

Coalescent Simulations of Yakut mtDNA Variation Suggest Small Founding Population

Mark Zlojutro,^{1*} Larissa A. Tarskaia,² Mark Sorensen,³ J. Josh Snodgrass,⁴ William R. Leonard,⁵ and Michael H. Crawford⁶

¹*Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX 78227*

²*Institute of Molecular Genetics, Russian Academy of Medical Sciences, Moscow 115478, Russia*

³*Department of Anthropology, University of North Carolina, Chapel Hill, NC 27599*

⁴*Department of Anthropology, University of Oregon, Eugene, OR 97403*

⁵*Department of Anthropology, Northwestern University, Evanston, IL 60208*

⁶*Department of Anthropology, University of Kansas, Lawrence, KS 66045*

KEY WORDS Yakut origins; mitochondrial DNA; coalescent simulation

ABSTRACT The Yakuts are a Turkic-speaking population from northeastern Siberia who are believed to have originated from ancient Turkic populations in South Siberia, based on archaeological and ethnohistorical evidence. In order to better understand Yakut origins, we modeled 25 demographic scenarios and tested by coalescent simulation whether any are consistent with the patterns of mtDNA diversity observed in present-day Yakuts. The models consist of either two simulated demes that represent Yakuts and a South Siberian ancestral population, or three demes that also include a regional Northeast Siberian population that

served as a source of local gene flow into the Yakut deme. The model that produced the best fit to the observed data defined a founder group with an effective female population size of only 150 individuals that migrated northwards approximately 1,000 years BP and who experienced significant admixture with neighboring populations in Northeastern Siberia. These simulation results indicate a pronounced founder effect that was primarily kin-structured and reconcile reported discrepancies between Yakut mtDNA and Y chromosome diversity levels. *Am J Phys Anthropol* 000:000–000, 2009. © 2009 Wiley-Liss, Inc.

The Yakuts of northeastern Siberia are a Turkic-speaking population of about 430,000 who are settled throughout the Sakha Autonomous Republic (also known as Yakutia) of the Russian Federation. Traditionally, Yakuts are a cattle- and horse-breeding people and represent one of the northernmost Turkic-speaking populations in the world, which distinctly contrasts the subsistence patterns and languages of other native groups in the region. Comparative linguistics has revealed close similarities between the Yakut language and Turkic languages spoken in the Altai-Sayan region (Ruhlen, 1987), which suggests that Yakuts have ancestral ties to southern Turkic groups and thus originally migrated to northeastern Siberia from areas further to the south. Other types of evidence support this notion of a southern origin, including similarities in material culture, elements in the pastoral economy, celebrated traditions, and religious beliefs (Tokarev and Gurvich, 1956; Okladnikov, 1970). Moreover, petroglyphs and artifacts attributed to the ancient Turkic-speaking Kurykans from the Lake Baikal region display close affinities to Yakut culture. Based on these archaeological finds, some scholars postulate that Yakuts stem from a Kurykan exodus that trekked north along the Lena River system in order to escape Mongolian incursions during the 11th to 13th centuries (Okladnikov, 1970; Konstantinov, 1975; Alekseev, 1996).

Although it is evident that the most parsimonious model for Yakut ancestry would necessitate a southern origin and founder event, the timing of the northern migration, the size of the founder group and the degree

of genetic admixture with non-Turkic Siberian populations is less apparent. At Russian contact in the 17th century, Yakuts were primarily settled in central Yakutia in the basins of the Lena and Aldan river systems and had a population size of approximately 30,000–40,000 (Dolgikh, 1960; Forsyth, 1992). In the following centuries, Yakuts expanded beyond this region into neighboring territories once occupied by various indigenous peoples—namely the Yukaghirs and Tungusic-speaking Evenks and Evens—and today represent one of the largest and most widespread populations in Siberia.

Several studies have characterized the mtDNA and Y-chromosome variation in the Yakuts in order to elucidate their demographic history and genetic relationship to other Asian populations. Federova et al. (2003) presented evidence for a Mongol/Central Asian origin for the Yakuts based on HVS-I data, whereas Puzyrev et al.

Grant sponsor: National Science Foundation.

*Correspondence to: Mark Zlojutro, Department of Genetics, Southwest Foundation for Biomedical Research, 7620 NW Loop 410 at Military Drive, San Antonio, TX 78229, USA.
E-mail: markz@sfbrgenetics.org

Received 24 May 2008; accepted 25 November 2008

DOI 10.1002/ajpa.21003
Published online in Wiley InterScience
(www.interscience.wiley.com).

