

ARTICLE

# The mutation spectrum of hyperphenylalaninaemia in the Republic of Ireland: the population history of the Irish revisited

Kate A O'Donnell<sup>1,2</sup>, Charles O'Neill<sup>1</sup>, Orna Tighe<sup>2</sup>, Giorgio Bertorelle<sup>3</sup>, Eileen Naughten<sup>4</sup>, Philip D Mayne<sup>1</sup> and David T Croke<sup>\*2</sup>

<sup>1</sup>Department of Pathology, The Children's Hospital, Temple Street, Dublin 1, Republic of Ireland; <sup>2</sup>Department of Biochemistry, The Royal College of Surgeons in Ireland, St. Stephen's Green, Dublin 2, Republic of Ireland; <sup>3</sup>Dipartimento di Biologia, Università di Ferrara, via L. Borsari 46, 44100 Ferrara, Italy; <sup>4</sup>National Centre for Inherited Metabolic Disorders, The Children's Hospital, Temple Street, Dublin 1, Republic of Ireland

Phenylketonuric and hyperphenylalaninaemic patients in the population of the Republic of Ireland were screened for mutations at the human phenylalanine hydroxylase (*PAH*) locus. A composite data set for the island of Ireland was generated by merging the findings of this study with extant data for Northern Ireland. Analysis of this data on the basis of the four historic provinces (Munster, Leinster, Connacht and Ulster) revealed genetic diversity that is informative in terms of demographic forces that shaped the Irish population. R408W, the predominant Irish *PAH* mutation associated with haplotype 1.8, reached its highest relative frequency in the most westerly province, Connacht. This suggests that the gradient of R408W-1.8 observed across north-western Europe continues into Ireland and peaks in Connacht. Spatial autocorrelation analysis demonstrated that the gradient is consistent with a localised cline of R408W-1.8 likely to have been established by human migration. This and parallel allele frequency clines may represent the genetic traces of the Palaeolithic colonisation of Europe, a pattern not substantially altered in north-western Europe by subsequent Neolithic migrations. An analysis of mutant allele distributions in Ulster, Scotland and the rest of Ireland confirmed that Ulster has been a zone of considerable admixture between the Irish and Scottish populations, indicating a proportion of Scottish admixture in Ulster approaching 46%. Mutations primarily associated with Scandinavia accounted for 6.1% of mutations overall, illustrating the influence of Viking incursions on Irish population history.

*European Journal of Human Genetics* (2002) 10, 530–538. doi:10.1038/sj.ejhg.5200841

**Keywords:** hyperphenylalaninaemia; phenylketonuria; phenylalanine hydroxylase; mutation; migration; admixture

## Introduction

Hyperphenylalaninaemia (HPA) is a group of inherited disorders in which plasma phenylalanine concentrations are abnormally elevated, giving a clinical presentation involving abnormal development of the central nervous system with cognitive impairment in untreated children.<sup>1</sup> Phenylketonuria (PKU; MIM No. 261600), the most severe form of

HPA, is caused by deficient activity of the hepatic enzyme phenylalanine hydroxylase (*PAH*) which results from mutation at the *PAH* locus on chromosome 12q23.1.<sup>2,3</sup> To date, more than 400 mutations in the *PAH* gene have been identified in human populations and catalogued in *PAHdb* [[http://data.mch.mcgill.ca/pahdb\\_new/](http://data.mch.mcgill.ca/pahdb_new/)], a public-domain database.<sup>4,5</sup> The *PAH* locus also contains a number of neutral polymorphisms useful in defining haplotypes, including seven Single Nucleotide Polymorphisms (SNPs), a Variable Number Tandem Repeat (VNTR) motif and a Short Tandem Repeat (STR) polymorphism.<sup>2,6–9</sup> *PAH* haplotypes, and their

\*Correspondence: Professor David T Croke, Department of Biochemistry, The Royal College of Surgeons in Ireland, 123, St. Stephen's Green, Dublin 2, Republic of Ireland. Tel: (353-1-) 402-2280; Fax: (353-1-) 402-2467; E-mail: dtcroke@rcsi.ie

Received 6 September 2001; revised 23 April 2002; accepted 13 May 2002

**Table 1** Mutation spectrum of hyperphenylalaninaemia in the Southern Irish population

Trivial name	Exon	Number of alleles	Frequency (%)	Detection method	Restriction enzyme
R408W §	12	229	41.0	ACRS	<i>MvaI</i>
F39L	2	68	12.2	RFLP	<i>MaellI</i>
I65T	3	58	10.4	ACRS	<i>TaqI</i>
L348V	10	29	5.2	RFLP	<i>NlaIV</i>
F299C	8	22	3.9	RFLP	<i>PvuII</i>
R243X §	7	19	3.4	ACRS	<i>XbaI</i>
IVS12nt1 §	–	13	2.3	ACRS	<i>RsaI</i>
Y414C	12	8	1.4	ACRS	<i>RsaI</i>
S349P	10	8	1.4	RFLP	<i>PfIM1</i>
IVS10nt-11	–	7	1.3	RFLP	<i>DdeI</i>
E280K §	7	6	1.1	ACRS	<i>XbaI</i>
A447D	13	6	1.1	Sequencing	–
L249F	7	6	1.1	Sequencing	–
A345S	10	4	0.7	Sequencing	–
R252W §	7	4	0.7	RFLP	<i>AvaI</i>
R261Q §	7	3	0.5	RFLP	<i>HinfI</i>
A104D	3	3	0.5	Sequencing	–
G46S	2	3	0.5	ACRS	<i>Sau96 I</i>
IVS10nt-3	–	3	0.5	Sequencing	–
R270K	7	2	0.4	Sequencing	–
S67P	3	2	0.4	ACRS	<i>XbaI</i>
P281L §	7	2	0.4	Sequencing	–
E390G	11	2	0.4	ACRS	<i>BstNI</i>
A246D	7	1	0.2	Sequencing	–
V245A	7	1	0.2	RFLP	<i>Fnu4HI</i>
R408Q §	12	1	0.2	RFLP	<i>HaeIII</i>
R158Q §	5	1	0.2	ACRS	<i>XbaI</i>
Y277D	7	1	0.2	RFLP	<i>AccI</i>
IVS7nt1	–	1	0.2	Sequencing	–
Unidentified	–	21	3.8	–	–
Undetected	–	24	4.3	–	–
Total		558	100.0		

