microbiology Comment

Genome Update: 161 prokaryotic genomes sequenced, and counting

Bacterial genomes are now being published almost every week. The purpose of this new column is to provide monthly updates on new genomes which have been published, as well as to discuss methods for genome comparison of the more than 100 bacterial genomes which are already publicly available. Thus, each column will be divided into two parts, with the first part devoted to a brief description of highlights of the recently sequenced genomes and the second part devoted to a discussion of relevant methods currently being used to compare sequenced genomes to each other. Each month a table will be presented, as shown in Table 1, with the name, length, AT content, number of rRNAs, tRNAs, CDS regions annotated and accession numbers for each new genome. The genome sequences of four species have been published in the first half of December 2003: *Vibrio vulnificus*, a marine pathogen, the cyanobacterium *Gloeobacter violaceus*, the soil bacterium *Geobacter sulfurreducens* and the plant pathogen *Phytoplasma asteris*.

Genomes of the month

The *V. vulnificus* YJ016 genome contains two circular chromosomes and a plasmid, with a total size of more than 5 Mbp (Chen *et al.*, 2003). *V. vulnificus* is a member of the γ-subdivision of the *Proteobacteria* and is responsible for seafood-borne infections, with a mortality rate of more than 50 % within a few days. The authors do a detailed comparison of the *V. vulnificus* chromosomes versus *Vibrio cholerae* chromosomes and discuss at length genetic variability. About a year ago, in December 2002, the complete genome sequence for a DIFFERENT strain (CMCP6) of *V. vulnificus* was deposited in GenBank and, although a genome report for this strain is still not available, the GenBank files are public and accession numbers are mentioned in an article discussing *V. vulnificus* antigens (Kim *et al.*, 2003). A very brief overview of the two strains is shown in Table 1. One can see that strain CMCP6 has slightly smaller chromosomes and fewer proteins than strain YJ016. A more detailed comparison of the two strains might reveal many interesting properties of *V. vulnificus* genomes, compared to other *Vibrio* species.

In contrast to the marine environment of *V. vulnificus*, the cyanobacterium *Gloeobacter violaceus* strain PCC 7421 was isolated from calcerous rock in Switzerland. The genome is 4·7 Mbp long, with an AT content of only 38 %, and encodes 4430 proteins (Nakamura *et al.*, 2003a, b). The *Gloeobacter* lineage is quite distinct from other cyanobacteria and has a relatively long generation time (about 72 h). *Gloeobacter* cells lack thylakoid membranes and the machinery for photosynthesis is located in the cytoplasm. Six hundred and ten genes had matches only to other cyanobacterial genomes, with about half of these having no known function. About 20 % (684 genes) were found to be unique to this organism. As in other cyanobacteria, genes with inteins were found in the *Gloeobacter violaceus* genome. [Inteins are self-splicing proteins and, in the case of *trans*-splicing, two genes can code for a single protein, as for the *dnaE* gene in cyanobacteria (Caspi *et al.*, 2003).] There were also 74 genes for putative transposases and traces of large-scale genomic rearrangements could be detected. More information can be found at the CyanoBase web site (http://www.kazusa.or.jp/cyanobase/).

The *Geobacter sulfurreducens* strain PCA genome contains a single circular chromosome of 3 814 139 bp (Methé *et al.*, 2003). *Geobacter sulfurreducens* is a member of the δ-*Proteobacteria* and, as such, its genome sequence is the first of this subdivision to be published. [*Desulfovibrio vulgaris* subsp. *vulgaris* str. Hildenborough is also a member of the δ-*Proteobacteria* and its genome was sequenced by TIGR more than 2 years ago (October 2001), but the genome report has yet to be published at the time of writing this article (December 2003). However, BLAST searches against this genome are possible (GenBank accession no. NC_002937)]. *Geobacter sulfurreducens* can be used in bioremediation of groundwater. This is due to the ability of this species of bacteria to precipitate soluble metals, such as uranium. Based on the genome sequence, it looks as though *Geobacter sulfurreducens* might not be immobile, nor a strict anaerobe, as had been thought previously. The genome also contains an '*unprecedented collection of newly reported c-type cytochromes*' (Methé *et al.*, 2003).

