Am J Hum Genet 37:635–649, 1985

A Systematic Approach for Detecting High-Frequency Restriction Fragment Length Polymorphisms Using Large Genomic Probes

JOHN FEDER,[1] LEI YEN,[1] ELLEN WIJSMAN,[1] LOUISE WANG,[1] LOUISE WILKINS,[2] JIM SCHRODER,[3] NIGEL SPURR,[4] HOWARD CANN,[1,5] MIROSLAV BLUMENBERG,[1] AND LUIGI L. CAVALLI-SFORZA[1]

SUMMARY

Thirteen phage clones containing low-copy sequences were isolated from a human DNA library and tested for their ability to detect restriction fragment length polymorphisms (RFLPs). Reported are the RFLPs revealed with each clone, all found in frequencies useful for linkage studies. Cytological data are available for five of the 13 clones, with regional assignments made for three of the markers by in situ hybridization. It is concluded that phage clones containing large unique DNA inserts detect multiple RFLPs with high efficiency.

An analysis of the relative efficiency of 20 restriction enzymes for detecting single nucleotide changes is discussed by comparing the observed data to those expected on the basis of recognition and potential site frequencies, as computed from the dinucleotide distribution. Finally, in an effort to facilitate linkage studies using polymorphic DNA sequences, experiments were made with pools of probes from various sources.

INTRODUCTION

The number of arbitrary polymorphic sequences in the human genome that have been cloned is now large enough to have made possible several recent

Received October 2, 1984; revised February 5, 1985.
This work was supported by grant 28428 from the National Institutes of Health.
[1] Department of Genetics, Stanford University, Stanford, CA 94305.
[2] Department of Medicine, Stanford University.
[3] Department of Pathology, Stanford University.
[4] Imperial Cancer Research Fund Laboratories, London, England WC2A 3PX.
[5] Department of Pediatrics, Stanford University, Stanford, CA 94305.

 0002–9297/85/3704–0002$02.00

important applications. Gusella et al. [1] used arbitrary RFLPs to demonstrate linkage to the locus of the poorly understood genetic disease: Huntington disease. Cavenee et al. [2] used a collection of arbitrary polymorphic loci from a chromosome 13 enriched library [3, 4] to elucidate some of the genetic mechanisms involved in retinoblastoma. In addition, four groups [5–8] have used DNA polymorphisms at both previously mapped known and arbitrary loci to similarly analyze the Wilms tumor locus.

Linkage detection with RFLPs has been enhanced by using specific chromosomal libraries, which are generated by either somatic cell hybrids or fluorescence-activated flow cytometry (see [9] for review). By using these techniques as well as previously cloned probes that detect RFLPs, linkage maps have been initiated on the X chromosome [10] and chromosome 11 [11, 12].

Following the discovery of the first arbitrary polymorphic locus [13], Skolnick and White discussed the strategies for detecting RFLPs [14] by considering the types of probes, restriction enzymes, and the number of individuals to be tested. Various groups have experimentally investigated the parameters involved. The types of probes that can be employed include large genomic fragments (12–20) kilobases [kb]), smaller genomic fragments (1–3 kb), and cDNA gene transcript probes (usually 1–2 kb). Schafer and White [15] and Barker et al. [16] studied small genomic probes, and, in the process, found the restriction enzymes *Msp*I and *Taq*I superior for detecting RFLPs. Helentjaris et al. [17] used cDNA clones to define DNA polymorphisms but concluded that as genetic markers these types of clones were not very productive. Recently, in a study to find RFLPs on the X chromosome, Aldridge et al. [18] suggested that the number of individuals analyzed be reduced in favor of increasing the number of restriction enzymes tested. The theoretical basis for this approach was presented as well as data supporting the selection of specific restriction enzymes.

The purpose of this study was threefold: First, it was to test the effectiveness of large genomic clones containing only low-copy number DNA sequences to detect RFLPs. In addition, to extend the usefulness of the described arbitrary polymorphic loci in linkage studies, chromosomal assignments were made for some of the probes. Second, data were collected concerning the relative efficiencies of various restriction enzymes for detecting variant sites. These data were then used to test a model proposed by one of us (E. M. W.) [19] that predicts the efficiency of restriction enzymes for detecting variant sites on the basis of their recognition and potential site frequencies as computed from the genomic dinucleotide distribution. Finally, we investigated the feasibility of using mixtures of probes, a procedure that has been also suggested by other authors for increasing the efficiency of linkage and populations studies.

MATERIALS AND METHODS

Isolation of Low-Copy Unique Sequences

The methodology used was essentially as was described by Wyman and White [13]. Recombinant phages from a Charon 4a *Eco*RI library (kindly supplied by T. Maniatis) containing inserts devoid of highly repetitive sequences were isolated by the Benton and

Davis technique [20] using 32-P-labeled total human DNA as a probe. Phages were plated at low density, appproximately 1,000 per 150-mm plate, and the resulting plaques that failed to hybridize were picked and purified through at least three rounds of plating and rehybridization. Phage DNA to be used as probes was obtained by scaling up the "plate lysate" method [21] to 150-mm plates. The phage particles were purified through a CsCl block gradient and the DNA isolated by standard procedures.

