

Anna Degioanni<sup>a,b</sup>  
Pierre Darlu<sup>b</sup>

<sup>a</sup> Istituto di Genetica Biochimica ed  
Evoluzionistica, Pavia, Italy, and

<sup>b</sup> INSERM, Unité 155  
d'épidémiologie génétique,  
Université Denis-Diderot, Paris,  
France

## Analysis of the Molecular Variance at the Phenylalanine Hydroxylase (PAH) Locus

### Abstract

The frequencies of the different haplotypes identified at the phenylalanine hydroxylase (PAH) locus were analyzed for both phenylketonuria (PKU) and normal haplotypes of various European, Asiatic, Polynesian and Black American populations. These molecular variants were studied by applying a specific model of multivariate analysis of variance, allowing an estimation of components of variance at different levels of hierarchical subdivisions (intrapopulation, among different geographical groups of populations, and between PKU and normal haplotypes within populations). The results indicate that the PAH polymorphism could be appropriate to study divergence between African, European and Asiatic population groups, but is not sufficient to explain the diversity among European populations. However, the differences in PAH haplotype frequencies between PKU and normal haplotypes are statistically significant over all European populations.

### Key Words

Phenylalanine hydroxylase  
Phenylketonuria  
Human diversity  
Analysis of variance

### Introduction

Classical phenylketonuria (PKU) is a heterogeneous disorder caused by inborn errors in amino acid metabolism which determine irreversible mental retardation in children if left untreated [1]. Penrose [2] observed that it is a genetic disorder transmitted as an autosomal recessive trait. Neonatal screening of PKU, which has been systematically performed for several years in developed countries, led to a well-documented geographical

distribution of the frequency of PKU. Its prevalence ranges from  $4 \times 10^{-4}$  in Turkey [3] to  $8 \times 10^{-6}$  in Japan [4]. The frequency of PKU among Caucasians is approximately  $1 \times 10^{-4}$  [5], so that the frequency of the autosomal gene is about 1%, according to the Hardy-Weinberg equilibrium.

PKU is primarily caused by a deficiency of the hepatic enzyme phenylalanine hydroxylase (PAH). The recent isolation of PAH cDNA [6] led to the identification of the gene structure [7] and its location in the 12q22q24.1 region

[8]. Restriction enzymes reveal the presence of eight RFLPs associated with the PAH gene [9]. These can be used in prenatal diagnosis and genetic counseling in PKU families [9–11]. Geographical distribution of the different mutant haplotypes defined by these RFLP at the PAH locus have already been described in Europe, Asia and in Black Americans from Africa [see review in ref. 12]. Some of the mutant haplotypes are reported to decrease in frequency from east to west and from north to south (haplotype 2), whereas others seem present mostly in specific areas (e.g. haplotype 18 in China, haplotype 38 in France). The founder effect, migration and selective pressure are generally invoked to explain patterns in the geographical distribution of alleles.

However, these results are generally not statistically tested using a multivariate method, raising the question of the informativeness of the various PAH polymorphisms in localizing the geographical origins of mutations and describing the possible pattern of diffusion or selection.

The purpose of this article is to compare haplotype frequencies in different clusters of populations and, when possible, the distribution of the PKU and normal haplotypes within the same population, by performing multivariate analyses of the genetic diversity at the PAH locus.

## Materials and Methods

The eight restriction enzymes [*Bgl*II, *Eco*RI, *Eco*RV, *Hind*III, *Msp*I, *Pvu*II (two separate polymorphisms) and *Xmn*I] have led to the identification of 60 different haplotypes. They are labeled 1–52 (table 1) according to the classification of Eisensmith and Woo [13]. Other haplotypes are labeled differently [according to ref. 14–16].

We used published data for haplotype frequencies of PAH alleles in normal individuals and in PKU patients of various populations (table 2): Norway [17], Denmark [18], Sweden [14], Scotland [19], Poland [15,

**Table 1.** RFLP haplotypes at the human PAH locus

| Haplo-type | abcdefgh | Haplo-type | abcdefgh |
|------------|----------|------------|----------|
| 1          | 01001000 | 31         | 01010110 |
| 2          | 01001011 | 32         | 10001000 |
| 3          | 01010100 | 33         | 10001001 |
| 4          | 01010111 | 34         | 10011000 |
| 5          | 10111001 | 35         | 10011101 |
| 6          | 10111000 | 36         | 10101000 |
| 7          | 10010100 | 37         | 00001000 |
| 8          | 01011001 | 38         | 10101020 |
| 9          | 11011001 | 39         | 01111000 |
| 10         | 01011000 | 40         | 10000100 |
| 11         | 10011001 | 41         | 01011011 |
| 12         | 01001021 | 42         | 10010000 |
| 13         | 01101021 | 43         | 01001100 |
| 14         | 00011000 | 44         | 10010111 |
| 15         | 00011001 | 45         | 10010110 |
| 16         | 01010011 | 46         | 11000111 |
| 17         | 01010000 | 50         | 01011111 |
| 18         | 10110111 | 52         | 00010100 |
| 19         | 01010101 | P1         | 00101000 |
| 20         | 00111000 | P2         | 00001010 |
| 21         | 10111021 | P3         | 01000100 |
| 22         | 01011021 | P5         | 00020000 |
| 23         | 10110100 | SA         | 00001001 |
| 24         | 01001010 | SB         | 10101001 |
| 25         | 01001020 | SC         | 00101001 |
| 26         | 11001000 | SE         | 10001011 |
| 27         | 01001001 | SF         | 10111011 |
| 28         | 01000111 | SG         | 01010010 |
| 29         | 10000000 | SH         | 10101011 |
| 30         | 11000100 | BA         | 10102000 |

