Annals of human genetics

doi: 10.1111/j.1469-1809.2009.00549.x

Frequency of the AGT Pro11Leu Polymorphism in Humans: does Diet Matter?

Laure Ségurel*, Sophie Lafosse, Evelyne Heyer and Renaud Vitalis‡

Muséum National d'Histoire Naturelle–Centre National de la Recherche Scientifique–Université Paris Diderot-Paris 7, UMR 7206 Éco-Anthropologie et Ethnobiologie, CP 139, 57 rue Cuvier, 75231 Paris Cedex 05, France

Summary

The Pro11Leu substitution in the *AGXT* gene, which causes primary hyperoxaluria type 1, is found with high frequency in some human populations (e.g., 5–20% in Caucasians). It has been suggested that this detrimental mutation could have been positively selected in populations with a meat-rich diet. In order to test this hypothesis, we investigated the occurrence of Pro11Leu in both herder and agriculturalist populations from Central Asia. We found a lower frequency of this detrimental mutation in herders, whose diet is more meat-rich, as compared to agriculturalists, which therefore challenges the universality of the previous claim. Furthermore, when combining our original data with previously published results, we could show that the worldwide genetic differentiation measured at the Pro11Leu polymorphism does not depart from neutrality. Hence, the distribution of the variation observed in the *AGXT* gene could be due to demographic history, rather than local adaptation to diet.

Keywords: adaptation, *AGXT*, AMOVA, diet, *F*-statistics, primary hyperoxaluria type 1

Introduction

Primary hyperoxaluria type 1 (PH1, MIM 259900) is a lethal autosomal recessive disease caused by the deficiency of the alanine:glyoxylate aminotransferase (AGT), a liver-specific enzyme usually targeted in the peroxysomes in humans (Danpure & Jennings, 1986). The AGT deficiency leads to the formation of insoluble calcium salts, resulting in progressive renal failure. It has been estimated that, before the introduction of modern therapies, 80% of PH1 patients died from renal failure before the age of 20 years (Williams & Smith, 1983). In more than a third of the patients, the disease is not due to the absence of AGT, but rather to its de-localisation from the peroxysomes to the mitochondria (Danpure et al., 1990). This mistargeting is due to the synergistic effect of two non-synonymous recessive mutations in the human *AGXT* gene that encodes AGT: the Pro11Leu and Gly170Arg substitutions (Purdue et al., 1990). Pro11Leu alone, in a homozygous state, is responsible for the de-localisation of 5% of the protein to the mitochondria (Purdue et al., 1990), although this has not been associated with any clinical sequela. On the contrary, the association of Pro11Leu with the most common PH1 mutation Gly170Arg results in the mistargeting of at least 90% of the AGT to the mitochondria, resulting in primary hyperoxaluria type 1 (Danpure et al., 1989). From analysis of mutant constructs in *Escherichia coli*, it seems likely that Gly170Arg alone has no pathological phenotype (Lumb & Danpure, 2000), yet it has never been found so far in humans without the Pro11Leu substitution. Pro11Leu also acts synergistically with other *AGXT* mutations which, like Gly170Arg, seem innocuous alone (Lumb & Danpure, 2000).

Although the Pro11Leu substitution is therefore likely to be detrimental, by sensitising AGT to the effect of a wide range of PH1 mutations, it is found at relatively high frequency in some populations, e.g., 5–20% in Caucasians (Danpure et al., 1994a, 1994b). How can this paradox be solved? It has been shown that, in mammals, the intracellular location of AGT depends upon natural diet (Danpure et al., 1990). AGT normally converts glyoxylate to glycine (Danpure & Jennings, 1986). However, the sites where the conversion occurs, as well as the precursors of glyoxylate, are both related to natural diet

*Corresponding author: Laure Ségurel, Muséum National d'Histoire Naturelle–Centre National de la Recherche Scientifique– Université Paris 7, UMR 7206 Éco-Anthropologie et Ethnobiologie, CP 139, 57 rue Cuvier, 75231 Paris Cedex 05, France. Tel: +33 (0)1 40 79 81 65; Fax: +33 (0)1 40 79 38 91; E-mail: lsegurel@mnhn.fr
‡Present address: Institut National de la Recherche Agronomique (INRA), UMR CBGP (INRA–IRD–CIRAD–Montpellier SupAgro), Campus International de Baillarguet, CS 30016, 34988 Montferrier-sur-Lez, France

(Danpure, 1997). In herbivores, the main precursor of glyoxylate is thought to be glycolate, which is converted to glyoxylate in the peroxysomes (Noguchi, 1987). In carnivores, the major precursor is hydroxyproline, which is converted to glyoxylate in the mitochondria (Takayama et al., 2003). Therefore, the fact that in herbivores, AGT is principally found in peroxysomes, while in carnivores it is found in mitochondria, might reflect a more effective glyoxylate detoxification by AGT at the site of glyoxylate synthesis (Danpure, 1997). In humans, AGT is usually targeted in the peroxysomes (Cooper et al., 1988), which might reflect the ancestral herbivorous diet of hominoids (Holbrook et al., 2000). Therefore, it might well be that the redirection of a small proportion of AGT from peroxysomes to mitochondria in humans leads to a sub-cellular distribution of AGT that is more compatible with an omnivorous (or carnivorous), rather than herbivorous, diet (Danpure, 1997).