Isolation of High Molecular Weight Human DNA

The buffy coats from one unit of whole blood were obtained from the Stanford Blood Bank. The samples used represented unrelated Caucasian individuals and equal numbers of males and females. The buffy coats were resuspended in 100 ml of 50 mM Tris-HCl, pH 7.5, 1 mM EDTA, pH 7.5, 100 mM NaCl, 0.5% SDS, 100 μg/ml of proteinase K (EM Reagents, Gibbstown, N.J.) and incubated at 55°C overnight. This was followed by phenol, phenol:$CHCl_3$, $CHCl_3$ extraction and overnight dialysis against 10 mM Tris-HCl, pH.7.5, 1 mM EDTA (TE). The nucleic acids were precipitated with ethanol, collected by centrifugation, and resuspended in 20 ml of TE and treated with 100 μg/ml of RNAase A (Pharmacia, Piscataway, N.J.) for 2 hrs at 37°C. Sodium dodecyl sulfate was added to 0.5% and Proteinase K to 100 μg/ml and the solution incubated at 55°C for 1 hr. After extraction with phenol, the DNA was precipitated in the presence of 0.3 M sodium acetate and resuspended in TE. Typically, 1–2 mg of DNA were obtained.

Digestion, Electrophoresis, Transfer, Hybridization

Twenty micrograms of DNA was digested with each of 20 different restriction enzymes (New England Biolabs, Beverly, Mass.) (table 1) using the low-, medium- and

TABLE 1

SUMMARY OF THE PROBE AND ENZYME COMBINATIONS TESTED

	PROBES												
ENZYMES	2	3	4	6	10	15	18	19	24	28	32	39	42
*Apa*I		N	N	N		N	N	N	P	N	N	N	N
*Ava*II ...		N	P	P	N	N	N	N		N		N	
*Aha*II ...				N		N		N		N		N	
*Bam*HI ..	P	N	N	N	N	P	N	N	N	N	N	N	N
*Bcl*I		N	N	N		N	N	N	N	N	N	N	
*Bgl*II	P	N	N	P	N	N	P	N	N	N	N	N	P
*Ban*II ...		N	N	N		N	N	N	N	N	N	N	N
*Eco*RI ..	N	N	N	N	N	N	N	N	N	N	N	P	P
*Hinc*II ..		P	N	N	N	N	N	N	N	N	P	N	N
*Hind*III .	P	N	N	P	N	N	N	N	N	N	N	N	N
*Hph*I ...		N	P	N		N	N			N	P	N	
*Mbo*I ...				P		N		N		N	P	N	N
*Msp*I ...	P	P	N	N		N		N		N	P	N	P
*Pst*I	P	N	N	N	P	N	N	N	P	N	N	N	N
*Pvu*II ...	P	N	N	N		N	N	N	N	N	N	N	N
*Rsa*I			N	N		N		N		P	N	P	N
*Stu*I		N	N			P	N	N		N	P	N	P
*Sst*I	P	N	N	N	N	N	N	P	P	N	N	N	N
*Taq*I	P	P	P	N		N	N	P	N	N	N	P	P
*Xmn*I ...		N	N			N	N	N		P	N	N	N

NOTE: Results obtained on screening six random individuals. N = nonpolymorphic; P = polymorphic (i.e., differences between individuals noted); space = not tested.

high-buffer scheme [22]. For most enzymes, 2-μg aliquots were electrophoresed in 0.85% agarose (IBI, New Haven, Conn.) using a 40-slot comb and a 20 × 20-cm "submarine" apparatus in 89 mM Tris-borate, 89 mM boric acid, 2 mM EDTA at 0.5 V/cm. The exceptions to this procedure were with the enzymes *Mbo*I, *Rsa*I, and *Hph*I, for which 5-μg aliquots of DNA were electrophoresed in 1.2% agarose.

Capillary transfer [23] of DNA fragments was carried out in 20 × SSPE buffer (1 × = 0.18 M NaCl, 10 mM NaH_2PO_4, 1 mM Na_2EDTA, pH 7.0) onto nitrocellulose (Schleicher and Schuell, Keene, N.H.) or Zetabid (AMF Cuno, Meriden, Conn.) filters without depurination. Probes (500 ng) were labeled by nick-translation [24] with 32-P[dCTP] and 32-P[dATP] (Amersham-Searle, Arlington Heights, Ill., > 3,000 Ci/mmol) to approximately 1×10^8 cpm/μg. Filters were prehybridized (0.03 ml/cm^2) overnight at 42°C in 50% formamide (MCB, Cincinnati, Ohio), 5 × SSPE, 5 × Denhardt's solution [25], 200 μg/ml yeast RNA, 100 μg/ml salmon sperm DNA, and .1% SDS, which had been filtered through a 0.45-μm cellulose acetate membrane. Hybridization was carried out by replacing the prehybridization mixture with filtered 50% formamide, 5 × SSPE, 1 × Denhardt's solution, 100 μg/ml salmon sperm DNA, and 10 ng/ml of 32-P-labeled probe (2×10^7 cpm) and incubating at 42°C for 48 hrs. The post-hybridization washes were in 0.1 × SSPE, 0.1% SDS at 60°C, twice for 30 min. The filters were autoradiographed on Kodak XAR-5 film with Du Pont lightning plus intensifying screens at −70°C for 7 days.

Screening for RFLPs

Phage clones containing unique inserts were screened for their ability to detect high-frequency RFLPs by hybridization to DNA from six random individuals that had been digested with up to 20 different restriction enzymes. The rationale for choosing six individuals was that with this number, which made maximum use of the gel, the probability of detecting a heterozygote was greater than 90% for a polymorphism with two alleles of frequency 20% and 80%. The enzymes were selected on the basis of empirical criteria as well as on a study performed by Wijsman [19] that used dinucleotide frequencies to predict the relative efficiency of diffferent enzymes for detecting single-base changes. All polymorphisms found were tested on an average of 20 individuals to determine allele frequencies as well as compliance with the Hardy-Weinberg equilibrium. The chi-square for all tests did not reach significance. All polymorphisms described at each locus were tested in several families to establish the allelic relationships among the fragments based on the segregation patterns and to insure Mendelian inheritance.