Absence, presence and variant of various restriction sites are coded 0, 1 and 2, respectively. a = *Bgl*II; b = *Pvu*II(a); c = *Pvu*II(b); d = *Eco*RI; e = *Msp*I (when coded 2, new *Msp* site); f = *Xmn*I; g = *Hind*III (when coded 2, 4.4-kb allele instead of 4.0-kb allele); h = *Eco*RV. [Listing of haplotypes 1–52 from ref. 13; P1–P5 from ref. 15; SA–SH are A–H alleles from ref. 14; BA is A from ref. 16].

**Table 2.** Geographical distribution of 60 haplotypes at the human PAH locus

| Haplotype | N  | n  | P  | p  | SC | sc | SW | sw | DK | dk | D  | d  | CH | ch | H  | h  | CZ | cz | F  | f  | I <sup>1</sup> | i <sup>2</sup> | TK | tk <sup>3</sup> | BU <sup>4</sup> | SI | si | J | j | py <sup>5</sup> | BL | bl |   |
|-----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----------------|----------------|----|-----------------|-----------------|----|----|---|---|-----------------|----|----|---|
| 1         | 18 | 24 | 14 | 39 | 10 | 10 | 20 | 41 | 12 | 23 | 21 | 22 | 19 | 15 | 5  | 12 | 0  | 3  | 23 | 18 | 27             | 18             | 18 | 8               | 8               | 0  | 1  | 1 | 0 | 135             | 2  | 4  |   |
| 2         | 9  | 6  | 78 | 10 | 3  | 1  | 31 | 10 | 13 | 3  | 23 | 8  | 4  | 2  | 22 | 6  | 15 | 7  | 13 | 4  | 4              | 0              | 1  | 5               | 19              | 1  | 0  | 0 | 0 | 0               | 0  | 0  |   |
| 3         | 15 | 3  | 3  | 2  | 6  | 2  | 19 | 3  | 25 | 2  | 10 | 3  | 2  | 0  | 2  | 3  | 0  | 1  | 7  | 2  | 2              | 2              | 1  | 1               | 0               | 1  | 1  | 0 | 0 | 5               | 0  | 0  |   |
| 4         | 8  | 23 | 16 | 26 | 2  | 4  | 37 | 25 | 9  | 21 | 17 | 14 | 7  | 9  | 4  | 6  | 5  | 3  | 6  | 10 | 6              | 15             | 15 | 26              | 5               | 30 | 27 | 8 | 7 | 366             | 1  | 0  |   |
| 5         | 5  | 12 | 5  | 16 | 1  | 3  | 9  | 10 | 0  | 7  | 3  | 12 | 0  | 4  | 1  | 0  | 0  | 1  | 0  | 6  | 3              | 6              | 2  | 7               | 1               | 0  | 0  | 0 | 0 | 0               | 0  | 2  |   |
| 6         | 0  | 2  | 6  | 1  | 0  | 1  | 2  | 1  | 2  | 0  | 1  | 2  | 2  | 0  | 0  | 0  | 0  | 0  | 2  | 1  | 12             | 1              | 25 | 1               | 11              | 0  | 1  | 0 | 1 | 5               | 0  | 0  |   |
| 7         | 17 | 7  | 3  | 9  | 0  | 3  | 5  | 9  | 1  | 7  | 2  | 7  | 0  | 4  | 0  | 2  | 0  | 0  | 3  | 8  | 0              | 3              | 0  | 8               | 0               | 2  | 0  | 2 | 1 | 100             | 0  | 1  |   |
| 8         | 5  | 0  | 0  | 0  | 1  | 3  | 0  | 0  | 0  | 1  | 1  | 1  | 1  | 0  | 2  | 4  | 0  | 1  | 0  | 0  | 3              | 2              | 0  | 1               | 0               | 0  | 0  | 0 | 0 | 0               | 3  | 2  |   |
| 9         | 1  | 2  | 1  | 0  | 1  | 0  | 1  | 1  | 1  | 0  | 4  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 5  | 2  | 1              | 0              | 1  | 1               | 0               | 0  | 0  | 0 | 0 | 0               | 1  | 1  |   |
| 10        | 0  | 1  | 0  | 1  | 0  | 0  | 1  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 1  | 0  | 0  | 0              | 0              | 0  | 0               | 4               | 0  | 1  | 0 | 0 | 0               | 0  | 0  |   |
| 11        | 1  | 2  | 0  | 1  | 0  | 0  | 2  | 5  | 1  | 1  | 0  | 2  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1              | 0              | 0  | 1               | 1               | 0  | 0  | 0 | 0 | 0               | 1  | 4  |   |
| 12        | 7  | 2  | 1  | 4  | 1  | 0  | 9  | 6  | 2  | 0  | 1  | 6  | 2  | 0  | 0  | 1  | 0  | 0  | 1  | 3  | 0              | 0              | 0  | 2               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 13        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 2               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 14        | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 15        | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 2               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 6 |
| 16        | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 4  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 17        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 1  | 0               | 0               | 0  | 1  | 0 | 0 | 0               | 0  | 0  | 0 |
| 18        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 1  | 3 | 0 | 0               | 0  | 0  | 0 |
| 19        | 0  | 0  | 0  | 1  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 3  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 20        | 0  | 0  | 0  | 1  | 2  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 1              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 3  | 0  | 0 |
| 21        | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 1              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 22        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 23        | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | 0  | 1  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 1  | 0 | 0 | 0               | 0  | 0  | 0 |
| 24        | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1              | 0              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 25        | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 26        | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 27        | 0  | 0  | 0  | 2  | 1  | 1  | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 4              | 1              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 28        | 0  | 0  | 1  | 4  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 3  | 0  | 0  | 1  | 0  | 1  | 0  | 1  | 0  | 0              | 2              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 29        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 30        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0              | 1              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 31        | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 2  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 2               | 0  | 0  | 0 |
| 32        | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 2  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 2  | 0 | 0 | 0               | 0  | 0  | 0 |
| 33        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 34        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 2  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 0  | 0 |
| 35        | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0 | 0 | 0               | 0  | 1  | 0 |

