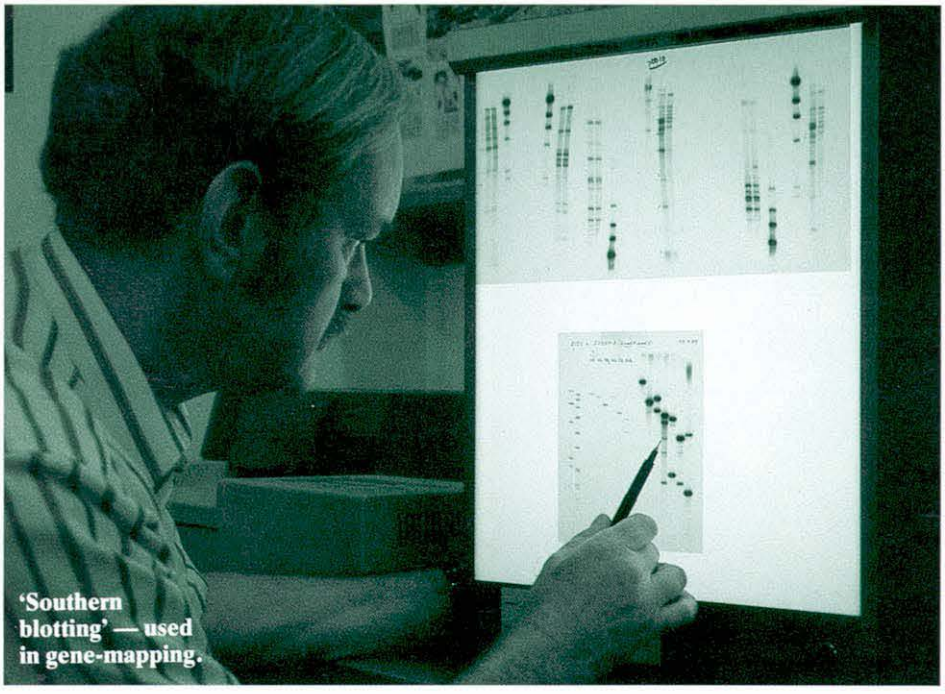


'Southern blotting' — used in gene-mapping.

The tale of life: mapping genes and chromosomes

A DNA-synthesiser — invaluable in fashioning small DNA sequences to probe an organism's genome.

Despite its central role in life, DNA is one of the most chemically monotonous of molecules.

Long stretches involve simple repeats of the bases cytosine (C), thymine (T), adenine (A), and guanine (G) and, at first glance, there is little fascination in a seemingly endless sequence like ATATAT... Similarly, there appears to be no rhyme or reason behind a several-hundred base stretch being repeated over and over, especially when such sequences can comprise up to 99% of the DNA in an organism's chromosomes and have, currently, no known function.

Yet we now know that set amid those repetitively boring bases are pearls of information. Three-base combinations of A, T, C, and G specify the 20 amino-acids; 500 of those triplets constitute a recipe for the average sort of gene that specifies the average sort of protein found in a cell. Link all these protein-determining combinations with regulatory sequences specifying such actions as stop, start, promote, and suppress, and that monotonous molecule orchestrates all life as we know it.

The amount of genetic information stored in each cell is quite formidable. The DNA — or genome — of the average mammalian cell is packaged in a variable number of chromosomes whose genes encode about 100 000 proteins. Those genes are hidden in among some 3 billion base pairs and surrounded and intruded upon by the so-called 'junk' DNA — the near-endless repeats of ATATAT and such-like. As well, genes can piggy-back on one another; in one recently unravelled human example, at least three genes are embedded in a mega-gene that may be as long as 2 million bases.

Finding a particular gene is a bit more daunting than finding that needle in the haystack, but it can be done. The on-going revolution in molecular biology has seen to that, and its tools now allow the dissection of all the chromosomes of any organism. It is now possible, in theory, to 'sequence' the genome of any animal, plant, or microbe — that is, to work out its precise base composition.

Indeed, the sequence of the 5000 bases in one of the smaller bacterial viruses was first reported as long ago as 1977. Within 5 years scientists had uncovered the tenfold-larger DNAs of other bacterial viruses, and the process continues; by 1990, the quarter-million base sequence of a member of the herpesvirus group was unravelled.

However, we are still a long way from physically mapping the genome of any organism higher than a virus. The difficulty is best illustrated by that most intensively studied microbe, *Escherichia coli*. This bacterium has been the work-horse of modern genetics and molecular biology, yet the sequence of only 800 000 base pairs (bp) of its 4 800 000-bp genome has been established.

