

# Investigating the Prehistory of Tungusic Peoples of Siberia and the Amur-Ussuri Region with Complete mtDNA Genome Sequences and Y-chromosomal Markers

Ana T. Duggan<sup>1\*</sup>, Mark Whitten<sup>2#a</sup>, Victor Wiebe<sup>1</sup>, Michael Crawford<sup>3</sup>, Anne Butthof<sup>2#b</sup>, Victor Spitsyn<sup>4</sup>, Sergey Makarov<sup>4</sup>, Innokentiy Novgorodov<sup>5</sup>, Vladimir Osakovsky<sup>6</sup>, Brigitte Pakendorf<sup>2#c\*</sup>

**1** Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, **2** MPRG on Comparative Population Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, **3** Department of Anthropology, University of Kansas, Lawrence, Kansas, United States of America, **4** Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russian Federation, **5** Institute of Foreign Philology and Regional Studies, North-Eastern Federal University, Yakutsk, Russia, **6** Institute of Health, North-Eastern Federal University, Yakutsk, Russia

## Abstract

Evenks and Evens, Tungusic-speaking reindeer herders and hunter-gatherers, are spread over a wide area of northern Asia, whereas their linguistic relatives the Udegey, sedentary fishermen and hunter-gatherers, are settled to the south of the lower Amur River. The prehistory and relationships of these Tungusic peoples are as yet poorly investigated, especially with respect to their interactions with neighbouring populations. In this study, we analyse over 500 complete mtDNA genome sequences from nine different Evenk and even subgroups as well as their geographic neighbours from Siberia and their linguistic relatives the Udegey from the Amur-Ussuri region in order to investigate the prehistory of the Tungusic populations. These data are supplemented with analyses of Y-chromosomal haplogroups and STR haplotypes in the Evenks, Evens, and neighbouring Siberian populations. We demonstrate that whereas the North Tungusic Evenks and Evens show evidence of shared ancestry both in the maternal and in the paternal line, this signal has been attenuated by genetic drift and differential gene flow with neighbouring populations, with isolation by distance further shaping the maternal genepool of the Evens. The Udegey, in contrast, appear quite divergent from their linguistic relatives in the maternal line, with a mtDNA haplogroup composition characteristic of populations of the Amur-Ussuri region. Nevertheless, they show affinities with the Evenks, indicating that they might be the result of admixture between local Amur-Ussuri populations and Tungusic populations from the north.

**Citation:** Duggan AT, Whitten M, Wiebe V, Crawford M, Butthof A, et al. (2013) Investigating the Prehistory of Tungusic Peoples of Siberia and the Amur-Ussuri Region with Complete mtDNA Genome Sequences and Y-chromosomal Markers. PLoS ONE 8(12): e83570. doi:10.1371/journal.pone.0083570

**Editor:** Toomas Kivisild, University of Cambridge, United Kingdom

**Received:** July 30, 2013; **Accepted:** November 5, 2013; **Published:** December 12, 2013

**Copyright:** © 2013 Duggan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was funded by the Max Planck Society. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

\* E-mail: ana\_duggan@eva.mpg.de (ATD); Brigitte.Pakendorf@cncrs.fr (BP)

#a Current address: dii Healthcare GmbH, Leipzig, Germany

#b Current address: Institute of Biochemistry, University of Leipzig, Leipzig, Germany

#c Current address: Laboratoire Dynamique du Langage, UMR5596, CNRS & Université Lyon Lumière 2, Lyon, France

## Introduction

Evenks and Evens are spread over a wide area of northern Asia from the Yenisey river in the west to the Chukotka and Kamchatka peninsulas in the east, and from the Taimyr Peninsula in the north to northern China in the south. They are linguistically and culturally closely related with a traditional lifestyle of highly nomadic hunting and gathering and reindeer herding; their languages belong to the North Tungusic branch of the Tungusic language family [1]. Other Tungusic-speaking groups are settled to the southeast of the Evenks and Evens, along the lower Amur and Ussuri rivers, as well as on Sakhalin

island. These include the linguistically closely related Negidal, whose North Tungusic language shows similarities to both Evenki and Even, as well as populations speaking languages classified as South Tungusic, such as the Udegey (also known as Udihe or Udeghe) and Ulchi. In contrast to the Evens and Evenks, the Tungusic peoples of the Amur-Ussuri region, who we here also refer to alternatively as Amur Tungusic, are traditionally sedentary fishermen and hunters rather than nomadic reindeer herders [2,3].

Different hypotheses exist concerning the origins of the North Tungusic Evenks and Evens and their relations with the Amur Tungusic peoples. Vasilevič [4] proposed a relatively ancient

separation some 3500 years before present (BP) and a split between the Evenks and Evens approximately 1500 years BP, when the North Tungusic groups migrated northwards from an area south of Lake Baikal. In contrast, Tugolukov [5] and Janhunen [6] propose a more recent separation of the North and Amur Tungusic groups some 800 years BP, with the ultimate split of Evenks and Evens possibly occurring as late as the 17<sup>th</sup> to 18<sup>th</sup> centuries CE. At this time Turkic-speaking cattle and horse pastoralists, the Yakuts, expanded over the large territory they occupy nowadays [7,8], in the process displacing the Tungusic reindeer herders. During this displacement, the ancestors of current-day Evenks moved west- and northwestwards, while the ancestors of present-day Evens moved to the east and northeast [9].

The spread of the ancestors of the North Tungusic groups over the large territory they occupy nowadays may have been accompanied by different degrees of intermarriage with local inhabitants. Thus, it is assumed that Yukaghir groups were assimilated by Evenks and especially Evens [9], a process that has continued until recent times [10]. Nowadays, the settlement pattern of these North Tungusic populations is highly fragmented, with small communities living interspersed with other peoples, such as Kets and Samoyedic groups in the west, Buryats in the southwest, Yakuts and Yukaghirs in the central regions, and Chukchi and Koryaks in the east. This has led to a large degree of dialectal diversification of both the Evenki and Even language [11,12], possibly due to contact with the languages spoken by their neighbours. The South Tungusic populations, on the other hand, live in the vicinity of the Nivkh and formerly of the Ainu. The Nivkh speak an isolate language and were traditionally fishermen and hunters of sea mammals [13].

The populations of northern Asia are characterized in the maternal line by high frequencies of mtDNA haplogroups C and D [10,14–18]; in contrast, the peoples of the Amur-Ussuri region carry high frequencies of haplogroups Y1 and N9b [10,18,19], while the peoples of Kamchatka are characterized by high frequencies of haplogroup G1, also common in the Negidal [18,20]. In the paternal line, Y-chromosomal haplogroup C is widespread over a large area encompassing both Siberia and the Amur-Ussuri region, being found in high frequency in North Tungusic and Amur Tungusic populations as well as in the Nivkh. In contrast, northern Siberian populations are characterized by high frequencies of haplogroup N, with N1c being the predominant haplogroup found in the Yakuts [14,15,17,19,21–26]. There are thus discrepancies between the maternal, paternal, and linguistic perspectives concerning the population history of northern Asia and the Amur-Ussuri region: whereas the mtDNA data point to an ancient divergence between peoples inhabiting the two regions, the Y-chromosomal haplogroup frequencies link the Tungusic peoples of central and northeastern Siberia with those of the Amur-Ussuri area, and linguistically the Evenks, Evens, Negidals, and South Tungusic populations such as the Udegey and Ulchi share a relatively recent common ancestor.

The historical expansion and the resulting fragmented settlement pattern of the Evenks and Evens raises the question to what extent different communities have intermarried with

their geographic neighbours rather than their linguistic and cultural relatives. A previous investigation into the relationship of Evenks and Evens found evidence for shared ancestry in both the mtDNA and Y-chromosomal data, with subsequent isolation of men belonging to different subgroups as well as indications of intermarriage with neighbouring populations in the maternal line [15]. In this study, we refine our analysis with complete mtDNA genome sequences for four subgroups of Evenks and five subgroups of Evens in comparison to their linguistic relatives, the South Tungusic Udegey, as well as to their geographic neighbours the Yakuts, Yukaghirs, and Koryaks. We also include the geographic neighbours of the Udegey, the Nivkh, to shed light on the question to which extent admixture rather than shared ancestry has shaped the genepool of the populations of Siberia and the Amur-Ussuri region. To explore the paternal relationships of the North Tungusic groups, Y-chromosomal haplogroup and short tandem repeat (STR) data are taken into account. With this data set, we are able to demonstrate that genetic drift and differential admixture have attenuated the signal of shared ancestry among the North Tungusic populations, and that the Udegey are likely to be the result of admixture in the maternal line between indigenous Amur-Ussuri populations and Siberian populations presumably speaking Tungusic languages.

## Results

### MtDNA diversity and haplogroup composition

The populations of the Amur-Ussuri region (Udegey and Nivkh) are characterized by low sequence diversity (0.88) and intermediate and low nucleotide diversity values (0.0016 and 0.0011, respectively; Table 1). This contrasts with the Yakuts, who are among the populations studied here with the highest sequence and nucleotide diversity (0.98 and 0.0022–0.0023, respectively). The Sebjan and Kamchatkan Evens and the Stony Tunguska Evenks have relatively low sequence (0.86, 0.90, and 0.94, respectively) and intermediate nucleotide diversity values (0.0016, 0.0018, and 0.0015, respectively), which contrast with the Sakkyryr Evens, whose sequence and nucleotide diversity are as high as those found in the Yakuts (0.98 and 0.0023).

As can be seen in Figure 1 and Table S1, haplogroups D4, C4b, and C4a1 are the most frequent haplogroups among the populations analysed here (16.3%–11.4% overall; cf. [17]). Although the Evens and Evenks are characterized by high frequencies of these common haplogroups, they also differ in their haplogroup composition: Evens have a much higher frequency of haplogroup Z (15.6%) than Evenks (4.6%), while Evenks have a much higher frequency of C4a2 (15.4%) than Evens (2.5%), as well as of haplogroup A (8.5%), which is absent among the latter. Subhaplogroup C4b3a is present only in Evens and Yukaghirs, as found previously [17]; in contrast, subhaplogroup C5d1, which was previously suggested to be common in Evens and Yukaghirs [17], is found in four Stony Tunguska Evenks as opposed to only one Yukaghir and two Evens (Table S1). Furthermore, the subhaplogroup of C4b carrying the T3306C transition defined by Fedorova et al. [17] as C4b9, which they suggest is common in Evenks, is found

**Table 1.** mtDNA diversity values in the populations studied here with their linguistic affiliation and rough geographic location.

Population	abbrev	linguistic affiliation	geography	N	n	S	Seq div	Seq div SD	$\pi$	$\pi$ SD
<b>all Evenk</b>		<b>N. Tungusic</b>	<b>Siberia</b>	<b>130</b>	<b>53</b>	<b>287</b>	<b>0,98</b>	<b>0,00</b>	<b>0,0020</b>	<b>0,0010</b>
Taimyr	TAI	N. Tungusic	Siberia	24	16	145	0,95	0,03	0,0020	0,0010
Stony Tunguska	STE	N. Tungusic	Siberia	39	17	124	0,94	0,02	0,0015	0,0007
Nyukzha	NYUK	N. Tungusic	Siberia	46	25	175	0,97	0,01	0,0021	0,0011
lengra	IENG	N. Tungusic	Siberia	21	13	107	0,96	0,02	0,0017	0,0009
<b>all Even</b>		<b>N. Tungusic</b>	<b>Siberia</b>	<b>122</b>	<b>54</b>	<b>232</b>	<b>0,98</b>	<b>0,00</b>	<b>0,0020</b>	<b>0,0010</b>
Sakkyryyr	SAK	N. Tungusic	Siberia	23	19	150	0,98	0,02	0,0023	0,0012
Sebjan	SEB	N. Tungusic	Siberia	18	8	71	0,86	0,06	0,0016	0,0008
Tompo	TOM	N. Tungusic	Siberia	27	16	128	0,95	0,02	0,0018	0,0009
Berezovka	BER	N. Tungusic	Siberia	15	11	85	0,95	0,04	0,0017	0,0008
Kamchatka	KAM	N. Tungusic	Kamchatka	39	13	103	0,90	0,02	0,0018	0,0009
<b>Udegey</b>	<b>UDI</b>	<b>S. Tungusic</b>	<b>Amur-Ussuri</b>	<b>31</b>	<b>14</b>	<b>96</b>	<b>0,88</b>	<b>0,05</b>	<b>0,0016</b>	<b>0,0008</b>
<b>all Yakut</b>		<b>Turkic</b>	<b>Siberia</b>	<b>169</b>	<b>94</b>	<b>446</b>	<b>0,98</b>	<b>0,00</b>	<b>0,0023</b>	<b>0,0011</b>
Vilyuy	VIL_YAK	Turkic	Siberia	49	35	252	0,98	0,01	0,0022	0,0011
Central	CNT_YAK	Turkic	Siberia	88	64	368	0,98	0,01	0,0023	0,0011
Northeast	NE_YAK	Turkic	Siberia	32	24	185	0,98	0,01	0,0023	0,0011
<b>Yukaghir</b>	<b>YUK</b>	<b>isolate</b>	<b>Siberia</b>	<b>20</b>	<b>13</b>	<b>105</b>	<b>0,94</b>	<b>0,03</b>	<b>0,0018</b>	<b>0,0009</b>
<b>Koryak</b>	<b>KOR</b>	<b>Chukotko-Kamchatkan</b>	<b>Kamchatka</b>	<b>15</b>	<b>8</b>	<b>75</b>	<b>0,90</b>	<b>0,05</b>	<b>0,0018</b>	<b>0,0009</b>
<b>Nivkh</b>	<b>NIV</b>	<b>isolate</b>	<b>Amur-Ussuri</b>	<b>38</b>	<b>14</b>	<b>60</b>	<b>0,88</b>	<b>0,04</b>	<b>0,0011</b>	<b>0,0006</b>

abbrev = abbreviation used in figures and tables; N = sample size; n = number of haplotypes; S = number of polymorphic sites; Seq div = sequence diversity; SD = standard deviation;  $\pi$  = nucleotide diversity

doi: 10.1371/journal.pone.0083570.t001

here only in Evens (belonging to the Sakkyryyr and Sebjan subpopulations).

