Am J Hum Genet 37:619–634, 1985

Extensive Restriction Site Polymorphism at the Human Phenylalanine Hydroxylase Locus and Application in Prenatal Diagnosis of Phenylketonuria

ALAN S. LIDSKY,[1] FRED D. LEDLEY,[1] ANTHONY G. DiLELLA,[1]
SIMON C. M. KWOK,[1] Stephen P. Daiger,[2] KATHRYN J. H. ROBSON,[3]
AND SAVIO L. C. WOO[1]

SUMMARY

A total of 10 restriction site polymorphisms have been identified at the human phenylalanine hydroxylase locus using a full-length human phenylalanine hydroxylase cDNA clone as a hybridization probe to analyze human genomic DNA. These polymorphic patterns segregate in a Mendelian fashion and concordantly with the disease state in various PKU kindreds. The frequencies of the restriction site polymorphisms at the human phenylalanine hydroxylase locus among Caucasians are such that the observed heterozygosity in the population is 87.5%. Thus, most families with a history of classical phenylketonuria can take advantage of the genetic analysis for prenatal diagnosis and carrier detection of the hereditary disorder.

INTRODUCTION

The major metabolic pathway for phenylalanine is conversion to tyrosine by the hepatic enzyme phenylalanine hydroxylase (PAH). Deficiency of PAH causes an accumulation of phenylalanine and its normally minor metabolites such as phenylpyruvic acid in serum and tissues [1]. Severe PAH deficiency,

Received October 16, 1984; revised February 1, 1985.

This work was partially supported by grant HD-17711, and grant AM-30471 (to S. P. D.), from the National Institutes of Health and grant 1-566 (to S. L. C. W.) from the March of Dimes. A. G. D. is the recipient of fellowship award HD-06495 from the National Institutes of Health. S. L. C. W. is an Investigator of the Howard Hughes Medical Institute.

[1] Howard Hughes Medical Institute, Department of Cell Biology, Baylor College of Medicine, Houston, TX 77030.

[2] Graduate School of Biomedical Sciences, The University of Texas Health Science Center, Houston, TX 77030.

[3] Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, United Kingdom.

 0002-9297/85/3704-0001$02.00

known clinically as phenylketonuria or PKU, is a genetic disorder that causes permanent mental retardation in children if left untreated [2]. Early diagnosis by newborn screening [3] followed by dietary intervention initiated within the first days of life can prevent postnatal brain damage and the mental retardation process associated with the disorder [4–7]. The low phenylalanine diet, however, must be implemented rigidly throughout the first decade of life to be effective, and there is a current trend to continue the diet indefinitely [8]. PKU is inherited as an autosomal recessive trait with a prevalence of about one in 10,000–15,000 births among Caucasians [9, 10].

PAH cDNA clones have been isolated from rat and human liver cDNA libraries [11, 12]. We recently reported the use of human PAH cDNA clones representing the 3′ half of the mRNA to identify restriction fragment length polymorphisms (RFLPs) at the PAH locus in man [12]. The polymorphisms, found using the restriction enzymes *Msp*I, *Sph*I, and *Hin*dIII, have been applied to trace the transmission of mutant PAH genes in informative PKU families with one or more affected children. These experiments demonstrated that within individual families, the mutant PAH alleles segregated concordantly with the disease state [12]. The frequencies of the three restriction site polymorphisms in the PAH gene detected by these partial-length cDNA clones are such that up to 75% of the PKU families are informative for genetic analysis [12]. The 75% estimate is theoretical and does not take into consideration the possibility of linkage disequilibrium between the polymorphic sites. If there is linkage disequilibrium, the percentage estimate would be significantly reduced. We recently isolated a full-length cDNA clone for human PAH and used it to identify additional RFLPs at the PAH locus. These additional RFLPs have enabled RFLP haplotype analysis of the PAH gene in 87.5% of PKU families and can be used for prenatal diagnosis and carrier detection of the genetic disorder in the population.

MATERIALS AND METHODS

The Recombinant Plasmid phPAH247

Plasmid phPAH247 is a recombinant containing a full-length human cDNA for PAH, which was identified from a human liver cDNA library. Its sequence characterization will be reported [13]. The cDNA insert was subcloned into the *Eco*RI site of pBR322 and isolated by digestion of phPAH247 DNA with *Eco*RI followed by preparative electrophoresis in low-melting agarose gels. The DNA was extracted and nick-translated to a specific activity of 2–3 $\times$ 10^8dpm/μg [14].

Southern Blotting and Hybridization

Total human genomic DNA was purified from peripheral leukocytes isolated as buffy coats from whole blood as described [12]. Ten micrograms of genomic DNA was digested to completion with various restriction enzymes in appropriate buffers. Agarose gel electrophoresis and transfer of DNA to nitrocellulose paper were performed as described by Southern [15]. Prehybridization and hybridization were done in 45% formamide, 5 $\times$ Denhardt's solution, 4 $\times$ SSC, 0.1 M sodium phosphate, pH 6.5, 0.1% SDS, 0.1% sodium pyrophosphate, and 250 μg/ml sheared, denatured herring sperm DNA at 42°C. Hybridization was done in the presence of 10% dextran sulfate and 2 $\times$ 10^6 cpm/ml of the ^{32}P-labeled cloned cDNA probe. Nitrocellulose papers were washed

twice in 2 × SSC, 0.5% SDS for 15 min at room temperature and twice in 0.1 × SSC, 0.5% SDS for 2 hrs at 68°C, followed by autoradiography for 1–5 days.

RESULTS

A Full-Length Human PAH cDNA Clone

phPAH247 is a recombinant plasmid that contains the full-length human cDNA for PAH inserted into pBR322. The cDNA insert is approximately 2.4 kilobases (kb) in length. It contains the entire coding region, 5′ and 3′ untranslated sequence, and a stretch of polyA's. Southern analysis of total human genomic DNA digested with a number of restriction enzymes using phPAH247 as the hybridization probe indicated that the chromosomal PAH gene is large and contains multiple intervening sequences. The cloning and preliminary characterization of the chromosomal PAH gene has shown that all fragments detected by Southern analysis of genomic DNA arise from a single-copy gene that is about 100 kb in length (A. G. DiLella, unpublished results, 1985).

