ANTHROPOLOGIE • LX/1 • pp. 173–186 • 2022

POLINA R. BUTOVSKAYA, M. ESTHER ESTEBAN, NIKOS AVRAMIDIS, OLEG E. LAZEBNY, JULIA N. FEDENOK, IRINA G. UDINA, MARINA L. BUTOVSKAYA, ARACELI ROSA

LARGE DIFFERENCES IN THE PROPORTION OF PEOPLE WITH REDUCED MTHFR ENZYMATIC ACTIVITY IN EURASIA: ALLELE AND HAPLOTYPE DISTRIBUTION OF MTHFR-C677T AND MTHFR-A1298C POLYMORPHISMS IN SPAIN AND SIBERIA

ABSTRACT: Adequate levels of folate are crucial for human development. MTHFR (Methylene tetrahydrofolate reductase) is involved in the metabolism of folic acid and other B vitamins. The aim of this work is to examine the allele frequency and haplotype distribution of MTHFR-C677T and MTHFR-A1298C polymorphisms in two independent samples (North-East Spanish and North-Western Siberia). The degree of population differentiation and signatures of positive selection were also examined in 18 world populations.
The samples comprised 623 Spanish individuals and 172 Ob-Ugric people (Khanty and Mansi).
The Spanish sample showed 44% individuals with 30% reduced enzyme activity, and 16% people with 70% reduced activity; whereas the Khanty sample revealed only 3% people with 70% reduced activity. In the 18 populations screened, we found significant South to North clines for SNPs and haplotypes. Overall, we failed to detect any traces of positive selection.
The high frequency of people with reduced MTHFR enzyme activity in Spain highlights the need for diets rich in folate and folic acid supplementation in some groups. The low frequency of individuals with reduced enzymatic activity among the Khanty together with their traditional diet indicates a protective combination against pathologies related to folic acid deficiencies.

KEY WORDS: MHTFR *gene – A1298C (rs1801131) and C677T (rs1801133) – Folic acid – Homocysteine – Positive selection – Cardiovascular risk.*

Received 12 November 2020; received in revised form 19 March 2021; accepted 5 April 2021. Available online 1 October 2021.

DOI: https://doi.org.10.26720/anthro.21.05.04.1

IINTRODUCTION

Methylenetetrahydrofolate reductase (MTHFR) catalyses the conversion of 5,10-methylenetetrahydrofolate (5,10-MTHF) to 5-methylenetetrahydrofolate, a predominant circulating form of folate which is also known as vitamin B9 (Goyette *et al.* 1994). This reaction is required for the multistep process that converts the amino acid homocysteine (Hcy) to methionine.

Mutations in *MTHFR*, a gene with 12 exons located on the short arm of chromosome 1 (1p36.22), can lead to hyperhomocysteinaemia, a condition with increased homocysteine concentrations in blood. The polymorphism C677T (rs1801133) in the coding region of *MTHFR* is a point mutation that replaces cytosine with thymine in exon 4, resulting in a change in alanine to valine (p.Ala222Val or A222V) in the N-terminal catalytic domain of MTHFR protein. The final product is thermolabile form with reduced enzymatic activity (Frosst *et al.* 1995). Another mutation leading to reduced MTHFR activity is the A1298C transversion in exon 7 (rs1801131). This polymorphism replaces adenine with cytosine resulting in a change in glutamic acid to alanine (p.Glu429Ala or E429A) in the C-terminal regulatory domain of the protein (Van Der Put *et al.* 1998). It is confirmed that the 677T and 1298C alleles result in loss of enzyme activity by approximately 70% and 32%, respectively (Weisberg *et al.* 1998). Individuals carrying the homozygous genotypes of the mutant variants (*TT* and *CC*, respectively) show lower levels of serum folate and higher levels of total Hcy compared with the other two genotypes. Decreased folate and increased plasma homocysteine levels are risk factors for cardiovascular disease, neural tube defects, cleft lip/palate, hypertension, preeclampsia, thrombosis, osteoporosis, dementia, Alzheimer's disease, Down syndrome, certain types of cancer, glaucoma, pregnancy complications, migraine, epilepsy, depression, and schizophrenia (Den Heijer *et al.* 2005, Katayama *et al.* 2007, Holmes *et al.* 2011, Shiao *et al.* 2018, Tinelli *et al.* 2019, Nefic *et al.* 2018).

These two single nucleotide polymorphisms (SNPs) exhibit different allele frequencies among different world populations. By continents, Africa shows the lowest frequencies for both 677T and 1298C alleles (9% and 15%, respectively), related to decreased enzyme activity. In contrast, some South-East Asian (42% for 1298C) populations exhibit the highest frequencies (1000 Genomes Project Consortium *et al.* 2015). The allele frequencies of 677T and 1298C variants have been associated with latitude in African and Eurasian populations (Jones *et al.* 1992). The low frequencies observed in African groups may be an adaptation to the differential degradation of folate depending on the degree of skin pigmentation, as suggested in the well-known vitamin D-folate Hypothesis (Jablonski, Chaplin 2000). In line with the above, the responsiveness of MTHFR to ultraviolet radiation (UVR) could explain the significant negative association between the 677T allele and UVR exposure (Jones *et al.* 2016).

The *MTHFR* gene has been recently found to show positive selection signatures in the form of a sweeping overlap in some European (CEU, IBS, GBR) and Asian populations (JPT) from the 1000 Genomes project (McLaren *et al.* 2016). This finding suggests a possible positive selection process underlying the variation of *MTHFR*, by which some alleles rapidly increased their frequency because of a fitness advantage under new selective pressures. The remarkably decreased enzyme activity of the 677T variant, and that of the 1298C variant to a lesser degree, combined with habitats having different exposures to UVR and/or diets with different folate contents may be the basis for such a fitness advantage (Hancock *et al.* 2010, Brown 2011).

The frequency of *MTHFR*-C677T and *MTHFR*-A1298C polymorphisms has been positively associated with latitude. These findings provide novel evidence suggesting that folate genotypes could have been selected to maintain homeostasis between folate-dependent *de novo* thymidylate synthesis and methylation pathways in environments with differing UV levels (Jones *et al.* 2016). The frequency of the polymorphic *T* allele in the *MTHFR*-C677T was found to be positively correlated with latitude, where increase in latitude degrees accompanied an increase in *T* allele frequency. Similarly, a positive latitude-dependent relationship was detected for the polymorphic *C* allele of the *MTHFR*-A1298C variant. Some studies have also reported the protective role of mutant *TT* genotype of the C677T polymorphism for colon cancer in some populations (Shiao *et al.* 2018).

Recent meta-analyses have revised the role of *MTHFR* polymorphisms together with other variables such as the ethnic group, diet, geography, and climate. A study found a larger effect of the 677T variant on Hcy concentration, and hence, an increased risk of stroke, in populations with low-folate diets (Holmes *et al.* 2011). It has been recently suggested that the beneficial or harmful effects of *MTHFR* polymorphisms

can be modulated by global warming and air pollution. In this manner, the wild type (*C* allele) of C677T showed a protective role for ischemic heart disease, particularly in highly polluted countries (Chen *et al.* 2018). The effect of climate also seemed to modulate the role of the *TT* genotype of C677T as either a protective factor or a risk factor for colorectal cancer. The *TT* genotype has been found to play a protective role in many world populations except in countries from southern regions and hot climates, where this genotype is a risk factor (Shiao *et al.* 2018).