Following this line of ideas, it has been suggested that the redirection of 5% of the AGT activity from peroxysomes to mitochondria, due to the Pro11Leu substitution, might give a selective advantage to individuals who have a larger proportion of meat in their diet (Danpure, 1997), although the precise link between the functional consequences of the human non-synonymous substitution Pro11Leu in *AGXT* and diet remains hypothetical. Interestingly, if this mutation is advantageous in populations with a meat-rich diet, it should have been selected for during the ice age periods, when our hunter-gatherer ancestors were highly dependent on meat-based resources. Hence, we should now observe a decrease of selective constraints in populations with a lower proportion of meat in their diet, like agriculturalists, as compared to hunter-gatherers and herders.

The prediction that the Pro11Leu detrimental substitution has been, and/or still is, advantageous for individuals who have a meat-rich diet has been tested by Caldwell et al. (2004). To do so, they first determined the frequency of the Pro11Leu substitution C → T in 11 human populations. They found the highest frequency of the derived T allele in the Saami (27.9%, see also Kozlov et al., 2008), a population of herders known to have a meat-rich ancestral diet (Haglin, 1991, 1999), and the lowest frequency in the Chinese (2.3%) and Indian populations (3.0%), which are supposed to have, respectively, mixed and vegetarian ancestral diets. They concluded that the frequency of the T allele in human populations has been shaped by dietary selection pressure. Yet, this apparent correlation between the Pro11Leu T allele frequency and the ancestral diet is not robust to the inclusion of, e.g., the Mongol sample, which shows a low frequency of the T allele (6.9%), although the diet of these traditionally nomadic herders mainly consists of meat, milk and dairy products (see Hruschka & Brandon, 2004). Caldwell et al. (2004) also compared the genetic differentiation (as measured by the parameter F_{ST}) at the Pro11Leu polymorphism for Nigerians, Chinese and Saami to that observed at 33,487 presumably neutral SNPs typed in related populations (African Americans, East Asians and European Americans). The rationale was to test whether the genetic differentiation at the Pro11Leu polymorphism exceeded that of the rest of the genome, which could be interpreted as a signature of divergent selection. They found that the F_{ST} estimate at the Pro11Leu polymorphism lay in the tail of the distribution of genome-wide F_{ST} estimates, although not significantly ($p = 0.074$ and $p = 0.266$, respectively, for the Saami vs. Chinese and Saami vs. Nigerians comparisons).

Caldwell et al.'s (2004) results are therefore tentative, and require further investigation. For example, it would be interesting to test whether the proportion of genetic variance explained by diet at the Pro11Leu polymorphism is significantly different from that observed in the rest of the genome. To address this question, hierarchical analysis of molecular variance (AMOVA, Excoffier et al., 1992) can be used to apportion the total observed genetic variation among nested hierarchical levels, with populations pooled into groups according to their ancestral diet. In this study, we aim at testing Danpure's (1997) hypothesis that the Pro11Leu replacement in *AGXT* is beneficial for individuals who have a predominantly meat-based diet. To that end, we provide new data collected from seven traditionally nomadic herder populations and four sedentary agriculturalist populations from Central Asia, that differ by their lifestyle and ancestral diet. In the following, as in Caldwell et al. (2004), we use the current lifestyle of populations as a proxy to the ancestral lifestyle. We do not necessarily assume a strict continuity in the subsistence pattern from the ice-age hunter-gatherer populations to the extant herder populations. We rather assume that, since the Neolithic transition, herder populations relied predominantly on meat and dairy products, as opposed to agriculturalist populations who had a mixed or more balanced diet. Analyzing these original data together with published data sets, we specifically ask (*i*) whether the Pro11Leu allele frequency correlates with ancestral diet, (*ii*) whether the genetic differentiation at the *AGXT* gene departs from the genome wide differentiation as a consequence of selection, and (*iii*) whether the proportion of genetic variance explained by diet at this gene is larger than that of the rest of the genome.

Materials and Methods

DNA Samples

We sampled seven populations of traditionally nomadic herders and four populations of sedentary agriculturalists from West Uzbekistan to East Kyrgyzstan, representing respectively 214 and 90 healthy adult men from five ethnic groups (Kyrgyz,

Table 1 Samples Description

Sampled populations (area)	Acronym	Location	Long.	Lat.	*n*	*p* (%)
Tajiks (Ferghana)	TJR	Tajikistan/Kyrgyzstan border	40.36	71.28	17	17.6
Tajiks (Gharm)	TJE	Northern Tajikistan	39.12	70.67	23	19.6
Tajiks (Penjinkent)	TDU	Uzbekistan/Tajikistan border	39.44	68.26	24	20.8
Tajiks (Yagnobs from Douchambe)	TJY	Western Tajikistan	38.57	68.78	26	26.9
Karakalpaks (Qongrat from Karakalpakia)	KKK	Western Uzbekistan	43.77	59.02	23	15.2
Kazaks (Karakalpakia)	KAZ	Western Uzbekistan	43.04	58.84	30	1.7
Kazaks (Bukara)	LKZ	Southern Uzbekistan	40.08	63.56	49	9.2
Kyrgyz (Andijan)	KRA	Tajikistan/Kyrgyzstan border	40.77	72.31	32	14.1
Kyrgyz (Narin)	KRG	Middle Kyrgyzstan	41.6	75.8	20	7.5
Kyrgyz (Narin)	KRB	Middle Kyrgyzstan	41.25	76	26	11.5
Turkmen (Karakalpakia)	TUR	Western Uzbekistan	41.55	60.63	34	13.2
Sichuan Chinese	*CHI*				*86*	*2.3*
Armenians	*ARM*				*73*	*19.2*
North Welsh	*NOR*				*82*	*14.6*
Nigerians	*NIG*				*62*	*8.9*
Mongols	*MON*				*80*	*6.9*

