J Mol Evol (2002) 55:127–137
DOI: 10.1007/s00239-001-2307-8

JOURNAL OF MOLECULAR EVOLUTION

Comparative Molecular Evolution of Primary (*Buchnera*) and Secondary Symbionts of Aphids Based on Two Protein-Coding Genes

Andrés Moya, Amparo Latorre, Beatriz Sabater-Muñoz, Francisco J. Silva

Instituto Cavanilles de Biodiversidad y Biología Evolutiva and Departamento de Genética, Universitat de València, P.O. Box 2085, E-46071, València, Spain

Received: 17 July 2001 / Accepted: 4 January 2002

Abstract. A + T content, phylogenetic relationships, codon usage, evolutionary rates, and ratio of synonymous versus non-synonymous substitutions have been studied in partial sequences of the *atpD* and *aroQ*/*pheA* genes of primary (*Buchnera*) and secondary symbionts of aphids and a set of selected non-symbiotic bacteria, belonging to the five subdivisions of the Proteobacteria. Compared to the homologous genes of the last group, both genes belonging to *Buchnera* behave in a similar way, showing a higher A + T content, forming a monophyletic group, a loss in codon bias, especially in third base position, an evolutionary acceleration and an increase in the number of non-synonymous substitutions, confirming previous results reported elsewhere for other genes. When available, these properties have been partly observed with the secondary symbionts, but with values that are intermediate between *Buchnera* and free living Proteobacteria. They show high A + T content, but not as high as *Buchnera*, a non-solved phylogenetic position between *Buchnera*, and the other γ-Proteobacteria, a loss in codon bias, again not as high as in *Buchnera* and a significant evolutionary acceleration in the case of the three *atpD* genes, but not when considering *aroQ*/*pheA* genes. These results give support to the hypothesis that they are symbionts at different stages of the symbiotic accommodation to the host.

Correspondence to: A. Moya; *email*: andres.moya@uv.es

Key words: Aphids — *Buchnera* — Primary endosymbiont — Secondary symbionts — A + T content — Codon bias — Relative rate tests — Synonymous versus non-synonymous substitution rates — Phylogenetic analyses

Introduction

Aphids are phloem-sap-sucking insects that balance their diet with the use of essential nutrients supplied by *Buchnera,* their primary endosymbiont. *Buchnera* is present in all the aphid species studied until now, with the exception of some species from the family Hormaphididae (Baumann et al. 1995) and it resides in specialized cells in the aphid hemocoel called bacteriocytes. It is maternally inherited through controlled infection of eggs and embryos (revised in Baumann et al. 2000). Phylogenetic analyses of *Buchnera* 16S rDNA sequences and aphid morphology are congruent, indicating an ancient infection followed by vertical transmission for 150–250 million years (Moran et al. 1993). Genetic studies have shown that *Buchnera* over expresses some essential amino acids (Baumann et al. 1999; Bracho et al. 1995; Jiménez et al. 2000; Lai et al. 1994; van Ham et al. 1997) that must be provided to the host (Douglas 1998). However, metabolic studies have shown that the contribution of essential amino acids is necessary but not sufficient for embryo growth (Wilkinson and Ishikawa 2000). The presence of a set of biosynthetic genes in *Buchnera* that are absent in intracellular

pathogenic bacteria, suggests that most of these genes may be involved in the provision of small molecules to the host (Shigenobu et al. 2000). This has been demonstrated at least in the case of the vitamin riboflavin (Nakabachi and Ishikawa 1997, 1999).

Besides *Buchnera* as primary simbiont, a number of aphids are known to contain other type of bacteria denoted as secondary symbionts (Buchner 1965; Fukatsu and Ishikawa 1993, 1998). The location, presence and pattern of inheritance of the secondary symbionts in aphids are much less studied and most of the current information has been obtained from those of *Acyrthosiphon pisum* and *Macrosiphum rosae* (Chen and Purcell 1997; Fukatsu et al. 1998; Fukatsu et al. 2000; Unterman et al. 1989). *In situ* hybridization and electron microscopy revealed that, at least in *A. pisum*, they are located in a different type of bacteriocytes adjacent to those formed by *Buchnera* (Fukatsu et al. 2000). Available data indicate that secondary symbionts have been acquired several times during aphid evolution and, contrary to the origin of *Buchnera*, they do not represent a monophyletic clade (Fukatsu and Ishikawa 1993; Moran and Telang 1998). Recently, three types of bacteria belonging to the Enterobacteriaceae have been identified in several species from the Macrosiphini tribe, named U-, R-, and T-type (Sandström et al. 2001). The phylogenetic distribution of secondary symbionts indicates occurrence of repeated horizontal transmissions.

Wernegreen and Moran (1999) have studied the effects of selection and random genetic drift on the evolution of several loci of *Buchnera*, as compared to free-living bacteria. The major conclusion of their study, which also supports previous findings (Moran 1996), is that random drift affects the potential effect of selection in the evolution of *Buchnera* in a crucial way. Due to its maternal inheritance, the effective population size of *Buchnera* is much lower than the effective population size of free-living bacteria. This has at least three important evolutionary consequences at the molecular level: a negative effect on the process of optimization of codon usage, an increase in the number of non-synonymous substitutions, and an acceleration of the evolutionary rate. According to Li (1987), the coefficient of selection needed to drive a population from a less to a more efficient codon usage (i.e., translational selection) should be larger than $2/N_e$. Hence, those populations having great effective size (N_e) will experience the effects of translational selection with weaker coefficients of selection than those others with less effective population size.

The aim of the present work is to make use of the previous results on *Buchnera* protein-coding genes, as well as new ones reported here for the first time and to test the three above mentioned evolutionary consequences of endosymbiotic processes on secondary symbionts.

Material and Methods

Aphid Species. Twenty representative aphid species belonging to the five major aphid families were used in this study (see aphid host in Table 1). Since most of this species cannot be cultured, total DNA was extracted and kept at −20°C for later analysis.

Sequences and Bacterial Symbionts. Partial sequences of the *atpD* and *aroQ*/*pheA* genes, some of them here reported for the first time, have been used in the present study. *atpD* locus codes for the β-subunit of the F-ATP complex, and *aroQ* /*pheA* is a bifunctional gene encoding the chorismate mutase (5′ end) and the prephenate dehydratase (3′ end), which play a central role in L-phenylalanine biosynthesis. Table 1 shows the aphid host, bacterial species and accession numbers of the 44 partial sequences used in this study. Twenty nine of them correspond to the central region (453 bp) of the *atpD* locus, 17 of which have been previously sequenced by our group (Buades et al. 1999) and the other 12 obtained from the databases. The remaining 15 partial sequences correspond to a fragment (372 bp) of the 3′-terminal part of the *aroQ*/*pheA* gene, eight of which have been sequenced for the present work, one more reported elsewhere (Jiménez et al. 2000) and six obtained from the databases.

