

ETHNOGENESIS AND POPULATION GENETICS

**On the origins of the Sakhas' paternal lineages:
Reconciliation of population genetic / ancient DNA data,
archaeological findings and historical narratives**

Tikhonov D.G., Gurkan C., Demirdov K.D., Beyoglu E

Abstract. Sakhas are Turkic-speaking people from Northeastern Siberia, constituting the largest ethnic population in Yakutia. According to popular legends, two heroes who arrived from the Asian Steppe during the late medieval ages, namely Elley Bootur and Omogoy Baay, are the progenitors of all Sakhas. While there is ample historical evidence towards the existence of such legendary characters, archaeological findings and ancient DNA studies provide further insights on actual Sakha ethnogenesis. This study aims to establish the genetic basis of the legendary characters Elley and Omogoy, at least through their paternal lineages, and then to reveal the prevalence of these Y-chromosomes among the contemporary Yakut population. To this end, an attempt was made to delineate fact from fiction with respect to the Sakhas' paternal lineages through a reconciliation of population genetics data on contemporary and ancient Sakhas, along with archaeological evidence and well-recorded historical narratives. To achieve this, 17-loci Y-chromosomal STR and haplogroup analyses were conducted on a contemporary Sakha who was presumably a direct descendant of Elley's paternal line. Furthermore, 367 Sakha Y-chromosomal STR haplotypes were compiled from the literature and elsewhere, and searched against the Y-chromosome STR Haplotype Reference Database to find potential matches with non-Sakha populations. Sakhas' paternal lineages were found to comprise 6 major descent clusters, each corresponding to an ancient clan. The most prevalent haplotype indeed corresponded to that of the contemporary Elley descendant. Furthermore, data presented in the current work suggests a Khitan origin for this paternal line. As shown before, Sakhas' paternal lineages were found to be very homogenous and exhibit signs of a strong population bottleneck. Reconciled genetic and archaeological data agree well with Sakhas' historical narratives, whereby, at least from a paternal lineage perspective, only a few individuals may have arrived from Central Asia and had reproductive success that led to the Sakha Y-chromosomal diversity today.

Key words: Geneology, Elley and Omogoy, haplogroup N, founder effect, Xiongnu, autochthons.

For citations: Tikhonov D.G., Gurkan C., Demirdov K. D., Beyoglu E. On the origins of the Sakhas' paternal lineages: Reconciliation of population genetic / ancient DNA data, archaeological findings and historical narratives // Siberian Research. 2019. 1(1). P. 91 -111. <http://doi.org/10.33384/26587270.2019.01.004e>

Received December 3, 2018; accepted for publication March 01, 2019; published April 15, 2019.

Introduction

Yakutia, or the Sakha Republic of the Russian Federation, spans an immense territory across the Central and Northeastern Siberia, with a land mass of about 3 million km² (Figure 1). According to the 2010 census, among a total population of around

1 million people, Yakuts (or Sakhas) constitute 49.9%, hence the most populous ethnic entity in the republic, followed by Russians (37.8%), Evenks (2.2%), Ukrainians (2.2%), Evens (1.6%) and Tatars (0.9%) [1].

Sakhas have a number of distinguishing ethnographic features over the other Northeastern Si-



Fig 1. A map showing the Russian Federation wherein the Sakha Republic and its capital Yakutsk, the Lena River and Lake Baikal are highlighted.

berian populations, such as their Turkic language, a traditional economy based on semi-nomadic horse- and cattle-breeding, and other customs that bear closer resemblances with those from the Asian Steppe instead. Yet, results from various population genetics investigations of contemporary Sakhas suggest a more complex picture. Sakhas' maternal lineages are heterogeneous and have an admixed origin, featuring close affinities with both the autochthonous Northeastern Siberian populations, such as Evenks, who were traditionally Tungusic-speaking reindeer-herders and hunter-gatherers, and Southern Siberian Turkic-speaking Tuvans [2, 3]. In sharp contrast, Sakhas' paternal lineages are very homogenous, both geographically and chronologically, possibly due to a population bottleneck or founder effect [3 – 6].

According to Sakhas' historical narratives, two legendary characters named Elley Bootur and Omogoy Baay arrived from the Asian Steppe during the late middle ages and revolutionarily changed the way of life in the middle Lena River. Through his monumental work titled *Elleyada*, Gavriel V. Ksenofontov, a well-known Siberian social anthropologist and folklore specialist, reported the story of Elley as it was vividly described in the Sakha oral tradition [7]. Apparently, behind such a choice for

the title of his work lied not only an impression with Homer's *Iliad* and the *Odyssey*, but also a deep fascination with the outcome of the excavations by Schliemann, which had formally confirmed the actual existence of otherwise mythical Troy and Mycenae. In the preface of *Elleyada*, which could only be published posthumously, the famous ethnographer Alexey P. Okladnikov highlighted that "The name in itself of a new publication on Yakut folklore was reminiscent of Homer's stories on the Trojan War. Now all he needed was to find his Troy like Schliemann!" [7].

Based on substantial historical evidence currently available, as well as popular legends that are often based at least in part on actual historical events and at times ornate with fictional anecdotes [8], Elley and Omogoy were not only legendary characters [9]. Some investigators hypothesize that Elley was indeed a real person, perhaps from the Khitan Dynasty [10, 11]. This study attempts to ascertain the genetic evidence towards the existence of the legendary Sakha characters such as Elley and Omogoy through the use of Y-chromosomal short tandem repeat (Y-STR) and haplogroup data on contemporary and ancient Sakhas, along with that of a contemporary descendant of Elley's paternal lineage according to the historical evidence available.

Zerjal et al. [12] were the first to describe a Y-chromosomal base substitution comprising a T→C transition at the RBF5 locus, subsequently named as the Tat C allele, at a very high-level among the Yakut population (86% of those tested). These authors also conducted 9-loci Y-STR analysis on the individuals bearing the Tat C allele, and were the first to reveal the presence of three main haplotypes among the Yakuts. Through Y-chromosomal SNP and STR (9-loci) analyses, Pakendorf et al. [5] also showed the presence of a strong founder-effect among the paternal lineages of the Yakut population. These findings were all later further confirmed

through the archaeogenetic studies, such as by 17-loci Y-STR analysis [6, 13]. Yet, despite all of these findings of immense significance, no attempt has so far been made to link the key Y-STR haplotypes with perhaps known historical personalities.

In the current study, an attempt is thus made for the very first time to establish the genetic affiliation of the main paternal lineages observed among the Yakut population, such as with the legendary Sakha characters of Elley and Omogoy, through the use of Y-STR and haplogroup data on contemporary and ancient Sakhas, along with a putative direct descendant of Elley's paternal lineage. Establishment of

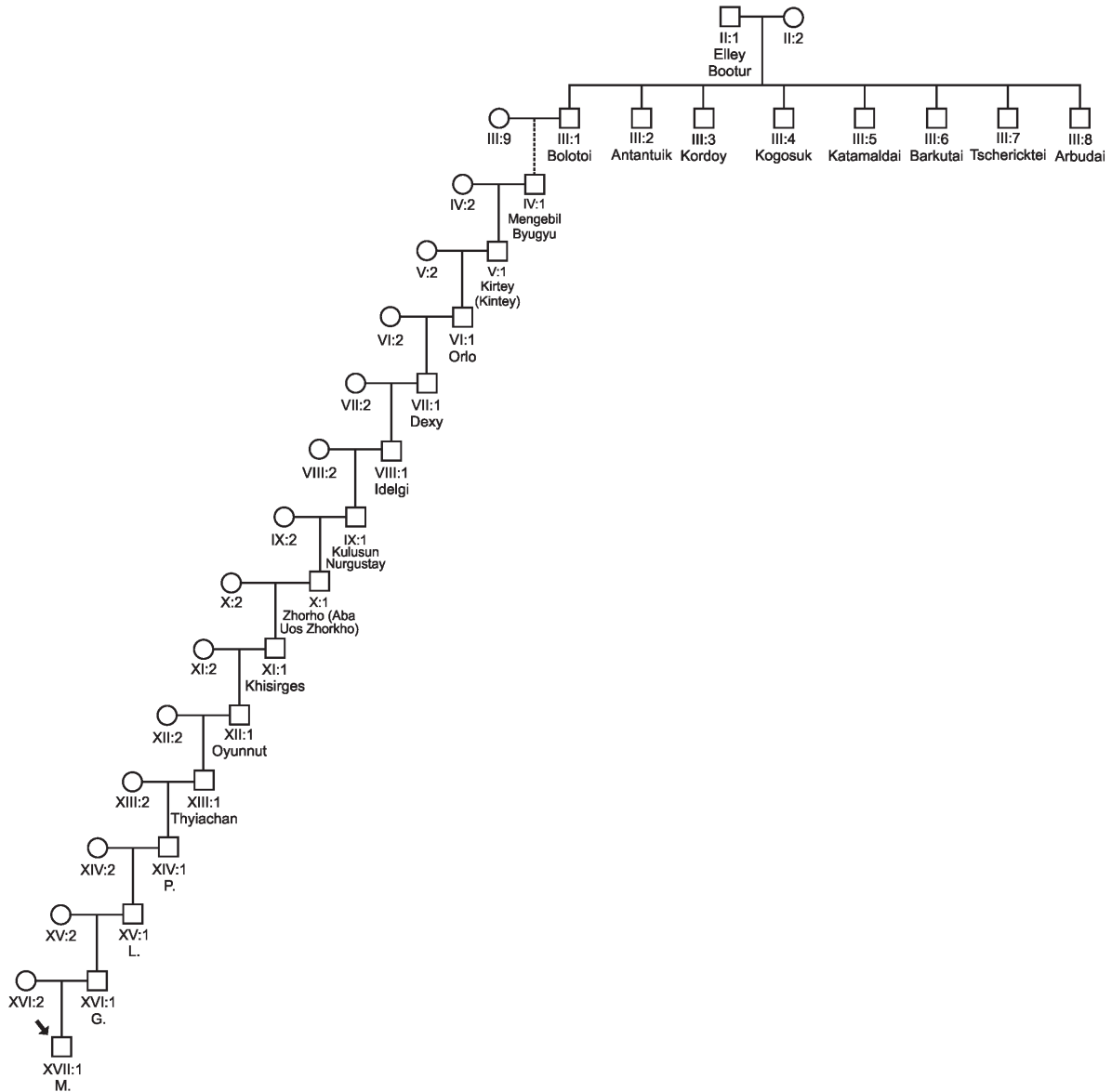


Fig. 2. Megino-Khangalassky branch of the Elley pedigree. Arrow indicates the actual individual sampled for the Y-chromosomal STR and haplogroup analyses.

these genetic connections would allow us to help clarify some important and hitherto unclear aspects on the Yakut ethnogenesis.

