Genome Mapping

Introductory article

Paul H Dear, *MRC Laboratory of Molecular Biology, Cambridge, UK*

Genome maps pinpoint the location of specific features on the chromosomes of an organism. They are essential tools in identifying genes responsible for diseases or traits, for comparing the genomes of different species, and for complete genome sequencing.

Article Contents

Introduction

A genome map, like any other map, defines the relative positions of features that are of interest, or which can serve as reference points for navigation. The features that are located on a genome map are collectively referred to as markers, and can include both genes and noncoding sequences, as outlined below.

Many methods exist for making genome maps, differing in the types of markers that can be mapped and in their accuracy. Only one technique (fluorescence *in situ* hybridization) allows the position of markers in the genome to be observed directly. In all other methods, a less direct approach is necessary. A common theme in these indirect methods is that some means is first used to isolate a portion of the genome from the remainder (for example, by cloning in bacterial cells), and to identify the markers present in that portion. The markers found can then be assumed to be physically linked in the genome (for example, all markers found on a single cloned fragment must lie consecutively). By analysing many such samples of the genome, the order and spacing of all markers can be inferred.

Good maps are invaluable in many aspects of genome analysis. Genetic linkage maps (see below), although often of low resolution, are unique in that they allow the position of a gene to be determined even when nothing about the gene is known except for its phenotypic effect. Such maps, therefore, are often the first step towards isolating the gene responsible for a disease or for a specific trait. Comparison of genome maps of different species can reveal conservation of the order of genes over large regions of chromosomes. Such comparative mapping can give insight into genome evolution, and can provide clues as to the location of a gene in one species if its position in the genome of another is known. Higher resolution maps can direct the early stages of genome sequencing or, if accurate enough, can help to guide the assembly of the complete sequence and ensure its long-range continuity and freedom from gaps.

Some confusion exists between the terms 'genome mapping' and 'gene mapping'. Although there is some overlap of meaning, we can define 'gene mapping' as the locating of one or a few genes of particular interest, within the framework of a more global 'genome map'. This article is concerned largely with genome mapping.

Markers for Genome Mapping

Almost any identifiable feature of the genome can serve as a marker in mapping (**Figure 1**). The markers most commonly used are simply short, unique fragments of known deoxyribonucleic acid (DNA) sequence, which can

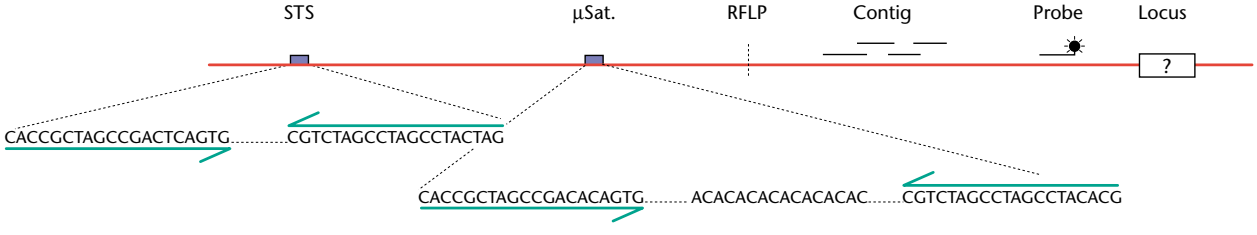

Figure 1 Markers used in genome mapping. This hypothetical composite map shows several types of marker located on the genome (heavy horizontal line). STS: sequence-tagged site, a region of known sequence which can be amplified and detected by the polymerase chain reaction using flanking sequence-specific primers (arrows); if the STS lies within a gene, it is an 'expressed sequence tag' or EST. μSat.: microsatellite, an STS that encompasses a simple, tandemly repeated sequence; the number of tandem repeats can vary from one individual to the next, making the marker polymorphic. The length of the PCR product produced using the flanking primers (arrows) will reflect the number of repeats. RFLP: a polymorphic restriction site, present in some individuals but absent in others. If genomic DNA is digested with the appropriate restriction enzyme, the sizes of the fragments will reflect the presence or absence of the site. Contig: a series of cloned DNA fragments that overlap one another in a contiguous series. The absolute position of the contig within the genome is known only if one or more of the clones contains a marker mapped by independent means. Probe: a labelled, cloned DNA fragment whose position in the genome can be observed directly by hybridizing it to metaphase chromosomes and observing it under a microscope. Locus: a gene or other sequence known only by its phenotypic effect. Its location in the genome can be inferred from the pattern of inheritance of the phenotype.