Compared to free living bacteria, total A + T content and, specially, at third base position is a distinctive and well documented feature of symbiotic ones (Moran and Werngreen 2000). Based on that we have chosen the criteria of considering symbiotic those bacteria that showed an A + T content at third base position higher than 48.4%, the value observed in *E. coli*.

We then removed from our analyses three *aroQ*/*pheA* sequences that correspond to Tca2, Tsu2, and Epa2b, with A + T content of 41.2%, 41.9%, and 30.6%, respectively (see Table 1).

Of *atpD* sequences (Table 1), 15 correspond to *Buchnera*, the primary endosymbiont of aphids, three to secondary symbionts of aphids and the remaining 11 are Proteobacteria belonging to the α (2), β (2), γ (3), δ (2), and ε (2) subdivisions. Of the 15 partial *aroQ*/*pheA* sequences (Table 1), three correspond to *Buchnera*, three to secondary symbionts, three more that were excluded and the remaining six belong to Proteobacteria of the β (1) and γ (5) subdivision. Proteobacteria species were chosen because most endosymbionts belong to this group (Moran and Wernegreen 2000). Alignments of the sequences are not shown but they are available upon request to the corresponding author.

DNA Extraction and Sequencing. Total DNA was extracted from the aphids following the method of Latorre et al. (1986) modified for a single individual (Sanchis et al. 2000). A fragment of the *atpD* gene was amplified by using the oligonucleotide primers as well as PCR conditions described by Clark and Baumann (1993). The PCR product was run in a 0.9% low melting point Agarose (Boehringer-Mannheim) and the band corresponding to the expected size was cut from the gel and ligated in the vector pUCBM20 (Boehringer-Mannheim). At least two clones from each individual were sequenced. During the analysis of the sequences we found in *M. rosae* two different sequences (see Table 1) that belong to *Buchnera* and a secondary symbiont respectively (see results). In order to discriminate, before sequencing, the existence of more than one sequence from a single individual, the clones obtained were digested with the restriction enzyme *Asp*700 that has at least a site in all the *Buchnera* sequences but none in less A + T rich sequences such as *E. coli* and secondary symbionts. After examining at least 20 clones from each DNA only the sequence corresponding to the secondary symbiont was found in the DNAs from *P. juglandis* and *C. pini*. For the rest of species only DNA corresponding to *Buchnera* was amplified.

Table 1. Accession numbers of the partial *atpD* and *aroQ*/*pheA* sequences corresponding to *Buchnera,* secondary symbionts and selected Proteobacteria

Type of bacteria	Aphid family or Proteobacteria subdivision	Species[a]	Code[b]	Accession number	
				atpD	*aroQ*/*pheA*
Buchnera	Pemphigidae	*Eriosoma lanuginosum*	Ela	AJ247134	:
		Geoica utricalaria	Gut	AJ247135	
		Pemphigus bursarius	Pbu	AJ247132	
		Pemphigus spyrothecae	Psp	AJ247133	
	Thelaxidae	*Thelaxes suberi*	Tsu	AJ247131	
	Drepanosiphidae	*Panaphis juglandis*	Pju	AJ247136	
	Lachnidae	*Lachnus roboris*	Lro	AJ247137	
		Tuberolachnus salignus	Tsa	AJ247138	
	Aphididae	*Aphis gossypii*	Ago	AJ247124	
		Acyrthosiphon pisum	Api		AJ239043
		Macrosiphum rosae	Mro	AJ247125	AJ291794
		Myzus persicae	Mpe	AJ247129	
		Pterocomma populeum	Ppo	AJ247126	
		Rhopalosiphum padi	Rpa	AJ247127	AJ291795
		Schizaphis graminum	Sgr	Z15147	
		Staticobium latifoliae	Sla	AJ247130	
Symbiont	Pemphigidae	*Tetraneura caerulescens*	Tca2		AJ291797
	Thelaxidae	*Thelaxes suberi*	Tsu2		AJ291792
	Drepanosiphidae	*Chaitophorus leucomelas*	Cle2	AJ247139	
		Panaphis juglandis	Pju2		AJ291796
	Lachnidae	*Cinara pini*	Cpi2	AJ247141	
		Eulachnus pallidus	Epa2a		AJ291799
		Eulachnus pallidus	Epa2b		AJ291793
		Tuberolachnus salignus	Tsa2		AJ291798
	Alphididae	*Macrosiphum rosae*	Mro2	AJ247140	
Proteobacteria	α-Subdivision	*Rhodobacter capsulatus*	Rca	X99599	
		Rhodospirillum rubrum	Rru	X02499	
	β-Subdivision	*Burkholderia cepacia*	Bce	X76877	’
		Neisseria meningitidis	Nme	AL162757	AL162757
	γ-Subdivision	*Escherichia coli*	Eco	V00267	M10431
		Erwinia herbicola	Ehe		X60420
		Salmonella typhi	Sty	Sanger[c]	Sanger[c]
		Shewanella putrefaciens	Spu		TIGR[d]
		Vibrio algenolyticus	Val	X16050	
		Vibrio cholerae	Vch		AE003852
	δ-Subdivision	*Desulfovibrio vulgaris*	Dvu	D84449	
		Stigmatella aurantiaca	Sau	X76879	
	ε-Subdivision	*Helicobacter pylori*	Hpy	AE001533	
		Wolinella succinogenes	Wsu	X76880	

[a] The first 24 species indicate the host aphid species where *Buchnera* or the putative secondary symbiont was obtained.
[b] Abbreviations used in other tables.
[c] http://www.sanger.ac.uk/Projects/S_typhi/.
[d] http://www.tigr.org/tdb/mdb/mdbinprogress.html.

A fragment of the *pheA* cistron was amplified using the degenerate oligonucleotide primers PheAd1 and PheAd2 described in Jiménez et al. (2000) which were based on the sequences of *E. coli* and related Proteobacteria. These primers amplified the expected PCR product in the reactions with DNA from *R. padi* and *M. rosae*. After sequencing, they were identified as belonging to *Buchnera* on the basis of their high A + T content and nucleotide similarity with de *aroQ*/*pheA* gene from *Buchnera* sp APS (from *A. pisum*, see Table 1). Since no amplification was obtained with DNA from other aphid species, a nested PCR strategy was used. First, a PCR reaction was carried out with the degenerate primers PheAd-6F (5′-TATAGTCATCCWCARCCNTTYCARC-3′) and PheAd-4R (5′-TWTTTTCWSWWGGATARCANCC-3′), followed by a second PCR reaction with primers *PheAd*1 and PheAd2. In five species the expected fragment (400 bp approx.) was obtained. The DNA was purified and ligated to the pGEM-T easy vector (Promega). After transformation, DNA from several clones was digested with the four-cutter enzyme *MspI*. In four out of five aphid species a single profile of digestion was obtained, whereas in *Eulachnus pallidus* two different kinds of clones were identified. All of them were sequenced, determining that they corresponded to the *aroQ*/*pheA* gene (Table 1).

