March 2021

The Distribution in Native Populations from Mexico and Central America of the C677T Variant in the MTHFR Gene

Lucio A. Reyes
University of South Florida

The Distribution in Native Populations from Mexico and Central America of the C677T Variant in the MTHFR Gene

by

Lucio A. Reyes

A thesis submitted in partial fulfillment
of the requirements for the degree of
Masters of Arts in Applied Anthropology
Department of Anthropology
College of Arts and Sciences
University of South Florida

Major Professor: Lorena Madrigal, Ph.D.
Elizabeth Miller, Ph.D.
David Himmelgreen, Ph.D.

Date of Approval:
March 11, 2021

Keywords: Folate, Evolution, Mutation, Persistence, Heterozygote Advantage

DEDICATION

This paper is dedicated to the Native Peoples of Mexico and Central America who inspired this project. This paper also recognizes the individuals who made the completion of this project possible: David Godfrey for data curation; project administration; software; visualization. Francisco Gonzalez for data curation; methodology; software; writing-original draft. Caroline MacLean for investigation; writing-review and editing. Li-June Ming for methodology; writing-review and editing. Lorena Madrigal for conceptualization; data curation; formal analysis; investigation; methodology; project administration; software; supervision; validation; visualization; writing-original draft; writing-review and editing.

ACKNOWLEDGMENTS

We appreciate the careful reviews this paper received from the American Journal of Human Biology. We appreciate the support from the department of Interlibrary Loan Service of the University of South Florida, which always gets any paper we need.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Objectives: To explore evolutionary hypotheses for the high frequencies of a substitution in the methylenetetrahydrofolate reductase (MTHFR) gene, in Mexican and Central American Indigenous populations.

Materials and methods: We obtained allele frequencies for the C677T variant in the MTHFR gene and ecological information for 37 indigenous samples from Mexico and Central America. We calculated Hardy–Weinberg equilibrium and computed Fst statistics. We computed correlations between the samples' allele frequencies and ecological and geochemical variables.

Results: Many of the samples have extremely high frequencies of the T allele ($\bar{q} = 0.62$, median = 0.66). In this region, the frequency of the T allele decreases from Southeast to Northwest and is significantly correlated with longitude, latitude, altitude, and insolation.

Conclusions: The native people of Central America and Mexico evolved high frequencies of an allele which has been shown to produce deleterious clinical effects including neural tube effects, cardiovascular events, and cancer. This allele has a clinal distribution in the region, perhaps associated with solar irradiation. As (Contreras-Cubas et al., 2016) noted, the traditional diet of these populations, which is high in folate, has likely mitigated the negative effect of the allele. It is of primary importance that their rights to their homeland and traditional diets be respected. It is a matter of Public Health to investigate whether this allele is a factor in the current wave of cardiovascular diseases affecting the majority population of this region, since it descends from

the Native peoples and the Mediterranean population, which also has high frequencies of the allele.

1 | INTRODUCTION

Human biologists and biological anthropologists interested in microevolutionary change in Homo sapiens will frequently frame their research within the premise that most human genetic variation is shared (Ramachandran et al., 2010). Therefore, researchers seek to explain the evolutionary reasons for population differences, where natural selection and genetic drift are the evolutionary forces which lead populations to diverge (Jobling et al., 2014).

It is within this evolutionary background that previous researchers have noted the highly polymorphic level of the methylenetetrahydrofolate reductase gene (MTHFR) and have asked if natural selection is responsible for the polymorphisms found in this gene (Le Marchand et al., 2004; Reyes-Engel et al., 2002; Wang et al., 2016; Yadav et al., 2017). A frequently studied variant in MTHFR is the C677T, a single nucleotide substitution in the 677th position where thymine takes the place of cytosine. The earlier studies made it clear that the T allele was highly frequent in Northern Asia and in Europe (particularly in the Mediterranean region) and that the T allele was virtually absent in the Southern hemispheres (Wilcken et al., 2003). More recently, researchers in Mexico have noted that the highest frequency of the T allele is found in Native populations of Mexico rather than in any population in Eurasia (Antonio-Vejar et al., 2014; Contreras-Cubas et al., 2016; Juarez-Velazquez et al., 2010).

If individuals inherit variants at the MTHFR gene, which cause the gene to produce a less-efficient enzyme, they may experience hyperhomocysteinemia, an unusually high level of homocysteine. Elevated homocysteine may be associated with negative clinical effects, particularly in low-folate environments (Schneede et al., 2000).

Studies demonstrate a decrease in the frequency of TT individuals in older age groups. These data suggest that TT individuals have higher mortality than do CC individuals and perhaps CT individuals. This diminished frequency of TT individuals was demonstrated in three folate-rich settings, namely Sicily, Burkina Faso, and Sardinia. Authors conclude that "…the reproducible trend that showed the prevalence of 677T in the young subjects of the three studied areas confirms the disadvantage of this polymorphism with the age" (Chillemi et al., 2005). In another study in India, where the sample was assessed to be deficient in folate by the researchers, the highest frequency of the 677T SNP was found in the 20–40 age category, after which the SNP frequency decreased from 40–60 to 60 years, in a highly significant trend (Naushad et al., 2014). Finally, a study which took place in Japan reported a decreased frequency of homozygous individuals from the <55 years to the 55–79 year and up to ≥80 years of age category. This reduction in frequency was significant and leads the authors to conclude that the T mutation was "a genetic factor that prevents the attainment of old age" (Matsushita et al., 1997). Matshushta et al. did not note their participants' folate status.

The T allele is associated with decreased fertility in CT and TT males (Gupta et al., 2013; Irfan et al., 2016; Kurzawski et al., 2015; Zhang et al., 2015; Zhu et al., 2016). Females with both TT and CT phenotypes frequently seek in vitro fertilization due to infertility problems and following birth defects, particularly neural tube defects (NTD) (Wang et al., 2015). As Imbard et al. (2013 and citations therein) note, there are several meta-analyses showing that mothers and even fathers with the CT or TT genotypes have a higher probability of producing NTD babies (Imbard et al., 2013). Studies have shown that naturally aborted conceptus have a significantly high frequency of the T variant alone or in conjunction with another MTHFR SNP (Zetterberg et al., 2002).

Importantly, whether people experience hyperhomocysteinemia depends in great part on the amount of dietary folate available to them. The level of dietary folate is contingent upon the traditional diet folate levels and whether the diet includes supplemented folic acid. Another environmental factor which impacts folate availability is solar irradiation, which depletes folate (Jablonski & Chaplin, 2010).

National campaigns to supplement basic foods with folic acid and to promote prenatal vitamins started in various countries in the late 1900s. However, not all women of reproductive age were equally successfully reached by the supplementation campaign (Crider et al., 2011). A reason why this may be the case is ethnicity: in countries such as Mexico and Costa Rica, the Indigenous peoples may not benefit from folic acid supplementation as much as the general population does. At the same time, once the Indigenous people enter the cash economy, they may give up their folate-rich traditional foods (Contreras-Cubas et al., 2016).

