

**Baimbetov G.<sup>1</sup>, Baimukhanov N.<sup>2</sup>**

<sup>1</sup>Analyst of the genetic genology of the research project «Shejire DNA»,  
Kazakhstan, Almaty, e-mail: g.baimbetov@list.ru

<sup>2</sup>Head of the research project «Shejire DNA»,  
Kazakhstan, Almaty, e-mail: baimukhan@gmail.com

**HAPLOGROUP C-L1370  
OF THE Y-CHROMOSOME DNA OR «STEPPE BRANCH»  
OF HAPLOGROUP C-M86,  
ITS SUB-BRANCHES IN PHYLOGENETIC TREE AND VARIETY  
OF HAPLOTYPES IN KAZAKH AND OTHER ETHNIC POPULATIONS**

As part of the applied material for the study of historical migration and their relationship with historical events in the territory of central Eurasia, the authors investigated the haplogroup C-L1370 of Y-chromosome DNA. Where through kinship in the direct paternal (male) genetic line, on the basis of this study and early scientific works, the locations of the carriers of its subclades were determined. In this paper, the authors identified the main 6 modal haplotypes, which were investigated by deep sequencing, which allowed us to identify three main genetic branches. Considering the age of subclades and the location of its carriers, the authors put forward a supposed connection with historical migrations that influenced the prevalence of the studied haplogroup.

**Key words:** history, genetic, migration processes, populations.

Баимбетов Г.<sup>1</sup>, Баймұханов Н.<sup>2</sup>

<sup>1</sup>«ShejireDNA» ғылыми-зерттеу жобасының генетикалық генеалогия зерделеушісі,  
Қазақстан, Алматы қ., e-mail: g.baimbetov@list.ru

<sup>2</sup>«ShejireDNA» ғылыми-зерттеу жобасының жетекшісі  
Қазақстан, Алматы қ., e-mail: baimukhan@gmail.com

**Ү-хромосоманың C-L1370 ДНҚ гаплоготы немесе  
C-M86 гаплоготының «Дала тармағы»,  
оның филогенетикалық ағаштың субтармақтары мен  
қазақ және басқа да этникалық популяцияларындағы  
гаплоготтерінің алуандығы**

Тарихи көші-қону саласы, сонымен қатар орталық Еуразия аумағындағы тарихи оқиғалармен байланыстарын зерттеуге арналған қолданбалы мәліметтің аясында, авторлар Y-хромосомасының ДНҚ-асының C-L1370 гаплоготын зерттеген. Осы зерттеу және бұрынғы жарияланған ғылыми мақалаларға сүйеніп, генетикалық аталық (әкелік) жағынан субкладтарды тасымалдаушылардың тұратын жерлері (локациясын) анықталған. Осы зерттеу және бұрынғы жарияланған ғылыми мақалаларға сүйеніп, аталық (әкелік) жағынан субкладтарды тасымалдаушылардың тұратын жерлері (локациясын) анықталған. Осы еңбекте авторлар негізгі 6 модальдік гаплоготтерін айырып алып, терең секвендеу арқылы зерттеп, үш негізгі генетикалық тармақты анықталды. Субкладтардың жасы мен оларды тасымалдаушылардың локациясын ескеріп, авторлар зерттелген гаплоготтың таралуына әсер еткен тарихи көші-қондармен болжалды байланысты ұсынды.

**Түйін сөздері:** тарих, генетика, көші-қон процестері, популяциялар.

Баимбетов Г.<sup>1</sup>, Баймуханов Н.<sup>2</sup><sup>1</sup>Аналитик генетической генеалогии научно-исследовательского проекта «Shejire DNA»,  
Казахстан, г. Алматы, e-mail: g.baimbetov@list.ru<sup>2</sup>Руководитель научно-исследовательского проекта «Shejire DNA»,  
Казахстан, г. Алматы, e-mail: baimukhan@gmail.com**Гаплогруппа C-L1370 ДНК Y-хромосомы или  
«Степная ветвь» гаплогруппы C-M86, ее субветви  
в филогенетическом древе и разнообразии гаплотипов  
в казахской и других этнических популяциях**

В рамках прикладного материала для исследования исторических миграций и их связи с историческими событиями на территории Центральной Евразии авторами была исследована гаплогруппа C-L1370 ДНК Y-хромосомы, где путем родственных связей по прямой отцовской (мужской) генетической линии, на основе данного исследования и ранних научных работ, были определены локации носителей ее субкладов. В данной работе авторы выделили основных 6 модальных гаплотипов, которые были исследованы путем глубокого секвенирования, что позволило определить три главные генетические ветви. Учитывая возраст субкладов и локацию ее носителей, авторы выдвинули предполагаемую связь с историческими миграциями, которые повлияли на распространённость исследуемой гаплогруппы.

