ($n = 50$ for each, Supplementary Table 1) were selected based on expression data obtained with the *Buchnera* microarray on AP3 diet. The highly expressed genes included 19 ribosomal proteins, 8 flagellar genes, *tuf*, *mopA*, *mopB* and several other genes coding for metabolic enzymes.

Codon usage indices (RSCU and N_c) were calculated using the codonw software developed by J. Peden (http://sourceforge.net/projects/codonw/). RSCU (relative synonymous codon usage) were computed by dividing the observed frequency for one codon by the frequency expected if all synonyms for that amino acid were used equally. Thus, RSCU values close to unity indicate a lack of bias. Differential RSCU were the ratio between RSCU obtained with the highly expressed genes versus RSCU of the poorly expressed genes. Comparisons were made using codon effectives (N_c, see below) and Pearson's Chi-squared tests.

N_c is the effective number of codons used by a gene, and varies between 20, when only one codon is used for each amino acid, and 61, when codons are randomly used for each amino acid. Expected values of N_c can be estimated under the assumption that codon bias is only due to mutational bias (GC3s) using the following formula:

$$\hat{N}_c = 2 + \mathrm{GC3s} + \left(\frac{29}{\mathrm{GC3s}^2 + (1 - \mathrm{GC3s})^2} \right).$$

RESULTS

Isoacceptors and codon usage

Boldface and underlined codons in Table 1 represent the perfect match codons for the 32 tRNA isoacceptors in *Buchnera*. The *Buchnera* tRNA repertory and comparison with other small genome bacteria have been described previously in Introduction. Codon usage is presented in Table 1 with the RSCU computed for each codon. It is very clear from the table that preferred codons in *Buchnera* are those ending with A or T. Codon usage in *Buchnera* is hence mainly governed by the mutational AT bias and most often the 'good choice' rule is violated.

tRNA expression and isoacceptor abundances

Unexpectedly, the expression of genes encoding tRNAs could only be quantified using manual hybridization. Figure 1C and D shows that when automated hybridization was performed tRNAs were systematically over-labelled with red fluorescence. The hypothesis of a direct interaction between Cy5 and tRNA can be rejected because when hybridization was performed manually (i.e. under a glass cover slip in a hybridization chamber, Material and Methods), the dye swap was observed for the tRNA genes (Figure 1A and B). We hypothesize that this artefact, which is specific to tRNA genes, is caused by an interaction between tRNAs and the mineral oil (LCS) used in Ventana automated hybridization. Indeed, the fluorescence of the oil at the same wavelength as Cy5 is described in the technical notes of the manufacturer. To explain this measurement artefact, secondary structures of tRNA molecules might be involved. However, examination of dye swap responses on the other *Buchnera* genes harbouring secondary structures (such as rRNA) did not show the same over-labelling problem (data not shown). As tRNAs are GC-rich sequences, the influence of GC content was tested, but again no significant effect was observed (data not shown). Specific properties of tRNAs, such as their link to amino acid, were suspected to be responsible for the artefact, but this possibility was not explored further here. Unexplained dye bias was also reported by Dittmar *et al*. (25) in the fluorophore-labelling reaction for some tRNA species.

g-Normalized expression levels for the 32 tRNA isoacceptors, corresponding to the control nutritional conditions for the host aphids (AP3 diet), are presented in Table 2 ('g-A' column). Figure 2 represents the correlation between g-normalized tRNA relative abundances and codon frequencies for the 50 most expressed genes of *Buchnera*. A significant correlation was observed ($r = 0.38$, $P = 0.02$), moreover removing the AT-rich codon (AAA) returns an higher correlation ($r = 0.61$, $P = 4 \times 10^{-4}$). Similar results were obtained with the whole set of *Buchnera* genes (data not shown). In *Buchnera*, the abundances of isoacceptors tRNA are therefore adjusted to their matching codon frequencies. These results confirm that rule 1 is respected in *Buchnera*. However, the expression of the $\mathrm{tRNA}_{\mathrm{Lys}}^{\mathrm{UUU}}$, which pairs the overused codons AAA (representing about 12% of all the *Buchnera* codons), has not been increased in the bacteria.

