Am. J. Hum. Genet. 42:143–159, 1988

Identification of More than 500 RFLPs by Screening Random Genomic Clones

James W. Schumm, Robert G. Knowlton, Jeffrey C. Braman, David F. Barker, David Botstein, Gita Akots, Valerie A. Brown, Thomas C. Gravius, Cynthia Helms, Kathy Hsiao, Kenneth Rediker, Joan G. Thurston, and Helen Donis-Keller

Department of Human Genetics, Collaborative Research, Inc., Bedford, MA

Summary

As part of our genome-mapping effort, we undertook a large-scale screening study to identify RFLPs useful as genetic markers. Some 1,664 single-copy or repeat-containing phage clones from a Charon 4A genomic library were tested for polymorphism against a panel of DNAs, from five unrelated individuals, digested with eight restriction enzymes. Approximately 30% (515) of the clones revealed polymorphism by Southern hybridization; 67 loci detected had PIC values >.5. Restriction enzymes *Msp*I, *Taq*I, and *Rsa*I were most efficient in detecting polymorphism within the 1–20-kb-fragment size range resolved. With only one exception each of the clones detected polymorphism originating from a single locus.

Introduction

Construction of a human genetic linkage map will greatly facilitate the identification and characterization of chromosomal loci involved in inherited disorders. Using an ordered, evenly spaced array of genetic markers to test for linkage to an unmapped inherited trait will generally be much more efficient than a trial-and-error approach with random markers. Furthermore, a genetic map will be essential to identify loci of diseases with complex modes of inheritance; for example, diseases that are heterogeneous (i.e., caused by any of several independent loci) or multifactorial can probably be mapped only by methods in which markers covering the entire genome are tested simultaneously for coinheritance with the disease (Lander and Botstein 1986*a*, 1986*b*).

The genetic markers constituting the linkage map are likely to be RFLPs, because of the number of such polymorphic loci in the genome and the high probability that informative RFLP markers occur in all families. The completion of a human genetic linkage map of polymorphic DNA markers as foreseen by Solomon and Bodmer (1979) and Botstein et al. (1980) depends on the concerted development of three resources. First, there must be a large collection of cloned DNA probes to identify polymorphic loci whose segregation can be followed in linkage studies. Second, there must be DNA from individual family members of large human pedigrees of a structure suitable for scoring meiotic segregation. Third, there must exist suitable methods of data collection and analysis to detect and to quantify the linkage relationships among the polymorphic loci. There has been sufficient progress in all three areas to demonstrate that comprehensive linkage mapping of all the human chromosomes with RFLP markers is a realistic goal (Skolnick 1980; Lathrop et al. 1985; White et al. 1985; Willard et al. 1985). The progress in human linkage mapping so far is especially encouraging given that the markers mapped are generally not informative in most families and that methods of multilocus linkage analysis have not been optimal for mapping the number of markers involved. Nevertheless, genetic linkage maps of the X chromosome and portions of several autosomes have been con-

Received November 17, 1986; revision received July 10, 1987.

Address for correspondence and reprints: Helen Donis-Keller, Ph.D., Department of Human Genetics, Collaborative Research, Inc., Two Oak Park, Bedford, MA 01730.

0002-9297/88/4201-0020$02.00

structed using large human pedigrees and computer programs that should be equally applicable to linkage analysis of polymorphic loci on other chromosomes (Drayna and White 1985; Kittur et al. 1985; Leach et al. 1986). The recent finding of more than 60 chromosome 7 RFLPs (Donis-Keller et al. 1986; Barker et al., in press) prompted the development of new multipoint data-analysis techniques (Lander and Green 1987) in order to refine a two-point chromosome 7 linkage map constructed using the RFLPs and traditional data-analysis methods (Ott 1974; Drayna and White 1985; Kittur et al. 1985; Lathrop et al. 1985; Donis-Keller et al. 1986; Leach et al. 1986; Lander and Green 1987). With continued improvement in methods of data analysis (Lander and Green 1987), the limiting resource in completing the human linkage map is therefore a sufficiently large set of DNA probes for polymorphic loci to cover the entire genome.

The number of RFLP markers required depends on the average map interval between loci and the intended application of the markers placed on the map. Both of these variables depend in turn on the extent of polymorphism of the loci being mapped, i.e., their PIC. The PIC is a measure of the probability that informative alleles are segregating in a family (Botstein et al. 1980). The more informative a locus is, the greater the map distance at which linkage with a second locus can be reliably measured. In addition, loci with higher PIC values are informative in a greater fraction of families and are therefore more likely to be useful in mapping other traits in different human pedigrees. Given a total of 3,300 cM in the human genome, a set of 150 markers is the minimum estimate for a linkage map consisting of fully informative loci, evenly spaced at ~22-cM intervals (Botstein et al. 1980). Constructing an equivalent genetic map with fewer polymorphic markers would require a greater number of families and a greater number of loci. With only two-allele polymorphisms (maximum PIC = .375) instead of fully informative markers, for example, a comparable map would require approximately twice the number of evenly spaced RFLP loci and 40% more families for analysis (Lander and Botstein 1986*a*). Thus, the distribution of PIC values of RFLP probes available and the acceptable level of efficiency in determining map distances on the basis of family inheritance data will determine the number of probes required.

To obtain the desired number of evenly spaced markers, it will be necessary to map many more random RFLP loci to approach saturation of all parts of the genome. To construct a map with a sufficient density of markers such that, with 95% probability, any locus falls within 10 cM of a RFLP marker, the required number of probes has been calculated to be between 500 and 1,700 (Bishop and Gardner 1980; Lange and Boehnke 1982; Ott 1985, pp. 56–59). In comparison with the number of markers needed to complete the genetic map, the number of polymorphic markers discovered in the human genome is relatively small. An increasing number of RFLPs detected with cloned DNA probes have been reported (Willard et al. 1985), but because of the diverse origins of these probes it is difficult to discern (1) the rate at which polymorphisms can be detected and, especially, (2) the relative frequency of RFLPs with high PIC values useful for mapping.

We report here the results of a large-scale program to screen genomic DNA clones for probes hybridizing to polymorphic loci. In this process we have identified more than 500 RFLP markers, a substantial fraction of the loci needed to construct a human genetic map. Because all of the probes tested were randomly isolated and tested by a standard procedure, we have also made a reliable measurement of the frequency and the distribution of PIC values of RFLPs detected using this method.

Material and Methods

Bacteriophage Probes

Human genomic clones were isolated from the bacteriophage Charon 4A library of Lawn et al. (1978) by selecting individual plaques on a lawn of *E. coli* CGE6. Phages putatively containing single-copy sequence inserts were isolated in two steps. The first step was plaque hybridization (Benton and Davis 1977) using nick-translated human placental DNA as probe according to the method described by Wyman and White (1977). Phage that did not hybridize extensively enough to show a signal in an autoradiograph were transferred by toothpick to a lawn of CGE6, grown overnight, and screened in a second plaque filter hybridization. Phage showing no hybridization signal in this second assay were candidate single-copy sequence clones and were analyzed for restriction-fragment patterns in gel-transfer hybridizations as described below. Bacteriophage stocks were prepared in CGE6 by the plate-stock method of Arber et al. (1983) to titers $> 10^{10}$ phage/ml, and

DNA was isolated by either of two methods (Davis et al. 1980; Helms et al. 1985). All work with recombinant bacteriophage was performed under BL1 conditions of containment in accordance with National Institutes of Health guidelines for research involving recombinant DNA molecules.

Isolation of High-Molecular-Weight Human DNA

Genomic DNA from unrelated Caucasians was isolated from white blood cells obtained as leukophoresis by-products from the American Red Cross. High-molecular-weight DNA was prepared from nuclei isolated following solubilization of the leukocytes in 1% Triton X-100 and centrifugation (Bell et al. 1981; Barker et al. 1984*a*). The nuclear pellet was resuspended in 0.25 mM EDTA, 1% SDS and incubated with proteinase K (Boehringer Mannheim), 10 μg/ml for 16–20 h at 42 C. After extraction with phenol and chloroform:isoamyl alcohol (24:1), DNA was precipitated with isopropanol (1.5 vol), redissolved in TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA), and precipitated with ethanol. DNA was redissolved in TE at a concentration of 250 μg/ml.

DNA samples from individual members of 2- and 3-generation families were provided by the Centre d'Etude du Polymorphisme Humain (CEPH), Paris.

Restriction, Digestion, Electrophoresis, Transfer, and Hybridization

DNA samples at 175 μg/ml were digested to completion with restriction endonucleases (5 units/μg) from New England Biolabs (Beverly, MA) under conditions recommended by the manufacturer. The extent of digestion was monitored in identical reaction mixtures containing equal amounts of human DNA and bacteriophage lambda DNA that were analyzed by agarose-gel electrophoresis with ethidium bromide staining. Complete digestion was assumed if the lambda-DNA band pattern represented a limit digest. Screening panels consisted of digested DNA of five unrelated individuals, with 2 μg of each sample/gel lane. Digests of the five DNA samples with six different restriction enzymes were loaded on each 16-cm × 19-cm 0.8% agarose gel and size-fractionated by electrophoresis at 3 V/cm for ~24 h. After denaturation in sodium hydroxide and neutralization, DNA was transferred to nylon membranes (Zeta-Bind®; AMF Cuno) in 25 mM sodium phosphate, pH 6.5, according to the method of Barker et al. (Arber et al. 1983). Filters were prehybridized for 2–24 h at 42 C in 5 × SSC; 40 mM sodium phosphate, pH 7; 5 × Denhardt's solution; denatured, sonicated salmon-sperm DNA (100 μg/ml); 10% dextran sulfate; and 50% formamide.

