

RESEARCH ARTICLE

Genetic Diversity Within the R408W Phenylketonuria Mutation Lineages in Europe

Orna Tighe,^{1,2†} Donncha Dunican,^{1,2†} Charles O'Neill,¹ Giorgio Bertorelle,³ Diane Beattie,⁴ Colin Graham,⁴ Johannes Zschocke,⁵ Francesco Cali,⁶ Valentino Romano,⁷ Eva Hrabincova,⁸ Libor Kozak,⁸ Marina Nechyporenko,⁹ Ludmilla Livshits,⁹ Per Guldborg,¹⁰ Monika Jurkowska,¹¹ Cezary Zekanowski,¹² Belen Perez,¹³ Lourdes Ruiz Desviat,¹³ Magdalena Ugarte,¹³ Vaidutis Kučinskas,¹⁴ Per Knappskog,¹⁵ Eileen Treacy,¹⁶ Eileen Naughten,¹⁶ Linda Tyfield,¹⁷ Susan Byck,¹⁸ Charles R. Scriver,¹⁸ Philip D. Mayne,¹ and David T. Croke^{2*}

¹Department of Pathology and National Centre for Newborn Screening, The Children's University Hospital, Dublin, Ireland; ²Department of Biochemistry, The Royal College of Surgeons in Ireland, Dublin, Ireland; ³Dipartimento di Biologia, Università di Ferrara, Ferrara, Italy; ⁴Department of Medical Genetics, Queens University Belfast, Belfast City Hospital, Belfast, Ireland; ⁵Institute of Human Genetics, Heidelberg, Germany; ⁶Istituto di Ricovero e Cura a Carattere Scientifico per lo Studio del Ritardo Mentale e dell'Involuzione Cerebrale, Troina, Italy; ⁷Dipartimento di Biopatologia e Metodologie Biomediche, Università di Palermo, Palermo, Italy; ⁸Department of Biochemical and Molecular Genetics, Research Institute of Child Health, Brno, Czech Republic; ⁹Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Kiev, Ukraine; ¹⁰The John F. Kennedy Institute, Glostrup, Denmark; ¹¹Department of Medical Genetics, National Research Institute of Mother and Child, Warsaw, Poland; ¹²International Institute of Molecular and Cell Biology, Warsaw, Poland; ¹³Centro de Biología Molecular "Severo Ochoa," Universidad Autónoma de Madrid, Madrid, Spain; ¹⁴Human Genetics Center, Vilnius University, Vilnius, Lithuania; ¹⁵Department of Medical Genetics, Haukeland Hospital, Bergen, Norway; ¹⁶National Centre for Inherited Metabolic Disorders, The Children's University Hospital, Dublin, Ireland; ¹⁷Southmead Hospital and University of Bristol, Bristol, UK; ¹⁸Department of Human Genetics, McGill University, Montreal, Quebec, Canada

For the PKU Special Issue

The R408W phenylketonuria mutation in Europe has arisen by recurrent mutation in the human phenylalanine hydroxylase (PAH) locus and is associated with two major PAH haplotypes. 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 Connacht, the most westerly province of Ireland. Spatial autocorrelation analysis has demonstrated that the R408W-2.3 cline, like that of R408W-1.8, is consistent with a pattern likely to have been established by human dispersal. Genetic diversity within wild-type and R408W chromosomes in Europe was assessed through variable number tandem repeat (VNTR) nucleotide sequence variation and tetranucleotide short tandem repeat (STR) allelic associations. Wild-type VNTR-8 chromosomes exhibited two major cassette sequence organizations: (a1)₅-b3-b2-c1 and (a1)₅-b5-b2-c1. R408W-1.8 was predominantly associated with (a1)₅-b5-b2-c1. Both wild-type VNTR-3 and R408W-2.3 chromosomes exhibited a single invariant cassette sequence organization, a2-b2-c1. STR allele distributions associated with the cassette variants were consistent with greater diversity in the wild-type VNTR-8 lineage and were suggestive of different levels of diversity between R408W-1.8 and R408W-2.3. The finding of greater genetic diversity within the wild-type VNTR-8 lineage compared to VNTR-3 suggests that VNTR-8 may be older within the European population. However, in the absence of a more extensive STR data-set, no such conclusions are possible for the respective R408W mutant lineages. *Hum Mutat* 21:387–393, 2003. © 2003 Wiley-Liss, Inc.

KEY WORDS: phenylketonuria; PKU; phenylalanine hydroxylase; PAH; Europe; population genetics; VNTR; STR

DATABASES:

PAH – OMIM: 261600; GenBank: NM_000277 (mRNA); Swiss-Prot: P00439
<http://www.pahdb.mcgill.ca> (PAHdb)

*Correspondence to: David T. Croke, Department of Biochemistry, The Royal College of Surgeons in Ireland, 123, St. Stephen's Green, Dublin 2, Republic of Ireland. E-mail: dtcroke@rcsi.ie

†Drs. Tighe and Dunican contributed equally to this work.

