Journal of Human Genetics (2014) 59, 145–152

www.nature.com/jhg

ORIGINAL ARTICLE

# The mutation spectrum of the *phenylalanine hydroxylase* (*PAH*) gene and associated haplotypes reveal ethnic heterogeneity in the Taiwanese population

Ying Liang[1], Miao-Zeng Huang[2], Cheng-Yi Cheng[2], Hung-Kun Chao[2], Victor Tramjay Fwu[2], Szu-Hui Chiang[2], Kwang-Jen Hsiao[2], Dau-Ming Niu[3] and Tsung-Sheng Su[1,2]

**Phenylalanine hydroxylase (PAH) deficiency is responsible for most cases of phenylketonuria (PKU). In this study of the *PAH* mutation spectrum in the Taiwanese population, 139 alleles were identified including 34 different mutations. The V190G, Q267R and F392I mutations are first reported in this study. The most common mutations, R241C, R408Q and Ex6-96A>G, account for 23.2%, 12.0% and 9.2%, of the mutant alleles, respectively. Haplotype analysis shows that R241C and Ex6-96A>G are exclusively associated with haplotype 4.3 to suggest founder effects. On the other hand, R408Q is found on two distinct haplotypes suggesting recurrent mutations. The spectrum of *PAH* mutations in Taiwan shows various links to those of other Asian regions, yet remarkable differences exist. Notably, R408Q, E286K and −4173_−407del, accounting for 21% of all mutant alleles in Taiwan, are very rare or are undetected among PKU cohorts of other Asian regions to suggest local founder effects. Moreover, the low homozygosity value of 0.092 hints at a high degree of ethnic heterogeneity within the Taiwanese population. Our study of *PAH* mutation spectrum and the associated haplotypes is useful for subsequent study on the origin and migration pattern via Taiwan, an island at the historical crossroad of migration of ancient populations.**
*Journal of Human Genetics* (2014) **59,** 145–152; doi:10.1038/jhg.2013.136; published online 9 January 2014

**Keywords:** haplotype; mutation spectrum; phenylalanine hydroxylase; phenylketonuria; STR; Taiwan; VNTR

## INTRODUCTION

Phenylketonuria (PKU; MIM#261600) is a genetic disorder resulting from the deficiency of phenylalanine hydroxylase (PAH). To date, more than 500 *PAH* mutations have been reported (online PAH database, http://www.pahdb.mcgill.ca/). These mutations vary in their impact on PAH enzymatic activity that lead to a range of clinical phenotypes from severe PKU to mild hyperphenylalaninemia (MHP). The spectrum of *PAH* mutations has been studied worldwide and a broad correlation of clinical phenotypes with mutational genotypes has been found.[1] Moreover, various types of polymorphic makers in the *PAH* locus have been reported including single-nucleotide polymorphism (SNP), variable number of tandem repeats (VNTRs) motif and a short tandem repeats (STRs) polymorphism (online PAH database, http://www.pahdb.mcgill.ca/). A prevalent *PAH* mutation in an area is generally shown to be the founding mutation that strongly associates with a particular polymorphic haplotype,[2] which has been used to understand origins and patterns of population migrations.[3–5] Moreover, the number of different *PAH* mutations identified have been shown to reflect the ethnic complexity in a population.[6]

Present-day Taiwanese residents came from multiple origins.[7,8] In addition to Austronesian aboriginals who settled in Taiwan dating back to many thousands of years ago, there were also descendants from southeast coast of China in the last few centuries, and the more recent immigrants from mainland China after the Second World War. Other ethnic groups, such as Spanish, Dutch and Japanese, have resided for short periods of time in the last 400 years. Chien *et al.*[9] have reported *PAH* mutations in 25 PAH-deficient families in Taiwan. They identified 18 different *PAH* mutations amongst which the R241C mutation accounted for 32% of the mutant chromosomes, and was proposed to be a founder mutation in the Taiwanese population.

To improve the representation of *PAH* mutations in Taiwan, we expanded the sample size to 71 unrelated PKU families in this study. In addition to the mutation spectrum, polymorphic haplotypes associated with prevalent *PAH* mutations were examined.

[1]Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei, Taiwan; [2]Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan and [3]Department of Pediatrics, Taipei Veterans General Hospital, Taipei, Taiwan
Correspondence: Dr D-M Niu, Department of Pediatrics, Taipei Veterans General Hospital, Taipei 112, Taiwan.
E-mail: dmniu@vghtpe.gov.tw
or Dr T-S Su, Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei 112, Taiwan.
E-mail: su31659@gmail.com
Received 1 October 2013; revised 29 November 2013; accepted 5 December 2013; published online 9 January 2014

## MATERIALS AND METHODS

### Patients

Seventy-one PAH-deficient families in Taiwan were included. Majority (63 families) of the subjects were identified through neonatal screening. The remaining subjects were from schools of special education (4 families) and from outpatient clinics (4 families). $BH_4$ deficiency has been ruled out in these patients. No consanguinity was known among PKU families. Informed consent for genetic analysis was obtained according to the regulation of Taipei Veterans General Hospital.

### Mutation analysis

Genomic DNA was extracted from leukocytes by a described method.[10] The *PAH* genomic sequences including promoter, 13 exons and their flanking exon–intron junctions were PCR amplified, and the sequence was determined using the Big Dye Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA, USA). Moreover, differential PCR was performed to detect mutation −4173_−407del, a large deletion occurred in the *PAH* enhancer.[11] When sequence alteration was identified, paternal or maternal origin was determined whenever the parental DNA was available.

### Analysis of VNTR, STR, SNP and RFLP haplotype

Structure of VNTR motif was analyzed by a PCR-based method.[12] The number of repeats was deduced through size determination of the PCR products by agarose gel electrophoresis using PCR products of known repeat number as a reference. To determine VNTR cassette sequences, the PCR product was sequenced directly if the sample was homozygous for the VNTR genotypes; if the samples were of heterozygote types, allelic PCR products were first separated and purified by agarose gel electrophoresis and gel band purification kit. The STR number variation was determined by PCR-based microsatellite genotyping with the fluorescent dye FAM attached to the 5′-end of the reverse primer.[13,14] PCR products were analyzed using the ABI PRISM 377 automated DNA sequencer (Applied Biosystems). STR genotypes were defined and edited using the Applied Biosystems GeneScan with PCR products of known repeat number as references. To analyze diallelic restriction fragment length polymorphism (RFLP), in addition to the Southern blot method,[10] the *Bgl*II, *Pvu*II(a), *Pvu*II(b), *Msp*I, *Xmn*I and *Eco*RI restriction enzyme sites of DNA samples were analyzed by the PCR-based method using primers described.[3] The PCR products were sequenced directly to determine the status of enzyme sites except *Eco*RI recognition sequence, which was determined by *Eco*RI digestion. Moreover, during the *PAH* mutation screening, intragenic SNPs including four in introns, IVS2 + 19T/C, IVS3 −22T/C, IVS4 + 47C/T and IVS10 + 97A/G, and three in exons, c.696G/A (Q232Q), c.735A/G (V245V) and c.1155G/C (L385L), were recorded accordingly.

