Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 6206–6210, September 1985
Genetics

A highly polymorphic locus in human DNA revealed by cosmid-derived probes

(restriction fragment length polymorphisms/prehybridization/Southern blotting)

M. LITT* AND R. L. WHITE†

*Departments of Biochemistry and Medical Genetics, School of Medicine, Oregon Health Sciences University, Portland, OR 97201; and †Howard Hughes Medical Institute, University of Utah Medical School, Salt Lake City, UT 84132

Communicated by David Botstein, May 6, 1985

ABSTRACT **Human gene mapping would be greatly facilitated if marker loci with sufficient heterozygosity were generally available. As a source of such markers, we have used cosmids from a human genomic library. We have developed a rapid method for screening random cosmids to identify those that are homologous to genomic regions especially rich in restriction fragment length polymorphisms. This method allows whole cosmids to be used as probes against Southern transfers of genomic DNA; regions of cosmid probes homologous to repeated genomic sequences are rendered unable to hybridize with Southern transfers by prehybridization of the probes with a vast excess of nonradioactive genomic DNA. From one cosmid identified by this procedure, we have isolated three single-copy probes that collectively identify seven polymorphic loci. Of 56 unrelated individuals, 52 were heterozygous at one or more of these marker loci.**

Restriction fragment length polymorphisms (RFLPs) are useful markers for gene mapping (1). Approximately 160 RFLPs in the human genome have been reported (2), and the rate of discovery of new RFLPs continues to increase. Although most of the reported markers have but two alleles, apparently due to single-base changes affecting restriction sites, a few have multiple alleles. Since frequent heterozygosity is a critical requirement for linkage studies, the number and frequency of alleles available at a polymorphic locus are important parameters, characterizing its usefulness as a genetic marker. The more alleles present in a population at noticeable frequency, the more useful will be the locus (1). Multi-allelic loci found in the human can result from multiple independent restriction site alterations at the locus, as has been characterized for the β-globin genes (3). Alternatively, very useful multi-allelic loci have also been defined in regions that include sets of short tandem repeats that vary in number in the population, as characterized at the insulin locus (4).

As an approach to the discovery of such multiply allelic loci, we have used cosmids from a human genomic library as sources of probes. The key to this approach stems from the observation of the uniquely high frequency of base-pair change polymorphism detected by the restriction enzymes *Msp* I and *Taq* I. It has been estimated that 1/20 to 1/10 of *Msp* I and *Taq* I sites are polymorphic in the human population (5). Hence, we have developed probes that allow us to examine multiple *Taq* I and *Msp* I sites at a single locus.

In principle, the longer the probe segment, the more sites are examined. The current practical limit is the 35- to 45-kilobase-pair (kb) segment cloned in cosmid vectors (6). Using values for the average distance between restriction sites calculated by Bishop *et al.* (7), we estimate that with a frequency of polymorphism per *Msp* I or *Taq* I site of 0.05–0.10, the average number of polymorphic sites per polymorphic locus revealed by using both enzymes would vary between 2.9 and 5.6 with a 35-kb probe. Even 3 sites could give as many as eight haplotypes, which with equal frequencies would yield a highly informative marker locus with 87.5% heterozygosity; five sites could give 32 haplotypes.

This approach might not always be as effective as it might seem, since over these genetic distances substantial linkage disequilibrium may exist (3). However, the results we report in this paper as well as those of others (5, 8) suggest that the degree of linkage disequilibrium between closely linked RFLPs may often be small.

The ubiquitous presence of interspersed repeated sequences, such as the *Alu* family, in the human genome poses a practical problem that has heretofore prevented the utilization of cosmids as probes to reveal polymorphisms. Since the average distance between neighboring *Alu* sequences is about 8 kb (9), nearly all cosmid inserts are expected to contain at least one such sequence. Hence, most cosmids are expected to hybridize at numerous different genomic locations, giving rise to smears on Southern transfers that totally obscure bands due to hybridization of single-copy sequences. This problem can be circumvented by the laborious process of subcloning small segments of cosmids and using as probes subclones that hybridize only to single-copy genomic sequences. However, in this paper we describe a method by which sequences in a cosmid probe that are homologous to genomic repeats can be prevented from hybridizing with Southern transfers by prehybridization of the cosmid with a vast excess of nonradioactive sonicated total human DNA. After prehybridization, cosmid probes are capable of revealing single-copy bands without interference from genomic repeats.

