microbiology Comment

Genome Update: tRNAs in sequenced microbial genomes

Genomes of the month – microbial genome evolution

Eight microbial genomes have been published in the four weeks since the last Genome Update was written (Ussery *et al.*, 2004). They represent five bacterial and three eukaryotic organisms, and provide several interesting aspects of genome evolution. A very brief overview of the new genomes will be presented below; this is meant merely to wet the appetite of the reader and to provide pointers to the relevant recent literature.

Two spirochaete genomes have been published this month, bringing the total number of genomes from three to five for this phylum. The genome of *Treponema denticola* strain ATCC 35405 (Seshadri *et al.*, 2004) is more than twice the size of the previously sequenced genome of *Treponema pallidum* (2·8 Mbp vs 1·1 Mbp), although the number of tRNAs and rRNAs are about the same in both genomes. The difference in genome size appears to be the result of a combination of three types of evolution: genome reduction, lineage-specific recombination and horizontal gene transfer (Seshadri *et al.*, 2004). The other newly sequenced spirochaete genome, of *Leptospira interrogans* serovar Copenhageni strain Fiocruz L1-130 (Nascimento *et al.*, 2004), has two chromosomes and encodes 3728 genes, two rRNA operons and 37 tRNAs, as shown in Table 1. This genome is nearly identical in size to that of *L. interrogans* serovar Lai (Ren *et al.*, 2003), which has 4727 annotated genes, or nearly 1000 extra genes. This is perhaps due to the difference in cut-off values for gene-finding from the two different groups.

Members of the *Chlamydiae* are amongst the most successful bacterial pathogens of humans, and there are currently eight sequenced pathogenic chlamydial genomes, ranging in size from 1·0 to 1·2 Mbp (see table on supplemental web page). Recently, it was discovered that *Chlamydia* and related species can also exist in free-living amoebae, and the genome of the *Acanthamoeba* sp. endosymbiont *Parachlamydia* sp. UWE25 has now been sequenced (Horn *et al.*, 2004); at 2·4 Mbp, it is about twice the size of the other chlamydial genomes. It is estimated that the last common ancestor for the pathogenic and symbiotic chlamydia was about 700 million years ago, and that this bacterium already contained many of the virulence factors found in modern pathogenic chlamydia (Horn *et al.*, 2004).

The thermophilic and halotolerant bacterium *Thermus thermophilus* has become a model organism for structural biology, as many of its proteins have been crystallized and their structures determined. Examination of the genome of *Thermus thermophilus* strain HB27, which can grow at temperatures up to 85 °C, has revealed some clues as to what it might take to live in a hot-spring environment (Henne *et al.*, 2004). Based on its genome sequence, it looks like this bacterium is a scavenger which lives on solid surfaces and takes up nutrients as they pass by.

The genome of the parasite *Wolbachia pipientis* wMel is unusual in that it is both streamlined and also contains high levels of repeats and mobile DNA elements (Wu *et al.*, 2004). Thus, for this bacterium, natural selection appears to be a bit inefficient, probably due to repeated population bottlenecks (Wu *et al.*, 2004).

Three eukaryotic genomes have also been sequenced this month. As usual, unfortunately the quality of the eukaryotic sequences is not as good as that of the prokaryotic genomes; there are many gaps in the sequences, and also the annotation (when present) is patchy at best (in our opinion). According to Kellis *et al.* (2004), the genome sequence of the yeast *Kluyveromyces waltii* strain NCYC 2644 compared to that of *Saccharomyces cerevisiae* provides '*the first comparison across an ancient whole genome duplication event and offers the opportunity to study the long-term fate of a genome after duplication*'. The intracellular pathogen *Cryptosporidium parvum* type II isolate has a genome of about 9·1 Mbp in length and encodes a mere 3800 proteins (Abrahamsen *et al.*, 2004). (Note that this is about the size of a medium to small bacterial proteome!) This parasite has undergone massive genome reduction and streamlining, even losing all of its mitochondrial DNA, which has been incorporated into the main chromosome. Finally, the genome of the alga *Cyanidioschyzon merolae* 10D (Matsuzaki *et al.*, 2004) is 16·5 Mbp long and spread over 20 chromosomes. There are very few introns, and only three rRNA operons (see Table 1). This genome