Grant sponsors: Research Committee of the Royal College of Surgeons in Ireland; Health Research Board (Ireland); Royal Irish

Academy Millennium Project on 'Irish Origins: the genetic history and geography of Ireland' jointly funded with the National Millennium Committee of Ireland.

DOI 10.1002/humu.10195

Published online in Wiley InterScience (www.interscience.wiley.com).

INTRODUCTION

Phenylketonuria (PKU; MIM# 261600) is caused by mutation at the phenylalanine hydroxylase (*PAH*) locus on chromosome 12q23.1 [Woo et al., 1983; Lidsky et al., 1984]. The *PAH* locus contains a number of neutral polymorphisms useful in defining haplotypes, including: 7 single nucleotide polymorphisms (SNPs); a VNTR motif; and a STR polymorphism [Lidsky et al., 1984; Woo, 1988; Goltsov et al., 1992, 1993]. More than 400 *PAH* mutations, their associated *PAH* haplotypes, and geographic distributions are documented in the public domain database *PAHdb* [Scriver et al., 2000, 2003]. In common with other rare, highly penetrant pathogenic alleles, *PAH* mutations have been used to infer human population history through the history of specific mutant alleles: for example, the distribution of the R408W allele in north America and Australasia is associated with the European diaspora in the New World [Scriver et al., 1996; Scriver, 2001].

Among the most common *PAH* mutations is R408W. R408W is found at relative allele frequencies as high as 84% in Europe [Zschocke, 2003]. The R408W mutation (c.1222C>T; [DiLella et al. 1987]), a C to T transition in exon 12 of the *PAH* gene, results in the substitution of tryptophan for arginine at amino-acid residue 408 and is a null mutation associated with <0.3% of normal activity and a severe PKU phenotype [Kayaalp et al., 1997]. R408W is observed in Europe on chromosomes of two distinct *PAH* haplotypes (R408W-2.3 and R408W-1.8) and is most likely to be the result of recurrent mutation [DiLella et al., 1987; John et al., 1990; Byck et al., 1994; Eisensmith and Woo, 1992; Eisensmith et al., 1995]. The two mutations differ markedly in their geographic distribution in Europe. R408W-2.3 was shown to exhibit a gradient of relative allele frequency increasing eastwards and peaking in the Baltic region, variously interpreted as indicative of range expansions of the Slavonic and Kurgan peoples and associated with a pre-Indo-European founder population [Kalydjieva et al., 1991; Eisensmith et al., 1992, 1995; Giannattasio et al., 1997]. In contrast, the R408W-1.8 mutation was found to occur at high relative frequency in Ireland and neighboring populations, leading to the suggestion that Ireland had been a second center of diffusion of R408W [Treacy et al., 1993; O'Neill et al., 1994; Eisensmith et al., 1995; Zschocke et al., 1997]. More recently, it has been shown that the gradient of R408W-1.8 observed across northwestern Europe continues into Ireland and peaks in Connacht (the most westerly province), while spatial autocorrelation analysis demonstrated that the gradient is consistent with a localized cline of R408W-1.8 likely to have been established by human migration [O'Donnell et al., 2002].

The observation of different clinal distributions of the two R408W mutations in Europe as the result of human dispersal raises the possibility of using them to probe the history of the European population. To this end, spatial autocorrelation analysis was used to investigate the significance of the R408W-2.3 distribution and genetic diversity within the two R408W mutant lineages. To this end, spatial autocorrelation analysis was used to investigate the significance of the R408W-2.3 distribution and genetic diversity within the two R408W mutant 'lineages' and their corresponding wild-type alleles was used to gain some insight into their relative ages.

PATIENTS, MATERIALS, AND METHODS

Data on the relative allele frequency and *PAH* haplotype associations of the R408W mutation in 23 European regions and countries were drawn from the literature [Eisensmith et al., 1995; Kozak et al., 1997; Zschocke et al., 1997; Desviat et al., 1999; Giannattasio et al., 2001; Nechyporenko et al., 2001; O'Donnell et al., 2002; Kasnauskiene et al., 2003]. Absolute allele frequencies were estimated 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 [Bodmer and Cavalli-Sforza, 1971], using data on the incidence of PKU drawn from the literature and from collaborating laboratories. The geographic distribution of the R408W-2.3 mutant *PAH* allele in Europe was subjected to spatial autocorrelation analysis by means of the SAAP software (Exeter Software, Setauket, NY) package as described previously [Sokal and Oden, 1978; O'Donnell et al., 2002]. Spatial autocorrelation analysis measures the average level of genetic similarity between populations in particular geographic distance classes expressed in a correlogram (Moran's autocorrelation coefficient plotted as a function of distance). The shape of this correlogram reflects the underlying genetic mechanism. The overall significance of correlograms was assessed by the Bonferroni criterion [Oden, 1984].

