microbiology

Genome Update: rRNAs in sequenced microbial genomes

Genomes of the month – three superkingdoms of life

Once again, this month, three microbial genomes have been published in the 4 weeks since the last Genome Update was written. However, in contrast to last month (Ussery & Hallin, 2004), when all three genomes were bacterial, this time each genome represents a different superkingdom of life. The three genomes include that of a methanogen isolated from a salt-marsh sediment (the archaean *Methanococcus maripaludis*), that of a soil-dwelling bacillus found to cause cheese spoilage (the bacterium *Bacillus cereus*) and a yeast genome (the eukaryote *Ashbya gossypii*).

Methanococcus maripaludis is a mesophilic, methane-producing, nitrogen-fixing member of the *Archaea*; it is a strict anaerobe and grows in the presence of hydrogen and carbon dioxide. The *Methanococcus maripaludis* S2 genome (Hendrickson *et al.*, GenBank accession no. BX950229) is 1 661 137 bp long (see Table 1), just a few thousand base pairs shorter than that of the thermophile *Methanocaldococcus jannaschii* DSM 2661^T, and in general it is one of the smaller archaeal genomes. The *Methanococcus maripaludis* genome contains few DNA repeats and has three rRNA operons.

The genome of *B. cereus* ATCC 10987 is 5·2 Mbp long and has about 5600 annotated genes (Rasko *et al.*, 2004); in many ways, the genome is more similar to the sequenced genome of *Bacillus anthracis* than to that of the other sequenced *B. cereus* strain. *B. cereus, Bacillus thuringiensis* and *B. anthracis* are all members of the *B. cereus sensu lato* group, and comparison of 16S rRNA gene sequences alone cannot distinguish between members of this group. The genome of *B. cereus* ATCC 10987 encodes 12 rRNA operons and 98 tRNAs, as shown in Table 1. The filamentous fungus A. gossypii is used by industry in vitamin production. The genome of A. gossypii ATCC 10895 is the smallest eukaryotic genome sequenced to date (Dietrich et al., 2004), encoding 4718 genes, or close to 1000 genes fewer than the B. cereus genome (see Table 1). Like many eukaryotic genomes, the presence of repeats makes the task of sequencing the whole A. gossypii genome quite difficult, and it is estimated that the genome is 9.2 Mbp long, although only 8.7 Mbp are present in the EMBL files. Furthermore, although it is known that there are about 40 copies of an rRNA repeat in the genome, only a single copy is found in the EMBL file. We have listed two rows in Table 1 for this genome – the first row lists what is reported in the sum of the EMBL files for all seven chromosomes and the second row lists the numbers given in the genome report (Dietrich *et al.*, 2004).

Method of the month – comparative sequence analysis of 16S rRNA genes in sequenced genomes

An unrooted phylogenetic tree of 165 sequenced prokaryotic genomes is shown in Fig. 1. The comparison of 16S rRNA gene sequences to infer the phylogenetic relationships among prokaryotes has been widely used for several decades, and the 16S rRNA molecule is generally accepted as the ultimate molecular chronometer,

Table 1. Summary of the published genomes discussed in this Update

Note that the accession number for each chromosome is the same for GenBank, EMBL and the DNA DataBase of Japan (DDBJ). Note that the 'Genome estimates' for the *A. gossypii* row are numbers from the published genome report, whilst the numbers given above are from the sum of the EMBL files for all seven chromosomes.

Genome	Size (bp)	AT content (%)	rRNA operons	tRNAs	CDS	Accession nos
Methanococcus maripaludis S2	1 661 137	66•9	3	37	1722	BX950229
Bacillus cereus ATCC 10987	5 224 283	64•4	12	98	5642	AE017194
Ashbya gossypii ATCC 10895	8 741 255	48.3	1	199	4718	AE016814– AE016821
(Genome estimates)	$(\sim 9.2 \text{ Mbp})$	(48)	(40)	(199)	(4718)	

Microbiology **Comment** provides a platform for readers of *Microbiology* to communicate their personal observations and opinions in a more informal way than through the submission of papers.