Sedentary agriculturalist populations are in white; traditionally nomadic herders are in gray. The five latter populations (in italics) correspond to the populations studied by Caldwell et al. (2004) for the Pro11Leu polymorphism, included in the present analysis. Long., longitude; Lat., latitude; *n*, sample size; *p*, frequency of the derived T allele of the Pro11Leu polymorphism in *AGXT*.

Karakalpaks, Kazaks, Turkmen and Tajiks; see Table 1). We collected ethnologic data prior to sampling, including the recent genealogy of the participants. Using this information, we retained only those individuals that were unrelated for at least two generations back in time. All individuals gave their informed consent for participation in this study. Total genomic DNA was isolated from blood samples or from saliva using a standard phenol-chloroform extraction protocol (Maniatis et al., 1982). We also used the data from 5 populations genotyped by Caldwell et al. (2004) at the Pro11Leu polymorphism: Sichuan Chinese, Mongols, Armenians, North Welsh and Nigerians, for which neutral data were also available on closely related populations (Han Chinese, Mongols, Adygeis, Orcadians and Yorubas, respectively).

Genotyping

The Pro11Leu substitution (nucleotide C in chimpanzees, C or T in humans) was detected by PCR-RFLP. PCR amplifications were performed in a 10 μL final volume containing 1X Eppendorf buffer, 125 μM of each dNTP, 0.5 U of Eppendorf Taq polymerase, 125 nM of forward (5′-GCACAGATAAGCTTCAGGGA-3′) and reverse (5′-CTTGAAGGATGGATCCAGGG-3′) primers, and 10 ng of DNA. The reactions were performed in an Eppendorf Mastercycler with an initial denaturation step at 94°C for 5 min, followed by 40 cycles at 94°C for 30 s, 59°C for 1 min and 72°C for 30 s, with a final extension step at 72°C for 10 min. PCR products were then incubated overnight at 37°C, with 10 units of the restriction endonuclease *Sty*I, 0.01 μg/μL acetylated BSA and NEB Buffer 3, in a 15 μL final volume. Digestion products were then electrophoresed at 4°C overnight on a 2% agarose gel stained with ethidium bromide.

There are two main forms of *AGXT* across human populations, the "major" allele (Ma) and the "minor" allele (Mi). Among other differences between the two alleles, the Mi allele presents a 74-bp duplication in intron 1 of the gene. In Caucasian populations, the Pro11Leu polymorphism is in strong linkage disequilibrium with the Mi allele, and it has been previously suggested that the presence of the 74-bp duplication would be indicative of the Pro11Leu polymorphism (Purdue et al., 1991). The Pro11Leu substitution C → T induces a *Sty*I restriction site loss. Therefore, after digestion of the PCR products with *Sty*I, one expects to find a 512-bp fragment for the Ma allele (which contains neither polymorphism), and a 619-bp fragment for the Mi allele (which contains both the 74-bp duplication and the Pro11Leu substitution) (Danpure & Rumsby, 1996). However, Coulter-Mackie et al. (2003) also found a 586-bp fragment in African populations, which they referred to as the Mi^A allele. This fragment contains the 74-bp duplication but not the Pro11Leu substitution. Importantly, Coulter-Mackie et al. (2003) never found Pro11Leu on the background of the Ma allele, i.e. without the duplication. Therefore, after digestion with *Sty*I, we expected to observe the three kinds of fragments, potentially: the Ma allele without Pro11Leu (512 bp), the Mi allele with Pro11Leu (619 bp), or the Mi^A allele (586 bp) without Pro11Leu.

Data Analysis

We aimed to evaluate whether the genetic differentiation observed at the Pro11Leu polymorphism departed from the genome-wide differentiation expected at neutrality among 16 populations whose ancestral diet was known: the 16 populations consisted of the 11 Central Asian populations described

above, and 5 populations genotyped by Caldwell et al. (2004) at the Pro11Leu polymorphism: Sichuan Chinese, Mongols, Armenians, North Welsh and Nigerians. In order to estimate the expected neutral differentiation, we used the genotypic data at 27 autosomal short tandem repeat (STR) markers from Ségurel et al. (2008) for the 11 Central Asian populations, as well as data from the HGDP-CEPH Human Genome Diversity Cell Line Panel dataset (Rosenberg et al., 2002) genotyped at the same STR markers as were used for the Han Chinese, the Mongols, the Adygeis, the Orcadians and the Yorubas. Although these latter populations are not strictly those for which the Pro11Leu frequency data were available, they are probably very closely related.

