Genetic Diversity and Balancing Selection within the Human Phenylalanine Hydroxylase (PAH) Gene Region in Iranian Population

A Haghighatnia [1, 2], *S Vallian [1], J Mowla [2], Z Fazeli [1]

1. *Division of Genetics, Dept. of Biology, Faculty of Sciences, University of Isfahan, Isfahan, Iran*
2. *Dept. of Molecular Genetics, Faculty of Basic Sciences, Tarbiat Modares University, Tehran, Iran*

Abstract

Background: Genetic diversity of three polymorphic markers in the phenylalanine hydroxylase *(PAH)* gene region including *Pvu*II (a), PAHSTR and *Msp*I were investigated.

Methods: Unrelated individuals (n=139) from the Iranian populations were genotyped using primers specific to *PAH* gene markers including *Pvu*II(a), *Msp*I and PAHSTR. The amplified products for *Pvu*II(a), *Msp*I were digested using the appropriate restriction enzymes and separated on 1.5% agarose. The PAHSTR alleles were identified using polyacrylamide gel electrophoresis followed by silver staining. The exact size of the STR alleles was determined by sequencing. The allele frequency and population status of the alleles were estimated using PHASE, FBAT and GENEPOP software.

Results: The estimated degree of heterozygosity for PAHSTR, *Msp*I and *Pvu*II (a) was 66%, 56% and 58%, respectively. The haplotype estimation analysis of the markers resulted in nine informative haplotypes with frequencies ≥5%. Moreover, the results obtained from Ewens-Watterson test for neutrality suggested that the markers were under balancing selection in the Iranian population.

Conclusion: These findings suggested the presence of genetic diversity at these three markers in the *PAH* gene region. Therefore, the markers could be considered as functional markers for linkage analysis of the *PAH* gene mutations in the Iranian families with the PKU disease.

Keywords: Phenylalanine hydroxylase (PAH), Phenylketonuria (PKU), Genetic diversity, Iran

Introduction

Deficiency in the human phenylalanine hydroxylase (PAH) enzyme due mutations in the coding region of *PAH* gene is associated with phenylketonuria (PKU) (1-4). The *PAH* gene is 90 kbp in length with 13 exons (5-7). More than 500 different mutations have been identified in this gene associated with the PKU disease (3, 7). Therefore, direct mutation analysis of the disease, especially for families with a diseased member (child) is time consuming and very expensive. Alternatively, indirect analysis of mutations, using linked markers have proven to be useful. However, linkage analysis requires the presence of informative haplotypes in the population (8). There are two multiallelic markers in the *PAH* gene region; a variable number of tandem repeats (VNTR) at the 3′ end of the gene, and an intra-genic short tandem repeat (STR) in intron 3. Moreover, eight biallelic markers are present along the gene region including *Bgl*II, *Pvu*II(a) and *Pvu*II(b), *Eco*RI, *Xmn*I, *Msp*I, *Hind*III and *Eco*RV (5,7,9,10).

*Corresponding Author: Tel: +983117932456, E-mail address: sadeqvallian@gmail.com

Although each marker can be analyzed independently, it is more informative to analyze them in groups. When multiple markers in a chromosomal region were used to assess their association with a disease, determination of haplotype was more efficient than separate analyses of individual markers (11). There are two categories of data which can unambiguously determine haplotypes, i) population-based data and ii) pedigree or family-based data (12). Several algorithms have been developed to infer haplotypes and estimate haplotype frequency by use of genotyping data obtained from unrelated individuals (unknown phase or population-based data) (11). These algorithms include parsimony, Expectation-Maximization, Bayesian, Perfect and Imperfect phylogeny (13-18). However, a proportion of the inferred haplotypes might be incorrect which could be considered as the main disadvantage of all these algorithms (17-20). In family-based methods, usually families with several generations were genotyped across several loci and the linkage phase could be determined by recording alleles passed from one generation to the next (12). It has been reported that switch errors where a segment of the maternal haplotype was incorrectly joined to the paternal one, occur extraordinarily rarely in the trio samples (parents and their offspring). The switch errors were found higher in the unrelated individuals due to the lack of information. However, even for the unrelated individuals, the estimation of haplotype frequency can be performed with relatively high accuracy (21)