Considering the marked frequency differences of C677T and A1298C MTHFR polymorphisms and their relevance in a large number of pathologies, we analysed their combined variability in two populations at both geographical ends of Europe. On one side, a North-East Spanish sample from Catalonia representative of a current urban South-European population was included, and on the other, a population from the Berezovsky region in North-Western Siberia, more precisely, a population from Sosvensky or Northern Khanty. The latter population is representative of the Ob-Ugric people (Khanty and Mansi) in Western Siberia along the basins of the Ob and the Irtysh rivers. According to the 2010 census, Ob-Ugric people are just over 43,000 in number (Khanty – 31,000 and Mansi – 12,000) (Butovskaya *et al.* 2016). The Khanty appear as a unique intermediate population living in the contact zone of genetically distinguishable Eastern and Western Eurasia (Pimenoff *et al.* 2008). Although most of the Ob-Ugric people currently live in villages, many of them are still practicing traditional occupations such as fishing, hunting, reindeer herding, and gathering (Butovskaya *et al.* 2017).

Other world populations have also been included in the analyses to determine the degree of population differentiation, particularly in Eurasia. Haplotypes of *MTHFR* have been described in some Eurasian populations (Trifonova *et al.* 2012). However, to our knowledge there are no previous studies assessing the haplotype distribution of A1298T and C677T in other world populations. We have also explored the existence of a cline for the two SNPs and their haplotypes, as well as the signatures of positive selection in *MTHFR*.

MATERIALS AND METHODS

Samples

The present study included samples of non-related healthy individuals (both sexes) from two different Eurasian populations: 1) a North-East Spanish sample from Catalonia, and 2) a sample from the Berezovsky region in North-Western Siberia. The first sample comprised 623 students recruited in Barcelona. The second sample included 172 Ob-Ugric people (Khanty and Mansi) from urban settlements located along the Ob and Sosva rivers in the Khanty-Mansiysky Autonomous District of North-Western Siberia, Russia. All participants from the second sample were born in urban localities and practiced traditional occupations including fishery, hunting and foraging, or reindeer herding. They usually keep reindeer close to their homes in the forest, and migrate with them to the mountain slopes or to the river-bank in summer (Vorobeva *et al.* 2015).

The participants volunteered to be included in the study and were not preselected based on any criteria. Institutional approvals from the Universitat Autónoma de Barcelona Ethics Committee and University and Moscow State University Ethics Committee for data collection in Russia from the Ob-Ugric people, were obtained prior to conducting this study. All subjects provided informed, written consent prior to participation in the study.

Laboratory methods

All participants provided buccal samples. Genomic DNA was isolated using a Real Extraction DNA kit (Durviz S.L.U., Valencia, Spain) (sample 1) and Diatom DNA Prep 200 (Isogen Lab, Moscow, Russia) (samples 2).

The two SNPs of *MTHFR* included in the study (rs1801131 and 1801133) were genotyped using the TaqMan allelic discrimination assay from Life Technologies (Thermo Fisher Scientific, California, USA). The assays were run in a 384-well plate on the ABI PRISM 7900HT Fast Real-Time PCR System (Applied Biosystems, California, USA) using standard conditions. The final volume of each well was 5 µl, which contained 5 ng of genomic DNA, 2.5 µl of TaqMan Master Mix, and 0.125 µl of the 40× genotyping assay (assays C_850486_20 and C_1202883_20, respectively). SDS v.2.4 software was used for genotype data analysis. To ensure the validity and accuracy of genotyping, we retested a 20% random sample. In all cases, the genotypes were reproducible.

Statistical analyses

A database of 20 populations with individual genotypic data for the rs1801131 and rs1801133 SNPs was assembled using original data from the two

populations genotyped here (Catalonia from NE Spain and Khanty-Mansi from NW Siberia) and additional data retrieved from the 1000Genomes Phase 3 Browser (http://phase3browser.1000genomes.org/index.html) and the Estonian Biocentre (http://evolbio.ut.ee/). The populations comprised 15 Eurasian and 5 African samples. Information about the origin of samples and their geographical location is indicated in *Figure 1*. Allele and haplotype frequencies, gene diversities, Hardy-Weinberg equilibrium fitting, linkage disequilibrium (LD) between rs1801131 and rs1801133, and molecular variation were determined using Arlequin v.3.5 (Excoffier, Lischer 2010). The existence of frequency clines in the 20 populations mentioned above was explored using GenoCline v1.0 software (http://www.didac.ehu.es/genocline). The spatial pattern distribution of allelic and haplotype frequencies was checked using the coefficient of determination (R^2) and Pearson's r correlation coefficient.

As a first attempt to identify positive selection signatures, we examined *MTHFR* using the population genomics-oriented genome browser, PopHuman (Casillas *et al.* 2018). The region of interest containing the gene was of 28.2 Kb (chr1:11841729-11869924), and the samples explored were from CEU (Utah residents (CEPH) with Northern and Western European ancestry), FIN (Finnish in Finland), IBS (Iberian Populations in Spain), Han Chinese in Beijing, China (CHB), and Yorubans from Nigeria (YRI). Signatures of positive selection were also checked in samples from Finland (FIN), an Iberian Population in Spain (IBS), Han Chinese in Beijing (CHB), and Yoruba in Nigeria (YRI) in the 1000Genomes project. The IBS and FIN samples were used as representatives of the South and North European variation, respectively. The sequence downloaded for the *MTHFR* gene comprised the region 11785723–11806920 of the first human chromosome (ensemble database GRH38). Two indices were examined with the Selscan program: iHS (integrated Haplotype Scores) and nSL (number of Segregating sites by Length). iHS refers to the decay of homozygosity between ancestral and derived haplotypes using recombination distance. nSL is a similar procedure that uses segregation sites as a distance measure. Extreme negative scores indicate that haplotypes on the derived allele are longer

TABLE 1: SNP variation (Minor Allele Frequencies) and heterozygosity values in the NE Spain and NW Siberian samples. Data from other European, Asian and African groups downloaded from 1000Genomes[1] and the Estonian Biocentre[2] are also included. The geographical position of the samples is shown in *Supplementary Figure 1*. [1]1000Genomes project (http://www.internationalgenome.org/), [2]Estonian Biocentre (http://evolbio.ut.ee/), [3]Kovacevic *et al.* 2014, [4]Tambets *et al.* 2018, [5]Yunusbayev *et al.* 2015.

		NE Spain	**Iberian P[1]**	**Italy[1]**	**Balkans[2,3]**	**Great Britain[1]**	**Finland[1]**	**Uralic speakers[2,4]**	**NW Siberia**
	N	**623**	**107**	**107**	**70**	**91**	**99**	**44**	**172**
rs1801131 A1298C	C freq	0.290	0.271	0.313	0.243	0.341	0.318	0.250	0.313
rs1801133 C677T	T freq	0.394	0.444	0.467	0.450	0.324	0.273	0.295	0.175
		ME Turkish speakers[2,5]	**Mongolian Turkic speakers[2,5]**		**China[1]**	**Pakistan[1]**	**India[1]**	**Bangladesh[1]**	**Vietnam[1]**
	N	**176**	**68**		**103**	**96**	**102**	**86**	**99**
rs1801131 A1298C	C freq	0.355	0.331		0.223	0.417	0.466	0.413	0.257
rs1801133 C677T	T freq	0.247	0.147		0.466	0.135	0.103	0.122	0.192
		Gambia[1]	**Sierra Leone[1]**		**Esan Nigeria[1]**	**Yoruba Nigeria[1]**		**Kenya[1]**	
	N	**113**	**85**		**99**	**108**		**99**	
rs1801131 A1298C	C freq	0.115	0.135		0.136	0.120		0.187	
rs1801133 C677T	T freq	0.062	0.076		0.081	0.106		0.071	

compared to the haplotypes associated with the ancestral allele at the same locus whereas extreme positive values indicate the opposite effect. Significant values are considered when the minimum value of the tests is less than -2 (Ferrer-Admetlla *et al.* 2014, Szpiech, Hernandez 2014).