Anonymized singleton DNA samples of known genotype from fifteen European populations (n = 182) were obtained with appropriate local ethical approval from the collaborating centers where they had been genotyped previously by standard methods: Republic of Ireland (n = 18); Northern Ireland (n = 16); Scotland (n = 8); England (n = 8); Wales (n = 4); Italy (n = 7); Spain (n = 14); Germany (n = 7); Denmark (n = 3); Faroe Islands (n = 3); Norway (n = 10); Czech Republic (n = 30); Poland (n = 19); Lithuania (n = 20); and Ukraine (n = 15). Those samples that were potentially informative for the present study (n = 148) were investigated to determine VNTR genotypes. These included samples from individuals of wild-type (n = 49), R408W homozygote (n = 75), compound heterozygote including R408W (n = 16), and R408W heterozygote (n = 8) genotypes. In practice, not all of the 148 samples proved to be fully informative (see below). For example, a proportion of the wild-type homozygotes were of VNTR-7/8 or -8/9 genotype and thus cassette sequences could not be obtained.

PAH VNTR sequences were amplified by PCR in a 50 μ l volume containing 50 to 100 ng of genomic DNA, 50 pmol of each primer (VNTR-U: GCTTGAAACTTGAAAGTTGC; VNTR-L: GGAAACTTAAGAATCCCATC; Golstov et al. [1992]), 200 μ M dNTPs, 50 mM KCl, 10 mM Tris HCl (pH 8.4), and 1.5 mM MgCl₂. Following an initial hot start

(95°C, 5 min), *Taq* DNA Polymerase was added (2U, Promega, Madison, WI) and samples were subjected to 34 cycles of 1 min at 95°C, 1 min at 55°C, 1 min at 72°C, and a final elongation step (72°C, 5 min). Aliquots (5 µl) of PCR product were electrophoresed on 12% acrylamide gels to confirm VNTR genotypes. Of the 148 samples, 116 (78%) were of VNTR 3/3, 3/8, or 8/8 genotype. The remaining 32 samples (wild-type homozygotes and carriers of mutant alleles other than R408W) were of VNTR-7/8 or -8/9 genotype and thus were intractable to the isolation of individual alleles for sequence analysis.

PCR product generated from samples homozygous for the 8/8 or 3/3 VNTR genotypes were ethanol precipitated for sequencing. PCR product generated from samples of the VNTR 3/8 heterozygote genotype were re-amplified and replicate samples were pooled before electrophoresis on a 1.5% low melting agarose gel (Sigma Aldrich, St. Louis, MO). The separated alleles were purified using a WizardTM PCR Prep. kit (Promega). DNA cycle sequencing was performed in forward and reverse orientations using the Big Dye Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) with primers VNTR-U and a nested reverse primer (VNTR-N: TAAAATTCACAAATACAA). Sequencing products were resolved on an ABI 310 automated fluorescent DNA sequencer (Applied Biosystems, Foster City, CA). Cassette sequences were identified according to the standard nomenclature [Byck et al., 1994]. VNTR sequences were determined for all 116 samples and their associations with wild-type and R408W alleles were assigned for 184 of the 232 alleles (79%). The remaining 48 samples were heterozygous for both PAH genotype (i.e., compound heterozygotes, carriers of R408W, and carriers of other mutations) and VNTR cassette sequence organization and, being phase-unknown, it was impossible to unambiguously determine the associations.

PAH tetranucleotide STR genotypes were determined according to the protocol of Zschocke et al. [1994] by PCR amplification and allele sizing by electrophoresis on an ALFexpress automated sequencer (Amersham Biosciences, Little Chalfont, UK). STR associations with VNTR genotypes were determined for 146 of 184 alleles (79%). Associations could not be assigned for the remaining 21% of alleles (n = 38) because they occurred in 19 phase-unknown individuals who were heterozygous for both VNTR cassette sequence organization and STR genotype.

RESULTS

The R408W-1.8 and R408W-2.3 mutant PAH alleles have different clinal distributions within Europe [Kalaydjieva et al., 1991; Eisensmith et al., 1992; Byck et al., 1994; Eisensmith et al., 1995; O'Donnell et al., 2002]. Using data on disease incidence, R408W relative frequency, and the proportions of the two R408W mutant lineages in a range of European populations (Table 1), the geographic distribution of the R408W-2.3 allele was subjected to spatial autocorrelation analysis producing a statistically significant correlogram ($p < 0.01$) consistent with a cline of R408W-2.3 in Europe (Fig. 1A and 1B). The shape of the correlogram is consistent with an allele frequency distribution resulting from human migration, with data points attaining significance at the $p < 0.05$ level at the extremes of the distance ranges analyzed (Fig. 1B).