Materials and Methods

A putative pedigree focusing on the paternal lineage of the Megino-Khangalassky Ulus of the Elley clan, starting with the progenitor Elley was constructed (Figure 2). The genealogical data were derived from various sources in the literature, historical documents and research from the current study: the immediate descendents of Elley were first described in the 18th century [14], from Bolotoi to Mengebil Byugyu, and further on to Kirtey (Kintey) and then Khisirges was according to anthropological literature [15, 16], from Khisirges to Oyunnut, and further on to Thyiachan was according to anthropological literature and archived documents [15, 17], and finally from Thyiachan down to the last four generations (for whom only the initials for the first names are provided) was according to research from the current study.

A contemporary representative of Elley's paternal lineage corresponding to a 59 year old Sakha was identified in the current study (Figure 2) and a buccal swab sample was collected along with informed consent and in full accordance with the principles of the Declaration of Helsinki by the World Medical

Association. 17 Y-chromosomal STR loci analysis (DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, Y-GATA-H4, DYS437, DYS438, and DYS448; the Life Technologies AmpFLSTR® Yfiler™ Kit) and Y-chromosomal SNP-based haplogroup assignment were carried out at the Gentis Laboratory (Moscow, The Russian Federation).

Table 1 provides the sources for the 367 contemporary Sakha Y-chromosomal STR haplotypes compiled from the literature, YHRD, and the current study [5, 13, 18 – 22]. Table 2 lists the corresponding nine-loci Sakha Y-chromosomal STR haplotypes, their frequencies and the associated Y-chromosomal haplogroup assignments. When Y-chromosomal SNP based haplogroup assignments were not readily available, the online 21-haplogroup version of the Whit Athey *in silico* haplogroup assignment algorithm was used instead (<http://www.hprg.com/hapest5/index.html>) [21].

Each Y-chromosomal STR haplotype listed in Table 2 were queried against Y chromosome STR Haplotype Reference Database (YHRD) for potential matches with those from non-Sakha populations using the 'search the database' option [21]. Accordingly, Table 3 lists the 17-loci Y-chromosomal STR haplotypes corresponding to the three most prevalent Sakha haplotypes observed in the current study and their near-perfect matches with

Table 1. Sakha Y-chromosomal STR haplotypes used in the current study

| Population Data | Number of haplotypes | Source or YHRD Accession No.: |
|----------------------|----------------------|-------------------------------|
| Sakha (Elley's line) | 1 | The present study |
| Sakha | 4 | [18] |
| Sakha | 4 | [19] |
| Sakha | 8 | [22] |
| Sakha | 21 | [20] |
| Sakha | 24 | YA004128 & YA00397 |
| Sakha | 133 | [13] |
| Sakha | 172 | [5] |
| Total | 367 | |

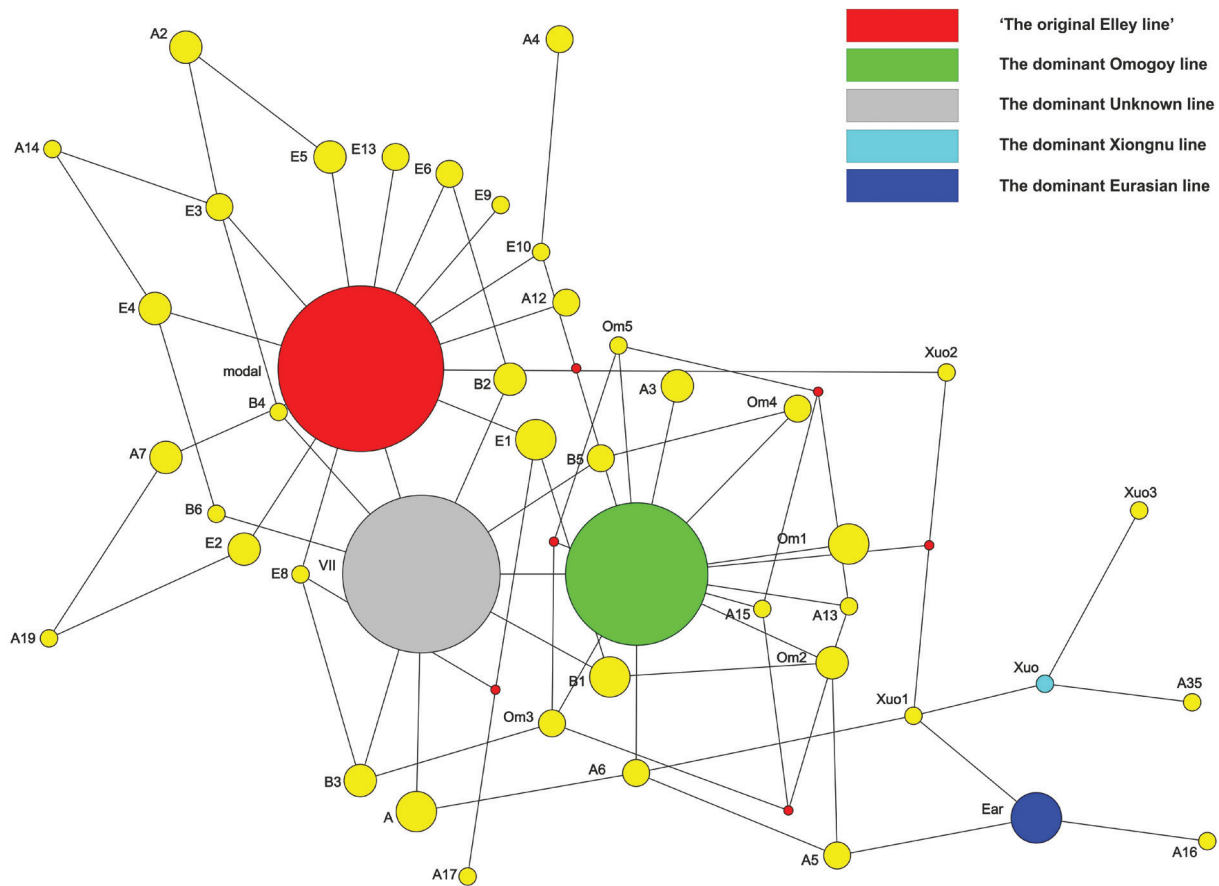


Fig. 3 Nine-loci median-joining network analysis of Sakha Y-chromosomal STR haplotypes belonging to the N haplogroup.

Each haplotype is denoted by a yellow circle, except for the proposed ancestral nodes for each of the five major descent clusters that are shown in different colours according to the legend inset provided. The size of each the node is indicative of the number of haplotypes represented. Median vectors are shown as tiny red circles.

those from non-Sakha population datasets.

Median-joining network analyses were carried out with the full 17-loci (Yfiler), 14-loci [Yfiler, minus DYS385a, DYS386b and DYS390], and nine-loci [DYS389I, DYS389II, DYS19, DYS385a, DYS385b, DYS390, DYS391, DYS392 and DYS393] Y-chromosomal STR datasets using the Network v.5.0.0.0 software (www.fluxusengineering.com). Either the whole network or a specific descent cluster in a given network was analyzed with the Network Time Estimates sub-program of the same software by selecting a proposed central ancestral node and then all the other descendant nodes. Y-chromosomal STR mutation rates of 0.003257, 0.003365 and 0.003307 per locus per generation were used for

the nine-, 14- and 17-loci median-joining network analyses, in that order, and along with a generation time of 32 years [23].

Results

Among the 367 nine-loci Y-chromosomal STR haplotypes corresponding to the contemporary Sakhas (Table 2), 66 different haplotypes were observed, thus affording a discrimination capacity of 17.9%, as well as only 35 unique haplotypes (9.5%). These results are in line with earlier reports on the high homogeneity of the Sakha paternal lineages [4 – 6]. The contemporary representative of Elley's paternal lineage (Ellyayevsky family), a 59 years old

