npg

ARTICLE

Genetic diversity patterns at the human clock gene *period 2* are suggestive of population-specific positive selection

Fulvio Cruciani[1], Beniamino Trombetta[1], Damian Labuda[2], David Modiano[3], Antonio Torroni[4], Rodolfo Costa[5] and Rosaria Scozzari*[,1]

[1]Dipartimento di Genetica e Biologia Molecolare, Sapienza Università di Roma, Rome, Italy; [2]Centre de Recherche, Hôpital Sainte-Justine, Montréal, Quebec, Canada; [3]Dipartimento di Scienze di Sanità Pubblica, Sezione di Parassitologia, Sapienza Università di Roma, Rome, Italy; [4]Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy; [5]Dipartimento di Biologia, Università di Padova, Padua, Italy

Period 2 (*PER2*) is a key component of the mammalian circadian clock machinery. In humans, genetic variation of clock genes or chronic disturbance of circadian rhythmicity has been implied in the onset of several phenotypes, ranging from periodic insomnias to advanced or delayed sleep phases, to more severe disorders. Peculiar geographic diversity patterns in circadian genes might represent an adaptive response to different light/dark cycles or environmental changes to which different human populations are exposed. To investigate the degree and nature of *PER2* gene variation in human populations of different geographic origin, and its possible correlation with different latitudes, we sequenced a 7.7 kb portion of the gene in 20 individuals worldwide. In total, 25 variable sites were identified. The geographic distribution of haplotypes defined by five polymorphic sites was analyzed in 499 individuals from 11 populations from four continents. No evidence for latitude-driven selective effects on *PER2* genetic variability was found. However, a high and significant difference in the geographic distribution of *PER2* polymorphisms was observed between Africans and non-Africans, suggesting a history of geographically restricted natural selection at this locus. In support of this notion, we found several signals for selection in the sequences. The putative selected haplotype showed a recent coalescent age (8.7 Kyr), and an unusually high frequency in non-African populations. Overall, these findings indicate that a human clock-relevant gene, *PER2*, might have been influenced by positive selection, and offer preliminary insights into the evolution of this functional class of genes.
European Journal of Human Genetics advance online publication, 25 June 2008; doi:10.1038/ejhg.2008.105

Keywords: *PER2* phylogeography; positive selection; latitude; human population genetics; nucleotide diversity; clock genes

*Correspondence: Professor R Scozzari, Dipartimento di Genetica e Biologia Molecolare, Sapienza Università di Roma, P.le Aldo Moro 5, Rome 00185, Italy.
Tel: +39 06 49912826 (office) and +39 06 49912924 (lab);
Fax: (39) 06 4456866; E-mail: rosaria.scozzari@uniroma1.it*

Introduction
In most living organisms, from cyanobacteria to fungi, plants and animals, many physiological and behavioral processes show rhythmic variation with a period close to 24 h (circadian rhythms). In mammals, daily rhythms are controlled by a central pacemaker localized in the suprachiasmatic nuclei (SCN) of the hypothalamus (reviewed in Ko and Takahashi[1]). A light signal transduction pathway from the retina through the retino-hypothalamic

tract adjusts the central pacemaker to external cues such as the daily light/dark cycle. In turn, the SCN master clock coordinates the phasing of multiple peripheral circadian oscillators present in many tissues/organs.[2]

Alterations in the circadian system have been shown to be responsible for disturbances in the sleep–wake cycle; it is also becoming more evident that long lasting or chronic disruption of circadian rhythmicity or of the coordination between SCN and peripheral clocks could have profound effects on general health and well-being, leading to pathologies not directly related to disturbances in daily rhythms (reviewed in Maywood *et al*[3] and Ebisawa[4]).

The molecular basis of the mammalian circadian clock has been investigated in the past few years[5] and has greatly profited from genetic studies in *Drosophila melanogaster*, as insects and mammals share orthologous genetic components of the circadian machinery. However, despite the relevance of circadian genes on a wide spectrum of biological processes, the degree and nature of clock-related gene variation in 'normal' human populations from different origins, and the possible correlation with diverse environmental conditions, has received scant attention.[6,7] Systematic spatial distributions of specific haplotypes could represent adaptive responses to the different photoperiodic and seasonally rhythmic environmental changes to which different human populations are exposed.

Period 2 (*PER2*) is a key component of the mammalian circadian clock machinery. Mutations in this gene have been associated with a wide spectrum of phenotypes that include not only clock-related disorders, such as those associated with the sleep–wake cycle,[4] but also other diseases such as tumors,[8–10] cardiovascular dysfunctions[11] and enhanced alcohol intake.[12]

In this study, we analyze the extent of diversity at the *PER2* gene in a worldwide human sample. Our findings suggest the involvement of positive selection in the evolutionary history of this gene.

Materials and methods
Samples
A total of 20 unrelated individuals from five continents (5 Africans, 4 Europeans, 4 Asian, 5 Native Americans and 2 from Oceania) were analyzed by sequencing (sequencing set). A total of 499 unrelated individuals from Africa (43 Amhara from Ethiopia, 41 Kung from South Africa, 47 Mossi from Burkina Faso), Europe (59 from southern Spain, 54 Sardinians, 41 Danes), Asia (50 Pakistani, 37 Chinese Li, 49 Yakuts from Siberia) and the Americas (45 Chipewyans from Canada, 33 Maya from Mexico; Table 1) were genotyped at five polymorphic sites (genotyping set). All DNAs were obtained as previously described ([13] and references therein;[14,15] and the ALlele FREquency Database at http://alfred.med.yale.edu/alfred/index.asp). Appropriate informed consent was obtained from all donors.