Multiple RFLPs at the Human PAH Locus

High molecular weight genomic DNA isolated from peripheral leukocytes of 18 unrelated normal Caucasian individuals was digested with a number of restriction enzymes followed by Southern blot analysis using phPAH247 as the hybridization probe. With 18 random individuals, the probability of detecting a restriction site polymorphism with a frequency of 10% or greater is 98.5%. The list of enzymes used for the analysis include the following: *Ava*II, *Bam*HI, *Bcl*I, *Bgl*II, *Bst*NI, *Eco*RI, *Eco*RV, *Hae*III, *Hin*cII, *Hin*dIII, *Kpn*I, *Msp*I, *Pst*I, *Pvu*II, *Sau*96AI, *Scr*FI, *Sph*I, *Sst*II, *Taq*I, *Xba*I, and *Xmn*I. In addition to *Sph*I, *Hin*dIII, and *Msp*I reported previously [12], five other enzymes that yield RFLPs were observed:

*Eco*RV: phPAH247 hybridizes to three fragments common to all individuals, plus a fourth fragment either 30 kb or 25 kb in length (fig. 1*A*). Individuals homozygous for the 30-kb fragment (fig. 1*A*, lane 1) lack a polymorphic *Eco*RV restriction site within the fragment, and those homozygous for the 25-kb fragment (fig. 1*A*, lane 2) have the polymorphic *Eco*RV site in each of their PAH genes. Heterozygotes for the restriction site have also been detected (fig. 1*A*, lane 3). The differential 5-kb fragment not detected by the full-length cDNA probe would indicate that it is comprised of either totally intronic or flanking DNA sequences. The other possibility is that the polymorphism is the result of insertion or deletion of a 5-kb segment within the hybridizing bands. These are phenomena that can also be responsible for polymorphisms observed with other enzymes.

*Eco*RI: Among the multiple *Eco*RI restriction fragments of the PAH gene, the 17-kb fragment contains a polymorphic *Eco*RI restriction site. The absence of the site in both copies of the gene results in a pattern shown in figure 1*B*, lane 1. The presence of the site in the 17-kb fragments results in its cleavage into a 11-kb fragment that hybridizes to phPAH247. Individuals homozygous for this site display an *Eco*RI restriction pattern as shown in figure 1*B*, lane 2. The pattern is complicated by the presence of an additional invariant but much

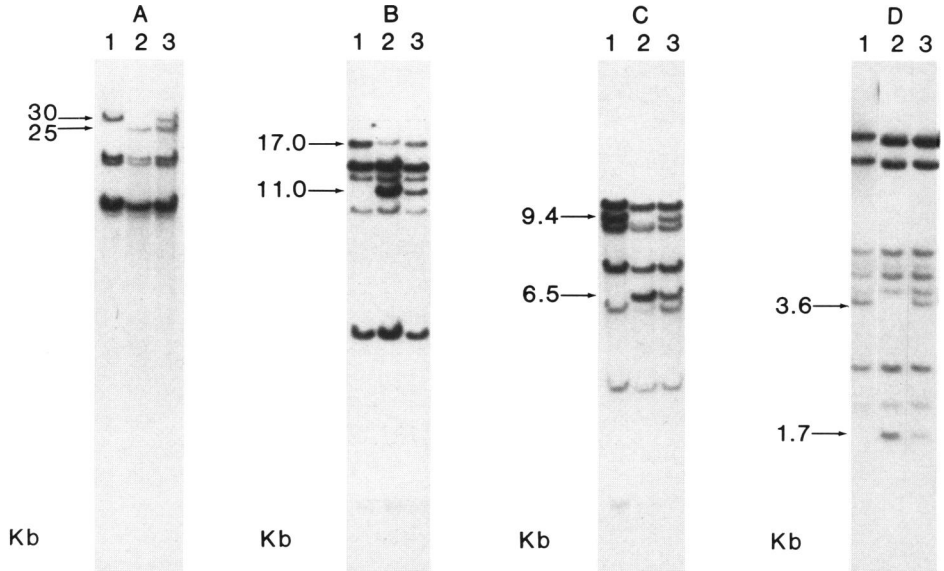

FIG. 1.—RFLP in the human PAH gene detected using a full-length human PAH cDNA clone, phPAH247, as the hybridization probe. Shown are genomic DNAs isolated from buffy coats of three normal Caucasians, which illustrate the variation when DNA is cleaved with *Eco*RV (*panel* A), *Eco*RI (*panel B*), *Xmn*I (*panel C*), and *Bgl*II (*panel D*).

fainter band that comigrates with the polymorphic 17-kb fragment (fig. 1*B*, lane 2). Individuals heterozygous for the polymorphic *Eco*RI site (fig. 1*B*, lane 3) have the 11- and 17-kb bands of almost equal intensity and are readily distinguishable from those homozygous for the 11-kb fragment (fig. 1*B*, lane 2). To avoid the necessity of having to determine whether the 17-kb band is the result of partial enzymatic digestion of genomic DNA, the *Eco*RI polymorphism can be better detected by a double digestion with *Bam*HI, which cleaves the 17- and 11-kb fragments to 8.3- and 6.5-kb fragments, respectively (data not shown).

*Xmn*I: A two-allelic system was also detected by this enzyme. Restriction site polymorphism results in individuals homozygous for either a 9.4-kb fragment (fig. 1*C*, lane 1) or a 6.5-kb fragment (fig. 1*C*, lane 2). Heterozygous individuals for the two fragments have also been identified (fig. 1*C*, lane 3).

*Bgl*II: A *Bgl*II site is polymorphic in the PAH gene and results in restriction fragments of either 3.6- or 1.7-kb in length. Individuals homozygous for the 3.6-kb fragment (fig. 1*D*,, lane 1) or for the 1.7-kb fragment (fig. 1*D*, lane 2) and those heterozygous for the two fragments (fig. 1*D*, lane 3) were readily detected in the panel of random individuals.

