

Evolution of a Length Polymorphism in the Human *PER3* Gene, a Component of the Circadian System

Nachiket A. Nadkarni,^{*,1} Michael E. Weale,[†] Malcolm von Schantz,[‡] and Mark G. Thomas^{*,1}

^{*}Department of Biology and [†]Department of Medicine,
University College London, United Kingdom,

[‡]Centre for Chronobiology, School of Biomedical and Molecular Sciences,
University of Surrey, Guildford, United Kingdom

Abstract Period homologue 3 (*PER3*) is a component of the mammalian circadian system, although its precise role is unknown. A biallelic variable number tandem repeat (VNTR) polymorphism exists in human *PER3*, consisting of 4 or 5 repeats of a 54-bp sequence in a region encoding a putative phosphorylation domain. This polymorphism has previously been reported to associate with diurnal preference (“morningness” and “eveningness”) and delayed sleep-phase syndrome. We have investigated the global allele frequencies of this variant in ethnically distinct indigenous populations. All populations were polymorphic, with the shorter (4-repeat) allele ranging in frequency from 0.19 (Papua New Guinea) to 0.89 (Mongolia). To investigate if allele frequency has been influenced by natural selection, the authors 1) tested for a correlation with latitude and mean annual insolation (incident sunlight energy), using classical markers to correct for historical population differentiation; and they 2) compared allele-frequency difference between European American, African American, and East Asian populations, as measured using F_{ST} , to an empirical null distribution of F_{ST} values based on a genome-wide dataset of single nucleotide polymorphisms (SNPs) of presumed neutral loci that were previously typed by The SNP Consortium. The variation in allele frequencies between indigenous populations did not show a pattern that would indicate selective pressure on *PER3* resulting from day-length variation or mean annual insolation, and the allele-frequency difference between European Americans, African Americans, and East Asians was not an outlier when compared to the distribution for presumed neutral SNPs. We therefore find no evidence for differential or balancing selection in the contemporary pattern of global *PER3* allele frequencies.

Key words *PER3*, circadian rhythms, polymorphism, natural selection, day length

The circadian oscillator is a complex series of intraorganismal interactions that result in self-sustaining, near-24-h-period oscillations in multiple biological activities. Such a system is possessed by

animals, fungi, plants, and cyanobacteria. On a molecular level, the core “clock” that drives the oscillations is made up of a series of interconnected and thermally compensated self-sustaining transcriptional feedback

1. To whom all correspondence should be addressed: Mark G. Thomas, Department of Biology, University College London, Gower Street, London, WC1E 6BT, UK; e-mail: m.thomas@ucl.ac.uk; Nachiket A. Nadkarni, Molecular Imaging Group, MRC Clinical Sciences Centre, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London, W12 0NN, United Kingdom.

loops (Lowrey and Takahashi, 2004). Some of the loops' transcription complexes (in addition to being part of the core clock) act as the clock's output by driving the expression of so-called clock controlled genes (CCGs) that effect the circadian phenotype (Dunlap, 1999; Reppert and Weaver, 2002). The *PER3* protein is thought to be a component of one of these transcriptional complexes, but the subtle effects of inactivating it in mice would indicate that rather than being a core clock component, *Per3* may be involved in "buffering" the other components or may even be a CCG (Bae et al., 2001; Shearman et al., 2000).

A screen of *PER3* in a Japanese population (Ebisawa et al., 2001) revealed a number of polymorphisms, of which one (a missense polymorphism defining a rare haplotype) was reported to be associated with delayed sleep-phase syndrome (DSPS), a circadian rhythm sleep disorder. Another polymorphism, which was described but not characterized in that publication, is a biallelic length change in a putative phosphorylation domain-encoding region. This region is composed of 4 indirect repeats of a 54-bp sequence, the novel allele having an added repeat in the middle. It is not located within any identified functional domain but is near to the putative binding site for casein kinase I (Archer et al., 2003). The repeat is entirely absent in mice, rats, and dogs, and is present with varying copy numbers in primates (Jenkins et al., 2005). It was reported in a UK population (Archer et al., 2003) that this polymorphism may influence diurnal preference ("morningness" and "eveningness"), the 4-repeat allele being associated with eveningness (a behavioral preference for evening activity) and DSPS (which may be considered an extreme form of eveningness), while the 5-repeat allele was associated with morningness. A recent report (Pereira et al., 2005) was able to replicate the association with diurnal preference in a Brazilian population, but the authors found the 5-repeat allele to be associated with DSPS, leading them to suggest the possibility of a latitudinal influence.