Chromosomal Localization by in Situ Hybridization

Mitotic chromosome spreads were prepared from peripheral blood lymphocyte cultures [26]. Phage DNA to be used as probes were prepared as described above and were nick-translated with [^{3}H]dATP and [^{3}H]dTTP. In situ hybridization was carried out essentially as described by Harper and Saunders [27].

RESULTS

Screening Human Library for Phages Containing Unique Inserts

Approximately 8,000 phages yielded 13 clones that contained inserts indicative of low-copy number DNA sequences when hybridized to genomic digests. This conclusion was based on the intensity of the hybridizing bands, the total amount of DNA annealing, and segregation results. In addition, all of the phages when hybridized to human DNA that had been digested with *Eco*RI produced bands corresponding to the internal *Eco*RI fragments in the phage, indicating that at least the internal fragments were not the result of a rear-

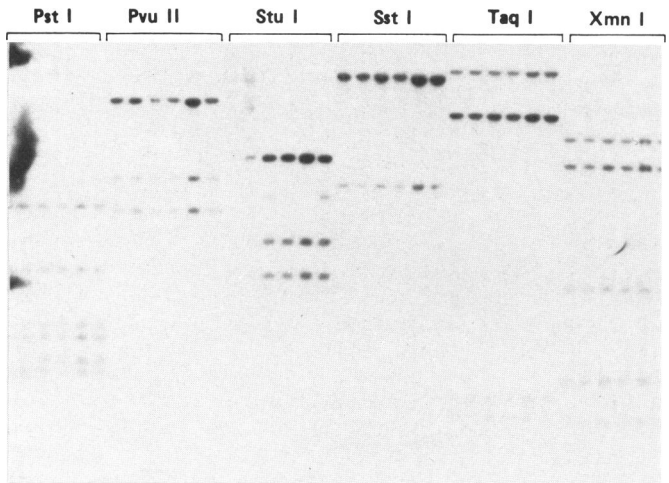

FIG. 1.—An example of an autoradiogram obtained in screening the phage probes for RFLPs. Shown here is phage 15, with variation apparent with the enzyme *Stu*I (*lanes 1* and *2* of *Stu*I contain incomplete digestion).

rangement in the library construction (data not shown). Therefore, clones containing low-copy unique DNA occurred in approximately 0.15% of the phage screened. This number is considerably lower than that predicted by Botstein (1%–3%) [28] and that found by Kao [29] (1%).

A typical result obtained in screening for RFLPs is shown in figure 1, and a summary of the probe/enzyme combinations tried and the results are presented in table 1. In table 2, a listing of the most useful polymorpisms revealed with each of the 13 probes is presented. Included in the table are the length of the probes (and subclones where applicable), the enzymes that reveal the polymorphisms, the allele sizes, and their frequencies. The "polymorphism information content" (PIC) [28] for each of the most frequent polymorphism at each locus is given in the last column. The three multiallelic polymorphisms (probe 2, 24, and32) have PIC values greater than .5, making them highly informative markers in linkages studies. The rest are all in the "reasonable" category. In addition to the data in table 2, three probes warrant further mention.

Phage 2 was the only clone for which portions were unequivocally present in the genome in more than one copy. This conclusion was based on the intensities of some of the bands, as well as on the large amount of hybridizing DNA and fragments obtained in addition to those expected from a restriction map of the clone. This 13.5-kb probe detects polymorphism at *Pvu*II, *Sst*I, *Bam*HI, *Hinc*II, and *Bgl*II sites. Probing with a 3.8-kb subclone of this region, pλ2-2, indicated that the polymorphism at these sites was being generated by a small deletion/insertion event. The three alleles present at this locus are best seen with the enzyme *Pvu*II (table 2). An additional *Pvu*II polymorphism is detected with this phage and has been localized to another subclone: pλ2-1. The polymorphism maps outside the cloned region so there is no way of knowing if the polymorphism is contiguous with the cloned segment or in a more distal copy.