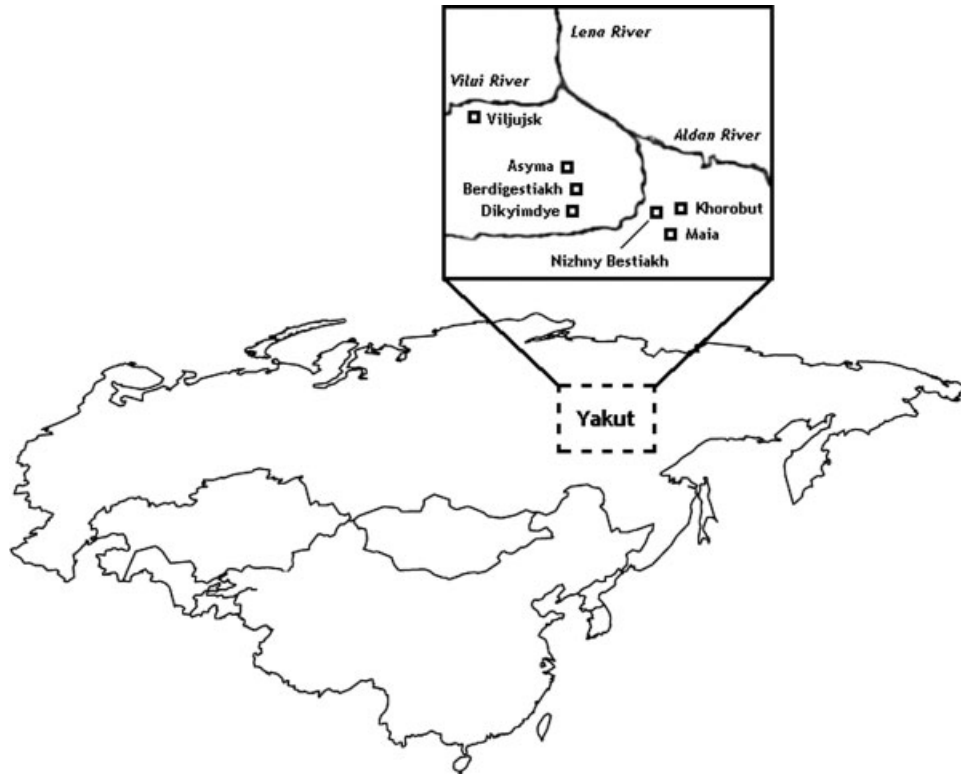


Fig. 1. Map of Asia showing the approximate geographic locations of the Yakut samples analyzed in this study.

(2003) identified common Yakut Y-chromosome haplotypes that are closest to haplotypes found in eastern Evenks and thus suggesting potential admixture with Tungusic groups. In a paper by Pakendorf et al. (2006), mtDNA results affirmed a southern ancestry for the Yakuts, but revealed no conclusive evidence for either a migration bottleneck or admixture with indigenous peoples. The Y-chromosome data, on the other hand, showed genetic signatures for a strong founder effect that was dated to approximately 880 years BP, although the exact origins of the Yakut paternal lines were not resolved. And in a recent study by Zlojutro et al. (2008), mtDNA variation was examined in seven communities from central Yakutia and revealed genetic features indicative of a founder event: fragmented MJ network dominated by high-frequency haplotypes within haplogroups C and D; nested cladistic analysis (NCA) that identified a significant geographic differentiation for subhaplogroup D5a; and type 1 deviation for the observed Ewens haplotype frequency spectrum.

To better understand Yakut origins, and specifically the nature of the Yakut founder event in northeastern Siberia, this paper evaluates a series of migration models through coalescent simulation. The models were defined by parameters such as effective population size, growth rate and gene flow, and the resulting simulated data were compared against the patterns of Yakut mtDNA diversity reported by Zlojutro et al. (2008). Of the various models tested, the one producing the highest likelihood defined a migration event at 1,000 years BP, a founder group with an effective female population size of only 150, and significant admixture with neighboring populations in Northeastern Siberia.

MATERIALS AND METHODS

Yakut sequence data

Measures of Yakut mtDNA diversity examined in this study are based on HVS-I sequences reported by Zlojutro et al. (2008). The Yakut samples were collected during the summer of 2000 from seven communities located in the central Lena River watershed of the Sakha Republic: Asyma, Berdigestiakh, Dikyimdye, Khorobut, Maia, Nizhny Bestiakh, and Viljujsk (see Fig. 1). DNA was extracted from whole blood using the Super Quik-Gene DNA isolation kit (LBA, University of Kansas), a salt-ing-out methodology. A total of 144 samples were successfully sequenced for the HVS-I and aligned using Bio-Edit software (Ibis Therapeutics), which produced a data set encompassing np 16050–16400 (a total length of 351 bp).

Coalescent simulations

The coalescent program SERIAL SIMCOAL (Anderson et al., 2005) was used to test different scenarios of Yakut demographic history. The program constructs simulated genealogies for demes of n sequences backwards in time for t generations. Mutations are then randomly distributed onto the trees by a user-specified mutation model. The demographic models considered in this study differ in terms of the number of demes, effective population sizes, growth rates, gene flow, timing of the Yakut founder event, and mutation rate. Based on recent estimates for the HVS-I (Meyer et al., 1999), a transition bias of 0.9375 and mutation rate-heterogeneity parameter of