Nucleotide sequencing was carried out with the AmpliTaqF Dye Deoxy Terminator cycle-sequencing kit and a 373 automated DNA sequencer (Perkin Elmer-Applied Biosystems) as recommended by the manufacturer.

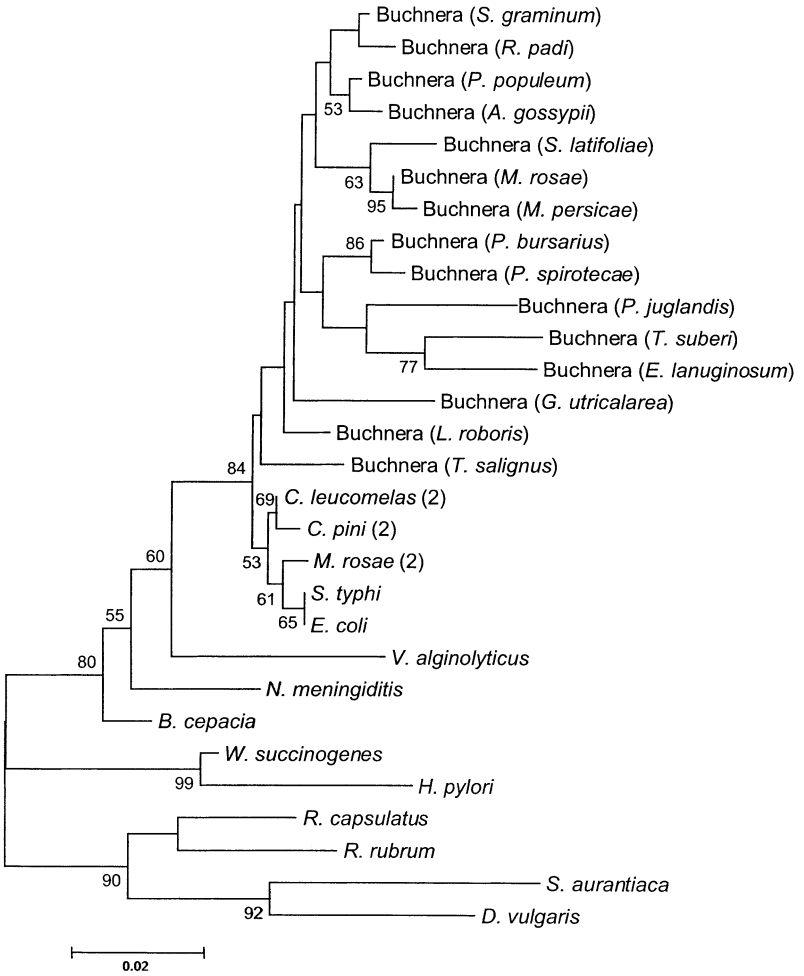

Fig. 1. Neighbor-joining tree obtained based on Tamura-Nei distances of first and second base positions among pairs of partial *atpD* sequences. The reliability of the different branches was evaluated by bootstrapping (1000 replicates). Only those branches with bootstrap values higher than 50% are represented.

Phylogenetic Analyses. Base composition is clearly different among free living and symbiotic bacteria (see below). For a proper distance based phylogenetic analysis both base composition and saturation effects on transition and transversion rates have to be considered. Thus, phylogenetic analyses were carried out with Tamura-Nei distance (Tamura and Nei 1993) using only transversions on first and second base positions. Neighbor-joining trees (Saitou and Nei 1987) have been constructed and the reliability of the clusters obtained evaluated by means of 1000 bootstrap replicates. In addition, the method of Galtier and Gouy (1995) for inferring phylogenies from DNA sequences of unequal base composition was also applied.

Analysis of Codon Usage at Fourfold Degenerate Sites. In order to make the results of the present study comparable to those reported by Wernegreen and Moran (1999) we have applied the same basic procedure to determine the existence of codon bias in selected Proteobacteria, in *Buchnera* and in six putative secondary endosymbionts. In essence, nonrandom use of U- and A-ending codons within fourfold degenerated codon families was evaluated by means of χ^2 analyses. With respect to *Buchnera* loci, C- and G-ending codons were excluded from the analysis due to their few or null presence. In this case expected relative frequencies of U- and A-ending codons for each fourfold codon family were estimated from the relative frequencies of A's and T's at fourfold degenerated sites. Observed and expected values of each one of these eight sites were compared by χ^2 tests. Codons for amino acids with fewer than five residues were not analyzed.

The χ^2 tests corresponding to selected Proteobacteria and secondary symbionts considered four instead of two classes. The comparison between *Buchnera* and the other two groups has been done on the basis of the proportion of significant cases and not on the values of the χ^2 test.

It should be noticed that the null hypothesis is that there is no loss in codon bias when comparing free living and symbiotic bacteria. On the other hand, it is well known the presence of codon bias among free living bacteria, specially in highly expressed coding genes.

Relative Rate Tests. To estimate whether primary and putative secondary endosymbionts have increased their evolutionary rate when compared with representative Proteobacteria species used as reference, relative rate tests were applied to triad of species (Sarich and Wilson 1973). Prior to the application of the test it is necessary to estimate the number K_{ij} of nucleotide substitutions per site among sequences i and j respectively. To make proper comparisons with previous results on rates of different *Buchnera* genes (Moran 1996) we applied Kimura's two parameters (K2p) method (Kimura 1983, data not shown). It should be mentioned, however, that the results obtained when using K2p gave in qualitative terms the same results than those obtained when distances were corrected using Tamura and Nei's method, the one used for phylogenetic reconstruction. Finally it was evaluated whether branches leading to sequences 1 and 2 (see results) were statistically different by applying the method of Wu and Li (1985) as implemented in program K2WULI (Muse and Weir 1992).

