Modern Linkage Mapping

with polymorphic DNA markers—a tool for finding genes

Problem: In "Classical Linkage Mapping" we showed how to construct maps that give the order of and genetic distances between gene pairs for variable, single-gene traits that are linked (lie on the same homologous chromosome pair). Prominent among the variable, single-gene traits of humans are inherited diseases. Several thousand such genetic disorders have been identified, and many of the genes for those disorders were mapped through classical linkage analysis. However, the maps included no reference to the physical reality of DNA, and therefore they did not provide the information necessary to isolate a segment of DNA containing a disease-causing gene. Then, in 1980, David Botstein, Raymond L. White, Mark Skolnick, and Ronald W. Davis transformed linkage mapping into a tool for finding genes.

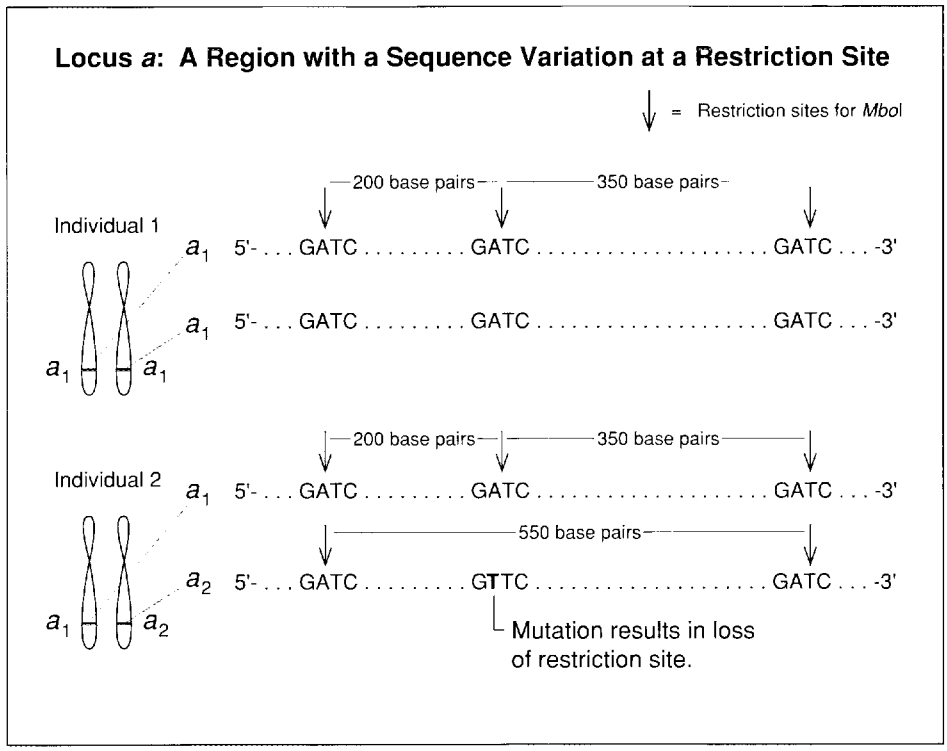
The Botstein Idea: If we could compare the base sequences of corresponding regions of the DNA from several individuals, we would find many regions with identical sequences—but we would also find many regions where the base sequence varies slightly from one individual to another. Those variable regions are called DNA polymorphisms. Now suppose we have available DNA probes that can not only reveal the presence of variable regions but also distinguish one sequence variation from another. Suppose further that some of the variable regions are fairly stable, so that a given sequence within such a region is transmitted from one generation to the next. In other words, each variable region exhibits only a limited number of sequence variations among the population. Such a variable region, together with the DNA probe that detects the sequence variations within that region, is called a polymorphic DNA marker.

Polymorphic DNA markers are very useful for several reasons. First, because they are variable, we can construct a linkage map of DNA markers just as we construct a linkage map of the genes that determine variable phenotypic traits. That is, we trace the co-inheritance of pairs of DNA markers to determine the genetic distances between them. Second, we can trace the co-inheritance of a marker and a variable phenotypic trait to determine the genetic distance between the marker and the gene responsible for the variable phenotypic trait. Finally, we can use the DNA probe for a marker to find the physical location of the marker on a chromosome. The physical loci of the polymorphic DNA markers can then serve as landmarks in the search for a specific gene. For example, if we know from the linkage map that a gene for a particular phenotypic trait lies between two particular DNA markers, then the gene of interest can be found in the stretch of DNA connecting the physical loci of the two markers. In summary, DNA markers provide a way to connect loci on linkage maps with physical loci in the human genome, which in turn, provides a way to find genes of interest.