be amplified and detected in a DNA sample using the polymerase chain reaction (PCR). Such sequence fragments are collectively called 'sequence tagged sites' or STSs. Often, the STS will be some arbitrarily chosen piece of DNA (obtained by sequencing a clone picked at random from a genomic clone library), and has no particular biological significance. However, an STS can originate from a coding sequence (for example, by sequencing a clone from a complementary DNA (cDNA) library), in which case it is an 'expressed sequence tag' or EST. A third category of STS consists of a tandem repetition of a simple, short motif (for example, CACACACA...) flanked by a unique sequence. Such 'microsatellites' can be detected by PCR using their unique flanking sequence but often display length polymorphism from one individual to the next, as the number of repeat elements varies. Such polymorphism is of key importance in genetic linkage mapping.

Cloned DNA fragments can be mapped, and therefore act as markers, even when little or nothing is known of their sequence. Techniques for mapping such cloned fragments include physical mapping and fluorescence *in situ* hybridization. If the anonymous cloned fragment contains a restriction site whose location differs from one allele to the next (a restriction fragment length polymorphism or RFLP), then genetic linkage mapping can be used.

Finally, phenotypic variations act as genetic linkage markers even when nothing whatever is known of the underlying DNA sequence. In these cases, it is the phenotypic trait that is the observable marker, from which the presence of the allele at the DNA level is inferred.

Methods for Genome Mapping

Genetic linkage mapping

In organisms that reproduce sexually, meiosis breaks the parental chromosomes at random, recombines the fragments and segregates the shuffled chromosomes into gametes, and thence to offspring. If two markers, A and B, lie close together on a chromosome, then it is unlikely that a meiotic break will occur between them. Hence, A and B will seldom recombine, and will usually segregate together (cosegregate) to the same gametes and hence to the same offspring. In contrast, if markers B and Z lie at opposite ends of a chromosome, it is much more likely that a meiotic break will fall between them, whereupon they may segregate independently. Markers lying on different chromosomes will also segregate independently. Hence, the distance between any two markers is reflected by their recombination frequency: closely linked markers recombine rarely (or cosegregate often) while distant markers recombine often, and hence cosegregate no more than expected by chance – 50% of the time. This provides a way to estimate the distances between pairs of markers and, if all pairwise distances are known, a map can be made.

There is one complicating factor, however. Meiotic recombination occurs by an exchange of homologous segments between chromosomes. Hence, some way must be found to distinguish between the copies of the markers on each parental chromosome, so that the pattern of segregation can be followed. This is only possible in the case of polymorphic markers, and recombination can be detected only in those cases where the parent is heterozygous at the relevant loci. In addition, three generations (grandparents, parents and offspring) are normally required to be able to deduce completely the pattern of segregation (**Figure 2**).

Many types of polymorphic markers exist. In the earliest genetic linkage maps, the markers were not defined DNA sequences but variable traits which were assumed to reflect underlying (and then undetectable) polymorphisms. Such 'trait mapping' is still performed, and has the great advantage that it can identify the relative positions of genes based solely on their phenotypic effect. This is exploited, for example, in mapping disease loci as a first step toward identifying the gene responsible.

More often, though, genetic linkage maps are made using markers characterized at the sequence level. The earliest such markers were restriction fragment length polymorphisms or RFLPs – regions of sequence in which the alleles differ in the distribution of sites for a particular restriction enzyme. RFLPs can be detected by digesting genomic DNA with the restriction enzyme, resolving the fragments electrophoretically, Southern blotting and probing with a labelled piece of DNA complementary to the region of interest. Different alleles then appear as bands of different sizes on the autoradiograph.