Another environmental factor which affects humans' folate availability is the photolysis of folate by solar radiation, which is mediated by skin color. According to Nina Jablonski, the distribution of dark skin in humans is best explained by the fact that more melanized skin prevents depletion of folate when humans inhabit areas of high solar radiation (Jablonski & Chaplin, 2000; Jones et al., 2018). Therefore, two environmental variables should be considered when studying cross-cultural access to (or loss of) folate: whether the individual has access to dietary folate and whether the individual's skin color protects them from UV photolysis of folate.

Before us, several researchers attempted to explain why the 677T allele has achieved high frequencies in some human populations. We summarize these proposals here:

1. Heterozygote advantage of CT individuals who have a cognitive advantage in old age in China (Tsai et al., 2011). This association was not replicated in a study with Scottish elderly patients (Schiepers et al., 2011)

2. Homozygote advantage of the TT fetus of pregnant mothers who receive folate supplementation (Mayor-Olea et al., 2008). Notably, Servy et al. offer data indicating that TT mothers who take folate supplements may develop un-metabolized folic acid syndrome (UMFA) which has negative effects for both mothers and fetuses (Servy et al., 2018).

3. An environmental-genetic interaction, which drove the T allele to high frequencies (Lucock et al., 2008, 2012, 2018; Lucock & Yates, 2006; Yates & Lucock, 2008).

4. Natural selection operating on human populations via UV-B rather than UV-A radiation (Wang et al., 2012).

5. Heterozygote and/or homozygote advantage in processing heavy metals. Specifically, heterozygote advantage in surviving platinum-based chemotherapy (Cui et al., 2011; Krawczyk et al., 2014; Li et al., 2014).

6. Decreased risk of TT individuals of suffering acute lymphoid leukemia (ALL), likely in the presence of adequate folate consumption (Robien & Ulrich, 2003).

7. A normal metabolism despite a thermolabile enzyme (Ojeda-Granados et al., 2017) in a folate-rich environment.

In this paper, we present data obtained from the literature, about the distribution of the T allele in indigenous populations from Mexico and Central America and its ecological correlates. We take

a population-genetics perspective to evaluate the distribution of the allele across the region. We discuss the public health implications of the high frequency of the allele for the region.

2 | MATERIALS AND METHODS

2.1 | Samples

We collected data from 37 samples from published literature after an exhaustive scan via Web of Science and PubMed. Our samples consist of 31 indigenous groups from Mexico and Guatemala and six from Costa Rica. The groups that originated in Guatemala are Mayan groups that were forced to flee to Mexico due to the political repression in their home territories in the 1980s. We show in Table S1 the samples sorted by country, by allele frequency and by their population names (where we used the name as it appeared in the publication).

2.1.1 | Environmental information

Our anthropological GIS expert (DG) analyzed the information given in the original paper reporting the genetic data to determine the most appropriate coordinates for the sample. With these coordinates, we obtained: 1. Latitude. 2. Longitude. 3. Elevation. 4. Geochemical data from the earthchem.org website as parts per volume of the following toxic metals: chromium (Cr), manganese (Mn), cobalt (Co), nickel (Ni), and lead (Pb). 5. A description of the ecosystem where the population evolved by the IUCN (International Union of Conservation of Nature). 6. Insolation, a measure of the incoming solar radiation that reaches the Earth's atmosphere and surface. Insolation stands for Incoming Solar Radiation (Ahrens, 2007; Mackay, 2015) and is a measure of incoming solar radiation in a particular area during a particular time. The advantage of measuring solar radiation with it rather than in UV-A or UV-B is that this measure of solar radiation takes into consideration a region's humidity's level and cloud coverage as a barrier to the solar radiation reaching the earth. This consideration is important in a study such as the

present one, in which some of the populations we considered evolved in desert, others evolved in tropical rainforests and others evolved in high altitude areas. These are ecological areas with different levels of humidity and exposure to solar irradiation. Insolation is usually expressed as kilowatt-hours per square meter per day (kWh/m2)(per day). In industrial settings, this equation may be simplified to kW/m2 (Husain et al., 2011).

2.2 | Methodology

2.2.1 | Analysis of population structure

If it was necessary to derive the allelic or the genotypic frequencies from the data provided by the original authors, we used the online Hardy–Weinberg calculator https://www.cog-genomics.org/software/stats. We only tested for departure of Hardy–Weinberg for sample sizes over 30. Sample sizes ranged from 17 to 220, with a mean of 67. The total number of individuals included in our analysis was 2512.

To evaluate the genetic structure of the samples in the region, we computed the three F statistics with the program GENEPOP (Raymond & Rousset, 1995; Rousset, 2008). We considered the review of Meirmans and Hedrick, who conclude that the F statistics are the best choice for this purpose in most cases (Meirmans & Hedrick, 2011). The F statistics program executed by GENEPOP is that proposed by Weir and Cockerman, in which sample sizes are standardized. This is particularly important in this study, in which sample sizes do vary (Weir & Cockerham, 1984). According to Hartl, an Fst value between 0.05 and 0.15 indicates moderate genetic differentiation among groups. One between 0.15 and 0.25 indicates great genetic differentiation, and one above 0.25 indicates very great genetic differentiation among groups (Hartl, 1988).

We computed a principal component (PC) analysis of the correlation matrix of the allele frequency and the ecological variables to reduce the variance of these variables into few uncorrelated multivariate indexes called principal components (PCs) (Manly, 2005).

2.2.2 | Analysis of ecological variables

We tested the normality of quantitative variables (latitude, longitude, elevation, and insolation) with the Shapiro–Wilk test. All statistical analyses were done with SAS 9.4 (SAS, 2013).

3 | RESULTS

3.1 | The distribution of the T allele in the region

We show the allele frequencies and the ecological variables in Table 1. In this region, the allele frequency ranges from 0.01 (Mazahuas) to a startling 0.92 (Trikís). In some samples (e.g., Bribrís, Cabecars, Trikís), there are more TT individuals than there are CC or CT individuals. To our knowledge, no other region in the world has this type of genotypic distribution (see Wang et al., 2016). The median allele frequency for the region is $M = 0.66$ while the mean is $\bar{q} = 0.62$. Only one population is not in equilibrium namely the Zapotecas ($X^2 = 8.93$, df = 2, p <.000). In this sample, we observe an excess of both homozygotes and a decrease in the number of observed heterozygotes in comparison with their expected numbers. For this sample, we observe 9 heterozygotes and expect 17. We show at the bottom of Table 1 the expected genotypic numbers of this group.