Genetic variation in components of the circadian clock allows natural selection to optimize the relationship between clock-influenced biological processes and the sun's diurnal cycle (Michael et al., 2003). The environmental factor that most directly affects variation in the timing of dawn and dusk is latitude, and previous studies in *Arabidopsis* have shown latitudinal clines both in the free-running (constant light) circadian period and in allele frequencies of genes thought to contribute to the clock's function (Michael et al.,

2003). In addition, allele frequencies of the polymorphic *Drosophila* clock component *period* (*per*) also show a latitudinal cline (Costa et al., 1992), although in this case, the cline is thought to be related to the clock's temperature compensation (Peixoto et al., 1998; Sawyer et al., 1997). A latitudinal variation in *PER3* allele frequency has already been hypothesized based on the opposite associations with DSPS reported in a British and a Brazilian population (Pereira et al., 2005). If *PER3* alleles do influence diurnal preference, then it is reasonable to hypothesize that such variation could have repercussions for fitness, possibly shaped by day-length variability or mean annual insolation (intensity of incoming solar radiation per unit surface area).

To investigate the hypothesis that natural selection has shaped the global *PER3* allele-frequency distribution, 2 approaches were employed. The first was to correlate allele-frequency differences (measured using the statistic F_{ST} ; Weir, 1996) with differences in potential environmental selective pressures (day-length variation and mean annual insolation), while correcting for the average degree of genetic relatedness among populations as evidenced by classical markers (Cavalli-Sforza et al., 1994). F_{ST} is a convenient measure of allele-frequency difference and is defined as the proportion of allele-frequency variance that is attributable to between-population differences (Weir, 1996). An F_{ST} of 0 would indicate that 2 populations have identical allele frequencies, while an F_{ST} of 1 would indicate no allele sharing and complete fixation of alleles within populations. Two models to explain the contemporary geographic distribution of 4/5-repeat allele frequency were compared: influence by drift shaped by demographic history alone, and influence by drift together with environmental selection pressure. The 2 models were formally compared using a partial-correlation permutation test similar in principle to the Mantel test (see Materials and Methods section for details).

The 2nd approach was a comparison of *PER3* allele-frequency differences (F_{ST}) to allele-frequency differences of a genome-wide panel of single nucleotide polymorphisms (SNPs) in 3 different populations (European American, African American, and East Asian). Allele-frequency differences among populations could arise as a result of differential selection pressures favoring different alleles in different populations. However, because of the stochastic nature of inheritance, non-zero F_{ST} s will also arise at strictly neutral polymorphic loci. A set of neutral loci will

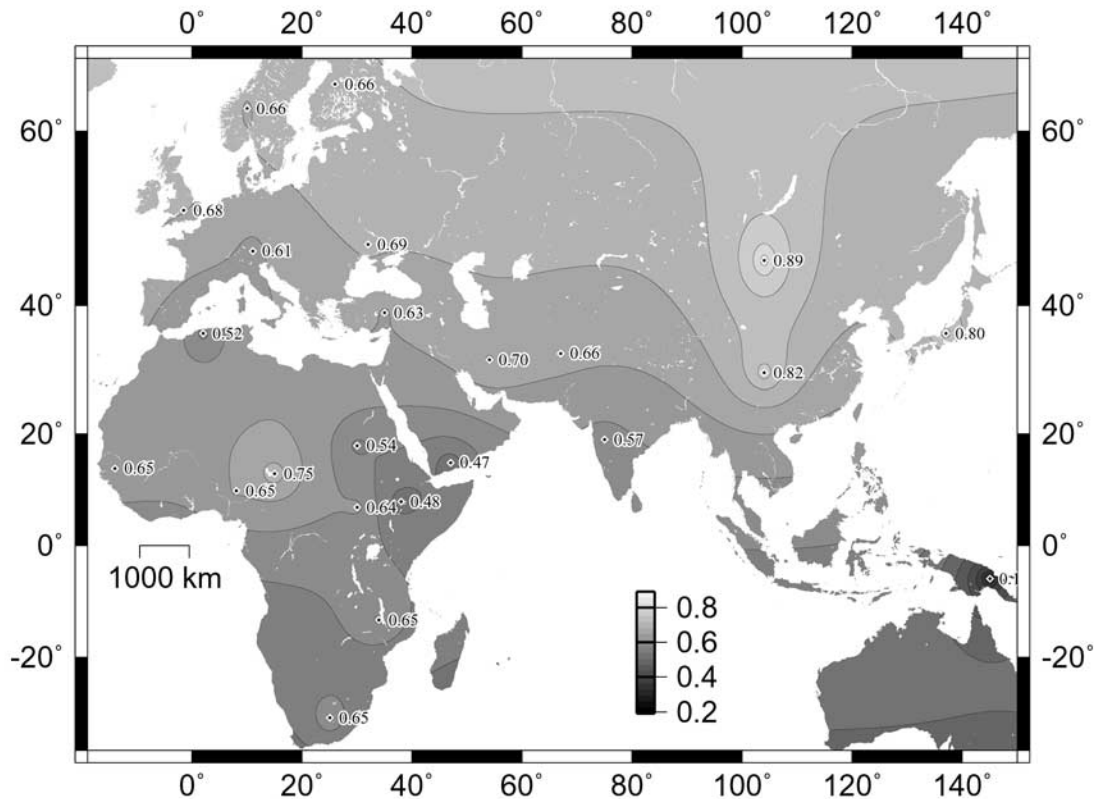


Figure 1. Interpolated contour map of the global distribution of the *PER3* 4/5-repeat polymorphism. The numbers indicate the frequencies of the *PER3* 4-repeat allele in the populations studied.