Higher organisms

A viral sequence may be far removed from the complex DNA sequences of bacteria, let alone those of mammals, yet by the mid '80s molecular biologists had started contemplating the possibility of mapping the human genome. The United States Department of Energy — which has a standing brief to investigate the subtle effects of nuclear irradiation on humans (and what better way to do so than look for changes, or mutations, in the human genome?) — became interested, and the ball began rolling. More scientists expressed curiosity, the United States National Institutes of Health (which fund much of the country's biomedical research) became involved, and, eventually, Congress passed legislation to initiate the Human Genome Project (HGP).

The aim of the HGP is to derive the ordered sequence of the 3 billion units of A,

T, G, and C that spell out exactly what a human is. In the words of Dr Jim Watson, co-discoverer of the double helix structure of DNA, and now the director of the project, 'a more important set of instruction books will never be found by human beings'.

Such instruction books will help us to understand how we function as healthy individuals, and explain, at the most basic chemical level, the role of genetic factors in disease.

Single-gene mutations leading to disorders such as cystic fibrosis, phenylketonuria (the PKU test given to each newborn babe indicates how commonly that occurs), Alzheimer's disease, and sickle-cell anaemia have long been recognised. However, finding the location of the responsible gene is a tough exercise — tracking down the location of the cystic fibrosis gene took 10 years of effort involving scientists from dozens of labs, at a cost of between US$50 and US$200 million.

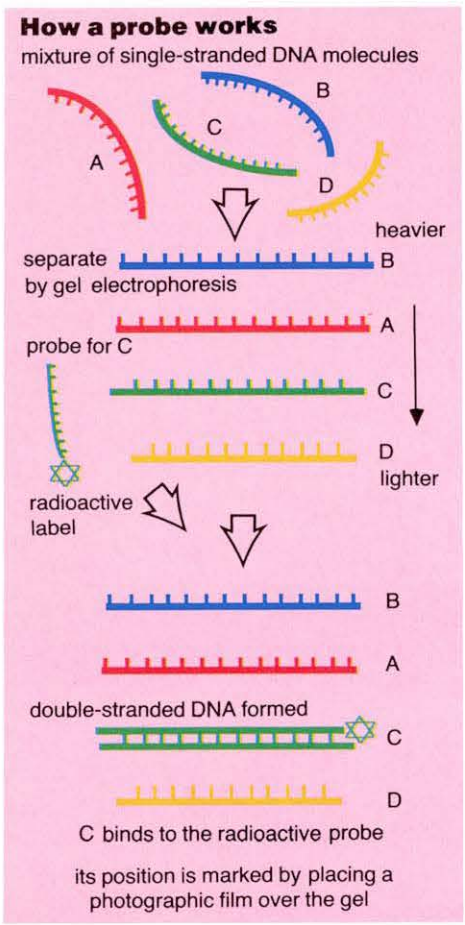
Multiple genetic factors are known to play an important role in modern society's major killers, heart disease and cancer. As with single-gene defects, the instruction book provided by the HGP should allow scientists to compare and match healthy and genetically-predisposed-to-disease individuals, monitor the onset and progress of the disease, and allow very precise medical intervention. It is not inconceivable that a new generation of drugs designed to substitute for the activities of faulty genes, or block the activity of genes that have been 'turned on' at inappropriate times, may emerge.

Success would also have implications for our domestic animals. Man and beast share many common genes and the project could change how we breed and rear our animals. More feed-efficient, disease-resistant, and commercially valuable animals would flow out of breeding programs using the results of genome mapping.

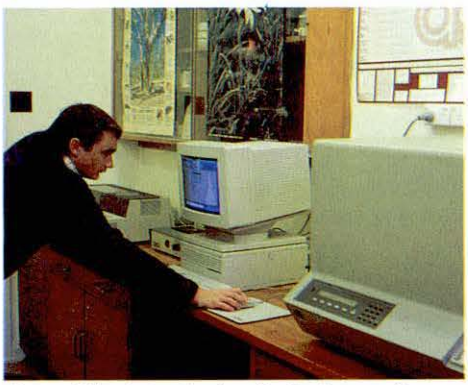
A major complication in this optimistic scenario is the sheer volume of information flowing out of the project. At present, a modern laboratory can routinely sequence 300–500 bp of DNA in a day at a cost of $3–5 per bp; a simple calculation around the 3 billion bp in the human genome produces some daunting figures. However, those involved in the HGP expect technological advances will bring the cost down to tens of cents per bp, and the project's expected budget is US$3 billion over the next 15 years.

The HGP is BIG science, on a par with the American Apollo project to put a man on the moon, and like that project it is going to have to exploit as-yet-undeveloped technologies. Apart from new sequencing techniques, its needs will include a huge input from computer scientists to handle the information flowing in; surprising as it may sound, today's supercomputers don't have the capacity necessary to process and match the sequence information.