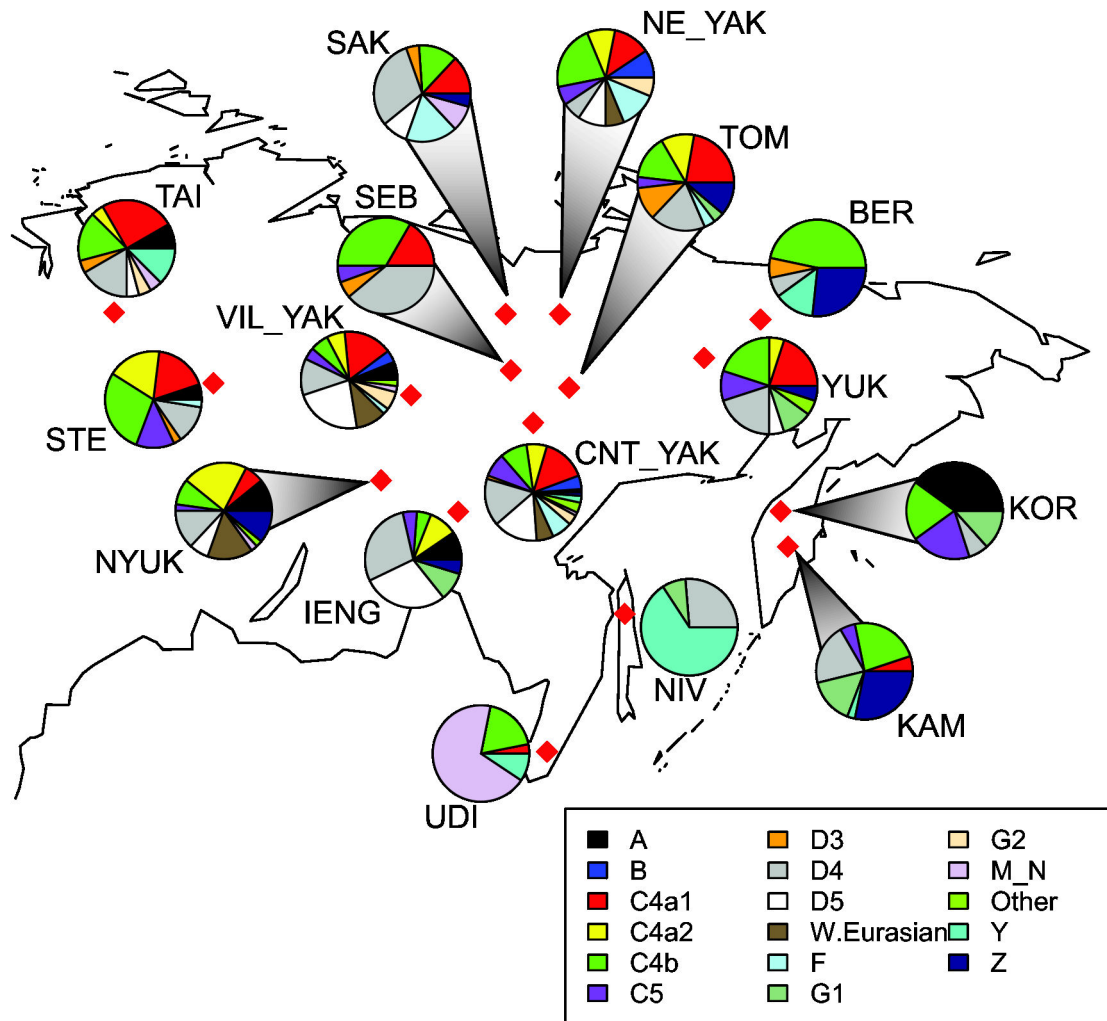
Differences exist among the individual Evenk and Even subgroups, too (Table S1). For instance, two sequences belonging to haplogroup A2a are found among the Taimyr Evens, but not in any of the other Tungusic or neighbouring populations. Haplogroup D5a2a2, which is frequent in Yakuts, is found in nearly 29% in the lengra Evens and nearly 9% in the Sakkyryyr Evens, but is absent in the Stony Tunguska Evens and other Even subgroups. Among the Evens, haplogroup Z is found at particularly high frequency in the eastern subgroups Kamchatka (28.2%) and Berezovka (26.7%) and in intermediate frequency in the Tompo Evens (11.1%), but is low to absent in Sakkyryyr and Sebjan, respectively. All the individuals classified here as Z1a carry the A11252G transition and thus correspond to what was called Z1a3 by Fedorova et al. [17], while all individuals classified in our study as Z1a1 carry the G7521A and G8251A mutations and thus correspond to what they defined as clade Z1a1b.

The South Tungusic Udegey, who are the linguistic relatives of the Evenks and Evens, have a very different haplogroup composition (Table S1): they lack D4, exhibit only low frequencies of C4a1, and have high frequencies of haplogroups N9b (32.3%), M7a2a (16.1%), and M9a1 (9.7%). Other than one M7a2a sequence in the Nyukzha Evens, these haplogroups are not found in any of the other populations studied here. In addition, the Udegey have 12.9% of haplogroup M8a1 which was not previously recorded in populations of Siberia or the Amur-Ussuri region. In their haplogroup composition, the Udegey also differ from their neighbours the Nivkh, who have 65.8% haplogroup Y1a and

26.3% D4m2, a haplogroup found in low frequencies in Evens and Yakuts, but lacking in the Udegey (Table S1).

### MtDNA haplotype sharing analyses and networks

An analysis of shared haplotypes (Table S2) demonstrates that most of the sharing involves the different Yakut subgroups: 35 of the 54 haplotypes that are shared across different (sub)populations involve Yakuts, and 14 are shared only among different Yakut subgroups. Five haplotypes are shared only between Evens and Yakuts, and five others are shared only between Evens and Yakuts. All five of the haplotypes shared solely between Evens and Yakuts are found in the Sakkyryyr Evens, and four are shared exclusively between Sakkyryyr Evens and Yakuts, demonstrating the considerable amount of admixture in the maternal line that has taken place between this Even subgroup and the Yakuts. Four haplotypes are shared only between Evens and Yukaghirs, while none are shared solely between Evens and Yukaghirs, which is in good accordance with the greater geographic proximity of the Yukaghirs and the Evens. Three haplotypes are shared exclusively between Evens and Evens; this contrasts with six haplotypes shared only among Evenk subgroups and four shared only among Evens. The Udegey, Nivkh, and Koryaks each share only one sequence with another population included here, although there is considerable haplotype sharing within these populations (Table S2). The analysis of haplotype sharing thus demonstrates the distinctiveness of the populations of the periphery (Amur-Ussuri region and Kamchatka), as also reflected in their divergent haplogroup composition (Table S1), a certain level of divergence between



**Figure 1. Map of Siberia showing approximate locations of sampled populations and their basic haplogroup composition.** (Sub)population abbreviations as in Table 1.

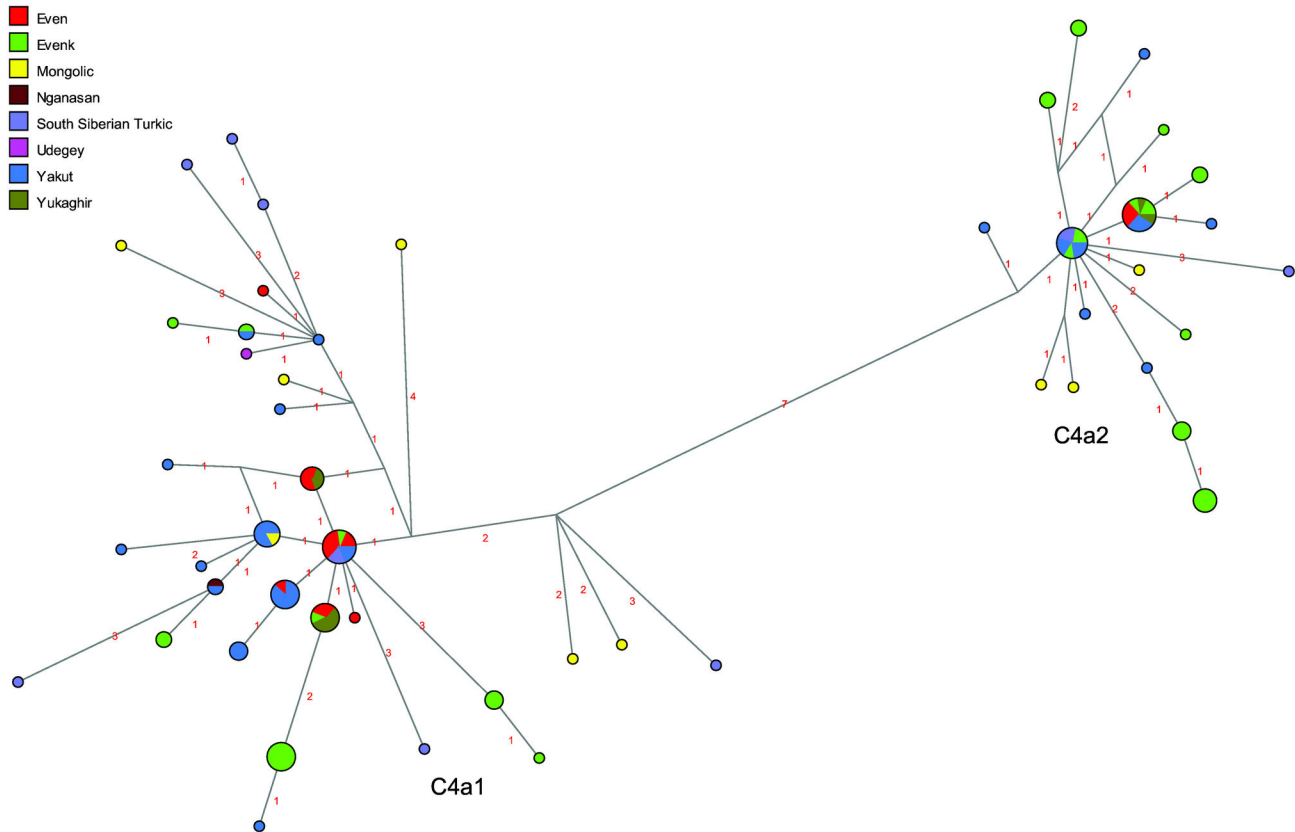
doi: 10.1371/journal.pone.0083570.g001

the Evenks and Evens, and the central position played by the Yakuts, who share haplotypes with Evenks, Evens, Yukaghirs, and even Nivkh and Udegey.

Analyses of haplotype sharing only provide information on identical sequences shared between populations; network analyses in addition show closely related, albeit not identical, sequences. Note, however, that the results of the two types of analysis are not entirely comparable, as they are based on different alignments, as described in the Material & Methods section below. The network of C4a (Figure 2) shows that whereas subhaplogroup C4a1 is widespread in Siberian populations, subhaplogroup C4a2 is dominated by Evenk and Yakut sequences. Thus, of 18 haplotypes, nine are found in Evenks and seven are found in Yakuts; only one haplotype, which is shared with Evenks, Yakuts, and Yukaghirs, is found in Evens. The network of haplogroup C4b (Figure 3) further illustrates the proximity of Evenks and Yakuts, who tend to fall upon the same branches with only one mutation between

haplotypes (shown by asterisks in the figure), while the Evens tend to share branches with the Yukaghirs (indicated by arrows in the figure), confirming the genetic proximity of these populations that emerged in the analysis of shared haplotypes. This network also demonstrates that the Udegey C4b sequences originate from a central haplotype shared with several Evenks as well as a Yakut and a Buryat; haplotypes found in Evens also derive from this central node.

The network of different M and N subhaplogroups (Figure 4) clearly illustrates the distinctiveness of the Udegey, and suggests different affinities of the sequence types found in this population. The sequences belonging to subhaplogroups M9 and M7a2a are relatively close to sequences found in southern Siberian populations (Nyukzha Evenks, Buryats, Khamnigan, and Tuvan), while the M8 and N9b sequences are quite distinct, with only one sequence found in the Ulchi, another South Tungusic population of the Amur-Ussuri region, being shared with the largest Udegey N9b haplotype.



**Figure 2. MJ-network of mtDNA haplogroup C4a.** The North Tungusic haplotypes are coloured by population (Evenks and Evens) rather than subgroup. The size of the nodes is proportional to the number of individuals carrying that node, and the number of mutations is indicated along the branches. The labelled subhaplogroups are discussed in the text.

