

2-1-2012

# The Y-chromosome C3\* star-cluster attributed to Genghis Khan's descendants is present at high frequency in the Kerey clan from Kazakhstan

Serikbai Abilev

*N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 119991 Moscow, Russia*

Boris Malyarchuk

*Institute of Biological Problems of the North, Russian Academy of Sciences, 685000 Magadan, Russia, malyarchuk@ibpn.ru*

Miroslava Derenko

*Institute of Biological Problems of the North, Russian Academy of Sciences, 685000 Magadan, Russia*

Marcin Wozniak

*Forensic Medicine Institute, the Ludwik Rydygier Medical College, the Nicolaus Copernicus University in Torun, 85-094 Bydgoszcz, Poland*

Tomasz Grzybowski

*Forensic Medicine Institute, the Ludwik Rydygier Medical College, the Nicolaus Copernicus University in Torun, 85-094 Bydgoszcz, Poland*

*See next page for additional authors*

---

## Recommended Citation

Open access pre-print, subsequently published as Abilev, Serikbai; Malyarchuk, Boris; Derenko, Miroslava; Wozniak, Marcin; Grzybowski, Tomasz; and Zakharov, Ilya (2012) "The Y-chromosome C3\* star-cluster attributed to Genghis Khan's descendants is present at high frequency in the Kerey clan from Kazakhstan," *Human Biology*: Vol. 84: Iss. 1, Article 12.  
Available at: [http://digitalcommons.wayne.edu/humbiol\\_preprints/13](http://digitalcommons.wayne.edu/humbiol_preprints/13)

This Open Access Preprint is brought to you for free and open access by the WSU Press at DigitalCommons@WayneState. It has been accepted for inclusion in Human Biology Open Access Pre-Prints by an authorized administrator of DigitalCommons@WayneState.

---

**Authors**

Serikbai Aibilev, Boris Malyarchuk, Mirosława Derenko, Marcin Wozniak, Tomasz Grzybowski, and Ilya Zakharov

**The Y-chromosome C3\* star-cluster attributed to Genghis Khan's descendants is present  
at high frequency in the Kerey clan from Kazakhstan**

Serikbai Abilev,<sup>1</sup> Boris Malyarchuk,<sup>2</sup> Mirosława Derenko,<sup>2</sup> Marcin Wozniak,<sup>3</sup> Tomasz Grzybowski,<sup>3</sup> and Ilya Zakharov<sup>1,4</sup>

<sup>1</sup>*N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 119991 Moscow, Russia;*

<sup>2</sup>*Institute of Biological Problems of the North, Russian Academy of Sciences, 685000 Magadan, Russia;*

<sup>3</sup>*Forensic Medicine Institute, the Ludwik Rydygier Medical College, the Nicolaus Copernicus University in Torun, 85-094 Bydgoszcz, Poland;*

<sup>4</sup>*Department of Genetics, Moscow State University, 119991 Moscow, Russia*

*Corresponding author: Dr. Boris Malyarchuk, Institute of Biological Problems of the North, Russian Academy of Sciences, Portovaya str., 18, Magadan, 685000, Russia; e-mail: malyarchuk@ibpn.ru*

**KEY WORDS: Y-CHROMOSOME, HAPLOGROUP C, KEREY CLAN, KAZAKHSTAN**

Running title: Y-chromosome C3\* star-cluster in the Kereys

## **Abstract**

In order to verify the possibility that the Y-chromosome C3\* star-cluster attributed to Genghis Khan and his patrilineal descendants is relatively frequent in the Kereys, who are the dominant clan in Kazakhstan and in Central Asia as a whole, polymorphism of the Y-chromosome was studied in Kazakhs, represented mostly by members of the Kerey clan. The Kereys showed the highest frequency (76.5%) of individuals carrying the Y-chromosome variant known as C3\* star-cluster ascribed to the descendants of Genghis Khan. C3\* star-cluster haplotypes were found in two sub-clans, Abakh-Kereys and Ashmaily-Kereys, diverged about 20-22 generations ago according to the historical data. Median network of the Kerey star-cluster haplotypes at 17 STR loci displays a bipartite structure, with two subclusters defined by the only difference at DYS448 locus. It is noteworthy that there is a strong correspondence of these subclusters with the Kerey sub-clans affiliation. The data obtained suggest that the Kerey clan appears to be the largest known clan in the world descending from a common Y-chromosome ancestor. Possible ways of Genghis Khan's relation to the Kereys are discussed.

