

## Genome Update: Chromosome Atlases

### Genomes of the month

Six new bacterial genomes have been published since last month's Genome Update. The list is shown in Table 1 and includes a species that can grow happily in a refrigerator (the  $\delta$ -proteobacterium *Desulfotalea psychrophila*, with an optimal growth temperature of 10 °C, although it can grow at temperatures slightly below freezing!) and the actinobacterium *Propionibacterium acnes*, which can cause acne in humans. Let us hope that these two never cross, because the thought of a fridge with complexion problems is not very pleasant! The other genomes include that of the actinobacterium *Leifsonia xyli*, those of the firmicutes *Mycoplasma mobile* and *Streptococcus pyogenes* and that of an  $\alpha$ -proteobacterium, *Rickettsia typhi*. A brief overview of each of these genomes is given below.

*D. psychrophila* strain LSv54<sup>T</sup> was isolated from permanently cold coastal sediments off the coast of Svalbard, Norway.

*D. psychrophila* is the type species of a genus of sulfate-reducing bacteria, *Desulfotalea*. Cells of strain LSv54<sup>T</sup> are thin, long and rod-shaped with an optimal doubling time of 27 h at 10 °C. This is the first bacterium to be sequenced that grows happily at such low temperatures. The genome consists of a main chromosome of 3.5 Mbp, a large plasmid of 121 kbp and a smaller plasmid of about 15 kbp (Rabus *et al.*, 2004). About 27 % of the proteins have no homologues in public databases. There are some differences in amino acid composition, such as a preference for cysteines, but any conclusions based on essentially one genome are difficult to draw (one 'datum point' compared to more than 100 other genomes); at any rate, it is not easy to tell whether this bias might also be explained by the difference in proteins used for sulfur metabolism. The genomes of more psychrophilic organisms need to be sequenced, as cold environments

represent a large part of the Earth's biosphere, especially when one takes into account the oceans.

The plant pathogen *L. xyli*, although not a psychrophile, has an optimal growth temperature of only about 20 °C, and has the largest number of pseudogenes for any bacterial plant pathogen sequenced to date (307 of 2326 predicted ORFs) (Monteiro-Vitorello *et al.*, 2004). *L. xyli* also has the lowest coding density (70.09 %) of the 196 prokaryotic genomes currently available to the public. Most bacterial genomes (167 of 177) have coding densities of at least 80 % (see supplemental materials for a table of genomes sorted by coding density). Surprisingly few genes associated with pathogenicity were found in this genome (85 of the 2030 predicted genes). It seems that *L. xyli* can synthesize the plant hormone abscisic acid, which is a growth inhibitor; this could explain how this bacterium can cause stunted growth in plants.

The genome of *M. mobile* strain 163K<sup>T</sup> is about 780 kbp long and encodes 633 genes (Jaffe *et al.*, 2004). It is very AT-rich (75 %), making it the third most AT-rich bacterial genome sequenced to date. *M. mobile* has only 28 tRNAs, the fewest number for any bacterium sequenced. More importantly, there is another 'first' for this genome, in that the authors also describe a detailed analysis of the

proteome. In fact, proteomics data were used to help in annotation, and 26 genes that were not annotated originally were added. An amazing 88 % of the predicted genes have been detected experimentally as proteins (Jaffe *et al.*, 2004). The *M. mobile* genome lacks all the genes for the citric acid cycle and for *de novo* nucleotide and amino acid synthesis. One interesting aspect of this genome is a set of five genes called 'lrp' (for long terminal repeat), which were found to be nearly perfect repeats, although the proteins have slightly different amino acid sequences and are expressed differently.

The genome of *P. acnes* strain KPA171202 is 2.6 Mbp long and is predicted to contain 2333 genes (Brüggemann *et al.*, 2004). Increased cellular and humoral immunity to *P. acnes* has been observed in patients with severe acne, and the genome of this bacterium contains many potential cell-surface proteins with antigenic potential, including 25 genes with a C-terminal LPXTG type cell-wall sorting signal required for attachment of surface proteins. Some of these genes contain hypervariable stretches of Gs or Cs that can cause phase variation in the protein sequence.

*R. typhi* is an obligate intracellular parasite that causes murine typhus. The genome of *R. typhi* strain Wilmington<sup>T</sup> is 1.1 Mbp long, with an AT content of 71 % (McLeod