Codon usage bias and codon–anticodon pairing rules

Differential RSCU values were computed using the two sets of highly and poorly expressed genes in *Buchnera* (Table 3, column Hi/Lo). Significant differences in codon effectives (Chi-squared test) were found for the following amino acids: leucine, isoleucine, valine, tyrosine, histidine, asparagine and arginine. For these amino acids (except Ile and Val),

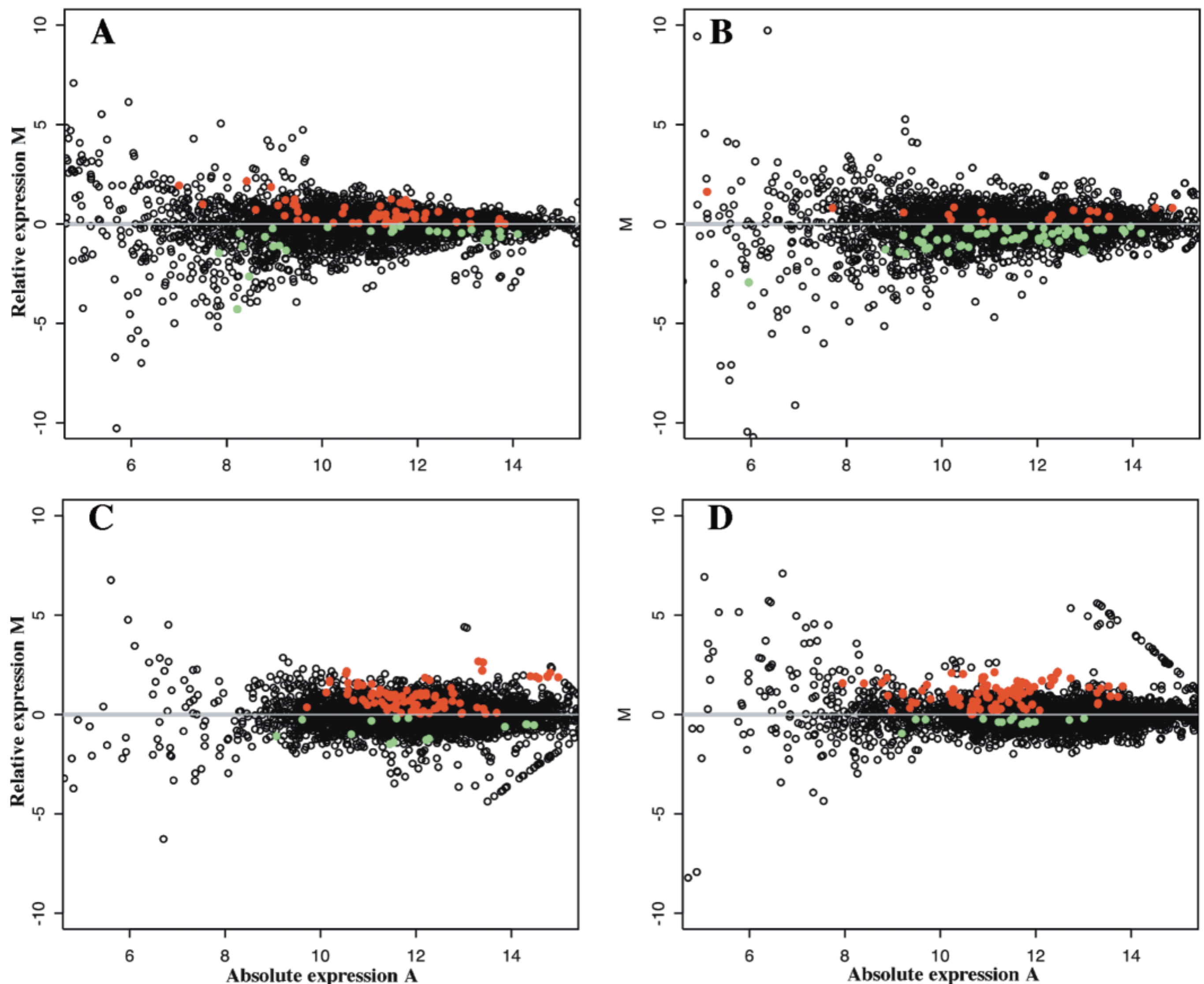

Figure 1. Normalized MA-plots for four slides in a dye swap experiment. (**A** and **B**) Manual hybridization; (**C** and **D**) automated hybridization. $A = \log 2\sqrt{R \times G}$ represents raw mean expression value. $M = \log 2(R/G)$ represents YF^0 versus AP3 logratio in (A) and (C) plots, whereas reverse hybridizations are found in (B) and (D) plots. tRNA genes are highlighted with red (over-expressed) and green (repressed) colours on graphs. Dye swap is clearly visible with manual hybridization, whereas in automated hybridization tRNA are always over-labelled in red (Cy5 fluorescence).

the highly expressed genes prefer C-ending codons as compared to the low expression group. It is also noteworthy that for all of the 16 C-ended codons in Table 3, the differential RSCU is above or equal to 1. This distribution is therefore highly biased (median test, $P = 10^{-5}$) and is partly explained by the optimal codon choice rule 4 for the amino acids Phe, Tyr, Asn, Asp and His. These C-ending codons are optimal for a majority of microorganisms (21). Although rare in *Buchnera* due to the AT bias, they could be selected for high expression due to their optimal pairing capacities.

Furthermore, the ATA Ile codon (requiring modified tRNA) is more rejected in the highly expressed genes. A similar rejection is observed for the arginine CGA codon which might be explained by the cost of the base modification required on the corresponding two tRNAs (rule 3).