*Msp*I: A polymorphic *Msp*I restriction site in the PAH gene resulting in fragments of 23 kb and 19 kb in length has been reported using a partial-length cDNA clone (re12). In addition to these fragments, two invariant fragments and three additional polymorphic fragments are detected with the full-length cDNA clone (fig. 2). The differential 4.0-kb fragment is detected when the polymorphic *Msp*I site (*Msp*Ia) in the 23-kb segment is cleaved to yield two frag-

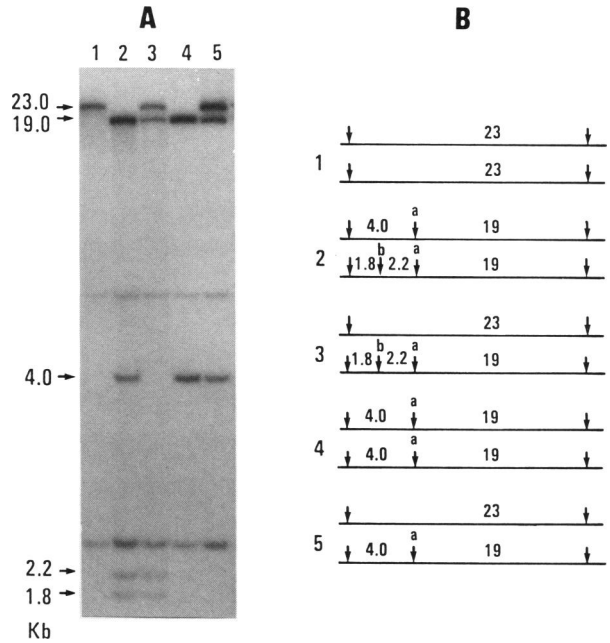

FIG. 2.—*A*, Southern blot analysis of five individuals' genomic DNA digested with *Msp*I and hybridized to phPAH247. *B*, Schematic diagram of the *Msp*I RFLP haplotypes of the PAH gene and the genotypes of the five individuals in *panel A* (1–5 correspond to *lanes 1–5*). *Arrows* show the cleavage sites of *Msp*I, and those labeled *a* and *b* are the polymorphic cleavage sites. The sizes of the fragments generated are shown in kilobases.

ments of 19 kb and 4 kb in length as shown in two individuals with the 19+4//19+4 and 23//19+4 genotypes (fig. 2, lanes 4 and 5). A second polymorphic *Msp*I site (*Msp*Ib) has been found within the 4-kb segment. The presence of this site causes the 4-kb fragment to be cleaved into two fragments of 2.2 kb and 1.8 kb, both of which hybridize with the cDNA probe (fig. 2, lanes 2 and 3). Individuals homozygous for the PAH gene that lack both of the polymorphic *Msp*I restriction sites will be homozygous for the 23-kb band (fig. 2, lane 1). Individuals homozygous for the *Msp*Ia site but heterozygous for the *Msp*Ib site have a genotype of 19+4//19+2.2+1.8, resulting in four hybridizing polymorphic bands (fig. 2, lane 2). An individual with one chromosome lacking both polymorphic sites and the other containing both has a genotype of 23//19+2.2+1.8 (fig. 2, lane 3). One possible haplotype that has not been found among the 18 random individuals is that of a gene containing only the *Msp*Ib but not the *Msp*Ia site. If the two polymorphic sites are in linkage equilibrium, this haplotype should have a frequency of 24%. Thus, the data suggest that these two sites may be in linkage disequilibrium.

*Pvu*II: Southern analysis of DNA digested with *Pvu*II shows four variant fragments of 19.0, 11.5, 9.1, and 6.0 kb in length, apparently created by two polymorphic *Pvu*II restriction sites. The polymorphic fragments are not the

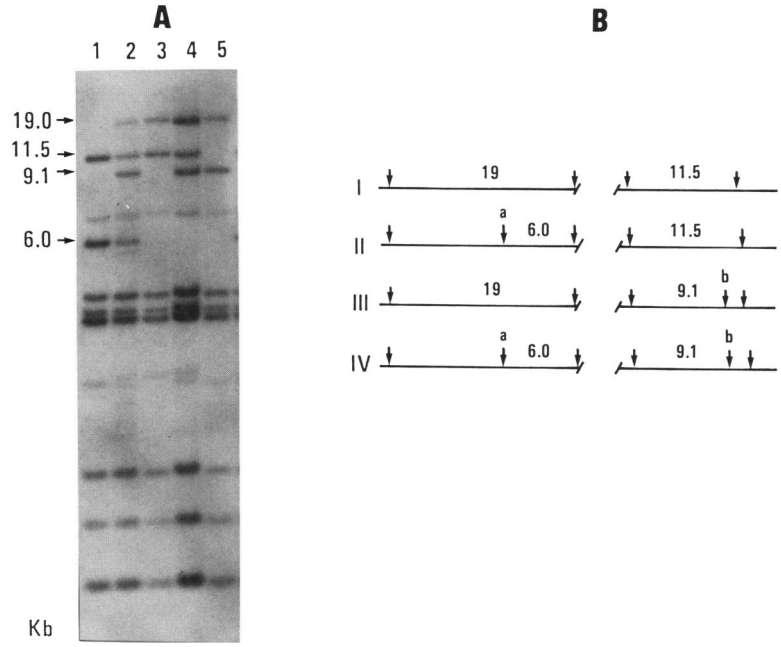

FIG. 3.—*A*, Southern blot analysis of the *Pvu*II polymorphisms in the human PAH gene. The genomic DNAs in *lanes 1–5* show five representative *Pvu*II cleavage patterns. *B*, A schematic drawing of the four possible *Pvu*II haplotypes of the PAH gene. The two variant sites *a* and *b* are labeled, and the sizes of the fragments detected with phPAH247 are given. *Slash marks* indicate that the polymorphic fragments are not contiguous in the PAH gene.