The three F statistics are as follows: F_{is}, a measure of within-population reduction in heterozygosity in comparison with the total population is $F_{is} = 0.008$. This value indicates that most groups have high within-group heterozygosity levels, confirming the Hardy–Weinberg results. F_{it}, a measure of individual reduction in heterozygosity regardless of sub-population membership and F_{st} the measure of population substructure due to group membership have virtually identical and moderate values at $F_{it} = 0.11$ and $F_{st} = 0.10$, respectively. The values of F_{it} and F_{st} indicate that most populations and most individuals in the different populations do not differ in their high levels of heterozygosity. In sum, there is little between-population

differentiation and among-individual differentiation. High levels of heterozygosity are found at the within-population level.

3.2 | The allele frequency in Mexico, Costa Rica, and its correlates

Of the quantitative variables, only longitude was normally distributed according to the Kolmogorov–Smirnoff test. As shown in Table S2, the descriptive statistics of latitude, longitude, elevation, and insolation, as well as these variables' normality tests. Allele frequencies were significantly negatively correlated with latitude ($r_s = -.52$, n = 37, p = .001), significantly negatively correlated with insolation ($r_s = -.40$, n = 37, p = .012), and positively significantly correlated with longitude ($r_s = .48$, n = 37, p = .0024). The allele frequency was not significantly correlated with elevation ($r_s = .16$, n = 36, p = .34). The negative correlation between allele frequency and latitude indicates that the populations with the higher allele frequency tend to be found toward the South. The positive correlation between allele frequency and longitude suggests that the populations with the higher frequencies are found toward the East (in the American continent the more Eastern the location, the closer to 0 the longitudinal values; while the more Western the location, the more negative the longitudinal values are). The distribution of the allele frequency itself and insolation is negative, as would be expected of an allele which produces an enzyme which interacts with folate and is depleted by UV radiation. The allele frequency was not correlated with any of the geochemical variables (data not shown).

We produced a principal component analysis to better understand the distribution of allele frequencies and ecology in our collection of samples. We obtained two PCs which explain 77% of the variation including five variables (latitude, longitude, insolation, elevation, and the allele frequency). The eigenvalues for the two eigenvectors are 2.77 and 1.11, respectively. The

eigenvalue of the third eigenvector was 0.62 (considerably under 1) which is why we did not interpret it.

We show a plot of the populations identified by their name and by their IUCN ecology notation along their two PC values in Figure 1. The first PC describes the geographical distribution of the allele, in which the populations with the lowest negative values of the first PC are the six Costa Rican samples and several of the Mayan groups such as the Chuj, Kanjobal, and Kaqchibel. Because these populations inhabit heavily forested areas, they do not receive as much direct sunlight as do populations who have a positive PC1 coefficient. Elevation does not play an important role in the first PC since its coefficient (0.13) is not as high as are the positive coefficients of latitude and insolation or the negative coefficients of allele frequency or longitude.

In contrast, the Seri, the Tarahumara, the Mayo, the Masahuas and the Yaquís are colored in blue because of their multivariate combination of variables causes them to have positive PC1 values. These populations have relatively lower frequencies of the T allele, inhabit Western areas, and live in northern latitudes. Because of their ecosystem, these populations receive more direct sunlight than do populations who have negative PC coefficients.

The first PC is : $PC1 = -0.4(\text{T freq.}) + 0.43(\text{Insolation}) + 0.21(\text{Elevation}) - 0.56(\text{Longitude}) + 0.54(\text{Latitude})$.

The second PC describes populations with high frequencies of the allele who inhabit regions with high values for insolation and altitude. These populations, (as opposed to those with high values in the first PC), are not sheltered from direct sunlight and inhabit areas of high elevation (and sun exposure not due to latitude, but due to elevation). In this PC, longitude does not

contribute at all and latitude hardly contributes with a coefficient of −0.10. In sum, PC2 is not about geography. It is about elevation and how elevation exposes populations to insolation.

The second PC is : PC2=0.49(Tfreq.) + 0.20(Insolation) + 0.84(Elevation) – 0.00 Longitude – 0.12 Latitude.

Among this array of populations, there are only two populations that could be considered to have a low frequency of the allele: the Mazahua (q = 0.01) and the Seri (q = 0.13) both of which inhabit a desert area. The next lowest allele frequencies are those of the Huave (p = .33), who inhabit an environment of dry broad leaves and the Tarahumara (p = .36), who inhabit a tropical conifer environment.

In Figure 2, we show a map of the region with the frequency of the allele superimposed. This figure confirms the Southeast to Northwest decline in the allele frequency.

Table 1 Summary of ecological variables for various populations.

Population	ID	Latitude	Longitude	Elevation	Country	Insolation	IUCN	CC	CT	TT	n	p	q
Bribrí	1	9.6255	-82.83	115.7	CostaRic	1700	Moist broad leaves	10	54	46	110	0.34	0.66
Cabecar	2	9.5191	-82.954	75.2	CostaRic	1800	Moist broad leaves	5	58	84	147	0.2297	0.7703
Chinanteco	3	17.3333	-95.967	690	Mexico	1643	Moist broad leaves	4	22	55	81	0.19	0.81
Oaxaca Chotal	4	16.5	-96.1	838	Mexico	2155	Dry broad leaves	10	16	18	44	0.49	0.59
Chorotega	5	10.0869	-85.421	334.5	CostaRic	2000	Moist broad leaves	18	74	41	133	0.41	0.59
Chuj	6	16.0817	-91.668	1519	Mexico	1863		1	4	12	17	0.29	0.71
Guatusos	7	10.6667	-84.817	59.1	CostaRic	1700	Moist broad leaves	1	14	14	29	0.28	0.72
Guaymí	8	8.82	-82.971	1000.6	CostaRic	1900	Moist broad leaves	9	38	110	157	0.18	0.82
Huastecos	9	21.9833	-99.017	80	Mexico	1863	Moist broad leaves	9	32	38	79	0.32	0.68
Huave	10	16.2	-94.983	3	Mexico	2227	Dry broad leaves	11	13	2	26	0.67	0.33
Huetar	11	9.8698	-84.237	1202.2	CostaRic	1900	Moist broad leaves	21	79	53	153	0.4	0.6
Huicholes	12	22.65	-103.58	1970.1	Mexico	2100	Deserts	8	28	14	50	0.44	0.56
Jakaltecos	13	15.6667	-91.733	660	Guatemal	2082	Tropical conifers	3	10	27	40	0.2	0.8
Kanjobal	14	15.6	-91.45	1750	Guatemal	1863	Tropical conifers	2	9	18	29	0.22	0.78
Kaqchibel	15	14.7358	-90.996	2277	Guatemal	2082	Tropical conifers	0	11	25	36	0.15	0.85
Mam	16	15.3833	-92.183	1083	Mexico	2155	Tropical conifers	4	17	24	45	0.28	0.72

Table 1 (Continued)

