PRINT: ISSN 0972-3757 ONLINE: 2456-6360
Int J Hum Genet, 9(2): 115-121 (2009)
DOI: 10.31901/24566330.2009/09.02.05

Estimation Haplotype Frequency of *Bgl*II/*Eco*RI/VNTR Markers at the *PAH* Gene Region in Iranian Population

Zahra Fazeli and Sadeq Vallian

Division of Genetics, Department of Biology, Faculty of Science, The University of Isfahan, Isfahan, IR Iran

KEYWORDS Phenylalanine Hydroxylase (PAH). Phenylketonuria (PKU). Haplotype. Polymorphism Markers. Iranian Population.

ABSTRACT Deficiency in phenylalanine hydroxylase (PAH) is the main molecular characteristic of phenylketonuria (PKU). So far over 500 different mutations in the *PAH* gene have been identified **as** associated with the disease. Mutation analysis of the *PAH* gene is a time-consuming and cost-effective procedure. Therefore, molecular markers which are highly linked to the *PAH* gene, have been used in carrier detection and prenatal diagnosis in PKU families. These markers show a population dependent based haplotype frequency. In the present study, the haplotype frequency of three markers including *Bgl*II, *Eco*RI and VNTR, at the *PAH* gene region were investigated in Iranian population. Nine different alleles for VNTR marker with 3, 5, 6, 7, 8, 9, 10, 11 and 13 core repeats were detected. Alleles 4 and 13 were found specific to the Iranian population. The haplotype frequency was calculated using FBAT, PHASE and Arlequin (haplotype frequency estimation computer programs). Among the 36 possible estimated haplotypes identified for 2 alleles (+ and -) for *Bgl*II and *Eco*RI; and nine alleles for VNTR, ten haplotypes showed relatively high frequency (e" 5%), based on the above programs. Therefore, a combination of *Bgl*II-*Eco*RI-VNTR could be suggested as an informative haplotype in performing carrier and prenatal diagnosis of the *PAH* gene mutations in Iranian population.

INTRODUCTION

Phenylalanine hydroxylase enzyme (PAH) catalyses conversion of the essential amino acid phenylalanine to tyrosine. Deficiency in this enzyme causes an accumulation of serum phenylalanine, resulting in postnatal brain damage and severe mental retardation, which was known as phenylketonuria (PKU) (Matthew 2004). More than 500 different mutations in the *PAH* gene have been identified and catalogued in the public-domain data base, PAHdb (http://www.pahdb.mcgill.ca/) (Scriver et al. 2003). Mutation analysis of the gene is a time-consuming and cost-effective procedure. Therefore, polymorphic markers which are highly linked to the *PAH* gene could facilitate prenatal diagnosis and carrier detection of the disease status in PKU families. The *PAH* gene region contains several well identified polymorphic markers, including eight restriction fragment length polymorphisms (RFLPs), an inter-genic variable number tandem repeat (VNTR) and an intra-genic short tandem repeat (STR). The restriction fragment length polymorphism sites were located in two clusters: *Bgl*II, *Pvu*II(a) and *Pvu*II(b) at the 5ó-end of the gene and *Eco*RI, *Xmn*I, *Msp*I, *Hind*III and *Eco*RV at the 3ó-end of the gene. These RFLPs except *Hind*III with three alleles are biallelic restriction site polymorphisms (DiLella et al. 1986). This *Hind*III polymorphism is the result of VNTR marker which is located in the 3ó region of the *PAH* gene. This VNTR marker contains an AT-rich (70%) tandemly repeated 30-bp cassettes of at least 10 alleles in different populations. These alleles differ by number of repeats (Goltsov et al. 1992). Interestingly, among the alleles identified, allele with 13 repeats seems to be specific to Iranian population (Kamkar et al. 2003; Vallian et al. 2005). The STR marker which is in intron III contains a series of tetra nucleotides $(TCTA)_n$ repeats harboring at least 9 alleles (Cali et al. 1997). Due to the known high degree of heterozygosity and allele frequency among different populations, currently these PAH markers are commonly used as informative tools in carrier detection and prenatal diagnosis of PKU disease. Moreover, these markers could be used to define haplotypes at the *PAH* gene. Informative PAH haplotypes are used in linkage analysis for investigation of the transmission of the affected alleles in the family members of the affected individual. **(**Eisensmith et al. 1992**).**

