Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2024;28(6):650-658

DOI 10.18699/vjgb-24-72

Genetic aspects of lactase deficiency in indigenous populations of Siberia

B.A. Malyarchuk

Institute of Biological Problems of the North of the Far Eastern Branch of the Russian Academy of Sciences, Magadan, Russia
malbor@mail.ru

Abstract. The ability to metabolize lactose in adulthood is associated with the persistence of lactase enzyme activity. In European populations, lactase persistence is determined mainly by the presence of the rs4988235-T variant in the *MCM6* gene, which increases the expression of the *LCT* gene, encoding lactase. The highest rates of lactase persistence are characteristic of Europeans, and the lowest rates are found in East Asian populations. Analysis of published data on the distribution of the hypolactasia-associated variant rs4988235-C in the populations of Central Asia and Siberia showed that the frequency of this variant increases in the northeastern direction. The frequency of this allele is 87 % in Central Asia, 90.6 % in Southern Siberia, and 92.9 % in Northeastern Siberia. Consequently, the ability of the population to metabolize lactose decreases in the same geographical direction. The analysis of paleogenomic data has shown that the higher frequency of the rs4988235-T allele in populations of Central Asia and Southern Siberia is associated with the eastward spread of ancient populations of the Eastern European steppes, starting from the Bronze Age. The results of polymorphism analysis of exons and adjacent introns of the *MCM6* and *LCT* genes in indigenous populations of Siberia indicate the possibility that polymorphic variants may potentially be related to lactose metabolism exist in East Asian populations. In East Asian populations, including Siberian ethnic groups, a ~26.5 thousand nucleotide pairs long region of the *MCM6* gene, including a combination of the rs4988285-A, rs2070069-G, rs3087353-T, and rs2070068-A alleles, was found. The rs4988285 and rs2070069 loci are located in the enhancer region that regulates the activity of the *LCT* gene. Analysis of paleogenomic sequences showed that the genomes of Denisovans and Neanderthals are characterized by the above combination of alleles of the *MCM6* gene. Thus, the haplotype discovered appears to be archaic. It could have been inherited from a common ancestor of modern humans, Neanderthals, and Denisovans, or it could have been acquired by hybridization with Denisovans or Neanderthals. The data obtained indicate a possible functional significance of archaic variants of the *MCM6* gene.

Key words: genetic polymorphism; lactase persistence; *MCM6* gene; LCT gene; human populations; Siberia; archaic variants of polymorphism.

For citation: Malyarchuk B.A. Genetic aspects of lactase deficiency in indigenous populations of Siberia. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(6):650-658. DOI 10.18699/vjgb-24-72

Генетические аспекты лактазной недостаточности у коренного населения Сибири

Б.А. Малярчук

Институт биологических проблем Севера Дальневосточного отделения Российской академии наук, Магадан, Россия
malbor@mail.ru

Аннотация. Способность метаболизировать лактозу во взрослом состоянии связана с сохранением активности фермента лактазы. В европейских популяциях персистенция лактазы детерминируется главным образом наличием варианта rs4988235-T в гене *MCM6*, который увеличивает экспрессию гена *LCT*, кодирующего лактазу. Наиболее высокие показатели персистенции лактазы характерны для европейцев, а самые низкие – для населения Восточной Азии. Анализ опубликованных данных о распределении варианта rs4988235-C, связанного с гиполактазией, у населения Центральной Азии и Сибири выявил, что частота этого варианта увеличивается в северо-восточном направлении. В Центральной Азии частота этого аллеля составляет 87 %, на юге Сибири – 90.6 % и на северо-востоке Сибири – 92.9 %. Соответственно, в таком же географическом направлении убывает способность населения метаболизировать лактозу. Анализ палеогеномных данных показал, что более высокая частота аллеля rs4988235-T в популяциях Центральной Азии и Южной Сибири связана с распространением на восток древнего населения восточноевропейских степей начиная с эпохи бронзового века. Результаты анализа полиморфизма экзонов и прилегающих к ним интронов генов *MCM6* и *LCT* у коренного населения Сибири свидетельствуют о возможности существования в восточноазиатских популяциях вариантов полиморфизма, потенциально связанных с метаболизмом лактозы. В популяциях Восточной Азии, в том числе в сибирских

этнических группах, обнаружен участок гена *MCM6* длиной ~26.5 тыс. пар нуклеотидов, включающий комбинацию аллелей rs4988285-A, rs2070069-G, rs3087353-T, rs2070068-A. Локусы rs4988285 и rs2070069 находятся в области энхансера, регулирующего активность гена *LCT*. Анализ палеогеномных последовательностей показал, что указанной выше комбинацией аллелей гена *MCM6* характеризуются геномы денисовцев и неандертальцев. Таким образом, обнаруженный гаплотип, по всей видимости, является архаичным. Он мог быть унаследован от общего предка современных людей, неандертальцев и денисовцев, или же был приобретен в результате гибридизации с денисовцами или неандертальцами. Полученные данные свидетельствуют о возможной функциональной значимости архаичных вариантов полиморфизма гена *MCM6*.
Ключевые слова: генетический полиморфизм; персистенция лактазы; ген *MCM6*; ген *LCT*; популяции человека; Сибирь; архаичные варианты полиморфизма.

Introduction

Lactose (milk sugar) is the main disaccharide in the milk of various mammals and its hydrolysis requires the enzyme lactase, encoded by the *LCT* gene, which is predominantly expressed in the small intestine. Lactase activity declines during ontogenesis, which can lead to difficulties digesting lactose in many adults (Ségurel, Bon, 2017). Primary hypolactasia (OMIM: 223100) is characterized by a range of symptoms (bloating, nausea, diarrhoea) after ingestion of milk and dairy products. However, ethnoregional populations around the world have been found to differ in their ability to metabolize lactose (Evershed et al., 2022). It is thought that this ability, or lactase persistence (LP), is inherited. One of the most important genetic polymorphisms that have been linked to LP is the T variant at the rs4988235 locus of the *MCM6* gene, which regulates the expression of the *LCT* gene (Enattah et al., 2002; Olds, Sibley, 2003; Troelsen et al., 2003). Although this genetic variant is about 14,000 nucleotide pairs away from the *LCT* gene (which is why it is often called –13910*T), it is responsible for increasing the enzymatic activity of lactase, which breaks down lactose into glucose and galactose molecules.

