

SHORT REVIEW

Chromosomal stasis *versus* plasmid plasticity in aphid endosymbiont *Buchnera aphidicola*

A Latorre, R Gil, FJ Silva and A Moya

Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado de Correos 2085, 46071 Valencia, Spain

The study of three genomes of the aphid endosymbiont *Buchnera aphidicola* has revealed an extraordinary stasis: conservation of gene order and genetic composition of the chromosome, while the chromosome size and number of genes has reduced. The reduction in genome size appears to be ongoing since some lineages we now know to have even smaller chromosomes than the first *B. aphidicola* analysed. The current sequencing by our group of one of these smaller genomes with an estimated size of 450 kb, and its comparison with the other three available genomes provide insights into the nature of processes involved in shrinkage. We discuss whether *B. aphidicola* might be driven to extinction and be replaced by secondary aphid endosymbionts. In some lineages, genes encoding key enzymes in

the pathways leading to tryptophan and leucine biosynthesis (*trpEG* and *leuABCD*, respectively) are located on plasmids, rather than the main chromosome. In contrast to the stasis of the main chromosome, plasmid genes have frequently been transferred to the main chromosome and undergone other gene rearrangements. We propose a two-step scenario to explain these contrasting modes of evolution. Essential genes may have escaped regulation by moving to plasmids in a moving *B. aphidicola* ancestor. *B. aphidicola* became polyploidy at a given stage of its evolution and plasmid genes have been transferred to the main chromosome through several independent events.

Heredity (2005) 95, 339–347. doi:10.1038/sj.hdy.6800716; published online 24 August 2005

Keywords: comparative genomics; gene order; leucine and tryptophan plasmids; endosymbionts

Introduction

Contrary to most prokaryotes, eukaryotes have rather limited metabolic capabilities, and hence, symbiosis has provided an evolutionary strategy through which eukaryotes gain access to a wider range of metabolic resources. Insects are particularly prone to symbiotic associations with microorganisms (Buchner, 1965). It has been estimated that at least 15–20% of insects live in such symbiotic relationships, allowing them to explore a great variety of ecological niches. Insects that live in close association with bacteria are characterized, in general, by feeding on unbalanced diets, poor in essential nutrients such as amino acids, sterols or vitamins, which are provided by the symbionts (Sasaki *et al*, 1991; Douglas, 1998; Baumann *et al*, 2000). Such is the case of members of the orders Homoptera (aphids, whiteflies, mealybugs, psyllids and cicadas), Blattaria (cockroaches) and Coleoptera (weevils). Insect endosymbionts live in a very closed environment, inside specialised host cells called bacteriocytes, which may form an organ-like structure called bacteriome, in the body cavity of the insects. The association is mutualistic and obligate for both partners: the bacteria cannot be cultured outside the host, whereas the host needs the bacteria for normal growth and reproduction.

In the last few years, five complete genomes of insect endosymbionts have been fully sequenced, all of them belonging to the γ -Proteobacteria: three strains of *Buchnera aphidicola* (Shigenobu *et al*, 2000; Tamas *et al*, 2002; van Ham *et al*, 2003), *Wigglesworthia glossinidia* (Akman *et al*, 2002) and *Blochmannia floridanus* (Gil *et al*, 2003), the primary endosymbionts of aphids, tsetse flies and carpenter ants, respectively. These genomes have a size ranging from 618 to 706 kb, and contain 545–661 genes, thus revealing a dramatic process of genome and gene number reduction when compared to close free-living relatives such as *Escherichia coli* (*E. coli* K12 has a 4639 kb genome, containing 4289 genes) (Blattner *et al*, 1997). The comparison of the genes present in these bacteria confirmed the nutritional role of the symbioses, since all of them contain species-specific genes involved in the supply of products that are needed by their particular hosts, in addition to a common set of genes necessary for intracellular life (Gil *et al*, 2003; Klasson and Andersson, 2004).

Aphids are plant phloem-feeding insects, whose diet is deficient in essential amino acids, which are provided by the primary endosymbiont *B. aphidicola*. The association between aphids and *B. aphidicola* is very ancient, and the congruence between the phylogenetic trees of hosts and symbionts indicates a unique infection event about 84–164 million years ago, followed by the coevolution of both partners (Moran *et al*, 1993; von Dohlen and Moran, 2000). The genus *Buchnera* contains one species, *B. aphidicola*, which designates all primary endosymbionts present in the different aphid species (Munson *et al*, 1991). *B. aphidicola* is maternally inherited by infection of

Correspondence: A Latorre, Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado de Correos 2085, 46071 Valencia, Spain. E-mail: amparo.latorre@uv.es

Received 26 November 2004; accepted 16 June 2005; published online 24 August 2005

the eggs or embryos at the blastoderm stage (Buchner, 1965). For the last few decades, many studies have been carried out regarding the biology of aphids, *B. aphidicola*, and the aphid-*B. aphidicola* association (revised in Douglas, 1998; Baumann et al, 2000; Latorre et al, 2003), as it is considered an excellent model to gain insights into the extreme evolutionary reductive process that takes place when bacterial lineages make the transition from an independent lifestyle to a permanent association with the host.

In addition to its drastic genome reduction and the proliferation of plasmids involved in the biosynthesis of the amino acids leucine and tryptophan (see below), *B. aphidicola* has undergone many other important molecular and biochemical changes, when compared to its free-living relatives. These changes can be characterised by the following features: amplification of the number of copies of the main chromosome per cell, an almost total absence of recombination, an increase in the rate of nucleotide substitution, high A + T content, an accumulation of deleterious mutations and loss of codon bias (Moran, 1996; Clark et al, 1999; Komaki and Ishikawa, 1999; Baumann et al, 2000; Moya et al, 2002).

Gene order and gene content in the *B. aphidicola* chromosome

In recent years, many studies performed on both symbionts and parasites have shown that, once the intracellular lifestyle is established, a massive process of genome degradation occurs, resulting in the inactivation and loss of many genes, reflecting the relaxation of selection to maintain genes that are rendered superfluous in the constant and rich environment provided by the host (Andersson and Kurland, 1998; Moran and Mira, 2001; Silva et al, 2001; Moran, 2002; Gómez-Valero et al, 2004a). The genomes of three *B. aphidicola* obtained from different aphid species are currently available: *B. aphidicola* (BAp) from *Acyrtosiphon pisum*, *B. aphidicola* (BSg) from *Schizaphis graminum*, and *B. aphidicola* (BBp) from *Baizongia pistaciae* (Table 1) (Shigenobu et al, 2000; Tamas et al, 2002; van Ham et al, 2003). In addition, we are currently sequencing the genome of *B. aphidicola* (BCc), from *Cinara cedri* (Pérez-Brocal et al, 2005). The comparison of the gene content and gene order in these four *B. aphidicola* strains associated with aphids from three different subfamilies can give some clues about specific adaptations of the bacteria to the particular environment provided by their respective hosts, leading to different structural changes in their genomes.