MATERIALS AND METHODS

DNA Isolation. Human DNAs were prepared from leukocytes or transformed lymphoblasts (4) with the addition of a second precipitation with ethanol in the presence of 2.5 M ammonium acetate.

Blot Transfer Hybridization. Restriction enzymes were obtained from Promega Biotec (Madison, WI), New England Biolabs, and Bethesda Research Laboratories and were used according to the suppliers' instructions. Human DNAs were digested with 5–10 units of enzyme per μg of DNA and completeness of digestion was assessed by agarose gel electrophoresis of parallel digests containing λ phage DNA in addition to human DNA (5). Blot transfer hybridization was performed with nylon membranes (AMF Cuno, Meriden, CT) according to Barker *et al.* (8).

Cosmids. A partial library of human genomic DNA cloned in the cosmid vector Kos I (10) was obtained from Ed Fritsch

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RFLP, restriction fragment length polymorphism; kb, kilobase pair(s); bp, base pair(s).

(The Genetics Institute, Boston). DNA was extracted from cosmids according to Ish-Horowicz and Burke (11) and was further purified by banding in CsCl/ethidium bromide gradients, treatment with RNase A, and centrifugation through 1 M NaCl (T. Gurney, personal communication).

Cosmids were screened for their utility in revealing multiple *Taq* I and/or *Msp* I polymorphisms by a modification of the method of Ardeshir *et al.* (12). Cosmids were nick-translated (5) in the presence of ^{32}P-labeled deoxynucleoside triphosphates (New England Nuclear) to give specific activities of at least 2×10^8 dpm/μg. After removal of unincorporated radioactivity by precipitation with spermine (13), cosmid probes were mixed with a vast excess of nonradioactive sonicated [average size, 500 base pairs (bp)] human placental DNA (Calbiochem), boiled, and prehybridized to a C_0t of *ca.* 100 mol·sec/liter. Prehybridization was carried out in 0.12 M sodium phosphate, pH 7.0, at 65°C for 4–6 hr. Prehybridized cosmid probes were hybridized with Southern blots made from *Taq* I or *Msp* I digests of genomic DNAs from a panel of unrelated individuals, as described by Barker *et al.* (5). In recent experiments, some probes that gave excessive lane background when hybridized at the usual temperature of 42°C were found to give satisfactory autoradiographs when hybridized at 55°C (14). Blots were washed as described (5) with the addition of a high-stringency wash at 70°C in 15 mM NaCl/1.5 mM sodium citrate for 30 min.

Subcloning. *Sau*3A digests of cosmid 1-5 were ligated into the *Bam*HI site of the plasmid pJB-8 (11). The ligation mixture was used to transform *Escherichia coli* strain HB101. Transformants were selected on ampicillin plates and were screened by colony hybridization, using as probes total human DNA as well as appropriate restriction fragments of the cosmid.

Restriction Mapping. Maps of sites for the enzymes *Eco*RI, *Taq* I, *Kpn* I, and *Hin*dIII in cosmid 1-5 were obtained by a modification of the partial digest method of Smith and Birnstiel (15). The cosmid was linearized with *Nru* I, which cuts once within the Kos I vector but does not cut within the insert, and portions of the linearized cosmid were partially digested with the aforementioned enzymes. Southern blots of partial digests were probed with a restriction fragment from the vector that hybridized to only one end of the linearized cosmid. The resultant ladder of partial digest fragments revealed the distribution of restriction sites within the cosmid. Data from double digests and from Southern blotting experiments in which fragments made with one enzyme were used to probe digests made with other enzymes were also used to construct the restriction map.

Linkage Analysis. The families used to determine linkage relationships between marker loci are large, three-generation nuclear families as previously described (16). In most cases, determination of grandparental genotypes revealed the allele distribution (linkage phase) in the parents, allowing recombinant and nonrecombinant children to be identified by examination of their genotypes. In those cases in which parental linkage phases could not be determined, recombination frequencies were calculated on the basis of each of the two possible linkage phases. Data from both phase-known and phase-unknown cases were used to calculate lod scores (17) for linkage at a recombination fraction of zero.