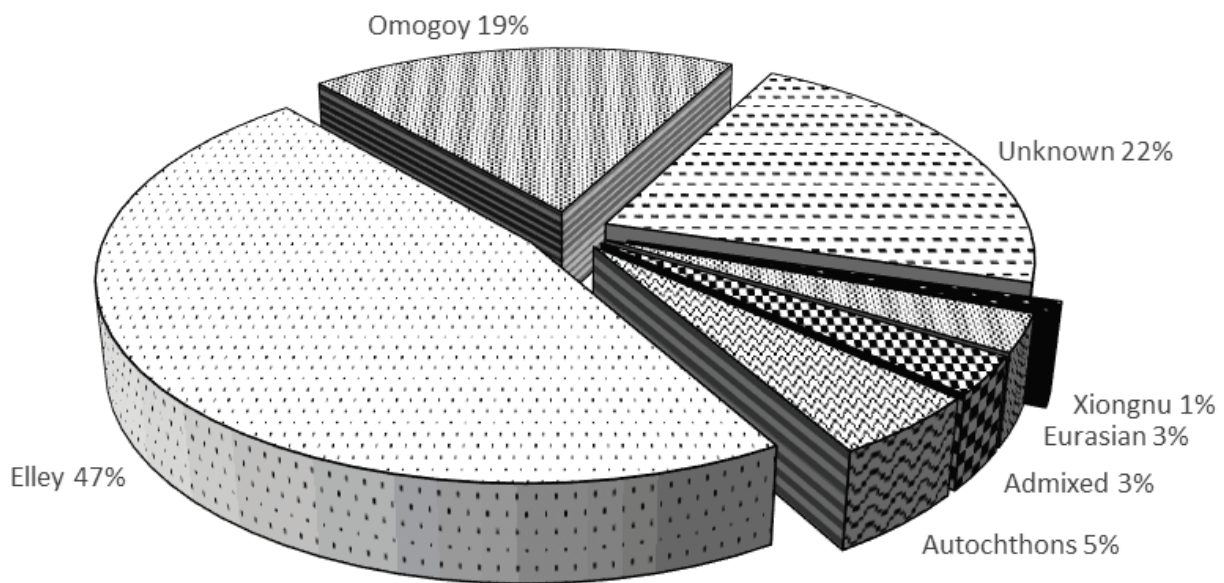


Fig. 4. The distribution of the Sakha Y-chromosomal STR haplotype clusters/groups

male individual (M.) (Figure 2), was found to have a Y-chromosomal haplogroup assignment of N1c1 [defined by the single nucleotide polymorphism (SNP) M178+] and a full 17-loci Y-chromosomal STR haplotype as shown in Table 3. At nine loci resolution, this haplotype corresponded to the most prevalent haplotype (37.3%) among all Sakhas. Taking into consideration the fact that the genealogical paternal lineage of the male M. tested in the current study presumably represented that of the legendary character Elley, it was classified as ‘the original Elley line’. Once again, all generations of the paternal lineage of the male individual M. were documented and confirmed by archival data and publications. The first information on the genealogy of the Elley family was described by the participant of the Academic Detachment of the Second Kamchatka Expedition (Russia) Yakov Lindenau in 1741-1745. The second most prevalent haplotype (16.9%) [13] with unknown origin, was classified as “dominant line Unknown”. The third most prevalent haplotype (13.9%) was supposedly attributed to Omogoy genus by Adamov [24]. Determination of the Y chromosomal STR haplotype to the mummified frozen body (ID: YAKa 69) from the Kous Tcharbyt site, conducted by Crubezy et al. [6] coincides with the third common haplotype of Sakha. The excavated

thomb at the Kous Tcharbyt site, was identified as that of a tojon (clan chief) and dated to the 16th Century/early 17th Century CE, presumably from the Bayaginskii family. According to Aprosimov and Popov [15], Bayaginskii nasleg (sub-district) belonged to the Omogoiisky family, thus this haplotype was classified as ‘the dominant Omogoy line’. Figure 3 depicts the median-joining network analysis for the nine-loci Y-chromosomal STR haplotypes belonging to the N haplogroup among the contemporary Sakhas.

As a direct comparison of the contemporary and ancient Sakha populations, out of the 27 successful Y-chromosomal STR profiles that could be obtained from mummified frozen bodies belonging to Sakhas from the 15th to 19th centuries, the two most prevalent haplotypes that were observed eight (29.6%) and seven (25.9%) times corresponded to ‘the original Elley line’ and the ‘dominant Omogoy line’ from the current study, respectively [6]. When the ancient DNA dataset was eventually increased to comprise data from 62 ancient Sakhas, the 17-loci Y-chromosomal STR haplotype corresponding to that of ‘the original Elley line’ from the current study was found to correspond to as high as 46.8% of all the male lineages analyzed [27].

Table 2. Frequencies of the 9-loci Y-STR haplotypes observed in the Sakha population

| Haplotype ID | DYS389 I | DYS390 | DYS389 II | DYS19 | DYS385a/b | DYS393 | DYS391 | DYS392 | Number of Haplotypes | Haplogroup* | Source |
|-----------------------------|----------|--------|-----------|-------|-----------|--------|--------|--------|----------------------|--|-----------------------|
| Elley Clan | - | - | - | - | - | - | - | - | 173 | N | |
| «the original Elley line» | 14 | 23 | 32 | 14 | 11, 13 | 14 | 11 | 16 | 137 | N1c1 ^C | a, 5, 13, 18 – 20, 22 |
| E1 | 14 | 23 | 32 | 14 | 11, 13 | 14 | 10 | 16 | 4 | N1c1 ^C , N ^b | 5, 13 |
| E2 | 14 | 23 | 32 | 14 | 12, 13 | 14 | 11 | 16 | 3 | N1c1 ^C | 5, 18 |
| E4 | 14 | 23 | 32 | 14 | 11, 13 | 14 | 12 | 16 | 3 | N ^b | 13 |
| E5 | 14 | 23 | 32 | 15 | 11, 13 | 14 | 11 | 16 | 3 | N1c1 ^C | 5, 13 |
| E6 | 14 | 24 | 32 | 14 | 11, 13 | 14 | 11 | 16 | 2 | N1c1 ^C , N1c1a1 ^C , | 19, 22 |
| E3 | 14 | 23 | 32 | 14 | 11, 12 | 14 | 11 | 16 | 2 | N1c1 ^C | 5 |
| E13 | 14 | 23 | 32 | 14 | 11, 13 | 14 | 11 | 17 | 2 | N1c1 ^C | 5, 22 |
| E8 | 14 | 23 | 32 | 14 | 11, 13 | 13 | 11 | 16 | 1 | N ^b | 13 |
| E9 | 14 | 22 | 32 | 14 | 11, 13 | 14 | 11 | 16 | 1 | N1c1 ^C | 5 |
| E10 | 14 | 23 | 32 | 14 | 11, 14 | 14 | 11 | 16 | 1 | N1c1 ^C | 5 |
| E14 | 14 | 23 | 32 | 14 | 11, 13 | 14 | 11 | 14 | 1 | N1c1 ^C | 5 |
| A2 | 14 | 23 | 32 | 15 | 11, 12 | 14 | 11 | 16 | 3 | N ^b | 13 |
| A7 | 14 | 23 | 33 | 14 | 11, 13 | 14 | 11 | 16 | 3 | N1c1 ^C , N ^b | 5, 13 |
| A12 | 15 | 23 | 33 | 14 | 11, 13 | 14 | 11 | 16 | 2 | N1c1 ^C | 5, 13 |
| A4 | 13 | 23 | 31 | 14 | 11, 14 | 14 | 11 | 16 | 2 | N1c1 ^C | 5 |
| A14 | 14 | 23 | 32 | 14 | 11, 12 | 14 | 12 | 16 | 1 | N1c1 ^C | 5 |
| A17 | 14 | 23 | 32 | 14 | 11, 13 | 12 | 10 | 16 | 1 | N ^b | 13 |
| A19 | 14 | 23 | 33 | 14 | 12, 13 | 14 | 11 | 16 | 1 | N ^b | 13 |
| Unknown Clan | - | - | - | - | - | - | - | - | 82 | N | |
| «the dominant Unknown line» | 14 | 23 | 31 | 14 | 11, 13 | 14 | 11 | 16 | 62 | N1c1 ^C , N ^b | 5, 13, YA003971 |

Continuation of table 2

| | | | | | | | | | | | |
|------------------------------|----|----|----|----|--------|----|----|----|-----------|------------------------------------|------------------------|
| B1 | 14 | 23 | 31 | 14 | 11, 13 | 14 | 10 | 16 | 5 | N1c1 ^C , N ^b | 5, 13, YA004128 |
| B2 | 14 | 24 | 31 | 14 | 11, 13 | 14 | 11 | 16 | 3 | N1c1 ^C | 5 |
| B3 | 14 | 23 | 31 | 14 | 11, 13 | 13 | 11 | 16 | 3 | N1c1 ^C , N ^b | 5, 13 |
| B5 | 13 | 23 | 31 | 14 | 11, 14 | 14 | 11 | 16 | 2 | N1c1 ^C | 5 |
| B4 | 14 | 23 | 31 | 14 | 11, 12 | 14 | 11 | 16 | 1 | N1c1 ^C | 5 |
| B6 | 14 | 23 | 31 | 14 | 11, 13 | 14 | 12 | 16 | 1 | N ^b | 13 |
| A | 14 | 23 | 30 | 14 | 11, 13 | 14 | 11 | 16 | 5 | N1c1 ^C , N ^b | 5, 13, 20 |
| Omogoy Clan | | - | - | - | - | - | - | - | 71 | N | |
| «the dominant Omogoy line» | 14 | 23 | 31 | 14 | 11, 13 | 14 | 11 | 15 | 51 | N1c1 ^C , N ^b | 5, 13, 22, YA003971 |
| Om1 | 14 | 23 | 31 | 14 | 12, 13 | 14 | 11 | 15 | 4 | N1c1 ^C , N ^b | 5, 13 |
| Om2 | 14 | 23 | 31 | 14 | 11, 13 | 14 | 10 | 15 | 3 | N1c1 ^C | 5 |
| Om3 | 14 | 23 | 31 | 14 | 11, 13 | 13 | 11 | 15 | 2 | N1c1 ^C | 5 |
| Om4 | 14 | 23 | 31 | 14 | 11, 14 | 14 | 11 | 15 | 2 | N ^b | 13 |
| Om5 | 14 | 23 | 31 | 15 | 11, 13 | 14 | 11 | 15 | 1 | N ^b | 13 |
| A3 | 13 | 23 | 30 | 14 | 11, 13 | 14 | 11 | 15 | 3 | N1c1 ^C , N ^b | 5, 13 |
| A6 | 14 | 23 | 30 | 14 | 11, 13 | 14 | 11 | 15 | 2 | N ^b | 13 |
| A13 | 15 | 23 | 32 | 14 | 11, 13 | 14 | 11 | 15 | 1 | N1c1 ^C | 5 |
| A15 | 15 | 23 | 32 | 15 | 11, 13 | 13 | 11 | 15 | 1 | N1c1 ^C | 5 |
| A22 | 13 | 23 | 30 | 14 | 11, 14 | 14 | 11 | 15 | 1 | N ^b | 13 |
| Eurasian Clan | | - | - | - | - | - | - | - | 10 | N | |
| «the dominant Eurasian line» | 14 | 23 | 30 | 14 | 11, 13 | 14 | 10 | 14 | 7 | N ^b | 13, YA004128 |
| A5 | 14 | 23 | 30 | 14 | 11, 13 | 14 | 10 | 15 | 2 | N1c1 ^C | 5 |
| A16 | 14 | 22 | 30 | 14 | 11, 13 | 14 | 10 | 14 | 1 | N1c1 ^C | 5 |
| Xioungnu line | | - | - | - | - | - | - | - | 4 | N | |