Table 1 Frequency distribution of *PER2* five-site haplotypes in 11 populations

	Population[a]										
		Africans			Europeans			Asian			Native Americans
	Amhara (43)	Kung (41)	Mossi (47)	S. Spaniards (59)	Sardinians (54)	Danes (41)	Pakistani (50)	Chinese Li (37)	Yakut (49)	Chipewyans (45)	Maya (33)
Haplotype[b]	8°N-38°E	29°S-25°E	12°N-2°W	37°N-5°W	41°N-9°E	55°N-10°E	26°N-67°E	19°N-110°E	65°N-140°E	59°N-109°W	20°N-90°W
Common haplotypes (> 1%)											
del C T G C (I)	0.01	0.01	0.01	0.04	0.02	0.00	0.02	0.00	0.11	0.00	0.00
del C T A C (II)	0.24	0.07	0.09	0.48	0.50	0.59	0.43	0.45	0.32	0.37	0.17
del C T A T (III)	0.16	0.01	0.00	0.14	0.19	0.10	0.19	0.27	0.14	0.36	0.41
ins T T G C (IV)	0.29	0.73	0.61	0.16	0.07	0.07	0.12	0.20	0.15	0.11	0.02
ins T C G C (V)	0.07	0.00	0.01	0.10	0.12	0.12	0.06	0.03	0.11	0.10	0.39
ins C T G C (VI)	0.12	0.04	0.09	0.05	0.07	0.06	0.16	0.04	0.15	0.02	0.00
ins T T A C (VII)	0.00	0.07	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rare haplotypes (< 1%)											
del T T G C (VIII)	0.02	0.05	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00
del T T A C (IX)	0.03	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
del T T A T (X)	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
del T C G T (XI)	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
del C C G C (XII)	0.01	0.00	0.00	0.02	0.01	0.00	0.00	0.01	0.00	0.03	0.00
del C C A C (XIII)	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
del C C A T (XIV)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
ins T T G T (XV)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
ins T T A T (XVI)	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
ins T C A C (XVII)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
ins C T G T (XVIII)	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
ins C T A C (XIX)	0.01	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00
ins C T A T (XX)	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00

[a]Numbers of individuals are in parentheses. Latitude and longitude at which populations live are also indicated.
[b]Allelic state at the polymorphic sites *PS1–PS5* is indicated in the order.

Sequencing

Overall 7.7 kb from the *PER2* (MIM no. 603426) gene in 2q37.3 were sequenced for each of the 20 subjects in the 'sequencing set'. About 2.2 kb of the sequenced region were from 11 of the 23 exons of the *PER2* gene, whereas the remaining 5.5 kb were from flanking intronic sequences and a portion of the 5′ region containing the promoter (Figure 1). PCR and sequencing primers were designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Purified PCR products were sequenced in both directions using the BigDye Terminator Cycle Sequencing Kit with AmpliTaq DNA polymerase (Applied Biosystems, Foster City, CA, USA) and an internal or PCR primer. Chromatograms were aligned and analyzed for mutations using Sequencher 3.0 (Gene Codes Corporation, Ann Arbor, MI, USA).

Polymorphism selection and genotyping

Among 25 variant sites identified by sequencing (Figures 1 and 2), we selected polymorphisms showing a frequency > 10% to be analyzed in the 'genotyping set'. Four polymorphic sites (*PS1-4*), were selected using this criterion (see column 3 in Figure 2). A fifth polymorphism in exon 23 (*PS5*, rs934945) was selected from the dbSNP build 126 (http://www.ncbi.nlm.nih.gov/projects/SNP/). Before genotyping, polymorphic sites (*PS1-5*) were PCR amplified using the primer pairs and methods reported in Table 2. For each polymorphism, a region flanking the
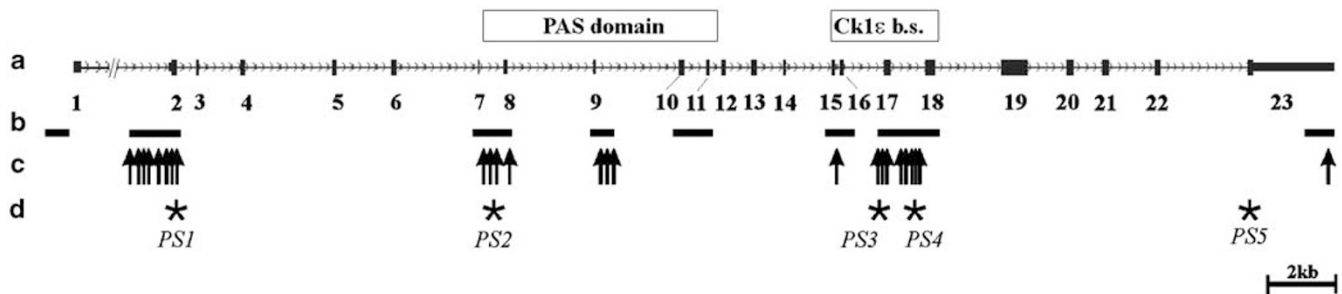

Figure 1 Schematic diagram of the human *period 2* (*PER2*) gene. (**a**) Exon-intron structure of the *PER2* gene. Exons are numbered and depicted as solid bars. The PAS domain and the CK1ε-binding site are also indicated. (**b**) Regions of the gene sequenced in 20 individuals. (**c**) Location of the 25 variant sites identified by sequencing (arrows). (**d**) Location of the five polymorphic sites (*PS1–PS5*) analyzed in 11 populations.