**Ключевые слова:** история, генетика, миграционные процессы, популяции.

**Introduction**

Haplogroup C-L1370 is one of the sub-branches of haplogroup C-M86, whose carriers live historically mainly in the steppe zone of eastern Eurasia. The very fact that carriers of this haplogroup spread from the practical territory of Manchuria to eastern Europe attests to historical mass migration processes in which they participated. The main carriers of this haplogroup are the peoples of the Turkic and Mongolian language families. In one of the earlier works by Kharkov V.N., relating to the study of haplogroup C-M86, where in the constructed network based on microsatellite STR-haplotypes, the authors identified two clusters, calling them «eastern» and «western». In the «eastern cluster», haplotypes predominantly belonged to the Tungus language family living in the taiga zone of Eastern Siberia

and adjacent regions, and the «western cluster» included the peoples of the Turkic and Mongolian language families living mainly in the steppe zone (Kharkov V.N., 2012). In this work, the authors gave a definition to the «western cluster», noting the frequent duplication at the DYS19 locus, however, since this particular value is not constant, we can note the variations at the DYS389 (I-II) loci (Table 1), by which the «Western cluster» can be defined. However, it should be taken into account that this definition is also conditional, it is compiled on the basis of most results, and rare exceptions and homoplasia are not excluded. Thus, the STR-haplotypes of C-M86 haplogroup showed a separate group of haplotypes that had common meanings, and as shown by the SNP-based study, markers shared mutations that separated them from the haplotypes of the «eastern cluster», thereby forming the «Steppe Branch».

**Table 1** – Difference of values in certain loci in STR-haplotypes in two clusters

Locus	Eastern cluster	Western cluster
DYS19 (duplication)	Not present	Frequent presence (but not permanent)
DYS389I – DYS389II	Modal haplotype: 13-29 (DYS389II max=30)	Modal haplotype: 14-31 (DYS389II min=30)

The main purpose of this work is to determine the haplotype diversity of C-L1370 haplogroups in the form of their modal forms, and, on the basis of SNP markers, the definition of their

sub-branches in a phylogenetic tree, the age of sub-branches, and the hypothesis of probable migration processes, particularly on Kazakhstan's territory.

## Materials and methods

In this paper, a total of 380 STR-haplotypes (12 markers) were analyzed from different scientific papers and the haplotypes of this study (Supplementary data). All the haplotypes of this study were tested in the commercial laboratory of FamilyTree DNA (hereinafter referred to as FTDNA) and all the haplotypes from other scientific papers were corrected to the organization format of the FTDNA loci. 128 STR-haplotypes with known values at loci in the FTDNA format were used to create the network. The network itself was built through the Network 5.0 program. For the study of single nucleotide polymorphisms in the Y chromosome DNA, a test was used with a deep sequencing analysis of the FTDNA commercial company with their product called «BigY» which has the following parameters (Big-Y Information):

- Targeted Non-recombining Y-DNA sequencing.
- Illumina HiSeq 200.
- 55X to 80X average coverage.
- Around 11.5 to 12.5 million base-pairs of reliably mapped positions of non-recombining Y chromosome.
- Analyzed using Arpeggi genome analysis technology for improved variant calls.

