MUTATIONS IN THE PAH GENE: A TOOL FOR POPULATION GENETICS STUDY

MAJA STOJILJKOVIĆ1, ANA STEVANOVIĆ1, MAJA DJORDJEVIĆ2, BRANKA PETRUČEV1, NATAŠA TOŠIĆ1,
TEODORA KARAN DIJURAŠEVIĆ1, SANJA AVEIĆ1, MILENA RADMILOVIĆ1, and SONJA PAVLOVIĆ1

1 Institute of Molecular Genetics and Genetic Engineering, 11000 Belgrade, Serbia
2 Dr Vukan Ćupić Mother and Child Healthcare Institute, 11000 Belgrade, Serbia

Abstract – Phenylketonuria (PKU), an inborn error of metabolism, is caused by mutations in the phenylalanine hydroxylase (PAH) gene. In the Serbian population, 19 different PAH mutations have been identified. We used PAH mutations as molecular markers for population genetics study. The low homozygosity value of the PAH gene (0.10) indicates that PKU in Serbia is heterogeneous, reflecting numerous migrations throughout Southeast Europe. The strategy for molecular diagnostics of PKU was designed accordingly. To elucidate the origin of the most common (L48S) PKU mutation in Serbia, we performed haplotype analysis by PCR-RFLP. Our results suggest that the L48S mutation was imported into Serbia from populations with different genetic backgrounds.

Key words: Phenylketonuria, phenylalanine hydroxylase gene mutations, homozygosity value, expected heterozygosity, haplotype analysis

INTRODUCTION

Phenylketonuria (PKU, MIM#261600) is the most common inborn error of amino acid metabolism in Caucasians (average incidence of 1/10000). It is transmitted in an autosomal recessive pattern and is caused by deficiency of the hepatic enzyme, phenylalanine hydroxylase (PAH, EC 1.14.16.1), which catalyzes the conversion of phenylalanine to tyrosine. Disfunction of the PAH enzyme results in an elevated serum level of phenylalanine and mental retardation, unless dietary intake of phenylalanine is restricted (Donlon et al., 2006).

Deficiency of PAH is mainly caused by mutations in the PAH gene (GenBank accession no. AF404777), mapped to human chromosome 12q24.1. The PAH gene contains 13 exons and 12 large introns spanning approximately 90 kb. Over 500 different mutations have been identified in the PAH gene so far and recorded in the PAH locus knowledgebase (http://www.pahdb.mcgill.ca) (Scriven et al., 2003).

Characterization of PAH mutations is a final confirmation of a defect in the PAH enzyme and could be a helpful tool for phenotype prediction in a newborn infant, for refining diagnosis, and for implementing optimal dietary therapy (Kayaalp et al., 1997; Guldberg et al., 1998; Guttler and Guldberg, 2000). Also, the correlation between mutations in the PAH gene and responsiveness to tetrahydrobiopterin, a cofactor of the PAH enzyme used as a specific therapeutic drug, is under investigation (Blau and Erlandsen, 2004).

In a previous study of the Serbian population, 19 different disease-causing mutations were identified, corresponding to a mutation detection rate of 97%. The most frequent mutations, L48S (21%), R408W (18%), P281L (9%), E390G (7%) and R261Q (6%), account for 60% of all mutant alleles. Less frequent ones are: R158Q (4.4%), I306V (4.4%), IVS12+1G>A (4.4%), Q20X (2.9%), R111X (2.9%), V177L (2.9%), P225T (2.9%), R261X (2.9%), p.S16>XfsX1 (1.5%), S231F (1.5%), R252Q (1.5%), R297H (1.5%), IVS10-11G>A (1.5%), and R413P (1.5%) (Stojiljković et al., 2006).
The most frequent mutation in PKU patients in Serbia and Montenegro is L48S, with a relative frequency of 21%. This mutation was initially reported in Turkey (Konecki et al., 1991). It was later reported in many European populations (Bulgaria, Romania, Czech Republic, Germany, and Belgium - 2%; Croatia and Sicily - 5%; Southern Italy - 11%; Turkish patients in Germany - 13.3%) (Zschocke et al., 2003). However, the exceptional prominence of L48S detected in Serbia suggests the influence of either a founder effect and genetic drift or its autochthonous origin. Haplotypes of PAH are nowadays widely used to determine the chromosomal background on which a mutation arose, as well as its geographic pattern.

The PAH gene sequence contains a large number of recognized polymorphisms. There are three forms of polymorphisms: i) seven biallelic restriction fragment length polymorphisms (RFLPs: BglII, PvuIIa, PvuIib, EcoRI, MspI, XmnI, and EcoRV); ii) multiallelic polymorphisms (variable number of tandem repeats, VNTR, and short tandem repeats, STR); and iii) silent single nucleotide polymorphisms, SNPs (e.g., Q232Q).

RFLP, VNTR and STR alleles can be combined to generate PAH haplotypes. Although several thousands of different polymorphic PAH haplotypes could be generated from combinations of these alleles, far fewer have actually been observed on human chromosomes. Particular PAH haplotypes are associated with disease-causing mutations in European populations (Scriver and Kaufman, 2006).

In this study, we analyzed allele frequencies of PKU mutations in Serbia in order to determine genetic variations at the PAH locus in the Serbian population. We have also performed haplotype analyses to elucidate the origin of the most common PKU mutation (L48S) in Serbia.

MATERIALS AND METHODS

Thirty four patients with PKU, identified through a neonatal screening program or during genetic counseling in Dr Vukan Ćupić Mother and Child Healthcare Institute in Belgrade, were further referred to the Institute of Molecular Genetics and Genetic Engineering in Belgrade for DNA analysis. Detection of PAH gene mutations was successfully performed using PCR-RFLP, DGGE, and sequencing methods (Stojilkovic et al., 2006).

Twelve unrelated patients, compound heterozygotes with L48S being one of two mutations, and one homozygote for L48S, were studied for polymorphism in the PAH gene.

Calculation of Homozygosity

Homozygosity \((j)\) at the PAH locus in the population was determined using the equation \(j = \sum x_i^2\), where \(x_i\) is the frequency of the \(i\)th allele. Here each of the uncharacterized alleles is defined as having a frequency of \(1/N\), where \(N\) is the total number of mutant chromosomes investigated.

This value is the theoretical frequency of patients carrying two identical mutations. Homozygosity values of different populations reflect their mutational heterogeneity for the particular locus (Guldberg et al., 1996).

Calculation of Expected Heterozygosity for Intron Polymorphisms

Expected heterozygosities for individual sites were estimated as one minus the sum of the squares of the allele frequencies \((1-\sum p_i^2)\). For a biallelic system, there are only two different possibilities (Kidd et al., 2000).

Polymorphism Analysis

Venous blood (5-10 mL) was collected in 3.8% Na-citrate anticoagulant and genomic DNA was extracted using standard procedure (Poncze et al., 1983).

Detection of EcoRI and XmnI biallelic restriction fragment length polymorphisms was performed by polymerase chain reaction (PCR) followed by restriction enzyme digestions (PCR-RFLP).

We performed PCR for PAH gene introns 5 and 8. PCR conditions being the same for both EcoRI and XmnI polymorphisms. Performance of PCR was in a 25-µl final volume for a reaction medium containing 100 ng of primers, 200 µM dNTPs, 2.25 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl (pH 8.4), and 1 U Taq Polymerase (Perkin Elmer). Amplification of DNA was achieved by 30 cycles at 95°C for 30 sec, 58°C for 30 sec, and 72°C for 30 sec, preceded by an initial denaturation step (95°C for 5 min) and followed by final elongation (72°C for 10 min). Amplified PCR products were electrophoretically analyzed on 2% agarose gel, stained with ethidium bromide and visualized under UV light.

Primers used for the EcoRI restriction enzyme poly-
morphic site were as follows: 5’-AGAGTTTTTAGCAT-GAAAGGC-3’; and 5’-CTACCCACCAAAAAAGTA-CA-3’. The PCR product was 458 bp long. Primers used for the XmnI restriction enzyme polymorphic site were as follows: 5’-CTGTACTTGTAAGATGCAGC-3’; and 5’-ACTGTCACCGAAATCAGAG-3’. The PCR product was 205 bp long (www.alfred.med.yale.edu).

The amplified products were digested with EcoRI or XmnI (Biolabs, England) according to manufacturer recommendations. In the presence of an EcoRI or XmnI site, the amplified fragment was digested into two fragments of 412 and 45 bp or 110 and 95 bp, respectively, easily distinguishable upon electrophoresis on 8% polyacrylamide gel. The undigested PCR product indicates the absence of an EcoRI or XmnI site. Detection of three fragments (undigested and digested PCR products) indicates heterozygosity at the polymorphic site.

The PAH gene intron 5 DNA fragment of a patient homozygous for the L48S mutation was sequenced for using the ABI PRISM 310 Sequence Analyzer (Applied Biosystems, Foster City, CA, USA).

RESULTS

Heterogeneity

The Serbian population is characterized by a high number of different mutations in the PAH gene: 19 disease-causing mutations have been identified (Stojiljković et al., 2006). This finding suggests that PKU in Serbia is heterogeneous. Also, the number of different mutations in relation to the number of mutant chromosomes is, in itself, a reflection of the allelic variation. In our study, one third of mutations were found on just one chromosome in the population, and only two mutations, L48S and R408W, were detected on more than 10 chromosomes (Fig. 1). Additionally, the three most common PAH gene mutations account for less than 50% of all mutant alleles. Homozygosity was observed in three patients only, which gives a frequency of homoallelic PKU genotypes of 8.82%. As expected, each of them was homozygous for one of the three most frequent mutations.

However, a more direct measure of the heterogeneity of PAH mutations was needed. The homozygosity value was therefore calculated on the basis of mutation frequencies \(j = \sum x_i^2\). For the population of Serbia, the homozygosity value, \(j\), is 0.10.

![Figure 1](image1.png)

Fig. 1. Number of different PAH mutations in relation to the number of mutant alleles detected in the Serbian population. 1 - fraction of PAH mutations detected on one chromosome (31%), 2 - fraction of PAH mutations detected on two chromosomes (26%), 3 - fraction of PAH mutations detected on three to 10 chromosomes (32%), 4 - fraction of PAH mutations detected on more than 10 chromosomes (11%).

![Figure 2](image2.png)

Fig. 2. Detection of EcoRI polymorphism by the PCR-RFLP method. Lanes: 1, 3 - heterozygote for EcoRI polymorphic site, 2 – 100-bp marker.
Polymorphism Analysis

In order to elucidate the origin of the most common PAH mutation in Serbia (L48S), we analyzed two polymorphic sites (EcoRI and XmnI) of 12 unrelated patients, compound heterozygotes with L48S being one of two mutations, as well as one homozygote for L48S.

Our research strategy was based on the finding that L48S was associated with four different haplotypes (3, 4, 16, and 28) in the populations studied so far (Table 1) (www.pahdb.mcgill.ca). Since the difference between these four haplotypes is in the polymorphic sites EcoRI, XmnI, and EcoRV, our research was focused on these polymorphisms. The exact position of the EcoRV polymorphic restriction site has not yet been determined. We therefore restricted our analysis to the XmnI and EcoRI polymorphisms. Using this experimental approach, we could distinguish between haplotypes 4, 16, and 28.

Our PCR-RFLP analysis of patients carrying the L48S mutation showed that all of them were heterozygous (+/-) for the EcoRI restriction site. The homozygote for L48S was also heterozygous for the EcoRI site (Fig. 2). Heterozygosity of the EcoRI site in this patient was confirmed by DNA sequencing (data not shown). The expected heterozygosity for the intron 5 EcoRI site therefore is 0.5.

Digestion with XmnI revealed three homozygotes (+/+), and nine heterozygotes (+/-). The homozygote for L48S was also homozygous XmnI (+/+). The expected heterozygosity for the intron 8 XmnI site is 0.46 (Fig. 3).

Based on high heterozygosity for the analyzed polymorphic sites, we can conclude that at least two different haplotypes associated with the L48S mutation exist in the Serbian population.
DISCUSSION

In this study, mutations in the PAH locus were used as markers for genetic variations in the Serbian population.

We analyzed a homozygosity value ($j$) because it indicates the extent of genetic variation at the PAH locus in a given population. The higher homozygosity value, the more homogeneous the population is with respect to PAH mutations (Guldberg et al., 1996).

The most homogeneous population described so far are Yemenite Jews, in whom a single molecular defect (deletion in the exon 3 of the PAH gene) is responsible for all the PKU cases in the population. Accordingly, the homozygosity value for this population is 1. Populations of Northeast and Eastern European countries (Latvia, Lithuania, Southern Poland) are also quite homogeneous. For the Serbian population, $j$ is 0.10. It is rather low, more similar to the situation in ethnically mixed populations (those of Germany and the United States) than to the case of a small isolated population (Yemenite Jews). It would be interesting to compare homogeneity of the Serbian population with that of other populations residing on the Balkan Peninsula. However, in the majority of Balkan populations, less than 90% ascertainment of mutations has been achieved. It does not give an idea of the level of heterogeneity, since a great number of mutations have not been characterized. The exception is Croatia, where the homozygosity value is 0.17, indicating that PKU is moderately homogeneous (Table 2).

Heterogeneity of the PAH gene in the Serbian population reflects numerous historically documented migrations and coexistence of different populations in this part of the Balkan Peninsula and confirms that gene flow occurred between populations.

The homozygosity value has to be considered when a PAH gene mutation detection strategy for a particular population is being created. Since the homozygosity value for the Serbian population is low, a diagnostic strategy has to be designed to identify a great number of mutations. We use denaturation gradient gel electrophoresis followed by sequencing analysis of all 13 exons of the PAH gene.

Table 1. Different PAH haplotype configurations associated with the L48S mutation. [Polymorphic restriction sites: BglII, PvuIIa, PvuIIb, EcoRI, MspI, XmnI, and EcoRV. +) Restriction site present; -) Restriction site absent].

