Worldwide Allele Frequencies of the Human Apolipoprotein E Gene: Climate, Local Adaptations, and Evolutionary History

Dan T.A. Eisenberg,[1,2]* Christopher W. Kuzawa,[1,2] and M. Geoffrey Hayes[1,3,4]

[1]*Department of Anthropology, Northwestern University, Evanston, IL 60208*
[2]*Cells to Society (C2S): The Center on Social Disparities and Health, Institute for Policy Research, Northwestern University, Evanston, IL 60208*
[3]*Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611*
[4]*Center for Genetic Medicine, Northwestern University, Chicago, IL 60611*

KEY WORDS cholesterol; metabolism; selection; population structure; HGDP-CEPH

ABSTRACT The ε4 allele of the apolipoprotein E (*APOE*) gene is associated with increased cholesterol levels and heart disease. Population allele frequencies of *APOE* have previously been shown to vary, with ε4 frequencies generally increasing with latitude. We hypothesize that this trend resulted from natural selection protecting against low-cholesterol levels. In high-latitude cold environments and low-latitude hot environments, metabolic rate is elevated, which could require higher cholesterol levels. To explore this hypothesis, we compiled *APOE* allele frequencies, latitude, temperature, and elevation from populations around the world. ε4 allele frequencies show a curvilinear relationship with absolute latitude, with lowest frequencies found in the mid-latitudes where temperatures generally require less expenditure on cooling/thermogenesis. Controlling for population structure in a subset of populations did not appreciably change this pattern of association, consistent with selection pressures that vary by latitude shaping ε4 allele frequencies. Temperature records also predict *APOE* frequency in a curvilinear fashion, with lowest ε4 frequencies at moderate temperatures. The model fit between historical temperatures and ε4 is less than between latitude and ε4, but strengthened after correcting for estimated temperature differences during the Paleolithic. Contrary to our hypothesis, we find that elevation did not improve predictive power, and an integrated measure of the cholesterol effect of multiple *APOE* alleles was less related to latitude than was ε4 alone. Our results lend mixed support for a link between past temperature and human *APOE* allele distribution and point to the need to develop better models of past climate in future analyses. Am J Phys Anthropol 143:100–111, 2010.

The apolipoprotein E (*APOE*) gene has been extensively studied in the biomedical literature, and its allele frequencies have been found to vary widely across population. The most commonly studied alleles/isoforms of *APOE* are ε2, ε3, and ε4, which are determined by two single-nucleotide polymorphisms that cause amino acid substitutions and result in functional changes to the apoE protein (Mahley et al., 2000). The ε4 allele of *APOE* has been associated with increased total cholesterol (TC) and low-density lipoprotein cholesterol (LDL-C), whereas its ε2 allele has been associated with similar-sized decreases in TC and LDL-C (relative to ε3 homozygotes; Mahley et al., 2000; Bennet et al., 2007). A primary feature of apoE is its involvement in the transport and clearance of lipids, including LDL-C, high-density lipoprotein cholesterol (HDL-C), very low-density lipoprotein cholesterol (VLDL-C), and triglycerides (TG) [reviewed in Mahley et al. (2000)]. *APOE* ε4 has been associated with increased cholesterol absorption in some (Kesaniemi et al., 1987; Sehayek et al., 2000; Tammi et al., 2000), but not all studies (Woollett et al., 1995; von Bergmann et al., 2003). Consistent with this, there is some evidence that carriers of the ε4 allele show enhanced responsivity to dietary fats (Masson et al., 2003). In turn, ε4 carriers face an increased risk for developing coronary heart disease (CHD; Bennet et al., 2007) and Alzheimer's disease (Bang et al., 2003).

These broad and generally deleterious effects on health have raised questions about the relatively high frequency of the ε4 allele in many populations. The *APOE* gene shows signs of positive selection in the human/chimp lineage (Vamathevan et al., 2008), pointing to the likely importance of phylogenetically recent selective pressures. Of the three common alleles of *APOE*, the ε4 allele is ancestral, whereas the ε3 allele is thought to have arisen within the last 220,000 years, and ε2 was later derived from ε3 (under a neutral model; Fullerton et al., 2000). It has been hypothesized that the ε3 and ε2 alleles may be adapted to meat eating and that they were related to lifespan extension and the capacity

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Dan T.A. Eisenberg, Department of Anthropology, Northwestern University, 1810 Hinman Ave, Evanston, IL 60208. E-mail: dtae@dtae.net

Received 24 October 2009; accepted 3 February 2010

DOI 10.1002/ajpa.21298
Published online 21 April 2010 in Wiley InterScience (www.interscience.wiley.com).

for intergenerational transfers (Finch and Stanford, 2004; Allen et al., 2005). It has also been noted that the ε4 allele is associated with improved perinatal health and survival outcomes (Nagy et al., 1998; cf. Wang et al., 2000; Zetterberg et al., 2002; Wright et al., 2003; Becher et al., 2006; Gaynor et al., 2007), which might act to maintain the ε4 allele despite its later life deleterious effects.

Although providing some possible insights into *APOE* function and evolution, these hypotheses fail to address a key feature of the genetic polymorphism—the tendency of ε4 allele frequency to vary by latitude. Past work suggests that ε4 allele frequency increases with latitude in Eurasia and North America but decreases moderately with latitude in Africa (Gerdes et al., 1992; Singh et al., 2006; Borinskaya et al., 2007). These past studies did not consider the potential role of population structure and, thus, do not speak to possible nonadaptive evolutionary explanations for the gradient. Nonetheless, adaptive evolutionary processes that might explain the worldwide allele distribution of *APOE* have long been and remain a source of broad speculation (e.g., Sing and Davignon, 1985; Corbo and Scacchi, 1999; Martin, 1999; Gerdes, 2003; Han, 2004, p. 1900; Allen et al., 2005; Prentice et al., 2005; Singh et al., 2006; Borinskaya et al., 2007; Finch, 2007, p. 361; Austad and Finch, 2008; Reser, 2009).

In the light of the latitudinal gradient, prominent hypotheses have pointed to climate as a driver of selection on the ε4 allele. Although they do not propose a specific mechanistic link, Borinskaya and colleagues (2007) suggested that the latitudinal gradient of ε4 in Eurasia could indicate selection for higher cholesterol levels in northern latitudes, where energy expenditure in both human populations and other animals are known to be higher (Lovegrove, 2003; Froehle and Schoeninger, 2006). In fact, there are good reasons to suspect such a link. TC levels scale to body mass across a wide range of mammals in a fashion similar to resting metabolic rate (Kuzawa et al., 2006), while mass-specific LDL-C turnover and synthesis rates are inversely related to body mass across several species (Dietschy et al., 1993; Dietschy and Turley, 2004). Cholesterol's involvement in cell proliferation (Cuthbert and Lipsky, 1987) and tissue growth (Schoknecht et al., 1994; Edison et al., 2007; Pond et al., 2008) corresponds to an increasing demand for cholesterol as elevated metabolic rates cause more rapid tissue turnover (Hulbert and Else, 2004).

Human studies in populations, varying in ecology and temperature, find evidence for a link between metabolic rate and cholesterol levels. Among native Siberians with basal metabolic rates (BMR) elevated above fat-free mass standards, BMR was inversely related to LDL-C levels (Leonard et al., 2008). The Pima of Arizona, United States, and Sonora, Mexico have low-BMRs and elevated cholesterol levels, and native Siberian populations with elevated BMRs have low-cholesterol levels (TC and LDL-C) (Leonard et al., 2008). Similarly, members of the US Navy subsisting on a standard diet increased their usual cholesterol consumption during temporary residence in Antarctica (Harford et al., 1993), perhaps indicating increased intake to compensate for increased cholesterol turnover and demand. These examples are consistent with the idea that high-metabolic rates and cellular turnover can strain the pool of circulating cholesterol in humans.

Higher demands for cholesterol with elevated metabolic rate might drive selection for higher circulating cholesterol concentrations, and thus, the ε4 allele. Indeed, very low-cholesterol levels could have broad and adverse functional and health impacts. In addition to its role as a membrane component (particularly in the central nervous system), cholesterol also serves as the precursor for steroid hormones (Dietschy and Turley, 2001; Vance et al., 2005). Lowered cholesterol levels are also likely to decrease the blood carrying capacity for fat-soluble vitamins, especially, vitamin E (Linton et al., 1993), and cholesterol appears to help maintain robust immune responses [reviewed in Jacobs and Iribarren (2000) and Muldoon et al. (1997)]. Low cholesterol has been related to altered cerebral development (Schoknecht et al., 1994; Pond et al., 2008), lower intelligence (Benton, 1995; Elias et al., 2005), and poorer memory (Henderson et al., 2003). In a piglet growth model, higher dietary cholesterol was associated with improved weight gain (Schoknecht et al., 1994).

