HYPOTHESIS

Is the Ala12 variant of the PPARG gene an ''unthrifty allele''?

E Ruiz-Narváez

....

Correspondence to:
Edward Ruiz-Narváez,
Department of Nutrition,
Room 202, Harvard
School of Public Health,
665 Huntington Ave,
Boston, MA 02115, USA;
eruiz@hsph.harvard.edu

Revised version received
1 October 2004
Accepted for publication
21 October 2004
....

Background: The thrifty genotype hypothesis proposes that genetic susceptibility to type 2 diabetes results from the positive selection of ''thrifty'' alleles in the past. A corollary of this hypothesis is that genetic variants protecting against the development of diabetes are ''unthrifty'' and thus subject to negative selection during human evolution.

Methods: It was assessed whether age estimates of the diabetes protective PPARG Ala12 allele indicate effects of natural selection. Based on published data from four populations, the date of origin of the diabetes protective PPARG Ala12 variant was estimated using both allele frequency and linkage disequilibrium (LD) with the C1431T single nucleotide polymorphism in exon 6 of the PPARG gene.

Results: The best LD based estimate of the age of the Ala12 allele gave an average of ~32 000 years with a maximum upper bound of ~58 000 years. Assuming a population with a growth rate of $r = 0.01$ per generation, the frequency based estimate of the age of the Ala12 variant gave an average of ~27 000 years with a maximum upper bound of ~42 000 years.

Discussion: The similarity of both time estimates is consistent with selective equivalence of the diabetes protective PPARG Ala12 allele and the diabetes susceptible PPARG Pro12 allele.

According to the thrifty genotype hypothesis, widespread genetic susceptibility to type 2 diabetes is the result of the positive selection of ''thrifty'' alleles in the past. A corollary of this hypothesis is that genetic variants protecting against the development of diabetes are ''unthrifty'' and thus subject to negative selection during human evolution. It was assessed whether age estimates of the diabetes protective PPARG Ala12 allele indicate effects of natural selection.

METHODS

Based on published data from four populations (Scottish, Chinese, Malays, and Indians), the date of origin of the diabetes protective PPARG Ala12 variant was estimated using both allele frequency and linkage disequilibrium (LD) with the C1431T single nucleotide polymorphism (SNP) in exon 6 of the PPARG gene.

RESULTS

Taking geographical distribution into account, the best LD based estimate of the age of the Ala12 allele gave an average of ~32 000 years with a maximum upper bound of ~58 000 years. Assuming a population with a growth rate of $r = 0.01$ per generation, the frequency based estimate of the age of the Ala12 variant gave an average of ~27 000 years with a maximum upper bound of ~42 000 years.

DISCUSSION

The similarity of both time estimates is consistent with selective equivalence of the diabetes protective PPARG Ala12 allele and the diabetes susceptible PPARG Pro12 allele. Evidence of *positive* selection was found in the Indian sample, suggesting that the selective value of the Ala12 allele may be modulated by the neighbouring T1431 variant.

The thrifty genotype hypothesis[1] states that susceptibility to type 2 diabetes is in part due to adaptation to ancient conditions of life during the course of human evolution. It is postulated that our hunter gatherer ancestors faced times of famine and thus genetic variants that allowed conservation of glucose and efficient storage of fat as an energy reserve would offer a selective advantage to their carriers. Given the modern conditions in affluent societies of plentiful availability of food, these ''thrifty genotypes'' became deleterious leading to the development of type 2 diabetes. A corollary of this hypothesis is that genetic variants that afford protection from disease are mutations that cause a ''loss of thriftiness'',[2] for example inefficient fat storage or renal loss of glucose. Another consequence, not yet recognised, of the thrifty genotype hypothesis is that these ''unthrifty alleles'' must have suffered negative selection during human evolution.

Because of its multiple roles in adipogenesis, energy homeostasis, and insulin sensitivity, the nuclear receptor PPARG is thought to be a major regulator of the ''thrifty gene response''.[3] The Ala12 variant of the PPARG gene has been associated with decreased receptor activity, lower body mass index, and higher insulin sensitivity compared to the wild type Pro12,[4] and it has been shown to be protective against type 2 diabetes in different ethnic groups.[5–7] These results suggest that the Ala12 variant could be an ''unthrifty allele'' of the PPARG gene and therefore exposed to negative selection in the past.

