Figure 1 **Pattern of discoveries for gas giant planets and brown dwarf stars found orbiting nearby stars similar to our Sun. In most cases, only the minimum mass of the companion is known, although this value is likely to be close to the true mass. The star HD168443 has now been found to have two massive companions (HD168443b and HD168443c) that blur the distinction between giant planets and brown dwarf stars. In particular, the mass of HD168443c falls into the observed gap (from about 15 M_J to 30 M_J) between gas giants and brown dwarfs. The radial-velocity technique preferentially detects massive, close-in planets, as indicated by the detection bias along the dashed line.**

HD168443 might be created. Could it have the same history as a multiple star system? These are thought to start with the collapse of a dense molecular cloud core to form a flattened protostellar disk, followed by fragmentation into multiple protostars. Our understanding of this process is still hazy, but it seems an unlikely way to give birth to a tight stellar system in which the mass of the primary star is 50 to 100 times more massive than its companions. This is because subsequent growth by gas accretion from the disk and an infalling envelope of gas would add mass preferentially to the smaller protostars, leading to more equal stellar masses[4]. Furthermore, fragmentation calculations do not favour the formation of low-mass protostars in stable, planet-like orbits around a single much more massive protostar. When multiple protostars form by fragmentation, they begin life in unstable configurations, which are then subject to gravitational scattering during close mutual encounters.

We are therefore left with trying to explain the formation of the HD168443 system by the process of planetary formation which unfolds well after the central protostar has gained the bulk of its mass. There are two competing mechanisms for the formation of gas giants in a protoplanetary disk: core accretion and disk instability. The currently favoured theory, core accretion, was developed to explain the formation of Jupiter and Saturn through the formation of

a rock and ice core, followed by the accretion of gas. Core accretion could account for the formation of planets with masses up to about 5 M_J (refs 5–7), but whether it could produce objects as massive as the companions of HD168443 within the lifetime[8] of a typical protoplanetary disk (a few million years) remains to be seen. Disk instability[9], on the other hand, is quite capable of rapidly forming massive protoplanets, as large as 17 M_J, with change to spare.

Theorists have their work cut out to explain the formation of HD168443's unexpected companions. The planet hunters, meanwhile, are discovering bizarre solar systems at an alarming rate. ∎

Alan P. Boss is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington DC 20015-1305, USA.
e-mail: boss@dtm.ciw.edu

1. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
2. Marcy, G. W. *et al. Astrophys. J.* (submitted). See http://exoplanets.org
3. Marcy, G. W., Butler, R. P., Vogt, S. S., Fischer, D. & Liu, M. C. *Astrophys. J.* **520**, 239–247 (1999).
4. Bate, M. R. *Mon. Not. R. Astron. Soc.* **314**, 33–53 (2000).
5. Bodenheimer, P., Hubickyj, O. & Lissauer, J. J. *Icarus* **143**, 2–14 (2000).
6. Kley, W. *Mon. Not. R. Astron. Soc.* **313**, L47–L51 (2000).
7. Nelson, R. P., Papaloizou, J. C. B., Masset, F. & Kley, W. *Mon. Not. R. Astron. Soc.* **318**, 18–36 (2000).
8. Briceño, C. *et al. Science* **291**, 93–96 (2001).
9. Boss, A. P. *Astrophys. J.* **536**, L101–L104 (2000).

Microbiology

Gastrogenomics

Jonathan A. Eisen

The genome of an *Escherichia coli* strain that is emerging as a severe threat to human health has been sequenced. Comparing it with that of a harmless strain suggests why some forms of this bacterium cause disease.

Humans and *Escherichia coli* normally live happily together: *E. coli* is a beneficial bacterium commonly found in the human gastrointestinal system. But it also exists in harmful forms. One of the most harmful of these, called O157:H7, was first linked to human disease in 1983 (ref. 1), when it was shown to have been the cause of two outbreaks of an unusual and severe gastrointestinal ailment in the United States the previous year. The number of documented human illnesses and deaths caused by O157:H7 strains has since increased steadily worldwide, and these strains are now considered to be both emerging pathogens and major threats to public health[2]. Studies of O157:H7 receive a boost from a paper on page 529 of this issue[3], in which Perna and colleagues describe and analyse the genome sequence of one strain of *E. coli* O157:H7.