Consistent with the idea that low-cholesterol levels can impair biological function, statins, a class of cholesterol lowering drugs, have been associated with decreased adrenal and sex steroid hormone levels, increases in corresponding stimulatory hormones [LH, FSH, and ACTH; reviewed in Kanat et al. (2007)] and decreased vitamin E levels (Galli and Iuliano, in press). Furthermore, low-cholesterol levels have been associated with decreased reproductive function in bovids (Anand and Prakash, 2008) and adverse birth outcomes in women (Edison et al., 2007). In addition, there is evidence that hypocholesteremia is associated with increased aggression, suicide, and accidents as well as decreased affiliative behavior (Kaplan et al., 1997; Lester, 2002; Zhang et al., 2005; Civelek et al., 2007; Lalovic et al., 2007). Cholesterol-lowering therapy among individuals with low-CHD risk has been associated with increased mortality rates (Smith et al., 1993). Consistent with this, low cholesterol has been associated with poorer recovery from physical disabilities (Onder et al., 2006), increased nonatherosclerotic deaths (Song et al., 2000), and increased all-cause mortality in prospective studies (Melton et al., 2006).

Although data are scarce, modern populations that subsist on wild game and foraged foods, as well as transitioning horticulturalists, have been found to have low-cholesterol levels compared to populations consuming a Western or higher fat diet (Miller et al., 1968; Mancilha-Carvalho and Crews, 1990; Kesteloot et al., 1997), suggesting that ancestral populations subsisting on similar foods were more likely confronted with low-cholesterol levels than the high levels typical of many contemporary societies. In light of the deleterious effects of very low cholesterol, environmental or climatic conditions that imposed especially high demands for circulating cholesterol could have selected for individuals with ε4 alleles, increasing ε4 frequencies in these populations.

Building from past work on *APOE* allele distributions in human populations, we seek to clarify the nature of this distribution and to evaluate the potential role of temperature as a selection pressure on this polymorphism. First, we evaluate the relationship of ε4 with latitude using an expanded dataset of *APOE* frequencies. Next, we used genome-wide data on a subset of populations [Human Genome Diversity Project (HGDP) populations] to see if *APOE* allele frequency trends are consistent with population structure, as an alternative to a

selection-based hypothesis. Finally, we examine associations with historical temperature, elevation, and other *APOE* alleles to more rigorously assess the role that temperature, as a proxy for metabolic rate, might have played in the evolution of the *APOE* polymorphism in modern humans.

We use these data to test the following hypotheses: [H1] ε4 allele frequency is increased in latitudes associated with higher metabolic rates; [H2] the worldwide distribution of ε4 frequencies is independent of population structure, thus supporting the idea that natural selection has created the latitudinal gradient. Under the assumption that metabolism is driving demand for higher cholesterol levels, we further hypothesize that [H3] temperature is more closely related to ε4 frequency than latitude and that [H4] increasing elevation, which causes colder temperatures, will also be associated with ε4 frequency. Because the locus under study is actually composed of three common alleles that influence cholesterol levels, we further predict [H5] that combining these alleles into one measure reflective of their association with cholesterol levels will reveal an even stronger association of *APOE* with absolute latitude than ε4 alone.

METHODS

Genetic dataset

Data on population *APOE* allele frequencies include those populations compiled by Singh and colleagues (2006), which were expanded in several ways: 1) we searched for allele frequencies in the ALFRED database posted as of 12/23/2007, 2) some errors in the preexisting dataset were corrected (e.g., geographical coordinates were made more specific/corrected, clinical populations were excluded), 3) elderly populations (mean age >65) were removed from all analyses because of past findings of decreasing ε4 alleles in elderly populations suggesting survival bias [reviewed in Smith (2002)], 4) ethnically heterogeneous populations were removed from all analyses, 5) ε4 allele frequencies (but not ε2 or ε3 frequencies, which were unavailable) from 53 populations (1,038 individuals) from the Human Genome Diversity Cell Line Panel (HGDP-CEPH; Cann et al., 2002) were added (Serre, 2008, personal communication; Serre et al., 2008), and 6) allele frequencies from new sources encountered in the course of research were added. Allele frequencies were selected without regard to genotyping methods, which included both direct genotyping and protein phenotyping methods. This is consistent with the previous worldwide analysis by Singh et al. (2006) and a meta-analysis by Bang et al. (2003). Furthermore, Song et al. (2004) showed that genotyping method had no effect on the conclusions reached by their meta-analyses of *APOE* effects.

Because recent migrants (generally resulting from colonial era population movements) to an area are not expected to have had enough time to genetically adapt to match the local optimum, they were excluded from all analyses. Where ethnicity was a clear signal of recent colonial or postcolonial era movements (e.g., white South-Africans), such samples were excluded from the analysis. Otherwise, ethnicity was not considered, except to mark distinct populations from the same locale as independent populations [e.g., Mexican Huicholes and Coras from the same location in Cruz-Fuentes et al. (2005)]. Clinical samples were excluded from all analyses, although randomly selected and healthy controls were included unless the healthy controls were selected for low-cholesterol levels. If two studies of the same population were available, only the one with the greater geographical specificity or larger sample size was included. Latitude and longitude coordinates for every population were determined from Google Earth and other maps.

Population structure control

The HGDP dataset (Cann et al., 2002) was used to control for possible influences of population structure on *APOE* ε4 allele frequency. Genotype data from the 660,918 markers derived from the HumanHap650Yv3 genotyping BeadChip (Illumina) were downloaded from CEPH (Human Genome Diversity Project, 2007). Plink version 1.05 (Purcell et al., 2007) and GLU version 1.0a5 (Jacobs, 2008) were used to manipulate the genotype data. The smartpca module of Eigensoft (version 2.0) was used for principal component analysis (PCA). We excluded sex chromosome and mitochondrial markers from being input into the PCA, leaving a total of 644,258 markers from which we derived the first 11 principal components (PCs). First, individuals were entered into the model, and then eigenvalues for each PC were averaged inside each population to create population level eigenvalues for each PC. We used the default control for linkage disequilibrium (LD) between two adjacent markers (Patterson et al., 2006; using values of 1, 3, and 4 yielded similar results). We used available data for the individuals recommended in the Rosenberg et al. "H971" subset of HGDP (2006), which excludes first degree relatives and obviously faulty samples. The Karitiana and Surui HGDP populations from South America have been noted as potentially problematic because of both a high degree of relatedness (Rosenberg, 2006) and being identified as outliers (Rosenberg et al., 2005). The Maya samples likely have substantial admixture (Rosenberg et al., 2005). Our analysis similarly showed the American populations as distinct, and because after the exclusion of the above three populations only two HGDP populations remained in the Americas, we excluded the Americas from the PCA. Populations with sample sizes less than or equal to five were also excluded from the PCA and regression analyses (San and Southern Bantu populations). In the primary analysis, the two populations from Oceania were excluded, because they were frequently outliers in PCAs and the small number of populations (two) in the sample from this ecologically distinct region of the world. After these exclusions, in the primary PCA analysis, there was no exclusion of individuals because of outlier eigenvalues. The final sample size was 845 individuals from 43 populations. In Supporting Information 1 and 2, two additional PCAs were conducted under slightly different parameters with generally similar results as those shown below.

Temperature and elevation dataset

An index of temperature histories was derived from the program GlobalTempSim Ver. 0.9 (Spokas, 2007). This program predicts average low and high-surface air temperatures for each day of the year for a specific geographical location based on 30 years (1961–1990) of worldwide temperature records. To reduce error, temperature data were retrieved in one batch using a custom-

ization of the GlobalTempSim program (customization programmed by James DeVona). Predicted low and high-air temperatures were averaged together and then averaged across all days of the year to derive a mean yearly surface air temperature for each location. We also calculated temperature variability by calculating the difference between the average high-daily temperature and average low-daily temperature as a measure of short-term temperature variability. By subtracting the lowest predicted temperature day of the year by the highest predicted temperature day, we derived an index of yearly temperature range (or seasonality).

Selection on the *APOE* locus likely occurred over many thousands of years, whereas our temperature data are from recent historical records. In an effort to correct recent temperature predictions for climatic differences in the past, we used two correction factors. The temperatures during the last glacial maximum (LGM), about 20,000 years ago, were characterized by a relatively greater latitudinal temperature gradient than today (Toracinta et al., 2004). That is, during the Pleistocene, while temperatures were colder globally, this cooling was greater at the poles than in the tropics. We corrected for this using estimates from different paleo-climate models. Specifically, we used the "Climate: Long-range Investigation, Mapping and Prediction study" (CL) model and an average (CL-AVG) of the CL, modified CL, and proxy models (Toracinta et al., 2004, p. 519). The mean yearly temperature variable used above was adjusted using the CL and CL-AVG models. For populations between 20°N and 20°S latitude 1.3°C was subtracted from the mean yearly temperature for the CL model and 3.73°C for the CL-AVG model. For populations between 20° and 40° latitude, 2.7°C and 4.63°C was subtracted. Finally, for those populations with latitudes greater than 40°, 7.1°C or 8.4°C was subtracted (for CL and CL-AVG, respectively).

Elevation data were retrieved from the USGS Seamless Elevation data sets (2008), using the highest-resolution data available and if not available there by consulting Google Earth. To reduce error, elevation data were retrieved in one batch from the USGS using a custom computer program [programmed by Jim DeVona (2008)].