According to the theory of neutral molecular evolution, most of the polymorphisms were selectively neutral in a population. A number of neutrality tests were devised that make general interferences about the causes of molecular evolution (22, 23). By using these tests, the impact of selection was studied at a number of genes, such as *ABO* blood group, the major histocompatibility antigens (*HLA*), lactase (24-26). Although, the more polymorphisms at the *PAH* gene was placed at non-coding region, but it was necessary to survey neutrality theory for these markers. These studies could increase our insight into evolutionary history of human populations.

In the present study, the genetic structure of three other markers of the *PAH* gene including *Pvu*II(a), STR and *Msp*I was investigated

Material and Methods

DNA sample
Overall, 139 individuals consist of 100 unrelated healthy persons and 13 family trios (with three persons, two parents, and one child) from Isfahan population of Iran were included in the study.

Genotyping
Genotyping was performed by first extracting genomic DNA from peripheral blood using a standard salting out procedure (27). The DNA samples were amplified for two biallelic restriction fragment length polymorphism (RFLP) markers (*Pvu*II (a) and *Msp*I), and one multiallelic short tandem repeat marker (PAHSTR). The *Pvu*II (a) site is located at the 5' end of intron 2 (GeneBank AF003966) (28). The position of *Msp*I site is in intron 7, downstream to exon 8 (GeneBank AF003967) (28). PCR reaction was performed in 25 μL total volume containing 50 ng DNA, 500 mM KCl, 100 mM Tris–HCl (pH 8.4), 50 mM MgCl$_2$, 200 μM dNTP, 5U Taq DNA polymerase and 10 pmole of each forward and reverse primers. Initial denaturation was performed at 94 °C for 4 minutes, followed by 30 cycles including 1 minute denaturation at 95 °C; different annealing temperature depending on the marker used (57°C for PAHSTR and 58°C for *Msp*I and *Pvu*II (a)) for 1 minute), 1 minute extension at 72 °C; followed by a final extension for 10 minutes at 72 °C. The amplified DNA for *Msp*I and *Pvu*II loci were subjected to enzymatic digestion. The digestion products were separated on 1.5% agarose gel electrophoresis (29, 30).

The exact size of the PAHSTR alleles were examined by separating of the alleles on PAGE, gel extraction using a melting and freezing procedure as described previously (31) and cloning into a PTZ57R/T using TA clone™ PCR cloning Kit (Fermentas, Germany). The cloned PAHSTR alleles were sequenced using an ABI 737 sequencer (Perkin Elmer/ABI).

Statistical analysis

Allele frequency of the observed and expected heterozygosity was determined using the GENEPOP software (32). The haplotype frequency was estimated using PHASE (33) and FBAT software (34). Genotype data in 100 unrelated individuals were used by PHASE software and haplotype frequency estimation of genotype data obtained from the families were performed by FBAT.

The genotyping data obtained from family trios was used to infer haplotype phase in each individual. The inferring of haplotype phase was done by use of PedPhase program (35, 36). In order to provide a better description of the obtained results from the haplotype estimation, the estimation of LD (linkage disequilibrium) was also performed for 100 unrelated individuals using 2LD computer program (37).

The Ewens–Watterson homozygosity test of neutrality was performed using PyPop program (38). Ewens-Watterson test was also performed using Popgene32 software (available at http://www.ualberta.ca/~fyeh/download.htm).

The observed F (sum of square of allelic frequency) and limit (upper and lower) at 95% confidence region were calculated using similar software.

Results

Three markers in the phenylanine hydroxylase (PAH) gene region including PAHSTR, *Pvu*II (a) and *Msp*I were genotyped. Genotyping of PAHSTR allele revealed the presence of 9 different alleles with 260, 252, 248, 244, 240, 236, 232, 228, 224 bp in size. The exact size of the alleles was confirmed by sequencing (Fig. 1, line M).