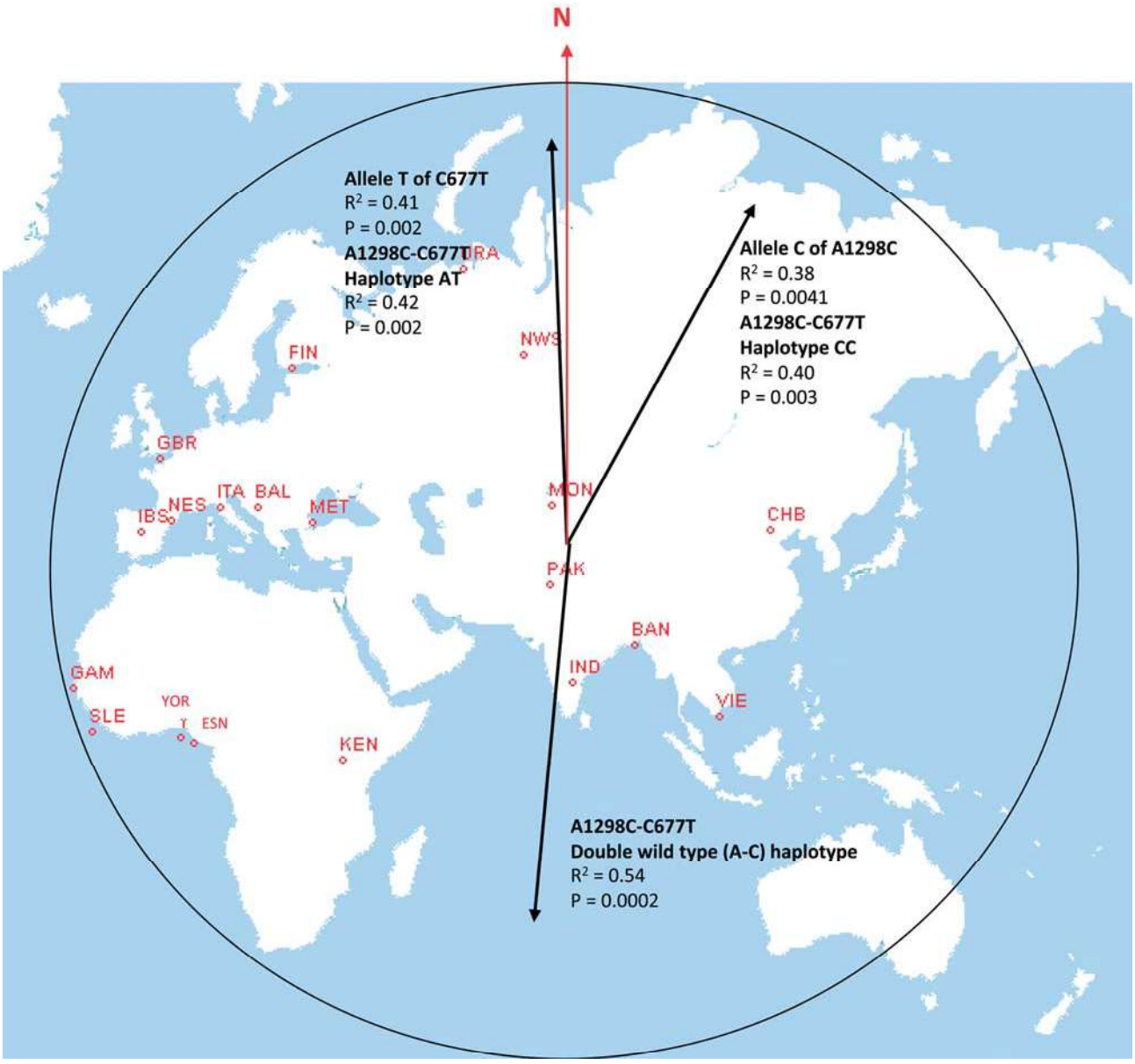

FIGURE 1: Increasing gradients of allele frequencies for C1298A, C677T and their haplotypes (as in *Table 2*) for 20 world populations. Black arrows departing from the circumference center indicate the orientation of the cline. The significant clines are expressed with the coefficient of determination of the linear regression (R2) and its statistical significance. *Population abbreviations:* IBS: Iberian Peninsula; NES: North-East Spain; ITA: Toscani in Italia; BAL: Balkans; FIN: Finland; GBR: British in England and Scotland; URA: Uralic speakers; NWS: North-West Siberia; MET: Middle-East Turkic speakers; MON: Mongolian Turkic speakers; CHB: Han Chinese in Beijing, China; PAK: Punjabi from Lahore, Pakistan; IND: Gujarati Indian from Houston, Texas; BAN: Bengali from Bangladesh; VIE: Kinh in Ho Chi Minh City, Vietnam; GAM: Gambian in Western Divisions, Gambia; SLE: Mende in Sierra Leone; YOR: Yoruba in Ibadan, Nigeria; ESN: Esan in Nigeria; KEN: Luhya in Webuye, Kenya.

TABLE 2: Frequencies (f) and standard deviation (SD) values (f/SD) for rs1801131 and rs1801133 haplotypes. Population references and sample sizes as in *Table 1*.

rs1801131 (A/C) and rs1801133 (C/T) Haplotypes	**NE Spain**	**Iberian P**	**Italy**	**Balkans**	**Great Britain**	**Finland**	**Uralic speakers**	**NW Siberia**
A-C (double wild type)	0.3360 / 0.0141	0.2850 / 0.0322	0.2196 / 0.0292	0.3071 / 0.0396	0.3352 / 0.0350	0.4091 / 0.0338	0.4545 / 0.0536	0.5117 / 0.0280
A-T	0.3743 / 0.0140	0.4439 / 0.0340	0.4673 / 0.0346	0.4500 / 0.0428	0.3242 / 0.0347	0.2727 / 0.0302	0.2955 / 0.0489	0.1754 / 0.0214
C-C	0.2700 / 0.0130	0.2710 / 0.0287	0.3131 / 0.0319	0.2429 / 0.0367	0.3407 / 0.0348	0.3182 / 0.0343	0.2500 / 0.0452	0.3129 / 0.0252
CT (double mutant)	0.0198 / 0.0048	- / -	- / -	- / -	- / -	- / -	- / -	- / -
LD probability	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