Bacteriophage probe DNAs were radioactively labeled with ^{32}P deoxynucleotide triphosphate (NEN) by nick-translation (Rigby et al. 1977) to a specific activity of $1–5 \times 10^8$ dpm/μg DNA.

Radioactively labeled single-copy sequence probes (see below) were heated to 100 C for 5 min before addition to the prehybridization solution and nylon membrane. After 16–24 h of incubation at 42 C, filters were washed for 30 min at 20 C with 2 × SSC, 0.1% SDS and for 60 min at 65 C in 0.1 × SSC, 0.5% SDS. Autoradiography was carried out for 1–5 days at −70 C with Kodak XAR-5 film and intensifying screens (DuPont Cronex® Lightning-Plus).

Repeat-containing radioactively labeled DNA probes were prehybridized with a 2,000-fold excess of total human placental DNA to a C_0t of 10 to 100 by denaturing at 100 C for 10 min and incubating the mixture in 0.12 M sodium phosphate, pH 6.5, at 65 C for 4–6 h (Barker et al. 1985; Litt and White 1985). The "repeat-blocked" probes were used in hybridization experiments as described above, except that hybridization was performed at 50 C.

Results

Screening Cloned DNA for Polymorphic Loci

We devised a strategy to isolate, by a relatively rapid and efficient process, probes revealing polymorphic loci in Southern transfer hybridizations. With the aim of minimizing the number of probes to be screened, we chose the genomic clone library of Lawn et al. (1978) as a source of random human sequences to test for polymorphism. The human DNA inserts in these Charon 4A bacteriophage clones are large (10–20 kb) and can therefore be used to screen several restriction fragments in each digest simultaneously.

Probe hybridizations were limited to single-copy sequences (i.e., sequences that are not highly repeated) by either of two methods. In the first method, phage clones free of human repeat sequences were identified in plaque hybridization with radiolabeled, total human DNA (Wyman and White 1980). Only plaques containing highly repeated human sequences hybridize with genomic DNA to a sufficient extent to produce an autoradiographic signal. The remaining plaques consist of phage with low-copy human se-

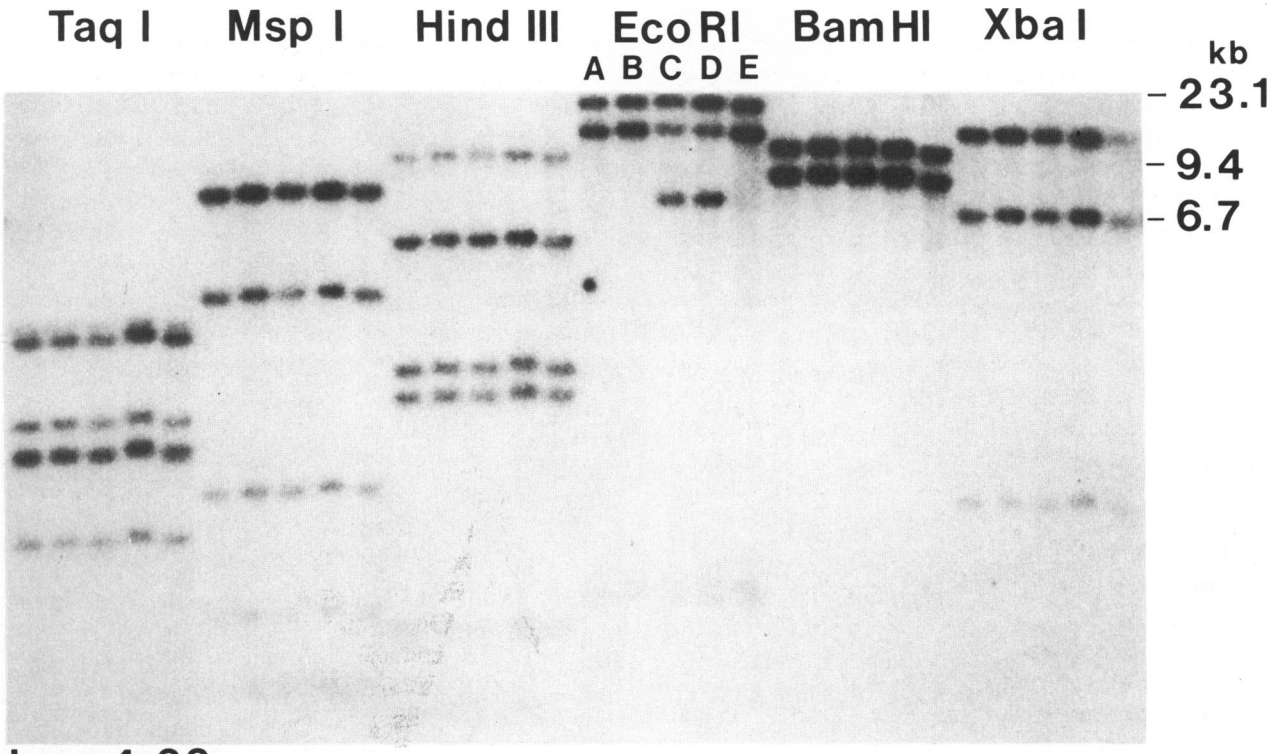
quences or no human insert. The second method of avoiding repeat sequences was to incubate the labeled probe DNA with excess human genomic DNA to saturate the repeat sequences in the probe before hybridization with the Southern transfer filter (Barker et al. 1985; Litt and White 1985). This prehybridization method allowed us to test the single-copy sequences in virtually any of the clones in the library, even though ~99% contain highly repeated sequences (see below).

We also sought to minimize the number of gels and filters to be used in screening probes for polymorphism. Because only polymorphic loci with frequently occurring alleles would be useful for genetic mapping, we reasoned that most useful RFLPs would be detected by testing each probe with just five unrelated individuals. By sampling this number of individuals, the probability of detecting an allele with a frequency of .5 is >99%, and the chance of detecting a rare allele (.1 frequency) is 70%. By testing only five individuals, we could then test as many as six different restriction enzymes on a single 16-cm × 19-cm agarose gel. Maximum use of each gel transfer was obtained by using nylon filters, which could be washed and rehybridized with as many as 10 different probes.

The panel of DNA samples for the initial RFLP screening of each probe was a set of five unrelated individual DNAs digested with six restriction enzymes and separated on a 0.8% agarose gel. Digests with *Msp*I and *Taq*I were always included because of observations that these restriction cleavage sites were polymorphic more often than other restriction sites (Barker et al. 1984*b*). Other restriction enzymes used were *Bam*HI, *Bcl*I, *Bgl*I, *Bgl*II, *Eco*RI, *Eco*RV, *Hae*III, *Hinc*II, *Hin*dIII, *Hinf*I, *Kpn*I, *Pst*I, *Pvu*II, *Rsa*I, *Sac*I, and *Xba*I. Figure 1*A* shows a typical screening panel in which the digested samples have been separated by agarose-gel electrophoresis, transferred to nylon membrane, and hybridized with radiolabeled probe Lam4-33. For five of the enzymes, the pattern of restriction fragments seen in the autoradiograph is identical in all five individuals. However, two individuals carry an *Eco*RI fragment representing an additional allele at this locus. Presumably a small sequence difference in one of the chromosomes of each of these individuals defines an additional cleavage site for the restriction endonuclease *Eco*RI. Figure 1*B* shows a similar screening panel hybridized with the repeat-containing probe RL4-59. The radiolabeled DNA of repeat-containing probes was prehybridized with an excess of unlabeled human placental DNA prior to use in Southern hybridizations (see Material and Methods). This procedure limited the contribution of repeat sequences in the hybridization with filter DNAs and allowed visualization of fragments containing unique sequences. In this example, five of the restriction-endonuclease digestions reveal the same pattern in all of the tested individuals. However, the *Msp*I-digested samples display five variable fragment lengths. Family-inheritance studies indicate that two independent two-allele polymorphisms segregate as a single locus.