The *Pvu*II (a) and *Msp*I biallelic markers were genotyped using PCR followed by restriction enzyme digestion (Fig. 2).

The allelic frequency and the expected and observed heterozygosity of the markers were estimated by GENEPOP software as depicted in Tables 1, 2 and 3.
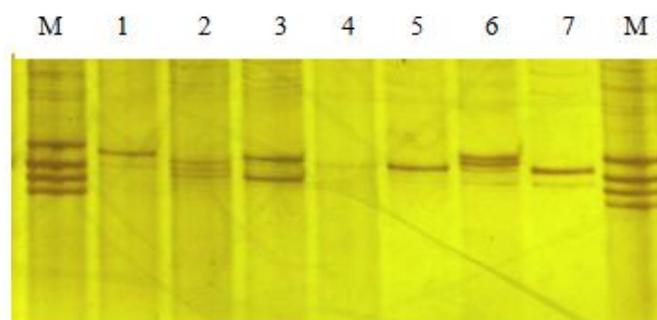

Fig. 1: Genotyping of the PAHSTR allele located in the *PAH* gene region. The PAHSTR alleles were genotyped and analyzed in 12% polyacrylamide gel electrophoresis followed by silver staining. The indicated alleles were as follows: Lane 1, allele number 7; Lane 2, alleles 8 and 9; Lane 3, alleles 7 and 4; Lane 4, allele 6, lane 5, allele 6; Lane 6, alleles 8 and 7; and Lane 7, allele 6. M, represents DNA ladder composed of four sequenced alleles
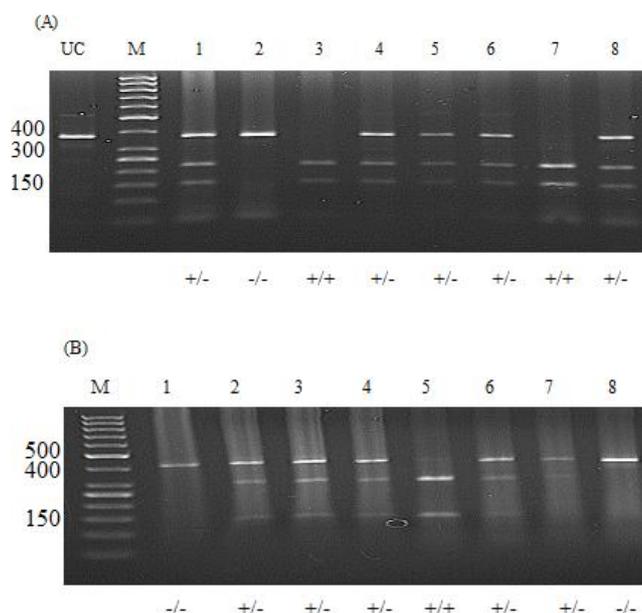

Fig. 2: Genotyping of *Pvu*II (a) and *Msp*I markers. A) For *Pvu*II (a) marker, PCR using specific primers resulted in a 375 bp fragment. In the presence of the restriction site, the amplified product was digested by *Pvu*II (a) to 225 and 150bp fragments. B) For *Msp*I marker, PCR using specific primers resulted in a 425bp fragment. In the presence of the restriction site, the amplified product was digested by *Msp*I into 300 and 125bp fragments. Symbols (+) and (-) represent the presence and absence of the restriction site, respectively. Line M, represents 100bp DNA ladder

The genotyping data for PAHSTR showed that alleles with 244 bp and 224 bp had the most and lowest frequency, respectively. The observed heterozygosity for the PAHSTR marker was lower than the expected heterozygosity (Table 3). The frequency of *Msp*I and *Pvu*II (a) restriction sites were 0.4800 and 0.6300 in the Iranian population.