Corresponding Author:
Sadeq Vallian, PhD
Division of Genetics, Department of Biology, Faculty of Science, The University of Isfahan, Isfahan, IR Iran
Telephone/Fax: +983117932456
E-mail: svallian@medinews.com

The estimation of haplotype frequency is usually performed using direct and indirect approaches. The direct approach is completely experimental and involves molecular methods such as cloning (Ruano et al. 1990), allele-specific polymerase chain reaction (Ruano et al. 1989) and single molecule dilution (Michalatos-Beloin et al. 1996). In contrast, in the indirect approach, at first, diploid sequence information is experimentally obtained from individuals in the population and then the haplotype is identified computationally. So far, over 40 computer programs have been developed for estimation of haplotype frequency. Most of these programs use the expectation-maximization (EM) algorithm or Bayesian algorithm to estimate haplotype frequencies in the population. The direct methods usually have 100% accuracy in haplotype determination, but they are slower, more labor-intensive and expensive than the indirect methods. Although, the indirect methods are fast and relatively cost-effective, but are prone to minor genotyping errors, which may result in missing some data (Xu 2006).

Non-random association between alleles at different loci was defined as linkage disequilibrium (LD). LD provides information on mapping of loci and population histories (Goldstein 2001a). The estimation of LD has a great importance when the haplotype frequency was evaluated by use of different algorithms. In the most cases, if all the markers were in linkage equilibrium, the number of possible haplotypes which were estimated by use of algorithms may be too large (Zhao et al. 2003). Recent studies showed that the pattern of LD between two loci differs notably among population. Therefore, it is necessary to estimate LD between a pair of loci in every population, independently (Goldstein et al. 2001b).

Our previous data and others indicated that the frequency of the PKU is relatively high in Iranian population (Koochmeshgi et al. 2002; Kamkar et al. 2003; Vallian et al. 2003). However, no study has been preformed on haplotype frequency of the *PAH* gene in this population. In the present study, the allele frequency of three markers in the *PAH* gene region including *Bgl*II, *Eco*RI and VNTR were determined, and their haplotype frequency estimation was made.

MATERIAL AND METHODS

Blood Samples: Blood samples were collected from one hundred and sixty unrelated healthy individuals and 20 families (consisting of both parents and one child). These individuals were from the Isfahan province of Iran. Genomic DNA was extracted from the leukocytes by salting out procedure (Miller et al. 1988).

Gyenotyping, Allele and Haplotype Frequency Estimation: DNA samples were genotyped for *Bgl*II and *Eco*RI, and the variable number of tandem repeat (VNTR) located at the 3ó end of the *PAH* gene using PCR. Primers were as described earlier (Dworniczak et al. 1991; Goltsov et al. 1992). PCR conditions were performed with a few modifications. The PCR reactions were carried out in reactions of 25 ìl total volume containing 50 ng DNA, 500 mM KCl, 100 mM Tris–HCl (pH 8.4), 50 mM $MgCl_2$, 200 ìM dNTP, 5 U *Taq* DNA polymerase and 10 pìco mole of each primer. Initial denaturation was performed at 94°C for 4 min, followed by 30 cycles of 95°C denaturation (1 min), different annealing temperature depending on the primer (1 min), 72°C extension (1 min), with a final 72 °C extension for 5 min. The optimum annealing temperature for amplification of *Bgl*II, *Eco*RI and VNTR markers was 58°C, 58°C and 55°C, respectively. The PCR products for *Bgl*II and *Eco*RI were digested with appropriate restriction enzymes. Amplified DNA fragments for VNTR and enzymatic digestion products were separated by electrophoresis on 1.5% agarose gel and visualized by UV trans-illumination following ethidium bromide staining in a UV gel documentation (Syngene, Cambridge, UK).

