(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

100 YEARS AGO

An interesting description of the ravages of white ants, or termites, in Rhodesia is furnished by the Rev. A. Leboeuf to the *Zambesi Mission Record* for January. The special interest of the contribution centres in the account of the damage done to property by white ants in Rhodesia, which seems to be even greater than in India. It is no uncommon thing, says the writer, for the colonist, on returning from his day's labour, to find the coat he left hanging on a nail of his cottage wall and the books on the table absolutely destroyed by these tiny marauders. Nor is this all. "On awaking next morning," writes Mr. Leboeuf, "you are astonished to see in the dim light a cone-shaped object rising from the brick floor a short distance from your bed, with two holes on the top like the crater of a miniature volcano. Upon closer examination you discover that the holes have just the size and shape of the inside of your boots, which you incautiously left on the brick floor the night before. They have given form and proportion to an ant heap, and nothing is left of them except the nails, eyelets and, maybe, part of the heels."

From *Nature* 24 January 1901.

50 YEARS AGO

May I support Prof. Alan Boyden's plea that parthenogenesis should not be classified as asexual reproduction? The habit of doing so presumably arose because the text-book definition of sexual reproduction excludes anything which does not involve fusion of gametes; but such a reliance on a definition instead of on the facts implies a degree of philosophical realism which has no proper place in science. Biology, unlike logic and mathematics, has to take the world as it finds it. Definitions are descriptions of concepts or phenomena, sometimes of arbitrary stages in a series, and when they are made short they inevitably become inaccurate… No satisfactory definition of 'species' has ever been made. One of the latest, that of Mayr, is quite inadequate; apart from being formally incorrect in stating that a species is a group of populations, which it certainly is not, his definition would, since it contains the terms 'interbreeding', exclude *Amoeba proteus* and all creatures without cross-fertilization. Similar difficulties arise with the words cell, reproduction, tissue, skeleton, parasite and many others.

From *Nature* 27 January 1951.