Outbreaks of O157:H7 infections in humans have been traced primarily to infected cattle[4], which are the source of contamination of ground beef, milk and — indirectly, through fertilizer — many fruit and vegetable products. Fortunately, proper cooking or pasteurization can prevent human infection by contaminated food. But infections can also be transmitted by sewage-contaminated water and person-to-person contact.

Genetically, the pathogenicity (ability to cause damage) and virulence (degree of pathogenicity) of O157:H7 strains depend on several factors. For example, these strains possess genes encoding the so-called shiga toxin, as well as small, circular DNA molecules that encode 'virulence factors'. And
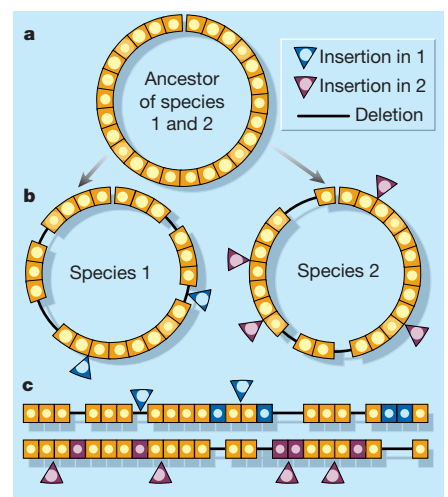

Figure 1 **Model of the generation of genome islands specific to just one of a pair of related strains or species. a, The ancestral circular genome. b, Modern genomes — the results of differential insertion and deletion of genomic segments in each lineage. c, Genomes 1 and 2 linearized and aligned. 'Islands' within each genome are highlighted on alignment (blue for species 1 and purple for species 2). Note that islands in one genome are generated both by insertion of genetic segments into that genome and deletion of segments from the other genome.**

O157:H7 has at least one pathogenicity island — a section of chromosomal DNA containing many genes that contribute to pathogenicity[5]. Evolutionary studies have shown that all O157:H7 strains are closely related, and that they share a common ancestry with pathogenic O55:H7 strains[6]

(the numbers and letters refer to the different types of variable antigen molecules produced by these bacteria). Many virulence factors in O157:H7 strains were acquired from other strains or species by a process known as horizontal gene transfer[5,7]. So, what can analysis of the whole genome add to our knowledge of this pathogen?

Perna *et al.*[3] have used the now standard 'whole-genome shotgun' approach to determine the genome sequence of the O157:H7 strain EDL933, which was isolated from ground beef linked to the 1982 outbreak. Their analysis of this genome, and in particular their comparison of this genome with that of a non-pathogenic laboratory *E. coli* strain, K-12 MG1655, is revealing.

The two genomes have a large amount of DNA — about 4.1 million base pairs (megabases) each — that was clearly derived from a common ancestor. This 'backbone' DNA is arranged similarly in the two strains: the two genomes can be lined up side by side along their lengths, except at one point, where part of the O157:H7 genome is reversed.

Although this conserved arrangement and inversion are not surprising for closely related strains[8], one feature is rather unusual. Scattered roughly evenly within each genome's backbone are hundreds of sections of DNA that are unique to one or the other strain. Sections found only in O157:H7 — 'O-islands' — total 1.34 megabases and 1,387 genes. K-islands, which are unique to the non-pathogenic *E. coli* strain, add up to 0.53 megabases and 528 genes. It remains to be seen which of these differences contribute to the virulence and pathogenicity of O157:H7. The O-islands include many known and predicted pathogenicity genes — for example, some of them may encode toxins, or factors needed to make the adhesive filaments (fimbriae) that help the bacterium to stick to the lining of the gut. But it is difficult to predict gene function accurately, and many of the O-islands might have no connection with pathogenicity.

Differences between the DNA backbones may also be important. Although most of the backbone differences do not result in changes in protein sequence, many do: about 75% of the backbone-encoded proteins differ by at least one amino acid between the two strains. A more thorough analysis of these patterns will help in determining which differences are the result of natural selection and which are merely neutral changes.

Interestingly, the patterns of variation within each genome differ between the coding and non-coding strands of backbone genes. Perna *et al.* suggest that this may result from transcription-coupled repair of oxidative damage in DNA. This process was originally discovered in *E. coli*: as the coding strand is copied into RNA (transcribed),

DNA damage in that strand is mended at a higher rate than normal[9]. Confirmation of whether this has caused the strand bias described here will require analysis of the genome sequence of another related species[10].