***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., Chromosomes.

Genome	Size (bp)	AT content (%)	rRNA operons	tRNAs	CDS	Accession nos
Leptospira interrogans serovar Copenhageni Fiocruz L1-130	5 260 086	64·9	2	37	3728	AE016823
Parachlamydia sp. UWE25	2 414 465	65·3	4	35	2031	BX908798
Thermus thermophilus HB27	1 894 877	31·6	2	47	1988	AE017221
Treponema denticola ATCC 35405	2 843 201	62·1	6	44	2786	AE017226
Wolbachia pipientis wMel	1 267 782	64·8	1	34	1270	AE017221
Cryptosporidium parvum type II (8 chr.)	~9 100 000	70·0	5	45	3807	AAEE01000000
Cyanidioschyzon merolae 10D (20 chr.)	16 520 305	45·0	3	30	5331	AP006483–AP006502
Kluyveromyces waltii NCYC 2644 (8 chr.)	10 613 225	55·6	2	244	5230	AADM01000000

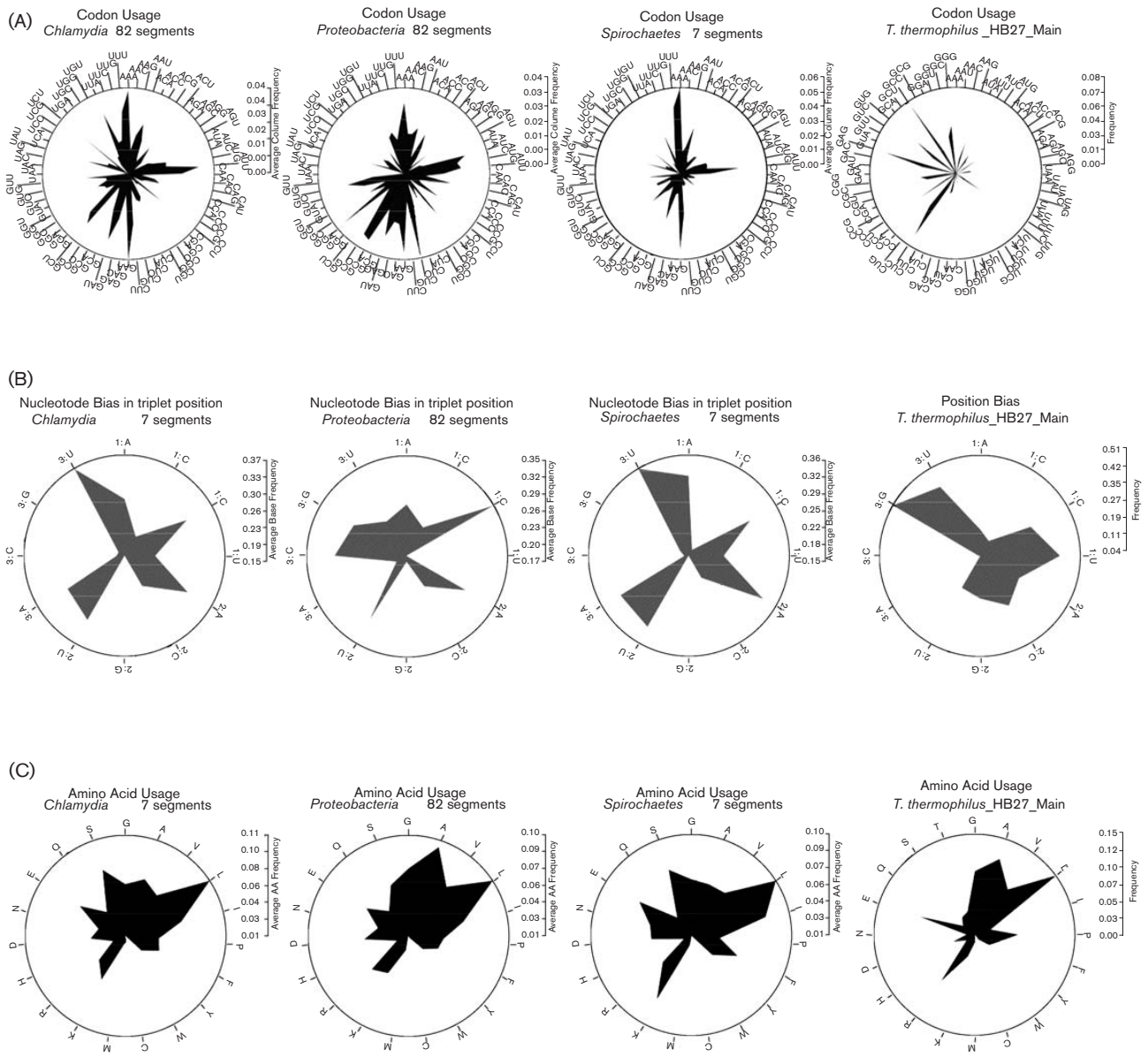

Fig. 1. Codon usage 'rose-plots' for genomes of the phyla *Chlamydiae*, *Proteobacteria* and *Spirochaetes*, and for *Thermus thermophilus*. (A) The first row is of codon usage for all the chromosomes within a given phylum, normalized to the most commonly used codon (outside circle). Note that several of the codons are rarely used, whilst others are quite common. (B) The bias of the third position of the codon. (C) Amino acid usage. Note that in all three phyla, leucine is the most commonly used amino acid.

provides '*a model system with a simple gene composition for studying the origin, evolution and fundamental mechanisms of* [photosynthetic] *eukaryotic cells*' (Matsuzaki *et al.*, 2004).

Method of the month – comparison of tRNA genes in sequenced genomes