### PAH activity analysis

An expression plasmid carrying the *PAH* cDNA (complementary DNA) sequence under the transcriptional control of SV40 promoter was used for PAH activity assay. Mutation was introduced into the *PAH* sequence by the QuikChange site-directed mutagenesis (Stratagene, La Jolla, CA, USA). All constructs were confirmed by DNA sequencing. The expression plasmids were individually transfected into COS-1 cells by the calcium phosphate method.[15] PAH activity was assayed in crude extracts of the transfected cells by measuring the conversion of L-[$^{14}$C]phenylalanine to L-[$^{14}$C]tyrosine using either natural cofactor $BH_4$ (6*R*-5,6,7,8-tetrahydrobiopterin, Schircks, Jona, Switzerland) or synthetic cofactor $6MPH_4$ (6-methyl-5,6,7,8-tetrahydropterin, Sigma, St Louis, MO, USA).[16] Residual PAH enzymatic activity of a particular allele was determined from at least three sets of lysates of transfected COS-1 cells using independently prepared plasmid DNAs.

### Analysis of RNA splicing efficiency

To study the effects of IVS7 + 5G>A base substitution on RNA splicing, modified cDNA expression plasmid carrying the 1058-bp *PAH* intron 7 was constructed. To this end, *PAH* genomic DNA fragment comprising intron 6 to exon 8 was generated by PCR amplification using the forward primer, 5′-CTCCTAGTGCCTCTGACTCA-3′, located in intron 6, and reverse primer, 5′-TCCTTACCTGGGAAAACTGGG-3′, located at the junction of exon 8 and intron 8. Subsequently, the DNA fragment containing intron 7 and the flanking exonic sequences was obtained by co-digesting the PCR product with restriction enzymes *Bam*HI and *Afl*III. The fragment obtained was used to replace the corresponding *Bam*HI/*Afl*III fragment on *PAH* cDNA to generate a modified expression plasmid carrying *PAH* cDNA and intron 7 of the *PAH* gene. Northern blot analysis of *PAH* RNA was performed on RNA from such plasmid-transfected COS-1 cells.

### Phenotype determination

Patients were assigned to three phenotype categories mainly according to the plasma phenylalanine (Phe) levels before clinical treatment: classical PKU (Phe $\geqslant$ 1200 μmol l$^{-1}$), mild PKU (Phe = 600–1199 μmol l$^{-1}$) and MHP (Phe = 240–599 μmol l$^{-1}$).[6] Data on dietary phenylalanine tolerance were used as references when available.[1]

## RESULTS AND DISCUSSION

### Identification of *PAH* mutations

The *PAH* DNA sequences including enhancer, promoter, 13 exons and flanking exon–intron junctions were determined by sequencing of PCR-amplified genomic DNAs from probands of 71 unrelated PKU families. Mutations were identified on both alleles in 68 families and on only a single allele in 3 families. Thus, mutations were determined in 139 of 142 *PAH* mutant chromosomes, a detection rate of 98%. The mutations represent 34 different types, including 20 missense mutations, 4 nonsense mutations, 4 deletion/insertion within structural gene, 1 deletion in enhancer region and 5 affecting splicing (Table 1). The most prevalent mutations in Taiwan are R241C, R408Q and Ex6-96A>G accounting for 23.2%, 12.0% and 9.2% of the 142 mutant chromosomes, respectively. The 17 mutations were found only once.

### Functional analysis of novel *PAH* mutations

Three previously unreported alterations, V190G, Q267R and F392I, were found in this study in the Taiwanese population. To access functional significance associated with such changes, residual PAH activity from COS-1 cells transfected with the respective *PAH* cDNA expression plasmid was examined (Figure 1). The results showed that PAH activity expressed from *PAH* sequences harboring V190G or Q267R had lower than 2% of the wild-type enzymatic activity level, indicating that these alterations could result in a severe phenotype. Indeed, patients harboring the V190G mutation in functionally hemizygous state (V190G/E228X) displayed classical PKU phenotype (Table 2, genotype 6). Likewise, patient carrying compound heterozygous mutation of Q267R/G257V manifested classical PKU phenotype (Table 2, genotype 20). On the other hand, PAH expressed from the F392I allele showed about half the wild-type activity when assayed with the synthetic cofactor $6MPH_4$ (Figure 1a, right panel) whereas reaching the wild-type level when assayed with natural cofactor $BH_4$ (Figure 1a, left panel).

In addition to being a coding sequence, exon has been shown to harbor *cis*-regulatory elements important for exon recognition during pre-mRNA splicing.[17] In fact, we have shown that about one-quarter of naturally occurring mutations on the human *PAH* exon 9 affected RNA processing.[16] Likewise, we found that the silent mutation V399V (c.1197A>T) affects exon-splicing enhancer element, resulting in exon 11 skipping during pre-mRNA splicing.[18] To assess whether alteration of F392I (c.1174T>A) affects *PAH* exon 11 recognition, a study similar to that of V399V (c.1197A>T) was carried out. No significant effects of such an alteration on pre-mRNA splicing were found (data not shown). Clinical manifestation of patients carrying the F392I allele was MHP (Table 2, genotype 40, F392I/R408Q and