§Mutations occurring at a CpG dinucleotide motif. 'Unidentified' indicates cases where a DGGE mobility shift, indicative of base sequence alteration, was detected but where further characterisation proved impossible. 'Undetected' indicates cases where no DGGE mobility shift was detected. Detection methods: RFLP indicates a naturally occurring restriction site created or destroyed by the mutation; ACRS indicates an artificially created restriction site generated during PCR amplification of the mutant or wild-type sequence. Sequencing indicates characterisation *via* direct DNA sequencing.

associations with specific PKU mutations, have been useful in studying the population genetics of the disorder and in probing the origins of mutations.<sup>10–15</sup>

The genetic basis of PKU and HPA in the Irish population has been investigated in studies which looked separately at the Republic of Ireland and Northern Ireland.<sup>16–18</sup> These reported that three mutations (R408W, F39L and I65T) accounted for over 50% of mutant alleles, although at different individual relative frequencies. The predominant haplotype associated with R408W in both populations was haplotype 1.8,<sup>16,17</sup> that is SNP haplotype 1 with the 8-copy VNTR allele according to the revised *PAH* haplotype nomenclature (R408W-1.8).<sup>14</sup> R408W-1.8 exhibits an east to west cline of increasing relative frequency across north-western Europe peaking in Ireland.<sup>19</sup> We undertook the present study with the objectives of reaching full ascertainment of HPA mutations in the population of the Republic of Ireland, and of investigating regional differences in the HPA mutation spectrum within the island of Ireland, looking in greater depth at the influence of neighbouring populations on the Irish gene-pool.

### Patients, materials and methods

Patients ( $n=386$ ) were recruited to the study through the National Centre for Inherited Metabolic Disorders and the National Newborn Screening Programme at The Children's Hospital (Temple Street, Dublin 1). Initial diagnosis of hyperphenylalaninaemia was made on neonatal screening using the bacterial inhibition assay<sup>20</sup> with confirmation by quantitative analysis of serum amino-acids. Among the 386 patients were 107 sibling pairs and, since only one member of each sib-pair was used for genotypic analysis, 279 unrelated cases in total were available for study. Patient samples were obtained as dried blood spots on Guthrie cards. Small nuclear families ( $n=24$ ) consisting of affected individuals, siblings and parents were recruited to the study for *PAH* haplotype determination. Blood samples were obtained with informed consent from parents by venepuncture and as dried blood spots from children. Anonymised control samples ( $n=50$ ) were obtained, again as dried blood spots, from the National Newborn Screening Programme. Ethical approval for the study was obtained from the Ethics Committee of The Children's Hospital.

Exon sequences of the *PAH* locus were scanned for the presence of mutation by denaturing gradient gel electrophoresis (DGGE) using primer sequences and conditions described by Guldberg *et al.*<sup>21</sup> Where altered electrophoretic mobility of amplicons was observed, the relevant exons were screened for known mutations by PCR–RFLP assays and/or DNA sequencing. The PCR–RFLP assays used either naturally occurring restriction sites or sites created during PCR amplification (*PAHdb*). DNA cycle sequencing<sup>22</sup> was performed using the Big Dye Terminator cycle sequencing kit (Perkin Elmer) and sequencing products were resolved on an ABI 310 automated fluorescent DNA sequencer (Perkin Elmer Applied Biosystems). Mutations and polymorphisms detected in the *PAH* gene were identified by comparison with the reference sequence and genotypic data archived in *PAHdb*.<sup>5</sup>

The distribution of mutant alleles within Ireland was investigated using data on the county of residence of the birth mother of each of the cases studied. Mutant allele counts were summed for the four provinces of Ireland (Connacht, Leinster, Munster and Ulster). Relative allele frequencies for the province of Ulster could only be assessed in this study for three of its nine counties (Cavan, Donegal and Monaghan) since the remaining six lie within the United Kingdom (Northern Ireland). In order to produce an estimate of relative allele frequencies within Ulster as a whole, allele counts for the Ulster counties in the Republic (this study) and in Northern Ireland<sup>17</sup> were summed. Mutant allele frequencies were presented either as relative frequencies or as absolute frequencies to account for variation in disease incidence between populations. Absolute frequency values were calculated as the product of the square root of the disease incidence and the relative frequency of the mutation on the assumption of Hardy-Weinberg equilibrium.<sup>23</sup> Statistical tests on the frequency distributions of mutant genotypes in the four Irish provinces were performed using the SPSS software (SPSS Inc., USA). Spatial autocorrelation analysis<sup>24</sup> of the geographic distribution of R408W-1.8 absolute mutant allele frequencies in north-western Europe was performed by means of the SAAP software package, which measures the average level of genetic similarity between populations in particular geographic distance classes, expressed in a correlogram of Moran's autocorrelation coefficient plotted as a function of distance. Different migration processes are reflected in correlograms of different shapes.<sup>25</sup> The overall significance of correlograms was assessed by the Bonferroni criterion.<sup>26</sup> Admixture calculations were carried out using a modification of the Bernstein model<sup>27,28</sup> which allows estimation of the relative contributions of two parental populations to a hybrid group assuming that the absolute allele frequencies are known. This is not the case, however, when we treat Ulster as a hybrid population formed between Ireland (excluding Ulster) and Scotland. The overall allele frequency of PKU mutants in Ulster,

necessary to estimate the absolute frequencies from their relative frequencies, is unknown, so we introduced a modification of the Bernstein model that allows the simultaneous estimation of the overall allele frequency of mutant alleles in the hybrid population and the admixture coefficients (see Appendix).