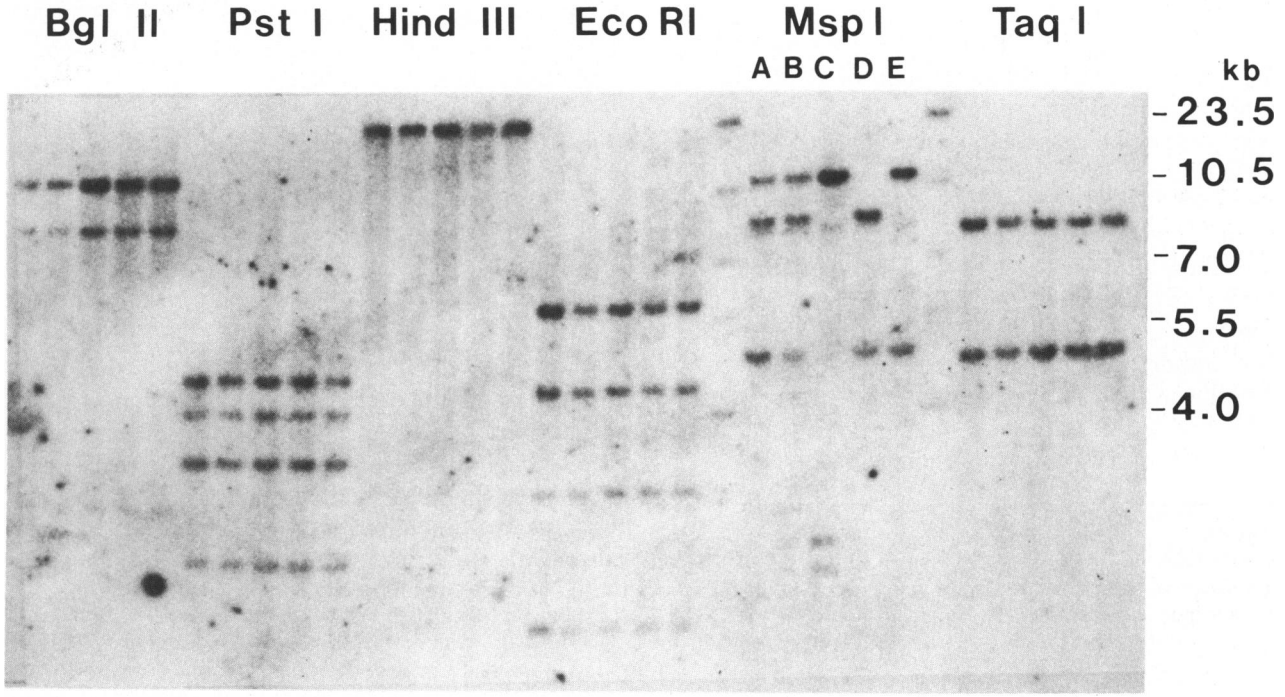
A total of 1,664 phage clones were tested in Southern hybridizations for polymorphic band patterns (table 1). Of this total, 1,025 were repeat-free probes selected from 80,000 clones in the library by plaque hybridization with labeled human genomic DNA. From the patterns of hybridization to Southern transfers, we found that 680 clones of the initial 1,025 actually contained single-copy human sequences and that the remaining 345 showed excessive background hybridization, band patterns too complex for a single-copy sequence, or no hybridization to human DNA. (Most of the clones with no human insert, ~20% of the repeat-free clones in the library, can be identified by plaque hybridization with the Charon 4A "stuffer" fragments containing the *E. coli lac* or *bio* genes [DeWet et al. 1980].) Our finding that 0.9% of the clones in the Maniatis library contained

Figure 1 *A*, Standard screening-blot single-copy clone. Human genomic DNA from five unrelated individuals (A–E) was digested with six restriction endonucleases and prepared for hybridization with single-copy clone Lam4-33 in a test for polymorphism (see Material and Methods). A simple two-allele polymorphism is seen with *Eco*RI. Individuals A, B, and E are homozygous for the 10.0-kb allele whereas individuals C and D have a 7.0-kb allele in addition to the 10.0-kb allele. An invariant fragment length (~20 kb) is also evident. The PIC for this polymorphism was calculated to be .30. *B*, Standard screening-blot repeat-containing clone. A human genomic-DNA-polymorphism screening blot was prepared as described above but hybridized with a probe, RL4-59, known to contain repeat sequences. RL4-59 was prehybridized with unlabeled total genomic DNA prior to the Southern blot test for polymorphism (see Material and Methods). Five variable fragment lengths (10.9, 8.9, 4.7, 2.5, and 2.2 kb) constitute the two independent two-allele polymorphisms revealed with *Msp*I. One polymorphism consists of a 10.9-kb allele and a 8.9-kb allele. The second polymorphism consists of a 4.7-kb and a 2.5-kb + 2.2-kb allele. Both polymorphisms appear to be site polymorphisms. All individuals examined display an allele from the first and second polymorphism, and all combinations (i.e., all four haplotypes) are observed.



Lam 4-33

A



RL4-59

B

Table 1

Screening Results

Clones Tested (*N*)	No. Not Polymorphic	No. Polymorphic	
		2 Alleles	⩾2 Alleles
Single Copy (1,025)	350	180	151
Random (639)	410	112	72
Total (1,664)	760	292	223

only unique-sequence DNA is consistent with the observation of Kao et al. (1982) that 1% of the clones in the library are single-copy sequences but is not consistent with the sixfold-lower estimate of Feder et al. (1985).

In addition to the repeat-free probes, 639 clones tested were random probes from the library that were prehybridized with human genomic DNA immediately before the Southern hybridization. On the basis of our results in screening for single-copy probes, we estimate that 99% of these random probes contained highly repeated human sequences.

Polymorphic loci were detected at the same frequency regardless of whether repeat-free clones or prehybridized random clones were used. As summarized in table 1, 331 of the repeat-free clones revealed polymorphism. This total represents 32% of the single-copy candidate probes or 49% of the probes ultimately found to hybridize only to single-copy restriction fragments. One hundred eighty of these displayed two-allele systems, and 151 revealed loci with three or more alleles as indicated by the number of restriction-fragment patterns observed with those probes. The 639 random clone probes revealed an additional 184 polymorphisms in single-copy fragments displayed in Southern hybridization. One hundred twelve of these were two-allele polymorphisms, and 72 appeared to be loci with three or more alleles. In all, 515 polymorphic loci were identified by screening 1,664 phage probes.

Basis of Polymorphisms

The RFLPs that were identified in this survey represent several different types of sequence variation in the human genome. The most common basis for variation in restriction-fragment lengths that could be inferred from the patterns of hybridizing bands is the simple mutation of restriction sites; that is, alleles most often differed only by the presence or absence of a single restriction site, attributable to single base substitutions in the restriction site or to minor sequence rearrangements limited to that region. Such restriction-site polymorphisms were expected to be common, given the 1% frequency of single base substitutions in comparisons of individual DNA sequences (Jeffreys 1979; Ullrich et al. 1980). Criteria for defining an RFLP as a restriction-site polymorphism include (1) individuals' genomic restriction maps that differ only by a restriction site, (2) a restriction fragment that is allelic with a pair of restriction fragments whose molecular weights sum to that of the first, or, more generally, (3) a simple two-allele polymorphism with no evidence of related variations in any other restriction-enzyme-fragment patterns. RFLPs that appear to result from rearrangement of DNA segments rather than from mutation of single restriction sites were also observed, although more rarely than the restriction-site polymorphisms (58 of the 515 polymorphic loci). This type of polymorphism was detected when the lengths of fragments produced with two or more enzymes varied coordinately in the DNA of 10 or more individuals, implying a common molecular basis. An example of such a polymorphism is shown in the hybridization of Lam4-945 to five individuals' DNA digested with different restriction enzymes (fig. 2). Allelic restriction fragments differ in molecular weight by the same absolute amount in each of several restriction digests, as if an extra DNA segment in one allele increases the spacing between restriction sites in otherwise homologous sequences. Although other explanations of such linkage disequilibrium are possible, DNA rearrangements (insertions, deletions, or inversions) are the most likely mechanisms for these RFLPs within a single probed locus.

Characterization of RFLPs as Genetic Markers

Probes that revealed RFLPs in the initial screening process were generally characterized in additional gel-transfer hybridizations, so as to (1) determine in-