20], Hungary [11], Czechoslovakia [11], Switzerland [19], Germany [21, 22], France [23], Italy [24], China [25], Japan [25], US Black [16], Turkey [26], Bulgaria [27] and Polynesia [28].

The data were first analyzed by a classical principal component analysis (PCA) of contingency tables [29].

Then statistical analyses [AMOVA; ref. 30] were performed on the basis of a linear model defining different effects on individual haplotypes:

$$p_{jig} = p + a_g + b_{ig} + c_{jig},$$

where  $p_{jig}$  is the s-dimensional vector in the jth haplotype in the ith population in the gth group of populations. This s-dimensional Boolean vector is of the form  $\{r_1, r_2, \dots, r_s\}$  where  $r_s = 1$  if the haplotype is cut at the restriction site s, and 0 otherwise. The dimension is  $s = 8$  RFLP sites (see table 1). For example, the Boolean vector corresponding to haplotype 1 is the following string of 0 and 1: {01001000}.  $p$  is the unknown expectation of  $p_{jig}$  averaged over the whole study,  $a$  the group effect,  $b$  the population effect within groups and  $c$  the individual effect within populations. Effects are as-

**Table 2** (continued)

| Ha-<br>plo-<br>type | N  | n  | P   | p   | SC | sc | SW  | sw  | DK | dk | D  | d  | CH | ch | H  | h  | CZ | cz | F  | f  | I <sup>1</sup> | i <sup>2</sup> | TK | tk <sup>3</sup> | BU <sup>4</sup> | SI | si | J  | j  | py <sup>5</sup> | BL | bl |
|---------------------|----|----|-----|-----|----|----|-----|-----|----|----|----|----|----|----|----|----|----|----|----|----|----------------|----------------|----|-----------------|-----------------|----|----|----|----|-----------------|----|----|
| 36                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 2   | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 1  | 0  |
| 37                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 2              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| 38                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 6  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| 39                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| 40                  | 0  | 0  | 0   | 1   | 0  | 0  | 0   | 1   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| 41                  | 0  | 0  | 0   | 3   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 1  | 1              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| 42                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 2  | 2  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| 43                  | 0  | 0  | 0   | 1   | 0  | 0  | 0   | 1   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 2  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| 44                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 2   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| 45                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| 46                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| 50                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 2   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| 52                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 1   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 1  | 0  |
| P1                  | 0  | 0  | 0   | 1   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| P2                  | 0  | 0  | 0   | 1   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| P3                  | 0  | 0  | 0   | 2   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| P5                  | 0  | 0  | 1   | 0   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| SA                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 1   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| SB                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 1   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| SC                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 1   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| SE                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 1   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| SF                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 1   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| SG                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 1   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| SH                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 1   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 0  | 0  |
| BA                  | 0  | 0  | 0   | 0   | 0  | 0  | 0   | 0   | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0              | 0              | 0  | 0               | 0               | 0  | 0  | 0  | 0  | 0               | 5  | 0  |
|                     | 86 | 85 | 130 | 130 | 33 | 31 | 136 | 132 | 66 | 66 | 88 | 88 | 38 | 36 | 40 | 38 | 22 | 21 | 74 | 68 | 66             | 59             | 64 | 64              | 50              | 36 | 35 | 11 | 96 | 16              | 20 |    |

Upper case = PKU haplotypes; lower case = normal haplotypes; N = Norway; P = Poland; SC = Scotland; SW = Sweden; DK = Denmark; D = Germany; CH = Switzerland; H = Hungary; CZ = Czechoslovakia; F = France; I = Italy; TK = Turkey; BU = Bulgaria; SI = China; J = Japan; py = Polynesia; BL = Black Americans.

