

Genomics update

The vibrio that sheds light

Michael Y. Galperin*

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

Among the relatively few microbial genomes completed in the beginning of 2005 (Table 1), one environmental organism clearly stands out. *Vibrio fischeri* is a marine gamma-proteobacterium and the first non-pathogenic representative of *Vibrio* spp. to have its genome sequenced. Owing to their health risk and financing priorities, the three *Vibrio* spp. sequenced previously, *V. cholerae*, *V. parahaemolyticus* and *V. vulnificus* (two sequenced strains), were all facultative human pathogens. These species, too, are common in marine environment. *Vibrio cholerae*, for example, has been isolated from planktonic samples as far north as Chesapeake Bay in Maryland, USA, and as far south as coastal waters of Peru and Argentina (Louis *et al.*, 2003; Gil *et al.*, 2004, and see Colwell, 2004 for a recent review). Still, *V. fischeri* is the first vibrio whose genome has been sequenced solely for its environmental importance (or, rather, aesthetic value). Quite appropriately, the sequencing project was supported not by US government money but by private funding, a grant from the W. M. Keck Foundation.

The sequenced strain inhabits the light organ of the Hawaiian nocturnal bobtail squid, *Euprymna scolopes*, where it is actually responsible for light emission. The mechanism of bioluminescence, produced by bacterial luciferase, is relatively well studied and the *lux* operon that encodes this enzyme is widely used in various biosensors. Bioluminescence is believed to be used by squid as an anti-predation measure, explained in a recent review in the following way: 'During its nocturnal feeding period, the squid emits light downward and modulates it to match the intensity of moonlight, thus preventing the formation of a tell-tale shadow on the ocean floor below' (Visick and McFall-Ngai, 2000). One might find such an explanation a bit too romantic for a squid, whose other, equally famous, anti-predation measure is releasing a blob of black ink that fools and disorients the potential predator.

In any case, there is no doubt that *V. fischeri* and its host both benefit from this relationship, which is therefore considered mutualistic symbiosis.

Analysis of the *V. fischeri* genome revealed its similarity to those of previously sequenced vibrios. It also consists of two chromosomes and, as in *V. cholerae*, the smaller chromosome carries a higher fraction of unique genes. *Vibrio fischeri* encodes many adhesive factors that help in its attachment to the epithelial cells of squid. Surprisingly, many of these adhesins (type IV pili, mannose-sensitive haemagglutinin, toxin-coregulated pili) are known virulence factors in *V. cholerae*. It even encodes homologues of several *V. cholerae* proteins that are presumed to have toxin activity. These observations show what a thin line separates *V. fischeri* from its pathogenic relatives and shed some light on the origin of pathogenesis among the Vibrionaceae.

The publication of the *V. fischeri* genome was followed by the description of the genome of another member of Vibrionaceae, the deep-sea bacterium *Photobacterium profundum* strain SS9 (Vezzi *et al.*, 2005), whose genome sequence had been deposited in the GenBank/EMBL/DBJ database a year earlier and already mentioned in this column. Strain SS9 has been originally isolated at the depth of 2500 m and is a model organism for studying adaptations to high-pressure environment. Despite its name, this bacterium does not encode luciferase and is not bioluminescent. Still, many features of its genome are similar to those of *V. fischeri* and other marine vibrios. It also consists of two chromosomes, the larger one that carries most of housekeeping genes and the smaller one carrying a larger fraction of unique genes. Early experiments on SS9 showed pressure-dependent expression of outer membrane proteins. Now, the authors were able to compare transcriptional profiles of SS9 cells grown under high and low pressure using a whole-genome microarray. Growth at high pressure activated a number of genes, including those coding for chitin, pullulan and cellulose degradation pathways. High pressure also activated expression of the selenoprotein-containing glycine reductase complex that catalyses reductive deamination of glycine, betaine or sarcosine and had previously been seen only in obligate anaerobes. While the authors interpret this as an evidence of anaerobic metabolism under high pres-

*For correspondence. E-mail galperin@ncbi.nlm.nih.gov; Tel. (+1) 301 435 5910; Fax (+1) 301 435 7794.