Table 1: The frequency distribution of the *Msp*I and *Pvu*II(a) markers in the Iranian population

Frequency		
Site	Allele (+)	Allele (-)
MspI	0.4800	0.5200
PvuII(a)	0.6300	0.3700

Table 2: The frequency distribution of the PAHSTR marker in the Iranian population

PAHSTR		
Allele	Frequency	Size (bp)
1	0.0300	260
2	0.1050	252
3	0.1850	248
4	0.2200	244
5	0.2150	240
6	0.1050	236
7	0.1000	232
8	0.0350	228
9	0.0050	224

Table 3: Expected and observed heterozygosity of three markers, PAHSTR, *Msp*I and *Pvu*II(a) at the *PAH* gene in the Iranian population

Heterozygosity	PAHSTR	*Msp*I	*Pvu*II(a)
Observed	66%	56%	58%
Expected	84.1%	50.17%	46.8%

The haplotype frequency for three markers was estimated. As shown in Table 4, among 36 possible haplotypes by PHASE software, nine haplotypes were considered as informative haplotypes with frequencies ≥5% in the Iranian population. Haplotype estimation of families was performed by FBAT software, which resulted in 7 informative haplotypes with frequencies ≥5% among 23 possible haplotypes. The presence and absence of restriction sites in *Pvu*II (a) and *Msp*I was indicated

as number 1 and 0, respectively. For the PAHSTR marker, number 1 to 9 was used to indicate different allele sizes (Table 5).

Table 4: The estimation of *Pvu*II(a)-PAHSTR-*Msp*I haplotypes frequency at the *PAH* gene by two programs, FBAT and PHASE in the Iranian population

Haplotype*	FBAT	PHASE
111	0.009615	0.004907
110	0.000000	0.004039
121	0.000000	0.026179
120	0.009615	0.057420
131	0.048077	0.055828
130	0.038462	0.059984
141	0.201923	0.053070
140	0.105769	0.089032
151	0.019231	0.070292
150	0.000000	0.088447
161	0.009615	0.027882
160	0.086538	0.036665
171	0.009615	0.012482
170	0.0000000	0.032121
181	0.019231	0.013548
180	0.000000	0.006676
191	0.000000	0.000301
190	0.000000	0.001126
011	0.009615	0.020803
010	0.000000	0.000251
021	0.057692	0.014405
020	0.009615	0.006995
031	0.067308	0.058781
030	0.038462	0.010408
041	0.086538	0.072781
040	0.048077	0.025116
051	0.000000	0.040881
050	0.019231	0.005380
061	0.009615	0.028945
060	0.067308	0.028945
071	0.009615	0.005851
070	0.019231	0.049546
081	0.000000	0.000350
080	0.000000	0.009426
091	0.000000	0.000150
090	0.000000	0.003422

*The numbers in each haplotype represent the alleles of *Pvu*II(a), STR and *Msp*I markers, respectively.

The haplotype frequency of *Pvu*II(a), *Msp*I, PAHSTR markers was estimated by use of two software, FBAT for families, and PHASE for unrelated individuals. The results of haplotype estimation showed the presence of 9 [120, 131, 130, 141, 140, 151, 150, 031, 041] and 7 [141, 140, 160, 021, 031, 041, 060] informative haplotypes by use of PHASE and FBAT, respectively. Four haplotypes including 141, 140, 031 and 041 were estimated as informative haplotypes using both computer programs. The data using PHASE program showed that five haplotypes including 120, 131, 130, 151, 150 could be considered as informative haplotypes in the Iranian population.

In the next step, the determination of haplotype phase at 13 family trios was done by use of Ped-Phase program. As presented in Table 4, the haplotype phase of all three members, two parents and their offspring, at five of thirteen families was only determined. The origin of haplotype phase (maternal or parental) was also distinguished in these families. At eight remainder pedigrees, the genotyping data of families was not enough for haplotype phasing. Estimation of haplotype frequency using PHASE or FBAT programs among

phase-known haplotypes indicated the presence of seven haplotypes 021, 151, 160, 031, 141, 140 and 041 with frequency >5% (Table 4).

Table 5: D' and χ^2 values for three possible pairing of markers in the *PAH* locus

pairing of markers	D'	χ^2	df
*Pvu*II(a)-PAHSTR	0.217190	16.89	8
*Pvu*II(a)-*Msp*I	0.179677	3.93	1
PAHSTR-*Msp*I	0.318775	27.75	8

The pattern of LD for the pairing of these three markers was assessed by calculating average D' and χ^2. As indicated in Table 5, the results of estimation for D' values for all three possible pairing of markers were smaller than 0.5. The calculated χ^2 values were also smaller than χ^2 values obtained from the chi-square table ($P < 0.05$).