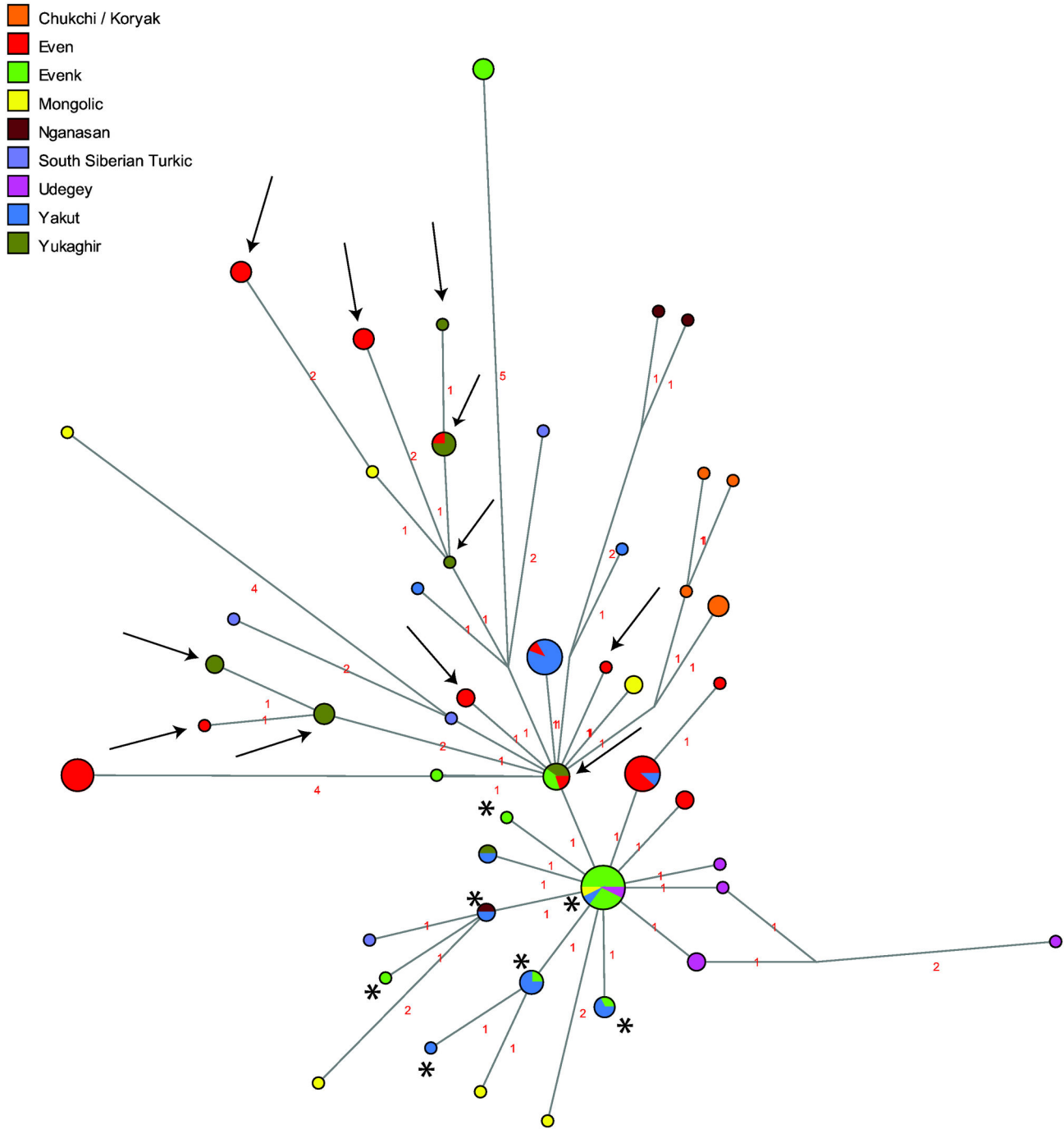
doi: 10.1371/journal.pone.0083570.g002

The network analyses also provide evidence of gene flow. The network of haplogroup D4 (Figure S1) demonstrates that the D4o2 sequence shared between Koryaks and Kamchatkan Evens (indicated by arrow 1 in the figure) is more likely to be due to gene flow from Kamchatkan Evens into Koryaks than the other way round, since haplogroup D4 is not at all characteristic of the populations of the Chukotka and Kamchatka Peninsulas. This network furthermore demonstrates that the Nivkh sequences belonging to subhaplogroup D4m2 stem from a haplotype shared with Yakuts (indicated by arrow 2 in Figure S1) and belong to the same branch as sequences found in Evenks, Evens, Yukaghirs, and South Siberian Buryats and Turkic speakers – all populations settled far to the north or west of the Nivkh. The network of D5a2 (Figure S2) demonstrates the effect of Yakut gene flow into neighbouring Tungusic populations: sequences belonging to this haplogroup are found especially in the Iengra Evenks, but are lacking from the Stony Tunguska Evenks, who are the most distant geographically from the Yakuts. Among the Even subgroups, it is found only in the Sakkyryr Evens, a group that has already largely lost its indigenous Even language and now speaks Yakut.

The haplogroup A sequences found in the Stony Tunguska, Nyukzha, and Iengra Evenks, which belong to haplogroup A4, fall with other haplogroup A4 sequences from southern and western Siberian populations (Figure S3). They thus form part of the general geographic region of their provenance, although several mutations separate the Evenk haplotypes from those found in other populations. In contrast, the two Taimyr Evenk individuals, who carry an identical haplogroup A2a sequence, cluster with Koryak, Chukchi, and Eskimo individuals. Even though there is no direct sharing, only one mutation separates this haplotype from a sequence type found in Eskimo. In a previous study based only on HVR1 sequences, a Yakut-speaking Evenk from northwestern Yakutia (in the vicinity of the Taimyr Peninsula) was also found to carry an A2 sequence [14]. This is notable given the great geographic distances separating these populations.

### Maternal population structure and genetic differentiation

As shown by the AMOVA analysis (Table 2), 9.5% of the total variance in the dataset is due to between-population differences. The proportion of the variance due to differences between four basic geographic groups (west, central,

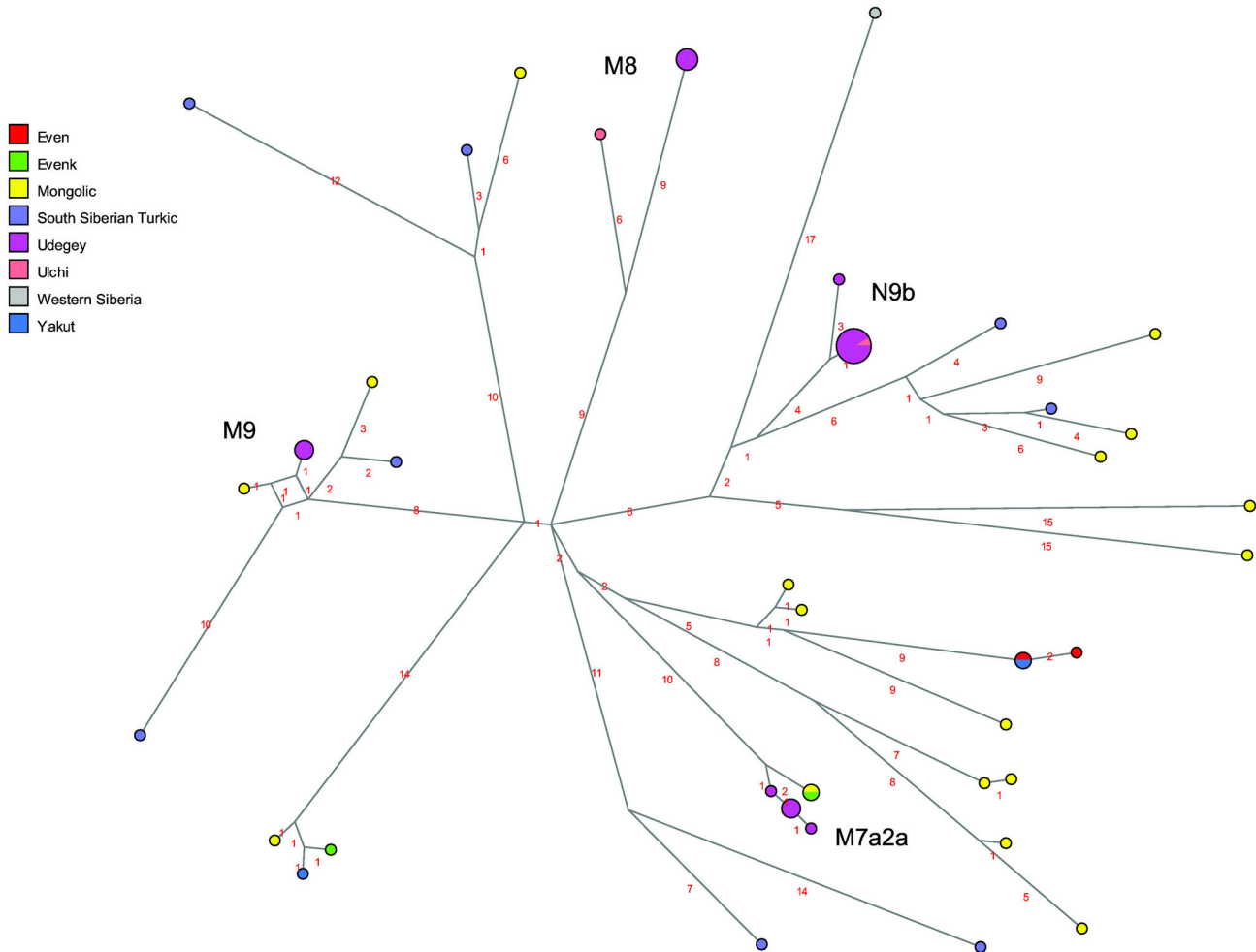


**Figure 3. MJ-network of mtDNA haplogroup C4b.** The North Tungusic haplotypes are coloured by population (Evenks and Evens) rather than subgroup. The size of the nodes is proportional to the number of individuals carrying that node, and the number of mutations is indicated along the branches. The haplotypes marked by arrows and asterisks are discussed in the text.

doi: 10.1371/journal.pone.0083570.g003

northeast, southeast) is smaller than the variance among the populations grouped by this criterion (4% vs. 6.3%), indicating that the geographic location of the populations has not had a major effect in shaping their genetic variation. This might be

due to the extreme mobility of the nomadic North Tungusic groups. In contrast, grouping the populations by linguistic affiliation appears to provide a better fit to the data, with 6.7% of the variance being found between linguistic groups as



**Figure 4. MJ-network of various mtDNA M and N subhaplogroups.** The North Tungusic haplotypes are coloured by population (Evenks and Evens) rather than subgroup. The size of the nodes is proportional to the number of individuals carrying that node, and the number of mutations is indicated along the branches. Subhaplogroups discussed in the text are labelled.

doi: 10.1371/journal.pone.0083570.g004

compared to 4.5% between populations within groups. However, it should be noted that in this grouping there are four groups composed of only one population each (South Tungusic, Chukotko-Kamchatkan, Yukaghir, and Nivkh) – which excludes between-population variance – and is thus not comparable to the grouping based on geography.

When investigating only the Tungusic populations (Evens, Evenks, and Udegey), 8.4% of the variance is due to differences between (sub)populations. Only 5.7% of the variance can be explained by differences between groups when comparing the North Tungusic populations (Evens and Evenks) with the South Tungusic Udegey, with 6.8% of the variance being found between populations in each group. With respect to the North Tungusic populations, 7% of the variation is due to differences between subgroups. That these are due to differences among the individual Evenk and Even subgroups, and not to differences between the Evenk and Even populations, is demonstrated by the non-significant and

negative proportion of variance due to between-group differences when comparing Evenks with Evens, while the proportion of between-population variance is highly significant and quite high when comparing only the Evenk subgroups (8.3%) and only the Even subgroups (6%) (Table 2).

In a three-dimensional MDS plot based on pairwise  $\Phi_{ST}$  values, the Nivkh and Udegey are the biggest outliers (Figure 5, Figure S4 and Figure S5), with all the other populations clustering fairly closely in the first and second dimension (Figure 5). It is noteworthy that neither the Evenk nor the Even subgroups cluster together in the first or second dimension; in the third dimension, however, the Evenks are in relative proximity to each other. There is geographic clustering among the Even subgroups in the second dimension: the westernmost Sebjan and Sakkyryr Evens are proximal to each other, as are the easternmost Kamchatkan and Berezovka Evens, and the Tompo Evens are located in an intermediate position. Nevertheless, the Tompo Evens are closer to the Yukaghirs

**Table 2.** AMOVA analyses based on mtDNA  $\Phi_{ST}$  values.

grouping criterion	between groups	percentage of variance	
		between populations	within populations
<b>1 group (16 (sub)populations)</b>		<b>9.47**</b>	<b>90.53</b>
4 geographic groups	4.01*	6.29**	89.71**
6 linguistic groups	6.66	4.54**	88.80**
<b>1 group (Tungusic family)</b>		<b>8.39**</b>	<b>91.61</b>
2 groups (N. Tungusic vs. S. Tungusic)	5.67	6.77**	87.55**
<b>1 group (North Tungusic)</b>		<b>7.04**</b>	<b>92.96</b>
2 groups (Evenks vs. Evens)	-0.21	7.16**	93.05**
1 group (all Evenks)		8.28**	91.72
1 group (all Evens)		6.00**	94

**4 geographic groups:** West (TAI, STE, NYUK, IENG, VIL\_YAK), Central (SAK, SEB, TOM, CNTRL\_YAK, NE\_YAK), Northeast (BER, KAM, YUK, KOR), Southeast (UDI, NIV)

**6 linguistic groups:** North Tungusic (TAI, STE, NYUK, IENG, SAK, SEB, TOM, BER, KAM), South Tungusic (UDI), Turkic (VIL\_YAK, CNTRL\_YAK, NE\_YAK), Chukotko-Kamchatkan (KOR), Yukaghir, Nivkh

\* P-value < 0.05

\*\* P-value < 0.01

doi: 10.1371/journal.pone.0083570.t002

than to other Even subgroups, as has been observed previously [15] (where the Tompo Evens were called Central Evens); the Kamchatkan Evens are close to the Koryaks in the first and second dimension, although the third dimension separates the Koryaks.

In a correspondence analysis (CA) based on the frequencies of basic haplogroups at the level of resolution shown in Figure 1, the distinct position of the Nivkh and Udegey due to their high frequencies of haplogroups Y1a and various M and N subgroups, respectively, causes all the other populations to cluster together (not shown). When excluding the Nivkh and Udegey, a central cluster is discernible (Figure 6), with the Koryaks, Berezovka Evens, and Kamchatkan Evens in outlying positions. Although the CA slightly distinguishes clusters of different North Tungusic subgroups, such as the Nyukzha and lengra Evenks, or the Tompo and Sebjan Evens with the Taimyr Evenks, overall this plot illustrates the common haplogroup composition present in the central Siberian Evens and Evenks.