Gene-order fossil in *B. aphidicola*

The sequencing of the first *B. aphidicola* genome from the aphid *A. pisum* (BAp) (Shigenobu et al, 2000), and its comparison to the genomes of *E. coli* (Blattner et al, 1997) and *Vibrio cholerae* (Heidelberg et al, 2000), two free-living bacteria closely related to *B. aphidicola*, indicated that the endosymbiont has not only experienced substantial genome reduction but also many chromosomal rearrangements since the divergence of these species, the most frequent ones being the inversions around the origin and terminus of replication (Silva et al, 2001).

Soon after, a second *B. aphidicola* genome was sequenced from *S. graminum* (BSg) (Tamas et al, 2002), an aphid belonging to the same subfamily as *A. pisum* (Aphidinae), but to a different tribe (Aphidini for *S. graminum*, and Macrosiphini for *A. pisum*) (see Table 2). The estimated divergence time between these tribes is 50–70 million years (Clark et al, 1999). The comparison between both genomes revealed an extreme degree of conservation, without either chromosomal rearrangements (translocations, inversions or duplications), or gene acquisition by horizontal gene transfer, thus being the most extreme case of genome stability to date (Tamas et al, 2002). Comparison with a third strain, *B. aphidicola* BBp, associated with the aphid *B. pistaciae* (subfamily Pemphiginae), confirmed that the genomic architecture is extremely similar. Near-perfect gene-order conservation was found, with only four minor rearrangements (two inversions and two translocations involving the leucine and tryptophan plasmid-contained genes) relative to the BAp and BSg strains (van Ham et al, 2003). It was suggested that *B. aphidicola* could be considered as an enterobacterial 'gene-order fossil', and that the onset of genomic stasis coincided with the establishment of the symbiosis with aphids. Although no consensus aphid phylogeny has yet been achieved, all of those published show that Aphidinae and Pemphiginae are very divergent lineages (Heie, 1987; Ortiz-Rivas et al, 2003), with estimated divergence times of 84–164 million years (von Dohlen and Moran, 2000). Therefore, considering the divergence time for the *B. aphidicola* strains, the level of genome stability is completely unexpected. On one hand, no horizontal gene transfer insertions can be detected along the *B. aphidicola* genomes, while the incidence in other bacteria reaches 20–30% in within-species genome comparisons, even over short periods of time, as it is the case of several *E. coli* strains (Welch et al, 2002). On the other hand, it is remarkable that there is an almost complete absence of inversions and other internal rearrangements in the genome, since inversions are relatively frequent evolutionary events in most of the γ -proteobacterial lineages (Belda et al, 2005). However,

Table 1 Main genomic features of four analysed *B. aphidicola* strains and comparison of two selected genome regions

Aphid host	Strain	Chromosome length (bp)	Genes	Accession number	rrl-arok and fpr-trxa regions				
					length (bp)	G+C (%)	ORFS	RNAS	IRs length ^a
<i>Acyrtosiphon pisum</i>	BAp	640 681	608	NC002528	47 547	29.2	60	5	3126
<i>Schizaphis graminum</i>	BSg	641 454	596	NC004061	46 895	28.2	57	5	2704
<i>Baizongia pistaciae</i>	BBp	615 980	544	NC004545	44 907	28.1	55	5	4118
<i>Cinara cedri</i>	BCc	~448 000	Unknown	AY744381; AY744382	34 576	25.4	48	5	2342

^aLength of the conserved intergenic regions.

Table 2 Taxonomic status, location and gene order of the leucine cluster and *trpEG* genes in different strains of *B. aphidicola* from aphids of the family Aphididae

Subfamily	Aphid host		Strain	<i>B. aphidicola</i>				
	Tribe	Species		Leucine cluster			<i>trpEG</i> genes	
				Plasmid	Chromosome	Accession no.	Location	Accession no.
Thelaxinae (T)		<i>Thelaxes suberi</i>	BTs	BCDA		Y11966	Unknown	Unknown
Lachninae (L)	Eulachnini	<i>Tuberolagnus salignus</i>	BTg	BCDA		AY546103	Unknown	Unknown
	Lachnini	<i>Cinara cedri</i>	BCc	BCDA		AY438025	Unknown	Unknown
Pterocommatinae (P)		<i>Pterocomma populeum</i>	BPp	ABCD		AJ006877	Unknown	Unknown
Aphidinae (A)	Aphidini	<i>Rhopalosiphum padi</i>	BRp	ABCD		X71612	Plasmid	L43551
		<i>Schizaphis graminum</i>	BSg	ABCD		AF041836	Plasmid	Z21938
	Macrosiphini	<i>Acyrtosiphum pisum</i>	BAP	ABCD		AJ006878	Plasmid	AP001070
		<i>Diuraphis noxia</i>	BDn	ABCD		AF041837	Plasmid	L46769
Pemphiginae (Pe)	Pemphigini	<i>Pemphigus spyrothecae</i>	BPs		ABCD (cp)	AJ426489; J404864	Plasmid	AJ012334
	Fordini	<i>Baizongia pistaciae</i>	BBp		BCDA (cp)	NC004545	Chromosome	NC004545
	Eriosomatini	<i>Tetraneura caerulea</i>	BTC		BCDA (cp)	AY375290; Y11974	Plasmid	AJ012333
Chaitophorinae (C)		<i>Chaitophorus populeti</i>	BCp		BCDA	AY375291	Unknown	Unknown

(cp), criptic plasmid.

B. aphidicola is not the only lineage with low inversion frequency. The frequency of internal rearrangements in some parts of the genome is extremely low in comparison of *Salmonella* and *Escherichia* lineages, which has been related to selection against changes in the position of the genes on the genome (Campo et al, 2004). In the case of *B. aphidicola*, this pattern is associated with an extremely reduced recombination frequency, probably due to the loss of an efficient homologous recombination system (absence of RecA) as well as to the loss of repeated sequences longer than 30 bp in its genome (Klasson and Andersson, 2004).

Gene content varies among lineages

Despite of the similar genome size of the three sequenced *B. aphidicola* strains, there are some differences in gene content, indicating that independent gene losses have occurred since the Last Common Symbiotic Ancestor of *B. aphidicola* (LCSA). Since *B. aphidicola* has maintained the gene order, it was possible to postulate the genome structure of the LCSA of all extant *B. aphidicola* strains, and to ascertain which specific genes have been lost in the different lineages.