Analytical method

In addition to using simple *APOE* allele frequencies, a combined measure (*APOE*-index) was constructed such that the approximate relative effects on cholesterol profiles are considered (Bennet et al., 2007), specifically: *APOE* index = $\varepsilon4$ frequency − $\varepsilon2$ frequency. Because $\varepsilon2$ and $\varepsilon4$ have approximately equal and opposite magnitude effects on cholesterol levels (relative to $\varepsilon3$) and exhibit an additive dose-response pattern, the *APOE*-index approximates the phenotypic effect of the *APOE* locus on population level lipid phenotypes (LDL, HDL, TC, and VLDL; Mahley et al., 2000). If *APOE*'s influence on cholesterol levels are in fact under simple and linear selection and mating is random with respect to *APOE* genotype, this index should encompass the relevant functional variation better than considering the allele frequencies individually. Higher values on this index should approximate higher population level serum cholesterol levels. Only $\varepsilon4$ frequencies were available for the HGDP sample, so the *APOE*-index could not be constructed for these populations nor could the HGDP populations be included in secondary analysis of $\varepsilon2$ and $\varepsilon3$ allele frequencies (Table 5, models 2–4). Allele frequencies, mainly of $\varepsilon4$, are considered as the outcome variables in OLS regression models run in STATA/IC 10 with an alpha of 0.05. All models were tested for heteroscedasticity using the Breusch–Pagan/Cook–Weisberg test and calculated with robust standard errors as indicated.

RESULTS

Description of dataset and variables

Our literature search yielded a dataset containing allele frequencies from 430 populations. Of those, 321 met the inclusion criteria (see methods) and 268 of these were unique populations used for the remainder of the analysis (median population sample size of 94 individuals; range of 8–3,067 individuals per population; and total number of subjects across all 268 studies was 61,035). The locations and $\varepsilon4$ allele frequencies of these 268 populations are shown in Figure 1. Forty-three of these are HGDP populations with genome-wide marker data, which permit us to control for population structure. However, these HGDP populations only have allele frequencies for $\varepsilon4$ and are missing data for the $\varepsilon2$ or $\varepsilon3$ alleles. The remaining 225 populations are considered "non-HGDP" populations. When possible, all 268 populations were analyzed together, but some analyses were limited to only HGDP or non-HGDP populations as noted below. Descriptions of the key genotype and environmental variables are given in Table 1 and their Pearson's pairwise correlations in Table 2. Absolute latitude was significantly correlated with all the climate variables and often strongly so.

Hypothesis testing

H1: $\varepsilon4$ allele frequency is increased in latitudes associated with higher metabolic rates. It has long been observed that $\varepsilon4$ allele frequencies increase with latitude (Gerdes et al., 1992), although a more recent analysis showed more ambiguous global patterns (Singh et al., 2006). Graphical analysis of our expanded dataset revealed that the relationship between latitude and $\varepsilon4$ frequencies is instead curvilinear (see Fig. 2). We modeled the curvilinear relationship by including both latitude and squared latitude terms in regression models (Table 3, models 1, 6, and 7; Tables 4 and 5). Population $\varepsilon4$ allele frequencies first decrease with distance from the equator, reach an inflection point (nadir) at roughly 35°, and increase again at higher latitudes (Table 3, model 1; Fig 2). The curvilinear relationship is consistent with the hypothesis that higher metabolic rate has selected for the $\varepsilon4$ allele: metabolic rates are lowest in moderate climates, but increase in high temperatures as well as low temperatures (reviewed below in discussion).

H2: The association of $\varepsilon4$ with absolute latitude is not due to underlying population structure (and thus likely due to natural selection). The association between $\varepsilon4$ and absolute latitude might be an artifact of human population histories rather than selection. Using the non-HGDP dataset, we controlled for the effect of world region, and the association with absolute latitude was little changed (Table 4, model 1 compared to model 2). To more rigorously assess whether the latitudinal gradient of $\varepsilon4$ is due to nonselective factors, we considered population structure using PCA across genome-wide data from 845 individuals from the 43 HGDP populations (see Methods section). This subset of HGDP populations had a relationship between $\varepsilon4$ and absolute

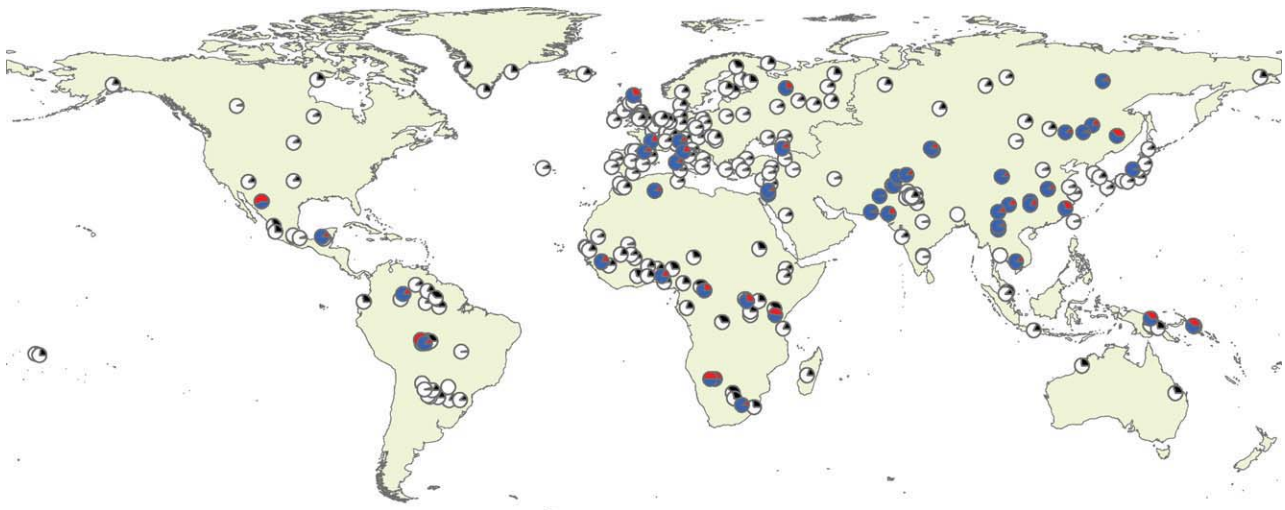

Fig. 1. Map of populations used in analyses. Each pie chart is a population-specific allele frequency estimate for ε4. The colored pie charts represent HGDP populations, with red representing percent ε4. The black and white pie charts are non-HGDP populations with black representing percent ε4.

TABLE 1. Descriptive statistics for key dependent and independent variables

Variable	N[a]	Mean	SD	Min	Max
ε2 (% frequency)	225	6.4	5.1	0.0	37.5
ε3 (% frequency)	225	78.3	12.1	8.5	98.0
ε4 (% frequency)	268	14.5	8.5	0.0	49.0
APOE-index[b]	225	8.1	9.6	−25.0	45.0
Mean temp (°C)	267	14.6	9.3	−11.6	28.5
Yearly temp range (°C)	267	27.7	10.8	8.5	65.2
Daily temp range (°C)	267	10.5	2.8	1.9	19.7
Elevation (m)	268	502.9	681.0	0.0	4139.0

[a] N, number of populations (not number of individuals).
[b] APOE-index = ε4 frequency − ε2 frequency. The index should approximate the effect of the APOE locus on cholesterol levels.

latitude that was similar to the remaining 225 non-HGDP populations (Fig. 2; Table 4, compare model 1 to model 3). Each PC represents a different underlying statistical construct reflecting shared population covariance evaluated across the 644,528 polymorphisms examined. As consistently observed in other studies using this method (Cavalli-Sforza, 2005; Lopez Herraez et al., 2009), the first two PCs accounted for the largest proportion of the shared variance in allele frequencies, with each subsequent component contributing only minimally. These first two components were included in a regression model in an effort to see if the pattern of association between ε4 and latitude was thereby changed (Table 4, model 4). The first four PCs are included heuristically (Table 4, model 5), and finally only those PCs significantly related to ε4 frequency in pairwise correlations (PC1: $r = 0.41$, $P = 0.007$; PC5: $r = -0.31$, $P = 0.041$; PC11: $r = -0.31$, $P = 0.041$) were included (Table 4, model 6). Although P values of absolute latitude and latitude-squared did increase when controlling for population structure, the beta values and inflection points remained little changed. Given the small sample size and increased degrees of freedom when introducing population structure controls, increased P values are not surprising. The results suggest that population variation in ε4 frequency is not a reflection of population structure and was instead likely shaped by selection.

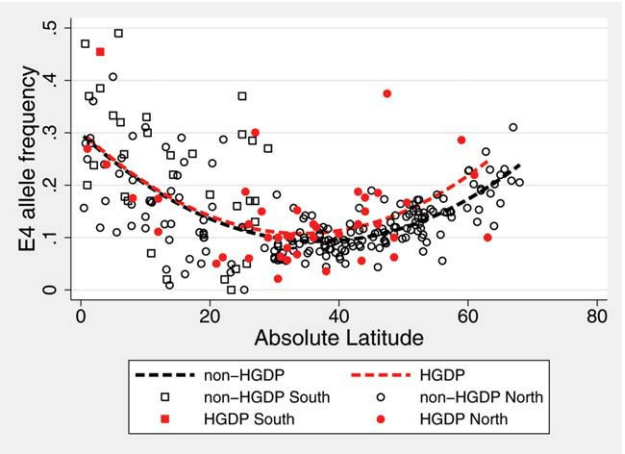

Fig. 2. ε4 allele frequencies (Y axis) by absolute latitude (X axis). Red filled in circles/squares represent HGDP populations, and black-opened squares/circles are non-HGDP populations. Squares are populations South of the Equator, and circles are North of the Equator. Curved lines are quadratic best fit lines. Red solid curve is for HGDP populations and black-dashed curve is for the non-HGDP populations. See Table 4, models 1 and 3 for equations of curves.

