

Genomics update

Sorting out the mix in microbial genomics

OnlineOpen: This article is available free online at www.blackwell-synergy.com

Michael Y. Galperin*

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

The relatively small number of microbial genomes completed in the past two months (Table 1) includes, however, representatives of two new bacterial phyla, *Dictyoglomi* and *Nitrospirae*. To highlight the first genome sequences from these poorly studied taxa, they have been placed in a new section at the top of Table 1.

So far, little is known about either *Dictyoglomus thermophilum* or *Thermodesulfovibrio yellowstonii*. Both are Gram-negative thermophilic heterotrophs with an extremely low (29 mol%) G+C content of their chromosomal DNA (Saiki *et al.*, 1985; Henry *et al.*, 1994). *Dictyoglomus thermophilum* is an obligate anaerobe that grows optimally at 70–73°C. It was isolated from the Tsuetate hot spring in Japan (Saiki *et al.*, 1985) and used to purify three extremely heat-stable amylases (Horinouchi *et al.*, 1988). *Thermodesulfovibrio yellowstonii* was isolated from a thermal vent in Yellowstone Lake in Wyoming. It contains c-type cytochromes and grows optimally at 65°C using lactate, pyruvate or formate plus acetate as substrates and can use sulfate, thiosulfate and sulfite as terminal electron acceptors (Henry *et al.*, 1994). Analysis of these genomes should provide a window into the physiology and evolutionary relationships of these new bacterial lineages.

Completion of these two genomes is a major step towards the goal of having at least one complete genome sequence from representatives of all major prokaryotic groups. Indeed, we now have at least one completely sequenced genome for 18 bacterial phyla out of the 24 listed in the taxonomic outline in the second volume of the *Bergey's Manual* (Garrity *et al.*, 2004; available at <http://www.bergeys.org/outlines.html>). Representatives of five more bacterial phyla are at various stages of genome sequencing: *Chrysiogenes arsenatis* (phylum *Chrysiogenetes*), *Denitrovibrio acetiphilus* (*Deferribacteres*),

Fibrobacter succinogenes (*Fibrobacteres*), *Thermodesulfovibrium commune* (*Thermodesulfobacteria*) and *Thermomicrobium roseum* (*Thermomicrobia*, which may be considered a class in the phylum *Chloroflexi*). Only one of those 24 phyla (*Gemmatimonadetes*) remains not a subject of any publicly announced genome sequencing project. This is obvious progress in coverage of microbial diversity compared with the status of the genome sequencing just two years ago (Galperin, 2006),

However, this nice and clear picture of the bacterial taxonomy and the corresponding genome sequencing efforts is complicated by several different factors. First of all, high-rank bacterial taxonomy is still in the state of flux: new candidate phyla are being identified and new genome sequencing projects are being planned to characterize their representatives. There are ongoing sequencing projects for *Lentisphaera araneosa* and *Thermanaerovibrio acidaminovorans*, cultured representatives, respectively, of the recently recognized phyla *Lentisphaerae* (Cho *et al.*, 2004) and *Synergistetes* (*Aminanaerobia*) (Hongoh *et al.*, 2007; Jumas-Bilak *et al.*, 2007). In addition, genomic sequencing is being performed on candidate phyla that were initially deduced based solely on the clustering of 16S rRNA sequences. Examples include a nearly complete genome from the candidate phylum TM7, which still has no cultivated members (Marcy *et al.*, 2007) and two recently sequenced genomes from representatives of the candidate phylum Termite group 1 (TG1), one of which, “*Elusimicrobium minutum*”, was in the meantime successfully cultivated. Sometimes genomic data reveal distant similarities between two or more phyla, which results in their unification into a group (e.g. *Bacteroidetes/Chlorobi*, *Fibrobacteres/Acidobacteria*) or a superphylum, e.g. *Chlamydiae/Verrucomicrobial/Planctomycetes/Lentisphaerae* (Wagner and Horn, 2006; Hou *et al.*, 2008). Besides, certain validly described bacterial groups still lack any sequence information (Yarza *et al.*, 2008). Finally, there are several alternative classifications of bacteria that made their way into taxonomic literature but, for a variety of reasons, failed to gain acceptance in the community (Gupta, 1998; 2000; Cavalier-Smith, 2002; 2006). Another such example is the already mentioned (Galperin, 2008) recent transfer of *Mollicutes* from the