Finally, G-ending codons are often rejected in highly expressed genes as compared to poorly expressed genes (with the exception of Leu, Ser, Thr and Val). This might be explained by rule 2 involving thiolated or carboxymethylated uridine facing the codon third position; however, the preference for A-ending codons is not clearly observable in our dataset.

The plot of the effective number of codons (N_c) versus GC content in the third position of synonymous codons (GC3s) is also well adapted to visualize codon usage bias (Figure 3). Gene distribution is nearly symmetrical around the expected curve (dashed line in Figure 3) revealing that codon usage is mainly directed by GC3s composition. However, the distribution of highly expressed genes is slightly biased toward small N_c values (codon usage bias) and higher GC3s, comforting

the hypothesis of C-ending preferred codons, revealed by the differential RSCU analysis.

Codon usage bias and GC skew

GC skew is high in *Buchnera* (http://pbil.univ-lyon1.fr/software/Oriloc/NC_002528.html) and a bias distribution of highly expressed genes on the leading strand might explain the observed codon usage bias (12). However, as the leading strand is more G+T rich, the C-ending codon preference in *Buchnera* should be attenuated (rather than accentuated) by strand distribution bias. Moreover, no significant differences were found when ratios of gene numbers located on leading strand versus lagging strand were compared for the two expression groups. Ratios were equal to 1.33 and 1.52 for the highly and poorly expressed genes, respectively. A more biased distribution, with numerous highly expressed genes located on the leading strand, might have been expected. However, this distribution is consistent with the work of Rocha and Danchin (39) showing, in different bacteria, that distribution of genes between the two replicating strands is driven by gene essentiality and not by gene expressivity.

To quantify GC skew influence on codon usage in *Buchnera*, the two datasets (highly and poorly expressed) were split following the strand localization of each gene, and differential RSCU analyses were performed within both strands. The C-ending preference was enhanced in the CT-rich lagging strand and attenuated in the AG-rich leading strand. However, the tendency was still significant in both groups (data not shown). The rejection of G-ending codons (except the GTG-Val) was also observed for the two strands.

The results presented here are contradictory with the work of Sharp *et al*. (11) who studied the strength of codon usage selection in 80 bacterial taxa. Based on a set of putative highly expressed genes mostly composed of ribosomal proteins, the authors observe that T-ending codons are unexpectedly preferred in *Buchnera* sp. and explain the bias by the GC skew in the bacteria. Two factors were probably misleading in this interpretation: (i) ribosomal protein genes are not the only and the highest expressed genes in *Buchnera* and (ii) no strand distribution asymmetry is observed with the highly expressed genes of *Buchnera*, whereas the essential ribosomal proteins are mainly located on the leading strand.

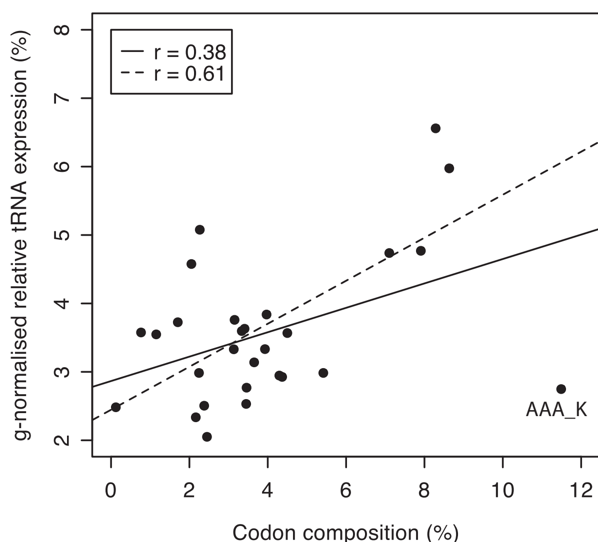

Figure 2. Plot of g-normalized tRNA relative abundances (expressed as percentage of total tRNA) versus matching codon compositions (i.e. the sums of codons associated to the corresponding tRNA species) of the 50 most expressed genes in *Buchnera*. Solid line, complete regression; dashed line, regression without the AAA-Lysine codon.