To confirm that mutation-haplotype associations for the predominant mutations were as reported previously, *PAH* haplotypes were analysed in a panel of small nuclear families and in a group of patients homozygous for the most common mutations. Five single nucleotide polymorphisms (SNPs) were analysed by means of PCR–RFLP assays, namely *BglII*<sup>29</sup>, *PvuII(a)*<sup>30</sup>, *PvuII(b)* (Eisensmith, personal communication), *MspI*<sup>31</sup> and *XmnI*.<sup>32</sup> The variable number tandem repeat (VNTR) polymorphism (formerly scored as a triallelic *HindIII* polymorphism) was detected by PCR amplification with flanking primers followed by visualisation and size determination of alleles via polyacrylamide gel electrophoresis.<sup>8</sup> Haplotypes were assigned according to the nomenclature of Eisensmith and Woo.<sup>14</sup> Mini-haplotypes<sup>17</sup> associated with specific mutant *PAH* alleles were determined by genotypic analysis of the VNTR marker, together with the short tandem repeat (STR) marker described by Goltsov *et al.*,<sup>9</sup> in homozygous patients. STR genotypes were determined by PCR amplification and allele sizing by electrophoresis on the ABI 310 automated sequencer.<sup>9,33</sup>

## Results

DNA samples from 279 unrelated hyperphenylalaninaemia cases, equivalent to 558 *PAH* alleles, were screened by DGGE for the presence of mutation in the 13 exons of the *PAH* gene. Electrophoretic mobility shifts, indicative of base sequence alteration, were detected for all except 24 (4.3%) alleles. No mobility shifts were detected in exons 1 or 4. Exons exhibiting shifts were then screened for known mutations by means of PCR–RFLP assays. Following restriction analysis, exons with mobility shifts that had not been characterised were analysed by DNA sequencing. For 21 alleles (3.8%), mobility shifts localised to exons 6 ( $n=1$ ), 7 ( $n=11$ ), 9 ( $n=2$ ) and 11 ( $n=7$ ) could not be sequenced because of poor DNA template quality. Mutations were thus characterised for 513 of the 558 alleles (92.0%), identifying 29 distinct mutations (Table 1). Four polymorphisms/silent substitutions were detected at relative frequencies of up to 2.8%, namely IVS2nt19, Q232Q, V245V and L385L (data not shown). The mutations were distributed across the *PAH* coding sequence with 34.5% (10 of 29) occurring in exon 7. The majority were missense mutations (24 of 29, 82.8%) with four splice-site mutations and one nonsense mutation (R243X). Nine mutations occurred at CpG dinucleotide motifs (Table 1). In all, 239 of the 279 cases (85.7%) were fully genotyped, 66 (27.6%) having homoallelic genotypes and the remaining 173 (72.4%) having heteroallelic genotypes. A substantial

proportion of mutant alleles (63.6%) were accounted for by the mutations R408W (41.0%), F39L (12.2%) and I65T (10.4%). All other mutant alleles were present at relative frequencies of 6% or less. No novel mutations were detected. *PAH* mutation-haplotype associations were confirmed by haplotype and minihaplotype analysis. Combined SNP and VNTR haplotypes were determined for the predominant mutations in a group of 22 affected individuals in small nuclear families and minihaplotype associations in a group of 38 homozygous patients (Table 2). Both R408W and F39L were found in association with haplotype 1.8 while I65T was primarily associated with haplotype 9.8 (seven of eight alleles; 87.5%). The predominant minihaplotype association for R408W was with the 8:242 minihaplotype (49 of 60 alleles, 81.7%). F39L was found in association with three minihaplotypes (8:238, 8:242 and 8:246) while I65T was predominantly associated with the 8:246 minihaplotype (five of eight alleles, 62.5%).

**Table 2** Associations with *PAH* haplotypes and minihaplotypes observed for the three most common mutations in the Southern Irish population

Mutation	SNP haplotype	Alleles*	Minihaplotype	Alleles §
R408W	1.8	31	8:238	9
			8:242	49
			8:246	2
			8:238	4
F39L	1.8	5	8:242	3
			8:246	1
			8:246	1
I65T	1.8	1	3:246	1
			9.8	7
	9.8	7	8:246	5
			8:250	1
Total		44	9:246	1
				76

\*Associations between mutations and SNP haplotypes were investigated in small nuclear families of affected individuals; §Minihaplotype associations were investigated in individuals homozygous for the three most common mutations.

The geographic distribution of mutant alleles within the island of Ireland was investigated using a combined data-set produced by merging the data generated in this study with extant data for Northern Ireland (Table 3).<sup>17</sup> Significant differences in frequency distributions of the eight most common mutant alleles were observed between the four provinces ( $P < 0.0001$ ). Based upon this analysis, R408W was found to occur with the highest relative frequency (55.8%) in Connacht, the most westerly province. Given that R408W is associated with haplotype 1.8 in Ireland, these data can be added to those previously reported by Eisensmith *et al.*<sup>19</sup> to show that the cline of R408W-1.8 relative frequency across north-western Europe peaks in

**Table 4** Ulster as a hybrid population formed by admixture between Scotland and Ireland. The proportion of Scottish admixture in the population of Ulster was estimated as the mean value of  $m = (q_2(r_h - r_2)) / (q_1 r_1 - q_2 r_2 - r_h(q_1 - q_2))$  using relative allele frequency data for 13 mutant *PAH* alleles in Ireland (excluding Ulster;  $r_1$ ), Scotland ( $r_2$ ) and Ulster ( $r_h$ )