<sup>1</sup> 2 haplotypes unclassified.

<sup>2</sup> 4 haplotypes unclassified.

<sup>3</sup> 3 haplotypes unclassified.

<sup>4</sup> 38% of normal haplotypes are unclassified.

<sup>5</sup> 13 haplotypes unclassified.

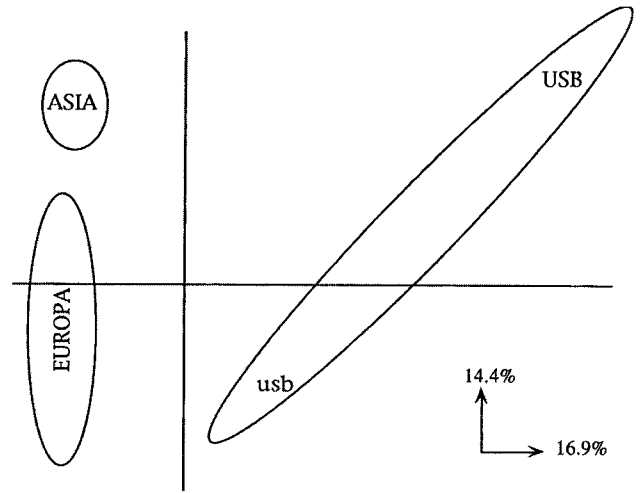
sumed to be uncorrelated, randomly distributed with variances  $s_a^2$ ,  $s_b^2$ ,  $s_c^2$  for effects  $a$ ,  $b$ , and  $c$  respectively, with  $s_1^2 = s_a^2 + s_b^2 + s_c^2$ .

To estimate these variances, the distance  $\delta_{jk}^2$  between two haplotypes  $j$  and  $k$  must first be defined [30].

In this study, we assume that RFLP sites are independent and equally informative. We also assume that the number of mutational steps separating two haplotypes is equal to the number of observed RFLP differences between them (for example haplotype 1 {01001000} and haplotype 3 {01010100} are separated

by 3 steps; see table 1). This simplification, which is not theoretically required by Excoffier's model, is essential since we cannot apply any evident probabilistic model to crossing-over events compared to mutation events. In other words, we assume that crossing-over producing a new haplotype from two common or rare haplotypes is an event which is less probable than the event producing the minimum number of steps required to obtain this new haplotype from another haplotype. For example, we assume that the distance between haplotypes 3 and 8 is more likely the result of

**Fig. 1.** PCA of the PAH haplotype frequencies on the whole samples. Percentages are the proportions of variance explained by the two first axes. USB and usb are PKU and normal haplotypes, respectively, of the US Black population. Both Asiatic (including the Polynesians) and European populations are clustered together.



three mutations than a single crossing-over of haplotypes 3 and 5 occurring between the restriction sites c and e (see table 1). This is a simplified model since at present we do not have accurate information on mutation rates characteristic of each restriction site or on the probability of each possible crossing-over between these restriction sites.

Under these conditions, the evolutionary distance  $\delta^2_{jk}$  between haplotype j and haplotype k is given by the Euclidean distance between the vectors  $p_{jig}$  and  $p_{kig}$ :

$$\delta^2_{jk} = \sum (p_{jig} - p_{kig})^2$$

where the sum is over all s restriction sites.

According to Li [31] and Excoffier et al. [30], the sum of squared differences  $\delta^2_{jk}$  between all pairs of N haplotypes can be broken down into: (a) the sum of squared differences within a population; (b) the sum of squared differences between populations within the same population group, and (c) the sum of squared differences between population groups. These different sums of squares lead to the estimation of  $s^2_a$ ,  $s^2_b$  and  $s^2_c$ . The significance of the different components is tested using a permutational approach, which is appropriate when the normality assumption is unwarranted, as is the case with molecular data [30].

As no information on the individual level of variability (within-individual haplotypic variability) is available, we do not take into consideration this level of variability. Consequently, instead of analyzing individuals, we applied Excoffier's method to the N haplotypes observed in various populations.

## Results

Table 2 shows the number of various normal and PKU haplotypes at the PAH locus observed in several populations.

The results given by PCA are plotted on figures 1 and 2. When PCA includes all populations, the frequencies of normal and PKU alleles of the US Black population account for the largest part (58.5%) of the variance on the first three axes. Clearly, this population is the most distant from the others, essentially because of the high frequencies of haplotypes 15, 35, 36 and BA (table 2, fig. 1), although normal and PKU frequencies of haplotypes, estimated on a very small sample, are very different within the US Black population. Another group of populations including China, Japan and Polynesia accounts for 43% of the variance on the second axis and 17% on the first three axes.