The authors also suggest that much of the DNA in the O-islands and K-islands was acquired by horizontal gene transfer. One of their lines of evidence is that many of the islands contain sequences related to those of bacterial viruses and other vectors that carry genes from one species to another. Another possibility is that the islands were present in the common ancestor of the two strains, and then lost in one lineage (Fig. 1).

Analyses of the genomes of related species will also help to answer this question. But if horizontal gene transfer has occurred, then the fact that so many genes have been transferred to O157:H7 supports the suggestion[11] that the continuing emergence of O157:H7 as a pathogen results from its ability to undergo rapid genetic change. This suggestion was made because a high proportion of O157:H7 strains have defects in genes involved in repairing DNA mismatches[11]. This tends to lead to higher rates of both mutation and acquisition of DNA from other strains. Many other pathogenic bacteria also have mismatch-repair defects[12], but so too do many non-pathogenic *E. coli* strains[13], and the existence of so many K-islands suggests that gene transfer is also common in non-pathogenic strains. Moreover, O157:H7 has an apparently normal long-term rate of sequence change[14].

Perna *et al.*'s work[3] emphasizes the power of comparing genomes from closely related strains or species — something that is becoming possible for more and more taxa. Such comparisons allow the detection and analysis of genetic processes that occur on relatively short timescales. They have led to discoveries such as the possible occurrence of transcription-coupled repair of oxidative damage, reported here[3], and the finding that inversions that are symmetrical around the start point of replication of a bacterial chromosome are common in bacterial genome evolution[8].

Many of these insights depend on knowing details such as gene location and orientation, and the absence of genes that are present in related species. This emphasizes the importance of having complete or nearly complete genome sequences. (The sequencing of the O157:H7 genome is nearly complete; only two gaps remain.) We should view with scepticism press releases announcing the completion of a genome sequence in a day[15] — they refer simply to the completion of the initial sequencing part of a project, but many gaps (frequently hundreds) always remain. Closing those gaps is difficult but essential.

There is still much to learn about the

biology and history of *E. coli* O157:H7, but the sequence and analysis presented by Perna *et al.*[3] will be an excellent starting point. One of the great things about genomics is that the data it provides allow nearly any group, anywhere in the world, to start seeking answers to outstanding questions immediately. ∎

Jonathan A. Eisen is at the Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA.
e-mail: jeisen@tigr.org

1. Riley, L. W. *et al. New Engl. J. Med.* **308,** 681–685 (1983).
2. Mead, P. S. *et al. Emerg. Infect. Dis.* **5,** 607–625 (1999).
3. Perna, N. T. *et al. Nature* **409,** 529–533 (2001).
4. Wells, J. G. *et al. J. Clin. Microbiol.* **29,** 985–989 (1991).
5. Reid, S. D., Herbelin, C. J., Bumbaugh, A. C., Selander, R. K. & Whittam, T. S. *Nature* **406,** 64–67 (2000).
6. Whittam, T. S. *et al. Infect. Immun.* **61,** 1619–1629 (1993).
7. Pupo, G. M., Karaolis, D. K., Lan, R. & Reeves, P. R. *Infect. Immun.* **65,** 2685–2692 (1997).
8. Eisen, J. A., Heidelberg, J. F., White, O. & Salzberg, S. L. *Genome Biol.* **1,** research0011.1-0011.9 (2000).
9. Mellon, I., Spivak, G. & Hanawalt, P. C. *Cell* **51,** 241–249 (1987).
10. Pilar-Francino, M., Chao, L., Riley, M. A. & Ochman, H. *Science* **272,** 107–109 (1996).
11. LeClerc, J. E., Li, B., Payne, W. L. & Cebula, T. A. *Science* **274,** 1208–1211 (1996).
12. Eisen, J. A. *Nucleic Acids Res.* **26,** 4291–4300 (1998).
13. Matic, I. *et al. Science* **277,** 1833–1834 (1997).
14. Whittam, T. S., Reid, S. D. & Selander, R. K. *Emerg. Infect. Dis.* **4,** 615–617 (1998).
15. http://news.bbc.co.uk/hi/english/sci/tech/newsid_743000/743329.stm

Carbon cycle

The age of river carbon

Wolfgang Ludwig

The organic carbon that runs into the oceans from rivers could be hundreds or thousands of years old. If so, aspects of our understanding of the global carbon cycle will have to change.

