# Decoupling of Genome Size and Sequence Divergence in a Symbiotic Bacterium

JENNIFER J. WERNEGREEN,* HOWARD OCHMAN, ISAAC B. JONES, AND NANCY A. MORAN

*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721*

In contrast to genome size variation in most bacterial taxa, the small genome size of *Buchnera* sp. was shown to be highly conserved across genetically diverse isolates (630 to 643 kb). This exceptional size conservation may reflect the inability of this obligate mutualist to acquire foreign DNA and reduced selection for genetic novelty within a static intracellular environment.

Bacteria may undergo significant changes in genome size through the duplication or loss of existing chromosome fragments and through the acquisition of exogenous DNA, including plasmids or other accessory elements. Because the vast majority of a bacterial chromosome consists of coding sequences (4), changes in genome size reflect differences in gene content. The variation among bacterial genome sizes, ranging from 0.6 to 9 Mb (15), reflects differences in biochemical capabilities and, hence, in the range of environments available to particular microbial lineages.

Based on early comparisons of genetic maps, the size, organization, and gene content of bacteria were viewed as being evolutionarily conserved (20). However, the physical mapping of bacterial genomes by pulsed-field gel electrophoresis (PFGE) has provided overwhelming evidence that the size and structure of the bacterial chromosome varies both within and among species (6). Genome sizes of very closely related taxa can differ by as much as 100%, often by more than a megabase in total length (Table 1). Recent evidence from these studies and from analyses of full genome sequences indicates that genome size and organization are more evolutionarily labile than gene sequences (12). This lability extends both to those bacteria with relatively large genomes, such as the enteric bacteria and *Bacillus* species, and to those with small genomes, such as the mycoplasmas and rickettsiae (Table 1).

Bacterial genome size is determined by a balance between the introduction of new sequences, through duplication or horizontal transfer, and the loss of genes, through mutations that eliminate function and subsequent deletions. Whether or not a given gene persists within a lineage depends upon the selection coefficient associated with its functional role, the effective population size, and mutation rate (13). Thus, in bacteria with small effective population sizes and high mutation rates, the selection required to maintain a given gene increases. The expected dependence of genome size on population size and mutation rate is supported by the repeated observation of reduced genome sizes in intracellular pathogenic bacteria (2, 25). Chronic pathogens and symbionts may experience small population sizes due to bottlenecks during infection of hosts (1, 16) and higher per-site mutation rates (19). These factors would increase the selection coefficient required to maintain a gene, so that only the most essential genes persist, resulting in small genomes as observed for these organisms.

Intracellular bacteria might be expected to show conservation of genome size if they retain only essential genes for a specialized lifestyle. However, although they show consistently

TABLE 1. Genome size range and divergence in 16S rDNA for representative bacterial taxa

| Organism | % 16S rDNA divergence | Genome size range (Mb) | % Difference in genome size | Source or reference (for genome size) |
|---|---|---|---|---|
| **Free-living/opportunistic pathogens** | | | | |
| *Escherichia coli* | <1 | 4.5–5.5 | 22 | 4 |
| *Salmonella typhi* | <1 | 3.9–4.9 | 25 | 23 |
| *Bacillus cereus* | 1 | 2.4–5.3 | 121 | 5 |
| **Chronic pathogens** | | | | |
| Spotted fever group of *Rickettsia* spp. | <3 | 1.2–2.1 | 75 | 21 |
| *Mycoplasma hominis* | <1 | 0.70–0.83 | 19 | 14, 18 |
| *Mycoplasma genitalium-M. pneumoniae* | 2 | 0.58–0.82 | 41 | 9, 11 |
| *Chlamydia psittaci-C. trachomatis* | 5 | 1.0–2.65 | 165 | 10, 22 |
| **Vertically transmitted symbiont** | | | | |
| *Buchnera aphidicola* | 6–8 | 0.63–0.64 | <3 | 7; this study |