	ME Turkish speakers	**Mongolian Turkic speakers**	**China**	**Pakistan**	**India**	**Bangladesh**	**Vietnam**
A-C (double wild type)	0.3977 / 0.0260	0.5221 / 0.0422	0.3107 / 0.0326	0.4479 / 0.0338	0.4314 / 0.0344	0.4783 / 0.0390	0.5505 / 0.0357
A-T	0.2472 / 0.0230	0.1471 / 0.0290	0.4660 / 0.0357	0.1354 / 0.0247	0.1029 / 0.0206	0.1089 / 0.0248	0.1919 / 0.0294
C-C	0.3551 / 0.0255	0.3309 / 0.0401	0.2233 / 0.0302	0.4167 / 0.0344	0.4657 / 0.0344	0.3996 / 0.0393	0.2576 / 0.0320
CT (double mutant)	- / -	- / -	- / -	- / -	- / -	0.0132 / 0.0127	- / -
LD probability	< 0.001	0.007	< 0.001	< 0.001	< 0.001	0.035	0.027

	Gambia	**Sierra Leone**	**Esan Nigeria**	**Yoruba Nigeria**	**Kenya**
A-C (double wild type)	0.8230 / 0.0260	0.8023 / 0.0322	0.7981 / 0.0301	0.7731 / 0.0301	0.7462 / 0.0315
A-T	0.0619 / 0.0167	0.0624 / 0.0204	0.0655 / 0.0193	0.1065 / 0.0216	0.0669 / 0.0179
C-C	0.1150 / 0.0218	0.1212 / 0.0259	0.1210 / 0.0244	0.1204/ 0.0237	0.1831 / 0.0285
CT (double mutant)	- / -	0.0141 / 0.0110	0.0153 / 0.0107	- / -	0.0038 / 0.0067
LD probability	0.172	0.921	0.729	0.132	0.515

RESULTS

Allele, haplotype frequencies, and geographical clines

The allele frequencies of rs1801131 and rs1801133 in the NE Spanish and NW Siberian samples are shown in *Table 1*, along with data from other world populations. Both samples fitted the HWE for rs1801133. For rs1801131, the NE Spanish population fitted the HWE, whereas the NW Siberians did not (Supplementary *Table 1*). The observed heterozygosity in this group largely exceeded the expected value for this SNP. To identify the underlying reason, we divided the original sample into two sub-samples according to the origin of the individuals: A sub-sample of Khanty-Mansi individuals with both parents from the same origin, and a sub-sample of individuals with mixed origins (only one parent was Khanty-Mansi). The mixed sub-sample fitted the Hardy-Weinberg equilibrium for rs1801131 ($p = 0.0796$) whereas the first sub-sample still showed deviation from the equilibrium ($p = 0.013$). Deviation from equilibrium for the whole sample could thus be then be attributed to the population substructure. As allele frequencies in the sub-samples did not show significant differences among them, we considered the rs1801131 data for the complete Khanty-Mansi sample.

With regard to variation ranges (*Table 1*), allele *C* of A1298C (rs1801131) showed the lowest values in African groups (11.5% in Gambia) and the highest values in Asians (46.5% in India). The NE Spanish (28.9%) and NW Siberian (31.2%) samples fitted within the Eurasian ranges. Allele *T* of C677T (rs1801133) also showed its lowest values in African groups (6.2% in Gambia), and the highest ones in some European (Italy 46.7%) and Asian samples (China 46.6%). NE Spanish samples showed similarity to the samples with high values (39.4%), whereas NW Siberians (17.5%) resembled groups with low values.

Both SNPs showed in linkage disequilibrium in all samples except for the African samples. The haplotype frequencies and LD probabilities are indicated in *Table 2*. The haplotype carrying the wild type alleles *A* (rs1801131) and *C* (rs1801133) showed the highest frequencies in African samples (from 74.6% to 82.3%), and was also the most frequent haplotype among samples from Asia (except China), North Eurasia (NW Siberian and the Uralic speakers) and some North Europe (Great Britain and Finland) with frequencies ranging from 33%–55%. The double mutant haplotype (*CT*) reached frequencies up to 1% only in the NE Spanish sample, Bangladesh, Esan Nigeria, and Sierra Leone.

The results for allelic and haplotypic AMOVA analyses for different population groups are indicated in Supplementary *Table 2*. Polymorphism A1298C failed to show any significant degree of internal genetic variation in the non-hierarchical AMOVAs. With an exception of the African group that was highly homogeneous, the internal genetic variation of C677T in the remaining population clusters was remarkable. Significant F_{ST} values ranged from 8.82% in Asia to 1.42% in Europe. The haplotypic genetic variation showed the same trend as C677T, but with much lower F_{ST} values: from 0.6% in Europe to 4.35% in Asia. When different groups were compared, the strongest genetic structure was found between Africans and the remaining populations, particularly for the haplotype frequencies (among groups F_{CT} 16.54%, among populations within groups F_{SC} 3.19%). Significant differences among groups for C677T and haplotypes were also observed when Europeans, Asians, and North Eurasians were compared.

The genetic structure observed for individual SNPs and haplotypes was also attested by the significant genetic clines observed for the 20 populations (*Figure 1*). The mutant alleles of both polymorphisms showed a significant South to North cline (allele *C* of A1298C $R^2 = 0.38$, $p = 0.004$; allele *T* of C677T $R^2 = 0.41$, $p = 0.002$). The double wild type haplotype showed the opposite trend ($R^2 = 0.54$, $p = 0.0002$). Because of the linkage disequilibrium between SNPs, the *A-T* and *C-C* haplotypes showed significant clines following the same geographical pattern described for individual SNPs.

When only the Eurasian populations were considered (*Figure 2*), C1298A failed to show a significant cline. The spatial pattern of allele frequencies of the allele *C* of C677T and the *AT* haplotype followed a SE to NW cline ($R^2 = 0.30$, $p = 0.035$ and $R^2 = 0.30$, $p = 0.036$, respectively). The double wild type haplotype showed a significant SW to NE cline ($R^2 = 0.42$, $p = 0.009$).

Traces of selection signatures in the *MTHFR* gene

To check for recent selective sweeps, we explored the nucleotide diversity, divergence levels, distribution of rare alleles, Tajima's D statistic, and degree of linkage disequilibrium of CEU, FIN, IBS, CHB, and YRI (see Supplementary *Table 2*) for the *MTHFR* gene using the PopHuman browser. No reduction in genetic diversity (Pi, and Theta) or haplotype diversity was observed. Tajima D showed negative values (suggesting rare derived alleles) only in IBS, CHB, and YRI. Although NI values ranged between 0.22 and 0.5, and

DoS values were higher than 0 (suggesting some signals of adaptive evolution), the non-significant values of Fisher's exact test and the negative values of Alpha-cor indicated no traces of positive selection for *MTHFR*.

With regard to the iHS and nSL values estimated with Selscan in Finns, Iberians, and Chinese populations, none of the samples showed values lower than -2. The Yoruban sample showed iHS values lower than -2 for the rs6419480 (iHS -2.50) and rs4846035 (iHS -2.42). These two SNPs are intron variants with no reported clinical significance.