ON THE ORIGINS OF THE SAKHAS' PATERNAL LINEAGES

Continuation of table 2

| | | | | | | | | | | | |
|-------------------|----|----|----|----|--------|----|----|----|-----------|--------------------------|----------|
| Xuo | 13 | 23 | 29 | 14 | 11, 13 | 14 | 11 | 14 | 1 | N ^b | 13 |
| Xuo1 | 14 | 23 | 30 | 14 | 11, 13 | 14 | 11 | 14 | 1 | N ^b | 13 |
| Xuo3 | 13 | 24 | 29 | 14 | 11, 13 | 14 | 11 | 14 | 1 | N ^b | YA004128 |
| A35 | 13 | 23 | 29 | 14 | 11, 13 | 15 | 11 | 14 | 1 | N1c1 ^C | 5 |
| Admixed | | - | - | - | - | - | - | - | 10 | various | |
| A8 | 14 | 22 | 30 | 15 | 9, 19 | 13 | 10 | 13 | 2 | L ^b | 13 |
| A9 | 13 | 24 | 30 | 16 | 14, 15 | 13 | 11 | 11 | 2 | l2a (xl2a1) _b | 13 |
| A21 | 12 | 23 | 28 | 14 | 14, 15 | 13 | 10 | 11 | 1 | l1 ^b | 13 |
| A28 | 13 | 25 | 31 | 16 | 11, 14 | 13 | 11 | 11 | 1 | R1a ^b | 13 |
| A29 | 13 | 25 | 29 | 16 | 12, 13 | 13 | 10 | 11 | 1 | l2a1 ^b | 13 |
| A30 | 13 | 25 | 29 | 16 | 12, 13 | 13 | 10 | 12 | 1 | l2a (xl2a1) ^b | 13 |
| A31 | 13 | 25 | 30 | 16 | 11, 15 | 13 | 12 | 11 | 1 | R1a ^b | 13 |
| A36 | 13 | 23 | 28 | 15 | 11, 19 | 14 | 10 | 11 | 1 | Jl ^b | YA003971 |
| Autochtons | | - | - | - | - | - | - | - | 17 | various | |
| A1 | 10 | 24 | 27 | 15 | 11, 12 | 13 | 9 | 11 | 4 | E1blb ^b | 13 |
| A11 | 10 | 24 | 27 | 15 | 11, 12 | 13 | 9 | 10 | 2 | E1blb ^b | 13 |
| A10 | 14 | 23 | 30 | 17 | 11, 18 | 13 | 10 | 11 | 1 | l2a (xl2a1) ^b | 20 |
| A18 | 14 | 23 | 32 | 14 | 13, 19 | 14 | 11 | 16 | 1 | Q ^b | 13 |
| A1 | 10 | 24 | 27 | 15 | 11, 12 | 13 | 9 | 11 | 1 | E1blb ^b | 13 |
| A20 | 12 | 23 | 28 | 14 | 12, 13 | 13 | 10 | 14 | 1 | N ^b | 13 |
| A23 | 13 | 23 | 28 | 15 | 11, 17 | 16 | 10 | 12 | 1 | G2a ^b | 13 |
| A24 | 13 | 23 | 31 | 15 | 11, 16 | 13 | 11 | 11 | 1 | R1a ^b | 13 |
| A25 | 10 | 24 | 27 | 15 | 12, 12 | 13 | 9 | 7 | 1 | l2a (xl2a1) ^b | 13 |
| A26 | 10 | 24 | 27 | 15 | 12, 19 | 13 | 9 | 11 | 1 | E1blb ^b | 13 |
| A27 | 10 | 24 | 27 | 15 | 12, 13 | 13 | 9 | 10 | 1 | G2a ^b | 13 |

Continuation of table 2

| | | | | | | | | | | | |
|-------|----|----|----|----|--------|----|----|----|-----|------------------------------|----|
| A32 | 13 | 23 | 30 | 14 | 13, 13 | 13 | 10 | 14 | 1 | N2a1a (N1b ^c) | 19 |
| A34 | 13 | 24 | 30 | 15 | 15, 15 | 13 | 11 | 11 | 1 | l2a (xl2a1) ^b | 20 |
| Total | | | | | | | | | 367 | | |

* Table 1 in the Supporting information file contains the parallel haplogroup assignments by the V. Urasin's Predictor and that by D. Adamov.

a Haplotype observed in the current study, the putative Elley line M., XVII:1 (Fig. 2).

b Haplogroup assignments made with the Whit Athey algorithm.

c the name of the haplogroup according to ISOGG 2016

‡ Alleles shown in *italics* correspond to those that differ from the original/dominant haplotype in the same cluster/group by only a few allelic repeat values.

Table 3. 17-loci Y-STR near-matches at YHRD between the original/dominant Elley, Unknown and Omogoy haplotypes and non-Sakha haplotypes (excluding the Athabaskan Indian haplotype).

Notably, the following haplotype pairs have the same 17-loci Y-STR profile:

Mg4 and Mg5; LiMo and ShC1.

| ID | Haplotype * | DYS456 | DYS389 I | DYS390 | DYS389 II | DYS458 | DYS19 | DYS385a/b | DYS393 | DYS391 | DYS439 | DYS635 | DYS392 | GATA_H4 | DYS437 | DYS438 | DYS448 | YHRD Accession No.: |
|------|--------------------------------------|--------|----------|--------|-----------|--------|-------|-----------|--------|--------|--------|--------|--------|---------|--------|--------|--------|---------------------|
| Ello | the original Elley line ^a | 14 | 14 | 23 | 32 | 16 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 19 | Current study |
| Vil | the dominant Unknown line | 14 | 14 | 23 | 31 | 16 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 19 | [13] |
| Omo1 | the dominant Omogoy line | 14 | 14 | 23 | 31 | 16 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 15 | 12 | 14 | 11 | 19 | [6] |
| TCy1 | Turkish Cypriot | 15 | 14 | 23 | 32 | 17 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 19 | YA003850 |
| LiKo | Liaoning, China (Korean) | 14 | 14 | 23 | 32 | 17 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 20 | YA003759 |
| Afg | Afghanistan (Afghan) | 14 | 14 | 23 | 32 | 16 | 14 | 10, 13 | 14 | 10 | 10 | 22 | 16 | 12 | 14 | 11 | 19 | YA003842 |
| JilH | Julin, Chian (Han) | 14 | 14 | 23 | 30 | 16 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 19 | YA004124 |
| LiMo | Liaoning, China (Mongolian) | 14 | 14 | 23 | 30 | 17 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 19 | YA003758 |

| | | | | | | | | | | | | | | | | | | |
|------|--|----|----|----|----|----|----|--------|----|----|----|----|----|----|----|----|----|------------------------------|
| ShC1 | Shandong, China (Han) | 14 | 14 | 23 | 30 | 17 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 19 | YA004081 |
| SSK1 | Seoul, South Korea (Korean) | 14 | 14 | 23 | 30 | 16 | 14 | 11, 13 | 14 | 11 | 10 | 21 | 16 | 12 | 14 | 11 | 19 | YC000251 |
| SSK2 | Seoul, South Korea (Korean) | 14 | 14 | 23 | 31 | 17 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 20 | YC000251 |
| SKK | South Korea (Korean) | 14 | 14 | 23 | 31 | 17 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 20 | YA003728, YC000239, YA003441 |
| Mg1 | Ulaanbaatar, Mongolia (Khalkh) | 14 | 14 | 23 | 31 | 17 | 14 | 10, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 19 | YA003733 |
| Mg4 | Northern Mongolia (Khalkh, Darkhad, Uriankhai) | 14 | 14 | 23 | 31 | 17 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 19 | YA003734 |
| Mg5 | Ulaanbaatar, Mongolia (Khalkh) | 14 | 14 | 23 | 31 | 17 | 14 | 11, 13 | 14 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 19 | YA003733 |
| LiHa | Liaoning, China (Northern Han) | 14 | 14 | 23 | 31 | 17 | 14 | 11, 13 | 13 | 11 | 10 | 22 | 16 | 12 | 14 | 11 | 19 | YA003756 |
| ChHa | Changchun, Jilin, China (Han) | 14 | 14 | 23 | 31 | 16 | 14 | 11, 13 | 15 | 11 | 10 | 22 | 15 | 12 | 14 | 11 | 19 | [31] |

a - Haplotype observed in the current study, the putative Elley line *M.*, XV:1 (fig. 2).

