

CORRECTED 2 MARCH 2007 AND 15 FEBRUARY 2008; SEE LAST 2 PAGES

The 160-Kilobase Genome of the Bacterial Endosymbiont *Carsonella*

Atsushi Nakabachi,^{1,2*} Atsushi Yamashita,^{3†} Hidehiro Toh,^{3,4†} Hajime Ishikawa,⁵ Helen E. Dunbar,² Nancy A. Moran,² Masahira Hattori^{6,7*}

Many bacterial lineages have evolved mutually obligate endosymbiotic associations with animal hosts. In such cases, the bacteria typically produce essential nutrients that are rare in the host diet, and the animal produces specialized cells (bacteriocytes) where bacteria are confined and, like organelles, continuously maintained through vertical transmission across host generations. These bacteriocyte-restricted bacteria have distinctive genomic features, including massive reduction in genome size and biased nucleotide base composition (1). However, all cases of genome reduction appear to reach limits of about 400 kb and about 20% GC, which are believed to be the minimal limits for cellular organisms (Fig. 1).

Here, we present a genome that has evolved far beyond these limits. *Carsonella ruddii* (Fig. 1) is a bacteriocyte-associated γ -proteobacterial symbiont that appears to be present in all species of phloem sap-feeding insects, psyllids (2). We determined the complete genome sequence of *C. ruddii* strain Pv (*Carsonella*-Pv) of the hackberry petiole gall psyllid, *Pachypsylla venusta*, which has no other microbial symbionts (2). The genome of *Carsonella*-Pv is a single circular chromosome of 159,662 base pairs (bp), averaging 16.5% GC content. The assembly analysis, using a large excess of sequence data, did not reveal any other symbionts or plasmids. The genome size, which was further confirmed by long-range electrophoresis, is only about one-third that of the archaeal parasite *Nanoarchaeum equitans* (having the smallest fully sequenced genome to date, at 491 kb) (3) and that of an unsequenced *Buchnera* strain (having the smallest known bacterial genome, at about 450 kb) (4).

The genome has only 182 open reading frames (ORFs) (fig. S1A), which were classified into the clusters of orthologous groups (COGs). Notably, more than half of ORFs are devoted to only two categories, translation (34.6%) and amino acid metabolism (17.6%) (fig. S1B). In the latter, *Carsonella* retains many genes for biosynthesis

of essential amino acids, as in *Buchnera*, the symbiont of aphids. Because both psyllids and aphids feed only on plant phloem sap that is poor in essential amino acids, the analogy of gene repertoires in *Carsonella* and *Buchnera* is an intriguing

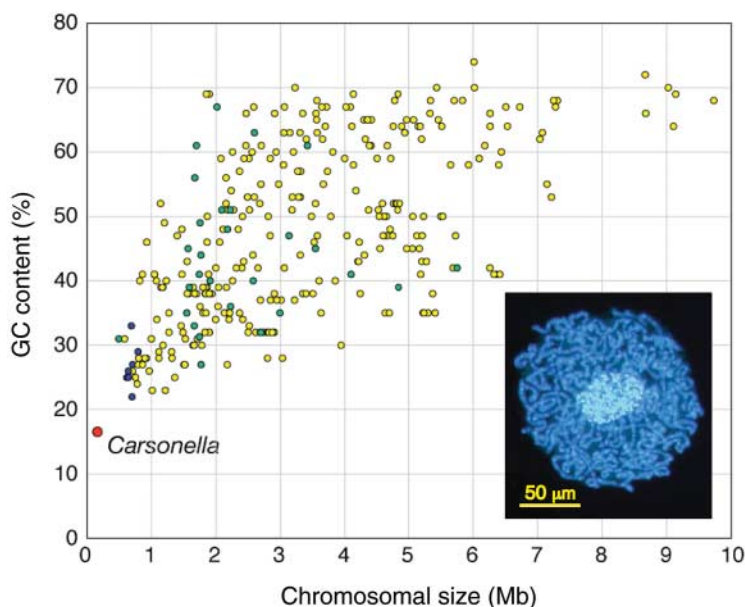


Fig. 1. Relationship between genome sizes and GC content of 358 complete genomes from Bacteria and Archaea: red indicates *Carsonella*; blue represents endosymbionts *Buchnera*, *Blochmannia*, *Wigglesworthia*, and *Baumannia*; yellow, other Bacteria; and green, Archaea. (Inset) A 4,6'-diamidino-2-phenylindole-stained bacteriocyte of *P. venusta*. Tubular cells surrounding the host nucleus (center) are *Carsonella*.

example of convergence. A remarkable feature of the genome is the total loss of genes for numerous categories, including cell envelope biogenesis and metabolisms of nucleotides and lipids (fig. S1B).