Sixteen samples were used for a deep sequencing analysis by the «BigY», of which 13 samples belong to Kazakhs from different regions (including the territory of Mongolia), 1 Siberian Tatar (Chat), 1 Kalmyk, and 1 Evenk. The study of predominantly samples from the Kazakh ethnos is due to the fact that in the variety of haplotypes of the C-L1370 haplogroup there are almost all types in the gene pool of the Kazakhs, with the exception of one that is inherent in Siberian (Chat) Tatars and possibly to an insignificant part of the Kalmyk ethnic group with similar haplotypes. However, since in the haplotypes of Kalmyks of this type there are no values at the key loci, we can only assume that they belong to this rare type of haplotype and do not exclude homoplasia. Another Kalmyk sample, which went through the deep sequencing test, has a haplotype that is characteristic of Kalmyks and western Mongols (Oirats). This haplotype is also found among Kazakhs in a small mass in different regions of Kazakhstan. To determine the degree of kinship and historical entry into the Kazakh ethnos, a sample of a Kazakh with the same haplotype was tested for the deep analysis. The Evenk sample (from the Baikal region) having the classical modal Tungus haplotype of the «eastern cluster» was

tested to determine the time of the divergence of the steppe branch from the taiga branch (of eastern Siberia), thereby determining the age of the C-L1370 haplogroup. TMRCA in this study are calculated by the method of Adamov D. et al (Adamov D. et al., 2015), which is an improved version of the calculation method by Poznik et al. (Poznik et al., 2013) in terms of selecting the region of coverage in Y chromosome DNA for age calculations. In order to add the main branches of the C-L1370 haplogroup to the geographic map, all 380 STR-haplotypes were used (Supplementary data).

## Results

**STR-haplotypes analysis.** Analyzing different haplotypes of the western cluster, taking into account stable values at some loci, and the mutation frequency in others, we identified 6 main modal haplotypes, of which only one is not found in the gene pool of the Kazakh population. HWCL1 haplotype is found to be practically in the entire steppe zone of eastern Eurasia, where the maximum frequency of its carriers is in western Mongolia and the European part of the Russian Federation (Kalmykia), i.e. its main carriers are the Mongols of the Oirat ethnic group, which also includes the Kalmyks as a separate ethnic group. It occurs with a low frequency in the gene pool of Kazakhs, Kirghiz, Uighurs, and Karakalpaks. Its main difference from other haplotypes is the value «12» at the DYS388 locus and frequent duplication at the DYS19 locus with the «16-17» values. HWCL2 haplotype meets with maximum frequency mainly in the southern Altai region and the north-eastern part of Kazakhstan, in connection with which it would be appropriate to call it the «Altai cluster» of the C-L1370 haplogroup. The maximum frequency falls on the Kazakhs of the Yergenekty-Naiman tribe and people of Southern Altai. Its main differences in the modal form are the values «15-17» at the DYS19 locus, taking into account that the DYS388 locus will be «13», in the non-modal form there are haplotypes with the value of «16» at the DYS19 locus. HWCL3 haplotype according to the analysis was revealed only among Kazakhs of the South Kazakhstan region in a small clan «Kokrek» of the «Dulat» tribe. Its main differences are the values of «23» at the DYS391 locus and «11-12» at the DYS385 locus. HWCL4 haplotype occurs throughout eastern Eurasia, but everywhere it is so with low frequency in different populations. Among the Kazakhs it is found in a small clan «Asan» of the «Tabyn» tribe. Its main difference is the value of «15» at the DYS19

locus and the lack of duplication at this locus, even in a non-modal form. HWCL5 haplotype is the only one that does not occur among the Kazakh population and has a difference at the DYS385 locus with a value of «12-13» with duplication at the DYS19 locus with a value of «16-17». It occurs mainly in the Chat Tatars, which are a small ethnic group of Siberian Tatars. A similar haplotype is found among Kalmyks, but

since haplotypes taken from other scientific papers have a PowerPlex kit that does not include the data of DYS388 locus, there is a possibility that this is a HWCL1 haplotype homoplasia. HWCL6 haplotype is characteristic only of Western Kazakhs, the bulk of which are from the «Alshyn» tribe. This haplotype has a value of «10» at the DYS391 locus and «25» on the DYS390 locus in modal form.