RESULTS

We have screened 43 cosmids for their ability to reveal multiple *Taq* I and/or *Msp* I polymorphisms. When entire prehybridized cosmids were used as probes, 36 of these 43 probes gave autoradiographs of acceptable quality—i.e., with numerous bands and minimal lane background. Twelve of these autoradiographs showed multiple *Taq* and/or *Msp* polymorphisms. Blots probed with three of these cosmids are shown in Fig. 1. Further studies on one of these cosmids, 1-5, are described below.

Identification of Fragments Revealing Polymorphic Restriction Endonuclease Sites. Cosmid 1-5 revealed *Taq* I poly-

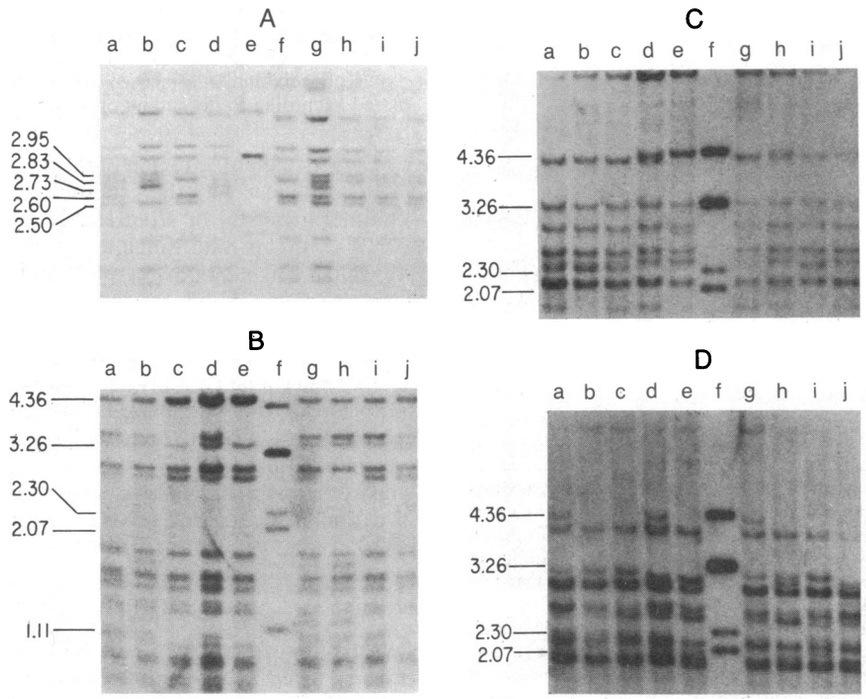

FIG. 1. Southern transfers of *Taq* I and *Msp* I digests of a panel of DNAs from unrelated individuals probed with whole cosmids prehybridized with total human DNA. (*A*) *Taq* digests probed with cosmid 1-5 (lanes a–d and f–j); lane e has size markers. (*B*) *Msp* digests probed with cosmid 1-11 (lanes a–e and g–j); lane f has size markers. (*C*) *Taq* digests (lanes a–e and g–j) probed with cosmid 1-13; lane f has size markers. (*D*) *Msp* digests (lanes a–e and g–j) probed with cosmid 1-13; lane f has size markers. Size markers were restriction fragments of pBR322 that hybridized to the cosmid vector. Sizes are indicated on the left in kb.

morphisms involving several fragments in the size range of 2.60–2.85 kb (Fig. 1). Therefore, *Taq* fragments 2.80, 2.60, and 2.55 kb long were isolated from cosmid 1-5 and used as probes of colony filters containing *Sau*3A subclones of 1-5. Duplicate filters were screened for repeated sequences, using total human DNA as a probe (5). Subclones of 1-5 that gave signals with one of the cosmid restriction fragment probes but that failed to give signals with the total human DNA probe were tested as probes against restriction digests of human genomic DNA samples from panels of unrelated individuals. Three subclones, 1-30, 1-32, and 2-96, revealed a total of seven polymorphic loci, as summarized in Table 1. Blots illustrating four of these loci are shown in Fig. 2. Evidence for the assignment of specific *Taq* fragments revealed by probe 1-32 to two distinct loci, called δ and γ, is presented in the legend to Table 1.