TABLE 2

DESCRIPTIONS OF THE MOST FREQUENT POLYMORPHISMS DETECTED WITH THE 13 PROBES STUDIED

Probe (kb)	(Subclone) (kb)	Enzyme	Alleles (kb)	Frequency	(SE)	PIC
Phage 2	(pλ2 – 2)	*Pvu*II	2.9	.59	.047	
(13.5)	(3.8)		2.6	.19	.036	.55
			2.8	.22	.04	
	(pλ2 – 1)	*Pvu*II	6.4	.75	.04	
	(6.0)		4.2	.25	.04	
Phage 3	(pλ3 – 1)	*Msp*I	3.7	.68	.06	
(17.3)	(5.8)		2.1, 1.6	.32	.06	.35
Phage 4		*Taq*I	2.1	.41	.08	
(16.7)			1.1, 1.0	.59	.08	.36
Phage 6		*Hin*dIII	9.8	.3	.06	
(11.6)			7.8	.7	.06	.33
Phage 10		*Pst*I	3.2	.53	.05	
(12)			2.6	.47	.05	.38
Phage 15		*Stu*I	5.1	.87	.03	
(17)			4, 1.1	.13	.03	.2
Phage 18	(pλ18 – 1)	*Bgl*II	6.4	.43	.07	
(17.6)	(4.4)		4, 2.6	.58	.07	.38
Phage 19	(pλ19 – 2)	*Sst*I	3.2	.63	.06	
(7)	(1.6)		1.8, 1.2	.37	.06	.35
Phage 24		*Sst*I	9	.58	.05	
(20)			3.8	.42	.05	.36
		*Pst*I	A1 2.3	.35	.04	
			A2 1.2, 1.15	.65	.04	
			B1 1.75	.08	.027	
			B2 1.70	.08	.027	
			B3 1.55	.5	.05	
			B4 1.50	.26	.04	
			B5 1.45	.08	.027	.62
Phage 28		*Xmn*I	7	.75	.05	
(17.4)			6, 1	.25	.05	.31
Phage 32	(pλ32 – 1)	*Msp*I	A1 11	.50	.063	
(13.3)	(8.9)		A2 7.6, 6, 4	.18	.048	
			A3 6, 4, 3	.32	.059	.54
		*Mbo*I	2.9	.32	.06	
			2.6	.68	.06	
Phage 39	(pλ39 – 1)	*Taq*I	A1 5.8	.79	.07	
(15)	(4.4)		A2 3.0, 2.8	.21	.07	
			B1 1.95	.41	.07	
			B2 1.3, .65	.59	.07	.36
Phage 42		*Taq*I	4.8	.37	.08	
(10.4)			3.1	.63	.08	.35

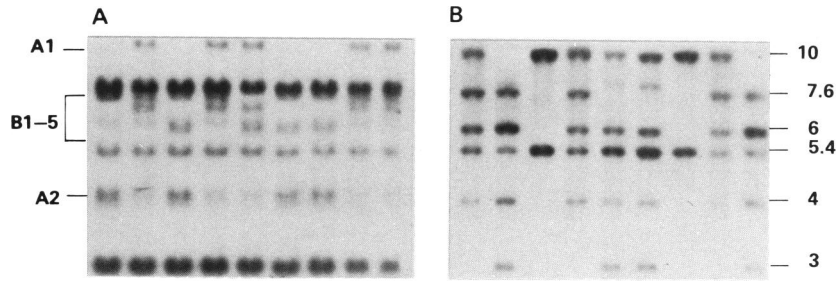

FIG. 2.—Autoradiograms of polymorphisms detected with phage 24 and 32. *A*, Phage 24 hybridized to a *Pst*I genomic digest of random individuals. *B*, Phage 32 hybridized to an *Msp*I genomic digest of random individuals.

However, family analysis between this locus and the polymorphism at pλ2-2 has shown no crossing over in the informative matings studied so far (10 gametes tested). That the putative multiple copies of phage 2 reside on the same chromosome and are very close, so that they are useful in linkage studies, has been demonstrated by both in situ hybridization and somatic cell hybrids (see below). This polymorphism was reported at the last gene mapping conference and was assigned the locus designation of D6S2 [9].

Phage 24 revealed polymorphism to two *Sst*I sites, the most frequent one showing alleles of 9 and 3.8 kb. *Pst*I also reveals two variable regions; in one, a small alteration in a recognition site produces one allele of 2.3 kb and another of 1.25 and 1.05 kb. In addition, a region between 1.75 and 1.45 kb shows high variability (fig. 2). Individuals have been observed to possess one to four bands differing in size from 50 to 100 bases. The intensity of the four bands is variable and is presumed to reflect different copy numbers. Although detailed restriction mapping of this probe has not been performed, family studies have shown the fragments to be allelic and inherited in a Mendelian fashion (data not shown).

With phage 32, frequent heterogeneity is observed with the enzyme *Mbo*I, producing alleles of 2.9 and 2.6 kb. *Msp*I reveals three strikingly different alleles (fig. 2). One has molecular length of approximately 10–11 kb, which is uninterrupted by any *Msp*I sites. Another allele contains two sites producing fragments of 7.6, 6, and 4 kb. In the third type, the 7.6-kb fragment is lost and a 3-kb fragment appears.

A tabulation of the number of recognition sites observed for each enzyme (number of fragments observed with a probe plus one), assuming each fragment represents a single copy, and of those found to be polymorphic is presented in table 3. Omitted are the polymorphisms due to insertion/deletion events. The observed data were fitted to the model suggested by Wijsman [19]. This model allows one to estimate the relative efficiencies of restriction enzymes in detecting single nucleotide differences on the basis of: (1) the total amount of DNA tested as determined by the probe lengths listed in table 2, (2) the enzymes actually tested for each probe (table 1), and (3) the predicted distribution of restriction and potential sites (restriction sites with one base difference) cal-

TABLE 3

SUMMARY OF RESTRICTION ENZYME RECOGNITION SITES OBSERVED (NO. FRAGMENTS PLUS ONE) AND THE NO. SITES OBSERVED TO BE VARIANT (SINGLE SITE CHANGES ONLY)