Statistical and LD Analysis: The estimation of allele frequency along with the observed and expected heterozygosity was performed using the GENEPOP website (Raymond et al. 1995). The haplotype frequency was estimated using the FBAT program on the data obtained from different families (Rabinowitz 2000). Genotype analysis was performed using Arlequin (Schneider et al. 2000) and PHASE (Marchini et al. 2006) computer programs for unrelated individuals. The programs Arlequin and PHASE implement estimation maximization (EM) and Bayesian algorithms, respectively.

LD was measured by using the standardized D' first introduced by Lewontin, 1964 (Lewontin 1964). D' which is the LD relative to its maximum value for a given set of allelic frequencies for the pair of sites, was calculated using 2LD computer program (Zhao 2004). In this sense, D' is a

normalized value of LD. Since D' alone could not provide good predictor of the potential usefulness for multiallelic markers. Therefore, Yamazaki's standardized Chi-square, χ^2, which is another multiallelic LD statistics, was used (Zhao et al. 2005). Chi-square provides a direct evaluation of the strength of the association between alleles at the two loci. Also, the estimation of χ^2 was performed using 2LD computer program.

RESULTS

The allele frequency of genetic markers including a multiallelic variable number of tandem repeats (VNTR) present at the 3'-end of the *PAH* gene and two biallelic *Bgl*II and *Eco*RI markers were estimated. Analysis of the VNTR was performed for 160 healthy individuals. The results as shown in Figure 1a indicate the presence of at least nine alleles with electrophoretic band size