Mutation	Position	Selected polymorphisms	dbSNP	Africa	Europe	Asia	Americas	Oceania
Intron 1 C•T	238851948		-					
Intron 1 A•T	238851897		-					
Intron 1 G•A	238851867		-					
Intron 1 G•A	238851697		-					
Intron 1 A•G	238851484		-					
Intron 1 insGA	238851467		-					
Intron 1 del CA(GA)₃	238851459	PS1	-					
Exon 2 C•G	238851328		rs2304672					
Intron 7 C•T	238841997		-					
Intron 7 G•A	238841965		-					
Intron 7 A•G	238841812	PS2	rs11894535					
Intron 8 C•T	238841391		-					
Intron 9 C•G	238838644		rs10201361					
Intron 9 G•A	238838592		-					
Intron 9 G•A	238838433		rs28434027					
Intron 15 C•T	238831870		-					
Exon 17 A•G	238830402	PS3	rs2304669					
Exon 17 G•A	238830375		rs2304670					
Exon 17 G•A	238830372		rs2304671					
Intron 17 G•T	238830273		-					
Intron 17 C•T	238830272		-					
Intron 17 A•G	238829983		rs7579064					
Intron 17 C•T	238829867	PS4	rs6431590					
Intron 17 T•G	238829497		-					
Exon 23 T•G	238817614		-					

Figure 2 Summary of the mutations identified in the human *period 2* (*PER2*) gene. Only synonymous substitutions were found in the coding part of the gene (first column). The position of the mutations on chromosome 2 is reported (March 2006 assembly of the human genome; http://genome.ucsc.edu/) (second column). The polymorphic sites chosen for the haplotype frequency distribution analysis in 11 populations are indicated (third column). The SNP identifier is reported if present in the dbSNP build 126 (fourth column). Each sub-column below Africa, Europe, Asia, Americas and Oceania represents one individual. Genotypes at variable sites are color coded. Homozygotes for the derived allele are black, heterozygotes are gray, and homozygotes for the ancestral allele are white. The ancestral allele at each locus was estimated on the basis of the *Pan troglodytes* sequence.

4

Table 2 Oligonucleotide primers used for the PCR amplification of the *PER2* polymorphic sites *PS1–PS5*

Polymorphic site	Primer FOR	Primer REV	Type of polymorphism
PS1	5′-CATTCCCTGGGCAACAGA	5′-CAGGTAGAGCACCCATTTGAC	Length polymorphism
PS2	5′-CAGCACCCCTTTTCTGTC	5′-TGATGTGTACCTCTGCTCTG	RFLP (*Bse*DI)
PS3	5′-GGGTGTCGGTTTCTCATCT	5′-GGCTTCTTGTCTCCCACA	RFLP (*Bsp*143II)
PS4	5′-GCCTGGGGGTTCTTTCCTCT	5′-ACACTCAGCCACCAATGCGaTC[a]	RFLP (*Pvu*I)
PS5	5′-ACGTCTGCTCTTCGATCCTGT	5′-GTAACAACTCAGATCTCAAGTTTG	RFLP (*Bam*HI)

Abbreviations: FOR, forward; REV, reverse.
[a]Lower case, non-complementary base that generates a restriction site for *Pvu*I.

Table 3 Analysis of molecular variance for 20 *PER2* haplotypes worldwide

Category (no. of populations)	Φ_{ST} (P)[a]	Φ_{CT} (P)[a]	Φ_{SC} (P)[a]
Three 'latitudinal groups'[b]			
Low (4)	0.20 ($<10^{-3}$)		
Intermediates (4)	0.21 ($<10^{-3}$)		
High (3)	0.04 ($<10^{-2}$)		
Overall (11)	0.15 ($<10^{-3}$)	−0.02 (0.60)	0.17 ($<10^{-3}$)
Four 'geographic groups'			
Africa (3)	0.16 ($<10^{-3}$)		
Europe (3)	0.00 (0.58)		
Asia (3)	0.02 (0.01)		
Americas (2)	0.06 (0.01)		
Overall (11)	0.17 ($<10^{-3}$)	0.13 (0.02)	0.05 ($<10^{-3}$)
Two 'geographic groups' (11)[c]	0.26 ($<10^{-3}$)	0.22 (0.01)	0.05 ($<10^{-3}$)
Two 'geographic groups' (10)[d]	0.35 ($<10^{-3}$)	0.33 ($<10^{-2}$)	0.03 ($<10^{-3}$)

[a]P-values for the Φ statistics were obtained through 1000 permutations of individuals into populations (for Φ_{ST}), individuals within groups (for Φ_{SC}) and populations across groups (for Φ_{CT}).
[b]Low: Amhara, Mossi, Chinese Li, Maya; Intermediate: Kung, Southern Spaniards, Sardinians, Pakistani; High: Danes, Yakut, Chipewyans (see also Table 1).
[c]Africans and non-Africans.
[d]As above, with Amhara excluded.

orthologous site in the chimpanzee was also sequenced. No differences at these sites were found with respect to the published chimp sequence.[16] The ancestral/derived state at each position was assessed depending on whether or not a correspondence with the chimp sequence was observed.

Data analysis
Population genotype data for the *PS1-5* polymorphic loci were used to obtain maximum-likelihood haplotype frequencies using the expectation-maximization algorithm[17] implemented in the Arlequin 3.0 software.[18]

Lewontin's D' normalized measure of disequilibrium[19] was used to estimate pair-wise linkage disequilibrium between five polymorphic sites in 11 populations. Haplotype-based pair-wise Φ_{ST} values were calculated using Arlequin, version 3.0. The significance of the correlation between Φ_{ST} and latitudinal distance matrices was evaluated by the Mantel's test implemented in Arlequin 3.0. The same program was used for the analysis of molecular variance (AMOVA[20]). Two hierarchical levels (individuals into populations and populations into groups) were considered and two different grouping schemes were used

(see Table 3). As a measure of molecular distance between haplotypes, we used the number of differences in the allelic state at the five polymorphic sites considered (Table 1, column 1).

Arlequin 3.0 was also used for locus-by-locus F_{ST} calculations. To obtain an empirical distribution of F_{ST} values we queried the HapMap Phase I database (http://www.hapmap.org/) and extracted the allele frequencies of 54 895 SNPs located in a telomeric region of chromosome 2 (about 42.9 Mb; chr2:200 000 000–242 951 149) containing *PER2*. We chose this region rather than the whole genome to minimize the effects of F_{ST} heterogeneity among different genomic regions.[21] SNPs that were monomorphic for the same allele in any pair of populations were removed from the analysis.