Question: What is an example of a base-sequence variation within a region that can turn the region into a DNA marker?

Answer: The base-sequence variation within a region must be easily detectable to make the region a candidate for a DNA marker. One type of detectable variation is a single base change that results in the creation or loss of a restriction-enzyme cutting site. Such sites are short sequences, four to eight base pairs in length, at which a restriction enzyme cuts a DNA molecule. For example, each cutting site for the restriction enzyme *Mbo*I has the base sequence 5′-GATC.

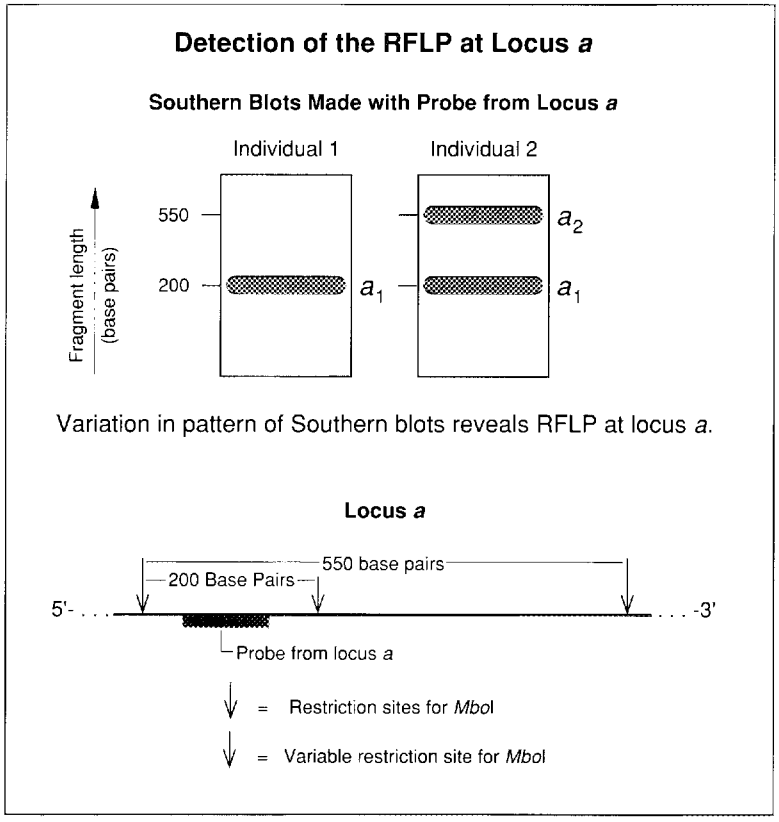
Example: Consider locus *a*, a variable region on a particular pair of of homologous chromosomes. The figure shows the DNA segments that compose locus *a* in the homologous chromosome pairs of two individuals. Also shown are the positions of the cutting, or restriction, sites for the restriction enzyme *Mbo*I within locus *a* and the distance between successive sites. Individual 1 carries two copies of a_1, a version, or allele, of locus *a* that has three restriction sites for *Mbo*I. Individual 2 carries one copy of a_1 and also a copy of another allele, a_2. Note that a_2 is missing the middle restriction site present in a_1. The absence of that restriction site is due to a change in a single base pair (shown in red). If *Mbo*I is allowed to cut the DNA from these two individuals, a_1 will be cut into two fragments of lengths 200 base pairs and 350 base pairs, whereas a_2 will be cut into one fragment of length 550 base pairs.



Question: How do we detect which alleles of locus* a *are present in the DNA molecules of two individuals?

Answer: We measure the lengths of the fragments from locus *a* produced by cutting the DNA with *Mbo*I and note the differences between the lengths of the fragments from the two individuals. We do so by making a Southern blot (see "Hybridization" in "Understanding Inheritance"). We begin by extracting many copies of the DNA from the blood cells of each individual. We then chop up, or digest, the DNA in each sample with the restriction enzyme *Mbo*I. The next step is to separate the resulting fragments (called restriction fragments) according to length by gel electrophoresis (see "Gel Electrophoresis" in "Understanding Inheritance"). Because shorter fragments travel farther through the gel than longer fragments, the lengths of the fragments can be determined from their final positions on the gel. We then transfer (blot) the fragments onto a filter paper in a manner that preserves their final gel positions.