have a distribution of interpopulation F_{ST} values (the neutral or null distribution), the shape of which is a result of genetic drift shaped by demographic history. Thus, in the absence of other evidence, the default position among evolutionary biologists is to use non-selection-based explanations for differences in allele frequencies. Stated more formally, neutrality is the null hypothesis against which predictions of the effect of adaptive evolution on allele frequency must be compared. However, if a locus has been under directional selection for 1 allele in 1 population but not the other, this would have the effect of driving allele-frequency difference (and increasing F_{ST}) beyond that expected under neutrality (Akey et al., 2002; Cavalli-Sforza, 1966; Lewontin and Krakauer, 1973). Conversely, under balancing selection, F_{ST} would be reduced relative to its neutral distribution. This provides the basis of a convenient test for hypotheses of selection acting on candidate loci. Recently, large data sets of SNP frequencies in 3 different populations (European American, African American, and East Asian) have become available and have been used to

compile empirical null distributions of F_{ST} (Akey et al., 2002; Sachidanandam et al., 2001).

MATERIALS AND METHODS

Samples

The DNA samples from 25 global populations (see Table 1 and Fig. 1 for locations and sample sizes) were used for global studies of allele-frequency distribution. All samples were collected from paternally unrelated men. Data for the Japanese population are from Ebisawa et al. (2001). Additionally, 84 DNA samples representing 42 African Americans and 42 European Americans (Coriell Cell Repositories, panel IDs: TSC_42_AA and TSC_42_CAU, respectively) that had been used by The SNP Consortium (TSC) were also selected for genotyping. The 84 Coriell Cell Repository samples will hereafter be referred to as the *Coriell Panel*.

Table 1. Population Data Used in Statistical Analyses

Population	Equivalent Population for F_{ST} Comparison	4-Repeat Allele Frequency	χ^2 p-Value ^a	N	Latitude (degrees)	Longitude (degrees)	Mean Annual Insolation	SD in Day Length (fraction of 24 h)
Papua New Guineans	New Guinean	0.191	0.999	84	6	325	229.4	0.0101
South African Bantu	Bantu	0.652	0.774	69	30	205	253.5	0.0556
Amharic Ethiopians	East African	0.481	0.995	77	8	218	281.1	0.0134
South Sudanese	Nilo-Saharan	0.640	0.957	50	7	210	264.8	0.0117
Nigerians	West African	0.658	0.928	79	10	188	242.9	0.0168
Algerians	Berber	0.527	0.334	75	32	182	206.9	0.0602
Iranians	Iranian	0.700	0.790	75	32	234	246.7	0.0602
Japanese ^b	Japanese	0.805	Unknown	100	36	317	161.7	0.0703
Sichuan Chinese	South Chinese	0.822	0.861	73	30	284	163.5	0.0556
Norwegians	Danish	0.664	0.838	70	62	196	106.0	0.2008
British	English	0.683	0.942	131	52	179	119.7	0.1276
Anatolians	Near Eastern	0.634	0.959	56	39	215	181.1	0.0786
Mongolians	Mongol Tungus	0.891	0.710	46	46	284	175.6	0.1016
Hindus from India	Indian	0.576	0.751	33	19	255	247.4	0.0330
Finns	Finns	0.669	0.759	77	64	206	94.2	0.2267
Lake Chad	Lake Chad	0.750	0.837	80	13	195	275.7	0.0221
Malawi	Malawi	0.651	0.958	93	14	214	244.8	0.0239
Manjak Senegalese	West African	0.622	1.000	49	14	166	271.3	0.0238
Wolof Senegalese	West African	0.685	0.958	46	14	166	271.3	0.0238
North Sudanese	North Sudanese	0.543	0.559	47	18	210	300.3	0.0311
Ladin-speaking Tyrolians	Ladin-speaking Tyrolians	0.635	0.016	48	43	191	150.5	0.0910
Italian-speaking Tyrolians	Italian-speaking Tyrolians	0.583	0.063	42	43	191	150.5	0.0910
Ukrainians	Ukrainians	0.694	0.988	72	48	212	153.9	0.1095
Yemenis	Yemenis	0.477	1.000	22	15	227	297.0	0.0256
Afghans	Afghans	0.661	0.147	31	33	247	238.4	0.0627
Caucasians	Caucasians	0.571	0.720	42	n/a	n/a	n/a	n/a
African Americans	African Americans	0.643	0.9716	42	n/a	n/a	n/a	n/a

NOTE: PER3 4-repeat allele frequencies for the 27 populations considered, together with geographic coordinates, mean annual insolation values, day-length variation values, and, where applicable, the name of the population from which classical-marker F_{ST} values were taken are from Cavalli-Sforza et al. (1994) to correct for general genetic relatedness. n/a = not applicable.

a. Based on the Hardy-Weinberg equilibrium.

b. Data for this row are from Ebisawa et al. (2001).