A few years ago, Zerjal et al. (2003) have identified a C3\* Y-chromosome lineage frequent in Central Asia (about 8%). They suggested that this lineage originated in Mongolia about 1,000 years ago and it is carried by likely male-line descendants of Genghis Khan. Population screening revealed the highest frequencies of C3\* star-cluster (more than 30%) in Kazakhs, Mongols and Hazaras of Pakistan (Zerjal et al. 2003; Derenko et al. 2007; Zhong et al. 2010), while in Siberia the frequency of star-cluster varies from 18% in Altaian Kazakhs to less than 3% in Tuvinians and Altaians (Derenko et al. 2007; Dulik et al. 2011).

It has been found recently that Kazakhs do not share a common paternal source (Dulik et al. 2011). It is possible that differences between sample sets of Kazakhs were mostly maintained by their tribal affiliations, because Kazakh society is characterized by a complex tribal structure (Dulik et al. 2011). However this possibility was not considered in previous studies of Y-chromosome variation in Kazakh populations (Perez-Lezaun et al. 1999; Wells et al. 2001; Zerjal et al. 2002; Chaix et al. 2007; Malyarchuk et al. 2010; Zhong et al. 2010; Zhong et al. 2011). Our preliminary unpublished data have suggested that C3\* star-cluster haplotypes are relatively frequent in the Kereys, who are the dominant clan in Kazakhstan and in Central Asia as a whole (Atwood 2004; Khoyt 2008). Therefore, to verify the possibility that C3\* star-cluster is a genetic marker of the Kereys we have investigated the Y-chromosome profiles for 17 STR loci in Kazakhs living both in Kazakhstan and Russia and belonging to different tribes, mostly to the Kerey clan.

## **Materials and Methods**

**Samples.** A total of 57 samples of hair roots were collected in males from different Kazakh tribes, including 51 members of the Kereys, with the ethical approval of the Forensic Medicine Institute (Bydgoszcz, Poland) and the Institute of Biological Problems of the North (Magadan,

Russia). All biological samples were collected using informed consent. Genomic DNA was prepared by means of ExtraGene DNA Prep 200 kit (Laboratory Isogen, Moscow).

**Y-chromosome polymorphism typing.** All DNA samples were analyzed at 17 STR loci (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, GATA-H4, DYS448, DYS456, DYS458, DYS635) using AmpF/STR®Yfiler® PCR Amplification kit (Applied Biosystems) according to the manufacturer's instructions. Products of amplification were analyzed on ABI 3100 and ABI 3130 Genetic Analyzers (Applied Biosystems). Electrophoresis results were analyzed using Genscan v. 3.7 and Genotyper v. 3.7 software (Applied Biosystems).

To search for haplotypes belonging to C3\*-paragroup, SNP markers defining the whole haplogroup C (RPS4Y<sub>711</sub>) and its subgroups C1 (M8), C2 (M38), C3 (M217), C3a (M93), C3b (P39), C3c (M77), C3d (M407), C3e (P53.1) and C3f (P62) were assayed using PCR primers summarized in Karafet et al. (2008) as described in Malyarchuk et al. (2010). Thus, C3\* paragroup includes haplotypes which are outside of SNP-defined subgroups C3a-f.

**Data analysis.** Median joining networks of C3\* STR-haplotypes were constructed using the Network 4.6 program (<http://www.fluxus-engineering.com>). For the network construction, STR variants were weighted (with a weight assigned to a range of variance values) following the distribution of the number of mutations at character (Bandelt et al. 2000). The allele sizes for locus DYS389II were determined with the subtraction of DYS389I. In addition, loci DYS385a and DYS385b were excluded from all analyses due to possibility of ambiguous assignment of the alleles to these loci.

The upper bound for divergence time of two groups of haplotypes was calculated as  $T_D$ , assuming STR variance in repeat number at the beginning of population subdivision ( $V_0$ ) equal to zero (Zhivotovsky 2001).

## Results

In the present study, we have studied Y-chromosome variability at 17 STR loci in 57 Kazakhs (Table 1). SNP typing and median network analysis of C3\* STR haplotypes demonstrates that 70% of individuals studied belong to the Y-chromosome star-cluster within paragroup C3\* (Table 2, Figure 1). However, our Kazakh sample is represented mostly by the Kereys, so 76.5% of them are carriers of haplotypes belonging to the C3\* star-cluster (Table 2). Noteworthy that frequency of this cluster is almost 90% in the Kerey-Abakh, whereas it is fairly less in the Kerey-Ashmaily (55%).