It would appear that the lowest LP values are characteristic of East Asian populations, while the highest are found in Europeans (Liebert et al., 2017). This is due to the fact that, according to archaeological data, dairy farming may have emerged in the steppe zone of the North Caucasus and the Black Sea region around 4–5 thousand years ago (kya) (Scott et al., 2022). Paleogenomic data suggests that the frequency of the LP-associated variant rs4988235-T began to increase around 6 kya within the ancestral EHG and CHG genomic components characteristic of Eastern European and Caucasian hunter-gatherers, respectively (Segurel et al., 2020; Irving-Pease et al., 2024). The linkage between the polymorphism variants in the rs4988235 and rs1438307 loci was also revealed, and the increase in frequency of the rs1438307-T allele may have begun much earlier than previously thought, around 12 kya (Irving-Pease et al., 2024). With regard to the rs1438307-T variant, it has been suggested that it may have arisen as a consequence of the adaptation of ancient humans to starvation and exposure to pathogens; this is based on the observation that it is involved in the regulation of the body's energy expenditure and the development of metabolic diseases (Evershed et al., 2022).

Despite the great interest of the genetic and medical communities in the genetic aspects of hypolactasia in human populations, many regions of the world remain poorly studied (Liebert et al., 2017; Anguita-Ruiz et al., 2020). The aim of this paper is to attempt to review the results of studies on the polymorphism of the *LCT* and *MCM6* genes, which are directly related to lactase persistence, in indigenous populations of Siberia, one of the least studied regions.

Distribution of rs4988235 locus polymorphisms in modern and ancient North Asian populations

Genetic and epidemiological studies have indicated that in populations of the European part of Russia, primary hypolactasia is determined predominantly or exclusively by the rs4988235-C allele of the *MCM6* gene (Borinskaya et al., 2006; Kovalenko et al., 2023), and accordingly, LP is defined by the rs4988235-T allele. However, in East Asian populations (including Siberian ones), this relationship is not as clearly evident – some populations (e. g., Buryats and Uyghurs) show very high (at the level of 95 %) frequency of the rs4988235-C variant, which is associated with a reduced prevalence of hypolactasia (Borinskaya et al., 2006; Sokolova et al., 2007). In this context, it has been proposed that the lower frequency of hypolactasia in some ethnic groups of Siberia and Central Asia may be associated with the presence of not only the rs4988235-T variant, but also some other genetic LP markers (Sokolova et al., 2007).

To date, it has been found that in addition to the rs4988235- T allele, several other genetic polymorphism variants that determine the ability to break down lactose are common in African and Middle Eastern ethnic groups, for example they include the rs41525747, rs41380347, rs145946881 and rs182549 loci of the *MCM6* gene (Ingram et al., 2007; Tishkoff et al., 2007). However, data on the association between genetic polymorphisms and LP in East Asian populations is somewhat more conflicting. For instance, in Central Asian populations, a mixed sample of Tajiks and Uzbeks, as well as Kazakhs, showed that the rs4988235-T variant (with frequencies of 10 and 16.5 %, respectively) correlated quite well with the ability to digest lactose (11–30 % in Tajiks and Uzbeks, and 25–32 % in Kazakhs) (Heyer et al., 2011). It would appear that Tibetans, who have a long-standing tradition of consuming yak milk, also digest lactose at a level of around 30 %, but they appear to lack the rs4988235-T and rs182549-T polymorphism variants that are found in neighbouring populations in northern China at a frequency of 3.8 and 6.9 %, respectively (Xu et al., 2010; Peng et al., 2012). Tibetans have been found to have their own spectrum of alleles of the enhancer region of the *MCM6* gene, which may be associated with LP, among which the –13838*A variant appears to predominate with a frequency of about 6.5 % (Peng et al., 2012).

A more detailed study of the genetic adaptation to milk consumption in Central Asian populations, distinguished by their economic patterns, has demonstrated that pastoralists

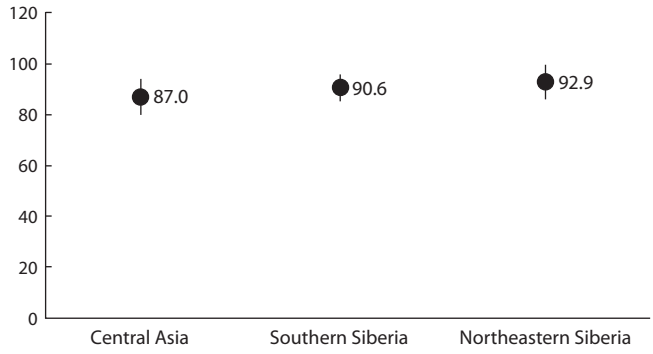
(Kazakhs, Kyrgyz, Karakalpaks, Buryats, Mongolians and Altaians), whose diets rely heavily on dairy products, do not have a higher ability to metabolize lactose than farmers (Turkmens, Tajiks and Uzbeks), who have a higher prevalence of the rs4988235-T variant (Sequrel et al., 2020). The relatively low frequency (~10 %) of this genetic variant is also observed in ethnic groups in southern Siberia (Khakasians, Shorians and Tubalars) who lead a semi-nomadic lifestyle and are engaged in forestry and taiga hunting (Sequrel et al., 2020). The data indicate that the frequency of occurrence of the rs4988235-T variant in Central Asian and Siberian populations is not significantly influenced by economic structure or milk consumption levels.

Table 1 presents the distribution of the rs4988235-C variant in various indigenous populations from Northeast China, Central Asia and Siberia. As can be observed, the frequency of this variant in the populations varies from 70 to 100 %. However, when the samples are divided into three regional groups, there is an increase in the frequency of the rs4988235-C variant from the south to the northeast of Siberia (see the Figure).