To identify the signatures of natural selection we used a modified version of the software package Dfdist (Beaumont & Nichols, 1996). This method is based on the principle that genetic differentiation among populations is expected to be higher for loci under divergent selection than for the rest of the genome. The rationale here was to compute F_{ST} and the overall heterozygosity of the pooled sample for the Pro11Leu substitution, and then to compare this value to a neutral distribution of F_{ST} conditional on heterozygosity, generated by means of coalescent simulations in a symmetrical island migration model at migration-drift equilibrium (Wright, 1951), given the observed level of differentiation measured at the 27 autosomal STRs ($F_{ST} = 0.019$). Since the software package Dfdist was specifically designed for the analysis of bi-allelic, dominant markers (see, e.g., Bonin et al., 2006), we modified it in order to simulate co-dominant, bi-allelic data (code available upon request). 500,000 coalescent simulations were performed with a 50-demes island model and $\theta = 2nN\mu = 0.2$ (where $n = 50$ is the number of demes of size N, and μ is the mutation rate). This particular θ value would correspond, e.g., to $N = 100$ and $\mu = 2 \times 10^{-5}$. Because this choice of parameter value might seem somewhat arbitrary, we checked that the distribution of F_{ST} conditional on heterozygosity was robust to alternative values ($\theta = 0.02$ and $\theta = 2.0$), particularly in the range of heterozygosity observed at $AXGT$ ($H_e = 0.207$) (data not shown).

The maximum frequency of the most common allele allowed was set to 0.99, for the simulated datasets. To obtain a close approximation of the expected joint distribution of F_{ST} and heterozygosity, we followed the algorithm detailed in the Appendix of Vitalis et al. (2001). This algorithm takes the pairs of (F_{ST}, H_e) estimates from the simulations to generate a 2D histogram, which is a close approximation to the bivariate probability distribution. We further used the Averaged Shifted Histogram (ASH) algorithm (Scott, 2002) to smooth the probability distribution, and to provide a continuous "high probability region" that contains 90%, 95%, or 99% of the total probability distribution. Finally, we derived an empirical p-value for the joint (F_{ST}, H_e) estimate at the Pro11Leu polymorphism observed in the real dataset. To do so, we calculated the relative frequency of this observation over the full set of simulated data, i.e. the height of the 2D histogram's cell that corresponds to the Pro11Leu (F_{ST}, H_e) estimate. The empirical p-value was then calculated as the proportion of the bivariate probability distribution of the full set of simulations, which is less probable than the Pro11Leu (F_{ST}, H_e) estimate. A so-obtained p-value equal to, say, 0.05 indicates that 95% of the simulated data are more probable than the observed Pro11Leu (F_{ST}, H_e) estimate (or, conversely, that only 5% of the simulated data show as "extreme" (F_{ST}, H_e) estimates as at Pro11Leu). In that case, the Pro11Leu (F_{ST}, H_e) estimate would lie at the limit of the 95% "high probability region".

Because we also aimed to test whether the proportion of genetic variance explained by diet was larger than that of the rest of the genome, hierarchical analyses of molecular variance (AMOVAs) were performed, using diet as an explaining factor. Sedentary agriculturalist populations from Central Asia (see Table 1), along with the Sichuan Chinese, the Armenians, the North Welsh and the Nigerians were considered as having a meat-poor diet, and the traditionally nomadic herders from Central Asia (see Table 1), along with the Mongols, as having a meat-rich diet. We used exactly the same simulated data as for the Dfdist analysis previously described, except that we considered diet as a fixed factor in the AMOVA. Hierarchical AMOVAs were all performed with the statistical software R (R Development Core Team, 2007) using the ade4 package (Dray & Dufour, 2007) to calculate the genetic variance components. For each simulated dataset, we calculated the parameter F_{CT}, which measures the part of the variation accounting for differences between diet groups. We then estimated the joint distribution of F_{CT} conditional upon heterozygosity, and proceeded in much the same way as for the joint distribution of F_{ST} and H_e.

Results and Discussion

AGXT Genetic Diversity

We found only two alleles at the Pro11Leu polymorphism in *AGXT* in the Central Asian samples: the major allele without Pro11Leu (Ma, 512 bp) and the minor allele with Pro11Leu (Mi, 619 bp), which differ by the presence of a 74-bp duplication in intron 1 (see the Materials and Methods section). We did not observe a 586-bp fragment, which would correspond to an allele with the 74-bp duplication but without Pro11Leu (Mi^A allele). We can therefore conclude that the Pro11Leu substitution is strongly associated to the 74-bp duplication in Central Asian populations, as it was found in Caucasian populations (Purdue et al., 1991), but not in South African populations (Coulter-Mackie et al., 2003).

Does the Pro11Leu Allele Frequency Correlate with Lifestyle?

Despite within-group heterogeneity (see Table 1), the frequency of the derived T allele of the Pro11Leu polymorphism is significantly higher among agriculturalists (17.6–26.9%) than among herders (1.7 – 15.2%) in Central Asia

(Wilcoxon signed-rank test, p-value = 0.003). This contradicts Danpure's (1997) well accepted hypothesis, which claims that this presumably detrimental mutation in *AGXT*, involved in primary hyperoxaluria type 1 disease, might be positively selected in human populations with a more carnivorous diet, such as herders. Yet our results are not the only ones to challenge Danpure's (1997) hypothesis. It is indeed worth noting that Caldwell et al. (2004) also found a low frequency of the T allele (6.9%) in their Mongol sample, although the diet of this population is known to be strongly meat-based (Hruschka & Brandon, 2004), and a high frequency of the T allele (19.7%) in their Norwegian sample, although Norway is a farming population with a mixed ancestral diet.