Table 2. A + T content of the *atpD* partial sequences in *Buchnera*, secondary symbionts and selected Proteobacteria

Symbiont	Aphid species	A + T content (%)			
		lst base	2nd base	3rd base	Total
Buchnera	Ela	54.3	60.9	86.8	67.3
	Gut	51.7	59.1	88.9	66.5
	Pbu	48.4	61.0	86.7	65.4
	Psp	49.7	60.9	87.4	66.0
	Tsu	52.3	61.0	84.1	65.8
	Pju	49.0	60.9	87.3	65.7
	Lro	49.7	61.6	90.7	67.4
	Tsa	49.7	59.6	94.0	67.8
	Ago	52.4	60.5	82.6	65.1
	Mro	49.0	61.6	85.4	65.3
	Mpe	49.1	61.0	86.1	65.3
	Ppo	49.1	58.9	88.0	65.3
	Rpa	47.7	61.0	84.2	64.2
	Sgr	47.7	60.3	85.5	64.4
	Sla	50.3	60.0	91.4	67.2
	Average	50.0	60.6	87.3	65.9
Secondary Symbiont	Cle2	43.4	59.1	62.5	54.9
	Cpi2	41.7	60.3	55.6	52.6
	Mro2	45.7	59.6	64.0	56.4
	Average	43.6	59.7	60.7	54.6
Proteobacteria	Rca	43.8	58.4	17.0	37.1
	Rru	33.6	58.9	17.1	36.5
	Bce	38.0	62.7	16.0	38.9
	Nme	45.9	60.5	39.5	48.6
	Eco	38.4	59.6	39.1	45.7
	Sty	38.7	60.0	37.3	45.3
	Val	36.1	60.1	56.9	51.0
	Dvu	34.2	58.6	19.1	37.3
	Sau	35.2	56.0	7.5	32.9
	Hpy	43.4	60.5	43.4	49.2
	Wsu	42.8	59.9	49.4	50.7
	Average	39.1	59.6	31.1	43.0

Ratio of Synonymous Versus Non-Synonymous Substitution Rates. Estimates of synonymous (K_s) and non-synonymous (K_a) pairwise divergence and standard deviations were calculated using Li's method (1993) (DAMBE, Data Analysis in Molecular Biology and Evolution version 4.0.17). In order to avoid inaccurate estimates due to saturation, only pairwise comparisons with $K_s < 1$ were included in the calculation of K_s/K_a ratios.

Results

Nucleotide Composition

Tables 2 and 3 show the A + T content (%) of the *atpD* and *aroQ*/*pheA* selected loci, respectively, where three groups are considered: *Buchnera*, secondary symbionts and selected Proteobacteria. On average, both *Buchnera* genes show the highest A + T content (65.9% and 76.0% for *atpD* and *oroQ*/*pheA* genes, respectively), which is even more pronounced when the third base is considered (87.3% and 84.2%, respectively). As it can be observed, as compared to *Buchnera* and Proteobacteria, secondary symbionts have an intermediate A + T content (60.7% and 59.1% for *atpD* and *aroQ*/*pheA* third base position, respectively). As expected, due to functional constraints, both genes show the smallest variation in A + T composition in second codon position among the three groups.

Phylogenetic Analyses

Tamura-Nei's pairwise distances were used to derive a NJ-tree for each gene (see Material and Methods). Figures 1 and 2 show the trees corresponding to *atpD* and *aroQ*/*pheA* genes, respectively. In the tree obtained with the *atpD* gene, *Buchnera* species, *E. coli*, *S. typhi*, and the three secondary symbionts form a single monophyletic cluster, being *V. algenolitycus* the closest outgroup of it. A different topology appears with the tree obtained with the *aroQ*/*pheA* gene (Fig. 2). The three *Buchnera* species belonging to the Aphididae family (Table 1) constitute a monophyletic cluster. A second monophyletic cluster is formed by *E. coli*, *S. typhi*, and *E. herbicola* as well as the secondary symbiont *E. pallidus2a*, whereas

Table 3. A + T content of the *aroQ/pheA* partial sequences in *Buchnera*, secondary symbionts and selected Proteobacteria

Symbiont	Aphid species	A + T content (%)			
		1st base	2nd base	3rd base	Total
Buchnera	Api	74.2	67.7	86.3	76.1
	Mro	71.3	68.0	81.0	76.8
	Rpa	73.8	66.4	85.2	75.1
	Average	73.1	67.4	84.2	76.0
Secondary endosymbiont	Pju2	49.2	60.5	66.9	58.9
	Epa2a	40.3	58.8	49.2	49.4
	Tsa2	49.2	63.7	61.3	58.1
	Average	46.2	61.0	59.1	55.5
Proteobacteria	Nme	40.6	48.0	27.6	38.8
	Eco	41.9	57.2	48.4	49.2
	Ehe	36.3	60.5	29.0	41.9
	Sty	40.3	56.4	29.8	42.2
	Spu	40.3	54.8	48.4	47.8
	Vch	46.8	56.4	50.8	51.3
	Average	41.0	55.6	39.0	36.7

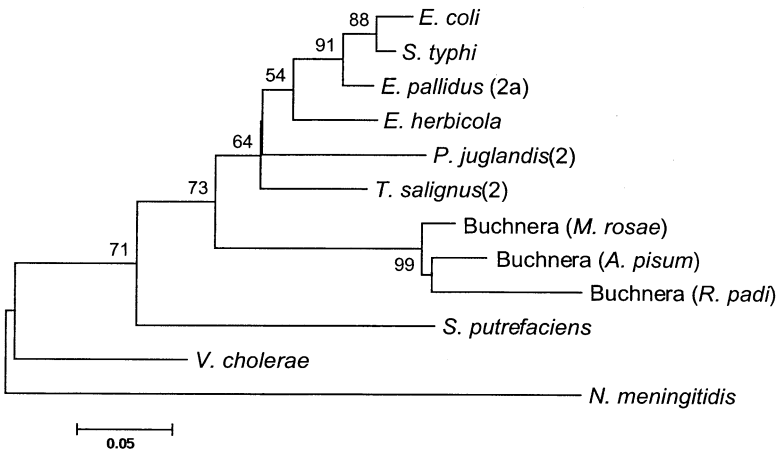

Fig 2. Neighbor-joining tree obtained based on Tamura-Nei distances of first and second base positions among pairs of partial *aroQ/pheA* sequences. The reliability of the different branches was evaluated by bootstrapping (1000 replicates). Only those branches with bootstrap values higher than 50% are represented.

P. juglandis2 and *T. salignus2* show an intermediate position. *S. putrefaciens* and *V. cholerae* behave as outgroup species with respect to these two clusters.