*For correspondence. E-mail galperin@ncbi.nlm.nih.gov; Tel. (+1) 301 435 5910; Fax (+1) 301 435 7793.

Re-use of this article is permitted in accordance with the Creative Commons Deed, Attribution 2.5, which does not permit commercial exploitation.

Table 1. Recently completed microbial genomes (August–September 2008).

Species name	Taxonomy	GenBank accession	Genome size (bp)	Proteins (total)	Sequencing centre ^a	Reference
New taxa						
<i>Dictyoglomus thermophilum</i>	<i>Dictyoglomi</i>	CP001146	1 959 987	1912	TIGR/JCVI	Unpublished
<i>Thermodesulfovibrio yellowstonii</i>	<i>Nitrospirae</i>	CP001147	2 003 803	2033	TIGR/JCVI	Unpublished
New organisms						
<i>Plasmodium knowlesi</i>	<i>Eukaryota,</i> <i>Apicomplexa</i>	AM910983– AM910996	23.5 Mbp	5188	Sanger institute	Pain <i>et al.</i> (2008)
<i>Plasmodium vivax</i>	<i>Eukaryota,</i> <i>Apicomplexa</i>	AAKM00000000	26.8 Mbp	5433	TIGR/JCVI	Carlton <i>et al.</i> (2008)
<i>Coprothermobacter proteolyticus</i>	<i>Firmicutes</i>	CP001145	1 424 912	1482	JCVI	Unpublished
<i>Streptococcus equi</i>	<i>Firmicutes</i>	CP001129	2 024 171	1893	Methodist hospital	Beres <i>et al.</i> (2008)
<i>Phenylobacterium zucineum</i>	<i>α-Proteobacteria</i>	CP000747, CP000748	3 996 255 382 976	3854	Zhejiang University	Luo <i>et al.</i> (2008)
<i>Acidithiobacillus ferrooxidans</i> ATCC 23270	<i>γ-Proteobacteria</i>		2 982 397	3217	TIGR/JCVI	Valdés <i>et al.</i> (2008)
<i>Acidithiobacillus ferrooxidans</i> ATCC 53993	<i>γ-Proteobacteria</i>	CP001132	2 885 038	2826	JGI	Unpublished
<i>Alteromonas macleodii</i> 'Deep ecotype'	<i>γ-Proteobacteria</i>	CP001103	4 412 282	4072	JCVI	Unpublished
<i>Anaeromyxobacter</i> sp. K	<i>δ-Proteobacteria</i>	CP001131	5 061 632	4457	JGI	Unpublished
<i>Geobacter bemidjiensis</i>	<i>δ-Proteobacteria</i>	CP001124	4 615 150	4018	JGI	Unpublished
<i>Borrelia duttonii</i>	<i>Spirochaetes</i>	CP000976– CP000992	1.57 (total)	1305	Genoscope	Lescot <i>et al.</i> (2008)
<i>Borrelia recurrentis</i>	<i>Spirochaetes</i>	CP000993– CP001000	1.24 (total)	990	Genoscope	Lescot <i>et al.</i> (2008)
New strains						
<i>Klebsiella pneumoniae</i> 342	<i>γ-Proteobacteria</i>	CP000964, CP000965, CP000966	5 641 239 187 922 91 096	5768	JCVI	Fouts <i>et al.</i> (2008)
<i>Salmonella enterica</i> ssp. <i>enterica</i> serovar Agona str. SL483	<i>γ-Proteobacteria</i>	CP001138, CP001137	4 798 660 37 978	4614	JCVI	Unpublished
<i>Salmonella enterica</i> ssp. <i>enterica</i> serovar Dublin str. CT_02021853	<i>γ-Proteobacteria</i>	CP001144, CP001143	4 842 908 74 551	4617	JCVI	Unpublished
<i>Salmonella enterica</i> ssp. <i>enterica</i> serovar Gallinarum str. 287/91	<i>γ-Proteobacteria</i>	AM933173	4 658 697	3965	Sanger Institute	Thomson <i>et al.</i> (2008)
<i>Salmonella enterica</i> ssp. <i>enterica</i> serovar Enteritidis str. P125109	<i>γ-Proteobacteria</i>	AM933172	4 685 848	4206	Sanger Institute	Thomson <i>et al.</i> (2008)
<i>Salmonella enterica</i> ssp. <i>enterica</i> serovar Paratyphi A str. AKU_12601	<i>γ-Proteobacteria</i>	FM200053	4 581 797	4078	Sanger Institute	Unpublished
<i>Aliivibrio (Vibrio) fischeri</i> MJ11	<i>γ-Proteobacteria</i>	CP001139, CP001133, CP001134	2 905 029 1 418 848 179 459	4039	JCVI	Unpublished
<i>Helicobacter pylori</i> G27	<i>ε-Proteobacteria</i>	CP001173, CP001174	1 652 982 10 031	1504	University of Oregon	Unpublished