**Table 1 PAH mutations identified in Taiwan and relative frequencies**

| *Mutation name*[a] | *Systematic name*[b] | *Location* | *Mutation type* | *Frequency (actual no. of chromosomes)* | *Phenotype assigned* |
|---|---|---|---|---|---|
| −4173_−407del | c.−4173_−407del 3767 | Enhancer | Deletion | 3.5 (5) | Classical PKU |
| R53H | c.158G>A | Exon 2 | Missense | 2.1 (3) | MHP |
| S70del | c.208_210delTCT | Exon 3 | Deletion | 0.7 (1) | Uncertain |
| E76G | c.227A>G | Exon 3 | Missense | 0.7 (1) | MHP |
| T81_R86>VfsX6 | c.241_256del16 | Exon 3 | Deletion | 0.7 (1) | Classical PKU |
| R111X | c.331C>T | Exon 3 | Nonsense | 3.5 (5) | Classical PKU |
| IVS4−1G>A | c.442−1G>A | Intron 4 | Splicing | 0.7 (1) | Classical PKU |
| R155H | c.464G>A | Exon 5 | Missense | 0.7 (1) | MHP |
| R155fsX43 | c.463_464insTGTGTACC c.509+15_+18del | Exon 5 | Insertion/deletion | 0.7 (1) | Classical PKU |
| V190G | c.569T>G | Exon 6 | Missense | 0.7 (1) | Classical PKU |
| EX6-96A>G | c.611A>G | Exon 6 | Splicing | 9.2 (13) | Classical PKU |
| E228X | c.682G>T | Exon 6 | Nonsense | 0.7 (1) | Classical PKU |
| R241C | c.721C>T | Exon 7 | Missense | 23.2 (33) | Mild |
| R243Q | c.728G>A | Exon 7 | Missense | 4.2 (6) | Classical PKU |
| R252Q | c.755G>A | Exon 7 | Missense | 1.4 (2) | Classical PKU |
| G257V | c.770G>T | Exon 7 | Missense | 2.1 (3) | Classical PKU |
| R261Q | c.782G>A | Exon 7 | Missense | 0.7 (1) | Uncertain |
| R261X | c.781C>T | Exon 7 | Nonsense | 0.7 (1) | Classical PKU |
| Q267R | c.800A>G | Exon 7 | Missense | 0.7 (1) | Classical PKU |
| IVS7+5G>A | c.842+5G>A | Intron 7 | Splicing | 0.7 (1) | Classical PKU |
| E286K | c.856G>A | Exon 8 | Missense | 5.6 (8) | Classical PKU |
| R297H | c.890G>A | Exon 8 | Missense | 0.7 (1) | Uncertain |
| F302fsX39 | c.904delT | Exon 8 | Deletion | 0.7 (1) | Classical PKU |
| IVS8−7A>G | c.913−7A>G | Intron 8 | Splicing | 1.4 (2) | Classical PKU |
| G312V | c.935G>T | Exon 9 | Missense | 2.1 (3) | Classical PKU |
| P314T | c.940C>A | Exon 9 | Missense | 0.7 (1) | MHP |
| Y356X | c.1068C>G | Exon 11 | Nonsense | 0.7 (1) | Classical PKU |
| V388M | c.1162G>A | Exon 11 | Missense | 0.7 (1) | Uncertain |
| F392I | c.1174T>A | Exon 11 | Missense | 1.4 (2) | MHP |
| V399V | c.1197A>T | Exon 11 | Splicing | 3.5 (5) | Classical PKU |
| R408W | c.1222C>T | Exon 12 | Missense | 2.8 (4) | Classical PKU |
| R408Q | c.1223G>A | Exon 12 | Missense | 12.0 (17) | Mild |
| R413P | c.1238G>C | Exon 12 | Missense | 4.9 (7) | Classical PKU |
| A434D | c.1301C>A | Exon 12 | Missense | 2.8 (4) | Classical PKU |
| Unidentified | | | | 2.1 (3) | |
| Total | | | | 100.0 (142) | |

Abbreviations: MHP, mild hyperphenylalaninemia; PKU, phenylketonuria.
[a]Sequence changes at the protein level.
[b]Sequence changes at the DNA level. Nucleotide+1 is the A of the ATG initiation codon.

genotype 41, F392I/?). As the associated *PAH* allele in these two patients was either a mild R408Q mutation or one of unknown identity, it was difficult to assign phenotype to F392I conclusively. It is noteworthy that no other mutations could be identified in the coding region of the *PAH* gene in these two patients, and none of the F392I (c.1174T>A) substitution was found among the 108 control alleles examined (data not shown), suggesting that the substitution may not be a neutral polymorphism but a mild *PAH* mutation.

### Further functional analysis of selected *PAH* mutations

To study the relationship between genotype and clinical phenotype of *PAH* mutations, knowledge of the residual activity of a mutant gene was important. In this aspect, nucleotide alteration predicted to completely abolish PAH activity, that is, nonsense and frame-shift mutations, and splice-site mutations at both the 5′- and 3′-end of an intron are counted as null. Also considered null in gene activity is the −4173_−407del deletion in the *PAH* enhancer region that has been shown to severely impair *PAH* hepatic transcriptional activity.[11]

To improve on accurate assignment of mutational phenotypes, several mutations were chosen for analysis of *in vitro* activities. First, residual enzyme activities of a set of missense mutations in transfected COS-1 cells were determined (Figure 1b) and their relative activities (RA), that is, percent of *in vitro* activity relative to that of wild-type control, were determined (Table 2). It is to be noted that residual enzyme activities of R261Q, P314T, R408Q and R408W were not assayed in this study but were taken from published reports.[16,19–21] We found that *in vitro* residual enzyme activities associated with mutations can vary considerably to that of published reports. For example, PAH protein carrying the R241C substitution showed 25% of residual activity in the COS cell expression system in a report by Okano *et al.*[22] however, we obtained a result of 49% for this mutant in our study. This may, in some extent, be due to the mutant protein

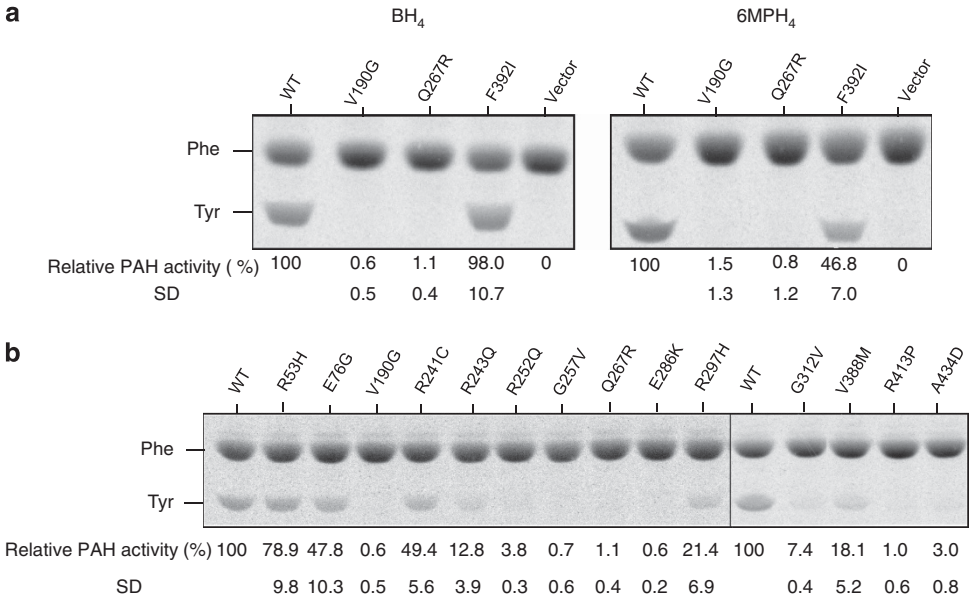


**Figure 1** Effects of *PAH* mutations on PAH enzymatic activities. *PAH* cDNA expression plasmid carrying a defined substitution was transfected to COS-1 cells. Constant amounts of crude extract from transfected cells were aliquoted for determination of PAH enzymatic activities using $BH_4$ (**a**, left panel, and **b**) or $6MPH_4$ (**a**, right panel) as a cofactor. The substrate L-[$^{14}$C] phenylalanine (Phe) and the product L-[$^{14}$C] tyrosine (Tyr) were separated on a TLC plate and visualized using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA, USA). PAH enzymatic activities were calculated as cpm of L-[$^{14}$C] Tyrosine/(cpm of L-[$^{14}$C] Phenylalanine + cpm of L-[$^{14}$C] Tyrosine), where the mutant enzymatic activity was expressed as a percentage of that of the wild type (WT). Mutations are indicated at the top. cDNA, complementary DNA; PAH, phenylalanine hydroxylase.

manifesting variable residual activities dependent on assay conditions, such as substrate concentration, or the type of cofactors used. In this regard, the residual activity of PAH carrying F392I was shown to vary considerably depending on whether cofactor $BH_4$ or $6MPH_4$ was used (Figure 1a). Therefore, the natural state of some mutant enzymes may not be able to be accurately reproduced *in vitro*.