Phylogenetic analysis of the Kerey star-cluster haplotypes demonstrates that median network displays a bipartite structure, with two subclusters recognized (Figure 1). One of them is determined by allele DYS448-22, and another one by allele DYS448-23 (Figure 1, Table 3). It is noteworthy that there is a strong correspondence of these subclusters with the Kerey sub-clans affiliation. We have found that all members of the Kereys-Abakh belong to DYS448-23 subcluster, while all but one the Kereys-Ashmaily belong to DYS448-22 subcluster. Unfortunately there is no reliable information on the distribution of DYS448 alleles in Mongol populations. According to the YHRD database ([www.yhrd.org](http://www.yhrd.org)), in the Khalkhs from Mongolia only founder haplotype with DYS448-22 allele is present (4/45 individuals), whereas haplotype with DYS448-23 allele is absent. However, this Mongolian data set is still unpublished, and in addition it cannot be accessible in full from the YHRD. So, it seems that at present we have no possibility to estimate distribution of C3\*-DYS448 alleles in Central Asia and in Mongolia in part.

According to the historical data, the split between two sub-clans of the Kereys occurred about 20-22 generations ago (Khalidullin 2005). Estimation of divergence time ( $T_D$ ) of two groups of 15 STR haplotypes (except for DYS385a,b loci) found in the Kereys sub-clans

demonstrates that  $T_D$  value equal to  $630 \pm 190$  years (or approximately  $21 \pm 6$  generations) is resulted when a mean of per-locus, per-generation mutation rate of 0.0033 and a 30-year generation time are used. Note that similar value of mutation rate (0.00324) has been calculated as optimal for 15 STR haplotypes by Busby et al. (2011) who have investigated the question on how average squared distance (ASD) estimates change within haplotype sets when using different combinations of Y-chromosome STRs. This mutation rate belongs to a class of so called genealogical STR mutation rates revealed by direct observation in father/son pairs (Kayser et al. 2000; Goedbloed et al. 2009). Genealogical rates differ by about factor of three from the evolutionary mutation rate based on microsatellite variation within Y-chromosome haplogroups in the populations with documented short-term histories (Zhivotovsky et al. 2004). Although it is suggested that genealogical rates do not consider the evolutionary consequences of population dynamics (Zhivotovsky et al. 2006), it seems that genealogical rates could be used for age estimations of some phylogenetic subclusters, associated, for instance, with the influence of social selection (Zerjal et al. 2003).

Figure 2 demonstrates a fragment of the median network of C3\* haplotypes based on variability of 7 STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393), as a part of the reference data did not include other STR markers, in different populations of North Eurasia (Table 4), as well as in Kazakhs studied here. As seen, C3\* star-cluster is present in different populations of Central and East Asia, but its highest frequencies are revealed in Kazakhs and Mongols (Table 4) as well as in the Kereys. Therefore, it is possible that C3\* star-cluster of Y-chromosome can be considered as a genetic marker of the Kerey clan.



## **Discussion**

The Keraits, a people related to modern Kereys, inhabited the area of modern Mongolia along the Onon, Kherulen, Selenga and Argun rivers. This area is considered as the motherland of the Keraits (Atwood 2004; Khoyt 2008). During the period of Mongolian tribe consolidation into one state ruled by Temujin (later Genghis Khan), the Keraits were led by Toghrul (Wang Khan) (Man 2004). In alliance with the Keraits, Genghis Khan destroyed the Naimans, and then redirected all the force against his former ally. Not all the Keraits were included in the Genghis Khan's empire – some of them were killed, some fled to the Altai, Central Asia, and the Irtysh and Volga rivers regions.

Consequently, after defeat a part of the Keraits was disseminated among Turkic peoples and tribes. They have been gradually assimilated, preserving a Turkicized form of the ethnonym 'Kerait' as Kerey, Kiray, Karait, Giray, Ghirai and so on. The Keraits were involved as a main element in ethnogenesis of many Turkic peoples, such as Crimean Tatars, Karachays, Nogays, Bashkirs, Kazakhs, Uzbeks, Kyrgyz, Altaians, Tatars (Khoyt 2008). Before the conquest of the Keraits by Genghis Khan, they were a powerful tribe in northern Mongolia and had a highly developed culture. They had a writing system and were Nestorian Christians (Moffett 1992). It is assumed that the Kerait descendants formed a large Kazakh clan of Kereys, which, as noted above, is divided into two sub-clans (Khalidullin 2005). The Kereys-Abakh live in Mongolia, China, Altai, and eastern regions of Kazakhstan Republic. As estimated, about 1.5 millions of Kazakhs live in China, and 20% of them belong to the Kereys and Naimans (Sadovskaya 2001). The Kereys-Ashmaily live in northern Kazakhstan and also in Omsk, Tyumen, Kurgan, and Chelyabinsk regions of the Russian Federation.