In order to ascribe the secondary symbionts to the three R-, T-, and U-types described in aphids (Sandström et al. 2001), we show in Fig. 3 a phylogenetic analysis based on 16S rDNA of the following 23 sequences: 10 corresponding to *Buchnera's* from major aphid families, seven corresponding to secondary symbionts of aphids and whiteflies (2 R-types, 2 T-types, 2 U-types and a secondary symbiont from *Bemisia tabaci*), and the remaining six are Proteobacteria also included in our study. As it can be observed the topology obtained allow us to ascertain *M. rosae2* (Fig. l) and *E. pallidus2a* (Fig. 3) as R-types. Regarding *C. leucomelas2*, C. *pini2* (Fig. 1) and *P. juglandis2* and *T. salignus2* (Fig. 2), they can be considered as either T- or U-types of secondary symbionts.

Codon Bias Analyses

Table 4 shows the χ^2 analysis of the non-random use of U- and A-ending codons for each of the eight fourfold degenerate families in both genes. On average, there is a clear decline in the number of significant cases of non-random use of codons when free living bacteria (37.9% and 37.2% for *atpD* and *aroQ/pheA* genes, respectively) are compared with *Buchnera* (1% and 0%, respectively), which is in agreement with previous results (Wernegreen and Moran 1999). In addition, the percentage of significant cases of the secondary symbionts is 13.0% and 11.1% for *atpD* and *aroQ/pheA*, respectively, which are intermediate to the ones found in *Buchnera* and free-living bacteria. Finally, in spite of their different functional roles, it is worth noticing the similarity of percentages between both genes for each one of the three groups considered.

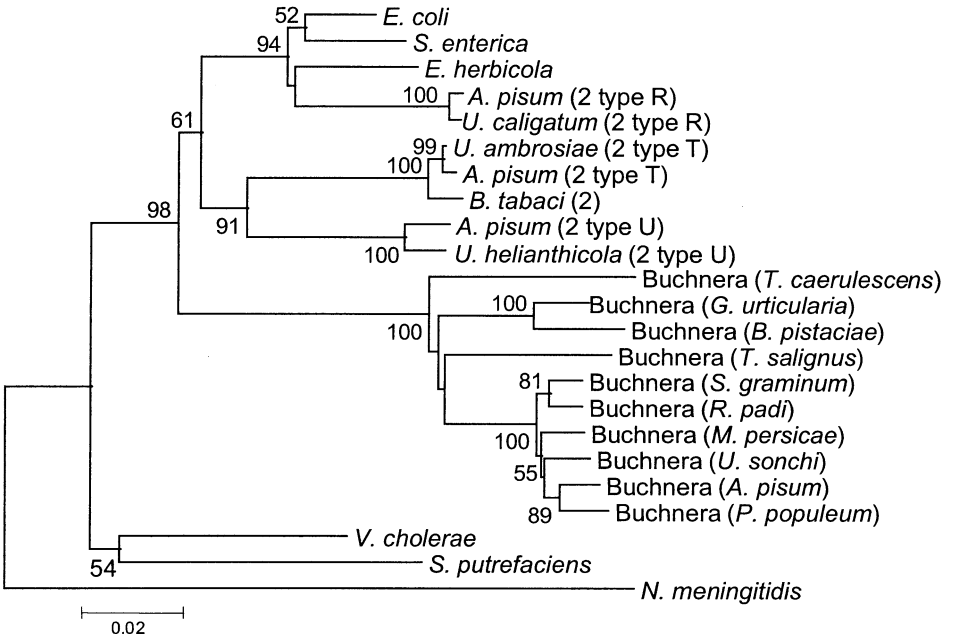

Fig. 3. Neighbor-joining tree obtained based on Tamura-Nei distances among pairs of 16S rDNA sequences of 23 species. The reliability of the different branches was evaluated by bootstrapping (1000 replicates). Only those branches with bootstrap values higher than 50% are represented. Accession numbers of the sequences as they appear on the figure (from top to down) are as follows: *E. coli* (AE000460); *S. enterica* (AF227869); *E. herbicola* (AF290417); *A. pisum2*, type R (AF293617); *Uroleucon caligatum2*; type R (AF293624); *Uroleucon ambrosiae2*, type T (AF293622); *A. pisum2,* type T (AF293616); *Bemisia tabaci2* (Z11926); *A. pisum2*, type U (AF293618); *U. helianthicola2*, type U (AF293625); *Buchnera* (*T. caerulescens*) (AJ296749); *Buchnera* (*G. urticularia*) (AJ296751); *Buchnera* (*T. salignus*) (AJ296754); *Buchnera* (*S. graminum*) (M63246); *Buchnera* (*R. padi*) (M63248); *Buchnera* (*M. persicae*) (M63249); *Buchnera* (*U. sonchi*) (M63250); *Buchnera* (*A. pisum*) (M27039); *Buchnera* (*P. populeum*) (AJ296747); *V. cholerae* (X74695); *S. putrefaciens* (X81623); *N. meningitidis* (AE002098).

Relative Rate Tests

Table 5 shows the results of the relative rate tests between *Buchnera* and the putative secondary endosymbionts with respect to *E. coli*. *V. alginolyticus* and *S. putrefaciens*, the species that first branched off the monophyletic clades including *Buchnera* and the secondary symbionts, were used as outgroups for *atpD* and *aroQ*/*pheA* genes, respectively (Figs. 1 and 2). As it can be observed by inspecting the tests of the *atpD* gene, *Buchnera* is evolving faster than *E. coli* since their divergence. Moran (1996) and Moran et al. (1995) have reported similar results studying the genes *trp*EG and 16S rDNA from *Buchnera* of different species. When comparing the three secondary symbionts and *E. coli*, the difference is smaller but still significant, thus indicating that the *atpD* gene is also evolving faster in these species than in *E. coli*.

As expected, *aroQ*/*pheA* genes from the three *Buchnera* species show acceleration with respect to *E. coli*. However, none of the three secondary endosymbionts show significant differences with *E. coli*. A particular feature when applying this test to *aroQ*/*pheA* is the high standard deviation of the difference K_{13}–K_{23} (see Table 5), doubling those estimated in the case of *atpD*. Only high K_{13}–K_{23} differences, as those estimated in the case of *Buchnera* comparisons, yielded significant evolutionary acceleration, but that is not the case when considering secondary symbionts.