Most of us feel, from time to time, that other authors have not acknowledged the work of our own or other groups or have omitted to interpret important aspects of their own data. Perhaps we have observations that, although not sufficient to merit a full paper, add a further dimension to one published by others, or we may have a useful piece of methodology that we would like to share.

Guidelines on how to submit a *Microbiology* Comment article can be found in the Instructions for Authors at http://mic.sgmjournals.org

It should be noted that the Editors of *Microbiology* do not necessarily agree with the views expressed in *Microbiology* Comment.

Chris Thomas, Editor-in-Chief



Fig. 1. Phylogenetic tree based on 16S rRNA gene sequences from 150 bacterial and 15 archaeal sequenced genomes. The tree was constructed using the neighbour-joining method. All major bacterial and archaeal groups have been indicated, as well as the phylogenetic positions of *B. cereus* and *Methanococcus maripaludis*. Bar, 10% sequence dissimilarity. The width of the triangle represents the relative number of rRNA sequences (i.e. the number of bacterial genomes sequenced) for a given phylum.

because it is functionally constant, shows a mosaic structure of conserved and more variable regions and because of its universal presence (Woese, 1987). As can be seen from Fig. 1, complete genomes are available for most major bacterial groups (including the five subdivisions of the *Proteobacteria*, the *Firmicutes*, the *Actinobacteria* and the *Cyanobacteria*) and for representatives of the two major archaeal groups, the phyla *Euryarchaeota* and *Crenarchaeota*. The large distance between the *Euryarchaeota* and *Crenarchaeota* can be readily seen in this figure.

The number of rRNA operons per sequenced bacterial genome varies from one to 13, and there is some occasional heterogeneity within the rRNA gene sequences of an organism (Coenye & Vandamme, 2003). The genome of B. cereus ATCC 10987 (Rasko et al., 2004) has 12 rRNA operons, which is a fairly large number when compared to the other sequenced bacterial genomes. The genome of B. cereus ATCC 14579^T has 13 rRNA operons, which is the largest number of rRNA operons observed among the sequenced bacterial genomes: the Methanococcus maripaludis genome contains only three rRNA operons. See supplemental web pages for a complete list of all 154 sequenced bacterial genomes, sorted by number of rRNA operons. So far, the prokaryote in which most rRNA operons have been detected is Clostridium paradoxum, which has 15 rRNA operons (Klappenbach et al., 2001).

Next month, the 'method' of genome comparison discussed will be a look at tRNA genes in prokaryotic genomes. For the sequenced bacterial genomes, the number of tRNA genes varies from 113 in the genome of *Vibrio parahaemolyticus* to 29 in that of *Mycoplasma pulmonis*.

Supplemental web pages

Web pages containing supplemental material related to this article can be accessed from the following url: http://www.cbs.dtu.dk/services/ GenomeAtlas/suppl/GenUp004/

Acknowledgements

This work was supported by a grant from the Danish Center for Scientific Computing (DCSC).

David W. Ussery,¹ Peter F. Hallin,¹ Karin Lagesen² and Tom Coenye³

¹Center for Biological Sequence Analysis, Department of Biotechnology, Building 208, The Technical University of Denmark, Lyngby, DK-2800, Denmark

²Department of Molecular Biology, Institute of Medical Microbiology, University of Oslo, The National Hospital, 0027 Oslo, Norway

³Laboratorium voor Microbiologie, Universiteit Gent, Ledeganckstraat 35, B-9000 Gent, Belgium

Coenye, T. & Vandamme, P. (2003).

Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol Lett* **228**, 45–49.

Dietrich, F. S., Voegeli, S., Brachat, S. & 11 other authors (2004). The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* Epub ahead of print, DOI: 10.1126/science.1095781

Klappenbach, J. A., Saxman, P. R., Cole, J. R. & Schmidt, T. M. (2001). rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res* **29**, 181–184.

Rasko, D. A., Ravel, J., Okstad, O. A.

& 12 other authors (2004). The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1. *Nucleic Acids Res* **32**, 977–988.

Ussery, D. W. & Hallin, P. F. (2004). Genome Update: AT content in sequenced prokaryotic genomes. *Microbiology* **150**, 749–752.

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**, 221–271.

DOI 10.1099/mic.0.27173-0