RFLPs have largely been superseded in linkage mapping by microsatellites, regions of simple repeated sequence (such as CACACACA...) in which the number of repeats differs between alleles. Alleles are scored by measuring (on gels) the length of a PCR product produced by primers which anneal to the unique sequence on either side of the microsatellite.

Sequence polymorphisms of a single nucleotide can be detected by PCR- or hybridization-based methods. Such single nucleotide polymorphisms or SNPs are extremely common in the human genome, occurring once every few hundred bases on average. However, SNPs are more often identified once a region of the genome has been mapped and sequenced, rather than serving as markers for the initial genetic linkage mapping.

The analysis of linkage data to produce maps is far from simple (especially where many markers are to be mapped) and several mathematical methods exist. The first step is normally to estimate the distances between all possible pairs of markers in the study. In the ideal case, this can be done by finding the proportion of meiotic events in which recombination occurs between the two markers. However,

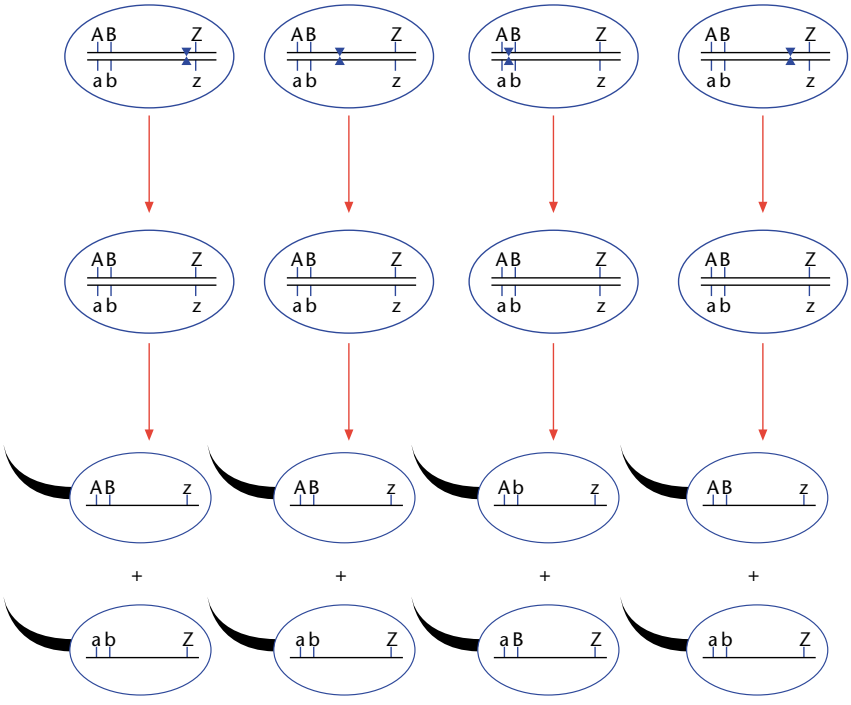

Figure 2 Principle of genetic linkage mapping (only one chromosome is shown). Diploid parental cells (top row) carry polymorphic marker alleles A, B and Z on one chromosome and alleles a, b and z on the other homologue. Meiotic recombination can occur at any point (indicated by the hourglass shape) along the chromosomes, causing an exchange of markers on either side of the site of crossover (second row). The recombined chromosomes are then segregated to gametes (bottom). Recombination is unlikely to occur in the small interval between closely-linked markers such as A and B – in this example, only one of the four parental cells (two of eight gametes) show recombined genotypes Ab or aB. Recombination between widely spaced markers B and Z is more common, producing many gametes with bZ or Bz genotypes. By genotyping the members of a pedigree, the pattern of recombination can be deduced, allowing the distances between markers to be estimated.

linkage studies seldom provide such complete data, for example because not all members of the pedigree are available for genotyping or heterozygous at the appropriate loci. A more approximate method must therefore be used, in which various possible intermarker distances are modelled, to find the distance that is most likely to give rise to the observed pattern of segregation. Once the distances between all possible pairs of markers have been estimated in this way, the arrangement of all the markers is found that best agrees with the complete set of pairwise distance estimates. Because the number of possible marker orders increases as the factorial of the number of markers (for example, $3 \times 2 \times 1 = 6$ possible orders for three markers, but $4 \times 3 \times 2 \times 1 = 24$ for four markers, and so on), rigorous analysis is seldom possible, and some degree of compromise or approximation is usually required.