The frequency of this allele is 87.0 ± 2.0 % in Central Asia, 90.6 ± 1.7 % in Southern Siberia and 92.9 ± 2.3 % in Northeastern Siberia. Consequently, the population's capacity to metabolize lactose diminishes in a similar geographic direction. It is also noteworthy that the observed differences in the allele frequency of the rs4988235 locus are statistically significant only between the populations of Central Asia and Siberia (both its southern and northeastern parts; $P < 10^{-5}$, Fisher's exact test). Furthermore, the two Siberian populations do not differ from each other ($P = 0.09$). It is important to note that allele frequencies may vary in different samples within the same ethnic group. This is exemplified by the case of the Kazakhs, Buryats, and especially Chukchi (Table 1). In addition to random factors, which are particularly relevant in small sample sizes, admixture with individuals belonging to ethnic groups exhibiting a higher frequency of the rs4988235-T variant may contribute to the observed heterogeneity in allele frequencies.

This was clearly demonstrated in the study of the rs4988235 locus polymorphism in the Nenets, who have been reindeer herders for generations and who practically do not drink milk (Khabarova et al., 2012). It appears that this is primarily due to a high prevalence of lactose intolerance. The frequency of the rs4988235-C variant in the Nenets with all four grandparents of Nenets origin is 92.7 % (the frequency of the rs4988235-CC genotype is 90 %). Concurrently, the prevalence of the rs4988235-C allele among the Nenets with at least one relative of Nenets origin has declined to 73 % (Khabarova et al., 2012). The majority of interethnic marriages with Nenets involve Komi and North Russians, in whom the frequency of the rs4988235-T allele is 35–42 % (Khabarova et al., 2011). A certain decrease in hypolactasia among the populations of the northernmost regions of Europe and Siberia can be attributed to intermarriage contacts with immigrant populations of Eastern European origin, which commenced in the 17th century in conjunction with the expansion of the Russian pioneers and reached its peak during the Soviet period.

The duration and intensity of contacts with Eastern Europeans were evidently greater on the territory of Southwestern Siberia and Central Asia, taking into account the migrations of the populations of the Eastern European steppes during the



Distribution of the rs4988235-C variant in regional groups of Siberia and Central Asia.

The average frequency of the genetic variant (in %) and the limits of the standard deviation of the frequency are shown.

Bronze Age. As previously stated, paleogenomic data indicates that the rs4988235-T mutation emerged approximately 6 kya in the ancient population of the northern Black Sea coast, and later this variant of the polymorphism spread throughout northern Eurasia, from Spain to Kazakhstan (Segurel et al., 2020; Irving-Pease et al., 2024). Furthermore, in Europe, the increased frequency of the rs4988235-T variant, which determines stable lactase activity required for milk digestion in adults, was favored by positive selection due to vitamin D deficiency at high latitudes and the need for increased calcium intake from milk (Kozlov, Vershubskaya, 2017).

A review of paleogenomic data in the AADR database (Allen Ancient DNA Resource, https://reich.hms.harvard.edu/) revealed that the first documented instances of the rs4988235-T allele were observed in ancient European populations, including those in Ukraine (~6 kya), Ireland (5.5 kya), and from 4.5 kya and later in Lithuania, Germany, Czech Republic, Estonia, and Russia. In East Asia, the rs4988235-T allele was first identified in an individual from the Botai archaeological culture in northern Kazakhstan (5.3 kya). In Central Asia (in the ancestors of the Kazakhs, Kyrgyz, Mongolians, Turkmens, Uzbeks, and Tajiks), the frequency of the rs4988235-T variant was 4.2 % between 0.5 and 5.3 kya. According to modern data (Table 1), its value is estimated to be approximately 13 %. It is hypothesized that this mutation was already prevalent in Central Asia with a frequency of approximately 5 % since the Iron Age (Segurel et al., 2020). Consequently, if this variant of genetic polymorphism was subject to strong selective pressure, it should have had sufficient time to reach high frequencies in modern populations. This is estimated by L. Segurel et al. (2020) to be 51 %. However, this was not the case, which leads to the reasonable conclusion that the rs4988235-T variant did not experience significant selective pressure in Central Asian populations, in contrast to Europeans and some populations in Africa and the Middle East (Segurel et al., 2020).

The AADR database indicates that in the ancient populations of Siberia and the Urals (between 0.5 to 10 kya), the rs4988235-T variant was distributed with a frequency of 1.8 %, although only in the westernmost region of this territory. All cases of this allele were registered around 3.1–3.8 kya in representatives of the Karasuk (Southwestern Siberia) and Sintashta (Southern Urals) archaeological cultures. The mean

Table 1. Frequency of the rs4988235-C allele in North Asian populations

Ethnic group	Sample size	Frequency of the rs4988235-C allele	References
Tajiks (Uzbekistan, Tajikistan)	254	0.82	Segurel et al., 2020
Uzbeks (Uzbekistan)	45	0.76	Segurel et al., 2020
Turkmens (Uzbekistan)	50	0.80	Segurel et al., 2020
Kazakhs (China)	94	0.95	Sun et al., 2007
Kazakhs (Uzbekistan)	83	0.83	Heyer et al., 2011
Kazakhs (Uzbekistan)	159	0.79	Segurel et al., 2020
Kazakhs (Kazakhstan)	34	0.88	Sokolova et al., 2007
Altaian Kazakhs	128	0.91	Pilipenko et al., 2016
Kyrgyz (Kyrgyzstan)	201	0.88	Segurel et al., 2020
Karakalpaks (Uzbekistan)	45	0.93	Segurel et al., 2020
Uyghurs (Kazakhstan)	30	0.95	Sokolova et al., 2007
Mongolians (China)	82	0.98	Sun et al., 2007
Mongolians (Mongolia)	32	0.88	Segurel et al., 2020
Manchurians (China)	75	1.0	Sun et al., 2007
Orochen (China)	45	0.99	Sun et al., 2007
Nanai (China)	77	1.0	Sun et al., 2007
Southern Altaians	24	0.85	Cardona et al., 2014
Southern Altaians	62	0.92	Segurel et al., 2020
Northern Altaians	29	0.93	Segurel et al., 2020
Shorians	24	0.94	Cardona et al., 2014
Shorians	29	0.90	Segurel et al., 2020
Khakasians	29	0.86	Segurel et al., 2020
Khakasians	64	0.92	Pilipenko et al., 2016
Buryats	78	0.95	Sokolova et al., 2007
Buryats	24	0.98	Cardona et al., 2014
Buryats	28	0.82	Segurel et al., 2020
Yakuts	22	0.93	Cardona et al., 2014
Yakuts	55	0.95	Liebert et al., 2017
Yakuts	25	0.94	Bersaglieri et al., 2004
Western Evenki	24	0.96	Cardona et al., 2014
Evens	24	0.96	Cardona et al., 2014
Koryaks	25	0.96	Cardona et al., 2014
Chukchi	35	0.94	Borinskaya et al., 2006
Chukchi	14	0.75	Cardona et al., 2014
Eskimo	19	0.97	Cardona et al., 2014