The reconstruction of the minimum gene content of the last common ancestor of *B. aphidicola* and its close relative *E. coli* (LCA) identified 1818 (Silva et al, 2001) or 2425 genes (Moran and Mira, 2001), depending on the bacterial outgroup being used in the analysis. The complete set of coding genes present in the three sequenced *B. aphidicola* strains is 640, obtained as the sum of all shared and nonshared genes and pseudo-genes, and it represents the most parsimonious reconstruction of the gene content of the LCSA (Silva et al, 2003; van Ham et al, 2003; Gómez-Valero et al, 2004a), indicating that the massive loss of genes occurred in the transition from the LCA to LCSA. The number of genes in the BAP, BSG and BBp genomes is 608, 596 and 544, respectively (Table 1). A minimum of 164 independent gene losses have occurred, 96 in the BBp lineage, eight before the split of the two Aphidini, and 24 and 36 in the BAP and BSG lineages, respectively. A set of 139 genes (22%) accounts for differences in gene content among the

species. It includes pseudogenes, entirely lost genes, and remnant DNA where a gene was located in the past, according to the three postulated steps of genome reduction during *B. aphidicola* genome evolution (Silva et al, 2001). The reason why certain genes are lost (or retained) in some lineages is a matter of debate. The differences in functional gene content between BBp on one hand, and BAP and BSG on the other hand, are substantially greater than the differences between BAP and BSG. When the nature of the lost genes protein products were categorised into the functional Clusters of Orthologous Groups (COGs, Tatusov et al, 1997), we found that the losses embrace all functional categories, although in a different proportion. The majority of the lost genes are involved in metabolic and cellular processes, while the most conserved are genes devoted to information processing (Gómez-Valero et al, 2004a). Moreover, we found that the majority of the losses were not convergent (80%), indicating that the differential losses are host-specific, probably due to its particular diet and/or lifecycle. In addition to genes involved in informational processes, the genes involved in the essential amino-acid biosynthetic pathways are also conserved in the three genomes, presumably due to their nutritional role in the symbiosis.

The genome of *B. aphidicola* BCc, from the aphid *Cinara cedri* (subfamily Lachninae), is currently being obtained in our laboratory. This bacterium possesses the smallest genome reported so far, with an estimated size of 450 kb (Gil et al, 2002). Our preliminary results showed that the gene order has also been maintained in this strain, confirming the hypothesis that *B. aphidicola* is a gene-order fossil. We have, therefore, been able to perform a detailed analysis of two selected, already sequenced, regions of BBc which have a very different gene content, and compare them with the corresponding regions in the three previously sequenced *B. aphidicola* genomes (Pérez-Brocá et al, 2005). The first region is located between the genes that encode 23S ribosomal RNA and shikimate kinase I (*rrl-aroK*), and mainly contains genes coding for ribosomal proteins. The second region is located between the coding genes for ferredoxine-NADP reductase and thioredoxin (*fpr-trxA*), and it contains mainly genes

involved in metabolic and cellular processes. The gene order is completely conserved among the four strains in both regions, in spite of a large reduction in gene number (Table 1). In the region *rrl-aroK*, five genes have been lost with respect to BAp and BSg (*argD*, *yhfC*, *yrdC*, *fjpA*, and *smg*). Among these genes, *argD* and *smg* are also absent in BBp. Since the region contains 47 genes in BAp and BSg (45 in the case of BBp), this means that 89% of the genes have been preserved in BCc. The region *fpr-trxA* contains 13 genes in BAp and BSg and 10 in BBp. More than half of the genes have been lost in BCc with respect to BAp and BSg (*ynfM*, *kdtB*, *hemC*, *hemD*, *yjeA*, *pitA*, *yhiQ*, *yba4* and *uspA*). These results indicate that gene loss has not occurred evenly throughout the genome, but is highly dependent on the nature of the gene function. Thus, the genome reduction in BCc appears mainly to be due to the loss of genes belonging to metabolic and cellular processes, which appear to be nonessential in the protected environment provided by the bacteriocyte, while significantly more genes involved in information processing have been retained in BCc than in the other three strains.

Genome reduction is still an ongoing process

In bacteria, gene content usually correlates to genome size. The reduction in genome size of intracellular bacteria (symbionts and parasites) is associated with the loss of a great number of genes, probably in response to the new environment, where many molecules can be obtained from the host and do not need to be synthesized. In addition, the endosymbiotic bacteria have lost most of the genes involved in the recombination processes and, consequently, the genome size cannot be increased by acquisition of foreign DNA. It was, therefore, expected that, after the massive gene loss occurred during adaptation to intracellular life from the LCA to the LCSA of *B. aphidicola*, the preserved genes would be those needed mostly for symbiotic life, with no significant changes in gene content and genome size in different lineages. In fact, the genome sizes of the three sequenced *B. aphidicola* genomes are quite similar (Table 1). However, the PFGE analysis of the chromosome size of nine *B. aphidicola* strains belonging to five aphid subfamilies (Gil et al, 2002) revealed that there is huge size variation among different lineages, ranging from 670 to 450 kb (Figure 1). *B. aphidicola* strains from aphids of the subfamilies Chaitophorinae, Thelaxinae and Lachninae have genomes of approximately 450–550 kb, smaller than the genome of *Mycoplasma genitalium*, the smallest sequenced bacterial genome (580 kb) (Fraser et al, 1995). The study also revealed great differences among aphid subfamilies, with the most extreme being the decrease of approximately 200 kb in *B. aphidicola* from aphids belonging to the subfamily Lachninae, compared to *B. aphidicola* from the Aphidinae. However, the differences within subfamilies are generally much smaller, being 20 kb on average (see Figure 1). These results suggest that, although the great majority of genome shrinkage probably occurred in the transition from LCA to LCSA, the different *B. aphidicola* lineages are still undergoing a reductive process, but at a slower rate. The comparative analysis of two analysed regions across the four *B. aphidicola* strains provides some further clues to address this issue (Pérez-Brocal et al, 2005). At least in

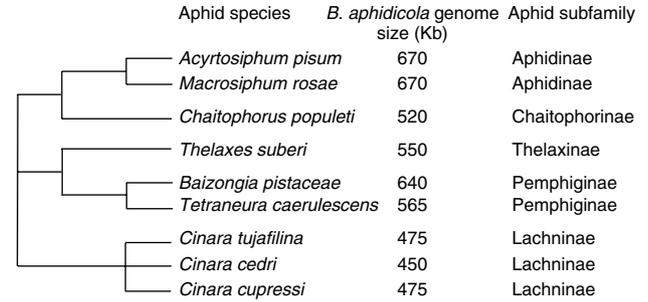


Figure 1 Genome sizes of *B. aphidicola* from nine aphid species belonging to five aphid subfamilies (Gil et al, 2002). The apparent genome sizes of BAp and BBp strains are slightly higher than those obtained after the sequencing of the complete genome (Shigenobu et al, 2000; van Ham et al, 2003).

these two regions, the reduction process is mainly due to the loss of genes by a process of gene disintegration (Silva et al, 2001), rather than to the shortening of coding regions and/or intergenic regions (IRs). As it can be seen in Table 1, the size of the conserved IRs in the two analysed regions is about the same in BSg and BCc. Gómez-Valero et al (2004a) have estimated that the half-life of a pseudogene in *B. aphidicola* (ie the period in which half of its nucleotides are lost) is 23.9 million years. It seems then that the size of the different *B. aphidicola* genomes is controlled by restrictions on the loss of function of crucial genes, because when any gene is inactivated, its nucleotides are progressively removed from the genome in a relatively short period of time. Finally, we found that the genome reduction is not random, because more divergent genes are significantly more prone to loss than more conserved ones. More studies should be carried out to test whether there is a correlation between the type of host (ie biological and ecological factors) and *B. aphidicola* genome size, or whether there are historical contingencies involved in this process. This subject has been approached by gene-array hybridisation experiments in other insect models. Using commercial *E. coli* arrays, a heterologous hybridisation experiment was carried out to infer the genomic composition of *Sodalis glossinidius*, secondary endosymbiont of the tsetse fly, and SOPE, *Sitophilus oryzae* primary endosymbiont (Rio et al, 2003). Both bacteria belong to the same clade, and have recently established symbiotic associations, while their hosts belong to different insect orders (Diptera and Coleoptera) and feed on animal blood and stored grain, respectively. The main differences between both genomes appear to reflect differences in the endosymbiotic relationship with their respective hosts, providing insights into the impact of biological and ecological factors on endosymbiont genome evolution.

