PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Genetic Maps

P. N. Goodfellow, L. Sefton and C. J. Farr

Phil. Trans. R. Soc. Lond. B 1993 **339**, 139-146
doi: 10.1098/rstb.1993.0010

Genetic maps

P. N. GOODFELLOW†, L. SEFTON AND C. J. FARR†

Laboratory of Human Molecular Genetics, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, U.K.

SUMMARY

The same genetic principles are common to all animals, plants and microorganisms. One consequence of the generality of genetic principles is that methods developed in one organism can often be used in other species. Recent advances in mammalian molecular genetics, genome analysis and gene mapping have been found to be directly applicable to other animals and even plants. This is facilitating the rapid construction of detailed genetic maps in a wide variety of species. Similar advances are being made in the development of manipulation techniques, which allow the genomic deletion and addition of specific genes as well as more subtle alterations. Armed with detailed maps and precise manipulation techniques it is possible to associate genes with phenotypes, test the function of genes and to unravel complex traits that depend on the interaction of different genes.

1. INTRODUCTION

Improvement of plant crops by altering yield, disease resistance and growth range has been a goal since the beginning of plant cultivation. Similarly improvements in fecundity, rates of growth and disease tolerance have been sought in animal husbandry. These agricultural aims have been approached traditionally by plant and animal breeding methods which are rooted in statistical analysis (Falconer 1960). The low resolution of the available genetic maps prevented the use of more precise methods of genetic analysis (Thoday 1961; Tanksley *et al.* 1989). The marriage of traditional genetics with molecular biology to produce molecular genetics has changed this outlook in three ways: first, by exploiting DNA sequence variation directly as a source of genetic markers, it is now relatively simple to construct a high resolution genetic map in any species; second, the map position of an identified locus can be used as the starting point for cloning any gene, a process variously called 'reverse genetics' (Ruddle 1984) or 'positional cloning' (Collins 1992), and third, the cloned gene can be used for genome modification by transgenesis.

Techniques for map creation and genome modification developed in one species are often directly applicable to another species. In closely related species, such as mammals, the maps themselves may be sufficiently similar to be applicable between species (O'Brien & Graves 1991; Searle *et al.* 1989). Even in distantly related organisms, such as fruit flies, mammals and flowering plants, both regulatory genes and structural genes may be sufficiently conserved that cloned genes from one organism can be used to recognize genes with similar biochemical functions in the other organisms (Frigerio *et al.* 1986; Burri *et al.* 1989).

Although the cloning of genes which cause major inherited diseases has been spectacular in the past ten years (Kunkel *et al.* 1985; Rommens *et al.* 1989), application of the same methods to either rare inherited diseases or multifactorial diseases is limited by both available family material and the major effort involved in cloning by positional cloning methods. The human genome project aims to provide a high resolution linkage map, identification and mapping of every gene and, eventually, the sequence of the genome (Dulbecco 1986). In the future, this will obviate much of the work needed in any specific study aimed at the analysis of any particular phenotype. Many of the techniques developed for genome mapping and analysis in humans have found a ready application in animal genetics and an echo in plant genetics.

In agricultural species, maps of the genome facilitate the identification of genes and this information can be exploited in breeding, however, even greater benefits can be expected if genes are cloned and then used for directed genome modification. The addition of genes by transgenesis and the subtle modification of genomes by homologous recombination will allow the rapid construction of plant and animal varieties designed specifically for different purposes and environments.

This review is intended to provide a brief summary of the use of molecular techniques in the construction of genetic maps and to provide a background for the following papers in this volume.

† Present address: Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, U.K.

Phil. Trans. R. Soc. Lond. B (1993) **339,** 139–146
Printed in Great Britain

2. POLYGENES AND QUANTITATIVE TRAIT LOCI

Many complex phenotypes are under the control of different genes acting together; these genes have been termed polygenes. In humans, several of the major diseases afflicting our society have complex inheritance patterns pointing to the existence of polygenes and environmental factors. In agricultural species, disease resistance, yield, quality and environmental tolerance are all subject to polygenic inheritance. Many of these traits are quantitative and the loci to which the corresponding polygenes map have been named quantitative trait loci or QTLs.

The contribution of any one 'polygene' to a specific phenotype may be small and dependent on environmental conditions. The identification of polygenes has been difficult because of the lack of suitable markers that could be used to track them in crosses (Thoday 1961; Tanksley *et al.* 1989). Moreover, the few markers that were available were often morphologically based and may have contributed to the phenotype under study. A number of techniques were developed to identify polygenes; for example, the construction of congenic lines by repeated backcrossing combined with selection for graft rejection was used to identify major and minor histocompatibility loci and recombinant inbred lines produced by inbreeding the progeny of an F2 generation were used to analyse drug resistance (Taylor 1976). These methods are laborious and tend to identify only those genes that make a large contribution to the phenotype. In practice, plant and animal breeding has been based on statistical methods with no attempt to identify the specific genes involved. This means that information gained in one cross cannot be applied to any future cross designed to yield further enhancement. High resolution linkage maps, composed of markers that are phenotypically neutral and that can be scored in any cross, provide the raw material for the identification of all the genes that contribute to a phenotype.