a. Sequencing centre names are abbreviated as follows: Genoscope, Centre National de Séquençage, Evry, France; JCVI, J. Craig Venter Institute, Rockville, Maryland, USA; JGI, US Department of Energy Joint Genome Institute, Walnut Creek, California, USA; Methodist Hospital, Center for Molecular and Translational Human Infectious Disease Research, The Methodist Hospital Research Institute, Houston, Texas, USA; Sanger Institute, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK; TIGR, The Institute for Genomic Research (currently J. Craig Venter Institute), Rockville, Maryland, USA; University of Oregon, Institute of Molecular Biology, University of Oregon, Eugene, Oregon, USA; Zhejiang University, Cancer Institute, The Second Affiliated Hospital, Zhejiang University Medical School, Hangzhou, Zhejiang, China.

phylum *Firmicutes* into a new phylum *Tenericutes* in the latest edition of *Bergey's Manual* (Ludwig *et al.*, 2008). The phylogenetic trees that served as the rationale for that move show numerous inconsistencies and hardly justify the decision to create this new phylum.

It must be noted that back in 1992, Sneath and Brenner stated 'There is no such thing as an official classification' (see <http://www.bacterio.cict.fr/Sneath-Brenner.html>). This point was recently reiterated by J.P. Euzéby, whose List of Prokaryotic names with Standing in Nomenclature

(<http://www.bacterio.cict.fr/>) includes an up-to-date listing of commonly recognized prokaryotic phyla (<http://www.bacterio.cict.fr/classifphyla.html>). For a quick look at the current state of microbial genome sequencing, the easiest tool might be the NCBI's Tax Tree (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html), which lists both completed and ongoing genome sequencing projects. However, for those interested in the emerging microbial diversity, the best source of information is probably the 'greengenes' website ([Journal compilation © 2008 Society for Applied Microbiology and Blackwell Publishing Ltd, *Environmental Microbiology*, **10**, 3187–3192
No claim to original US government works](http://</p>
</div>
<div data-bbox=)

greengenes.lbl.gov), which includes six different classification schemes, from the most conservative ones (the Ribosomal Database project) to the schemes by Pace and Hugenholtz that include up to 100 bacterial lineages (see http://greengenes.lbl.gov/Download/Taxonomic_Outlines/five_way_venn.png for a nice graphical representation of the relation between these schemes).

