

Genome Update: Protein secretion systems in 225 bacterial genomes

Genomes of the month

Seven new microbial genomes have been published since last month's Genome Update column was written. As usual, there is a heavy bias towards members of the *Proteobacteria* (which constitute about half of all the bacterial genomes sequenced so far; 107 out of 225).

The list of new genomes includes five *Proteobacteria*. Two are members of the α -*Proteobacteria*, the rickettsia *Ehrlichia ruminantium*, and *Gluconobacter oxydans*, which is of interest to the food industry. Three are γ -*Proteobacteria*, the 'warfare germ' *Francisella tularensis*, the plant pathogen *Xanthomonas oryzae* and finally *Vibrio fischeri*. The other genomes are of the probiotic *Lactobacillus acidophilus* and the archaeon *Thermococcus kodakaraensis*. A brief overview of these genomes is given below and a summary is presented in Table 1.

Ehrlichia ruminantium is an obligate intracellular bacterium that causes heartwater, a tick-borne disease with high mortality in livestock. Although this infectious disease is currently restricted to Africa, it threatens to invade the Americas. It is hoped that the genome sequence of *Ehrlichia ruminantium* strain Welgevonden will facilitate the development of potential vaccines against heartwater. The genome is 1.5 Mb long, with an A+T content of 73 mol%, and encodes 888 proteins. There are 36 tRNA and 3 rRNA genes (Collins *et al.*, 2005). One remarkable property is the large number of tandemly repeated and duplicated sequences. Thirty-two pseudogenes were also found, most of them are truncated fragments of genes associated with repeats. It seems likely that the identified pseudogenes are products of ongoing sequence duplication events (Collins *et al.*, 2005). The pathogenicity genes are likely to be secreted by a Type IV secretion system (more on this below in Method of the month).

Francisella tularensis is a Gram-negative aerobic bacterium with two main serotypes: Jellison Type A and Type B. Type A is the more virulent form. It is one of the most infectious human pathogens and has long been considered as a potential biological weapon. Especially in the 1950s and 1960s, *F. tularensis* was examined by the US military and as recently as the early 1990s the Soviet Union displayed a special interest in tularaemia. The genome of *F. tularensis* strain SCHU S4 is 1.9 Mb long, with a high A+T content of 77 mol% (Larsson *et al.*, 2005). There are 38 predicted tRNAs and 3 rRNAs operons. The genome encodes a total of 1804 predicted genes, including 201 pseudogenes; 1281 of these predicted genes have homologues in one or more γ -proteobacterial genomes. Potential virulence or virulence-associated genes were identified as part of a putative pathogenicity island that is duplicated in the genome. The genome sequence should increase understanding of metabolic pathways in this organism and contribute to development of strategies against this potential biological warfare and bioterrorism agent (Larsson *et al.*, 2005). A type I secretion system could be identified, but complete gene clusters encoding type III, IV or V secretion systems were not found.

Gluconobacter oxydans belongs to the family *Acetobacteraceae*. These organisms have been used since ancient times in

biotechnological processes like the production of vinegar and are still used for industrial applications which take advantage of their ability to incompletely oxidize a great variety of carbohydrates, alcohols and related compounds. The genome of *Gluconobacter oxydans* strain 621H (Prust *et al.*, 2005) consists of one main 2.7 Mb circular chromosome and an additional five plasmids (pGOX1–pGOX5, ranging from 163 to 2.6 kb), comprising 232 ORFs. The genome has an A+T content of 39 mol% and contains 2432 ORFs (1877 of these have an assigned function). The unique metabolism of *Gluconobacter oxydans* makes it an ideal model organism to study microbial processing of food.

Lactobacillus acidophilus is a representative member of the lactic acid bacteria, which are used in food and feed fermentations, such as dairy and silage. Lactic acid bacteria include probiotic strains, several of which have been sequenced. The 2.0 Mb genome of *Lactobacillus acidophilus* NCFM (Altermann *et al.*, 2005) with 1864 predicted CDSs is relatively small and also has limited biosynthetic capabilities; most amino acids, cofactors and vitamins cannot be synthesized. On the other hand, the genome contains a relatively large fraction of genes involved in taking up various sugars and amino acids/peptides. This is consistent with its relatively nutrient-rich habitat of the gastrointestinal (GI) tract.

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