Finally, *P. asteris* is an obligate intracellular plant pathogen, transmitted by insects, and its genome is the first of a bacterium inhabiting both plants and insects to be sequenced. The genome of the *P. asteris*

***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

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). chr., Chromosome.

Genome	Size (bp)	AT content (%)	rRNA operons	tRNAs	CDS	Accession no.
Geobacter sulfurreducens PCA	3 814 139	39·1	2	49	3466	AE017180
Gloeobacter violaceus PCC 7421	4 659 019	38·0	1	45	4430	BA000045
Phytoplasma asteris OY	860 631	72·0	2	32	754	AP006628
Vibrio vulnificus YJ016	5 260 086	53·3	9	111	5086	–
Vibrio vulnificus CMCP6	5 126 798	53·3	9	112	4586	–
Vibrio vulnificus YJ016 chr. 1	3 354 505	53·6	8	100	3297	BA000037
Vibrio vulnificus CMCP6 chr. 1	3 281 945	53·6	8	98	3002	AE016795
Vibrio vulnificus YJ016 chr. 2	1 857 073	52·8	1	12	1718	BA000038
Vibrio vulnificus CMCP6 chr. 2	1 844 853	52·8	1	13	1584	AE016796
Vibrio vulnificus YJ016 plasmid	48 508	55·1	0	0	71	AP005352

strain OY contains a circular chromosome of 860 631 bp (Oshima *et al.*, 2003). The genome has undergone reductive evolution and there are many genes thought to be essential for life, such as the ATP-synthase subunits, which are missing in this organism. Although almost twice as large as the smallest genome sequenced so far, the contents of the *P. asteris* genome are important to take into consideration when contemplating the '*minimal genome of life*'.

Method of the month

Perhaps one of the simplest ways of comparing many genomes is simply to list them in a table, with rows of various numbers describing each genome (such as length, AT content, number of genes, etc., as in Table 1). Of course, it is obvious that a genome cannot be fully reduced to a single number. But some characteristics, such as the length of the chromosome, AT content, fraction of repeats and number of genes, are at least a good place to start in comparison of a large number of genomes. As an example, of the genomes presented in Table 1, looking at the number of rRNA operons, the *Gloeobacter violaceus* genome has only a single rRNA operon, compared to nine rRNA operons in the *V. vulnificus* genome. There is certainly a difference in lifestyle of the two organisms – whilst a single *Gloeobacter violaceus* cell is lazily replicating under the alpine sun, over a period of about 3 days (72 h), *V. vulnificus* can have an *in vivo* doubling time of about 45 min (Starks *et al.*, 2000). Thus, a single *V. vulnificus* cell can undergo about 100 cell divisions, yielding about 6×10^{29} cells, during the same time it would take one *Gloeobacter violaceus* cell to divide into two cells. Obviously, the *Vibrio* cells will need to make much more rRNA for such rapid growth. Thus, the number of rRNA operons in a genome can be reflective of information about the relative doubling time of the organism.

At the time of writing this article, there are 144 sequenced bacterial genomes and 17 archaeal genomes that are available to the public. These can be viewed on our web page and sorted by various data, including the date of publication (http://www.cbs.dtu.dk/services/GenomeAtlas). About 50 of the 144 public bacterial genomes have been made available during 2003 and it is very likely that more than 100 bacterial genomes will become available in 2004. There are about twice as many bacterial genome sequences (250) that can be searched against using BLAST (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?) than can be downloaded from GenBank or EMBL (128 genomes; http://www.ebi.ac.uk/genomes/bacteria.html). Currently, it takes a year (or longer) between the time a genome is sequenced and the time that it is fully annotated and a detailed report is written about the project, so there is a lag between