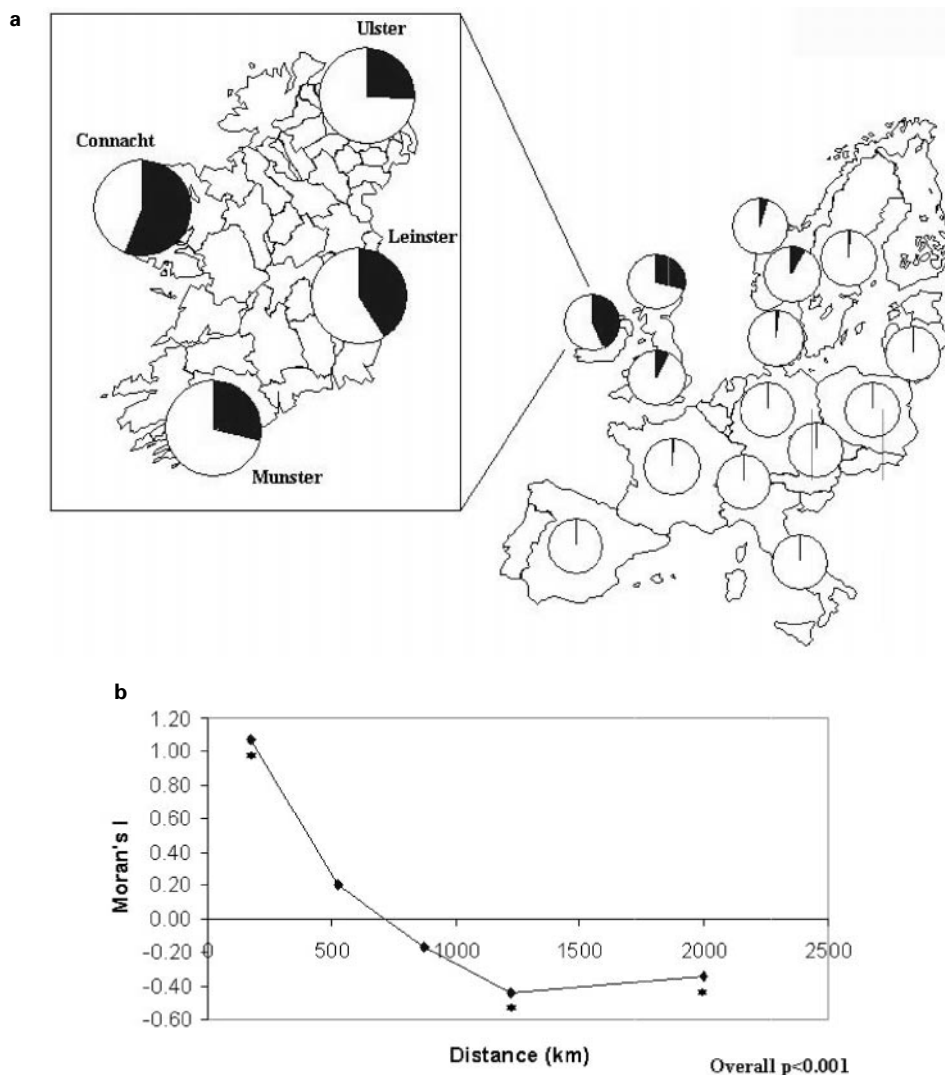
Mutation	$r_1$	$r_2$	$r_h$
R408W	0.4032	0.3447	0.2837
F39L	0.1230	0.0631	0.0957
I65T	0.0867	0.1650	0.2021
L348V	0.0484	0.0777	0.0532
IVS12nt1	0.0181	0.0534	0.0319
E280K	0.0161	0.0097	0.0142
P281L	0.0020	0.0146	0.0035
R158Q	0.0020	0.0049	0.0035
R243X	0.0343	0.0097	0.0177
R261Q	0.0060	0.0097	0.0106
R408Q	0.0020	0.0194	0.0284
S349P	0.0121	0.0097	0.0071
Y414C	0.0161	0.0388	0.0426

Sources of data: Ulster and rest of Ireland – this study (Table 3), western Scotland – Tyfield *et al*, 1997.  $q_1$  and  $q_2$ , respectively the overall mutant allele frequencies for Ireland (excluding Ulster) and Scotland, are 0.0149 and 0.0116.

**Table 3** Geographic distribution of mutant *PAH* alleles within the island of Ireland. Allele counts for eight mutations are given together with their relative frequencies expressed as percentage values (in brackets)

Mutant allele	Munster	Leinster	Connacht	Ulster* Combined	Ulster§ N. Ireland	Ulster* Republic
R408W	31 (29.2)	140 (41.4)	29 (55.8)	80 (28.4)	67 (28.0)	13 (32.5)
F39L	13 (12.3)	46 (13.6)	2 (3.8)	27 (9.6)	23 (9.5)	4 (10.0)
I65T	13 (12.3)	26 (7.7)	4 (7.7)	57 (20.2)	49 (20.0)	8 (20.0)
L348V	8 (7.5)	15 (4.4)	1 (1.9)	15 (5.3)	13 (5.4)	2 (5.0)
F299C	6 (5.7)	11 (3.3)	1 (1.9)	12 (4.3)	9 (3.7)	3 (7.5)
R243X	3 (2.8)	11 (3.3)	3 (5.8)	5 (1.8)	4 (1.7)	1 (2.5)
IVS12nt1	4 (3.8)	2 (0.6)	3 (5.8)	9 (3.2)	8 (3.3)	1 (2.5)
Y414C	2 (1.9)	6 (1.8)	0 (0.0)	12 (4.3)	12 (5.0)	0 (0.0)
Other	26	81	9	65	46	8
Total	106	338	52	282	242	40

Variation in allele frequency (counts) between the four provinces (Munster, Leinster, Connacht, Ulster combined):  $\chi^2 = 62.01$  with 24 d.f.;  $P < 0.0001$ . \*Allele frequencies summed for the nine counties of the province of Ulster; §Allele frequencies for the six Ulster counties located in Northern Ireland (Antrim, Armagh, Down, Fermanagh, Londonderry and Tyrone; Zschocke *et al.*, 1995); \*Allele frequencies for the three Ulster counties located in the Republic of Ireland (Cavan, Donegal and Monaghan; this study).



**Figure 1** A cline of frequency of the R408W-1.8 Phenylketonuria mutation across north-western Europe peaking in Connacht, the most westerly province of Ireland. (A) Depicts relative frequencies of R408W-1.8 in each of fifteen European populations<sup>18,19</sup>, represented by the shaded areas of the relevant pie-charts superimposed on the outline map; values for the provinces of Ireland derived from this study are similarly represented on the inset map. (B) Presents a correlogram obtained by spatial autocorrelation analysis (using the SAAP software) of R408W-1.8 absolute allele frequencies for 13 populations (the four Irish provinces, Scotland, England, France, Germany, Spain, Denmark, Sweden, NW Norway and SE Norway). Individual data points that attained significance at the  $P < 0.05$  level are indicated by asterisks (\*). The overall significance of the correlogram was  $P < 0.001$  (calculated by the Bonferroni criterion).