Linkage data are most powerful when all markers have been analysed in a common set of individuals. Only in this way can all of the pairwise distances between markers be estimated. For example, if markers A, B and C are mapped on one pedigree, and markers C, D and E on another, then the distances A–D, A–E, B–D and B–E cannot be directly inferred. The CEPH panel (Centre d'Etudes du Polymorphisme Humain) is a collection of DNA samples from large multigeneration families, available to researchers worldwide. By typing polymorphic markers on this common resource, the resulting linkage data can be integrated effectively. Equivalent shared resources have been established for many other species.

Genetic linkage mapping is a powerful method, but suffers from some limitations. In particular, only polymorphic markers can be mapped, and the resolution that can be achieved is limited. In humans, recombination occurs only every 100 Mb or so on average, limiting the resolution to around 1 Mb if a few hundred individuals are genotyped. Finally, it has been found that certain small regions of genomes (including the human) are more prone to meiotic recombination than others. Markers situated on either side of such 'recombination hotspots' will recombine more often than would otherwise be the case, leading to an overestimation of the distance between them.

Radiation hybrid mapping

In the mid-1970s, Stephen Goss and Henry Harris discovered that, if cultured human cells were subjected to high dose of radiation (enough to fragment their

chromosomes) and then fused to unirradiated hamster cells, hybrids would be produced. Such 'radiation hybrids' would initially contain many fragments of human chromosomes (in addition to the complete hamster genome), and hence would express many human proteins. Over time, however, the unstable human fragments would be lost (and with them, the human proteins) until only a few were stably retained. If, in many such hybrids, two human proteins were often retained together, it was likely that their genes lay on the same human chromosome fragment. In this way, the arrangements of genes on the human chromosomes could be deduced.

This method was little exploited until the late 1980s when it was updated using modern molecular genetic techniques by researchers including David Cox, Mike Walter and Peter Goodfellow. In its current form, hybrids are formed in essentially the same way between irradiated donor cells and unirradiated host cells. Hybrids containing donor chromosome fragments are then identified using selectable markers (for example, drug resistance genes) from each species. Many such hybrids are isolated and cultured, to give a mapping panel of perhaps a hundred distinct hybrid cell lines, each containing a different set of donor fragments.

DNA is prepared from each cell line in the panel, and is screened to determine which donor-derived genome markers it contains. The markers are usually STSs (including ESTs), which can easily be detected using PCR. In this way a table of results is produced showing which markers are present in each member of the mapping panel. As in Goss and Harris' original method, two markers that are frequently found together in the same hybrid (i.e. which cosegregate often) can be inferred to lie close together in the donor genome. In this way, the order and distance of markers in the genome can be inferred.

Radiation hybrid (or RH) mapping is closely analogous to genetic linkage mapping. Instead of recombination, it is the radiation that randomly breaks the chromosomes; hybrid cell formation replaces meiotic segregation as a way to shuffle and segregate these fragments. As with linkage mapping, cosegregation frequencies reflect the proximity of markers to one another. Therefore, the analytical techniques used to construct the maps are fundamentally similar and, with minor modifications, interchangeable.

RH mapping differs from linkage mapping in several important ways, however. Because the donor fragments segregate directly into the hybrids (rather than being exchanged between homologous chromosomes as in linkage mapping), the method is not restricted to polymorphic markers. This means that almost any STS can be mapped. Moreover, the resolution afforded by RH mapping is far higher than that attainable by linkage mapping. This resolution depends only upon how finely the donor chromosomes are broken at the outset, which in turn can be controlled by varying the dose of radiation. If the dose is low, the donor fragments are large and therefore the resolution is poor. A higher dose will break the chromosomes into smaller fragments, allowing the pattern of segregation to be observed on a far smaller scale and permitting finer maps to be made. Resolutions down to a few tens of kilobases can be achieved in this way.