The data reported above, taken together with the known arguments in favor of the possible Genghis Khan's descent of Y-chromosome C3\* star-cluster (Zerjal et al. 2003), allow us to suggest two hypotheses.

(1) The star-cluster is not directly related to the descendants of Genghis Khan, but rather is associated with the Kerait clan members. Mongol conquest with participation of the Keraites as special Khan's military forces allowed them to disseminate the Kerait-specific Y-chromosomes in the vast area inhabited by various peoples.

(2) Genghis Khan by himself belonged to the Keraites. This is supported by the following historical evidence (Man 2004; Khalidullin 2005). The Keraites inhabited the banks of the Onon River, where the camp of Genghis Khan's father Yesukhei was located. Yesukhei was declared as a blood brother of the Keraites' Khan Toghrul (Wang Khan). Toghrul then declared Genghis Khan his son-in-law. Fraternization of the Genghis Khan family with the Keraites' Khan suggests that a real blood relationship, though probably not approved officially, existed between them.

These hypotheses can be verified hardly based on currently available data. However, our study demonstrates that the Y-chromosome C3\* star-cluster, early ascribed to Genghis Khan, is a genetic marker of the Kerey clan, which consequently appears to be the largest known in the world clan descending from a common male ancestor.

*Acknowledgements* We are grateful to all the voluntary donors of DNA samples used in this study. The study was supported by the Program of Presidium of Russian Academy of Sciences "Biological diversity" (09-I-P23-10) and the Russian Foundation for Basic Research (11-04-00620).

## Literature Cited

Atwood, C.P. 2004. *Encyclopedia of Mongolia and the Mongol Empire*. New York: Facts on File.

Bandelt, H.-J., V. Macaulay, and M. Richards. 2000. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol. Phylogenet. Evol.* 16:8-28.

Busby, G.B., F. Brisighelli, P. Sánchez-Diz et al. 2011. The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. *Proc. Biol. Sci.*, in press.

Chaix, R., L. Quintana-Murci, T. Hegay et al. 2007. From social to genetic structures in central Asia. *Curr. Biol.* 17: 43-48.

Derenko, M.V., B.A. Malyarchuk, M. Wozniak et al. 2007. Distribution of the male lineages of Genghis Khan's descendants in northern Eurasian populations. *Russ. J. Genet.* 43:422-426.

Dulik, M.C., L.P. Osipova, and T.G. Schurr. 2011. Y-chromosome variation in Altaian Kazakhs reveals a common paternal gene pool for Kazakhs and the influence of Mongolian expansions. *PLoS One.* 6:e17548.

Goedbloed, M., M. Vermeulen, R.N. Fang et al. 2009. Comprehensive mutation analysis of 17 Ychromosomal short tandem repeat polymorphisms included in the AmpFISTR Yfiler PCR amplification kit. *Int. J. Legal Med.* 123: 471-482.

Karafet, T.M., F.L. Mendez, M.B. Meilerman et al. 2008. New binary polymorphisms reshape and increase resolution of the human Y-chromosomal haplogroup tree. *Genome Res.* 18:830-838.

Khalidullin, O.K. 2005. *The Kazakh genealogy. The Keraits*. Almaty: Daik-Press. (in Russian)

- Khoyt, S.K. 2008. *Kereits in ethnogenesis of peoples of Eurasia: historiography of the problem*. Elista: Kalmyk State University Press. (in Russian)
- Malyarchuk, B., M. Derenko, G. Denisova, et al. 2010. Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. *Ann. Hum. Genet.* 74: 539-546.
- Man, J. 2004. *Genghis Khan: Life, Death and Resurrection*. London, New York: Bantam Press.
- Moffett, S.H. 1992. *A History of Christianity in Asia: Beginnings to 1500*. San Francisco: Harper.
- Perez-Lezaun, A., F. Calafell, D. Comas et al. 1999. Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am. J. Hum. Genet.* 65:208-219.
- Sadovskaya, E.Y. 2001. *Migrations in Kazakhstan on the boundary of XXI century: basic tendencies and perspectives*. Almaty: Galym. (in Russian)
- Wells, R.S., N. Yuldasheva, R. Ruzibakiev et al. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA.* 98:10244-10249.
- Zerjal, T., R.S. Wells, N. Yuldasheva et al. 2002. A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am. J. Hum. Genet.* 71:466-482.
- Zerjal, T., Y. Xue, G. Bertorelle et al. 2003. The genetic legacy of the Mongols. *Am. J. Hum. Genet.* 72:717-721.
- Zhivotovsky, L.A. 2001. Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. *Mol. Biol. Evol.* 18:700-709.