frequency of the rs4988235-T variant in the contemporary indigenous population of Siberia is 0.8 % (Table 1). This suggests that, over the past 3,000 years, the frequency of the allele responsible for the enhancement of lactase enzymatic activity has remained unchanged in Siberian populations, despite changes in dietary habits and increased consumption of dairy products. This evidence indicates that in Siberian populations, the rs4988235-T allele behaves as a neutral variant of genetic polymorphism.

In light of the possibility of additional variants of the enhancer polymorphism of the *MCM6* gene in the East Asian population, it is worth noting that this kind of screening was

Table 2. Frequencies of the rs4988235-T and rs182549-A alleles in Siberian populations (according to Cardona et al., 2014)

Ethnic group	Sample size, *N*	Frequency of the rs4988235-T allele	Frequency of the rs182549-A allele
Southern Altaians	24	0.15	0.15
Shorians	24	0.06	0.06
Buryats	24	0.02	0.02
Yakuts	22	0.07	0.07
Western Evenki	24	0.04	0.04
Evens	24	0.04	0.04
Koryaks	25	0.04	0.04
Chukchi	14	0.25	0.25
Eskimo	19	0.03	0.05

performed for several loci, including rs41525747, rs41380347, rs869051967, rs145946881, and rs182549 (Xu et al., 2010; Liebert et al., 2017; Anguita-Ruiz et al., 2020). The enhancer element of the *LCT* gene was also investigated in two indigenous populations of Southern Siberia – the Altaian Kazakhs and Khakasians (Pilipenko et al., 2016). Nevertheless, the frequencies of alleles potentially associated with LP were generally quite low. The sole exception to this is the rs182549 locus. In some East Asian populations, it has been reported that the rs182549-A allele is more informative than rs4988235-T. This is due to the observed occurrence of the rs182549-A allele in the absence of the rs4988235-T allele (Sun et al., 2007; Mattar et al., 2010; Xu et al., 2010). A similar conclusion has been reached for some populations of African, European, and West Asian origin (Bersaglieri et al., 2004; Coelho et al., 2005; Raz et al., 2013). However, this is at odds with the previous conclusion that a complete linkage disequilibrium exists between the rs4988235-T and rs182549-A alleles (Enattah et al., 2002; Troelsen et al., 2003), which included East Asian populations (Kato et al., 2018).

The data on the frequency of distribution of rs4988235-T and rs182549-A variants in ethnic groups of Siberia (Cardona et al., 2014) also indicate that these alleles are linked. The only exception is the Eskimo group, where the frequency of the rs182549-A allele exceeds that of the rs4988235-T allele (Table 2). Therefore, it is highly unlikely that the rs182549-A allele is responsible for maintaining lactase persistence in the indigenous Siberian population.

Polymorphic variants (including archaic ones) of the *LCT* and *MCM6* genes in indigenous Siberians

The results of the polymorphism analysis of exons and adjacent introns of the *MCM6* and *LCT* genes in the indigenous populations of Siberia clearly indicate the existence of polymorphic variants potentially related to lactose metabolism in East Asian populations. Table 3 presents data on the distribution of *LCT* and *MCM6* gene polymorphisms in 102 individuals from various regions of Siberia. These include the in digenous populations of Northeastern (Eskimo, Chukchi, Koryaks), Central (Evens, Evenki, Yakuts), Southern (Tuvinians, Shorians, Altaians, Buryats), and Western (Kets, Khanty, Mansi, Selkups, and Nenets) Siberia. The data were obtained from a full-genome variability study (Pagani et al., 2016). The *LCT* gene contains 21 polymorphic loci, while the *MCM6* gene has seven. The majority of polymorphic variants identified in indigenous Siberians belong to alleles that are commonly found in both East Asian and European populations. Rare variants characteristic only of the East Asian population were found in the rs201668742, rs144864087, and rs3739021 loci. Similarly, variants characteristic only of Europeans were revealed in the rs34307240 locus.

However, it is noteworthy that a group of polymorphic variants in the rs79023654 locus of the *LCT* gene and the rs4988285, rs2070069, rs3087353, and rs2070068 loci of the *MCM6* gene (highlighted in bold in Table 3) warrant further investigation. The rs79023654, rs4988285, and rs2070069 loci are located in the noncoding region of the genes. The rs3087353 and rs2070068 loci are situated within exons, yet nucleotide substitutions within them fail to result in amino acid substitutions. The alleles listed in Table 3 are linked in both indigenous Siberians and other East Asian populations, including Japanese, Koreans, and Vietnamese (Tables 3 and 4). In individuals from Siberia, all these alleles are known to be present in the rs4988235-CC genotype, which is associated with primary hypolactasia. The rs79023654 locus of the *LCT* gene is located at a distance of ~29.7 thousand nucleotide pairs from the rs4988285 locus of the *MCM6* gene. The polymorphic loci within the *MCM6* gene are located at a distance of ~26.5 thousand nucleotide pairs from each other. Furthermore, the rs4988285 and rs2070069 loci are located in the vicinity of the enhancer that regulates the activity of the *LCT* gene. This suggests the potential functional significance of the identified haplotype's polymorphic variants.