H3: Temperature will be a stronger predictor of ε4 frequency than latitude. We predicted that the measure of average temperature would be a better proxy for temperatures during our evolutionary past than absolute latitude and thus a better predictor of metabolic rate and ε4 frequency. As apparent from comparing Figures 2 and 3 and corroborated by regression analyses (Table 3, model 1 compared to model 2), latitude is a stronger predictor of ε4 frequency than is temperature. As a result of this finding contrary to our expectation, we explored some reasons why these data, reflecting the average interpolated yearly temperatures derived from data from 1961 to 1990, might be a weak predictor of past metabolic demands of the climate. Using corrections for differences in climate during the LGM (see Methods section), we found that the models improved substantially

TABLE 2. Pairwise Pearson correlations between key dependent and independent variables[a]

Variable	ε2	ε3	ε4	Index	Temp	ytemp[b]	dtemp[b]	Elevation
ε2 (% frequency)	1							
ε3 (% frequency)	**−0.46**	1						
ε4 (% frequency)	0.06	**−0.68**	1					
APOE-index	**−0.47**	**−0.36**	**0.85**	1				
Mean temp (°C)	0.09	−0.10	0.07	0.02	1			
Yearly temp range (°C)	**−0.16**	**0.24**	**−0.27**	**−0.17**	**−0.63**	1		
Daily temp range (°C)	−0.11	−0.03	0.00	0.11	**0.31**	**0.23**	1	
Elevation (m)	−0.10	−0.01	0.05	**0.14**	−0.06	0.02	**0.44**	1
Absolute latitude[c]	−0.02	**0.19**	**−0.26**	**−0.24**	**−0.85**	**0.59**	**−0.42**	**−0.24**

[a] Bolded correlation coefficients have $P < 0.05$.
[b] ytemp stands for yearly temperature range and dtemp for daily temperature range.
[c] Absolute latitude is the number of degrees north OR south of the equator.

TABLE 3. Primary regression output[a]

Model number	1	2	3	4	5	6	7
Model description	Baseline	Temp	Temp CL	Temp CL-AVG	Temp + range	Lat + range	+Elevation[b]
Absolute latitude[c]	−0.011***					−0.011***	−0.010***
Latitude^2	0.00015***					0.00015***	0.00013***
Temperature (Kelvin)		−0.101***	−0.094***	−0.074***	−0.159***		
Temperature^2		0.00018***	0.00017***	0.00013***	0.00027***		
Temperature: mean daily range					0.008*	0.003	
Temperature: yearly range					−0.005***	−0.001	
Constant	0.297***	14.500***	13.240***	10.711***	23.527***	0.276***	0.317***
Inflection point (nadir)[d]	36.7°	10.79°C	6.66°C	16.40°C	19.75°C	35.25°	40.7°
Observations	268	267	266	266	267	267	268
Adjusted R^2	0.392[e]	0.038	0.063	0.109	0.199	0.392	0.272

[a] All models are heteroscedastic and calculated with robust standard errors; models 2 and 5 use mean annual temperatures based upon climate records from 1961 to 1990; model 3 modifies temperatures in model 2 with the CL correction for the last glacial maximum; model 4 uses the CL-AVG correction for the last glacial maximum; *$P < 0.05$, ***$P < 0.001$.
[b] Model 7 adjusts absolute latitude by elevation. In a regression model predicting mean annual temperature, the coefficient of elevation (in meters) is 0.82% of the coefficient of absolute latitude (in degrees). As a result, latitude is adjusted by adding 0.82% of the elevation of the location.
[c] Absolute latitude is the number of degrees north OR south of the equator.
[d] Inflection point of latitude or temperature predicting APOE allele frequencies in curvilinear models; inflection points in models 1, 6, and 7 are in degrees absolute latitude; inflection points in models 2–5 are in temperature (°C).
[e] Nonadjusted $R^2 = 0.397$.

TABLE 4. Regressions assessing population structure[a]

Model number	1	2	3	4	5	6	7
Model description	Non-HGDP	Non-HGDP	HGDP	HGDP + PCA	HGDP + PCA	HGDP + PCA	Non-HGDP ε2
Absolute latitude	−0.011***	−0.009***	−0.011***	−0.009	−0.010	−0.008*	−0.001
Latitude^2	0.00015***	0.00013***	0.00017***	0.00014*	0.00015	0.00010	0.00001
Europe		0.017**					0.014*
North America		0.031					−0.054***
South America		0.031					−0.056***
Africa		0.066***					0.031*
Oceania		0.162***					0.035
PC1				0.420	0.349	0.140	
PC2				0.136	0.081		
PC3					−0.146		
PC4					−0.018		
PC5						−0.909	
PC11						−0.524	
Constant	0.298***	0.226***	0.302***	0.256*	0.268*	0.269***	0.071***
Inflection point (nadir)[b]	36.9°	34.01°	34.21°	32.1°	32.77°	38.75°	38.06°
Observations	225	225	43	43	43	43	225
Adjusted R^2	0.417	0.518	0.285	0.255	0.216	0.322	0.282

[a] Models 1, 2, and 7 are heteroscedastic and run with robust standard errors, whereas models 3–6 are homoscedastic; models 1 and 2 are run on non-HGDP populations only, whereas models 3–6 are run on HGDP populations only; model 2 controls for world-region with an out-group of Asia; model 7 is this same as model 2, but the dependent variable is ε2 rather than ε4; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.
[b] Inflection point of latitude predicting APOE allele frequencies in curvilinear models.

TABLE 5. Regressions examining different APOE alleles correlated with latitude[a]

Model number	1	2	3	4
Model description	ε4	APOE-index	ε2	ε3
Absolute latitude	−0.011***	−0.011***	−0.000	0.010***
Latitude^2	0.00015***	0.00015***	0.00000	−0.00015***
Constant	0.298***	0.229	0.069	0.642***
Inflection Point (nadir)[b]	36.9°	36.84°	38.78°	37.37°
Observations	225	225	225	225
Adjusted R^2	0.417	0.301	−0.008	0.165

[a] All models besides model 4 are heteroscedastic and calculated with robust standard errors; all models use only populations for which allele information for ε2, ε3, and ε4 are available; model 8's dependent variable is ε4, model 9's is the combined APOE-index, model 10's is ε2, and model 11's ε3; ***$P < 0.001$.
[b] Inflection point of latitude predicting APOE allele frequencies in curvilinear models.

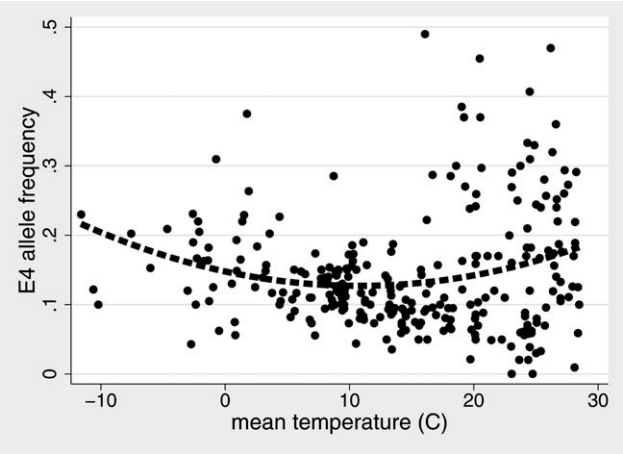

Fig. 3. ε4 allele frequency (*Y* axis) by mean temperature in °C (*X* axis). Dots represent populations and curve is a quadratic best fit line. See Table 3, model 2 for equation of curve.

from an adjusted R^2 of 0.038 to between 0.063 and 0.109 depending on the correction factor (Table 3, models 2–4). This suggests that recently recorded temperature measures might be inadequate proxies for temperature during our evolutionary past.

Humans and other animals have abilities to acclimate to temperature and find variable temperatures more energetically stressful than less variable conditions (Davis, 1964; Al-Haboubi, 1996). As a result, it might be that increased temperature variability, not just mean temperature, places increased metabolic demands on individuals. Incorporating temperature variability did considerably increase the predictive power of the model (Table 3, model 5 compared to model 2). The direction of the association was significant and in the expected direction for daily temperature variability, but in the opposite direction of expected for yearly temperature variability. That is, with an increased daily temperature range ε4 frequencies increased, while an increased range of temperatures for an average year predicted decreases in ε4 frequencies. Adding temperature variability measures to a model that includes latitude did not increase the explanatory power of the latitude model nor were the measures of temperature variability significant (Table 3, model 6 compared to model 1). Mean yearly temperature and yearly temperature range are strongly inversely correlated (Table 2; $r = -0.63$, $P < 0.0001$)—regions with lower temperatures had higher yearly temperature ranges. Similarly, absolute latitude and yearly temperature variability are directly correlated (Table 2; $r = 0.59$, $P < 0.0001$)—higher latitudes have higher yearly temperature ranges. However, this association between latitude and yearly temperature range is considerably more marked at latitudes below 35° (<35°: $r = 0.84$, $P < .0001$; >35°: $r = 0.14$, $P = 0.10$). Together, this suggests that the relationship between absolute latitude and ε4 worldwide incorporates the association between ε4 and temperature variability.