**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 DDBJ.

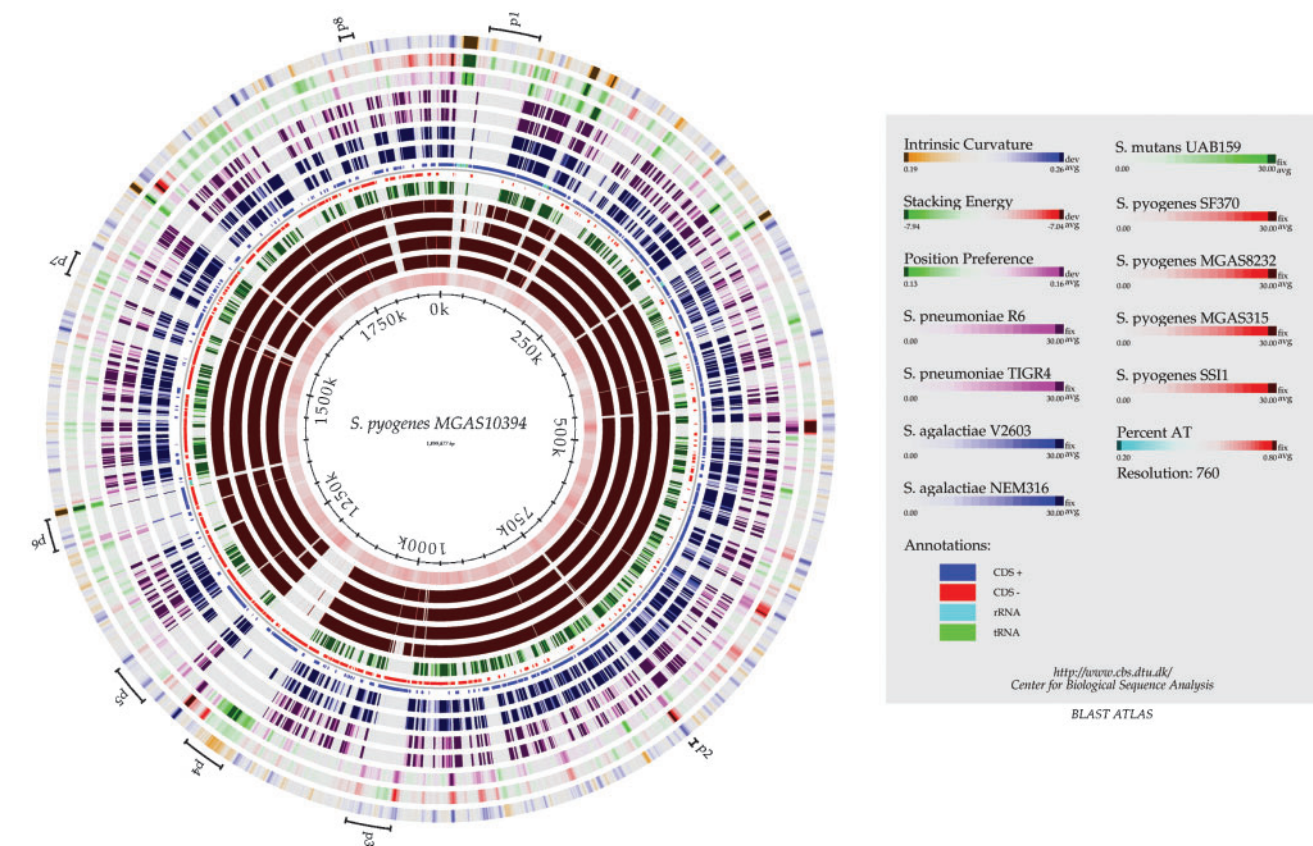
Name	Length	AT content (%)	No. of genes	tRNAs	rRNAs	Accession no.
<i>Desulfotalea psychrophila</i> LSV54 <sup>T</sup>	3 523 383	53.2	3118	64	7	CR522870
<i>Leifsonia xyli</i> CTCB07	2 584 158	32.3	2030	45	1	AE016822
<i>Mycoplasma mobile</i> 163K <sup>T</sup>	777 079	75.1	633	28	1	AE017308
<i>Propionibacterium acnes</i> KPA171202	2 560 265	40.0	2297	45	3	AE017283
<i>Rickettsia typhi</i> Wilmington <sup>T</sup>	1 111 496	71.1	838	33	1	AE017197
<i>Streptococcus pyogenes</i> MGAS10394	1 899 877	61.3	1886	67	6	CP000003

*et al.*, 2004). A three-way comparison of *R. typhi* with *Rickettsia prowazekii* and *Rickettsia conorii* revealed only 24 genes unique to *R. typhi*.

Finally, *S. pyogenes* strain MGAS10394 is the first *Streptococcus* M6 genome to be sequenced. The genome is 1 899 877 bp long, making it the largest Group A

*Streptococcus* (GAS) genome sequenced to date (Banks *et al.*, 2004). It includes eight prophage-like regions and an 8.3 kb prophage relic that encodes the SpeA4 variant of *S. pyogenes* exotoxin A. One of the prophage-like elements contains a transposon encoding the *mefA* gene and a surface-exposed protein (the ‘R6 protein’, which corresponds to M6\_Spy1173)

thought to be responsible for the M6 serotype. Interestingly, virtually all the serotype M6 strains examined (104 strains) had this gene, while none of the strains examined with the 11 other M protein serotypes (112 strains) had the R6 gene (Banks *et al.*, 2004). Also, a SpeA4 variant was found, and 19 distinct combinations of prophage element genes encoding



**Fig. 1.** DNA Atlas plot for the *S. pyogenes* MGAS10394 chromosome. Each circle represents a different property mapped along the chromosome; for example, the outermost circle plots intrinsic curvature along the sequence, with regions more curved than average in blue, and regions less curved in orange. For the BLAST comparisons, a strong match is darkly coloured; regions with weak or no matches are not coloured.

proven or putative virulence proteins were present. This result is in agreement with previous work from M1 and M18 strains (Banks *et al.*, 2004).