Connacht (Table 3, Figure 1A). Spatial autocorrelation analysis was used to test the significance of this cline, analysing R408W-1.8 absolute allele frequencies. The correlogram obtained (Figure 1B) is statistically significant ( $P < 0.001$ ) and is consistent with a regional cline of R408W-1.8.

In contrast to R408W, the F39L and I65T mutations exhibited maximal relative frequencies in Leinster (13.6%) and Ulster (20.0%) respectively (Table 3). IVS12nt1, the most common English PKU mutation (27%),<sup>34</sup> was found to occur at relative frequencies of 3–6% in Munster,

Connacht and Ulster but was relatively uncommon (0.6%) in Leinster. In addition, three mutations (F299C, R408Q and Y414C) which are prevalent on the Atlantic coast of Norway<sup>35</sup> were found to be under-represented in Connacht as compared to the other provinces (Table 3).

The sub-division of data on the basis of the historic provinces of Ireland facilitated an examination of the influence on Ulster of its neighbouring population, Scotland. Admixture calculations were thus performed using the frequencies of 13 mutant *PAH* alleles in Scotland<sup>34</sup>, Ulster (as the hybrid population) and the rest of Ireland (Table

4). By the method described in the Appendix, the proportion of Scottish admixture in Ulster was estimated as 0.4561 (46%). The overall mutant *PAH* allele frequency in Ulster was also estimated in these calculations as 0.0131, corresponding to a disease incidence of approximately one in 6000 (assuming Hardy–Weinberg equilibrium).

## Discussion

The mutation spectrum of hyperphenylalaninaemia in the Irish Republic has been elucidated, mutations being detected in 95.7% of alleles (534 of 558), a detection efficiency comparable with other recent studies.<sup>17,34,36</sup> Mutations were fully characterised in 513 of the 558 alleles (92%), identifying 29 distinct mutations in 11 of the 13 *PAH* exons. Mobility shifts were not detected in the case of 24 alleles (4.3%) despite scanning of the entire *PAH* coding sequence by DGGE. These mutations may have been either nucleotide substitutions localised deep within intronic sequences or in the flanking 5' and 3' untranslated regions of the *PAH* gene, or exonic deletions. In a further 21 alleles, mobility shifts localised to specific exons could not be characterised due to insufficient DNA template quantity or quality. In some cases, the age of the Guthrie card resulted in poor PCR product yields that precluded sequence analysis. Mutations were distributed across the coding sequence but occurred largely in the highly conserved region spanning exons 4 to 11 of the *PAH* gene.<sup>37</sup> The three most common mutations were R408W (41.0%), F39L (12.2%) and I65T (10.4%), together accounting for 63.6% of mutant alleles (Table 1). All other mutations were present at relative frequencies of less than 6.0%. These findings confirm the results of a previous study by O'Neill *et al.*<sup>16</sup> A similar situation was observed in the Northern Irish population where the same three mutations were found to account for 57.5% of mutant alleles, although at different individual relative frequencies.<sup>17</sup> While no novel mutations were detected in the study, this is the first report in north-western Europe of A447D, a mutation previously detected in the U.S. population.<sup>38</sup> SNP haplotypes<sup>14</sup>, minihaplotypes<sup>17</sup> and their associations with the predominant HPA mutations were investigated using a cohort of small nuclear families and unrelated homozygous individuals. The most common mutation, R408W, was associated with haplotype 1.8 and principally with the 242 bp allele of the short tandem repeat (STR) marker (Table 2). F39L was found to be associated with haplotype 1.8 and three STR alleles (238, 242 and 246 bp). I65T showed the greatest variation in haplotype associations, being associated mainly with haplotype 9.8 (but also with 1.8) and exhibiting a range of STR allele associations. These haplotype and minihaplotype associations are similar to those previously reported for Northern Ireland, western Scotland and south-western England.<sup>17,19,34,39</sup>

Relative allele frequencies were calculated for the eight most common mutations in the four provinces of Ireland

(Table 3). The provinces have been recognised as distinct regions for longer than the modern geopolitical subdivisions, having been established in the early historic period (400–1000 AD) with their boundaries fixed in the early modern period.<sup>40</sup> This, coupled with relatively low mobility within Ireland until recent times,<sup>41</sup> makes the provinces a reasonable basis upon which to look for evidence of stratification within the Irish population. Our data show that the relative frequencies of the most common mutations vary considerably between the provinces. For example, F39L showed maximal frequency in Leinster (the east coast), whereas I65T peaked in Ulster (the north/north-west). Perhaps most interestingly, R408W occurred with the highest relative frequency (55.8%) in Connacht, the most westerly province (Table 3).

R408W is known to have arisen by recurrent mutation on chromosomes of two distinct haplotype backgrounds: R408W-2.3 exhibits a west-to-east cline of relative frequency reaching its maximum in the Balto-Slavic region while R408W-1.8 exhibits an east-to-west cline peaking in Ireland.<sup>19,39,42,43</sup> Our data demonstrate that the east to west cline of R408W-1.8 continues across Ireland to peak in Connacht (Figure 1A). Spatial autocorrelation analysis<sup>24</sup> (Figure 1B) has determined that the data are consistent with a regional cline of R408W-1.8, indicating that genetic drift has not erased a pattern of genetic variation most likely established by human migration. Although selection can not be excluded in establishing this pattern, it is unlikely that it would have acted differently on R408W-1.8 and -2.3, given that they produce identical phenotypes.