Nucleotide diversity (π), and its sampling variance, was calculated according to Nei.[22]

The following neutrality test statistics were used: Tajima's D,[23] Fu and Li's D and F[24] and Fay and Wu's H.[25]

The ρ statistic[26,27] was used to obtain the time to the most recent common ancestor (TMRCA) for the chromosomes carrying the derived allele at the polymorphic sites

PS1, *2* and *4*. Due to the relatively low number of sequences available and the high number of singletons observed, only chromosomes from individuals homozygous for the $PS1_{del}$, $PS2_C$, $PS4_A$ alleles were considered (seven subjects, see Figure 2). The ρ-value was converted into years by dividing by the mutation rate. This latter – one mutational event every 121 thousand years (Kyr) for the entire 7.7 kb sequenced region – was obtained by comparing the human *PER2* sequence with the orthologous chimp sequence (http://genome.ucsc.edu/, March 2006 assembly) and by considering an approximate human-chimp divergence time of 6 Myr.[28]

Phylogenetic relationships among the seven most common *PER2* haplotypes were obtained by using the median-joining procedure[29] through the use of the network 4.2 program (http://www.fluxus-engineering.com).

The mean and the mean square age of a neutral allele with a given frequency in a population of effective size $N_e = 10\,000$[30] were obtained using formulas (13) and (14), respectively, as given by Kimura and Ohta.[31]

Results
Pattern of allele/haplotype frequency distribution
Five polymorphic sites (*PS1–PS5*, Figure 1d), selected either by sequencing or from the dbSNP, were analyzed in a sample of 499 subjects belonging to 11 populations from four continents (see Materials and methods section). For all the polymorphisms, the African populations showed the highest frequency of the ancestral allele, with the alleles at two sites (*PS3* and *PS5*) being very near to fixation in the Kung from South Africa and the Mossi from western Africa (Table 1). By contrast, the derived allele at three sites (*PS1*, *PS2* and *PS4*) was the most common one for the non-African populations. Out of 20 haplotypes (haplotypes I–XX, Table 2), the 7 most frequent haplotypes (I–VII) accounted for about 96.1% of the chromosomes in the whole sample. The relatively low number of high frequency haplotypes is the consequence of the high degree of pair-wise linkage disequilibrium observed (Supplementary Table 1). Five of the most common haplotypes (haplotypes II, III, IV, V and VII) showed an uneven geographic distribution, with pronounced differences in their frequencies between Africans and non-Africans (Table 1 and Figure 3). The Amhara population from Ethiopia showed haplotype frequencies that were intermediate between those of the sub-Saharan African and non-African populations, likely as a result of gene flow from western Asia.[32,33]

To evaluate whether the observed pattern of inter-population haplotypic diversity at the *PER2* gene has been shaped by variation in clock-relevant environmental conditions, such as the seasonal profile of photoperiod and/or temperature, we employed two related approaches by taking advantage of the fact that both these potential
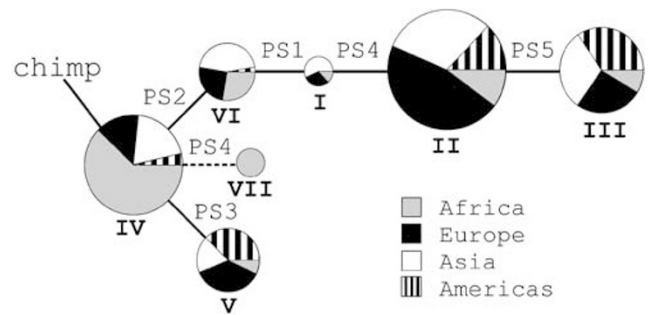

Figure 3 Median joining network of human *period 2* (*PER2*) haplotypes. The network was generated using the network 4.2 program[29] (http://www.fluxus-engineering.com). The most common *PER2* haplotypes (haplotypes I–VII, frequency > 1%) are shown. Each haplotype is represented by a circle and the area of the circle is proportional to the haplotype frequency. Mutations are indicated by solid lines, with the name of the mutation above the line. A dashed line denotes a possible recurrent mutational event at *PS4*. It should be noted that the 'African' haplotype VII could most likely have also been generated by a recombinational event between haplotypes II and IV, both found at relatively high frequencies in Africa.

selective factors vary with latitude. The first one was to correlate haplotype-based pair-wise Φ_{ST} distances with absolute latitudinal distances between populations. No significant correlation between the two distance matrices was found ($r = -0.07$, $P = 0.40$) (see Supplementary Figure 1). In the second approach, Φ statistics[20] were calculated to assess the intra- and intergroup variation, after the populations were grouped by a latitudinal criterion (Table 3). Whereas a high and statistically significant level of intragroup genetic differentiation was found ($\Phi_{SC} = 0.17$, $P < 10^{-3}$), no differentiation was observed among the three 'latitudinal' groups of populations ($\Phi_{CT} = 0.02$, $P = 0.60$). Overall these results provide no signal of differential selection resulting from latitude variation.