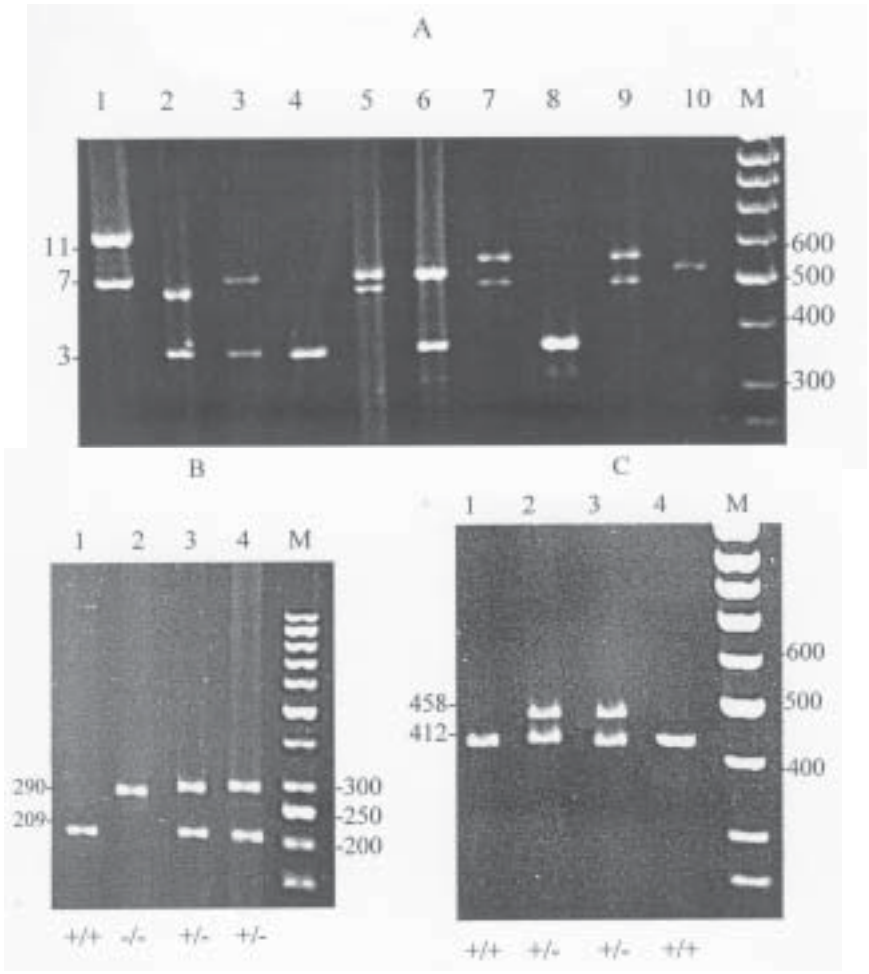

Fig.1. Genotyping of VNTR, *Bgl*II and *Eco*RI alleles in Iranian population. A) The VNTR allele: the numbers on the left represent the number of repeats. Lane 1: 7/11; Lane 2: 3/6; Lane 3: 3/7; Lane 4: 3/3; Lane 5: 6/8; Lane 6: 3/8; Lane 7: 7/9; Lane 8: 3/3; Lane 9: 7/9; Lane 10: 8/8; Lane M: 50 bp DNA ladder. B) The *Eco*RI marker: PCR yields a 458 bp fragment. If the restriction site is present, amplified product can be cleaved by *Eco*RI into 412 and 45 bp fragments. C) The *Bgl*II marker: PCR yields a 290 bp fragment. If the restriction site is present, amplified product can be cleaved by *Bgl*II into 209 and 81 bp fragments. The presence and absence of the restriction site are represented by the (+) and (-) symbols.

of 380, 410, 470, 500, 530, 560, 590, 620 and 680 bp with the specific primers used (Goltsov et al. 1992). The bands correspond to alleles with core repeat unit of 3, 5, 6, 7, 8, 9, 10, 11 and 13 copies, respectively. As indicated in the Figure 1a, samples number 4, 8 and 10 are homozygous and the rest are heterozygous. Alleles with 5, 10 and 13 core repeat units were found with a very low frequency in the population studied. Interestingly, the allele 13 was found specific to Iranian population. The allele 13 was previously found in our study and others (Kamkar et al. 2003; Vallian et al. 2005). Moreover, allele 5 which was only found in one heterozygote individual has not been observed in Iranian population previously.

Genotyping of *Bgl*II and *Eco*RI markers were performed using specific primers. Amplification of the genomic region of the markers yielded in a 290 bp fragment for *Bgl*II and a 458 bp fragment for *Eco*RI (data not shown). Following amplification, the PCR products were digested with the corresponding restriction enzymes. In Figures 1b and c, the results for digestion of *Bgl*II and *Eco*RI markers were represented. The allele frequency, expected and observed frequency of heterozygosity of the markers were estimated using GENEPOP website (Raymond et al. 1995). As shown in Table 1, the VNTR allele with 3 core repeat units had the highest frequency, and the VNTR allele with 5 core repeat units had the lowest. Furthermore, the observed heterozygosity for the VNTR marker was 70% which was calculated based on the identified unrelated heterozygous individuals. This indicated a relatively high degree of heterozygosity for the marker in Iranian population.

Table 1: The frequency distribution and the expected and observed heterozygosity frequency of VNTR marker in Iranian population.
PAHVNTR

Allele	*Size, bp*	*Freq.*
3	380	0.388
5	410	0.003
6	470	0.028
7	500	0.224
8	530	0.171
9	560	0.112
10	590	0.034
11	620	0.019
13	680	0.022

The frequency of *Bgl*II and *Eco*RI restriction sites were 0.27 and 0.52, repectively (Table 2). The observed heterozygosity for *Bgl*II was 41% and for *Eco*RI was 83%. The data indicated that for *Eco*RI, the observed heterozygosity was much higher than the expected heterozygosity. However, for the VNTR and *Bgl*II markers, similar data was found for both observed and expected heterozygosity. Together, the results indicated that the *Eco*RI and VNTR are more informative than *Bgl*II.