It has been suggested that the R408W-1.8 mutation arose independently in Ireland and that Ireland acted as its centre of diffusion, based upon the observation of 'Celtic' *PAH* alleles in Scotland, Scandinavia, and the New World.<sup>17–19,44–46</sup> However, these observations are linked to population movements in the historic period, such as Irish settlement in Scotland and England in the 3rd to 5th centuries AD,<sup>40</sup> population movements associated with the Vikings in the 8th to 12th centuries AD<sup>47–50</sup> and Irish emigration in the last three centuries.<sup>44</sup> R408W-1.8 is likely to be an ancient mutation, given that it is associated with at least four STR allelic variants (*PAHdb*), one less than the I65T mutation which is believed to be Palaeolithic.<sup>18</sup> Therefore, the distribution of the R408W-1.8 allele in Europe is more likely to reflect earlier human dispersals. A more plausible hypothesis is that the cline of R408W-1.8 in north-western Europe reflects an older pattern of genetic variation substantially unaltered by the Neolithic dispersal of individuals of near-eastern origin carrying other alleles into Europe. In support of this, it should be noted that the cline of R408W-1.8 parallels a number of frequency clines in north-western Europe for alleles at unlinked loci, including the Q188R mutant allele in transferase-deficient galactosaemia,<sup>46</sup> blood group O<sup>51</sup> and Y chromosome haplogroup 1,<sup>52</sup> all of which reach maximal frequency in

Ireland (the latter two in Connacht). Where parallel allele frequency clines are observed for a number of unlinked genes, they are more likely to have been generated simultaneously by a single population migration process rather than by several independent processes of dispersal.<sup>23</sup> In this case the peak of R408W-1.8 frequency in Ireland, and of other markers in north-westerly areas least affected by the expansion of Neolithic populations, should be regarded as the genetic traces of the Palaeolithic Europeans.<sup>52–54</sup> Contrary therefore to previous views<sup>18</sup>, R408W-1.8 may indeed be a *PAH* mutation which links Ireland with Europe.

Some argue that Ulster should be considered separately since its history has been marked by distinct demographic forces including Ulster–Irish settlement in western Scotland (3rd to 5th Centuries AD), Viking settlement of and exchange between western Scotland and Ireland (8th to 12th Centuries AD), and immigration from Scotland in the Plantation of Ulster (17th Century AD).<sup>40,47,55,56</sup> To investigate the extent of Scottish admixture in Ulster, calculations were performed using a modification of the Bernstein model<sup>27</sup> and data for thirteen mutant *PAH* alleles in Scotland, Ulster (as the hybrid population) and the rest of Ireland (Table 4). The estimated proportion of Scottish genes in Ulster (46%) is consistent with Ulster having been a zone of significant Irish–Scottish admixture. In addition, a number of mutations previously observed at low frequencies ( $\leq 2.5\%$ ) in the Ulster population (T380M, R243Q, S273F, L194P and IVS8nt1) have been reported at comparable frequency in western Scotland.<sup>34</sup> None of these were detected by DGGE or PCR–RFLP assay in this study of the Republic of Ireland.

Scandinavian *PAH* mutations (F299C, R408Q, Y414C and G46S) accounted for 6.1% of the mutant alleles detected in this study while G272X, the most common Norwegian mutation, was not observed (Table 1). Mutations associated with the north-west coast of Norway (F299C, R408Q and Y414C) were more common than those associated with south-eastern Norway.<sup>35,57</sup> This supports the view that the Scandinavian contribution to the Irish gene-pool was significant and mediated largely by Viking incursions from north-western Norway.<sup>18</sup> Although the major direction of gene-flow was from Norway to Ireland, the possibility of flow in the opposite direction cannot be ruled out. Analysis of the geographic distribution of Scandinavian mutations within Ireland (Table 4) has shown that they account for the lowest proportion of mutations in Connacht (1.9%) compared to the other provinces (Munster 7.5%, Leinster 5.9%, Ulster 12.4%).

This study was undertaken to produce a detailed picture of the mutation spectrum underlying HPA and PKU in the Southern Irish population. By merging this with the data for Northern Ireland, the composite data-set yields insights into factors which have shaped the population of the island, underlining the utility of recessive disease-associated alleles in the study of human populations.

## Acknowledgements

We thank our colleagues Dr Donncha Dunican, Dr Peter McWilliam, Professor Stanley Monkhouse, Dr Daniel Bradley (Smurfit Institute of Genetics, Trinity College Dublin), Dr Barra O Donnabhain (Department of Archaeology, University College Cork) and Dr Linda Tyfield (University of Bristol, UK) for helpful discussions and for critical reading of the manuscript. We gratefully acknowledge financial support from the Research Committee of the Royal College of Surgeons in Ireland and from the Royal Irish Academy Millennium Project on 'Irish Origins: the genetic history and geography of Ireland' jointly funded with the National Millennium Committee of Ireland.