Among archaea, Bergey's taxonomic outline recognized two phyla, *Crenarchaeota* and *Euryarchaeota*, which are now represented by 50 complete genomes (16 and 34 respectively). In addition, there are two newly suggested but not yet officially recognized phyla, each represented by a single genome, "*Candidatus* Korarchaeum cryptofilum" (*Korarchaeota*) and *Nanoarchaeum equitans* (*Nanoarchaeota*). Again, classifications by Pace and Hugenholtz at <http://greengenes.lbl.gov> include up to 40 archaeal lineages.

Among eukaryotic microorganisms, few genomes have been sequenced to any significant degree, and most of these genomes represent just a handful of lineages: *Amoebozoa* (*Dictyostelium*, *Entamoeba*), *Apicomplexans* (*Cryptosporidium*, *Plasmodium*, *Toxoplasma*), *Ciliophora* (*Paramecium*, *Tetrahymena*) and *Kinetoplasts* (*Leishmania*, *Trypanosoma*). There are partially sequenced representatives of choanoflagellates (*Monosiga*), diplomonads (*Giardia*), parabasalids (*Trichomonas*), rhodophytes (*Cyanidioschyzon*) and stramenopiles (*Phytophthora*, *Thalassiosira*). Several more eukaryotic lineages (*Apusozoa*, *Haptophyceae* *Heterolobosea*) are going to be covered by ongoing sequencing projects, but the overall coverage of major eukaryotic groups is extremely poor and will remain that way at least for the next several years.

In summary, recent genomic projects are successfully covering the diversity of cultured bacteria and archaea. Any significant coverage of the main eukaryotic lineages or of prokaryotic lineages identified through cultivation-independent methods will remain the challenge for years to come.

In eukaryotic genomics, the biggest news was the completion of the genome sequences of two malaria parasites, *Plasmodium knowlesi* and *Plasmodium vivax*. Their back-to-back publication in *Nature* (Carlton *et al.*, 2008; Pain *et al.*, 2008) was accompanied by an excellent commentary (Winzeler, 2008) on the contribution of the genomic data to the progress of malaria research in the 6 years that have passed since the completion of the genomes of *Plasmodium falciparum* (Gardner *et al.*, 2002) and *Plasmodium yoelii* (Carlton *et al.*, 2002).

Coprothermobacter proteolyticus (formerly known as *Thermobacteroides proteolyticus*) is a moderately thermophilic proteolytic bacterium, originally isolated from the fermentation sludge of an anaerobic digester treating cattle manure mixed with tannery waste (Ollivier *et al.*, 1985). It can grow at temperatures of up to 75°C with an

optimum at 63°C. Although *C. proteolyticus* was initially described as a Gram-negative bacterium and therefore suggested to belong to a deep bacterial lineage, potentially at the phylum level (Rainey and Stackebrandt, 1993), analysis of its 16S rRNA revealed that it is related to *Thermoanaerobacter* sp. It is currently assigned to the family *Thermodesulfobiaceae* (Mori *et al.*, 2003) within the order *Thermoanaerobacteriales* and is the first sequenced genome from that family.

Phenylobacterium zucineum is an α -proteobacterium recently isolated from a human erythroleukemia cell line (Zhang *et al.*, 2007). All close relatives of *P. zucineum* are free-living environmental organisms, and its 4.4 Mb genome is much larger than that of any intracellular parasite or symbiont characterized so far. Indeed, the genome sequence (Luo *et al.*, 2008) revealed similarities with the genome of *Caulobacter crescentus*. However, fragments of *P. zucineum* genomic DNA were found among the EST libraries from breast cancer and lymphatic cell lines, suggesting that this organism might survive in proliferative tissues.