A Mantel test of correlations between mtDNA  $\Phi_{ST}$  and geographic distances reveals that this correlation is not significant when considering all populations (Table 3), in good accordance with the results of the AMOVA analysis. The correlation between the geographic and genetic distances among the Evenk and Even populations is also not significant; neither is the correlation only for the Evenks or Evens. However, the correlation is significant for the Even subgroups when excluding the Sakkyryr Evens (Table 3), confirming the geographic cline apparent in the MDS plot (Figure S4).

### Bayesian Skyline Analysis of mtDNA sequences

The Bayesian Skyline plots show three different patterns, illustrated here with the Yakuts (Figure 7A), the Evens (Figure 7B), and the Udegey (Figure 7C); the plots for the Evenks, Yukaghirs, Koryaks, and Nivkh are included in the Material (Figure S6). In the Yakuts, there is a notable decline in population size 3000-5000 years ago (ya) followed by a sharp increase ~1000 ya. This contrasts with the Evenks and Evens, where a notable decline in population size 1000-2000 ya is not followed by an increase. The Udegey, Yukaghirs, Nivkh, and Koryaks all have relatively flat curves with at most a gentle decline close to the present. Furthermore, the effective population size of the Yakuts, Evenks, and Evens is larger than that of the Udegey, Yukaghirs, Nivkh, and Koryaks.

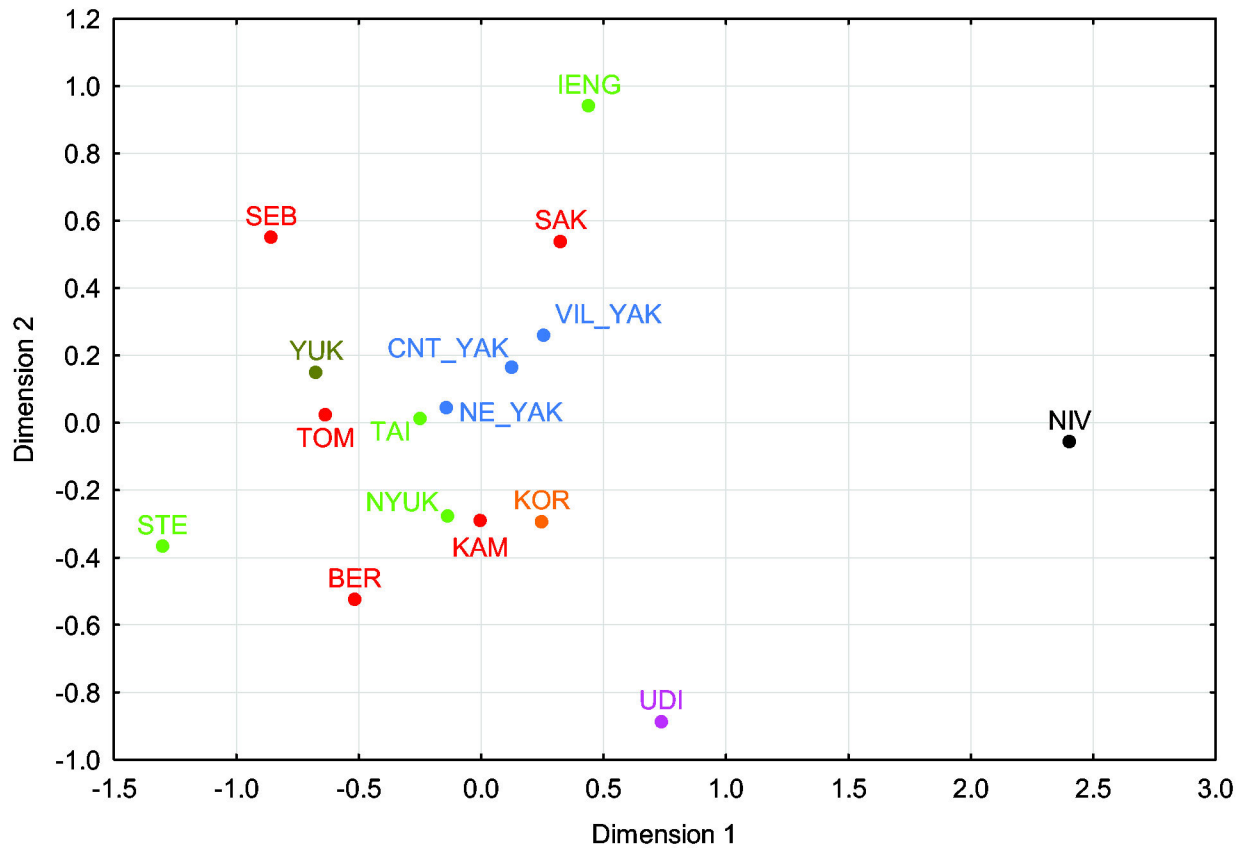
### Y-chromosome analyses

As can be seen in Table 4, haplogroup C3c1 (defined by SNP M86) is the most frequent Y-chromosomal haplogroup found in both the Evenks (56.6%) and the Evens (48.3%). As with the mtDNA haplogroup composition, there are notable differences between subgroups. The Taimyr Evenks, Stony Tunguska Evenks, and Tompo Evens carry high frequencies of haplogroup N1b (38.9%, 27.5%, and 42.9%, respectively), which is otherwise found at low frequency only in the Sakkyryr Evens. Haplogroup N1c, which is highly characteristic of Yakuts [14,17], is found at very high frequency in the two western Even subgroups Sakkyryr (76%) and Sebjan (64.3%) and in high frequency in the lengra Evenks (22.2%) and Nyukzha Evenks (20.5%). In contrast, the eastern Even subgroups Berezovka and Kamchatka carry only haplogroup C3c1.

In a network of Y-STR haplotypes belonging to haplogroup C3c based on nine STRs and including published data, three major haplotypes are apparent in C3c1 (Figure 8). One of these consists entirely of Evenk individuals from different subgroups as well as Yakut-speaking Evenks from northwest Yakutia, in direct neighbourhood to the Taimyr Peninsula; one consists entirely of eastern Evens (Berezovka and Kamchatka) plus one Koryak individual, and the third consists of Even individuals from all five subgroups plus Yukaghirs and one Northeastern Yakut. Only one haplotype, belonging to subgroup C3c\* (defined by the derived state at M48 and the ancestral state at M86, marked by an arrow in Figure 8), is shared between Evenks and Evens (the lengra and the Sakkyryr subgroup, respectively) as well as Yukaghirs. This parallels the findings of the mtDNA haplotype sharing analyses, where only three haplotypes were shared between Evenks and Evens that were not also shared with Yakuts or Yukaghirs.

As becomes apparent from the network of STR haplotypes belonging to haplogroup N1b (Figure 9), the 13 Evens carry only one haplotype, which is shared with Stony Tunguska Evenks; it is also found in one Yakut and one Tuvan. There is somewhat more diversity in the Evenks, with four different haplotypes, of which three are found in the Stony Tunguska Evenks, and two are found in the Taimyr Evenks; only one haplotype is shared between the two Evenk subgroups. The haplotypes found in the Turkic-speaking Tuvans from southern Siberia are distinct from those found in the Tungusic-speaking





**Figure 5. Dimensions 1 vs. 2 of a three-dimensional MDS analysis.** Based on pairwise  $\Phi_{ST}$  values between populations; stress = 0.06. (Sub)population abbreviations as in Table 1, with colours distinguishing the different populations: green = Evenks, red = Evens, pink = Udegey, blue = Yakuts, olive = Yukaghirs, orange = Koryaks, black = Nivkh.

doi: 10.1371/journal.pone.0083570.g005

Evenks and Evens, with the exception of one individual who carries the haplotype also found in Evenks and Evens.

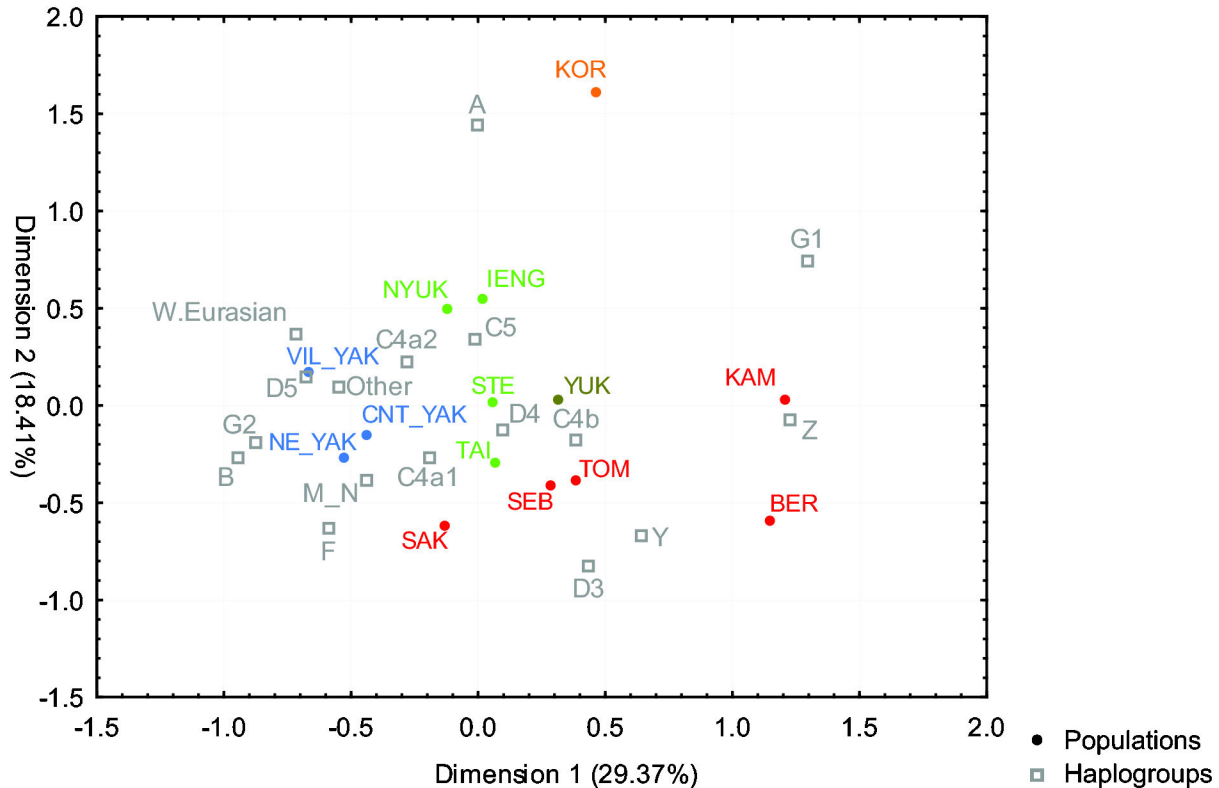
The network of STR haplotypes belonging to haplogroup N1c is highly complex with several multidimensional cycles (Figure 10). As seen previously [15], the network is dominated by haplotypes found in the Yakuts and the Yakut-speaking Evenks. The most common Yakut haplotype is shared by seven Sakkyryyr and two Sebjan Evens as well as one lengra Evenk; the second most common Yakut haplotype is shared by two Sebjan Evens and one lengra Evenk. A further haplotype is shared by one Yakut, one Tompo and one Sebjan Even. A small portion of the network completely lacks Yakut individuals; this is removed from the closest Yakut haplotypes by at least two mutational steps. This contains a haplotype found in one Sebjan and 10 Sakkyryyr Evens, a haplotype shared by Koryaks and Tuvans, and a Yukaghir haplotype. Thus, the network analysis of haplogroup N1c demonstrates that both the lengra Evenk haplotypes are shared with Yakuts, as are one out of two Tompo Even, five out of eight Sebjan Even, and seven out of 18 Sakkyryyr Even haplotypes. Furthermore, one Tompo and one Sakkyryyr Even haplotype differ by only one mutational step from Yakut haplotypes.

## Discussion

### Relationships of Tungusic populations

The major hypotheses about the origins of the North Tungusic Evenks and Evens postulate either a deep split from the Tungusic peoples of the Amur-Ussuri region and a fairly ancient divergence of Evenks and Evens [4] or a much shallower divergence of Evenks and Evens and a fairly recent split of the Northern Tungusic and Amur Tungusic populations [5,6]. The results of the current investigation show that while the Evens and Evenks indeed stem from a common ancestral population, genetic drift and differential admixture with neighbouring but unrelated populations have played a role in shaping their diversity, as was already found in a previous study [15].

The shared paternal ancestry is apparent in the high frequencies of Y-chromosomal haplogroup C3c1 which characterise these populations, cf. [15,17,22,23]. The shared maternal ancestry of Evenks and Evens is evident in the presence at relatively high frequency of mtDNA haplogroups C4a, C4b, and D4, which leads to a lack of differentiation of most subgroups in the CA plot, and especially in the non-



**Figure 6. CA plot.** Based on mtDNA haplogroup frequencies at the level of resolution depicted in Figure 1, excluding Nivkh and Udegey. (Sub)population abbreviations as in Table 1, with colours distinguishing the different populations: green = Evenks, red = Evens, blue = Yakuts, olive = Yukaghirs, orange = Koryaks; haplogroups in grey.

doi: 10.1371/journal.pone.0083570.g006

significant and negative between-group variance in the AMOVA analysis comparing Evenks with Evens. The effect of drift can be seen in the low mtDNA diversity in these populations as well as by their scattered location in the MDS analysis. The effect of gene flow is evident in the affinities of Evenks with Yakuts, and Evens with Yukaghirs emerging in the networks of mtDNA haplogroups C4a and C4b and the haplotype sharing analysis, as well as in the sharing of haplotypes belonging to Y-chromosomal haplogroup N1c with Yakuts.