Regulation of tRNA gene expression

Depletion of the aromatic amino acids tyrosine and phenylalanine in aphid artificial diet results in up-regulation of most tRNA genes of *Buchnera* (Table 2, M column). Fold change measurements reveal positive values for 26 tRNA out of 32. This proportion is highly significant (median test,

Table 3. tRNA g-normalized expression and differential RSCU

	T cod[a]	AA	Exp[b]	Hi/Lo	C cod	AA	Exp	Hi/Lo	A cod	AA	Exp	Hi/Lo	G cod	AA	Exp	Hi/Lo
T	TTT	Phe	**12.53**	0.97[c]	TCT	Ser	—	0.98	TAT	Tyr*	—	0.94	TGT	Cys	—	1
	TTC			**1.35**	**TCC**		16.43	**1.25**	**TAC**		15.4	**1.42**	**TGC**[c]		15.31	1
	TTA	Leu*	20.41	0.97	**TCA**		21.73	0.99	TAA	STOP	—	0.92	TGA	STOP	—	1.35
	TTG		—	**0.77**	TCG		—	**0.56**	TAG		—	1.13	**TGG**	Trp	10.18	1
C	CTT	Leu*	—	**1.25**	CCT	Pro	—	0.9	CAT	His*	—	0.93	**CGT**	Arg*	12.77	**1.36**
	CTC		15.95	**1.55**	CCC		—	1.11	**CAC**		8.78	**1.6**	CGC		—	**1.35**
	CTA		15.19	1.1	**CCA**		19.59	1.1	**CAA**	Gln	15.27	1.01	CGA	Arg*	—	**0.87**
	CTG		—	1.1	CCG		—	1.08	CAG		—	0.93	**CGG**		10.63	**0.5**
A	ATT	Ile*	—	1.06	ACT	Thr	—	1.03	AAT	Asn*	—	0.96	AGT	Ser	—	1.04
	ATC		28.08	1.04	**ACC**		10.84	1.1	**AAC**		25.57	**1.28**	**AGC**		10.72	**1.34**
	ATA	Ile*	—	**0.89**	**ACA**		16.09	1.01	**AAA**	Lys	11.76	1	**AGA**	Arg*	10	**0.87**
	ATG	Met	10.81	1	ACG		—	**0.66**	AAG		—	1	AGG		—	**0.33**
G	GTT	Val*	—	0.98	GCT	Ala	—	0.91	GAT	Asp	—	0.97	GGT	Gly	—	0.99
	GTC		13.44	1	**GCC**		14.25	**1.26**	**GAC**		12.77	**1.23**	**GGC**		12.61	1.19
	GTA		11.85	0.93	**GCA**		15.54	1.04	**GAA**	Glu	20.28	1.01	**GGA**		20.28	0.98
	GTG		—	**1.71**	GCG		—	1.08	GAG		—	0.95	GGG		—	0.96

[a]Perfect match codons are bold-underlined (codons with a corresponding perfect match anticodon-harbouring isoacceptor tRNA in *Buchnera*).
[b]Expression measurement (g-normalized fluorescence) of the tRNA isoacceptor.
[c]Differential ratio of RSCU between the high and low expressed genes in *Buchnera* (boldface values are extreme values and asterisks are amino acids associated with significant Chi-squared test, $\alpha = 0.05$).

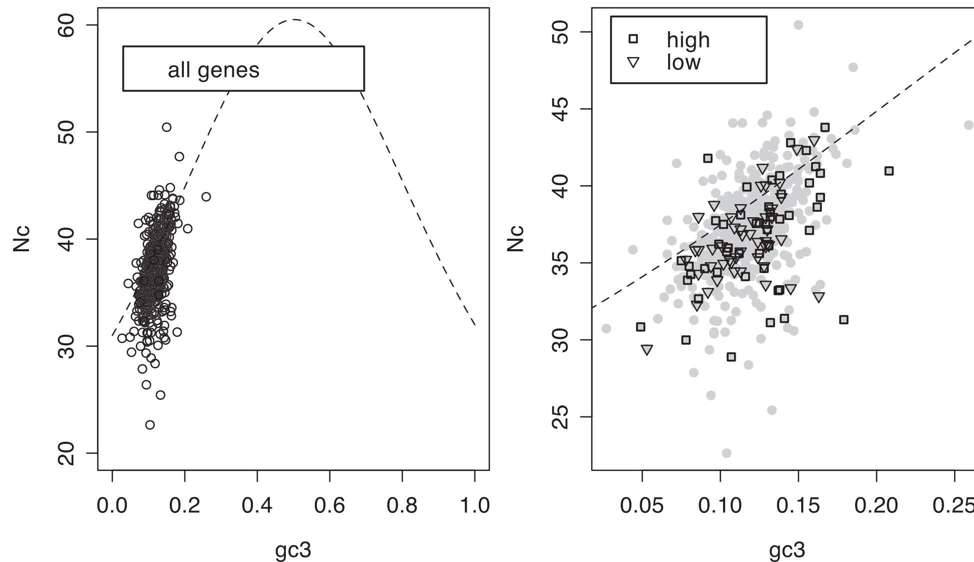

Figure 3. Effective number of codons used (N_c) plotted against GC3s (gc3) for 571 *Buchnera* genes. Dashed line represents the expected value of N_c if bias is only due to GC content; left panel: all *Buchnera* genes; right panel: open squares are the 50 genes with the highest expression rates and open triangles are the 50 genes with the lowest expression rates. Note that the square distribution is slightly decentred in the bottom and right area of the graph.