In addition to the above-mentioned publication concerning the *Streptococcus* M6 genome sequence, another recent article describes an extensive comparison of 255 different *Streptococcus* serotype M3 strains, cultured from different patients over a period of 11 years (Beres *et al.*, 2004). A variety of different molecular biology techniques were utilized to explore genetic diversity in these strains (including PFGE, DNA–DNA microarray and prophage genotyping). Variation in gene content (including virulence genes) in different strains can result from acquisition or loss of prophages, and the conclusion of this study is that the assigned M3 genotypes show that phage-induced population changes can be responsible for different severities of infections (Beres *et al.*, 2004).

### Method of the month – Chromosome Atlases

In previous Genome Updates, we have brought forward methods for complete chromosome analysis such as predictions of tRNAs and rRNAs, analysis of AT content, analysis of homology by using the Artemis Comparison Tool (ACT), and promoter profiles. Each method is important in itself, but this month we will use Chromosome Atlases to join different kinds of information (properties) in a single view of the chromosome. This enables a fast visual overview and reveals possible correlations between properties. The genus *Streptococcus* is well represented (11 strains sequenced from four different species) in the Genome Atlas Database. We have performed a nucleotide BLAST of all ORFs in strain MGAS10394 against nine other *Streptococcus* chromosomes and mapped all  $-\log E$ -values along the chromosome, in a colour scale ranging from nearly 0 (no homology) to 30 ( $E = 1 \times 10^{-30}$ ). (The *S. pyogenes* strain M5 Manfredo was left out since it lacks annotations.) Along with these nine data lanes, AT content, Intrinsic Curvature, Stacking Energy and Position Preference are displayed, as described previously (Pedersen *et al.*, 2000). The Atlas and the lane descriptions can be seen in Fig. 1.

The region marked P4 is low in intrinsic curvature, position preference and stacking energy, while there is no visible change in AT content compared to the mean value for the chromosome. This could suggest an area of foreign DNA, phage insertion or DNA uptake. The lanes of BLAST results clearly show that P4 features are not present in any of the genomes except for that of *S. pyogenes* strain MGAS10394. Phage P1 is only partly present in *S. pyogenes* strain SSI1 and *S. pyogenes* strain MGAS315, while absent in all other chromosomes except that of *S. pyogenes* strain MGAS10394. Phage P2, on the other hand, is present in *S. pyogenes* strains SSI1, MGAS315 and MGAS8232, while it is partly present in strain SF370. Phage P2 is absent in all other species of the genus *Streptococcus*. The Atlas approach is useful for combining different kinds of information to build a genome comparison. The program used to construct these atlases (GeneWiz) has been implemented in a web-based service, where users can make their own Custom Atlases, at the CBS Genome Atlas Database site <http://www.cbs.dtu.dk/services/GenomeAtlas/myAtlas>

### 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/GenUp009/>

### Acknowledgements

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

**Peter F. Hallin, Tim T. Binnewies and David W. Ussery**

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

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

**Banks, D. J., Porcella, S. F., Barbican, K. D. & 7 other authors (2004).** Progress toward characterization of the group A *Streptococcus* metagenome: complete genome sequence of a macrolide-resistant serotype M6 strain. *J Infect Dis* **190**, 727–738.

**Beres, S. B., Sylva, G. L., Sturdevant, D. E. & 11 other authors (2004).** Genome-wide molecular dissection of serotype M3 group A *Streptococcus* strains causing two epidemics of invasive infections. *Proc Natl Acad Sci U S A* **101**, 11833–11838.

**Brüggemann, H., Henne, A., Hoster, F., Liesegang, H., Wiezer, A., Strittmatter, A., Hujer, S., Durre, P. & Gottschalk, G. (2004).** The complete genome sequence of *Propionibacterium acnes*, a commensal of human skin. *Science* **305**, 671–673.

**Jaffe, J. D., Stange-Thomann, N., Smith, C. & 16 other authors (2004).** The complete genome and proteome of *Mycoplasma mobile*. *Genome Res* **14**, 1447–1461.

**McLeod, M. P., Qin, X., Karpathy, S. E. & 19 other authors (2004).** Complete genome sequence of *Rickettsia typhi* and comparison with sequences of other *Rickettsiae*. *J Bacteriol* **186**, 5842–5855.

**Monteiro-Vitorello, C. B., Camargo, L. E., Van Sluys, M. A. & 41 other authors (2004).** The genome sequence of the Gram-positive sugarcane pathogen *Leifsonia xyli* subsp. *xyli*. *Mol Plant Microbe Interact* **17**, 827–836.

**Pedersen, A. G., Jensen, L. J., Brunak, S., Staerfeldt, H. H. & Ussery, D. W. (2000).** A DNA structural atlas for *Escherichia coli*. *J Mol Biol* **16**, 907–930.

**Rabus, R., Ruepp, A., Frickey, T. & 15 other authors (2004).** The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ Microbiol* **6**, 887–902.

DOI 10.1099/mic.0.27582-0