The Ala12 allele of the PPARG gene has a wide geographical distribution. Table 1 shows a non-exhaustive list of the frequency of this allele in different populations. It is clear that European populations have the highest frequency of the Ala12 allele and East Asian populations have the lowest frequency. With the available data it is unclear whether this genetic variant is present in Africa. The low frequency of the allele in African-Americans does not rule out the possibility of recent genetic flow from non-African groups. Age estimates of the Ala12 allele may provide some light about its origin and, more importantly, they can be used to test the hypothesis of neutrality of this allele.

Age estimates can be obtained using both linkage disequilibrium (LD) and allele frequency. By using haplotypic

....

Table 1 Frequency of the PPARG Ala12 allele in different human populations

Population	Sample size	Allele frequency (SD)
Europe		
Finland[4]	1306	0.145 (0.007)
Scotland[7]	973	0.142 (0.008)
Poland[11]	500	0.198 (0.013)
Spain[19]	309	0.115 (0.013)
US		
Caucasian[20]	2092	0.123 (0.005)
African-American[21]	46	0.011 (0.011)
Canada	332	0.084 (0.011)
Oji-Cree (Ontario)[22]		
Asia		
Chinese from Singapore[6]	2730	0.037 (0.003)
Malays from Singapore[6]	740	0.032 (0.005)
Indians from Singapore[6]	568	0.119 (0.010)
Japan[23]	215	0.030 (0.008)

information in two linked microsatellite loci, Stephens et al[8] estimated that a 32 base pair (bp) deletion in the CCR5 locus that confers resistance to infection by HIV occurred about 700 years ago. However, the high frequency of this deletion in Europe (~10%) suggested an age greater than 100 000 years under constant population size. Because the pattern of LD suggested a much younger age than that indicated by the allele frequency, this large difference in age estimates was interpreted to be the result of positive selection of the deletion in Europeans.[8] This same approach can be applied to the Ala12 allele of the PPARG gene. In particular, if the Ala12 allele is an ''unthrifty allele'' that suffered negative selection in the past, then the pattern of LD should indicate an older age than is consistent with its frequency.

The C1431T SNP is located in exon 6 of the PPARG gene separated by 82 kb from the Pro12Ala polymorphism. Tai et al[6] showed the haplotypic frequencies of both variants in Chinese, Malays, and Indians, and Doney et al[7] reported these frequencies in a Scottish cohort. Thus, it is possible to estimate the age of the Ala12 allele using both LD and allele frequency.

Allele age based on linkage disequilibrium

By using a first order approximation of the likelihood[9] and the single linked C1431T marker locus, a maximum likelihood estimate of the age (in generations) of the Ala12 allele is

$$\hat{t}_{MLE,1} = \frac{1}{\theta} \log\left(\frac{p_2(k_1 + k_2)}{k_1 p_2 - k_2 p_1} \right),$$

where θ is the recombination rate per generation between both polymorphisms, k_1 is the number of chromosomes carrying both the T1431 and the Ala12 alleles, k_2 is the number of chromosomes having both the C1431 and the Ala12 alleles, p_1 is the frequency of the T1431 allele on Pro12 chromosomes, and p_2 is the frequency of the C1431 allele on Pro12 chromosomes. Because in all the surveyed populations the Ala12 allele is in high linkage disequilibrium with the T1431 variant (D'>0.500), it is assumed that the mutation originating the Ala12 allele first arose in a chromosome carrying the T1431 allele. Three different recombination rates were used: θ = 0.0008 which corresponds to an average rate of 1 Mb–1 cM, θ = 0.0016 which is equal to double the average rate, and θ = 0.0004 which corresponds to one half of the average rate. 95% Confidence intervals (95% CI) were estimated based on the composite likelihood given by Guo and Xiong[9] (see equation 12 in their paper).

Allele age based on frequency

Slatkin[10] found that a maximum likelihood estimate of age as a function of allele frequency in a growing population is equal to

$$\hat{t}_{MLE,2} = \frac{1}{r} \log(1 + 2N_0 rx),$$

where r is the population growth rate per generation, N_0 is the current size of the population, and x is the frequency of the Ala12 allele. 95% CI can be estimated using the following cumulative probability distribution of allele age[12]:

$$P(\hat{t} \le t) = (1 - x)^{-1 + n/(1 + n\tau(t)/2)},$$

where n is the sample size and $\tau(t)$ is a scaled time that takes into account the changes in population size at t generations in the past. A growth rate of r~0.02 per generation for the last 10 000 years has been estimated for European populations.[12] Data from Biraben[13] show historic growth rates in the range of 0.01–0.04 for Asian populations. Therefore, to estimate the age of the Ala12 allele values of r = 0.01, r = 0.02, and r = 0.04 were used.