* - Allelic values shown in italics denote allelic repeat variations from 'the original Elley line'

Based on Y-chromosomal SNP or in silico haplogroup assignments (Table 2), nearly all Y-chromosomal STR haplotypes in the Elley, Unknown, Omogoy, Eurasian and Xiongnu descent clusters exclusively comprised haplogroup N (92.7% of all haplotypes), while those from the 'Admixed' and 'Autochthons' groups exhibited more heterogenous haplogroup distributions. Table 4 shows the 'Time to Most Recent Common Ancestor' or TMRCA estimates for each of the five major Sakha clans classified based on the descent clusters described in Table 2 and Figure 3. In close agreement with the hypotheses of the current study, the Xiongnu

clan seems to be the oldest, followed by Omogoy, Elley, Unknown and Eurasian clans, in that order. However, it should be noted that, TMRCA calculations can only be rough estimates due to the uncertainty with respect to the precise Y-chromosomal STR mutation rate(s) that should be used during such calculations [28]. When each of the different nine-loci Y-chromosomal STR haplotypes in Table 2 were queried against YHRD, the following major observations could be made: (i) haplotypes associated with the Elley, Omogoy, Unknown and Autochthons descent clusters had 20, seven, three and two matches in total and in that order (2.5%

Table 4. TMRCA analyses for the nine-loci Sakha Y-chromosomal STR clusters

| Sakha Clan(s) | TMRCA (in years) | SD (in years) | Time period (Mean) |
|--------------------------|------------------|---------------|-----------------------|
| Elley | 283 | 77 | 1656 - 1807 CE (1733) |
| Unknown | 271 | 117 | 1628 - 1862 CE (1745) |
| Omogoy | 360 | 121 | 1535 - 1777 CE (1656) |
| Elley / Unknown / Omogoy | 767 | 166 | 1083 - 1465 CE (1249) |
| Eurasian | 334 | 249 | 1433 - 1931 CE (1682) |
| Xiongnu | 1334 | 588 | 94 - 1270 CE (682) |

of the total matches) with the population datasets corresponding to other ethnic groups from Yakutia, namely Evenks and Yakut-speaking Evenks, (ii) the haplotypes associated with the Elley, Omogoy and Unknown descent clusters had 4, eleven and twenty five matches in total and in that order (2.9% of the total matches) with outside-Yakutia population datasets, largely in the Eurasian context, (iii) vast majority of haplotype matches (94.7% of the total matches) were associated with the Eurasian, Xiongnu and Admixed/Autochthons descent clusters/groups and the outside-Yakutia population datasets, and (iv) 67.0% of the haplotype matches with the outside-Yakutia population datasets were associated with the Admixed and Autochthons groups, where the matches were either over the entire 17-loci or had differences at only a few loci by one-to-two allelic repeat(s).

Notably, the original Elley line and the dominant Omogoy line haplotypes themselves merely differed by a single allelic repeat at two loci, namely DYS389II and DYS392 (Tables 2 and 3). In turn, the dominant Unknown line haplotype differed only by a single allelic repeat at one locus each with those from the Omogoy and Elley lines at DYS389II and DYS392, respectively (Tables 2 and 3). Upon querying the 17-loci Y-chromosomal STR haplotypes corresponding to 'the original Elley line' and 'the dominant Unknown' and 'the dominant Omogoy lines' against YHRD, 14 near-perfect matches were observed in total, each of which differed from the queried haplotypes by single allelic value differ-

ences at one-to-four different loci and largely originating from populations spanning a wide range of Eurasian geography (Table 3). One further match corresponded to a haplotype from the Athabaskan Indian dataset from Central Alaska, U.S.A. (YHRD Accession No.: YA003683), which had a single allelic repeat difference at only one locus from 'the original Elley line', hence bringing about the possibility for the former to be also a descendant of the Elley family.

Figure 5 depicts the 17-loci median-joining network for the 14 near-perfect YHRD matches with the Y-chromosomal STR haplotypes corresponding to the original Elley, the dominant Unknown and Omogoy lines, and those from the outside-Yakutia population datasets. TMRCA is estimated at 1127 ± 317 years or 889 ± 317 CE, covering the period of 572 – 1206 CE, which coincides with the rise and fall of Turkic, Uyghur and Rouran Khaganates.

When only 9-loci Y-chromosomal STR haplotypes corresponding to the original Elley and the dominant Omogoy lines were queried against YHRD, the original Elley line haplotype had perfect matches with two non-Sakha haplotypes, namely one from the Turkish Cypriot and another from the Korean minority in Liaoning, China datasets (haplotype IDs TCy1 and LiKo, respectively, in Table 3) [29, 30], while the dominant Omogoy line had none. In comparison, among the 13 different nine-loci Y-chromosomal STR haplotypes in the Autochthons group, only three had 19 matches in total with the outside-Yakutia datasets at YHRD

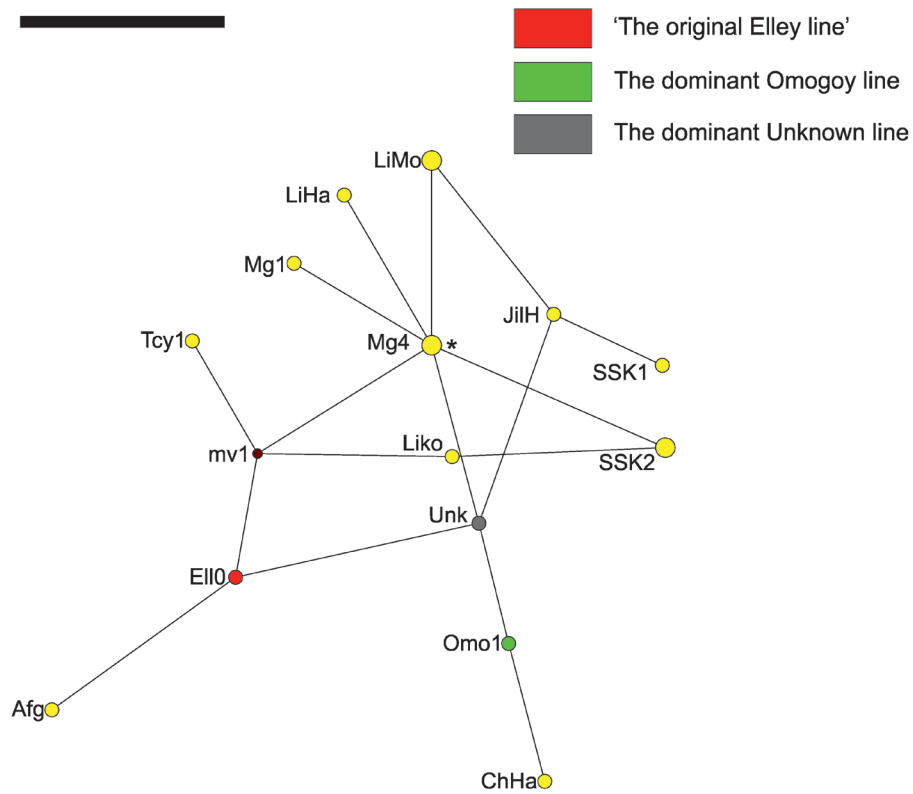


Fig. 5. 17-loci median-joining network analysis of the original/dominant Elley, Unknown and Omogoy Y-chromosomal STR haplotypes with the YHRD matches from outside Yakutia populations.

Each haplotype is denoted by a yellow circle, except for the Elley, Unknown and Omogoy haplotypes shown in different colored circles according to the legend provided. The size of each circle is indicative of the number of haplotypes represented. Median vectors are shown as small black circles and the proposed ancestral node is marked with *. Refer to Table 4 for the haplotype IDs, etc.

(Table 2). Notably, one of these Sakha haplotypes (A20) had nine such matches in total: three with the Central Anatolia, Turkey (YHRD Accession No.: YP000080), three with the Zhejiang, China [Han] (YHRD Accession No.: YP000506), two with the Phillipines [Fillipino] (YHRD Accession No.: YP000229) and one with the Taiwan [Paiwan] (YHRD Accession No.: YP000552) datasets. YHRD matches between the two other nine-loci Y-chromosomal STR haplotypes from the Autochthons group and outside-Yakutia datasets corresponded to those from U.S.A., Thailand, China and Eastern European countries. When repeated at 17-loci, none of the 13 Y-chromosomal STR haplotypes from the Autochthons group had any matches at YHRD with the outside-Yakutia datasets.

Discussion

Yakut ethnogenesis has been investigated by several generations of scientists. These investigations comprised ethnographic, archaeological, anthropological, and genetic studies, as well as those on comparative linguistics and folklore of Sakha. Nevertheless, up until now, these studies have not allowed us to fully understand many aspects of the Yakut ethnogenesis. A breakthrough in the study of this issue was made with the discovery of the Kulun-Atakh culture (13th-15th centuries CE) [32]. A.I. Gogolev was the first to prove the emergence of Sakha as a nationality in the territory of the middle Lena, and his work still remains as the only scientific work that comprehensively analyzed the ethnogenesis of the Yakuts using a wide range of scientific

disciplines: historical, cultural, linguistic, archaeological and anthropological [32]. Notably, the author finally refuted the possibility for the migration of Sakha from the south to an already formed ethnos in Middle Lena. Yet, population genetic analyses still remain a discipline not reconciled with the established facts of Yakut ethnogenesis. This is perhaps partly due to the fact that genetic studies have begun to be involved in the study of the ethnogenesis of the Yakuts only relatively recently [6, 33].

In the present work, an attempt is made to utilize the accumulated Y-chromosomal data that has already been in use in population and forensic genetics to help identify the paternal ancestors of the Yakut people. The solution of these questions is of great importance in understanding the ethnogenesis of the Sakha. Each population is heterogeneous in a genetic sense, and the elucidation of the population genetic structure can help determine the precise origins and potential contributors in a given population. According to the views of a number of authoritative researchers, the Yakut ethnos was formed in the territory of Yakutia as a result of the mixing of people from the south and the autochthonous population [34].

Starting with Zerjal et al. [12], a number of studies on the Yakut Y-STR haplotypes were carried out, but none of these studies can cover a truly representative diversity of the haplotypes of the studied population. Thus, in the current work, an attempt was made to compile from all these sources 367 9-loci Y-STR haplotype data in total (see Table 1), currently the most complete collection of such data on the Sakha population. Such an approach enabled the grouping of similar Sakha haplotypes in an attempt to identify their ancestral paternal lineages (see Table 2).

The relationship among the Sakha Y-STR haplotypes belonging to the haplogroup N are now shown by a median-joining network analysis (Figure 3). The presence of three, closely related star-shaped clusters is immediately notable in the network. Similar structures in phylogenetic patterns reflect the facts of the long historical isolation of relatively small groups of the population. In addition, two small centers of divergence are noted in the peripheral part of the star-shaped structures.