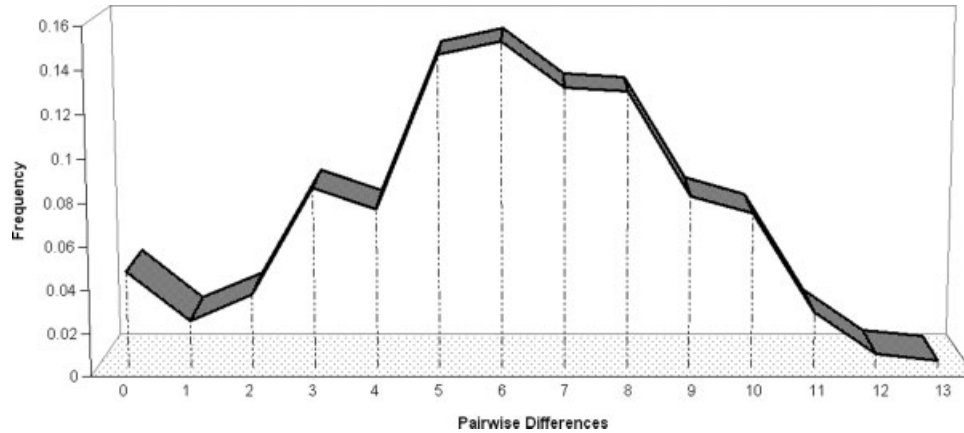


Fig. 2. Mismatch distribution for Yakut HVS-I sequences (Zlojutro et al., 2008).

0.26 were used, allowing for variation at each of the 351 nucleotide sites in the 144 simulated sequences.

The SERIAL SIMCOAL output for each of the demographic models consisted of 1,000 sets of sequence data, which included statistics such as haplotype number, segregating sites, haplotype diversity, pairwise differences and Tajima's D generated in Excel spreadsheets. In order to evaluate the overall significance of the individual models, the method described by Belle et al. (2006) was employed. The empirical likelihood (P) for each of the observed diversity measures given the parameters of the different demographic scenarios were computed relative to the diversity statistics obtained from the simulated data. For instance, if the observed statistic x ranks as the k th among 1,000 simulated values with a median of m , the empirical likelihood of that value is represented by the number simulated values greater than x (assuming $x > m$) and doubling that number to obtain a two-tailed test. Fisher's method was used to combine the probabilities of the individual statistics in order to obtain an overall test of significance for each model (note, there is not complete independence between the different genetic diversity measures). Fisher's test statistic has a χ^2 distribution, where $\chi^2 = -2\sum \ln(P)$. A total of 25 different models were tested (of which 14 are discussed in detail in the Results section) and under a Bonferroni correction the significance level becomes $P = 0.05/25 = 0.002$. For 10 degrees of freedom (derived from $2k$, where k is the number of measures), the critical value of the χ^2 for the demographic models is 27.72.

RESULTS

Yakut mtDNA diversity

The Yakut HVS-I data reported by Zlojutro et al. (2008) comprise 53 different haplotypes (k) characterized by 64 variant sites. The haplotype diversity for this data set is 0.955, with values of 11.365 and 6.025 for the θ_S and θ_π estimators, respectively. These scores are intermediate to those exhibited by northeastern Siberian populations and the genetically diverse Asian populations further to the south. The neutrality test statistics Tajima's D and Fu's F_S have values of -1.462 and -25.025 , respectively, and both are significant at the 0.05 level. The mismatch distribution for the Yakut sequences is primarily unimodal with a peak at six pairwise differences (see Fig. 2), which is considered a signature of pop-

ulation expansion (Rogers and Harpending, 1992; Sherry et al., 1994). In addition, a minor peak or lip is present at zero differences.

The observed haplotype frequency spectrum for the Yakut HVS-I sequences exhibits a type 1 deviation from the Ewens (1972) sampling distribution (given $k = 53$ and $n = 144$) where the most common haplotypes are at higher frequencies than expected (see Fig. 3). The homozygosity of the Ewens distribution is 0.0384, which is significantly less than the observed value of 0.0513 ($P = 0.045$ based on the Ewens-Watterson homozygosity test) and is largely due to the high frequencies of the six most common Yakut haplotypes. This deviation likely represents the genetic consequence of the postulated Yakut founder event in northeastern Siberia for which a limited number of maternal lineages appears to have dominated the Yakut founder population (Zlojutro et al., 2008). Overall, haplogroups C and D make up 75.7% of the Yakut sample, a mtDNA composition that is generally consistent with those of other neighboring Siberian populations. The seven most common Yakut haplotypes all belong to haplogroups C and D.

Coalescent simulation models

The general design of the various demographic models tested in this study is illustrated in Figure 4. The three cone-like structures represent simulated populations or demes, specifically the Yakuts, Northeast Siberians and South Siberians. The varying cone diameters correspond to effective population sizes (N_e) and its change over time. Historical and archaeological benchmarks are provided on the left-hand side (e.g., Russian contact), along with the number of generations from the present day (T). The models in this study are based on either: two demes representing Yakuts and a South Siberian meta-population responsible for giving rise to the Yakut founders; or three demes that also include Northeastern Siberians that serve as a source of local gene flow into the Yakut deme. Assuming an isolation-by-distance (IBD) pattern for mtDNA diversity, which is a reasonable expectation for prehistoric pastoralist populations, Yakut admixture would have mostly involved surrounding Northeast Siberian groups, such as Yukaghirs and Tungus, and not the highly diverse Turkic- and Mongolic-speaking populations located approximately 2,000 km to the south in the Lake Baikal and Altai-Sayan