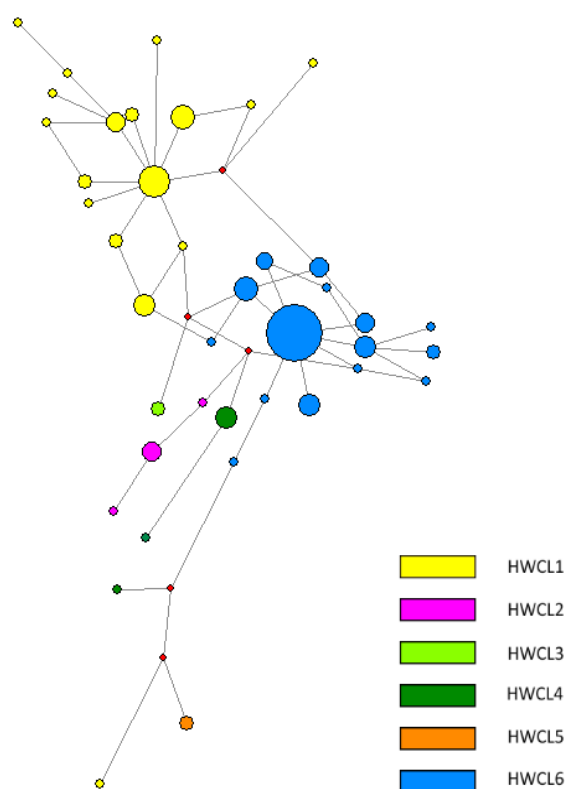
**Table 2** – Modal haplotypes of the C-L1370 haplogroup

	DYS393	DYS390	DYS19	DYS391	DYS385	DYS426	DYS388	DYS439	DYS389I	DYS392	DYS389II
HWCL1	13	24	16-17	9	12-12	11	12	11	14	11	41
HWCL2	13	24	15-17	9	12-12	11	13	11	14	11	41
HWCL3	13	23	16	9	11-12	11	13	11	14	11	41
HWCL4	13	24	15	9	12-12	11	13	11	14	11	41
HWCL5	13	24	16-17	9	12-13	11	13	11	14	11	41
HWCL6	13	25	16	10	12-12	11	13	11	14	11	41

According to the haplotypes described above, taking into account all probable mutations, a network was created where two large clusters of haplotypes were clearly distinguished, which we classified as HWCL1 and the other five. (Figure 1).

#### **Analysis of SNP markers and TMRCA.**

This analysis was based on selected by us modal haplotypes of the C-L1370 haplogroup, by deep analysis of the Y chromosome DNA. TMRCA was calculated for accumulated private mutations. The overall result showed that the C-L1370 haplogroup has three main branches, where the HWCL1 haplotypes form a separate branch of C-SK1066, the HWCL2, HWCL3, and HWCL4 haplotypes form C-F12970 branch, and the HWCL5 and HWCL6 haplotypes form the third branch of C-Y15550. The total time of divergence of the three branches was: 2355ybp (CI 95%: 3096 – 1743ybp). The overall discrepancy between C-L1370 and C-B469 of one of the branches of the «eastern cluster» was 4431 ybp (CI 95%: 5677 – 3398 ybp). (Supplementary data: Table 3: TMRCA)



**Figure 1** – Haplotype network of the C-L1370 haplogroups  
\*HWCL – haplotype of western cluster

**Table 3** – The main branches of the C-L1370 haplogroup, their modal haplotypes, and general TMRCA discrepancies

Main branches of C-L1370	Modal haplotypes	TMRCA
C-SK1066	HWCL1	2355 ybp (CI 95%:3096 – 1743 ybp)
C-F12970	HWCL2, HWCL3, HWCL4	
C-Y15550	HWCL5, HWCL6	

**Table 4** – The main branches of the C-L1370 haplogroup, tested samples, and their TMRCA

Main branches of C-L1370	Samples #	TMRCA
C-SK1066	235724 (western kazakh)	1059 ybp (CI 95%: 1711 -586 ybp)
	K1830 (kalmyk)	
C-F12970	K1960 (mongolian kazakh )	2353 ybp (CI 95%: 3279-1611 ybp)
	277579 (southern kazakh)	
	250592 (western kazakh)	
C-Y15550	249567 (sibirian tatar (Chat))	2114 ybp (CI 95%: 2982-1403 ybp)
	277577 (western kazakh)	
	250586 (western kazakh)	
	K2249 (western kazakh)	
	K2142 (western kazakh)	
	268426 (western kazakh)	
	182517 (western kazakh)	
	K2232 (western kazakh)	
	K2223 (western kazakh)	
	K2239 (western kazakh)	

**Table 5** – The main branches of the C-M86 haplogroup and TMRCA of their discrepancies

Main branches of C-M86	Clusters	TMRCA
C-B469	HECL (Estern cluster)	4431 ybp (CI 95%: 5677 – 3398 ybp)
C-L1370	HWCL (Western cluster)	

### Conclusions

By the revealed genealogical branches it can be concluded that each of them in one way or another relates to the territory of Kazakhstan at different stages of history, from which it is possible, based on the data of TMRCA and according to the population map of other nations, to give

several hypotheses about migratory processes in the carriers of the C-L1370 haplogroup. In the opinion of the authors, taking into account the fact that each sub-branch of the haplogroup under investigation has most likely a separate migration history and is associated with different nations and tribes, the authors considered each sub-branch separately.