Are the *B. aphidicola* genomes driven to extinction?

It is clear that the loss of some essential metabolic capabilities during the genome reduction process forced intracellular bacteria to depend irreversibly on their host for survival. The question that remains to be solved is whether the final step of the reduction process is the extinction of these endosymbionts.

Owing to their maternal inheritance, only a small fraction of bacteria pass to the next generation, thus experiencing continuous bottlenecks, and accumulating slightly deleterious mutations in an irreversible way, a phenomenon known as Muller's ratchet (Moran, 1996). The same scenario is true for the oldest known cases of endosymbiosis, which lead to the formation of modern eukaryotic mitochondria and chloroplasts, as a result of symbiogenic events between prokaryotes and primitive eukaryotes around 2000 and 1000 million years ago, respectively (Margulis, 1981). However, in these cases, the establishment of the symbiosis took place at a monocellular stage, and the irreversible bacterial loss of control over their own cellular processes involved the loss of many redundant pathways plus a massive gene transfer from the symbionts to the host nuclear genome, while the existing organelles have retained the functions that make them essential for the survival of the eukaryotic cells. In the present case of bacterial endosymbionts of multicellular organisms, a similar process of horizontal gene transfer from a newly acquired prokaryotic cell to the eukaryotic host must be very difficult. The only reported case has been the transfer of a genome fragment of a *Wolbachia* endosymbiont to the X chromosome of its insect host *Callosobruchus chinensis* (Kondo *et al*, 2002). In the absence of gene transfer to the host, the loss of essential functions to preserve host fitness might end with the extinction and replacement of these endosymbionts by healthier endosymbiotic bacteria that are present in the same environment. The most direct evidence of secondary symbionts overtaking the role of *B. aphidicola* has been obtained by Koga *et al* (2003), revealing that, when *B. aphidicola* was eliminated, the secondary symbionts invaded the bacteriocyte space, establishing a novel endosymbiotic system. Moreover, the infection with secondary symbionts enabled the survival and reproduction of *B. aphidicola*-free aphids. Furthermore, evidence of natural endosymbiotic replacement has been reported recently in some weevil species of the family Dryophoridae (Lefèvre *et al*, 2004).

If the prediction is correct, a good aphid-endosymbiont candidate to reach extinction is *B. aphidicola* BCc, since its genome is so reduced that the impact of Muller's ratchet must be even higher than in the other *B. aphidicola* genomes. The preserved genes must still be accumulating deleterious mutations, and the compensatory mechanisms, such as the overproduction of GroEL (Fares *et al*, 2002), would not be enough to maintain the essential functions, which need to be preserved for host fitness. We have recently postulated that the abundant secondary symbionts, which coexist with *B. aphidicola* BCc in different bacteriocytes of the aphid, might have taken over at least some of the functions lost by the primary endosymbiont, and might eventually replace it (Gómez-Valero *et al*, 2004b). However, it still remains to be proven that *B. aphidicola* BCc has lost its ability to provide essential nutrients to the host.

Leucine and tryptophan plasmids go back and forward

In recent years, genes encoding key enzymes in the pathways leading to tryptophan and leucine biosynthesis (*trpEG* and *leuABCD*, respectively) were found to be

translocated from the chromosome to plasmids (Lai *et al*, 1994; Bracho *et al*, 1995). Amplification of plasmids for essential amino-acid biosynthesis was considered as an indication that *B. aphidicola* is able to overproduce these nutrients for the benefit of its host and, hence, an adaptation to its symbiotic lifestyle. However, the discovery that the main chromosome is present in multiple copies in each cell (Komaki and Ishikawa, 1999), together with the findings that ratios of plasmid-borne *trpEG* and *leuABCD* copies to chromosomal copies could vary, both within and between species (Thao *et al*, 1998; Plague *et al*, 2003), casts doubt on the idea that plasmid location is a means of leucine and tryptophan overproduction. Alternatively, if we consider that the transfer to plasmids occurred when the *B. aphidicola* genome contained regulatory elements, this could have been a way of skirting genome regulation of leucine and tryptophan biosynthesis by negative feedback, thus allowing a continued supply of these amino acids to the insect, despite the high amino-acid concentration in the bacterial cells.

The bacterial studies carried out on the location of the *trpEG* and, mainly, on the *leuABCD* genes in several aphid subfamilies, either on plasmids or on the main chromosome (Lai *et al*, 1995, 1996; Rouhbakhsh *et al*, 1996; van Ham *et al*, 1997, 1999, 2000; Sabater-Muñoz *et al*, 2002, 2004), revealed a great plasticity throughout *B. aphidicola* evolution, showing that the genomic stasis hypothesis is not supported when plasmids are taken into account. In fact, the evolutionary history of plasmids is puzzling, since not all of the bacterial lineages carry plasmids, and not all plasmids have the same gene content and/or gene order. In addition, when the leucine cluster is located in the chromosome, it is flanked by different genes in different lineages (Sabater-Muñoz *et al*, 2004), in contrast with the colinearity found in the three sequenced *B. aphidicola* genomes.

The Leucine cluster

Up to seven plasmids and four chromosomal locations have been described for the leucine cluster in species belonging to six aphid subfamilies: Thelaxinae, Lachninae, Pterocommatinae, Aphidinae, Pemphiginae and Chaitophorinae (Table 2, Figure 2).

Leucine plasmids, ranging in size from 6.3 to 8.2 kb, are present in the subfamilies Thelaxinae, Lachninae, Pterocommatinae and Aphidinae (van Ham *et al*, 1997, 2000; Silva *et al*, 1998; Baumann *et al*, 1999; Soler *et al*, 2000). They contain the four structural genes for the synthesis of leucine (*leuA*, *leuB*, *leuC* and *leuD*) plus one or two genes coding for a plasmid replicase (*repA1* and/or *repA2*). Some plasmids also contain the *ihp* gene, coding for a heat-shock protein, and *yqhA*, coding for a putative membrane protein of unknown function. However, the genes *leuA*, *B*, *C* and *D* are located on the main chromosome in *B. aphidicola* strains associated with the subfamilies Pemphiginae and Chaitophorinae (Sabater-Muñoz *et al*, 2002, 2004; Van Ham *et al*, 2003). In addition, all analysed *B. aphidicola* strains from the Pemphiginae possess a cryptic leucine plasmid (cp) of 1.7–2.4 kb (van Ham *et al*, 2000) that does not contain the leucine genes.