Next, we allow a radioactively labeled DNA probe from locus *a* to hybridize, or bind by complementary base pairing, to the restriction fragments. The probe hybridizes only to fragments from locus *a* and thereby reveals their positions and therefore their lengths. Finally, we make an autoradiogram of the filter paper in which the positions of the fragments that have hybridized to the probe are imaged as dark bands.



Example: The figure shows Southern blots for the DNA of individuals 1 and 2 made with the enzyme *Mbo*I and a probe for locus *a*. The position of the probe is shown in the diagram of locus *a*. That particular probe binds to the restriction fragments of length 200 base pairs from allele a_1 and to the restriction fragments of length 550 base pairs from allele a_2. Since individual 1 carries allele a_1 only, the Southern blot of individual 1 shows one band at a position corresponding to a length of 200 base pairs. Individual 2 carries alleles a_1 and a_2 and therefore has a Southern blot showing two bands, one at 200 base pairs and one at 550 base pairs. The variation within locus *a* that causes this difference between the two Southern blots (the presence or absence of a restriction site) is called a restriction fragment length polymorphism, or RFLP, which is one type of polymorphic DNA marker. (Another type of polymorphic DNA marker is described in "The Polymerase Chain Reaction and Sequence-tagged Sites.")

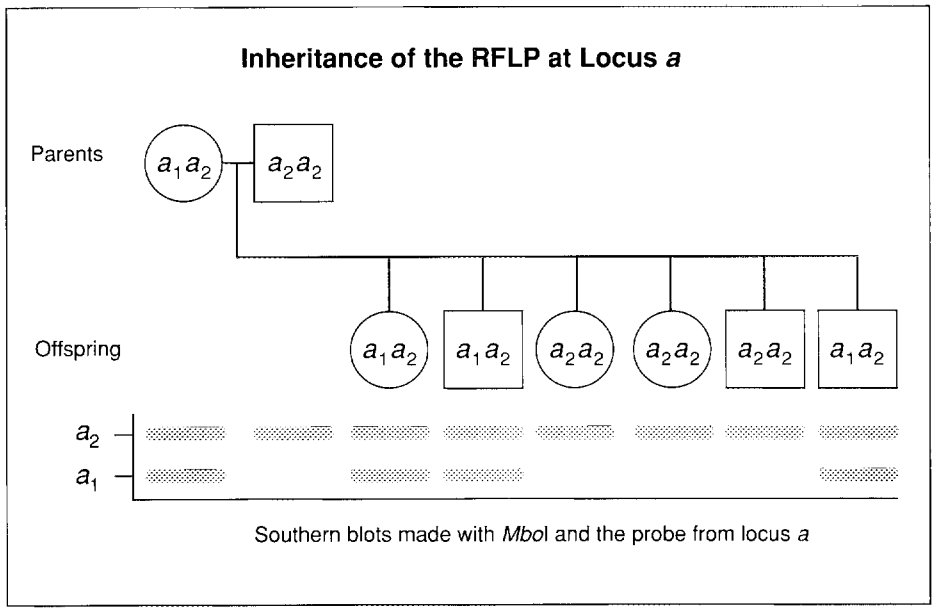
Question: How do we find polymorphic DNA markers?

Answer: Originally, this was done by a process involving patience and preferably luck. We randomly choose one clone from a collection of human DNA clones, use it as a probe in the making of Southern blots of the DNA of many individuals, and see whether the Southern blots vary from one individual to the next. A variation implies that the probe is part of a variable region of the genome and therefore defines that region as a polymorphic DNA marker. If the clone chosen does not reveal a difference, we continue choosing clones until a difference does show up. More recently, with the wide application of the polymerase chain reaction (PCR) and the discovery that there are a large number of highly variable, short di-, tri-, and tetranucleotide repeat sequences flanked by unique DNA sequences, it has become possible to select such regions of DNA and then develop them into highly polymorphic markers.

Question: How are polymorphic DNA markers used in linkage analysis?

Answer: In linkage analysis a polymorphic DNA marker is analogous to a gene that has two or more alleles. Each parent carries a pair of alleles of the marker, one on each member of a chromosome pair, so each parent may be either homozygous or heterozygous for the marker. Also, each parent transmits only one allele of the marker to each offspring.