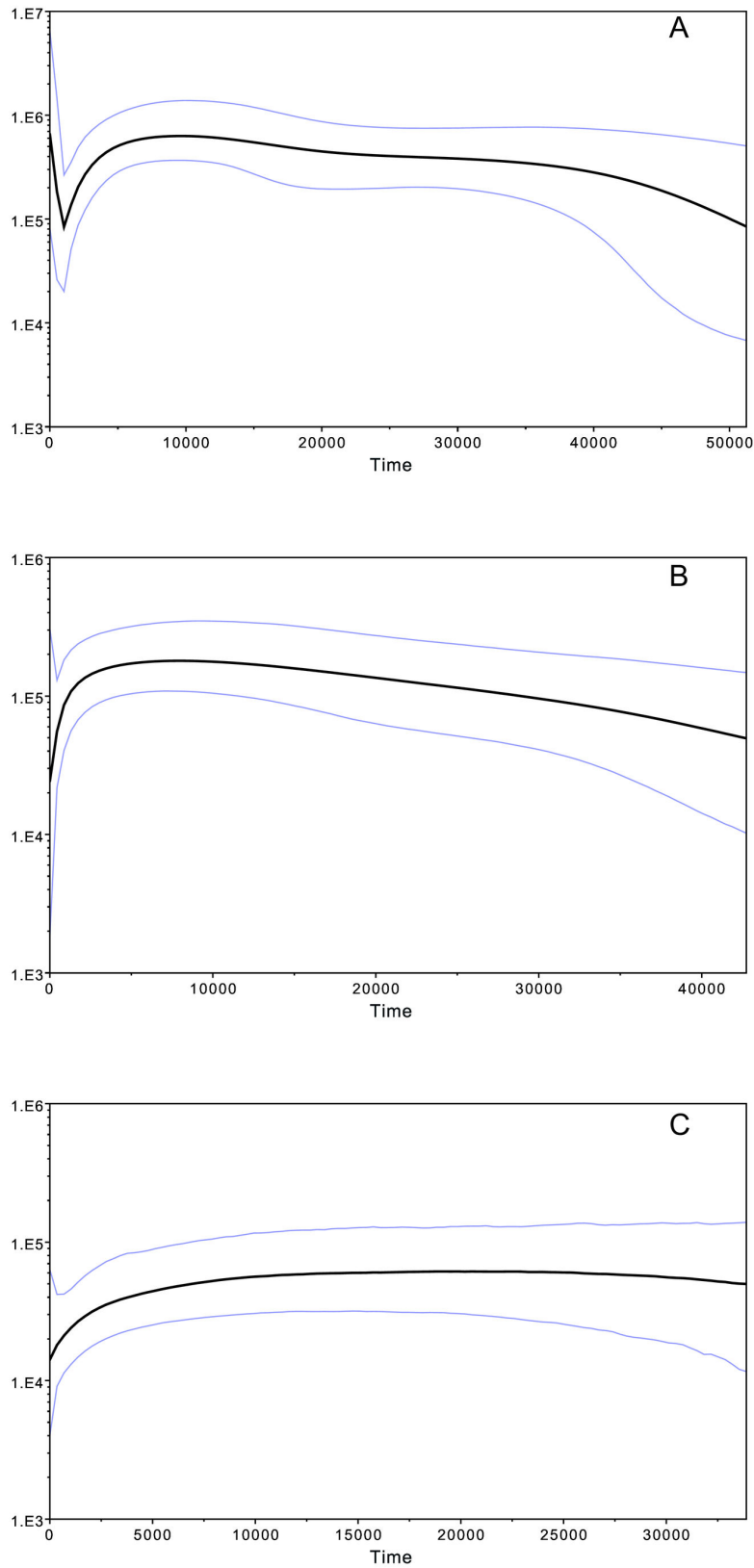
These differential effects of drift and admixture have resulted in significant differences between individual Evenk and Even subgroups, as demonstrated by the AMOVA analysis (Table 2). Surprisingly, these differences are larger among the Evenk subgroups than among the Even subgroups, even though the latter are more widely dispersed geographically. Together with the significant correlation between geographic and genetic distances for four of the Even subgroups (Table 3), this might be an indication that the break-up of the ancestral Even population occurred at a later time than that of the Evenk ancestral population, possibly as a direct result of the Yakut expansion. As for the paternal lineages, although all North Tungusic subgroups have high frequencies of Y-chromosomal haplogroup C3c1, there is no sharing of C3c1 haplotypes between Evenks and Evens, indicating a fairly deep split

**Table 3.** Results of Mantel tests between mtDNA  $\Phi_{ST}$  and geographic distances.

comparison	r	P
all populations	0,24	0,12
Evens_Evens	0,07	0,38
all Evenks	0,12	0,41
all Evens	0,46	0,11
4 Evens (no SAK)	0,71	0,04

doi: 10.1371/journal.pone.0083570.t003

between these populations and lack of subsequent admixture (Figure 8). Using the rho statistic [27] implemented in Network and a mutation rate of 0.0025 per locus per generation [28], we estimate the age of the STR diversity shared by Evenks and Evens to be  $1916 \pm 985$  years. The age of the STR diversity specific to Evens is estimated to be  $615 \pm 418$  years, that specific to Evenks  $303 \pm 177$  years. The split of the Even and Evenk populations is thus estimated to have taken place between approximately 1900 BP and 500 BP – an estimate which brackets both the time of divergence proposed by Vasilevich [4] (~1500 BP) and that proposed by Tugolukov [5] and Janhunen [6] (sometime after 800 BP and possibly as



**Figure 7. Bayesian Skyline Analysis plots.** Based on complete mtDNA genome sequences and a strict clock model. A: All Yakuts; B: All Evens; C: Udegey.

doi: 10.1371/journal.pone.0083570.g007

**Table 4.** Y-chromosomal haplogroup frequencies.

population	N	C*	C3*	C3c*	C3c1	F*	I*	J2*	N1b*	N1c	O*	Q1*	R1a*	R1a1*	reference
<b>all Evenk</b>	<b>127</b>		<b>13</b>	<b>2</b>	<b>82</b>	<b>1</b>	<b>5</b>	<b>1</b>	<b>18</b>	<b>18</b>	<b>1</b>		<b>4</b>		
TAI	18				8				7				3		this study
STE	40				28		1 <sup>§</sup>		11						Pakendorf et al. 2006
NYUK	78		13	n/a	42	1	4	1		16			1		Karafet et al. 2002
IENG	9			2	4					2	1 <sup>§§</sup>				Pakendorf et al. 2007
<b>all Even</b>	<b>89</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>43</b>				<b>13</b>	<b>30</b>					
SAK	25			1	4				1	19					Pakendorf et al. 2007, this study
SEB	14	1			4					9					Pakendorf et al. 2007, this study
TOM	28		1		13				12	2					Pakendorf et al. 2007, this study
BER	7				7										this study
KAM	15				15										this study
<b>all Yakut</b>	<b>184</b>		3		1			1	1	173		1	4		Pakendorf et al. 2006
<b>Yukaghir</b>	<b>13</b>		1	1	2	1				4		4			Pakendorf et al. 2006

<sup>§</sup> note that in the original publication, this STE individual had been erroneously genotyped as belonging to haplogroup F; retyping showed he belonged to haplogroup I-M170

<sup>§§</sup> note that in the original publication, this IENG individual had been erroneously genotyped as belonging to haplogroup K; retyping showed he belonged to haplogroup O-M175

doi: 10.1371/journal.pone.0083570.t004

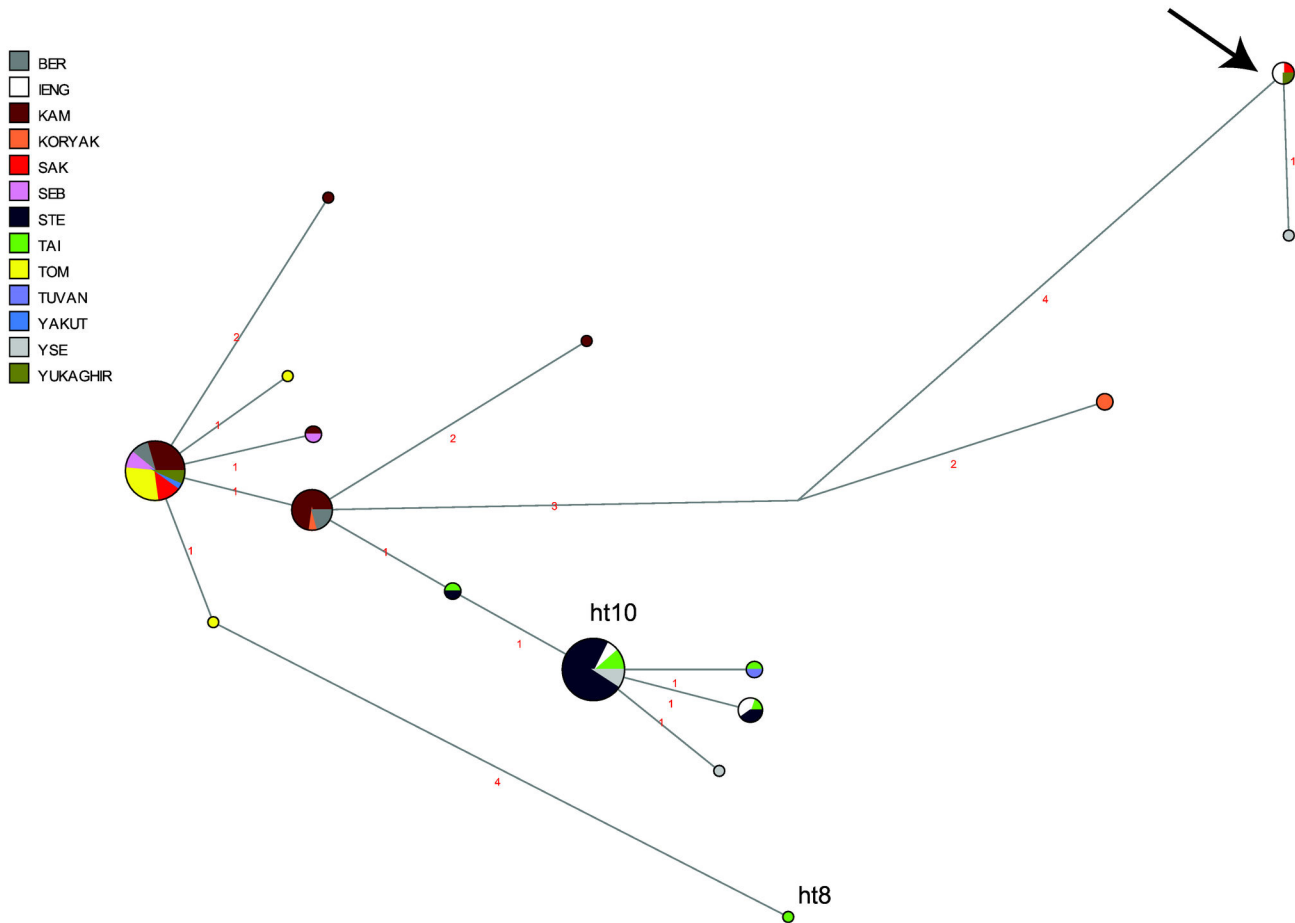
recently as 400 BP). In a study of haplogroup C Y-chromosomes in northern Asia, Malyarchuk et al. [29] date the split between the Evenk and Even haplotypes to either 1400 ( $\pm$  1060) or 390 ( $\pm$  290) years ago, depending on whether the evolutionary or the pedigree-based mutation rate is used. The question of which mutation rate to use is still not settled, but it has been suggested for mtDNA data that it is preferable to use the pedigree rate in studies of relatively recent history, while the phylogenetic rate is preferable in studies of deep history [30,31]. Furthermore, a previous study of Yakut origins using the same set of Y-chromosomal STRs [14] found that the age resulting from the pedigree-based mutation rate was more compatible with linguistic, historical, and archaeological data for the origins of the Yakuts. Taking these circumstantial arguments into account lends greater weight to the younger date estimated by Malyarchuk et al. [29] and thus favours the late split between the Evenks and Evens suggested by Janhunen [6] and Tugolukov [5], and contradicts the early split proposed by Vasilevič [4].

As to the Tungusic populations of the Amur-Ussuri region, in the maternal line the Udegey appear quite distinct from their linguistic relatives the Evenks and Evens with respect to their haplogroup composition (Table S1). Nevertheless, although the second dimension of an MDS analysis separates the Udegey from the Siberian populations, in the first and third dimension they cluster with Yakuts and the Iengra Evenks (Figure S4). Similarly, in the AMOVA analysis, less than 6% of the variance is due to differences between the South Tungusic Udegey and the North Tungusic Evenks and Evens, in contrast to nearly 7% of the variance found among the different North Tungusic subgroups (Table 2), indicating that there are some affinities of the Udegey with their linguistic relatives. A direct link between the Udegey and the Evenks is evident in mtDNA haplogroup C4b: as shown in the network (Figure 3), the Udegey haplotypes are derived from a sequence type that is shared with all the Evenk subgroups. Links between the Udegey and

southern Siberian populations are also apparent in the presence of haplogroup M7 and M9 (Table S1, Figure 4). These are found at low frequency in Mongolic-speaking populations, Tuvans from southern Siberia, and the Ulchi, a South Tungusic population of the Amur river; furthermore, M7 is found in Koreans with a frequency of nearly 10% [10,32]. The high frequency of haplogroup N9b and the presence of haplogroup Y1a, however, connects the Udegey with other populations of the Amur-Ussuri region rather than Siberia, as also shown in another study: in a PC analysis based on frequencies of mtDNA subhaplogroups the Udegey cluster with the Nivkh and the South Tungusic Ulchi of the Amur-Ussuri region, as well as, at a slight distance, with the North Tungusic Negidal and the Kamchatkan Koryaks and Itelmen [10]. This might indicate that the Udegey resulted from admixture between indigenous populations of the Amur-Ussuri region, such as those represented by the prehistoric "Okhotsk people" [33], and Tungusic-speaking immigrants from southern Siberia. This hypothesis of large-scale mixing in the prehistory of the Udegey is further confirmed by their combination of very low sequence diversity, which is as low as that found in the Nivkh, with intermediate levels of nucleotide diversity (Table 1). Alternatively, the shared haplotypes found in the Evenks and Udegey might be retentions from an earlier shared ancestral population; this would be in accordance with the proposed origin of the Tungusic languages and populations in the Amur region [6].