Contrary to the gene-order conservation observed in the main chromosome, the leucine genes are found in

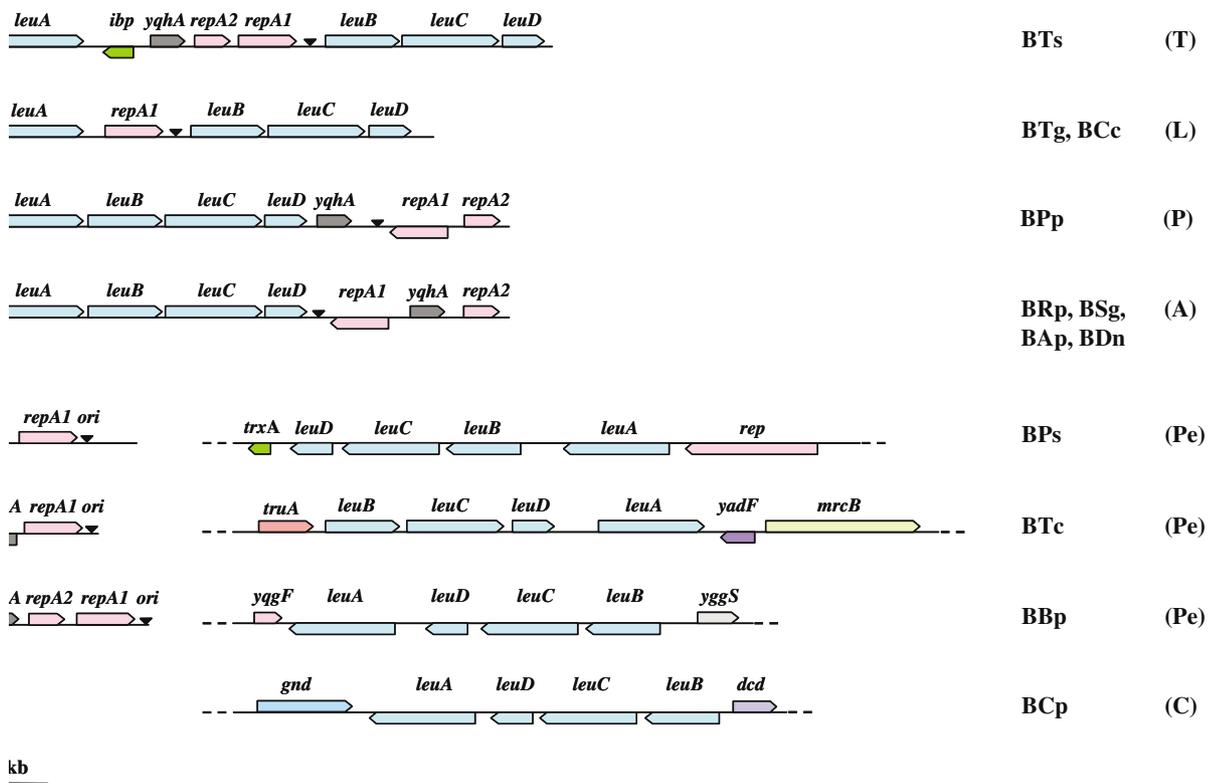


Figure 2 Structures of the leucine cluster found in the different *B. aphidicola* strains analysed. (a) Leucine plasmid in species from the subfamilies (T) Thelaxinae, (L) Lachninae, (P) Pterocommatinae and (A) Aphidinae. (b) Chromosomal version of the leucine cluster and flanking regions in species belonging to different tribes of the Pempighinae (Pe), and in one species of the subfamily Chaitophorinae (C). The structures of the cryptic *repA* plasmid (cp) of the Pempighinae are also shown. See Table 1 for species code.

different gene orders and flanked by different genes, both when they are located on a plasmid or on the chromosome (Table 2, Figure 2). Bacteria from members of the subfamilies Aphidinae and Pterocommatinae contain the leucine genes on plasmids in the same order as in *E. coli* (*leuABCD*), but the companion genes are in a different order, proving the existence of rearrangements, even in those plasmids belonging to endosymbionts of very closely related aphid subfamilies. The other two leucine plasmids present in *B. aphidicola* strains from aphids of the subfamilies Thelaxinae (BTs) and Lachninae (BTg and BCc) show a different order for the leucine genes (*leuBCDA*). BTg and BCc contain the simplest leucine plasmids, with only a *repA* gene in addition to the four leucine genes. Conversely, the leucine plasmid of BTs contains all the genes found in any of the other seven plasmids, probably representing the original structure of the ancestral plasmid. Finally, there are also some differences among the three cryptic plasmids found in the subfamily Pempighinae. They are phylogenetically related to the leucine plasmids (van Ham *et al*, 2000), but instead of having the structural leucine genes, they only code for *ibp* or *yqhA* genes, plus one or two replicase genes (Figure 2), depending on the tribe within the subfamily.

In the *B. aphidicola* strains from aphids of the subfamilies Chaitophorinae and Pempighinae, the leucine cluster is located at different chromosome positions (Figure 2) (Sabater-Muñoz *et al*, 2004). In BCp (subfamily

Chaitophorinae), it is flanked by the genes *gnd* and *dcd*. In the subfamily Pempighinae, the location of the leucine cluster varies depending on the tribe. In BBp (tribe Fordini), it is flanked by *yqgF* and *yqgS*, in BPs (tribe Pempighini), by *trxA* and *rep*, and in BTc (tribe Eriosomatini), the flanking genes are *truA* and an ORF homologous to the *yadF* gene in *E. coli* (Blattner *et al*, 1997), upstream of the gene *mrcB*. This ORF is not present in any of the three *B. aphidicola* sequenced genomes, so it must have been lost in these lineages.

