microbiology Comment

Genome update: sigma factors in 240 bacterial genomes

Genomes of the month

Ten new microbial genomes were published since the last 'Genome Update' column was written. The collection of this month's prokaryotic genomes, listed in Table 1, consists of one archaeon (*Sulfolobus acidocaldarius*) and five bacteria (*Corynebacterium jeikeium*, *Haemophilus influenzae*, *Pseudomonas fluorescens*, *Rickettsia felis* and *Xanthomonas campestris*). In addition, four microbial eukaryotic genomes have been published: *Dictyostelium discoideum* AX4 (Eichinger *et al.*, 2005), *Theileria annulata* (Gardner *et al.*, 2005), *Theileria parvarich* (Pain *et al.*, 2005) and *Toxoplasma gondii* (Khan *et al.*, 2005).

Corynebacterium jeikeium is an opportunistic pathogen and causes systemic infections, particularly in immunocompromised patients/hosts. Broad-spectrum resistance to antimicrobial agents is a common feature of *C. jeikeium* clinical isolates. Tauch *et al.* (2005) have published the genome sequence of the clinical isolate *C. jeikeium* K411. This strain contains a circular chromosome of ~2·5 Mbp and a ~15 kbp bacteriocin-producing plasmid (pKW4). About half the *C. jeikeium* genes (~52 %) constitute a 'chromosomal backbone' of conserved genes found in all four *Corynebacteria* species sequenced to date (*C. glutamicum*, *C. efficiens*, *C. diphtheriae* and *C. jeikeium*).

Haemophilus influenzae strain Rd was the first bacterium to have its genome completely sequenced (Fleischmann *et al.*, 1995) and was also the first bacterial genome to be patented (O'Malley *et al.*, 2005), lending to this genome sequence a certain historical significance. In the decade since 1995, about 250 bacterial genomes have been sequenced and for many bacterial species multiple genome sequences have become available (for example, note that for all six of the genomes listed in Table 1, at least one other genome has been sequenced from the same genus). A second *H. influenzae* isolate has now been sequenced (Harrison *et al.*, 2005). It is not generally appreciated that the originally sequenced *H. influenzae* strain was a rough form of a serotype not normally associated with disease. This was not stated as such in the publication (Fleischmann *et al.*, 1995) and in this light it is an improvement that the sequence of a pathogenic, non-typeable serotype of *H. influenzae* has now been completed.

The gene content of the newly sequenced *H. influenzae* strain 86-028NP was compared with the *H. influenzae* rough serotype d strain KW20 (Rd). In total, 280 ORFs were identified in strain 86-028NP that are absent in the previously sequenced Rd strain, and 169 of the genes found in the Rd strain were missing in strain 86-028NP. However, the Rd sequence had been annotated when genome sequences were largely *terra incognita*, and annotation by comparative genomic methods was not possible. Annotation techniques have improved substantially as hundreds of genomes have been sequenced and annotated. Several studies have indicated that bacterial genomes can be overannotated, and in the *H. influenzae* Rd genome there could be as many as 200 genes annotated that might not be real (Skovgaard *et al.*, 2001). In addition to the 'extra' genes, there could well be missing genes, for example, any small non-coding RNAs that can play important regulatory roles. In our opinion, using the same set of 1709 genes annotated 10 years ago in the original *H. influenzae* GenBank file as 'gospel' is a missed chance to take advantage of the progress that has been made in 10 years of microbial genome annotation.

Of most interest are genes found in the new genome that are related to virulence. These include genes whose products are involved in adherence, of which 5 of 12 are present in strain Rd as well, and two others have contingency repeats (short stretches of simple base repeats, for example GGGGG, which can slip during replication, resulting in addition or deletion of a single base, changing the reading frame of the gene; Bayliss *et al.*, 2001). A total of 52 genes were identified for LPS biosynthesis, only four of which are unique to 86-028NP. At least eight LPS genes have contingency repeats. Iron acquisition and the oxidative stress response are important processes in determining virulence of *H. influenzae*. Strain 86-028NP has 21 genes involved in iron acquisition (of which 20 have homologues in Rd) and the genes thought to be involved in the oxidative stress response are relatively conserved between the two species. While these simple comparisons suggest that the difference in virulence may lie in differences in adherence, this needs to be confirmed with experimental data.

***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.

Charles Dorman, Editor-in-Chief