The number of tRNA genes in bacterial genomes ranges from 126 in *Vibrio parahaemolyticus* to 29 in *Mycoplasma pulmonis*. Since there are a maximum of 61 possible codons (and hence different tRNA genes), some genomes obviously have missing tRNAs, although all of the genomes can code for the use of all 20 amino acids. The use of base wobble in the third position allows for a given tRNA gene to utilize certain codons which differ only in the third position. Thus, for example, in the case of the *Mycoplasma pulmonis* genome, even though there are only 29 tRNA genes, all 61 codons are found within the protein-coding sequences. However, the frequency of usage within the coding regions varies considerably – for example, of the six possible codons for leucine, UUA is used 13 272 times, whilst CUG is only used 165 times, or nearly 100-fold less.

Codon usage plots for three different phyla and one species are shown in Fig. 1(A). Note that some codons (such as AAA and GAA) are used frequently in all phyla, whilst other codons, such as UAA, UAC and UAU, are used infrequently. A change in the third position in the codon often will code for the same amino acid, and bias in this position is correlated with changes in the AT content of the genome. For example, in the *M. pulmonis* genome mentioned above, the CUN codon usage is strongly biased towards U or A (CUC is only used 399 times, compared to 7523 for CUU and 3932 for CUA). Thus, an AT-rich genome and a GC-rich genome might code for a similar amino acid composition, but each genome would have a different third position bias, as can be seen in Fig. 1(B). Finally, the overall amino acid composition of the genomes from the three different phyla look quite similar, with the amino acids leucine, alanine, glycine and serine being most abundant, and tryptophan, cytosine, histidine and methionine being used infrequently.