Data from the current study suggest that the Yakut modal Y-STR haplotype, which belonged to the haplogroup N1c1, could be that of the legendary character Elley. The remaining four major clusters could also be classified in line with the anthropological/archaeological/archaeogenetic data available in the literature. Yet, even prior to that, each Y-STR haplotype could be classified as mixed and autochthons based on the fact that whether matching haplotypes could be observed or not, respectively, with the populations outside the territory of modern Yakutia.

In forensic genetics, 17-loci Y-STR haplotypes analyses are commonly used to identify the paternal lineage of a given individual. In an effort to help identify similar Y-STR haplotypes to the three main Yakut haplotypes, the search function of the YHRD database was used. Table 3 lists the 14-loci haplotypes that are presumably related to these three major Yakut haplotypes, which were then analyzed through the construction of median-joining networks (Figure 5). According to the results presented here, very similar Y-STR haplotypes to that of the original Elley line were found in the west: Afghanistan and northern Cyprus, and in the east: Liaoning Province, China and Ulaanbaator, Northern Mongolia. In the case of the dominant Omogoy line, very closely matching haplotypes differing by a single mutational step were found in the city of Chifen of the Jirin Province, China. The widest range of similar haplotypes was found for the Yakut haplotype Unknown: In Mongolia, China and South Korea. For instance, haplotypes differing by a single step mutation were found in Northern Mongolia (Khalk, Darhad, Uryankhai populations), Ulaanbaator (Khalk) and in the province of Jirin, China (Han population).

Notably, Tat-C-bearing Y-chromosomes were also observed in ancient DNA samples from the 2700-3000 years-old Upper Xiajiadian culture in Inner Mongolia, as well as those from the Serteya II site at the Upper Dvina region in Russia and the 'Devichyi gory' culture of long barrow burials at the Nevel'sky district of Pskovsky region in Russia [35, 36]. A 14-loci Y-chromosomal STR median-joining network of the most prevalent Sakha haplotypes and a Tat-C-bearing haplotype from one of

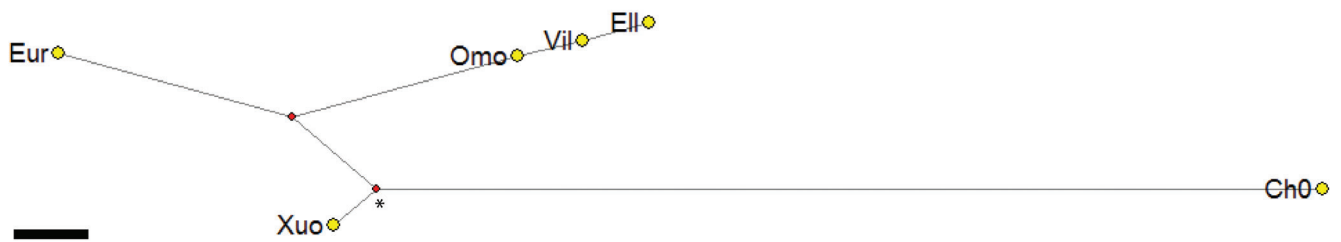


Fig. 6. 14-loci median-joining network analysis for the original/dominant Elley (Ell), Unknown Clan (Vil), Omogoy (Omo), Eurasian (Eur) and Xiongnu (Xuo) Y-chromosomal STR haplotypes and that for a representative ancient DNA sample (Ch0 or DSQ04) from the Upper Xiajiadian Culture recovered from the Inner Mongolia Autonomous Region, China.

*Each haplotype is denoted by a yellow circle, while the median vectors are shown as small red diamonds. The proposed ancestral node is marked with *. A scale bar whose length is roughly proportional to a single mutation event between any of the two neighboring haplotypes is also provided.*

the ancient DNA samples recovered from the Upper Xiajiadian culture in Inner Mongolia (DSQ04) revealed that the contemporary Sakha haplotype ‘Xuo’ (Table 2, Haplotype ID “Xuo”) classified as that of ‘the Xiongnu clan’ in our current study, was the closest to the ancient Xiongnu haplotype (Figure 6). TMRCA estimate for this 14-loci Y-chromosomal STR network was 4357 ± 1038 years or 2341 ± 1038 BCE, which correlated well with the Upper Xiajiadian culture that was dated to the Late Bronze Age (700-1000 BCE).

The three most prevalent 17-loci Y-chromosomal STR haplotypes observed among contemporary Sakhas (79% in total), namely ‘the original Elley line’, ‘the dominant Unknown line’ and ‘the dominant Omogoy line’ themselves seem to be closely related after all (Table 3). This is because from ‘the original Elley line’ to ‘the dominant Omogoy line’, through the apparent intermediary of ‘the dominant Unknown line’, there are only two incremental one-step mutations at only two loci, namely DYS389II and DYS19, or vice versa. A TMRCA estimate for the corresponding descent clusters around these three haplotypes (Figure 3) was calculated as 767 ± 166 years or 1083 - 1415 CE (Table 4).

These three major Sakha paternal lineages may have also arrived in Yakutia at different times and/or from different places and/or with a difference in several generations instead, or perhaps Y-chromo-

somal STR mutations may have taken place in situ in Yakutia. Nevertheless, the immediate common ancestor(s) from the Asian Steppe of these three most prevalent Sakha Y-chromosomal STR haplotypes possibly lived during the prominence of the Turkic Khaganates, hence the near-perfect matches observed across a wide range of Eurasian geography, including as far as from Cyprus in the West to Liaoning, China in the East, then Middle Lena in the North and Afghanistan in the South (Table 3 and Figure 5). There may also be haplotypes closely-related to ‘the dominant Elley line’ among Karakalpaks, Uzbeks and Tajiks, however, limitations in the loci coverage for the available dataset (only eight Y-chromosomal STR loci) precludes further conclusions on this matter [25].

A potential focus for the Westward proliferation of ‘the original Elley line’ could be the Zhetysu territory in Central Asia. After the fall of the Liao Dynasty in 1125 CE following the Jurchen invasion, Elyui Dashi’s together with 20 thousand warriors of the Chotan fortress, fled from the Jurchen and westward to establish the Qara Khitai Khaganate or Western Liao Empire in the Zhetysu territory [37]. Apparently, with the descendants of this army, the related Elley haplotype reached Afghanistan and Cyprus. In Balaesque P. et al. [25] we found polymorphisms close to Elley haplotype in Karakalpaks, Uzbeks and Tajiks. Liao Empire of the Elyui Dynas-

ty lasted from 916 to 1125 CE, mentions about Khitan people in the Chinese chronicles gradually disappear by the time of the Ming Dynasty (1638-44 CE). Ancient DNA studies based on mitochondrial DNA on the remains belonging to the members of the Elyui Dynasty of the Liao Empire near towns Chifens in Inner Mongolia and Fuxin in the Liaoning province of China revealed that Khitans were genetically closest to Daur, Northern Chinese and Mongols, and also bore similarities to Orochons, Buryats and Evenks [38]. Daur are considered to be direct descendants of Khitans, and the former were also found to have paternal lineages belonging to the N1c haplogroup at a frequency of 7.7% [39]. Although nuclear DNA has so far not been successfully isolated from the burials of the Khitan nobility, data from contemporary populations in Liaoning already allows us to postulate a potential Khitan origin for the most prevalent Y-chromosomal STR haplotypes among Sakhas (Table 3).

In one way or the other, the arrival of perhaps only a few people from the Asian Steppe in Yakutia between the 12th and 14th centuries CE coincides with the introduction hitherto unknown technologies and cultural traditions to Northeastern Siberia, which ranged from advanced animal husbandry and milk production (including preparation and consumption of koumiss, a form of fermented product made with mare's milk), transition to a semi-nomadic lifestyle and formation of farms, new house building technologies and introduction of new household equipments [7]. These revolutionary changes in Yakutia coincided with emergence of the Kulun-Atakhs kaya Culture of Sakhas around the 13th - 14th centuries CE [32].

A novel Y-chromosomal STR haplotype that was observed at a high frequency across Eurasia corresponds to those Y-chromosomal lineages again belonging to haplogroup N1 [25]. Balarsque et al. concluded that this high frequency haplotype spread along the Silk Road in around 700 CE and corresponded to the reproductive success of men from the Rouran and Uighur Khaganates, as well as the West Liao Empire of the Khitan Dynasty. Similar haplotypes were also observed among contemporary Sakhas and classified under 'the Eurasian Clan' (Table 2), whose descent cluster (Figure 3)

was dated as 334 ± 249 years old and corresponded to a period between 1433 - 1931 CE (Table 4). This brings about the possibility for the proliferation of these haplotypes as a result of the 200 years old tea trade between Russia and China. The Tea Trade Route stretched from Hankow, China and passed through Mongolia, then Kyakhta in modern day Buryatia, and finally reached Moscow [40]. A Northern branch of the Tea Route also stretched from Kyakhta to Yakutsk. Support for this theory comes from the observation of 35 perfect and near perfect-matches at YHRD (differing by only one allelic difference at a single locus of 17-loci haplotypes) from China, Mongolia and Russia (at central Russian cities of Sverdlovsk, Bryansk and Vologda), and as far as Switzerland, Poland, Norway and Finland, in a way recapitulating the tea trading path.

A potential connection between Sakhas and Xiongnu culture was also observed, albeit in the maternal lineages instead, whereby the mtDNA hypervariable sequence 1 of a woman from grave 25A from the Egyin Gol necropolis was observed in contemporary Sakhas, as well as in an ancient DNA sample belonging to a male individual recovered from the 2200-2400-years-old Pokrovsk grave site in Yakutia [41, 42].