Yucatán Maya	17	20.97	-89.62	12.3	Mexico	1925	Moist broad leaves	7	28	19	54	0.39	0.61
Mayo	18	27.2298	-108.986	156	Mexico	2155	Dry broad leaves	10	13	5	28	0.62	0.38
Mazahuas	19	19.7125	-99.951	2568	Mexico	2155	Deserts	72	15	1	88	0.99	0.01
Mazatecos	20	18.1314	-96.841	1672	Mexico	1935	Moist broad leaves	4	21	34	59	0.25	0.75
Mixe	21	17.2167	-95.983	1323	Mexico	1716	Moist broad leaves	11	43	35	89	0.38	0.62
Mixtecas	22	17.55	-99.5	.	Mexico	2190	Tropical conifers	12	50	62	124	0.3	0.7
Nahuas	23	19.4333	-99.133	2230.2	Mexico	2075	Tropical conifers	7	58	70	135	0.27	0.73
Nahuat1	24	19.4333	-99.133	2249	Mexico	2150	Desert	7	25	20	52	0.37	0.63
Nahuatl2	25	18.5598	-99.271	880	Mexico	2227	Dry broad leaves	6	26	12	44	0.43	0.57
Nahuatl3	26	19.5521	-97.488	2560	Mexico	2150	Desert	3	25	24	52	0.3	0.7
Nahuatl4	27	21.2667	-98.783	284	Mexico	1863	Moist broad leaves	4	23	17	44	0.35	0.65
Otomí	28	20.4861	-99.259	1741	Mexico	2227	Deserts	28	84	108	220	0.32	0.68
Popoluca	29	18.2333	-94.867	441	Mexico	1789	Moist broad leaves	5	18	13	36	0.39	0.61
Purepecha1	30	19.68	-101.95	2426.3	Mexico	1900	Tropical conifers	4	10	7	21	0.43	0.57
Purepecha2	31	19.6833	-101.95	2405.8	Mexico	2000	Tropical conifers	2	13	14	29	0.29	0.71
Seri1	32	29.015	-112.162	7	Mexico	2227	.	14	5	0	19	0.87	0.13
Tarahumara	33	27.7523	-107.635	2345.6	Mexico	2150	Tropical conifers	17	15	6	38	0.64	0.36
Tojolabal	34	16.3153	-91.982	1521	Mexico	2082	Tropical conifers	6	28	12	46	0.53	0.57
Trikís	35	17.2667	-97.683	2063.1	Mexico	2075	Tropical conifers	0	14	75	89	0.08	0.92

Table 1 (Continued)

Yaquís	36	27.4707	-110.3318	9	Mexico	2155	Dry broad leaves	9	21	7	27	0.537	0.463
Zapotecas	37	17.0678	-96.72	1569	Mexico	2225	Dry broad leaves	7	9	26	42	0.27	0.73
	Expected genotypic values:					**ExpCC**	**ExpCT**	**ExpTT**		**Chi square**		**p**	
	*** Zapotecas**					**3**	**17**	**22**		**8.9**		**p<0.004**	

Table 2 Descriptive statistics and normality tests of quantitative variables.

Variable	n	Mean	Median	Stand Deviation	S-W for normality test p
Latitude	37	17.91	17.55	5.06	<0.054
Longitude	37	-95.95	-96.72	7.52	>0.15
Elevation*	36	1170.01	1142.6	903.44	<0.0041
Insolation	37	2011	2075	175.2	<0.0065

* Elevation is not available for the Mixteca sample.

4 | DISCUSSION

Several research groups have attempted to understand from a microevolutionary perspective why the T allele in the MTHFR gene would reach high frequencies despite the allele's harmful effects. While it is commendable that different groups have investigated the C677T variant from different perspectives, some of us (including our own group), were unaware of other groups' work, which is an omission we deeply regret. Only by including the contributions of others, and by researching diverse human populations, will our field be able to understand the evolutionary origin and further microevolutionary history of the C677T SNP.

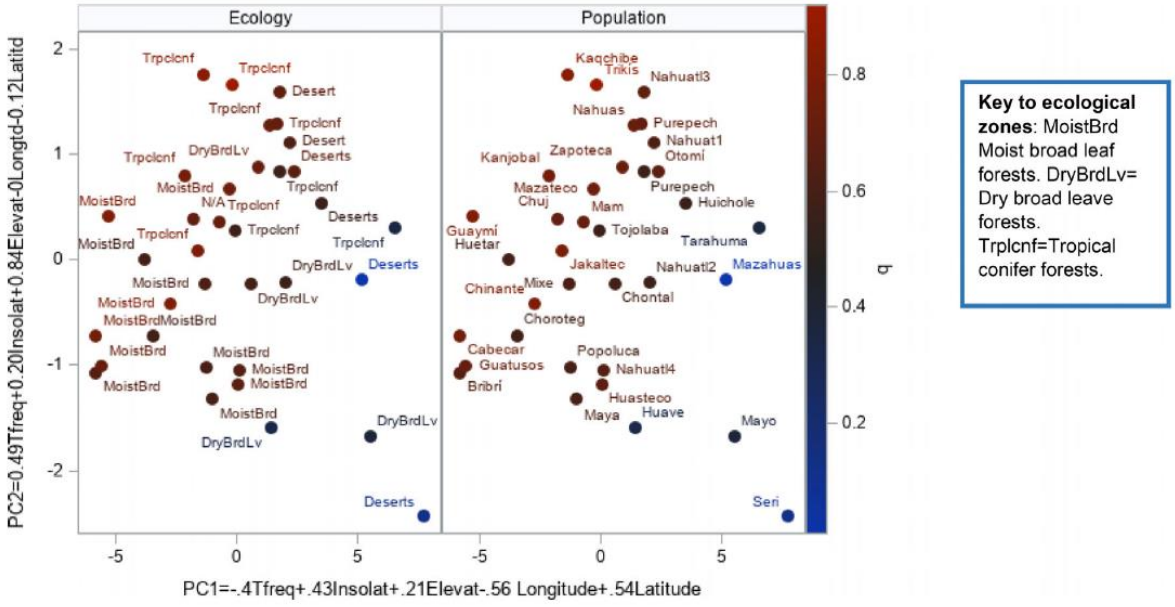

Figure 1 Samples plotted along their values of the first and second principal component vector. Populations shown by their name and by their ecology.

This paper has limitations. We are relying on secondary data for obtaining the allele frequencies. In addition, some of the populations moved away from their place of origin. The Maya people we included in this study lived in the region now known as Guatemala and Mexico, and they were displaced by the Guatemalan government during the 1980s Civil War from Guatemala into Mexico. These groups claim specific places in Guatemala as their place of origin, but they all descend from the great Maya Civilization (Fash, 1994). It is unfortunate that we could not find information about the frequencies of the allele among the Indigenous people from Honduras, Belize, Nicaragua, El Salvador, and Panamá. Of course, such research must be done only under the guidance, approval, and consent of the Indigenous people themselves and must not be imposed on them.