3. GENETIC MARKERS AND LINKAGE MAPS

Linkage maps are the fundamental resource for genetic analysis: they allow the association of genes with phenotypes. The ideal map would be comprised of a set of highly polymorphic markers spread evenly throughout the genome. The major problem in map creation is the number of markers that can be scored in any one cross. The major problem in map exploitation is the probability that a marker is polymorphic in a test cross. Sequence variation at the DNA level provides an unlimited source of polymorphic markers for map creation and exploitation. The three most commonly used sequence-based markers are:

1. Restriction fragment length polymorphisms (RFLPs). These markers are based on Southern blotting and simple sequence changes leading to the loss and gain of restriction sites (Solomon & Bodmer 1979; Botstein *et al.* 1980). The probes used include cDNA clones and random genomic fragments. The disadvantages of these markers are their limited polymorphism and the labour-intensive methods needed to identify and score them.

2. Minisatellites and microsatellites. These consist of a variable number of tandem repeats (VNTRs) and are frequently highly polymorphic with many different alleles at each locus. The core of the repeat sequence in many minisatellites is related to the sequence GGCAG-GAXG (Jeffreys *et al.* 1985). The repeat units in microsatellites are mono-, di-, tri- and tetranucleotides. In humans, minisatellites tend to cluster towards the telomeric ends of chromosomes, whereas, microsatellites are apparently more evenly distributed (Royle *et al.* 1990). A further advantage of microsatellites is that they can be scored by the polymerase chain reaction (PCR). The best studied class of minisatellites are the CA dinucleotide repeats. With an estimated 12 000 polymorphic $(dC-dA)_n \cdot (dG-dT)_n$ loci in the human genome, these markers alone would yield a genetic map with an average resolution of 0.3–0.5 cM (Weber 1990). Microsatellites are found universally in eukaryotes (Tautz & Renz 1984) and their high levels of polymorphism make them the markers of choice for many purposes. The disadvantage of microsatellite markers is the expense of isolating large numbers.

3. Random amplified polymorphic DNA (RAPDs). A single, short, oligonucleotide of random sequence is used as a primer in each PCR reaction (Williams *et al.* 1990). This is an inexpensive method for generating large numbers of markers very quickly, however, the patterns produced are strain specific and it is often not possible to recognize pairs of alleles in a cross. A further disadvantage is the difficulty in accessing the genome at the locus corresponding to the defined marker as this requires further cloning and sequencing. Nevertheless, for many purposes, especially in plant genetics, RAPDs are the markers of choice (Waugh & Powell 1992).

There are many other methods for detecting sequence variation some of which can be used for identifying or scoring polymorphism: allele specific hybridization with oligonucleotides (ASO, Wallace *et al.* 1979); single strand conformation polymorphisms (SSCP, Orita *et al.* 1989); denaturing gradient gel electrophoresis (DGGE, Myers *et al.* 1985); chemical cleavage at mismatches (Cotton *et al.* 1988) and two-dimensional gel electrophoresis (Hatada *et al.* 1991).

The possibility of generating unlimited numbers of highly polymorphic genetic markers by DNA based methods has changed genetics and map construction is occuring at a rapid rate in many different plant and animal species. We have chosen a few examples to illustrate the progress and problems in the creation of linkage maps.

(*a*) *The human map*

Constructing linkage maps of the human genome poses special problems. Humans are an outbred species with a long life cycle, random mating behaviour and small families. Although, Bayesian statistical methods can be used to obviate some of the worst aspects of random mating and small families (Ott

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

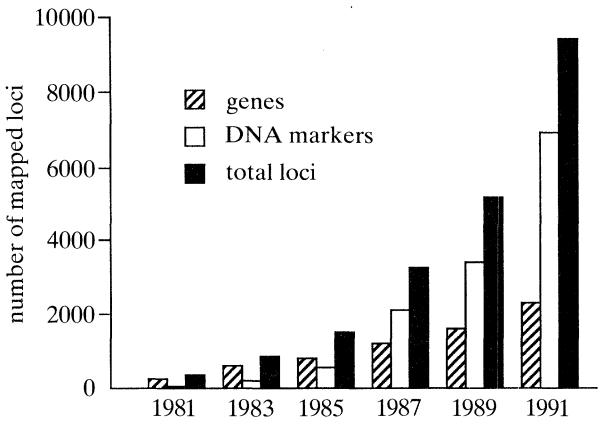

Figure 1. Progress in mapping the human genome. Legend data taken from the Human Gene Mapping Workshop (Human Gene Mapping 1991).

1974), these are poor circumstances for creating maps. Nevertheless, the biggest barrier to map creation for most of this century was the lack of phenotypic traits suitable for use as genetic markers. The impact of molecular techniques on human genetics can be seen in figure 1. The first human autosomal genes were not mapped to a specific chromosome until 1969 (Donahue *et al.* 1968); ten years later 360 markers had been placed on the human gene map by somatic cell genetic techniques (Shows & McAlpine 1979). However, the chance of finding linkage between a disease locus and a known polymorphic marker was negligible. The introduction of molecular probes resulted in a dramatic increase in the number of mapped loci. In 1992, 2325 genes have been mapped, 10 000 loci defined by DNA probes and framework linkage maps for the whole genome have been published (Human Gene Mapping 11 1991).