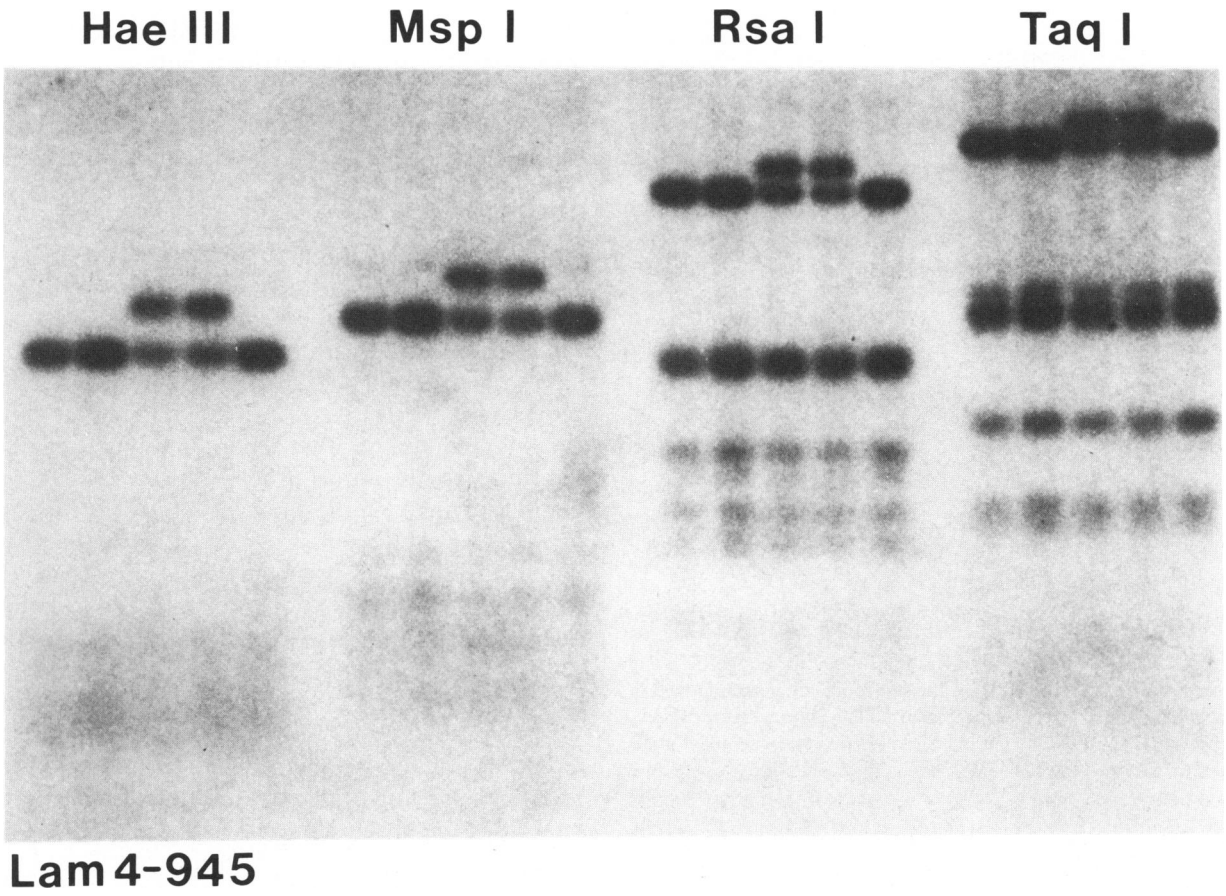


Figure 2 DNA rearrangement polymorphism. A standard screening blot was prepared, and hybridization with cloned single-copy probe Lam4-945 was carried out as described in Material and Methods. Allelic restriction fragments differ in molecular weight by the same absolute amount in each of the restriction digests shown, indicating the insertion of a segment of DNA that increases the spacing between restriction sites.

formation content by analysis of the genotypes of a larger sample of individuals, (2) screen for additional polymorphism at the same locus with different restriction enzymes, and (3) determine, by analysis of their segregation in 3-generation families with large sibships, whether RFLP alleles detected with each probe are inherited in Mendelian fashion (i.e., as a single genetic locus).

The relative usefulness of probes as genetic markers can be quantified by PIC (Botstein et al. 1980). The PIC is a measure of how often the occurrence of distinguishable parental alleles will allow the ascription of the pattern of segregation of the polymorphic locus to the offspring of a particular mating. PIC values are determined on the basis of the allele frequencies at each locus and range between 0 and 1, with a value of 1 indicating that the inheritance of every parental chromosome can be traced. The most informative two-allele systems, with both alleles occurring with equal frequency in the population, have a PIC value of .375. The proportion of polymorphic loci that have been reported as having PIC values >.7 is relatively small (Willard et al. 1985).

We have determined the PIC values of the genetic loci defined by 176 of our DNA probes. These determinations were based on allele frequencies inferred from the restriction-fragment patterns of 10–50 unrelated individuals. For loci with complex patterns of variable bands, alleles could not be defined from the fragments observed in single individuals. In those cases allelic combinations of bands (haplotypes) were identified by segregation analysis in families. Haplotype frequencies were then estimated from their occurrence in unrelated individuals drawn from the families analyzed. Figure 3 shows the distribution of PIC values among this set of probes. Twenty-nine

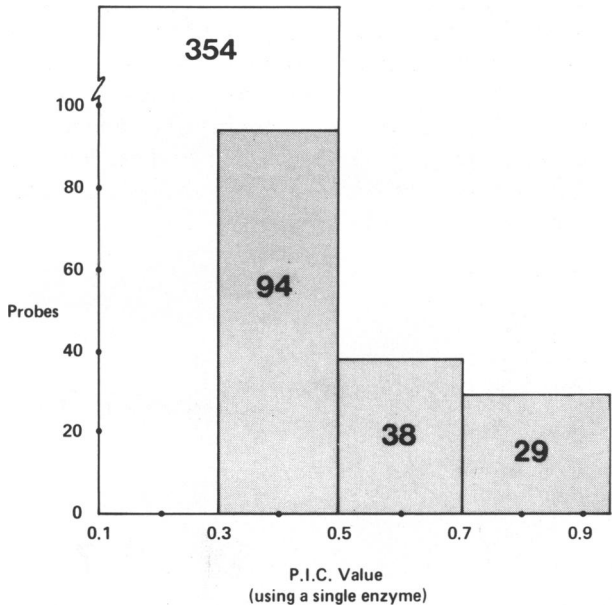


Figure 3 Distribution of RFLP probes according to PIC. The polymorphisms identified from the screening study summarized in table 1 were characterized further, and PIC values were calculated. Approximately 10–50 unrelated individuals were scored for alleles that formed the basis upon which the calculations were made.

probes identify with PIC values >.7; 38 are .5–.7; 94 are .3–.5; and 15 <.3. The remaining 339 RFLP probes are from loci that have not been characterized sufficiently for PIC estimates, but most appear to identify two-allele polymorphisms. The distribution of PIC values shown in figure 3 does not necessarily represent the distribution of polymorphic sequence variation in random genomic clones. Rare polymorphisms (i.e., loci of low PIC) are more likely to escape detection in a screen of only five individuals, and we have generally selected the more polymorphic loci for PIC determination. Thus, the distribution of loci in the different PIC categories in figure 3 is skewed toward more highly polymorphic loci, and we expect that the bulk of the uncharacterized probes have PIC values <.3.

Approximately 40% of the 515 RFLP probes reveal independent polymorphisms with two or more different restriction enzymes. Because the alleles seen with each enzyme appear to be at least partly in linkage equilibrium, we presume that these coincident polymorphisms are due simply to independent, variable restriction sites rather than to DNA rearrangements affecting several different restriction-fragment patterns. Of the 176 probes for which we have calculated PIC in single-enzyme digests (fig. 3), there are 68 with additional, independent RFLPs with other restriction enzymes. For some of these probes, the PIC increase derived from additional restriction digests has been shown to be substantial. In most cases, however, the PIC calculation for the combination of polymorphisms at these loci must await measurement of both linkage equilibrium and haplotype frequencies in the population.

The use of RFLPs as genetic markers is based on the premise that all the restriction-fragment-length variation observed with each probe is attributable to stably inherited alleles of a single genetic locus. When more than one RFLP is detected with a single clone, the possibility exists that separate polymorphic loci hybridize to the same probe. We have rarely encountered this complication with this set of probes, however. We have analyzed the inheritance of the variable restriction fragments for more than 100 probe-enzyme combinations in 3-generation human pedigrees. In every case except one, the segregation pattern of all RFLPs detected with each probe was consistent with single-locus inheritance. In addition, mapping results for 15 of the most polymorphic clones in somatic-cell hybrids indicate a single chromosomal location for the restriction fragments detected by each probe (O. Cohen-Haguenauer, N. Van Cong, R. G. Knowlton, V. A. Brown, M. F. deTand, C. Jegou, M.-S. Gross, J. Frezal, and H. Donis-Keller, unpublished data). In the one exception noted so far (*Pst*I and *Rsa*I polymorphisms detected with probe RL4-397), the inheritance patterns appear to be generated by the segregation of two unlinked polymorphic loci cross-hybridizing to the same probe.

Probes for Highly Polymorphic Loci

Several of the random probes hybridized to genetic loci displaying unusual variability in their restriction-fragment patterns. Hybridization to unrelated individuals' samples digested with just a single restriction enzyme produced very distinctive sets of bands that could not have resulted from variation at one or even a small number of restriction sites (fig. 4). Twenty probes that define highly polymorphic loci have been characterized by study of their inheritance in 21 CEPH reference families and are listed in table 2. These data include the restriction enzymes used to detect the polymorphism, the estimated PIC heterozygosity, and the assessment of whether the polymorphism is caused by mutations limited to specific restriction sites or by more extensive rearrangements of the DNA sequence.

One property of many of the most informative genetic markers is that two or more variable fragments define each allele. In two cases (Lam4-159 and Lam4-427) we have determined that variable restriction fragments arise from rearrangements in two different regions separated by a segment of constant length (Knowlton et al. 1986; J. W. Schumm, unpublished data). An allele of one of these loci is therefore a combination or haplotype of variable fragments, all of which cosegregate in any given family. Because polymorphism in the two or more variable fragments at each locus is independent and there is no apparent linkage disequilibrium, the number of fragment combinations defining alleles at some of these loci is so high that the markers are almost always fully informative for inheritance.