TABLE 1. PKU Disease Incidence, Relative and Estimated Absolute Frequencies of R408W-1.8 and R408W-2.3 in the European Populations Surveyed

Population	Incidence ^a	R408W relative frequency	R408W-2.3 proportion	R408W-2.3 absolute frequency ^b	R408W-1.8 proportion	R408W-1.8 absolute frequency ^b
Connacht	4,500	0.558	0.00	0.000	1.00	0.832
Leinster	4,500	0.414	0.00	0.000	1.00	0.617
Munster	4,500	0.292	0.00	0.000	1.00	0.435
Ulster	4,500	0.284	0.09	0.038	0.91	0.385
Scotland	7,500	0.323	0.09	0.034	0.91	0.339
England	12,000	0.111	0.31	0.031	0.69	0.070
France	10,000	0.141	0.95	0.134	0.05	0.007
Germany	8,000	0.257	1.00	0.287	0.00	0.000
Denmark	12,000	0.182	0.93	0.155	0.07	0.012
NW Norway	18,600	0.200	0.75	0.110	0.25	0.037
SE Norway	12,800	0.110	0.25	0.024	0.75	0.073
Sweden	20,000	0.219	0.92	0.142	0.05	0.008
Iceland	10,000	0.032	1.00	0.032	0.00	0.000
Spain	10,000	0.007	0.00	0.000	0.00	0.000
Portugal	14,500	0.016	0.00	0.000	0.00	0.000
Italy	13,500	0.011	0.67	0.006	0.00	0.000
Hungary	8,800	0.486	1.00	0.518	0.00	0.000
Poland	8,000	0.622	0.99	0.688	0.00	0.000
Czech Republic	9,000	0.549	0.99	0.573	0.00	0.000
Bulgaria	15,000	0.350	0.95	0.271	0.00	0.000
Lithuania	9,000	0.734	1.00	0.774	0.00	0.000
Ukraine	8,300	0.570	1.00	0.626	0.00	0.000
Russia	10,000	0.565	1.00	0.565	0.00	0.000

^aIncidence data are expressed as one case per *n* live-births. Data sources: Eisensmith et al. [1995]; Kozak et al. [1997]; Zschocke et al. [1997]; Desviat et al. [1999]; Giannattasio et al. [2001]; Nechyporenko et al. [2001]; O'Donnell et al. [2002]; Kasnauskienė et al. [2003]; Prof. Gyoergy Fekete (personal communication).

^bEstimated absolute allele frequency values multiplied by 100.

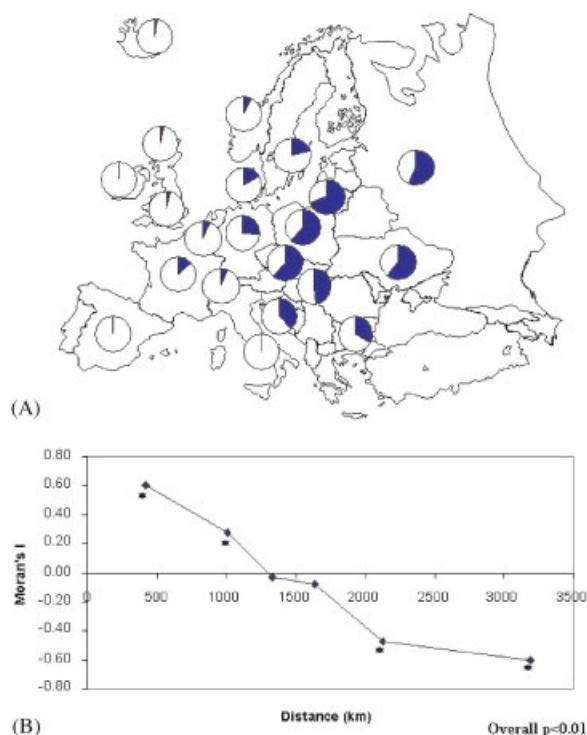


FIGURE 1. A cline of frequency of the R408W-2.3 phenylketonuria mutation across eastern Europe. **A:** Depiction of relative frequencies of R408W-2.3 in each of 21 European populations, represented by the shaded areas of the relevant pie-charts superimposed on the outline map. **B:** Presentation of a correlogram obtained by spatial autocorrelation analysis (using the SAAP software) of R408W-2.3 absolute allele frequencies for the 21 populations. Individual data points that attained significance at the $p < 0.05$ level are indicated by asterisks (*). The overall significance of the correlogram was $p < 0.01$, calculated by the Bonferroni criterion. A similar figure for R408W-H1.8 is provided in O'Donnell et al. [2002]. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

In order to begin to investigate the relative ages of R408W-1.8 and R408W-2.3, genetic diversity within the two mutant lineages and their underlying haplotypes in Europe was investigated by DNA sequence analysis of the VNTR motif in 184 alleles of wild-type, R408W-1.8, and R408W-2.3 genotypes [Goltsov et al., 1992; Byck et al., 1994]. VNTR-8 wild-type chromosomes ($n = 38$) exhibited two major cassette sequence organizations (Table 2): $(a1)_5$ -b3-b2-c1 (58%; 22 of 38) and $(a1)_5$ -b5-b2-c1 (40%; 15 of 38). A novel VNTR cassette, designated a4 (Table 3), was identified on a single wild-type VNTR-8 chromosome of cassette organization $(a1)_6$ -a4-c2. R408W-1.8 chromosomes ($n = 60$) exhibited an almost exclusive cassette sequence organization of $(a1)_5$ -b3-b3-c1 (98%; 59 of 60) with a single R408W-1.8 allele having an $(a1)_5$ -b3-b3-c1 cassette sequence organization. Both wild-type VNTR-3 chromosomes ($n = 20$) and R408W-2.3 mutant chromosomes ($n = 66$) exhibited a single invariant cassette sequence organization, a2-b2-c1. It is interesting to note that, among the wild-type heterozygous samples (carriers of R408W or another mutation), several other examples of the $(a1)_6$ -a4-c2 and $(a1)_5$ -b3-b3-c1 cassette sequence organizations were observed. Allelic associations could not be assigned unequivocally for these because of double heterozygosity at nucleotide positions 26 and 28 of cassettes 6 and 7 within the 8-cassette VNTR motif.