Australian scientists are contributing to the HGP. The recent Australasian Gene Mapping Workshop, a meeting sponsored by the federal Department of Industry, Technology and Commerce, canvassed many of the organisational and technical issues involved in genome mapping in an attempt to co-ordinate Australian and New Zealand efforts, particularly with regard to genome mapping of livestock species.



Bonding between the complementary bases adenine (A) and thymine (T), and guanine (G) and cytosine (C), produces the double-stranded DNA molecule. Similarly, a single-stranded DNA molecule will bind with a complementary DNA probe — for example, a probe with the sequence TACG will selectively bind to a molecule that includes ATGC.



One of the latest technological marvels — an automated DNA-sequencer that can sequence up to 12 000 bases a day.

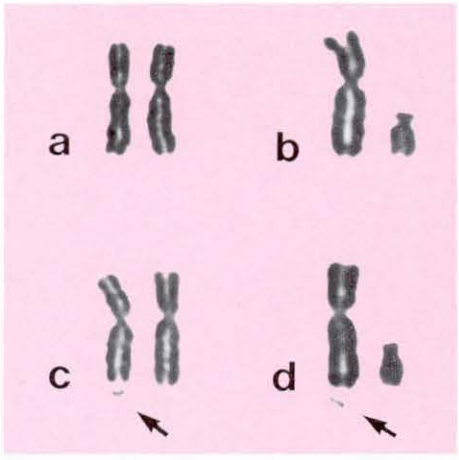
Human genetics

Scientific interest in human genetics has been centred on disease, best illustrated by those diseases that are sex-linked and impinge on the vulnerable male. The male has no compensating gene in his XY chromosome configuration (unlike the female, who is XX) when his mother passes on a mutant gene on the X chromosome, and suffers from the activity (or lack of it) of the affected enzyme. By determining just a part of the amino-acid sequence in the protein, researchers can infer the corresponding DNA sequence. Construction of a DNA-probe that is complementary to a short stretch of that sequence (see the diagram on this page) then allows them to define the precise location of the gene on the X chromosome.

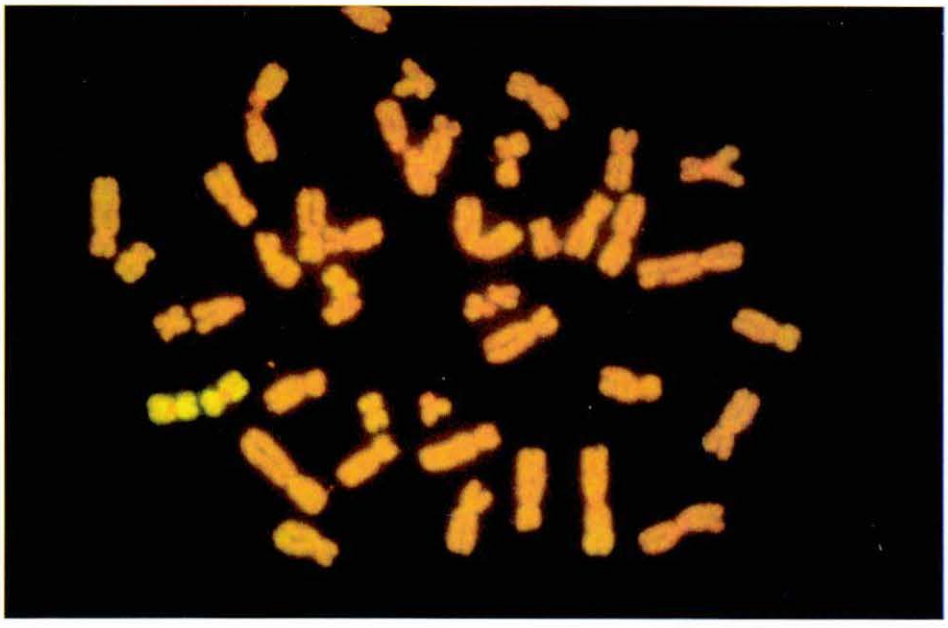
Identifying disease genes on chromosomes other than the sex chromosomes is much more difficult. Where it has been done, the information is invaluable in a mapping exercise because it helps orientate all the other genetic information that has accrued. To date the sequences (but not necessarily accurate chromosomal locations) of approximately 5000 human genes are known; of these, the vast majority have been assigned to the sex chromosomes.

Within Australia, a group led by Dr Grant Sutherland at the Adelaide Children's Hospital has, with the aid of United States Department of Energy and Australian National Health and Medical Research Council funds, taken on the study of human chromosome 16. The work grew out of Dr Sutherland's interest in a sex-linked chromosomal disorder giving rise to the Fragile-X syndrome. This is the most common genetic defect leading to familial mental retardation, affecting 1 in 2500 live births; the defect is obvious under the microscope, with the normal chromosomal shape being distorted by a 'pinching' around the fragile area.