In the creation of maps of the human genome, linkage studies have been extensively supplemented with information gained from somatic cell genetics and cytogenetics. For many years, the only realistic option for mapping a gene to a specific human chromosome was to use human–rodent hybrids. Subchromosomal localization by this approach was restricted by the availability of appropriate translocation chromosomes. Some of the limitations of somatic cell methods have been overcome by radiation hybrid mapping (also known as irradiation and fusion gene transfer). Radiation hybrids are generated by the lethal irradiation of a donor cell line (usually a human–hamster hybrid containing a single human chromosome) and subsequent rescue by fusion to a second cell line (Goss & Harris 1975; Cox *et al.* 1990). The hybrids produced retain fragments of the donor chromosomes including the human chromosome. Marker order and inter-marker distance can be deduced from the frequency of retention of two human markers within the same cell line (Lawrence *et al.* 1991). A potential advantage of this method is that the frequency of chromosome breakage is related to the radiation dose and it is possible to create high resolution maps of chromosomes in regions of low recombination.

New techniques in cytogenetics are also contributing to genome mapping. Fluorescence *in situ* hybridization (FISH) is a sensitive technique which allows the simultaneous sublocalization and ordering of several DNA sequences. The resolution of FISH depends on the degree of condensation of the target chromatin: metaphase chromosomes are tightly packed and permit a resolution of 1 Mb; with the use of less condensed chromatin from interphase nuclei or pronuclei, the resolution may be increased to 50 kb (Brandriff *et al.* 1991).

The human map has been exploited to locate many of the genes causing common single-gene defect human diseases (e.g. Murray *et al.* 1982; Gusella *et al.* 1983) and this has led to the direct benefits of prenatal and presymptomatic diagnosis. The map location has also been exploited to clone several of these genes including those causing muscular dystrophies (Kunkel *et al.* 1985; Brook *et al.* 1992), cystic fibrosis (Rommens *et al.* 1989) and the fragile-X syndrome (Verkerk *et al.* 1991). Nevertheless, the difficulty of this approach can be gauged by the frustrations in cloning the Huntington disease gene which was among the first to be mapped using molecular markers. The increasing saturation of the human map has resulted in the identification of genes causing disease by co-localization with plausible candidate genes encoding proteins of known biochemical function (the 'candidate gene' method, e.g. Marfan syndrome, Dietz *et al.* 1991) and this type of serendipity will become increasingly common as the Genome Project proceeds.

One of the major problems in constructing maps is integration of those constructed by different groups using different techniques. Map integration can only be achieved if groups cooperate by using common sets of markers (reference markers). The construction of maps of the human genome has been aided by two cooperative organisations: the Human Gene Mapping Workshops and the Centre d'Etude du Polymorphisme Humaine (CEPH). Human Gene Mapping Workshops have provided a forum for collaboration on concensus map construction (HGM11 1991). Recently, these meetings have undergone a metamorphosis under the sponsorship of HUGO (Human Genome Organisation) to the Chromosome Co-ordinating Committee Meetings. CEPH maintains and distributes DNA from reference families allowing different groups to type the same families. The immortalised cell lines from the reference families are a renewable source of DNA and the nearest human equivalent to recombinant inbred strains.

(*b*) *The mouse map*

As a model for creating genetic maps of agricultural species, the mouse is a much more appropriate model than humans. The laboratory mouse is an experimental animal with a short life span and a wealth of phenotypic markers many of which affect coat colour and form. In consequence, the mouse linkage map was, for many years, much more advanced than the human map. Even in the mouse, however, DNA

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

markers have been needed for the creation of high resolution maps of general utility.

The most productive approach for creating linkage maps in the mouse has been the interspecific backcross between inbred laboratory mice and inbred strains of *Mus spretus* (Copeland & Jenkins 1991). The use of inbred strains reduces the number of alleles segregating in the cross from four to two and helps the unambiguous definition of the origin of any allele. The use of genetically diverse strains or subspecies increases the probability of detecting sequence variation at any locus. Potential problems in exploiting crosses between genetically divergent strains include suppression or enhancement of recombination due to chromosomal rearrangements, however, with a few minor exceptions these problems have not been observed. Nearly 3000 markers have been placed on the mouse linkage map and a map with an average resolution of 1 cM should be reached soon (Chapman *et al.* 1991).

Although the interspecific cross is now the workhorse of mouse gene mapping, recombinant inbred strains and congenic strains are still used extensively. Recombinant inbred strains are constructed by crossing two inbred strains followed by repeated brother sister mating of individuals of the F2 generation. Once homozygosity has been achieved, individual meiotic events from the F2 mice and subsequent rounds of inbreeding are fixed and these mice provide a renewable resource of meiotic events for map construction (Dietrich *et al.* 1992). Recombinant inbred strains also have a significant advantage for analysing quantitative traits because each genotype is present in all animals in the inbred line. This allows accurate measurement of phenotypic variability for each genotype (Taylor 1976; Lander & Botstein 1989).

Several groups have used high resolution maps to identify the genes contributing to polygenic diseases in the mouse. Genetic crosses between non-obese diabetic (NOD) mice and disease-resistant strains have identified 4–5 loci outside the MHC contributing to diabetes in the mouse (Todd *et al.* 1991; Prochazka *et al.* 1991; Cornall *et al.* 1991; Garchon *et al.* 1991). A similar analysis of hypertension has been performed in the rat (Hilbert *et al.* 1991).