RH mapping, nonetheless, suffers from a few limitations. First, the production and growth of suitable hybrids is far from easy, and has not yet been possible for a wide variety of species. Second, reliable RH mapping requires that the radiation-induced breaks and the segregation of donor chromosome fragments amongst the hybrids should be uniform for all parts of the donor genome. In fact, this is known not to be the case: for example, donor fragments that contain centromeres are more likely to be stably retained in the hybrid cells than those that do not. Donor telomeres and certain genes can also bias the retention of fragments. These biases can lead to distortions and errors in the resulting maps. Third, problems can arise in mapping markers that are closely homologous between the host and donor genomes (as is often the case with ESTs derived from conserved genes). In such cases, it may be difficult to distinguish the donor marker in the hybrid from the host's counterpart, complicating the analysis.

Despite these shortcomings, however, RH mapping is an invaluable tool in human and in many other vertebrate species.

HAPPY mapping

RH mapping overcomes many of the limitations of genetic linkage mapping. However, its few remaining problems (difficulties in making hybrids, distortions arising from the effects of donor centromeres and other sequences, and difficulty in distinguishing some donor markers from host sequences) all arise from the *in vivo* step – the use of hybrid cells to propagate the donor fragments. HAPPY mapping overcomes these limitations, being an entirely *in vitro* technique. Indeed, it can be thought of as 'radiation hybrid mapping without radiation hybrids'.

In HAPPY mapping, DNA is first prepared from the species whose genome is to be mapped. To prevent unwanted mechanical breakage to the naked, fragile DNA molecules, this is normally done by embedding the living cells in agarose gel, and then treating them with a solution of detergents and proteases. These diffuse into the agarose, lysing the cells and stripping away proteins and other cellular debris, which diffuse out of the agarose; the long molecules of chromosomal DNA remain trapped and protected within the agarose.

The DNA is now broken randomly, either by melting the agarose and mechanically shearing the DNA solution, or by treating the embedded DNA with gamma radiation (much as in the initial stages of RH mapping). The result is a pool of random DNA fragments, whose average size

depends upon the violence of the mechanical shearing or on the dose of radiation used.

This broken DNA is then diluted to a very low concentration, and about one hundred samples are dispensed into separate tubes. The complete set of samples is referred to as a 'mapping panel'. Most importantly, each sample is exceedingly small, containing less than one genome's worth of DNA fragments. For example, a mammalian genome contains about 3 pg (3×10^{-12} g) of DNA per haploid copy; in this case, each sample would contain only 1–2 pg of DNA fragments.

These minuscule samples are screened by PCR, to determine which markers (STSs or ESTs) are present in each one. (The fact that roughly *hap*loid samples of DNA are screened using the *po*ly*m*erase chain reaction gives the method its name.) Because the samples are so small, each one will contain only a randomly sampled subset of the markers, rather than the complete genome. Hence, any particular marker will be present in some, but not all of the members of the mapping panel. If two markers are close together in the genome (compared to the average size of the random fragments), then they will seldom be broken apart and hence will tend to be found together in the same members of the mapping panel – they cosegregate. As the distance between markers increases, it is more likely that the random breakage will separate them, so that they lie on separate DNA fragments and are hence less likely to cosegregate. If the markers are very far apart, they will always be broken apart during the random breakage, and hence will show no particular tendency to cosegregate.

As in RH mapping, statistical analysis of the cosegregation frequencies between markers allows the distance between them to be estimated. Once all the pairwise distances are known, the order and spacing of the markers in the genome can be deduced. The analytical procedures are effectively identical to those used in RH mapping, and the same software packages can be used to analyse either type of data.

The resolution that can be attained depends only upon how finely the DNA is broken at the outset. Since naked DNA can be broken into fragments as small as a few hundred base pairs, the resolution of this method can be very high, and maps with resolutions of a few kilobases have been produced.