When the 11 populations were grouped into four continental groups, the inter-group diversity was high and statistically significant ($\Phi_{CT} = 0.13$, $P = 0.02$). An even higher level of differentiation was detected when the populations were grouped as Africans and non-Africans ($\Phi_{CT} = 0.22$, $P < 0.01$) and when the likely admixed Amhara sample was removed from the analysis ($\Phi_{CT} = 0.33$, $P < 10^{-3}$). Both geographically restricted positive selection and genetic drift could be the determinants of differentiation among populations.[34] However, natural selection is expected to act on specific loci, thus, by sampling a large number of SNPs throughout the genome, it is in principle possible to detect selection by identifying candidate genes as outliers in the extreme tails of the F_{ST} distribution.[34] By taking advantage of the large number of SNPs that have been genotyped by the HapMap Consortium,[35] the observed degree of differentiation at the five polymorphic loci here analyzed was compared with an empirical distribution of F_{ST} calculated from the HapMap allele

frequency data (see Materials and methods section). Since the F_{ST} values strongly depend on the populations analyzed, we tried to minimize this confounding effect by choosing populations that are geographically close to those analyzed by the HapMap. We used the Mossi, the Chinese Li and the Danes as proxies for the HapMap Yoruba, Chinese and CEPH populations, respectively. On the whole, the observed F_{ST} values at the five *PER2* polymorphic sites analyzed here tend to cluster toward the highest values of the empirical distribution of F_{ST} values both by comparing Africans *vs* Europeans (Mann–Withney *U*-test, $P = 0.016$) and Africans *vs* Asian (Mann–Withney *U*-test, $P = 0.036$). In particular, *PS1* and *PS2* showed F_{ST} values lying in the top 5% tail of the empirical F_{ST} value distribution in both comparisons (Africans *vs* Europeans, $F_{ST} = 0.34$ and 0.32, respectively; and Africans *vs* Asian, $F_{ST} = 0.35$ and 0.33, respectively; Figure 4). *PS4* also showed high F_{ST} values, although lower than *PS1* and 2 (Africans *vs* Europeans, $F_{ST} = 0.17$ and Africans *vs* Asian, $F_{ST} = 0.18$; Figure 4). Importantly, the pair-wise F_{ST} estimates obtained from a HapMap-based analysis of a region of about 150 kb flanking the *PER2* gene (two genes and one gene, 5′ and 3′ of *PER2,* respectively) were all found to lie below the top 5% line of the F_{ST} distribution. The marked population differentiation of *PER2* is accompanied by high frequencies of the derived alleles at *PS1, 2* and *4* in non-Africans, suggesting a history of geographically restricted natural selection at this locus.

Pattern of sequence diversity

To better understand the pattern of genetic diversity at *PER2*, we went deeper into the analysis of its sequence variation (Figure 2). The allele frequency spectrum was strongly skewed as the large majority (17 out of 23) of the single nucleotide substitutions observed were singletons, one was observed twice, and only five were observed more than twice. The observed nucleotide diversity was low ($\pi = 3.3 \times 10^{-4} \pm 0.6 \times 10^{-4}$) when compared to the average sequence diversity of the human genome ($\pi = 7.5 \times 10^{-4}$),[36] and had a rank value of 274 among 286 high-to-low π-values from as many autosomal genes reported in the Seattle SNPs Variation Discovery Resources (http://pga.gs.washington.edu/). A low nucleotide diversity can be the consequence of either local gene history (for example natural selection) or low mutation rate.[37] We estimated a sequence divergence value between human and chimp *PER2* equal to 1.3%, very similar to that reported for the human-chimp genome-wide comparison (1.2%).[16] Thus, based on the assumption that in any genomic region, the amount of observed inter-species sequence divergence largely depends on the mutation rate,[37] a low mutation rate for the human *PER2* gene seems to be unlikely. To assess whether the observed pattern of sequence diversity at the *PER2* gene conformed to neutral expectations, we used four different neutrality
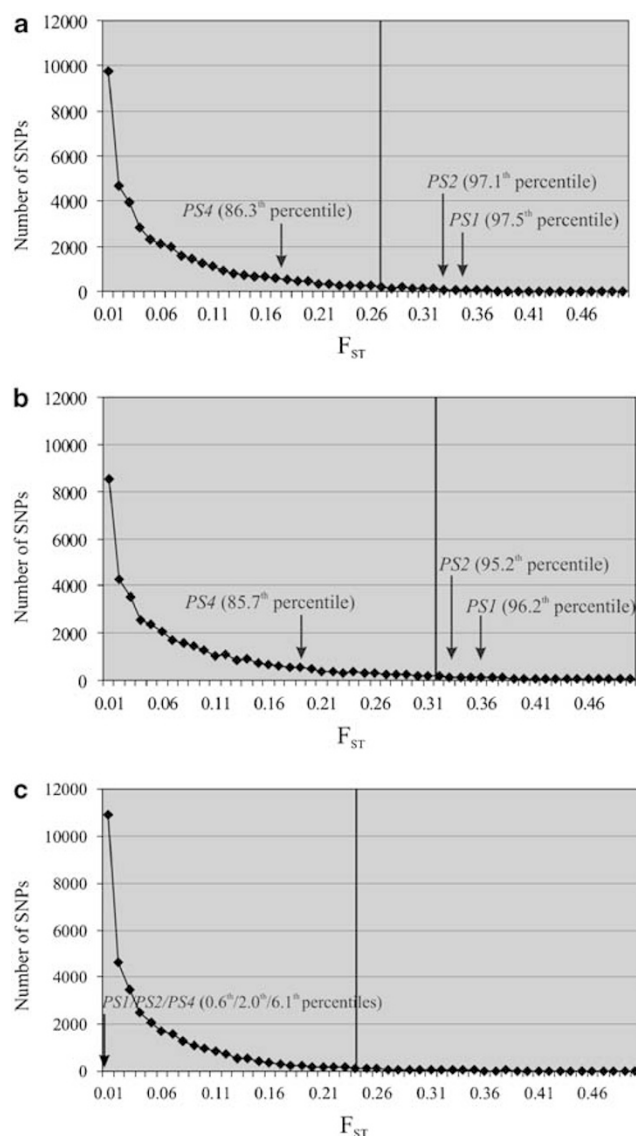