Genotyping

The region of the *PER3* gene containing the 4/5-repeat polymorphism was amplified by polymerase chain reaction (PCR) in a volume of 10 μ L containing 200 μ M dNTPs, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100, 0.01% gelatin, 50 mM KCl, 1.5 mM MgCl₂, 0.13 U *Taq* polymerase (HT Biotech, Cambridge, UK), 9.3 nM TaqStart Monoclonal Antibody (Clontech, Palo Alto, CA), and 0.15 μ M of each of the primers described by Ebisawa et al. (2001). The *Taq* polymerase and TaqStart Monoclonal Antibody were premixed prior to being added to the other PCR components. Cycling parameters were a preincubation at 94 °C for 3 min followed by 38 cycles of 94 °C for 45 sec, 63.5 °C for 45 sec, and 72 °C for 1 min. The PCR products were electrophoresed in a 1.5% agarose gel and visualized by ethidium bromide staining.

Data Analysis

Data on presumed neutral SNPs typed in 42 African Americans, 42 European Americans, and 42 East Asians (Akey et al., 2002; Sachidanandam et al., 2001) were taken from a dataset of 33,487 SNPs typed by the Orchid Laboratory, publicly available at the TSC Web site (http://snp.cshl.org/allele_frequency_project/panels.shtml). We carefully selected SNPs from the larger dataset, such that each SNP was 1) variable, either within or between populations; 2) mapped only once onto the genome; and 3) separated by at least 50 kb from the next-nearest SNP, to minimize correlation in F_{ST} values. This left a total of 11,970, 11,024, and 11,952 SNPs active in the African American versus European American, East Asian versus European American, and African American versus East Asian comparisons, respectively. We note that this SNP set will, by chance, contain some loci that are under selection, but unless the proportion of loci under balancing selection is large, then this will have only a conservative effect on the comparisons presented here. All F_{ST} values were calculated using the unbiased "random populations" formula for haploid data given by Weir (1996).

Mean surface solar irradiance (insolation) was estimated from data obtained from the NASA Surface Solar Irradiance Web site (<http://daac.gsfc.nasa.gov/interdisc/readmes/srfrad.shtml>). These data con-

sisted of mean monthly insolation estimates covering a period from July 1983 to June 1991 on a 1° × 1° grid over the entire earth. These data were averaged to provide a mean annual insolation. Day-length variation across 1 year at the sample locations was calculated using a Web-based NASA algorithm (<http://aom.giss.nasa.gov/srlocat.html>). Population pairwise distance matrices of mean annual insolation and day-length variation were created by squaring the population pairwise differences for the 2 parameters between the populations used. Global population pairwise F_{ST} values, calculated from the isoform frequencies of polymorphic proteins (also known as classical markers), were taken from Cavalli-Sforza et al. (1994).

Mantel-type permutation tests carried out on partial correlation coefficients were performed using routines written by one of the authors (M.E.W., available on request). The test takes 3 square matrices (A, B, and C) of population pairwise difference values (i.e., 1 row and 1 column for each population sample). First, the observed partial correlation statistic of the off-diagonal matrix elements of A on B given C is found. Then, the null distribution of this statistic is found by a Mantel-type permutation procedure, in which rows (and corresponding columns) of the A matrix are shuffled at random 10,000 times, and the statistic is recalculated. The estimated 1-tailed p -value is the proportion of times the permuted partial-correlation values are equal to or exceed the observed value, and twice this proportion is the estimated 2-tailed p -value. In the analyses performed here, matrix A contained *PER3* F_{ST} values, matrix C contained classical-marker F_{ST} values obtained from Cavalli-Sforza et al. (1994), and matrix B contained pairwise squared differences in either day-length variation or mean annual insolation (see Results section for further details).

The exact test of population differentiation (Raymond and Rousset, 1995) was performed using the population genetics data analysis software Arlequin (Schneider et al., 2000). All other statistical analyses were carried out using the statistical package "R" (R Development Core Team, 2003). A contour map of the geographic distribution of the *PER3* 4-repeat allele was estimated and visualized by surface interpolation (tension factor = 0.9) using the Generic Mapping Tools software (Wessel and Smith, 1998).