Comparison of the human and mouse maps has revealed a large degree of conservation in map order despite the genetic distance between rodents and primates. On average the blocks of synteny are about 10 cM for autosomes (Searle 1989). The conserved blocks of synteny can be used for predicting homologous genetic diseases and can be helpful for suggesting candidate genes for human diseases. In a recent example, a gene causing Waardenburg syndrome (WS1) was mapped to a region of the long arm of chromosome 2 that is homologous to the proximal part of mouse chromosome 1. Waardenburg syndrome is responsible for over 2% of adult deafness and the tissues affected by this disease are related by their embryonic origin in the neural crest. The mouse mutation *Splotch* maps to the proximal part of chromosome 1 and the mutant mice are characterized by dysgenesis of neural crest cell derived tissues. As the mutation in *Splotch* is due to a deletion in *Pax-3* (Epstein *et al.* 1991) the human homologue HuP2 (Burri *et al.* 1989) was investigated and mutations in this gene were identified in patients with Waardenburg syndrome (Baldwin *et al.* 1992; Tassabehji *et al.* 1992).

(*c*) *Maps in other mammals*

There are 19 recognized orders of mammals that include 4000 species. Gene maps of 28 species have been published representing eight orders (O'Brien & Graves 1991). The high degree of syntenic conservation observed between mammalian species has allowed the human and mouse gene maps to be used as references for general gene map construction. The large number of cloned human and mouse sequences has provided a ready source of DNA markers.

For breeders of domestic animal species the availability of large numbers of genetic markers offers the possibility of mapping the genes that determine production traits and of using this information to increase the speed and accuracy of selection. One advantage which animal breeders have is that phenotypic information on thousands of animals has been carefully recorded over the years for use in classical biometrical breeding programmes, allowing them to design and generate the ideal family material required for mapping studies.

For linkage analysis of cattle, reference families have been constructed in Australia, Texas and Kenya and are being distributed to the bovine mapping community following the CEPH model. The combined use of VNTRs, microsatellites, *in situ* hybridization and of bovine–hamster somatic cell hybrids has resulted in the majority of the bovine genome being included in a low resolution linkage map. The remarkable conservation of mini- and microsatellites within Bovidae will substantially accelerate the construction of genetic maps in sheep and goats (Georges *et al.* 1991). In general it has been found that the extent of human–bovine chromosomal conservation exceeds that of human–mouse (O'Brien & Graves 1991).

Pig chromosomes were first studied at the beginning of this century and the correct chromosome number of 38 was described as early as 1931 over 20 years before man (Krallinger, 1931). The Pig Gene Mapping project (PiGMap) is based on a shared reference population (three generations) produced from crosses between the Chinese Meisham and the European Large White or European Wild Boar and improved European breeds. Recently, remarkable conservation between human and pigs has been reported for dinucleotide repeats (Wintero *et al.* 1992). The polydisperse nature of the porcine karyotype allows the chromosomes to be sorted effectively by FACS analysis, providing a powerful resource for physical mapping and for the construction of chromosome-specific libraries (Archibald & Haley 1992).

(*d*) *Maps in non-mammalian vertebrates*

Gene maps in non-mammalian vertebrates are underdeveloped despite the evolutionary insights that

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

might acrue from comparative mapping. Comparisons of fish and amphibian gene maps with those of mammals have identified several syntenic associations apparently conserved through more than 400 million years of vertebrate divergence. Expansion of fish gene maps should provide a valuable tool for the reconstruction of chromosome rearrangement events in mammalian lineages. Fish of the genus *Xiphophorus* (platyfishes and swordtails) are ideal for gene mapping because of their almost complete interfertility, the high level of genetic variability and numerous interspecific hybrids have been established, which should allow for rapid expansion of the *Xiphophorus* gene map (Morizot *et al.* 1991).

A preliminary linkage map of the chicken genome has recently been reported based on RFLP analysis of crosses between two lines of white leghorns (Bumstead & Palyga 1992). The chicken has a number of advantages for the construction of linkage maps: the genome is relatively small and the reproductive potential is high. However, the chicken karyotype is complex, with eight pairs of macrochromosomes and 30 pairs of microchromosomes (most of which are cytologically indistinguishable), as well as the sex chromosomes.

(*e*) *Linkage maps of plants*

The large numbers of morphological markers, the existence of ecotypes, the ease of creation of new variants by mutagenesis and the large numbers of offspring obtainable from individual plants facilitated the construction of detailed linkage maps in several plants (e.g. *Arabidopsis thaliana*, and maize). Nevertheless linkage maps constructed with DNA markers have been a major focus of research; access to the genome is offered by each marker and testing of large numbers of markers can be achieved in a single cross. In construction of maps it has also been possible to exploit a large number of other genetic tools including recombinant inbred strains, aneuploid variants that allow chromosomal localization by dosage, congenic strains constructed by introgression of chromosomal fragments from related species and translocation stocks. A list of some of the available plant linkage maps based on molecular markers is presented in table 1.