Acidithiobacillus ferrooxidans (previously known as *Thiobacillus ferrooxidans*; Kelly and Wood, 2000), is an obligately acidophilic chemolithoautotrophic γ -proteobacterium, a popular model organism to study bacterial membrane energetics at acidic pH values (see Ferguson and Ingledew, 2008 for a recent review). It gains energy by oxidizing ferrous iron and is able to grow in the pH range from 1.3 to 4.0 using CO₂ as the sole source of carbon. *Acidithiobacillus ferrooxidans* is a major component of microbial consortia used in bio-mining to extract copper, zinc and other metals from low-grade ores. With the recent increase in the price of gold, *A. ferrooxidans*-based microbial consortia are increasingly used to improve recovery of gold from arsenopyrite ores. Despite the importance of this organism (or maybe because of it), sequencing of the *A. ferrooxidans* genome had a long and convoluted history. The first (incomplete or 'gapped') genome sequence of the type strain *A. ferrooxidans* ATCC 23270 was produced at the Integrated Genomics in 1999. It consisted of 1353 contigs covering 2611 kb and coding for 2712 proteins; it was estimated to lack ~100 kb (Selkov *et al.*, 2000). This sequence was used for an analysis of the amino acid metabolism in *A. ferrooxidans*, which allowed an almost complete reconstruction of its metabolic pathways, leaving just 10 unassigned (missing) enzymes. Despite the initial intent of the authors to demonstrate that 'gapped' microbial genomes were almost as good as complete ones (Selkov *et al.*, 2000), this paper actually succeeded in proving the opposite: a meaningful and unequivocal analysis is only possible with a complete genome sequence. Furthermore, only small pieces of the genome have been submitted to the GenBank, which prevented others from analysing this genome.

Shortly after that, sequencing of *A. ferrooxidans* genome has been undertaken at the Institute of Genomic Research (TIGR). The resulting incomplete genome sequence of 3081 kb was made publicly available in 2001 as RefSeq entry NC_002923 and was subsequently used for a variety of genome analyses (e.g. Valdés *et al.*, 2003; Quatrini *et al.*, 2007). Over the next two years, this sequence was updated more than a dozen times and was finally withdrawn at the end of 2003. Since 2006, a complete genome sequence of 2982 kb coding for 3217 predicted proteins has been available on the TIGR website but was not submitted to GenBank. Finally, a recent joint paper by Chilean and TIGR scientists (Valdés *et al.*, 2008) reported a detailed analysis of this genome and its availability to the public.

Meanwhile, JGI scientists have released a 2885 kb genome sequence of another strain of *A. ferrooxidans*, which encodes 2826 proteins. This strain *A. ferrooxidans* ATCC 53993 was isolated from mine water of the Alaverda copper deposit in Armenia and initially assigned the name *Leptospirillum ferrooxidans* (Balashova *et al.*, 1974; Hippe, 2000). Although its relation to the type strain ATCC 23270 is not known at this time, their 16S rRNA sequences are 100% identical. Thus, after 10 years of struggling with unfinished genome sequences, the public now has access to two complete genomes of *A. ferrooxidans*. This should allow further analyses of the properties of this remarkable organism and stimulate its use in energy research and bio-mining.

The list of completely sequenced spirochaete genomes has grown to include genomes of *Borrelia duttonii* and *Borrelia recurrentis* (Lescot *et al.*, 2008). Both organisms are important human pathogens causing relapsing fevers. The first one is transmitted by the tick *Ornithodoros moubata* and is found primarily in east Africa. *Borrelia recurrentis* is transmitted by human body lice *Pediculus humanus* and is found in around the world. The sequenced strain *B. duttonii* Ly was isolated from a 2-year-old girl with tick-borne relapsing fever in Tanzania, whereas *B. recurrentis* strain A1 was isolated from an adult patient with louse-borne relapsing fever in Ethiopia.