Figure 4 F_{ST} distribution. Distribution of F_{ST} values observed for (**a**) Africans *vs* Europeans, (**b**) Africans *vs* Asian and (**c**) Europeans *vs* Asian. The curves show the frequency distribution of F_{ST} based on (**a**) 43 158, (**b**) 42 269 and (**c**) 36 265 SNPs from the HapMap Phase I database (see Materials and methods section). The vertical line cuts off the top 5% of the distribution.

tests: Tajima's D,[23] Fu and Li's D and F[24] and Fay and Wu's H[25] statistics. The rationale for using different neutrality tests is that they are complementary in their ability to detect departure from neutrality.[25,38,39] The Tajima's D, and the Fu and Li's D and F tests indicated a statistically significant deviation from selective neutrality, both when a worldwide sample of 40 chromosomes and a sample of 30 non-African chromosomes were considered (Table 4). The statistically significant negative values for Tajima's D and

Table 4 Neutrality tests[a] for *PER2* gene

	Whole sample	Africans	Non-Africans
Tajima's *D*-test	−1.704*	−0.789	−1.528*
Fu and Li's *D*-test	−3.888**	−1.386	−3.654*
Fu and Li's *F*-test	−3.457**	−1.272	−3.224*
Fay and Wu's *H*-test	0.706	1.770	0.110

Statistical significance levels were obtained by simulation as described in Fu and Li[24] and Fay and Wu.[25]
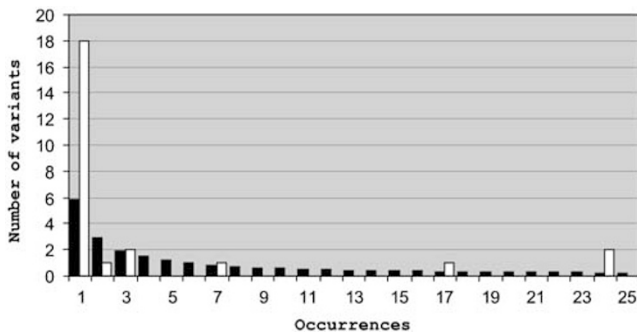[a]*$P < 0.05$; **$P < 0.01$.

Figure 5 Frequency spectrum of derived alleles at the *period 2* (*PER2*) locus. The expected[40,41] and the observed number of mutations are depicted as solid and open bars, respectively.

Fu and Li's *D* and *F* reflect an excess of rare variants with respect to the neutrality expectations (Figure 5). It should be noted that the sample analyzed included 20 subjects from five different continents, most likely resulting in an inflated average pair-wise difference and a positively biased Tajima's *D*.[38] The results of the Fay and Wu's *H* test, which compares the fit of the observed derived allele frequency spectrum with that expected under neutrality, were not statistically significant for any population sample (Table 4). As it has been noted by others,[42] a possible reason explaining the failure of this test to reject the hypothesis of neutrality is that the power of the test may be low for regions characterized by a small number of polymorphic sites.

Under the hypothesis that positive selection has driven *PS1, 2* and *4* derived alleles to high frequency, we would expect to observe a low degree of sequence variation in the associated haplotypes. The derived alleles at these loci define a haplogroup ($PS1_{del}$-$PS2_C$-$PS4_A$, harboring haplotypes II–III in Figure 3) that shows low associated sequence diversity ($\pi = 0.2 \times 10^{-5} \pm 0.2 \times 10^{-5}$) with an estimate of the TMRCA for the haplogroup of 8.7 ± 8.7 Kyr. The coalescence-based estimate of its age places the TMRCA of this haplogroup in recent times, in marked contrast with the date computed under a strictly neutral genetic-drift model (869 ± 617 Kyr), based on an average allele frequency = 0.54, $N_e = 10\,000$ and 30 years per generation.[31]

Discussion

Circadian rhythmicity is a ubiquitous feature of living systems and circadian clocks are of fundamental importance for the adaptation of organisms to the alternating light/dark cycles produced by the rotation of the earth around its axis. Day length varies with latitude and seasonality, so variation in clock and/or clock-related genes, which have a patterned geographical distribution may reflect local adaptation. Indeed, there are several reports showing a correlation between clock gene polymorphisms and latitude-associated environmental variables. A clinal distribution of period allele frequencies at a microsatellite coding for a threonine-glycine repeat was observed in *D. melanogaster*, suggesting a possible role of this polymorphism in fine tuning a cardinal property of the circadian oscillation, its temperature compensation.[43–46] Moreover, a latitudinally structured length polymorphism at the timeless clock gene of *D. melanogaster* has been recently implicated in the photoperiod- and temperature-dependent adaptive life-history trait of diapause.[47,48] The finding of latitudinally structured clock gene polymorphisms exclusively in poikilotherms might suggest that thermal selection more than latitudinal changes in day length would have played the major role in such polymorphisms (but see Sandrelli *et al*[47] and Tauber *et al*[48]). However, a latitudinal cline for a length polymorphism in the *Clock* gene of a homeothermal avian species, the nonmigratory passerine *Cyanistes caeruleus*, has been recently described, suggesting that, at least in this case, an additional significant role for other environmental parameters, that is photoperiodicity, could be important.[49] Interestingly, a similar cline was not detectable in another passerine bird, *Luscinia svecica* which, in contrast to *Cyanistes*, has a pronounced migratory behavior.[49] These clinally distributed differences in length variation of repetitive regions in clock genes might reflect the enhanced mutagenic potential of repetitive/modular motifs,[50,51] compared to single point mutations, for generating potentially 'adaptable' genetic substrates for fine tuning circadian rhythms to the environment. In humans, a possible influence of latitude on the association between a circadian phenotype and a *PER3* length polymorphism has been recently suggested.[6] However, the analysis of the variation in allele frequencies at the same *PER3* polymorphism in different human populations provided no indication of any geographically structured selective phenomena on this polymorphism.[7]

To investigate the hypothesis that genetic variation in components of the human circadian clock has been shaped by environmental variables such as photoperiod and/or temperature profiles, we analyzed the possible relationship between *PER2* haplotype frequencies and different latitudes where human populations live. No evidence for latitudinal selective effects on the *PER2* gene was found, either by testing for correlation between latitudinal and

genetic distances or by analyzing the *PER2* molecular diversity among groups of latitudinally homogeneous populations. Investigating the extent to which the latitudinal dependencies, if any, are linked to the interaction of many clock genes rather than to the variation of a single circadian clock component, will require further experimental studies.