Enzyme	No. sites observed	No. sites variant	No. expected variant	χ^2
*Msp*I	64	7	3.85	2.50
*Taq*I	77	7	5.06	.75
*Bgl*II	65	5	1.68	6.54
*Sst*I	75	5	1.64	6.89
*Mbo*I	46	4	2.47	.95
*Stu*I	47	3	1.51	1.48
*Rsa*I	45	2	3.25	.48
*Hph*I	43	2	2.50	.10
*Pst*I	78	2	1.92	.00
*Ava*II	62	2	3.59	.71
*Hinc*II	90	2	3.13	.41
*Xmn*I	60	1	1.24	.05
*Bam*HI	52	1	1.64	.25
*Hin*dIII	72	1	1.92	.44
*Eco*RI	72	1	1.58	.21
*Pvu*II	76	1	1.78	.35
*Apa*I	45	0	1.80	1.8
*Aha*II	30	0	0.75	.75
*Bcl*I	42	0	1.44	1.44
*Ban*II	63	0	3.92	3.92
Total	1,204	46	46	30.08

NOTE: The expected no. variant sites was calculated by the Wijsman method [19]. Contribution to χ^2 of individual enzymes are given in the last column.

culated from the dinucleotide frequencies listed in table 4. The relative efficiencies thus obtained (see Wijsman [19] for details of the calculation) were converted into expected numbers of variant sites by rescaling them so as to make them directly comparable with the observed numbers (table 3). The correlation coefficient between the observed and expected numbers of variant sites is .49, which is significant at $P = .0266$. The observed numbers of RFLPs detected with the enzymes are significantly different from expectation only at a marginal level as indicated with a test ($\chi^2 = 30.88$, with 19 degrees of freedom

TABLE 4

DINUCLEOTIDE FREQUENCIES DERIVED FROM THE GENBANK SEQUENCE DATABASE USED IN CALCULATING THE PROBABILITIES OF RESTRICTION AND POTENTIAL SITES OCCURRENCE ON THE BASIS OF THE METHOD OUTLINED IN [19]

	SECOND BASE			
FIRST BASE	A	T	G	C
A	.0747	.0594	.0685	.057
T	.0490	.0684	.0713	.0638
G	.0608	.0455	.0664	.0575
C	.0751	.0794	.0239	.0794

TABLE 5

CHROMOSOMAL ASSIGNMENTS OF PHAGE PROBES

Probe	Somatic cell hybrid assignment	In situ assignment
Phage 2	6p21-6qter	6qter
Phage 6	7pter-7q22	7p11-p15
Phage 10	15	Not done
Phage 15	15	15q12-q24 and 13q
Phage 32	11	Not done

[df], $P = .05$). It should be noted that the test was applied to data where the expected number of observations was very low, thereby tending to increase the P value. A comparison of observed vs. expected variants under the assumption that the number of variants is Poisson-distributed yields three enzymes with individual P values less than .05; none of these is as low as the necessary rejection level of $P = .05/20 = .0025$, which takes into account the fact that 20 enzymes are being independently tested [30]. In fact, three out of 20 are expected to deviate at the 5% level or, less approximately, 7.5% of the time under the null hypothesis.

Chromosomal Localization

Five of the polymorphic loci described above were assigned to chromosomes, and three of the five were regionally localized. Two techniques were employed: (1) assignment with somatic cell hybrids and (2) in situ hybridization to metaphase chromosomes. The results are in table 5, and details of the somatic cell hybrid assignment will be published elsewhere.

Three of the somatic cell hybrid assignments have been confirmed by in situ hybridization. Figure 3 shows the distribution of the grains on the chromosomes obtained with each probe. The expected distribution, as indicated by the continuous line, was calculated on the basis of the observed total number of grains and the relative length of each chromosome in a metaphase spread [31]. Phage 2 was localized to 6qter, where in 10 metaphase spreads analyzed, 29% (11/38) of the total grains could be accounted for. The χ^2 test for agreement between the observed and expected grain distribution on all chromosomes was 66.2 with 22 df ($P < .001$). When chromosome 6 was excluded, the χ^2 dropped to 18.2 with 21 df (not significant). This shows that chromosome 6 is the only one with a high significant proportion of grains. The assignment of phage 2 to chromosome 6 is strongly backed by the data, and there is no indication of affinity of the probe to any other chromosome since the distribution of the grains on the other chromosomes is as expected. Phage 6 was mapped to 7p15-p11, where in 20 metaphase spreads, 29% of the total grains were observed (20/68). The χ^2 test including chromosome 7 is 92.5 with 22 df ($P < .001$), and without chromosome 7, χ^2 is 15.2 with 21 df (not significant). Thus, the distribution of the grains on all chromosomes except no. 7 is as expected. Phage 15

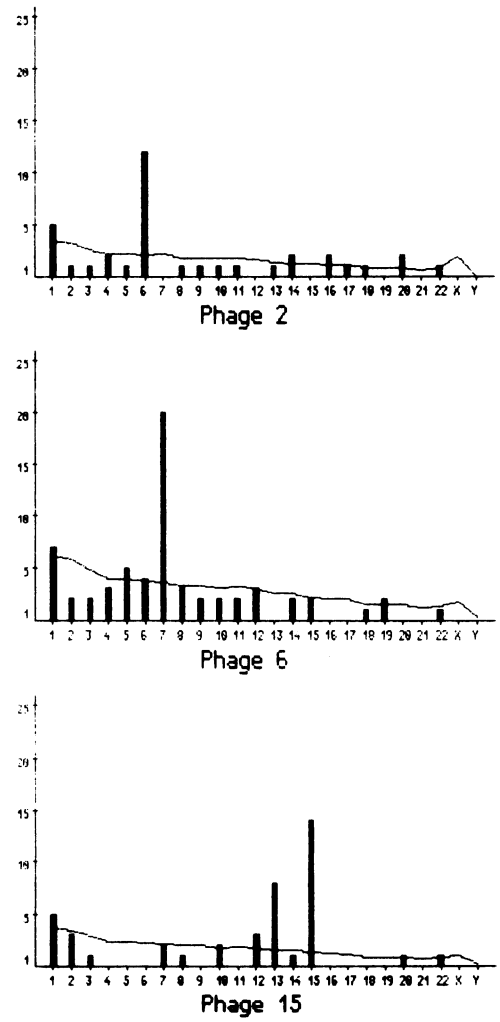

FIG. 3.—Histograms of the grain distribution obtained for the three probes mapped by in situ hybridization. The continuous line represents the expected no. grains for each chromosome.