result of partial digestion of the DNAs, because the profile of any given individual remained constant after repeated enzymatic digestion (data not shown). In an attempt to determine which fragments are allelic, homozygous individuals are the most informative because they necessarily have the least number of hybridizable fragments that can be excluded as being alleles. Thus, the 11.5-kb fragment cannot be allelic with either the 6-kb fragment (fig. 3*A*, lane 1) or the 19-kb fragment (fig. 3*A*, lane 3). The 19-kb and the 9.1-kb fragments can also be excluded from being allelic (fig. 3*A*, lane 5). The two polymorphic *Pvu*II sites, designated *a* and *b*, are illustrated schematically in figure 3*B*. PAH genes that lack both polymorphic sites would yield two bands of 19 and 11.5 kb or a haplotype of 19+11.5 (fig. 3*B*, type I). The presence of only site *a* causes cleavage of the 19-kb fragment into a 6-kb hybridizable fragment and results in a haplotype of 6+11.5 (fig. 3*B*, type II). The presence of only site *b* causes cleavage of the 11.5-kb fragment into a 9.1-kb hybridizable fragment, resulting in a haplotype of 19+9.1 (fig. 3*B*, type III), and the presence of both polymorphic sites gives a haplotype of 6+9.1 (fig. 3*B*, type IV). As with *Msp*I, there can be a total of four haplotype and 10 genotype combinations between the two sites of *Pvu*II alleles. The Southern blot in figure 3*A* shows five such cleavage patterns of *Pvu*II-digested DNA. Individuals homozygous for allele types I, II,

and III are shown in figure 3*A*, lanes 3, 1, and 5, respectively. Individual no. 4 shows three hybridizing bands at 19, 11.5, and 9.1 kb (fig. 3*A*, lane 4) and must be a heterozygote of the allele types I/III. All four hybridizing bands are present in individual no. 2 (fig. 3*A*, lane 2), who must be a heterozygote of the allele types I/IV or II/III. Consequently, it can be concluded that the four polymorphic fragments are the results of the presence of two independent sets of *Pvu*II RFLP alleles, which is confirmed by kindred studies that show Mendelian segregation of the two *Pvu*II RFLPs as presented in the following section.

In all the individuals tested, there was no 100% concordance between any of the RFLPs described above, and the alteration in DNA fragment lengths was different for all pairs of alleles. These observations rule out the possibility that the variation in restriction fragment lengths observed with different enzymes is due to a single insertion/deletion event or a base-pair substitution causing alteration of multiple restriction sites. Thus, each of the RFLPs described in this section are independent of each other and are unique in origin.

Mendelian Segregation of the RFLPs at the PAH Locus

Prior to the application of RFLP analysis to trace the transmission of PAH genes in PKU families, it must be demonstrated conclusively that the RFLPs are indeed authentic and are not the result of artifacts such as partial digestion of genomic DNA samples by the restriction enzymes used in the analyses. Such artifacts are not likely to be transmissible as Mendelian traits and their authenticity can be verified by performing kindred analysis. The Mendelian segregation of the *Hin*dIII, *Sph*I, and *Msp*I RFLPs in PKU kindreds has been demonstrated [12]. Kindred analysis using each of the RFLPs presented in this study was thus performed. Figure 4 shows the Mendelian segregation of the *Bgl*II, *Pvu*IIa, *Pvu*IIb, *Eco*RI, *Eco*RV, *Msp*Ia, *Msp*Ib, and *Xmn*I polymorphisms in a PKU family with one affected child and three unaffected siblings. The father is homozygous for the 3.6-kb *Bgl*II fragment, and the mother is homozygous for the 1.7-kb fragment (fig. 4*A*, lanes 1 and 2). This results in all of the offspring being heterozygous (3.6/1.7) for the *Bgl*II polymorphism as expected (fig. 4*A*, lanes 3–6). Similarly, the RFLPs using *Eco*RI, *Eco*RV, and *Xmn*I also show Mendelian segregation in this kindred (fig. 4*B–D*).

The transmission of the PKU trait in this family can be traced by *Eco*RV RFLP analysis. Both parents are RFLP heterozygotes for the two *Eco*RV bands, with a genotype of 30/25 (fig. 4*C*, lanes 1 and 2). Since the proband is also an RFLP heterozygote (fig. 4*C*, lane 3), the PKU traits must be segregating with the 30-kb allele in one parent and the 25-kb allele in the other. Two of the unaffected siblings have inherited the gene containing the 25-kb fragment from both parents (fig. 4*C*, lanes 4 and 6). One of these genes must contain the PKU mutation, and the two siblings are thus PKU carriers. A third sibling is homozygous for the 30-kb fragment (fig. 4*C*, lane 5) and must have inherited the PKU trait from the other parent.

Similarly, the analysis of this family with *Msp*I illustrates the inheritance of both polymorphic *Msp*I sites and the segregation of the mutant PAH genes. The father's DNA contains only the 19- and 4.0-kb fragments and is therefore a