Strengths of this paper are its holistic perspective and methodology. We considered numerous variables in attempting to explain the distribution of the variable. We asked if the T allele could have evolved to these high levels as a response to toxic metals, solar irradiation, altitude, latitude, and longitude. We computed the correlation of the allele with these variables. We applied population genetics and GIS methods to understand the distribution of the allele in Mexico and Central America.

Some researchers have looked for UV-related explanations for the evolution of high frequencies of the T allele (Lucock et al., 2015, 2017, 2018; Wang et al., 2012). This is understandable, since the MTHFR gene affects the folate cycle. The MTHFR enzyme is active in the folate pathway and is also closely associated with the homocysteine (Hcy) metabolism. According to Lai and Kan (2015), any disturbances to either pathway will result in accumulations of homocysteine (Lai & Kan, 2015). Differences in access to folate in traditional cuisine or in supplemented foods may ameliorate such disturbances. We note that Servy et al. challenge the notion that folate

supplementation is advantageous to TT mothers and refer the reader to their paper (Servy et al., 2018). Cross-cultural and bio-cultural approaches to the study of this allele are necessary because different human populations differ in their skin color and because UV radiation depletes folate.

In evolutionary population genetics studies, it is not always obvious that an allele found at high frequencies in a population arrived at its high frequency due to natural selection rather than due to genetic drift or even gene flow. If said allele is found in multiple populations (as is the case for C677T), then there is reason to suspect natural selection. If these populations are not small (as is the case here), then there is reason to suspect natural selection. If the allele in the homozygote state (and possibly in the heterozygote state) is accompanied by deleterious effects, as discussed in the introduction, then there is even more reason to suspect natural selection (Jobling et al., 2014; Werren et al., 2020).

In this paper, we focused on the region with the highest frequencies of the T allele: Mexico and Central American Indigenous populations. Some of the populations in this region have allele frequencies (to our knowledge) unseen anywhere else in the world. We asked how this allele is distributed across space, within populations and how it is correlated with ecological variables. Our principal component and GIS analyses show us that the main explanatory variable of the variation of the allele is geography, where the populations with the highest T allele frequencies cluster in the South East of the region. The second orthogonal PC is one in which the allele frequencies are explained by altitude and insolation. Both PCs suggest solar radiation as the causal factor of the allele's distribution.

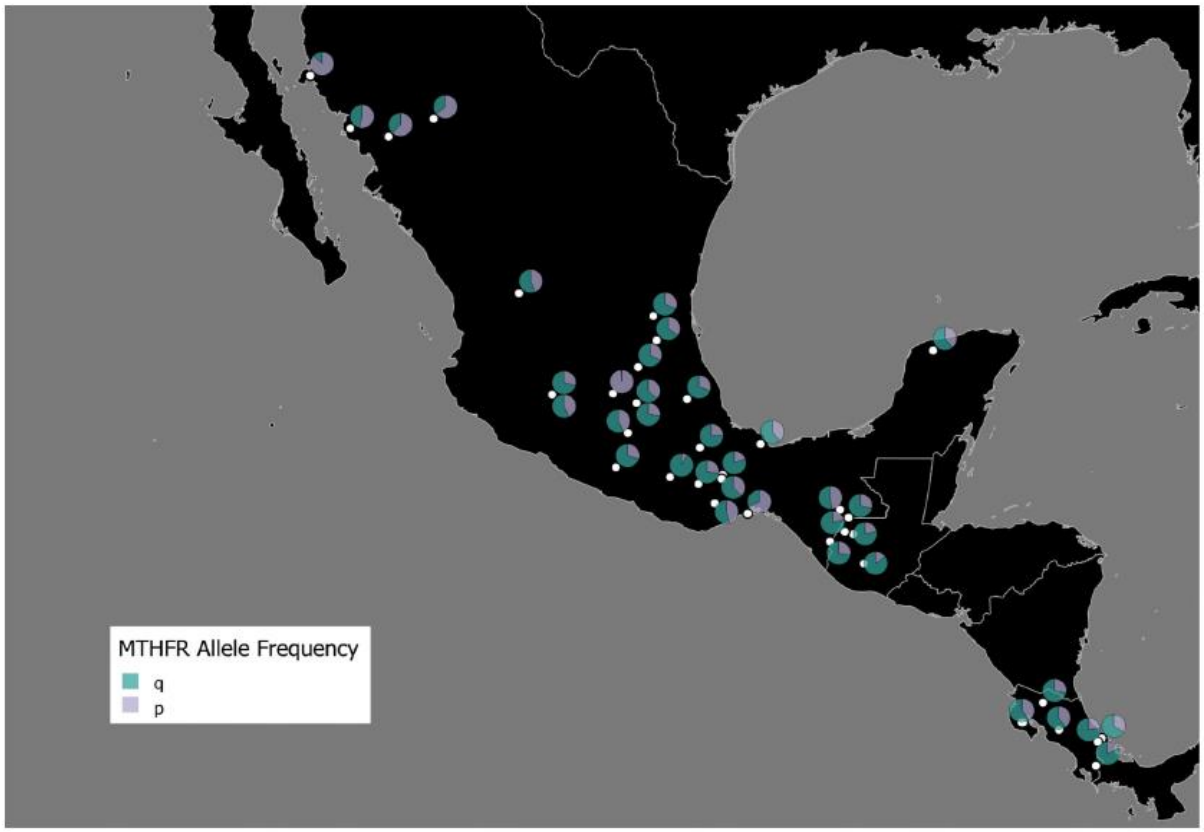

Figure 2 Pie charts of the allele frequencies of the C and T alleles in the 37 populations.

Although our team had previously entertained the possibility that the T allele had evolved in this region as a response to toxic levels of environmental lead (Reyes et al., 2019, 2020) we no longer think that is the case. Rather, we think that this allele evolved as a response to geographically mediated ecology, as demonstrated by our map.

The F statistics computations provide a view into the population structure of these Native American population allele frequencies. Seen as an array of points within a meta population, the individuals within these populations are found to differ moderately among them. It is quite interesting that population membership hardly differentiated these groups. Instead, what is quite

remarkable is the amount of within-population diversity, which is obvious from the Hardy–Weinberg tests and from the extremely low value of F_{is} (0.0082). This value reflects the high levels of heterozygosity found within groups. Indeed, the only sample not in Hardy–Weinberg equilibrium had a deficit in heterozygotes. The distribution of the allele frequencies, the Hardy–Weinberg equilibrium results and the F statistics analyses all paint a vision of a group of populations which evolved to achieve their present level of genotypic frequencies and are currently in equilibrium. In their current level of equilibrium, high levels of heterozygosity are favored and maintained.