appears to have homology with sequences on two chromosomes. It was localized to 15q12-q24, where 34% of the total grains in 10 metaphase spreads were accounted for (14/41) and to chromosome 13q21-ter, where 20% of the grains could be found (8/41). Here, the χ^2 with chromosome 15 is 162.3 with 22 df, and without it, it is 68.7 with 21 df, which is still highly significant.

Probe Mixture Hybridization

To test the feasibility of mixing several subclones polymorphic for the same enzyme that possess nonoverlapping hybridizing fragments, the following probes were combined in equal proportions, nick-translated, and hybridized to

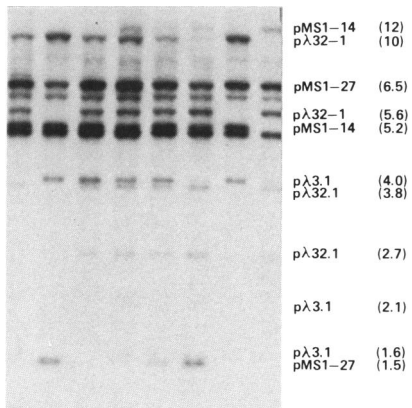

FIG. 4.—An example of an autoradiograph obtained using a pool of plasmid probes hybridized to an *Msp*I genomic digest of unrelated individuals. Included are pλ3−1, pλ32−1, pMS1-14, and pMS1-27 (pMS series probes kindly provided by R. White).

an *Msp*I genomic digest: pλ3-1, pλ32-1, pMS1-27, pMS-14 (pMS series probes provided by R. White). An example of the resulting autoradiograms is present in figure 4.

DISCUSSION

Here, we report 13 restriction fragment length polymorphic loci that display frequencies useful as markers for the human linkage map. In addition, we present cytological data that provisionally assign two of the loci and confirm assignments for three others. Our efforts have been concentrated on long stretches of unique sequences found in the Maniatis *Eco*RI library. Our previous experience with the use of small 1–2-kb unique sequence probes proved to be considerably less effective in rapidly defining RFLPs.

Our results demonstrate that large genomic probes are highly efficient in detecting RFLPs under the condition used. Even though this is intuitively obvious and was the basis for the first arbitrary polymorphism found [13], it had not been shown in large-scale screenings, nor have screenings for polymorphisms generally used larger probes. The usefulness of large probes is emphasized by comparing the data to a study conducted by Aldridge et al. [18] that defined RFLPs along the X chromosome and used procedures similar to those used here except for differences in the number of individuals screened. When one compares the enzymes used in our study and that by Aldridge et al., the latter authors found that with a total of 22 probes from the X chromosome and chromosome 13 nine reveal polymorphisms and 13 do not. In our study, all 12 single-copy phage probes tested detect polymorphisms. This difference is highly significant ($\chi^2 = 12.2$, $P < .001$, df = 1), indicating that if the X chromosome and autosomes have similar levels of polymorphism larger probes are substantially better at detecting polymorphisms than are smaller probes.

It was our desire to define RFLPs with all of the phage probes isolated as

considerable work is invested in obtaining such clones. Work of others [15, 16] indicated that some restriction enzymes detected RFLPs more readily than others. We wanted to investigate and extend this analysis. In presenting our data, we choose to list all results, polymorphic and nonpolymorphic, with each probe/enzyme combination tested (table 1). We feel that these data are valuable for future analysis in evaluating not only the efficiency of various enzymes in detecting polymorphisms but also estimates of the overall heterozygosity at the sequence level [32]. When these data are analyzed with the model described by Wijsman [19], the correlation coefficient between the observed and expected numbers indicates that the model provides a reasonable predictor of the relative abilities of different enzymes to detect variability. Larger samples of data taken from the literature give an even higher correlation: .84 [19]. Although the χ^2 test for goodness of fit is barely significant at the 5% level, 45% of the total can be contributed by the enzymes *Sst*I and *Bgl*II. In addition, when each of these enzymes is tested individually for agreement with the expectation according to a Poisson distribution, the probability of a deviation as large or larger than observed is smaller than 5% for the three enzymes *Sst*I, *Bgl*II, and *Ban*II. Even though one enzyme out of 20 is expected to deviate at the 5% level, the probability that three out of 20 deviate to at least this degree is about 7.5%. This suggests but does not prove that *Sst*I and *Bgl*II might detect single nucleotide changes more readily and *Ban*II less readily than anticipated by the model. However, none of these enzymes deviates significantly enough from expectation to conclude that it is really aberrant. The suggestion that *Bgl*II may be efficient in finding RFLPs has previously been made by Pearson [33], but neither *Sst*I nor *Bgl*II shows a significant excess in its ability to detect RFLPs in the compilation made by Wijsman [19]. The data also suggest that with continued testing the enzyme *Stu*I may have enhanced ability to detect single nucleotide changes. It has been suggested that enzymes with recognition sequences of three purines, three pyrimidines (like *Bgl*II, *Sst*I, and *Stu*I) readily detect polymorphism [18]. A mechanism for this phenomenon has not been proposed; however, the model by Wijsman indicates that it is probably not due to the respective dinucleotide frequencies of the enzymes' recognition sequences or the frequencies of their potential restriction sites. It must be noted that the other enzymes with this type of recognition structure, namely, *Bam*HI, *Ban*II, *Eco*RI, *Hin*dIII, and *Pvu*II, do not in our current analysis exhibit any special quality in defining RFLPs. The model predicts, as suggested by others [14], that the enzymes *Msp*I and *Taq*I are efficient in detecting single nucleotide changes because of their particular recognition and potential site sequences with respect to the genomic composition. Our results for these enzymes, which are in close agreement with those of Barker et al. [16], indicate that they are indeed efficient, but not significantly more than the model predicts.