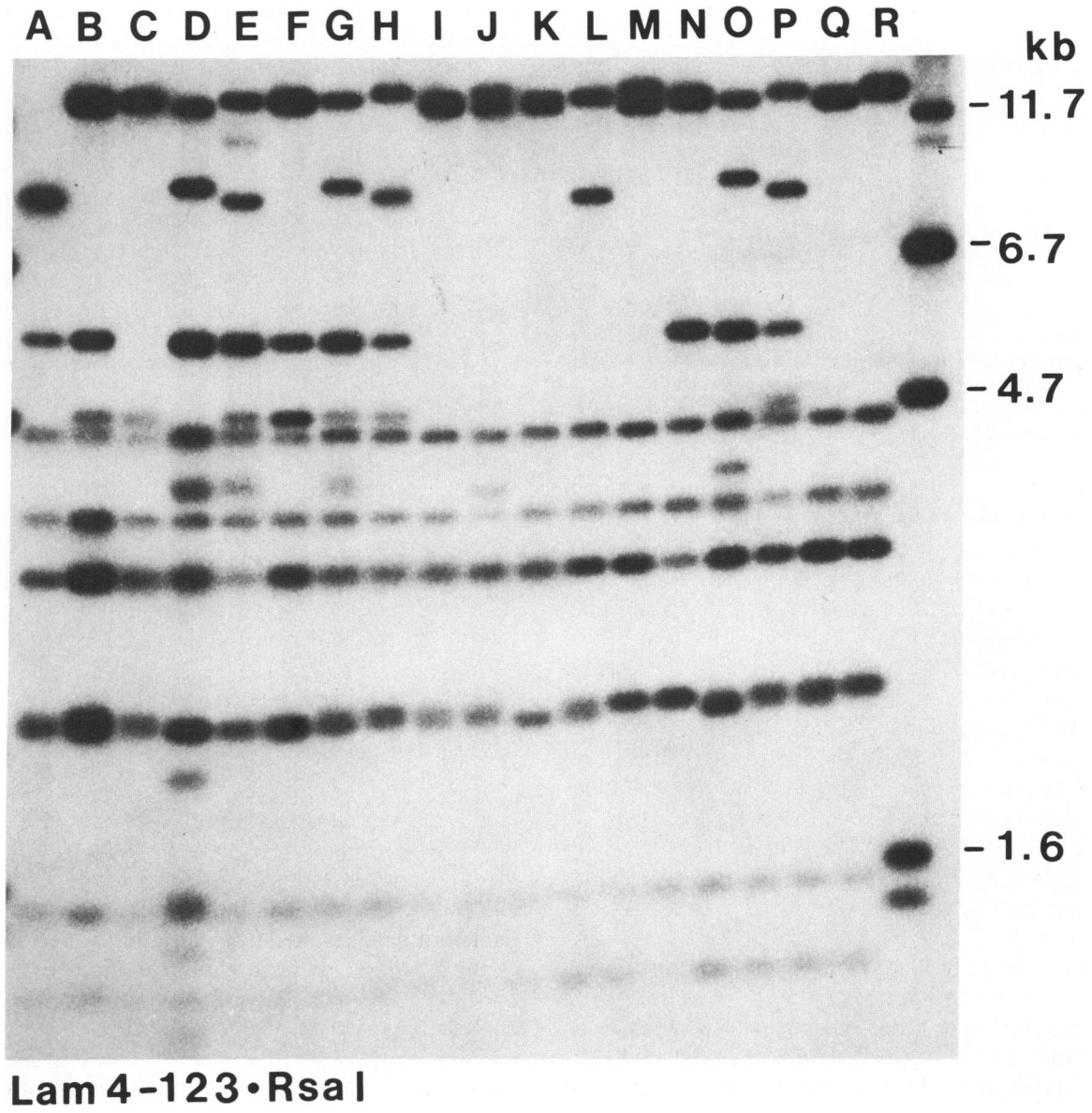


Figure 4 Highly polymorphic RFLP loci. Human genomic DNA from 23 unrelated individuals (A–W) was digested with restriction enzymes *Msp*I, *Bgl*II, or *Rsa*I and prepared for hybridization with the RFLP probes indicated (see Material and Methods). Additional unmarked lanes contain lambda-DNA restriction digests used as size markers (indicated in kilobases to the right of each figure). See table 2 and text for additional information.