The analysis of the dbSNP data (https://www.ncbi.nlm.nih.gov/snp/) indicated that the rs79023654-A, rs4988285-A, rs2070069-G, and rs3087353-T variants were characteristic of the East Asian population and were observed with a low frequency (approximately 1 %) in the South Asian populations (Table 4). However, the fifth allele from this group, rs2070068-A, was detected with a relatively high frequency (24.7 %) in African populations (Table 4). From this distribution, it can be concluded that the East Asian haplotype rs79023654-A, rs4988285-A, rs2070069-G, and rs3087353-T

Table 3. Polymorphic variants of exons and adjacent introns of the *LCT* and *MCM6* genes and their frequency (in %) in Eurasian populations

Polymorphic variant	Gene	NES (*N* = 25)	CS (*N* = 29)	SS (*N* = 28)	WS (*N* = 20)	EAS	EUR
rs62170085-G	*LCT*	6.0	0	0	2.5	0.26	2.84
rs1042712-C	*LCT*	0	6.9	14.3	10.0	21.7	19.4
rs2278544-G	*LCT*	60.0	51.7	64.3	67.5	43.3	68.6
rs3213890-A	*LCT*	0	6.9	14.3	10.0	20.1	19.5
rs2322659-C	*LCT*	62.0	58.6	64.3	70.0	45.6	66.5
rs2304371-G	*LCT*	0	8.6	21.4	12.5	22.6	23.3
rs3739022-A	*LCT*	36.0	32.8	12.5	20.0	21.5	13.9
rs201668742-T	*LCT*	0	5.2	0	0	0.03	0
rs144864087-C	*LCT*	4.0	6.9	7.1	5.0	1.03	0
rs79023654-A	***LCT***	**4.0**	**10.3**	**10.7**	**7.5**	**16.2**	**0**
rs35093754-C	*LCT*	0	1.7	7.1	2.5	5.04	2.92
rs6719488-T	*LCT*	60.0	48.3	57.1	62.5	39.7	62.6
rs2322812-G	*LCT*	36.0	32.8	12.5	17.5	21.5	13.9
rs2874874-C	*LCT*	36.0	32.8	12.5	17.5	21.5	13.9
rs7579771-A	*LCT*	40.0	51.7	44.6	37.5	60.4	37.4
rs2164210-C	*LCT*	60.0	48.3	55.4	62.5	39.7	62.6
rs60376570-A	*LCT*	36.0	32.8	12.5	17.5	21.5	13.9
rs3816088-C	*LCT*	0	1.7	7.1	2.5	5.04	3.0
rs3754689-T	*LCT*	4.0	19.0	26.8	20.0	37.7	20.2
rs2236783-A	*LCT*	54.0	48.3	51.8	57.5	37.01	62.7
rs34307240-A	*LCT*	2.0	0	1.8	0	0	0.95
rs4988285-A	***MCM6***	**4.0**	**10.3**	**10.7**	**7.5**	**16.2**	**0**
rs3739021-A	*MCM6*	0	1.7	3.6	2.5	0.17	0
rs3087350-T	*MCM6*	0	1.7	7.1	2.5	5.2	3.0
rs2070069-G	***MCM6***	**4.0**	**10.3**	**10.7**	**7.5**	**16.2**	**0**
rs3087353-T	***MCM6***	**4.0**	**10.3**	**12.5**	**7.5**	**15.7**	**0**
rs2070068-A	***MCM6***	**4.0**	**10.3**	**12.5**	**7.5**	**15.8**	**0**
rs1057031-A	*MCM6*	0	8.6	14.3	7.5	21.3	20.5

Note. Designations: NES – Northeastern Siberia; CS – Central Siberia; SS – Southern Siberia; WS – Western Siberia; EAS – East Asia; EUR – Europe. For Siberian populations, frequencies are given according to Pagani et al. (2016), for East Asia and Europe, according to the dbSNP database.

Table 4. Frequency (in %) of variants rs79023654-A, rs4988285-A, rs2070069-G, rs3087353-T and rs2070068-A in world populations

Region/country	rs79023654-A	rs4988285-A	rs2070069-G	rs3087353-T	rs2070068-A
Europe	0	0	0	0	0
Siberia	8.3	8.3	8.3	8.8	8.8
East Asia	16.2	16.2	16.2	15.7	15.8
Japan	13.9	13.9	13.9	13.9	13.9
Vietnam	17.1	17.1	11.2	12.8	13.1
South Korea	18.4	18.5	18.6	18.5	18.5
South Asia	1.1	0.58	0.58	0.58	0.58
Africa	0	0	0	0	24.7

Note. Population frequencies are given according to the dbSNP database; for Siberian populations, according to Pagani et al. (2016).

was formed on the basis of ancestral (African) haplotypes, which were characterized by the rs2070068-A variant. However, an analysis of paleogenomic data (AADR database) revealed that the rs2070068-A variant emerged in Africa at a later point in time than in Eurasia. The earliest documented occurrence of this allele in Africa is associated with the northern region of the continent (in the territory of Morocco) at approximately 14.5 kya. Subsequent cases were identified at approximately 9 kya and later. However, it became evident that in Eurasia, this variant of *MCM6* gene polymorphism was observed in both Denisovans and Neanderthals (individuals who lived between ~40 and 110 kya), as well as in numerous most ancient representatives of *Homo sapiens* in Europe and East Asia (aged between ~34 and 44 kya).

Further analysis of paleogenomic sequence databases (Denisova Variants Track Settings; https://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=dhcVcfDenisovaPinky) revealed that the *MCM6* haplotype rs4988285-A, rs2070069-G, rs3087353-T, rs2070068-A was common among Denisovans and Neanderthals. The rs79023654 locus of the *LCT* gene fell into a sequencing region with low coverage. Hence the presence of polymorphism at this locus in Denisovans and Neanderthals remains uncertain.