DISCUSSION

In this article, we have provided new data on the A1298C and C677T SNPs of the *MTHFR* gene in two populations at both extremities of Eurasia, a sample from Catalonia in North-East Spain as representative

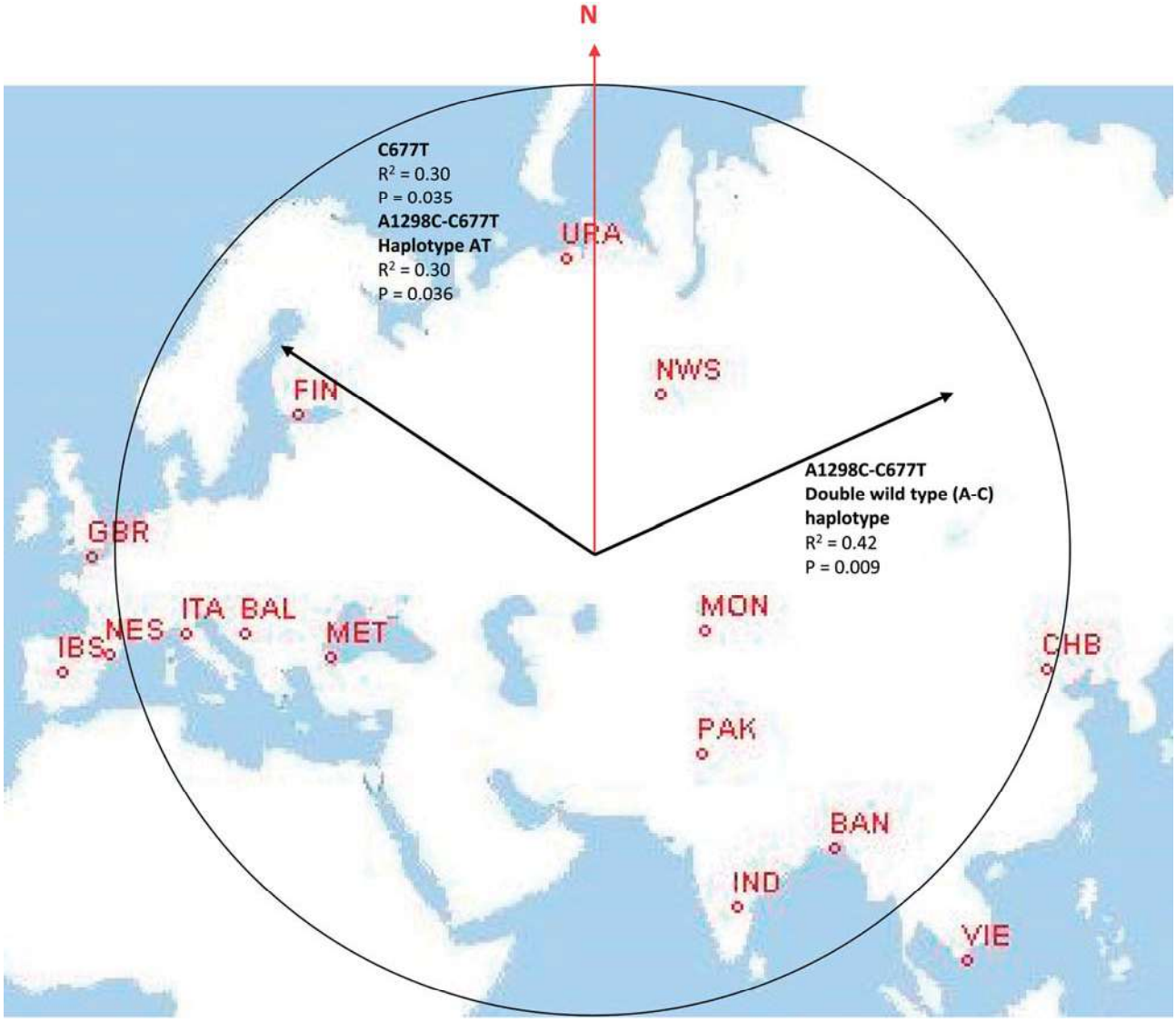

FIGURE 2: Increasing gradients of allele frequencies for C1298A, C677T and the AT and double wild type haplotype for 15 Eurasian populations. Black arrows departing from the circumference center indicate the orientation of the cline. The significant clines are expressed with the coefficient of determination of the linear regression (R^2) and its statistical significance. Population abbreviations as in *Figure 1*.

of urban and intermixed South European populations, and the Khanty population from North-West Siberia. Additionally, we downloaded and analysed the data regarding these two polymorphisms in 18 world populations to evaluate the genetic structure of these two SNPs and their haplotypes, particularly in Eurasia.

Polymorphisms A1298C and C677T are of high biomedical significance because they reduce the enzyme activity by approximately 30% (1298C) and 70% (677T). Depending on the combined frequency of both alleles, folate concentration among individuals and/or populations may differ significantly, putting them at greater risk of several diseases (Van Der Put *et al.* 1998).

A large body of evidence underlines the importance of having adequate folate levels at all stages of human ontogeny (McKinley *et al.* 2001). Maintaining adequate levels of folate during pregnancy avoids neural tube defects (NTDs) and other pregnancy complications, and ensures correct brain development. After birth, folate requirements increase at the same rate as metabolism during childhood to adolescence. In older adults, adequate levels of folate and other related B vitamins maintain healthy nervous system function, and reduction of cardiovascular diseases, dementia, depression, and other neurocognitive pathologies. Lastly and most importantly, folate plays a crucial role in improving the quality of life among the elderly (Kehoe *et al.* 2019).

With regard to the *T* variant of C677T (associated with the greatest loss of enzyme activity), our NE Spanish sample showed a value (39.4%) close to the highest frequencies reported in South Europe, whereas NW Siberians exhibited a value (17.5%) similar to those reported in other Siberian groups (Trifonova *et al.* 2012). We have found that this polymorphism shows a marked genetic structure in Eurasia, where 5.3% of its genetic variation is evident from the comparison of European, North Eurasian and Asian groups (see supplementary *Table 2*). The internal genetic variation of populations within these groups is also noteworthy (4.1%).

In contrast, the *C* allele of A1298C, the one associated with a lower drop in loss of enzyme activity, showed much less population variation. Our two samples had similar values (29% in NE Spain, 31.3% in NW Siberia), similar to other Eurasian populations. In fact, this polymorphism showed no population differentiation in Eurasia.

The haplotype carrying the two wild type alleles, and hence, without any attributable reduction of enzyme activity, shows the highest values in Africans. In Eurasians, except in South Europeans, it is also the most frequent haplotype. In African samples, the percentage of homozygote individuals for the two wild type alleles ranged from 55% in Kenya to 68% in Gambia. In Eurasia, this percentage ranged from 5% in Italy to 27% in Mongolian Turkic speakers. NE Spaniards showed 11% homozygotes whereas NW Siberians showed 26%, and was the second population with the highest value.

Seen in another way, and paying special attention to the number of people carrying haplotypes associated with a notable loss of enzyme activity, all African samples represent less than 1% of the population. NW Siberians, like most Asian samples, except China, exhibit a remarkably low percentage (3%). NE Spanish samples with 16% are in line with the values observed in southern Europeans, without reaching the high value as observed in Italy (22%).

We have found significant South to North clines for both SNPs and haplotypes for the 18 populations (see *Figure 1).* When only Eurasians were included in the analyses, C1298C failed to show a significant cline, whereas C677T showed a Southeast to Northwest cline (see *Figure 2*). The double wild type haplotype cline runs in the opposite direction. Previous works have already suggested a geographic North-South gradient for the C677T polymorphism (Wilcken *et al.* 2003, Mansoor *et al.* 2009) but to our knowledge, we are the first to estimate clines for both SNPs and their haplotypes. These clines may respond to selective pressures related with habitats having differential sun exposures, skin pigmentation, ultraviolet radiation, and availability of folate-rich foods (Jablonski, Chaplin 2000, 1000 Genomes Project Consortium *et al.* 2015). We failed to find recent selective signatures in the *MTHFR* gene, but our results may be limited by the type and availability of data. We are aware that the analysed data, downloaded from population repositories, are not designed for a thorough analysis of the *MHTFR* gene.