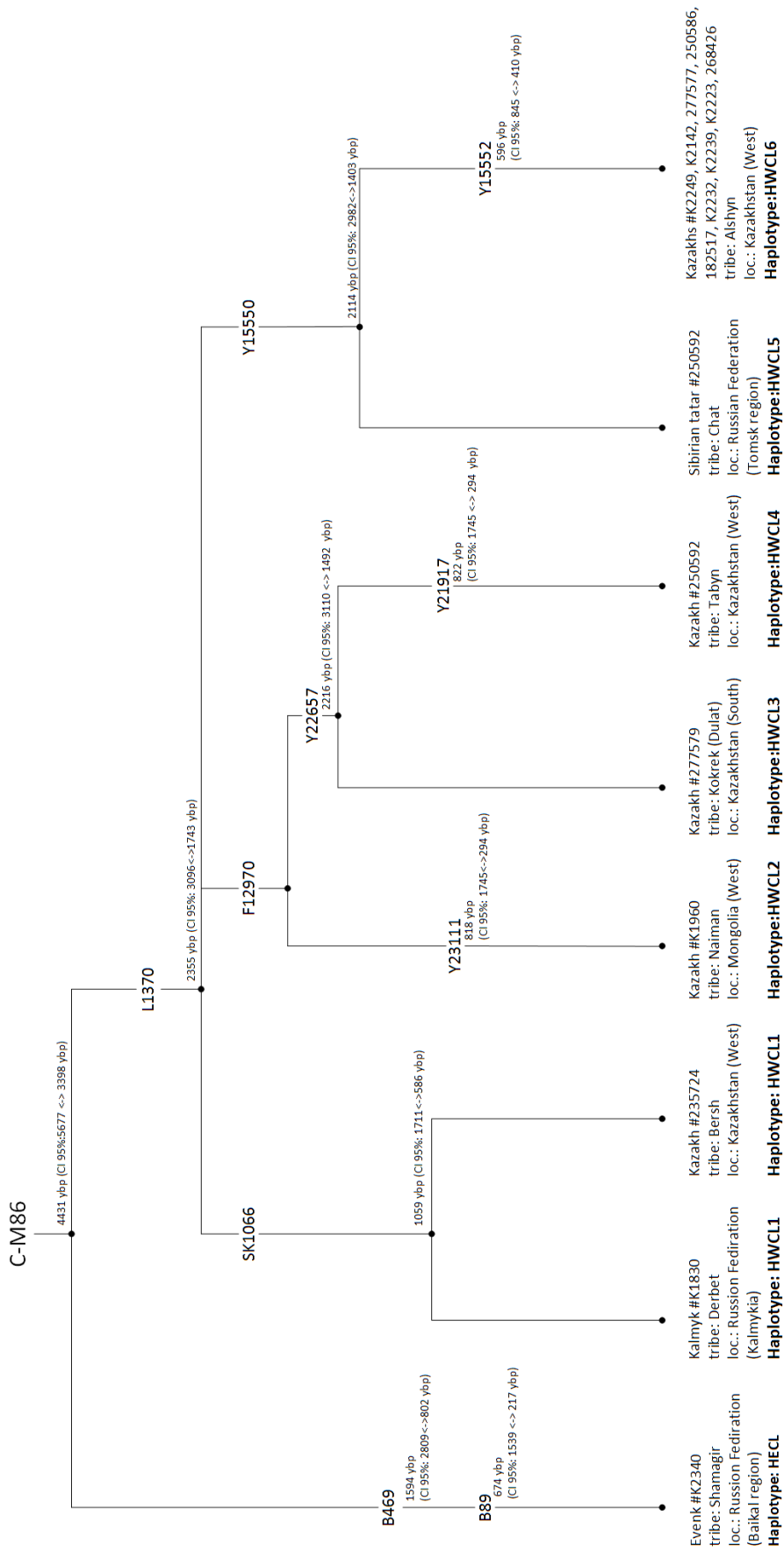