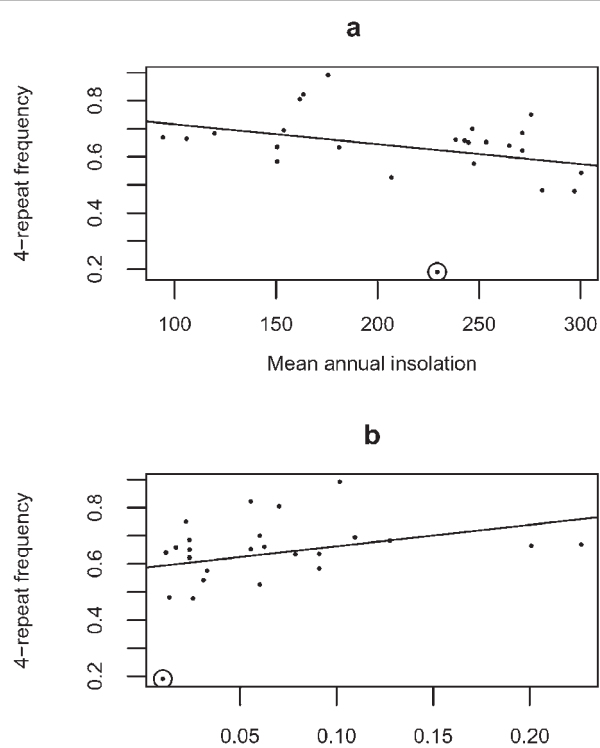


Figure 2. Plot of *PER3* 4-repeat allele frequency against (a) mean annual insolation and (b) standard deviation in day length during 1 year. The r^2 and corresponding p -values are, respectively, (a) 0.107 and 0.110 and (b) 0.1015 and 0.121. The data point for the Papua New Guinea sample (excluded as an outlier in some analyses) is circled in both (a) and (b).

RESULTS

Global Allele-Frequency Data

The observed frequencies of the *PER3* gene 4-repeat allele are given in Table 1 and are displayed as an interpolated contour map in Figure 1. The 4-repeat allele frequency observed in our UK sample was the same as that observed by Archer et al. (2003) (exact test of population differentiation, $p = 1.00$) (Raymond and Rousset, 1995). Genotype frequencies deviated significantly ($p \leq 0.05$) from the Hardy-Weinberg equilibrium in the Ladin-speaking Tyrolean sample ($p = 0.016$). This may be due to the Ladin-speaking populations being isolated, having a small effective population size, and being highly structured (M.G.T., unpublished data). There was a great deal of heterogeneity in

gene frequency, with populations forming 3 groupings and 2 outliers. A total of 14 populations (Finns, Iranians, Malawis, Nigerians, Norwegians, Manjak Senegalese, Wolof Senegalese, Southern Sudanese, Ladin-speaking Tyroleans, British, Ukrainians, South African Bantu-speakers, Anatolians, and Afghans) had 4-repeat allele frequencies between 0.6 and 0.7. The Mongolian, Sichuan Chinese, and Japanese populations had values between 0.8 and 0.9. Five other populations (Algerians, Northern Sudanese, Amharic Ethiopians, Yemenis, and Hindus from India) formed a grouping with 4-repeat allele frequencies of about half to 0.6. The Lake Chad population had a 4-repeat allele frequency of 0.75. The major outlier was the Papua New Guinean sample, with a 4-repeat allele frequency of 0.19.

Clinal and Pairwise Matrix Analysis of Population Allele Frequencies

To investigate clinal relationships between *PER3* allele frequency, day-length variation, and mean annual insolation, simple linear regressions were performed on population allele frequency against day-length variation and against mean annual insolation at that population's location (Fig. 2). These analyses were performed twice, with and without the Papua New Guinea data, since it was an influential outlier that could disproportionately alter the overall results for the rest of the data. There was no statistically significant trend in allele frequency against day-length variation ($p = 0.12$ with Papua New Guinea, $p = 0.24$ without Papua New Guinea). There was a stronger negative trend between allele frequency and insolation ($p = 0.11$ with Papua New Guinea, $p = 0.05$ without Papua New Guinea).

Although the trend of allele frequency against insolation borders on significance when Papua New Guinea is removed, the above analyses may be misleading, as they do not take into account the potentially confounding effect of human demographic history. To account for this, we took advantage of the large dataset available on the frequencies of classical markers in global populations (Cavalli-Sforza et al., 1994). We used F_{ST} values derived from these markers as an indication of genetic distance due to demographic history, under the assumption that most classical markers are neutral. An *PER3* F_{ST} matrix was cal-

