

Genomes on the shrink

Howard Ochman*

Department of Biochemistry and Molecular Biophysics, University of Arizona, Tucson, AZ 85721

Because genetic events do not fossilize, we are forced to deduce the evolution of bacterial genomes by comparing the features of contemporary organisms. This comparative approach has exposed many of the alterations endured by bacterial genomes (bouts of expansion and contraction or changes in base compositions), but such reconstructions are often imprecise, and sometimes incorrect, because they are limited by the spectrum and relationships of the sequenced genomes that are available. Determining whether changes in gene repertoires and genome size were gradual or episodic is not feasible when the genomes being considered diverged several hundred million years ago. Fortunately, the results reported by Nilsson *et al.* (1) in this issue of PNAS indicate how such transformations proceed and help to explain one of the most interesting and pervasive trends in the evolution of bacterial genomes.

When analyzed in a molecular phylogenetic perspective, every clade of bacteria with genome sizes of <2 Mb was derived from ancestors with substantially larger genomes (Fig. 1). This pattern dispels the long-held notion that bacteria evolved by the successive doubling of small-genomed progenitors (2, 3) and raises numerous questions about an evolutionary process that seems to affect all bacterial lineages. Among the groups best suited for investigating the progression toward reduced genomes are the γ -proteobacteria, due principally to the large number of fully sequenced constituents (53 at last count). Within this phylum, which includes the workhorses of bacterial genetics and pathogenesis, *Escherichia coli* and *Salmonella typhimurium*, the sizes of already-sequenced genomes vary over an order of magnitude, from 600 kb in *Buchnera aphidicola* (4) to 7,000 kb in *Pseudomonas fluorescens* (5).

With the near-perfect correlation between genome size and gene number in bacteria (6), reductions in genome size will usually result in the loss of some functional capabilities. Given the observed range of genome sizes, what circumstances might allow elimination of 80% or more of the coding capacity of an organism? Those bacteria with the smallest genomes are intracellular pathogens and symbionts that maintain obligate associations with eukaryotic hosts. In these cases, the hosts provision

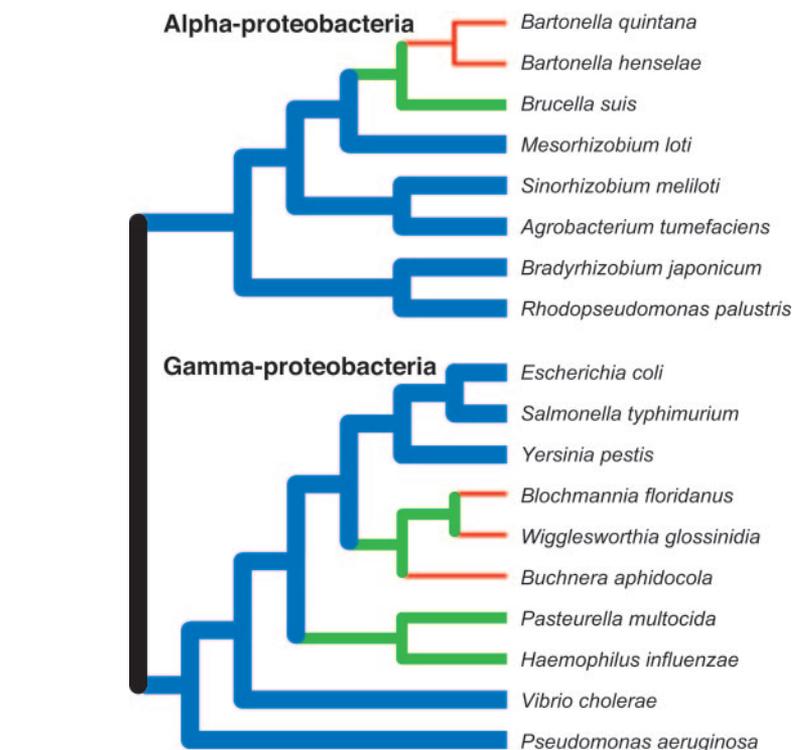


Fig. 1. Relationships of sequenced bacterial genomes showing that in both the α - and γ -proteobacteria, lineages with smaller genome sizes are derived from ancestors that had larger genomes. Branch widths and colors correspond to relative genome sizes as follows: red, <2 Mb; green, 2–4 Mb; blue, >4 Mb. Within the γ -proteobacteria, the insect endosymbionts *Buchnera*, *Wigglesworthia*, and *Blochmannia* form a clade in which all genomes are <700 kb. Differences in the gene repertoires of these symbionts indicate that the extreme reduction in genome size occurred independently after the lineages diverged. Figure adapted from phylogenies presented in refs. 18 and 19.