Included in the study were several enzymes with frequently occurring recognition sites: *Mbo*I, *Rsa*I, and *Hph*I. Although approximately one-half of the potential information generated with these enzymes is lost because of undetectable fragments, usable polymorphisms have been found with each. Therefore, we are reluctant to reject these enzymes in spite of the technical difficulties involved with their use.

The technique of in situ hybridization has confirmed and regionally localized three of the five assignments made by somatic cell hybrids. To our knowledge, phage 2(D6S2), which is mapped to the distal end of 6q, represents a marker in a region previously unavailable for linkage studies. Phage 6, which has been mapped to the short arm of chromosome 7, represents the second arbitrary clone to be defined on this chromosome and the first to be assigned regionally to the short arm. The results with phage 15 suggest the existence of two copies residing on different chromosomes, 15 and 13, a fact confirmed by repeated independent hybridization. However, Southern blot analysis does not allude to such an arrangement in respect to the number of fragments obtained with any given enzyme, the total amount of DNA hybridized, or the segregation results with polymorphisms. The possibility exists that under the reduced stringency of the in situ hybridization technique (done at 37°C), a partially homologous sequence was detected. From results obtained with somatic cell hybrids (to be discussed elsewhere), it is more likely that polymorphisms detected with this probe are located on chromosome 15. Phage 32 and 10 have been provisionally assigned to chromosome 11 and 15, respectively, by somatic cell hybridization.

The polymorphisms described in this paper are mostly the result of sequence alterations that affect one restriction site. Only two of the probes (phage 2 and phage 24) detect polymorphisms that can be attributed to DNA rearrangements. Even in these cases, the unit of change is small, on the order of 50–100 bases, differences that are at the limit of resolution under the conditions employed. This result supports the conclusion that major DNA rearrangements like those found with D14S1 [13] and the region 5′ to the insulin gene are rare [34].

Finally, several of the phage clones have been subcloned and the segment detecting polymorphism identified. Eliminating nonpolymorphic hybridizing fragments from each clone facilitates our intentions of combining several subclones polymorphic for the same enzyme in multiprobe hybridizations, decreasing by an important factor the labor involved in linkage studies with arbitrary polymorphic probes and undefined loci. In our experience, it is clear that polymorphic clones can be organized into "pools" and tested as such, without the need of ligating probes into a single DNA fragment. This is in agreement with results by other investigators (R. White, personal communication, 1984). We are currently pursuing this line of investigation.

In summary, we conclude that phage probes containing totally unique sequences are capable of defining multiple polymorphic restriction sites efficiently, thereby increasing the chances of finding a frequent variant and extending their uselfulness in linkage studies.

The generation of random polymorphic loci in the human genome by recombinant DNA techniques will undoubtedly be of enormous benefit in the study of human genetics. It is our hope that the polymorphisms described here will contribute to this advancement.

ACKNOWLEDGMENTS

We gratefully acknowledge Berta Young and Bing Huey for their assistance with DNA extraction and blotting, Juliana Hwang for her help with computer graphics, and

Mary Harper for her assistance in setting-up the in situ hybridization technique. The genBank℠ sequences are publicly available: GenBank, c/o Computer Systems Division, Bolt Beranek and Newman, Inc., 10 Moulton St., Cambridge, MA 02238.