Analysis of neutrality of all three markers showed that the F value (sum of square of allelic frequency) located outside the lower and upper limit of 95% confidence region of expected F value at two markers of *PAH* gene, PAHSTR and *Msp*I (Table 6).

Table 6: The Ewens-Watterson test for neutrality at three markers of the *PAH* gene in the Iranian population

The marker	k	Observed F	Expected F	Fnd	L95	U95
*Pvu*II(a)	2	0.5392	0.8314	-1.7443	0.5018	0.9901
PAHSTR	9	0.1631	0.3436	-1.4452	0.1865	0.6798
*Msp*I	2	0.5004	0.8314	-1.9756	0.5032	0.9901

k: Number of alleles; Observed F: Observed sum of the squared of allelic frequency; L95, U95: The 95% confidence interval upper and lower limit; Fnd: Normalized deviate of F

Discussion

Genetic diversity plays an important role in the application of genetic markers in linkage analysis in carrier diagnosis of genetic diseases. In the present study, genetic diversity of three markers *Pvu*II (a), *Msp*I, PAHSTR located in the *PAH* gene region were examined in the Iranian population. Genotyping of the markers showed the presence of nine different alleles for the PAHSTR marker in this population. Comparison of the frequency of the PAHSTR alleles in the Iranian

population with other populations indicated that the frequency of these alleles was very similar to those of other populations (Table 2) (39). However, minor differences were observed as to alleles with sizes of 216 and 256 bp were absent in the Iranian population. Similar situation was observed in other Asian populations, which were studied previously (39). As in other populations like South America, North America, Oceania and some part of East Asia, the 244 bp allele had the most fre-

quency in the Iranian population (39). The allele with 224 bp was absent in some populations such as South and North America, Europe and Africa. Moreover, in some other populations, e.g. some Asian regions, the allele with 224 had the lowest frequency.

Among the biallelic markers examined, *Pvu*II (a) showed higher degree of heterozygosity (Table 1), and therefore, this marker could be suggested as an informative and applicable marker for carrier detection applications. The observed heterozygosity for PAHSTR marker was found lower than the expected one. Therefore, this may indicate that this marker could not be optimal for diagnostic purposes. However, combination of PAHSTR and PAHVNTR polymorphic systems could significantly increase the informativeness of the STR/VNTR haplotype in linkage analysis in PKU families. Moreover, the data showed that the observed heterozygosity of *Msp*I and *Pvu*II (a) was higher than their expected heterozygosity. The degree of observed heterozygosity of *Pvu*II (a) marker was higher than *Msp*I, suggesting that this marker could be more applicable than *Msp*I marker in carrier detection and prenatal diagnosis of the PKU disease in the Iranian population.

Estimation of haplotype frequency by use of FBAT program, indicated a frequency of <5%. Because, the estimation of haplotype frequency was performed by use of only genotype data obtained from 13 family trios, it is likely that the low frequency of these haplotypes was resulted from insufficient number of individuals, which were used for estimation of haplotype frequency using the FBAT program.

The phase of *Pvu*II (a)-PAHSTR-*Msp*I haplotype was also determined at 13 family trios by use of PedPhase program. The inferring of *Pvu*II (a)-PAHSTR-*Msp*I haplotype phase indicated that haplotype diversity was high at the Iranian population. Determination of haplotype phase at *PAH* gene requires genotype data obtained from other family members. Similar results were also reported previously for *Bgl*II-*Eco*RI-PAHVNTR haplotypes (40). As shown in Table 3, for the markers studied, the observed heterozygosity was >50%. Therefore, it is expected that more indi-

viduals in the Iranian population to be heterozygous for more than one marker and haplotype phasing at *PAH* gene.