Based on our results, we have drawn some conclusions about the possible clinical relevance of *MTHFR* allele frequencies in NE Spaniards and NW Siberians.

It is evident that the both studied polymorphisms have biomedical significance. The connection between these polymorphisms and chronic human diseases was suspected earlier, as they are associated with hyperhomocysteinaemia and toxicity. Our NE Spanish sample showed 44% individuals with 30% of reduced enzyme activity and, more importantly, 16% people with 70% reduced enzyme activity. This fact, that put these individuals at a higher risk for a large number of pathologies, can be compensated by the Mediterranean

diet, which is rich in folate, and by practices such as periconceptional folic acid supplementation.

In Spain, folic acid fortification of food is not mandatory, and hence, we suggest vigilance in compliance with folic acid supplementation when prescribed, particularly because a relevant percentage of the Spanish population (Samaniego-Vaesken *et al.* 2017) do not meet the recommended intake of folate-rich foods. Great care must be devoted to the Spanish elderly, in particular, to residents of long-term care homes. In this group, with a growing tendency in Spain, low intake of folic acid can reach 80%, with malnutrition in 56%, and sarcopenia in 63% (Rodríguez-Rejón *et al.* 2019).

Our data on the Khanty population revealed a low frequency of risk alleles as well as haplotypes (3% of people with combinations of 70% reduced activity) suggesting that this population retained the natural adaptation to minimize the risk for pathologies associated with folate levels at all stages of human ontogeny. Despite living in a harsh, cold climate with limited sunlight during most of the year, the traditional diet of Khanty people provides an example of optimal adaptation to prevent nutritional deficiencies, especially of folate and vitamin D. River fishing, hunting and gathering, and reindeer herding are their leading occupations. Reindeer, along with fish, have traditionally provided a basic source of food. Traditionally, all meat is boiled, dried, or smoked, whereas fish is mainly eaten raw (Vorobeva *et al.* 2015). The kidney, liver, marrow, and eyes of animals are also eaten uncooked. Animal blood is either drunk fresh or mixed with other food. Flesh is dried and powdered for eating during hunting. In the forest, the Khanty people gather wild onions, cedar nuts, berries, and bird-cherries, as well as other herbs (Vorobeva *et al.* 2015).

The favourable *MTHFR* genetic profile of the Khanty combined with their nutritionally rich traditional diet, protects pregnancies and offspring health in addition to decreasing susceptibility to chronic diseases. Given the negative effect of nutritional transition in similar populations, as some Inuit groups from Canada, it is crucial for the Khanty to maintain their traditional diet (Schaefer *et al.* 2011).

ACKNOWLEDGEMENTS

We thank all the respondents that participated in this study, both from Russia and Spain. We also thank state assignment for the Institute of General Genetics, RAS, Moscow (PB, IU); the work of OEL was conducted under the IDB RAS Government basic research program in 2020, № 0108-2019-0007. This research was conducted in line with the state assignment (01201370995) for the Institute of Ethnology and Anthropology, RAS, Moscow (MB, JF) and by the Comissionat per a Universitats i Recerca of the Generalitat 2017SGR1577 and 2017SGR1630.

REFERENCES

1000 Genomes Project Consortium, AUTON A., BROOKS L. D., DURBIN R. M., GARRISON E. P., KANG H. M., KORBEL J. O., MARCHINI J. L., MCCARTHY S., MCVEAN G. A., ABECASIS G. R., 2015: A global reference for human genetic variation. *Nature* 526: 68–74. https://doi.org/10.1038/nature15393

BROWN A. S., 2011: The environment and susceptibility to schizophrenia. *Progress in Neurobiology* 93: 23–58. https://dx.doi.org/10.1016%2Fj.pneurobio.2010.09.003

BUTOVSKAYA M., SOROKOWSKA A., KARWOWSKI M., SABINIEWICZ A., FEDENOK J., DRONOVA D., NEGASHEVA M., SELIVANOVA E., SOROKOWSKI P., 2017: Waist-to-hip ratio, body-mass index, age and number of children in seven traditional societies. *Scientific Reports* 7: 1622. https://doi.org/10.1038/s41598-017-01916-9

BUTOVSKAYA P. R., LAZEBNY O. E., SUKHODOLSKAYA E. M., VASILIEV V. A., DRONOVA D. A., FEDENOK J. N., ROSA A., PELETSKAYA E. N., RYSKOV A. P., BUTOVSKAYA M. L., 2016: Polymorphisms of two loci at the oxytocin receptor gene in populations of Africa, Asia and south Europe. *BMC Genetics* 17: 17. https://doi.org/10.1186/s12863-015-0323-8

CASILLAS S., MULET R., VILLEGAS-MIRÓN P., HERVAS S., SANZ E., VELASCO D., BERTRANPETIT J., LAAYOUNI H., BARBADILLA A., 2018: PopHuman: The human population genomics browser. *Nucleic Acids Research* 46: D1003–D1010. https://doi.org/10.1093/nar/gkx943

CHEN Z. F., YOUNG L., YU C. H., SHIAO S. P. K., 2018: A meta-prediction of methylenetetrahydrofolate-reductase polymorphisms and air pollution increased the risk of ischemic heart diseases worldwide. *International Journal of Environmental Research and Public Health* 15: 1453. https://dx.doi.org/10.3390%2Fijerph15071453

DEN HEIJER M., LEWINGTON S., CLARKE R., 2005: Homocysteine, MTHFR and risk of venous thrombosis: A meta-analysis of published epidemiological studies. *Journal of Thrombosis and Haemostasis* 3: 292–299. https://doi.org/10.1111/j.1538-7836.2005.01141.x

EXCOFFIER L., LISCHER H. E. L., 2010: Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

FERRER-ADMETLLA A., LIANG M., KORNELIUSSEN T., NIELSEN R., 2014: On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular Biology and Evolution* 31: 1275–1291. https://doi.org/10.1093/molbev/msu077

FROSST P., BLOM H. J., MILOS R., GOYETTE P., SHEPPARD C. A., MATTHEWS R. G., BOERS G. J. H., DEN HEIJER M., KLUIJTMANS L. A. J., VAN DEN HEUVE L. P., ROVEN R., 1995: A candidate genetic risk factor for vascular disease: A common mutation in methylenetetrahydrofolate reductase. *Nature Genetics* 10: 111–113. https://doi.org/10.1038/ng0595-111

GOYETTE P., SUMNER J. S., MILOS R., DUNCAN A. M. V., ROSENBLATT D. S., MATTHEWS R. G., ROZEN R., 1994: Human methylenetetrahydrofolate reductase: Isolation of cDNA, mapping and mutation identification. *Nature Genetics* 7: 195–200. https://doi.org/10.1038/ng0694-195

HANCOCK A. M., ALKORTA-ARANBURU G., WITONSKY D. B., DI RIENZO A., 2010: Adaptations to new environments in humans: The role of subtle allele frequency shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 2459–2468. https://doi.org/10.1098/rstb.2010.0032

HOLMES M. V., NEWCOMBE P., HUBACEK J. A., SOFAT R., RICKETTS S. L., COOPER J., BRETELER M. M., BAUTISTA L. E., SHARMA P., WHITTAKER J. C., *et al.*, 2011: Effect modification by population dietary folate on the association between MTHFR genotype, homocysteine, and stroke risk: A meta-analysis of genetic studies and randomised trials. *Lancet* 378: 584–594. https://doi.org/10.1016/s0140-6736(11)60872-6

JABLONSKI N. G., CHAPLIN G., 2000: The evolution of human skin coloration. *Journal of Human Evolution* 39: 57–106. https://doi.org/10.1006/jhev.2000.0403

JONES S., MARTIN R., PILBEAM D., 1992: The encyclopedia of Human Evolution. Cambridge University Press, Melbourne.