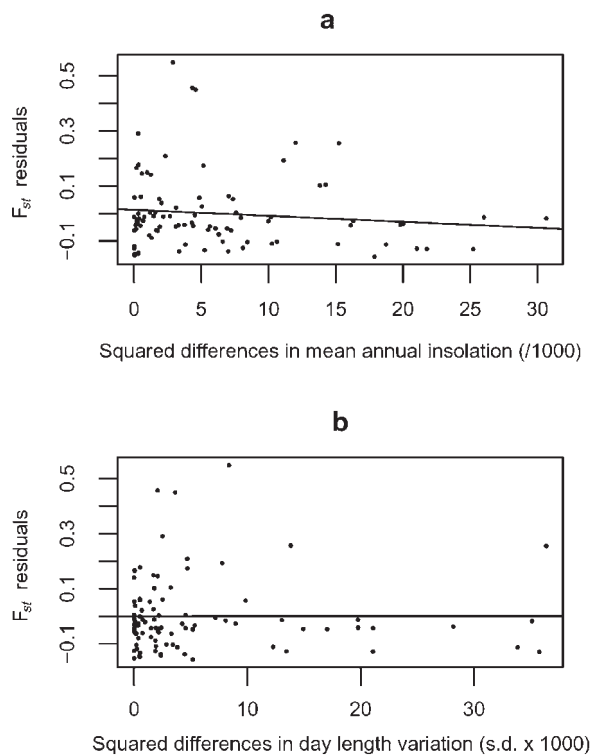


Figure 3. Plot of residuals (from a linear regression of *PER3* F_{ST} values on classical marker-based general genetic relatedness) of pairwise *PER3* F_{ST} values based on 4-repeat allele frequency. These are plotted against (a) pairwise squared differences in mean annual insolation and (b) pairwise squared differences in standard deviation in day length during 1 year. The r^2 and corresponding p -values are, respectively, (a) 0.013 and 0.563 and (b) 7.13×10^{-7} and 0.607.

culated from 4-repeat allele frequencies, and a comparable matrix was calculated from classical marker F_{ST} values (Cavalli-Sforza et al., 1994). The populations for which classical marker data were available do not completely coincide with those genotyped here. We therefore identified 14 populations that had exact or near-exact equivalents to those we have genotyped for *PER3* and used them for comparison (see Table 1). We also calculated pairwise distance matrices for day-length variation and annual insolation, based on squared differences. We then tested, by permutation of the *PER3* F_{ST} matrix (see Materials and Methods section), whether the partial correlations between *PER3* F_{ST} and the 2 solar variables were significant after classical marker F_{ST} values were taken into account (Fig. 3). In both cases, after correcting for classical marker genetic variation, the remaining partial correlations were low and not significant (day-length variation: partial correlation = 0.000847, $p =$

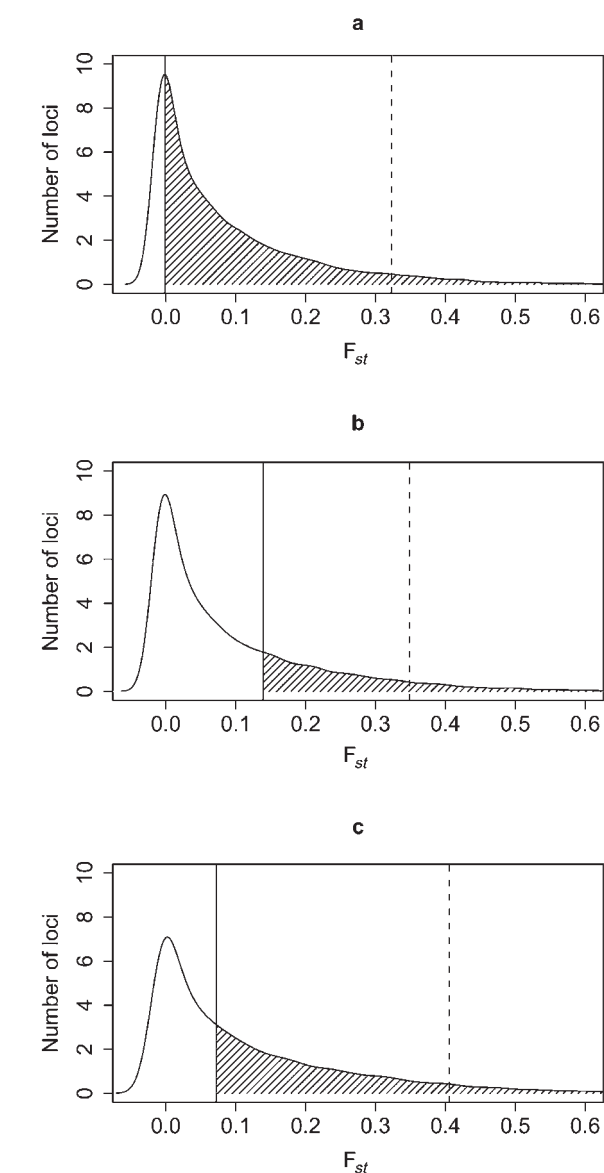


Figure 4. F_{ST} outlier analysis for (a) African American vs. European American, (b) European American vs. East Asian, and (c) African American vs. East Asian population comparisons. The curves show the frequency distribution of F_{ST} based on (a) 11,970, (b) 11,024, and (c) 11,952 SNPs typed by the Orchid Laboratory in 42 African Americans, 42 European Americans, and 42 East Asians. The vertical solid lines indicate the F_{ST} values for the *PER3* 4/5-repeat polymorphism, and the vertical dashed lines indicate the top 5% of the null-distribution values. The cross-hatched areas indicate the top (a) 20.5%, (b) 74.9%, and (c) 52.8% of F_{ST} values in the null distribution, corresponding to the areas above the observed *PER3* F_{ST} values.

0.607; mean annual insolation: partial correlation = -0.113 , $p = 0.563$).