Zhivotovsky, L.A., P.A. Underhill, C. Cinnioglu et al. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* 74: 50-61.

Zhivotovsky, L.A., P.A. Underhill, and M.W. Feldman. 2006. Difference between evolutionarily effective and germ line mutation rate due to stochastically varying haplogroup size. *Mol. Biol. Evol.* 23: 2268-2270.

Zhong, H., H. Shi, X.B. Qi et al. 2010. Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. *J. Hum. Genet.* 55:428-435.

Zhong, H., H. Shi, X.B. Qi et al. 2011. Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol. Biol. Evol.* 28:717-727.

**Table 1.** Y-chromosome STR haplotypes in Kazakhs

Sample Code	Clan/ Sub-clan	DYS19	DYS385a	DYS385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	GATA_H4
Gz28	?	16	12	13	13	16	25	10	11	13	14	10	10	22	15	18	21	11
Gz33	?	15	12	14	13	16	24	10	11	13	14	11	12	del	17	16	22	11
Gz49	Argyn	13	13	17	14	15	23	10	12	13	16	10	14	22	16	15	20	11
Gz18	Kerey/?	16	12	13	13	16	25	10	11	13	14	10	10	22	15	18	21	11
Gz19	Kerey/?	16	12	13	13	16	25	10	11	13	14	10	10	22	15	18	21	11
Gz25	Kerey/?	16	12	13	13	16	25	10	11	13	14	10	10	22	15	18	21	11
Gz16	Kerey/Abakh	14	12	13	13	16	23	10	14	13	14	10	10	19	15	17	25	11
Gz17	Kerey/Abakh	15	12	20	14	16	23	11	11	15	14	10	11	21	15	15	21	11
Gz20	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	17	21	11
Gz21	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	16	18	21	11
Gz22	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	16	18	21	11
Gz23	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz24	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz26	Kerey/Abakh	16	12	14	13	16	25	10	11	13	14	10	10	23	15	17	21	11
Gz37	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz38	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz39	Kerey/Abakh	13	13	17	14	15	23	10	12	13	16	10	13	22	16	15	21	11
Gz40	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz70	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	14	18	21	11
Gz71	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz72	Kerey/Abakh	15	12	13	13	16	25	10	11	13	14	10	10	23	15	17	21	11
Gz73	Kerey/Abakh	15	12	13	13	16	25	10	11	13	14	10	10	23	15	17	21	11

Gz74	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz75	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	14	18	21	11
Gz76	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	14	18	21	11
Gz77	Kerey/Abakh	15	12	13	13	16	25	10	11	13	14	10	10	23	15	17	21	11
Gz78	Kerey/Abakh	15	12	13	13	16	25	10	11	13	14	10	10	23	15	17	21	11
Gz79	Kerey/Abakh	15	12	13	13	16	25	10	11	13	14	10	10	23	15	17	21	11
Gz80	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz81	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz87	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	24	15	18	21	11
Gz88	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz89	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz91	Kerey/Abakh	16	12	13	13	16	25	10	11	13	14	10	10	23	15	18	21	11
Gz1	Kerey/Ashmaily	16	12	13	13	16	25	10	11	13	14	10	10	22	15	18	21	11
Gz34	Kerey/Ashmaily	13	13	17	14	15	23	10	12	13	16	10	13	22	16	15	20	11
Gz35	Kerey/Ashmaily	16	12	13	13	16	25	10	11	13	14	10	10	22	15	18	21	11
Gz36	Kerey/Ashmaily	15,16	12	12	14	17	24	9	11	13	14	10	11	20	15	18	24	10
Gz56	Kerey/Ashmaily	16	12	13	13	16	25	10	11	13	14	10	10	22	15	18	21	11
Gz59	Kerey/Ashmaily	15	12	13	13	16	25	10	11	13	14	10	10	22	15	18	19	11
Gz60	Kerey/Ashmaily	16	12	13	14	16	25	10	11	13	14	10	10	22	15	17	21	11
Gz62	Kerey/Ashmaily	16	12	13	13	16	25	10	11	13	14	10	10	22	15	18	21	11
Gz64	Kerey/Ashmaily	16	12	13	13	16	25	10	11	13	14	10	10	22	15	18	21	11
Gz65	Kerey/Ashmaily	16	12	13	13	16	25	10	11	13	15	10	10	22	15	18	21	11
Gz68	Kerey/Ashmaily	16	12	13	13	16	26	10	11	13	14	10	10	22	15	19	21	11
Gz84	Kerey/Ashmaily	16	12	13	13	16	25	10	11	13	14	10	10	22	16	19	21	11
Gz85	Kerey/Ashmaily	16	12	14	14	16	25	10	11	13	14	10	10	23	15	18	21	12
Gz57	Kerey/Ashmaily	13	13	17	14	15	23	10	12	13	16	10	12	22	16	15	20	11
Gz58	Kerey/Ashmaily	15	12	15	13	16	24	10	11	13	14	11	12	del	18	17	21	11
Gz61	Kerey/Ashmaily	15	13	14	12	17	23	10	10.2	14	16	10	11	20	17	17	21	12