Is Genetic Differentiation Stronger at AGXT as Compared to the Rest of the Genome?

We used the overall F_{ST} estimate among 16 worldwide populations, calculated from the 27 presumably neutral STR loci (F_{ST} = 0.019), to perform the coalescent simulations and generate the distribution of F_{ST} conditional upon heterozygosity expected for neutral, bi-allelic co-dominant markers. As depicted in Figure 1a, we found that the Pro11Leu polymorphism (F_{ST} = 0.025) does not depart significantly from the neutral expectation, given the STR data (p-value = 0.214). This suggests that demography mainly shapes the observed variation at the Pro11Leu polymorphism. It might well be that the present sampling scheme better fits a hierarchical island model (Slatkin & Voelm, 1991, Vigouroux & Couvet, 2000), with some populations exchanging more migrants within ethnic groups than between ethnic groups. This is the case, e.g., of Central Asian herder populations, which are less differentiated on average, as compared to agriculturalist or worldwide populations (see Table 2). Yet, Excoffier et al. (2009) have shown that taking the hierarchical structure of populations into account results in a much wider distribution of F_{ST} vs. H_e than expected with a strict island model. Hence, ignoring the hierarchical structure, when it exists, tends to generate an excess of false positive outlier loci. Here, since the Pro11Leu polymorphism does not depart from neutrality when we assume a simple island model, we also expect Pro11Leu to lie within the F_{ST} vs. H_e distribution, when accounting for a more complex hierarchical structure.

Interestingly, this result somehow confirms that of Caldwell et al. (2004) who also found that the genetic differentiation measured at Pro11Leu was not significantly different from the genome-wide (presumably neutral) differentiation. Although they claimed that the genetic differentiation at Pro11Leu lay in the tail of the presumably neutral distribution, they based most of their analyses on the pairwise comparisons between the Saami population and both the Chinese and the Nigerian populations. Yet, it is worth stressing that the Saami population is commonly considered as a genetic isolate in Europe (Tambets et al., 2004) and is therefore not easily comparable to the HGDP-CEPH sample. This is why we discarded this sample from the present study. As it is very likely that the Saami population underwent strong genetic drift, demography could indeed account for the high frequency of the Pro11Leu T allele observed in this population.

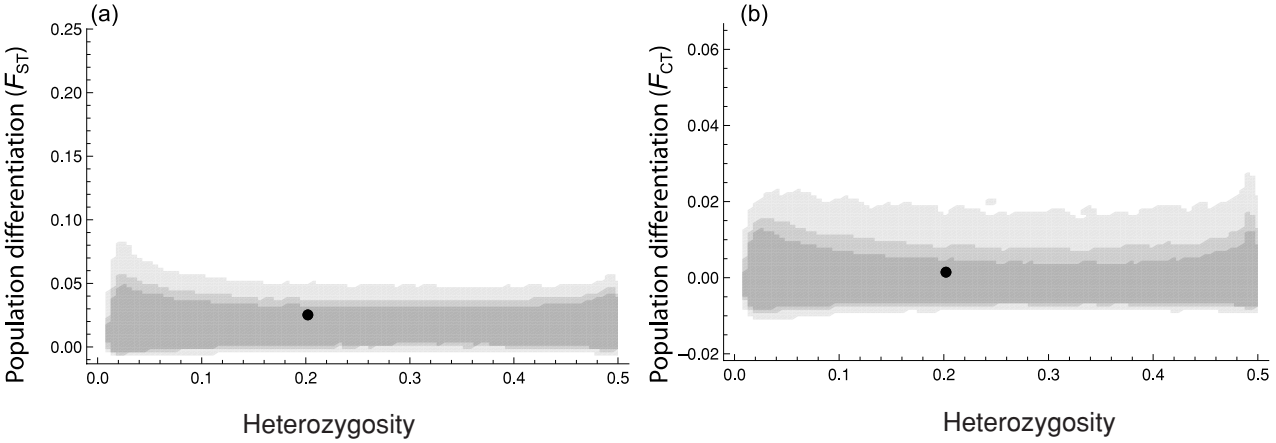

Figure 1 Joint distribution of F_{ST} (a) and F_{CT} (b) vs. the overall heterozygosity of the pooled sample (H_e) expected under neutrality, across 16 worldwide populations. The 90%, 95% and 99% confidence regions of the null distribution are shown from darker to lighter grey, respectively. The smoothed density was obtained using the Average Shifted Histogram (ASH) algorithm (Scott, 2002) with smoothing parameter m = 2. The black dots represent the observed measures of differentiation and heterozygosity at the Pro11Leu polymorphism.

