Bacterial Evolution: Chromosome Arithmetic and Geometry

Dispatch

Howard Ochman

Recent sequencing projects have characterized bacterial genomes that are organized onto elements of various sizes, shapes and numbers. Aside from its biological relevance and curiosity, this diversity calls into question the way that we define bacterial chromosomes.

In the beginning, bacteria were simple. The two workhorses of bacterial genetics — *Escherichia coli* and *Bacillus subtilis* — were both shown to have circular genetic maps and chromosomes of about 4.5 million base pairs. Because these organisms were known to be rather distantly related, it was reasonable to assume that the shape and numbers of chromosomes would be constant across bacterial species. The occurrence of plasmids in virtually all bacterial species actually helped to reinforce this view. These 'extrachromosomal' elements were typically small, dispensable, sporadically distributed and of variable copy numbers, implying that all of the genes essential for housekeeping functions were encoded on the single, circular chromosome present in all members of a species.

As genetic and physical maps began to accumulate for additional species, the exceptions began to emerge. The spirochaete *Borrelia burgdorferi* [1], the proteobacterium *Agrobacterium tumefaciens* [2], and the actinomycete *Streptomyces coelicolor* [3] were all found to have linear chromosomes, as do many of the close relatives of these species. Chromosome linearity arose independently in these groups, each of which has apparently solved the telomere-replication problem in its own way: for example, the ends of the *Borrelia* chromosome are closed hairpin structures, whereas the *Streptomyces* chromosome terminates with covalently bound proteins [4]. Furthermore, the plasmids recovered from these genomes comprise both linear and circular molecules; and because of the mobile nature of these elements, linear plasmids have occasionally been detected in bacteria that have only circular chromosomes.

Along with the finding that bacteria can have linear chromosomes were reports that they may also contain multiple chromosomes. From the application of physical mapping techniques, the genomes of several bacterial species were compartmentalized into multiple chromosomes, including *Rhodobacter sphaeroides* [5], *Rhizobium* (now *Sinorhizobium*) *meliloti* [6], *Bacillus thuringiensis* [7], *Pseudomonas* (now *Burkholderia*) *cepacia* [8], *Brucella melitensis* [9] and *Agrobacterium tumefaciens* [2], which actually has one linear and one circular chromosome. The presence of multiple chromosomes in a cell is perhaps not especially surprising, given that bacterial genomes have long been known to harbor additional replicons in the form of plasmids. Are the added chromosomes just exceedingly large plasmids, or are there objective criteria that support the distinctions between these two categories of heritable elements?

Insights into the classification of these elements can be distilled from the published genome sequences, noting that some authors tackle this nomenclature problem with explicit rules, whereas others apply a more visceral approach. To be fair, most of these elements were named well before their sequences were determined, so there is an underlying deference to history and expediency in their current designations. Nonetheless, chromosomes are always the largest repositories of genetic material in the cell and contain the bulk of genes supplying housekeeping functions, as supported by the presence of ribosomal RNA operons (Table 1). Conversely, plasmids, though attaining sizes that might garner classification as 'megaplasmids', are always smaller than the cell's chromosome and never (with one exception) harbor ribosomal genes.

Among the fully sequenced genomes included in Table 1, the partitioning of replicons into plasmids and chromosomes is straightforward for *Borrelia* [10], *Clostridium* [11], *Salmonella* [12] and *Yersinia* [13]. In these cases, all of the ribosomal operons are contained on a single large element, designated as the chromosome. By extension, this would justify the present assignment of two chromosomes in both *Agrobacterium* [14] and *Brucella* [15], and of the 1.4 Mb and 1.7 Mb elements in *Sinorhizobium* [16] as megaplasmids.

Whereas relative replicon size and rDNA number are convenient metrics for distinguishing plasmids from chromosomes, the biological significance of these features still needs to be assessed. Implicit to this scheme is that chromosomes represent the ancestral genetic material, encoding central housekeeping functions that are common to all cells; and functional analyses confirm that the overwhelming majority of universally conserved genes are indeed limited to the chromosomes. By focusing on such properties, some of the more troublesome issues have been averted. This approach allows plasmids to be essential and distributed among all members of a species but implies that they are involved in traits appropriated over the evolutionary history of a lineage.

Given the dynamics of bacterial genomes, we might expect there to be some equivocal cases — and indeed there are. *Vibrio* [17] and *Deinococcus* [18] each have small 'chromosomes' that lack rRNA operons, and *Ralstonia* [19] harbors a 2 Mb 'megaplasmid' containing a ribosomal gene cluster complete with tRNAs. That the authors are impelled to justify their designations — typically based on the putative functions of replicon-specific genes — suggests that the distinction between

Department of Biochemistry & Molecular Biophysics, University of Arizona, Tucson, Arizona 85721, USA.
E-mail: hochman@email.arizona.edu

Table 1. Replicons In Some Sequenced Bacterial Genomes With Multiple Heritable Elements.