Furthermore, the estimation of D' and chi-square ($\chi2$) for the above markers showed that they were not in linkage disequilibrium. The results obtained from LD indicated that these markers at the *PAH* gene were independent to each other. In Ewens-Watterson test for neutrality for the markers, PAHSTR and *Msp*I, at the *PAH* gene lied outside the limit of 95% confidence region. Therefore, these two markers of the *PAH* gene were not neutral and may be linked with some selection traits or genes. Therefore, these two markers might be under genetic hitchhiking, which is a potent force in changing allelic frequency and heterozygosity (41, 42).

The observed and expected frequency (F) was estimated for *Pvu*II (a), *Msp*I and PAHSTR markers by use of Pypop program. The comparison of these data revealed that the observed F was lower than the expected F for all three markers at the *PAH* gene in the Iranian population. These results and negative value of F indicated that these markers were under balancing selection in the Iranian population.

In conclusion, the results obtained from haplotype frequency estimation study suggested that the combination of *Pvu*II (a)/PAHSTR/*Msp*I markers could be used as an informative tool in diagnostic purposes of *PAH* gene mutations in the Iranian population. The results of this study could increase our insight into diversity and evolution history of these markers at the PAH gene in the Iranian population.

Ethical considerations

Ethical issues (Including plagiarism, Informed Consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc) have been completely observed by the authors.

Acknowledgements

References

1. Centerwall SA, Centerwall WR (2000). The discovery of phenylketonuria: the story of a young couple, two affected children, and a scientist. *Pediatrics*, 105: 89-103.
2. Madden M (2004). Phenylketonuria: Defects in amino acid metabolism. *Mol Med (SCJMM)*, 5: 57-61.
3. Scriver CR (2007). The PAH gene, phenylketonuria, and a paradigm shift. *Hum Mutat*, 28 (9): 831-845.
4. Scriver CR, Kaufman S (2001). Hyperphenylalanemia: phenylalanine hydroxylase deficiency. In: *The metabolic and molecular bases of inherited disease*. 8th ed, Mc Graw-Hill, New York. pp. 1667-1709.
5. 5.DiLella AG, Kwok SC, Ledley FD, Marvit J, Woo SL (1986). Molecular structure and polymorphic map of the human phenylalanine hydroxylase gene. *Biochemistry*, 25 (4): 743-749.
6. Tyfield LA (1997). Phenylketonuria in Britain: genetic analysis gives a historical perspective of the disorder but will it predict the future for affected individuals? *Mol Pathol*, 50 (4): 169-179.
7. Scriver CR, Hurtubise M, Konecki D, Phommarinh M, Prevost L, Erlandsen H, et al (2003). PAHdb 2003: What a locus-specific knowledgebase can do. *Hum Mut*, 21: 333-344.
8. Fazeli Z, Vallian S (2009a). Estimation haplotype frequency of BglII/EcoRI/VNTR markers at the PAH gene region in Iranian population. *Int J Hum Gent*, 9(2): 115-121.
9. Goltsove A, Eisensmith R, Konecki D, Lichter-Konecki U, Woo S (1992). Associations between mutations and a VNTR in the human phenylalanine hydroxylase gene. *Am J Hum Genet*, 51: 627-36.
10. Goltsove A, Eisensmith R, Naughton E, Jin L, Chakraborty R, Woo S (1993). A single polymorphic STR system in the human phenylalanine hydroxylase gene permits rapid prenatal diagnosis and carrier screening for phenylketonuria. *Hum Mol Genet*, 2: 577-81.
11. Zhao H, Pfeiffer R, Gail MH (2003). Haplotype analysis in population genetics and association studies. *Pharmacogenomics*, 4 (2): 171-178.
12. Crawford DC, Nickerson DA (2005). Definition and clinical importance of haplotypes. *Annu Rev Med*, 56: 303-320.
13. Boehnke M, Arnheim N, Li H, Collins FS (1989). Fine-structure genetic mapping of human chromosomes using the polymerase chain reaction on single sperm: Experimental design considerations. *Am J Hum Genet*, 45: 21-32.
14. Excoffier L, Slatkin M (1995). Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol*, 12 (5): 921-927.
15. Michalatos-Beloin S, Tishkoff SA, Bentley KL, Kidd KK, Ruano G (1996). Molecular haplotyping of genetic markers 10kb apart by allele-specific long-range PCR. *Nucleic Acids Res*, 24: 4841-4843.
16. Stephens M, Donnelly P (2000). Inference in molecular population genetics. *J R Statist Soc B*, 62 (4): 605-655.
17. Stephens M, Smith NJ, Donnelly P (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*, 68: 978-989.
18. Stephens M, Donnelly P (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet*, 73: 1162-1169.
19. Niu T, Qin ZS, Xu X, Liu JS (2002). Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am J Hum Genet*, 70: 157–169.
20. Zhang S, Pakstis AJ, Kidd KK, Zhao H (2001). Comparisons of two methods for haplotype reconstruction and haplotype frequency estimation from population data. *Am J Hum Genet*, 69: 906–914.
21. Paul B (2006). Origins of the Human Genome Project: Why sequence the human genome