Another feature of this genome is an extremely high gene density. The protein-coding sequences and RNA genes (one 16S-23S-5S ribosomal RNA operon and 28 tRNA genes for all 20 amino acids) cover 97.3% of the genome, which is a gene density higher than those in known bacterial genomes. This density is attributable to numerous overlapping genes. Of 182 ORFs, 164 (90%) overlap with at least one of the two adjacent ORFs, and the average length of all 132 overlaps is 10.7 bases. The majority (92%) are tandem overlaps on the same strand, all of which are out of frame. Moreover, the average length of *Carsonella* ORFs (826 bp) is notably shorter than that of other bacteria. Indeed, a comparison

of 89 orthologous ORFs conserved in *Carsonella* and in seven bacteriocyte-restricted endosymbionts revealed that the average length of the ORFs in *Carsonella* is 17.8 to 18.4% shorter than the average ORF lengths of the other endosymbionts.

This genome is by far the most streamlined studied to date. Its gene inventory seems insufficient for most biological processes that appear to be essential for bacterial life, and possibly the host bacteriocyte compensates. Although some psyllids possess additional secondary endosymbionts that might be a source of specialized gene products, no other symbionts are present in *P. venusta*, based on several lines of evidence [e.g., (2)]. The genome also lacks many genes for bacterium-specific processes. One of several possible explanations for the absence of these genes is that, as in the case of organelles (5), some genes were transferred from the genome of a *Carsonella* ancestor to the genome of a psyllid ancestor and are now expressed under control of the host nucleus.

References and Notes

1. P. Baumann, *Annu. Rev. Microbiol.* **59**, 155 (2005).
2. M. L. Thao *et al.*, *Appl. Environ. Microbiol.* **66**, 2898 (2000).
3. E. Waters *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12984 (2003).
4. R. Gil, B. Sabater-Muñoz, A. Latorre, F. J. Silva, A. Moya, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4454 (2002).
5. S. D. Dyal, M. T. Brown, P. J. Johnson, *Science* **304**, 253 (2004).
6. We thank K. Oshima (Kitasato University), H. Ochman and P. Tran (University of Arizona), and H. Kuroiwa and T. Kuroiwa (Rikkyo University) for technical support. This research was supported by a Grant-in-Aid for Young Scientists (B) (A.Y.) and a Grant-in-Aid for Scientific Research on Priority Areas "Comprehensive Genomics" (M.H.) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (M.H.). GenBank, DNA Data Bank of Japan, and European Molecular Biology Laboratory accession number is AP009180 for the chromosome of *C. ruddii* strain Pv.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5797/267/DC1

Materials and Methods

Fig. S1

References

22 August 2006; accepted 18 September 2006

10.1126/science.1134196

¹Environmental Molecular Biology Laboratory, RIKEN, Wako, Saitama 351-0198, Japan. ²Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA.

³Kitasato Institute for Life Sciences, Kitasato University, Sagamihara, Kanagawa 228-8555, Japan. ⁴Center for Basic Research, Kitasato Institute, Minato-ku, Tokyo 108-8641, Japan. ⁵School of General Education, University of the Air, Mihama, Chiba 261-8586, Japan.

⁶Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba 277-8561, Japan. ⁷Genome Core Technology Facility, RIKEN Genomic Sciences Center, Yokohama, Kanagawa 230-0045, Japan.

*To whom correspondence should be addressed. E-mail: nakabachi@email.arizona.edu (A.N.); hattori@k.u-tokyo.ac.jp (M.H.)

†These authors contributed equally to this work.

ERRATUM

Post date 2 March 2007

Brevia: "The 160-kilobase genome of the bacterial endosymbiont *Carsonella*" by A. Nakabachi *et al.* (13 Oct. 2006, p. 267). The last sentence of the second paragraph is incorrect. It should read, "The genome size, which was further confirmed by long-range electrophoresis, is only about one-third that of the archaeal parasite *Nanoarchaeum equitans* (which is 491 kb) (3) and that of a *Buchnera* strain (which has the second smallest bacterial genome, at 422.4 kb) (4)." The current reference (4) should be replaced by the following reference: V. Pérez-Brocal *et al.*, *Science* **314**, 312 (2006).

ERRATUM

Post date 15 February 2008

Brevia: "The 160-kilobase genome of the bacterial endosymbiont *Carsonella*," by A. Nakabachi *et al.* (13 October 2006, p. 267). The scale in the Fig. 1 inset should have been 25 μm instead of 50 μm .