Likewise, we examined how severe the IVS7 + 5G > A base substitution that alters intron 7 sequence outside GT–AG dinucleotides affected pre-mRNA splicing. COS-1 cells were transfected with a modified *PAH* cDNA expression plasmid carrying the intron 7 sequence (Figure 2a). Northern blot analysis of *PAH* RNA showed that when a construct carrying the G–A substitution at the fifth nucleotide of intron 7 (IVS7 + 5G > A), pre-mRNA splicing was completely abolished; only the 3.5-kb primary transcript was detected (Figure 2b, lane 1). In contrast, properly spliced mRNA of 2.4 kb, in addition to the 3.5-kb primary transcript, could be detected if a construct carrying the wild-type intron sequence was used (Figure 2b, lane 2), suggesting that the IVS7 + 5G > A substitution severely impairs *PAH* pre-mRNA splicing and hence the PAH enzymatic activity (Figure 2c, lane 1). Thus, the IVS7 + 5G > A mutation is classified as null. It is noted that the IVS7 + 5G > A base substitution does not involve the invariant GT dinucleotides at the splice-donor site, and yet such a substitution severely affects pre-mRNA splicing. This may be because the splice-donor sequence of intron 7 is 'weak' (CC/GTGAGT, where the slash '/' indicates the exon–intron border) in that it matches only six bases of the eight-base consensus sequence (AG/GTRAGT, where R = purine). In such a configuration, a base change in the consensus sequence may lead to near-complete abolishment of pre-mRNA splicing.[23]

### Correlation of mutation genotype to clinical phenotype

The 34 different PAH mutations identified are variously combined into 50 genotypes in the 71 PKU families studied (Table 2). Homozygous mutations were found in 10 families, which gave an observed frequency of homoallelic PKU genotypes of 14% (Table 2, genotypes 1–4). On the basis of the pretreated plasma phenylalanine levels also taken as references to the phenylalanine tolerance data when available, three clinical phenotypes, that is, classical PKU, mild PKU and MHP, were assigned.[1,6] Among the 71 PKU families, 22 were assigned to classical PKU, 26 to mild PKU and 19 to MHP, whereas, 4 families were not assigned due to lack of sufficient clinical data (Table 2). In our PKU cohort, patients possessing the same mutation genotype were found to fall mostly into a single clinical category; the exception was genotype E286K/R241C, which was assigned to the categories of mild PKU and MHP (Table 2, genotype 26). Nevertheless, more cases are needed to firmly establish such a correlation, especially when many genotypes occurred in a single case.

It is noted that a number of mutations, although showed substantial *in vitro* activities, were found to manifest severe clinical phenotypes. As an example, PAH carrying the R243Q substitution showed 13% residual enzyme activity in our study or 10% in the study of Okano *et al*.[22] Yet, this mutation appears to be associated with a severe phenotype as deduced by its homozygous state (Table 2, genotype 3) or functionally hemizygous state (Table 2, genotypes 10 and 30). Likewise, substantial enzymatic activities were shown for R297H and R261Q, and yet patients carrying this allele displayed severe clinical phenotype (Table 2, genotype 31, R297H/R413P and genotype 35, R261Q/?). It is noteworthy that patients carrying R261Q or R297H were identified at teenage when symptoms of mental retardation were apparent. More studies such as measurement of their phenylalanine tolerance level are needed to re-evaluate phenotype assignments. In particular, the R297H mutation in functional hemizygous state in a PKU cohort of Poland was found to be associated with the MHP phenotype.[24] One would envision that if a mutant protein is displaying kinetic instability, the protein can be stabilized *in vitro* if substrates or cofactors are exceeding physiological

**Table 2 Genotype–phenotype association**

| | | | | No. of patients with clinical phenotype | | | |
|---|---|---|---|---|---|---|---|
| | *Mutation genotype* | *RA (%)* | *AV1 + AV2 (predicted phenotype)* | *Classical* | *Mild* | *MHP* | *NA* |
| (1) | EX-96A>G + EX-96A>G | ?+? | 2+2 (Moderate/mild) | 2 | | | |
| (2) | R241C + R241C | 49+49 | 8+8 (MHP) | | | 5 | 1 |
| (3) | R243Q + R243Q | 13+13 | 1+1 (Classical) | 1 | | | |
| (4) | R413P + R413P | 1+1 | | 1 | | | |
| (5) | −4173_−407del + E286K | Null+<1 | | 1 | | | 1 |
| (6) | E228X + V190G | Null+<1 | | 1 | | | |
| (7) | R261X + R413P | Null+1 | | 1 | | | |
| (8) | T81_R86>VfsX6 + V399V | Null+2 | | 1 | | | |
| (9) | R111X + G312V | Null+7 | 1+? | 1 | | | |
| (10) | IVS7 + 5G>A + R243Q | Null+13 | 1+1 (Classical) | 1 | | | |
| (11) | −4173_−407del + E76G | Null+48 | | | | 1 | |
| (12) | IVS4 −1G>A + R241C | Null+49 | 1+8 (MHP) | | 1 | | |
| (13) | R111X + R241C | Null+49 | 1+8 (MHP) | | 1 | | |
| (14) | −4173_−407del + R408Q | Null+55 | ?+4 | | 2 | | |
| (15) | R111X + R408Q | Null+55 | 1+4 (Mild) | | 2 | | |
| (16) | Y356X + R408Q | Null+55 | 1+4 (Mild) | | 1 | | |
| (17) | R111X + R155H | Null+? | 1+8 (MHP) | | | 1 | |
| (18) | R155fsX43 + EX-96A>G | Null+? | 1+2 (Moderate) | 1 | | | |
| (19) | F302fsX39 + EX-96A>G | Null+? | 1+2 (Moderate) | 1 | | | |
| (20) | G257V + Q267R | <1+1 | | 1 | | | |
| (21) | R408W + V399V | <1+2 | 1+? | 1 | | | |
| (22) | E286K + R252Q | <1+4 | ?+1 | 1 | | | |
| (23) | G257V + G312V | <1+7 | | 1 | | | |
| (24) | R408W + R241C | <1+49 | 1+8 (MHP) | | 2 | | |
| (25) | G257V + R241C | <1+49 | ?+8 (MHP) | | 1 | | |
| (26) | E286K + R241C | <1+49 | ?+8 (MHP) | | 2 | 1 | |
| (27) | E286K + R408Q | <1+55 | ?+4 | | 1 | | |
| (28) | R408W + R408Q | <1+55 | 1+4 (Mild) | | 1 | | |
| (29) | E286K + ? | <1+? | | 1 | | | |
| (30) | R413P + R243Q | 1+13 | ?+1 | 1 | | | |
| (31) | R413P + R297H | 1+21 | | 1 | | | |
| (32) | R413P + R408Q | 1+55 | ?+4 | | 2 | | |
| (33) | G312V + IVS8 −7A>G | 7+? | ?+1 | 1 | | | |
| (34) | V399V + EX-96A>G | 2+? | ?+2 | 1 | | | |
| (35) | R261Q + ? | 30+? | 2/1+? | 1 | | | |
| (36) | P314T + V399V | 25+2 | | | | 1 | |
| (37) | R53H + S70del | 79+? | | | | 1 | |
| (38) | R53H + R243Q | 79+13 | ?+1 | | | 1 | |
| (39) | R53H + R252Q | 79+4 | ?+1 | | | 1 | |
| (40) | F392I + R408Q | 98+55 | ?+4 | | | 1 | |
| (41) | F392I + ? | 98+? | | | | 1 | |
| (42) | R241C + EX-96A>G | 49+? | 8+2 (MHP) | | 5 | | |
| (43) | R241C + IVS8 −7A>G | 49+? | 8+1 (MHP) | | | 1 | |
| (44) | R241C + R243Q | 49+13 | 8+1 (MHP) | | | | 1 |
| (45) | R241C + V388M | 49+18 | 8+2 (MHP) | | 1 | | |
| (46) | R241C + R408Q | 49+55 | 8+4 (MHP) | | | 2 | 1 |
| (47) | R241C + A434D | 49+3 | 8+? (MHP) | | 2 | | |
| (48) | R408Q + EX-96A>G | 55+? | 4+2 (Mild) | | 1 | | |
| (49) | R408Q + V399V | 55+2 | 4+? | | 1 | | |
| (50) | R408Q + A434D | 55+3 | 4+? | | | 2 | |
| | | | Total | 22 | 26 | 19 | 4 |