when 96% of it is junk? *Am J Hum Genet,* 79 (4): 603–605.

22. Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, Rieder MJ, et al (2005). Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res,* 15: 1553-1565.

23. Nielsen R (2001). Statistical tests of selective neutrality in the age of genomics. *Heredity,* 86: 641- 647.

24. Bersaglieri T *et al* (2004). Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Gene,* 74: 1111–1120.

25. Hughes AL, Yeager M (1998). Natural selection and the evolutionary history of major histocompatibility complex loci. *Front Biosci,* 3: d509-d516.

26. Seltsam A, Hallensleben M, Kollmann A, Blasczyk R (2003). The nature of diversity and diversification at the ABO locus. *Blood,* 102: 3035–3042.

27. Miller S, Dykes D, Polesky H (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res,* 16: 1215.

28. Fazeli Z, Vallian S (2011). Phenylketonuria from genetics to clinics: An Iranian prospect. *Iranian J Biotech,* 9(3): 163-172.

29. Dworniczak B, Wedemeyer N, Eigel A, Horst J (1991). PCR detection of the PvuII(Ea) RFLP at the human phenylalanine hydroxylase (PAH) locus. *Nucleic Acids Res,* 19: 1958-60.

30. Wedemeyer N, Dworniczak B, Horst J (1991). PCR detection of the MspI (Aa) RFLP at the human phenylalanine hydroxylase (PAH) locus. *Nucleic Acids Res,* 19: 1959-61.

31. Sambrook J, Russell DW (2006). Molecular Cloning: A Laboratory Manual, 3th ed, Cold Spring Harbor Laboratory.

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

33. Marchini J *et al* (2006). A comparison of phasing algorithms for trios and unrelated individuals. *Am J Hum Genet,* 78: 437- 450.

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

35. Li J, Jiang T (2003a). Efficient rule-based haplotyping algorithms for pedigree data. *Proc RECOMB,* 3: 197-206.

36. Li J, Jiang T (2003b). Efficient inference of haplotypes from genotypes on a pedigree. *J Bioinfo Comp Biol,* 1: 41-69.

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

38. Lancaster A, Nelson MP, Single RM, Meyer D, Thomson G (2003). PyPop: a software framework for population genomics, analyzing large-scale multi-locus genotype data. *Pacific Symp Biocomput,* 8: 514–525.

39. Cheung KH, Osier MV, Kidd JR, Pakstis AJ, Miller PL, Kidd KK (2000). ALFRED: an allele frequency database for diverse populations and DNA polymorphisms. *Nucleic Acids Res,* 28 (1): 361-363.

40. Fazeli Z, Vallian S (2009b). The investigation of haplotype phasing efficiency at the *PAH* gene region in Iranian family trios. *Iranian J Pub Health,* 38(4): 136-139.

41. Kumar S *et al* (2009). Genetic diversity analysis of the gohilwari breed of indian goat (*Capra hircus*) using microsatellite markers. *Am J Animal Vet Sci,* 4 (3): 49-57.

42. Maynard-Smith J, Haigh J (1974). The hitchhiking effect of a favorable gene. *Genet Res,* 23: 23-35.