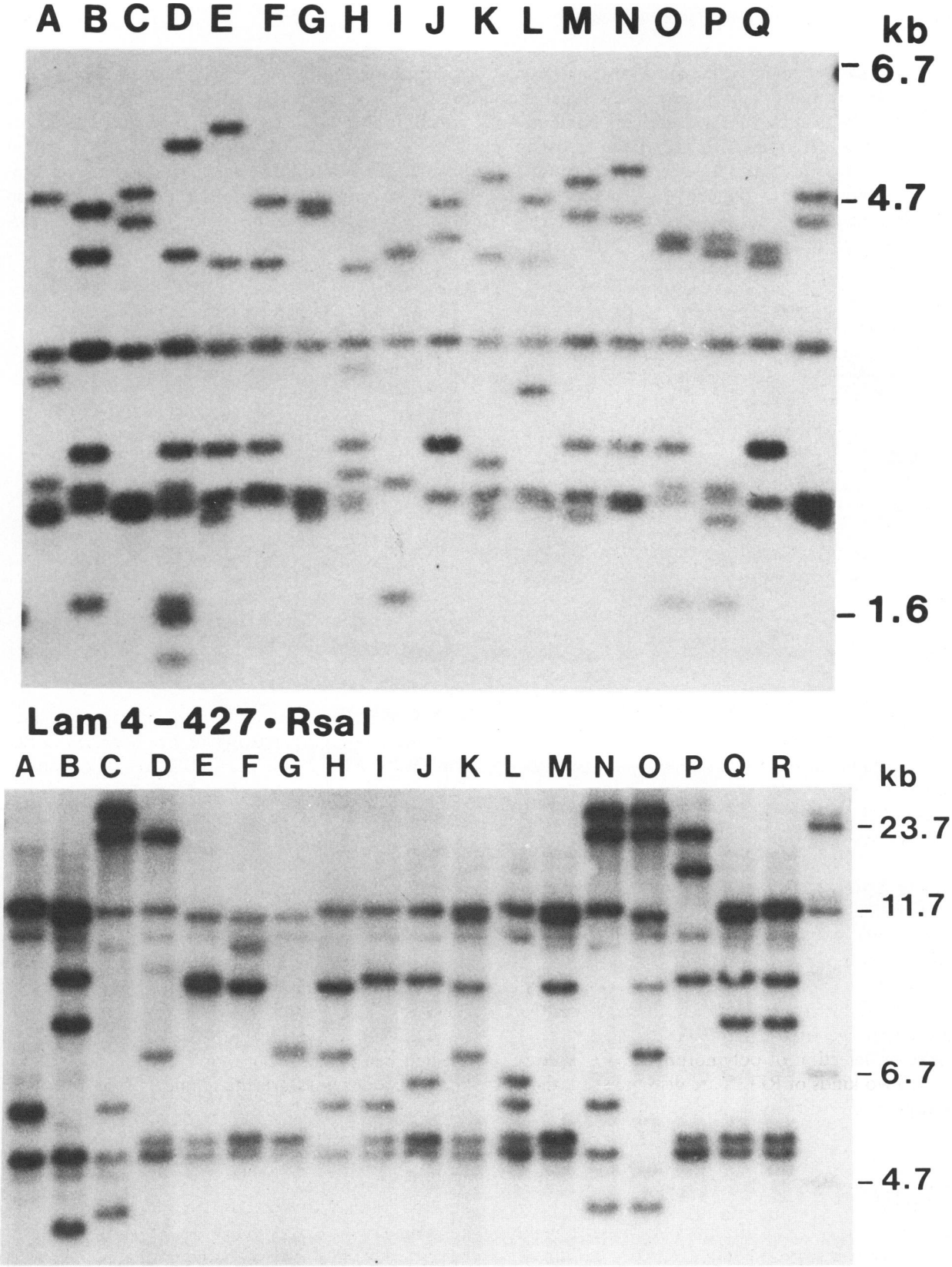


Figure 4 Continued

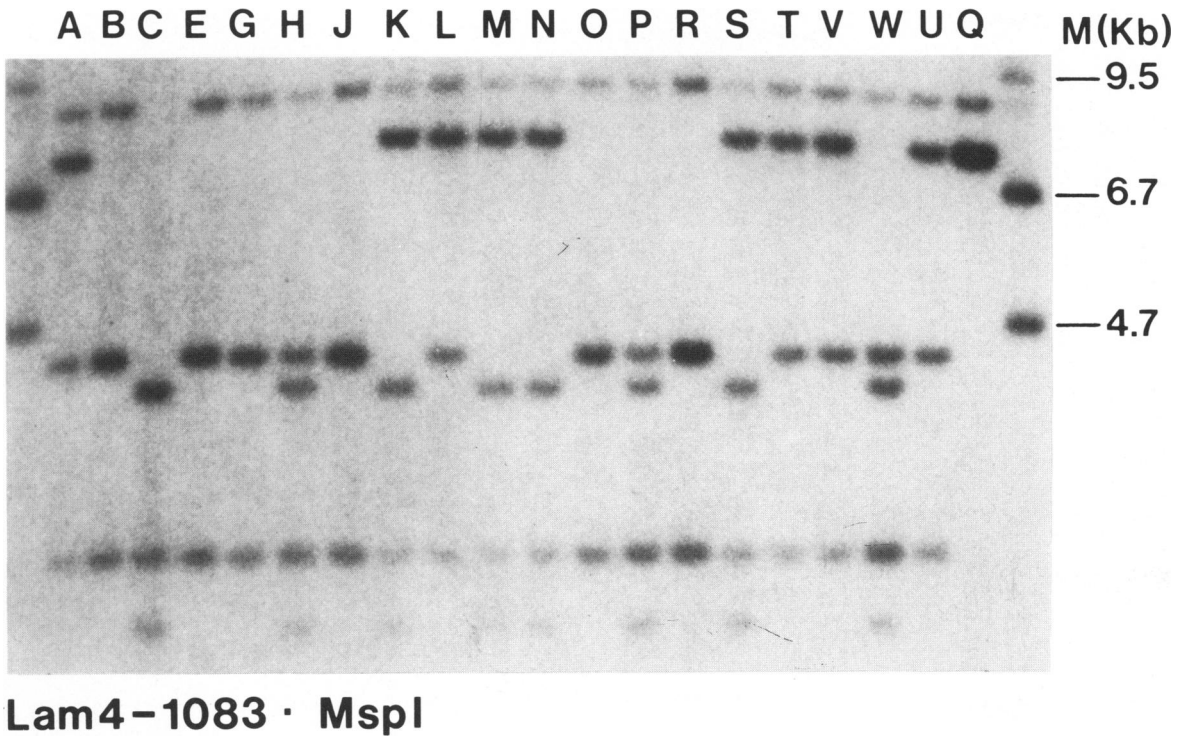


Figure 4 Continued

Relative Efficiencies of Restriction Endonucleases in Detecting Polymorphism

The accumulated results of screening for polymorphism with several different restriction enzymes are a measure of the frequency of variation at each of the restriction sites in the genome and can also be used to predict the most efficient strategies for identifying RFLPs in untested DNA sequences. By using a standard procedure to test large numbers of restriction sites in the genome, we were able to make statistically significant comparisons of RFLP frequencies observed with a variety of restriction enzymes.

For eight different restriction enzymes, the number of probes and genomic restriction sites screened and the rate of detection of polymorphisms are listed in table 3. Two kinds of RFLPs are distinguished, those that appear to result from mutation of single restriction sites and those that apparently result from more extensive DNA rearrangements affecting fragment lengths of more than one restriction-enzyme digest. When these eight restriction enzymes are compared in terms of the rate of detection of restriction-site variants per site screened (table 3), they can be divided into two classes. The first group, consisting of *Msp*I and *Taq*I, generated RFLPs at the highest rate (40.5 and 33.1 RFLPs/1,000 sites, respectively), and the second group, consisting of five enzymes (*Pst*I, *Bam*HI, *Eco*RI, *Bgl*II, and *Hin*dIII), reveal polymorphism at approximately half the rate seen with *Msp*I or *Taq*I. As determined by χ^2 analysis, the rates with *Taq*I and *Pst*I, the nearest neighbors between the high and the low group, are significantly different ($P < .05$). The rate of restriction-site polymorphism seen with *Rsa*I falls closer to that of the lower group but, because of the smaller sample size of *Rsa*I sites tested, cannot be distinguished from the *Taq*I frequency with 95% confidence.

When the efficiencies of the eight restriction enzymes are compared on the basis of polymorphisms detected per phage probe (table 3), the differences among the enzymes are maintained. Even with the inclusion of polymorphisms resulting from insertions or deletions, *Msp*I and *Taq*I polymorphisms are still observed most often. *Rsa*I polymorphisms are found almost as frequently, at a rate significantly higher ($P < .05$) than that seen with any of the five remaining enzymes. We had expected that it would be more efficient to screen with enzymes with frequent restriction sites because more sites could be screened per probe hybridization. However, the data in table 3

Table 2

Probes for Highly Polymorphic Loci

Probe	Enzyme	Rearrangement or Site PM?	HET	PIC
Lam4-1214	*Bgl*II	R	.98	.98
Lam4-427	*Rsa*I	R	.95	.93
RL4-227	*Eco*RI/*Taq*I	S	.81	.76
Lam4-281	*Msp*I/*Taq*I	S	.79	.71
Lam4-1020	*Taq*I/*Hin*dIII	S	.79	.76
Lam4-45	*Msp*I	S	.76	.74
Lam4-892	*Taq*I	R	.76	.69
RL4-208	*Bgl*II/*Msp*I	S	.76	.71
Lam4-336	*Rsa*I	R	.76	.71
Lam4-1169	*Msp*I	S	.74	.69
RL4-191	*Msp*I/*Taq*I	S	.74	.64
Lam4-1065	*Rsa*I	R	.74	.74
Lam4-368	*Hin*dIII	S	.74	.74
Lam4-159	*Pst*I	R	.74	.74
Lam4-966	*Rsa*I	R	.74	.69
Lam4-355	*Bgl*II	R	.71	.71
Lam4-123	*Rsa*I/*Msp*I	R	.71	.71
Lam4-1046	*Msp*I/*Taq*I	R	.71	.71
Lam4-962	*Msp*I	S	.71	.62
Lam4-605	*Msp*I	S	.71	.50

show that, in terms of the number of fragments detected per probe hybridization, the differences among these eight enzymes are not as great as the differences in number of sites in the genome. In practice, fragments <1 kb in length are not generally detected by our transfer and hybridization procedure, so that many small *Rsa*I, *Msp*I, and *Taq*I fragments complementary to the probes are not screened. The data also indicate that repeat-containing and single-copy probes display the same number of fragments and the same frequency of polymorphic sites.

Discussion

In identifying 515 RFLP loci we have shown that the collection of a sufficient number of markers to construct a linkage map of the human genome is feasible. This set of probes alone defines polymorphic loci at an average spacing of ~6 cM. If these markers are randomly distributed, then ~95% of the genomic sequences are ≤10 cM from at least one of the RFLPs in this set. Of course, a generally useful map will consist of markers that are sufficiently polymorphic

Table 3

Restriction-Enzyme Efficiency in Detecting Polymorphism

Enzyme	No. of Probes	No. of Restriction Sites		Restriction-Site PMs		PMs/ 100 Probes
		Screened	Per Probe	Per 1,000 Sites	Per 100 Probes	
*Msp*I	827	3,427	4.1	40.5	16.8	19.8
*Taq*I	823	3,623	4.4	33.1	14.6	17.5
*Rsa*I	171	790	4.6	22.8	10.5	14.6
*Pst*I	395	1,599	4.0	19.4	7.8	8.6
*Bam*HI	286	983	3.4	17.3	5.9	8.0
*Eco*RI	649	2,617	4.0	14.9	6.0	8.0
*Hin*dIII ...	835	3,341	4.0	12.9	5.1	7.9
*Bgl*II	492	1,887	3.8	13.2	5.1	6.9

to permit linkage analysis with a manageable number of families; we have defined a minimum PIC value of .3 as the criterion for a useful RFLP marker. Of the 515 RFLP probes, we have already identified 161 with PIC values >.3, and we estimate that at least 90 additional useful RFLP loci will be found by further characterization of probes in this collection. Thus, by screening only 0.1% of the clones in the Maniatis Charon 4A library (including 10% of the repeat-free clones), we have found usefully informative markers at an average spacing of 13 cM, so additional markers to achieve saturation of the genomic map can be found by continuation of the same process.

The genetic variability displayed by these probes also indicates the feasibility of constructing a human linkage map consisting mainly of highly informative RFLP markers. Botstein et al. (1980) calculated that a useful human linkage map could consist of ~150 evenly spaced markers if the markers were highly informative. We have identified 29 probes that reveal RFLPs with PIC values >.7 and 38 more with PIC values >.5. Linkage studies with more than 100 of the most polymorphic probes in 21 3-generation pedigrees (data not shown) and chromosomal localization of more than 20 probes with rodent-human somatic-cell hybrid panels (O. Cohen-Haguenauer, N. Van Cong, R. G. Knowlton, V. A. Brown, M. F. deTand, C. Jegou, M.-S. Gross, J. Frezal, and H. Donis-Keller, unpublished data) indicate that the highly polymorphic loci are dispersed throughout the genome.