At least 30 fragile sites exist on chromosomes other than the X; these are not implicated in any known defect. Nevertheless Dr Sutherland believed the two fragile sites on chromosome 16 could serve as a useful model for what happens on the X chromosome, and so, in 1985, he began mapping the area around those sites.



The normal female complement, XX (a), and male, XY (b). Below them one X-chromosome of the female (c) and the X-chromosome of the male (d) display the Fragile-X syndrome.



The 46 chromosomes of the human cell. The chromosome-16 pair is highlighted.

From right around the world, cell cultures from individuals with abnormal chromosome 16s are sent to the Adelaide laboratory; there they are grown and fused with mouse tumour cells. Most of the human chromosomes are lost as the hybrid continues to grow, but sometimes parts of a chromosome fuse with those of the mouse.

Dr Sutherland's group now hold many hybrid cell-lines and they probe the chromosome-16 fragments in a variety of ways to help define the gene sequence along the chromosome. First, they cut these with a restriction enzyme (see the box on page 22) and probe the separate pieces with cloned genes — the number of known genes on chromosome 16 was only five in 1985, but increased to more than 30 by early 1990 — and with anonymous DNA fragments (sequences known to be on the chromosome, but that is all). Systematic analysis allows overlaps to be identified and the order of genes and anonymous DNA fragments on chromosome 16 to be deduced.

The first aim of the group is to construct a map involving about 500 genetic markers, mapped at intervals of about 2 million bp. To date, mapping of chromosome 16 has proceeded as far as breaking it up into 30 different intervals. This appears to be slow progress, but the pace reflects the complexity of the human genome and the Adelaide project is as well-advanced as any other of the HGP's mapping efforts.

Some of the Adelaide group's markers are sufficiently close to allow sequencing of the intervening DNA, and they regularly send out information and probes to international collaborators who are helping prepare the physical map — or complete base sequence (see the box) — of chromosome 16. As the work proceeds, it uncovers other sequences and genes and the information is fed back to generate fresh probes for the chromosome.

Mammalian links

Inevitably, once substantial parts of the human genome have been defined, biologists will apply the results to other animals. This has already begun, and it has also become clear that animal mapping efforts will also benefit the HGP.

Mapping the relatively small (or well-known, in the case of the fruit fly) genomes of the lower animals will provide a good testing ground for the new technologies and computing approaches needed for the human genome. In the case of the mammals, the HGP will benefit because evolutionary history means man and beast share many genetic features.

The common ancestor of all mammals appeared on Earth around 80 million years ago, and subsequent evolution led to the emergence of the mouse, cow, and man. While these very different creatures have experienced subtle evolutionary changes so that they differ in the amino acid sequence of proteins like haemoglobin, their genomes remain remarkably similar — put simply, during those millions of years the arrangement and numbers of chromosomes have been scrambled, but the order of their (slightly different) genes is essentially unchanged.

Chromosomes are very mobile elements. Sometimes, during meiosis (the cell division involved in the creation of eggs and sperm) a mis-pairing may see one arm of a chromosome brought into contact with an unrelated chromosome, and gametes with one shorter (and one longer) chromosome can result. If this rearrangement leads to a new DNA sequence that improves the structure of an existing protein, or gives rise to an advantageous new gene and protein product, natural selection will see the translocation perpetuated.

By comparing what is known of the human, mouse, and cattle genomes, researchers estimate that 150–200 chromosomal rearrangements differentiate those of mouse and man. As few as 100 rearrangements mark the differences between cattle and man. Recent studies suggest, for example, that parts of the human chromosome 1 are now found on three different cattle chromosomes.

Mapping cattle

With such major similarities between the mammals, DNA probes and markers that have been isolated from the human genome can be used in mapping the cattle genome and *vice versa*. But cattle — in fact any domestic animals — offer a major advantage over humans in that genome mappers can control the parentage of their test animals. By making wide crosses, such as between Zebu (*Bos indicus*) and British (*Bos taurus*) cattle breeds, they can markedly increase the variety of genetic material found in the progeny and thus improve their chances of detecting stretches of DNA that differ between individuals; these bits of DNA can then serve as markers in the mapping exercise.

A group at the CSIRO Division of Tropical Animal Production, led by Dr Jay Hetzel, has already established such a set of cattle families and is in the process of developing a linkage map for cattle in collaboration with a group from the Texas A&M University led by Professor Jim Womack and with the American biotechnology company

The basics of mapping

Gene mappers seem to be working in the dark when they take out the most fundamental of their mapping tools — one of the many restriction enzymes derived from microbes. These enzymes cut DNA at very precise places — for example, only after it came across the base sequence AATG... would a particular enzyme make a cut.