H4. Increasing elevation, by causing colder temperatures, will be associated with ε4 allele frequencies. To further examine the hypothesis that increasing metabolic rate drives increasing ε4 allele frequencies, we examined the association between elevation and ε4 frequencies. Based on the curvilinear association between ε4 and absolute latitude and the fact that high temperatures as well as low temperatures increase energy expenditure, modeling the effects of elevation on ε4 frequencies is a complex matter. For example, in low latitudes, the cooling effect of increasing elevation on temperature is expected to decrease metabolic rate up to a certain elevation, above which cold temperatures will increase requirements for thermogenesis and increase metabolic expenditure. That is, the thermoneutral zone is contingent not just upon latitude, but also elevation.

To evaluate the possible moderating influence of elevation, we constructed an index which puts elevation on the same scale as latitude by incorporating the effect of elevation on temperature. First, a multivariate regression model was used to compare the effects of absolute latitude and elevation on mean annual temperature across the locations used in this study. Both latitude and elevation were highly significantly related to temperature ($P < 0.001$), and the beta-coefficient of elevation (in meters) was 0.82% of the beta-coefficient of absolute latitude (in degrees latitude; model not shown). In other words, a one meter increase in elevation is equivalent in effects on temperature change to a 0.082° latitude increase. Based on these results, an index of absolute latitude and elevation was created as follows: index = absolute latitude + 0.0082 * elevation. This index estimates the expected effect of elevation on average yearly temperature for given latitude by putting elevation on a latitudinal scale. A squared term of the index was also created to model the expected curvilinear relationship between temperature/latitude and metabolic rate. This index and its squared term are expected to increase the model fit by more accurately predicting the temperature related metabolic expenditure at each location. Contrary to this expectation, the results showed that the adjusted

R^2 decreased after the inclusion of elevation in the model (Table 3, model 1 compared to model 7). Reasons for this result, including adaptations to hypoxia, are discussed below in the discussion.

H5: A combined APOE measure, more strongly predictive of cholesterol levels, will be more strongly related to latitude than ε4 alone. If cholesterol levels are the selectively relevant phenotype, there is reason to expect that a combined index of *APOE* alleles that considers the effects of all alleles on cholesterol simultaneously, rather than one at a time, will be more closely related to climate variables than the allele frequency of any one allele. That is, because ε2 is associated with decreased cholesterol levels and ε4 increased cholesterol levels (both relative to ε3) in an allele dose-response manner (Bennet et al., 2007), the combined allele frequencies of ε2 and ε4 should more closely approximate the effect of *APOE* on cholesterol levels. Contrary to this, ε4 was more closely related to absolute latitude than the combined *APOE* index (Table 5, model 2 compared to model 1; using only non-HGDP populations). Similarly, ε2 and ε3 allele frequencies were also less strongly related to latitude than was ε4 (Table 5, models 3 and 4 compared to model 1). Post hoc analysis, aimed at finding the determinants ε2 allele frequencies, suggests that 28% of population ε2 allele frequency variability was accounted for by continental origin of the population (Table 4, model 7).

DISCUSSION

Our results suggest strong geographical patterning of ε4 allele frequencies that does not mimic population structure. With respect to absolute latitude, ε4 frequencies show a curvilinear relationship with a high frequency near the equator, decreasing to a nadir in mid-latitudes (about 35°) and then increasing again closer to the poles (see Fig. 2). Because metabolic rates are elevated in hot and cold environments, this pattern is broadly consistent with the hypothesis that elevated metabolic rate has selected for the ε4 allele. The associations of ε4 with temperature and elevation as well as the association of other *APOE* alleles with latitude provide mixed support for our primary hypothesis. First, contrary to our expectations, absolute latitude was a stronger predictor of ε4 frequencies than was temperature. Using a correction factor that compensates for temperature changes during the Pleistocene roughly doubled the explanatory power of temperature, hinting at limitations of using historical records to infer selective processes that occurred over longer timescales. Second, including the effects of elevation on climate, in addition to latitude, did not improve the variance explained in ε4 frequencies. Finally, adding consideration of the phenotypic effect of the ε2 allele, which reduces cholesterol levels, and thus by our hypotheses should have been subjected to opposing gradients of selection did not improve the model fit with absolute latitude. Thus, our analyses provide mixed support for our hypothesis that temperature-driven differences in metabolic rate would have influenced requirements for cholesterol, thereby selecting for *APOE* alleles in human populations.

This is the first study to rigorously control for population structure while examining the worldwide patterning of *APOE* allele frequencies. Our results suggest that the observed association of ε4 with latitude is not due to population structure and adds additional evidence that natural selection has shaped population differences in ε4 frequencies.

ε4 frequencies follow a curvilinear relationship with absolute latitude, consistent with the tendency for metabolic expenditure to be elevated in both hot and cold climates. For humans, the thermoneutral zone at rest with minimal clothing is between ~25 and 31°C (77–89°F; e.g. Davis, 1964). ε4 frequencies are lowest at 35° absolute latitude, corresponding to a mean annual temperature of 14°C (57°F). This seems a reasonable approximation of a thermoneutral temperature given that human populations were often physically active and used other means of staying warm such as clothing, fire, and dwellings. For example, one thick layer of clothing compensates for a 10°C decrease in temperature (Hassi et al., 2005), sufficient to decrease the thermoneutral zone from 25–31°C to 15–21°C (59–70°F). These behavioral means serve to broaden the effective thermoneutral zone, but primarily toward colder, not hotter temperatures. Thus, the curvilinear latitudinal gradient in ε4 frequency and the latitudinal nadir of this gradient are consistent with the hypothesis that latitudes associated with high-metabolic rates would have selected for the ε4 allele.

Consistent with the notion that high temperatures increase metabolism and thus demands for cholesterol, high-heat exposure in humans (39.8°C or 103.6°F) has been found to be associated with decreases in serum TC, LDL-C, and TG levels, consistent with the depletion of peripheral cholesterol stores (controlling for plasma volume; Yamamoto et al., 2003). However, other studies not controlling for plasma volume do not show this trend (Keatinge et al., 1986; Al-Harthi et al., 1990). Similarly, lactating cows show decreases in TC in hot periods of the summer (Abeni et al., 2007). In river buffalo in India, the hot humid season is associated with the decreased cholesterol and TG levels compared to the more mild winter months, which decrease the available substrate for ovarian steroidogenesis and result in low-estrus behavior and conception rates at this time of year (Anand and Prakash, 2008). Thus, although relevant findings are relatively few, past studies suggest that extreme temperatures—whether hot or cold—can elevate metabolic rate and the demand for cholesterol in mammals, conditions that could select for the ε4 allele.

Although our findings are generally supportive of an ε4 gradient that parallels latitudinal changes in temperature, historical temperature records were a poor predictor of ε4 compared to latitude alone. It is notable, however, that, for many populations, modern *APOE* allele frequencies will have been shaped by local selection operating over thousands of years or longer. Thus, historical temperature records, which reflect only the latest Holocene interglacial conditions, may be poor proxies for average temperatures experienced by populations over long time scales, given that glacial conditions have predominated in the Quaternary. For instance, during the LGM (~20,000 years ago), it is estimated that temperatures were ~7°C colder than present in regions greater than 40° latitude, whereas the equatorial regions (20°N–20°S) were estimated to be only ~2°C colder (Toracinta et al., 2004). We adjusted our measure of historical temperatures for this and found that this correction improved the model fit substantially, roughly doubling the fraction of variance in ε4 frequency explained. Our correction for past temperature was coarse and assumed uniform temperature changes within broad latitudinal belts. This assumption is unrealistic; climate change

exhibits strong regional variability and it is known, for instance, that the magnitude of glacial-to-interglacial temperature changes is modulated by factors such as proximity to oceans and ice sheets. The fact that our correction improved model fit, despite these limitations, provides an incentive to pursue improved strategies to model past temperature as a selective force on human populations. Factors that might be important to consider include locale-specific estimates of temperature history, the possible modulating influence of humidity, and population migratory history. In addition, if metabolic rate is a driver of cholesterol requirements, the strength of selection related to temperature might also be expected to vary by season, with strongest selection in the coldest winter months at high latitudes and strongest selection among heat-selected populations during months of maximal temperature.

Including the effect of elevation on temperature decreased the variance explained in $\varepsilon 4$ frequency—contrary to what we predicted based upon the metabolic rate hypothesis (Table 3, model 7 compared to model 1). The reasons for this counter-intuitive result are not completely clear. One explanation is that daily temperature variability is greater at higher altitudes (Table 2; $r = 0.44$, $P < 0.05$). Another possible explanation is that while elevation is inversely correlated with temperature, it is also related to atmospheric pressure, which engages another suite of selective pressures on human physiology. Adaptations to hypoxia are complex and have involved distinct mechanisms in different populations (Beall, 2007). Aitbaev (1985) suggested that hypoxia might account for differences in serum cholesterol levels observed in different populations in the Kirghiz Alpine. In some populations, increasing elevation has been related to increased HDL-C levels (Coello et al., 2000; Mohanna et al., 2006), while, in another, no difference in serum cholesterol levels were observed (Fiori et al., 2000), and one study found low levels of HDL-C in a high-altitude population (Bellido et al., 1993; cited in Mohanna et al. (2006)). Differential susceptibility of different APOE isoforms to oxidative stress (Leininger-Muller et al., 1998), which may increase with altitude (Askew, 2002) might also explain the results. Although it is difficult to parcel out the effects of cold, hypoxia, and variable nutrition across populations, diverse means of adaptation to hypoxia might differentially influence the fitness value of $\varepsilon 4$.