Most of the polymorphisms that we have observed appear to result from nucleotide changes or small rearrangements that can be detected only by digestion with a single restriction endonuclease. However, the most polymorphic loci display variation in restriction fragments with several different enzymes and appear to result from DNA rearrangements. In some cases, such as Lam4-159 and Lam4-427, it is clear that independent DNA rearrangements have occurred at two different sites within the same locus. We have also observed that at these and other complex loci (e.g., Lam4-1214, Lam4-336, and Lam4-71) the DNA sequences surrounding the variable insertion/deletion are not included within the cloned bacteriophage probe. This result may reflect a growth disadvantage of phage clones containing unusual, unstable sequence arrangements, such as inverted or tandem repeats. Wyman et al. (1985) have demonstrated the occurrence in the human genome of a significant fraction (~10%) of sequences that require recombination-deficient hosts for propagation. The host and cloning vector system used in these experiments would not support the growth of this class of clones.

The use of the probe prehybridization method (Bell et al. 1981; Barker et al. 1985) to saturate repetitive sequences in DNA probes prior to their use in hybridization experiments makes it possible to survey unique-sequence DNA for polymorphisms even in the presence of interspersed repetitive DNA. Approximately 70% of the randomly isolated clones tested in this way gave readable patterns, and these clones were comparable to single-copy sequence clones in the frequency with which they displayed polymorphism. However, only three of the 20 highly polymorphic loci in table 2 (RL4-227, RL4-208, and RL4-191) were identified with random, prehybridized probes, even though those probes represented 40% of the clones screened for polymorphism. We do not know why the repeat-free clones detected highly variable loci at greater frequency. Because the number of restriction fragments displayed per probe hybridization is the same in both methods, the discrepancy is not due to sensitivity or other artifactual differences. It may be that hypervariable sequences tend to occur in regions free of interspersed, high-copy-number repeat sequences.

Searching for RFLPs simply by testing individual random clones from a human genomic library has proved to be a satisfactory approach to identifying polymorphic loci for mapping. Approximately 30% of the probes that we tested were found to reveal polymorphic loci. The usefulness of large genomic fragments as probes was not impaired by complexity of the restriction-fragment patterns detected, and hybridization signals of interspersed repetitive sequences could be avoided. Phage clones with large human inserts and cosmid clones with even longer human sequences have been used successfully by other laboratories to find polymorphism (Feder et al. 1985; Litt and White 1985). Screening for polymorphism with cDNA probes (Helentjaris and Gesteland 1983) has been less successful, presumably because the probes hybridize to fewer genomic fragments and perhaps because the coding sequences are more conserved than random sequences.

Even though this screening procedure has been successful in identifying useful polymorphisms, more-efficient strategies would still be an advantage. Any other plan, however, would have to provide a substantial increase in the rate of identification of RFLPs,

to compensate for the loss of simplicity of a random screening procedure. The best alternative appears to be to exploit the length variation of clusters of tandemly repeated sequences. Jeffreys et al. (1985) found that a polymorphic sequence adjacent to the myoglobin gene and consisting of a set of tandemly repeated sequences cross-hybridized with similar clusters elsewhere in the genome and found that these homologous regions also tended to be polymorphic. Although it is clear that the collective polymorphism displayed by each of the "mini-satellite" probes is extensive, the identification of alleles of individual loci contributing to the pattern is difficult. At least one of the chromosomal loci displaying mini-satellite polymorphism proved to be highly informative (Wong et al. 1986), but others were not (Jeffreys et al. 1985). Nakamura et al. (1987) have recently shown that genomic clones hybridizing to mini-satellite oligonucleotides represent an enriched source of single loci with variable-number tandem repeat sequences that tend to be polymorphic. Another strategy for finding polymorphic tandem repeat–sequence clones is to select bacteriophage clones that are viable only in *rec*$^-$ *E. coli* strains, on the presumption that sequences that are unstable during propagation in *E. coli* might also be prone to rearrangement in the human genome. Again, preliminary results from such a search procedure do not indicate an advantage over random screening for the identification of useful single-locus polymorphisms (Wyman et al. 1987).

Our systematic screening program with random clones has allowed us to make a statistically significant comparison of the efficiencies of different restriction enzymes in detecting polymorphism. Our results provide the basis for quantitative comparison of different restriction enzymes, not only because we have tested a large number of sites with each enzyme but also because we adhered to a single standard procedure for screening DNA clones for polymorphism. Of the eight restriction enzymes for which we had screened more than 500 independent recognition sites, the most efficient were *Msp*I, *Taq*I, and *Rsa*I, followed by the set of enzymes *Pst*I, *Bgl*II, *Eco*RI, *Hin*dIII, and *Bam*HI, which were indistinguishable in their results. To examine specifically the variability of different restriction sites in the genome, we subtracted those polymorphisms that appeared to result from DNA rearrangements and compared the enzyme efficiencies on a per-site basis. *Msp*I and *Taq*I sites were more variable than the other sites tested, confirming indications given by the results of others (Barker et al. 1984*b*; Cooper et al. 1985; Feder et al. 1985). Barker et al. (1984*b*) argue that the high polymorphism rate with *Msp*I and *Taq*I is a consequence of the CG dinucleotide in the recognition sites for these enzymes. In mammalian genomes, a 5-methyl cytosine is often found in the CG dinucleotide sequences and is probably subject to a higher mutation frequency. An additional consequence of unequal mutation rates that might contribute to the high frequency of *Msp*I and *Taq*I polymorphisms found is the dinucleotide sequence composition of the human genome. Wijsman (1984) has calculated expected frequencies of RFLPs for different restriction enzymes on the basis of existing dinucleotide frequencies in the human genome and on the basis of the practical limitations of detecting restriction fragments and has predicted that *Msp*I and *Taq*I would be among the most successful at revealing polymorphism. Her model also predicts that *Rsa*I polymorphisms would be found at high frequency. In our screening this enzyme was third in efficiency, behind only *Msp*I and *Taq*I.

The set of RFLPs described here increases substantially the number of such markers in the human genome. The most recent tabulation (Willard et al. 1985) listed 333 RFLPs identified with various DNA probes, with at least three polymorphic loci on each human chromosome. The value of the reported RFLP markers for linkage mapping varies greatly. The vast majority are two-allele systems, but an increasing number of highly polymorphic loci have also been found (Wyman and White 1980; Bell et al. 1981; Higgs et al. 1981; Litt and White 1985; Oberle et al. 1985; Nakamura et al. 1987). Directed efforts to obtain RFLPs from selected chromosomes have led to probe collections adequate for both detection of linkage among RFLPs on portions of several human chromosomes (Kittur et al. 1985; Lathrop et al. 1985; White et al. 1985; Leach et al. 1986) and construction of linkage maps of the X chromosome and chromosome 7 (Drayna and White 1985; Donis-Keller et al. 1986; Barker et al., in press).

The large RFLP probe resource has also proved to be useful for other applications besides linkage studies. In particular, the very highly polymorphic loci are very informative genetic markers in paternity testing (Kazazian et al. 1986) and in determining the genetic origin of cells of the lympho-hematopoietic system in recipients of bone marrow transplants (Knowlton et al. 1986).

The construction of a linkage map of RFLP mark-

ers will obviously have many important applications (Botstein et al. 1980; Lander and Botstein 1986*a*, 1986*b*). Informative RFLP loci that have been mapped can be used subsequently to map other inherited traits, particularly genetic loci responsible for inherited diseases. Linkage of RFLP markers to single-gene disorders can be detected even without a genetic map by testing for cosegregation with random polymorphisms, although selecting probes from a mapped collection would be more efficient. Already loci responsible for several diseases have been localized by their association with linked RFLPs. The cystic fibrosis locus was first mapped by linkage to DOCRI-917 (Tsui et al. 1985), one of the random polymorphic probes in the collection described here. Chromosomal localization of this probe and close genetic linkage to other mapped probes allowed assignment of the CF gene to chromosome 7 (Knowlton et al. 1985; Wainwright et al. 1985; White et al. 1985). For more complex disorders, however, a complete genetic map will be essential to track simultaneously the multiple genetic components involved. With such a genetic map and the necessary data-analysis tools, we will be equipped to begin to define precisely the inherited components (if present) in such diverse disorders as cancer, diabetes, schizophrenia, and coronary artery disease.

Probes from this collection will be made available to interested investigators for research purposes.

References

Arber, W., L. Enquist, B. Hohn, N. Murray, and K. Murray. 1983. Experimental methods for use with lambda. Pp. 433–466 *in* R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg, eds. Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Barker, D., C. Helms, G. Akots, and H. Donis-Keller. 1985. RFLP analysis of probes containing repetitive DNA using probe prehybridization. Eighth International Workshop on Human Gene Mapping. Cytogenet. Cell. Genet. 40:576.

Barker, D., T. Holm, and R. White. 1984*a*. A locus on 11p with multiple restriction site polymorphisms. Am. J. Hum. Genet. 36:1159–1171.

Barker, D., M. Schafer, and R. White. 1984*b*. Restriction sites containing CpG show a higher frequency of polymorphisms in human DNA. Cell 36:131–138.

Barker, D. F., P. Green, R. G. Knowlton, J. W. Schumm, E. S. Lander, A. Oliphant, G. Akots, V. A. Brown, T. Gravius, C. Helms, C. Nelson, C. Parker, K. Rediker, C. Parker, D. Wat, B. Weiffenbach, and H. Donis-Keller. A genetic linkage map of 63 chromosome 7 DNA RFLP markers. Proc. Natl. Acad. Sci. USA (in press).