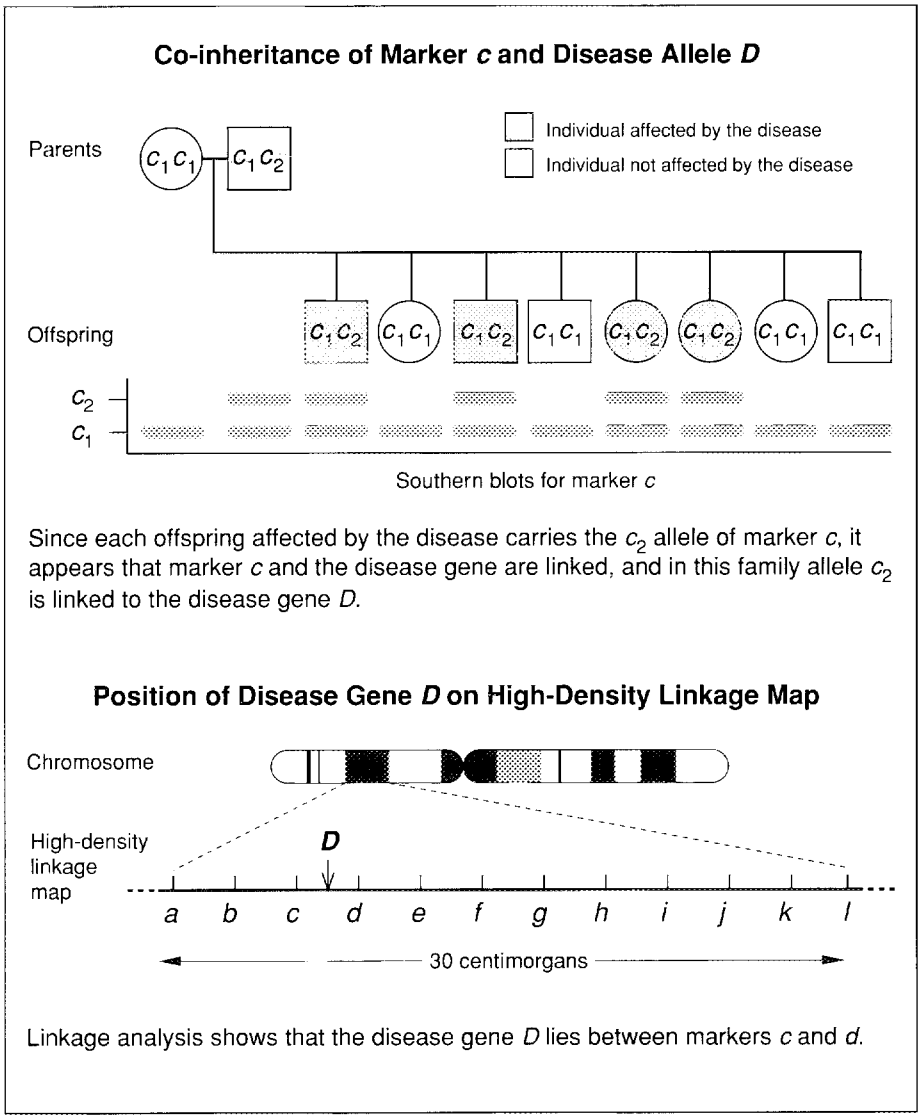
Example: The figure at right shows an example of the inheritance of the RFLP at locus *a*. Beneath each parent and each of their six children is shown the Southern blot for the marker. The father is heterozygous for the marker, carrying alleles a_1 and a_2. Among the offspring three are heterozygous and three are homozygous for a_2. The heterozygous offspring have inherited the allele a_1 from their father. Note that the alleles of a polymorphic DNA marker are inherently easier to trace than the alleles of a gene because the alleles of a polymorphic DNA marker are codominant. That is, none of them are recessive and each is directly observable.



We can also trace the inheritance of two markers, find out whether they are linked (on the same chromosome), and determine the recombination fraction for the two markers and thus the genetic distance between their loci. The linkage analysis exactly parallels that described for phenotypic traits in "Classical Linkage Mapping." In particular, an informative mating, one that reveals linkage between a pair of markers, must involve a parent who is heterozygous for both markers.

Question: Why does the Genome Project have as one of its top priorities the construction of a high-density linkage map of polymorphic DNA markers?

Answer: By 1996 the Genome Project hopes to have produced a set of linkage maps, each containing polymorphic DNA markers spaced along each human chromosome at intervals of 2 to 5 centimorgans, genetic distances that roughly correspond to physical distances of 2 to 5 million base pairs of DNA. Such a set of maps will enable researchers to find any gene of interest relative to the loci of approximately 1500 markers. In other words, the markers will form a set of reference points along the genome.