The *in vitro* nature of HAPPY mapping eliminates some of the potential limitations of RH maps. Since no cell culture or hybrid formation is required, the method can be applied to any genome. The segregation of DNA fragments is done by simple dilution, and is therefore not affected by centromeres, telomeres or other biologically active sequences. Since there is no 'host genome', the samples can be screened for markers without interference from homologous host sequences. As a result, HAPPY maps are relatively immune to artefacts.

The *in vitro* nature of HAPPY mapping, however, also imposes certain limitations of its own. The first is that it is difficult to prepare DNA fragments of more than a few megabases in size. This in turn means that intermarker distances of more than a megabase or so cannot be measured, making it unsuitable for producing very low-resolution maps in which the markers are widely spaced. (In contrast, most other techniques are restricted in the maximum resolution they can attain.) The second limitation is that the DNA samples in the mapping panel are exceedingly small, yet must be screened (by PCR) for all of the markers to be mapped. In the analogous stage of RH mapping, the hybrid cells are cultured to produce large quantities of DNA before marker screening begins. In HAPPY mapping, the DNA samples cannot be 'grown' in the same way. Instead, *in vitro* techniques must be used to 'preamplify' the minute samples, replicating many-fold any marker sequences which they contain. Only after this preamplification can the samples be treated as a bulk resource, and screened in turn for many markers. Several techniques are available for the preamplification step, including PEP (primer-extension preamplification) and ligation-mediated whole-genome PCR. This is the most critical step in HAPPY mapping, as the loss of any markers during preamplification will make them unmappable. Nevertheless, HAPPY mapping has proved effective in making high-resolution maps in a wide variety of species.

Physical mapping

Physical mapping involves finding a contiguous series (or 'contig') of cloned DNA fragments which contain overlapping portions of the genome. The overlaps define the positions of the clones relative to one another. If at least some of the clones contain markers that have been independently mapped by other means, then the position of the entire contig in the genome is also known.

The starting point for physical mapping is a library of cloned genomic fragments, normally prepared by either random mechanical breakage or partial restriction digestion of genomic DNA. (Complete restriction digestion of the DNA would mean that no clones could be found that overlapped across the restriction sites for that enzyme.) The fragments are usually cloned in bacterial hosts (normally *Escherichia coli*), using bacteriophage, cosmid, plasmid or other vector systems. For physical mapping of large genomes, it is desirable to use clones containing large inserts, such as P1 artificial chromosomes (PACs) or bacterial artificial chromosomes (BACs) which can carry inserts in excess of 100 kb. Even larger fragments (over 1 Mb) can be cloned in yeast (*Saccharomyces cerevisiae*) using yeast artificial chromosome (YAC) vectors; however, such clones are often unstable, undergoing deletions or internal rearrangements.

Once the library is established, overlapping clones can be identified using several different approaches. In STS content mapping, the library is screened to identify all

clones that contain a specific STS marker. Screening can be done either by PCR using the appropriate primers (in which case only clones containing the STS will give a PCR product); or by arraying the clones on a nylon membrane, lysing them to expose their DNA, and probing the membrane with labelled DNA corresponding to the STS in question. In either case, all clones containing a given STS marker are identified. If two clones are found to contain the same STS, then they must represent overlapping parts of the genome (with the STS lying in the region of the overlap). If one clone contains two different STSs, then those two STSs must be consecutive in the genome. If enough clones are characterized for enough STSs, then a contiguous overlapping path of clones is defined.

A second common approach to physical mapping is by the use of restriction fingerprinting. In this case, each member of the clone library is digested in turn with one or more restriction enzymes, and the sizes of the resulting fragments are measured by gel electrophoresis. If two clones are found that have several fragment sizes in common, then they must represent overlapping parts of the genome, with the shared fragments coming from the region of overlap. Again, if enough clones are fingerprinted then the hope is to find a contiguous series of overlapping clones.