Results and implications

Table 2 shows the age estimates in years of the Ala12 allele. These time estimates were calculated assuming a generational length of 25 years.

Except for the results from the Indian sample, all the allele age estimates based on linkage disequilibrium are consistent with the age estimates assuming neutrality of the Ala12 allele in a growing population. As expected, time estimates greatly depend on the assumptions used. However, the presence of the Ala12 allele in Europeans and Asians indicates that this allele first arose, at the latest, about the time of the split between both population groups ~43 000 years ago,[14] a fact that agrees with the estimated allele ages assuming either a recombination rate of θ = 0.0004 (average age ~32 000 years, maximum upper bound of ~58 000 years) or a population growth rate of r = 0.01 (average age ~27 000 years, maximum upper bound of ~42 000 years). If the true recombination rate between the Pro12Ala and the C1431T SNPs is ~0.0004, then, only by assuming a fast growing population (r = 0.04), it is possible to reject the hypothesis of neutrality. In particular, in this last scenario, the LD pattern indicates an older age of the Ala12 variant compared to the time estimates based on allele frequency, therefore suggesting negative selection against the Ala12 allele.

Although the estimated ages of the Ala12 allele are consistent with the hypothesis of neutrality, other explanations must be considered. First, it is possible that selection over the Ala12 allele was not sufficiently strong to produce evidence using available data; the use of just one linked locus in the present analysis may be insufficient. Information regarding both more linked loci and the nucleotide sequence of the PPARG gene in several populations would be helpful to settle this question concerning selection of the Ala12 allele. Also, it may be argued that there was insufficient time for the Ala12 allele to undergo negative selection given the fact that agriculture originated ~12 000 years ago during the Neolithic period. However, if the present age estimates are correct and based on its geographical distribution, the Ala12 allele may be at least ~43 000 years old, 1000 human generations before the dawn of agriculture, providing enough time for natural selection. In addition, a consequence of the agricultural revolution was that Neolithic humans depended mostly on just one or a few plants, and often suffered from starvation due to crop failure,[15] an event also common during

Table 2 Estimated age in years of the Ala12 allele of the PPARG gene using both linkage disequilibrium and allele frequency

Population*	MLE (95% CI) using linkage disequilibrium			MLE (95% CI) using frequency of the Ala12 allele		
	$\theta = 0.0016$	$\theta = 0.0008$	$\theta = 0.0004$	$r = 0.01$	$r = 0.02$	$r = 0.04$
Scottish,[7] $N_0 \sim 5 \times 10^6$, D' = 0.608, n = 973	6845 (5500–8405)	13 690 (11 000–16 815)	27 375 (22 000–33 625)	23 900 (22 400–30 125)	12 825 (12 050–15 950)	6850 (6475–8400)
Chinese,[6] $N_0 \sim 1 \times 10^9$, D' = 0.555, n = 2730	8960 (6795–11 665)	17 920 (13 595–23 330)	35 845 (27 190–46 655)	33 775 (32 175–42 475)	17 750 (16 950–22 100)	9325 (8900–11 475)
Malays,[6] $N_0 \sim 23 \times 10^6$, D' = 0.572, n = 740	8565 (4795–14 415)	17 125 (9595–28 830)	34 250 (19 190–57 655)	24 000 (22 050–32 875)	12 850 (11 900–17 300)	6875 (6375–9100)
Indians,[6] $N_0 \sim 1 \times 10^9$, D' = 0.799, n = 568	3195 (2030–4740)	6390 (4065–9485)	12 780 (8125–18 970)	36 700 (35 100–43 300)	19 225 (18 425–22 525)	10 050 (9650–11 700)

MLE, maximum likelihood estimator. *D', standardised LD coefficient between Pro12Ala and C1431T SNPs; n, sample size; N_0, current population size.

historical times. For example, Japan endured more than 200 famines after the sixth century and France had ~75 crop failures between 500 and 1500 AD.[15] Therefore, selection pressure did not disappear following the Neolithic agricultural revolution.