Restriction Map. A restriction map of cosmid 1-5 is shown in Fig. 3. Also shown are the map locations of the three fragments in subclones 1-30, 1-32, and 2-96 as well as the locations of polymorphisms. As predicted by the map, 1-30 and 1-32 hybridize to a common 13-kb *Kpn* I fragment in digests of genomic or cosmid DNA (Fig. 4). Subclone 1-32 hybridizes to a 5.9-kb *Kpn* I fragment in digests of genomic DNA but hybridizes to a 4.9-kb *Kpn* I fragment in digests of the cosmid (Fig. 4*B*). We suggest that, during construction or propagation of the cosmid library, approximately 1 kb was deleted from the region between the two leftmost *Kpn* sites of the cosmid. This interpretation is supported by the hybridization of 1-32 to a 4.5-kb *Taq* fragment in digests of cosmid DNA, about 1 kb smaller than *Taq* fragments that hybridize to this probe in genomic digests (Fig. 5). Lane a in Fig. 5 shows that 1-32 hybridizes weakly to a 3.3-kb *Taq* fragment of the cosmid (map location 12–15.3 kb). The location of this secondary site of hybridization was confirmed by probing Southern transfers of *Hin*dIII- and *Eco*RI-digested cosmid with 1-32 (data not shown).

Table 1. Probe–enzyme pairs that reveal polymorphism

Probe	Insert size, kb	Enzyme	Fragment pattern	Locus	Allele frequencies*
1-30	1.4	*Taq* I	2.60,† 2.48†	β	0.18, 0.82 (57)
		Msp I	1.87,† 1.72†	β	0.18, 0.82 (57)
		*Hin*dIII	9.3,† 4.6†	κ	0.18, 0.82 (28)
1-32	1.9	*Taq* I	5.60,† 5.50,† 5.40,† 5.00,† 3.30,	δ‡	0.32, 0.49, 0.18, 0.01 (45)
			3.05,† 2.90†	γ‡	0.82, 0.18 (53)
		*Hin*dIII	14.5,† 12.5,† 3.4	λ	0.23, 0.77 (43)
2-96	2.4	*Taq* I	2.75,† 2.60†	ϕ	0.21, 0.79 (51)
		Msp I	2.85, 0.80, 0.57,† 0.55,† 0.52,† 0.40,† 0.42, 0.35	ε	0.20, 0.20, 0.41, 0.19 (43)

The lengths of restriction fragments observed in Southern transfers of DNA from unrelated individuals after digestion with the indicated restriction enzyme are shown for each probe.

*Allele frequencies were determined by examining the number of unrelated individuals shown in parentheses. Alleles are listed in descending order of fragment lengths.

†Indicates fragments observed to vary among DNA samples tested.

‡The assignment of specific bands to the δ and γ loci was made possible by our discovery of a fourth single-copy subclone (1-25) which revealed only the δ locus. Subclone 1-25 has a 3.2-kb insert, hybridizes with the 6.8-kb *Eco*RI fragment of cosmid 1-5, and reveals a *Pst* I polymorphism with allelic fragments 2.90, 2.75, and 2.60 kb long whose presence correlates perfectly with the presence of the 5.6-, 5.5-, and 5.4-kb allelic fragments, respectively, on *Taq* I blots of the same individuals probed with subclone 1-32.