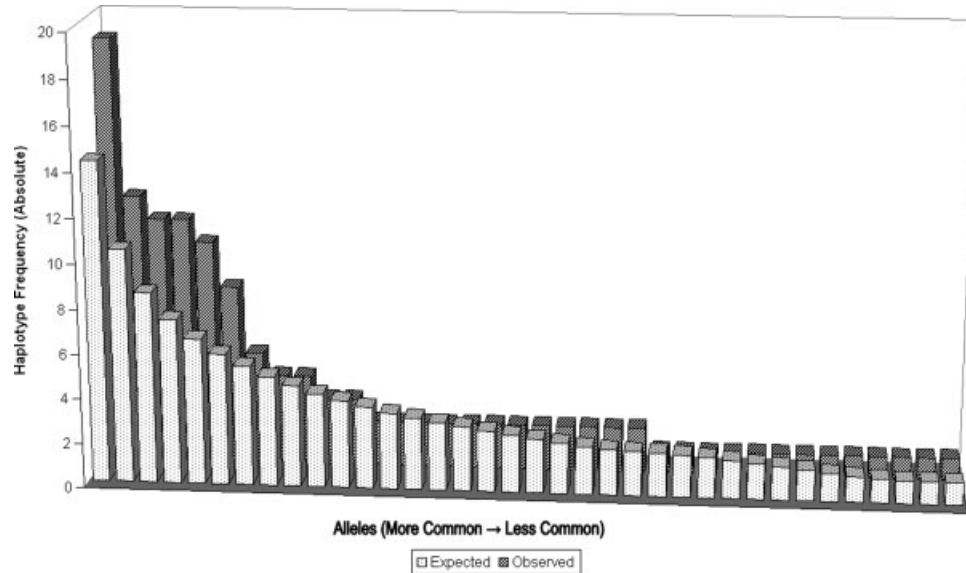


Fig. 3. The expected Ewens haplotype frequency spectrum plotted against the observed haplotype frequency spectrum for Yakut HVSI data (Zlojutro et al., 2008).

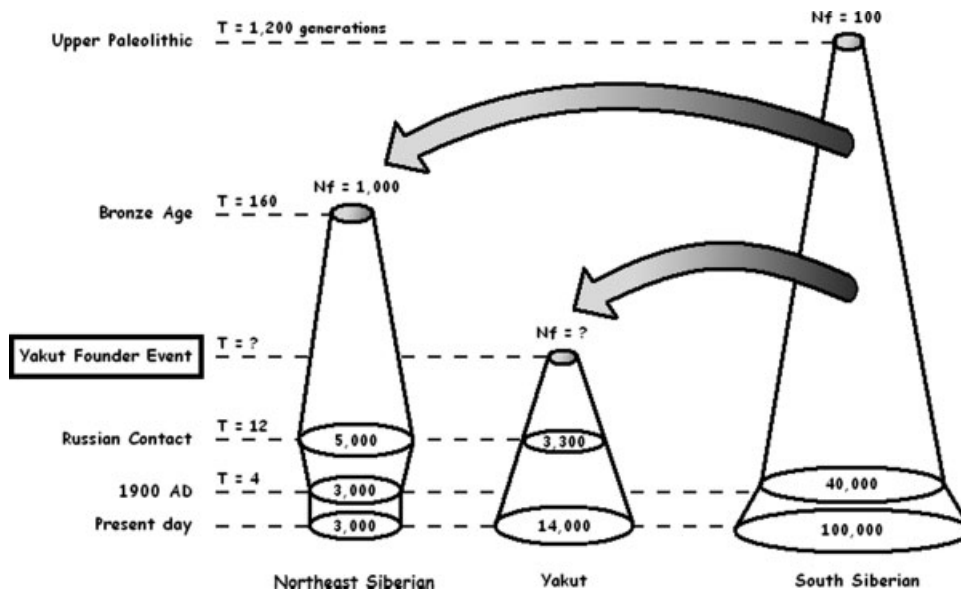


Fig. 4. Schematic of the general design of the various demographic models tested in this study. The cone-like structures represent the change in effective population sizes (N_f) of demes over time for the Yakuts, Northeast Siberians, and South Siberians. The left-hand margin indicates the timing of historical periods and demographic events, which is estimated by the number generations to the present day (T). The N_f sizes are provided within the demes at certain time periods. The arrows stemming from the south Siberian gene pool indicate the source of founder events related to the origins of Northeast Siberians and Yakuts.

areas. Therefore, a third deme was defined in a subset of the demographic models that approximate the prehistoric conditions of localized gene flow into the Yakut population.

Based on the archaeological record, the first modern humans settled in South Siberia during the Early Upper Paleolithic period approximately 30,000–40,000 years BP and later colonizing the mammoth-steppes of subarctic Siberia by means of various behavioral adaptations (Goebel, 1999). This benchmark (30,000 years BP or 1,200 generations, assuming 25 years per generation) was used as the evolutionary endpoint for the simulated

South Siberian gene pool (note, the coalescence process is simulated backwards in time). The population size of the South Siberian founders is not known with any degree of certainty; however human settlement during the Middle Pleniglacial was sparse given the environmental constraints of the region (e.g., long cold winters and patchwork of vegetation zones). Thus, a conservative effective population size of 100 was used in the models (note, N_f is approximately $N/12$ because effective population size is typically estimated as one third the census size and the effective size of mtDNA is one fourth of that for autosomes).