Typically, a mapper begins by establishing a genomic library of the target organism's DNA by exposing it to a restriction enzyme that breaks it up into, in the case of mammals, as many as 1 000 000 fragments. Individual fragments can then be incorporated into a bacterial plasmid (a small circular piece of extra-chromosomal DNA) and multiplied many times over as the host bacterium divides and grows.

Mappers can extract and analyse this DNA fragment whenever they desire, using additional restriction enzymes to break it down; and gel electrophoresis — which involves running the DNA pieces through a porous gel under the influence of an electrical field — sees the even-smaller fragments separate out on the basis of their molecular weight. A process known as southern blotting (see the diagram at the top of the page) allows the mapper to define the position of specific DNA sequences in the gel.

At this stage all the mapper has is a series of long stretches of DNA that are virtually featureless and cannot even be traced back to the chromosome from whence they came. It is here that classical genetics intrude.

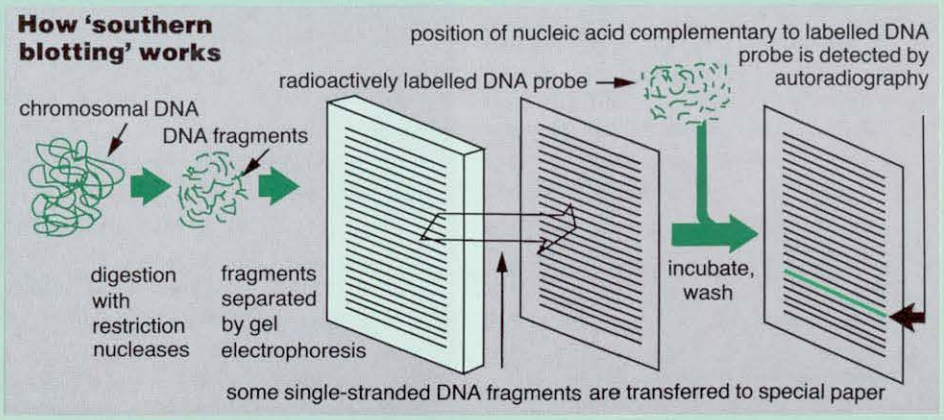
For successful sexual reproduction the chromosome complement of the mature animal has to be halved when eggs and sperm are created via meiosis. During meiosis, crossing-over occurs — adjacent arms of the chromosomes an individual has inherited from each parent come into close contact and the mechanics of the process demand the formation of stabilising links between the parental chromosomes. These are sturdy links, and, when the chromosomes part, arms of the paternal and maternal chromosome can be exchanged.

The illustration on the right shows how a female with two genes we will call *A* and *B* on her father's chromosome and *a* and *b* on the mother's contribution can, after crossing-over has occurred, produce eggs with the genetic arrangement *Ab* and *aB*. Crossing-over is a random, albeit common, process and eggs could still retain the *AB* and *ab* configuration. How often they do — the frequency of recombination — offers an indirect measure of how far apart those *A* and *B* genes are on the chromosome.

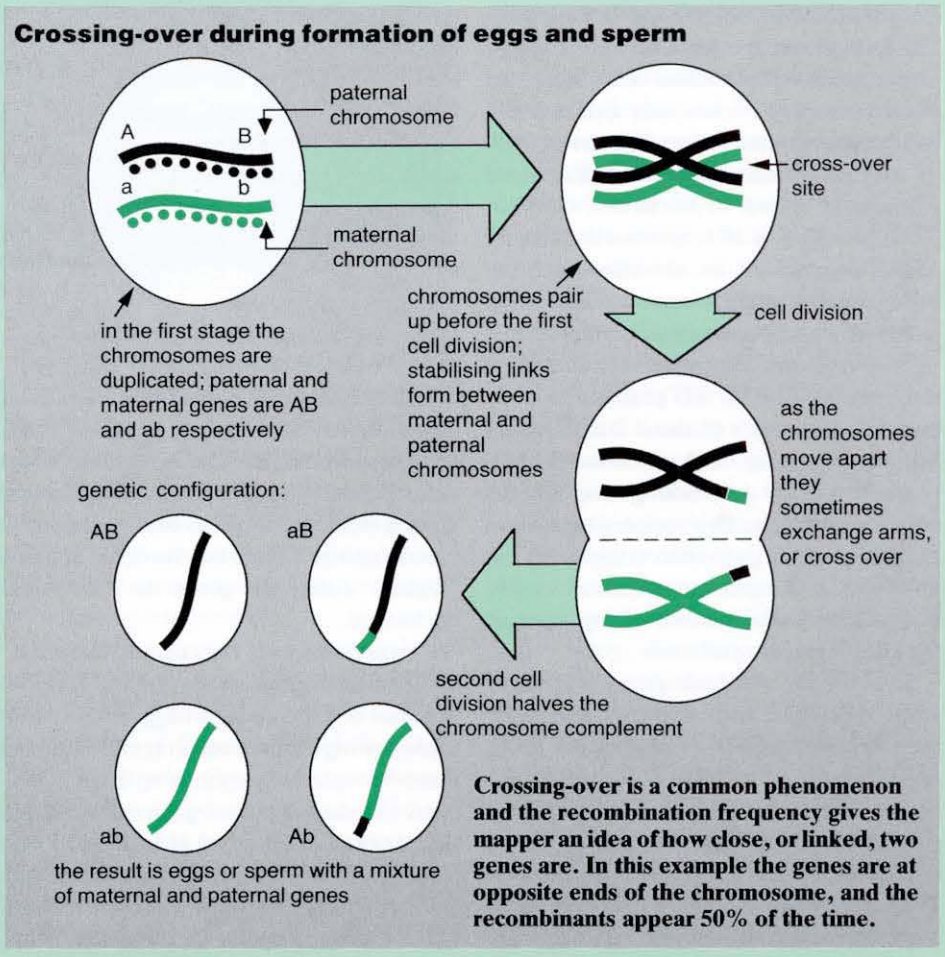
If they are at opposite ends of the chromosome, cross-overs will see the *Ab* and *aB* arrangements come up nearly half the time. Conversely, if they are very close to one another recombinations will be very rare — in other words, the genes are very closely linked.