The results obtained thus suggest that the *MCM6* haplotype detected in the population of East Asia (and, to a much lesser extent, South Asia) is archaic. It is possible that this haplotype was inherited from the common ancestor of *H. sapiens*, Neanderthals, and Denisovans (approximately 600 kya, at the time of the divergence of the ancestor of *H. sapiens* from the ancestors of Neanderthals and Denisovans, as reported by H. Zeberg et al. (2024)). Alternatively, it may have been acquired as a result of hybridization with Neanderthals or Denisovans. Given the distribution of the archaic haplotype in East Asia, it seems more probable that introgression from Denisovans occurred. It has been demonstrated that Neanderthals and Denisovans also exchanged genes – for example, approximately 80–90 kya in southern Siberia (Slon et al., 2018). Consequently, the hypothesis that polymorphic variants were transferred from Denisovans to Neanderthals is also a plausible one.

In recent years, there has been a considerable amount of work done to catalogue archaic variants of genetic polymorphisms that have been identified in the gene pool of modern humans (https://bioinf.eva.mpg.de/catalogbrowser), but the incompleteness of this type of information may be dependent on the extent to which populations have been studied (Zeberg et al., 2024). It seems likely that this database will become much more comprehensive as genomic research continues to expand geographically. There are already some interesting findings of rare ancestral polymorphism variants in widely separated populations – for example, identical alleles of a number of genes in the South African Khoisan and the Philippine Aeta (Zeberg et al., 2024).

There is a great deal of information about the genetic variants that modern humans have inherited from Neanderthals. In particular, there is much to be gained from an understanding of the advantages that humans have gained from admixture, in terms of metabolism, sensory function (especially pain perception), immunity (including SARS-CoV-2), and the expression of some genes (Telis et al., 2020; Zeberg et al., 2020; Pairo-Castineira et al., 2021; Haeggström et al., 2022; Zeberg et al., 2024).

Much less is known about the functional manifestations of Denisovan genetic influence. The main examples of such influence are related to adaptation to high altitude and cold conditions. For instance, a ~33,000 bp fragment of Denisovan DNA has been found in Tibetans that encodes the hypoxia-inducible transcription factor EPAS1, which is involved in adaptation to low oxygen levels (Zhang et al., 2021). In Greenland Eskimos, a ~28,000 bp fragment of Denisovian DNA containing the *WARS* and *TBX15* genes has been identified with high frequency – it is believed that these polymorphic variants may play a role in the adaptation of Arctic indigenous peoples to low temperatures (Racimo et al., 2017). It seems plausible to suggest that the archaic haplotype of the *MCM6* gene found in East Asian populations may be used to implement a specific programme for regulating the enzymatic activity of lactase, which is still relevant today. Further studies are needed to gain a deeper understanding of the specific role of this haplotype in regulating lactase activity. These studies should consider a range of factors, including medical genetics, biochemical and physiological aspects.

Conclusion

Thus, the results of the review of the data on the variability of the *LCT* and *MCM6* genes indicate that from ancient times the indigenous populations of Siberia have been characterized by a low frequency of the rs4988235-T variant, which may contribute to the enhancement of the enzymatic activity of lactase. A certain increase in the frequency of this allele over time in the populations of Central Asia and Southwestern Siberia is associated with the eastward expansion of the ancient populations of the Eastern European steppes starting from the Bronze Age (Heyer et al., 2011; Pilipenko et al., 2016; Segurel et al., 2020). However, it seems that the rs4988235-T variant did not reach high frequencies in Central Asian populations, in contrast to Europe. This may suggest that there is no significant selective pressure on this variant of polymorphism in Central Asian populations (Segurel et al., 2020). It is still unclear why different groups of East Asian populations that traditionally consume dairy products have not developed specific variants of genetic polymorphisms for lactose metabolism. One possible explanation is the hypothesis of cultural adaptation of Central Asian populations, including the development of a culture using bacteria to digest lactose during fermentation, which may have contributed to the establishment of specific microflora in the gut (Segurel et al., 2020).

It is also worth noting that some epigenetic mechanisms (mainly DNA methylation) may also be involved in regulating the expression of lactose metabolism genes (Labrie et al., 2016). It has also been suggested that the type of DNA methylation in the enhancer and promoter regions of the *LCT* gene may be a useful indicator of lactase phenotypes, and it appears that epigenetic modifications may play an important role in the regulation of lactase deficiency (Leseva et al., 2018). Thus, both genetic and epigenetic approaches should be used to investigate the functional significance of polymorphic variants potentially associated with LP, including archaic genetic variants, which the present study has shown to still have some prevalence in human populations.

References

Anguita-Ruiz A., Aguilera C.M., Gil Á. Genetics of lactose intolerance: an updated review and online interactive world maps of phenotype and genotype frequencies. *Nutrients.* 2020;12(9):2689. DOI 10.3390/nu12092689

Bersaglieri T., Sabeti P.C., Patterson N., Vanderploeg T., Schaffner S.F., Drake J.A., Rhodes M., Reich D.E., Hirschhorn J.N. Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* 2004;74:1111-1120. DOI 10.1086/421051

Borinskaya S.A., Rebrikov D.V., Nefedova V.V., Kofiadi I.A., Sokolova M.V., Kolchina E.V., Kulikova E.A., Chernyshov V.N., Kutsev S.I., Polonikov A.V., Ivanov V.P., Kozlov A.I., Yankovsky N.K. Molecular diagnosis and frequencies of primary hypolactasia in populations of Russia and neighboring countries. *Mol. Biol.* 2006; 40(6):931-935. DOI 10.1134/S0026893306060124

Cardona A., Pagani L., Antao T., Lawson D.J., Eichstaedt C.A., Yngvadottir B., Shwe M.T.T., Wee J., Romero I.G., Raj S., Metspalu M., Villems R., Willerslev E., Tyler-Smith C., Malyarchuk B.A., Derenko M.V., Kivisild T. Genome-wide analysis of cold adaption in indigenous Siberian populations. *PLoS One.* 2014;9:e98076. DOI 10.1371/journal.pone.0098076

Coelho M., Luiselli D., Bertorelle G., Lopes A.I., Seixas S., Destro-Bisol G., Rocha J. Microsatellite variation and evolution of human lactase persistence. *Hum. Genet.* 2005;117(4):329-339. DOI 10.1007/s00439-005-1322-z