Klebsiella pneumoniae ssp. *pneumoniae* is a well-known human pathogen, and the first genome of its clinical isolate MGH 78578 was sequenced more than two years ago. A very interesting paper from the JCVI scientists now reports the genome sequence of an environmental N₂-fixing strain of *K. pneumoniae* (Fouts *et al.*, 2008). Such strains are commonly found as endophytes that colonize tissues of rice, maize, sugarcane, banana and various grasses and improve the growth of the host plants by supply them with ammonia. The sequenced strain *K. pneumoniae* 342 was isolated from maize and later shown to colonize wheat and alfalfa sprouts. Comparative analysis of the two strains pro-

vides interesting clues to the adaptation to the endophytic lifestyle, as well as into the evolution of pathogenicity in *K. pneumoniae*.

The list of organisms with recently sequenced genomes also includes the marine γ -proteobacteria *Alteromonas macleodii* and *Vibrio fischeri*, δ -proteobacteria *Anaeromyxobacter* sp. and *Geobacter bemedjiensis*, five new strains of *Salmonella enterica* ssp. *enterica* that include four new serovars (Thomson *et al.*, 2008), *Streptococcus equi* ssp. *zooepidemicus* (also known as *Streptococcus zooepidemicus*), the cause of an acute nephritis acute epidemic in Brazil (Beres *et al.*, 2008), and the well-studied *Helicobacter pylori* strain G27 (Table 1).

Acknowledgements

M.Y.G. is supported by the Intramural Research Program of the NIH, National Library of Medicine. The author's opinions do not reflect the views of NCBI, NLM or the National Institutes of Health.

References

- Balashova, V.V., Vedenina, I.Y., Markosyan, G.E., and Zavarzin, G.A. (1974) The autotrophic growth of *Leptospirillum ferrooxidans*. *Mikrobiologiya* **43**: 491–494.
- Beres, S.B., Sesso, R., Pinto, S.W., Hoe, N.P., Porcella, S.F., Deleo, F.R., and Musser, J.M. (2008) Genome sequence of a Lancefield group C *Streptococcus zooepidemicus* strain causing epidemic nephritis: new information about an old disease. *PLoS ONE* **3**: e3026.
- Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Perte, M., Silva, J.C., *et al.* (2002) Genome sequence and comparative analysis of the model rodent malaria parasite. *Plasmodium yoelii yoelii*. *Nature* **419**: 512–519.
- Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., *et al.* (2008) Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* **455**: 757–763.
- Cavalier-Smith, T. (2002) The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int J Syst Evol Microbiol* **52**: 7–76.
- Cavalier-Smith, T. (2006) Rooting the tree of life by transition analyses. *Biol Direct* **1**: 19.
- Cho, J.C., Vergin, K.L., Morris, R.M., and Giovannoni, S.J. (2004) *Lentisphaera araneosa* gen. nov., sp. nov., a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, *Lentisphaerae*. *Environ Microbiol* **6**: 611–621.
- Ferguson, S.J., and Ingledew, W.J. (2008) Energetic problems faced by micro-organisms growing or surviving on parsimonious energy sources and at acidic pH. I. *Acidithiobacillus ferrooxidans* as a paradigm. *Biochim Biophys Acta* (in press). 10.1016/j.bbabi.2008.08.012
- Fouts, D.E., Tyler, H.L., DeBoy, R.T., Daugherty, S., Ren, Q., Badger, J.H., *et al.* (2008) Complete genome sequence of the N₂-fixing broad host range endophyte *Klebsiella pneu-*