JONES P., BECKETT E., YATES Z., VEYSEY M., LUCOCK M., 2016: Converging evolutionary, environmental and clinical ideas on folate metabolism. *Exploratory Research and Hypothesis in Medicine* 1: 34–41. https://dx.doi.org/10.14218/ERHM.2016.00003b

KATAYAMA Y., HARAOKA J., HIRABAYASHI H., KAWAMATA T., KAWAMOTO K., KITAHARA T., KOJIMA J., KUROIWA T., MORI T., MORO N., NAGATA I., OGAWA A., OHNO K., SEIKI Y., SHIOKAWA Y., Akira TERAMOTO A., TOMINAGA T., YOSHIMINE T., 2007: A randomized controlled trial of hydrocortisone against hyponatremia in patients with aneurysmal subarachnoid hemorrhage. *Stroke* 38: 2373–2375. https://doi.org/10.1161/strokeaha.106.480038

KEHOE L., WALTON J., FLYNN A., 2019: Nutritional challenges for older adults in Europe: Current status and future directions. *Proceedings of the Nutrition Society* 78: 221–233. https://doi.org/10.1017/s0029665118002744

MANSOOR A., MAZHAR K., ALI L., MUAZZAM A. G., SIDDIQI S., USMAN S., 2009: Prevalence of the C677T single-nucleotide polymorphism in the methylenetetrahydrofolate reductase gene among Pakistani ethnic groups. *Genetic Testing and Molecular Biomarkers* 13: 521–526. https://doi.org/10.1089/gtmb.2009.0012

MCKINLEY M. C., MCNULTY H., MCPARTLIN J., STRAIN J. J., PENTIEVA K., WARD M., WEIR D. G., SCOTT J. M., 2001: Low-dose vitamin B-6 effectively lowers fasting plasma homocysteine in healthy elderly persons who are folate and riboflavin replete. *The American Journal of Clinical Nutrition* 73: 759–764. https://doi.org/10.1093/ajcn/73.4.759

MCLAREN W., GIL L., HUNT S. E., RIAT H. S., RITCHIE G. R. S., THORMANN A., FLICEK P., CUNNINGHAM F., 2016: The Ensembl variant effect predictor. *Genome Biology* 17: 122. https://doi.org/10.1186/s13059-016-0974-4

NEFIC H., MACKIC-DJUROVIC M., EMINOVIC I., 2018: The frequency of the 677C>T and 1298A>C polymorphisms in the methylenetetrahydrofolate reductase (MTHFR) gene in the population. *Medical archives (Sarajevo, Bosnia and Herzegovina)* 72: 164–169. https://dx.doi.org/10.5455%2Fmedarh.2018.72.164-169

PIMENOFF V. N., COMAS D., PALO J. U., VERSHUBSKY G., KOZLOV A., SAJANTILA A., 2008: Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *European Journal of Human Genetics* 16: 1254–1264. https://doi.org/10.1038/ejhg.2008.101

RODRÍGUEZ-REJÓN A. I., RUIZ-LÓPEZ M. D., ARTACHO MARTÍN-LAGOS R., 2019: Diagnosis and prevalence of sarcopenia in long-term care homes: EWGSOP2 versus EWGSOP1. *Nutrición Hospitalaria* 36: 1074–1080. http://dx.doi.org/10.20960/nh.02573

SAMANIEGO-VAESKEN M. DE L., PARTEARROYO T., OLZA J., ARANCETA-BARTRINA J., GIL Á., GONZÁLEZ-GROSS M., ORTEGA R. M., SERRA-MAJEM L., VARELA-MOREIRAS G., 2017: Iron intake and dietary sources in the Spanish population: Findings from the ANIBES study. *Nutrients* 9: 1–14. https://dx.doi.org/10.3390%2Fnu9030203

SCHAEFER S. E., ERBER E., TRZASKOS J. P., ROACHE C., OSBORNE G., SHARMA S., 2011: Sources of food affect dietary adequacy of Inuit women of childbearing age in Arctic Canada. *Journal of Health, Population and Nutrition* 29: 454–464. https://dx.doi.org/10.3329%2Fjhpn.v29i5.8899

SHIAO S. P. K., LIE A., YU C. H., 2018: Meta-analysis of homocysteine-related factors on the risk of colorectal cancer. *Oncotarget* 9: 25681–25697. https://dx.doi.org/10.18632%2Foncotarget.25355

SZPIECH Z. A., HERNANDEZ R. D., 2014: Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution* 31: 2824–2827. https://doi.org/10.1093/molbev/msu211

TINELLI C., DI PINO A., FICULLE E., MARCELLI S., FELIGIONI M., 2019: Hyperhomocysteinemia as a risk factor and potential nutraceutical target for certain pathologies. *Frontiers in Nutrition* 6: 49. https://dx.doi.org/10.3389%2Ffnut.2019.00049

TRIFONOVA E. A., EREMINA E. R., URNOV F. D., STEPANOV V. A., 2012: The genetic diversity and structure of linkage disequilibrium of the MTHFR gene in populations of northern Eurasia. *Acta Naturae* 4: 53–69.

VAN DER PUT N. M., GABREËLS. F., STEVENS E. M., SMEITINK J. A., TRIJBELS F. J., ESKES T. K., VAN DEN HEUVEL L. P., BLOM H. J., 1998: A second common mutation in the methylenetetrahydrofolate reductase gene: An additional risk factor for neural-tube defects? *American Journal of Human Genetics* 62: 1044–1051. https://doi.org/10.1086/301825

VOROBEVA V., FEDORINOVA Z., KOLESNIK E., 2015: Three crucial crises in the development of the Khanty and Mansi unique culture. *Procedia - Social and Behavioral Sciences* 206: 108–113. https://doi.org/10.1016/j.sbspro.2015.10.037

WEISBERG I., TRAN P., CHRISTENSEN B., SIBANI S., ROZEN R., 1998: A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity. *Molecular Genetics and Metabolism* 64: 169–172. https://doi.org/10.1006/mgme.1998.2714

WILCKEN B., BAMFORTH F., LI Z., ZHU H., RITVANEN A., RENLUND M., STOLL C., ALEMBIK Y., DOTT B., CZEIZEL A. E., *et al.*, 2003: Geographical and ethnic variation of the 677C>T allele of 5,10 methylenetetrahydrofolate reductase (MTHFR): Findings from over 7000 newborns from 16 areas world wide. *Journal of Medical Genetics* 40: 619–625. https://doi.org/10.1136/jmg.40.8.619

Polina R. Butovskaya[1]
M. Esther Esteban[2,8]
Nikos Avramidis[2]
Oleg E. Lazebny[3]
Julia N. Fedenok[4]
Irina G. Udina[5]
Marina L. Butovskaya[4, *]
Araceli Rosa[2, 6, 7, *]

*Corresponding author.