Another line of evidence towards potential ancestral links between Xiongnu and Sakhas came from 3D geometric-morphometric analyses based on 44 craniofacial biometric points on 1558 skulls from the collections of various museums from around the world, which included 68 Xiongnu and 31 Sakha skulls from the Musée de l'Homme collection in Paris, France [43]. In this study, Xiongnu skulls were found to be closest to those of Mongols, as well as those from the Iron Age (2nd - 3rd century BCE) proto-Mongolic Xianbei people from the Liaoning province in China and Sakhas, albeit the direct connection between Xiongnu and Sakha remained mysterious at the time. Although the current evidence from ancient DNA studies may so far fail to substantiate such a claim, our results from median-joining network and associated TMRCA analyses for the most prevalent Y-chromosomal STR Sakha haplotypes (Table 2-4) suggest that the penetration of haplogroup N to the Northeastern Siberia possibly started as early as the period of

Xiongnu domination in the Asian Steppe around 200 BCE. In addition to the matches between four contemporary Sakha Y-chromosomal STR haplotypes ('the Xiongnu clan') and that from the remains of a Xiongnu nobility from the Egyin Gol necropolis (grave 25A), other lines of scientific evidence have also been accumulating towards cultural, economic and historical links between Middle Lena and Xiongnu dominated Asian Steppe [34, 44, 45].

According to the Sakha legends and this study, Elley and Omogoy were not Turkic per se. The core population forming the basis for Sakha ethnogenesis may be represented by the 'autochthonic' descent cluster as classified in the current study instead (Table 2 and Figure 3). Such a classification for a subset of the contemporary Sakha paternal lineages was made based on an earlier hypothesis, which also considered the contribution of an aboriginal stratum of unknown origin during the ethnogenesis of Sakhas [34]. Among such 'Autochthons' haplotypes observed among Sakhas, there were 21 matches at YHRD, 23.8% of which were Turkic, along with 3 matches from Turkey and 2 matches with Yakut-speaking Evenks. One 'autochthons' Y-chromosomal STR haplotype had matches at 4 Y-chromosomal STR loci with an ancient DNA sample from the 2200-2400-years-old Pokrovsk grave site in Yakutia [41]. Ancient DNA analyses from the Balychtakh site dated between 1420-1470 CE in Yakutia revealed a 12-loci Y-chromosomal STR haplotype and a haplogroup assignment of K (excluding N, O and P) [6]. While a corresponding paternal lineage has not so far been detected among contemporary Sakhas, there were 28 matches in total at YHRD from Mongolia, China, Russia and Turkey, for the last one of which there were 10 matches from 4 different geographic regions. Therefore, the 'autochthons' cluster from the current study may correspond to a core Sakha population where the agglutinative language structure was based on. Current consensus is that the Yakut language arose as a result of Turkicisation of an unknown language that also went Mongolization in the process of development. Loaning of Mongolian words to the Yakut language could only be dated as far back as 12th to 13th centuries CE somewhere in Central Yakutia [46], curiously coinciding with the legend of Elley

and Omogoy and data from population genetics of ancient and contemporary Sakhas.

The scientific reconciliation of all the data from genetic studies on contemporary and ancient Sakhas, as well as the current archaeological evidence towards the emergence of the Sakha culture seemingly correlate well with the popular historical narratives, such as the legends of Elley and Omogoy. According to these legends, Elley is also credited for the introduction of novel handicrafts and farming technologies to the Middle Lena. As in the case of traditional fairy tales [47], legends are known to be based at least in part on some historical anecdotes [48]. In any case, these stories were possibly associated with the events that occurred in the Asia Steppe spanning the Southern part of Siberia from the Okhotsk coast to the Caspian Sea and in periods of the dawning and decline of the state administrative formations of Xiongnu, Turkic Khaganates and the Khitan Empire. A large proportion of the contemporary Sakha paternal lineages belong to the haplogroup N and seemingly fall under five major descent clusters based on 9-loci Y-chromosomal STR data as described in Table 2 and Figure 2. The beginning of these paternal lineages was perhaps laid by influential Xiongnu nobility, closely related to a territory of the middle Lena River since the ancient times. In one way or the other, 79 to 89% of all contemporary Sakha paternal lineages could be traced down to only three haplotypes in total, which are in turn closely related themselves and exhibit only minute differentiations of an ancestral 17-loci Y-chromosomal STR haplotype (Table 4). Curiously, even when the number of Y-chromosomal STR loci analyzed were increased up to 23, no significant increase in the resolution of the Sakha paternal lineages could be observed, which constituted a further proof towards the actual homogeneity of these lineages [27]. Unfortunately, the current data still does not have the resolution to help deduce whether such Y-chromosomal STR mutation among the three most prevalent Sakha paternal lineages took place in situ in Yakutia or prior to their arrival in the Asian Steppe. In either case, the legends of Elley and Omogoy and how they have together changed the way of life in Middle Lena forever are still in agreement with the reconciled scientific data in the

current study. Finally, around 8% of the contemporary Sakha Y-chromosomal STR haplotypes, which can not be classified under the five major Sakha paternal lineages discussed in this manuscript may be considered as a mixed Sakha cluster instead, more than half of which (4,9%) constitute autochthonic

polymorphisms. Apparently these paternal lineages may have constituted a core for forming an agglutinative basis for the formation of the Yakut language which subsequently underwent processes of Turkisation and Mongolization.

Acknowledgements. The authors are grateful to F. Guo for providing access to Y-chromosomal STR data on the Korean minority from Liaoning Province; Popov V.V. and Dyakonov V.M. for useful advice, Adamov D. for reading the manuscript and for making valuable comments.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosure statement. The authors reported no potential conflict of interest.

Contributions. Conceived and designed the article Tikhonov D.G., Gurkan C. Coordination and sample preparation for genotyping: Tikhonov D.G. Added analysis tools: Tikhonov D.G., Gurkan C. Demirdov K. D. Data analysis: Tikhonov D.G., Gurkan C. Demirdov K. D. Manuscript drafting: Tikhonov D.G., Gurkan C. Demirdov K. D., Beyoglu E. Discussed and critically edited manuscript: Gurkan C., Tikhonov D. G., Beyoglu E. All authors have read and approved the final version of the manuscript before publication.

REFERENCES

1. RFSS, 2011. (The Russian Federation, the Federal State Statistics Service) 'Vserossiyskaya perepis naseleniya 2010 goda ' [2010 All-Russian Population Census, vol. 1]. (Retrieved June 29, 2012) [in Russian].
2. Pakendorf B., Wiebe V., Tarskaia L.A., Spit-syn V.A., Soodyall H., Rodewald A., Stoneking M. Mitochondrial DNA evidence for admixed origins of central Siberian populations // *Am. J. Phys. Anthropol.* 2003. 120. P. 211–224.
3. Puzyrev V.P., Stepanov V.A., Golubenko M.V., Puzyrev K.V., Maksimova N.R. et al. MtDNA and Y-chromosome lineages in the Yakut population // *Genetika.* 2003. 39. P. 975–981.
4. Pakendorf B., Morar B., Tarskaia L.A., Kay-ser M., Soodyall H., Rodewald A., Stoneking M. Y-chromosomal evidence for a strong reduction in male population size of Yakuts // *Hum. Genet.* 2002. 110. P. 198–200.
5. Pakendorf B., Novgorodov I.N., Osakovskij V.L., Danilova A.P., Protod'jakonov A.P., Stoneking M. Investigating the effects of prehistoric migrations in Siberia : genetic variation and the origins of Yakuts // *Hum. Genet.* 2006, 120. P. 334–353.
6. Crubézy E., Amory S., Keyser C., Bouakaze C., Bodner M., Gibert M. et. al. Human evolution in Siberia: from frozen bodies to ancient DNA // *BMC Evol. Biol.* 2010. 10. 25 p.
7. Ksenofontov, G.V. 'Elleyada: materialyi po mifologii i legendarnoy istorii yakutov' [Elleyada: Materials on the mythology and the legendary history of the Yakuts]. Novosibirsk: Nauka. 1977. 275 p. [In Russian].
8. TDK. Türk Dil Kurumu Büyük Türkçe Sözlük [Turkish Language Institution Big Turkish Dictionary]. Türk Dil Kurumu Yayınları. 2011 [in Turkish]
9. Romanova, E.N. 'G.V. Ksenofontov: mif o stranstvuyuschem geroe' [G.V. Ksenofontov: The myth about a travelling hero], in: Tumarkin, D.D.