P-value $= 2 \times 10^{-4}$). If genes with fold changes close to 1 (i.e. >0.83 or <1.2) are grouped in an 'invariant class', the proportion of up-regulated genes remains significant (*P*-value $= 2 \times 10^{-3}$).

The metabolic analysis of the experiment is not detailed in this paper. However, we can mention that six other genes related to tRNA metabolism were differentially expressed in the YF^0 conditions. The genes *glnS* (Glutaminyl-tRNA synthetase), *gltX* (Glutamyl-tRNA synthetase) and *glyS* (Glycyl-tRNA synthetase beta chain) were up-regulated in YF^0 condition, whereas *pheT* (Phenylalanyl-tRNA synthetase beta chain) was significantly down-regulated on the aromatic-depleted medium. Two genes involved in tRNA modification, *trmE* and *trmD*, were both significantly induced in the YF^0 condition. Several ribosomal proteins were also regulated in the experiment. The TrmE protein is involved in the biosynthesis of the hypermodified nucleoside 5-methylaminomethyl-2-thiouridine, found in the wobble position on some tRNAs. Translation regulation efficiency might therefore be regulated via the modification of the wobble position of isoacceptor tRNAs. Overexpression of tRNA synthase genes is consistent with tRNA overexpression. Moreover, in the context of limiting aromatic amino acids, the specific repression of the *pheT* gene encoding the Phenylalanyl-tRNA synthetase might appear as a regulatory mechanism to increase the flux of the free phenylalanine exportation toward the host rather than incorporating this free amino acid for its own protein biosynthesis ('altruistic' response of endosymbiont). However, the latter two hypotheses are as yet speculative, and the mechanisms underlying this regulatory response are completely unknown.

Expression regulation of genes encoding tRNA isoacceptors is usually determined by their association to the different transcription units (e.g. rRNA operons). In *E.coli*, genome location of tRNA genes explains ~15% of their expression variation at a given growth rate (40). In *Buchnera*, tRNA genes are located on 18 different contiguous chromosome locations (pTU, putative transcription units) visualized by dashed lines in Table 2. Although not significant ($P = 0.12$), between-pTU variance of M values is always bigger than within-pTU variance (data not shown). To obtain more robust data, the M column was reduced to the up/down column of Table 2, and a contingency table was built between the two qualitative variables, 'pTU' and 'up/down'. A Pearson's Chi-squared test was then applied with simulated *P*-values to prevent bias of small effectives. A significant dependency was observed (*P*-value $= 0.05$) between the two variables, indicating that there is a tendency for tRNA belonging to the same transcription unit, and most probably controlled by the same promoter, to be regulated similarly.

DISCUSSION

To our knowledge, the results presented here are the first analysis of tRNA gene expression in *Buchnera*, the previously published microarray work having been developed on a cDNA-chip lacking this set of short genes (41). Also, this is the first time that real sets of high/low expression genes have been used in the same species, characterized by extreme GC content composition, to analyse the link between expressivity and codon usage biases. In *Buchnera*, previous studies either analysed *E.coli* codon adaptation of orthologous genes to detect high expression (42) or used limited sets of genes from different species to infer those of the *BuchneraAPS* set (20).

With only 32 molecular species, *Buchnera* set of tRNAs is almost minimal. Further reduction may have been possible, as was occurred in some mollicutes (2 tRNA-Met, 1 tRNA-Ala or 2 tRNA-Arg in *M.mobile*). Two reductions in tRNA numbers, observed in *Buchnera*, seem very widespread in small genomes, namely the charging of proline and isoleucine through a single tRNA. This implies, for these two amino acids (and for arginine), that a part of the tRNA population

is hypermodified to ensure proper codon–anticodon recognition. Hence, all the enzymes involved in the hypermodification of the tRNAs (*rnpA*, *pth*, *trmU*, *cca*, *trmD*, *gidA*, *trmE* and *truA*) are conserved in *Buchnera*, as well as in other intracellular endosymbionts, whereas they are not all essential in *E.coli* (43). Owing to reduction in tRNA sets, we propose that this set of genes is trully essential in small genomes, confirming their inclusion in the minimal cell gene set proposed by Gil *et al*. (43).

In bacteria, translational control of gene expression is partly regulated by changes in the charging of tRNA due to differential expression of single isoacceptors (44). Indeed during amino acid limitation, the concentration of the charged tRNAs corresponding to the most abundant codons tends to zero whereas concentrations of the other charged-tRNA will be unchanged. The theory of selective tRNA charging is the basis of the attenuation system described in *E.coli* (45), but in *Buchnera*, most if not all attenuation systems have been lost, and the reduction of tRNA species might be an adaptation to the metabolic constraints of the intracellular environment characterized by extreme limitations of several amino acids: reducing tRNA species and limiting codon usage would reduce the waste of free amino acids for charging useless tRNAs.