The estimation of haplotype frequency was obtained using three computer programs including FBAT, Arlequine and PHASE (Rabinowitz 2000; Schneider et al. 2000; Marchini et al. 2006). First the FBAT program was used to analyze haplotypes in 20 families. Each haplotype was represented as a three-digit number, in which the order of the numbers indicated *Bgl*II, *Eco*RI and the VNTR marker, respectively. The first two digits 1 and 2 refers to the presence (+) or absence (-) of restriction sites for *Bgl*II and *Eco*RI, and the third digit indicated the VNTR alleles. The VNTR alleles were arranged from 1 (3 repeats) to 9 (13 repeats) on the basis of number of repeat units. Among 36 possible haplotypes estimated, ten different haplotypes named 211, 221, 215, 111, 214, 115, 121, 216, 224, and 114 had the frequency equal or above 5 percent (Table 3). However, by estimation maximization (EM) algorithm using the Arlequin program, the frequency of only eight haplotypes including 211, 221, 215, 115, 121, 224, 114 and 226 was ≥5%. The estimation of haplotype frequency using Bayesian algorithm in the PHASE program suggests the following eight haplotypes- 211, 221, 215, 111, 214, 121, 216 and 225- to be in frequency ≥5% in Isfahan population. Comparison of the data produced by three programs, indicated that at least seven haplo-

Table 2: The frequency distribution and the expected and observed heterozygosity frequency of *Bgl*II and *Eco*RI restriction sites in Iranian population.

BglII				*EcoRI*			
Allele	*Freq*	*Heterozygosity*		*Allele*	*Freq*	*Heterozygosity*	
+	0.276	Expected	Observed	+	0.526	Expected	Observed
-	0.724	40%	41%	-	0.474	50%	83%

Note: The presence and absence of the restriction sites are represented by the (+) and (-) symbols in this table.

Table 3: Haplotypes frequency determined by three programs.

Haplotype[1]	*FBAT*	*PHASE*	*Arlequin*
211	0.150000	0.175146	0.186092
221	0.131249	0.104542	0.109383
215	0.087500	0.086896	0.099117
111	0.075000	0.072702	0.023497
214	0.074999	0.063582	0.042675
115	0.062499	0.039286	0.062160
121	0.056250	0.055944	0.081029
216	0.056250	0.059551	0.042525
224	0.050001	0.043059	0.053845
114	0.050000	0.036599	0.078480
226	0.037500	0.048349	0.060973
116	0.031250	0.028598	0.040455
124	0.025000	0.023426	0.000000
126	0.025000	0.021835	0.014381
129	0.025000	0.002109	0.000000
225	0.012500	0.060403	0.038723
127	0.012500	0.010201	0.016667
229	0.012500	0.006156	0.016667
217	0.012500	0.010317	0.008333
125	0.012499	0.013415	0.000000

1 For each haplotype, the numbers represent *Bgl*II, *Eco*RI and VNTR alleles, respectively, so that 1 and 2 refers to presence and absence of restriction site in *Bgl*II and *Eco*RI, and the VNTR alleles are arranged from 1 (3 repeats) to 9 (13 repeats) on basis of the number of their core repeat unit.

types were informative in the population studied when the estimation of haplotype frequency were performed by use of EM and Bayesian algorithms.

Given the presence of relatively high proportion of informative haplotypes in the population, D' and χ^2 values for the pairing of these three markers were calculated using 2LD program (Zhao 2004). D' ranges between 0 and 1. When D'=1 the LD was completed. When D' was estimated zero, it indicated linkage equilibrium between two loci. The null hypothesis of random association between pairs of alleles at the two loci (D'=0) was tested by χ^2 value (Gaunt et al. 2006). The χ^2 value for three pair wise marker obtained by means of 2LD programs and is compared with χ^2 obtained from the chi-square table ($p \leq 0.05$). As illustrated in Table 4, the results of estimation for D' values were smaller than 0.5. The 2LD program calculated degree of freedom from alleles present and calculated χ^2 have (m-1) (n-1) degree of freedom, where m and n is the number of alleles at markers 1 and 2 (Zhao 2004).

Table 4: D' and χ^2 values for three possible pairing of markers in the *PAH* locus.

Pairing of markers	*D'*	χ^2	*df*
*Bgl*II- *Eco*RI	0.021	0.03	1
*Bgl*II-VNTR	0.315	24.19	8
*Eco*RI-VNTR	0.342	34.06	8

DISCUSSION

Linkage analysis using polymorphic markers present in the *PAH* gene region, facilitated carier detection and prenatal diagnosis in PKU families (Tyfield 1997). Linkage analysis requires the use of at least two informative polymorphic markers. The degree of heterozygosity of the *PAH* markers, and their informative haplotypes have been reported to be population specific (Kidd et al. 2000). Among the markers present in the *PAH* gene region, two intragenic markers including *Bgl*II and *Eco*RI, and the VNTR exargenic marker were selected for the present study. The allele and haplotype frequency of the markers were estimated in Isfahan population.