Enattah N.S., Sahi T., Savilahti E., Terwilliger J.D., Peltonen L., Järvelä I. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* 2002;30(2):233-237. DOI 10.1038/ng826

Evershed R.P., Davey Smith G., Roffet-Salque M., Timpson A., Diekmann Y., Lyon M.S., Cramp L.J.E., … Tasić N., van Wijk I., Vostrovská I., Vuković J., Wolfram S., Zeeb-Lanz A., Thomas M.G. Dairying, diseases and the evolution of lactase persistence in Europe. *Nature.* 2022;608(7922):336-345. DOI 10.1038/s41586-022-05010-7

Haeggström S., Ingelman-Sundberg M., Pääbo S., Zeberg H. The clinically relevant CYP2C8*3 and CYP2C9*2 haplotype is inherited from Neandertals. *Pharmacogenomics J.* 2022;22(4):247-249. DOI 10.1038/s41397-022-00284-6

Heyer E., Brazier L., Ségurel L., Hegay T., Austerlitz F., Quintana-Murci L., Georges M., Pasquet P., Veuille M. Lactase persistence in Central Asia: phenotype, genotype, and evolution. *Hum. Biol.* 2011; 83(3):379-392. DOI 10.3378/027.083.0304

Ingram C.J., Elamin M.F., Mulcare C.A., Weale M.E., Tarekegn A., Raga T.O., Bekele E., Elamin F.M., Thomas M.G., Bradman N., Swallow D.M. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum. Genet.* 2007;120(6):779-788. DOI 10.1007/s00439-006-0291-1

Irving-Pease E.K., Refoyo-Martínez A., Barrie W., Ingason A., Pearson A., Fischer A., Sjögren K.G., … Korneliussen T., Werge T., Allentoft M.E., Sikora M., Nielsen R., Racimo F., Willerslev E. The selection landscape and genetic legacy of ancient Eurasians. *Nature*. 2024;625(7994):312-320. DOI 10.1038/s41586-023-06705-1

Kato K., Ishida S., Tanaka M., Mitsuyama E., Xiao J.Z., Odamaki T. Association between functional lactase variants and a high abundance of *Bifidobacterium* in the gut of healthy Japanese people. *PLoS One*. 2018;13(10):e0206189. DOI 10.1371/journal.pone.0206189

Khabarova Y., Tornianen S., Tuomisto S., Järvelä I., Karhunen P., Isokoski M., Mattila K. Lactase non-persistent genotype influences milk consumption and gastrointestinal symptoms in Northern Russians. *BMC Gastroenterol.* 2011;11:124. DOI 10.1186/1471-230X-11-124

Khabarova Y., Grigoryeva V., Tuomisto S., Karhunen P.J., Mattila K., Isokoski M. High prevalence of lactase non-persistence among indigenous nomadic Nenets, north-west Russia. *Int. J. Circumpolar Health.* 2012;71(1):1-6. DOI 10.3402/ijch.v71i0.17898

Kovalenko E., Vergasova E., Shoshina O., Popov I., Ilinskaya A., Kim A., Plotnikov N., Barenbaum I., Elmuratov A., Ilinsky V., Volokh O., Rakitko A. Lactase deficiency in Russia: multiethnic genetic study. *Eur. J. Clin. Nutr.* 2023;77(8):803-810. DOI 10.1038/s41430-023-01294-8

Kozlov A.I., Vershubskaya G.G. D-vitamin status and lactase persistence in European populations (review with the elements of meta-analysis. *Vestnik Moskovskogo Universiteta. Seria XXIII. Antropologia = Moscow University Anthropology Bulletin.* 2017;3:68-75 (in Russian)

Labrie V., Buske O.J., Oh E., Jeremian R., Ptak C., Gasiunas G., Maleckas A., Petereit R., Žvirbliene A., Adamonis K., Kriukienė E., Koncevičius K., Gordevičius J., Nair A., Zhang A., Ebrahimi S., Oh G., Šikšnys V., Kupčinskas L., Brudno M., Petronis A. Lactase nonpersistence is directed by DNA-variation-dependent epigenetic aging. *Nat. Struct. Mol. Biol.* 2016;23(6):566-573. DOI 10.1038/nsmb.3227

Leseva M.N., Grand R.J., Klett H., Boerries M., Busch H., Binder A.M., Michels K.B. Differences in DNA methylation and functional expression in lactase persistent and non-persistent individuals. *Sci. Rep.* 2018;8(1):5649. DOI 10.1038/s41598-018-23957-4

Liebert A., López S., Jones B.L., Montalva N., Gerbault P., Lau W., Thomas M.G., Bradman N., Maniatis N., Swallow D.M. World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection. *Hum. Genet*. 2017;136(11-12):1445-1453. DOI 10.1007/s00439-017-1847-y

Mattar R., Monteiro M., Silva J.M., Carrilho F.J. LCT-22018G>A single nucleotide polymorphism is a better predictor of adult-type hypolactasia/lactase persistence in Japanese-Brazilians than LCT-13910C>T. *Clinics* (*Sao Paulo*). 2010;65(12):1399. DOI 10.1590/s1807-59322010001200030

Olds L.C., Sibley E. Lactase persistence DNA variant enhances lactase promoter activity *in vitro*: functional role as a *cis* regulatory element. *Hum. Mol. Genet.* 2003;12(18):2333-2340. DOI 10.1093/hmg/ddg244

Pagani L., Lawson D.J., Jagoda E., Mörseburg A., Eriksson A., Mitt M., Clemente F., … Thomas M.G., Manica A., Nielsen R., Villems R., Willerslev E., Kivisild T., Metspalu M. Genomic analyses inform on migration events during the peopling of Eurasia. *Nature.* 2016; 538(7624):238-242. DOI 10.1038/nature19792