## References

- 1 Scriver CR: The Hyperphenylalaninaemias; in Scriver CR, Beaudet AL, Sly WS, Valle D (eds): *The Metabolic and Molecular Bases of Inherited Disease*. 7th edn. New York: McGraw-Hill Inc., 1995; pp 1015–1075.
- 2 Woo S, Lidsky A, Guttler F, Chandra T, Robson K: Cloned human phenylalanine hydroxylase gene permits prenatal and carrier detection of classical phenylketonuria. *Nature* 1983; **306**: 151–155.
- 3 Lidsky A, Robinson K, Thirumalachary C, Barker P, Ruddle F, Woo S: The PKU locus in man is on chromosome 12. *Am J Hum Genet* 1984; **36**: 527–533.
- 4 Novacki P, Byck S, Prevost L, Scriver C: The *PAH* Mutation Analysis Consortium Database: Update 1996. *Nucleic Acids Res* 1997; **25**: 139–142.
- 5 Scriver CR, Waters PJ, Sarkissian C *et al*: *PAHdb*: A locus-specific knowledgebase. *Hum Mutat* 2000; **15**: 99–104.
- 6 DiLella A, Kwok S, Ledley F, Marvit J, Woo S: Molecular structure and polymorphic map of the human phenylalanine hydroxylase gene. *Biochemistry* 1986; **25**: 749–753.
- 7 Woo S: Collation of RFLP haplotypes at the human phenylalanine hydroxylase (*PAH*) locus. *Am J Hum Genet* 1988; **43**: 781–783.
- 8 Goltsov AA, Eisensmith RC, Konecki DS, Lichter-Konecki U, Woo SL: Associations between mutations and a VNTR in the human phenylalanine hydroxylase gene. *Am J Hum Genet* 1992; **51**: 627–636.
- 9 Goltsov A, Eisensmith R, Naughten E, Jin L, Chakraborty R, Woo S: A single polymorphic STR system in the human phenylalanine hydroxylase gene permits rapid prenatal diagnosis and carrier screening for phenylketonuria. *Hum Molec Genet* 1993; **2**: 577–581.
- 10 Chakraborty R, Lidsky A, Daiger S *et al*: Polymorphic DNA haplotypes at the human phenylalanine hydroxylase locus and their relationship with phenylketonuria. *Hum Genet* 1987; **76**: 40–46.
- 11 Daiger S, Chakraborty R, Reed L *et al*: Polymorphic DNA haplotypes at the phenylalanine hydroxylase (*PAH*) locus in European families with phenylketonuria (PKU). *Am J Hum Genet* 1989; **45**: 310–318.
- 12 Sullivan SE, Moore SD, Connor JM *et al*: Haplotype distribution of the human phenylalanine hydroxylase locus in Scotland and Switzerland. *Am J Hum Genet* 1989; **44**: 652–659.
- 13 Konecki D, Lichter-Konecki U: The phenylalanine hydroxylase locus: current knowledge about alleles and mutations of the phenylalanine hydroxylase gene in various populations. *Hum Genet* 1991; **87**: 377–388.
- 14 Eisensmith R, Woo S: Updated listing of haplotypes at the human phenylalanine hydroxylase (*PAH*) locus. *Am J Hum Genet* 1992; **51**: 1445–1448.
- 15 Carter KC, Byck S, Waters PJ *et al*: Mutation at the phenylalanine hydroxylase gene (*PAH*) and its use to document population genetic variation: the Quebec experience. *Eur J Hum Genet* 1998; **6**: 61–70.
- 16 O'Neill C, Eisensmith RC, Croke D, Naughten ER, Cahalane SF, Woo SLC: Molecular analysis of PKU in Ireland. *Acta Paediatrica* 1995; **407**: 43–44.