A brief word should be mentioned about alternative genetic codes, where 'stop codons' can actually code for an amino acid. First, of course, in the genomes of *Mycoplasma* spp., the stop codon UGA can code for tryptophan (Yamao *et al.*, 1985). Furthermore, selenium is incorporated into some enzymes and has been shown to be incorporated as selenocysteine, again utilizing the UGA stop codon which, with the right enzymic machinery, can code for selenocysteine in other bacterial genomes (Zinoni *et al.*, 1987). About one-quarter of a set of bacterial genomes examined (13/54) contained potential genes incorporating selenocysteine (Wassenaar & Meinersmann, 2003). A 22nd amino acid has also been proposed, which utilizes the stop codon UAG to code for pyrrolysine (Srinivasan *et al.*, 2002). Finally, tRNA editing describes the post-transcriptional modification of a tRNA so that it can only recognize a particular triplet; this has been described for an *Escherichia coli* or a *Bacillus subtilis* tRNA with anticodon CAU (normally encoding Met), which is modified to translate codon AUA exclusively, and is loaded with Ile (Grosjean & Björk, 2004). This may be an explanation as to why a Met tRNA is frequently found duplicated in bacterial genomes, although Met is not a frequently used amino acid.

Next month, the number of genes per genome will be discussed. At the time of writing, the bacterial genome with the fewest genes is that of *Mycoplasma genitalium*, with a mere 480 genes, whilst the largest is that of *Bradyrhizobium japonicum*, with 8317 genes.

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/GenUp005/

Acknowledgements

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

David W. Ussery,[1] Peter F. Hallin,[1] Karin Lagesen[2] and Trudy M. Wassenaar[3]

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

[2]Department of Molecular Biology, Institute of Medical Microbiology, University of Oslo, The National Hospital, NO-0027 Oslo, Norway

[3]Molecular Microbiology and Genomics Consultants, Tannestrasse 7, D-55576 Zotzenheim, Germany

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

Abrahamsen, M. S., Templeton, T. J., Enomoto, S. & 17 other authors (2004). Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441–445.

Grosjean, H. & Björk, G. R. (2004). Enzymatic conversion of cytidine to lysidine in anticodon of bacterial $tRNA^{Ile}$ – an alternative way of RNA editing. *Trends Biochem Sci* **29**, 165–168.

Henne, A., Bruggemann, H., Raasch, C. & 17 other authors (2004). The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nat Biotechnol* Epub ahead of print, DOI: 10.1038/nbt956

Horn, M., Collingro, A., Schmitz-Esser, S. & 10 other authors (2004). Illuminating the evolutionary history of *Chlamydiae*. *Science* Epub ahead of print, DOI: 10.1126/science.1096330

Kellis, M., Birren, B. W. & Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624.

Matsuzaki, M., Misumi, O., Shin, I. T. & 39 other authors (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428**, 653–657.

Nascimento, A. L., Verjovski-Almeida, S., Van Sluys, M. A. & 9 other authors (2004). Genome features of *Leptospira interrogans* serovar Copenhageni. *Braz J Med Biol Res* **37**, 459–477.

Ren, S. X., Fu, G., Jiang, X. G. & 36 other authors (2003). Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature* **422**, 888–893.

Seshadri, R., Myers, G. S., Tettelin, H. & 36 other authors (2004). Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes. *Proc Natl Acad Sci U S A* **101**, 5646–5651.

Srinivasan, G., James, C. M. & Krzycki, J. A. (2002). Pyrrolysine encoded by UAG in *Archaea*: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459–1462.

Ussery, D. W., Hallin, P. F., Lagesen, K. & Coenye, T. (2004). Genome Update: rRNAs in sequenced microbial genomes. *Microbiology* **150**, 1113–1115.

Wassenaar, T. M. & Meinersmann, R. J. (2003). The TGA stop codon and the phylogeny of the selenocysteine pathway. *Genome Lett* **2**, 127–138.

Wu, M., Sun, L. V., Vamathevan, J. & 27 other authors (2004). Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* **2**, E69.

Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y. & Osawa, S. (1985). UGA is read as tryptophan in *Mycoplasma capricolum*. *Proc Natl Acad Sci U S A* **82**, 2306–2309.

Zinoni, F., Birkmann, A., Leinfelder, W. & Bock, A. (1987). Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. *Proc Natl Acad Sci U S A* **84**, 3156–3160.