Gz63	Kerey/Ashmaily	14	13	16	13	16	24	10	11	12	15	10	11	19	16	20	21	11
Gz66	Kerey/Ashmaily	14	13	20	12	16	23	11	14	12	15	11	12	20	15	20	20	12
Gz67	Kerey/Ashmaily	14	11	13	14	16	23	10	14	14	14	10	10	19	15	17	22	12
Gz69	Kerey/Ashmaily	13	13	17	14	15	23	10	12	13	16	10	13	22	16	15	20	11
Gz55	Naiman	13	12	20	12	16	25	10	13	13	14	10	12	18	14	18	23	12
Gz82	Tore/Kogadai	15	13	14	14	15	23	10	10	14	14	11	11	18	15	17	26	11
Gz83	Tore/Kogadai	15	13	14	14	15	23	10	10	14	14	11	11	18	15	17	26	11

C3\* star-cluster haplotypes are shown in grey. ‘?’ denotes unknown tribal affiliation, ‘del’ - deletion.



**Table 2.** Frequency of the Y-chromosome C3\* star-cluster in Kazakhs based on the present study data

Population	Sample size	C3* star-cluster frequency (%)
Kazakhs	57	40 (70.2)
The Kereys, including:	51	39 (76.5)
the Kereys-Abakh	28	25 (89.3)
the Kereys-Ashmaily	20	11 (55.0)

**Table 3.** Structure of founder haplotypes in the Y-chromosome C3\* subclusters defined by different variants of DYS448 locus

Subcluster	Haplotype (loci order: DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, <b>DYS448</b> , DYS456, DYS458, DYS635, GATA-H4)
DYS448-22	16-12-13-13-16-25-10-11-13-14-10-10- <b>22</b> -15-18-21-11
DYS448-23	16-12-13-13-16-25-10-11-13-14-10-10- <b>23</b> -15-18-21-11

**Table 4.** Frequency of the Y-chromosome C3\* star-cluster in different populations of northern Eurasia

Populations	N	C3* star-cluster (%)	Reference
Kazakhs (from Xingjiang, China)	53	40 (75.5) <sup>1</sup>	Zhong et al. 2011
Kazakhs (from Kazakhstan)	181	73 (40.3) <sup>2</sup>	YHRD data (www.yhrd.org)
Altaiian Kazakhs	36	3 (8.3) <sup>3</sup>	Derenko et al. 2007
Altaiian Kazakhs	119	21 (17.6) <sup>3</sup>	Dulik et al. 2011
Mongols (from Mongolia)	46	16 (34.8) <sup>3</sup>	Derenko et al. 2007
Mongols (from Mongolia)	60	11 (18.3) <sup>2</sup>	YHRD data (www.yhrd.org)
Mongols (Khalkh)	84	14 (16.7) <sup>2</sup>	YHRD data (www.yhrd.org)
Mongols (from Inner Mongolia, China)	22	5 (22.7) <sup>1</sup>	Zhong et al. 2011
Altaiians	89	3 (3.4) <sup>3</sup>	Derenko et al. 2007
Teleuts	44	0 <sup>3</sup>	Derenko et al. 2007
Khakassians	64	0 <sup>3</sup>	Derenko et al. 2007
Shors	38	0 <sup>3</sup>	Derenko et al. 2007
Todjins	26	0 <sup>3</sup>	Derenko et al. 2007
Tuvinians	108	2 (1.9) <sup>3</sup>	Derenko et al. 2007
Yakuts	10	0 <sup>3</sup>	Derenko et al. 2007
Sojots	28	0 <sup>3</sup>	Derenko et al. 2007
Nogays	29	4 (13.8) <sup>3</sup>	Derenko et al. 2007
Buryats	217	5 (2.3) <sup>3</sup>	Derenko et al. 2007
Khamnigans	51	0 <sup>3</sup>	Derenko et al. 2007
Kalmyks	91	1 (1.1) <sup>3</sup>	Derenko et al. 2007
Evenks	41	0 <sup>3</sup>	Derenko et al. 2007