Number of Non-Synonymous Versus Synonymous Substitutions

The effect of purifying selection should be smaller in populations of haploid organisms having a strong random genetic drift. This might be the expectation when comparing *Buchnera* with selected Proteobacteria, i.e. using pairs of species whose divergence times are similar or higher than those of *Buchnera* species. For this reason we have estimated K_s and K_a between *E. coli* and *S. typhi* in both genes. These two species diverged more than 100–150 MYA. Other relevant estimates of divergence date, according to the fossil record (see Clark et al. 1999), are between 50–70 MYA for *Buchnera* from Pemphigidae or Aphididae families (Table 1). For this reason, we have estimated average K_s and K_a for *Buchnera* from Aphididae and Pemphigidae in the case of *atpD* and only from Aphididae for *aroQ*/*pheA* (as we did not amplify any *Buchnera* sequence of species from the Pemphigidae family). Table 6 shows the corresponding average values of K_s and K_a for both genes among *Buchnera* and the pair of selected Proteobacteria. As it can be observed there is

Table 4. χ^2 values for testing the nonrandom use of U- and A-ending codons for each of the eight fourfold degenerate families

		A		G		L		P		R		S		T		V	
	Species	*atpD*	*aroQ/pheA*	*atpD*	*aroQ/pheA*	*atpD*	*aroQ/pheA*	*atpD*	*aroQ/pheA*	*atpD*	*aroQ/pheA*	*atpD*	*aroQ/pheA*	*atpD*	*aroQ/pheA*	*atpD*	*aroQ/pheA*
Buchnera	Ela	0.07		0.10		—		0.02		—		0.97		0.07		0.07	
	Gut	0.08		0.45		—		1.98		0.63		0.06		0.14		0.12	
	Pbu	1.38		0.24		—		0.91		0.94		0.20		0.56		0.06	
	Psp	2.09		0.51		—		0.22		1.67		1.04		0.15		0.15	
	Tsu	0.05		0.20		—		1.01		—		0.03		0.03		1.15	
	Pju	0.08		0.65		0.17		0.50		0.50		0.07		0.07		0.65	
	Lro	0.70		0.34		—		4.45*		2.84		0.25		0.25		1.71	
	Tsa	1.01		0.03		—		—		0.03		0.01		1.12		0.00	
	Ago	0.10		1.37		—		0.69		—		0.44		0.18		0.04	
	Api		2.04		1.50		—		—		—		2.45		1.50		—
	Mro	1.03	0.38	0.01	0.01	—	—	0.02	—	—	—	0.00	—	0.00	1.28	0.05	—
	Mpe	0.28		0.41		—		0.00		—		0.02		0.06		0.00	
	Ppo	0.09		0.00		—		1.92		0.05		0.00		0.00		0.00	
	Rpa	0.08	1.64	0.21	0.23	—	—	0.03	—	—	—	0.08	0.25	0.01	0.07	0.00	—
	Sgr	0.00		0.00		—		—		0.05		0.41		0.02		0.04	
	Sla	0.36		0.32		—		0.04		1.40		0.00		0.84		1.52	
Secondary symbiont	Cle2	6.00		1.05		1.08		1.05		6.22		2.37		1.63		2.74	
	Pju2		3.91		2.71		2.30		1.85		—		—		2.67		2.22
	Cpi2	1.65		1.66		6.21		5.67		17.03*		3.70		11.60*		1.68	
	Epa2a		2.95		5.51		8.76*		—		7.73		—		8.35*		1.30
	Tsa2		6.94		5.68		2.27		3.68		—		—		0.42		3.49
	Mro2	9.97*		1.98		—		4.44		5.63		0.52		2.11		2.36	
Proteobacteria	Rca	6.55		8.78*		3.32		1.76		5.26		14.49*		10.80*		7.84*	
	Rru	7.05		7.21		15.34*		2.80		7.32		14.50*		8.61*		1.07	
	Bce	13.19*		11.27*		3.91		4.25		24.46*		5.71		2.15		0.40	
	Nme	3.88	7.12	7.43	11.04*	—	7.93*	10.28*	2.11	8.67*	0.52	4.75	2.70	8.17*	1.00	20.94*	21.23*
	Eco	4.04	2.06	6.24	2.18	30.53*	10.05*	6.05	4.81	12.67*	5.78	8.27*	—	12.69*	11.73*	21.97*	2.66
	Ehe		3.88		8.78*		12.29*		2.15		10.32*		—		11.02*		0.94
	Sty	10.59*	8.29*	7.62	4.12	24.15*	7.96*	5.54	3.35	15.27*	7.92*	12.58*	—	12.17*	0.86	25.02*	2.08
	Spu		4.03		3.17		1.61		1.25		—		2.30		—		7.35*
	Val	15.31*		14.50*		1.17		0.38		10.92*		3.02		1.54		7.25	
	Vch		1.60		2.66		4.63		3.03		9.14*		9.26*		8.33*		1.11
	Dvu	3.10		4.63		2.53		1.72		6.87		5.93		7.41		2.39	
	Sau	1.72		5.39		1.23		1.33		6.16		1.21		1.06		15.14*	
	Hpy	1.20		3.52		4.11		10.19*		6.23		0.07		2.61		7.07	
	Wsu	12.74*		2.21		2.15		4.64		15.09*		1.09		4.17		11.25*	

*$p < 0.05$

Table 5. Relative rate tests for *atpD* and *aroQ/pheA* genes of *Buchnera*, and secondary endosymbionts (taxon 1) compared to *E. coli* (taxon 2). The outgroups (taxon 3) used have been *V. algenolyticus* and *S. putrefaciens* for *atpD* and *aroQ/pheA*, respectively

		atpD			*aroQ/pheA*		
	Taxon 1	K_{13}–K_{23}[a]	SD[b]	Z-score[c]	K_{13}–K_{23}	SD	Z-score
Buchnera	Ela	0.2084	0.0413	5.05*			
	Gut	0.1703	0.0390	4.31*			
	Pbu	0.1727	0.0384	4.49*			
	Psp	0.1878	0.0389	4.83*			
	Tsu	0.1732	0.0388	4.47*			
	Pju	0.1837	0.0406	4.52*			
	Lro	0.1459	0.0378	3.86*			
	Tsa	0.1535	0.0370	4.14*			
	Ago	0.1567	0.0369	4.24*			
	Api				0.2713	0.0913	2.97*
	Mro	0.1443	0.0361	4.00*	0.2201	0.0902	2.44*
	Mpe	0.1555	0.0367	4.24*			
	Ppo	0.1485	0.0367	4.05*			
	Rpa	0.1457	0.0362	4.03*	0.3026	0.0988	3.06*
	Sgr	0.1426	0.0361	3.95*			
	Sla	0.1845	0.03900	4.73*			
Secondary symbiont	Cle2	0.1057	0.0329	3.21*			
	Pju2				−0.0305	0.0622	−0.49
	Cpi2	0.0911	0.0290	3.14*			
	Epa2a				−0.0296	0.0496	−0.60
	Tsa2				0.1052	0.0641	0.83
	Mro2	0.1390	0.0351	3.96*			

[a] K_{13}–K_{23} give us an estimate of the net nucleotide substitution per site difference between taxon 1 and 2 since its divergence from its more common ancestor. For more details see Wu and Li (1985) and Moran (1996).
[b] SD: estimated standard deviation for the difference K_{13}–K_{23}.
[c] Z-score were estimated as described in Wu and Li (1985).
* $p < 0.05$.