DOI 10.1099/mic.0.27260-0

Methylotrophy versus heterotrophy: a misconception

I still remember the confusion in my mind when, in 1992, as a first-year PhD student, I attended the 7th 'C_1-meeting' at the University of Warwick. I was working with methanol and methane oxidizers and several times speakers in the conference used the word *heterotroph* to quickly describe non-methylotrophic organisms or non-methylotrophic metabolism within facultative methylotrophs. Based on my first-degree knowledge, I was pretty sure the bugs I was studying were heterotrophic, but I humbly thought I had a lot to learn and that my doubts would be solved by reading more about the subject. However, a few years later now, after making some order in the facts about the types of methylotrophic metabolism that have been described, I am sure that *heterotroph* is not the antonym (word of opposite meaning) of *methylotroph*, and that my original sense of confusion was not caused by my ignorance.

Let us start with three simple textbook definitions.

- Autotroph: an organism that derives its cell carbon from CO_2 (inorganic carbon) by fixation and reduction.
- Heterotroph: an organism that obtains its cell biomass by incorporating directly reduced (organic) molecules.
- Methylotroph: an organism that derives energy and, in many cases, cell carbon from reduced molecules that have no C–C bond (also called C_1 compounds).

From these standard definitions it is clear that methylotrophy is not the opposite of heterotrophy. Nevertheless, in many reports in the literature (Megraw & Knowles, 1989; Kraffzik & Conrad, 1991; Spivak & Rokem, 1994, 1995; Thompson *et al.*, 1995; Nanba *et al.*, 1999; Goodwin *et al.*, 2001; Bothe *et al.*, 2002; Korotkova *et al.*, 2002; Chistoserdova *et al.*, 2003; Van Dien *et al.*, 2003) and at congresses, the term heterotroph has been constantly used to define non-methylotrophs (contaminants or symbionts for instance) even by some of the most pre-eminent scholars in the field. In some cases (Levering *et al.*, 1981; Levering & Dijkhuizen, 1985) the uneasiness with this choice surfaced in the usage of inverted commas ('heterotrophic').

As a matter of fact, methylotrophy does not describe one type of metabolism; it includes under a common name a bunch of different ways of utilizing C_1 compounds. I see at least four reasons that can account for this inappropriate usage of the term heterotroph.

- Some micro-organisms do indeed grow on C_1 compounds autotrophically, fixing the CO_2 produced (*Ralstonia*, *Xanthobacter*, *Paracoccus*, the methylotrophic *Archaea*, the methylotrophic clostridia). Thus, for these organisms (exclusively) *heterotrophic* would be the proper antonym of *methylotrophic*. Furthermore, other autotrophic micro-organisms (lithotrophs or phototrophs) can utilize C_1 compounds (methanol, formate) as a supplementary source of energy besides their more typical ones.
- Some methylotrophs are mixotrophic. *Methylococcus capsulatus* fixes carbon derived from methane mostly at the level of formaldehyde, through the ribulose monophosphate (RuMP) pathway (heterotrophic), but also in part at the level of CO_2 through the Calvin–Benson–Bessham (CBB) cycle (autotrophic) (Taylor *et al.*, 1980; Baxter *et al.*, 2002). Also, the serine cycle for carbon fixation is intrinsically mixotrophic in that it incorporates one molecule of CO_2 for every two of formaldehyde fixed.
- Many of the reactions that compose the RuMP pathway are common to the CBB cycle. Some authors (Quayle & Ferenci, 1978) have hypothesized that the CBB cycle actually originated from the RuMP pathway and that organisms like *Methylococcus capsulatus*, in which both pathways work simultaneously, may be 'transition organisms' between the two metabolic types (Taylor *et al.*, 1980).
- Formaldehyde, which is the molecule fixed by the heterotrophic methylotrophs, is obviously not inorganic, but it is not at all 'far' from CO_2 (just two dehydrogenation steps away); it is easy to slip into the mistake of considering it as 'almost inorganic'. Or as Foster put it back in 1951 (as cited in Quayle, 1961): '...*It is more the fact that these organisms synthesize their complex cell constituents from simple 1-carbon compounds chemically analogous to carbon dioxide that has resulted in their association with*