#### Gene flow in the history of individual North Tungusic subgroups

With respect to the individual North Tungusic subgroups, the Taimyr Evenks stand out in showing connections in the maternal line with populations far to the east, notwithstanding their settlement in northwestern Siberia. Thus, two individuals carry a sequence type belonging to mtDNA haplogroup A2a,



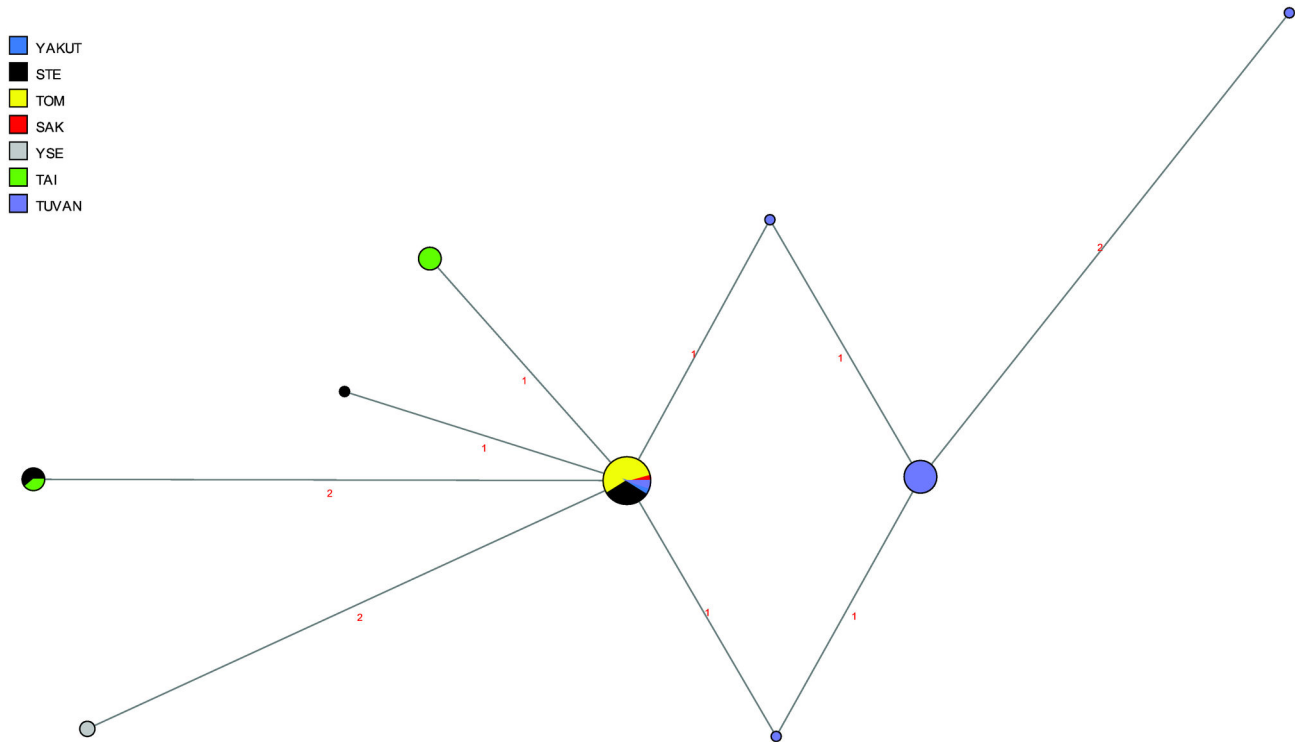
**Figure 8. MJ-network of Y-chromosomal haplogroup C3c.** Based on nine Y-STRs. The North Tungusic haplotypes are coloured by subgroup. The size of the nodes is proportional to the number of individuals carrying that node, and the number of mutations is indicated along the branches. The haplotype marked by an arrow is discussed in the text. Note that the placement of the Taimyr Evenk haplotype #8 as undertaken by Network is erroneous; as evident from Table S3, this haplotype is only two mutational steps distant from haplotype #10, also labelled in the figure.

doi: 10.1371/journal.pone.0083570.g008

which is practically absent in western and central Siberia, but found in high frequency in the Siberian Eskimos and Chukchi and in low frequency in Koryaks and Chuvantsi, all populations settled far to the east of the Taimyr Peninsula [10,34]. Another haplogroup that links the Taimyr Evenks with more easterly populations is Y1a, present in 12.5% of the Taimyr Evenks, but in no other Evenk subgroup of our sample. This haplogroup is found in very high frequency in the Nivkh (65.8%); it is also found in the Udegey and in the eastern Even subgroups Kamchatka and Berezovka (Table S1). It is furthermore present in high frequency in the Ulchi and Negidal, Tungusic populations of the Amur-Ussuri region and in 9%-10% in eastern Evenks as well as Koryaks from Kamchatka [10,20,32], as well as in low frequency in Central and Vilyuy Yakuts [17] (Table S1). How the presence of these haplogroups of far eastern provenance can be explained among the Taimyr Evenks is not clear; they are not found in their neighbours on the Taimyr Peninsula, the Turkic-speaking Dolgans or the

Samoyedic Nganasan [10,17]. In their paternal lineages, the Taimyr Evenks do not differ from the Stony Tunguska Evenks, who are settled to their south: they have high frequencies of haplogroups C3c and N1b, and share three out of five C3c-STR haplotypes with the Stony Tunguska and Iengra Evenks and one of their two N1b-STR haplotypes. In addition, they show evidence of European admixture in the paternal line: three individuals carry haplogroup R1a (Table 4).

The Iengra Evenks are settled in the south of Yakutia in a region where intense contact and intermarriage between Yakuts and Evenks has been recorded [35], and they show clear evidence of Yakut admixture in both the paternal and the maternal line, sharing two Y-chromosomal N1c-STR haplotypes and two out of three mtDNA haplogroup D5a2a2 sequences with Yakuts. Although they are separated in the second dimension of the MDS analysis, probably in result of genetic drift [15], their close maternal affinities with their linguistic and geographic neighbours, the Nyukzha Evenks, are



**Figure 9. MJ-network of Y-chromosomal haplogroup N1b.** Based on nine Y-STRs. The North Tungusic haplotypes are coloured by subgroup. The size of the nodes is proportional to the number of individuals carrying that node, and the number of mutations is indicated along the branches.

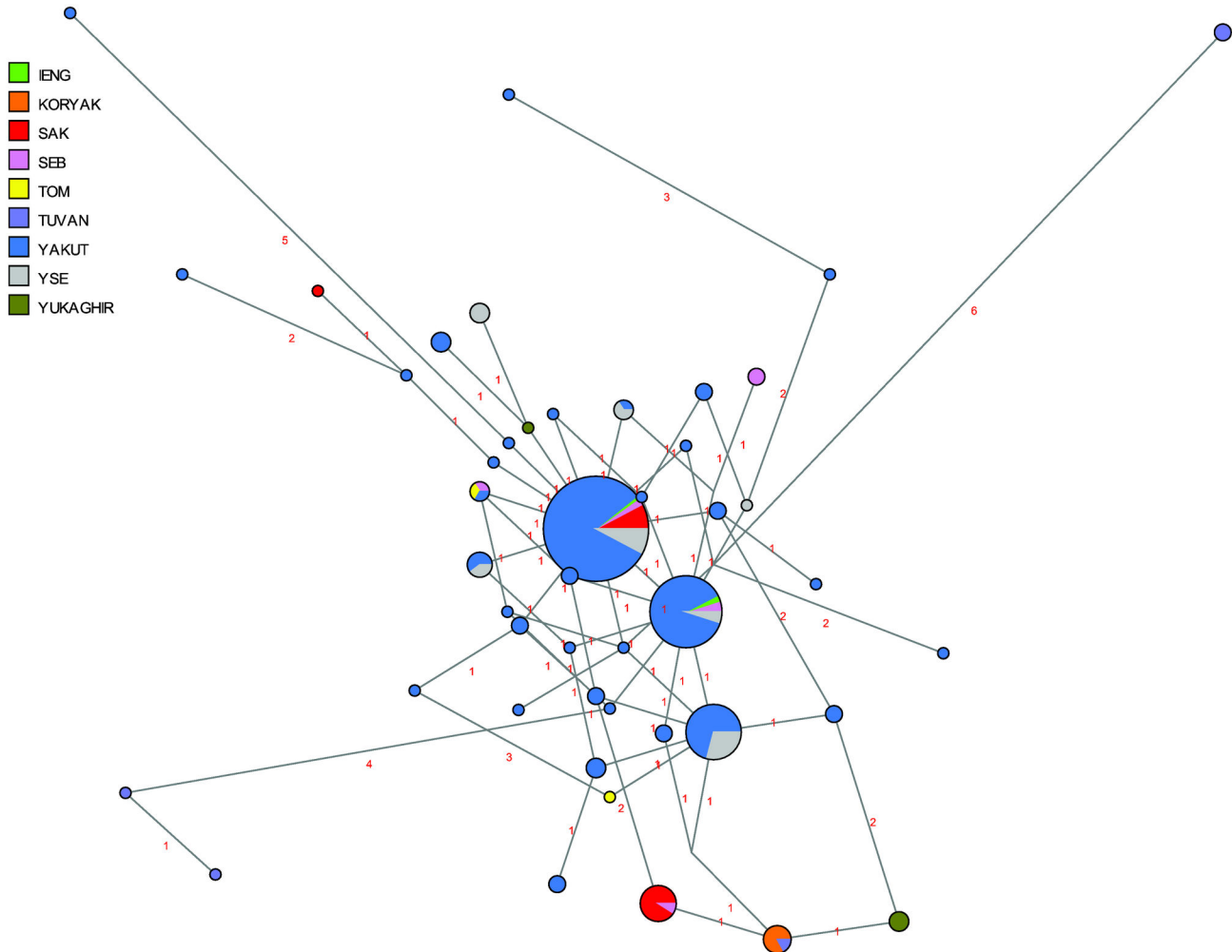
doi: 10.1371/journal.pone.0083570.g009

evident in the four haplotypes (31% of the lengra mtDNA haplotypes) shared exclusively with the latter (Table S2).

The Sakkyryyr Evens also show evidence of Yakut admixture in both the maternal and the paternal lines. Thus, rather than clustering with their closest geographic and linguistic neighbours, the Sebjan Evens, they are pulled towards the Yakuts in the MDS and CA plots, and they share nine out of their 19 mtDNA haplotypes (47.4%) with the Yakuts, four of these uniquely (Table S2). Two of these belong to haplogroup F1b and one to haplogroup M7, which are characteristic of populations of southwestern and southern Siberia rather than those of northeastern and central Siberia [10,18,32,36–38]. Furthermore, one shared haplotype belongs to haplogroup D5a2a2, which is characteristic of the Yakuts. This indicates that the direction of gene flow in these cases is likely to have been from Yakuts – who retain a signature of their southern origins [14,17] – into the Sakkyryyr Evens. This Yakut maternal admixture might explain the high sequence and nucleotide diversity values found in the Sakkyryyr Evens in comparison to other North Tungusic subgroups (Table 1). Likewise, the network analysis of Y-chromosomal haplogroup N1c (Figure 10) provides a clear indication of gene flow from Yakuts into the Sakkyryyr Evens, in that seven STR haplotypes are shared with Yakuts, with an additional haplotype separated by only one mutational step. Furthermore, a haplotype that was not included in the analysis because of a duplication in DYS393 is

also shared with Yakuts (with the exception of the duplication [15]). Thus, 32% - 36% of the Sakkyryyr Even Y-chromosomes are most likely to be of Yakut origin. This indicates that the Sakkyryyr Even population incorporated entire Yakut communities, both men and women, during their history, and this might partly explain why the Sakkyryyr Evens have given up their North Tungusic language in favour of Yakut.

In contrast, the neighbouring Sebjan Evens do not show evidence of substantial amounts of Yakut gene flow in the maternal line: they are located at a distance from the Yakuts in the MDS plot and cluster with the Taimyr Evenks and the Tompo Evens in the CA analysis. Furthermore, they share only three of their eight haplotypes (37.5%) with Yakuts, none uniquely, and two of these shared haplotypes belong to haplogroups C4a1c and D4I2 characteristic of North Tungusic populations (Table S2). These shared haplotypes found in the Sebjan Evens might therefore be due to maternal Tungusic admixture into Yakuts during the expansion of the latter [14]. However, the proportion of Yakut male admixture in the Sebjan Evens is higher than that detected in the Sakkyryyr Evens: five Y-STR haplotypes belonging to haplogroup N1c are directly shared with Yakuts, and two more are separated from a common Yakut haplotype by two mutational steps. Furthermore, a haplotype not included in the network analysis because it contained a duplication at DYS393 is also shared with Yakuts (again excluding the duplication [15]). This



**Figure 10. RM-MJ network of Y-chromosomal haplogroup N1c.** Based on nine Y-STRs. The North Tungusic haplotypes are coloured by subgroup. The size of the nodes is proportional to the number of individuals carrying that node, and the number of mutations is indicated along the branches.

doi: 10.1371/journal.pone.0083570.g010

indicates that approximately half (43% - 57%) of the Sebjan Even paternal genepool is of Yakut origins and indicates male-biased admixture from Yakuts to Evens rather than the incorporation of both maternal and paternal lineages detected in the Sakkyryr Evens. Interestingly, although the Sebjan Evens have retained their Even language, this has undergone striking changes under the influence of Yakut [39].