At this stage the mapper encounters major problems, since the arrangement of genes along the chromosome isn't as simple as the A and B described above. The essence of the problem is that there are only a very few genes available to study linkage.

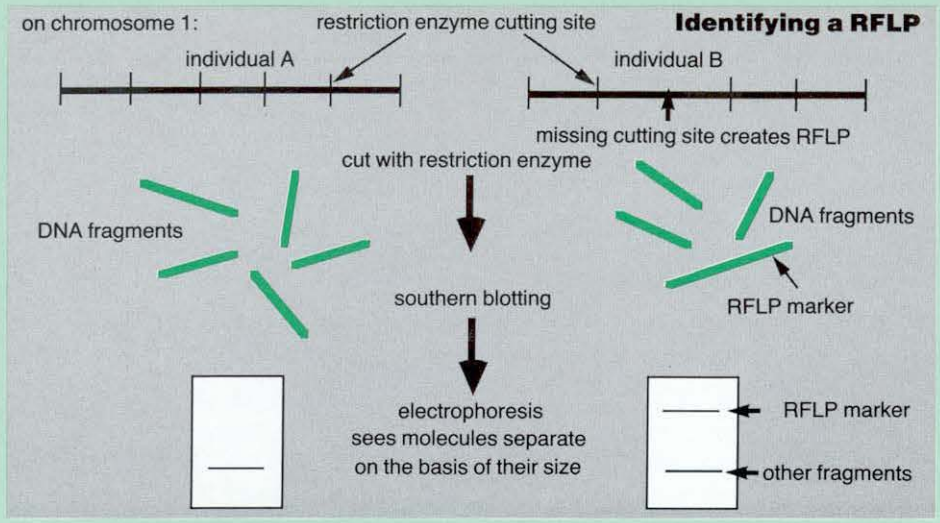
In the absence of defined genes, what the mappers need is well-spaced genetic markers that differ between individuals in the same way that the *A*/*a* and *B*/*b* genes in our earlier example did. The search for such markers is a major preoccupation in modern molecular biology.



DNA of interest, and the surrounding sequence, can be isolated by 'southern blotting'. Making use of a radioactively labelled DNA probe, this technique allows a desired stretch of DNA to be separated for further manipulation — such as insertion into a bacterial plasmid for multiplication.



Crossing-over is a common phenomenon and the recombination frequency gives the mapper an idea of how close, or linked, two genes are. In this example the genes are at opposite ends of the chromosome, and the recombinants appear 50% of the time.



Individuals differ in their restriction enzyme cutting sites, and this can be exploited in genome mapping and animal breeding. For example, if the RFLP in B is close to a known gene, the inheritance of that gene in B's progeny can be followed.