Table 6. Average number of K_s, synonymous and K_a, non-synonymous, substitution rates, and K_s/K_a ratio for *Buchnera* from different aphid families and *E. coli–S. typhi*

Gene	Lineage	$K_s \pm$ SE[a]	$K_a \pm$ SE	K_s/K_a
atpD	*Buchnera* (Aphididae)	0.7209 ± 0.3933	0.0308 ± 0.0096	23.39
	Buchnera (Pemphigidae)	0.7234 ± 0.6540	0.0592 ± 0.0137	12.22
	E. coli–S. typhi	0.1373 ± 0.03549	0.0000 ± 0.0000	∞
aroQ/pheA	*Buchnera* (Aphididae)	0.5589 ± 0.1609	0.1326 ± 0.0233	4.21
	E. coli–S. typhi	0.9899 ± 0.2096	0.0632 ± 0.0155	17.15

[a] Standard error.

a substantial decrease in the ratio K_s and K_a in *Buchnera* with respect to the *E. coli–S. typhi* in both genes due to a higher number of non-synonymous substitutions. In fact this number is zero between *E. coli* and *S. typhi* in the case of *atpD*. It should be considered that we have used the central domain of the gene, which is the most conserved. For the whole gene Clark et al. (1999) obtained K_a estimated for two pairs of *Buchnera* of 0.11 and 0.13, whereas for *E. coli–S. typhimurium* the K_a was as low as 0.03.

Because it is not known a cluster of secondary endosymbiont with divergence time similar to the cluster of *Buchnera,* either from Pemphigidae or Aphididae or to the enteric bacteria cluster formed by *E. coli* and *S. typhi,* we are not able to test if secondary endosymbionts have accumulated a higher or similar number of non-synonymous and synonymous substitutions than Proteobacteria or *Buchnera,* respectively.

Discussion

Compared to free living bacteria, a high A + T content, loss of codon bias, accelerated evolutionary rate and increased number of non-synonymous substitutions are some features of the accommodation process of endosymbiotic bacteria to the intracellular life of other organisms. According to Wernegreen and Moran (1999) the last three of these features result from severe population bottlenecks of endosymbiotic

bacteria during maternal transmission through their hosts. These authors reported evidence of the abovementioned features in a study where a large set of protein coding genes was tested. In the present study we report additional evidence of such predictions based on partial sequences of two genes that have been examined in a large set *of Buchnera* species representative of the major aphid families (Table 1). Table 4 shows that the two *Buchnera* genes present, in general, loss of the non-random use of U- and A-ending codons. As discussed at large by Wernegreen and Moran (1999), following the same procedure suggested by these authors and only for comparative purposes, we have tested the hypothesis that translational selection is completely inefficient in the primary (less than 1% of significant cases) and highly non efficient in the secondary (less than 13%) symbionts. These results support the hypothesis that A + T mutational bias and fixation of non-optimal codons through random drift modulate codon usage in *Buchnera* and other symbionts. On the other hand, K_a and K_s gave comparative values that are the ones expected (Table 6). Considering *atpD* and *aroQ*/*pheA* genes, K_a for *Buchnera* higher than K_a for the pair *E. coli–S. typhi,* and the K_s/K_a ratio is lower for *Buchnera* than for free living bacteria.

High A + T content is a characteristic feature of the resident genomes, i.e. genome of a bacteria that live in close, often intracellular, association with a eukaryotic host genome (Andersson and Kurland 1998; Moran and Wernegreen 2000). Accordingly, the intermediate A + T content of the *atpD* and *aroQ*/*pheA* genes of the secondary symbionts compared to *Buchnera* and free living bacteria might be compatible with an active and younger endosymbiotic accommodation involving other bacteria. The timing and way of this process is, however, not uniform. We have detected significant cases of evolutionary acceleration in the case of the *atpD* gene, but not in *aroQ*/*pheA,* what might be interpreted, apart from sampling problems to detect it in the last case, as giving support to the hypothesis of random drift in evolving populations of small but different effective population size. In addition, these two factors, are large enough as to promote a disruption in the effectiveness of translational selection, and then to induce the loss of codon bias, as mentioned above.

Regarding the origin of secondary symbionts, clearly they do not constitute a monophyletic group, and based on phylogenetic analyses carried out (Figs. 1–3), we cannot give support to the hypothesis that all these secondary symbionts result from a single and new symbiotic event. In fact, within this group of non-related secondary symbiotic bacteria, *M. rosae2* and *E. pallidus2* can be considered as R-type, if we follow Sandström et al. (2001), whereas the rest can be considered as either T- or U-types. Sandström et al. (2001), working with fifteen species of macrosiphine aphids and based on DNA sequencing, PCR diagnostic, phylogenetic analyses and electron microscopy of the T-type, have reported the existence of at least three groups of secondary symbionts maternally transmitted and of independent origin. The cytological study was not possible in our case because the specimens were not available once DNA was extracted. In the two species that are currently cultivated in our lab *(A. pisum* and *R. padi,* see Table 1), only *Buchnera* was amplified. The association between secondary symbionts and their hosts is not as stable as that between *Buchnera* and aphids. According to these authors the distribution of secondary symbionts indicates frequent horizontal transmissions and occasional infidelity of vertical transmission. On the contrary, detected cases of horizontal transmission in *Buchnera* are scarce (van Ham et al. 2000).

Acknowledgments. We appreciate the critical reading of Dr. Fernando González-Candelas. We are grateful to 'Servicio Central de Soporte a la Investigation Cientifica (Universitat de Valencia)' for bioinformatics and sequencing facilities. The work has been supported by grants GV-3216/95 from Generalitat Valenciana (Spain) and PB96-0793-C04-01 from DGES (Spain) to AM.

