Gene size reduction in the bacterial aphid endosymbiont, Buchnera.
H. Charles, Dominique Mouchiroud, J.R. Lobry, I. Gonçalves, Y. Rahbe

To cite this version:

HAL Id: hal-00402189
https://hal.archives-ouvertes.fr/hal-00402189
Submitted on 26 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright
Aphids that feed solely on phloem sap, a diet poor in nitrogenous compounds, harbor intracellular bacteria of the genus *Buchnera* in specialized cells, called bacteriocytes, within their body cavities (Buchner 1965). *Buchnera* are members of the *Enterobacteriaceae* (Munson, Baumann, and Kinsey 1991), and they are transmitted to successive host insect generations via oocyte reinfection (Buchner 1965). Congruent host/bacterium phylogenies suggest that this symbiotic association developed 200–250 MYA (Moran and Baumann 1994), and it is now so close that neither partner could survive independently (Houk and Griffith 1980).

The work of Moran (1996) and Brynnel et al. (1998) demonstrated that several *Buchnera* genes exhibit accelerated rates of sequence evolution and comparatively low ratio of synonymous-to-nonsynonymous substitutions for the coding sequences. These findings were attributed to an increased rate of nonsynonymous substitutions, resulting from the population dynamics of *Buchnera* combined with a mutational bias toward A+T. *Buchnera* populations are isolated within the bacteriocytes and experience recurrent bottlenecks during transmission from one host generation to the next, leading to low recombination frequencies between individuals (Moran 1996).

The genome of *Buchnera* contains about 30% G+C (Ishikawa 1987). Most intracellular symbiotic or parasitic bacteria seem to accumulate AT bases in their genomes (reviewed in Heddi et al. 1998). The most striking example is perhaps that of *Mycobacterium leprae*. This bacterium is the only obligate intracellular species of the GC-rich *Mycobacterium* group, characterized by 16S rDNA gene sequences between 56.5% and 59.6% GC, and it is also the most AT-rich species of the genus (56.5%). AT accumulation in obligate intracellular symbiotic or parasitic bacteria could result from an AT mutational bias in conjunction (or not) with the relaxation of selective pressure due to intracellular habitat (Heddi et al. 1998).

The *Buchnera* genome is extremely small, 657 kb compared with 4–5 Mb in related free-living *Enterobacteriaceae* (Charles and Ishikawa 1999). This seems to be a common characteristic of many intracellular symbiotic bacteria, like the endosymbionts of the para-
kb of *Buchnera* gene. *Buchnera* genes were also smaller than those of *H. influenzae*. However, this difference was not significant (*P* = 0.13, *d* = 5.3 ± 0.04 bp/kb), probably because gene shortening also occurred in *H. influenzae*, characterized by a small AT-rich genome (1.83 Mb, 38% GC) (Oliver and Marin 1996).

In order to determine if deletions occur preferentially at the 3′ end (Oliver and Marin hypothesis) or within sequences (bias in deletion versus insertion rate), we analyzed separately 3′-end deletions and within-sequence deletions in *Buchnera* genes. The 3′ ends of *Buchnera* genes seem to have undergone more deletions than the 3′ ends of *E. coli* genes (*P* = 0.01, *d* = 2.8 ± 1.0 bp/kb). The difference between *Buchnera* and *H. influenzae* 3′-end lengths was not significant (*P* = 0.34, *d* = 1.0 ± 0.09 bp/kb).

The within-sequence deletion and insertion ratios were calculated using the interface program JaDis (Gonçalves et al. 1999). This program allows discrimination between insertion and deletion events based on phylogenetic relationships between *Buchnera*, *E. coli*, and *H. influenzae* (fig. 1). In *Buchnera*, the mean deletion ratio was 6.8 ± 1.5 deletions/kb, with a mean deletion size *s* = 7.7 ± 1.0 bp (two or three amino acids). The mean insertion ratio was estimated to be 3.6 ± 1.2 insertions/kb (*s* = 5.9 ± 0.9 bp), giving a net ratio of 3.2 ± 1.9 deletions/kb. As alignments were performed with amino acid sequences and translated afterward, indel sizes are always multiples of 3 bp. However, visual inspection of our data set revealed that single-base indels, which would have generated “out-of-phase” stretches before being restored by a compensatory deletion/insertion event, seem to be very rare. The corresponding values were also calculated for the *E. coli* genes. The mean deletion rate was 2.59 ± 0.7 deletions/kb (*s* = 5.5 ± 1 bp), and the mean insertion rate was 3.07 ± 0.2 insertions/kb (*s* = 11.4 ± 4.5 bp), indicating a net ratio of 0.48 ± 0.7 insertions/kb. The latter result indicates that gene shortening in *E. coli* does not seem to have occurred via insertion accumulation since the divergence with *Buchnera*.

Figure 2A and B suggests that both 3′-end deletions and within-sequence deletions in *Buchnera* may be longer in the AT-rich genes than in the GC-rich genes. These results may be explained by the high correlation between GC contents and evolutionary rates (data not shown), with the less conserved genes being the most AT-rich genes in *Buchnera*. Similar findings have been reported for eukaryotes, for which pseudogenes are shorter and have higher AT contents than their functional homologs (Graur, Shuali, and Li 1989; Gu and Li 1995).

Taken together, these results imply that *Buchnera* genes accumulate both deletions and AT bases in their sequences. Nevertheless, gene shortening cannot account for genome reduction in *Buchnera*, and as in other symbiotic or parasitic bacteria, whole genes or even chromosome regions have probably been deleted, resulting in the drastically reduced number of genes in these bacteria. The *Buchnera* genome (657 kb) has shrunk about 86% compared with the *E. coli* genome (Charles and Ishikawa 1999), and deletion bias can only account for 0.8% of this. Actually, deletion accumulation probably reflects defective DNA synthesis and repair mechanisms in the bacterium, resulting from low recombination frequencies (Muller’s ratchet), as previously suggested by Moran (1996). Indeed, it is hard to believe that single deletions (corresponding to <0.0005% genome size reduction) could confer any decisive selective advantage on individuals within the small populations of *Buchnera*.

Oliver and Marin (1996) observed that AT-rich genes are shorter than GC-rich ones in bacteria. We have demonstrated here that stop codon density (i.e., AT bias) may not be the sole neutral factor governing gene size variation in bacteria and that gene shortening may result from a combined effect of AT bias (3′-end deletions) and deletion bias (within-sequence deletions), both enhanced by the relaxation of the selective pressure due to the intracellular habitat of *Buchnera*.

### Acknowledgments

This work was supported by a grant from the CNRS-INRA “Programme national de Diversité” (AIP 98/P00202). We thank INIST for correcting the English text.

### Literature Cited


Howard Ochman, reviewing editor

Accepted August 30, 1999