The detailed linkage maps have been exploited for mapping genes affecting disease resistance (e.g. Landry *et al.* 1987) and development (e.g. Weigel *et al.* 1992) and have been used to study the evolutionary relationships between different plant species (Bonierbale *et al.* 1988). The largest impact, however, has been on the study of complex phenotypes and polygenes. In several experiments it has been possible to define the Mendelian factors underlying quantitative traits affecting fruit size and quality (Paterson *et al.* 1988; 1991; Edwards *et al.* 1987) and water use (Martin *et al.* 1989). The ability to map a polygene depends critically on both the contribution it makes to the phenotype and the density of the linkage map. Once identified further localization may require construction of congenic or near isogenic lines specifically designed to study a particular polygene. In the short term the mapping of polygenes should facilitate manipulation of plant species to produce variants of improved agricultural performance; in the longer term cloning of polygenes and the direct manipulation of plant genomes should greatly improve the efficiency of producing new varieties.

Table 1. *RFLP linkage maps in plants*

plant species	reference
Arabidopsis thalia	Chang *et al.* (1988)
barley (*Hordeum vulgare*)	Heun *et al.* (1991)
lettuce (*Lactuca sativa*)	Landry *et al.* (1987)
maize (*Zea mays*)	Helentjaris (1987)
potato (*Solaneum tuberosum*)	Bonierbale *et al.* (1988)
rice (*Oryza sativa*)	McCouch *et al.* (1988)
soybean (*Glycine max*)	Apuya *et al.* (1988)
tomato (*Lycopersicon esculentum*)	Bernatzky & Tankersly (1986)

4. PHYSICAL MAPS AND CONSTRUCTION OF CONTIGS

Genome analysis has been transformed by the introduction of long-range restriction mapping (Brown & Bird 1986) and by cloning systems based on yeast artificial chromosomes (YACs, Burke *et al.* 1987). Long-range restriction mapping can be used to measure the precise physical distance between loci defined by DNA probes and has been applied to the genomes of prokaryotes as well as eukaryotes. Before the introduction of long-range restriction mapping, physical distances between loci were estimated indirectly by reference to linkage maps or by cloning of the intervening sequences: a feat that was rarely achieved in higher eukaryotes. Long-range restriction mapping utilizes restriction enzymes ('rare-cutters') that cleave rarely in genomic DNA because of the number of bases in their recognition sites. In many cases, rare cutters are also 'methylation-sensitive' and will not cut DNA in the recognition site has been methylated. Cleaved DNA is separated by pulsed field gel electrophoresis (PFGE) or by related methods such as field inversion gel electrophoresis (FIGE); these methods are capable of separating molecules as large as a few megabases (Schwartz & Cantor 1984; Carle & Olson 1984; Orbach *et al.* 1988). The distribution of CpG methylation and the corresponding long range restriction maps provides clues to genome organization. In mammals, methylation occurs only at CpG dinucleotides and is genome wide except for the CpG rich regions (also known as HTF islands; Bird 1987) found at the 5′ ends of many genes; this means that there is a direct relationship between the distribution of genes and the construction of the long range map (Bird & Brown 1986). Methylation at cytosine in plants occurs at CpXpG trinucleotides as well as CpG nucleotides (Gruenbaum *et al.* 1981). The distribution of methylation in plants has been studied less. However, blocks of methylation free genome appear to correspond to gene

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

rich regions (Antequera & Bird 1988; Moore *et al.* 1992).

The fragments of DNA that can be cloned in YACs are sufficiently large that they can integrate directly with linkage maps. In humans, 1 cM is approximately equivalent to 1 Mb of DNA and YACs with inserts this size are being produced by several groups (Larin *et al.* 1991). By comparison, in *Arabidopsis* 1 cM corresponds to only 140 kb (Chang *et al.* 1988). YACs can be used to rapidly isolate the DNA from target region of the genome and have greatly simplified chromosome walking. The major 'bottleneck' in positional cloning experiments is probably now in gene identification rather than chromosome walking.

Contigs are contiguous, overlapping cloned DNA fragments conceptually arranged in the same order as they are found in the genome. A complete contig of an organism's genome would allow high resolution mapping of any DNA based marker and could be used for instant access to any part of the genome. Initial attempts to construct contigs were based on cosmids (Somerville 1989; Stallings *et al.* 1990) using 'fingerprinting' techniques pioneered in studies on *C. elegans* (Coulson *et al.*, 1986); most recent studies have utilised YACs. In model organisms, such as *C. elegans* (Coulson *et al.* 1991) and *Arabidopsis* (Grill & Somerville 1991) near closure of contigs has been achieved.

The construction of a contig, besides providing an important mapping tool, is an essential first step in sequencing of the genome (Sulston *et al.* 1992).

5. CONCLUSIONS

The construction of genetic maps in plants and animals used to be regarded as an esoteric pastime with few practical consequences. With the introduction of molecular techniques this is no longer the case.

We would like to thank all those who took the time to try and educate us, in particular P.N.G. is grateful for conversations with Dr I. Furner, Dr G. Moore and Dr E. Coen. N. Ellis read the manuscript and made many perceptive comments, most of which we ignored.

REFERENCES

Antequera, F. & Bird, A.P. 1988 Unmethylated CpG islands associated with genes in higher plant DNA. *EMBO J.* **7**, 2295–2299.

Apuya, N.R., Frazier, B.L., Keim, P., Roth, E.J. & Lark, K.G. 1988 Restriction fragment length polymorphisms as genetic markers in soybean *Glycine max*. *Theor. appl. Genet.* **75**, 889–901.