- (Ed.), 'Repressirovannyye etnografyyi' [Repressed Ethnography]. Vost. lit., Moscow. 2003. P. 78-104 [in Russian].
10. Vasilyev Yu., I., 2008. 'Moy rodnoy drevniy yazyik' [My native ancient language]. Bichik, Yakutsk. 2008. P. 20-25 [in Yakut language].
11. Tikhonov D.G. Legendarnyyi rodonachalnik Elley i effekt osnovatelya v populyatsii saha. *Mezhdunarodnyy zhurnal prikladnykh i fundamentalnykh issledovaniy* [Elley legendary ancestor and founder effect in a population of Sakha]. 2016. № 9 – 2. P. 241– 243 [in Russian].
12. Zerjal T., Dashnyam B., Pandya A., Kayser M., Roewer L., Santos F.R. et al. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis // *Am. J. Hum. Genet.* 1997, 60. P. 1174-1183.
13. Thèves C., Balaesque P., Evdokimova L.E., Timofeev I.V., Alekseev A.N., Sevin A. et al. Population genetics of 17 Y-chromosomal STR loci in Yakutia // *Forensic Sci. Int. Genet.* 2010. 4. e129–130.
14. Lindenau Ia., I. 'Opisanie narodov Sibiri (pervaya polovina XVIII veka): Istoriko-etnograficheskie materialy o narodakh Sibiri i Severo-Vostoka' [A description of people of Siberia (the first half of the XVIIIth century): The historicaethnographic materials about people of Siberia and the North-East]. Magadan Publishing House, Magadan. 1983. P. 18 [in Russian, translated from German].
15. Aprosimov A.M., Popov G.V. 'Rodoslavnaya po naslegam yakutskikh ulusov' [A pedigree according to the naslegs of Yakut uluses]. Yakutsk: Bichik. 2015. 62 p. [In Yakut language].
16. Tokarev S.A. 'Obschestvennyy stroy yakutov XVII – XVIII vv.' [Public formation of Yakuts in the XVII – XVIIIth centuries]. Yakutsk: Yakut State Publishing House. 1945. 240 p. [In Russian].
17. CNAYASSR. (Central National Archive of the Yakut ASSR) 'Vedomost' o razdele pokosnykh mest Meginskogo ulusa Moyrutskago naslega. Gen'varya 19-go dnya 1776 g.' [Register about division of harvest places in Meginsky ulus of Moruksky nasleg, January 19,1776]. 1776. Fond No. 43, List No. 1. Case No. 2154 [in Russian].
18. Rootsi S., Zhivotovsky L.A., Baldovic M., Kayser M., Kutuev I.A., Khusainova R. et al. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe // *Eur. J. Hum. Genet.* 2007. 15. P. 204–211.
19. Karmin M., Saag L., Vicente M., Wilson Sayres M.A., Järve M., Talas U.G. et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture // *Genome Res.* 2015. 25. P. 459–466.
20. Gao T., Yun L., Gu Y., He W., Wang Z., Hou Y. Phylogenetic analysis and forensic characteristics of 12 populations using 23 Y-STR loci // *Forensic Sci. Int. Genet.* 2015. 19. P. 130–133.
21. Willuweit S., Roewer L. The new Y Chromosome Haplotype Reference Database // *Forensic Sci. Int. Genet.* 2014. 15. P. 43–48.
22. Derenko M., Malyarchuk B., Denisova G., Wozniak M., Grzybowski T., Dambueva I., Zakharov I. Y-chromosome haplogroup N dispersals from south Siberia to Europe // *J Hum Genet.* 2007. 52. P. 763–770.
23. Ballantyne K.N., Goedbloed M., Fang R., Schaap O., Lao O., Wollstein A. et al. Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications // *Am. J. Hum. Genet.* 2010. 87. P. 341–353.
24. Adamov D. O gaplogruppe N1c1 sakha (yakutov). [About haplogroup N1c1 of Sakha people (Yakuts)]. *The Russian Journal of Genetic Genealogy (Русская версия)*. 2010. 2, 31-42 [in Russian].
25. Balaesque P., Poulet N., Cussat-Blanc S., Gerard P., Quintana-Murci L., Heyer E., Jobling M.A. Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations // *Eur. J. Hum. Genet.* 2015. 23. P. 1413–1422.
26. Keyser-Tracqui C., Blandin A.P., Ricaut A.F.X., Petkovski B.E., Crubézy A.E., Ludes B.B. Does the Tat polymorphism originate in northern Mongolia // *Int. Congr. Ser.* 2004. 1261. P. 325–327.
27. Keyser C., Hollard C., Gonzalez A., Fausser J.L., Rivals E., Alexeev A.N. et al. The ancient Yakut's: a population genetic enigma // *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 2015. 370.
28. Jobling M.A., Tyler-Smith C. The human Y-chromosome: an evolutionary marker comes of age // *Nat. Rev. Genet.* 2003. 4. P. 598–612.
29. Guo F., Song L., Zhang L. Population genetics

for 17 Y-STR loci in Korean ethnic minority from Liaoning Province, Northeast China // *Forensic Sci. Int. Genet.* 2016. 22. e9–e11.

30. Gurkan C., Sevay H., Demirdov D.K., Hossoz S., Ceker D., Terali K., Erol A.S. Turkish Cypriot paternal lineages bear an autochthonous character and closest resemblance to those from neighboring Near Eastern populations // *Ann. Hum. Biol.* 2017, Mar; 44(2):164 – 174.

31. Chen W.Q., Feng Z., Jin W., Zhang Y.J., Wang Q., Li P. et al. Genetic analysis of 17 Y-STR loci from 1026 individuals of Han populations in Jilin Province, Northeast China // *Int. J. Legal Med.* 2018. 132(5). P. 1309–1311.

32. Gogolev A.I. 'Yakutyi (problemyi etnogeneza i formirovaniya kulturyi)' [The Yakut (Problems of ethnic genesis and formation of culture)]. Ilin, Yakutsk: Ilin. 1993. 200 p. [In Russian].

33. Fedorova S.A. 'Geneticheskie portretyi narodov Respubliki Saha (Yakutiya): analiz liniy mitohondrialnoy DNK i Y-hromosomyi' [Genetic portraits of the peoples of the Republic of Sakha (Yakutia): analysis of mitochondrial DNA and Y-chromosome lines]. Yakutsk: Publishing House of the Yakutsk Scientific Center SB of the RAS. 2008. 182 p. [In Russian].

34. Alekseev A.N., 2013. 'Rannie kochevniki v Yakutii' [Early nomads in Yakutia] // 'Vestnik Severo-Vostochnogo federalnogo universiteta im. M.K. Ammosova.' [A journal of M.K. Ammosov North-Eastern Federal University]. 2013. P. 62-69 [in Russian].

35. Cui Y., Li H., Ning C., Zhang Y., Chen L., Zhao X. et al. Y chromosome analysis of prehistoric human populations in the West Liao River Valley Northeast China // *BMC Evol. Biol.* 2013. 13. P. 216.

36. Chekunova E.M., Yartsev, N.V., Chekunov, M.K. et al. 'Pervyye rezultatyi genotipirovaniya korennyih zhiteley i chelovecheskih kostnyih ostankov iz arheologicheskikh pamyatnikov Verhnego Podvinya' [The first results of genotyping of indigenous peoples and human bone remains from archaeological monuments of Verkhnee Podvinie] // 'Arheologiya ozYorniyh poseleniy IV—II tyis. do n.e.: hronologiya kultur i prirodno-klimaticheskie ritmyi' [The archeology of lake settlements of IV-II centuries B.C.: A chronology of cultures and nat-

uralclimatic rhythms]. "Periphery" LLC, Saint Petersburg, 2014. P. 287-294 [in Russian].

37. Biran M. *The Empire of the Qara Khitai in Eurasian History: Between China and the Islamic World.* Cambridge University Press, Cambridge. 2005. 279 p.

38. Xu Y., Zhang X., Cuy Y., Zhang Q., Zhou H., Zhu H.. Genetic Structure analysis of Human Remains from Khitan Noble Necropolis // *Chem. Res. Chin. Univ.* 2006. 22. P. 123–128.

39. Xue Y., Zerjal T., Bao W., Zhu S., Shu Q., Xu J. et al. Male demography in East Asia: a north-south contrast in human population expansion times // *Genetics.* 2006. 172. P. 2431–2439.

40. Tszaitzi, L. 'Chaynaya trgovlya mezhdou Kittaem i Rossiey' [Tea trade between China and Russia] // 'Teoriya i praktika obschestvennogo razvitiya' [Collection: A theory and practice of society development]. 2009. P. 63-75 [in Russian].

41. Amory S., Crubézy E., Keyser C., Alekseev A.N., Ludes B. Early influence of the steppe tribes in the peopling of Siberia // *Hum. Biol.* 2006. 78. P. 531–549.

42. Keyser-Tracqui C., Crubézy E., Pamzav H., Varga T., Ludes B. Population origins in Mongolia : genetic structure analysis of ancient and modern DNA // *Am J Phys. Anthropol.* 2006. 131. P. 272–281.

43. Schmidt R.W., Seguchi N. Craniofacial variation of the Xiongnu Iron Age nomads of Mongolia reveals their possible origins and population history // *Quat. Int.* 2016. 405. Part B. P. 110–121.

44. Gogolev A.I., 2014 'Otrazhenie istokov hunnskogo vremeni v yakutskoy kulture' [Reflection of Hunnish time sources in Yakut culture] (in Russian) // 'Vestnik Severo-Vostochnogo federalnogo universiteta im. M.K. Ammosova' [A journal of M.K. Ammosov North-Eastern Federal University], 2014. 60-65.

45. Ubryatova, E.I., 1960. 'Yakutskiy yazyik v ego otnoshenii k drugim tyurkskim yazyikam, a takzhe k yazyikam mongolskim i tunguso-manchzhurskim' [Yakut in its relation to other Turkic languages, as well as to Mongol, Tungus and Manchurian languages], 'Sb. trudov. Sektsiya: Altaistika' [Collected papers. Section: Altaic languages studies]. Moscow: Vost. Lit. 1960 [in Russian].

46. Beyoglu E. Cyprus Turkish Fairy Tales: Glimpse of a Harmonious Past // *Am. J. Psychoanal.* 2015. 75. P. 382–393.

47. Duranli M. Yakut Efsaneleri' [Yakut Legends] (in Turkish). Social Sciences Institute. T.C. Ege University. Zmir. 2004. P. 505.

About the authors

TIKHONOV Dmitry Gavrilievich, MD, Professor, Senior Research Officer of the Scientific research Center of the Medical Institute of the North-Eastern Federal University, 677009, Yakutsk, St. Bld. 8, Russia, <https://orcid.org/0000-0003-3385-9471>, e-mail: Tikhonov.dmitri@yandex.ru.

GURKAN Cemal, Director of the Turkish Cypriot DNA Laboratory / Academic Staff, Dr Fazil Kucuk Faculty of Medicine, Eastern Mediterranean University, Nicosia (North Cyprus), Turkey, <http://orcid.org/0000-0001-7379-4559>, e-mail: cemal.gurkan@gmail.com.

DEMIRDOV Damla, Turkish Cypriot DNA Laboratory, Committee on Missing Persons in Cyprus Turkish Cypriot Member Office, Nicosia (North Cyprus), Turkey, damla0330@hotmail.com.

BEYOGLU Erdem, Baris Mental and Neurological Disorders State Hospital, Nicosia (North Cyprus), Turkey erdembey@yahoo.com.

Supporting Online Material

Refer to Web version of the journal for supplementary material.