Owing to the extreme gene-order conservation of the *B. aphidicola* genomes, the variability in the position of the leucine cluster in the chromosome can only be interpreted as resulting from four independent insertions from an ancestral plasmid that contained the leucine genes. The pairs of genes that flank the leucine cluster in each strain (*gnd-dcd*; *yqgF-yqgS*; *trxA-rep* and *truA-mrcB*) are adjacent in BAp, BSg and BCc (unpublished results). Although in the past our research group proposed a chromosomal location for the leucine genes in the LCSA, as in free-living enterobacteria (van Ham *et al*, 1997), the new findings suggest that a leucine plasmid must have been present in the *B. aphidicola* LCSA that preceded the diversification of all the endosymbionts, and that the chromosomal location of the leucine genes observed in some *B. aphidicola* strains arose through a back-transfer from the plasmid to the main chromosome. We have performed a phylogenetic reconstruction with the strains shown in Table 2, and with some free-living γ -proteo-

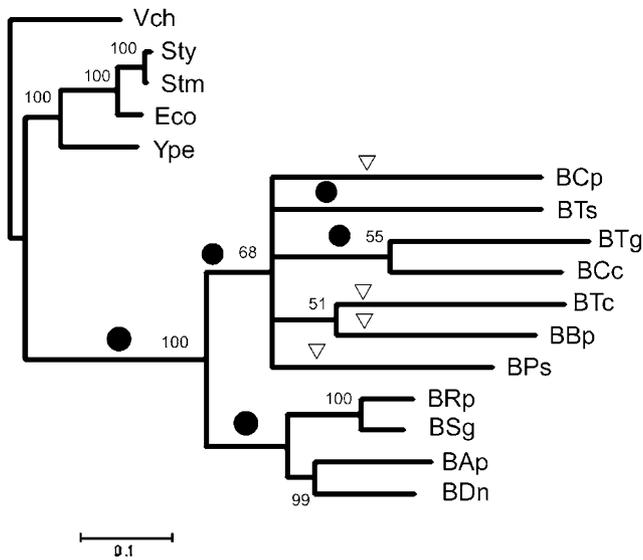


Figure 3 Phylogenetic tree obtained by maximum likelihood from a concatenated amino-acid alignment of LeuA, LeuB, LeuC and LeuD proteins. Abbreviations of the names of the *B. aphidicola* strains are shown in Table 2. Free-living bacterial species included are *E. coli* K12 (Eco, accession number: NC_000913), *Salmonella typhimurium* LT2 (Stm, acc. number: NC_003197), *S. enterica* subsp. *enterica* serovar Typhi str. CT18 (Sty, acc. number: NC_003198), *Yersinia pestis* CO92 (Ype, acc. number: NC_003143) and the outgroup *V. cholerae* O1 biovar eltor str. N16961 (Vch, acc. number: NC_002505). The initial alignment was edited using G-BLOCKS (Castresana, 2000) to select the amino-acid positions with the greater phylogenetic information. This edition reduced the alignment to 1075 amino-acid positions (67% of the original one). The phylogenetic tree was inferred by maximum likelihood (ML) combined with the quartet puzzling algorithm implemented in TREEPUZZLE 5.2 (Schmidt et al, 2002). Values at nodes indicate proportion of quarters supporting the corresponding inner branch as determined by the quartet-puzzling method. Close circles indicate the presence of a plasmid in the lineage. Open triangles indicate the insertion of the leucine cluster in the bacterial chromosome.

bacterial species, using a concatenated alignment of the LeuA, LeuB, LeuC and LeuD proteins. The tree topology obtained in this phylogenetic reconstruction (Figure 3) agrees with previous ones (even showing the same uncertainties at the subfamily level, Ortiz-Rivas et al, 2003), indicating the common evolutionary origin of the genes located in the plasmids and in the chromosome. A hypothetical scenario of independent back-transfer from an ancestral leucine plasmid to the main chromosome is also shown.

The *trpEG* genes

B. aphidicola from aphids of the subfamily Aphidinae and some tribes of the subfamily Pemphiginae also contain tryptophan plasmids (Lai et al, 1996; Rouhbakhsh et al, 1996; van Ham et al, 1999), ranging in size from 3.0 to 12.8 kb, and containing the two first genes of the tryptophan pathway (*trpEG*) that encode the rate-limiting enzymes for the biosynthesis of tryptophan. The variability in size is mainly due to variation in the number of tandem repeats of these genes or pseudogenes (Lai et al, 1996). The other genes of the tryptophan pathway (*trpDC(F)BA*) remain in the main chromosome.

In the subfamily Pemphiginae, where the endosymbionts of the three studied tribes have the leucine cluster located in the chromosome, the *trpEG* genes are plasmids located in two cases: *B. aphidicola* BPs (tribe Pemphigini) and BTc (tribe Tetraneurini). However, in BBp and BSc (tribe Fordini), the *trpEG* genes are chromosomally encoded, thought unlinked to the remaining genes of the pathway (Lai et al, 1995; van Ham et al, 2003) (Table 2; BSc, the endosymbiont of *Schlechtendalia chinensis*, has not been included in the table because no data for leucine cluster are available). The two *trpEG* plasmids found in endosymbionts of the Pemphiginae are also structurally different. BTc contains a 3.0 kb plasmid that carries a single copy of the *trpEG* and an origin of replication *ori3.6*. The BPs plasmid contains the genes *trpEG* and *repAC* plus a variable number of tandem repeats of a 1.8 kb unit carrying *repAC*, *trpG* and remnants of *trpE* (van Ham et al, 1999).

In the symbionts of the subfamily Aphidinae, two different *trpEG* plasmids have been found, which are similar to the plasmids found in the Pemphiginae. All of them hold at least one copy of the *trpEG* genes plus the *ori3.6* region, and resemble the plasmid found in BTc. Some of them include a variable number of the *trpEG* genes or pseudogenes (Lai et al, 1996; Rouhbakhsh et al, 1996). The plasmid of *B. aphidicola* BRm (from *Ropalosiphum maidis*) also contains the gene *repAC*, and therefore it is similar to the plasmid found in BPs. In BRp, *repAC* appears as a pseudogene. These findings suggest a common ancestor for all these tryptophan plasmids, which probably contained a *repAC* gene that has been lost and replaced by the *ori3.6* region in some modern lineages (van Ham et al, 1999).

Phylogenetic analyses carried out by van Ham et al (1999) with TrpE and TrpG sequences from several Aphidinae endosymbionts and the Pemphiginae endosymbionts BTc, BPs and BSc showed that the plasmid-borne sequences of BTc and BPs are more closely related to the chromosomal *trpEG* sequence of BSc than to the Aphidinae plasmid-borne sequences. Although only *trpEG* genes from *B. aphidicola* belonging to two aphid subfamilies have been studied at present (Aphidinae and Pemphiginae), these results support a common origin of the genes located either in a plasmid or in the chromosome, as for the leucine genes.

A proposed evolutionary scenario

Since the LCSA, the evolutionary dynamics have been very different for the *B. aphidicola* chromosome and plasmids. The extraordinary gene-order conservation of the main chromosome contrasts starkly with the versatility of leucine and tryptophan plasmids, regarding gene content and location, in the different *B. aphidicola* lineages.

We propose that the leucine cluster was transferred from the main chromosome to a *repA* plasmid at some stage between the LCA and the LCSA of *B. aphidicola*. Since then, at least four independent back-transfers have occurred throughout *B. aphidicola* evolution (Figure 3). The ancestral LCSA leucine plasmid would contain the four leucine genes and at least one *repA*, plus two additional genes, *ibp* and *yqhA*, resembling the BTs leucine plasmid (Figure 2). Regarding the *trpEG* genes,

although the analysis was carried out in *B. aphidicola* from only two aphid subfamilies, an evolutionary scenario similar to the one we propose for the leucine genes has also been postulated. The transfer to a plasmid carrying a repA/C-like replicon would have occurred in the ancestor of all present-day *B. aphidicola* strains, followed by independent events of replicon replacement, and back-transfer of *trpEG* to the chromosome in different lineages (van Ham et al, 1999). Thus, the ancestral LCSA tryptophan plasmid would contain the gene *repAC*, and at least one copy of *trpEG*.