Archibald, A. & Haley, C.S. 1992 The pig gene mapping project (PiGMap). *Gnome News* **10**, 26–28.

Baldwin, C.T., Hoth, C.F., Amos, J.A., da-Silva, E.O. & Milunsky, A. 1992 An exonic mutation in the *HuP2* paired domain gene causes Waardenburg's syndrome. *Nature, Lond.* **355**, 637–638.

Bernatzky, R. & Tanksley, S.D. 1986 Towards a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics* **112**, 887–898.

Bird, A.P. 1987 CpG islands as markes in the vertebrate nucleus. *Trends Genet.* **3**, 342–347.

Bonierbale, M.W., Plaisted, R.L. & Tanksley, S.D. 1988 RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* **120**, 1095–1103.

Botstein, D., White, R.L., Skolnick, M.H. & Davis, R.W. 1980 Construction of a genetic linkage map in man using restriction fragment length polymorphisms (RFLPs). *Am. J. hum. Genet.* **32**, 314–331.

Brandriff, B., Gordon, L. & Trask, B. 1991 A new system for high-resolution DNA sequence mapping in interphase pronuclei. *Genomics* **10**, 75–82.

Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K. *et al.* 1992 Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell* **68**, 799–808.

Brown, W.R. & Bird, A.P. 1986 Long-range restriction site mapping of mammalian genomic DNA. *Nature, Lond.* **322**, 477–481.

Bumstead, N. & Palyga, J. 1992 A preliminary linkage map of the chicken genome. *Genomics* **13**, 690–697.

Burke, D.T., Carle, G.F. & Olson, M.V. 1987 Cloning of large fragments of exogenous DNA in yeast by means of artificial chromosome vectors. *Science, Wash.* **236**, 806–812.

Burri, M., Tromvoukis, Y., Bopp, D., Frigerio, G. & Noll, M. 1989 Conservation of the paired domain in metazoans and its structure in three isolated human genes. *EMBO J.* **8**, 1183–1190.

Carle, G.F. & Olson, M.V. 1984 Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis. *Nucl. Acids Res.* **12**, 5647–5664.

Chang, C., Bowman, J.L., Dejohn, A.W., Lander, E.S. & Meyerowitz, E.M. 1988 Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6856–6860.

Chapman, V.M., Klein, J., Nadeau, J.H. & Silver, L.M. 1991 Encyclopedia of the mouse genome. *Mamm. Genome*, **1**, 1–532.

Collins, F.S. 1992 Positional cloning: Let's not call it reverse anymore. *Nature Genetics* **1**, 3–6.

Copeland, N.G. & Jenkins, N.A. 1991 Development and applications of a molecular genetic linkage map of the mouse genome. *Trends Genet.* **7**, 113–118.

Cornall, R.J., Prins, J.-B., Todd, J.A., Pressey, A., De Larato, N.H., Wicker, L.S. & Peterson, L.B. 1991 Type 1 diabetes in mice is linked to the interleukin-1 receptor and *Lsh/Ity/Bcg* genes on chromosome 1. *Nature, Lond.* **353**, 262–265.

Cotton, R.G.H., Rodrigues, N.R. & Campbell, R.D. 1988 Reactivity of cytosine and thymidine in single base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4397–4401.

Coulson, A., Kozono, Y., Lutterbach, B., Shownkeen, R., Sulston, J. & Waterston, R. 1991 YACs and the *C. elegans* genome. *BioEssays* **13**, 413–417.

Coulson, A., Sulston, J., Brenner, S. & Karn, J. 1986 Towards a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7821–7825.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S. & Myers, R.M. 1990 Radiation Hybrid Mappng: A somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science, Wash.* **250**, 245–250.

Dietrich, W., Katz, H., Lincoln, S.E., Sin, H.-S., Friedman, J., Dracopoli, N.C. & Lander, E.S. 1992 A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131**, 423–447.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Dietz, H.C., Cutting, G.R., Pyeritz, R.E., Maslen, C.L., Sakai, L.Y., Corson, G.M. & Puffenberger, E.G. *et al.* 1991 Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature, Lond.* **352**, 337–339.

Donahue, R.P., Bias, W.B., Renwick, J.H. & McKusick, V.A. 1968 Probable assignment of the Duffy blood group locus to chromosome 1 in man. *Proc. natn. Acad. Sci. U.S.A.* **61**, 949–955.

Dulbecco, R. 1986 A turning point in cancer research: sequencing the human genome. *Science, Wash.* **231**, 1055–1056.

Edwards, M.D., Stuber, C.W. & Wendel, J.F. 1987 Molecular-marker-facilitated investigations of quantitative loci in maize. 1. Numbers, genomic distribution and gene action. *Genetics* **116**, 113–125.

Epstein, D.J., Vekemans, M. & Gross, P. 1991 Splotch (Sp^{2H}), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of *Pax-3*. *Cell* **67**, 767–774.

Falconer, D.S. 1960 *Introduction to quantative genetics*. New York: Raven Press.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. & Noll, M. 1986 Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. *Cell* **47**, 735–746.