The Tompo Evens are close to Yukaghirs in the MDS analysis based on mtDNA sequences, as was seen previously [15] (where this subgroup was called Central Evens). However, only two haplotypes are shared between Tompo Evens and Yukaghirs, and these are also shared with Stony Tunguska, Nyukzha, and Iengra Evenks and Kamchatkan Evens (Table S2). In the paternal line, the Tompo Evens are characterized by high frequencies of Y-chromosomal haplogroup N1b (Table 4), with all the individuals belonging to this haplogroup carrying a STR haplotype that is shared with Evenks. This haplogroup is

found at highest frequency in the Samoyedic-speaking populations of northwestern Siberia, the Nganasan, Enets, and Nenets [22,24], and in high frequency in some Turkic-speaking populations of southern Siberia [26]. It is furthermore characteristic of North Tungusic populations of western Siberia, but is absent in those of the east: it is found in 39% in the Taimyr Evenks and 27.5% in the Stony Tunguska Evenks (Table 4), in 16.7% in a sample of 'Western Evenks' and in 24.4% in a geographically undefined sample of 'Evenks', but in only 2.5% in Chinese Evenks and is absent in Eastern Evenks, Evens, and Oroqen, who are closely related culturally and linguistically to Evenks [22,26]. The presence of this haplogroup in the Tompo Evens therefore appears indicative of paternal gene flow from Evenks. However, it is intriguing that this Evenk admixture led only to an incorporation of Y-chromosomal haplogroup N1b in the Tompo Evens, without any C3c1 Y-chromosomes being transferred. Possibly this

presumed admixture event was restricted to one individual, with subsequent expansion of his Y-chromosomal lineage, as indicated by the fact that all the Tompo Evens carrying haplogroup N1b share the same STR haplotype.

The mtDNA haplogroup composition of the eastern Even subgroups, Berezovka and Kamchatka, with substantial frequencies of the characteristic eastern Siberian and Kamchatkan mtDNA haplogroups G1, Y1a and Z [10,20], indicates a certain amount of admixture with local populations during their eastward expansion. In contrast, their Y-chromosomal gene pool consists solely of haplogroup C3c1, which is typical of North Tungusic populations, indicating that these interactions with local groups were biased towards the female line [15]. On Kamchatka, although the Evens have 15.4% of mtDNA haplogroup G1, which is found in very high frequencies in Kamchatkan Koryaks and Itelmen [20], there is no direct evidence for recent gene flow from Koryaks into the Even population. In contrast, gene flow from Evens into Koryaks can be shown to have taken place, both in the presence of a shared sequence type belonging to mtDNA haplogroup D4 and in the sharing of a Y-chromosomal haplotype belonging to haplogroup C3c1.

### Demographic history of the populations

The mtDNA diversity values and Bayesian Skyline analyses (Table 1, Figure 7, Figure S6) show substantial differences in demographic history between the populations included in this study. The Yakuts stand out as having higher diversity values than most of the other populations; this is in good accordance with their much larger census population size in comparison to that of the other populations (478,000 Yakuts vs 38,000 Evenks, 21,000 Evens, 1500 Udegey, 1600 Yukaghir, 4600 Nivkh, and 7900 Koryaks in 2010 [40]). They furthermore show an increase in population size approximately 1000 years BP (Figure 7A); this fits with previous estimates based on the Y-chromosome and mtDNA [14,41] and is also in relatively good accordance with archaeological data showing that the ancestors of the Yakuts migrated to the north in the 13<sup>th</sup> or 14<sup>th</sup> centuries CE [42,43]. The Evenk and Even subgroups generally have lower diversity values than the Yakuts, with the exception of the Sakkyryyr Evens. The Udegey, Yukaghirs, Nivkh, and Koryaks, who are traditionally predominantly sedentary hunter-gatherers and especially fishermen, all have relatively low mtDNA diversity values and basically flat BSP curves with a lower effective population size than that found in the Evenks and Evens and especially Yakuts. This pattern, which indicates that they have had a constantly low effective population size, differs from the pattern found in the North Tungusic populations, who not only have a higher effective population size, but where a slight increase is apparent before the recent decline. This might reflect the differences in life-style between the sedentary hunter-gatherers and the nomadic Evenks and Evens, whose exceptional mobility would have increased their pool of potential spouses. The decline in population size evident in the BSPs for the Evens and Evenks overlaps the increase appearing in the Yakuts; it is therefore possible that the northward expansion of the Yakuts resulted in

a decrease in population size of the North Tungusic populations.

### Conclusions

Our analysis of complete mtDNA genome sequences and Y-chromosomal SNP and STR variation has revealed a differentiated picture of the relationships among Tungusic populations. Although the Amur Tungusic Udegey are linguistic relatives of the North Tungusic Evenks and Evens, in the maternal line most traces of their possible genetic relationship have been erased by the effects of drift and admixture both in the Siberian populations as well as in the Udegey. As to the Evenks and Evens, while they retain evidence of their shared ancestry, admixture with neighbouring populations, especially in the maternal line, has led to their genetic differentiation. In-depth investigations of the autosomal variation in these populations will enable deeper insights into their population history and relationships among each other and with other populations of Siberia.

### Materials & Methods

#### Ethics statement

The collection of the samples was approved by the Ethics Committee of the University of Leipzig, the Ethics Committee of the Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, and the Institutional Review Board of the University of Kansas. Written and/or oral consent to use their samples was obtained from all donors after explanation of the aims of the study. The Institutional Review Board of the University of Kansas approved the substitution of oral informed consent in lieu of written consent in the case of the Udegey, Koryaks, and Kamchatkan Evens, since donors were suspicious of signing written statements. The oral consent was witnessed by the Russian team members of the sampling expedition.

#### mtDNA analyses

We generated a total of 525 full genome mtDNA sequences from 130 Evenks belonging to four subgroups (from northwest to southeast: Taimyr, Stony Tunguska, Nyukzha, and lengra), 122 Evens belonging to five subgroups (from west to east: Sakkyryyr, Sebjan, Tompo, Berezovka, and Kamchatka), 31 Udegey, 169 Yakuts belonging to three subgroups (Vilyuy, Central, and Northeast), 20 Yukaghirs, 15 Koryaks, and 38 Nivkh (Figure 1) using the method described in the Supplementary Materials of Barbieri et al. [44]. The Taimyr Evenk samples were collected in the village of Khantayskoe Ozero, the Sebjan and Berezovka samples in the villages of Sebjan-Küöl and Berezovka (Yakutia), respectively, the Kamchatkan Evens and Koryaks in the villages of Esso and Anavgai, the Nivkh were collected in northern Sakhalin, and the Udegey were collected in the village of Gavsjugi (Khabarovsk Region). For details on the sample locations for the other (sub)populations see 14,15.

The sequences were generated with an Illumina Genome Analyzer Ix sequencer to an average coverage of 274x and full



sequences were deposited in GenBank under accession numbers KF148067-KF148359 and KF148361-KF148592. Three different alignments were generated for analysis. For calculation of pairwise  $\Phi_{ST}$ , standard mtDNA diversity indices, and Analysis of Molecular Variance (AMOVA) in Arlequin v3.5 [45] all indels, positions with unclear and missing data (coded as N's) and the poly-C region (16183 - 16194) were removed, leaving an alignment of length 16,507 bp. For the analysis of haplotype sharing using an in-house Python script as well as Bayesian Skyline Plots (BSP) produced with BEAST, a second alignment was generated which omitted only the poly-C region and positions containing Ns, but included all indels, resulting in an alignment length of 16,521 bp. The third alignment was used to construct Median-Joining networks; it was 16,478 bp in length to accommodate published data downloaded from GenBank. Haplogroups were assigned using the online tool Haplogrep [46] in reference to PhyloTree Build 15 [47], using rCRS [48] as a reference.

For each population and subgroup the ideal substitution model was calculated using jModelTest v 2.1 [49]. With the information as to the appropriate substitution model, each (sub)population was then tested for adherence to a molecular clock in MEGA v5 [50]; for none of the (sub)populations was the null hypothesis of a molecular clock rejected. Bayesian Skyline plots (BSP) were then generated using the BEAST package v1.6 [51,52], partitioning the data between the coding (577 - 16,023) and non-coding (16,024 - 576) regions and applying the mutation rates from Soares et al. [53] ( $1.708E^{-8}$  and  $9.883E^{-8}$ , respectively), using a strict clock model.

Median Joining (MJ) networks [54] were generated using the Network 4.6 and Network Publisher v1.3 programmes ([www.fluxus-engineering.com](http://www.fluxus-engineering.com)), with transversions given a threefold higher weight than transitions. Correspondence analysis (CA) and multi-dimensional scaling (MDS) plots were generated with STATISTICA v10 [55]. Mantel tests of the correlation between geographic great-circle distances – calculated with the R package “geosphere” [56] – and  $\Phi_{ST}$  genetic distances between (sub)populations were performed in R using the “ade4” package [57].

Published mtDNA genome sequences were downloaded from GenBank for inclusion in the network analyses; these included additional Yukaghir, Evenk, Even, Udegey, Koryak, and Yakut sequences as well as Mongolic-speaking Buryats and Khamnigans, Turkic-speaking Tofalar, Tubalar, Altai-Kizhi, Shor, Teleut, Tuvan, and Kazakh from South Siberia, Kets, who speak an isolate language, and Finno-Ugric-speaking Mansi from western Siberia, Samoyedic-speaking Nganasan from the Taimyr Peninsula, Tungusic-speaking Negidal and Ulchi from the Amur-Ussuri region, Chukchi from Chukotka who speak a Chukotko-Kamchatkan language, and Eskimos [10,16,18,32,34].

### Y-chromosomal analyses

To investigate the paternal relationships of the Evenks and Evens, Y-chromosomal SNPs were genotyped in the Sebjan, Berezovka, and Kamchatkan Evens as well as in the Taimyr Evenks following the protocol of de Filippo et al. [58]; subhaplogroups of C and N were defined by genotyping the

SNPs M48, M86, and P43 following the method of Nasidze et al. [59] and Pakendorf et al. [14], and were named according to the nomenclature of Karafet et al. [60]. In addition, 12 short tandem repeat (STR) loci were genotyped using the Promega Y-Powerplex kit (<http://www.promega.com>). The newly generated Y-chromosomal genotypes were compared to previously published Y-chromosomal data [14,15,61]; for this purpose, the Even individuals were reassigned from the previous subdivision of Western and Central Evens [15] into the subdivision followed here, namely Sakkyryyr, Sebjan, and Tompo. While the Y-STR data published by Rubicz et al. [61] were not associated with haplogroup information, this was obtained from Rubicz [62] for the Koryaks and Evens. For network analyses of STR haplotypes belonging to haplogroups C3c, N1b, and N1c only nine STRs overlapping between the datasets were used, and individuals with missing data or duplications at DYS393 were excluded. This concerned one Berezovka Even belonging to haplogroup C3c1 with missing data, one Sakkyryyr and one Sebjan Even belonging to haplogroup N1c with a duplication at DYS393, and two Tompo Evens belonging to haplogroup C3c1 with a duplication at DYS393. For the network analyses, the STRs were assigned weights inversely proportional to their variance within each haplogroup [63]. Since the network of haplogroup N1c was very complex with a large number of multidimensional reticulations, a joined Reduced Median and Median Joining analysis was performed. The newly generated Y-chromosome genotypes are available in Table S3.

The age of the Y-STR diversity accumulated in Evenks and Evens belonging to haplogroup C3c1 was estimated with the rho statistic [27] implemented in Network, using a mutation rate of 0.0025 per locus per generation [28]. Assuming a male generation interval of 30 years [64] and including nine loci, this amounts to one mutation every 1333 years.

### Supporting Information

**Figure S1. MJ-network of mtDNA haplogroup D4.** The North Tungusic haplotypes are coloured by population (Evenks and Evens) rather than subgroup. The size of the nodes is proportional to the number of individuals carrying that node, and the number of mutations is indicated along the branches. The haplotypes marked by arrows are discussed in the text. (EPS)

**Figure S2. MJ-network of mtDNA haplogroup D5a2.** The North Tungusic haplotypes are coloured by subgroup. The size of the nodes is proportional to the number of individuals carrying that node, and the number of mutations is indicated along the branches. (EPS)

**Figure S3. MJ-network of mtDNA haplogroup A.** The North Tungusic haplotypes are coloured by population (Evenks) rather than subgroup. The size of the nodes is proportional to the number of individuals carrying that node, and the number of mutations is indicated along the branches. Subhaplogroups discussed in the text are labelled.