Although our study provides no indication of any signal of latitudinal-driven effects on *PER2* diversity, nevertheless the high levels of inter-population genetic differentiation observed for this gene, as compared with other loci, strongly suggest a role for geographically restricted positive selection, a scenario also supported by the intra-population sequence diversity analysis. Noteworthy, the estimate of the coalescence age of the putatively selected *PER2* haplogroup gave a much lower value than that expected under a model of neutrality. Although the role of confounding demographic effects cannot be completely ruled out due to large uncertainties in predictions of the neutral model parameters, considering all the present findings it is conceivable that selection affected the genetic history of the *PER2* locus after the out-of-Africa expansion.

Since the polymorphic sites in the two adjacent regions around the *PER2* locus display relatively low F_{ST} values in both Africans *vs* Europeans and Africans *vs* Asian comparisons (see Results section), it is conceivable that *PER2* itself has been under the selection process, rather than being hitchhiked by adaptive variation at a nearby locus. However, the actual sites of natural selection remain to be identified. Whereas no missense mutations were found in *PER2* in our sequencing set, one nonsynonymous mutation (rs934945, *PS5* in the present study), resulting in a glycine to glutamine amino-acid substitution, was included in the analysis of our genotyping set. The haplotype identified by the derived allele at the *PS5* site (haplotype III in Figure 3) represents a nonnegligible fraction of the chromosomes belonging to the putatively selected haplogroup $PS1_{del}$-$PS2_C$-$PS4_A$ (22% in Europeans, 33% in Asian, 57% in Native Americans). Given the potential functional importance of the *PS5* mutation, such a finding would suggest an adaptative role of this mutation, which could at least partly account for the rise in frequency of the haplogroup $PS1_{del}$-$PS2_C$-$PS4_A$. However, the possibility that the selection signature is due to another *PER2* polymorphic site in linkage disequilibrium with $PS1_{del}$, $PS2_C$ and $PS4_A$ alleles cannot be ruled out. Recently, Carpen *et al*[52] found a mutation in the 5′ UTR of the *PER2* gene, C111G (corresponding to rs2304672 in the SNP database), which was significantly associated with extreme diurnal preference, and offered a possible explanation of the molecular mechanisms responsible for this sleep–wake phase disorder. This mutation was not included in the analysis of our genotyping set, and the subjects who were available for the present investigation were collected without regard to relevant sleep phenotypes such as

morning or evening preference. However, given the relatively low frequency of the derived allele at the C111G site (one G allele found among 40 chromosomes in our sequencing set, see Figure 2; 7–10% in the HapMap Phase I data at www.hapmap.org), and also the very low level of inter-population diversity revealed by this polymorphism (HapMap Phase I data at www.hapmap.org), it seems unlikely that selection at this site, if any, may be the factor that drove the haplogroup $PS1_{del}$-$PS2_C$-$PS4_A$ to high frequency.

The identification of the genetic and molecular bases of the circadian clock has brought a completely new perspective on the relationships between circadian clock, health and disease. The present study reports on significative differences in the pattern of distribution of *PER2* haplotypes between Africans and non-Africans, suggestive of natural selection. The functional meaning of *PER2* diversity on human physiology and circadian rhythmicity remains to be investigated and requires more targeted studies, which can be guided by the present findings.

References
1 Ko CH, Takahashi JS: Molecular components of the mammalian circadian clock. *Hum Mol Genet* 2006; **15**: R271–R277.
2 Stratmann M, Schibler U: Properties, entrainment, and physiological functions of mammalian peripheral oscillators. *J Biol Rhythms* 2006; **21**: 494–506.
3 Maywood ES, O'Neill J, Wong GKY, Reddy AB, Hastings MH: Circadian timing in health and disease. *Prog Brain Res* 2006; **153**: 253–269.
4 Ebisawa T: Circadian rhythms in the CNS and peripheral clock disorders: human sleep disorders and clock genes. *J Pharmacol Sci* 2007; **103**: 150–154.
5 Reppert SM, Weaver DR: Coordination of circadian timing in mammals. *Nature* 2002; **418**: 935–941.
6 Pereira DS, Tufik S, Louzada FM *et al*: Association of the length polymorphism in the human Per3 gene with the delayed sleep-phase syndrome: does latitude have an influence upon it? *Sleep* 2005; **28**: 29–32.
7 Nadkarni NA, Weale ME, von Schantz M, Thomas MG: Evolution of a length polymorphism in the human *PER3* gene, a component of the circadian system. *J Biol Rhythms* 2005; **20**: 490–499.
8 Fu L, Pelicano H, Liu J, Huang P, Lee CC: The circadian gene *Period2* plays an important role in tumor suppression and DNA damage response *in vivo*. *Cell* 2002; **111**: 41–50.
9 Hua H, Wang Y, Wan C *et al*: Circadian gene *mPer2* overexpression induces cancer cell apoptosis. *Cancer Sci* 2006; **97**: 589–596.
10 Ross K: Circadian rhythms play role in cancer research. *J Natl Cancer Inst* 2006; **98**: 806–807.

11 Viswambharan H, Carvas JM, Antic V *et al*: Mutation of the circadian clock gene *Per2* alters vascular endothelial function. *Circulation* 2007; **115**: 2188–2195.

