nature genetics March

Lift-off for bovine genome project

A linkage map for of the cattle genome points to mechanisms of the reassortment of mammalian genes while a new technique raises questions about the structure of gene-starts in chromatin.

THE genome of the cow (otherwise called the bovine genome) is this month's claim on public attention in *Nature Genetics* for several reasons. For one thing, bovids are also mammals. For another, they are animals of great economic importance in which advantageous and distinguishable traits command high prices at rural auction sales, with the result that the fate of individuals is closely watched by fond owners. Then the practice of artificial insemination and the relatively short time-lag in cattle between birth and sexual maturity means that hugely informative pedigrees can readily be available if the record-keeping is reasonably good.

So it is only natural that a detailed genetic map of the bovine genome should be a by-product of state-sponsored cattle-breeding (a similar map from the US Department of Agriculture has been published in the February issue of *Genetics*). Almost predictably, Australia, and the Commonwealth Scientific and Industrial Research Organization (CSIRO), seems to have taken the lead in this project. Six out of 23 authors of the map (*Nature Genet.* **6**, 227–237; 1994) are from the CSIRO laboratory of the Tropical Animal Production Division at the University of Queensland at Brisbane. But collaborating laboratories span the world, from Texas A&M, the International Laboratory for Research on Animal Diseases (ILRAD) at Nairobi and the University of Wisconsin at Madison (all predictable) to Trinity College, Dublin, the University of Georgia, the Hebrew University of Jerusalem and Erdegnische Technische Hochschule (ETH) at Zurich (in varying degrees, less so).

The product of the collaboration is, at this stage, a linkage map only. In other words, the collaboration has considered a subset of 202 conveniently polymorphic loci of the bovine genome out of the more than 300 genetic so far assigned to chromosomes and has sought to put them in topographical order by analysis of the frequency with which putative nearest-neighbour variants are linked in different individuals. The result is not so much a map, as the starting-point for the construction of a map. It is both sobering and encouraging that the mapping of the human genome was only a little further advanced at the beginning of this decade — sobering, because the collaboration illustrates again the difficulty of making a map of any kind when the topography is devoid of landmarks, and encouraging because, in the human case, there has since been a ten-fold refinement.

Cattle-breeders will naturally await the more detailed maps that lie ahead, but others will be surprised at the guesses at the mechanism of mammalian speciation that the authors have been able to make by comparing their crude linkage map for cattle with the much more detailed maps now available for mice and people. The issue is simply put. At least to a first approximation, there is an uncanny resemblance between the genes of different mammals. Human genes are usually homologous with those of the mouse (whence the case, sometimes put, for making the Murine Genome Project a mode for the human version). So is the rudimentary bovine map more like that of mouse or Man?

Cattle have had 60 chromosomes to the genome (29 pairs of autosomal chromosomes and an X and a Y), compared with 46 in people. Given the apparently explosive radiation of very different super-families of mammals in the Cenezoic, it would be folly to expect the any one of the bovine, human and murine genomes to be ancestral to the other two. At least on the resolution of the new bovine map, the three species are equally alike (or different). In each of the three two-way comparisons now possible, while some stretches of chromosomes appear to have been simply translocated, other groups of linked genes seem to have been shuffled without rhyme or reason. But that in itself is a challenge; it is tempting to believe that there must be a pattern of some kind.

Meanwhile, there is also a potentially important improvement of sequencing technique to report, based on an observation of the non-random distribution of CG dinucleotide pairs in the mammalian genome. If it were that nucleotide sequences were strictly random, the chance that any pair of consecutive nucleotides would be a CG would be 1 in 16, but in reality the occurrence of CG pairs is far from random. Mostly they corp up once every 50 to 100 base-pairs and are heavily methylated, but they occur much more frequently (1 in 10 base-pairs or so) upstream of the 5'- ends of mammalian genes, when they are usually not methylated. So much has been demonstrated during the past decade by Adrian Bird and successive colleagues at the University of Edinburgh. Now *Nature Genet.* **6**, 236; 1994), Bird's group shows how the non- random distribution of CG pairs may be used in at least two ways in large-scale sequencing efforts.

The starting-point is a technique for separating short lengths of DNA containing methylated and unmethylated CG pairs, based on the use of a polypeptide (MBD) which is itself a portion of a protein associated with and isolated from rat chromosomes and which has the property of binding to symmetrically (or equally) methylated CG pairs. The goal is to isolate the CG "islands" which Bird's group (and others) have shown to span the upstream and early-exon parts of 60 per cent or so of mammalian genes. So why not incorporate the MBD peptide into a retention column, chop up the genomic DNA into fragments with a restriction enzyme cutting at sites that rarely occur within CG islands and use the retention column to fragments with methylated CG pairs?

The goal is to isolate the estimated 45,000 CG islands in the human genome, each typically 1,000 base-pairs long. The retention column works both as a means of removing non-island (or methylated) DNA as a means of fractionating the island sequences when they are artificially methylated after being segregated as a mixture. One potential prize is that the relatively simple task of sequencing 45,000 CG-islands, each roughly a kilobase in length, will yield as many distinctive genetic markers. But because the islands span both the promoters of the genes to which they correspond and at least the beginning of the transcribed mRNA, they should be distinctive means of isolating intact cDNA molecules, not the partial sequences usually produced. E

Obviously, there are snags. The restriction enzymes chosen so as to cut DNA only outside CG islands are not as specific as they might be, presumably because their target cutting-sequences occasionally occur within islands (which helps to explain why the average length of the fragments is less than expected. And some of the fragments are from ribosomal or mitochondrial DNA. It is also, of course, the plain truth that, if unmethylated CG-islands are associated with active genes, 40 per cent of the genes in the human genome will not be so identified. But that, it may be thought, is a lot better than nothing. □