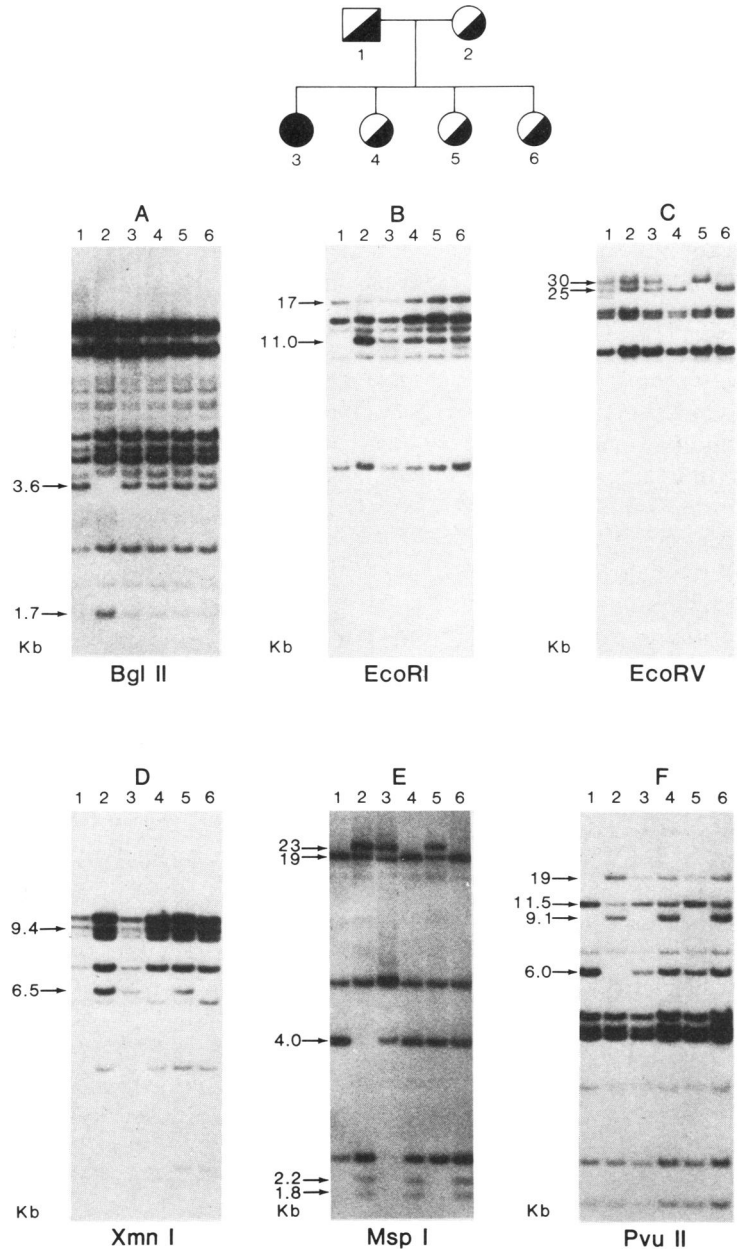

FIG. 4.—Mendelian segregation of the polymorphic restriction fragments and RFLP analysis of a PKU family. Pedigree drawing illustrates the carrier status of each family member, as determined by RFLP analysis. Southern blots show the segregation of the *Bgl*II (*panel A*), *Eco*RI (*panel B*), *Eco*RV (*panel C*), *Xmn*I (*panel D*), *Msp*I (*panel E*), and *Pvu*II (*panel F*) RFLPs in this kindred. *In each panel,* the family members' DNA are ordered as follows: father, *lane 1*; mother, *lane 2*; proband, *lane 3*; unaffected sibling, *lane 4*; unaffected sibling, *lane 5*; and an unaffected sibling, *lane 6*.

haplotype homozygote (fig. 4*E*, lane 1). His genotype is 19+4//19+4, being the result of the presence of the *Msp*Ia site and the absence of the *Msp*Ib site in both of his PAH genes. The mother's DNA shows four bands and has a genotype of 23//19+2.2+1.8 (fig. 4*E*, lane 2), the result of one PAH gene lacking both polymorphic sites (23 kb) and the other having both (19+2.2+1.8 kb). The proband has inherited the mutant gene containing the 23-kb fragment from the mother and has a genotype of 23//19+4 (fig. 4*E*, lane 3). One of the phenotypically unaffected siblings has the same genotype as the proband (fig. 4 *E*, lane 5). This individual has inherited the mutant allele from the mother and the normal gene from the father whose DNA is noninformative for *Msp*I RFLP analysis. The two other unaffected siblings have inherited the normal allele from the mother and have genotypes of 19+2.2+1.8//19+4 (fig. 4*E*, lanes 4 and 6). Consequently, these two individuals are either carriers or free of the PKU trait. The same analysis can also be performed for the *Pvu*II RFLPs (fig. 4*F*), and the result is in complete agreement with the *Msp*I polymorphism data.

Similar analyses of a number of additional kindreds have clearly demonstrated Mendelian inheritance of these RFLPs. The genotype frequencies are in Hardy-Weinberg equilibrium, and each allelic type was observed in more than one individual. These observations clearly demonstrate that all of these RFLPs are inherited in an autosomal manner and are not the result of a laboratory artifact.

RFLP Haplotype Analysis and Prenatal Diagnosis

From the above data, the RFLP haplotypes of the four PAH genes in this particular family can be readily deduced (table 1), which can be used to trace the transmission of the PKU genes and to perform prenatal diagnosis if the need arises. Prenatal diagnosis by RFLP analysis depends on the ability to distinguish the mutant genes from the normal ones. In this family, each PAH gene has a unique haplotype, and complete prenatal diagnosis will be possible. In practice, since the amount of fetal DNA obtainable is usually limited, it is not necessary to determine the complete RFLP haplotypes of the fetal PAH genes. Instead, it may be preferable to use the fewest number of informative restriction enzymes necessary for the diagnosis. For instance, a complete diagnosis can be made using *Msp*I plus either *Xmn*I or *Eco*RV for this family. Since both parents are heterozygous for the 9.4- and 6.5-kb *Xmn*I fragments (table 1), the fetal DNA can either be also heterozygous for the two fragments or homozygous for either fragment. Because the proband is an *Xmn*I heterozygote, a homozygous 9.4- or 6.5-kb pattern for the fetus would indicate the inheritance of one normal and one mutant gene for either parent. Thus, the fetus would be a carrier of the PKU trait. On the other hand, if the fetus is also an *Xmn*I heterozygote, it may have inherited both mutant genes or both normal genes. The same analysis can be performed using *Eco*RV, since the polymorphic fragments are also in repulsion phase in this family.

To complete the diagnosis, a separate *Msp*I digestion of fetal DNA can be performed. In this family, the father is an *Msp*I 19+4//19+4 homozygote and the mother is a 23//19+2.2+1.8 heterozygote (table 1). Thus, the fetal *Msp*I

TABLE 1

RFLP HAPLOTYPES OF THE PAH GENES IN A PKU KINDRED

INDIVIDUAL	RELATIONSHIP		RFLP HAPLOTYPE							
			*Bg*1II	*Eco*RI	*Eco*RV	*Xmn*I	*Msp*Ia	*Msp*Ib	*Pvu*IIa	*Pvu*IIb
1	Father	Normal	3.6	17	30	6.5	19	4	6	11.5
		Mutant	3.6	17	25	9.4	19	4	6	11.5
2	Mother	Normal	1.7	11	25	9.4	19	2.2 + 1.8	19	9.1
		Mutant	1.7	11	30	6.5	23	...	19	11.5
3	Proband	Mutant F	3.6	17	25	9.4	19	4	6	11.5
		Mutant M	1.7	11	30	6.5	23	...	19	11.5
4	Sibling	Mutant F	3.6	17	25	9.4	19	4	6	11.5
		Normal M	1.7	11	25	9.4	19	2.2 + 1.8	19	9.1
5	Sibling	Normal F	3.6	17	30	6.5	19	4	6	11.5
		Mutant M	1.7	11	30	6.5	23	...	19	11.5
6	Sibling	Mutant F	3.6	17	25	9.4	19	4	6	11.5
		Normal M	1.7	11	25	9.4	19	2.2 + 1.8	19	9.1

genotype must either be 23//19+4 or 19+2.2+1.8//19+4. Since the maternal mutant PAH gene segregates with the 23-kb fragment, the presence of this band and the lack of the 2.2+1.8-kb bands would indicate that the fetus has inherited both mutant genes from the parents and would therefore be affected. The opposite observation would indicate that the fetus has inherited both normal genes and would be free of the PKU trait.