Table 2 Pairwise F_{ST} between the studied populations, for 27 autosomal STRs

	Han	Mongola	Orcadian	Adygei	Yoruba	KAZ	KKK	KRA	KRB	KRG	LKZ	TUR	TDU	TJE	TJR	TJY
Han		-	***	***	***	***	***	***	***	***	***	***	***	***	***	***
Mongola	−0.004		***	***	***	*	***	***	_	***	*	***	***	***	***	***
Orcadian	0.115	0.100		-	***	***	***	***	***	***	***	***	***	***	***	***
Adygei	0.090	0.081	0.009		***	***	***	***	***	***	***	***	***	***	***	***
Yoruba	0.132	0.119	0.065	0.059		***	***	***	***	***	***	***	***	***	***	***
KAZ	0.020	0.010	0.054	0.037	0.078		***	***	**	_	_	***	***	***	***	***
KKK	0.026	0.019	0.043	0.024	0.077	0.005		***	**	***	_	***	***	***	***	***
KRA	0.018	0.012	0.062	0.044	0.085	0.004	0.007		**	***	***	***	***	***	***	***
KRB	0.012	0.005	0.063	0.041	0.087	0.002	0.004	0.002		***	_	***	***	***	***	***
KRG	0.028	0.021	0.058	0.041	0.083	0.003	0.010	0.011	0.008		*	***	***	***	***	***
LKZ	0.022	0.013	0.043	0.027	0.075	0.001	0.000	0.004	0.002	0.004		***	***	***	***	***
TUR	0.051	0.042	0.022	0.012	0.057	0.016	0.012	0.020	0.021	0.021	0.010		***	***	***	***
TDU	0.063	0.054	0.029	0.012	0.064	0.022	0.020	0.030	0.029	0.025	0.016	0.008		***	***	***
TJE	0.068	0.058	0.025	0.011	0.060	0.022	0.020	0.031	0.030	0.027	0.017	0.011	0.016		***	***
TJR	0.053	0.046	0.031	0.012	0.062	0.017	0.013	0.021	0.022	0.018	0.010	0.011	0.012	0.009		***
TJY	0.093	0.082	0.029	0.014	0.060	0.043	0.033	0.048	0.053	0.048	0.033	0.024	0.022	0.022	0.023	

Refer to Table 1 for population acronyms. F_{ST} estimates (below the diagonal) were calculated with the software package Genepop v. 4.0 (Rousset, 2008). Results from exact tests of differentiation are given above the diagonal. -: non significant, *: $p < 0.05$, **: $p < 0.01$ and ***: $p < 0.001$.

Is Genetic Variance Due to Diet Larger for AGXT as Compared to the Rest of the Genome?

To answer this question, we calculated from each simulated dataset the parameter F_{CT}, which measures the part of the variation accounting for differences between diet groups. The results are depicted in Figure 1b: the part of the variation accounting for differences between diet groups, measured at the Pro11Leu polymorphism ($F_{CT} = 0.002$), does not depart significantly from neutral expectation, given the STR data (p-value = 0.187).

Alternative Evolutionary Hypotheses

Even though our results show that the differentiation at the Pro11Leu polymorphism does not significantly depart from neutral expectation, we cannot rule out that the observed pattern has been shaped by two opposite selective forces. On the one hand, the Pro11Leu mutation has a presumably deleterious effect, due to the PH1 disease, while on the other hand this mutation is argued to be advantageous in association with a meat-rich diet. Both forces could therefore compensate each other and wipe out a signature of selection, at least among herder populations. This, however, barely explains the extreme differences in allele frequency observed between, e.g., the Saami and the Mongol populations (Caldwell et al., 2004), which both share a meat-rich diet.

As the Neolithic transition to agriculture took place 12,000–9,000 BC in Asia and 7,000–5,000 BC in Europe (and probably even later in Northern Europe), the change in diet may be so recent that the genetic signature of an adaptation to an ancestral predominantly meat-based diet still remains. This may explain why the Pro11Leu is found at such high frequency in, e.g., the Norwegian population. It might also be that in this latter agriculturalist population, individuals have more meat in their diet, as compared to other agriculturalist populations. Yet, neither hypothesis provides a satisfactory interpretation of the low frequency of Pro11Leu observed in Mongolian and Central Asian herders.

One alternative interpretation could be that the associated causative mutation (Gly170Arg) is not present at the same frequency in European and Asian populations, therefore allowing the Pro11Leu mutation to reach higher frequencies in Europe. It would therefore be interesting to collect worldwide frequency data for Gly170Arg. It might also be that another, yet unknown, advantageous mutation in *AGXT* in Mongolian and Central Asian populations provides alternative means to detoxify glyoxylate from meat, in the absence of the Pro11Leu substitution. Such a convergent adaptation has been described for lactase persistency, with two different mutations in the *LCT* gene providing the same phenotype in Africa and Europe (Tishkoff et al., 2007). Yet it is not clear how this hypothesis would account for the high frequency of the Pro11Leu T allele observed in the agriculturalist populations of Central Asia. Furthermore, because of its implication in PH1 disease, the *AGXT* gene has been extensively investigated (Danpure, 2004), and it seems unlikely that causative amino acid substitutions have been overlooked.

In conclusion, our results show that the differentiation at the Pro11Leu polymorphism does not depart from neutral expectation, and do not support the general idea that positive selection has been, or still is, shaping the genetic variation at the *AGXT* locus. Yet, importantly, we cannot exclude some complex scenarios that would involve a combination of the previous hypotheses: for example, the presence of another advantageous allele in Mongolian and Central Asian herder populations (distinct from Pro11Leu), together with a remnant signature of adaptation of to a meat-rich diet among agriculturalist populations in Europe, might explain the pattern observed. There is no doubt that sequence data in the *AGXT* locus would provide powerful tools to test such complex scenarios, by allowing the use of statistical tests of neutrality within populations, e.g., based on the frequency spectrum of DNA polymorphic sites. However, our approach, which aimed to evaluate whether worldwide genetic variation at the Pro11Leu substitution was better explained by demographic history or by adaptation to diet, has been shown to be robust to the vagaries of demographic history (Beaumont, 2005), even in the case of partial selection (see Beaumont & Balding, 2004). Finally, our conclusion that the observed allele frequency differences at the Pro11Leu substitution may result from demographic history rather than selection, supports the recent claim that the observed large allele frequency differences at a number of genetic polymorphisms result from neutral processes rather than from local adaptation (Hofer et al., 2009, Coop et al., 2009).