Abbreviations: AV, assigned value for individual PAH as described in Guldberg *et al.*,[1] i.e., classical PKU, Total AV = 1; moderate PKU, AV = 2; mild PKU, AV = 3; MHP, AV = 8; MHP, mild hyperphenylalaninemia; NA, not available; PAH, phenylalanine hydroxylase mutation; PKU, phenylketonuria; RA, percent of *in vitro* activity relative to the wild-type control.
Predicted phenotypes in parentheses are determined by the sum of the AVs of the two PAH mutations: classical PKU = 2; moderate PKU = 3; moderate/mild PKU = 4; mild PKU = 5–6; mild PKU/MHP = 8; MHP = 9–16. Clinical phenotype is assigned mainly by pretreatment plasma phenylalanine level as described in Guldberg *et al.*[6]

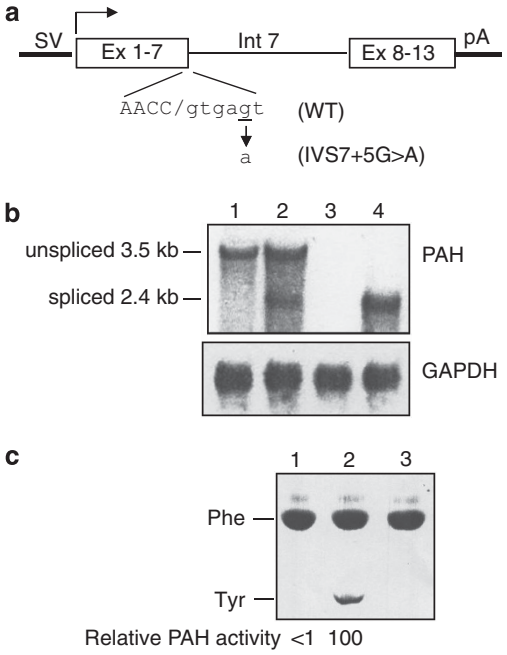


**Figure 2** Effects of the *PAH* IVS7 + 5G > A mutation on PAH activity. (**a**) Schematic representation of a *PAH* expression plasmid under transcriptional control of SV40 promoter (SV) and polyadenylation signal (pA). The bent arrow indicates transcriptional start site. The *PAH* sequence includes exons 1 to 13 and intron 7 (Int 7); the nucleotide sequences flanking the 5′ splice junction of intron 7 are shown with lowercase letters representing the intron sequence. The IVS7 + 5G > A mutation carrying nt 5 mutation in intron 7 is highlighted. (**b**) Northern blot analysis. Total RNA isolated from COS-1 cells transfected with *PAH* expression plasmid with intron 7 carrying the IVS7 nt 5 mutation (IVS + 5G > A) (lane 1), intron 7 carrying the wild-type (WT) sequence (lane 2), untransfected COS-1 control (lane 3) or wild-type *PAH* cDNA (lane 4) were hybridized with $P^{32}$-labeled *PAH* cDNA probe (upper panel) and subsequently to *GAPDH* cDNA probe (lower panel) after stripping. The unspliced and spliced *PAH* mRNAs are indicated. (**c**) PAH enzymatic activities assay. Constant amounts of crude extract from COS-1 transfected cells were aliquoted for determination of PAH enzymatic activities using $6MPH_4$ as a cofactor. The substrate L-[$^{14}$C] phenylalanine (Phe) and the product L-[$^{14}$C] tyrosine (Tyr) were separated on a TLC plate and visualized using a PhosphorImager. The samples in each lane are as in lanes 1–3 in (**b**). cDNA, complementary DNA; PAH, phenylalanine hydroxylase.

concentrations. Therefore, to make a better correlation between residual enzyme activities and clinical severity, detailed biochemical characterization such as determinations of specific activity, substrate activation, enzyme kinetics and stability of the mutant PAH proteins may be required.[25,26]

On the basis of the phenotype of the patient and data of *in vitro* expression study, phenotype of *PAH* mutation was assigned (Table 1). Among the 34 mutations, 23 were assigned to classical PKU, 2 to mild PKU, 5 to MHP and 4, namely, S70del, R261Q, R297H and V388M, remained to be assigned. It is noted that patients displayed classical PKU phenotype when carrying the S70del and R261Q mutations in their functionally hemizygous state, or R261Q in a homozygous state according to the study of Song *et al.*[27] Therefore, both S70del and R261Q are likely associated with severe PKU phenotype. As for the V388M mutation, the mutant protein showed 18% of the normal enzyme activity in our work (Figure 1b), in agreement with 15% of a mutant in the study of Okano *et al.*,[22] which is known to be a kinetic variant.[28] PKU patients in Japan and Korea carrying V388M in their functionally hemizygous state showed classical PKU.[29–31] Thus, the V388M mutation is most likely associated with severe phenotype. In line with this, mild PKU phenotype was observed in one patient in our PKU cohort carrying compound heterozygous mutation of V388M/R241C (Table 2, genotype 45). Furthermore, when phenotype assignments of *PAH* mutation were compared among Oriental PKU cohorts,[22,27,29–31] a good agreement was found among the available mutations that were studied, except for that of R241C. The R241C mutation was found to be associated with mild PKU phenotype in our cohort as well as in that from Korea and Japan.[29–31] However, in China as studied by Song *et al.*,[27] R241C was found to be associated with mild PKU in three patients carrying the compound heterozygote mutation of R241C/R243Q, or with MHP in one patient with the compound heterozygote mutation of R241C/V399V. It is conceivable that clinical phenotype of R241C may vary depending on residual activity of the associated allele when present in a near-functional hemizygous state. It is noted that none of the patients from this study and that of Korea and Japan carried R241C/V399V genotype. In summary, among the *PAH* mutations identified in this study, classical PKU type occupied the highest proportion (23 out of 31 mutations, 74%), but more than 60% of the PAH-deficient families (45 out of 67 families) belonged to milder forms probably due to the fact that the most common R241C and R408Q mutations belong to mild forms of PAH deficiency.