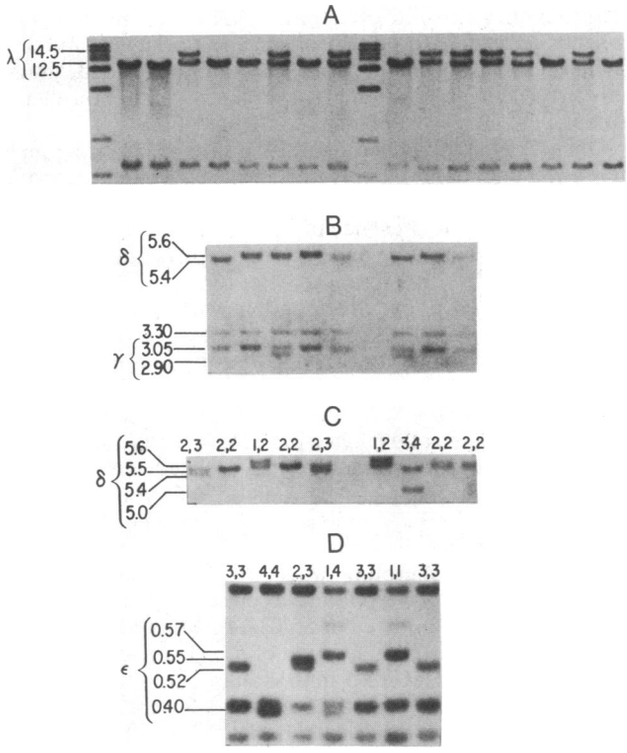

FIG. 2. RFLPs revealed by subclones of cosmid 1-5. Loci and fragment lengths (kb) are on the left. (*A*) The insert from subclone 1-32 was used as a hybridization probe on Southern transfers of *Hin*dIII-digested DNA from 16 unrelated individuals. (*B*) The insert from 1-32 was used as a hybridization probe on Southern transfers of *Taq* I-digested DNA from 8 unrelated individuals. The 0.8% agarose gel from which this blot was made had been run 24 hr at 40 V. (*C*) The 5- to 6-kb region of Southern transfers similar to those of *B* except that the gel from which this blot was made was run for 48 hr at 40 V. Genotypes of individuals are shown above the lanes. (*D*) The insert from subclone 2-96 was used as a hybridization probe on Southern transfers of *Msp* I-digested DNA from 7 unrelated individuals; the fragments were resolved on a 1.5% agarose gel. Genotypes of individuals are shown above the lanes, with alleles numbered according to decending order of fragment lengths.

Evidence that the region of the cosmid containing 2-96 and 1-30 is colinear with the genome was obtained by probing a Southern transfer of *Hin*dIII-digested genomic and cosmid DNA with an 8-kb single-copy *Eco*RI cosmid fragment, revealing fragments of 5.1, 4.8, and 4.4 kb present in both digests (data not shown).

Linkage Analysis. To examine whether the loci revealed by probes isolated from cosmid 1-5 were closely linked, we performed a linkage study using two large Utah families, each with four living grandparents and seven children. Fourteen pairs of markers were tested. The lod scores for linkage at a recombination fraction of zero reached +3.6 for the six marker pairs: β–κ, β–ε, γ–ε, ϕ–ε, λ–ε, and κ–ε. In no case did we find evidence for recombination between any of the marker loci.

DISCUSSION

We have described a set of RFLPs derived from a single cosmid insert that constitute a highly polymorphic marker in the human genome. Seven polymorphic loci are revealed by probes derived from this cosmid. Of 56 unrelated individuals examined, 52 (93%) were heterozygous at one or more of these loci, indicating that the compound locus will be a useful marker for human gene mapping.

