

A bug with excess gastric avidity

Russell F. Doolittle

The 1,667,867-base-pair genome of the bacterium that is responsible for peptic ulcers has been completely sequenced. Among the many features revealed is machinery for existence in an acidic environment.

Yet another completely sequenced bacterial genome! The 1,667,867 base pairs of *Helicobacter pylori*, reported by Tomb *et al.*¹ on page 539 of this issue, constitute the seventh completely sequenced bacterial genome (Table 1), and the fourth by The Institute for Genome Research (TIGR). With another 30 such projects likely to be completed in various labs around the world during the next two years, some readers may be wondering when enough is enough. Is this new sequence big news? The answer has to be a resounding “yes” — *Helicobacter pylori* (a.k.a. *Campylobacter pylori*) is the organism that causes peptic ulcers and, astonishingly, it may infect as much as half of the world's population.

In 1983 an Australian physician, Robin Warren, and his colleague Barry Marshall, a biochemist, wrote separate letters to *The Lancet* under a single title². Warren described how, over the previous five years or so, he had observed spiral-shaped bacteria on the gastric epithelium of patients with peptic ulcers. Marshall then discussed the history of pyloric bacteria in mammals, and offered some thoughts on why these *Campylobacter*-like organisms had not been reported previously. He noted that they had been found in non-human mammals, and that they were even observed long ago in human cadavers, but their existence had been rationalized as a post-mortem consequence.

In 1984, the two wrote an article together in *The Lancet*³ offering the hypothesis that these bacteria were actually the cause of peptic ulcers, a very common malady attributed to stress, diet and excess gastric acidity. At about the same time, a large epidemiological study of peptic ulcers was published in the *Medical Journal of Australia*⁴. The authors ascribed the varying incidence of peptic ulcers in different Australian states to environmental factors, such as contaminated drinking water. An accompanying editorial⁵ stated that “the agent is probably not a bacterium”. Warren responded immediately, stating that the agent probably was a bacterium⁶.

Interestingly, the epidemiological study was based entirely on records of state-mandated prescriptions for the antihistamine drug cimetidine (better known by its brand name Tagamet) which, according to news-

paper accounts, was the most prescribed drug in the world during the 1980s. Imagine the impact of a finding that penicillin or other antibacterial agents might be a more appropriate treatment. Henrik Ibsen would have loved it!

Bemused by the situation, I searched Medline for reports involving *H.* or *C. pylori*, as well as for cimetidine and, for reasons that will become apparent, for the enzyme urease. These were plotted against five-year time periods, and the graph (Fig. 1) speaks for itself — papers dealing with *H. pylori* and the enzyme urease have greatly increased in number, whereas cimetidine citations are in precipitous decline. But why is urease suddenly so popular? The answer is that this enzyme is crucial to the survival of *H. pylori* in the very acid pH of the stomach. (Incidentally, urease is also a sentimental favourite among biochemists because, in 1926, it was the first enzyme ever to be crystallized.)

It has long been known that resident microorganisms in the stomach use the enzyme urease to convert urea to ammonia and carbon dioxide⁷. Few other organisms can survive in this acidic environment, but *H. pylori* has an electropositive internal milieu which helps it to fend off the onslaught of protons in the surrounding medium. Tomb *et al.*¹ have now shown that the proteins in *H. pylori* contain twice as many of the basic amino acids arginine and lysine as proteins in other organisms. Beyond that, the positively charged ammonium ions must contribute greatly to this effect.

Urease is the main antigenic activity associated with *H. pylori* and it is a convenient diagnostic tool⁸, which is why there has been

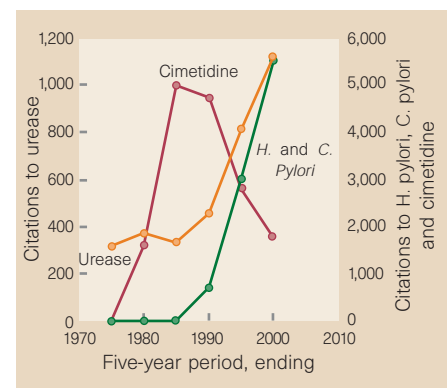


Figure 1 Citations to *Helicobacter* (or *Campylobacter*) *pylori*, cimetidine (Tagamet) and urease plotted in five-year increments. Note the abrupt changes in the period following Marshall and Warren's report³ in 1984, that *H. pylori* probably causes peptic ulcers (data for the year 2000 extrapolated from 1997).

such an enormous increase in citations involving urease. Many of the previous observations are now put into a proper context by having the entire genomic sequence. For example, all indications are that the ‘pathogenicity island’ is the main virulence factor. This 40-kilobase segment contains a set of accessory factors for secretion of the cytotoxins that damage the host gut⁹. The region is flanked by insertion elements and is probably the result of horizontal gene transfer from some other organism.