References

Andersson SGE, Kurland CG (1998) Reductive evolution of resident genomes. Trends Microbiol 6:263–268

Baumann P, Baumann L, Lai CY, Rouhbakhsh D, Moran NA, Clark MA (1995) Genetics, physiology, and evolutionary relationships of the genus *Buchnera:* intracellular symbionts of aphids. Annu Rev Microbiol 49:55–94

Baumann L, Baumann P, Moran NA, Sandstrom J, Thao ML (1999) Genetic characterization of plasmids containing genes encoding enzymes of ieucine biosynthesis in endosymbionts (*Buchnera*). J Mol Evol 48:77–85

Baumann P, Moran NA, Baumann L (2000) Bacteriocyte-associated endosymbionts of insect. In: M. Dworkin (ed.), The prokaryotes. [Online] Springer-Verlag, New York. http://link.springer.de/link/service/books/10125/

Bracho AM, Martinez-Torres D, Moya A, Latorre A (1995) Discovery and molecular characterization of a plasmid localized in *Buchnera* sp. bacterial endosymbiont of the aphid *Rhopalosiphum padi*. J Mol Evol 41:67–73

Buades C, Michelena JM, Latorre A, Moya A (1999) Accelerated evolution in bacterial endosymbionts of aphids. Int Microbiol 2:11–14

Buchner P (1965) Endosymbiosis of animals with plant microorganisms. John Wiley and Sons, New York

Chen D-Q, Purcell AH (1997) Occurrence and transmission of facultative endosymbionts in aphids. Curr Microbiol 34:220–225

Clark MA, Baumann P (1993) Aspects of energy-yielding metabolism in the aphid *Schizaphis graminum,* and its endosymbionts: detection of gene fragments potentially coding for the ATP synthase β-subiunit and glyceraldeyde-3-phosphate dehydrogenase. Curr Microbiol 26:233–237

Clark MA, Moran NA, Baumann P (1999) Sequence evolution in bacterial endosymbionts having extreme base compositions. Mol Biol Evol 16:1586–1589.

Douglas AE (1998) Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. Ann Rev Entomoli 43:17–37

Fukatsu T, Ishikawa H (1993) Occurrence of chaperonin 60 and chaperonin 10 in primary and secondary bacterial symbionts of aphids: implications for the evolution of an endosymbiotic system in aphids. J Mol Evol 36:568–577

Fukatsu T, Ishikawa H (1998) Differential immunohistochemical visualization of the primary and secondary intracellular symbiotic bacteria of aphids. Appl Entomol Zool 33:321–326

Fukatsu T, Watanabe K, Sekiguchi Y (1998) Specific detection of intracellular symbiotic bacteria or aphids by oligonucleotide-probed in situ hybridization. Appl Entomol Zool 33:461–472

Fukatsu T, Nikoh N, Kawai R, Koga R (2000) The secondary endosymbiotic bacterium of the pea aphid Acyrtosiphon pisum (Insecta: Homoptera). Appl Environ Microbiol 66:2748–2758s

Galtier N, Gouy M (1995) Inferring phylogenies from DNA sequences of unequal base compositions. Proc Natl Acad Sci USA 92:11317–11321

Jiménez N, González-Candelas F, Silva FJ (2000) Prephenate dehydratase from the aphid endosymbionts (*Buchnera*) displays changes in the regulatory domain that suggest its desensitization to inhibition by phenylalanine. J Bacteriol 182:2967–2969

Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge

Lai C-Y, Baumann L, Baumann P (1994) Amplification of *trpEG*: adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. Proc Natl Acad Sci USA 91:3819–3823

Latorre A, Moya A, Ayala FJ (1986) Evolution of mitochondrial DNA in *Drosophila subobscura*. Proc Natl Acad Sci USA 83:8649–8653

Li W-H (1987) Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J Mol Evol 24:337–345

Li W-H (1993) Unbiased estimation of the rates of synonymous and non-synonymous substitution. J Mol Evol 36:96–99

Moran NA (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. Proc Natl Acad Sci USA 93:2873–2878.

Moran NA, Munson MA, Baumann P, Ishikawa H (1993) A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc R Soc London 253:167–171

Moran NA, von Dohlen CD, Baumann P (1995) Faster evolutionary rates in endosymbiotic bacteria than in cospeciating insect hosts. J Mol Evol 41:727–731

Moran NA, Telang A (1998) Bacteriocyte-associated symbionts of insects. BioScience 48:295–304

Moran NA, Wernegreen JJ (2000) Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends Ecol Evol 15:321–326

Muse SV, Weir BS (1992) Testing for equality of evolutionary rates. Genetics 132:269–276

Nakabachi A, Ishikawa H (1997) Differential display of mRNAs related to amino acid metabolism in the endosymbiotic system of aphids. Insect Biochem Molec Biol 27:1057–1062

Nakabachi A, Ishikawa H (1999) Provision of riboflavin to the host aphid Acyrtosiphon pisum, by ensosymbiotic bacteria, *Buchnera*. J Insect Physiol 45:l–6

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:4l8–426

Saitou N, Nei M (1987) The neighbor-joining method: new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sanchis A, Latorre A, González-Candelas F, Michelena JM (2000) An 18SrDNA based molecular phylogeny of Aphidiinae (Hymenoptera: Braconidae). Mol Phylogenet Evol 14:180–194

Sandström JP, Russell JA, White JP, Moran NA (2001) Independent origins and horizontal transfer of bacterial symbionts of aphids. Mol Ecol 10:217–228

Sarich, VM, Wilson AC (1973) Generation time and genomic evolution in primates. Science 179:1144–1147

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbionts of aphids *Buchmera* sp. APS Nature 407:81–86

Tajima F (1993) Simple methods for testing molecular clock hypothesis. Genetics 135:599–607

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in human and chimpanzees. Mol Biol Evol 10:512–526

Unterman BM, Baumann P, McLean DL (1989) Pea aphid symbiont relationships established by analysis of 16SrRNAs. J Bacteriol 171:2970–2974

van Ham RCHJ, González-Candelas F, Silva FJ, Sabater B, Moya A, Latorre A (2000) Postsymbiotic plasmid acquisition and evolution of the repAl-replicon in *Buchnera aphidicola*. Proc Natl Acad Sci USA 97:10855–10860

van Ham RCHJ, Moya A, Latorre A (1997) Putative evolutionary origin of plasmids carrying the genes involved in leucine biosynthesis in *Buchnera aphidicola* (endosymbiont of aphids). J Bacteriol 179:4768–4777

Wernegreen JJ, Moran NA (1999) Evidence for genetic drift in endosymbionts (*Buchnera*): Analyses of protein-coding genes. Mol Biol Evol 16:83–97

Wilkinson TL, Ishikawa H (2000) Injection of essential amino acids substitutes for bacterial supply in aposymbiotic pea aphids (*Acyrthosiphon pisum*). Entomol Exp Appl 94: 85–91

Wu, C-I, Li W-H (1985) Evidence of higher rates of nucleotide substitution in rodents than in man. Proc Natl Acad Sci USA 82:1741–l745