The frequency distribution of the VNTR alleles was compared with the data from other populations (Goltsov et al. 1992; Akhmetova et al. 2000; Korshunova et al. 2004) (see Table 1 and 5). Intriguingly, the results showed a similar discontinuous distribution of the allele sizes at the VNTR marker in Isfahan population compared to other populations. Moreover, the data showed that the VNTR allele with 3 core repeat units had the highest frequency in Isfahan population which was similar to other populations studied around the world. In view of the evolutionary aspects, this notion may suggest that the VNTR allele with 3 repeats has been the first allele in every population, and that the other alleles were formed through endo-replication presumably by unequal crossing over.

The VNTR allele with 8 core repeat units was the third frequent-allele in Isfahan population. This allele has been reported to be the second frequent-allele in other populations as well (Kidd et al. 2000). Among the rest of the VNTR alleles detected in this study, the allele with 4 and 13 repeats seems to be specific to the Iranian population. At present, there is no report for these two alleles in other populations. The presence of such a wide spectrum of the VNTR alleles in the present study suggests that the Iranian population could be considered as one of the populations with a high degree of heterogeneity.

Allele frequency at *Bgl*II and *Eco*RI suggest that the latter was in much higher frequency (Table 2), which makes this marker more

informative in carrier diagnosis applications. No significant difference between the observed and expected hetreozygosity is evident except the *Eco*RI. Therefore, all three markers could be considered useful for linkage analysis in PKU families (see Table 1 and 2). Due to the highest level of observed heterozygosity of *Eco*RI, this marker could be more informative than other markers, and therefore could be considered as the first marker to be used in carrier detection and prenatal diagnosis of the PKU disease in Iranian population.

The FBAT program works on family data to estimate haplotype frequency, but the other programs are able to estimate haplotype frequency on unrelated individuals. Estimation of haplotype frequency of the markers using FBAT program showed that almost ten haplotypes could be introduced as common haplotypes in Iranian population. The haplotype frequency estimation using Arlequin and PHASE programs implement estimation maximization (EM) and Bayesian algorithm, respectively. In the EM and Bayesian algorithm, the estimation of hapltype frequency was performed on 60 unrelated individuals. Seven out of ten common haplotypes which were identified through genotyping data on different families using FBAT, also showed a frequency of ≥5% in both Arlequin and PHASE programs. These result indicate that the estimated frequencies of common haplotypes do not differ significantly using phase-known versus phase-unknown data. The comparison of the obtained results from the three programs showed that the estimation of haplotype frequency by use of Bayesian algorithm in the PHASE program was similar to the obtained result from FBAT program in comparison to EM algorithm in Arlequin program. In fact, this data inducted that the PHASE program was more suitable than Arlequin in cases where knowledge of rare haplotypes was critical. For instance, haplotypes that were not predicted by the EM algorithm (i.e., those that had frequency estimation of zero), were identified when haplotype frequency was estimated by means of data obtained from family genotyping (e.g. haplotypes 124, 129 and 125). Therefore, when common haplotypes are most important, frequency estimation based on the phase-unknown data of unrelated individuals will be sufficient. Furthermore, the estimation of D' and chi-square (χ^2) for the above markers showed that they were not in linkage disequilibrium. Therefore, a large percentage of the chromosomes in this population may display uncommon or rare haplotypes. In such cases, the overall level of uncertainty from phase-unknown data will be larger.

Together, the results from the present study demonstrated that the degree of heterozygosity for *Bgl*II and *Eco*RI, and the VNTR markers were high in Isfahan population. These markers could be used in families with at least one affected child

Table 5: A comparison of frequency distributions of the PAH VNTR alleles in Iranian population and the populations from Northern Caucasus and Volga–Ural region.