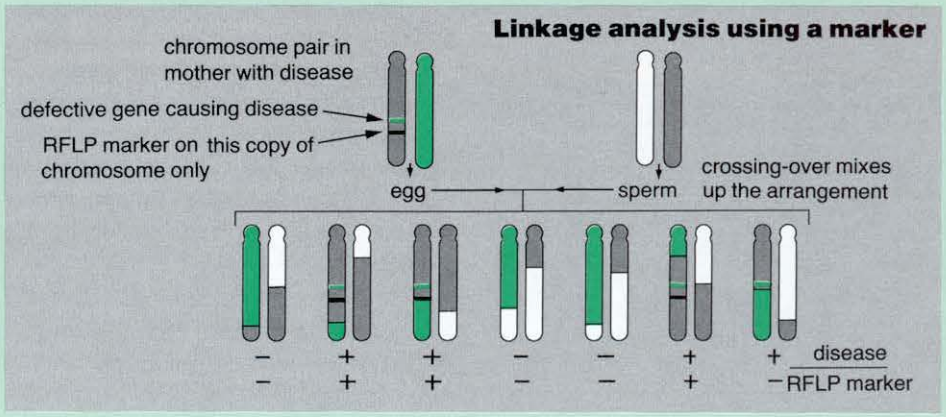
Almost perversely, the results of this search suggest that some of the best markers are to be found in the 'junk' DNA referred to at the beginning of the article. As an example, the sequence TGTGTG..., many times over, is common in mammals. The mapper makes good use of the fact that individuals can vary markedly in the length of such sequences.

In practical terms this means that, if a gene containing a restriction enzyme cutting site adjoins one of the variable lengths of TGTGTG..., a cut by that enzyme will give DNA fragments that differ between individuals. Those individuals are considered to be polymorphic, and that particular variable length of DNA is known as a Restriction Fragment Length Polymorphism, or RFLP.

Forensic scientists have made very good use of RFLPs — they are the basis of the 'genetic fingerprints' derived from the DNA in blood, semen, or hair found at a crime-scene — but genome mappers also use RFLPs to trace the exchange of genetic material during the development of families.

Markers cross over just like genes and, by following their progress, genome mappers can gauge how closely different markers are linked. By systematically analysing their co-inheritance, mappers can derive the relationship of these markers to one another and draw a *linkage map*. The use of DNA-probes derived from the markers then allows them to define which particular chromosome supports this array of markers.

The final stage of the HGP involves drawing the *physical map*. In its purest form this map would show just how the 3 billion bases in the genome are arranged, but for most practical purposes knowing the precise order of an array of genes along a particular chromosome, or even being able to order the clones in a yeast artificial chromosome (or YAC — see the box on page 24) library will do. Once this basic map is in place, biologists will be able to focus on the most important parts to more accurately define the sequence, the corresponding proteins, and their role in making us what we are.



Markers cross over just like genes. By following the transmission of a marker and associated gene, the gene mapper can estimate how closely they are linked. In this example, the gene causing disease in the mother is co-inherited with the RFLP marker in 75% of the progeny, indicating that they are relatively close on the chromosome.

GENMARK. The initial aim is to derive a useful (for breeding purposes) map with markers spaced approximately 20 million bases apart; this means that about 150 markers will be needed.

They have progressed so well that Dr Hetzel expects the map to be completed within the next 3 years. This rapid development is greatly assisted by the information and probes flowing out of the HGP; indeed, 75% of human gene probes can be applied to the cattle map. And information from the cattle project — particularly on the location of genes — is flowing back to the HGP.

For example, a rather obscure gene was initially defined as being 'somewhere' on human chromosome 1. However, researchers probing mouse and cattle genomes found it between two other genes. Extrapolating this knowledge back to the human chromosome quickly led to a more accurate definition of its location.

Animal breeding

Extrapolating information from one species to another will be common during the HGP. It will also prove useful in animal breeding.

As an example, a United States' group has identified mouse strains that differ by a factor of 10 in their response to the protozoan pathogen that parasitises red blood cells causing the disease leishmaniasis. This microbe is a representative of a large group of parasites, including those responsible for malaria and the cattle diseases East Coast fever and sleeping sickness. East Coast fever periodically sweeps through East African herds, killing up to 90% of a susceptible population, while sleeping sickness means that a very large swath of Central Africa cannot be grazed by cattle.

A prize bull. Genetic markers will alter the attributes that breeders follow in its progeny.

Some new technologies

To complete the Human Genome Project in a reasonable time, scientists will need an array of powerful new technologies. Since we are witnessing a revolution in molecular genetics this is not an unrealistic expectation, and a stream of innovative technologies is flowing out of the current intellectual ferment.

The article describes the fundamentals of genetic manipulation using bacterial plasmids. Unfortunately, these can only be used on stretches of DNA of about 5000 bases. Inserting the DNA to be mapped into bacterial viruses gives a tenfold increase, but this is still too small for efficient mapping.

New technology developed in the last 2 years has overcome that problem. It involves the construction of entities known as 'yeast artificial chromosomes' (YACs). A YAC begins life as a bacterial plasmid that has the genetic elements of yeast necessary for chromosome maintenance and duplication introduced into it. Further manipulation allows a large stretch — up to 1 million bases — of mammalian DNA to be inserted and, finally, incorporated into its yeast host. Almost certainly yeast (*Saccharomyces cerevisiae*) will become the favoured vehicle for manipulating the human genome.