Acknowledgements

We thank Frédéric Austerlitz for helpful comments on this manuscript. This work was supported by the Centre National de la Recherche Scientifique (CNRS) ATIP programme (to EH), by the CNRS interdisciplinary programme ''Origines de l'Homme du Langage et des Langues'' (OHLL), by the European Science Foundation (ESF) EUROCORES programme ''The Origin of Man, Language and Languages'' (OMLL), by the MNHN department ''Homme, Nature, Société'' and by the ANR-funded programme NUTGENEVOL. We also thank the ''Fondation pour la Recherche Médicale'' (FRM) for financial support. LS was financed by the French Ministry of Higher Education and Research.

References

Beaumont, M. A. (2005) Adaptation and speciation: what can F_{ST} tell us? *Trends Ecol Evol* **20**, 435–440.

Beaumont, M. A. & Balding, D. J. (2004) Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* **13**, 969–980.

Beaumont, M. A. & Nichols, R. A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc Lon* **263**, 1619–1626.

Bonin, A., Taberlet, P., Miaud, C. & Pompanon, F. (2006) Exploratory genome scan to detect loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol Biol Evol* **23**, 773–783.

Caldwell, E. F., Mayor, L. R., Thomas, M. G. & Danpure, C. J. (2004) Diet and the frequency of the alanine:glyoxylate aminotransferase Pro11Leu polymorphism in different human populations. *Hum Genet* **115**, 504–509.

Coop, G., Pickrell, J. K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., Myers, R. M., Cavalli-Sforza, L. L., Feldman, M. W. & Pritchard, J. K. (2009) The role of geography in human adaptation. *PLoS Genet* **5**, e1000500.

Cooper, P. J., Danpure, C. J., Wise, P. J. & Guttridge, K. M. (1988) Immunocytochemical localization of human hepatic alanine: glyoxylate aminotransferase in control subjects and patients with primary hyperoxaluria type 1. *J Histochem Cytochem* **36**, 1285–1294.

Coulter-Mackie, M. B., Tung, A., Henderson, H. E., Toone, J. R. & Applegarth, D. A. (2003) The AGT gene in Africa: a distinctive minor allele haplotype, a polymorphism (V326I), and a novel PH1 mutation (A112D) in Black Africans. *Mol Genet Metab* **78**, 44–50.

Danpure, C. J. (1997) Variable peroxisomal and mitochondrial targeting of alanine: glyoxylate aminotransferase in mammalian evolution and disease. *Bioessays* **19**, 317–326.

Danpure, C. J. (2004) Molecular aetiology of primary hyperoxaluria type 1. *Nephron Exp Nephrol* **98**, e39-e44.

Danpure, C. J., Birdsey, G. M., Rumsby, G., Lumb, M. J., Purdue, P. E. & Allsop, J. (1994a) Molecular characterization and clinical use of a polymorphic tandem repeat in an intron of the human alanine:glyoxylate aminotransferase gene. *Hum Genet* **94**, 55–64.

Danpure, C. J., Cooper, P. J., Wise, P. J. & Jennings, P. R. (1989) An enzyme trafficking defect in two patients with primary hyperoxaluria type 1: peroxisomal alanine/glyoxylate aminotransferase rerouted to mitochondria. *J Cell Biol* **108**, 1345–1352.

Danpure, C. J., Guttridge, K. M., Fryer, P., Jennings, P. R., Allsop, J. & Purdue, P. E. (1990) Subcellular distribution of hepatic alanine:glyoxylate aminotransferase in various mammalian species. *J Cell Sci* **97**, 669–678.

Danpure, C. J. & Jennings, P. R. (1986) Peroxisomal alanine:glyoxylate aminotransferase deficiency in primary hyperoxaluria type 1. *FEBS Letters* **201**, 20–24.

Danpure, C. J., Jennings, P. R., Fryer, P., Purdue, P. E. & Allsop, J. (1994b) Primary hyperoxaluria type 1: genotypic and phenotypic heterogeneity. *J Inherit Metab Dis* **17**, 487–499.

Danpure, C. J. & Rumsby, G. (1996) Strategies for the prenatal diagnosis of primary hyperoxaluria type 1. *Prenat Diagn* **16**, 587–598.

Dray, S. & Dufour, A. B. (2007) The ade4 package: implementing the duality diagram for ecologists. *J Stat Soft* **22**, 1–20.

Excoffier, L., Hofer, T. & Foll, M. (2009) Detecting loci under selection in a hierarchically structured population. *Heredity advance online publication*, doi:10.1038/hdy.2009.74.

Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Haglin, L. (1991) Nutrient intake among Saami people today compared with an old, traditional Saami diet. *Arctic Med Res* (Suppl), 741–746.