In an attempt to locate more accurately the various European populations in the space of the allele frequencies, a PCA was then performed excluding US Black, Asiatic and Polynesian populations. Since Bulgarian data are



**Table 3.** Hierarchical analysis of molecular variance at the PAH locus

|   | Variance | Percentage of total | p    |
|---|----------|---------------------|------|
| <b>All normal haplotypes (except US blacks)<sup>1</sup></b> |          |                     |      |
| Among groups  | 0.154    | 9.06                | 0.01 |
| Among populations within groups                             | 0.023    | 1.35                | 0.03 |
| Within populations  | 1.524    | 89.59               | 0.99 |
| <b>All normal European haplotypes<sup>2</sup></b>           |          |                     |      |
| Among groups  | 0.000    | 0.00                | 0.60 |
| Among populations within groups                             | 0.015    | 0.85                | 0.02 |
| Within populations  | 1.732    | 99.14               | 0.96 |
| <b>All PKU haplotypes (except US blacks)<sup>3</sup></b>    |          |                     |      |
| Among groups  | 0.176    | 10.79               | 0.01 |
| Among populations within groups                             | 0.051    | 3.14                | 0.01 |
| Within populations  | 1.400    | 86.07               | 0.99 |
| <b>All PKU European haplotypes<sup>4</sup></b>              |          |                     |      |
| Among groups  | 0.095    | 5.95                | 0.01 |
| Among populations within groups                             | 0.054    | 3.40                | 0.01 |
| Within populations  | 1.450    | 90.65               | 0.99 |
| <b>All populations (except US blacks)<sup>5</sup></b>       |          |                     |      |
| Among populations   | 0.079    | 4.67                | 0.01 |
| Among PKU and normals within population                     | 0.072    | 4.29                | 0.01 |
| Within populations  | 1.530    | 91.04               | 0.99 |
| <b>All European populations<sup>6</sup></b>                 |          |                     |      |
| Among populations   | 0.015    | Darlu 0.09          | 0.14 |
| Among PKU and normals within population                     | 0.078    | 4.71                | 0.01 |
| Within populations  | 1.572    | 94.39               | 0.99 |

<sup>1</sup> Six groups of European populations (Norway and Sweden; Scotland and France; Denmark and Germany; Poland and Czechoslovakia; Turkey and Hungary; Italy and Switzerland) and one Asiatic group (China, Japan and Polynesia).

<sup>2</sup> Six groups of European populations identified in footnote 1.

<sup>3</sup> Six groups of European populations (Norway and Sweden; Scotland and France; Denmark and Germany; Poland, Hungary and Czechoslovakia; Turkey and Bulgaria; Italy and Switzerland) and one Asiatic group (China and Japan). These analyses are significant only when Poland, Hungary and Czechoslovakia are included in the same group.

<sup>4</sup> Six groups of European populations identified in footnote 3. These analyses are significant only when Poland, Hungary and Czechoslovakia are included in the same group.

<sup>5</sup> Populations formed by normal and PKU haplotypes. Twelve European populations and two Asiatic populations (China and Japan).

<sup>6</sup> Populations formed by normal and PKU haplotypes. Twelve European groups.

of PAH haplotype 6 in PKU populations. However, it should be noted that the sample sizes are small and that haplotypes are not always typed for these two populations, so that their location on the diagram is still uncertain.

AMOVA was applied in different ways. First, seven population groups were formed, one including, as clearly suggested by the PCA (fig. 1), all Asian and Polynesian populations (China, Japan, Polynesia) and the other six each composed of two or three European populations. The US Black population was removed from the analysis. Analyses were separately performed for normal and PKU haplotypes.

Several distributions of European populations into the six groups were tested. All of them led to roughly the same result when Asiatic and Polynesian populations composed the seventh group (table 3): the variance among groups, which represents about 10% of the total variance, is significantly different from 0 at  $p < 0.01$ . This result holds for both normal and PKU alleles. Results of the following distribution are given in table 3: Norway and Sweden, Scotland and France, Denmark and Germany, Poland and Czechoslovakia, Turkey and Hungary, Italy and Switzerland, Polynesian, China and Japan.

When the Asiatic and Polynesian group is removed, the variance among the six European groups is no longer significantly different from 0 with normal alleles (table 3), however the European groups are distributed (two populations per group). The overall variance is almost totally explained by the within-population variance. Twenty different distributions of the twelve European populations into six groups were tested, all of them leading to the same results. These partitions were built either at random or on the basis of geographical proximity (or following PCA results). This last method would be expected to give the best

results according to the hypothesis of the gradient of haplotype frequencies in Europe. However, the only significant results ( $p < 0.01$ ) are observed with PKU alleles and when Czechoslovakia and Hungary are assigned to the same group (as given in table 3). In this case, both the variance among groups and the variance between populations within a group are significantly different from 0 even though explaining small parts of the whole variance (6.0 and 3.4%, respectively). When these two populations are not assigned to the same group, the variance among groups is no longer significantly different from 0.