bacteria with a constant supply of nutrients, thereby rendering unnecessary many genes that were previously needed in less certain environments, such as those encountered by free-living bacteria.

Just because a gene is superfluous does not assure its removal from a genome. For example, the human genome maintains hundreds of nonfunctional olfactory receptor genes, including some that date to the origin of tetrapods (7, 8). However, as evident from comparisons of bacterial pseudogenes with their functional counterparts, the mutational process in bacterial genomes is strongly biased toward deletions (6, 9). Although nonfunctional regions can be maintained in a bacterial genome for some time, they gradually erode and are eventually eliminated.

The deletional bias observed in bacterial pseudogenes goes a long way toward explaining why bacterial genomes are compact and gene-rich, and why large

nonfunctional regions do not accumulate within their genomes. However, the extent to which this process has been responsible for the extreme reduction of symbiont genomes is difficult to evaluate using conventional methods that align and compare homologous sequences. Because the majority of genes present in their large-genome relatives are missing from highly reduced genomes, there is no information about the manner in which these sequences were eliminated: the extreme genome reduction could proceed by a slow and continual erosion of individual genes, or, alternatively, by expansive deletions that jettison numerous genes with each event. The presence of hundreds of pseudogenes scattered around the genomes of some recent pathogens (10–13), as well as the fact

See companion article on page 12112.

*E-mail: hochman@email.arizona.edu.

© 2005 by The National Academy of Sciences of the USA

that large deletions would often eliminate essential genes, support a scenario whereby genome reductions occur on a gene-by-gene basis. However, the wholesale disappearance of large stretches of genes suggests the broad-scale events may be integral to evolution of small genomes and establishing reliance on a host environment.

To determine the scale of deletion events, in terms of both their magnitude and their frequency, Nilsson *et al.* (1) monitored the changes occurring in wild-type and repair-defective (*mutS*⁻) lines of *Salmonella typhimurium* during propagation in the lab. *S. typhimurium*, a facultative pathogen with a genome size of 4,900 kb, contains virtually all of the genes now present in the drastically reduced genomes of *Buchnera*, *Blochmannia*, and *Wigglesworthia*, suggesting that it might be genetically similar to the ancestor of endosymbiotic lineages. Therefore, its pattern of genome evolution in a well supplemented growth environment might closely mimic the events that occur during the formation of symbioses.

In 4 of the 60 *mutS*⁻ lines, Nilsson *et al.* (1) detected deletions of regions up to 173 kb in length, and in a separate assay in which they selected for the simultaneous loss of two marker loci, they recovered strains with individual deletions spanning 100–200 kb. Given that

most bacterial genomes consist of genes that average ≈1 kb in length and that align almost contiguously along the chromosome, such deletions can instantly remove 5% of the genes in a genome. Although gene erosion is certainly operating on the contents and coding potential of bacterial genomes, large-scale deletions also are likely to play a crucial role, and even remove the majority of genes, during the initial stages of genome reduction.

Large-scale deletions are likely to play a crucial role during the initial stages of genome reduction.

Defining the precise endpoints of these deletions provides additional insights into the process of reductive evolution. Because the genomes of many pathogens are laden with mobile and repetitive elements, it has been suggested that deletions mediated by RecA-dependent exchange at long homologous repeats have been instrumental in shaping the genomes of endosymbionts (14).