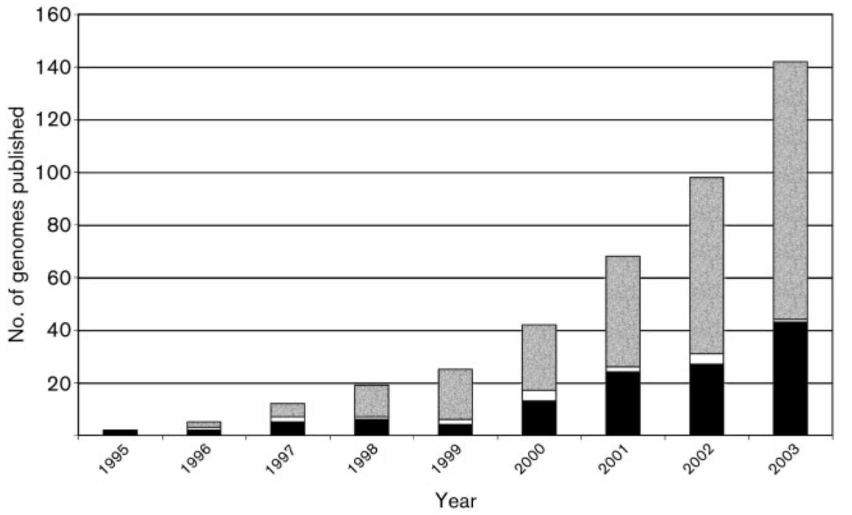

Fig. 1. Number of prokaryotic genomes sequenced each year since 1995. Black, bacterial genomes; white, archaeal genomes; grey, running total.

the time a genome is sequenced and when it is published. The growth in the number of prokaryotic genome sequences published in recent years is shown in Fig. 1. It is worth noting that, although archaeal species probably exist in equal abundance with bacterial species, to date comparatively few archaeal genomes have been sequenced. The publication of so many genome sequences has led to an explosion of genomic information, which can be difficult to keep up with. In fact, as can be seen from the *V. vulnificus* genomes of the two different strains mentioned above, sometimes even people within the field can miss projects where other groups have sequenced the same organism and deposited it in GenBank.

Next month's article will focus on comparison of lengths of genomes, which differ by nearly 20-fold among the prokaryotes, from 490 885 bp for the *Nanoarchaeum equitans* genome to 9 105 828 bp for the *Bradyrhizobium japonicum* genome.

Acknowledgements

This work was supported by a grant from the Danish National Research Foundation.

David W. Ussery

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

Correspondence: David W. Ussery (dave@cbs.dtu.dk)

Caspi, J., Amitai, G., Belenkiy, O. & Pietrokovski, S. (2003). Distribution of split DnaE inteins in cyanobacteria. *Mol Microbiol* **50**, 1569–1577.

Chen, C. Y., Wu, K. M., Chang, Y. C. & 12 other authors (2003). Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. *Genome Res* **13**, 2577–2587.

Kim, Y. R., Lee, S. E., Kim, C. M. & 9 other authors (2003). Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. *Infect Immun* **71**, 5461–5471.

Methé, B. A., Nelson, K. E., Eisen, J. A. & 31 other authors (2003). Genome of *Geobacter sulfurreducens*: metal reduction in subsurface environments. *Science* **302**, 1967–1969.

Nakamura, Y., Kaneko, T., Sato, S. & 16 other authors (2003a). Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. *DNA Res* **10**, 137–145.

Nakamura, Y., Kaneko, T., Sato, S. & 16 other authors (2003b). Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids (supplement). *DNA Res* **10**, 181–201.

Oshima, K., Kakizawa, S., Nishigawa, H. & 8 other authors (2003). Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat Genet* Epub ahead of print, http://www.nature.com/cgi-taf/DynaPage.taf?file=ng/journal/vaop/ncurrent/abs/ng1277.html

Starks, A. M., Schoeb, T. R., Tamplin, M. L., Parveen, S., Doyle, T. J., Bomeisl, P. E., Escudero, G. M. & Gulig, P. A. (2000). Pathogenesis of infection by clinical and environmental strains of *Vibrio vulnificus* in iron-dextran-treated mice. *Infect Immun* **68**, 5785–5793.

DOI 10.1099/mic.0.26984-0