The first question we addressed in this study concerned the forces driving codon usage in *Buchnera*. Mutational bias has a major influence on codon choice in the bacteria and is independent of gene expressivity. However, a significant correlation between codon composition of *Buchnera* genes and tRNA production was observed with the exception of the isoacceptor $tRNA_{Lys}^{UUU}$. For the latter, adjustment of its expression has been overcome by the effect of mutational bias increasing drastically the frequencies of the corresponding lysine codon in proteins. These results indicate that growth rate is probably not the main factor governing evolution of *Buchnera*. Indeed, *Buchnera*'s environment is highly constrained by the metabolic demand from the aphid host. In that context, Rocha and Danchin (46) suggested that the higher energy cost and limited availability of G and C over A and T/U could explain the AT-enrichment of all obligate intracellular organisms including bacteria, viruses and plasmids. Such metabolic constraints are not restricted to the bacteria and occur on the whole symbiocosme (e.g. the associated organisms). Hence, the AT-enrichment of the pea-aphid genome observed by Sabater-Munoz *et al*. (47) might also be explained by the strict intracellular feeding of the insect.

In *Buchnera*, a significant codon usage bias was also observed for the choice of rare codons: C-ending codons are preferred for the majority of amino acids in highly expressed genes. There is experimental evidence that rare codons can reduce the rate of translation in *E.coli* (48,49). In organisms with efficient selection for translational efficiency, one might expect rejection of rare codons in highly expressed genes. It has been demonstrated in *Drosophila* that accuracy of translation is increased when using perfect match codon–anticodons pairs (50). Conservation of C-ending rare codons within essential or highly expressed genes might correspond to selection for translation efficiency of functionally important protein amino acids in *Buchnera* (C. Rispe, personal communication).

To test this hypothesis, we analysed the codon composition of the *groEL Buchnera* gene (one of the most highly expressed in the bacteria) and no clear evidence for selection was found. Of the 28 C-ending codons of the protein 3 were functionally important and highly conserved codons (51). When compared to the proportion of all the C-ending codon possibilities within the gene, the distribution was not significantly biased. Surprisingly, more C-ending codons accumulated within the positively selected regions of the protein described by Fares *et al*. (52) (P-value = 0.06). However, the distributions were not significantly biased in some other *Buchnera* species (e.g. those from the aphids *Schizaphis graminum* and *Baizongia pistacea*) and further studies are needed to analyse whether this codon bias is functional (selected) or whether it is a remnant trait derived from ancestral bacteria.

The second question was directed towards the tRNA transcriptional regulation capacities of the bacteria. We have shown here that *Buchnera* is able to respond to the variation in nutritional demand of the aphid host, even though most specific transcriptional regulators are absent from its genome. Overexpression of tRNA genes is generally correlated with increase of metabolic activities and cell division in bacteria. In our experiment, the biological parameters of aphids feeding on the depleted diet are good. Indeed, only larval growth was slightly affected (−20%), but neither survival nor fecundity is affected by the lack of phenylalanine and tyrosine (Supplementary Figure 1). These results are consistent with a global increased compensatory metabolism of *Buchnera* upon phenolic amino acid deficiency in the aphid diet and the correlation between differential gene expression and transcription unit organization in the chromosome of *Buchnera* may indicate a functional regulation activity in the bacteria.

In conclusion, this paper addresses the question of determinants that shape the evolution of codon usage, as well as tRNA isoacceptors abundance and diversity. Translation speed is optimized when tRNA abundances are adjusted to fit to codon usage of highly expressed genes for the highest growth rates in *E.coli* (53). However, as reviewed in Bailly-Bechet *et al*. (54), the efficiency of translation is probably not the only force driving codon usage and tRNA evolution. Metabolism limitation, proper folding of the nascent protein, horizontally transferred genes and local recycling of rare tRNAs are other factors that might influence selection of the translational machinery. These factors might be even more crucial for slow growers and intracellular bacteria (55). In the aphid-*Buchnera* association, the strong metabolic constraints affecting the two partners (i.e. both are intracellular feeders) have probably shaped the evolution of the translation machinery, explaining both the codon usage and the genome AT-enrichment.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank colleagues from the PRABI bioinformatics plateforme (http://www.prabi.fr/), J. Bernillon from the

Rhône-Alpes Genopole transcriptomic plateforme (DTAMB, Université Claude Bernard Lyon 1) and G. Duport for their support in this work. We are grateful to A. Douglas and J. Bermingham for critical reading and English corrections. Funding to pay the Open Access publication charges for this article was provided by Programme Fédérateur INRA de Biologie Intégrative, AgroBI 2006.

Conflict of interest statement. None declared.