Finally, if increased metabolic rate drives demands for cholesterol, we expect regions with higher metabolic rate not only to exert positive selection for the $\varepsilon 4$ allele, which increases circulating cholesterol, but also to select against $\varepsilon 2$, which is associated with lower cholesterol (Mahley et al., 2000; Bennet et al., 2007). In contrast to this expectation, $\varepsilon 2$ frequency showed no linear or curvilinear association with absolute latitude (Table 5, model 3), nor did using a combined index of *APOE* allele frequencies scaled to the predicted effect on cholesterol levels (Table 5, model 2) show a stronger relationship with absolute latitude than $\varepsilon 4$ alone (Table 5, model 1). This suggests that the $\varepsilon 2$ allele distribution is explained by factors other than those that explain $\varepsilon 4$. $\varepsilon 2$ homozygosity is associated with increased TG levels (Bennet et al., 2007), which might cause different factors to be selectively relevant. Despite being generally associated with low cholesterol, $\varepsilon 2$ homozygotes are also at increased risk for pathologically high-cholesterol levels, manifesting as part of a syndrome known as type III hyperlipoproteinemia (HLP) (Mahley et al., 2000). Type III HLP is likely caused by epistatic interactions or linkage of $\varepsilon 2$ with rare genetic polymorphisms that vary by population (Fullerton et al., 2000; Mahley et al., 2000; Nickerson et al., 2000) and thereby changes the fitness value of $\varepsilon 2$ in a complex and population dependent manner. It is also possible that $\varepsilon 2$ frequencies vary for nonadaptive reasons as the result of drift or underlying population structure. Although we do not have the $\varepsilon 2$ allele frequencies in the HGDP sample to test this hypothesis rigorously, continental origin accounts for 28% of the population variability in population level $\varepsilon 2$ allele frequencies (Table 4, model 7).

Other factors besides metabolic rate should be explored to determine if they explain *APOE* allele frequencies. It has been proposed that *APOE* allele frequencies could be shaped by the selection for the $\varepsilon 4$ allele due to protection it provides against infectious diseases (Martin, 1999; Wozniak et al., 2002, 2007; Oria et al., 2007). Others have argued that the $\varepsilon 4$ allele was selected against, because it is a risk factor for other infections (Allen et al., 2005; Ewald, 2008). A few studies have found associations between malaria susceptibility and the *APOE* locus studied here, but the results have not been consistent (Wozniak et al., 2003; Aucan et al., 2004; Vignali et al., 2008). *APOE* genotypes have also been found to correlate with the severity of symptoms of infections, but again studies are sparse and not wholly consistent (Corder et al., 1998; Wozniak et al., 2002; Itzhaki et al., 2004; Oria et al., 2005; Burt et al., 2008). As a result, these studies do not yet yield coherent predictions for comparative analyses.

Because distinguishing between these hypotheses more definitively will require rigorous empirical analysis, beyond the scope of the current analysis, using additional environmental/cultural variables, the dataset used to conduct the present analysis is made available as Supporting Information 3. Fruitful avenues of future analysis might include generating more fine scaled population-specific paleo-climate estimates, examining the worldwide distribution of other genetic polymorphisms related to cholesterol function, using indices of environmental pathogen loads (Murray and Schaller, 2010), and/or controlling for dates of origin of agriculture and animal domestication [as in Putterman (2008)] or subsistence strategy [as in Porter and Marlowe (2007)] to see if dietary stability or population density explain *APOE* distributions.

ACKNOWLEDGMENTS

Thanks to Bill Leonard, Thom McDade, Julienne Rutherford, Marco Aiello, Jared Bragg, and Aryeh Jacobsohn for valuable discussion and comments; Yarrow Axford for paleoclimate advice and related citations; Ricardo Godoy for econometrics consultations and discussions; David Serre and colleagues for sharing unpublished allele frequencies from the HGDP-CEPH samples; Sarabjit Mastana for sharing the digital dataset from his worldwide *APOE* analysis; Robert Mahley for sharing unpublished allele frequencies by region in Turkey; Svetlana Borinskaya for sharing unpublished data; Kurt Spokas for his fine climate software and support; Kevin Jacobs for help with glu; Nick Patterson for help with Eigenstrat; and Jim DeVona for his customization of Kurt Spokas's software and programming the batch USGS elevation query module. DTAE is funded by a

National Science Foundation Graduate Research Fellowship and is funded by a National Science Foundation Graduate Research Fellowship and had research assistance from the Newton Free Library, Newton, Massachusetts. An earlier version of this analysis was presented to at the 2009 Human Biology Association meetings and published in abstract form (Eisenberg et al., 2009).

LITERATURE CITED

Abeni F, Calamari L, Stefanini L. 2007. Metabolic conditions of lactating Friesian cows during the hot season in the Po valley. I. Blood indicators of heat stress. Int J Biometeorol 52:87–96.

Aitbaev KA. 1985. The levels of high density lipoprotein cholesterol and other lipids in the native population of the mountain region of Kirghizia. Vopr Med Khim 31:58–61.

Al-Haboubi MH. 1996. Energy expenditure during moderate work at various climates. Int J Ind Ergon 17:379–388.

Al-Harthi SS, Karrar O, Almashhadani SA, Saddique AA. 1990. Metabolite and hormonal profiles in heat-stroke patients at mecca pigrimage. J Intern Med 228:343–346.

Allen JS, Bruss J, Damasio H. 2005. The aging brain: the cognitive reserve hypothesis and hominid evolution. Am J Hum Biol 17:673–689.

Anand LN, Prakash BS. 2008. Comparative profiles of different lipoprotein cholesterol parameters and growth hormone during hot humid and winter season in Murrah Buffaloes. Livest Sci 115:294–300.

Askew EW. 2002. Work at high altitude and oxidative stress: antioxidant nutrients. Toxicology 180:107–119.

Aucan C, Walley AJ, Hill AVS. 2004. Common apolipoprotein E polymorphisms and risk of clinical malaria in the Gambia. J Med Genet 41:21–24.

Austad SN, Finch CE. 2008. Chapter 23: the evolutionary context of human aging and degenerative disease. In: Stearns SC, Koella JC, editors. Evolution in health and disease, 2nd ed. Oxford: Oxford University Press. p 301–311.

Bang OY, Kwak YT, Jool IS, Huh K. 2003. Important link between dementia subtype and apolipoprotein E: a meta-analysis. Yonsei Med J 44:401–413.

Beall CM. 2007. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proc Natl Acad Sci USA 104 (Suppl 1):8655–8660.

Becher JC, Keeling JW, McIntosh N, Wyatt B, Bell J. 2006. The distribution of apolipoprotein E alleles in Scottish perinatal deaths. J Med Genet 43:414–418.

Bellido D, Barragán M, Aguilar M, Colque N. 1993. Perfil lipídico de la población de Viacha. Acta Andina 2:28–29.

Bennet AM, Di Angelantonio E, Ye Z, Wensley F, Dahlin A, Ahlbom A, Keavney B, Collins R, Wiman B, de Faire U, Danesh J. 2007. Association of apolipoprotein E genotypes with lipid levels and coronary risk. JAMA 298:1300–1311.

Benton D. 1995. Do low-cholesterol levels slow mental processing. Psychosom Med 57:50–53.

Borinskaya SA, Kal'ina NR, Sanina ED, Kozhekbaeva ZM, Gupalo EY, Garmash IV, Ogurtsov PP, Parshukova ON, Bojko SG, Veselovsky EM, Vershubskaya GG, Kozlov AI, Rogaev EI, Yankovsky NK. 2007. Polymorphism of the apolipoprotein E gene (APOE) in the populations of Russia and neighboring countries. Russ J Genet 43:1201–1207.

Burt TD, Agan BK, Marconi VC, He W, Kulkarni H, Mold JE, Cavrois M, Huang Y, Mahley RW, Dolan MJ, McCune JM, Ahuja SK. 2008. Apolipoprotein (apo) E4 enhances HIV-1 cell entry in vitro, and the APOE ε4/ε4 genotype accelerates HIV disease progression. Proc Nat Acad Sci 105:8718–8723.

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu JY, Carcassi CY, Contu L, Du RF, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang XY, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian YP, Shu QF, Xu JJ, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL. 2002. A human genome diversity cell line panel. Science 296:261–262.

Cavalli-Sforza LL. 2005. The Human Genome Diversity Project: past, present and future. Nat Rev Genet 6:333–340.

Civelek T, Turan MK, Birdane FM, Demirkan I. 2007. Serum lipid concentrations in military patrol dogs trained for controlled aggression. Revue Med Vet 158:7–12.

Coello SD, Cabrera De Leon A, Bosa Ojeda F, Perez Mendez LI, Diaz Gonzalez L, Aguirre-Jaime AJ. 2000. High density lipoprotein cholesterol increases with living altitude. Int J Epidemiol 29:65–70.