The lack of evidence of negative selection of the Ala12 allele could be also due to fitness modulation by neighbouring polymorphisms. In particular, the T1431 variant has been associated with opposing phenotypes such as higher body mass index[16] and greater total carotid plaque volume[17] as compared to the Ala12 allele. Also, it has been shown that the diabetes protective effect of the Ala12 allele is weakened by the presence of the T1431 allele.[7] These observations raise the hypothesis that the fitness of the Ala12 allele depends on the presence of the T1431 variant. Chromosomes carrying both Ala12 and T1431 alleles could be ''thrifty haplotypes'' that have undergone *positive* selection, and chromosomes with both Ala12 and C1431 variants could be ''unthrifty haplotypes'' that have experienced *negative* selection.

The last hypothesis may explain in part the anomalous results from the Indian sample. The LD based age estimates are consistently younger than estimates based on allele frequency and therefore suggest *positive* selection of the Ala12 allele. The Indian sample showed the highest LD (D' = 0.799) of the four samples analysed. Around 83% of the chromosomes carrying the Ala12 allele also carried the T1431 variant in the Indian sample as compared to ~66% in the other three samples. This means that in the Indian population Ala12−T1431 ''thrifty haplotypes'' are in the majority, indicating *positive* selection of the Ala12 allele in this particular sample.

Overall, time estimates based on LD are similar to estimates based on allele frequency. The geographical distribution of the Ala12 allele in European and Asian populations is consistent with a recombination rate with the C1431T SNP of $\theta \sim 0.0004$ and a population growth rate of $r \sim 0.01$. If these age estimates are correct, they indicate that the mutation giving rise to the Ala12 variant occurred after the genetic split between African and non-African populations, which is why this allele may be absent in Africans. Because there are no published frequencies in Africans, it is unclear whether the Ala12 allele is present or not in African populations. Hara *et al*[5] found just one heterozygous individual out of 46 non-diabetic African-American women with polycystic ovary syndrome, implying a frequency of ~1% of the Ala12 allele in that sample. Since African-Americans have an average of ~25% of Caucasian genes,[18] the presence of the Ala12 allele in African-Americans may be due to recent genetic flow from US Caucasians.

The thrifty genotype hypothesis[1] was first postulated to explain the widespread susceptibility to type 2 diabetes in modern human populations. A corollary of this hypothesis is that genetic variants conferring resistance to disease are ''unthrifty alleles''[2] subject to natural selection in the past. By examining published data on the Ala12 allele of the PPARG gene, it was not possible to reject the hypothesis of neutrality of this genetic variant. Under a reasonable range of population scenarios, available data are compatible with neutrality of the Ala12 variant. However, it cannot be ruled out that the Ala12 allele has undergone natural selection in the past, with its selective value depending on neighbouring polymorphisms. It is likely that the T1431 variant modulates the selective value of the Ala12 allele. Evidence of *positive* selection of the Ala12 variant in the Indian sample supports this hypothesis, and indicates the difficulty of assigning selective values to single genetic variants. Information about more linked loci and about the nucleotide sequence of the PPARG gene in several human populations, as well as from other genes, will be required to elucidate the evolutionary history of genetic variants associated with a predisposition to type 2 diabetes.

REFERENCES

1 **Neel JV**. Diabetes mellitus: a ''thrifty'' genotype rendered detrimental by ''progress''? *Am J Hum Genet* 1962;**14**:353–62.
2 **Sharma AM**. The thrifty-genotype hypothesis and its implications for the study of complex genetic disorders in man. *J Mol Med* 1998;**76**(8):568–71.
3 **Auwerx J**. PPARgamma, the ultimate thrifty gene. *Diabetologia* 1999;**42**(9):1033–49.
4 **Deeb SS**, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, Kuusisto J, Laakso M, Fujimoto W, Auwerx J. A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat Genet* 1998;**20**(3):284–7.
5 **Hara K**, Okada T, Tobe K, Yasuda K, Mori Y, Kadowaki H, Hagura R, Akanuma Y, Kimura S, Ito C, Kadowaki T. The Pro12Ala polymorphism in PPAR gamma2 may confer resistance to type 2 diabetes. *Biochem Biophys Res Commun* 2000;**271**(1):212–6.
6 **Tai ES**, Corella D, Deurenberg-Yap M, Adiconis X, Chew SK, Tan CE, Ordovas JM. Differential effects of the C1431T and Pro12Ala PPARgamma gene variants on plasma lipids and diabetes risk in an Asian population. *J Lipid Res* 2004;**45**(4):674–85.
7 **Doney AS**, Fischer B, Cecil JE, Boylan K, McGuigan FE, Ralston SH, Morris AD, Palmer CN. Association of the Pro12Ala and C1431T variants of PPARG and

their haplotypes with susceptibility to Type 2 diabetes. *Diabetologia* 2004;**47**(3):555–8.