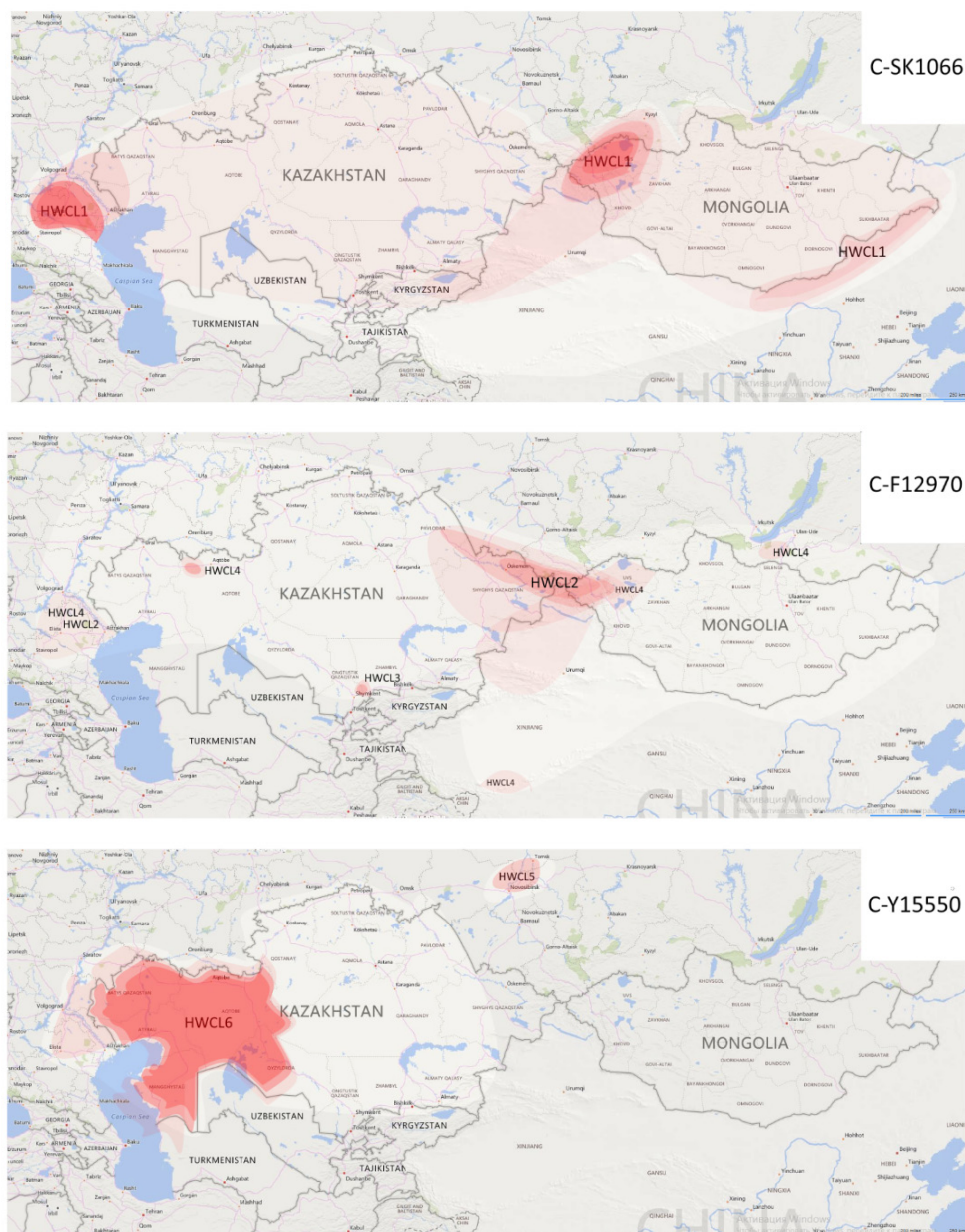