[1] Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia
E-mail: butovskaya@hotmail.com.
[2] Secció de Zoologia i Antropologia Biològica, Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia, Universitat de Barcelona (UB), Barcelona, Spain
E-mails: mesteban@ub.edu; nabramidis@gmail.com; araceli.rosa@ub.edu
[3] Koltzov Institute of Developmental Biology, Russian Academy of Sciences, Moscow, Russia.
E-mail: oelazebny@gmail.com
[4] Institute of ethnology and anthropology, Russian Academy of Sciences, Moscow, Russia.
E-mails: fedenok.julia@gmail.com; marina.butovskaya@gmail.com
[5] Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow. Russia.
E-mail: irina_udina@mail.ru
[6] Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Spain
[7] Centre for Biomedical Research Network on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Barcelona, Spain
[8] Grup de Recerca en Antropologia Biològica (GREAB), Department of Evolutionary Biology, Ecology and Environmental Sciences, Institut de Recerca de la Biodiversitat (IRBio), University of Barcelona, Spain

SUPPLEMENTARY TABLE 1: Observed and Expected Heterozygosity values and Hardy-Weinberg P-values for rs1801131 and rs1801133.

		EUROPEANS							
		NE Spain	**Iberian P**	**Italy**	**Balkans**	**Great Britain**	**Finland**	**Uralic speakers**	**NW Siberia**
rs1801131	Observed Het.	0.390	0.430	0.402	0.371	0.417	0.454	0.363	0.532
	Expected Het.	0.412	0.397	0.432	0.370	0.452	0.436	0.379	0.431
	P-Value	0.207	0.469	0.500	1.000	0.490	0.816	1.000	0.002
rs1801133	Observed Het.	0.496	0.551	0.4490	0.500	0.472	0.424	0.409	0.292
	Expected Het.	0.478	0.496	0.500	0.499	0.440	0.399	0.422	0.290
	P-Value	0.355	0.325	0.332	1.000	0.633	0.615	1.000	1.000
		ASIANS							
		ME Turkish speakers	**Mongolian Turkic speakers**	**China**	**Pakistan**	**India**	**Bangladesh**	**Vietnam**	
rs1801131	Observed Het.	0.449	0.544	0.369	0.479	0.520	0.500	0.454	
	Expected Het.	0.459	0.446	0.348	0.489	0.500	0.486	0.384	
	P-Values	0.869	0.097	0.776	1.000	0.841	0.827	0.111	
rs1801133	Observed Het.	0.358	0.265	0.485	0.271	0.167	0.244	0.303	
	Expected Het.	0.373	0.253	0.500	0.235	0.186	0.216	0.312	
	P-Values	0.686	1.000	0.844	0.205	0.275	0.351	0.750	
		AFRICANS							
		Gambia	**Sierra Leone**	**Esan Nigeria**		**Yoruba Nigeria**		**Kenya**	
rs1801131	Observed Het.	0.159	0.247	0.232		0.241		0.273	
	Expected Het.	0.204	0.235	0.237		0.213		0.305	
	P-Values	0.038	1.000	1.000		0.355		0.320	
rs1801133	Observed Het.	0.124	0.153	0.162		0.213		0.141	
	Expected Het.	0.117	0.142	0.149		0.191		0.132	
	P-Values	1.000	1.000	1.000		0.603		1.000	

SUPPLEMENTARY TABLE 2: Allelic and haplotypic AMOVA analyses of A1298C and C677T *MTHFR* polymorphisms and their haplotypes. Population groups are indicated with population abbreviations as in *Figure 1*.

	Non-hierarchical AMOVA (F_{ST}, probability) for different population groups					
	Africa	**Europe**	**North Eurasia**	**Asia**	**Europe plus North Eurasia**	**Asia plus North Eurasia**
	GAM, SLE, YOR, ESN, KEN	**IBS, NES, ITA, BAL, FIN, GBR**	**URA, NWS**	**MET, MON, PAK, IND, BAN, VIE, CHB**	**IBS, NES, ITA, BAL, FIN, GBR, URA, NWS**	**MET, MON, PAK, IND, BAN, VIE, CHB, URA, NWS**
A1298C	0.0017, 0.494	0.0001, 0.757	0.0036, 0.263	0.0285, < 0.001	-0.0017, 0.868	0.0238, < 0.001
C677T	-0.0004, 0.777	0.0142, < 0.001	0.0369, 0.016	0.0882, < 0.001	0.0420, < 0.001	0.0695, < 0.001
Haplotypes	-0.0040, 0.757	0.0060, 0.045	0.0031, 0.306	0.0435, < 0.001	0.0215, < 0.001	0.0288, < 0.001
	Hierarchical AMOVA for different population clusters					
	Among groups (F_{CT}, probability) / Among populations within groups (F_{SC} probability)					
	Africa - the rest	**Europe vs North Eurasia**	**Asia vs North Eurasia**	**Europe vs Asia**	**Europe vs North Eurasia vs Asia**	
A1298C	0.0715, < 0.001 / 0.0127, < 0.001	-0.0015, 0.792 / 0.0004, 0.353	-0.0065, 0.556 / 0.0264, < 0.001	0.0034, 0.150 / 0.0141, < 0.001	0.0006, 0.259 / 0.0136, < 0.001	
C677T	0.1126, < 0.001 / 0.0709, < 0.001	0.0673, 0.085 / 0.0160, < 0.001	-0.0367,1.000 / 0.0828. < 0.001	0.0620, 0.016 / 0.0411, < 0.001	0.0533, 0.014 / 0.0415, < 0.001	
Haplotypes	0.1654, < 0.001 / 0.0319, < 0.001	0.0413, 0.039 / 0.0058, 0.051	-0.0116, 0.794 / 0.0333, < 0.001	0.0289, 0.015 / 0.0198, < 0.001	0.0260, 0.016 / 0.0191, < 0.001	

SUPPLEMENTARY TABLE 3: Statistics to detect traces of positive selection signatures in the *MTHFR* gene. Pi: Nucleotide diversity; Theta: Nucleotide polymorphism; Tajima D: Tajima's D test statistic; Hap Diversity within: Haplotype diversity within the population; NI: Neutrality index; DoS: Direction of Selection; Fisher 2: Fisher exact test p-value for the McDonald and Kreitman test (McDonald, Kreitman 1991); Alpha-cor: Fraction of new mutations that are adaptive.

	Pi	**Theta**	**Tajima D**	**Hap Diversity within**	**NI**	**DoS**	**Fisher 2**	**Alpha-cor**
CEU	0.000869	0.000860	0.092	0.8121	0.2424	0.1894	0.574	-0.485
CHB	0.001593	0.000696	-0.256	0.7245	0.2727	0.0167	0.530	-0.864
FIN	0.000822	0.000819	0.073	0.7640	0.4091	0.0985	0.447	-1.455
IBS	0.000778	0.001044	-0.687	0.7653	0.2273	0.2024	1.000	-0.352
YRI	0.001096	0.001347	-0.470	0.9574	0.5000	0.0705	0.349	-2.500