Associations between particular VNTR cassette organizations and STR alleles were also investigated in 146 (79%) of the 184 PAH alleles (Table 2, Fig. 2). For 19 individuals who were heterozygous at both the VNTR and STR loci, it was impossible to assign STR associations to individual VNTR genotypes due to the absence of phase information. The wild-type

TABLE 2. Genetic Diversity in the R 408W-1.8 and R 408W-2.3 Lineages in Europe at the Level of PAH VNTR Cassette Sequence and of VNTR-STR Allele Association

PAH genotype	n	VNTR-8 cassette sequence	Associated STR allele		n	VNTR-3 cassette sequence	Associated STR allele	
			n	size (bp)			n	size (bp)
Wild-type	22	$(a1)_5$ -b3-b2-c1	1	230	20	a2-b2-c1	3	234
			1	234			10	238
			2	238			3	242
			4	242			3	246
			2	246			1	250
	15	$(a1)_5$ -b5-b2-c1	2	230				
			3	234				
			1	238				
			2	242				
			1	246				
1	$(a1)_6$ -a4-c2	1	238					
		1	246					
		1	246					
R408W	0	$(a1)_5$ -b3-b2-c1	0		66	a2-b2-c1	53	238
			1	238			3	242
	59	$(a1)_5$ -b5-b2-c1	44	242			1	246
			4	246				
1	$(a1)_5$ -b3-b3-c1	ND						

Data from 15 European populations: Ireland, Northern Ireland, England, Wales, Scotland, Germany, Spain, Italy, Denmark, Faroe Islands, Norway, Poland, Czech Republic, Lithuania, and Ukraine. ND, not determined.

TABLE 3. PAH VNTR Cassette Sequence Variants

Cassette	Sequence
a1	CACATATATGTATATGCATATGTAC CGTATG
a2	CACATATATGTATATGCATAC CGTACGTATG
a3	CACATATATGTATATGCATATGTACATAT G
a4	CACATATATGTATATGCATATGTACATAG G
b1	CACATAIATGTATGTGCATATGTAC CGTATA
b2	CACATATATGTATGTGCATATGTACATAG G
b3	CACATAIATGTATGTGCATATGTAC CGTAGG
b4	CACATATATGTATGTGCATATGTAA AGTAGG
b5	CACATAIATGTATGTGCATATGTAC CGTATG
c1	CACATAIATGTATGTGCATATGTATGT ATA
c2	CACATAIATGTATATGCATATGTATGT ATA

Cassette sequence definitions are as originally specified by Byck et al. [1994] with a1 arbitrarily designated the canonical sequence; a4 is a novel cassette sequence (this work).

(a1)₅-b3-b2-c1 VNTR (n = 10) exhibited a unimodal distribution of STR alleles centered on the 242bp allele, while the wild-type (a1)₅-b5-b2-c1 VNTR (n = 8) exhibited an overlapping but possibly bimodal distribution. The R408W (a1)₅-b5-b2-c1 VNTR (n = 49) exhibited a unimodal STR allele distribution centered on the 242bp allele. The wild-type (n = 20) and R408W (n = 57) a2-b2-c1 VNTRs exhibited unimodal STR allele distributions both centered on the 238bp allele.

DISCUSSION

In Europe, the R408W mutation is observed on chromosomes of two major haplotype backgrounds. R408W-2.3 exhibits a west-to-east cline of relative frequency reaching its maximum in the Balto-Slavic

region. R408W-1.8 exhibits an east-to-west cline in northwestern Europe peaking in Connacht, the most westerly province of Ireland [Kalaydjieva et al., 1991; O'Donnell et al., 2002]. These clinal distributions are suggestive of geographic patterns resulting from human migration. The R408W-2.3 cline has been interpreted as a dispersal from an eastern European founder population variously associated with the range expansions of the Slavic and Kurgan peoples [Eisen-smith et al., 1992, 1995; Giannattasio et al., 1997]. While it has been suggested that the R408W-1.8 cline represents diffusion from a founder population in the British Isles [Treacy et al., 1993; Zschocke et al., 1997], the distribution may also be associated with the early colonization of the island from Europe [O'Donnell et al., 2002]. In an effort to further elucidate the history of the two R408W lineages in Europe, the significance of the R408W-2.3 cline was investigated. Spatial autocorrelation analysis of the R408W-2.3 distribution (Table 1) produced a correlogram consistent with a pattern of genetic variation most likely to have been established by human migration (Fig. 1). A previous analysis of the northwest European R408W-1.8 cline demonstrated that it too is likely to have been established by human migration [O'Donnell et al., 2002]. These analyses therefore confirm previous suggestions that the geographic distributions of R408W-1.8 and R408W-2.3 observed in Europe are the result of human dispersals.