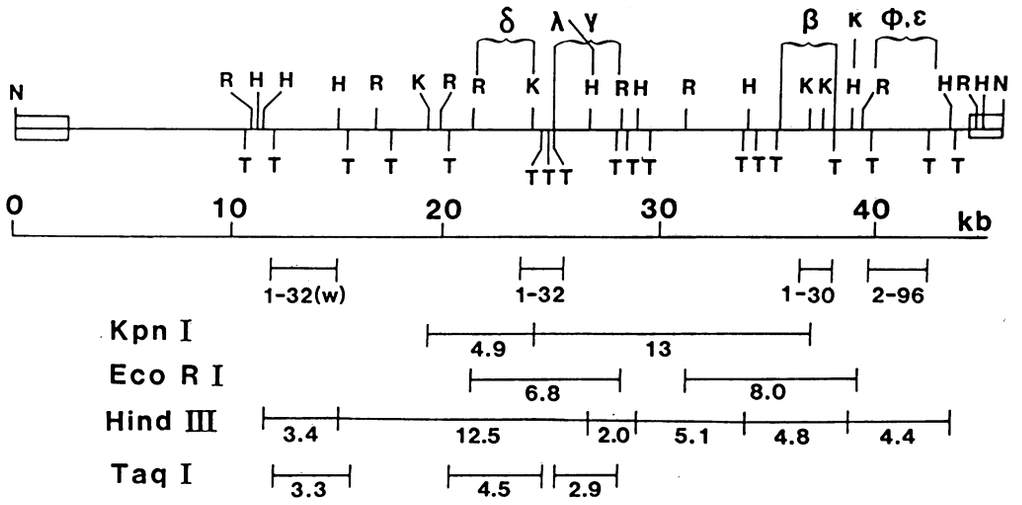

FIG. 3. Map of *Eco*RI (R), *Hin*dIII (H), *Kpn* I (K), *Taq* I (T), and *Nru* I (N) sites in cosmid 1-5. Open bars represent vector segments. Regions to which subclones 1-30, 1-32, and 2-96 hybridize are shown below the map. The region from 12 to 15.3 kb to which subclone 1-32 hybridizes [labeled 1-32(w)] has a weaker homology to this subclone than the region from 24 to 26 kb. Locations of polymorphisms are shown above the map. The κ and λ loci are *Hin*dIII site polymorphisms whose positions are precisely known. The β and δ loci are insertion/deletion polymorphisms, which map within the bracketed regions. Brackets also indicate the map locations of the δ, φ, and ε loci; it is not known whether these loci represent polymorphic restriction sites or insertion/deletion polymorphisms.

Close linkage of the RFLPs that constitute the compound locus has been verified by family studies as well as by restriction mapping. These precautions are essential because of the occasional presence in cosmid libraries of clones containing inserts derived from noncontiguous regions of genomic DNA. Indeed, cosmid 3-3B, derived from the same library as cosmid 1-5, yielded two probes that revealed polymorphisms residing on genetically unlinked loci and that were shown by restriction mapping to be noncolinear in the genome.

In studies that will be reported elsewhere, we have used a somatic cell hybrid panel to map probes 1-30, 1-32, and 2-96 to chromosome 2q. This result has been confirmed and extended by an *in situ* hybridization study with probe 2-96 showing hybridization to region 2q 35→37 (unpublished work).

The approach described here appears to be generally useful for obtaining probes capable of revealing highly heterozygous polymorphic loci in the human genome. Although single-copy probes from cosmids usually must be tested individually for their ability to reveal polymorphisms, the use of whole nick-translated cosmids after prehybridization with a vast excess of nonradioactive human DNA is extremely helpful in selecting cosmids capable of yielding several such probes. If, as in the case of cosmid 1-5, appropriate cosmid restriction fragments can be found that hybridize to single-copy subclones likely to reveal polymorphisms, screening such subclones is greatly simplified. If appropriate fragments cannot be found, it becomes necessary to screen cosmid subclones at random until the polymorphisms are revealed. We suggest that the labor of screening subclones might be considerably reduced if large (*ca.* 5- to 10-kb) restriction fragments of the cosmid were subcloned instead of the much smaller *Sau*3A fragments that we have been using. If such large insert subclones were nick-translated and prehybridized with total human DNA prior to blot hybridization, they should frequently yield autoradiographs with much simpler band patterns than those observed with whole cosmid probes, allowing polymorphic alleles to be recognized without isolation of single-copy subclones.

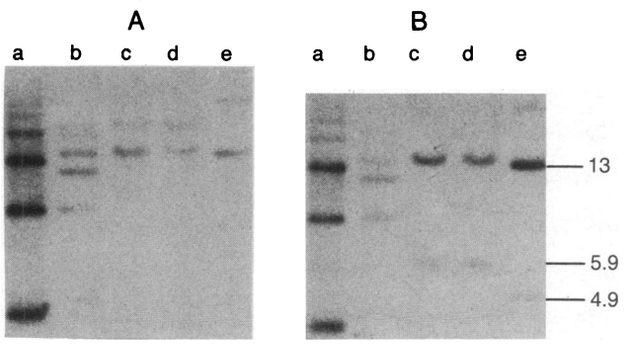

FIG. 4. Southern transfers of *Kpn* I-digested DNAs from unrelated individuals (lanes c and d) and *Kpn* I-digested DNA from cosmid 1-5 (lane e) probed with inserts from subclones 1-30 (*A*) or 1-32 (*B*). Lanes a and b contain size markers (kb).