Restriction enzymes also pose problems through cutting DNA into shorter pieces than required; the human genome in a million pieces is, to say the least, a bit unwieldy. A way to make fewer cuts, leaving larger pieces, is desirable, and just this year research has devised a way to achieve that goal.

Scientists at the University of Wisconsin fashion a DNA-regulatory protein (these proteins help suppress or initiate action on the DNA) so that it binds to a particular region of the genome, covering up some of the sites a restriction enzyme can cut. Chemical treatment of the whole genome then renders every other cutting site resistant to the enzyme. After release of the protein, the restriction enzyme can only cut at those previously protected sites.

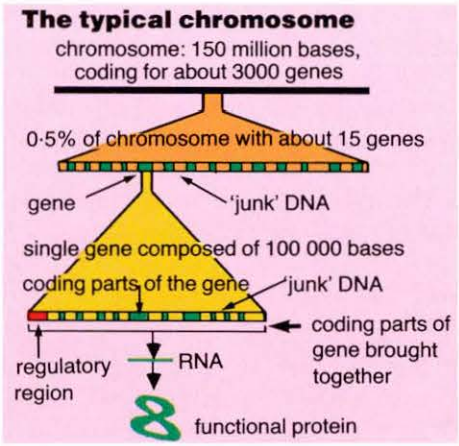
Theoretically, it is now possible to precisely cut just one chromosome out of the 46 found in a human cell, aiding both the separation and study of that single chromosome.

Given the pace of the molecular biology revolution, the many nascent technologies in the labs and minds of molecular biologists should see the mapping enterprise achieve its goals around the turn of the century.

The researchers have found that the genes, or gene, responsible for leishmaniasis resistance in the mouse lie between two known genes on mouse chromosome 1. Researchers are now homing in on the corresponding DNA sequence on bovine chromosome 8, with the implications for breeding disease-resistant animals, and perhaps a better understanding of human malaria, being very obvious.

However, for the immediate future domestic economic considerations will predominate in the thinking of the Australian cattle-genome mappers. Once having established an array of markers, the team will identify bulls with desirable traits — such as leanness or marbling in their meat cuts, or superior disease resistance or heat tolerance — by their specific marker patterns. Those patterns will be followed in the progeny, and individuals in which they appear will continue in the breeding program — it is being assumed that the desirable genes are swept along with the markers.

The marker-selection process offers the prospect of an unprecedented improvement in genetic gain and will have a great impact on traits that are currently difficult to select for. As an example, the meat-marbling characteristic has a heritability of about 50%, meaning that environmental effects account for about 50% of the variability observed in a population supporting the marbling genes. It is practically impossible to precisely control the nutritional and general environment in which cattle are raised, and selection for this important economic trait is made even more difficult because animals have to be slaughtered before measurements are made. Genetic markers linked to the character will, however, simplify breeding for the trait.



'Junk' DNA makes up the great bulk of the DNA in an organism's genome. Production of a functional protein from a gene involves 'editing out' the 'junk'.

Another example concerns characters with low heritability. Turning to sheep, the fine-wool character has a heritability of approximately 20% and progress through conventional breeding is very slow. Again, genetic markers related to the trait will simplify selection of desirable individuals.

Extending the book

With few other nations having such an interest in sheep, Australian and New Zealand researchers are planning to apply and extend the results of human and cattle mapping to our ovine friends. Plans are also afoot to map the genomes of chickens and goats. And the action is not confined to animals — the genomes of many plants are also being mapped, again with the same goal of improving breeding progress.

Clearly we have a technology that will give us, in the decades ahead, a series of 'instruction books' detailing the facts of life for many plants and animals. The human one will be the most detailed, and the prospect of the full base sequence of our DNA prompts memories of a statement made by a pioneering molecular biologist who, referring to the imminent release of the full sequence of a viral nucleic acid, suggested that this would be 'the first organism about which nothing remains to be learnt'.

While exaggerated, his statement contains more than a grain of truth: DNA is life. Within limits, it tells us who we are and what we will become — and having the instruction book of life in front of us will inspire the biologists of the next century to explore the limits of just who and what we are.

Wayne Ralph

More on the topic

The Human Genome Project: past, present, and future. J.D. Watson. *Science*, 1990, **248**, 44–9.

The use of reference families for genome mapping in domestic livestock. D.J.S. Hetzel. In 'Gene Mapping: Strategies, Techniques and Applications', ed. L.B. Schook, H.A. Lewin, and D.G. McLaren. (Marcel Dekker: New York 1990.)

Proceedings of the Australasian Gene Mapping Workshop, Macquarie University, Sydney, June 1990.