Haglin, L. (1999) The nutrient density of present-day and traditional diets and their health aspects: the Sami- and lumberjack families living in rural areas of Northern Sweden. *Int J Circ Health* **58**, 30–43.

Hofer, T., Ray, N., Wegmann, D. & Excoffier, L. (2009) Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Ann Hum Genet* **73**, 95–108.

Holbrook, J. D., Birdsey, G. M., Yang, Z., Bruford, M. W. & Danpure, C. J. (2000) Molecular adaptation of alanine:glyoxylate aminotransferase targeting in primates. *Mol Biol Evol* **17**, 387–400.

Hruschka, D. J. & Brandon, A. K. (2004) Mongolia. In: *Encyclopedia of medical anthropology. Volume II: Cultures* (eds. C. R. Ember & M. Ember), pp. 850–862. New York: Springer.

Kozlov, A., Borinskaya, S., Vershubsky, G., Vasilyev, E., Popov, V., Sokolova, M., Sanina, E., Kaljina, N., Rebrikov, D., Lisitsyn, D. & Yankovsky, N. (2008) Genes related to the metabolism of nutrients in the Kola Sami population. *Int J Circ Health* **67**, 56–66.

Lumb, M. J. & Danpure, C. J. (2000) Functional synergism between the most common polymorphism in human alanine:glyoxylate aminotransferase and four of the most common disease-causing mutations. *J Biol Chem* **275**, 36415–36422.

Maniatis, T., Fritsh, E. F. & Sambrook, J. (1982) *Molecular cloning. A laboratory manual*. New York: Cold Spring Harbor Laboratory.

Noguchi, T. (1987) Aromatic-amino-acid aminotransferase from small intestine. *Methods Enzymol* **142**, 267–273.

Purdue, P. E., Lumb, M. J., Allsop, J. & Danpure, C. J. (1991) An intronic duplication in the alanine: glyoxylate aminotransferase gene facilitates identification of mutations in compound heterozygote patients with primary hyperoxaluria type 1. *Hum Genet* **87**, 394–396.

Purdue, P. E., Takada, Y. & Danpure, C. J. (1990) Identification of mutations associated with peroxisome-to-mitochondrion mistargeting of alanine/glyoxylate aminotransferase in primary hyperoxaluria type 1. *J Cell Biol* **111**, 2341–2351.

R Development Core Team (2007) R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, URL http://www.R-project.org.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A. & Feldman, M. W. (2002) Genetic structure of human populations. *Science* **298**, 2381–2385.

Rousset, F. (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Res* **8**, 103–106.

Scott, D. W. (2002) Multivariate Density Estimation: Theory, Practice, and Visualization. New York: Wiley.

Ségurel, L., Martinez-Cruz, B., Quintana-Murci, L., Balaresque, P., Georges, M., Hegay, T., Aldashev, A., Nasyrova, F., Jobling, M. A., Heyer, E. & Vitalis, R. (2008) Sex-specific genetic structure and social organization in Central Asia: insights from a multi-locus study. *PLoS Genet* **4**, e1000200.

Slatkin, M. & Voelm, L. (1991) F_{ST} in a hierarchical island model. *Genetics* **127**, 627–629.

Takayama, T., Fujita, K., Suzuki, K., Sakaguchi, M., Fujie, M., Nagai, E., Watanabe, S., Ichiyama, A. & Ogawa, Y. (2003) Control of oxalate formation from L-hydroxyproline in liver mitochondria. *J Am Soc Nephrol* **14**, 939–946.

Tambets, K., Rootsi, S., Kivisild, T., Help, H., Serk, P., Loogvali, E. L., Tolk, H. V., Reidla, M., Metspalu, E., Pliss, L., Balanovsky, O., Pshenichnov, A., Balanovska, E., Gubina, M., Zhadanov, S., Osipova, L., Damba, L., Voevoda, M., Kutuev, I., Bermisheva, M., Khusnutdinova, E., Gusar, V., Grechanina, E., Parik, J., Pennarun, E., Richard, C., Chaventre, A., Moisan, J. P., Barac, L., Pericic, M., Rudan, P., Terzic, R., Mikerezi, I., Krumina, A., Baumanis, V., Koziel, S., Rickards, O., De Stefano, G. F., Anagnou, N., Pappa, K. I., Michalodimitrakis, E., Ferak, V., Furedi, S., Komel, R., Beckman, L. & Villems, R. (2004) The western and eastern roots of the Saami–the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes. *Am J Hum Genet* **74**, 661–682.

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., Lema, G., Nyambo, T. B., Ghori, J., Bumpstead, S., Pritchard, J. K., Wray, G. A. & Deloukas, P. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* **39**, 31–40.

Vigouroux, Y. & Couvet, D. (2000) The hierarchical island model revisited. *Genet Sel Evol* **32**, 395–402.

Vitalis, R., Dawson, K. & Boursot, P. (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics* **158**, 1811–1823.

Williams, H. E. & Smith, L. H. (1983) Primary Hyperoxaluria. In: *The Metabolic Basis of Inherited Disease* (eds. J. B. Stanbury, J. B. Wyngaarden, D. S. Frederickson, J. L. Goldstein & M. S. Brown), pp. 204–228. New York: McGraw-Hill.

Wright, S. (1951) The genetical structure of populations. *Ann Eugen* **15**, 323–354.

Received: 19 May 2009
Accepted: 7 September 2009