AMOVA was also applied to different geographical clusters of populations, each of them including two different samples, normal and PKU haplotypes, of the same population. The variance between populations (table 3) is significantly different from 0 only when the Asiatic group (China, Japan and Polynesia) is considered, and not when only European populations are taken into account. Moreover, the variance between normal and PKU samples within populations is highly significant ( $p < 0.01$ ) although this variance represents less than 5% of the total variance.

## Discussion

Population differences for haplotype frequencies at the PAH locus are now well documented [12, 32, 33]. However, as far as we know, no inclusive statistical analysis of all available haplotypes has been performed. The PAH locus is highly polymorphic, since at least 71 different haplotypes have been listed recently [13]. A few haplotypes are found in all populations (haplotypes 1–4), whereas the majority are found in few, sometimes distantly related populations (table 2). Therefore, a specific statistical analysis is required to validate inferences as to the relationships be-



tween populations or a possible migratory flux between them.

Both PCA (giving heuristic descriptions of the data) and hierarchical analysis of variance (which tests different clustering hypotheses) lead to the same conclusions.

PKU and normal haplotypes clearly show a substantial divergence among Asian populations (including Polynesians), the US Black population and European populations. However this intercontinental divergence explains no more than 6–7% of the molecular variance at this locus, with the within-population variance accounting for about 90% of the total variance.

Looking at the European populations, hierarchically structured groups of populations cannot be established significantly either for normal or PKU haplotypes. The only significant results indicate that Czechoslovakia and Hungary (and Poland) share the same profile of PAH haplotypes in only the PKU population. One can also note that the European differences between the distributions of normal haplotypes and PKU haplotypes within the same population are very low (less than 5% of the whole variance) albeit significant ( $p < 0.01$ ). In other words, PKU haplotype frequencies and normal haplotype frequencies at the PAH locus diverge slightly systematically for all populations, as shown in figure 2. This can be explained by a higher frequency of haplotypes 2, 3 and 6, and lower frequencies of haplotype 1 and 5 in PKU than in normal samples, whatever the geographical origin of the population.

However, attempts to interpret, at least in part, the multifactorial analysis plotted on figure 2 in terms of geography failed. PKU populations of Denmark, France and Scotland are close together on the first two axes. Poland, Hungary and Czechoslovakia are clearly plotted into the same cluster only for PKU populations. This result is statistically demon-

strated by the variance analysis. No trend from east to west or from south to north, as suggested by several authors for some haplotypes [34], seems to be convincingly inferred when taking into account all haplotypes.

One can seriously argue that the sample size for each population is far from large enough to allow clear statistical conclusions, due to lack of power. If the haplotype frequencies are too inaccurate to use a methodology built on estimation of frequencies, another approach may be possible by only noting, for each population, the presence or absence of each haplotype whatever its frequency and then looking for a parsimonious tree of populations, following a method suggested by Mickevitch and Mitter [35]. Unfortunately, this method leads to totally unresolved trees, since more than 100 equally parsimonious trees are obtained with the same data.

Moreover, the model used to evaluate distances between haplotypes before performing variance analyses is probably too simple and needs to be improved by integrating information on mutation rates and crossing-over probabilities as soon as they become available, and by following, for instance, the suggestions of Templeton et al. [36].

As far as European populations are concerned, one can also argue that we did not try all possible distributions of twelve populations into various groups of various sizes, in order to obtain at least one with significant results. However, as explained, obvious and logical distributions based on geographical criteria or on PCA did not lead to significant results. In European populations, the proportion of variance within populations is clearly higher than both variance among groups and variance among populations within a group. This result holds even when one population group is very distant from the others (analysis with Japan, China and Polynesian as a supplementary group; table 3). Thus, there is no

chance of inverting this trend by chance (i.e. by just randomly distributing populations). Moreover, the advantage of CPA analysis is that it gives a clear distribution of populations in the space of haplotype frequencies. As can be seen, the structure shown by the CPA is weak so there is no chance that a better distribution can be found. Thus, if we obtain one distribution giving significant results by chance (about 50 among 1,000 random distributions for a probability level of 5% with no structured data) it will be explained by other factors than known geographical location or known historical events.

Obviously, the RFLP polymorphism at the PAH locus does not suggest any significant heterogeneity among European populations (except haplotype 2) and, consequently, cannot provide unequivocal inferences on the past evolution of European populations. To answer this question, we need more information about the mutations at the PAH gene. Moreover, different mechanisms can explain

the lack of geographical distribution of haplotype frequencies, such as the founder effect, historical variations in the effective population size, sampling bias, selective pressure or migration flux occurring differently both for different haplotypes and in various spatial directions. As a result, one can only point out that PKU and normal haplotypes show significantly different frequencies. On the other hand, this locus will provide useful information on the divergence between Asiatic, European and African populations as soon as the present lack of extensive data on Africa and Asia is acquired. For inferring the evolution of human populations using polymorphism at the PAH locus, future work will no doubt be increasingly fruitful.