To provide a further test of geographical differential selection, using a different comparative dataset, the *PER3* 4/5-repeat F_{ST} values between the European

American, African American, and East Asian samples were estimated and compared to null- F_{ST} distributions derived from a genome-wide SNP dataset typed by the Orchid Laboratory. For the African American and European American populations, the same (Coriell panel) samples were used to type both the *PER3* 4/5-repeat allele and the genome-wide SNP set. For the East Asian population, we combined the Japanese (Ebisawa et al., 2001) and Southern Chinese (this study) data to establish an "East Asian" sample for *PER3* typing, and we compared this to the East Asian Coriell panel sample used for the genome-wide SNP dataset. Consistent with no selection, in none of the 3 possible comparisons were the *PER3* F_{ST} values outliers in their respective genome-wide null- F_{ST} distributions. The African American versus European American, African American versus East Asian, and European American versus East Asian *PER3* F_{ST} s were found to be -0.0013, 0.072, and 0.14, respectively. These values were in the lower 21%, 53%, and 75% of the empirical null- F_{ST} distributions, respectively (Fig. 4). The comparisons with East Asia are complicated by the fact that different samples were used to define "East Asians"; nevertheless, the results provide no indication of any signal of geographical differential selection.

DISCUSSION

We have shown that there is considerable variation in *PER3* 4/5-repeat allele frequencies among global populations. This accounts for the differences reported in previous studies between Japanese (Ebisawa et al., 2001) and a group mostly consisting of UK inhabitants (Archer et al., 2003). However, our analyses also indicate that genetic drift, molded by human demographic patterns of migration, is sufficient on its own to explain this variation. Given the importance of accounting for human demographic history, as demonstrated here, we would recommend that the analytical methods we have used to correct for this problem are also employed in future studies investigating human differential selection.

While our study provides no evidence for the influence of natural selection, this does not exclude the possibility that natural selection has acted. Our study was aimed at detecting geographical effects of selection, either through large differences across continents, correlations with day length (which would influence the circadian system directly) and mean annual insolation, or through unusual lack of variation, which

would indicate balancing selection (although this would be contrary to our expectation given the variation in environmental factors). Our study would not, however, reveal purifying selection acting equally on all human populations studied or any bouts of selection that had acted prior to the geographical dispersal of humans, while weak differential selection might have too slight an influence to be detectable. In addition, our study would not reveal signs of selection acting on other genetic markers not in strong linkage disequilibrium with the *PER3* 4/5-repeat variant. This includes the possibility of SNP variation within the repeat polymorphism itself, although in practice, one might expect strong linkage disequilibrium with the repeat variant in this case.

Our study did not involve the collection of relevant phenotypes such as DSPS prevalence. So far, data about the prevalence of DSPS are only available for Europeans (Schrader et al., 1993) and Japanese (Yazaki et al., 1999); thus for the majority of the populations in this study, nothing is known about the prevalence of DSPS, nor about the distribution of diurnal preference in general. More data are required on the global distribution of such phenotypes. However, we note that a lack of evidence for selection does not preclude a functional role for the *PER3* 4/5-repeat variant, provided the phenotypes have, until relatively recently, been selectively neutral. While differences in circadian period (and the clock gene polymorphisms that affect them) have been observed to affect reproductive fitness in *Arabidopsis* (Michael et al., 2003), this need not necessarily apply in humans. It is possible that even extreme aspects of diurnal preference, as found in DSPS patients, may only be selectively disadvantageous in the context of societies with strictly defined working hours. More anthropological work is needed to investigate this theory. A recent demonstration that the 4/5-repeat genotype is a potential marker for breast cancer risk in premenopausal women (Zhu et al., 2005) provides further evidence that its phenotypic consequences are far from trivial. It remains to be determined, however, whether this association is directly due to differences in chronotype, to which this polymorphism contributes (Archer et al., 2003), or to a noncircadian pleiotropic process, such as the role in tumor suppression and DNA damage response reported for *Per2* (Fu et al., 2002).

While the profound influence of day length on circadian and circannual processes is very well established, the relationship between insolation and circadian rhythmicity has not been investigated. Although

it is unlikely to be as important as day length in terms of photoentrainment, it may affect the temperature compensation of the clock or the sleeping habits of the individual or the population. However, it is difficult to posit quantitative environmental or behavioral factors that would favor morning or evening preference outside of the parameters listed above. The process of elucidating these factors is made more complicated by evidence that ancient human sleep patterns were different from those of modern humans (Wehr, 1999), and that sleep patterns also vary with culture and latitude (Pereira et al., 2005). Further studies of the *PER3* locus are required to investigate the possible existence of other signals of selection.

ACKNOWLEDGMENTS

The authors thank all sample donors, Neil Bradman for his comments on the article, Abigail Jones for all her assistance, and all the past and present members of the The Centre for Genetic Anthropology for their painstaking work collecting and purifying DNA samples.