Dissolved organic carbon in the oceans is one of the biggest reservoirs in the global carbon cycle — it is comparable in size to all of the carbon in terrestrial plants, or to all of that in the form of CO_2 in the atmosphere. The input of terrestrial organic carbon from rivers, the main source of most constituents of sea water[1], could fill this marine reservoir in just a few thousands of years[2] — which, according to radiocarbon dating[3], is also about the average age of marine organic carbon. But although there should be lots of terrestrial-derived carbon in the ocean, geochemical studies indicate that there seems to be very little. So what happens to the riverine carbon once it enters the oceans?

On page 497 of this issue[4] Raymond and Bauer address the question by presenting new data on the average ages of organic matter in rivers. There may in fact be much more terrestrial carbon in the oceans than we thought, but we cannot see it. To understand this, one has to know how geochemists usually distinguish between organic carbon from terrestrial and marine sources.

Almost all of the organic carbon on Earth is created through photosynthesis, whether on land or in water. But on land the process leaves characteristic fingerprints, which mean that this carbon should be traceable after it has entered the oceans. Many land plants synthesize certain compounds, such as lignin or tannin, which are absent in marine phytoplankton. In principle, then, detecting these biomarkers in the sea can reveal if carbon had a terrestrial origin. The other widely applied method involves measuring the ratio between the two stable carbon isotopes, ^{13}C and ^{12}C, in the bulk organic matter. Most land plants produce carbon that is

Figure 1 **The Amazon delta, seen from space. Raymond and Bauer[4] find that, contrary to general belief, most of the organic carbon entering the Atlantic Ocean from the Amazon is not 'fresh' but has aged and been degraded.**

more depleted in the heavy carbon isotope (^{13}C) than carbon produced by marine phytoplankton, leading to higher isotopic ratios in marine than in terrestrial carbon.

When going into details, however, things can be more complicated. Biomarker concentrations in different plant materials are variable, reducing their potential for the accurate assessment of sources. And stable carbon-isotope ratios can sometimes give ambiguous results because there are exceptions to the general trend[5]. A broader problem arises when comparing relatively old organic material in the sea with freshly synthesized material of terrestrial and marine plants, because chemical and biological degradation may have altered the composition of the fresher material and make it difficult to trace its origin.

This is where Raymond and Bauer's study[4] comes in. They analysed organic matter in the Amazon (Fig. 1) and three North American rivers by radiocarbon dating, and

found that it was up to several thousand years old. This contrasts starkly with the general belief that most of the organic carbon in rivers should be relatively 'fresh'[6]. The particulate organic carbon (that is, the fraction retained on a filter) was especially old, the dissolved carbon being younger on average. But even the dissolved fraction revealed relatively old ages in some samples, and long-term laboratory experiments with samples from one of the rivers showed that the slow bacterial oxidation of part of this organic matter can markedly increase the average age of the carbon that remained in solution.

From these results Raymond and Bauer conclude that ageing and degradation of terrestrial organic matter in river basins and coastal zones may significantly alter its structure, distribution and quantity before it reaches the open oceans. The implication is that we simply have not been able to distinguish it from marine-generated carbon. Future studies need to test whether the old radiocarbon ages found by the authors can be confirmed in rivers from other parts of the world, however, and whether ageing can wipe out the fingerprints of terrestrial carbon. Furthermore, old average ages of riverine carbon can also be explained by the presence of small amounts of 'fossil' organic carbon[7] that originated in ancient sedimentary rocks. This phenomenon may be restricted to rivers draining these kind of rocks, and that has to be tested as well.

But it remains possible that dissolved organic carbon in the sea is mainly composed of old marine carbon, and that the terrestrial component is not hidden but largely absent. What if that is the case? A process would then be needed that rapidly removes the terrestrial carbon after it enters the oceans. It is generally accepted that part of this carbon is buried in the coastal zones where sedimentation rates are high. Biological activity is also high in these waters, and some carbon may be released as CO_2 to the atmosphere through biological oxidation. But the two processes do not seem to be powerful enough to remove all riverine carbon[5], and an additional mechanism for its rapid withdrawal from the sea would remain on the most-wanted list in the offices of many carbon-cycle researchers.

Nevertheless, at the least the results of Raymond and Bauer[4] remind us that the organic matter that runs from rivers into the sea is not necessarily identical to the organic matter of the plants and soils upstream in the river catchments. Certainly, other processes taking place on the floodplains or in the river channels, such as slowed transport caused by cycles of deposition and movement downstream, may also have to be considered to better understand the composition and fluxes of river carbon[8]. ∎

Wolfgang Ludwig is at the Centre de Formation et