(EPS)

**Figure S4. Dimensions 1 vs. 3 of a three-dimensional MDS analysis.** Based on pairwise  $\Phi_{ST}$  values between populations; stress = 0.06. (Sub)population abbreviations as in Table 1, with colours distinguishing the different populations: green = Evenks, red = Evens, pink = Udegey, blue = Yakuts, olive = Yukaghirs, orange = Koryaks, black = Nivkh.  
(EPS)

**Figure S5. Dimensions 2 vs. 3 of a three-dimensional MDS analysis.** Based on pairwise  $\Phi_{ST}$  values between populations; stress = 0.06. (Sub)population abbreviations as in Table 1, with colours distinguishing the different populations: green = Evenks, red = Evens, pink = Udegey, blue = Yakuts, olive = Yukaghirs, orange = Koryaks, black = Nivkh.  
(EPS)

**Figure S6. Bayesian Skyline Analysis plots.** Based on complete mtDNA genome sequences and a strict clock model. A: All Evenks, B: Nivkh, C: Koryaks, D: Yukaghir.  
(EPS)

**Table S1. List of mtDNA (sub)haplogroups found in the dataset and their frequencies in individual (sub)populations.**

(XLS)

**Table S2. Analysis of shared haplotypes.**  
(XLS)

**Table S3. Y-chromosomal genotypes generated for this study.**  
(XLS)

## Acknowledgements

The goal of this study is to elucidate the prehistory of populations as reflected in their genetic variation. It does not intend to evaluate the self-identification or cultural identity of any group, which consist of more than simply genetic ancestry. We sincerely thank all the donors of genetic samples for their participation in this investigation, Rem Sukernik and Dejan Matic for assistance with sample collection, Mark Stoneking for helpful discussion, Torsten Blass and Gabriel Renaud for assistance with R and the haplotype sharing analysis, and two anonymous reviewers for comments.

## Author Contributions

Conceived and designed the experiments: BP MW. Performed the experiments: MW AB. Analyzed the data: ATD BP. Contributed reagents/materials/analysis tools: VW MC VS SM IN VO BP. Wrote the manuscript: BP ATD.

## References

- Spevakovsky AB (1994) Even. In: P FriedrichN Diamond. Encyclopedia of World Cultures; —Russia and Eurasia/China. New York: GK Hall & Company. pp. 115–119
- Black LT (1988) Peoples of the Amur and Maritime Regions. In: W FitzhughA Crowell. Crossroads of Continents: Cultures of Siberia and Alaska. Kensington, MD: Smithsonian Institution Press. pp. 24–31.
- Smoljak AV (2001) Narody Nižnego Amura i Saxalina. Fotoal'bom. Moscow: Nauka. 317 pp.
- Vasilevič GM (1969) Èvenki. *Istoriko-ètnografičeskie očerki (XVIII-načalo XX v.)*. Leningrad: Nauka Leningradskoe otdelenie. 303 p
- Tugolukov VA (1980) Ètničeskie korni tungusov. In: IS Gurvič. Ètnogenez narodov severa. Moscow: Nauka. pp. 152–176.
- Janhunen J (1996) Manchuria. An Ethnic History. Helsinki: The Finno-Ugrian Society. 335 pp.
- Dolgiĥ BO (1960) Rodovoy i plemennoy sostav narodov Sibiri v XVII veke. Moscow: Izdatel'stvo Akademii Nauk SSSR. 620 pp.
- Wurm SA (1996) Siberia: 1650-1950 ethnic and linguistic changes. In: SA WurmP MühlhäuslerDT Tryon. Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas. Berlin, NY: Mouton De Gruyter. pp. 969–974.
- Kočeškov NV, Samar AP, Tugolukov VA (1997) Istorija i kul'tura evenov. St Petersburg: Nauka. 179 p.
- Sukernik RI, Volodko NV, Mazunin IO, El'tsov NP, Dryomov SV et al. (2012) Mitochondrial genome diversity in the Tubalar, Even, and Ulchi: contribution to prehistory of native Siberians and their affinities to Native Americans. *Am J Phys Anthropol* 148(1): 123–138. doi:10.1002/ajpa.22050. PubMed: 22487888.
- Vasilevič GM (1948) Očerki dialektov èvenkijskogo (tungusskogo) jazyka. Leningrad: Gosudarstvennoe učebno-pedagogičeskoe izdatel'stvo Ministerstvo prosvěščenija RSFSR, Leningradskoe otdelenie. 352 p.
- Burykin AA (2004) Jazyk maločislennogo naroda v ego pis'mennoj forme. Sociolingvističeskie i sobstvenno lingvističeskie aspekty. St Petersburg: Peterburgskoe Vostokovedenie: 373.
- Taksami ĆM (2007) Nivxi. Ètničeskaja kul'tura nivxov – aborigenov tixookeanskogo severa i dal'nego vostoka. Novosibirsk: Nauka. 226 p.
- Pakendorf B, Novgorodov IN, Osakovskij VL, Danilova AP, Protod'jakonov AP et al. (2006) Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum Genet* 120(3): 334–353. doi:10.1007/s00439-006-0213-2. PubMed: 16845541.
- Pakendorf B, Novgorodov IN, Osakovskij VL, Stoneking M (2007) Mating patterns amongst Siberian reindeer herders: Inferences from mtDNA and Y-chromosomal analyses. *Am J Phys Anthropol* 133: 1013-1027. doi:10.1002/ajpa.20590. PubMed: 17492671.
- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Rogalla U et al. (2010) Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern. Asia - PLOS ONE 5(12): e15214. doi: 10.1371/journal.pone.0015214.
- Fedorova SA, Reidla M, Metspalu E, Metspalu M, Rootsi S et al. (2013) Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol Biol* 13(1): 127. doi:10.1186/1471-2148-13-127. PubMed: 23782551.
- Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E et al. (2005) Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet* 69(1): 67–89. doi:10.1046/j.1529-8817.2003.00127.x. PubMed: 15638829.
- Jin H-J, Kim K-C, Kim W (2010) Genetic diversity of two haploid markers in the Udegey population from southeastern Siberia. *Am J Phys Anthropol* 142(2): 303–313. doi:10.1002/ajpa.21232. PubMed: 19953529.
- Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC (1999) Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea-Bering Sea region during the Neolithic. *Am J Phys Anthropol* 108(1): 1–39. doi:10.1002/(SICI)1096-8644(199901)108:1. PubMed: 9915299.
- Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L et al. (2002) The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* 70(1): 192–206. doi:10.1086/338457. PubMed: 11731934.
- Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL et al. (2002) High levels of Y-chromosome differentiation among native

- Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum Biol* 74(6): 761–789. doi:10.1353/hub.2003.0006. PubMed: 12617488.
23. Malyarchuk BA, Derenko M, Denisova G (2012) On the Y-chromosome haplogroup C3c classification. *J Hum Genet* 57(10): 685–686. doi: 10.1038/jhg.2012.93. PubMed: 22810113.
  24. Rootsi S, Zhivotovskiy LA, Baldovic M, Kayser M, Kutuev IA et al. (2007) A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur J Hum Genet* 15(2): 204–211. doi:10.1038/sj.ejhg.5201748. PubMed: 17149388.
  25. Derenko M, Malyarchuk B, Denisova GA, Wozniak M, Dambueva I et al. (2006) Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai-Sayan regions. *Hum Genet* 118(5): 591–604. doi:10.1007/s00439-005-0076-y. PubMed: 16261343.
  26. Derenko M, Malyarchuk B, Denisova G, Wozniak M, Grzybowski T et al. (2007) Y-chromosome haplogroup N dispersals from south Siberia to Europe. *J Hum Genet* 52(9): 763–770. doi:10.1007/s10038-007-0179-5. PubMed: 17703276.
  27. Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59: 935–945. PubMed: 8808611.
  28. Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A et al. (2009) Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFISTR(R) Yfiler(R) PCR amplification kit. *Int J Legal Med* 123(6): 471–482. doi:10.1007/s00414-009-0342-y. PubMed: 19322579.
  29. Malyarchuk B, Derenko M, Denisova G, Wozniak M, Grzybowski T et al. (2010) Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. *Ann Hum Genet* 74(6): 539–546. doi:10.1111/j.1469-1809.2010.00601.x. PubMed: 20726964.
  30. Pääbo S (1996) Mutational hot spots in the mitochondrial microcosmos. *Am J Hum Genet* 59: 493–496. PubMed: 8751848.
  31. Macaulay VA, Richards MB, Forster P, Bendall KE, Watson E et al. (1997) mtDNA mutation rates - no need to panic. *Am J Hum Genet* 61(4): 983–986. doi:10.1016/S0002-9297(07)64211-6. PubMed: 9382113.
  32. Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I et al. (2007) Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet* 81(5): 1025–1041. doi: 10.1086/522933. PubMed: 17924343.
  33. Sato T, Amano T, Ono H, Ishida H, Kodera H et al. (2007) Origins and genetic features of the Okhotsk people, revealed by ancient mitochondrial DNA analysis. *J Hum Genet* 52(7): 618–627. doi: 10.1007/s10038-007-0164-z. PubMed: 17568987.
  34. Volodko NV, Starikovskaya EB, Mazunin IO, Eltsov NP, Naidenko PV et al. (2008) Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and Pleistocene peopling of the Americas. *Am J Hum Genet* 82(5): 1084–1100. doi:10.1016/j.ajhg.2008.03.019. PubMed: 18452887.
  35. Romanova AV, Myreeva AN (1962) Očerki tokkingskogo i tommotskogo dialektov. Moscow, Leningrad: Izdatel'stvo akademii nauk SSSR. 107 p.
  36. Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA et al. (2003) Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet* 67(5): 391–411. doi:10.1046/j.1469-1809.2003.00035.x. PubMed: 12940914.
  37. Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik RI (2002) Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet* 70(4): 1009–1014. doi: 10.1086/339524. PubMed: 11845409.
  38. Derenko MV, Malyarchuk BA, Dambueva IK, Shaikhaev GO, Dorzhu CM et al. (2000) Mitochondrial DNA variation in two South Siberian Aboriginal populations: implications for the genetic history of North Asia. *Hum Biol* 72(6): 945–973. PubMed: 11236866.
  39. Pakendorf B (2009) Intensive Contact and the Copying of Paradigms: An Even Dialect in Contact with Sakha (Yakut). *Journal of Language Contact Varia* 2: 85–110.
  40. Russian census (2010). Available: [http://www.gks.ru/free\\_doc/new\\_site/perepis2010/croc/perepis\\_itogi1612.htm](http://www.gks.ru/free_doc/new_site/perepis2010/croc/perepis_itogi1612.htm).
  41. Zlojutro M, Tarskaia LA, Sorensen M, Snodgrass JJ, Leonard WR et al. (2009) Coalescent simulations of Yakut mtDNA variation suggest small founding population. *Am J Phys Anthropol* 139(4): 474–482. doi: 10.1002/ajpa.21003. PubMed: 19235790.
  42. Gogolev AI (1993) Jakuty. Problemy étnogeneza i formirovanija kul'tury. Yakutsk: Izdatel'stvo JaGU. 199 p.
  43. Alekseev AN (1996) Drevnjaja Jakutija - Zheleznyj vek i epokha srednevekov'ja. Novosibirsk: Izdatel'stvo Instituta Arxeologii i Etnografii SO RAN. 95 p.
  44. Barbieri C, Whitten M, Beyer K, Schreiber H, Li M et al. (2012) Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. *Mol Biol Evol* 29(4): 1213–1223. doi:10.1093/molbev/msr291. PubMed: 22130972.
  45. Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1: 47–50.
  46. Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R et al. (2011) HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* 32(1): 25–32. doi:10.1002/humu.21382. PubMed: 20960467.
  47. Van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30(2): E386–E394. doi:10.1002/humu.20921. PubMed: 18853457.
  48. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA - Nat Genet 23: 147. doi: 10.1038/13779.
  49. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9(8): 772. doi:10.1038/nmeth.2109.
  50. Tamura K, Peterson D, Peterson N, Stecher G, Nei M et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10): 2731–2739. doi:10.1093/molbev/msr121. PubMed: 21546353.
  51. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214. doi: 10.1186/1471-2148-7-214. PubMed: 17996036.
  52. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22(5): 1185–1192. doi:10.1093/molbev/msi103. PubMed: 15703244.
  53. Soares P, Ermini L, Thomson N, Mormina M, Rito T et al. (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84(6): 740–759. doi:10.1016/j.ajhg.2009.05.001. PubMed: 19500773.
  54. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16(1): 37–48. doi: 10.1093/oxfordjournals.molbev.a026036. PubMed: 10331250.
  55. StatSoft. Inc. (2011) STATISTICA. Tulsa, OK.
  56. Hijmans RJ, Williams E, Vennes C (2012) geosphere: Spherical Trigonometry.
  57. Dray S, Dufour AB, Chessel D (2007) The ade4 package-II: Two-table and K-table methods. *R NEWS* 7(2): 47–52.
  58. De Filippo C, Barbieri C, Whitten M, Mpoloka SW, Gunnarsdóttir ED et al. (2011) Y-chromosomal variation in Sub-Saharan Africa: insights into the history of Niger-Congo groups. *Mol Biol Evol* 28(3): 1255–1269. doi: 10.1093/molbev/msq312. PubMed: 21109585.
  59. Nasidze I, Quinque D, Dupanloup I, Cordaux R, Kokshunova L et al. (2005) Genetic evidence for the Mongolian ancestry of Kalmyks. *Am J Phys Anthropol* 128(4): 846–854. doi:10.1002/ajpa.20159. PubMed: 16028228.
  60. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL et al. (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosome haplogroup tree. *Genome Res* 18(5): 830–838. doi:10.1101/gr.7172008. PubMed: 18385274.
  61. Rubicz R, Melton PE, Spitsyn V, Sun G, Deka R et al. (2010) Genetic structure of native circumpolar populations based on autosomal, mitochondrial, and Y chromosome DNA markers. *Am J Phys Anthropol* 143(1): 62–74. doi:10.1002/ajpa.21290. PubMed: 20333712.
  62. Rubicz R (2007) Evolutionary consequences of recently founded Aleut communities in the Commander and Pribilof Islands. University of Kansas PhD dissertation
  63. Bosch E, Calafell F, González-Neira A, Flaiz C, Mateu E et al. (2006) Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet* 70(4): 459–487. doi:10.1111/j.1469-1809.2005.00251.x. PubMed: 16759179.
  64. Fenner JN (2005) Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* 128(2): 415–423. doi:10.1002/ajpa.20188. PubMed: 15795887.