Table 1. Recently completed microbial genomes (February–March 2005).

Species name	Taxonomy	GenBank accession	Genome size (bp)	Proteins ^a (% known)	Sequencing centre	Reference
New genomes <i>Entamoeba histolytica</i>	Eukaryota, Entamoeba	AAFB00000000	23 751 783	9938 (68%)	The Institute for Genomic Research, Rockville, MD, USA, and the Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK	Loftus <i>et al.</i> (2005)
<i>Vibrio fischeri</i>	γ-Proteobacteria	CP000020 CP000021 CP000022	2 906 179 1 332 022 45 849	3802 (75%)	University of Hawaii, Honolulu, HI, University of Iowa, Iowa City, IA, and Integrated Genomics, Chicago, IL, USA	Ruby <i>et al.</i> (2005)
<i>Neisseria gonorrhoeae</i>	β-Proteobacteria	AE004969	2 153 922	2002 (77%)	University of Oklahoma, Oklahoma City, OK, USA	Lewis <i>et al.</i> (unpubl.)
<i>Wolbachia</i> endosymbiont of <i>Brugia malayi</i>	α-Proteobacteria	AE017321	1 080 084	805 (79%)	New England Biolabs, Beverly, MA, and Integrated Genomics, Chicago, IL, USA	Foster <i>et al.</i> (2005)
New strains <i>Bacteroides fragilis</i> NCTC 9343	Bacteroidetes	CR626927	5 205 140	4236 (64%)	The Sanger Institute, Hinxton, Cambridge, UK	Cerdeno-Tarraga <i>et al.</i> (2005)

a. Encoded in all chromosomes and megaplasmids; includes pseudogenes.

sure, it seems more likely that in this case glycine reductase actually works in reverse direction to produce glycine and betaine, well-known osmoprotectors. This could mean that *P. profundum* has adopted to its needs an unusual system for generating compatible solutes using genes laterally transferred from its neighbours.

Remarkably, atmospheric pressure was perceived by *P. profundum* as a stress condition, leading to activation of stress-response genes, such as *htpG*, *dnaK*, *dnaJ*, *groEL*, and genes involved in DNA repair. These observations confirmed that *P. profundum* is a true piezophile that needs increased expression of chaperones to correctly fold its proteins at low pressure.

Another bacterial genome that has been released long before publication is that of *Neisseria gonorrhoeae*, the causative agent of infamous sexually transmitted disease. In this case, the genome sequence has been available on the web for many years (Ferretti *et al.*, 1997) but only recently has been submitted to GenBank in its final form. Genome comparisons of *N. gonorrhoeae* and *Neisseria meningitidis* could help in delineating the determinants of tissue tropism of these common pathogens. This organism's genome will be discussed in more detail when its description is published.

In February, researchers from TIGR and the Sanger Institute reported results of a joint project on sequencing the genome of the protozoan *Entamoeba histolytica* (Loftus *et al.*, 2005). This organism is an intestinal parasite of humans and the causative agent of amoebiasis, a serious

health risk in developing countries. As an anaerobic amitochondrial eukaryote, it is also extremely interesting from the evolutionary point of view. The genome of *E. histolytica* is not yet complete, but the released set of 1819 contigs (the paper says 888 scaffolds) represents 23 752 kb with 12.5-fold coverage, which was sufficient for prediction of encoded proteins and analysis of metabolic pathways. Coding DNA comprises almost a half of *E. histolytica* genome, and only a quarter of genes appear to contain introns. A significant fraction of genes have closest homologues in bacteria and could have been acquired by *E. histolytica* through lateral gene transfer. This organism has a typical bacterial-like fermentative metabolism and lacks enzymes of tricarboxylic acid cycle. However, it encodes flavoproteins and rubrerythrin that could be involved in oxygen detoxification. Similarly to other amitochondrial eukaryotes, *Giardia lamblia* and *Trichomonas vaginalis*, *E. histolytica* lacks biosynthetic pathways for most amino acids (except for Ser and Cys), purines and pyrimidines, although it encodes certain salvage pathways. A surprising discovery in the genome sequence was the presence of as much as 270 Ser/Thr/Tyr-type protein kinases and more than 100 protein phosphatases. Many of these protein kinases have predicted extracellular domains and may serve as environmental receptors. Although *E. histolytica* apparently lacks the endoplasmic reticulum and the Golgi complex, its genome encodes certain components of the vesicle transport machinery. This and other observations make *E. histolytica* genome a valuable

tool for understanding the organization of a eukaryotic cell.