On the other hand, Guldberg *et al.*[1] have established a system for genotype-based prediction of metabolic phenotype using PKU cohort of Europe. They assigned each of the 105 *PAH* mutations an 'assigned value' (AV) to one of four phenotypic classes, that is, mutation-causing severe PKU (AV = 1), moderate PKU (AV = 2), mild PKU (AV = 4) and MHP (AV = 8), based on phenotypes of *PAH* mutations in association with functionally hemizygous state. These authors further demonstrated that in the majority of cases studied, phenotypic effects of two mutant *PAH* alleles could be determined by the sum of their AVs: classical PKU, AV = 2; moderate PKU, AV = 3; moderate/mild PKU, AV = 4; mild PKU, AV = 5–6; mild PKU/MHP, AV = 8; and MHP, AV = 9–16. To examine whether such a system is valid in our PKU cohort, phenotype assignment of the *PAH* mutations by Guldberg *et al.*,[1] if available, is included in Table 2. The results showed that observed clinical phenotype matched the predicted phenotype in patients, except those carrying the EX6-96A > G and R241C alleles (Table 2). EX6-96A > G was assigned as moderate PKU by Guldberg *et al.*,[1] yet patients homozygous (Table 2, genotype 1) or functionally hemizygous for EX-96A > G (Table 2, genotypes 18 and 19) manifested classical PKU phenotype in our cohort as in patients from China, Korea and Japan.[27,29–31] Likewise, R241C, the most frequent mutations in Taiwan, has been assigned by Guldberg *et al.*[1] as MHP. However, the majority of our patients in or near functionally hemizygous state of R241C (Table 2, genotypes 12, 13, 24, 25 and 42) belonged to the mild PKU category. Mutations EX6-96A > G and R241C are prevalent among oriental populations,[22,27,29–32] yet are rare in European populations.[1] Assignment of EX6-96A > G and R241C by Guldberg *et al.*[1] was done from only one allele in a PKU patient, conceivably leading to discrepancy due to a small number of patients being analyzed. It is also possible that the R241C mutation may manifest MHP or mild PKU phenotype dependent on the associated allele as shown in the study of Song *et al.*[27] Thus, in the majority of PKU mutations under study, a good agreement of phenotype assignment was found not only in Oriental PKU cohort but also in that of Caucasians, indicating that disease severity is primarily determined by *PAH* genotype as have also been suggested by other studies.[1,22]

**Table 3 Genetic diversity of the *PAH* common mutations in Taiwan**

| | | | | Associated SNP sequence | | | | | | | Associated STR allele | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *PAH genotype* | n | *RFLP haplotype* | *VNTR cassette sequence* | *IVS2+19 T/C* | *IVS3−22 T/C* | *IVS4+47 C/T* | *c.696G/A Q232Q* | *c.735A/G V245V* | *IVS10+97 A/G* | *c.1155G/C L385L* | n | *size (bp)* |
| Wild type | 49 | 4 | a2-b2-c1 | T | T | C | G | A | A | G | 1 | 230 |
| | | | | | | | | | | | 2 | 234 |
| | | | | | | | | | | | 17 | 238 |
| | | | | | | | | | | | 24 | 242 |
| | | | | | | | | | | | 5 | 246 |
| R241C | 33 | 4 | a2-b2-c1 | T | T | C | G | A | A | G | 2 | 234 |
| | | | | | | | | | | | 28 | 238 |
| | | | | | | | | | | | 3 | 242 |
| R408Q | 15 | 4 | a2-b2-c1 | T | C | T | G | A | A | G | | 238 |
| | 2 | 11 or 34 | (a1)3-b5-c1 | C | C | C | A | G | G | G | | 238 |
| EX-96A>G | 13 | 4 | a2-b2-c1 | T | T | C | G | A | A | G | | 242 |

Abbreviations: RFLP, restriction fragment length polymorphism; SNP, single-nucleotide polymorphism; VNTR, variable number of tandem repeat.

**Genetic diversity of the common *PAH* mutations in Taiwan**

In this study, mutations R241C, R408Q and EX6-96A>G were found to be most common in Taiwan, accounting 23.2%, 12.0% and 9.2% of the mutant chromosomes, respectively (Table 1). To study whether their prevalence was due to enrichment through the founder effect or the result of the recurrent mutation, the associated haplotypes of these mutant alleles were determined by analysis of RFLP, VNTR, STR and seven intragenic SNP markers (Table 3).

The R241C mutation involving a CpG dinucleotide could be a mutational hotspot. Among the 33 chromosomes harboring R241C, all were found to be associated with RFLP haplotype 4 (Table 3), a major haplotype in both wild type and mutant alleles in Taiwan.[10] This haplotype was found to be associated with three-repeat VNTR (VNTR 3) with a cassette sequence of a2-b2-c1, designated as haplotype 4.3.[33] Variations on the associated STR number, including 234, 238 and 242, were observed for R241C with unimodal distribution centered on the 238 allele (Table 3). In contrast, a greater STR allelic diversity was found in 49 wild-type RFLP haplotype 4.3 alleles examined, where the STR variation exhibited a broad unimodel distribution with the 238 and 242 alleles at comparable frequencies (35% vs 49%) (Table 3). The study by Chien *et al.*[9] have found that R241C is the most common PAH mutation in Taiwan suggesting a strong founder effect. In our studies of RFLP haplotype, VNTR configuration and SNPs pattern, we showed that all R241C alleles ($n=33$) fell into the same *PAH* chromosomal background to support that they were identical by descent (Table 3). However, variations on STR allele distribution were observed (Table 3). The STR loci evolve rapidly and have been shown to have high mutation rates.[34] The association of the R241C mutation with chromosomes of various STR numbers suggests that R241C has occurred rather early in population history. Nonetheless, one cannot rule out that R241C has occurred independently in this chromosomal background, especially when haplotype 4 is the most common haplotype among Chinese.[10] The R241C allele is also found in China, Korea and Japan with reported frequencies of 3.7%, 5.7% and 10.1%, respectively.[30–32] It would be of interest to study their associated haplotypes to explore whether they originated from the same founder.

R408Q that accounted for 12% of mutant alleles in Taiwan has been reported to be in linkage disequilibrium with *PAH* RFLP haplotype 4.[35] Interestingly, in our PKU cohort, 2 among a total of 17 R408Q alleles studied were found to be associated with VNTR of five repeats (VNTR 5) of the cassette sequence $(a1)_3$-b5-c1 (Table 3). The RFLP haplotype of this allele was deduced to be either haplotype 11 or 34 depending on the status of the EcoRV recognition sequence, which cannot be analyzed by the current PCR-based method. Genotype of intragenic SNP markers including four in introns IVS2+19T/C, IVS3−22T/C, IVS4+47C/T and IVS10+97A/G, and three in exons c.696G/A (Q232Q), c.735A/G (V245V) and c.1155G/C (L385L), were determined by sequencing the PCR-amplified genomic DNA. Substantial SNP variations between R408Q alleles associated with VNTR 3 and VNTR 5 were found (Table 3). Apparently, the R408Q mutation that appeared on two chromosomal backgrounds is most likely due to the recurrent mutation. In contrast, the EX6-96A>G mutation that was associated with the RFLP 4.3 haplotype appeared in a single SNP haplotype with STR of 242, and is likely derived from a founder mutation (Table 3). It is noteworthy that although the major R408Q allele was found to associate with chromosome of the RFLP haplotype 4.3 similar to that of R241C and EX6-96A>G, R408Q SNP patterns at IVS3−22 and IVS4+47 are different (Table 3). R408Q is the only mutation associated with the minor subtype of RFLP haplotype 4.3.