- moniae* 342 and virulence predictions verified in mice. *PLoS Genet* **4**: e1000141.
- Galperin, M.Y. (2006) Sampling of microbial diversity by complete genomes. *Environ Microbiol* **8**: 1313–1317.
- Galperin, M.Y. (2008) The quest for biofuels fuels genome sequencing. *Environ Microbiol* **10**: 2471–2475.
- Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**: 498–511.
- Garrity, G.M., Bell, J.A., and Lilburn, T.G. (2004) Taxonomic outline of the *Prokaryotes*. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, Vol. 2: *The Proteobacteria*. Brenner, D.J., Krieg, N.R., and Staley, J.T. (eds). New York, USA: Springer.
- Gupta, R.S. (1998) Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaeobacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* **62**: 1435–1491.
- Gupta, R.S. (2000) The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol Rev* **24**: 367–402.
- Henry, E.A., Devereux, R., Maki, J.S., Gilmour, C.C., Woese, C.R., Mandelco, L., *et al.* (1994) Characterization of a new thermophilic sulfate-reducing bacterium *Thermodesulfobacterium yellowstonii*, gen. nov. and sp. nov.: its phylogenetic relationship to *Thermodesulfobacterium commune* and their origins deep within the bacterial domain. *Arch Microbiol* **161**: 62–69.
- Hippe, H. (2000) *Leptospirillum* gen. nov. (ex Markosyan 1972), nom. rev., including *Leptospirillum ferrooxidans* sp. nov. (ex Markosyan 1972), nom. rev. and *Leptospirillum thermoferrooxidans* sp. nov. (Golovacheva *et al.* 1992). *Int J Syst Evol Microbiol* **50**: 501–503.
- Hongoh, Y., Sato, T., Dolan, M.F., Noda, S., Ui, S., Kudo, T., and Ohkuma, M. (2007) The motility symbiont of the termite gut flagellate *Caduceia versatilis* is a member of the 'Synergistes' group. *Appl Environ Microbiol* **73**: 6270–6276.
- Horinouchi, S., Fukusumi, S., Ohshima, T., and Beppu, T. (1988) Cloning and expression in *Escherichia coli* of two additional amylase genes of a strictly anaerobic thermophile, *Dictyoglomus thermophilum*, and their nucleotide sequences with extremely low guanine-plus-cytosine contents. *Eur J Biochem* **176**: 243–253.
- Hou, S., Makarova, K.S., Saw, J.H.W., Senin, P., Ly, B.V., Zhou, Z., *et al.* (2008) Complete genome sequence of the extremely acidophilic methanotroph isolate V4, '*Methylophilum inferorum*', a representative of the bacterial phylum *Verrucomicrobia*. *Biol Direct* **3**: 26.
- Jumas-Bilak, E., Carlier, J.P., Jean-Pierre, H., Citron, D., Bernard, K., Damay, A., *et al.* (2007) *Jonquetella anthropi* gen. nov., sp. nov., the first member of the candidate phylum 'Synergistetes' isolated from man. *Int J Syst Evol Microbiol* **57**: 2743–2748.
- Kelly, D.P., and Wood, A.P. (2000) Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen. nov., *Halothiobacillus* gen. nov. and *Thermithiobacillus* gen. nov. *Int J Syst Evol Microbiol* **50**: 511–516.
- Lescot, M., Audic, S., Robert, C., Nguyen, T.T., Blanc, G., Cutler, S.J., *et al.* (2008) The genome of *Borrelia recurrentis*, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne *Borrelia duttonii*. *PLoS Genet* **4**: e1000185.
- Ludwig, W., Schleifer, K.-H., and Whitman, W.B. (2008) Revised road map to the phylum *Firmicutes*. In: *Bergey's Manual of Systematic Bacteriology*, 2nd edn, Vol. 3: *The Firmicutes*. De Vos, P., Garrity, G., Jones, D., Krieg, N.R., Ludwig, W., Rainey, F.A., *et al.* (eds). New York, USA: Springer.