The transfer of the bacterial rate-limiting genes for the biosynthesis of leucine and tryptophan from the main chromosome to plasmids must have been a key event for the successful adaptation of the ancestor of modern aphids to unbalanced diets, by avoiding the regulatory feedback control of both operons. However, extant aphid species are extremely diverse biologically and ecologically, and some of them may require relatively little leucine or tryptophan provisioning from their endosymbionts. During the coevolution of *B. aphidicola* and the aphids, it must have been advantageous to adjust the amino-acid biosynthesis to the insects' needs. A first step in the regulation of the levels of such amino acids might be related to the control of plasmid copy number (Thao et al, 1998; Plague et al, 2003), as well as the number of tandem repeats of the plasmid-borne genes and its later pseudogenisation (in the case of the *trpEG* genes). At some point, the loss of *B. aphidicola* genes involved in the control of DNA replication and segregation led to polyploidy of the bacterial cells. In addition, other regulatory elements upstream to structural genes for the biosynthetic enzymes were also lost or altered (Moran et al, 2003). In this new scenario, the back-transfer of the plasmid-borne genes to the chromosome might have been advantageous, which would explain why, in several *B. aphidicola* lineages, these genes are present in the main chromosome. At least four independent events of back-transfer have been identified for the leucine cluster. Since *recA* is absent from the three sequenced *B. aphidicola* genomes, we have postulated that the insertions would have been mediated by the *recBCD* system that has been retained (Sabater-Muñoz et al, 2004). This same system could be responsible for the gene-order rearrangements found in the different leucine plasmids. Finally, the cryptic plasmid found in the *B. aphidicola* associated with the Pemphiginae would be a remnant of the back-transfer of the leucine genes to the chromosome, instead of a cryptic *repA* plasmid present in the LCSA, as previously proposed (van Ham et al, 2000). Probably, these plasmids are retained because they contain essential genes (*ibp* or *yqhA*) that were not transferred to the chromosome in the recombination event.

Acknowledgements

We thank Dr C Biéumont for his kind invitation to write this review and to an anonymous reviewer for his/her valuable comments. This research was funded by grants Grupos03/204 from Govern Valencià, and BFM2003-00305 from Ministerio de Ciencia y Tecnología (MCyT) to A Moya. RG is a recipient of a contract in the Ramon y Cajal Program from the MCyT, Spain.

References

- Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M et al (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat Genet* **32**: 402–407.
- Andersson SG, Kurland CG (1998). Reductive evolution of resident genomes. *Trends Microbiol* **6**: 263–268.
- Baumann L, Baumann P, Moran NA, Sandström J, Thao ML (1999). Genetic characterization of plasmids containing genes encoding enzymes of leucine biosynthesis in endosymbionts (*Buchnera*) of aphids. *J Mol Evol* **48**: 77–85.
- Baumann P, Moran NA, Baumann L (2000). Bacteriocyte-associated endosymbionts of insect. In: Dworkin M (ed) *The Prokaryotes*. Springer-Verlag: New York, pp 1–67 (Online). <http://link.springer.de/link/service/books/10125/>.
- Belda E, Moya A, Silva FJ (2005). Genome rearrangement distances and gene order phylogeny in γ -proteobacteria. *Mol Biol Evol* **22**: 1456–1467.
- Blattner FR, Plunkett III G, Bloch CA, Perna NT, Burland V, Riley M et al (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **227**: 1453–1474.
- Bracho AM, Martínez-Torres D, Moya A, Latorre A (1995). Discovery and molecular characterization of a plasmid localized in *Buchnera* sp. bacterial endosymbionts of the aphid *Rhopalosiphum padi*. *J Mol Evol* **41**: 67–73.
- Buchner P (1965). *Endosymbiosis of Animals with plants Microorganisms*. Interscience Publishers, Inc: New York.
- Campo N, Dias MJ, Daveran-Mingot ML, Ritzenthaler P, Le Bourgeois P (2004). Chromosomal constraints in Gram-positive bacteria revealed by artificial inversions. *Mol Microbiol* **51**: 511–522.
- Castresana J (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**: 540–552.
- Clark M, Moran NA, Baumann P (1999). Sequence evolution in bacterial endosymbionts having extreme base composition. *Mol Biol Evol* **16**: 1586–1598.
- Douglas AE (1998). Nutritional interactions in insect–microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Ann Rev Entomol* **43**: 17–37.
- Fares MA, Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E (2002). GroEL buffers against deleterious mutations. *Nature* **417**: 398.
- Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD et al (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397–402.
- Gil R, Sabater-Muñoz B, Latorre A, Silva FJ, Moya A (2002). Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. *Proc Nat Acad Sci USA* **99**: 4454–4458.
- Gil R, Silva FJ, Zientz E, Delmotte F, González-Candelas F, Latorre A et al (2003). The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc Nat Acad Sci USA* **100**: 9388–9393.
- Gómez-Valero L, Latorre A, Silva FJ (2004a). The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont *Buchnera aphidicola*. *Mol Biol Evol* **21**: 2172–2181.
- Gómez-Valero L, Soriano-Navarro M, Pérez-Brocá V, Heddi A, Moya A, García-Verdugo JM et al (2004b). Coexistence of *Wolbachia* with *Buchnera aphidicola* and a secondary symbiont in the aphid *Cinara cedri*. *J Bacteriol* **186**: 6626–6633.
- Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ et al (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**: 477–483.
- Heie OE (1987). Paleontology and Phylogeny. In: Minks AK, Harrewijn P (eds) *Aphids. Their Biology, Natural Enemies and Control*. Elsevier: Amsterdam, pp 367–391.
- Klasson L, Andersson SGE (2004). Evolution of minimal-genesets in host-dependent bacteria. *Trends Genet* **12**: 37–43.
- Koga R, Tsuchida T, Fukatsu T (2003). Changing partners in an obligate symbiosis: a facultative endosymbiont can compen-