*Population**	*Allele (number of repeats)*										
	13	*12*	*11*	*10*	*9*	*8*	*7*	*6*	*5*	*4*	*3*
Karachais	0.000	0.000	0.052	0.000	0.148	0.371	0.085	0.010	0.010	0.000	0.324
Kumyks	0.000	0.000	0.045	0.000	0.096	0.373	0.068	0.000	0.009	0.000	0.409
Kuban Nogais	0.000	0.017	0.017	0.000	0.091	0.329	0.136	0.006	0.006	0.000	0.398
Karanogais	0.000	0.003	0.017	0.000	0.084	0.250	0.186	0.008	0.003	0.000	0.449
Adygs	0.000	0.013	0.044	0.000	0.146	0.364	0.070	0.006	0.000	0.000	0.357
Bashkirs	0.000	0.030	0.000	0.000	0.109	0.277	0.169	0.003	0.003	0.000	0.409
Tatars	0.000	0.060	0.000	0.000	0.180	0.270	0.050	0.010	0.000	0.000	0.430
Chuvashes	0.000	0.030	0.000	0.000	0.050	0.400	0.080	0.000	0.000	0.000	0.440
Maris	0.000	0.080	0.000	0.000	0.070	0.210	0.170	0.000	0.000	0.000	0.470
Mordovians	0.000	0.030	0.000	0.000	0.130	0.370	0.030	0.000	0.000	0.000	0.440
Udmurts	0.000	0.060	0.000	0.000	0.220	0.220	0.130	0.000	0.000	0.000	0.370
Komis	0.000	0.020	0.000	0.000	0.040	0.250	0.220	0.000	0.000	0.000	0.470
Russians	0.000	0.010	0.000	0.000	0.060	0.290	0.190	0.000	0.000	0.000	0.450
Europeans	0.000	0.040	0.000	0.000	0.110	0.450	0.120	0.000	0.000	0.000	0.280
Chinese	0.000	0.000	0.000	0.000	0.000	0.160	0.000	0.000	0.000	0.000	0.840
Iranian**	0.022	0.000	0.019	0.034	0.112	0.171	0.224	0.028	0.000	0.003	0.388

*The population data from studies in references Goltsov et al. 1992; Akhmetova et al. 2000; Korshunova et al. 2004
**Data from the present study.

for prenatal diagnosis and carrier detection with a high degree of accuracy regardless of the nature of mutation. Finally, the data suggested that the combination of *Bgl*II-*Eco*RI-VNTR could be used as an informative tool in performing carrier and prenatal diagnosis of *PAH* gene mutations in Iranian population.

ACKNOWLEGMENTS

The work was supported by grant number 851119 from the department of research of the University of Isfahan, Isfahan, IR Iran.

REFERENCES

Akhmetova VL, Viktorova TV, Khusnutdinova EK 2000. Molecular analysis of the VNTR polymorphism at the phenylalanine hydroxylase gene in populations of the Volga-Ural region. *Genetika,* 36: 1161-1165.

Cali F, Dianzani I, Desviat LR, Perez B, Ugarte M, et al. 1997. The STR252-IVS10nt546-VNTR7 phenylalanine hydroxylase in five mediterranean samples. *Hum Genet,* 100: 350-355.

DiLella AG, Kwok SCM, Ledley FD, Marvit J, Woo SLC 1986. Molecular structure and polymorphic map of the human phenylalanine hydroxylase gene. *Biochemistry,* 25: 743-749.

Dworniczak B, Wedemeyer N, Horst J 1991. PCR detection of the *Bg*II RFLP at the human phenylalanine hydroxylase (*PAH*) locus. *Nucleic Acids Res,* 19: 1958.

Eisensmith RC, Woo SLC 1992. Updated listing of haplotypes at the human phenylalanine hydroxylase (PAH) locus. *Am J Hum Genet,* 51: 1445-1448.

Goldstein DB 2001a. Islands of linkage disequilibrium. *Nature Genetics,* 29: 109-111.

Goldstein DB, Weale ME 2001b. Population genomics: Linkage disequilibrium holds the key. *Current Biology,* 11: R576-R579.