Figure 2 – Phylogenetic tree of tested samples and their TMRCA





**Figure 3** – Settling of the carriers of the three main sub-branches of the C-L1370 haplogroup

SK1066 is the most common branch, which clearly has its haplotype with a value of «12» at the DYS388 locus, is mainly in the western Mongols «Oirat ethnic group», which also includes Kalmyks. The carriers of this polymorphism most likely had two waves of migration to the west. One is connected with the Genghis Khan Mongols invasion (Rashid-ad-Din, 1952) (The Secret Legend of the Mongols, 1941) , the other is linked with the fall of the Junggar Khanate. In this study, attention was drawn to its presence among the Kazakhs, namely one of the Bersh family subdivision, which already existed before the Junggar Khanate.(Usenbaev.

T., 2003) In connection with this, we researched a Kalmyk sample, who is a descendant of the former Junggar Khanate and an above-mentioned sample of the Kazakh tribe with the same haplotype, according to TMRCA, whose common ancestor lived about 1060 years ago; that indicates that its presence in the Kazakh is not connected with the Junggar Khanate, and is most likely associated with the Genghis Khan Mongols invasion, although the early migration of his ancestor to the territory of Kazakhstan is not excluded. Due to the fact that the TMRCA data and the high frequency of the presence of this branch in certain areas indicate that this genealogical branch is

directly connected with Oirat and related tribes that historically were recorded in the Middle Ages as forest people west of Lake Baikal and were related to Mongolian ethnic group (de Bridia, 2002).

F12970 branch has two sub-branches: Y23111 and Y22657, which were divided in the early period and associated with different peoples. Y23111 can be called the «Altai branch» because its main carriers are found among the people of Southern Altai, Altai Kazakhs and Kazakhs of the Naiman tribe living in the north-eastern part of Kazakhstan and western Mongolia. The Y22657 branch does not have a certain territory and is scattered throughout eastern Eurasia, while its young branches indicate that the genealogical line passed through the bottleneck, a period that coincides with the Genghis Khan Mongols invasion, and apparently for the same reason was scattered throughout eastern Eurasia.

Y15550 branch does not found on the territory of Mongolia and is the third separate branch

concerning the above mentioned. It is found in Siberian Tatars, and its Y15552 sub-branch is found in western Kazakhs, mainly among the «Alshin» tribe (Usenbaev. T., 2003). Among the identified haplotypes, there is a likelihood that it exists in Kalmyks, but given that such haplotypes do not occur in western Mongolia, it can be assumed that this is a later infusion into the gene pool of the Kalmyks, after the collapse of the Junggar Khanate, or it is only «homoplasia» as the subclade has not been verified (Supplementary data). Considering the fact that Y15550 representatives are not numerous among Siberian Tatars and the young age of its Y15552 sub-branch (about 600 years), it can be asserted that this branch passed the «bottleneck» most likely during the Genghis Khan Mongols invasion and its absence in the territory of Mongolia indicates its early migration to the territory of Kazakhstan, before the appearance of the Mongols of Genghis Khan time.