* Corresponding author. Mailing address: Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, The Marine Biological Laboratory, 7 MBL St., Woods Hole, MA 02543. Phone: (508) 548-3705, ext. 6650. Fax: (508) 457-4727. E-mail: jwernegreen @mbl.edu.
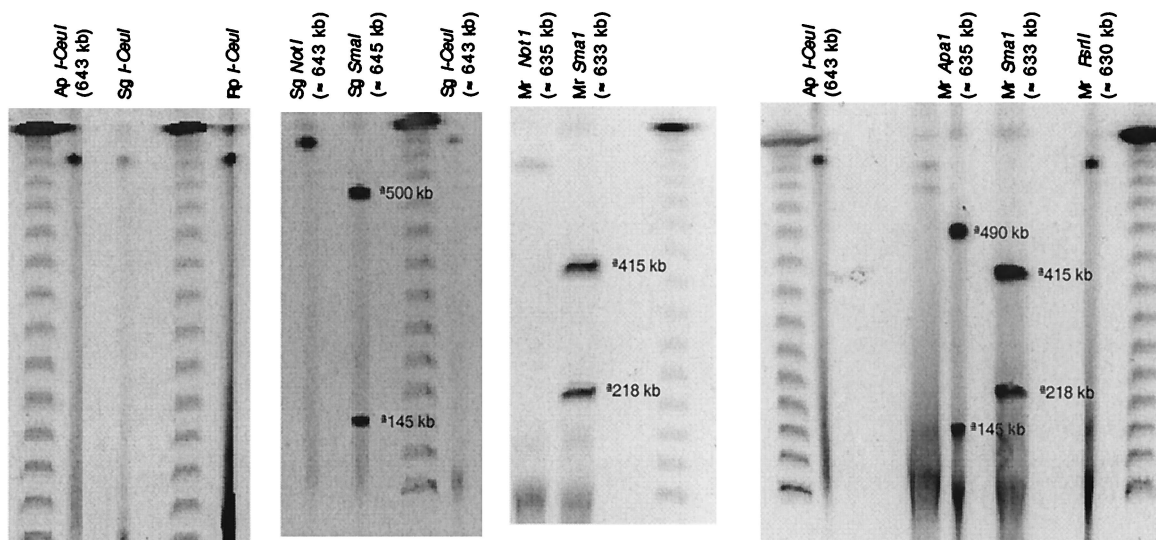
FIG. 1. Conservation of genomes sizes in *Buchnera*. *Buchnera* chromosomes digested with rare-cutting restriction enzymes were run on pulsed field gels to estimate genome sizes. Lanes are labeled by the abbreviation for the host name (Ap, *A. pisum*; Rp, *R. padi*; Sg, *S. graminum*; Mr, *M. rhois*). Note that *I-CeuI* cuts once and linearizes the *Buchnera* chromosome. With this enzyme, genomes sizes of *Buchnera* from *S. graminum*, *R. padi*, and *A. pisum* are indistinguishable. The finding of a slightly smaller genome size of *Buchnera* from *M. rhois* is supported by several independent enzyme digests.

small genomes, intracellular pathogens display wide variation in chromosome length (Table 1). This variation suggests the incidence of lateral transfer of DNA, as supported by findings of gene transfer between distantly related intracellular pathogens (24).

The intracellular mutualistic symbiont of aphids, *Buchnera aphidicola*, has an extremely reduced genome (643 kb) based on PFGE of symbionts of the host species *Acyrthosiphon pisum* (7; H. Ishikawa, personal communication). The clade corresponding to *Buchnera* is over 100 million years old and has cospeciated with aphids during this period (17). Based on observations for pathogenic bacteria with small genomes, *Buchnera* spp. of different aphid lineages might be expected to differ in genome size. Applying methods developed by Charles and Ishikawa (7), genome sizes of *Buchnera* lineages, including isolates from *A. pisum*, *Rhopalosiphum padi*, *Schizaphis graminum*, and *Melaphis rhois*, were determined using PFGE. Symbiont DNA was prepared using filtration techniques described previously (7), modified by the addition of 100 mM EDTA and

the omission of $MgCl_2$ in buffer A. *Buchnera* chromosomes were digested with several rare-cutting restriction enzymes (*I-CeuI*, *NotI*, *SmaI*, *ApaI*, and *RsrII*) and run on pulsed field gels. Fragment sizes were calculated by comparison to concatemers of lambda DNA (Fig. 1).