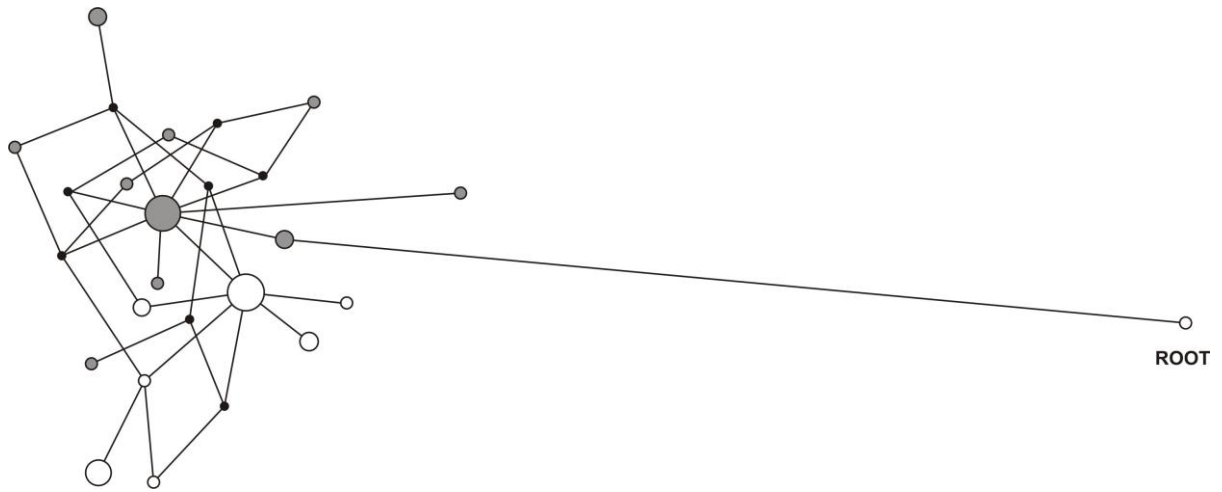
Evans	63	0 <sup>3</sup>	Derenko et al. 2007
Koreans	52	0 <sup>3</sup>	Derenko et al. 2007
Koryaks	39	0 <sup>3</sup>	Derenko et al. 2007
Tibetans	222	5 (2.3) <sup>1</sup>	Zhong et al. 2011
Manchu	109	7 (6.4) <sup>1</sup>	Zhong et al. 2011
Uyghurs	187	6 (3.2) <sup>1</sup>	Zhong et al. 2011