Bacteroides fragilis NCTC 9343 is the second strain of that species and fifth representative of the Bacteroidetes phylum (aka CFB group) with a sequenced genome. Although this obligately anaerobic bacterium is a normal component of human gut microflora, *B. fragilis* is isolated in greater numbers during intestinal inflammation, which attracts heightened attention to its highly variable surface structures. Genome sequencing revealed a large number of invertible regions, which significantly complicated assembly of the shotgun data. Some of these invertible regions contained a predicted promoter and were located upstream of gene clusters coding for biosynthesis of specific surface polysaccharides. These observations are similar to those made on the previously sequenced *B. fragilis* strain YCH46 (Kuwahara *et al.*, 2004) and suggest a likely mechanism for the antigenic variation in *B. fragilis*.

The recent sequencing of *Wolbachia* strain TRS, an endosymbiont of *Brugia malayi* (Foster *et al.*, 2005), offers another interesting example of mutualistic symbiosis. Filarial nematode *B. malayi* is a noxious human parasite that causes lymphatic filariasis, a terrible disease that affects millions of people in tropical countries. Still, there is place for mutualism even in this parasite's organism: its metabolism largely depends on an alpha-proteobacterial symbiont that belongs to the *Wolbachia* genus. *Wolbachia* symbionts inhabit oocytes and hypodermis of *B. malayi* and are required for worm fertility and survival. Treatment of filariasis patients with tetracycline and other antibacterial drugs has shown some promise, which prompted a detailed study of the *Wolbachia*'s role. Although related to intracellular parasites *Rickettsia* spp., *Wolbachia* has retained certain biosynthetic pathways and appears to provide its host with riboflavin, haem and nucleotides. In turn, the worm probably supplies the bacterium with amino acids. The *Wolbachia* symbiont of *B. malayi* has a smaller genome than the previously sequenced *Wolbachia pipiens* endosymbiont of *Drosophila melanogaster* (Wu *et al.* 2004), which might reflect a closer association with the host.

The *Wolbachia* story got an interesting twist when Salzberg *et al.* (2005) looked at the unfinished genome sequences for several *Drosophila* spp., available at the NCBI Trace Archive (<http://www.ncbi.nih.gov/Traces>). Suspecting that at least some *Drosophila* sequences might contain *Wolbachia* DNA, the authors used *Wolbachia* genome in a BLAST search to identify such bacterial sequences. Their expectation proved to be correct and yielded *Wolbachia*-like sequences from three different species of *Drosophila*, *D. ananassae* and *D. mojavensis* that are being sequenced by Agencourt Bioscience and *D. simulans*, which is being sequenced at the Washington University Genome Sequencing Center. The amount of

Wolbachia-like sequence was so large that it allowed Salzberg and colleagues to assemble a nearly complete genome of one of these organisms and compile significant chunks of two others. In its current form, the genome of *Wolbachia* endosymbiont of *D. ananassae* (GenBank accession No. AAGB00000000) contains 1 440 650 bp in 329 separate scaffolds, at approximately eightfold coverage. The authors estimate that this assembly covers up to 98% of the whole genome. Two other assemblies, *Wolbachia* endosymbiont of *D. simulans* and *Wolbachia* endosymbiont of *D. mojavensis*, have also been deposited in GenBank with Accession numbers AAGC00000000 and AY897435–AY897548 respectively. This work certainly deserves an honourable mention for the most creative use of someone else's data.