To begin to gauge the relative ages of R408W-1.8 and R408W-2.3, genetic diversity within the mutant and corresponding wild-type lineages was investigated, focusing on cassette sequence variation in the VNTR motif located at the 3' end of the PAH locus and on the tetranucleotide STR marker located in intron 3 [Goltsov et al., 1992, 1993]. The PAH VNTR, originally detected as a triallelic *Hind*III RFLP, was characterized by Goltsov et al. [1992] who demonstrated that it consisted of tandemly repeated 30bp cassettes with some evidence of sequence divergence between them. Cassette sequences on wild-type and mutant chromosomes of VNTR-3 and VNTR-8 genotypes were investigated in 116 samples from a variety of European regional populations (Table 2). Wild-type VNTR-8 chromosomes were found to exhibit two major cassette organizations: (a1)₅-b3-b2-c1 and (a1)₅-b5-b2-c1 at relative frequencies of 58% and 40%, respectively, while wild-type VNTR-3 chromosomes were associated with a single invariant cassette organization, a2-b2-c1. This finding demonstrates greater genetic diversity within the VNTR-8 lineage as compared with VNTR-3 and may suggest that VNTR-8 is older than VNTR-3. R408W-1.8 chromosomes were predominantly associated with the (a1)₅-b5-b2-c1 cassette organization (98%), though a single R408W-1.8 chromosome was observed with an unusual (a1)₅-b3-b3-c1 organization. R408W-2.3

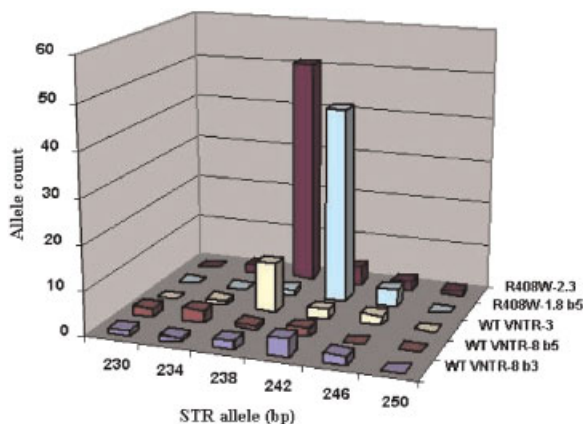


FIGURE 2. Tetranucleotide STR allele frequency distributions associated with the wild-type and mutant VNTR cassette sequence organizations: wild-type (a1)₅-b3-b2-c1 [WT VNTR-8 b3], wild-type (a1)₅-b5-b2-c1 [WT VNTR-8 b5], wild-type a2-b2-c1 [WT VNTR-3], R408W (a1)₅-b5-b2-c1 [R408W-1.8 b5], and R408W a2-b2-c1 [R408W-2.3]. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

chromosomes were associated exclusively with the a2-b2-c1 cassette organization. It is interesting to note that the VNTR-8 cassette variants observed in this study differ consistently from those reported by Byck et al. [1994], which were wild-type (a1)₅-b2-b3-c1 and R408W-1.8 (a1)₅-b5-b3-c1. Our initial hypothesis that Byck et al. [1994] may have inadvertently mislabeled cassette variant b3 as b2, and vice versa, was confirmed by resequencing of a number of the samples used in the Byck et al. [1994] study (data not shown).

The tetranucleotide STR marker, a complex repeat containing several regions of tandemly repeated [TCTA]_n sequences, has been shown to vary by gain or loss of 4bp repeat units and to exhibit distinct allele frequency distributions associated with wild-type and mutant PAH alleles [Goltsov et al., 1993]. Analysis of associations between VNTR cassette sequences and STR alleles yielded further evidence of genetic diversity (Table 2; Fig. 2). Wild-type VNTR-8 (a1)₅-b3-b2-c1 and (a1)₅-b5-b2-c1 chromosomes differed in their STR allele distributions. VNTR-8-b3 chromosomes exhibited a unimodal distribution (242 bp) whereas VNTR-8-b5 chromosomes exhibited an apparently bimodal distribution (234 bp and 242 bp). Wild-type VNTR-3 chromosomes had a unimodal distribution (238 bp). These observations are similar to those of Byck et al. [1994] but, given that wild-type allele numbers were low in both studies, must be interpreted cautiously. R408W-1.8-b5 and R408W-2.3 chromosomes both had unimodal STR allele distributions centered on the 242 bp and 238 bp alleles, respectively, though in both cases the STR allelic diversity observed was less than in the corresponding wild-type groups. These STR associations are similar to those reported by Byck et al. [1994].