Garchon, H.-J., Bedossa, P., Eloy, L. & Bach, J.-F. 1991 Identification and mapping to chromosome 1 of a susceptibility locus for periinsulitis in non-obese diabetic mice. *Nature, Lond.* **353**, 260–262.

Georges, M., Gunawardana, A., Threadgill, O.W., Lathrop, M., Olsaker, I., Mishra, A. & Sargeant, L.L. *et al.* 1991 Characterisation of a set of variable number of tandem repeat markers conserved in Bovidae. *Genomics* **11**, 24–32.

Goss, S.J. & Harris, H. 1975 New method for mapping genes in human chromosomes. *Nature, Lond.* **255**, 680–683.

Grill, E. & Somerville, C. 1991 Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. *Molec. Gen. Genet.* **226**, 484–490.

Gruenbaum, Y., Naveh-Many, T., Cedar, H. & Razin, A. 1981 Sequence specificity of methylation in higher plant DNA. *Nature, Lond.* **292**, 860–862.

Gusella, J.F., Wexler, N.S., Conneally, P.M., Naylor, S.L., Anderson, M.A., Tanzi, R.E. & Watkins, P.C. *et al.* 1983 A polymorphic DNA marker genetically linked to Huntington's disease. *Nature, Lond.* **306**, 234–236.

Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H. & Mukai, T. 1991 A genomic scanning method for higher organisms using restriction sites as landmarks. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9523–9527.

Helentjaris, T. 1987 A genetic linkage map for maize based on RFLPs. *Trends Genet.* **3**, 217–221.

Heun, M., Kennedy, A.E., Anderson, J.A., Lapitan, N.L.V., Sorrells, M.E. & Tanksley, S.D. 1991 Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*) *Genome* **34**, 437–447.

Hilbert, P., Lindpaintner, K., Beckmann, J.S., Serikawa, T., Soubrier, F., Dubay, C., Cartwright, P. *et al.* 1991 Chromosomal mapping of two genetic loci associated with blood-pressure regulation in hereditary hypertensive rats. *Nature, Lond.* **353**, 521–529.

Human Gene Mapping 11. 1991 *Cytogenet. Cell Genet.* **56**, 1–2200.

Jeffreys, A.J., Wilson, V. & Thein, S.L. 1985 Hypervariable 'minisatellite' regions in human DNA. *Nature, Lond.* **314**, 67–73.

Krallinger, H.F. 1931 Cytologische Studien in einigen Haussaugetieren. Habilitationsschrift 1988. Arch. F. Tierernahrwig und Tierzucht. Abt. B5 Breslau. (Cited in Gustavsson, I. *Hereditas* **109**, 151–157.

Kunkel, L.M., Monaco, A.P., Middlesworth, W., Ochs, H.D. & Latt, S.A. 1985 Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosomal deletion. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4778–4782.

Lander, E.S. & Botstein, D. 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Landry, B.S., Kesseli, R.V., Farrara, B. & Michelmore, R.W. 1987 Genetic map of lettuce (*Lactuca sativa*) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. *Genetics* **116**, 331–337.

Larin, Z., Monaco, A.P. & Lehrach, H. 1991 Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4123–4127.

Lawrence, S., Morton, N.E. & Cox, D.R. 1991 Radiation hybrid mapping. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7477–7480.

Li, Y.K. & Ponce-de-Leon, F.A. 1992 Partitioning of the chicken genome by microcell hybridisation. *Poult. Sci.* **71**, 151–160.

Ludecke, H.-J., Senger, G., Claussen, U. & Horsthemke, B. 1989 Cloning of defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification. *Nature, Lond.* **338**, 348–350.

Martin, B., Nienhuis, J., King, G. & Schaefer, A. 1989 Restriction fragment length polymorphisms associated with water use efficiency in tomato. *Science, Wash.* **243**, 1725–1728.

McCouch, S.R., Kochert, G., Yu, Z.H., Wang, Z.Y., Khush, G.S., Coffman, W.R. & Tankersly, S.D. 1988 Molecular mapping of rice chromosomes. *Theor. appl. Genet.* **76**, 815–829.

Moore, G., Cheung, W., Foote, T., Gale, M., Koebner, R., Leitch, A. & Leitch, I. 1992 Key features of cereal genome organisation as revealed by the use of cytosine methylation-sensitive restriction endonucleases. *Genomics* (In the press.)

Morizot, D.C., Slaugenhaupt, S.A., Kallman, D.K. & Chakravarti, A. 1991 Genetic linkage map of fishes of the genus Xiphophorus (Teleostei: Poeciliidae). *Genetics* **127**, 399–410.

Murray, J.M., Davies, K.E., Harper, P.S., Meredith, L., Mueller, C.R. & Williamson, R. 1982 Linkage relationships of a cloned DNA on the short arm of the X chromosome to Duchenne muscular dystrophy. *Nature, Lond.* **300**, 69–70.

Myers, R.M., Fischer, S.G., Lerman, L.S. & Maniatis, T. 1985 Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucl. Acids Res.* **13**, 3131–3146.

Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D., Ledbetter, D.H. & Caskey, C.T. 1989 Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6686–6690.

O'Brien, S.J. & Graves, J.A.M. 1991 Report of the committee on comparative gene mapping. *Cytogenet. Cell Genet.* **58**, 1124–1151.