REFERENCES

1. Liadouze,I., Febvay,G., Guillaud,J. and Bonnot,G. (1995) Effect of diet on the free amino acid pools of symbiotic and aposymbiotic pea aphids, *Acyrthosiphon pisum*. *J. Insect. Physiol*., **41**, 33–40.
2. Douglas,A.E. (1998) Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria *Buchnera*. *Annu. Rev. Entomol*., **43**, 17–37.
3. Shigenobu,S., Watanabe,H., Hattori,M., Sakaki,Y. and Ishikawa,H. (2000) Genome sequence of the endocellular bacterial symbiot of aphids *Buchnera sp. APS*. *Nature*, **407**, 81–86.
4. Gil,R., Sabater-Muñoz,B., Latorre,A., Silva,F. and Moya,A. (2002) Extreme genome reduction in *Buchnera spp*.: toward the minimal genome needed for symbiotic life. *Proc. Natl Acad. Sci. USA*, **99**, 4454–4458.
5. Ishikawa,H. (1987) Nucleotide composition and kinetic complexity of the genomic DNA of an intracellular symbiont in the pea aphid *Acyrthosiphon pisum*. *J. Mol. Evol*., **24**, 205–211.
6. Moran,N.A. (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl Acad. Sci. USA*, **93**, 2873–2878.
7. 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.
8. Moran,N.A. and Mira,A. (2001) The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol*., **2**, 54.1–54.12.
9. Koonin,E. (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature Rev. Microbiol*., **1**, 127–136.
10. Lambert,J.D. and Moran,N.A. (1998) Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. *Proc. Natl Acad. Sci. USA*, **95**, 4458–4462.
11. Sharp,P.M., Bailes,E., Grocock,R.J., Peden,J.F. and Sockett,R.E. (2005) Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res*., **33**, 1141–1153.
12. Lobry,J.R. (1996) Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol*., **13**, 660–665.
13. Ochman,H., Lawrence,J.G. and Groisman,E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature*, **405**, 299–304.
14. 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.
15. Ikemura,T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol*., **2**, 13–34.
16. Wright,F. and Bibb,M.J. (1992) Codon usage in the G+C-rich *Streptomyces* genome. *Gene*, **113**, 55–65.
17. Andersson,S.G. and Sharp,P.M. (1996) Codon usage and base composition in *Rickettsia prowazekii*. *J. Mol. Evol*., **42**, 525–536.
18. Kerr,A.R., Peden,J.F. and Sharp,P.M. (1997) Systematic base composition variation around the genome of *Mycoplasma genitalium*, but not *Mycoplasma pneumoniae*. *Mol. Microbiol*., **25**, 1177–1179.
19. Wernegreen,J.J. and Moran,N.A. (1999) Evidence for genetic drift in endosymbionts (*Buchnera*): analyses of protein-coding genes. *Mol. Biol. Evol*., **16**, 83–97.
20. Rispe,C., Delmotte,F., van Ham,R.C. and Moya,A. (2004) Mutational and selective pressures on codon and amino acid usage in *Buchnera*, endosymbiotic bacteria of aphids. *Genome Res*., **14**, 44–53.
21. Kanaya,S., Yamada,Y., Kudo,Y. and Ikemura,T. (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene*, **238**, 143–155.
22. Grosjean,H.J., de Henau,S. and Crothers,D.M. (1978) On the physical basis for ambiguity in genetic coding interactions. *Proc. Natl Acad. Sci. USA*, **75**, 610–614.
23. Grantham,R., Gautier,C., Gouy,M., Jacobzone,M. and Mercier,R. (1981) Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res*., **9**, r43–74.
24. Gourse,R.L., de Boer,H.A. and Nomura,M. (1986) DNA determinants of rRNA synthesis in *E.coli*: growth rate dependent regulation, feedback inhibition, upstream activation, antitermination. *Cell*, **44**, 197–205.
25. Dittmar,K., Mobley,E., Radek,A. and Pan,T. (2004) Exploring the regulation of tRNA distribution on the genomic slide. *J. Mol. Biol*., **337**, 31–47.
26. Emilsson,V., Naslund,A.K. and Kurland,C.G. (1993) Growth-rate-dependent accumulation of twelve tRNA species in *Escherichia coli*. *J. Mol. Biol*., **230**, 483–491.
27. Dong,H., Nilsson,L. and Kurland,C.G. (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol*., **260**, 649–663.
28. Dorman,C.J. and Deighan,P. (2003) Regulation of gene expression by histone-like proteins in bacteria. *Curr. Opin. Genet. Dev*., **13**, 179–184.
29. Febvay,G., Delobel,B. and Rahbé,Y. (1988) Influence of the amino acid balance on the improvement of an artificial diet for a biotype of *Acyrthosiphon pisum* (Homoptera:Aphididae). *Can. J. Zool*., **66**, 2449–2453.
30. Charles,H. and Ishikawa,H. (1999) Physical and genetic map of the genome of *Buchnera*, the primary endosymbiont of the pea aphid *Acyrthosiphon pisum*. *J. Mol. Evol*., **48**, 142–150.
31. Laforest,F., Tchounikine,A., Chaari,T., Charles,H. and Calevro,F. (2005) SITRANS: a web information system for microarray experiments. *Stud. Health Technol. Inform*., **116**, 33–38.
32. Calevro,F., Charles,H., Reymond,N., Dugas,V., Cloarec,J.-P., Bernillon,J., Rahbe,Y., Febvay,G. and Fayard,J.-M. (2004) Assessment of 35mer amino-modified oligonucleotide based microarray with bacterial samples. *J. Microbiol. Methods*, **57**, 207–218.
33. Reymond,N., Charles,H., Duret,L., Calevro,F., Beslon,G. and Fayard,J.-M. (2004) ROSO: optimizing oligonucleotide probes for microarrays. *Bioinformatics*, **20**, 271–273.
34. Tseng,G., Oh,M.-K., Rohlin,L., Liao,J. and Wong,W. (2001) Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessement of gene effects. *Nucleic Acids Res*., **29**, 2549–2557.
35. Dudoit,S., Yang,Y., Speed,T. and Callow,M. (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Stat. Sinica*, **12**, 111–139.
36. Tusher,V., Tibshirani,R. and Chu,G. (2001) Signifiance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA*, **98**, 5116–5121.
37. Wu,H., Kerr,K. and Cui,X. (2002) MAANOVA: a software package for the analysis of spotted cDNA microarray experiments. In Parmigiani,G., Garrett,E.S., Irizarry,R.A. and Zeger,S.L. (eds), *The Analysis of Gene Expression Data: Methods and Software*. Springer, NY, pp. 455.
38. Talaat,A.M., Howard,S.T., Hale,W.t., Lyons,R., Garner,H. and Johnston,S.A. (2002) Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. *Nucleic Acids Res*., **30**, e104.
39. Rocha,E.P. and Danchin,A. (2003) Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Res*., **31**, 6570–6577.
40. Ardell,D.H. and Kirsebom,L.A. (2005) The genomic pattern of tDNA operon expression in *E.coli*. *PLoS Comput. Biol*., **1**, e12.
41. 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.
42. Schaber,J., Rispe,C., Wernegreen,J., Buness,A., Delmotte,F., Silva,F.J. and Moya,A. (2005) Gene expression levels influence amino acid usage and evolutionary rates in endosymbiotic bacteria. *Gene*, **352**, 109–117.
43. Gil,R., Silva,F.J., Pereto,J. and Moya,A. (2004) Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev*., **68**, 518–537.