VNTR cassette sequences and STR allelic associations indicate the presence of greater genetic diversity within the VNTR-8 lineage. Wild-type VNTR-8 chromosomes exist in two distinct cassette organizations, VNTR-8-b3 and VNTR-8-b5, differing by a single substitution (g>t) at nucleotide position 29 in the sixth cassette. The fact that these wild-type variants are observed at comparable frequencies (58% vs. 40%) and appear to have different associated STR allele distributions is suggestive of considerable age for both. The VNTR-3 lineage exhibited less diversity being associated with a single invariant cassette organization (a2-b2-c1) and with a unimodal STR allele distribution centered on the 238bp allele. If one assumes that single nucleotide substitution rates are unaffected by overall VNTR allele length, this relative lack of diversity suggests that the VNTR-3 lineage may have arisen more recently than VNTR-8 in the European population. R408W-1.8, associated primarily with the 242 bp allele (44 of 49 alleles; 90%), is likely to have arisen by mutation on a VNTR-8-b5 chromosome of 242bp STR genotype. The observa-

tion of a single R408W-1.8 allele with an unusual (a1)₅-b3-b3-c1 cassette organization may indicate that the R408W-1.8 mutation is of sufficient age to have accumulated some nucleotide sequence variation within the VNTR and, thus, may potentially be older than R408W-2.3. However, given the limited STR data presented here, firm conclusions regarding the relative ages of the mutant alleles must await a more extensive study involving the analysis of other informative STR markers within and flanking the PAH gene.

ACKNOWLEDGMENTS

We thank our colleagues Dr. Jonathon Flanagan, Prof. Stanley Monkhouse, and Dr. Dan Bradley (Smurfit Institute of Genetics, Trinity College Dublin) for helpful discussions and for critical reading of the manuscript. We thank Prof. Gyögy Fekete for information on the incidence of PKU in Hungary. We acknowledge constructive comments made by the anonymous reviewer.

REFERENCES

- Bodmer W, Cavalli-Sforza LL. 1971. The genetics of human populations. San Francisco: W.H. Freeman & Company. p. 39–70.
- Byck S, Morgan K, Tyfield L, Dworniczak B, Scriver CR. 1994. Evidence for origin, by recurrent mutation, of the phenylalanine hydroxylase R408W mutation on two haplotypes in European and Quebec populations. *Hum Mol Genet* 3:1675–1677.
- Desviat LR, Perez B, Gamez A, Sanchez A, Garcia MJ, Martinez-Pardo M, Marchante C, Boveda D, Badellou A, Arena J, Sanjurjo P, Fernandez A, Cabello ML, Ugarte M. 1999. Genetic and phenotypic aspects of phenylalanine hydroxylase deficiency in Spain: molecular survey by regions. *Eur J Hum Genet* 7:386–392.
- DiLella A, Marvit J, Brayton K, Woo S. 1987. An amino-acid substitution involved in phenylketonuria is in linkage disequilibrium with DNA haplotype 2. *Nature* 327:333–336.
- Eisensmith R, Woo S. 1992. Updated listing of haplotypes at the human phenylalanine hydroxylase (PAH) locus. *Am J Hum Genet* 51:1445–1448.
- Eisensmith RC, Okano Y, Dasovich M, Wang T, Guttler F, Lou H, Guldberg P, Lichter-Konecki U, Konecki DS, Svensson E, Hagenfeldt L, Rey F, Munnich A, Lyonnet S, Cockburn F, Connor JM, Pembrey ME, Smith I, Gitzelmann R, Steinmann B, Apold J, Eiken HG, Giovannini M, Riva E, Longhi R, Romano C, Cerone R, Naughten ER, Mullins C, Cahalane S, Ozalp I, Kekete G, Schuler D, Berencsi GY, Nasz I, Brdicka R, Kamaryt J, Pijackova A, Cabalska B, Boszkowa K, Schwartz E, Kalinin VN, Jin L, Chakraborty R, Woo SLC. 1992. Multiple origins for phenylketonuria in Europe. *Am J Hum Genet* 51:1355–1365.
- Eisensmith RC, Goltsov AA, O'Neill C, Tyfield LA, Schwartz EI, Kuzmin AI, Baranovskaya SS, Tsukerman GL, Treacy E, Scriver CR, Guttler F, Guldberg P, Eiken HG, Apold J, Svensson E, Naughten E, Cahalane SF, Croke DT, Cockburn