Genome sizes in *Buchnera* vary by <5%, even among strains that diverged 100 to 200 million years ago and that show substantial divergence in rRNA sequences (Fig. 2). In contrast, other bacterial clades with less sequence divergence consistently have more variation in genome size. The uniformity of genome size in *Buchnera* strongly suggests that genome shrinkage occurred early in the evolution of the symbiosis and that modern lineages retain the genome size of a common ancestor.

Two factors may contribute to the conservation of genome size in *Buchnera*. First, unlike pathogens that move horizontally among hosts and experience coinfection, *Buchnera* is entirely vertically transmitted. As a result, bacteria within a single host are genetically homogeneous, preventing opportunity for the uptake of novel genes. Second, *Buchnera* is confined permanently to a single host lineage and to a mutualistic lifestyle that depends on a static set of physiological capabilities (3). In contrast, pathogens sometimes acquire new host taxa or new ways of exploiting hosts, providing a selective context for the stable incorporation of new genes.

The upcoming availability of full genome sequences of *Buchnera* strains (H. Ishikawa, personal communication) will show whether the conservation of genome size corresponds to a uniformity in gene inventories and genome organization. Furthermore, these full genome sequences will reveal whether specific genetic features, such as lack of translocatable elements or the loss of recombinase pathways, contribute to this unusual stability in genome size.
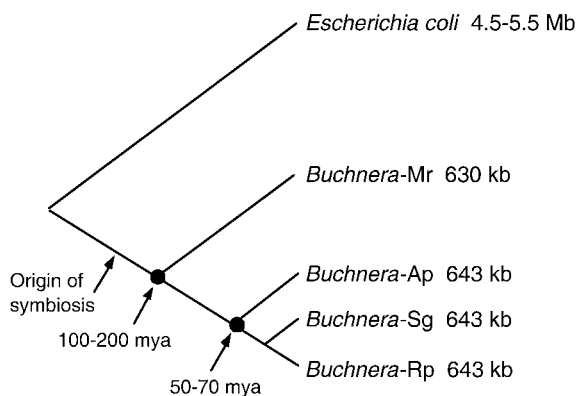


FIG. 2. Relationships and genome sizes of *Buchnera* taxa. Host abbreviations are as given in Fig. 1. Divergence times are based on the aphid fossil record (8, 17).