The peptic ulcer–urease story is, however, hardly the only reason that the sequence reported by Tomb *et al.*¹ is news. This is a genome that has something for everyone — clinicians, sociologists, epidemiologists, biochemists, ecologists, molecular biologists, immunologists and, last mentioned but hardly least interested, evolutionary biologists. Tomb *et al.* reveal the entire restriction-modification system for recognizing and destroying foreign DNA, and outline a complete scheme of metabolism on the basis of the resident genes. They have also worked out how *H. pylori* mimics blood-group antigens, as well as a likely molecular mechanism for encouraging immunogenic variation.

Of the many details reported, the generalizations about molecular evolution are

Table 1 Fully sequenced genomes

Organism	Genome size (kbp)	ORFs*	URFst	Year reported
Prokaryotes				
<i>Haemophilus influenzae</i>	1,830	1,743	732 (42%)	1995
<i>Mycoplasma genitalium</i>	580	470	187 (39%)	1995
<i>Methanococcus jannaschii</i>	1,665	1,738	1,078 (62%)	1996
<i>Mycoplasma pneumoniae</i>	816	679	344 (46%)	1996
<i>Synechocystis</i> sp.	3,570	3,168	1,766 (50%)	1996
<i>Helicobacter pylori</i>	1,668	1,590	499 (31%)	1997
Eukaryote				
<i>Saccharomyces cerevisiae</i>	12,068	5,885	2,942 (50%)	1996

* ORF (open reading frame); potential for encoding a protein.
† URF (unidentified reading frame). The numbers are approximate, many having been identified since the initial publications.

the most interesting. *H. pylori* is a Gram-negative bacterium, so most of its protein sequences would be expected to resemble other Gram-negative bacteria such as *Escherichia coli* and *Haemophilus influenzae*. Most of them do, of course, but a considerable number have sequences that are most similar to other, more distantly related, bacteria. One enzyme that is involved in chorismate biosynthesis is even reported to be most closely related to an equivalent in chloroplasts.

Although the significance of these anomalies is not clear, the possibility of rampant horizontal gene-transfer is unnerving to a community that hopes to reconstruct the history of life on the basis of amino-acid sequence comparison¹⁰. To test the point, I examined a number of bacterial urease sequences including *H. pylori* (Fig. 2). I found it unsettling not only that the generic boundaries were so ill-defined, but that the plant (jackbean) sequence was almost as similar to each of the bacterial sequences (an average of 65 per cent identity) as most of them were to one another.

There are other frustrations about the data. As in the previous studies, many of the potential gene-encoding regions have not been identified with regard to likely

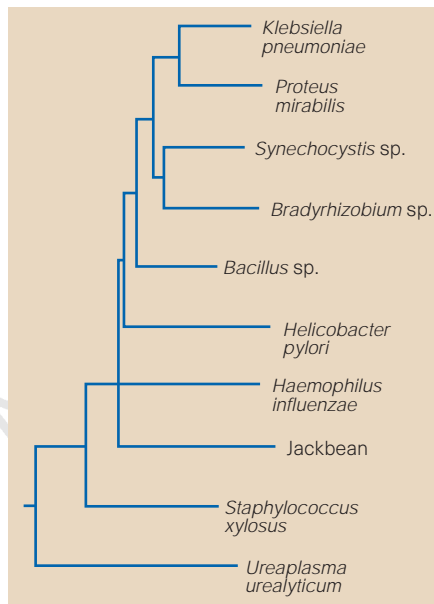


Figure 2 The genome of the Gram-negative bacterium *H. pylori*, sequenced by Tomb *et al.*¹, has thrown up a few surprises. One is that certain protein sequences do not most closely resemble those of other Gram-negative bacteria, and this may sound a warning for reconstruction of evolutionary relationships. The phylogenetic tree shown was constructed from a 480-residue, highly conserved region of the enzyme urease. On the average, the plant (jackbean) sequence is 65 per cent identical with the bacterial sequences used. The closest relationship shown is between *Klebsiella pneumoniae* and *Proteus mirabilis*, which are 77 per cent identical.

function. Given that the genes for all new proteins come from other genes by duplication, in whole or in part, followed by descent with modification, I am surprised that so many open reading frames remain as unidentified reading frames. More rigorous searching will doubtless establish more relationships, as has proved to be the case in the earlier genome reports.