### Acknowledgement

We wish to thank L. Excoffier for providing us helpful comments and for making available to us his computer program AMOVA.

---

### References

- 1 Fölling A: Über Ausscheidung von Phenylbrenztraubensäure in den Harn als Stoffwechsellanomalie in Verbindung mit Imbezillität. *Z Physiol Chem* 1934;227:169-176.
- 2 Penrose LS: Inheritance of phenylpyruvic amentia (phenylketonuria). *Lancet* 1935;iii:192-194.
- 3 Özalp I, Coskun T, Ceyhan M, Tokol S, Oran O, Erdem G, Tekinalp G, Durmus Z, Tarikahya Y: Incidence of phenylketonuria and hyperphenylalaninaemia in a sample of the Turkish newborn population. *J Inherit Metab Dis* 1986;9:237-239.
- 4 Aoki K, Wada Y: Outcome of the patients detected by newborn screening in Japan. *Acta Paediatr* 1988;30:429-434.
- 5 Bickel H, Bachmann C, Beckers R: Neonatal mass screening for metabolic disorders. *Eur J Pediatr* 1981; 137:133-139.
- 6 Kwok SCM, Ledley FD, DiLella AG, Robson KJH, Woo SLC: Nucleotide sequence of a full-length complementary DNA clone and amino acid sequence of human phenylalanine hydroxylase. *Biochemistry* 1985;24:556-561.
- 7 DiLella AG, Kwok SCM, Ledley FK, Marvit J, Woo SLC: Molecular structure and polymorphic map of the human phenylalanine hydroxylase gene. *Biochemistry* 1986;25: 743-749.
- 8 Lidsky AS, Law ML, Morse HG, Kao FT, Rabin M, Ruddle FH, Woo SLC: Regional mapping of the phenylalanine hydroxylase gene and the phenylketonuria locus in the human genome. *Proc Natl Acad Sci USA* 1985;82:6221-6225.
- 9 Lidsky AS, Ledley FD, DiLella AG, Kwok SCM, Daiger SP, Robson KJH, Woo SLC: Extensive restriction site polymorphism at the human phenylalanine hydroxylase locus and application in prenatal diagnosis of phenylketonuria. *Am J Hum Genet* 1985;37:619-634.
- 10 Woo SLC, Lidsky AS, Güttler F, Chandra T, Robson KJH: Cloned human phenylalanine hydroxylase gene allows prenatal detection of classical phenylketonuria. *Nature* 1983;306:151-155.