## REFERENCES

1. **Andersson, J. O., and S. G. E. Andersson.** 1999. Genome degradation is an ongoing process in *Rickettsia*. Mol. Biol. Evol. **16:**1178–1191.
2. **Andersson, S. G. E., and C. G. Kurland.** 1998. Reductive evolution of resident genomes. Trends Microbiol. **6:**263–268.
3. **Baumann, P., N. A. Moran, and L. Baumann.** 1997. The evolution and genetics of aphid endosymbionts. Bioscience **47:**12–20.
4. **Bergthorsson, U., and H. Ochman.** 1998. Distribution of chromosome length variation in natural isolates of *Escherichia coli*. Mol. Biol. Evol. **15:**6–16.
5. **Carlson, C. R., and A. B. Kolsto.** 1994. A small (2.4 Mb) *Bacillus cereus* chromosome corresponds to a conserved region of a larger (5.3 Mb) *Bacillus cereus* chromosome. Mol. Microbiol. **13:**161–169.
6. **Casjens, S.** 1998. The diverse and dynamic structure of bacterial genomes. Annu. Rev. Genet. **32:**339–377.
7. **Charles, H., and I. Ishikawa.** 1999. Physical and genetic map of the genome of *Buchnera*, the primary endosymbiont of the pea aphid, *Acyrthosiphon pisum*. J. Mol. Evol. **48:**142–150.
8. **Clark, M. A., N. A. Moran, and P. Baumann.** 1999. Sequence evolution in bacterial endosymbionts having extreme base composition. Mol. Biol. Evol. **16:**1586–1598.
9. **Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. G. Sutton, J. M. Kelley, J. L. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. L. Fuhrmann, D. T. Nguyen, T. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J. Tomb, B. A. Dougherty, K. F. Bott, P. C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, and J. C. Venter.** 1995. The minimal gene complement of *Mycoplasma genitalium*. Science **270:**397–403.
10. **Frutos, R., M. Pages, M. Bellis, G. Roizes, and M. Bergoin.** 1989. Pulsed-field gel electrophoresis determination of the genome size of obligate intracellular bacteria belonging to the genera *Chlamydia*, *Rickettsiella*, and *Porochlamydia*. J. Bacteriol. **171:**4511–4513.
11. **Himmelreich, R., H. Hilbert, H. Plagens, E. Pirkl, B. C. Li, and R. Herrmann.** 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucleic Acids Res. **24:**4420–4449.
12. **Huynen, M. A., and P. Bork.** 1998. Measuring genome evolution. Proc. Natl. Acad. Sci. USA **95:**5849–5856.
13. **Lawrence, J. R., and J. G. Roth.** 1999. Genomic flux: genome evolution by gene loss and acquisition, p. 263–289. *In* R. Charlesbois (ed.), Organization of the prokaryotic genome. American Society for Microbiology, Washington, D.C.
14. **Ladefoged, S. A., and G. Christiansen.** 1992. Physical and genetic mapping of the genomes of five *Mycoplasma hominis* strains by pulsed-field gel electrophoresis. J. Bacteriol. **174:**2199–207.
15. **Maniloff, J.** 1996. The minimal cell genome: "on being the right size." Proc. Natl. Acad. Sci. USA **93:**10004–10006.
16. **Moran, N. A.** 1996. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **93:**2873–2878.
17. **Moran, N. A., M. A. Munson, P. Baumann, and H. Ishikawa.** 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. R. Soc. Lond. B **253:**167–171.
18. **Mygind, T., S. Birkelund, and G. Christiansen.** 1998. DNA sequencing reveals limited heterogeneity in the 16S rRNA gene from the *rrnB* operon among five *Mycoplasma hominis* isolates. Int. J. Syst. Bacteriol. **48:**1067–1071.
19. **Ochman, H., S. Elwyn, and N. A. Moran.** 1999. Calibrating bacterial evolution. Proc. Natl. Acad. Sci. USA **96:**12638–12643.
20. **Riley, M., and A. Anilionis.** 1978. Evolution of the bacterial genome. Annu. Rev. Microbiol. **32:**519–560.
21. **Roux, V., M. Drancourt, and D. Raoult.** 1992. Determination of genome sizes of *Rickettsia* spp. within the spotted fever group, using pulsed-field gel electrophoresis. J. Bacteriol. **174:**7455–7457.
22. **Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin, and R. W. Davis.** 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science **282:**754–759.
23. **Thong, K. L., S. D. Puthucheary, and T. Pang.** 1997. Genome size variation among recent human isolates of *Salmonella typhi*. Res. Microbiol. **148:**229–235.
24. **Wolf, Y. I., L. Aravind, and E. V. Koonin.** 1999. Rickettsiae and chlamydiae: evidence of horizontal gene transfer and gene exchange. Trends Genet. **15:**173–175.
25. **Zomorodipour, A., and S. G. E. Andersson.** 1999. Obligate intracellular parasites: *Rickettsia prowazekii* and *Chlamydia trachomatis*. FEBS Lett. **452:**11–15.