The results of our analyses all point to natural selection rather than genetic drift as the force which drove the allele to its current levels in this region of the world. However, we are not ready to propose a natural selection scenario which initially chose the T allele and drove it into its current frequencies. Here, we wish to quote Jobling et al., who say (p. 159): "Selection often substantially increases the probability that an advantageous allele becomes fixed compared to a neutral allele; in humans, most new advantageous alleles are still far more likely to be eliminated than fixed. If new alleles are almost exclusively deleterious, the optimal allele can persist unchanged over very long time-scales." (Jobling et al., 2014). We agree with Jobling et al. that it is improbable that a deleterious allele was naturally selected under any circumstance. The high frequency of the C677T allele must have resulted from natural selection because it was advantageous under some condition to gain the highest evolutionary fitness. After all, the protein produced by the SNP needs to be fed with folate to rescue its proper structure and function to result in positive adaptation.

Our next aim is to estimate the fitness of the phenotypes by differential mortality and fertility in some populations within the region. As Richard Lewontin notes, the better adapted individuals in

a specific environment will have higher fitness (Lewontin, 1957). We would like to know if (and how) the Native American populations of this region are able to maintain the high frequency of this allele. We do not wish to assume that the allele is adaptive, given the overwhelming clinical literature indicating that it is not.

This paper also has public health applications. The majority population of Mexico and Central America descends from the two populations with the highest frequencies in the world of the T allele: The Native people of the region and the Mediterranean. The entire region is amid an epidemiological transition. It is possible that the current epidemiological transition into cardiovascular diseases and cancer that the continent is experiencing is due not only to environmental changes but also to its genetic structure.

The high frequency of the T allele in the Mexican and Central American Native peoples should be a reason for them to remain sovereign of their rights, lands, and traditions. Evolutionary theory predicts that if these people evolved the allele as an adaptation to an environment, then they are likely to suffer its deleterious effects outside of the environment.

5 | CONCLUSIONS

The T allele of the C677T SNP in the MTHFR gene achieves its highest frequency in the world in Mexico and Central American Native populations. In this region, the allele has a Southeast to Northwest cline, also associated with insolation and altitude. Our results suggest that the allele in this region evolved because of natural selection rather than genetic drift. However, we cannot explain the adaptive reason which lead to this allele been selected to its current frequencies. Although it is possible that the culinary traditions of these peoples helps alleviate the deleterious aspects associated with the TT and CT phenotypes, as proposed by others (Binia et al., 2014; Contreras-Cubas et al., 2016), evolutionary theory does not predict that this cultural environment would have selected a deleterious allele. As Richard Lewontin says (Lewontin, 1957), there must have been an initial problem which the allele "solved" and for which it was naturally selected and brought to its current levels. At this point, we are not ready to propose what that problem was.

6 | ETHICS STATEMENT

The data analyzed in this paper were obtained from published papers. The authors of each and every paper wrote that they obtained consent from the participants and authorization for their respective Institutional Review Boards. Because the data were already published and de-identified, they were no longer considered to be human subject data. For that reason, according to the University of South Florida IRB, it is not necessary to obtain Human Research approval (https://arc.research.usf.edu/).

7 | REFERENCES

Ahrens, C. D. (2007). Meteorology today: An introduction to weather, climate and the environment. Thomson/Brooks/Cole. ISBN-13:978-1305113589.

Antonio-Vejar, V., del Moral-Hernandez, O., Alarcon-Romero, L. C., Flores-Alfaro, E., Leyva-Vazquez, M. A., Hernandez-Sotelo, D., & Illades-Aguiar, B. (2014). Ethnic variation of the C677T and A1298C polymorphisms in the methylenetetrahydrofolate-reductase (MTHFR) gene in southwestern Mexico. Genetics and Molecular Research, 13(3), 7950–7957. https://doi.org/10.4238/2014.September.29.8

Binia, A., Contreras, A. V., Canizales-Quinteros, S., Alonzo, V. A., Tejero, M. E., & Silva-Zolezzi, I. (2014). Geographical and ethnic distribution of single nucleotide polymorphisms within genes of the folate/homocysteine pathway metabolism. Genes and Nutrition, 9(5), 421. https://doi.org/10.1007/s12263-014-0421-7

Chillemi, R., Angius, A., Persico, I., Sassu, A., Prodi, D., & Musumci, S. (2005). Methylenetetrahydrofolate reductase (MTHFR) from Mediterranean to sub-Saharan areas. Online Journal of Biological Sciences, 6(1), 28–34.

Contreras-Cubas, C., Sanchez-Hernandez, B. E., Garcia-Ortiz, H., Martinez-Hernandez, A., Barajas-Olmos, F., Cid, M., Mendoza- Caamal, E., Centeno-Cruz, F, Jiménez-López, J.C., Córdova, E.J., Salas-Bautista, E.V., Saldana-Alvarez, Y., Fernandez-Lopéz, J.C, Mutchinick, O.M., & Orozco, L. (2016). Heterogenous distribution of MTHFR gene variants among mestizos and diverse Amerindian groups from Mexico. PLoS One, 11(9), 13. https://doi.org/10.1371/journal.pone.0163248

Crider, K., Bailey, L., & Berry, R. (2011). Folic acid food fortification. Its history, effects, concerns and future directions. Nutrients, 3, 370–384.

Cui, L. H., Yu, Z., Zhang, T. T., Shin, M. H., Kim, H. N., & Choi, J. S. (2011). Influence of polymorphisms in MTHFR 677 C -> T, TYMS 3R -> 2R and MTR 2756 a -> G on NSCLC risk and response to platinum-based chemotherapy in advanced NSCLC. Pharmacogenomics, 12(6), 797–808. https://doi.org/10.2217/pgs.11.27

Fash, W. L. (1994). Changing perspectives on Maya civilization. Annual Review of Anthropology, 23, 181–208. https://doi.org/10.1146/annurev.an.23.100194.001145

Gupta, N., Sarkar, S., David, A., Gangwar, P. K., Gupta, R., Khanna, G., … Rajender, S. (2013). Significant impact of the MTHFR polymorphisms and haplotypes on male infertility risk. PLoS One, 8(7), e69180. https://doi.org/10.1371/journal.pone.0069180

Hartl, D. (1988). A primer of population genetics. Sinauer Associates, Inc. ISBN13: 9780878933013.

Husain, N., Zainal, A., BSM, S., Mohamed, N., & Nor, N. (2011). Integrated PV based solar insolation measurement and performance monitoring system. In: Paper presented at the IEEE colloquium on humanities, Science and Engineering Research, Penang, China.

Imbard, A., Benoist, J., & Blom, J. (2013). Neural tube defects, folic acid and methylation. International Journal of Environmental Research and Public Health, 10, 4352–4389.

Irfan, M., Ismail, M., Beg, M. A., Shabbir, A., Kayani, A. R., & Raja, G. K. (2016). Association of the mthfr C677T (rs1801133) polymorphism with idiopathic male infertility in a local Pakistani population. Balkan Journal of Medical Genetics, 19(1), 51–61. https://doi.org/10.1515/bjmg-2016-0007

Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. Journal of Human Evolution, 39(1), 57–106. https://doi.org/10.1006/jhev.2000.0403

Jablonski, N. G., & Chaplin, G. (2010). Human skin pigmentation as an adaptation to UV radiation. Proceedings of the National Academy of Sciences of the United States of America, 107, 8962–8968. https://doi.org/10.1073/pnas.0914628107

Jobling, M., Hollox, E., Hurles, M., Kivisild, T., & Tyler-Smith, C. (2014). Human evolutionary genetics (2nd ed.). Garland Science. https://www.taylorfrancis.com/books/human-evolutionarygenetics-mark-jobling-edward-hollox-matthew-hurles-toomaskivisild-chris-tyler-smith/10.1201/9781317952268.