Bell, G. I., J. H. Karam, and W. J. Rutter. 1981. Polymorphic DNA region adjacent to the 5′ end of the human insulin gene. Proc. Natl. Acad. Sci. USA 78:5759–5763.

Benton, W. D., and R. W. Davis. 1977. Screening lambda-gt recombinant clones by hybridization to single plaques in situ. Science 196:180–182.

Bishop, D. T., and E. J. Gardner. 1980. Pp. 389–406 *in* J. Cairns, J. Lyon, and M. Skolnick, eds. Banbury report 4: cancer incidence in defined populations. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Botstein, D., R. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314–331.

Cooper, D. N., B. A. Smith, H. H. Cooke, S. Niemann, and J. Schmidtke. 1985. An estimate of unique DNA sequence heterozygosity in the human genome. Hum. Genet. 69:201–205.

Davis, R. W., D. Botstein, and J. R. Roth. 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

DeWet, J. R., D. L. Daniels, J. L. Schroeder, B. G. Williams, K. Denniston-Thompson, D. D. Moore, and F. R. Blattner. 1980. Restriction maps for twenty-one Charon vector phages. J. Virol. 33:401–410.

Donis-Keller, H., D. F. Barker, R. G. Knowlton, J. W. Schumm, J. C. Braman, and P. Green. 1986. Highly polymorphic RFLP probes as diagnostic tools. Pp. 317–324 *in* Cold Spring Harbor symposia. Vol. 51. Molecular biology of *Homo sapiens*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Drayna, D., and R. White. 1985. The genetic linkage map of the human X chromosome. Science 230:753–758.

Feder, J., L. Yen, E. Wijsman, L. Wang, L. Wilkins, J. Schroder, N. Spurr, H. Cann, M. Blumenberg, and L. L. Cavelli-Sforza. 1985. A systematic approach for detecting high frequency restriction fragment length polymorphisms using large genomic probes. Am. J. Hum. Genet. 37:635–649.

Helentjaris, T., and R. Gesteland. 1983. Evaluation of random cDNA clones as probes for human restriction fragment polymorphisms. J. Mol. Appl. Genet. 2:237–247.

Helms, C., M. Y. Graham, J. E. Dutchik, and M. V. Olson. 1985. A new method for purifying lambda from phage lysates. DNA 4:39–49.

Higgs, D., S. Goodbourn, J. Wainscoat, J. Clegg, and D. Weatherall. 1981. Highly variable regions of DNA flank the human alpha-globin genes. Nucleic Acids Res. 9:4213–4224.

Jeffreys, A. J. 1979. DNA sequence variants in γ-, α-, δ-, and β-globin genes of man. Cell 18:1–10.

Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. Hypervar-

iable "minisatellite" regions in human DNA. Nature 314:67–73.

Kao, F., J. A. Hartz, M. L. Law, and J. N. Davidson. 1982. Isolation and chromosomal localization of unique DNA sequences from a human genomic library. Proc. Natl. Acad. Sci. USA 79:865–869.

Kazazian, H. H., Jr., S. H. Orkin, C. D. Boehm, S. C. Goff, C. Wong, C. E. Dowling, P. E. Newburger, R. G. Knowlton, V. Brown, and H. Donis-Keller. 1986. Characterization of a spontaneous mutation to a β-thalassemia allele. Am. J. Hum. Genet. 38:860–867.

Kittur, S. D., J. W. M. Hoppener, S. E. Antonarakis, J. D. J. Daniels, D. A. Meyers, M. Jensen, R. G. Korneluk, B. D. Nelkin, and H. H. Kazazian, Jr. 1985. Linkage map of the short arm of the human chromosome 11: location of the genes for catalase, calcitonin, and insulin-like growth factor II. Proc. Natl. Acad. Sci. USA 82:5064–5067.

Knowlton, R. G., V. A. Brown, J. C. Braman, D. Barker, J. W. Schumm, C. Murray, T. Takvorian, J. Ritz, and H. Donis-Keller. 1986. Use of highly polymorphic DNA probes for genotypic analysis following bone marrow transplantation. Blood 68:378–385.

Knowlton, R. G., O. Cohen-Haguenauer, N. Van Cong, J. Frezal, V. A. Brown, D. Barker, J. C. Braman, J. W. Schumm, L. C. Tsui, M. Buchwald, and H. Donis-Keller. 1985. A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. Nature 318:380–382.

Lander, E. S., and D. Botstein. 1986*a*. Strategies for studying heterogeneous genetic traits in humans using a linkage map of restriction fragment length polymorphisms. Proc. Natl. Acad. Sci. USA 83:7353–7357.

———. 1986*b*. Mapping complex traits in humans: new methods using a complete RFLP linkage map. Pp. 49–61 *in* Cold Spring Harbor symposia. Vol. 51. Molecular biology of *Homo sapiens*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Lander, E. S., and P. Green. 1987. Construction of multilocus genetic linkage maps in humans. Proc. Natl. Acad. Sci. USA 84:2363–2367.

Lange, K., and M. Boehnke. 1982. How many polymorphic marker genes will it take to span the human genome? Am. J. Hum. Genet. 34:842–845.

Lathrop, G. M., J. M. Lalouel, C. Julier, and J. Ott. 1985. Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am. J. Hum. Genet. 37:482–498.

Lawn, R. M., E. F. Fritsch, R. C. Parker, G. Blake, and T. Maniatis. 1978. The isolation and characterization of linked delta- and beta-globin genes from a cloned library of human DNA. Cell 15:1151–1174.

Leach, R., R. DeMars, S. Hasstedt, and R. White. 1986. Construction of a map of the short arm of human chromosome 6. Proc. Natl. Acad. Sci. USA 83:3909–3913.

Litt, M., and R. White. 1985. A highly polymorphic locus in human DNA revealed by cosmid-derived probes. Proc. Natl. Acad. Sci. USA 82:6206–6210.

Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, and R. White. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616–1622.

Oberle, I., D. Drayna, G. Camerino, R. White, and J. L. Mandel. 1985. The telomeric region of the human X chromosome long arm: presence of a highly polymorphic DNA marker and analysis of recombination frequency. Proc. Natl. Acad. Sci. USA 82:2824–2828.

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.

———. 1985. Analysis of human genetic linkage. Johns Hopkins University Press, Baltimore.

Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. 113:237–251.

Skolnick, M. H. 1980. The Utah genealogical data base: a resource for genetic epidemiology. Pp. 285–297 *in* J. Cairns, J. L. Lyon, and M. Skolnick, eds. Banbury report 4: cancer incidence in defined populations. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Solomon, E., and W. F. Bodmer. 1979. Evolution of sickle variant gene. Lancet 1:923.

Tsui, L. C., M. Buchwald, D. Barker, J. C. Braman, R. G. Knowlton, J. W. Schumm, H. Eiberg, J. D. Kennedy, N. Plavsic, M. Zsiga, D. Markiewicz, G. Akots, V. A. Brown, C. Helms, T. Gravius, C. Parker, K. Rediker, and H. Donis-Keller. 1985. Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. Science 230:1054–1057.

Ullrich, A., T. J. Dull, A. Gray, J. Brosius, and I. Sures. 1980. Genetic variation in the human insulin gene. Science 209:612–614.

Wainwright, B. J., P. J. Scambler, J. Schmidtke, E. A. Watson, H. Y. Law, M. Farral, H. J. Cooke, H. Eiberg, and R. Williamson. 1985. Localization of cystic fibrosis locus to human chromosome 7cen-q22. Nature 318:384–386.

White, R., M. Leppert, D. T. Bishop, D. Barker, D. Berkowitz, C. Brown, P. Callahan, T. Holm, and L. Jerominski. 1985. Construction of linkage maps with DNA markers for human chromosomes. Nature 313: 101–105.

Wijsman, E. M. 1984. Optimizing selection of restriction enzymes in the search for DNA variants. Nucleic Acids Res. 12:9209–9226.

Willard, H. F., M. H. Skolnick, P. L. Pearson, and J. L. Mandel. 1985. Report of the Committee on Human Gene Mapping by Recombinant DNA Techniques.

Eighth International Workshop on Human Gene Mapping. Cytogenet. Cell Genet. 40:360–489.

Wong, Z., V. Wilson, A. J. Jeffreys, and S. L. Thein. 1986. Cloning a selected fragment from a human DNA "fingerprint": isolation of an extremely polymorphic minisatellite. Nucleic Acids Res. 14:4605–4616.

Wyman, A. R., D. F. Barker, K. F. Wertman, C. Helms, and W. H. Petri. 1987. Factors which equalize the representation of genome segments in recombinant libraries. Gene 49:263–271.

Wyman, A. R., and R. White. 1980. A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. USA 77:6754–6758.

Wyman, A. R., L. B. Wolfe, and D. Botstein. 1985. Propagation of some human DNA sequences in bacteriophage lambda vectors requires mutant *Escherichia coli* hosts. Proc. Natl. Acad. Sci. USA 82:2880–2884.