**Ethnic heterogeneity in the Taiwanese population**

Among the 34 different *PAH* mutations identified, nine mutations, R241C, R408Q, EX6-96A>G, E286K, R413P, R243Q, −4173_−407 del, R111X and V399V, accounting for 70% of all mutant alleles, showed a relative frequency of 3% or more and are regarded as prevalent. Amongst them, R241C, EX6-96A>G, R413P, R243Q and R111X are also considered as prevalent in Orientals, whereas, V399V reaches a relative frequency of 3.2% in Chinese Han population.[32] The remaining mutations, R408Q (12.0%), E286K (5.6%) and −4173_−407del (3.5%), accounting for 21% of all mutant alleles in Taiwan, are either very rare or undetected among PKU cohorts of Chinese, Koreans and Japanese,[22,27,29–32] suggesting local founder effects. Therefore, the *PAH* mutation spectrum in Taiwan shows remarkable differences despite displaying various links to other Asian regions.

On the other hand, the R408W mutation, the most common European PKU mutation, is rare among PKU patients of Chinese, Koreans and Japanese ethnic groups,[22,27,29–32] yet R408W accounts for 2.8% of mutant alleles ($n=4$) in Taiwan. Particularly, this allele was found to be associated with haplotype 44 ($n=2$), 41 ($n=1$) and 4

($n=1$).[35,36] The high recurrent rates of R408Q (c.1223G>A) and R408W (c.1222C>T) have been attributed to CpG dinucleotide methylation.[37] The association of the R408Q and R408W alleles with various haplotypes supports the complexity of general population in Taiwan. On the other hand, homozygosity value at the *PAH* locus in a given population (determined by $j=\Sigma x_i^2$, where $x_i$ is the frequency of the $i$th allele) has been used as an indicator of population heterogeneity.[6] The homozygosity value was calculated to be 0.092 in Taiwanese, which is comparable to the value of 0.089 in Chinese.[32] Such a low value supports the high degree of ethnic heterogeneity in the Taiwanese population. In this aspect, Lin *et al.*[38] based on a HLA study, showed that the main ethnical constituents of the population in Taiwan, namely Minnan and Hakka, are mainly the descendants from tribal groups of the southeast coast of China with extensive intermarriage to the aboriginal people of Taiwan. Therefore, in addition to the shared ancestry between Chinese Han and Taiwanese populations, our finding supports the notion of a unique ancestral origin of the residents of Taiwan.[38]

## CONCLUDING REMARKS

Taiwan, a small island, has been at the crossroads of population movements.[7,8] Researches based on studying genetic variations in human gut bacteria *Helicobacter pylori*, and on linguistic evidences, indicate that most Pacific populations originated from Taiwan ~5000 years ago.[39,40] However, by examining genetic signature on mitochondrial DNA, Soares *et al.*[41] found that Polynesian people may have spread to the Pacific via New Guinea rather than via Taiwan as previously thought. In this work, we showed that the Taiwanese population has a unique spectrum of *PAH* mutations. It would be of interest to further study the *PAH* mutation spectrum in Pacific populations to clarify their origins and migration pattern in relation to Taiwan.

## ACKNOWLEDGEMENTS

We thank Dr K-B Choo for critical reading and editing of the manuscript. This work was supported by grants NSC 89-2320-B-075-051, NSC 90-2320-B-075-010 and NSC 91-2316-B-075-001 from the National Science Council and by grants VGH94-370-15, V95S4-021, V96S4-007, V97S4-015 and V98S4-010 from the Taipei Veterans General Hospital, Taiwan.