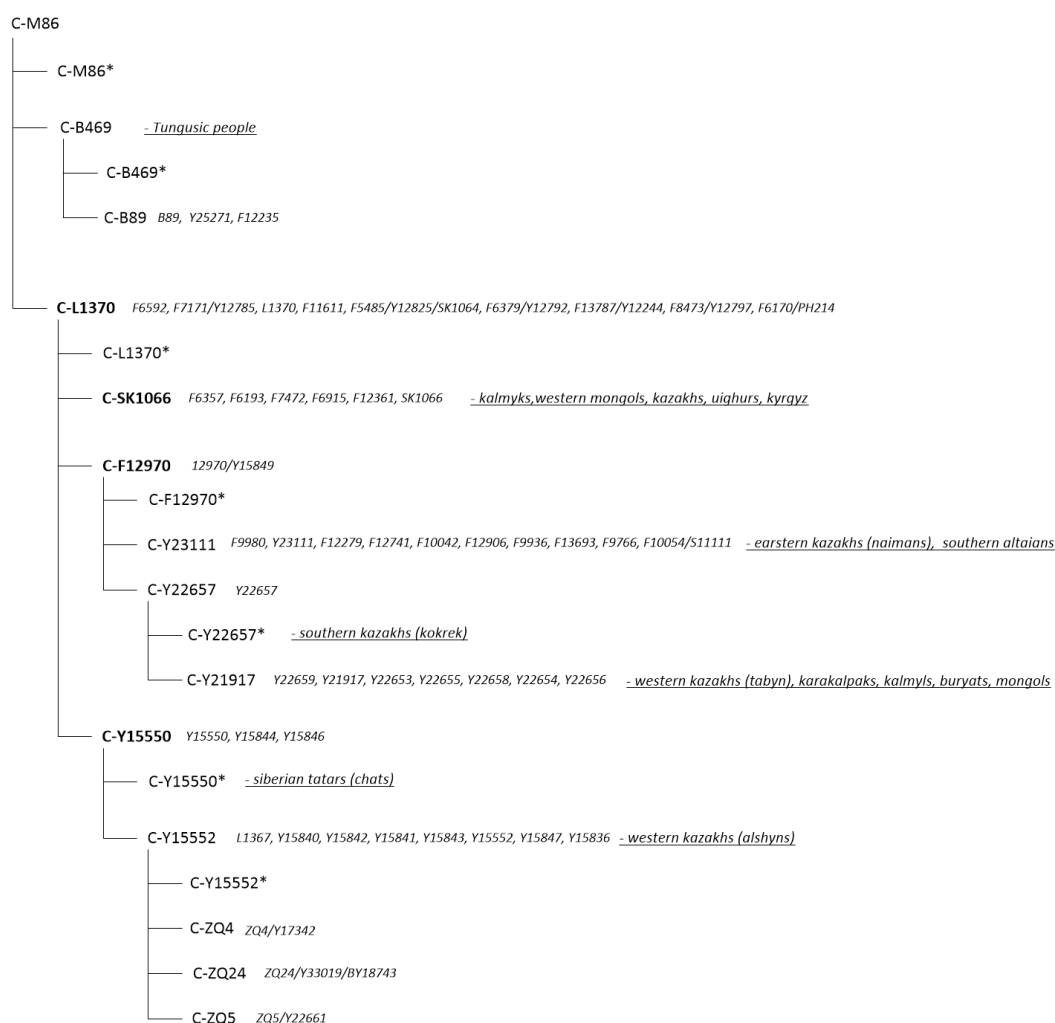


Figure 4 – Phylogenetic tree of haplogroup M86



### References

- Adamov D., Guryanov V., Karzhavin S., Tagankin V., Urasin V.. (2015). Defining a New Rate Constant for Y-Chromosome SNPs based on Full Sequencing Data. *The Russian Journal of Genetic Genealogy: Том 7, №1, ISSN: 1920-2997*  
Big-Y Information <https://www.familytreedna.com/learn/y-dna-testing/big-y/big-y/>  
C-M217 genetic project of FTDNA <http://www.familytreedna.com/public/C3/default.aspx?section=yresults> (Date 10.09.2018).
- de Bridia C. (202). *History of the Tatars. The Christian world and the Great Mongolian Empire. Preparation of the Latin text and the translation of S. V. Aksenov and A. G. Yurchenko. S.-Petersburg, «Eurasia».*
- Kharkov V.N. (2012). Structure and phylogeography of the gene pool of the indigenous population of Siberia and Y-chromosome markers. The author's abstract dissertation on competition of a scientific degree of the doctor of biological sciences. Tomsk. (pages 32-34);
- Poznik D.G., Brenna M. Henn, Muh-Ching Yee, Elzbieta Sliwerska, Ghia M. Euskirchen, Alice A. Lin, Michael Snyder, Lluís Quintana-Murci, Jeffrey M. Kidd, Peter A. Underhill, Carlos D. (2013). Bustamante. Sequencing Y Chromosomes Resolves Discrepancy in Time to Common Ancestor of Males Versus Females. *Science* 341, 562.
- Rashid-ad-Din.(1952). *Collection of Annals. Volume I. The Publishing House of the Academy of Sciences of the USSR. Moscow.*
- The Secret Legend of the Mongols. (1941). Epics, legends and stories. Moscow-Leningrad.*
- Togolukov V.A. (1985). *Tungus (Evenki and Evens) of Middle and Western Siberia. Moscow. «Science» Publishing house.*
- Usenbaev. T. (2003). *Alshyn shezhiresi. Qyzylorda. Tumar.*
- Supplementary data <https://drive.google.com/file/d/18OqmIK-Gqb8qB8TPzuNZKPRybZJTVyYF/view?usp=sharing> (Date 10.09.2018).