8 **Stephens JC**, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, Winkler C, Huttley GA, Allikmets R, Schriml L, Gerrard B, Malasky M, Ramos MD, Morlot S, Tzetis M, Oddoux C, di Giovine FS, Nasioulas G, Chandler D, Aseev M, Hanson M, Kalaydjieva L, Glavac D, Gasparini P, Kanavakis E, Claustres M, Kambouris M, Ostrer H, Duff G, Baranov V, Sibul H, Metspalu A, Goldman D, Martin N, Duffy D, Schmidtke J, Estivill X, O'Brien SJ, Dean M. Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. *Am J Hum Genet* 1998;**62**(6):1507–15.

9 **Guo SW**, Xiong M. Estimating the age of mutant disease alleles based on linkage disequilibrium. *Hum Hered* 1997;**47**(6):315–37.

10 **Slatkin M**. Allele age and a test for selection on rare alleles. *Philos Trans R Soc Lond B Biol Sci* 2000;**355**(1403):1663–8.

11 **Ardlie KG**, Lunetta KL, Seielstad M. Testing for population subdivision and association in four case-control studies. *Am J Hum Genet* 2002;**71**(2):304–11.

12 **Slatkin M**, Rannala B. Estimating allele age. *Annu Rev Genomics Hum Genet* 2000;**1**:225–49.

13 **Biraben JN**. An essay concerning mankind's evolution. *Population* 1980;**4**:1–13.

14 **Straus LG**. Age of the modern Europeans. *Nature* 1989;**342**:476–7.

15 **Kiple KF**, Ornelas KC, eds. *The Cambridge world history of food.* Cambridge, UK: Cambridge University Press, 2000.

16 **Doney A**, Fischer B, Frew D, Cumming A, Flavell DM, World M, Montgomery HE, Boyle D, Morris A, Palmer CNA. Haplotype analysis of the PPARgamma Pro12Ala and C1431T variants reveals opposing associations with body weight. *BMC Genet* 2002;**3**(1):21.

17 **Al-Shali KZ**, House AA, Hanley AJ, Khan HM, Harris SB, Zinman B, Mamakeesick M, Fenster A, Spence JD, Hegele RA. Genetic variation in PPARG encoding peroxisome proliferator-activated receptor gamma associated with carotid atherosclerosis. *Stroke* 2004;**35**(9):2036–40.

18 **Chakraborty R**, Kamboh MI, Nwankwo M, Ferrell RE. Caucasian genes in American blacks: new data. *Am J Hum Genet* 1992;**50**(1):145–55.

19 **Landi S**, Moreno V, Gioia-Patricola L, Guino E, Navarro M, de Oca J, Capella G, Canzian F. Association of common polymorphisms in inflammatory genes interleukin (IL)6, IL8, tumor necrosis factor alpha, NFKB1, and peroxisome proliferator-activated receptor gamma with colorectal cancer. *Cancer Res* 2003;**63**(13):3560–6.

20 **Ridker PM**, Cook NR, Cheng S, Erlich HA, Lindpaintner K, Plutzky J, Zee RYL. Alanine for proline substitution in the peroxisome proliferator-activated receptor gamma-2 (PPARG2) gene and the risk of incident myocardial infarction. *Arterioscler Thromb Vasc Biol* 2003;**23**(5):859–63.

21 **Hara M**, Alcoser SY, Qaadir A, Beiswenger KK, Cox NJ, Ehrmann DA. Insulin resistance is attenuated in women with polycystic ovary syndrome with the Pro(12)Ala polymorphism in the PPARgamma gene. *J Clin Endocrinol Metab* 2002;**87**(2):772–5.

22 **Hegele RA**, Cao H, Harris SB, Zinman B, Hanley AJ, Anderson CM. Peroxisome proliferator-activated receptor-gamma2 P12A and type 2 diabetes in Canadian Oji-Cree. *J Clin Endocrinol Metab* 2000;**85**(5):2014–9.

23 **Mori Y**, Kim-Motoyama H, Katakura T, Yasuda K, Kadowaki H, Beamer BA, Shuldiner AR, Akanuma Y, Yazaki Y, Kadowaki T. Effect of the Pro12Ala variant of the human peroxisome proliferator-activated receptor gamma 2 gene on adiposity, fat distribution, and insulin sensitivity in Japanese men. *Biochem Biophys Res Commun* 1998;**251**(1):195–8.