TABLE 1. Parameter values for tested demographic models and empirical likelihood values (P) for summary statistics computed from 1,000 simulations

Model	N_0	T_0	μ	M	Haplotype no.	Seg. sites	Hap. div.	Pair. diff.	Tajima's D	χ^2
Observed					53	64	0.9553	6.0389	-1.4849	
1	100	1500	0.5	0	18.679 (0.000)	30.307 (0.000)	0.8369 (0.000)	5.9455 (0.922)	0.2631 (0.004)	80.28
2	100	1000	0.5	0	18.465 (0.000)	32.147 (0.000)	0.8542 (0.000)	6.2341 (0.946)	0.2370 (0.000)	92.21
3	500	1500	0.5	0	28.669 (0.000)	44.603 (0.002)	0.9226 (0.004)	6.8309 (0.540)	-0.4500 (0.006)	57.96
4	500	1000	0.5	0	30.055 (0.000)	47.514 (0.006)	0.9299 (0.010)	6.9868 (0.418)	-0.5562 (0.002)	56.64
5	1000	1500	0.5	0	34.269 (0.000)	51.223 (0.028)	0.9410 (0.168)	7.1790 (0.332)	-0.6780 (0.004)	46.99
6	1000	1000	0.5	0	36.604 (0.000)	54.392 (0.094)	0.9474 (0.396)	7.2047 (0.344)	-0.8122 (0.020)	39.57
7	1000	1500	0.05	0	7.700 (0.000)	7.162 (0.000)	0.5198 (0.000)	0.7834 (0.000)	-0.8525 (0.218)	95.15
8	1000	1000	0.05	0	8.417 (0.000)	7.861 (0.000)	0.5294 (0.000)	0.8044 (0.000)	-0.9722 (0.328)	94.33
9	1000	1500	0.1	0	13.223 (0.000)	13.642 (0.000)	0.7409 (0.000)	1.5371 (0.000)	-0.9551 (0.250)	94.88
10	1000	1000	0.1	0	14.273 (0.000)	14.721 (0.000)	0.7439 (0.000)	1.5339 (0.000)	-1.0962 (0.368)	94.10
11	1000	1000	0.5	0.005	43.208 (0.034)	58.797 (0.334)	0.9539 (0.960)	7.1849 (0.312)	-0.9918 (0.088)	<i>16.22</i>
12	1000	1000	0.5	0.05	60.346 (0.140)	68.742 (0.374)	0.9686 (0.068)	7.3289 (0.232)	-1.2739 (0.462)	<i>15.74</i>
13	1000	1000	0.4	0.05	54.676 (0.680)	59.640 (0.412)	0.9621 (0.352)	5.8752 (0.818)	-1.4028 (0.844)	<i>5.37</i>
14	150	1000	0.4	0.10	53.008 (0.972)	58.871 (0.352)	0.9518 (0.906)	5.8427 (0.804)	-1.3870 (0.760)	<i>3.33</i>

Model parameters: N_0 , effective population size at time of founder event; T_0 , time of founder event in years BP; μ , mutation rate per million years per nucleotide; M , proportion of Tungus-Yukaghir migrants per generation.

Abbreviations: haplotype no., haplotype number; seg. sites, segregating sites; hap. div., haplotype diversity, pair. diff., pairwise differences. Likelihood values are provided in parentheses. χ^2 values for Fisher's test is defined by $-2\ln(P)$. Italicized scores are those that are not significant.

Within Northeastern Siberia, human colonization occurred later, with Late Neolithic and Bronze Age cultures exhibiting continuity with the traditional cultures of present day Tungus and Yukaghir groups (Okladnikov, 1956). According to Vasilevich (1969), the northern Tungus are the descendants of Neolithic populations that were settled in the Lake Baikal region in South Siberia approximately 3,000–4,000 years BP and later migrated northwards in the face of Turkish expansion into the area. Other scholars argue that Tungus origins lie further to the southeast in Manchuria and the Amur River region (Shirokogoroff, 1966; Janhunen, 1996). The Yukaghirs, who speak a language classified as a linguistic isolate (Comrie, 1981), are considered the remnants of a once widespread Paleo-Siberian people that experienced extensive admixture with Tungus cultural elements during the past millennium. For the models, a deme representing a population aggregate of the Tungus-Yukaghir peoples of northeastern Siberia was simulated by defining a founder event in the Bronze Age (4,000 years BP or 160 generations) with an N_f of 1,000 that stems from the South Siberian meta-population.