Goltsov AA, Eisensmith RC, Konecki DS, Lichter-Konecki U, Woo SLC 1992. Association between mutations and a VNTR in the human phenylalanine hydroxylase gene. *Am J Hum Genet,* 51: 627-636.

Kamkar M, Saadat M, Saadat I, Haghighi G 2003. Report of VNTR with 13 repeats linked to phenylalanine hydroxylase locus in unaffected members of two PKU families. *Iran Biomed J,* 7: 89-90.

Kidd JR, Pakstis AJ, Zhao H, Lu RB, Okonofua FE, et al. 2000. Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. *Am J Hum Genet,* 66: 1882-1899.

Koochmeshgi J, Bagheri A, Hosseini-Mazinani SM 2002. Incidence of phenylketonuria in Iran estimated from consanguineous marriages. *J Inherit Metab Dis,* 25: 80-81.

Korshunova TY, Akhmetova VL, Kutuev IA, Khusainova RI, Guseinov GG, Khusnutdinova EKK 2004. Analysis of the VNTR polymorphism at the *PAH* and eNOS genes and the CCR5 gene deletion in populations of Northern Caucasus. *Russian J Genet,* 40: 321- 325.

Lewontin RC 1964. The interaction of selection and linkage. *Genetics,* 49: 49–67.

Marchini J, Cutler D, Patterson N, Stephens M, Eskin E, et al. 2006. A comparison of phasing algorithms for trios and unrelated individuals. *Am J Hum Genet,* 78: 437-450.

Matthew M 2004. Phenylketonuria: Defects in amino acid metabolism. *SCJMM,* 5: 57-61.

Michalatos-Beloin S, Tishkoff SA, Bentley KL, Kidd KK, Ruano G 1996. Molecular haplotyping of genetic markers 10 kb apart by allele-specific long-range PCR. *Nucleic Acids Res,* 24: 4841-4843.

Miller SA, Dykes DD, Polesky HF 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res,* 16: 1215.

Rabinowitz L 2000. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Human Heredity,* 50: 211-223.

Raymond M, Rousset F 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Heredity,* 86: 248-249.

Ruano G, Kidd KK 1989. Direct haplotyping of chromosomal segments from multiple heterozygotes via allele-specific PCR amplification. *Nucleic Acids Res,* 17: 8392.

Ruano G, Kidd KK, Stephens JC 1990. Haplotype of multiple polymorphisms resolved by enzymatic amplification of single DNA molecules. *Pro Nat Acad Sci,* 87: 6296-6300.

Schneider S, Roesli D, Excofier yL 2000. *Arlequin: A Software for Population Genetics Data Analysis, Version 2.000.* Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Scriver CR, Hurtubise M, Konecki D, Phommarinh M, Prevost L, et al. 2003. PAHdb 2003: what a locus-specific knowledgebase can do. *Hum Mut,* 21: 333-344.

Tyfield LA 1997. Phenylketonuria in Britain: genetic analysis gives a historical perspective of the disorder but will it predict the future for affected individuals? *J Clin Pathol: Mol Pathol,* 50: 169-174.

Vallian S, Barahimi E, Moeini H 2003. Phyenylketonuria in Iranian population: a study in Institutions for mentally retarded in Isfahan. *Mut Res,* 526: 45-52.

Vallian S. Barahimi E 2005. Analysis of the importance of PAHVNTR marker in carrier detection of phenylketonuria disease in Isfahan population. *Genetics in the 3rd Millennium,* 3: 572-576.

Xu J 2006. Extracting haplotypes from diploid organisms. *Curr Issues Mol Biol,* 8: 113-122.

Zhao JH 2004. 2LD, GENECOUNTING and HAP: computer programs for linkage disequilibrium analysis. *Bioinformatics,* 20: 1325-1326.

Zhao H, Pfeiffe R, Gail MH 2003. Haplotype analysis in population genetics and association studies. *Pharmacogenomics,* 4: 171-178.

Zhao H, Nettleton D, Soller M, Dekkers JCM 2005. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genet Res,* 86: 77-87.