Species	Appellation	Size (kb)	Shape	rDNA no.
Agrobacterium tumefaciens	Chromosome	2842	Circular	2
	Chromosome	2057	Linear	2
	Plasmid	543	Circular	0
	Plasmid	214	Circular	0
Borrelia burgdorferi	Chromosome	911	Linear	1
	Plasmids (n = 11)	9–54	Circular/Linear	0
Brucella melitensis	Chromosome	2117	Circular	2
	Chromosome	1178	Circular	1
Clostridium acetobutylicum	Chromosome	3941	Circular	11
	Megaplasmid	192	Circular	0
Deinococcus radiodurans	Chromosome	2649	Circular	3
	Chromosome	412	Circular	0
	Megaplasmid	177	Circular	0
	Plasmid	46	Circular	0
Ralstonia solanacearum	Chromosome	3716	Circular	3
	Megaplasmid	2095	Circular	1
Salmonella typhi	Chromosome	4809	Circular	7
	Plasmid	218	Circular	0
	Plasmid	107	Circular	0
Sinorhizobium meliloti	Chromosome	3654	Circular	3
	Megaplasmid	1683	Circular	0
	Megaplasmid	1354	Circular	0
Vibrio cholerae	Chromosome	2941	Circular	8
	Chromosome	1072	Circular	0
Yersina pestis	Chromosome	4654	Circular	6
	Plasmids (n = 3)	10–96	Circular	0

mega-plasmid and mini-chromosome has blurred. There is, however, a convenient way out of this bind: when a questionable genomic element harbors the only copy of a gene conserved among all bacteria (one of the hundred or so genes constituting the minimal genome complement [20]), one can assume to be looking at a chromosome. Of course, it is possible to conceive of an exception and of insufficiencies in such criteria, but it will probably require another 50 sequenced genomes to uncover one.

References

1. Ferdows, M.S. and Barbour, A.G. (1989). Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi,* the Lyme disease agent. Proc. Natl. Acad. Sci. U.S.A. *86*, 5969–5973.
2. Allardet, A., Michaux-Charachon, S., Jumas-Bilak, E., Karayan, L. and Ramuz, M. (1993). Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. J. Bacteriol. *175*, 7869–7874.
3. Keiser, H.M. Kieser, T. and Hopwood, D.A. (1992). A combined genetic and physical map of the *Streptomyces coelicolor* A3(2) chromosome. J. Bacteriol. *174*, 5496–5507.
4. Casjens, S. (1998). The diverse and dynamic structure of bacterial chromosomes. Annu. Rev. Genet. *32*, 339–377.
5. Suwanto, A. and Kaplan, S. (1989). Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: presence of two unique circular chromosomes. J. Bacteriol. *171*, 5850–5859.
6. Honeycutt, R. and McClelland, M. (1995). Physical map of the genome of *Rhizobium meliloti* 1021. J. Bacteriol. *175*, 6945–6952.
7. Carlson, C. R., Johansen, T. LeCadet, M and Kolstø, A.-B. (1996). Genomic organization of the entomopathogenic bacterium *Bacillus thuringiensis* subsp. *berliner* 1715. Microbiology *142*, 1625–1634.
8. Cheng, H.P. and Lessie, T.G. (1994). Multiple replicons constituting the genome of *Pseudomonas cepacia* 17616. J. Bacteriol. *176*, 4034–4042.
9. Michaux-Charachon, S., Paillisson, J. Carles-Nurit, M.J., Bourg, G., Allardet-Servent, A. and Ramuz, M. (1993). Presence of two independent chromosomes in the *Brucella melitensis* 16M genome. J. Bacteriol. *175*, 701–705.
10. Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K. *et al*. (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature *390*, 580–587.
11. Nolling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., Zeng, Q., Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J. *et al*. (2001). Genome sequencing and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. J. Bacteriol. *183*, 4823–4838.
12. Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T. *et al*. (2001). Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature *413*, 848–852.
13. 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). Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature *413*, 523–527.
14. Wood, D.W., Setubal, J.C., Kaul, R., Monks, D.E., Kitajima, J.P., Okura, V.K., Zhou, Y., Chen, L., Wood, G.E., Almeida, N.F. Jr. *et al*. (2001). The genome sequence of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science *294*, 2317–2323.
15. DelVecchio, V.G., Kapatral, V., Redkar, R.J., Patra, G., Mujer, C., Los, T., Ivanova, N. anderson, I., Bhattacharyya, A., Lykidis, A. *et al*. (2002). The genome sequence of the facultative intracellular pathogen *Brucella melitenisis*. Proc. Natl. Acad. Sci. U.S.A. *99*, 443–448.
16. Galibert, F., Finan, T.M., Long, S.R., Puhler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M.J., Becker, A., Boistard, P. *et al*. (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science *293*, 668–672.
17. Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L. *et al*. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature *406*, 477–483.
18. White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L. *et al*. (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science *286*, 1571–1577.
19. Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L. *et al*. (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature *415*, 497–502.
20. Koonin, E.V. (2000). How many genes can make a cell: the minimal-gene-set concept. Annu. Rev. Genomics Hum. Genet. *1*, 99–116.