- Luo, Y., Xu, X., Ding, Z., Liu, Z., Zhang, B., Yan, Z., *et al.* (2008) Complete genome of *Phenylobacterium zucineum* – a novel facultative intracellular bacterium isolated from human erythroleukemia cell line K562. *BMC Genomics* **9**: 386.
- Marcy, Y., Ouverney, C., Bik, E.M., Losekann, T., Ivanova, N., Martin, H.G., *et al.* (2007) Dissecting biological 'dark matter' with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc Natl Acad Sci USA* **104**: 11889–11894.
- Mori, K., Kim, H., Kakegawa, T., and Hanada, S. (2003) A novel lineage of sulfate-reducing microorganisms: *Thermodesulfobiaceae* fam. nov., *Thermodesulfobium narugense*, gen. nov., sp. nov., a new thermophilic isolate from a hot spring. *Extremophiles* **7**: 283–290.
- Ollivier, B.M., Mah, R.A., Ferguson, T.J., Boone, D.R., Garcia, J.L., and Robinson, R. (1985) Emendation of the genus *Thermobacteroides*: *Thermobacteroides proteolyticus* sp. nov., a proteolytic acetogen from a methanogenic enrichment. *Int J Syst Bacteriol* **35**: 425–428.
- Pain, A., Bohme, U., Berry, A.E., Mungall, K., Finn, R.D., Jackson, A.P., *et al.* (2008) The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* **455**: 799–803.
- Quatrini, R., Lefimil, C., Veloso, F.A., Pedroso, I., Holmes, D.S., and Jedlicki, E. (2007) Bioinformatic prediction and experimental verification of Fur-regulated genes in the extreme acidophile *Acidithiobacillus ferrooxidans*. *Nucleic Acids Res* **35**: 2153–2166.
- Rainey, F.A., Stackebrandt, E. (1993) Transfer of the type species of the genus *Thermobacteroides* to the genus *Thermoanaerobacter* as *Thermoanaerobacter acetothyllicus* (Ben-Bassat and Zeikus 1981) comb. nov., description of *Coprothermobacter* gen. nov. and reclassification of *Thermobacteroides proteolyticus* as *Coprothermobacter proteolyticus* (Ollivier *et al.* 1985) comb. nov. *Int J Syst Bacteriol* **43**: 857–859.
- Saiki, T., Kobayashi, Y., Kawagoe, K., and Beppu, T. (1985) *Dictyoglomus thermophilum* gen. nov., sp. nov., a chemoorganotrophic, anaerobic, thermophilic bacterium. *Int J Syst Bacteriol* **35**: 253–259.
- Selkov, E., Overbeek, R., Kogan, Y., Chu, L., Vonstein, V., Holmes, D. *et al.* (2000) Functional analysis of gapped microbial genomes: amino acid metabolism of *Thiobacillus ferrooxidans*. *Proc Natl Acad Sci USA* **97**: 3509–3514.
- Thomson, N.R., Clayton, D.J., Windhorst, D., Vernikos, G., Davidson, S., Churcher, C., *et al.* (2008) Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res* **18**: 1624–1637.

- Valdés, J., Veloso, F., Jedlicki, E., and Holmes, D. (2003) Metabolic reconstruction of sulfur assimilation in the extremophile *Acidithiobacillus ferrooxidans* based on genome analysis. *BMC Genomics* **4**: 51.
- Valdés, J., Pedroso, I., Quatrini, R., Tettelin, H., Blake, R.A., Eisen, J.A., and Holmes, D.S. (2008) *Acidithiobacillus ferrooxidans* metabolism: from genome sequence to industrial applications. *BMC Biol* **6**, in press.
- Wagner, M., and Horn, M. (2006) The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr Opin Biotechnol* **17**: 241–249.
- Winzeler, E.A. (2008) Malaria research in the post-genomic era. *Nature* **455**: 751–756.
- Yarza, P., Richter, M., Peplies, J., Euzéby, J., Amann, R., Schleifer, K.-H., *et al.* (2008) The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* **31**: 241–250.
- Zhang, K., Han, W., Zhang, R., Xu, X., Pan, Q., and Hu, X. (2007) *Phenylobacterium zucineum* sp. nov., a facultative intracellular bacterium isolated from a human erythroleukemia cell line K562. *Syst Appl Microbiol* **30**: 207–212.