Allele Frequencies and Percent Heterozygosity

The usefulness of DNA probes that detect RFLP for genetic analysis depends on the frequency of polymorphism in the general population. To determine the frequencies of all 10 restriction site polymorphisms and estimate the percent RFLP heterozygosity of the PAH gene in the population, the genotypes of the 18 unrelated Caucasian individuals with respect to the 10 polymorphic restriction sites are shown in table 2. The RFLP haplotypes of their PAH genes, however, cannot be established without kindred analysis. Although multiple enzyme digestion experiments can also establish haplotypes in these individuals, it requires a complete genomic map of the polymorphic sites, which is not yet available. Nevertheless, the frequencies of each of the polymorphic alleles can be calculated from these data (table 3). The frequencies of the minor alleles exceed .3 for all of the enzymes, indicating the existence of a very high degree of polymorphism at the human PAH locus. The expected genotype frequencies calculated from the individual allele frequencies do not significantly differ from the observed genotype frequencies among this panel of random Caucasian individuals, indicating that many of these RFLPs are in Hardy-Weinberg equilibrium.

From the 10 polymorphic restriction sites identified at the PAH locus, all of which are two-allele systems except for *Hin*dIII, which has a rare third allele [12], it can be calculated that there are a total of 1,536 possible different combinations of the polymorphisms, or RFLP haplotypes. The haplotype heterozygosity, which is an indication of how widely the analysis can be applied for genetic analysis among PKU kindreds in the population, was estimated from the individual allele frequencies listed in table 3 to be well in excess of 95%. This high level of heterozygosity was estimated assuming that the 36 chromosomes analyzed are representative of the Caucasian population and that there is completely random association between the polymorphic restriction sites. If linkage disequilibrium exists between any of the observed polymorphic sites, the percent heterozygosity will be reduced. In the panel of random Caucasian individuals, 15 out of 18 are RFLP heterozygotes, which yields an observed heterozygosity of 83.3%. The analysis of 40 unrelated Caucasian PKU kindreds shows that 70 out of 80 parents are RFLP haplotype heterozygotes, or an observed heterozygosity of 87.5% (A. Lidsky, unpublished results, 1985). The fact that the observed heterozygosity is significantly lower than the theoretical value would suggest that some of the polymorphic sites are in linkage disequilibrium. Nevertheless, the observed heterozygosity of 87.5% would suggest that most of the PKU families in the Caucasian population can take advantage

TABLE 2

GENOTYPES OF THE RANDOM PANEL OF NORMAL CAUCASIANS AT THE PAH LOCUS

	GENOTYPE								
INDIVIDUAL	*Bgl*II	*Eco*RI	*Eco*RV	*Hin*dIII	*Msp*Ia	*Msp*Ib*	*Pvu*IIa	*Pvu*IIb	*Xmn*I
1	N.D.†	17/17	30/30	4.2/4.2	19/19	4.0/4.0	6/6	11.5/11.5	9.4/9.4
2	3.6/1.7	11/11	25/25	4.2/4.0	23/19	(−)/(+)	19/6	11.5/9.1	9.4/6.5
3	1.7/1.7	N.D.	30/30	4.2/4.2	23/23	(−)/(−)	19/19	11.5/11.5	6.5/6.5
4	1.7/1.7	11/11	30/30	4.2/4.2	23/19	(−)/(+)	19/19	11.5/9.1	9.4/6.5
5	3.6/1.7	N.D.	25/25	4.2/4.2	19/19	(+)/(+)	19/19	9.1/9.1	9.4/9.4
6	3.6/3.6	17/17	30/30	4.2/4.2	19/19	4.0/4.0	6/6	11.5/11.5	9.4/9.4
7	3.6/3.6	17/11	30/25	4.2/4.0	23/19	(−)/4.0	6/6	11.5/11.5	9.4/6.5
8	3.6/1.7	17/11	25/25	4.2/4.0	23/19	(−)/(+)	19/6	11.5/9.1	9.4/6.5
9	3.6/3.6	17/11	30/25	4.2/4.0	23/19	(−)/4.0	6/6	11.5/11.5	9.4/6.5
10	3.6/3.6	17/17	30/25	4.2/4.0	19/19	4.0/4.0	6/6	11.5/11.5	9.4/9.4
11	3.6/1.7	17/17	25/25	4.2/4.0	23/19	(−)/(+)	19/6	11.5/9.1	9.4/6.5
12	3.6/1.7	17/17	30/25	4.4/4.2	19/19	4.0/(+)	19/6	11.5/9.1	9.4/9.4
13	3.6/1.7	17/11	30/25	4.2/4.0	23/23	(−)/(−)	19/6	11.5/11.5	6.5/6.5
14	3.6/1.7	11/11	25/25	4.2/4.0	23/19	(−)/(+)	19/6	11.5/9.1	9.4/6.5
15	N.D.	17/11	30/25	4.2/4.2	N.D.	N.D.	19/6	11.5/9.1	9.4/9.4
16	3.6/3.6	17/17	30/25	4.4/4.2	N.D.	N.D.	6/6	11.5/11.5	9.4/9.4
17	3.6/1.7	17/17	30/30	4.0/4.0	23/19	(−)/4.0	19/6	11.5/11.5	9.4/6.5
18	3.6/1.7	11/11	25/25	4.0/4.0	19/19	(+)/(+)	19/19	9.1/9.1	9.4/9.4

* The presence and absence of the 2.2 + 1.8 DNA fragments are represented by the (+) and (−) symbols in this column, respectively.
† N.D. = not determined.