44. Sorensen,M.A., Elf,J., Bouakaz,E., Tenson,T., Sanyal,S., Bjork,G.R. and Ehrenberg,M. (2005) Over expression of a tRNA(Leu) isoacceptor changes charging pattern of leucine tRNAs and reveals new codon reading. *J. Mol. Biol.*, **354**, 16–24.
45. Yanofsky,C. (1981) Attenuation in the control of expression of bacterial operons. *Nature*, **289**, 751–758.
46. Rocha,E.P. and Danchin,A. (2002) Base composition bias might result from competition for metabolic resources. *Trends Genet.*, **18**, 291–294.
47. Sabater-Munoz,B., Legeai,F., Rispe,C., Bonhomme,J., Dearden,P., Dossat,C., Duclert,A., Gauthier,J.P., Ducray,D.G., Hunter,W. *et al.* (2006) Large-scale gene discovery in the pea aphid *Acyrthosiphon pisum* (Hemiptera). *Genome Biol.*, **7**, R21.
48. Sharp,P.M. and Li,W.H. (1986) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res.*, **14**, 7737–7749.
49. McNulty,D.E., Claffee,B.A., Huddleston,M.J., Porter,M.L., Cavnar,K.M. and Kane,J.F. (2003) Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein Exp. Purif.*, **27**, 365–374.
50. Akashi,H. (1994) Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics*, **136**, 927–935.
51. Fenton,W.A., Kashi,Y., Furtak,K. and Horwich,A.L. (1994) Residues in chaperonin GroEL required for polypeptide binding and release. *Nature*, **371**, 614–619.
52. Fares,M.A., Moya,A. and Barrio,E. (2005) Adaptive evolution in GroEL from distantly related endosymbiotic bacteria of insects. *J. Evol. Biol.*, **18**, 651–660.
53. Berg,O.G. and Kurland,C.G. (1997) Growth rate-optimised tRNA abundance and codon usage. *J. Mol. Biol.*, **270**, 544–550.
54. Bailly-Bechet,M., Danchin,A., Iqbal,M., Marsili,M. and Vergassola,M. (2006) Codon usage domains over bacterial chromosomes. *PLoS Comput. Biol.*, **2**, e37.
55. Rocha,E.P. (2004) Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res.*, **14**, 2279–2286.