- 11 Daiger SP, Chakraborty R, Reed L, Fekete G, Schuler D, Berencsi G, Nasz I, Brdicka R, Kamaryt J, Pijackava A, Moore S, Sullivan S, Woo SLC: Polymorphic DNA haplotypes at the phenylalanine hydroxylase (PAH) locus in European families with phenylketonuria (PKU). *Am J Hum Genet* 1989;45:310-318.
- 12 Konecki DS, Lichter-Konecki U: The phenylketonuria locus: Current knowledge about alleles and mutations of the phenylalanine hydroxylase gene in various populations. *Hum Genet* 1991;87:377-388.
- 13 Eisensmith RC, Woo SLC: Updated listing of haplotypes at the human phenylalanine hydroxylase (PAH) locus. *Am J Hum Genet* 1992;51:1445-1448.
- 14 Svensson E, Döbeln U von, Hagenfeldt L: Polymorphic DNA haplotypes at the phenylalanine hydroxylase locus and their relation to phenotype in Swedish phenylketonuria families. *Hum Genet* 1991;87:11-17.
- 15 Zygulska M, Eigel A, Aulehla-Scholz C, Pietrzyk JJ, Horst J: Molecular analysis of PKU haplotypes in the population of southern Poland. *Hum Genet* 1991;86:292-294.
- 16 Hofman KJ, Steel G, Kazazian HH, Valle D: Phenylketonuria in US Blacks: Molecular analysis at the phenylalanine hydroxylase gene. *Am J Hum Genet* 1991;48:791-798.
- 17 Apold J, Eiken HG, Odland E, Fredriksen A, Bakken A, Lorens JB, Boman H: A termination prevalent in Norwegian haplotype no. 7 PKU genes. *Am J Hum Genet* 1990;47:1002-1007.
- 18 Chakraborty R, Lidsky AS, Daiger SP, Güttler F, Sullivan S, DiLella AG, Woo SLC: Polymorphic DNA haplotypes at the human phenylalanine hydroxylase locus and their relationship with phenylketonuria. *Hum Genet* 1987;76:40-46.
- 19 Sullivan SE, Moore SD, Connor JM, King M, Cockburn F, Steinmann B, Gitzelmann R, Daiger SP, Woo SLC: Haplotype distribution of the human phenylalanine hydroxylase locus in Scotland and Switzerland. *Am J Hum Genet* 1989;44:652-659.
- 20 Jaruzelska J, Henriksen KF, Güttler F, Riess O, Borski K, Blin N, Slomski R: The codon 408 mutation associated with haplotype 2 is predominant in Polish families with phenylketonuria. *Hum Genet* 1991;86:247-250.
- 21 Aulehla-Scholz C, Vorgerd M, Sautter E, Leupold D, Mahlmann R, Ullrich K, Olek K, et al: Phenylketonuria: Distribution of DNA diagnostic patterns in German families. *Hum Genet* 1988;78:353-355.
- 22 Lichter-Konecki U, Schlotter M, Konecki DS, Labeit S, Woo SLC, Trefz FK: Linkage disequilibrium between mutation and RFLP haplotype at the phenylalanine hydroxylase locus in the German population. *Hum Genet* 1988;78:347-352.
- 23 Rey F, Berthelon M, Caillaud C, Lyonnet S, Abadie V, Blandin-Savaja F, Feingold J, Saudubray J, Frézal J, Munnich A, Rey J: Clinical and molecular heterogeneity of phenylalanine hydroxylase deficiencies in France. *Am J Hum Genet* 1988;43:914-921.
- 24 Dianzani I, Devoto M, Camaschella C, Saglio G, Ferrera GB, Cerone R, Romano C, Romeo G, Giovannini M, Riva E, Angemeydt F, Trefz FK, Okano Y, Woo SLC: Haplotype distribution and molecular defects at the phenylalanine hydroxylase locus in Italy. *Hum Genet* 1990;86:69-72.
- 25 Daiger SP, Reed L, Huang SS, Zeng YT, Wang T, Lo WHY, Okano Y, Hase Y, Fukuda Y, Oura T, Tada K, Woo SLC: Polymorphic DNA haplotypes at the phenylalanine hydroxylase (PAH) locus in Asian families with phenylketonuria (PKU). *Am J Hum Genet* 1989;45:319-324.
- 26 Lichter-Konecki U, Schlotter M, Yaylak C, Ozguc M, Coskun T, Ozalp I, Wendel U, Batzler U, Trefz FK, Konecki DS: DNA haplotype analysis at the phenylalanine hydroxylase locus in the Turkish population. *Hum Genet* 1989; 81:373-376.
- 27 Kalaydjieva L, Dworniczak B, Aulehla-Scholz C, Kremensky I, Bronzova J, Eigel A, Horst J: Classical phenylketonuria in Bulgaria: RFLP haplotypes and frequency of the major mutations. *J Med Genet* 1990; 27:742-745.
- 28 Hertzberg M, Jahromi K, Ferguson V, Dahl HHM, Mercer J, Mickleston DNP, Trent RJ: Phenylalanine hydroxylase gene haplotypes in Polynesians: Evolutionary origins and absence of alleles associated with severe phenylketonuria. *Am J Hum Genet* 1989;44:382-387.
- 29 Benzecri JP: L'analyse des données: Paris, Dunod, 1973.
- 30 Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 1992;131:479-491.
- 31 Li CC: Population Genetics. Pacific Grove, Boxwood, 1976.
- 32 Woo LC: Molecular basis and population genetics of phenylketonuria. *Biochemistry* 1989;28:1-7.
- 33 Scriver CR, John SMW, Rozen R, Eisensmith R: Associations between populations, phenylketonuria mutations and RFLP haplotypes at the phenylalanine hydroxylase locus: An overview. *Dev Brain Dysfunct* 1993;6:11-25.
- 34 Eisensmith RC, Okano Y, Dasovich M, Wang T, Güttler F, Lou H, Guldberg P, Lichter-Konecki U, Konecki DS, Svensson E, Hagenfeldt L, Rey F, Munnich A, Lyonnet S, Cockburn F, Connor JM, Pembrey ME, Smith I, Gitzelmann R, Steinmann B, Apold J, Eiken HG, Giovannini M, Riva E, Longhi R, Romano C, Cerone R, Naughten ER, Mullins C, Cahalane S, Ozalp I, Fekete G, Schuler D, Berencsi GY, Nasz I, Brdicka R, Kamaryt J, Pijackova A, Cabalska B, Boszkowa K, Schwartz E, Kalinin VN, Jin L, Chakraborty R, Woo SLC: Multiple origins for phenylketonuria in Europe. *Am J Hum Genet* 1992;51:1355-1365.
- 35 Mickevich MF, Mitter C: Treating polymorphic characters in systematics: A phylogenetic treatment of electrophoretic data; in Funk VA, Brooks DR (eds): *Advances in Cladistics*. New York, The New York Botanical Garden, 1981, pp 45-58.
- 36 Templeton AR, Crandall KA, Sing CF: A cladistic analysis of phenotypic associations with haplotypes and inferred from restriction endonuclease mapping and DNA sequence data. III: Cladogram estimation. *Genetics* 1992;132:619-633.