Jones, P., Lucock, M., Veysey, M., Jablonski, N., Chaplin, G., & Beckett, E. (2018). Frequency of folate-related polymorphisms varies by skin pigmentation. American Journal of Human Biology, 30(2), 1–9. https://doi.org/10.1002/ajhb.23079

Juarez-Velazquez, R., Canto, P., Canto-Cetina, T., Rangel-Villalobos, H., Rosas-Vargas, H., Rodriguez, M., Coral-Vazquez, R. M., Canizalez-Quintero, S., Velazquez-Wong, A.C, Ordoñez-Razo, A.M, Vilchis-Dorantes, A.M, & Coral-Vazquez, R.M. (2010). Analysis of polymorphisms in genes (AGT, MTHFR, GPIIIa, and GSTP1) associated with hypertension, thrombophilia and oxidative stress in mestizo and Amerindian populations of Mexico. Disease Markers, 28(5), 323–331. https://doi.org/10.1155/2010/716542

Krawczyk, P., Kucharczyk, T., Kowalski, D., Powrozck, T., Ramlau, R., Kalinka-Warzocha, E., Milanowski, J. (2014). Polymorphisms in TS, MTHFR and ERCC1 genes as predictive markers in first-line platinu and pemetrexed therapy in NSCLC patients. Journal of Cancer Research Clinical Oncology, 140, 2047–2057.

Kurzawski, M., Wajda, A., Malinowski, D., Kazienko, A., Kurzawa, R., & Drozdzik, M. (2015). Association study of folate-related enzymes (MTHFR, MTR, MTRR) genetic variants with non-obstructive male infertility in a polish population. Genetics and Molecular Biology, 38(1), 42–47. https://doi.org/10.1590/s1415-475738120140179

Lai, W., & Kan, M. (2015). Homocysteine-induced endothelial dysfunction. Annals of Nutrition and Metabolism, 67, 1–12.

Le Marchand, L., Haiman, C. A., Wilkens, L. R., Kolonel, L. N., & Henderson, B. E. (2004). MTHFR polymorphisms, diet, HRT, and breast cancer risk: The multiethnic cohort study. Cancer Epidemiology Biomarkers & Prevention, 13(12), 2071–2077.

Lewontin, R. (1957). The adaptations of populations to varying environments. Cold Spring Harbor Symposia on Quantitative Biology, 22, 395–408. https://doi.org/10.1101/sqb.1957.022.01.037

Li, X., Shao, M., Wang, S., Zhao, X., Chen, H., Ojan, J., Song, X., Wang, J., Jin, J.,Wu, J., Li, Q., Bai, C., Han, B., Gao, Z., & Lu, D. (2014). Heterozygote advantage of methylenetetrahydrofolate reductase polymorphisms on clinical outcomes in advanced non-small cell lung cancer (NSCLC) patients treated with platinum-based chemotherapy. Tumor Biology, 35, 11159–11170.

Lucock, M., Beckett, E., Martin, C., Jones, P., Furst, J., Yates, Z., Jablonski, N., Chaplin, C., & Veysey, M. (2017). UV-associated decline in systemic folate: Implications for human nutrigenetics, health, and evolutionary processes. American Journal of Human Biology, 29(2), 1–13. https://doi.org/10.1002/ajhb.22929

Lucock, M., Glanville, T., Yates, Z., Walker, J., Furst, J., & Simpson, N. (2012). Solar cycle predicts folate-sensitive neonatal genotypes at discrete phases of the first trimester of pregnancy: A novel folate-related human embryo loss hypothesis. Medical Hypotheses, 79(2), 210–215. https://doi.org/10.1016/j.mehy.2012.04.039

Lucock, M., Thota, R., Garg, M., Martin, C., Jones, P., Furst, J., Yates, Z., Yablonski, N., Chaplin, G, Veysey, M., Sutherland, J.M., & Becket, E. (2018). Vitamin D and folate: A reciprocal environmental association based on seasonality and genetic disposition. American Journal of Human Biology, 30(5), e23166.https://doi.org/10.1002/ajhb.23166

Lucock, M., & Yates, Z. (2006). Synergy between 677 TT MTHFR genotype and related folate SNPs regulates homocysteine level. Nutrition Research, 26(4), 180–185. https://doi.org/10.1016/j.nutres.2006.01.001

Lucock, M., Yates, Z., Martin, C., Choi, J., Beckett, E., Boyd, L., Veysey, M. (2015). Methylation diet and methyl group genetics in risk for adenomatous polyp occurrence. BBA Clinical, 3, 107–112. https://doi.org/10.1016/j.bbacli.2014.11.005

Lucock, M., Yates, Z., Ng, X., Veysey, M., Blades, B., Travers, C., Roach, P. (2008). Preliminary evidence for genetic selection of 677T-MTHFR by natural annual cycle of folate abundance. Journal of Nutrigenetics and Nutrigenomics, 1(1–2), 24–29. https://doi.org/10.1159/000109872

Mackay, M. E. A. (2015). Solar radiation. Oxford University Press. https://global.oup.com/ukhe/product/solar-energy-9780199652112?cc=ca&lang=en&.

Manly, B. (2005). Multivariate statistical methods: A primer (3rd ed.). Chapman & Hall, CRC.
Matsushita, S., Muramatsu, T., Arai, H., Matsui, T., & Higuchi, S. (1997). The frequency of the methylenetetrahydrofolate reductase-gene mutation varies with age in the Normal population. American Journal of Human Genetics, 61, 1459–1460.

Mayor-Olea, A., Callejon, G., Palomares, A. R., Jimenez, A. J., Jesus Gaitan, M., Rodriguez, A., Reyes-Engel, A. (2008). Human genetic selection on the MTHFR 677C > T polymorphism. BMC Medical Genetics, 9, 1–7. https://doi.org/10.1186/1471-2350-9-104

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: F-ST and related measures. Molecular Ecology Resources, 11(1), 5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Naushad, S. M., Krishnaprasad, C., & Devi, A. R. R. (2014). Adaptive developmental plasticity in methylene tetrahydrofolate reductase (MTHFR) C677T polymorphism limits its frequency in south Indians. Molecular Biology Reports, 41(5), 3045–3050. https://doi.org/10.1007/s11033-014-3163-0

Ojeda-Granados, C., Panduro, A., Gonzalez-Aldaco, K., Sepulveda-Villegas, M., Rivera-Iniguez, I., & Roman, S. (2017). Tailoring nutritional advice for Mexicans based on prevalence profiles of diet-related adaptive gene polymorphisms. Journal of Personalized Medicine, 7(4), 18. https://doi.org/10.3390/jpm7040016

Ramachandran, S., Tang, H., Gutenkunst, R., & Bustamante, C. (2010). Genetics and genomics of human population structure. In M. E. A. Speicher (Ed.), Vogel and Motulsky's human genetics: Problems and approaches. Springer-Verlag.