Corbo RM, Scacchi R. 1999. Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele? Ann Hum Genet 63:301–310.

Corder EH, Robertson K, Lannfelt L, Bogdanovic N, Eggertsen G, Wilkins J, Hall C, Gy. 1998. HIV-infected subjects with the E4 allele for APOE have excess dementia and peripheral neuropathy. Nat Med 4:1182–1184.

Cruz-Fuentes CS, Gonzalez-Sobrino BZ, Gomez-Sanchez A, Rueda HM, Chavez-Eakle RA, Sanchez CS. 2005. Distribution of apolipoprotein E alleles in Coras and Huicholes from Nayarit and Nahuas and Mestizos from Veracruz, Mexico. Hum Biol 77:867–872.

Cuthbert JA, Lipsky PE. 1987. Provision of cholesterol to lymphocytes by high-density and low-density lipoproteins-requirement for low-density-lipoprotein receptors. J Biol Chem 262:7808–7818.

Davis TR. 1964. The influence of climate on nutritional requirements. Am J Public Health Nations Health 54:2051–2067.

Dietschy JM, Turley SD. 2001. Cholesterol metabolism in the brain. Curr Opin Lipidol 12:105–112.

Dietschy JM, Turley SD. 2004. Cholesterol metabolism in the central nervous system during early development and in the mature animal. J Lipid Res 45:1375–1397.

Dietschy JM, Turley SD, Spady DK. 1993. Role of liver in the maintenance of cholesterol and LDL homeostasis in different animal species, including humans. J Lipid Res 34:1637–1659.

Edison RJ, Berg K, Remaley A, Kelley R, Rotimi C, Stevenson RE, Muenke M. 2007. Adverse birth outcome among mothers with low serum cholesterol. Pediatrics 120:723–733.

Eisenberg D, Hayes M, Kuzawa C. 2009. Evidence for cholesterol function and local climatic adaptation based on worldwide allele frequency variations of the apoliprotein E (APOE) gene. Am J Hum Biol 21:253.

Elias PK, Elias MF, D'Agostino RB, Sullivan LM, Wolf PA. 2005. Serum cholesterol and cognitive performance in the Framingham Heart Study. Psychosom Med 67:24–30.

Ewald P. 2008. An evolutionary perspective on the causes of chronic diseases: atherosclerosis as an illustration. In: Trevathan W, Smith EO, McKenna JJ, editors. Evolutionary medicine and health: new perspectives. New York: Oxford University Press. p 350–367.

Finch CE. 2007. The biology of human longevity: inflammation, nutrition, and aging in the evolution of life spans , Vol.xiii. Burlington, MA: Academic Press.626p.

Finch CE, Stanford CB. 2004. Meat-adaptive genes and the evolution of slower aging in humans. Quart Rev Biol 79:3–50.

Fiori G, Facchini F, Pettener D, Rimondi A, Battistini N, Bedogni G. 2000. Relationships between blood pressure, anthropometric characteristics and blood lipids in high- and low-altitude populations from Central Asia. Annals Hum Biol 27: 19–28.

Froehle AW, Schoeninger MJ. 2006. Intraspecies variation in BMR does not affect estimates of early hominin total daily energy expenditure. Am J Phys Anthropol 131:552–559.

Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, Stengard JH, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF. 2000. Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. Am J Hum Genet 67:881–900.

Galli F, Iuliano L. In press. Do statins cause myopathy by lowering vitamin E levels? Med Hypotheses.

Gaynor JW, Wernovsky G, Jarvik GP, Bernbaum J, Gerdes M, Zackai E, Nord AS, Clancy RR, Nicolson SC, Spray TL. 2007. Patient characteristics are important determinants of neurodevelopmental outcome at one year of age after neonatal and

infant cardiac surgery. J Thorac Cardiovasc Surg 133:U1344–U1391.
Gerdes LU. 2003. The common polymorphism of apolipoprotein E: geographical aspects and new pathophysiological relations. Clin Chem Lab Med 41:628–631.
Gerdes LU, Klausen IC, Sihm I, Faergeman O. 1992. Apolipoprotein-E polymorphism in a Danish population compared to findings in 45 other study populations around the world. Genet Epidemiol 9:155–167.
Han X. 2004. The role of apolipoprotein E in lipid metabolism in the central nervous system. Cell Mol Life Sci 61:1896–1906.
Harford RR, Reed HL, Morris MT, Sapien IE, Warden R, Dalesandro MM. 1993. Relationship between changes in serum thyrotropin and total and lipoprotein cholesterol with prolonged antarctic residence. Metab Clin Exp 42:1159–1163.
Hassi J, Rytkonen M, Kotaniemi J, Rintamaki H. 2005. Impacts of cold climate on human heat balance, performance and health in circumpolar areas. Int J Circumpolar Health 64:459.
Henderson VW, Guthrie JR, Dennerstein L. 2003. Serum lipids and memory in a population based cohort of middle age women. J Neurol Neurosurg Psychiatry 74:1530–1535.
Hulbert AJ, Else PL. 2004. Basal metabolic rate: history, composition, regulation, and usefulness. Physiol Biochem Zool 77: 869–876.
Human Genome Diversity Project. 2007. Stanford University HGDP-CEPH SNP Genotyping Data SUPPLEMENT 1 FLAT FILES. September 20, 2007 ed: Centre d'Etude du Polymorphisme Humain.
Itzhaki RF, Dobson CB, Shipley SJ, Wozniak MA. 2004. The role of viruses and of APOE in dementia. Strategies for engineered negligible senescence: why genuine control of aging may be foreseeable. New York: New York Academic Sciences. p 15–18.
Jacobs DR, Iribarren C. 2000. Invited commentary: low cholesterol and nonatherosclerotic disease risk: a persistently perplexing question. Am J Epidemiol 151:748–751.
Jacobs K. 2008. GLU: BioInformed LLC and the U.S. Department of Health and Human Services.
Kanat M, Sipahioglu M, Arinc H, Serin E, Yildiz O, Tunckale A, Celebi H. 2007. Is lipid lowering treatment aiming for very low LDL levels safe in terms of the synthesis of steroid hormones? Med Hypotheses 69:104–112.
Kaplan JR, Klein KP, Manuck SB. 1997. Cholesterol meets Darwin: public health and evolutionary implications of the cholesterol-serotonin hypothesis. Evol Anthropol 6:28–37.
Keatinge WR, Coleshaw SRK, Easton JC, Cotter F, Mattock MB, Chelliah R. 1986. Increased platelet and red-cell counts, blood-viscosity, and plasma-cholesterol levels during heat-stress, and mortality from coronary and cerebral thrombosis. Am J Med 81:795–800.
Kesaniemi YA, Ehnholm C, Miettinen TA. 1987. Intestinal cholesterol absorption efficiency in man is related to apoprotein-e phenotype. J Clin Invest 80:578–581.
Kesteloot H, Ndam ECN, Kowo M, Njoya O, Cobbaert C, Sasaki S, Seghers V. 1997. Serum lipid levels in a Pygmy and Bantu population sample from Cameroon. Nutr Metab Carbiovasc Dis 7:383–387.
Kuzawa CW, Snodgrass JJ, Leonard WR. 2006. Evolutionary medicine and human cholesterol. Am J Hum Biol 18:261–261.
Lalovic A, Levy E, Luheshi G, Canetti L, Grenier E, Sequeira A, Turecki G. 2007. Cholesterol content in brains of suicide completers. Int J Neuropsychopharmacol 10:159–166.
Leininger-Muller B, Jolivalt C, Bertrand P, Siest G. 1998. Oxidation of human apolipoprotein E and cerebral enzymatic systems. Alzheimers Rep 1:95–99.
Leonard WR, Snodgrass JJ, Sorensen MV. 2008. Health consequences of social and ecological adversity among indigenous Siberian populations: biocultural and evolutionary interactions. In: Panter-Brick C, Fuentes A, editors. Health, risk, and adversity. New York: Berghahn Books. p 26–52.
Lester D. 2002. Serum cholesterol levels and suicide: a meta-analysis. Suicide Life Threat Behav 32:333–346.
Linton M, Farese R Jr, Young S. 1993. Familial hypobetalipoproteinemia. J Lipid Res 34:521.
Lopez Herraez D, Bauchet M, Tang K, Theunert C, Pugach I, Li J, Nandineni MR, Gross A, Scholz M, Stoneking M. 2009. Genetic variation and recent positive selection in worldwide human populations: evidence from nearly 1 million SNPs. PLoS One 4:e7888.
Lovegrove BG. 2003. The influence of climate on the basal metabolic rate of small mammals: a slow-fast metabolic continuum. J Comp Physiol B-Biochem Syst Environ Physiol 173:87–112.
Mahley RW, Rall SC, Ng. 2000. Apolipoprotein E: far more than a lipid transport protein. Annu Rev Genomics Hum Genet 1:507–537.
Mancilha-Carvalho JJ, Crews DE. 1990. Lipid profiles of Yanomamo indians of Brazil. Prev Med 19:66–75.
Martin GM. 1999. APOE alleles and lipophylic pathogens. Neurobiol Aging 20:441–443.
Masson LF, McNeill G, Avenell A. 2003. Genetic variation and the lipid response to dietary intervention: a systematic review. Am J Clin Nutr 77:1098–1111.
Melton PE, Zlojutro M, Kimminau K, Crawford MH. 2006. Biological aging and Cox hazard analysis of mortality trends in a Mennonite community from south-central Kansas. Am J Hum Biol 18:387–401.
Miller K, Rubenstein A, Astrand PO. 1968. Lipid values in Kalahari bushmen. Arch Intern Med 121:414–417.
Mohanna S, Baracco R, Seclen S. 2006. Lipid profile, waist circumference, and body mass index in a high altitude population. High Alt Med Biol 7:245–255.
Muldoon MF, Marsland A, Flory JD, Rabin BS, Whiteside TL, Manuck SB. 1997. Immune system differences in men with hypo-or hypercholesterolemia. Clin Immunol Immunopathol 84:145–149.
Murray D, Schaller M. 2010. Historical prevalence of infectious diseases within 230 geopolitical regions: a tool for investigating origins of culture. J Cross-Cultural Psychol 41:99–108.
Nagy B, Rigo J, Fintor L, Karadi I, Toth T. 1998. Apolipoprotein E alleles in women with severe pre-eclampsia. J Clin Pathol 51:324–325.
Nickerson DA, Taylor SL, Fullerton SM, Weiss KM, Clark AG, Stengard JH, Salomaa V, Boerwinkle E, Sing CF. 2000. Sequence diversity and large-scale typing of SNPs in the human apolipoprotein E gene. Genome Res 10:1532–1545.
Onder G, Volpato S, Liperoti R, D'Arco C, Maraldi C, Fellin R, Bernabei R, Landi F, Invest G. 2006. Total serum cholesterol and recovery from disability among hospitalized older adults. J Gerontol Ser A-Biol Sci Med Sci 61:736–742.
Oria RB, Patrick PD, Blackman JA, Lima AAM, Guerrant RL. 2007. Role of apolipoprotein E4 in protecting children against early childhood diarrhea outcomes and implications for later development. Med Hypotheses 68:1099–1107.
Oria RB, Patrick PD, Zhang H, Lorntz B, Costa CMD, Brito GAC, Barrett LJ, Lima AAM, Guerrant RL. 2005. APOE4 protects the cognitive development in children with heavy diarrhea burdens in northeast Brazil. Pediatr Res 57:310–316.
Patterson N, Price AL, Reich D. 2006. Population structure and Eigen analysis. PLoS Genet 2:e190.
Pond WG, Mersmann HJ, Su DR, McGlone JJ, Wheeler MB, Smith EO. 2008. Neonatal dietary cholesterol and alleles of cholesterol 7-α hydroxylase affect piglet cerebrum weight, cholesterol concentration and behavior. J Nutr 138:282–286.
Porter CC, Marlowe FW. 2007. How marginal are forager habitats? J Archaeol Sci 34:59–68.
Prentice AM, Rayco-Solon P, Moore SE. 2005. Insights from the developing world: thrifty genotypes and thrifty phenotypes. Proc Nutr Soc 64:153–161.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575.
Putterman L. 2008. Agriculture, diffusion and development: ripple effects of the Neolithic revolution. Economica 75:729–748.
Reser JE. 2009. Alzheimer's disease and natural cognitive aging may represent adaptive metabolism reduction programs. Behav Brain Funct 5:13.