1 Guldberg, P., Rey, F., Zschocke, J., Romano, V., Francois, B., Michiels, L. *et al.* A European multicenter study of phenylalanine hydroxylase deficiency: classification of 105 mutations and a general system for genotype-based prediction of metabolic phenotype. *Am. J. Hum. Genet.* **63,** 71–79 (1998).
2 Eisensmith, R. C., Okano, Y., Dasovich, M., Wang, T., Güttler, F., Lou, H. *et al.* Multiple origins for phenylketonuria in Europe. *Am. J. Hum. Genet.* **51,** 1355–1365 (1992).
3 Kidd, J. R., Pakstis, A. J., Zhao, H., Lu, R. B., Okonofua, F. E., Odunsi, A. *et al.* Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. *Am. J. Hum. Genet.* **66,** 1882–1899 (2000).
4 Zschocke, J. Phenylketonuria mutations in Europe. *Hum. Mutat.* **21,** 345–356 (2003).
5 Scriver, C. R. The PAH gene, phenylketonuria, and a paradigm shift. *Hum. Mutat.* **28,** 831–845 (2007).
6 Guldberg, P., Levy, H. L., Hanley, W. B., Koch, R., Matalon, R., Rouse, B. M. *et al.* Phenylalanine hydroxylase gene mutations in the United States: report from the Maternal PKU Collaborative Study. *Am. J. Hum. Genet.* **59,** 84–94 (1996).
7 Blundell, D. *Taiwan-Linguistics, History and Prehistory* (SMC Publishing, Taipei, Taiwan, 2000).
8 Wu, D. Y. H. In *China Off Center: Mapping the Margins of the Middle Kingdom* (eds Blum, S. D. & Jenson, L. M.) 167–187 (University of Hawaii Press, 2002).
9 Chien, Y. H., Chiang, S. C., Huang, A., Chou, S. P., Tseng, S. S., Huang, Y. T. *et al.* Mutation spectrum in Taiwanese patients with phenylalanine hydroxylase deficiency and a founder effect for the R241C mutation. *Hum. Mutat.* **23,** 206 (2004).
10 Chen, S. H., Hsiao, K. J., Lin, L. H., Liu, T. T., Tang, R. B. & Su, T. S. Study of restriction fragment length polymorphisms at the human phenylalanine hydroxylase locus and evaluation of its potential application in prenatal diagnosis of phenylketonuria in Chinese. *Hum. Genet.* **81,** 226–230 (1989).
11 Chen, K. J., Chao, H. K., Hsiao, K. J. & Su, T. S. Identification and characterization of a novel liver-specific enhancer of the human phenylalanine hydroxylase gene. *Hum. Genet.* **110,** 235–243 (2002).
12 Goltsov, A. A., Eisensmith, R. C., Konecki, D. S., Lichter-Konecki, U. & Woo, S. L. Associations between mutations and a VNTR in the human phenylalanine hydroxylase gene. *Am. J. Hum. Genet.* **51,** 627–636 (1992).
13 Goltsov, A. A., Eisensmith, R. C., Naughton, E. R., Jin, L., Chakraborty, R. & Woo, S. L. A single polymorphic STR system in the human phenylalanine hydroxylase gene permits rapid prenatal diagnosis and carrier screening for phenylketonuria. *Hum. Mol. Genet.* **2,** 577–581 (1993).
14 Zschocke, J., Graham, C. A., McKnight, J. J. & Nevin, N. C. The STR system in the human phenylalanine hydroxylase gene: true fragment length obtained with fluorescent labelled PCR primers. *Acta Paediatr. (Suppl.)* **407,** 41–42 (1994).
15 Sambrook, J. & Russell, D. W. *Molecular Cloning- A Laboratory Manual*. 3rd edn (Cold Spring Harbor Laboratory Press, 2001).
16 Ho, P. Y., Huang, M. Z., Fwu, V. T., Lin, S. C., Hsiao, K. J. & Su, T. S. Simultaneous assessment of the effects of exonic mutations on RNA splicing and protein functions. *Biochem. Biophys. Res. Commun.* **373,** 515–520 (2008).
17 Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* **3,** 285–298 (2002).
18 Chao, H. K., Hsiao, K. J. & Su, T. S. A silent mutation induces exon skipping in the phenylalanine hydroxylase gene in phenylketonuria. *Hum. Genet.* **108,** 14–19 (2001).
19 Okano, Y., Eisensmith, R. C., Guttler, F., Lichter-Konecki, U., Konecki, D. S., Trefz, F. K. *et al.* Molecular basis of phenotypic heterogeneity in phenylketonuria. *N. Engl. J. Med.* **324,** 1232–1238 (1991).
20 Svensson, E., Eisensmith, R. C., Dworniczak, B., von Dobeln, U., Hagenfeldt, I., Horst, J. *et al.* Two missense mutations causing mild hyperphenylalaninemia associated with DNA haplotype 12. *Hum. Mutat.* **1,** 129–137 (1992).
21 Dilella, A. G., Marvit, J., Brayton, K. & Woo, S. L. C. An amino-acid substitution involved in phenylketonuria is in linkage disequilibrium with DNA haplotype 2. *Nature* **327,** 333–338 (1987).
22 Okano, Y., Asada, M., Kang, Y., Nishi, Y., Hase, Y., Oura, T. *et al.* Molecular characterization of phenylketonuria in Japanese patients. *Hum. Genet.* **103,** 613–618 (1998).
23 Ketterling, R. P., Drost, J. B., Scaringe, W. A., Liao, D. Z., Liu, J. Z., Kasper, C. K. *et al.* Reported *in vivo* splice-site mutations in the factor IX gene: severity of splicing defects and a hypothesis for predicting deleterious splice donor mutations. *Hum. Mutat.* **13,** 221–231 (1999).
24 Zekanowski, C., Nowacka, M., Cabalska, B. & Bal, J. Molecular basis of mild hyperphenylalaninaemia in Poland. *J. Med. Genet.* **34,** 1035–1036 (1997).
25 Pey, A. L., Desviat, L. R., Gamez, A., Ugarte, M. & Perez, B. Phenylketonuria: genotype-phenotype correlations based on expression analysis of structural and functional mutations in PAH. *Hum. Mutat.* **21,** 370–378 (2003).
26 Dobrowolski, S. F., Pey, A. L., Koch, R., Levy, H., Ellingson, C. C., Naylor, E. W. *et al.* Biochemical characterization of mutant phenylalanine hydroxylase enzymes and correlation with clinical presentation in hyperphenylalaninaemic patients. *J. Inherit. Metab. Dis.* **32,** 10–21 (2009).
27 Song, F., Qu, Y. J., Zhang, T., Jin, Y. W., Wang, H. & Zheng, X. Y. Phenylketonuria mutations in Northern China. *Mol. Genet. Metab.* **86,** S107–S118 (2005).
28 Leandro, P., Rivera, I., Lechner, M. C., de Almeida, I. T. & Konecki, D. The V388M mutation results in a kinetic variant form of phenylalanine hydroxylase. *Mol. Genet. Metab.* **69,** 204–212 (2000).
29 Okano, Y., Hase, Y., Kawajiri, M., Nishi, Y., Inui, K., Sakai, N. *et al. In vivo* studies of phenylalanine hydroxylase by phenylalanine breath test: diagnosis of tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency. *Pediatr. Res.* **56,** 714–719 (2004).
30 Okano, Y., Kudo, S., Nishi, Y., Sakaguchi, T. & Aso, K. Molecular characterization of phenylketonuria and tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency in Japan. *J. Hum. Genet.* **56,** 306–312 (2011).
31 Lee, D. H., Koo, S. K., Lee, K. S., Yeon, Y. J., Oh, H. J., Kim, S. W. *et al.* The molecular basis of phenylketonuria in Koreans. *J. Hum. Genet.* **49,** 617–621 (2004).
32 Zhu, T., Qin, S., Ye, J., Qiu, W., Han, L., Zhang, Y. *et al.* Mutational spectrum of phenylketonuria in the Chinese Han population: a novel insight into the geographic distribution of the common mutations. *Pediatr. Res.* **67,** 280–285 (2010).
33 Eisensmith, R. C. & Woo, S. L. Updated listing of haplotypes at the human phenylalanine hydroxylase (PAH) locus. *Am. J. Hum. Genet.* **51,** 1445–1448 (1992).
34 Weber, J. L. & Wong, C. Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2,** 1123–1128 (1993).
35 Lin, C. H., Hsiao, K. J., Tsai, T. F., Chao, H. K. & Su, T. S. Identification of a missense phenylketonuria mutation at codon 408 in Chinese. *Hum. Genet.* **89,** 593–596 (1992).
36 Tsai, T. F., Hsiao, K. J. & Su, T. S. Phenylketonuria mutation in Chinese haplotype 44 identical with haplotype 2 mutation in northern-European Caucasians. *Hum. Genet.* **84,** 409–411 (1990).
37 Murphy, B. C., Scriver, C. R. & Singh, S. M. CpG methylation accounts for a recurrent mutation (c.1222C>T) in the human PAH gene. *Hum. Mutat.* **27,** 975 (2006).
38 Lin, M., Chu, C. C., Chang, S. L., Lee, H. L., Loo, J. H., Akaza, T. *et al.* The origin of Minnan and Hakka, the so-called 'Taiwanese', inferred by HLA study. *Tissue Antigens* **57,** 192–199 (2001).
39 Moodley, Y., Linz, B., Yamaoka, Y., Windsor, H. M., Breurec, S., Wu, J. Y. *et al.* The peopling of the Pacific from a bacterial perspective. *Science* **323,** 527–530 (2009).
40 Gray, R. D., Drummond, A. J. & Greenhill, S. J. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323,** 479–483 (2009).
41 Soares, P., Rito, T., Trejaut, J., Mormina, M., Hill, C., Tinkler-Hundal, E. *et al.* Ancient voyaging and Polynesian origins. *Am. J. Hum. Genet.* **88,** 239–247 (2011).