Orbach, M.J., Vollrath, D., Davis, R.W. & Yanofsky, C. 1988 An electrophoretic karytype of Neurospora crassa. *Molec. Cell Biol.* **8**, 1469–1473.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. 1989 Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874–879.

Ott, J. 1974 Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am. J. hum. Genet.* **26**, 588–597.

Paterson, A.H., Damon, S., Hewitt, J.D., Zamir, D., Rabinowtch, H.D., Lincoln, S.E., Lander, E.S. & Tanksley, S.D. 1991 Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. *Genetics* **127**, 181–197.

Paterson, A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E. & Tanksley, S.D. 1988 Resolution of quantitative traits into Mendelian factors, using a complete linkage map of restriction fragment length polymorphisms. *Nature, Lond.* **335**, 721–726.

Prochazka, M., Serreze, D.V., Frankel, W.M. & Leiter, E.H. 1991 NOR/Lt Mice MHC-matched diabetes-resistant control strain for NOD mice. *Diabetes* **41**, 98–106.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R. *et al.* 1990 A novel rapid method for the isolation of terminal sequences from artificial chromosome (YAC) clones. *Nucl. Acids Res.* **18**, 2887–2890.

Rommens, J.M., Iannuzzi, M.C., Kerem, B.-S., Drumm, M.L., Melmer, G., Dean, M. & Rozmahel, R. *et al.* 1989 Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science, Wash.* **254**, 1059–1065.

Royle, N.J., Clarkson, R.E., Wong, Z. & Jeffreys, A.J. 1990 Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* **3**, 352–360.

Ruddle, F.H. 1984 The William Allan Memorial Award Address: Reverse genetics and beyond. *Am. J. hum. Genet.* **36**, 944–953.

Schwartz, D.C. & Cantor, C.R. 1984 Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67–75.

Searle, A.G., Peters, J., Lyon, M.F., Hall, J.G., Evans, E.P., Edwards, J.H. & Buckle, V.J. 1989 Chromosome maps of man and mouse IV. *Ann. hum. Genet.* **53**, 89–140.

Sefton, L. & Goodfellow, P.N. 1992 The human gene map. *Curr. Top. Genet. Devl Biol.* (In the press.)

Shows, T.B. & McAlpine, P.J. 1979 The 1979 catalogue of human genes and chromosome assignments. *Cytogenet. Cell Genet.* **25**, 117–127.

Solomon, E. & Bodmer, W.F. 1979 Evolution of a sickle variant gene. *Lancet* **1**, 923.

Somerville, C.R. 1989 Arabidopsis blooms. *Plant Cell* **1**, 1135–1137.

Stallings, R.L., Torney, D.C., Hildebrand, C.E., Longmire, J.L., Deaven, L.L., Jett, J.H., Doggett, N.A. & Moyzis, R.K. 1990 Physical mapping of human chromosomes by repetitive sequence fingerprinting. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6218–6222.

Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N. *et al.* 1992 The C. elegans genome sequencing project: a beginning. *Nature, Lond.* **356**, 37–41.

Tanksley, S.D., Young, N.D., Paterson, A.H. & Bonierbale, M.W. 1989 RFLP mapping in plant breeding: new tools for an old science. *Biotechnology* **7**, 257–266.

Tassabehji, M., Read, A.P., Newton, V.E., Harris, R., Ballings, R., Gruss, P. & Strachan, T. 1992 Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature, Lond.* **355**, 635–636.

Tautz, D. & Renz, M. 1984 Simple sequences are ubiquitous elements of eukaryotic genomes. *Nucl. Acids Res.* **12**, 4127–4138.

Taylor, B.A. 1976 Genetic analysis of susceptability to isonazid-induced seizures in mice. *Genetics* **83**, 373–377.

Thoday, J.M. 1961 Location of polygenes. *Nature, Lond.* **191**, 368–370.

Todd, J.A., Aitman, T.J., Cornall, R.J., Ghosh, S., Hall, J.R.S., Hearne, C.M., Knight, A.M. *et al.* 1991 Genetic analysis of autoimmune type 1 diabetes mellitus in mice. *Nature, Lond.* **351**, 542–547.

Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., Fu, Y.-H., Kuhl, D.P.A., Pizzuti, A., Reiner, O. *et al.* 1991 Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in Fragile X syndrome. *Cell* **65**, 905–914.

Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T. & Itakura, K. 1979 Hybridization of synthetic oligonucleotides to ox174 DNA: the effect of single base pair mismatch. *Nucl. Acids. Res.* **6**, 3543–3557.

Waugh, R. & Powell, W. 1992 Using RAPD markers for crop improvement. *Trends Biotechnol.* **10**, 186–191.

Weber, J.L. 1990 Informativeness of human $(dC-dA)_n$–$(dG-dT)_n$ polymorphisms. *Genomics* **7**, 524–530.

Weigel, D., Alvarez, J., Smyth, D.R., Yanofsky, M.F. & Meyerowitz, E.M. 1992 *LEAFY* controls floral meristem identity in arabidopsis. *Cell* **69**, 843–859.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. & Tingey, S.V. 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* **18**, 6531–6535.

Wintero, A.K., Fredholm, M. & Thomsen, P.D. 1992 Variable $(dG-dT)_n-(dC-dA)_n$ sequences in the porcine genome. *Genomics* **12**, 281–288.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES