Identification of a single ancestral CYP1B1 mutation in Slovak Gypsies (Roms) affected with primary congenital glaucoma

Martina Plášílová, Ivaylo Stoilov, Mansoor Sarfarazi, Ludovit Kádasi, Eva Feráková, Vladimír Ferák

Department of Molecular Biology, Comenius University, Mlynská dolina B2, SK-842 15 Bratislava, Slovakia
M Plášílová
E Feráková
V Ferák

Molecular Ophthalmic Genetics Laboratory, Surgical Research Center, Department of Surgery, University of Connecticut Health Center, Farmington, CT, USA
I Stoilov
M Sarfarazi

Institute of Molecular Physiology and Genetics, Slovak Academy of Sciences, Bratislava, Slovakia
I Kádasi

Correspondence to: Dr Ferák.

Received 11 June 1998
Revised version accepted for publication 27 August 1998

Abstract
Primary congenital glaucoma (PCG) is an autosomal recessive eye disease that occurs at an unusually high frequency in the ethnic isolate of Roms (Gypsies) in Slovakia. Recently, we linked the disease in this population to the GLC3A locus on 2p21. At this locus, mutations in the cytochrome P4501B1 (CYP1B1) gene have been identified as a molecular basis for this condition. Here, we report the results of CYP1B1 mutation screening of 43 PCG patients from 26 Slovak Rom families. A homozygous G→A transition at nucleotide 1505 in the highly conserved region of exon 3 was detected in all families. This mutation results in the E387K substitution, which affects the conserved K helix region of the cytochrome P450 molecule. Determination of the CYP1B1 polymorphic background showed a common DNA haplotype in all patients, thus indicating that the E387K mutation in Roms has originated from a single ancestral mutational event. The Slovak Roms represent the first population in which PCG is found to result from a single mutation in the CYP1B1 gene, so that a founder effect is the most plausible explanation of its increased incidence. An ARMS-PCR assay has been developed for fast detection of this mutation, thus allowing direct DNA based prenatal diagnosis as well as gene carrier detection in this particular population. Screening of 158 healthy Roms identified 17 (10.8%) mutation carriers, indicating that the frequency of PCG in this population may be even higher than originally estimated.

Keywords: primary congenital glaucoma (PCG); cytochrome P4501B1; Roms (Gypsies); founder effect

The glaucomas are a heterogeneous group of eye diseases characterised by an optic neuropathy in which degeneration of retinal ganglion cells brings about a characteristic excavation of the head of the optic nerve. According to the WHO estimate, 5.2 million people in the world are blind as a consequence of glaucoma, so it is thus one of the leading causes of blindness world wide.

Success in gene mapping within the last few years made it possible to map, so far, seven loci responsible for different forms of glaucoma, as well as several other loci responsible for developmental anomalies associated with this condition. For two glaucoma loci, the defective molecule has been identified.10 11

Primary congenital glaucoma (PCG, infantile glaucoma, hydrophthalmia, or buphthalmos; gene symbol GLC3) is an autosomal recessive, severe form of glaucoma which manifests itself in the perinatal or early postnatal period. The disease is characterised by increased intraocular pressure, which, if uncontrolled, results in damage to the optic nerve and subsequent permanent loss of vision. The ocular hypertension is likely to be the result of improper development of the anterior eye segment interfering with the absorption of the aqueous humour.12

The estimated incidence of primary congenital glaucoma in western countries ranges from 1:500013 14 to 1:22 000,15 with most figures being close to 1:10 000.16 The disease is much more frequent in the Middle East (1:2500),17 but the highest incidence of 1:1250 has been reported for the population of Roms (Gypsies) in Slovakia.15 In this population, PCG represents the major cause of blindness.

By means of genetic linkage analysis, the PCG phenotype has so far been assigned to two different loci. Locus GLC3A has been mapped to 2p21 in Turkish families6 with subsequent confirmation in Saudi Arabian families,18 and locus GLC3B has been localised to the 1p36 region.7 In a sample of Rom families from Slovakia, we have recently shown that in this population the disease maps to the GLC3A locus without any apparent heterogeneity.19

Recently, three different truncating mutations in the human cytochrome P4501B1 (CYP1B1) gene were shown to segregate with the disease in five PCG families linked to the GLC3A locus,20 thus providing evidence that defects in this enzyme represent the molecular cause of the disease. To date, 17 different truncating or missense mutations in the CYP1B1 gene have been detected, 12 of which were in exon 3.18 20 21

Given its severity and unusually high incidence, PCG in Slovak Roms presents a serious health problem. The present work aims to identify, by direct sequencing of the CYP1B1 gene, the mutation(s) causing PCG in this population, which would make it possible to perform direct DNA based diagnosis of the disease as well as heterozygote detection and screening in this specific population.
CYP1B mutations in Roms with glaucoma

Figure 1 Diagram showing genomic structure of the CYP1B1 gene and position of primers used in this study in relation to the exons. The CYP1B1 gene consists of three exons (345, 1044, and 3707 bp in length). The coding regions are shaded. Mutation causing PCG in Slovak Roms is indicated by a vertical black arrow. Single nucleotide polymorphisms (SNPs) are indicated by vertical white arrows. Nucleotide position numbers correspond to the complete gene sequence DDBJ/EMBL/GenBank accession No. U56438. SNPs at nucleotide positions 3947, 4160, 8131, and 8195 result in amino acid polymorphisms. Horizontal arrows denote approximate positions of primers used in sequencing reactions, SSCP, and ARMS-PCR assays.

Materials and methods

SUBJECTS

Forty-three PCG patients belonging to 26 nuclear families were ascertained through various Departments of Ophthalmology and Genetics in Slovakia. Ten families were grouped together as they belonged to five larger kindreds, while the remaining families were unrelated to one another. Seven of these families were used previously for linkage studies, and the disease was shown to segregate with the GLC3A locus. DNA samples from 158 healthy, unrelated subjects with no known family history of PCG were also collected from different Rom communities. All subjects identified themselves as Roms and all provided their informed consent before inclusion in the study.

MUTATION SCREENING

The translated genomic regions of the CYP1B1 gene were PCR amplified in three different fragments using the following primer pairs. Fragment 1, 786 bp (1F 5'-TCTCCAAGAGTCAGCTCCG-3'; 1R 5'-GGGTTGTCGTCGTCGTC-3), fragment 2, 787 bp (2F 5'-ATGGCTTTCGGCCACTACT-3'; 2R 5'-GATCTTGGTTTTGAGGGGTG-3), and fragment 3, 885 bp (3F 5'-TCCCCAGAAAATTAATTTGACCTACTG-3'; 3R 5'-TGGAGAGTCAGCTCCG-3'; 1R 5'-GGGTCGTCGTCGTC-3'). The PCR reactions were carried out in a 40 µl volume containing 100 ng of genomic DNA, 1.0 µl of 20 mmol/l stock solution for each primer, 1 U AmpliTag™ polymerase (Perkin Elmer), 0.1 mmol/l of each dNTP, 50 mmol/l KCl, 10 mmol/l Tris-HCl (pH 9.3), and, for fragments 1 and 2, 1.5 mmol/l MgCl₂, and 10% DMSO, or, for fragment 3, 2.0 mmol/l MgCl₂, without DMSO. Cycling conditions were 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for one minute, for 35 cycles. PCR products were purified by QIAquick PCR purification kit (Qiagen). dRhodamine terminator cycle sequencing using AmpliTag™ DNA polymerase FS (Perkin Elmer) was carried out in accordance with the manufacturer’s protocol. The products of sequencing reactions were analysed on an ABI-377 automated DNA sequencer (Perkin Elmer).

SSCP ANALYSIS

Single strand conformational polymorphism (SSCP) analysis was carried out on a 249 bp long PCR fragment of exon 3, where no known polymorphic sites are present. PCR was performed in a total volume of 25 µl containing 100 ng of genomic DNA, 10 pmol of primers CYP4F (5'-TAAGAATTTTGTCTCATTGC-3') and 113R (5'- AATGTTGTTAGCCAAGACACG-3'), 1 U AmpliTag™ polymerase (Perkin Elmer), 0.1 mmol/l of each dNTP, 50 mmol/l KCl, 10 mmol/l Tris-HCl (pH 9.3), and 1.0 mmol/l MgCl₂. Cycling conditions were 94°C (30 seconds), 50°C (30 seconds), and 72°C (one minute), for 35 cycles. Denaturing buffer (25 µl) was added to each reaction and the resulting mix was heated for one minute at 94°C. Subsequently, 8 µl were loaded onto a 6% polyacrylamide gel (19:1 acrylamide:bisacrylamide, 10% urea, 5% glycerol). The gels were run at 15 W for 10 hours. The PCR products were silver stained using Promega protocol for Silver Sequence™ DNA Sequencing System. Gels were dried and photographed by Silver Sequence™ Automatic Processor Compatible (APC) Film.

ARMS-PCR ASSAY

We designed an ARMS-PCR assay to detect the E387K mutation which makes use of two allele specific PCR primers, differing from each other at the 3’ terminus (mutation site). The wild type forward primer (FW) for this assay was designed as 5’-TGTCCTGGCCTTCCTTTA-3’ and the forward mutant (FM) primer bearing an “A” instead of “G” at the very 3’ end. When used with a reverse primer R (5’-TCACTCTGTCGGTCAGG-3’) in a PCR reaction, a 245 bp product is obtained. After optimisation, PCR composed of 100 ng DNA, 10 pmol of FW and R (or FM and R) primers, 10 pmol of each primer for the D95938 marker (used as positive control), 200 µmol/l dNTPs, 1.5 mmol/l MgCl₂, 50 mmol/l KCl, 10 mmol/l Tris-HCl (pH 9.5), and 1 U Taq polymerase (Promega) was performed. The cycling conditions consisted of 94°C (40 seconds), 60°C (40 seconds), and 72°C (one minute).
Table 1 Polymorphic haplotypes composed of six SNP sites within the CYP1B1 gene and of two flanking STR markers on 40 chromosomes harbouring the E387K mutation and on 27 chromosomes without mutation. Polymorphic sites: 5'-D2S177 (STR) - nt 3793 - 3947 - 4160 - 8131 - 8184 - 8195 - D2S1346 (STR) - 3'. The asterisk denotes two haplotypes with unknown phase in one patient. Last column displays frequencies of intragenic haplotypes

<table>
<thead>
<tr>
<th>Chromosomes with mutation</th>
<th>Affected</th>
<th>Healthy</th>
<th>Intragenic haplotype %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No %</td>
<td>No %</td>
<td></td>
</tr>
<tr>
<td>5'-1 - T-G-T-C-C-A-1 -3'</td>
<td>30 75.0</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>1 - T-G-T-C-C-A-2</td>
<td></td>
<td>2 5.0</td>
<td></td>
</tr>
<tr>
<td>1 - T-G-T-C-C-A-4</td>
<td></td>
<td>1 2.5</td>
<td></td>
</tr>
<tr>
<td>3 - T-G-T-C-C-A-1</td>
<td></td>
<td>1 2.5</td>
<td></td>
</tr>
<tr>
<td>3 - T-G-T-C-C-A-3</td>
<td></td>
<td>2 5.0</td>
<td></td>
</tr>
<tr>
<td>1 or 4 - T-G-T-C-C-A-2 or 4*</td>
<td>2 5.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-G-T-C-C-A</td>
<td></td>
<td>2 5.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chromosomes without mutation</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3 - T-G-T-C-C-A-3</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>4 - T-G-T-C-C-A-5</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>4 - T-G-T-C-C-A-6</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>5 - T-G-T-C-C-A-3</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>5 - T-G-T-C-C-A-4</td>
<td>3</td>
<td>11.2</td>
</tr>
<tr>
<td>8 - T-G-T-C-C-A-3</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>9 - T-G-T-C-C-A-3</td>
<td>2</td>
<td>7.4</td>
</tr>
<tr>
<td>3 - C-C-G-C-C-A-6</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>4 - C-C-G-C-C-A-6</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>4 - C-C-G-C-C-A-7</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>9 - C-C-G-C-C-A-7</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>2 - C-C-G-C-G-5</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>3 - C-C-G-C-G-3</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>4 - C-C-G-C-G-3</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>5 - C-C-G-C-G-4</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>9 - C-C-G-C-G-5</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>4 - C-C-G-C-T-A-6</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>3 - C-C-G-C-T-A-8</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>6 - C-C-G-G-T-A-4</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>2 - C-C-G-G-T-A-4</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>6 - C-C-G-G-T-A-6</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>3 - C-C-G-G-T-A-6</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>3 - C-C-G-G-T-A-6</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>4 - C-C-G-G-T-A-5</td>
<td>1</td>
<td>3.7</td>
</tr>
<tr>
<td>Total</td>
<td>40 100</td>
<td>27 100</td>
</tr>
</tbody>
</table>

Results

MUTATIONS IN THE CYP1B1 GENE

The CYP1B1 genomic region spans >12 kb and contains three exons, 5,6,22,23 with exons 2 and 3 only coding for a protein of 543 amino acids (fig 1). Sequence analysis of the translated regions was carried out in 20 patients from 20 families. All patients were found to be homozygous for a single mutation, a G>A transition at nucleotide 1505 in exon 3 (fig 2), numbering corresponds cDNA sequence of the gene; DDBJ/EMBL/GenBank accession No U03688(22). This mutation brings about a substitution of glutamic acid for lysine at position 387 (E387K) of cytochrome P4501B1.

The segregation of this mutation in the families investigated was studied by means of SSCP analysis of the amplified exon 3, and was found to be consistent in all instances with autosomal recessive inheritance and full penetrance of the PCG phenotype (data not shown).

SINGLE NUCLEOTIDE POLYMORPHISMS IN THE CYP1B1 GENE

Six different diallelic polymorphic sites, one in intron 1 and five within exons 2 and 3 of the CYP1B1 gene, have previously been reported(21) (fig 1). We have determined these single nucleotide polymorphisms (SNPs) on 40 mutation bearing chromosomes and on another 27 chromosomes without any mutation. The results, summarised in table 1, show that, on chromosomes without any mutation, these six SNP sites are very variable and combine into five (out of 64 theoretically possible) different haplotypes. However, all chromosomes harbouring the E387K mutation carried only a single common haplotype, 5'-T (at position 3793) - G (3947) - T (4160) - C (8131) - C (8184) - A (8195) -3', which is also the most frequent one on chromosomes without mutation (40.8%).

To characterise the polymorphic background of the E387K mutation further, two flanking STR markers, D2S177 and D2S1346, were genotyped in affected patients and their parents when available (table 1). These markers (and CYP1B1) are located on a CEPH YAC clone 806-F8 approximately 1.7 Mb in size (http://carbon.wi.mit.edu:8000/cgi-bin/ contig/yac_info),24 but their precise distances from the gene are not known.

The haplotype 5'-1(D2S177)-T-G-T-C-C-A-1(D2S1346)-3' was found to account for 75% of chromosomes with the mutation, but it was not present on any of the 27 normal chromosomes. This indicates that this extended haplotype represents the ancestral polymorphic background for the E387K mutation in the population examined in this study. Analysis of 60 additional chromosomes from unrelated, healthy, homozygous Roms did not identify any allele 1 (0%) for D2S1346 and only three alleles 1 (5%) for D2S177. Therefore, this indicates a high degree of linkage disequilibrium between the E387K mutation and these two polymorphic loci, and strongly supports the view that the E387K mutation in Slovak Roms has originated from a single ancestral mutational event.

ARMS-PCR MUTATION SCREENING

Using the ARMS-PCR screening method (fig 3) described in the Materials and methods section, 158 healthy Roms from several communities were screened and 17 carriers of the E387K mutation were detected (10.8%). These results have been subsequently confirmed by direct sequencing of CYP1B1 exon 3. Besides that, homozygosity for the E387K mutation was found in all 23 additional patients whose CYP1B1 gene had not been sequenced. Analysis of control DNA samples from 50 healthy Slovaks (non-Roms) did not detect any instance of this particular mutation.
Discussion
Genček et al\textsuperscript{29} performed the first clinical and genetic study of PCG in the Slovak Roms and noticed that, in this population, the disease was clinically much more uniform, its course was more severe, and its prognosis worse than in the non-Rom patients. In Roms with PCG, the eyes are always bilaterally affected (while in the non-Roms this is the case in only $<75\%$), the intraocular pressure is almost always $>29$ mm Hg (while in non-Roms, it is almost always $<24$ mm Hg), the disease can be always diagnosed at birth or within the first months of life (while in non-Roms, only one third of cases were diagnosed at birth, and less than $90\%$ by the 12th month of life), and the response to surgical intervention as well as to other forms of treatment is much less favourable compared with that in non-Roms.

The finding of allelic homogeneity of PCG in Slovak Roms provides an explanation in molecular terms for the clinical uniformity of the disease in this population. Population studies performed by other research groups indicated that PCG is attributable to mutations in at least two different loci (GLC3A and GLC3B),\textsuperscript{17} and the former locus alone (GLC3A/CYP1B1) exhibits extensive allelic heterogeneity.\textsuperscript{18} 21 Stoilov et al\textsuperscript{21} reported a comprehensive CYP1B1 sequence analysis in 22 PCG families of German-American, Hispanic, French-Canadian, British, Pakistani, and Turkish origin. In total, they identified 16 different mutations, 11 of them in 16 Turkish families. Patients from 24 out of 25 Saudi Arabian families were shown to be either homozygous or compound heterozygous for three mutations, G61E (in 78% of the PCG chromosomes), R469W (10%), and D374N (6%).\textsuperscript{18} This allelic heterogeneity corresponds with the observed variable expressivity of the PCG phenotype in these populations.\textsuperscript{21} In contrast to our observation in this study, and those reported by Stoilov et al\textsuperscript{21} in Saudi Arabian families the disease exhibits incomplete penetrance (80%).

All 17 mutations identified so far were shown to bring about substitution of amino acids that are highly conserved across the whole P450 superfamily, or they cause truncation of the P4501B1 protein in such a way that at least the haem binding region is deleted from the C-terminal part of the molecule.\textsuperscript{18} 20 21 The E387K mutation affects the K helix, a structure that determines the correct folding and haem binding activity of the cytochrome P4501B1 molecule.\textsuperscript{28} The normal Glu-387 position is absolutely conserved among all members of the cytochrome P450 superfamily. Hence, mutation affecting this essential structure is likely to have a profoundly deleterious effect on the P4501B1 activity. This is why homozygosity for the E387K mutation is invariably associated with severe disease.

So far, the E387K mutation has been found twice among 28 PCG chromosomes examined,\textsuperscript{21} once in an American patient of Hispanic origin and once in a French-Canadian subject. However, in these two instances this mutation was carried on a different extended haplotype from that which is reported here.

The unfavourable prognosis of this form of PCG justifies prenatal diagnosis, and the unusually high frequency of the disease among Slovak Roms warrants carrier testing in this particular population. Our identification of a single prevalent mutation in Roms will allow direct DNA based prenatal diagnosis in at risk Rom families, and the ARMS-PCR assay will provide an easy and reliable carrier screening method for this population.

Sequence analysis of the coding region of the CYP1B1 gene in 150 randomly selected healthy subjects (50 Turkish,\textsuperscript{21} 50 British,\textsuperscript{21} and 50 Saudis,\textsuperscript{21} as previously\textsuperscript{27} failed to detect a carrier of any known mutation. Our finding of 17 carriers in a sample of 158 healthy Roms (10.8%), assuming Hardy-Weinberg equilibrium, would imply an allele frequency of 5.7% and a homozygote frequency of 0.3% (1:308), which is four times more than the estimated incidence of PCG in Slovak Roms by Genček et al.\textsuperscript{17} This shows that the E387K mutation is indeed widespread in the Rom ethnic group, and provides an explanation for the occurrence of several cases of parent-child ("pseudodominant") transmission of the disease reported previously.\textsuperscript{27} Stoilov et al\textsuperscript{21} reported the occurrence of a large kinship, containing sometimes $>30$ affected members in several generations.\textsuperscript{28} The difference between the original estimate of the mutation frequency\textsuperscript{15} and that reported here is significant ($p<0.01$), and may be the result of the randomly increased incidence of the mutation in the communities from which our sample originates. A study is now in progress aimed at identifying the local subpopulations within the Gypsy ethnic group exhibiting the highest incidence of the E387K mutation.

The plausible explanation of the remarkably high frequency of the E387K allele is a founder effect. Roms, whose total number in Slovakia is nowadays estimated to be $>450$ 000$^{10}$ arrived in central Europe during the late Middle Ages from their Indian homeland, which they had left at the turn of the millenium.\textsuperscript{31} Though no data exist on the size of the founder population, it probably did not exceed a few hundred people. As late as 1893, from when first reliable census data exist, the whole population of Roms in Slovakia was about 36 000.\textsuperscript{11}

The majority of Roms still preserve their language, traditions, and lifestyle, and their communities remain almost totally genetically isolated not only from the surrounding population but also from one another. Endogamy is a strict rule, consanguineous marriages are frequent (15-45%), and the inbreeding coefficient ranges among the highest worldwide.\textsuperscript{12} This typical Gypsy founder demography, history, and population structure is likely to be responsible also for the increased incidence of other autosomal recessive diseases in different European Rom communities.\textsuperscript{33-35}

The authors would like to thank the families of the patients for their cooperation in the project. Thanks are due to several Ophthalmologists and Clinical Geneticists, particularly Professor A Gerinec from the Clinic for Child Ophthalmology, University Hospital in Bratislava, Dr D Dunajčíková from the Ophthalm
32 Ferak V, Stolova D, Sieglova Z. Slovak Gypsies (Roms) - a population with the highest coefficient of inbreeding in Europe. Bratil Lok Listy 1987;87:168-75.