- sate for loss of the essential endosymbiont *Buchnera* in an aphid. *Proc R Soc Lond B* **270**: 2543–2550.
- Komaki K, Ishikawa H (1999). Intracellular bacterial symbionts of aphids possess many genomic copies per bacterium. *J Mol Evol* **48**: 717–722.
- Kondo N, Nikoh N, Ijichi N, Shimada M, Fukatsu T (2002). Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proc Nat Acad Sci USA* **99**: 14280–14285.
- Lai CY, Baumann L, Baumann P (1994). Amplification of *trpEG*: adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. *Proc Nat Acad Sci USA* **91**: 3819–3823.
- Lai CY, Baumann P, Moran NA (1995). Genetics of the tryptophan biosynthetic pathway of the prokaryotic endosymbiont (*Buchnera*) of the aphid *Schlechtendalia chinensis*. *Insect Mol Biol* **4**: 47–59.
- Lai CY, Baumann P, Moran NA (1996). The endosymbiont (*Buchnera* sp.) of the aphid *Diruaphis noxia* contains plasmids consisting of *trpEG* and tandem repeats of *trpEG* pseudo-genes. *Appl Environ Microbiol* **62**: 332–339.
- Latorre A, Gil R, Silva FJ, Martínez-Torres D, Moya A (2003). Tempo and mode of genome evolution in endosymbiotic bacteria of insects: the case of *Buchnera aphidicola*. *Symbiosis* **34**: 301–316.
- Lefèvre C, Charles H, Vallier A, Delobel B, Farell B, Heddi A (2004). Endosymbiont phylogenies in the Dryophoridae weevils: evidence for bacterial replacement. *Mol Biol Evol* **21**: 965–973.
- Margulis L (1981). *Symbiosis in Cell Evolution*. Freeman: San Francisco.
- Moran NA (1996). Accelerated evolution and Muller's ratchet endosymbiotic bacteria. *Proc Nat Acad Sci USA* **93**: 2873–2878.
- Moran NA (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**: 583–586.
- Moran NA, Mira A (2001). The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol* **2** RESEARCH 0054.
- Moran NA, Munson MA, Baumann P, Ishikawa H (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc R Soc Lond B* **253**: 167–171.
- Moran NA, Plague GR, Sandström JP, Wilcox JL (2003). A genomic perspective on nutrient provisioning by bacterial symbionts of insects. *Proc Nat Acad Sci USA* **100**: 14543–14548.
- Moya A, Latorre A, Sabater B, Silva FJ (2002). Comparative molecular evolution of primary (*Buchnera*) and putative secondary endosymbionts of aphids based on two protein-coding genes. *J Mol Evol* **54**: 127–137.
- Munson MA, Baumann P, Kinsey MG (1991). *Buchnera* gen. Nov. and *Buchnera aphidicola* sp. Nov., a taxon consisting of the mycetocyte-associated, primary endosymbionts of aphids. *Int J Syst Bacteriol* **41**: 566–568.
- Ortiz-Rivas B, Moya A, Martínez-Torres D (2003). Molecular systematic of aphids (Homoptera: Aphididae): new insights from the long-wave length Opsina gene. *Mol Phyl Evol* **30**: 24–37.
- Pérez-Brocá V, Latorre A, Gil R, Moya A (2005). Comparative analysis of two genomic regions among four strains of *Buchnera aphidicola*, primary endosymbiont of aphids. *Gene* **345**: 73–80.
- Plague GR, Dale C, Moran NA (2003). Low and homogeneous copy number of plasmid-borne symbiont genes affecting host nutrition in *Buchnera aphidicola* of the aphid *Uroleucon ambrosiae*. *Mol Ecol* **12**: 1095–1100.
- Rio VM, Lefèvre C, Heddi A, Aksoy S (2003). Comparative genomics of insect-symbiotic bacteria: influence of host environment on microbial genome composition. *Appl Environ Microbiol* **69**: 6825–6832.
- Rouhbakhsh D, Lai CY, von Dohlen C, Clark MA, Baumann L, Baumann P et al (1996). The tryptophan biosynthetic pathway of aphids endosymbionts (*Buchnera*): genetics and evolution of plasmid-associated anthranilate synthase (*trpEG*) within the Aphididae. *J Mol Evol* **42**: 414–421.
- Sabater-Muñoz B, Gómez-Valero L, van Ham RCHJ, Silva FJ, Latorre A (2002). Molecular characterization of the leucine cluster in *Buchnera* sp. Strain PSY, a primary endosymbiont of the aphid *Pemphigus spyrothecae*. *Appl Environ Microbiol* **68**: 2572–2575.
- Sabater-Muñoz B, van Ham RCHJ, Moya A, Silva FJ, Latorre A (2004). Evolution of the leucine gene cluster in *Buchnera aphidicola*: insights from chromosomal version of the cluster. *J Bacteriol* **186**: 2646–2654.
- Sasaki T, Hayashi H, Ishikawa H (1991). Growth and reproduction of symbiotic and aposymbiotic pea aphids, *Acyrtosiphon pisum*, maintained on artificial diets. *J Insect Physiol* **37**: 85–92.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**: 502–504.
- Shigenobu S, Watanabe H, Hattori M, Ishikawa H (2000). Genome sequence of the endocellular bacterial symbionts of aphids *Buchnera* sp. APS. *Nature* **407**: 81–86.
- Silva FJ, Latorre A, Moya A (2001). Genome size reduction through multiple events of gene disintegration in *Buchnera* APS. *Trends Genet* **17**: 615–618.
- Silva FJ, Latorre A, Moya A (2003). Why are the genomes of endosymbiotic bacteria so stable? *Trends Genet* **19**: 176–180.
- Silva FJ, van Ham RCHJ, Sabater-Muñoz B, Latorre A (1998). Structure and evolution of the leucine plasmids carried by the endosymbionts (*Buchnera aphidicola*) from aphids of the family Aphididae. *FEMS Microbiol Lett* **168**: 43–49.
- Soler T, Latorre A, Sabater B, Silva FJ (2000). Molecular characterization of the leucine plasmid from *Buchnera aphidicola*, primary endosymbionts of the aphid *Acyrtosiphon pisum*. *Curr Microbiol* **40**: 264–268.
- Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ et al (2002). 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**: 2376–2379.
- Tatusov RL, Koonin EV, Lipman DJ (1997). A genomic perspective on protein families. *Science* **278**: 631–637.
- Thao ML, Baumann L, Bauman P, Moran NA (1998). Endosymbionts (*Buchnera*) from the aphids *Schizaphis graminum* and *Diuraphis noxia* have different copy numbers of the plasmid containing the leucine biosynthetic genes. *Curr Microbiol* **36**: 238–240.
- van Ham R, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U et al (2003). Reductive genome evolution in *Buchnera aphidicola*. *Proc Nat Acad Sci USA* **100**: 581–586.
- van Ham RCHJ, Gonzalez-Candelas F, Silva FJ, Sabater B, Moya A, Latorre A (2000). Postsymbiotic plasmid acquisition and evolution of the *repA1*-replicon in *Buchnera aphidicola*. *Proc Nat Acad Sci USA* **97**: 10855–10860.
- van Ham RCHJ, Martínez-Torres D, Moya A, Latorre A (1999). Plasmid-encoded anthranilate synthase (*trpEG*) in *Buchnera aphidicola* from aphids of the family Pemphigidae. *Appl Environ Microbiol* **65**: 117–125.
- 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.
- von Dohlen CD, Moran NA (2000). Mitochondrial ribosomal sequences and fossils support a rapid radiation of aphids in the Cretaceous and multiple origins of a complex life cycle. *Biol J Linn Soc* **71**: 689–717.
- Welch RA, Burland V, Plunkett III G, Redford P, Roesch P, Rasko D et al (2002). Extensive mosaic structure revealed by the genome sequence of urophthogenic *Escherichia coli*. *Proc Nat Acad Sci USA* **99**: 17020–17024.