Pairo-Castineira E., Clohisey S., Klaric L., Bretherick A.D., Rawlik K., Pasko D., Walker S., … Maslove D., Ling L., McAuley D., Montgomery H., Walsh T., Pereira A.C., Renieri A.; GenOMICC Investigators; ISARIC4C Investigators; COVID-19 Human Genetics Initiative; 23andMe Investigators; BRACOVID Investigators; GenCOVID Investigators; Shen X., Ponting C.P., Fawkes A., Tenesa A., Caulfield M., Scott R., Rowan K., Murphy L., Openshaw P.J.M., Semple M.G., Law A., Vitart V., Wilson J.F., Baillie J.K. Genetic mechanisms of critical illness in COVID-19. *Nature*. 2021; 591(7848):92-98. DOI 10.1038/s41586-020-03065-y

Peng M.S., He J.D., Zhu C.L., Wu S.F., Jin J.Q., Zhang Y.P. Lactase persistence may have an independent origin in Tibetan populations from Tibet, China. *J. Hum. Genet.* 2012;57(6):394-397. DOI 10.1038/jhg.2012.41

Pilipenko I.V., Pristyazhnyuk M.S., Kobzev V.F., Voevoda M.I., Pilipenko A.S. Polymorphism of the *LCT* gene regulatory region in Turkic-speaking populations of the Altay-Sayan region (southern Siberia). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2016;20(6):887-893. DOI 10.18699/VJ16.209 (in Russian)

Racimo F., Gokhman D., Fumagalli M., Ko A., Hansen T., Moltke I., Albrechtsen A., Carmel L., Huerta-Sanchez E., Nielsen R. Archaic adaptive introgression in TBX15/WARS2. *Mol. Biol. Evol.* 2017; 34(3):509-524. DOI 10.1093/molbev/msw283

Raz M., Sharon Y., Yerushalmi B., Birk R. Frequency of LCT-13910C/T and LCT-22018G/A single nucleotide polymorphisms associated with adult-type hypolactasia/lactase persistence among Israelis of different ethnic groups. *Gene.* 2013;519(1):67-70. DOI 10.1016/j.gene.2013.01.049

Scott A., Reinhold S., Hermes T., Kalmykov A.A., Belinskiy A., Buzhilova A., Berezina N., … Krause R., Karapetian M., Stolarczyk E., Krause J., Hansen S., Haak W., Warinner C. Emergence and intensification of dairying in the Caucasus and Eurasian steppes. *Nat. Ecol. Evol*. 2022;6(6):813-822. DOI 10.1038/s41559-022-01701-6

Ségurel L., Bon C. On the evolution of lactase persistence in humans. *Annu. Rev. Genomics Hum. Genet.* 2017;8:297-319. DOI 10.1146/annurev-genom-091416-035340

Segurel L., Guarino-Vignon P., Marchi N., Lafosse S., Laurent R., Bon C., Fabre A., Hegay T., Heyer E. Why and when was lactase persistence selected for? Insights from Central Asian herders and ancient DNA. *PLoS Biol.* 2020;18(6):e3000742. DOI 10.1371/journal.pbio.3000742

Slon V., Mafessoni F., Vernot B., de Filippo C., Grote S., Viola B., Hajdinjak M., Peyregne S., Nagel S., Brown S., Douka K., Higham T., Kozlikin M.B., Shunkov M.V., Derevianko A.P., Kelso J., Meyer M., Prüfer K., Pääbo S. The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature.* 2018;561(7721):113-116. DOI 10.1038/s41586-018-0455-x

Sokolova M.V., Vasilyev E.V., Kozlov A.I., Rebrikov D.V., Senkeeva S.S., Kozhekbaeva Zh.M., Lyundup A.V., Svechnikova N.S., Ogurtsov P.P., Khusnutdinova E.K., Yankovsky N.K., Borinskaya S.A. Polymorphism C/T-13910 of the *LCT* gene regulatory region and lactase deficiency in Eurasian populations. *Ekologicheskaya Genetika = Ecological Genetics.* 2007;5:25-34. DOI 10.17816/ecogen5325-34 (in Russian)

Sun H.M., Qiao Y.D., Chen F., Xu L.D., Bai J., Fu S.B. The lactase gene *–13910T* allele can not predict the lactase-persistence phenotype in north China. *Asia Pac. J. Clin. Nutr.* 2007;16(4):598-601

Telis N., Aguilar R., Harris K. Selection against archaic hominin genetic variation in regulatory regions. *Nat. Ecol. Evol.* 2020;4(11):1558-1566. DOI 10.1038/s41559-020-01284-0

Tishkoff S., Reed F., Ranciaro A., Voight B.F., Babbitt C.C., Silverman J.S., Powell K., Mortensen H.M., Hirbo J.B., Osman M., Ibrahim M., Omar S.A., Lema G., Nyambo T.B., Ghori J., Bumpstead S., Pritchard J.K., Wray G.A., Deloukas P. Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet*. 2007;39(1):31-40. DOI 10.1038/ng1946

Troelsen J.T., Olsen J., Møller J., Sjöström H. An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology*. 2003;125(6):1686-1694. DOI 10.1053/j.gastro.2003.09.031

Xu L., Sun H., Zhang X., Wang J., Sun D., Chen F., Bai J., Fu S. The *–22018A* allele matches the lactase persistence phenotype in northern Chinese populations. *Scand. J. Gastroenterol.* 2010;45(2):168-174. DOI 10.3109/00365520903414176

Zeberg H., Dannemann M., Sahlholm K., Tsuo K., Maricic T., Wiebe V., Hevers W., Robinson H.P.C., Kelso J., Pääbo S. A Neanderthal sodium channel increases pain sensitivity in present-day humans. *Curr. Biol.* 2020;30(17):3465-3469.e4. DOI 10.1016/j.cub.2020.06.045

Zeberg H., Jakobsson M., Pääbo S. The genetic changes that shaped Neandertals, Denisovans, and modern humans. *Cell*. 2024;187(5):1047-1058. DOI 10.1016/j.cell.2023.12.029

Zhang X., Witt K.E., Banuelos M.M., Ko A., Yuan K., Xu S., Nielsen R., Huerta-Sanchez E. The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans. *Proc. Natl. Acad. Sci. USA*. 2021;118(22):e2020803118. DOI 10.1073/pnas.2020803118

Conflict of interest. The author declares no conflict of interest.

Received April 26, 2024. Revised May 31, 2024. Accepted June 3, 2024.