---

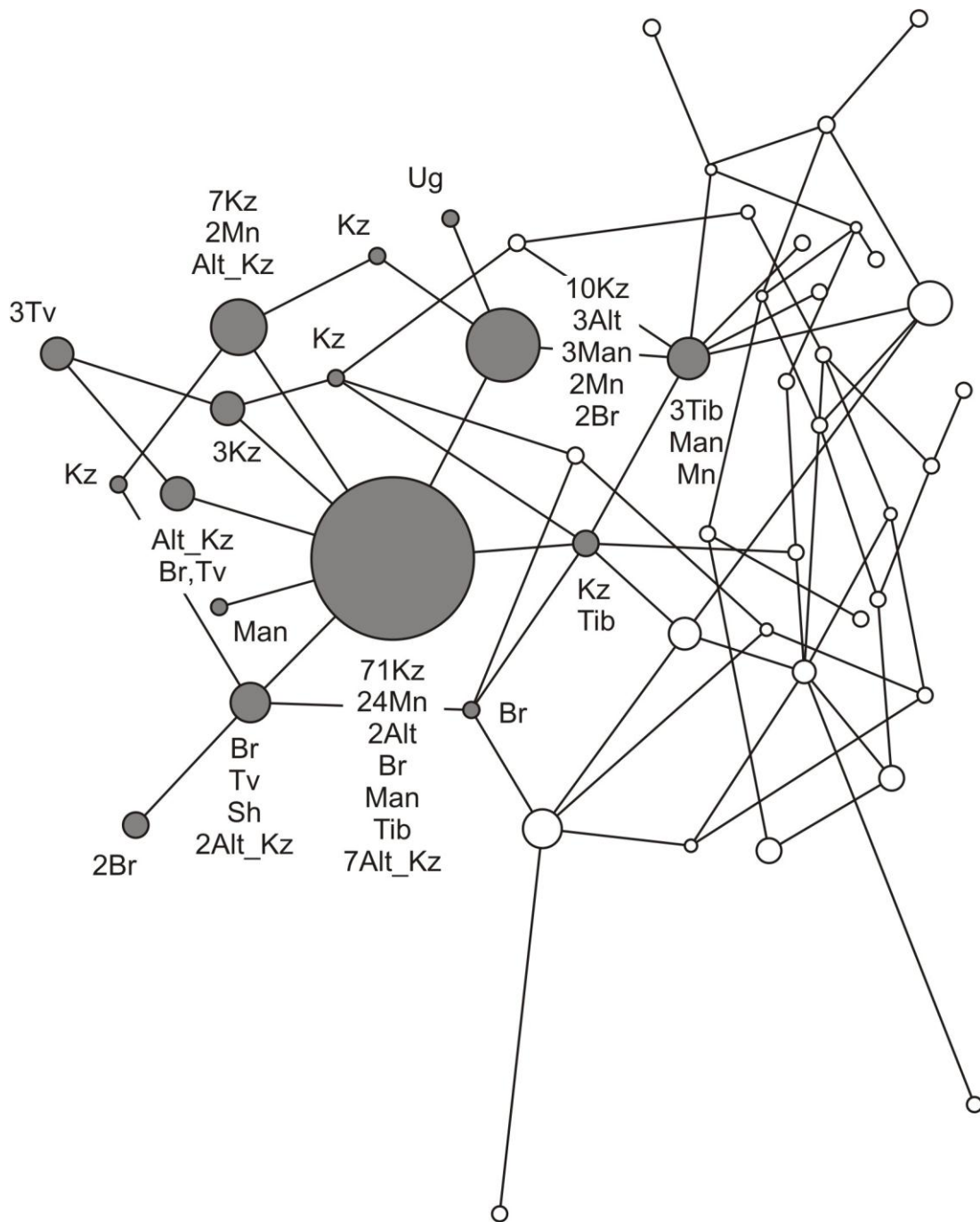
<sup>1</sup> Frequency of star-cluster was estimated based on results of the median network analysis of 7-loci STR haplotypes (DYS19-DYS389I- DYS389II- DYS390- DYS391- DYS392-DYS393).

<sup>2</sup> Frequency of star-cluster was estimated based on frequencies of the central star-cluster STR haplotype 16-12-13-13-16-25-10-11-13 for 9 loci (DYS19-DYS385a- DYS385b- DYS389I- DYS389II- DYS390- DYS391- DYS392-DYS393) and its one-step derivatives revealed in the YHRD 3.0 database ([www.yhrd.org](http://www.yhrd.org); release 37 built at 21 June 2011; 97575 haplotypes within 739 world populations).

<sup>3</sup> Frequency of star-cluster was estimated based on results of the median network analysis of 12-loci STR haplotypes (DYS19- DYS385a- DYS385b-DYS389I- DYS389II- DYS390- DYS391- DYS392-DYS393-DYS437-DYS438-DYS439).



**Figure 1.** Median joining network of the Y-chromosome C3\* star-cluster haplotypes based on 15 STR loci in the Kereys. Each circle represents a haplotype, defined by a combination of STR markers. Circle size is shown proportional to haplotype frequency and the smallest circle represents one haplotype. The lines between circles represent mutational distance, the shortest distance being a single mutational step. ROOT is the outgroup C3c-haplotype used for the network rooting. Haplotypes found in the Kereys-Abakh and the Kereys-Ashmaily are shown in white and grey, respectively. Median vectors are indicated by black points.



**Figure 2.** Fragment of the median joining network of C3\* haplotypes showing position of the central star-cluster haplotype 16-13-16-25-10-11-13 and its one- and two-step derivatives (for 7 STR loci: DYS19- DYS389I- DYS389II- DYS390- DYS391- DYS392-DYS393). Each circle represents a haplotype defined by a combination of STR markers. Haplotypes belonging to C3\* star-cluster are shown in grey. Circle size is proportional to haplotype frequency and the smallest circle represents one haplotype. The lines between circles represent mutational distance, the

shortest distance being a single mutational step. Populations coded as: Kz – Kazakhs, Alt\_Kz – Altaian Kazakhs, Mn – Mongols, Alt – Altaians, Br – Buryats, Man – Manchu, Tib – Tibetans, Ug – Uyghurs, Tv – Tuvinians, Sh – Shors.