TABLE 3

EXPECTED AND OBSERVED FREQUENCIES OF VARIOUS ALLELES OF THE HUMAN PAH GENE

RESTRICTION ENZYME	FRAGMENT SIZE (kb)	ALLELE FREQUENCY	GENOTYPE FREQUENCY			
				p^2	$2pq$	q^2
*Bgl*II.....	3.6	.59	Observed	.31	.56	.13
	1.7	.41	Expected	.35	.48	.17
*Eco*RI.....	17.0	.59	Observed	.44	.31	.25
	11.0	.41	Expected	.35	.48	.17
*Eco*RV	30.0	.47	Observed	.28	.39	.33
	25.0	.53	Expected	.22	.50	.28
*Hin*dIII.....	4.2	.61	Observed	.33	.55	.11
	4.0	.39*	Expected	.37	.48	.15
*Msp*Ia.....	23.0	.38	Observed	.13	.50	.37
	19.0	.62	Expected	.14	.48	.38
*Msp*Ib.....	4.0	.69	Observed	.50	.38	.12
	2.2	.31	Expected	.48	.42	.10
*Pvu*IIa.....	19.0	.44	Observed	.22	.44	.33
	6.0	.56	Expected	.19	.49	.32
*Pvu*IIb	11.5	.69	Observed	.50	.39	.11
	9.1	.31	Expected	.47	.43	.10
*Xmn*I.....	9.4	.67	Observed	.44	.45	.11
	6.5	.33	Expected	.45	.44	.11

* This no. represents a sum of the frequencies of the minor 4.0-kb allele and a rare 4.4-kb allele.

of this genetic analysis for prenatal diagnosis and carrier detection of the hereditary disorder.

Frequency of Nucleotide Substitutions at the Human PAH Locus

The identification of multiple RFLPs at the PAH locus is indicative of the high degree of polymorphic nucleotide substitutions in the population. To estimate the frequency of nucleotide substitutions throughout the PAH gene, the number of restriction sites and nucleotide pairs analyzed in this study with 20 different restriction enzymes was calculated. The number of restriction sites analyzed by the polymorphism study depends on whether or not the fragments represent a contiguous stretch of DNA. In the case of all contiguous fragments, the number of sites analyzed is equal to n + 1, where n is the number of DNA fragments detected. If the hybridizable fragments are separated by nonhybridizable fragments such as those located completely within intervening sequences, the number of restriction sites analyzed becomes 2n. In this study, the number of nucleotide pairs analyzed for polymorphism with the use of the 20 enzymes ranges from 849 to 1,482, assuming completely contiguous or completely noncontiguous hybridizable polymorphic DNA fragments. If each of the 10 polymorphic sites represents one polymorphic nucleotide, the frequency of polymorphic nucleotides in the human phenylalanine hydroxylase locus can be estimated to range from .012 to .007, or one in every 85 to 148 nucleotide pairs.

DISCUSSION

Extensive restriction fragment length polymorphism has been detected at the human PAH locus using a full-length PAH cDNA clone as the hybridization

probe. The potential application of RFLPs to the diagnosis of PKU prenatally and the detection of PKU carriers has been discussed [12]. It is possible to distinguish the normal and mutant phenylalanine hydroxylase genes in families with one or more previously affected children by identifying the RFLP haplotypes associated with the PKU genes in each family. In all of the families we analyzed, there was uniform concordance of segregation between the RFLP haplotype at the PAH locus and the PKU trait. Since these polymorphisms are at or near the structural PAH gene, the frequency of homologous recombination during each meiosis event is extremely low. In fact, in more than 100 informative meioses, no recombination event has been observed between PKU and any of the RFLPs (manuscript in preparation). Thus, RFLP analysis can be used in families with at least one affected child for prenatal diagnosis in future pregnancies and for carrier detection among siblings with a high degree of accuracy. Indeed, we have recently performed prenatal diagnosis for PKU in several affected families, and the accuracy of the method has been confirmed by phenotypic analysis of the neonates [16].

Application of this genetic analysis relies on the ability to identify the PKU alleles in the family, which in turn depends on the ability to distinguish the RFLP haplotypes of the PAH genes in both parents. Ideally, both parents ought to be RFLP haplotype heterozygotes so that complete diagnosis can be achieved. In the case only one parent is an RFLP haplotype heterozygote, the other parent's PAH genes are noninformative for RFLP analysis. Nevertheless, the genotypes of the children will be such that 50% of them can be excluded from having inherited both mutant genes from the parents, thereby permitting exclusion analysis. Fortunately, the frequencies in the population studied are high for many of these RFLPs. In fact, we have observed a haplotype heterozygosity of 87.5.%, which suggests that this probe can be used as a polymorphic marker for prenatal diagnosis and carrier detection of the genetic disorder for most PKU families in the population.

It is logical to ask why is there such a high degree of polymorphism at the human PAH locus. It has been suggested that the methylated cytosine residues that are present in the *Msp*I and *Taq*I recognition sequence are hotspots for mutation because of the large number of *Msp*I and *Taq*I RFLPs detected with other probes [17]. This alone cannot provide a satisfactory answer because in this report we describe seven new polymorphic sites identified with six restriction enzymes. Although *Msp*I identifies two polymorphic sites at the PAH locus, no polymorphism was observed with *Taq*I. A more plausible explanation may be the physical size and the complex molecular structure of this gene. Cosmid cloning and preliminary characterization of the human PAH gene have shown that the enzyme is encoded by a unique gene that is 90–100 kb in length with multiple introns (A. DiLella, unpublished results, 1985). The fact that there are a minimum of 10 exons scattered throughout this gene would allow the full-length cDNA to detect restriction sites present in a stretch of genomic DNA that is in excess of 100 kb when RFLP analysis is performed, which is much greater in scale than the entire 60-kb β-globin locus in man. Consequently, it is not too surprising to find that the frequency of polymorphic

nucleotides at the PAH locus ranges from .007 to .012, which is comparable to those observed in the β-globin and albumin loci in man [18, 19]. Nevertheless, the full-length PAH cDNA clone represents one of the most polymorphic markers for RFLP analysis of the human genome.

Questions remain unanswered as to the molecular lesions in the PAH gene that cause the enzyme deficiency. None of the genetic variations observed among normal individuals by Southern blotting using the cloned PAH cDNA as a probe represent the underlying mutations for PKU. Nevertheless, the extensive RFLPs observed in the chromosomal phenylalanine hydroxylase gene can be used as a powerful tool to identify mutations in the PAH gene that cause PKU. In β-thalassemia, distinct mutations in the β-globin locus are associated with specific RFLP haplotypes within a given population [20, 21]. Characterization of RFLP haplotypes and their correlation with individual mutations in a population have provided a valuable guide for delineating an extensive list of different β-globin mutations [20, 21]. We are investigating whether or not similar association of specific RFLP haplotypes and distinct mutations exist at the PAH locus in various ethnic backgrounds.