One final comment: there is certain irony in the publication timing of the *P. profundum* and *B. fragilis* genome papers. The former one was submitted to *Science* magazine in July 2004, 3 months after the *P. profundum* genome sequence was sent to the EMBL Nucleotide Database. It took more than 5 months for the paper to be accepted by *Science* and 2 months more for it to be published. All information on this paper has been embargoed until 4 March 2005, i.e. for almost a year after the release of the genome sequence. The *V. fischeri* genome paper, submitted to PNAS much later, was published when the *P. profundum* paper was still under wraps. In the case of two *B. fragilis* genome papers, the delay in publication was even more startling. The *B. fragilis* NCTC 9343 genome paper was submitted to *Science* in August 2004, just 2 months after Kuwahara and colleagues submitted their manuscript on *B. fragilis* YCH46 genome (Kuwahara *et al.*, 2004) to PNAS. However, the Japanese paper got approved in 2 months and was published, both online and in print, in October 2004. As an Open Access article, it immediately became available to the entire community. In contrast, due to the much longer editorial process in *Science*, a longer publication schedule, and the standard journal policy of embargoing the news on papers in print, there has been no information on the *B. fragilis* NCTC 9343 paper until March 2005, long after the paper by Kuwahara and colleagues has become part of public record. This certainly does not mean that one should never send time-sensitive papers to *Science*. However, the common notion that really important papers should be published only in a handful of highly prestigious (but slow and possessive) journals has probably outlived itself. The list of journals that have published genomics papers clearly shows that if the primary goal of a publication is to report the results to the scientific community, there are ample opportunities to bring good papers into the public eye without wasting so much time. This is, of course, just a personal opinion of the author, not necessarily shared by his employer or any of his supervisors.

References

- Cerdeno-Tarraga, A.M., Patrick, S., Crossman, L.C., Blakely, G., Abratt, V., Lennard, N., *et al.* (2005) Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* **307**: 1463–1465.
- Colwell, R.R. (2004) Infectious disease and environment: cholera as a paradigm for waterborne disease. *Int Microbiol* **7**: 285–289.
- Ferretti, J.J., Dyer, D.W., and Roe, B.A. (1997) Data available. *Nature* **386**: 320.
- Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., *et al.* (2005) The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* **3**: e121.
- Gil, A.I., Louis, V.R., Rivera, I.N., Lipp, E., Huq, A., Lanata, C.F., *et al.* (2004) Occurrence and distribution of *Vibrio cholerae* in the coastal environment of Peru. *Environ Microbiol* **6**: 699–706.
- Kuwahara, T., Yamashita, A., Hirakawa, H., Nakayama, H., Toh, H., Okada, N., *et al.* (2004) Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. *Proc Natl Acad Sci USA* **101**: 14919–14924.
- Loftus, B., Anderson, I., Davies, R., Alsmark, U.C., Samuelson, J., Amedeo, P., *et al.* (2005) The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**: 865–868.
- Louis, V.R., Russek-Cohen, E., Choopun, N., Rivera, I.N., Gangle, B., Jiang, S.C., *et al.* (2003) Predictability of *Vibrio cholerae* in Chesapeake Bay. *Appl Environ Microbiol* **69**: 2773–2785.
- Ruby, E.G., Urbanowski, M., Campbell, J., Dunn, A., Faini, M., Gunsalus, R., *et al.* (2005) Complete genome sequence of *Vibrio fischeri*: a symbiotic bacterium with pathogenic congeners. *Proc Natl Acad Sci USA* **102**: 3004–3009.
- Salzberg, S.L., Hotopp, J.C., Delcher, A.L., Pop, M., Smith, D.R., Eisen, M.B., and Nelson, W.C. (2005) Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol* **6**: R23.
- Vezi, A., Campanaro, S., D'Angelo, M., Simonato, F., Vitulo, N., Lauro, F.M., *et al.* (2005) Life at depth: *Photobacterium profundum* genome sequence and expression analysis. *Science* **307**: 1459–1461.
- Visick, K.L., and McFall-Ngai, M.J. (2000) An exclusive contract: specificity in the *Vibrio fischeri*-*Euprymna scolopes* partnership. *J Bacteriol* **182**: 1779–1787.
- Wu, M., Sun, L.V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J.C., *et al.* (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* **2**: E69.