- F, Woo S. 1995. Recurrence of the R408W mutation in the phenylalanine hydroxylase locus in Europeans. *Am J Hum Genet* 56:278–286.
- Giannattasio S, Jurgelevicius V, Lattanzio P, Cimbalistiene L, Marra E, Kucinskas V. 1997. Phenylketonuria mutations and linked haplotypes in the Lithuanian population: origin of the most common R408W mutation. *Hum Hered* 47:155–160.
- Giannattasio S, Dianzani I, Lattanzio P, Spada M, Romano V, Cali F, Andria G, Ponzone A, Marra E, Piazza A. 2001. Genetic heterogeneity in five Italian regions: analysis of PAH mutations and minihaplotypes. *Hum Hered* 52:154–159.
- Goltsov AA, Eisensmith RC, Konecki DS, Lichter-Konecki U, Woo SL. 1992. Associations between mutations and a VNTR in the human phenylalanine hydroxylase gene. *Am J Hum Genet* 51:627–636.
- Goltsov A, Eisensmith R, Naughten E, Jin L, Chakraborty R, Woo S. 1993. 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.
- John SWM, Rozen R, Scriver CR, Laframboise R, Laberge C. 1990. Recurrent mutation, gene conversion or recombination at the human phenylalanine hydroxylase locus: evidence in French-Canadians and a catalogue of mutations. *Am J Hum Genet* 46:970–974.
- Kalaydjieva L, Dworniczak B, Kucinskas V, Yurgeliavicius V, Kunert E, Horst J. 1991. Geographical distribution gradients of the major PKU mutations and the linked haplotypes. *Hum Genet* 86:411–413.
- Kasnauskienė J, Giannattasio S, Lattanzio P, Cimbalistiene L, Kucinskas V. The molecular basis of phenylketonuria in Lithuania. *Hum Mutat* 21:398.
- Kayaalp E, Treacy E, Waters PJ, Byck S, Nowacki P, Scriver CR. 1997. Human phenylalanine hydroxylase mutations and hyperphenylalaninaemia phenotypes: a metanalysis of genotype-phenotype correlations. *Am J Hum Genet* 61:1309–1317.
- Kozak L, Blazkova M, Kuhrova V, Pijackova A, Ruzickova S, St'astna S. 1997. Mutation and haplotype analysis of phenylalanine hydroxylase alleles in classical PKU patients from the Czech Republic: identification of four novel mutations. *J Med Genet* 34:893–898.
- Lidsky A, Robinson K, Thirumalachary C, Barker P, Ruddle F, Woo S. 1984. The PKU locus in man is on chromosome 12. *Am J Hum Genet* 36:527–533.
- Nechyporenko MV, Kravchenko SA, Livshits LA. 2001. Mutation and VNTR-polymorphism analysis of phenylalanine hydroxylase gene. *Ukr J Biochem* 73:63–67.
- Oden NL. 1984. Assessing the significance of a spatial correlogram. *Geogr Anal* 16:1–16.
- O'Donnell KA, O'Neill C, Tighe O, Bertorelle G, Naughten E, Mayne PD, Croke D. 2002. The mutation spectrum of hyperphenylalaninaemia in the Republic of Ireland: the population history of the Irish revisited. *Eur J Hum Genet* 10:530–538.
- O'Neill C, Eisensmith RC, Croke D, Naughten ER, Cahalane SF, Woo SLC. 1994. Molecular analysis of PKU in Ireland. *Acta Paediatr* 407:43–44.
- Scriver CR, Byck S, Prevost L, Hoang L, and the PAH Mutation Analysis Consortium. 1996. The phenylalanine hydroxylase locus: a marker for the history of phenylketonuria and human genetic diversity. In: Chadwick D, Cardew G, editors. *Variation in the human genome*. CIBA Found Symposium 197. Chichester: J Wiley & Sons. p. 73–96.
- Scriver CR, Waters PJ, Sarkissian C, Ryan S, Prevost L, Cote D, Novak J, Teebi S, Nowacki PM. 2000. PAHdb: a locus-specific knowledgebase. *Hum Mutat* 15:99–104.
- Scriver CR. 2001. Human genetics: lessons from Quebec populations. *Annu Rev Genomics Hum Genet* 2:69–101.
- Scriver CR, Hurtubise M, Konecki D, Phommavanh M, Prevost L, Erlandsen H, Stevens R, Waters PJ, Sarkissian C, McDonald D, Ryan S. 2003. PAHdb: updates on a locus-specific knowledgebase. *Hum Mutat* 21:333–344.
- Sokal RR, Oden NL. 1978. Spatial autocorrelation analysis in biology. *Biol J Linn Soc Lond* 10:199–249.
- Treacy E, Byck S, Clow C, Scriver C. 1993. 'Celtic' phenylketonuria chromosomes found? Evidence in two regions of Quebec Province. *Eur J Hum Genet* 1:220–228.
- Woo S, Lidsky A, Guttler F, Chandra T, Robson K. 1983. Cloned human phenylalanine hydroxylase gene permits prenatal and carrier detection of classical phenylketonuria. *Nature* 306:151–155.
- Woo S. 1988. Collation of RFLP haplotypes at the human phenylalanine hydroxylase (PAH) locus. *Am J Hum Genet* 43:781–783.
- Zschocke J. 2003. Phenylketonuria mutations in Europe. *Hum Mutat* 21:345–356.
- Zschocke J, Graham C, McKnight J, Nevin N. 1994. The STR system in the human phenylalanine hydroxylase gene: true fragment length obtained with fluorescent labelled PCR primers. *Acta Paediatr (Suppl)* 407:41–42.
- Zschocke J, Mallory JP, Eiken HG, Nevin N. 1997. Phenylketonuria and the peoples of Northern Ireland. *Hum Genet* 100:189–194.