(That the aphid endosymbiotic bacterium *Buchnera* has one of the smallest sequenced genomes and does not contain the *recA* gene might seem contrary to this proposal, but its loss of *recA* may well have succeeded the extreme genome reduction.) Only one of the deletions characterized by Nilsson *et al.* (1) was bounded by long regions of homology; in the rest of the cases, the deletion endpoints featured little, if any, homology, despite the presence of several sizeable repeats within each of the deleted regions.

Deletions of the size range detected in these experiments are not unknown in bacterial genomes. Highly reduced genomes have been engineered through the systematic deletion of very large regions in attempts to define the numbers of essential genes and the minimal gene set required for cellular life (15–17). The experiments of Nilsson *et al.* (1) show that changes of this sort will occur naturally, and frequently, even over the course of several weeks, so their impact on evolutionary time scales is apt to be enormous. The rapidity of these large-scale changes in genome content in laboratory experiments substantiates the view that some gene loss occurs can occur quickly in bacterial lineages that adopt chronically pathogenic or symbiotic lifestyles.

1. Nilsson, A., Koskineniemi, S., Eriksson, S., Kugelberg, E., Hinton, J. C. D. & Andersson, D. I. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 12112–12116.
2. Wallace, D. C. & Morowitz, H. J. (1973) *Chromosoma* **40**, 121–126.
3. Sparrow, A. H. & Nauman, A. F. (1976) *Science* **192**, 524–529.
4. van Ham, R. C., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernández, J. M., Jiménez, L., Postigo, M., Silva, F. J., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100**, 581–586.
5. Paulsen, I. T., Press, C. M., Ravel, J., Kobayashi, D. Y., Myers, G. S., Mavrodi, D. V., Deboy, R. T., Seshadri, R., Ren, Q., Madupu, R., *et al.* (2005) *Nat. Biotechnol.* **23**, 873–878.
6. Mira, A., Ochman, H. & Moran, N. A. (2001) *Trends Genet.* **17**, 589–596.
7. Malnic, B., Godfrey, P. A. & Buck, L. B. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 2584–2589.
8. Glusman, G., Yanai, I., Rubin, I. & Lancet, D. (2001) *Genome Res.* **11**, 685–702.
9. Andersson, J. O. & Andersson, S. G. E. (2001) *Mol. Biol. Evol.* **18**, 829–839.
10. Cole, S. T., Eiglmeier, K., Parkhill, J., James, K. D., Thomson, N. R., Wheeler, P. R., Honore, N., Garnier, T., Churcher, C., Harris, D., *et al.* (2001) *Nature* **409**, 1007–1011.
11. Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T., Prentice, M. B., Sebaihia, M., James, K. D., Churcher, C., Mungall, K. L., *et al.* (2001) *Nature* **413**, 523–527.
12. Jin, Q., Yuan, Z., Xu, J., Wang, Y., Shen, Y., Lu, W., Wang, J., Liu, H., Yang, J., Yang, F., *et al.* (2002) *Nucleic Acids Res.* **30**, 4432–4441.
13. Lerat, E. & Ochman, H. (2005) *Nucleic Acids Res.* **33**, 3125–3132.
14. Klasson, L. & Andersson, S. G. E. (2004) *Trends Microbiol.* **12**, 37–43.
15. Itaya, M. & Tanaka, T. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 5378–5382.
16. Kolisnychenko, V., Plunkett, G., 3rd, Herring, C. D., Fehér, T., Posfai, J., Blattner, F. R. & Pósfai, G. (2002) *Genome Res.* **12**, 640–647.
17. Westers, H., Dorenbos, R., van Dijk, J. M., Kabel, J., Flanagan, T., Devine, K. M., Jude, F., Seror, S. J., Beekman, A. C., Darmon, E., *et al.* (2005) *Mol. Biol. Evol.* **20**, 2076–2090.
18. Gil, R., Silva, F. J., Zientz, E., Delmotte, F., Gonzalez-Candelas, F., Latorre, A., Rausell, C., Kamerbeek, J., Gadau, J., Hölldobler, B., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100**, 9388–9393.
19. Boussau, B., Karlberg, E. O., Frank, A. C., Legault, B. A. & Andersson, S. G. E. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 9722–9727.