<table>
<thead>
<tr>
<th>Haplotype (position)</th>
<th>BglII (intron 1)</th>
<th>PvuIIa (intron 2)</th>
<th>PvuIIb (intron 3)</th>
<th>EcoRI (intron 5)</th>
<th>MspI (intron 7)</th>
<th>XmnI (intron 8)</th>
<th>EcoRV (gene 3’ downstream)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>16</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2. Homozygosity at the PAH locus in different populations. $\Sigma x_i^2$ - homozygosity value; $\ast$ $\Sigma x_i^2$ calculated in this study.

<table>
<thead>
<tr>
<th>Population</th>
<th>$\Sigma x_i^2$</th>
<th>No. of chromosomes</th>
<th>Mutation detection rate (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yemenite Jews</td>
<td>1.00</td>
<td>44</td>
<td>100</td>
<td>Avigad et al., 1990</td>
</tr>
<tr>
<td>Latvia</td>
<td>0.58$\ast$</td>
<td>96</td>
<td>98</td>
<td>Pronina et al., 2003</td>
</tr>
<tr>
<td>Lithuania</td>
<td>0.54$\ast$</td>
<td>184</td>
<td>95</td>
<td>Kasnauskiene et al., 2003</td>
</tr>
<tr>
<td>Southern Poland</td>
<td>0.44</td>
<td>80</td>
<td>91.3</td>
<td>Zygulska et al., 1994</td>
</tr>
<tr>
<td>Romania</td>
<td>0.26$\ast$</td>
<td>44</td>
<td>88</td>
<td>Popescu et al., 1998</td>
</tr>
<tr>
<td>Croatia</td>
<td>0.17</td>
<td>78</td>
<td>99.0</td>
<td>Zschocke et al., 2003</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>0.14</td>
<td>242</td>
<td>99.6</td>
<td>Zschocke et al., 1995</td>
</tr>
<tr>
<td>Serbia</td>
<td>0.10</td>
<td>68</td>
<td>97.0</td>
<td>Stojiljković et al., 2006</td>
</tr>
<tr>
<td>Germany</td>
<td>0.08</td>
<td>90</td>
<td>95.6</td>
<td>Zschocke and Hoffmann., 2003</td>
</tr>
<tr>
<td>United States</td>
<td>0.06</td>
<td>294</td>
<td>94.9</td>
<td>Guldberg et al., 1996</td>
</tr>
</tbody>
</table>
PAH gene. The detection rate of 97% achieved in our previous study confirms that our diagnostic approach was well designed.

Polymorphic haplotypes at the PAH locus are used to study human evolution and the history of human populations. They are markers that can be used to follow migrations between different populations. In addition to this, PAH haplotypes can also be used to determine the origin of PAH mutations.

In a previous study of PAH gene mutations, we speculated that high frequency of the L48S mutation in the Serbian population is possibly attributable to mixing of South Slavic populations with autochthonous populations residing in the region, influenced by east-to-west Neolithic migrations. However, we could not exclude additional factors, such as the founder effect and/or genetic drift (Stojiljković et al., 2006). This hypothesis had to be tested by determining the haplotype of chromosomes affected by the L48S mutation.

A unique haplotype associated with L48S would suggest its independent (Serbian) origin. Otherwise, the association of L48S with different haplotypes would suggest that it was imported during migrations on the territory of Serbia.

Association of L48S with haplotypes 3, 4, 16, and 28 has been reported to date (www.pahdb.mcgill.ca). Our preliminary results exclude the possibility that the L48S mutation originates from Serbia and suggest that it was imported from populations with different genetic backgrounds.

Our preliminary results exclude the possibility that the L48S mutation originates from Serbia and suggest that it was imported from populations with different genetic backgrounds.

Family study of all patients with the L48S mutation will provide full and final information about haplotypes in the Serbian population and definitely elucidate its origin.

Acknowledgments – This work was supported by the Ministry of Science and Environment Protection, Republic of Serbia (Grant No.143051).

REFERENCES


Фенилкетонурија је урођена метаболичка болест проузрокована мутацијама у гену за фенилаланин хидроксилазу (PAH). У спремној популацији је идентификовано 19 различитих PAH мутација. PAH мутације коришћене су као молекуларни маркери за популационо-генетичко истраживање. Ниска вредност хомозиготности PAH гена 0,10 указује на хетерогеност фенилкетонурије у Србији и одржава бројне миграционе у регион у југоисточне Европе. У складу са тим, осмишљена је стратегија молекуларне дијагностике фенилкетонурије за Србију. У циљу расветлевања порекла најчешће мутације која узрокује фенилкетонурију у Србији, L48S, урађена је хаплотипска анализа за PCR-RFLP методом. Наши резултати су теретна да је L48S мутација пореклом из више популација са различитим генетичким карактеристикама.

1 Институт за молекуларну генетику и генетичко инжењерство, 11000 Београд, Србија
2 Институт за здравствену заштиту мајке и детета "Др Вукан Чупић", 11000 Београд, Србија