The N_f values from Russian contact up to the present day are based on Russian/USSR census statistics (1897, 1989, and 2002) and historical accounts (cf. references in Forsyth, 1992). The Yakut effective sizes were estimated from population data for the districts of central Yakutia, the region from which the Yakut samples in the Zlojutro et al. study were collected from. Other mtDNA data sets from districts outside of central Yakutia were not included in the present study (Federova et al., 2003; Pakendorf et al., 2003, 2006) in order to remove any genetic substructure stemming from the far-ranging geographic expansions of Yakuts throughout Northeastern Siberia since Russian contact. Such effects would have important genetic implications for diversity levels of periphery versus central demes and would necessitate the simulation of additional Yakut demes within a spatial context to accommodate this additional Yakut data (Ray et al., 2003), and thus the present study focuses on the population growth in central Yakutia and the timing and size of the Yakut ancestral migration to this region from South Siberian sources.

A total of 25 models were tested, of which 14 are discussed below. For each model, numbered 1 through 14, 1,000 simulations were performed and the means for the following genetic statistics were calculated from the sequence data sets (Table 1): haplotype number, segregating sites, haplotype diversity, pairwise differences, and Tajima's D . Models 1 through 10 are basic two-deme scenarios characterizing the Yakut founder event derived from the south Siberian gene pool. The adjusted parameters for models 1 and 2 include a Yakut founder size (N_0) of 100 individuals and a mutation rate (μ) of 0.5 mutations per million years per nucleotide. The timing of the Yakut founder event is varied between 1,500 years BP (60 generations) and 1,000 years BP (40 generations) for the two models, as well as models 3 through 10. Although the Yakut migration to the north is assumed to have taken place no earlier than the 11th to 13th centuries AD based on both the Kurykan archaeological record and the absence of artifacts associated with cattle or horse-breeding in Yakutia prior to this period (Gogolev, 1993), South Siberia and the Asian steppe lands experienced a series of large-scale population movements of Turkic peoples during the first millennium AD (e.g., T'u-Chueh in the sixth century AD) and thus an earlier date for the Yakut migration (1,500 years BP) was considered. Overall, models 1 and 2 produced patterns of sequence diversity that differed substantially from the observed Yakut sequences. For instance, the average number of haplotypes for the simulated data from both models is only about 18, whereas 53 different haplotypes are observed in the Yakut sample. The only statistic that resulted in high P values is pairwise differences. The large, significant χ^2 values for the models indicate poor fits to the observed data.

For models 3 through 6, N_0 was increased to either 500 or 1,000 individuals. Model 6 has the best fit (significant χ^2 of 39.57), which defines the Yakut founder event at 1,000 years BP and N_0 of 1,000 individuals. However, the diversity levels of the simulated data remain low relative to the observed sequences (e.g., average haplotype number ranges from 28.67 to 36.60). These low levels became more pronounced in models 7 through 10 that set N_0 at 1,000 individuals and used more conservative

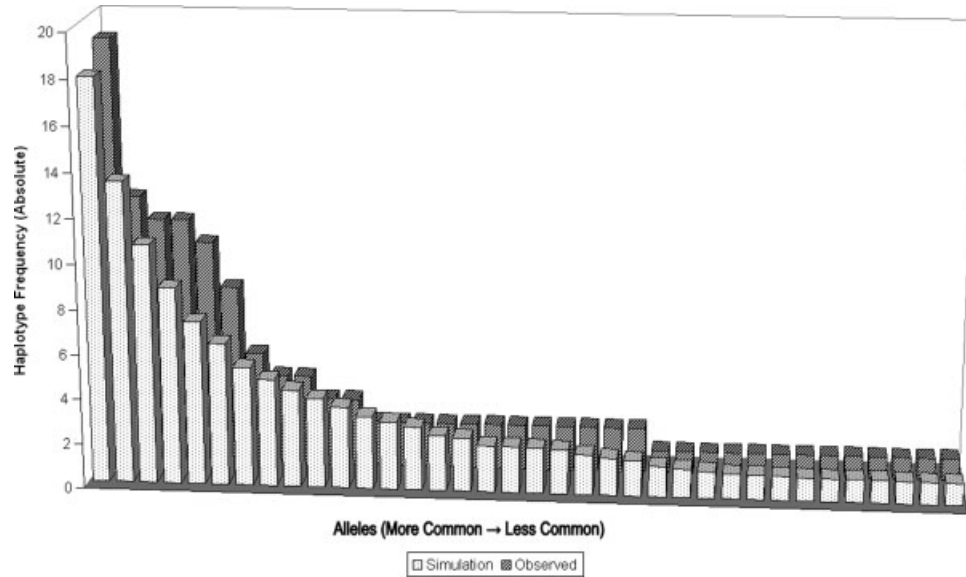


Fig. 5. The observed Yakut frequency spectrum plotted against the simulated haplotype frequency spectrum for demographic Model 14.

HVS-I mutation rate estimates that derive from phylogenetic considerations (either 0.05 or 0.10) (Ward et al., 1991; Hasegawa et al., 1993; Tamura and Nei, 1993), producing very high χ^2 scores. At lower N_0 values, genetic variation was almost absent (data not shown). Clearly, the elevated mutation rate of 0.5 used in models 1 through 6, which is based on estimates derived from pedigree studies (Howell et al., 2003; Pakendorf and Stoneking, 2005), generated data that was more concordant to the observed diversity patterns.