Raymond, M., & Rousset, F. (1995). Genepop (Version-1.2)—Population-genetics software for exact tests and ecumenicism. Journal of Heredity, 86(3), 248–249. https://doi.org/10.1093/oxfordjournals.jhered.a111573

Reyes, L., Gonzalez-Hernandez, F., Godfrey, D., Jean, M., Bidal, M., & Madrigal, L. (2019). The world-wide distribution of the C677T allele of the MTHFR gene. American Journal of Physical Anthropology, 168, 203–203.

Reyes, L., MacLean, C., Godfrey, D., Gonzalez-Hernandez, F., & Madrigal, L. (2020). The association among the C677T mutation in the MTHFR gene, skin color measures and UV radiation suggests local adaptation rather than sweeping clines. American Journal of Physical Anthropology, 171, 233–233.

Reyes-Engel, A., Muñoz, E., Gaitan, M. J., Fabre, E., Gallo, M., Dieguez, J. L., Ruiz, M., & Morell, M. (2002). Implications on human fertility of the 677C -> T and 1298A -> C polymorphisms of the MTHFR gene: Consequences of a possible genetic selection. Molecular Human Reproduction, 8(10), 952–957. https://doi.org/10.1093/molehr/8.10.952

Robien, K., & Ulrich, C. M. (2003). 5,10-methylenetetrahydrofolate reductase polymorphisms and leukemia risk: A HuGE minireview. American Journal of Epidemiology, 157(7), 571–582. https://doi.org/10.1093/aje/kwg024

Rousset, F. (2008). GENEPOP'007: A complete re-implementation of the GENEPOP software for windows and Linux. Molecular Ecology Resources, 8(1), 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

SAS Institute Inc 2013. SAS/ACCESS® 9.4 Interface to ADABAS: Reference. Cary, NC: SAS Institute Inc.

Schiepers, O. J. G., van Boxtel, M. P. J., Harris, S. E., Gow, A. J., Pattie, A., Brett, C. E., … Deary, I. J. (2011). MTHFR polymorphisms and cognitive ageing in the ninth decade: The Lothian birth cohort 1921. Genes Brain and Behavior, 10(3), 354–364. https://doi.org/10.1111/j.1601-183X.2010.00675.x

Schneede, J., Refsum, H., & Ueland, P. (2000). Biological and environmental determinants of plasma homocysteine. Seminars in Thromobosis and Hemostasis, 26, 263–279.

Servy, E. J., Jacquesson-Fournols, L., Cohen, M., & Menezo, Y. J. R. (2018). MTHFR isoform carriers. 5-MTHF (5-methyl tetrahydrofolate) vs folic acid: A key to pregnancy outcome: A case series. Journal of Assisted Reproduction and Genetics, 35(8), 1431–1435. https://doi.org/10.1007/s10815-018-1225-2

Tsai, S.-J., Hong, C.-J., Yeh, H.-L., Liou, Y.-J., Yang, A. C., Liu, M.E., & Hwang, J.-P. (2011). Heterozygote advantage of the MTHFR C677T polymorphism on specific cognitive performance in elderly Chinese males without dementia. Dementia and Geriatric Cognitive Disorders, 32(3), 159–163. https://doi.org/10.1159/000333074

Wang, L.-J., Lee, S.-Y., Chen, S.-L., Chang, Y.-H., Chen, P. S., Huang, S.-Y., … Lu, R.-B. (2015). A potential interaction between COMT and MTHFR genetic variants in Han Chinese patients with bipolar II disorder. Scientific Reports, 5, 1–6. https://doi.org/10.1038/srep08813

Wang, X. M., Fu, J. J., Li, Q. X., & Zeng, D. Y. (2016). Geographical and ethnic distributions of the MTHFR C677T, A1298C and MTRR A66G gene polymorphisms in Chinese populations: A meta-analysis. PLoS One, 11(4), e0152414. https://doi.org/10.1371/journal.pone.0152414

Wang, Y. F., Pei, L. J., Wang, J. F., & Zheng, X. Y. (2012). Is the prevalence of MTHFR C677T polymorphism associated with ultraviolet radiation in Eurasia? Journal of Human Genetics, 57 (12), 780–786. https://doi.org/10.1038/jhg.2012.113

Weir, B. S., & Cockerham, C. C. (1984). Estimating f-statistics for the analysis of population structure. Evolution, 38(6), 1358–1370. https://doi.org/10.2307/2408641

Werren, E. A., Garcia, O., & Bigham, A. W. (2020). Identifying adaptive alleles in the human genome: From selection mapping to functional validation. Human Genetics, 1–36. https://doi.org/10.1007/s00439-020-02206-7

Wilcken, B., Bamforth, F., Li, Z., Zhu, H., Ritvanen, A., Redlund, M., Botto, L. D. (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(8), 619–625. https://doi.org/10.1136/jmg.40.8.619

Yadav, U., Kumar, P., Gupta, S., & Rai, V. (2017). Distribution of MTHFR C677T gene polymorphism in healthy north Indian population and an updated meta-analysis. Indian Journal of Clinical Biochemistry, 32(4), 399–410. https://doi.org/10.1007/s12291-016-0619-0

Yates, Z., & Lucock, M. (2008). Folate and the C677T-MTHFR variant—Impact on population health. Journal of Nutrigenetics and Nutrigenomics, 1(4), 189–189.

Zetterberg, H., Regland, B., Palmer, M., Ricksten, A., Palmqvist, L., Rymo, L., … Blennow, K. (2002). Increased frequency of combined methylenetetrahydrofolate reductase C677T and A1298C mutated alleles in spontaneously aborted Embrys. European Journal of Human Genetics, 10, 113–118.

Zhang, W., Lin, W. Q., Cao, H. F., Li, C. Y., & Li, F. (2015). Association of a miR-34b binding site single nucleotide polymorphism in the 3'-untranslated region of the methylenetetrahydrofolate reductase gene with susceptibility to male infertility. Genetics and Molecular Research, 14(4), 12196–12204. https://doi.org/10.4238/2015.October.9.8

Zhu, X. D., Liu, Z. G., Zhang, M. C., Gong, R. H., Xu, Y. J., & Wang, B. M. (2016). Association of the methylenetetrahydrofolate reductase gene C677T polymorphism with the risk of male infertility: A meta-analysis. Renal Failure, 38(2), 185–193. https://doi.org/10.3109/0886022x.2015.1111086