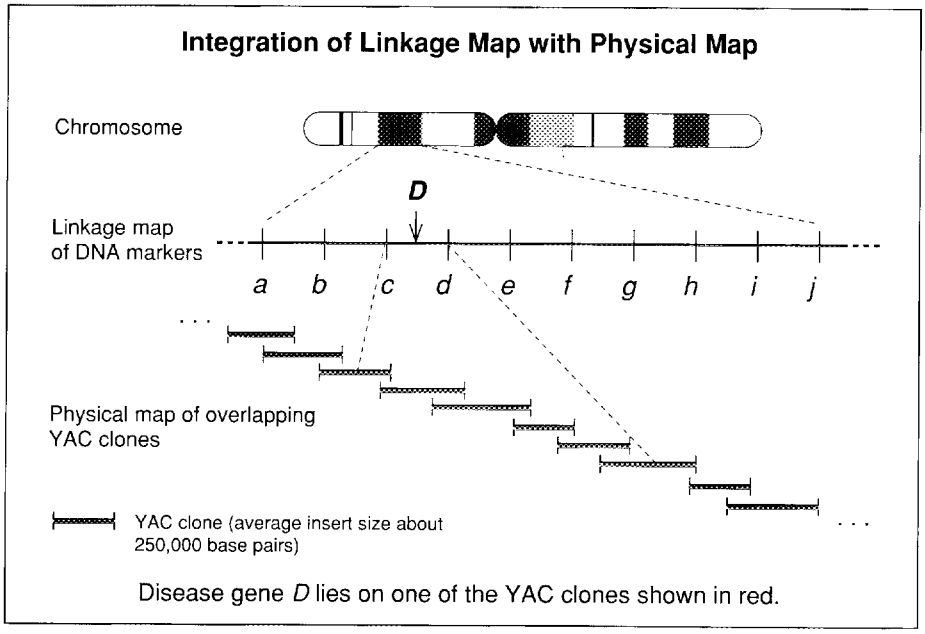


Example: Suppose we are interested in locating a mutant gene D that causes an inherited disease. We can find families affected by the disease and trace the co-inheritance of the disease with the reference markers on a linkage map. If we have a 2-centimorgan linkage map of highly informative markers (see "Informativeness and Polymorphic DNA Markers"), we can find markers flanking the gene that are less than 2 centimorgans away on either side. The pedigree in the figure shows the type of data needed to establish that the marker c and the disease gene D are tightly linked, that is, c and D are so close together that recombination events between them are rarely observed. Similar data between marker d and D would allow us to infer that D lies between c and d, as indicated in the lower part of the figure. This example shows the characteristic pattern of inheritance of an autosomal dominant disorder identified by allele c_2 of marker c.

Question: Once we have found DNA markers flanking a disease gene, how do we localize the disease gene on the DNA itself?

Answer: In addition to creating a linkage map of polymorphic DNA markers, the Genome Project is creating a physical map for each human chromosome. A physical map consists of an ordered set of overlapping cloned fragments that spans the entire length of the DNA molecule in the chromosome. As the physical maps and the linkage maps are constructed, the linkage map for each chromosome is being integrated with the physical map for that chromosome. That is, each locus on the linkage map will be associated with a locus on the physical map. Thus, if we find two markers that flank a disease gene, we will be able to ascertain how many base pairs of DNA separate the markers, and we will also have all that DNA available as cloned fragments. We therefore know that the disease gene is in one of those cloned fragments, and we can employ various methods to find the DNA segment that contains the gene. (Those methods are not necessarily straightforward, as explained on pages 111 and 142 of "Mapping the Genome.")

Example: The figure at right shows a schematic representation of a human metaphase chromosome (dark bands indicate A-T rich regions), a portion of a linkage map of polymorphic DNA markers, the position of a disease gene *D* on that map (as determined by linkage analysis), and the corresponding physical map of cloned fragments. Dotted lines connect the loci on the linkage map with the corresponding loci on the physical map and on the metaphase chromosome. Highlighted in red are the clones that must be searched to find the disease gene.



CAVEAT: In practice we need flanking markers that are within 1 centimorgan of the gene on either side so that the search for the disease gene will involve no more than about 2 million base pairs of DNA. Consequently, the long-term goal of the Genome Project is to find enough highly polymorphic DNA markers so that they are spaced at intervals of 1 centimorgan on the linkage maps, or a total of about 3300 markers. If they are found by a random search, we will have to find about ten times that number to achieve the 1-centimorgan map. The search for markers has been accelerated in several ways. For example, new types of markers are being systematically sought (see pages 133–134 in "The Polymerase Chain Reaction and Sequence-tagged Sites"), and automated techniques are being developed to detect DNA markers in large numbers of individuals. ■