Whatever the means used to make them, a variety of problems and artefacts can afflict clone contigs. Many clones must be screened to find ones that cover as much of the genome as possible. As more clones are screened, the number of contigs at first increases (since most new clones do not overlap with earlier ones, and hence form independent contigs), and then declines as more clones are found that overlap and link the earlier ones. Typically, a coverage of between 5- and 10-fold is required, meaning that the total DNA content of the screened clones must be 5- to 10-fold greater than the genome to be covered. (For example, if a genome of 100 Mb is cloned in 100 kb fragments, then 5000 clones would give 5-fold coverage.) Even then, there are likely to remain some gaps in the contig, corresponding to regions of the genome that resist cloning in the particular host/vector system used. Such gaps can often be filled by using clones from a different library (made using a different host/vector system). A process known as chromosome walking is often used to help close gaps: sequences are obtained from the ends of clones flanking the gap, and are used as probes to specifically identify new clones that extend into or across the gap.

Physical mapping is also vulnerable to a variety of clone artefacts. Chimaeric clones (that is, clones in which two unrelated pieces of DNA from different parts of the genome have become ligated together) can cause the contig to 'jump' from one part of the genome to another. Clones that have suffered internal deletions or rearrangements can also cause errors. Finally, sequences that are duplicated in the genome can disrupt contig assembly, since two clones that contain different copies of the repeated motif may wrongly be assumed to overlap.

Despite its limitations, however, physical mapping has the unique advantage that, as well as producing a map giving the relative locations of the clones (and of any STS markers they are known to contain), it also produces a tangible resource – the clones themselves – representing known segments of the genome. These mapped clones can be returned to for more detailed analysis, or can serve as the substrate for genome sequencing.

Fluorescence *in situ* hybridization

All of the preceding methods of genome mapping are indirect: the order of markers in the genome must be inferred by analysing the marker content of meiotic products, hybrid cells, DNA samples or cloned DNA fragments. In contrast, fluorescence *in situ* hybridization (FISH) is an elegantly direct way of observing the physical arrangement of markers along the chromosomes. Fluorescently labelled probes are hybridized to metaphase chromosomes on a glass slide, and their position observed using a fluorescence microscope.

The first requirement for conventional FISH is a population of cells in which a high proportion are in metaphase, the stage in the cell cycle where the chromosomes are condensed as distinct bodies. This is typically achieved by culturing lymphocytes or fibroblasts in the presence of a mitogen such as phytohaemagglutinin, then treating them with colcemid to block the cell cycle at metaphase. Treatment with bromodeoxyuridine (which is incorporated into the replicating chromosomes) and ethidium bromide (which intercalates into the DNA) improves the morphology of the chromosomes for FISH. The cells are swollen in a hypotonic salt solution and stabilized with a fixative. Drops of the cell suspension are placed on glass slides, to which the chromosomes adhere. Their DNA is denatured, and is then ready to be hybridized to the probe.

Probes are normally cloned fragments of genomic DNA or cDNA, which have been labelled (for example, using nick-translation) by the incorporation of a hapten such as biotin or digoxigenin. They are denatured, allowed to hybridize to the corresponding sequences of the metaphase chromosomes, and the surplus washed away. Nonspecific hybridization is normally suppressed by using an unlabelled competitor such as total genomic DNA. To detect the probes, the slide is incubated with fluorescently labelled proteins (such as avidin or anti-digoxigenin antibodies) which bind to the hapten in the probe and render it visible in the fluorescence microscope. At the same time, the chromosomes themselves can be visualized by a fluorescent counterstain, such as DAPI (4′,6′-diamidino-2-phenylindole), of a different colour to the fluorophore used to detect the probe.