In Models 11 and 12, a third deme was simulated, representing the Tungus-Yukaghir peoples and a source of gene flow into the Yakut deme. The parameter settings of Model 6 were used in these two scenarios, with the proportion of migrants per generation from the third deme, M , set at 0.005 and 0.05, respectively. As a result of the gene flow, both models produced higher levels of diversity and non-significant χ^2 values that indicate a good fit to the observed Yakut data. Higher migration rates were also tested, but this resulted in weaker fits (data not shown). In Model 13, the mutation rate was adjusted to 0.4 in order to depress the genetic variation of Model 12 to be more akin to the observed data, which achieved a low χ^2 of 5.37.

Of the models discussed up to this point, Model 13 exhibits the best fit to the observed Yakut data. However, when the simulated sequences were then used to construct an average mismatch distribution and haplotype frequency spectrum, this apparent concordance was not supported. As previously noted, the observed Yakut sequences exhibit a significant type 1 deviation from the Ewens haplotype frequency spectrum and a minor peak centered on zero differences in the mismatch distribution, two features that are presumably signatures of a pronounced founder effect and neither of which are evident in the simulated data for Model 13 (not shown). The homozygosity for the frequency distribution of the simulated data is 0.0353, which is less than the values for both the expected (0.0384) and observed distributions (0.0513; $P = 0.000$). Therefore, in order to generate a limited number of high-frequency haplotypes in the

simulated sequences, but without depressing the level of diversity achieved in Model 13, an additional series of models were tested in which N_0 was decreased and M increased in a stepwise fashion.

From this series of simulations, Model 14 in Table 1 produced the best fit (χ^2 of 3.33). This model is defined by a relatively small N_0 of 150 and M set at 0.10 migrants per generation. The timing of the founder event is 1,000 years BP, although the earlier date of 1,500 BP was also tested and produced strong fits. The means of the genetic statistics for the simulated data are remarkably close to the observed diversity patterns, especially for haplotype number (53.008; $P = 0.972$) and haplotype diversity (0.9518; $P = 0.906$). More importantly, Model 14 generated a haplotype frequency spectrum that exhibits very high frequencies in the front end of the distribution (see Fig. 5) and a mismatch distribution with a minor peak at zero pairwise differences (see Fig. 6). The homozygosity for the haplotype frequencies is 0.0499, which produced a high likelihood to the observed value ($P = 0.880$).

DISCUSSION

The demographic models tested in this study are not exhaustive of all possibilities. The focus of the simulation study was to characterize the Yakut founder event in terms of its timing and the number of female founders, as well as the degree to which subsequent gene flow on a regional level was responsible for shaping contemporary Yakut mtDNA variation. However, the SERIAL SIMCOAL program does not permit the identification of particular source populations for Yakut admixture, and thus detailed admixture scenarios were not considered in this paper, such as the potential genetic impact of Mongolic-Buryat contact prior to the Yakut migration to Northeastern Siberia (Pakendorf et al., 2006). Nonetheless, two main conclusions can be drawn from the simulation results. First, the higher HVS-I mutation rates estimated from direct observations of mutations arising in families or deep-rooted pedigrees generated levels of

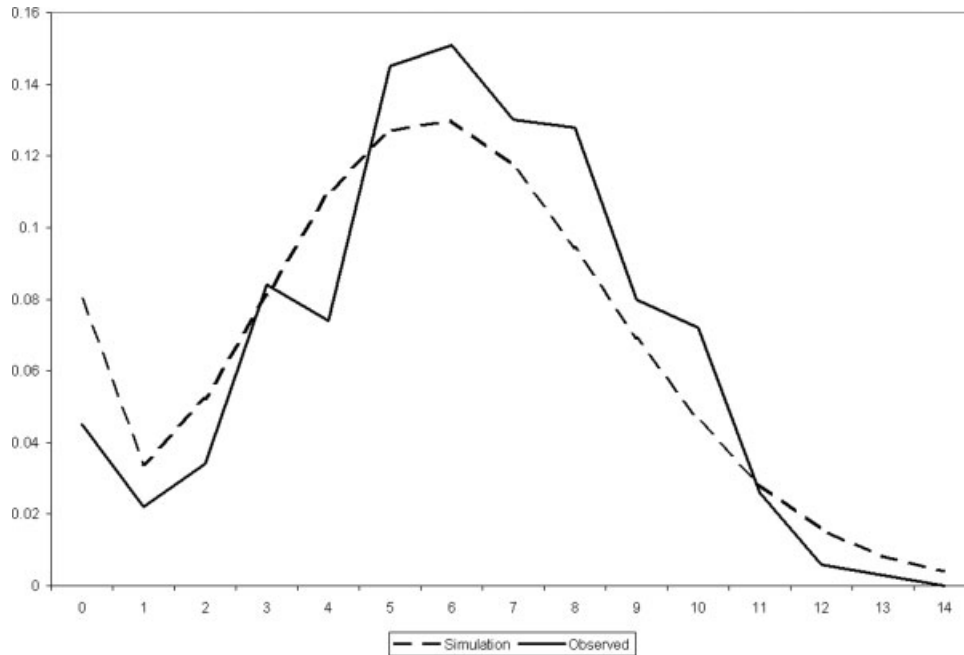


Fig. 6. Mismatch distributions for observed Yakut HVS-I sequences and simulated data for demographic Model 14.