REFERENCES

- Akey JM, Zhang G, Zhang K, Jin L, and Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 12:1805-1814.
- Archer SN, Robilliard DL, Skene DJ, Smits M, Williams A, Arendt J, and von Schantz M (2003) A length polymorphism in the circadian clock gene *Per3* is linked to delayed sleep phase syndrome and extreme diurnal preference. *Sleep* 26:413-415.
- Bae K, Jin X, Maywood ES, Hastings MH, Reppert SM, and Weaver DR (2001) Differential functions of *mPer1*, *mPer2*, and *mPer3* in the SCN circadian clock. *Neuron* 30:525-536.
- Cavalli-Sforza LL (1966) Population structure and human evolution. *Proc R Soc Lond B Biol Sci* 164:362-379.
- Cavalli-Sforza LL, Menozzi P, and Piazza A (1994) *The History and Geography of Human Genes*. New Haven, NJ: Princeton University Press.
- Costa R, Peixoto AA, Barbujani G, and Kyriacou CP (1992) A latitudinal cline in a *Drosophila* clock gene. *Proc R Soc Lond B Biol Sci* 250:43-49.
- Dunlap JC (1999) Molecular bases for circadian clocks. *Cell* 96:271-290.
- Ebisawa T, Uchiyama M, Kajimura N, Mishima K, Kamei Y, Katoh M, Watanabe T, Sekimoto M, Shibui K, Kim K, et al. (2001) Association of structural polymorphisms in the human *period3* gene with delayed sleep phase syndrome. *EMBO Rep* 2:342-346.
- Fu L, Pelicano H, Liu J, Huang P, and Lee C (2002) The circadian gene *Period2* plays an important role in tumor suppression and DNA damage response in vivo. *Cell* 111:41-50.
- Jenkins A, Archer SN, and von Schantz M (2005) Expansion during primate radiation of a variable number tandem repeat in the coding region of the circadian clock gene *Period3*. *J Biol Rhythms* 20:470-472.
- Lewontin RC and Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74:175-195.
- Lowrey PL and Takahashi JS (2004) Mammalian circadian biology: elucidating genome-wide levels of temporal organization. *Annu Rev Genomics Hum Genet* 5:407-441.
- Michael TP, Salome PA, Yu HJ, Spencer TR, Sharp EL, McPeck MA, Alonso JM, Ecker JR, and McClung CR (2003) Enhanced fitness conferred by naturally occurring variation in the circadian clock. *Science* 302:1049-1053.
- Peixoto AA, Hennessy JM, Townson I, Hasan G, Rosbash M, Costa R, and Kyriacou CP (1998) Molecular coevolution within a *Drosophila* clock gene. *Proc Natl Acad Sci U S A* 95:4475-4480.
- Pereira DS, Tufik S, Louzada FM, Benedito-Silva AA, Lopez AR, Lemos NA, Korczak AL, D'Ameida V, and Pedrazzoli M (2005) Association of the length polymorphism in the human *Per3* gene with the delayed sleep-phase syndrome: does latitude have an influence upon it? *Sleep* 28:29-32.
- RDevelopment Core Team (2003) *R: A Language and Environment for Statistical Computing*. 1.9.0 ed. Vienna, Austria: R Foundation for Statistical Computing.
- Raymond M and Rousset F (1995) An exact test for population differentiation. *Evolution* 49:1280-1283.
- Reppert SM and Weaver DR (2002) Coordination of circadian timing in mammals. *Nature* 418:935-941.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928-933.
- Sawyer LA, Hennessy JM, Peixoto AA, Rosato E, Parkinson H, Costa R, and Kyriacou CP (1997) Natural variation in a *Drosophila* clock gene and temperature compensation. *Science* 278:2117-2120.
- Schneider S, Roessli D, and Excoffier L (2000) *Arlequin Ver 2.000: A Software for Population Genetics Data Analysis*. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva.
- Schrader H, Bovim G, and Sand T (1993) The prevalence of delayed and advanced sleep phase syndromes. *J Sleep Res* 2:51-55.
- Shearman LP, Jin X, Lee C, Reppert SM, and Weaver DR (2000) Targeted disruption of the *mPer3* gene: subtle effects on circadian clock function. *Mol Cell Biol* 20:6269-6275.
- Wehr TA (1999) The impact of changes in nightlength (scotoperiod) on human sleep. In *Regulation of Sleep and Circadian Rhythms*, Turek FW and Zee PC, eds, pp 266-273, New York, M. Dekker.

- Weir M (1996) *Genetic Data Analysis II*. Sunderland, MA: Sinauer Associates Inc.
- Wessel P and Smith W (1998) New, improved version of Generic Mapping Tools released. *EOS Transactions* 79:579.
- Yazaki M, Shirakawa S, Okawa M, and Takahashi K (1999) Demography of sleep disturbances associated with circadian rhythm disorders in Japan. *Psychiatry Clin Neurosci* 53:267-268.
- Zhu Y, Brown HN, Zhang Y, Stevens RG, and Zheng T (2005) *Period3* structural variation: a circadian biomarker associated with breast cancer in young women. *Cancer Epidemiol Biomarkers Prev* 14:268-270.