- 17 Zschocke J, Graham C, Carson DJ, Nevin N: Phenylketonuria mutation analysis in Northern Ireland: a rapid stepwise approach. *Am J Hum Genet* 1995; **57**: 1311–1317.
- 18 Zschocke J, Mallory JP, Eiken HG, Nevin N: Phenylketonuria and the peoples of Northern Ireland. *Hum Genet* 1997; **100**: 189–194.
- 19 Eisensmith RC, Goltsov AA, O'Neill C *et al*: Recurrence of the R408W mutation in the Phenylalanine Hydroxylase locus in Europeans. *Am J Hum Genet* 1995; **56**: 278–286.
- 20 Guthrie R, Susi A: A simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. *Pediatrics* 1963; **32**: 338–343.
- 21 Guldborg P, Guttler F: 'Broad range' DGGE for single-step mutation scanning of entire genes: application to human phenylalanine hydroxylase gene. *Nucleic Acids Res* 1994; **22**: 880–881.
- 22 Rao V: Direct sequencing of polymerase chain reaction amplified DNA. *Anal Biochem* 1994; **216**: 1–14.
- 23 Bodmer W, Cavalli-Sforza LL: *The genetics of human populations*. San Francisco, USA: WH Freeman & Company, 1971.
- 24 Sokal RR, Oden NL: Spatial autocorrelation analysis in biology. *Biol J Linn Soc* 1978; **10**: 199–249.
- 25 Sokal RR: Ecological parameters inferred from spatial correlograms; in Patil GP, Rosenzweig M (eds): *Contemporary Quantitative Ecology and Related Econometrics*. Fairland, MD, USA: International Co-operative Publishing House, 1979, pp 167–196.
- 26 Oden NL: Assessing the significance of a spatial correlogram. *Geogr Anal* 1984; **16**: 1–16.
- 27 Bernstein F: Die geographische verteilung der blut-gruppen und ihre anthropologische bedeutung; in: *Comitato Italiano per lo studio dei problemi della popolazione*. Roma, Italia: Istituto Poligrafico dello Stato, 1931; pp 227–243.
- 28 Bertorelle G, Excoffier L: Inferring admixture proportions from molecular data. *Mol Biol Evol* 1998; **15**: 1298–1311.
- 29 Dworniczak B, Wedemeyer N, Horst J: PCR detection of the *Bgl*II RFLP at the human phenylalanine hydroxylase (*PAH*) locus. *Nucleic Acids Res* 1991; **19**: 1958.
- 30 Iyengar S, Seaman M, Deinard AS *et al*: Analyses of cross-species polymerase chain reaction products to infer the ancestral state of human polymorphisms. *DNA Sequence* 1998; **8**: 317–327.
- 31 Wedemeyer N, Dworniczak B, Horst J: PCR detection of the *Msp*I (Aa) RFLP at the human phenylalanine hydroxylase (*PAH*) locus. *Nucleic Acids Res* 1991; **19**: 1959.
- 32 Goltsov AA, Woo SL: Detection of the *Xmn*I RFLP at the human phenylalanine hydroxylase locus by PCR. Genbank accession number Z11537, 1992.
- 33 Zschocke J, Graham C, McKnight J, Nevin N: The STR system in the human phenylalanine hydroxylase gene: true fragment length obtained with fluorescent labelled PCR primers. *Acta Paed Suppl* 1994; **407**: 41–42.
- 34 Tyfield L, Stephenson A, Cockburn F *et al*: Sequence variation at the phenylalanine hydroxylase gene in the British Isles. *Am J Hum Genet* 1997; **60**: 388–396.
- 35 Eiken HG, Knappskog PM, Boman H *et al*: Relative frequency, heterogeneity and geographic clustering of PKU mutations in Norway. *Eur J Hum Genet* 1996; **4**: 205–213.
- 36 Mallolas J, Vilaseca MA, Campistol J *et al*: Mutational spectrum of phenylalanine hydroxylase deficiency in the population resident in Catalonia: genotype-phenotype correlation. *Hum Genet* 1999; **105**: 468–473.
- 37 Dworniczak B, Kalaydjieva L, Pankoke S, Aulehla-Scholz C, Allen G, Horst J: Analysis of exon 7 of the human phenylalanine hydroxylase gene: a mutation hot spot? *Hum Mutat* 1992; **1**: 138–146.
- 38 Guldborg P, Levy H, Handley W *et al*: Phenylalanine hydroxylase gene mutations in the United States: Report from the Maternal PKU collaborative study. *Am J Hum Genet* 1996; **59**: 84–94.
- 39 Byck S, Morgan K, Tyfield L, Dworniczak B, Scriver CR: Evidence for origin, by recurrent mutation, of the phenylalanine hydroxylase R408W mutation on two haplotypes in European and Quebec populations. *Hum Molec Genet* 1994; **3**: 1675–1677.
- 40 Charles-Edwards TM: *Early Christian Ireland*. Cambridge, UK: Cambridge University Press, 2000.
- 41 Relethford JH, Lees FC, Crawford MH: Population structure and anthropometric variation in rural Western Ireland: migration and biological differentiation. *Ann Hum Biol* 1980; **7**: 411–428.
- 42 Kalaydjieva L, Dworniczak B, Kucinskas V, Yurgeliavicius V, Kunert E, Horst J: Geographical distribution gradients of the major PKU mutations and the linked haplotypes. *Hum Genet* 1991; **86**: 411–413.
- 43 Giannattasio S, Jurgelevicius V, Lattanzio P, Cimbaliene L, Marra E, Kucinskas V: Phenylketonuria mutations and linked haplotypes in the Lithuanian population: origin of the most common R408W mutation. *Hum Hered* 1997; **47**: 155–160.
- 44 Treacy E, Byck S, Clow C, Scriver C: 'Celtic' phenylketonuria chromosomes found? Evidence in two regions of Quebec Province. *Eur J Hum Genet* 1993; **1**: 220–228.
- 45 Treacy, E: Phenylketonuria, a Celtic condition revisited. *Ir Med J* 1994; **87**: 100.
- 46 Murphy M, McHugh B, Tighe O. *et al*: The genetic basis of transferase-deficient galactosaemia in Ireland and the population history of the Irish Travellers. *Eur J Hum Genet* 1999; **7**: 549–554.
- 47 O Corrain D: Ireland, Wales, Man and the Hebrides; in Sawyer P. (ed): *The Oxford Illustrated History of the Vikings*. Oxford: Oxford University Press, 1997; pp 83–109.
- 48 Helgason A, Sigurdardottir S, Gulcher J, Ward R, Stefansson K: mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* 2000; **66**: 999–1016.
- 49 Helgason A, Sigurdardottir S, Nicholson J *et al*: Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *Am J Hum Genet* 2000; **67**: 697–717.
- 50 Helgason A, Hickey E, Goodacre S *et al*: mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 2001; **68**: 723–737.
- 51 Dawson, GWP: The frequencies of the ABO and Rh (D) blood groups in Ireland from a sample of 1 in 18 of the population. *Ann Hum Genet* 1964; **28**: 49–59.
- 52 Hill EW, Jobling MA, Bradley DJ: Y-chromosome variation and Irish origins. *Nature* 2000; **404**: 351–352.
- 53 Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press, 1994.
- 54 Barbuiani G, Bertorelle G: Genetics and the population history of Europe. *Proc Natl Acad Sci USA* 2001; **98**: 22–25.
- 55 Flanagan MT: The Vikings; in Loughrey P (ed): *The people of Ireland*. Belfast, Northern Ireland: Appletree, 1988; pp 55–84.
- 56 Mallory JP, McNeill TE: *The Archaeology of Ulster*. Belfast, N. Ireland: Institute of Irish Studies, 1989.
- 57 Apold J, Eiken H, Odland E *et al*: A termination mutation prevalent in Norwegian haplotype 7 phenylketonuria genes. *Am J Hum Genet* 1990; **47**: 1002–1007.

## Appendix

Assume that the overall mutant allele frequencies  $q_1$  and  $q_2$  can be estimated in the parental populations 1 and 2, for example from the square root of the disease incidence in these populations assuming Hardy–Weinberg equilibrium. The incidence of the disease in the hybrid group, however, is not known and the overall mutant allele frequency  $q_h$  cannot be estimated. Assume also that the relative frequency of a specific mutant, ie the frequency of that allele within a sample of mutants, is known for the parental populations ( $r_1$  and  $r_2$ ) and for the hybrid population ( $r_h$ ). The Berstein model<sup>27</sup> suggests that the frequency of an allele in the hybrid population ( $r_h q_h$  in our case) is a linear combination of the frequencies of the allele in the parental populations ( $r_1 q_1$  and  $r_2 q_2$ ), that is:

$$r_h q_h = m r_1 q_1 + (1-m) r_2 q_2 \quad (1)$$

where  $m$  and  $(1-m)$  are the relative contributions of parental population 1 and parental population 2 into the hybrid

respectively. In our case, however,  $q_h$  is unknown and  $m$  thus cannot be estimated directly from equation 1. A possible solution is to assume that the incidence of the disease in the hybrid group, and therefore  $q_h$ , is also a linear combination of the incidences observed in the parental populations. That is:

$$q_h = m q_1 + (1-m) q_2 \quad (2)$$

Combining equations (1) and (2) we obtain the following estimate of  $m$ :

$$m = (q_2(r_h - r_2)) / (q_1 r_1 - q_2 r_2 - r_h(q_1 - q_2)) \quad (3)$$

As suggested for other estimators of a ratio<sup>28</sup>, we propose to combine the information from different mutants averaging first the numerator and the denominator of (3) across mutations, and taking then the ratio of these averages. The estimate of  $m$  can be finally used to compute back an estimate of  $q_h$  from equation (2).