12 Spanagel R, Pendyala G, Abarca C *et al*: The clock gene *Per2* influences the glutamatergic system and modulates alcohol consumption. *Nat Med* 2005; **11**: 35–42.

13 Scozzari R, Cruciani F, Malaspina P *et al*: Differential structuring of human populations for homologous X and Y microsatellite loci. *Am J Hum Genet* 1997; **61**: 719–733.

14 Bortolini M-C, Salzano FM, Thomas MG *et al*: Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 2003; **73**: 524–539.

15 Cruciani F, La Fratta R, Santolamazza P *et al*: Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* 2004; **74**: 1014–1022.

16 The Chimpanzee Sequencing and Analysis Consortium: Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 2005; **437**: 69–87.

17 Excoffier L, Slatkin M: Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995; **12**: 921–927.

18 Excoffier L, Laval G, Schneider S: Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinformatics Online* 2005; **1**: 47–50.

19 Lewontin RC: The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 1964; **49**: 49–67.

20 Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; **131**: 479–491.

21 Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG: Measures of human population structure show heterogeneity among genomic regions. *Genome Res* 2005; **15**: 1468–1476.

22 Nei M: *Molecular Evolutionary Genetics*. New York: Columbia University Press, 1987.

23 Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; **123**: 585–595.

24 Fu Y-X, Li W-H: Statistical tests of neutrality of mutations. *Genetics* 1993; **133**: 693–709.

25 Fay JC, Wu C-I: Hitchhiking under positive Darwinian selection. *Genetics* 2000; **155**: 1405–1413.

26 Forster P, Harding R, Torroni A, Bandelt H-J: Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 1996; **59**: 935–945.

27 Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S: mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 2000; **67**: 718–726.

28 Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D: Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 2006; **441**: 1103–1108.

29 Bandelt H-J, Forster P, Röhl A: Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 1999; **16**: 37–48.

30 Takahata N: Allelic genealogy and human evolution. *Mol Biol Evol* 1993; **10**: 2–22.

31 Kimura M, Ohta T: The age of a neutral mutant persisting in a finite population. *Genetics* 1973; **75**: 199–212.

32 Kivisild T, Reidla M, Metspalu E *et al*: Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the Gate of Tears. *Am J Hum Genet* 2004; **75**: 752–770.

33 Lovell A, Moreau C, Yotova V *et al*: Ethiopia: between sub-Saharan Africa and western Eurasia. *Ann Hum Genet* 2005; **69**: 275–287.

34 Akey JM, Zhang G, Zhang K, Jin L, Shriver MD: Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 2002; **12**: 1805–1814.

35 The International HapMap Consortium: A haplotype map of the human genome. *Nature* 2005; **437**: 1299–1320.

36 The international SNP Map Working Group: A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; **409**: 928–933.

37 Reich DE, Schaffner SF, Daly MJ *et al*: Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet* 2002; **32**: 135–142.

38 Simonsen KL, Churchill GA, Aquadro CF: Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 1995; **141**: 413–429.

39 Przeworski M: The signature of positive selection at randomly chosen loci. *Genetics* 2002; **160**: 1179–1189.

40 Watterson GA: On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 1975; **7**: 256–276.

41 Zhou G, Zhai Y, Dong X *et al*: Haplotype structure and evidence for positive selection at the human *IL13* locus. *Mol Biol Evol* 2004; **21**: 29–35.

42 Thompson EE, Kuttab-Boulos H, Witonsky D, Yang L, Roe BA, Di Rienzo A: *CYP3A* variation and the evolution of salt-sensitivity variants. *Am J Hum Genet* 2004; **75**: 1059–1069.

43 Costa R, Peixoto AA, Barbujani G, Kyriacou CP: A latitudinal cline in a *Drosophila* clock gene. *Proc R Soc B-Biol Sci* 1992; **250**: 43–49.

44 Sawyer LA, Sandrelli F, Pasetto C *et al*: The *period* gene Thr-Gly polymorphism in Australian and African *Drosophila melanogaster* populations: implications for selection. *Genetics* 2006; **174**: 465–480.

45 Sawyer LA, Hennessy JM, Peixoto AA *et al*: Natural variation in a *Drosophila* clock gene and temperature compensation. *Science* 1997; **278**: 2117–2120.

46 Kyriacou CP, Peixoto AA, Sandrelli F, Costa R, Tauber E: Clines in clock genes: fine-tuning circadian rhythms to the environment. *Trends Genet* 2008; **24**: 124–132.

47 Sandrelli F, Tauber E, Pegoraro M *et al*: A molecular basis for natural selection at the *timeless* locus in *Drosophila melanogaster*. *Science* 2007; **316**: 1898–1900.

48 Tauber E, Zordan M, Sandrelli F *et al*: Natural selection favors a newly derived *timeless* allele in *Drosophila melanogaster*. *Science* 2007; **316**: 1895–1898.

49 Johnsen A, Fidler AE, Kuhn S *et al*: Avian *Clock* gene polymorphism: evidence for a latitudinal cline in allele frequencies. *Mol Ecol* 2007; **16**: 4867–4880.

50 Rosato E, Peixoto AA, Gallippi A, Kyriacou CP, Costa R: Mutational mechanisms, phylogeny, and evolution of a repetitive region within a clock gene of *Drosophila melanogaster*. *J Mol Evol* 1996; **42**: 392–408.

51 Rosato E, Peixoto AA, Costa R, Kyriacou CP: Linkage disequilibrium, mutational analysis and natural selection in the repetitive region of the clock gene, *period*, in *Drosophila melanogaster*. *Genet Res* 1997; **69**: 89–99.

52 Carpen JD, Archer SN, Skene DJ, Smits M, von Schantz M: A single-nucleotide polymorphism in the 5′-untranslated region of the *hPER2* gene is associated with diurnal preference. *J Sleep Res* 2005; **14**: 293–297.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)