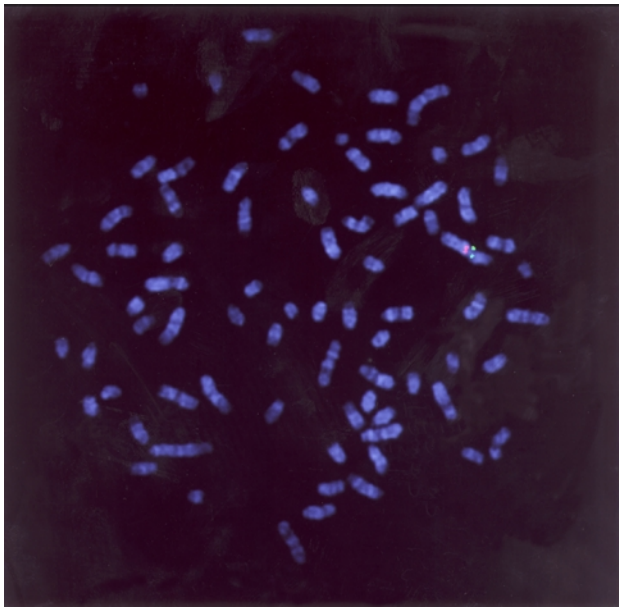

Figure 3 Dual-colour fluorescence *in situ* hybridization. The image shows a spread of male canine metaphase chromosomes, probed with two plasmid clones which hybridize to the X-chromosome (towards the right of image). One clone has been visualized using a red fluorophore (Texas red), the other with green (FITC). Chromosomes have been counterstained with DAPI to reveal their characteristic banding pattern. Each probe produces two adjacent spots, corresponding to the two sister chromatids. The distance between the probes is approximately 20 Mb. Picture courtesy of Dr H. F. Spriggs.

The resulting images clearly reveal the location of the bound probe, typically as two adjacent spots corresponding to the two chromatids (**Figure 3**). The characteristic banding pattern of the chromosomes gives the absolute location of the probe in the genome. By using different combinations of haptens and fluorophores, two or more probes may be hybridized and detected simultaneously. Such multicolour FISH allows the order and spacing of the probes to be determined more accurately than is possible with independent, single-probe experiments.

Metaphase chromosomes are relatively condensed (typically containing about 10 Mb of DNA per micrometre of their length), limiting the resolution of FISH to about 1 Mb. However, hybridization of probes to interphase nuclei (in which the chromatin is less condensed) provides considerably higher resolution, although there is no longer a visible banding pattern to provide an absolute chromosomal location. Chromatin fibres released from interphase nuclei are more extended still, and hybridization to these (known as 'fibre FISH') can allow the relative positioning of probes to within a few kilobases.

Map Integration

Each of the above approaches to genome mapping has its own strengths and weaknesses, and the same genome will often be mapped by several groups using different techniques. One of the greatest problems in genome analysis, therefore, is the integration of these disparate maps to produce a unified picture of the genome.

Alignment of different maps of the same region is hampered in the first instance because different markers will usually have been placed on each map. Even where there are some markers in common, placement of the remaining, unshared markers can be problematic. For example, one map may show markers A-X-B in that order, while another shows markers A-Y-B. Should the integrated map be A-X-Y-B or A-Y-X-B?

Even when most of the markers are common to two maps, problems remain because the accuracy of different maps varies, as does the nature of the errors likely to be present. For this reason, it is common to find that maps disagree over the local (or sometimes not so local) order and spacing of those markers that they share. Resolution of these conflicts requires a detailed consideration of the strengths and weaknesses of each type of map.

The problem of effective map integration has not been fully addressed, and researchers interested in a particular region of the genome will often have to use their best judgement to make sense of the conflicting maps of that area. Nevertheless, attempts have been made to present and unify maps as far as possible; a notable example in the human genome is the Genome Database (see Further Reading).

Further Reading

Coulson A and Sulston J (1988) Genome mapping by restriction fingerprinting. In: Davies KE (ed.) *Genome Analysis – A Practical Approach*, pp. 19–39. Oxford: IRL Press.

Dear PH (ed.) (1997) *Genome Mapping – A Practical Approach*. Oxford: IRL Press.

Dear PH and Cook PR (1993) HAPPY mapping: linkage mapping using a physical analogue of meiosis. *Nucleic Acids Research* **21**: 13–20.

Genome Database. [http://www.gdb.org/]

McCarthy LC (1996) Whole genome radiation hybrid mapping. *Trends in Genetics* **12**: 491–493.

Ott J (1986) A short guide to linkage analysis. In: Davies KE (ed.) *Human Genetic Disease – A Practical Approach*, pp. 19–32. Oxford: IRL Press.

Strachan T and Read AP (1999) *Human Molecular Genetics*, 2nd edn, pp. 244–250. Oxford: BIOS Scientific.