REFERENCES

1. GUSELLA JF, WEXLER NS, CONNEALLY PM, ET AL.: A polymorphic marker genetically linked to Huntington's disease. *Nature* 306:234–238, 1983
2. CAVENEE WK, DRYJA TP, PHILLIPS RA, ET AL.: Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* 305:779–784, 1983
3. CAVENEE WK, LEACH R, MOHANDAS T, PEARSON P, WHITE R: Isolation and regional localization of DNA segments revealing polymorphic loci from human chromosome 13. *Am J Hum Genet* 36:10–24, 1984
4. DRYJA TP, RAPAPORT JM, WEICHSELBAUM R, BRUNS GA: Chromosome 13 restriction fragment length polymorphisms. *Hum Genet* 65:320–324, 1984
5. KOUFOS A, HANSEN MF, LAMPKIN BC, ET AL.: Loss of alleles at loci on human chromosome 11 during genesis of Wilms' tumour. *Nature* 309:170–172, 1984
6. ORKIN SH, GOLDMAN DS, SALLAN SE: Development of homozygosity for chromosome 11p markers in Wilms' tumour. *Nature* 309:172–174, 1984
7. REEVE AE, HOUSIAUX PJ, GARDNER RJM, CHEWINGS WE, GRINDLEY RM, MILLOW LJ: Loss of a Harvey ras allele in sporadic Wilms' tumour. *Nature* 309:174–176, 1984
8. FEARON ER, VOGELSTEIN B, FEINBERG AP: Somatic deletion and duplication of genes on chromosome 11 in Wilms' tumours. *Nature* 309:176–178, 1984
9. SKOLNICK MH, WILLARD HF, MENLOVE LA: Report of the Committee on Human Gene Mapping by Recombinant DNA Techniques. *Cytogenet Cell Genet* 37:210–273, 1984
10. DRAYNA D, DAVIES K, HARTLEY D, ET AL.: Genetic mapping of the human X chromosome by using restriction fragment length polymorphisms. *Proc Natl Acad Sci USA* 81:2836–2839, 1984
11. GERHARD DS, KIDD KK, HOUSMAN D, GUSELLA JF, KIDD JR: Data on the genetic map of the short arm of chromosome 11(11p). *Cytogenet Cell Genet* 37:478, 1984
12. ANTONARAKIS SE, PHILLIPS JA, MALLONEE RL, ET AL.: β-Globin is linked to the parathyroid hormone (PTH) locus and lies between the insulin and PTH loci in man. *Proc Natl Acad Sci USA* 80:6615–6619, 1983
13. WYMAN AR, WHITE R: A highly polymorphic locus in human DNA. *Proc Natl Acad Sci USA* 77:6754–6758, 1980
14. SKOLNICK MH, WHITE R: Strategies for detecting and characterizing restriction fragment length polymorphisms (RFLPs). *Cytogenet Cell Genet* 32:58–67, 1982
15. SCHAFER M, WHITE R: Three random loci in the human genome with base pair change polymorphisms. *Cytogenet Cell Genet* 32:314–315, 1982
16. BARKER D, SCHAFER M, WHITE R: Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. *Cell* 36:131–138, 1984
17. HELENTJARIS T, GESTELAND R: Evaluation of random cDNA clones as probes for human restriction fragment length polymorphisms. *J Mol Appl Genet* 2:237–247, 1983
18. ALDRIDGE J, KUNKEL L, BRUNS G, ET AL.: A strategy to reveal high-frequency RFLPs along the human X chromosome. *Am J Hum Genet* 36:546–564, 1984
19. WIJSMAN EM: Optimizing selection of restriction enzymes in the search for DNA variants. *Nucleic Acid Res* 12:9209–9226, 1984
20. BENTON WD, DAVIS RW: Screening λgt recombinant clones by hybridization to single plaques in situ. *Science* 196:180–182, 1977
21. MANIATIS T, FRITSCH EF, SAMBROOK J: *Molecular Cloning, A Laboratory Manual.* Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory, 1982, pp 371–372
22. MANIATIS T, FRITSCH EF, SAMBROOK J: Ibid., p 104

23. SOUTHERN EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503–517, 1975
24. RIGBY PWJ, DIECKMANN M, RHODES C, BERG P: Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J Mol Biol* 113:237–251, 1977
25. DENHARDT D: A membrane filter technique for the detection of complementary DNA. *Biochem Biophys Res Commun* 23:641–652, 1966
26. YUNIS JJ, CHANDLER ME: High resolution chromosome analysis in clinical medicine, in *Progress in Clinical Pathology*, vol VII, edited by STEFANINI M, HOSSAINI AA, ISENBERG HD, New York, Grune & Stratton, 1977, pp 267–288
27. HARPER ME, SAUNDERS GF: Localization of single copy DNA sequences on G band human chromosomes by in situ hybridization. *Chromosoma* 83:431–439, 1981
28. BOTSTEIN D, WHITE R, SKOLNICK M, DAVIS RW: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331, 1980
29. KAO F, HARTZ JA, LAW ML, DAVIDSON JN: Isolation and chromosomal localization of unique DNA sequences from a human genomic library. *Proc Natl Acad Sci USA* 79:865–869, 1982
30. SNEDECOR GW, CHOCHRAN WG: *Statistical Methods*, 7th ed. Ames, Iowa State Univ. Press, 1980, pp 166–167
31. FRANCKE U, OLIVER N: Quantitative analysis of high-resolution trypsin Giemsa bands on human prometaphase chromosomes. *Hum Genet* 45:137–165, 1978
32. COOPER DN, SCHMIDTKE J: DNA restriciton fragment length polymorphisms and heterozygosity in the the human genome. *Hum Genet* 66:1–16, 1984
33. PEARSON PL, BAKKER E, FLAVELL RA: Considerations in designing an efficient strategy for localizing unique sequence DNA fragments to human chromosomes (abstr.). *Cytogenet Cell Genet* 32:308, 1982
34. BELL GI, KARAM JH, RUTTER WJ: Polymorphic DNA region adjacent to the 5′ end of the human insulin gene. *Proc Natl Acad Sci USA* 78:5759–5763, 1981

Erratum

In the paper "The Use of Multiple Restriction Fragment Length Polymorphisms in Prenatal Risk Estimation. I. X-Linked Diseases," by A. G. Clark (*Am J Hum Genet* 37:60–72, 1985), equation (13) that appeared on p. 67 was printed incorrectly. The corrected equation (13) appears below.

$$P(\text{phase}_i \mid \text{observed family}) = \frac{P(\text{observed family} \mid \text{phase}_i)P(\text{phase}_i)}{\sum_{j=0}^{NP} P(\text{observed family} \mid \text{phase}_j)P(\text{phase}_j)} \quad (13)$$