Rosenberg NA. 2006. Standardized subsets of the HGDP-CEPH human genome diversity cell line panel, accounting for atypical and duplicated samples and pairs of close relatives. Ann Hum Genet 70:841–847.

Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW. 2005. Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet 1:e70.

Schoknecht PA, Ebner S, Pond WG, Zhang SD, McWhinney V, Wong WW, Klein PD, Dudley M, Goddardfinegold J, Mersmann HJ. 1994. Dietary-cholesterol supplementation improves growth and behavioral-response of pigs selected for genetically high and low serum-cholesterol. J Nutr 124:305–314.

Sehayek E, Shefer S, Nguyen LB, Ono JG, Merkel M, Breslow JL. 2000. Apolipoprotein E regulates dietary cholesterol absorption and biliary cholesterol excretion: studies in C57BL/6 apolipoprotein E knockout mice. Proc Nat Acad Sci USA 97:3433–3437.

Serre D, Montpetit A, Paré G, Engert JC, Yusuf S, Keavney B, Hudson TJ, Anand S. 2008. Correction of population stratification in large multi-ethnic association studies. PLoS ONE 3:e1382.

Sing CF, Davignon J. 1985. Role of the apolipoprotein-e polymorphism in determining normal plasma-lipid and lipoprotein variation. Am J Hum Genet 37:268–285.

Singh PP, Singh M, Mastana SS. 2006. APOE distribution in world populations with new data from India and the UK. Ann Hum Biol 33:279–308.

Smith GD, Song FJ, Sheldon TA. 1993. Cholesterol-lowering and mortality—the importance of considering initial level of risk. Br Med J 306:1367–1373.

Smith JD. 2002. Apolipoproteins and aging: emerging mechanisms. Ageing Res Rev 1:345–365.

Song YM, Sung JH, Kim JS. 2000. Invited commentary: low cholesterol and nonatherosclerotic disease risk: a persistently perplexing question—response. Am J Epidemiol 151:752–753.

Song YQ, Stampfer MJ, Liu SM. 2004. Meta-analysis: apolipoprotein E genotypes and risk for coronary heart disease. Ann Intern Med 141:137–147.

Spokas K. 2007. GlobalTempSim Ver. 0.9: United States Department of Agriculture—Agricultural Research Service.

Tammi A, Ronnemaa T, Viikari J, Jokinen E, Lapinleimu H, Ehnholm C, Simell O. 2000. Apolipoprotein E4 phenotype increases non-fasting serum triglyceride concentration in infants—the STRIP study. Atherosclerosis 152:135–141.

Toracinta ER, Oglesby RJ, Bromwich DH. 2004. Atmospheric response to modified CLIMAP ocean boundary conditions during the last glacial maximum. J Clim 17:504–522.

Vamathevan J, Hasan S, Emes R, Amrine-Madsen H, Rajagopalan D, Topp S, Kumar V, Word M, Simmons M, Foord S, et al. 2008. The role of positive selection in determining the molecular cause of species differences in disease. BMC Evol Biol 8:273.

Vance JE, Hayashi H, Karten B. 2005. Cholesterol homeostasis in neurons and glial cells. Semin Cell Dev Biol 16:193–212.

Vignali M, McKinlay A, LaCount D, Chettier R, Bell R, Sahasrabudhe S, Hughes R, Fields S. 2008. Interaction of an atypical plasmodium falciparum ETRAMP with human apolipoproteins. Malaria J 7:211.

von Bergmann K, Lutjohann D, Lindenthal B, Steinmetz A. 2003. Efficiency of intestinal cholesterol absorption in humans is not related to apoE phenotype. J Lipid Res 44:193–197.

Wang HY, Zhang FC, Gao JJ, Fan JB, Liu P, Zheng ZJ, Xi H, Sun Y, Gao XC, Huang TZ, Ke J, Guo GR, Feng GY, Breen G, St Clair D, He L. 2000. Apolipoprotein E is a genetic risk factor for fetal iodine deficiency disorder in China. Mol Psych 5:363–368.

Woollett LA, Osono Y, Herz J, Dietschy JM. 1995. Apolipoprotein E competitively inhibits receptor-dependent low density lipoprotein uptake by the liver but has no effect on cholesterol absorption or synthesis in the mouse. Proc Nat Acad Sci USA 92:12500–12504.

Wozniak MA, Faragher EB, Todd JA, Koram KA, Riley EM, Itzhaki RF. 2003. Does apolipoprotein E polymorphism influence susceptibility to malaria? J Med Genet 40:348–351.

Wozniak MA, Itzhaki RF, Faragher EB, James MW, Ryder SD, Irving WL, Trent HCVSG. 2002. Apolipoprotein E-epsilon 4 protects against severe liver disease caused by hepatitis C virus. Hepatology 36:456–463.

Wozniak MA, Shipley SJ, Dobson CB, Parker SP, Scott FT, Leedham-Green M, Breuer J, Itzhaki RF. 2007. Does apolipoprotein E determine outcome of infection by varicella zoster virus and by Epstein Barr virus? Eur J Hum Genet 15:672–678.

Wright RO, Hu H, Silverman EK, Tsaih SW, Schwartz J, Bellinger D, Palazuelos E, Weiss ST, Hernandez-Avila M. 2003. Apolipoprotein E genotype predicts 24-month Bayley scales infant development score. Pediatr Res 54:819–825.

Yamamoto H, Zheng KC, Ariizumi M. 2003. Influence of heat exposure on serum lipid and lipoprotein cholesterol in young male subjects. Ind Health 41:1–7.

Zetterberg H, Palmer M, Ricksten A, Poirier J, Palmqvist L, Rymo L, Zafiropoulos A, Arvanitis DA, Spandidos DA, Blennow K. 2002. Influence of the apolipoprotein E epsilon 4 allele on human embryonic development. Neurosci Lett 324:189–192.

Zhang J, Muldoon MF, McKeown RE, Cuffe SP. 2005. Association of serum cholesterol and history of school suspension among school-age children and adolescents in the United States. Am J Epidemiol 161:691–699.