Nonetheless, one cannot help but be impressed by the tremendous accomplishments of Tomb *et al.*¹. Not only have they generated the data at a breathtaking pace, but the analyses are both insightful and thorough. The organizational problems associated with searching 1,600 open reading frames against existing databases,

and then making proper judgements about the findings, are awesome. Bring on the next bacterial genome sequence — don't hold that TIGR! □

Russell F. Doolittle is at the Center for Molecular Genetics, University of California at San Diego, La Jolla, California 92093-0634, USA.

1. Tomb, J.-F. *et al.* *Nature* **388**, 539–547 (1997).
2. Warren, J. R. & Marshall, B. *Lancet* **1**, 1273–1275 (1983).
3. Marshall, B. & Warren, J. R. *Lancet* **1**, 1311–1315 (1984).
4. Hugh, T. B. *et al.* *Med. J. Aust.* **141**, 81–85 (1984).
5. Piper, D. W. *Med. J. Aust.* **141**, 76 (1984).
6. Warren, J. R. *Med. J. Aust.* **141**, 477–478 (1984).
7. Kornberg, H. H. & Davies, R. E. *Physiol. Rev.* **35**, 169–177 (1955).
8. Vaira, D. *et al.* *J. Clin. Pathol.* **41**, 355–356 (1988).
9. Censini, S. *et al.* *Proc. Natl Acad. Sci. USA* **93**, 14648–14653 (1996).
10. Doolittle, R. F. *et al.* *Science* **271**, 470–476 (1996).

Asteroids

Eros's extended family

Richard P. Binzel

Lost in space and far from home. That's the fate suffered by the thousands of small asteroids wandering through the inner Solar System near the Earth, cast adrift from the main asteroid belt between Mars and Jupiter by collisions and by Jupiter's gravity. All hope of discovering their heritage would seem to be lost due to the untraceable chaotic evolution of their orbits; and although the unique spectral signatures of some near-Earth asteroids have provided links to their asteroid-belt origins^{1–3}, most near-Earth wanderers have no distinctive genealogical traits. But a new approach by Zappalà *et al.*⁴ has given new insights to the pedigree of the two largest near-Earth asteroids, 433 Eros and 1036 Ganymed (20 km and 32 km across, respectively). They find that Eros and Ganymed may be siblings, originating from a single collision that formed an entire 'family' of small asteroids, most of which still reside in the main belt.

The placement work culminating in the new paper by Zappalà and colleagues was begun seven decades ago by the Japanese astronomer K. Hirayama, who noticed that main-belt asteroids are clustered in their orbital dimensions. Hirayama called these groupings 'families' because he believed their members shared a common origin, and the collisional disruption of asteroid 'parent bodies' is now recognized as the process responsible for creating families.

One family recognized by Hirayama^{5,6} was called Maria, named after the member with the lowest catalogue number (170 Maria). Orbiting 2.55 AU from the Sun, the Maria family has about 70 identified members, and is precariously poised on the edge of one of the major gaps in the asteroid belt, a 0.05-AU-wide band of unoccupied orbits centred at 2.50 AU (1 AU is the mean Earth–Sun distance). Jupiter probably creates this

gap: a circular heliocentric orbit at 2.50 AU has a period exactly one-third of Jupiter's, and any orbit near this 1:3 resonance undergoes chaotic changes in its eccentricity⁷ (becoming more elongated). Eventually, it will have a high enough eccentricity to cross the orbits of the inner planets, so a consequence of this chaotic orbital evolution is the potential for delivering asteroids and meteorites to the inner Solar System from the main asteroid belt⁸.

A critical step in linking Eros and Ganymed to the Maria family was reconstructing the velocity field of bodies emanating from the disrupted parent body. This problem was solved using numerical techniques to determine the orbital parameters of the parent body at the time of the break-up, developed earlier by Zappalà and colleagues⁹. From this velocity field and the size distribution of the recognized family members, the authors now estimate⁴ that on the order of ten objects in the 15–30-km range should have been injected into the 1:3 resonance zone — objects about the size of Eros and Ganymed. It is not clear whether those two objects were the only ones to be successfully injected, or whether their brethren have already met their ultimate fate (collision with the Sun or a terrestrial planet, or gravitational ejection resulting from a close planetary encounter). No other asteroid families are candidates for the heritage of Eros and Ganymed. Zappalà *et al.* also compare the spectral properties of Eros and Ganymed with those of about a dozen members of the Maria family, and find agreeable matches.

Although the authors make an excellent set of plausibility arguments for the pedigree of Eros and Ganymed, a convincing case is extremely difficult. Injecting such large asteroids into a chaotic zone undoubtedly