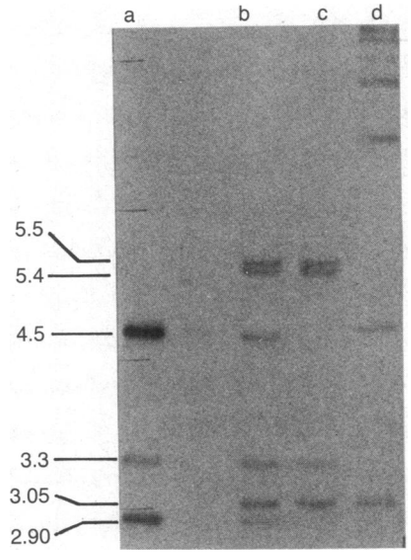

FIG. 5. Southern transfers of *Taq* I-digested DNA from cosmid 1-5 (lane a), *Taq* I-digested genomic DNA (lane c), and a mixture of *Taq* digests of genomic and cosmid DNAs (lane b) probed with subclone 1-32. Lane d contains size markers (kb).

Data to be published elsewhere indicate that cosmids 1-11 and 1-13 (Fig. 1) also yield probes that reveal compound polymorphic loci characterized by high extents of heterozygosity (unpublished work). Nine additional cosmids have revealed multiple polymorphisms when screened by using the prehybridization protocol but have not yet been complete- ly characterized. If the usefulness of these cosmids is not degraded by strong linkage disequilibrium or by cloning artefacts such as described above for cosmid 3-3B, our data suggest that approximately 25% of random human cosmids are convenient sources of probes for revealing highly polymorphic loci.

Although other approaches have been suggested for the discovery of probes that reveal highly polymorphic loci (18, 19), these approaches do not address the problem of increasing the heterozygosity at a specific locus. Solution of this problem should be greatly facilitated by use of the prehybridization protocol to screen clones selected from cosmid libraries by single-copy probes.

We thank Ruth H. Litt and Pat Dolph for excellent technical assistance. This work was supported by grant GM 32500 from the National Institute of General Medical Sciences and by the Howard Hughes Medical Institute.

1. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980) *Am. J. Hum. Genet.* **32,** 314–331.
2. Skolnick, M. H., Willard, H. F. & Menlove, L. A. (1984) *Cytogenet. Cell Genet.* **37,** 210–273.
3. Antonarakis, S. E., Boehm, C. D., Giardina, P. J. V. & Kazazian, H. H. (1982) *Proc. Natl. Acad. Sci. USA* **75,** 137–141.
4. Bell, G. I., Karam, J. H. & Rutter, W. J. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5759–5763.
5. Barker, D., Schafer, M. & White, R. L. (1984) *Cell* **36,** 131–138.
6. Collins, J. & Hohn, B. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4242–4246.
7. Bishop, D. T., Williamson, J. A. & Skolnick, M. H. (1983) *Am. J. Hum. Genet.* **35,** 795–815.
8. Barker, D., Holm, T. & White, R. (1985) *Am. J. Hum. Genet.* **36,** 1159–1171.
9. Jelinek, W. R. & Schmid, C. W. (1982) *Annu. Rev. Biochem.* **51,** 813–844.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), p. 50.
11. Ish-Horowicz, D. & Burke, J. (1981) *Nucleic Acids Res.* **9,** 2989–2998.
12. Ardeshir, R., Giulotto, E., Zieg, J., Brison, O., Liao, W. S. L. & Stark, G. R. (1983) *Mol. Cell. Biol.* **3,** 2076–2088.
13. Hoopes, B. C. & McClure, W. R. (1981) *Nucleic Acids Res.* **9,** 5493–5504.
14. Fisher, J. H., Gusella, J. F. & Scoggin, C. H. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 520–524.
15. Smith, H. O. & Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3,** 2387–2401.
16. White, R., Leppert, M., Bishop, D. T., Barker, D., Berkowitz, J., Brown, C., Callahan, P., Holm, T. & Jerominski, L. (1985) *Nature (London)* **313,** 101–105.
17. Morton, N. E. (1962) in *Methodology in Human Genetics*, ed. Burdette, W. J. (Holden Day, San Francisco).
18. Wyman, A. R. & White, R. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 6754–6758.
19. Goodbourn, S. E. Y., Higgs, D. R., Clegg, J. B. & Weatherall, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5022–5026.