Finally, a linkage map of the human genome is being constructed by a number of laboratories utilizing DNA probles that can detect RFLPs with reasonably high frequencies [22]. Such polymorphic probes can be used as linked genetic markers to analyze genetic disorders with unidentified biochemical lesions. This type of application is best exemplified by the linkage of anonomous polymorphic DNA probes to Duchenne muscular dystrophy on the short arm of the X chromosome [23] and to Huntington disease on the short arm of chromosome 4 [24]. Future success in this approach to the characterization of genetic disorders relies on collection of a sufficient number of polymorphic markers that are distributed throughout the human genome [22]. In this regard, we have recently reported the assignment of the PAH locus to human chromosome 12 [25]. More recently, we have further regionalized the locus to 12q22-q24.1 (manuscript in preparation). Thus, the human PAH cDNA clone can also serve as a highly polymorphic marker for this region of chromosome 12 in linkage studies with other genetic loci in man.

ACKNOWLEDGMENT

We thank Ms. Sharon Moore for technical assistance.

REFERENCES

1. Jervis GA: Deficiency of phenylalanine oxidizing system. *Proc Soc Exp Biol Med* 82:514–515, 1953
2. Folling A: Uber Ausscheidung von Phenylbrenztraubensaure in den Harn als Stoffweshselanomalie in Verbindung mit Imbezillitat. *Z Physiol Chem* 227:169–176, 1934
3. Guthrie R, Susi A: A simple method for the detection of phenylketonuria in large populations of newborn infants. *Pediatrics* 32:338–343, 1963
4. Bickel H, Gerrard J, Hickmans EM: The influence of phenylalanine intake on the chemistry and behaviour of a phenylketonuric child. *Acta Paediatr Scand* 43:64–67, 1954

5. ARMSTRONG MD, TYLER FH: Studies on phenylketonuria: restricted phenylalanine intake in phenylketonuria. *J Clin Invest* 34:565–580, 1955
6. SMITH I, LOBASCHER ME, STEVENSON JE, ET AL: Effect of stopping low-phenylalanine diet on intellectual progress of children with phenylketonuria. *Br Med J* 2:723–726, 1978
7. KOCH R, FRIEDMAN EG, WILLIAMSON ML, AZEN CG: Preliminary report on the effects of diet discontinuation in phenylketonuria. *J Inherit Dis* 5(Suppl. 1):36–64, 1982
8. SCHUETT VE, GURDA RF, BROWN ES: Diet discontinuation policies and practices of PKU clinics in the United States. *Am J Public Health* 70:498–503, 1980
9. THALHAMMER O, ET AL.: Frequency of inborn errors of metabolism, especially PKU, in some representative newborn screening centers around the world. A collaborative study. *Humangenetik* 30:273–286, 1975
10. BICKEL H, BACHMANN C, BECKERS R, ET AL: Neonatal mass screening for metabolic disorders. A collaborative study. *Eur J Pediatr* 137:133–139, 1981
11. ROBSON KJH, CHANDRA T, MACGILLIVRAY RTA, WOO SLC: Polysome immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning of its cDNA. *Proc Natl Acad Sci USA* 79:4701–4705, 1982
12. WOO SLC, LIDSKY AS, GUTTLER F, CHANDRA T, ROBSON KJH: Cloned human phenylalanine hydroxylase gene permits prenatal diagnosis and carrier detection of classical phenylketonuria. *Nature* 306:151–155, 1983
13. KWOK S, LEDLEY FD, DILELLA AG, ROBSON KJH, WOO SLC: Nucleotide sequence of a full-lenth cDNA clone of human phenylalanine hydroxylase. *Biochemistry* 24:556–561, 1985
14. MANIATIS T, JEFFREY A, KLEID DG: Nucleotide sequence of the rightward operator of phage λ. *Proc Natl Acad Sci USA* 72:1184–1188, 1975
15. SOUTHERN EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503–597, 1975
16. LIDSKY AS, GUTTLER F, WOO SLC: Prenatal diagnosis of classic phenylketonuria by DNA analysis. *Lancet* i:549–551, 1985
17. BARKER D, SCHAFER M, WHITE R: Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. *Cell* 36:131–138, 1984
18. JEFFREYS AJ: DNA sequence variants in the $^{G}\gamma$-, $^{A}\gamma$-, δ- and β-globin genes of man. *Cell* 18:1–10, 1979
19. MURRAY JC, DEMOPOLUS CM, LAWN RM, MOTULSKY AG: Molecular genetics of human serum albumin: restriction enzyme fragment length polymorphisms and analbuminemia. *Proc Natl Acad Sci USA* 80:5951–5955, 1983
20. ORKIN SH, KAZAZIAN HH JR, ANTONARAKIS SE, ET AL.: Linkage of β-thalassemia mutations and β-globin gene polymorphisms in the human β-globin gene cluster. *Nature* 296:627–631, 1982
21. KAZAZIAN HH JR, ORKIN H, MARKHAM AF, CHAPMAN CR, YOUSSOUFIAN H, WABER PG: Quantification of the close association between DNA haplotypes and specific β-thalassaemia mutations in Mediterraneans. *Nature* 310:152–154, 1984
22. BOTSTEIN D, WHITE RL, 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
23. MURRAY JM, DAVIES KE, HARPER PS, ET AL.: Linkage relationship of a cloned DNA sequence on the short arm of the X chromosome to Duchenne muscular dystrophy. *Nature* 300:69–71, 1982
24. GUSELLA JF, WEXLER NS, CONNEALLY PM, ET AL.: A polymorphic DNA marker genetically linked to Huntington disease. *Nature* 306:234–238, 1983
25. LIDSKY AS, ROBSON KJH, THIRUMALACHARY C, BARKER PE, RUDDLE FH, WOO SLC: The PKU locus in man is on chromosome 12. *Am J Hum Genet* 36:527–533, 1984