genetic variation that are more consistent with the observed Yakut data. The alternate mutation rate based on phylogenetic evidence is approximately 10-fold lower and could possibly be accommodated by the coalescent models if the rate of migration into the Yakut deme was extremely high and human settlement of Siberia occurred much earlier and in larger numbers, two conditions that are not supported by historical or archaeological evidence. Clearly, this discrepancy in mutation rates has important implications for reconstructions of human evolutionary history and the dating of demographic events based on mtDNA variation, and as a result has generated intense debate (Macaulay et al., 1997; Sigurdóttir et al., 2000; Heyer et al., 2001; Hagelberg, 2003; Howell et al., 2003). Many explanations have been offered to account for these differences, including paternal mtDNA leakage and recombination and the effects of purifying selection and/or genetic drift on genealogical versus geological timescales. In addition, mutation rate heterogeneity within the HVS-I (i.e., “hot spots”) is a likely factor as fast-evolving sites may be preferentially detected in pedigree studies, whereas phylogenetic comparisons generally involve mutations at slowly evolving sites, with a significant number of the mutations observed in pedigree data eliminated at the population level. Therefore, it may be necessary for studies of deep history to utilize phylogenetically based mutation rates, while the pedigree-based rates may be more appropriate for studies of recent history (Macaulay et al., 1997; Ho and Larson, 2006), such as Yakut ethnogenesis.

The second conclusion that can be drawn from the simulation results is that the Yakut founder event in northeastern Siberia likely involved a small female population with an effective size as few as 150 individuals that experienced notable gene flow from surrounding indigenous peoples. This finding is particularly interesting given the results of previous genetic studies of Yakut samples that revealed marked differences in the diversity of HVS-I sequences and Y-STR haplotypes. Based on

research by Puzyrev et al. (2003) and Pakendorf et al. (2006), Yakut Y-STR variation was found to be among the lowest for Siberian populations, with the vast majority belonging to closely related haplotypes within haplogroup N-TatC. This has been interpreted as evidence for a very strong founder effect in the Yakut paternal line. On the other hand, Yakut mtDNA sequence diversity is high relative to other Siberian populations (Federova et al., 2003; Puzyrev et al., 2003; Pakendorf et al., 2006; Zlojutro et al., 2008). Pakendorf et al. provide three possible explanations for this contradiction between the two genetic systems: (1) substantial Yakut admixture with local women; (2) the Yakut founder population was dominated by related patrilineal clans that practiced strict exogamy and thus comprised of women originating from diverse South Siberian groups; and (3) a large proportion of Yakut men practicing polygyny. But when considering the significant type 1 deviation in the haplotype frequency spectrum observed for the Yakut sequences (see Fig. 3), the second and third explanations are less satisfactory because it is apparent that the Yakut mtDNA structure is dominated by two classes of haplotypes: high-frequency ones within haplogroups C and D; and low-frequency haplotypes or singletons in the tail of the distribution that contribute to the overall high diversity. Complementing the major Y-STR haplotypes identified in the Yakuts, the high-frequency matrilineal lines are suggestive of a pronounced founder effect that was kin structured not only for men but also women. In particular, subhaplogroup D5a likely represents one of these founding matrilineal lines, dating to approximately 1,286 to 800 years BP (Pakendorf et al., 2006). This lineage is dominated by a single high-frequency haplotype (8.3%) that is closely related to South Siberia mtDNA variants and has been characterized in both contemporary samples and the majority of Yakut skeletal specimens dating from 300 to 400 years BP (Ricaud et al., 2004, 2006). Therefore, Pakendorf et al.’s first explanation appears to be the most parsimonious one given the genetic data.

During the Yakuts' successful settlement and expansion throughout the Lena River basin after the northward migration, admixture with Tungus and Yukaghir females would have had the effect of introducing greater mtDNA variation into the Yakut founder population and obscuring the genetic signature of a founder effect that is clearly evident in the Y chromosome data.

In addition to the diversity levels, Yakut mtDNAs and Y chromosomes differ in their phylogeographic relationships and potential origins. The mtDNA data is consistent with linguistic and archaeological evidence that point to a southern origin for the Yakut people. Many of the Yakut haplogroups are common in South Siberia and Central Asia, and this is reflected in MDS plots of genetic distances and SAMOVA trials that demonstrate close ties between the Yakuts and these southern populations (Pakendorf et al., 2006; Zlojutro et al., 2008). In contrast, most of the Yakut Y chromosomes are distinct from all Asian populations that have been studied to date. The reasons for this ambiguity in origins are not entirely clear, but genetic drift coupled with the elevated mutation rate of Y-STR loci (Kayser et al., 2000) may be responsible for differentiating the Y chromosomes from their antecedents in other Siberian populations. Another possibility is that in the face of repeated Mongol-Turkic incursions in the steppe lands of Asia, such as Genghis Khan's ruthless military campaigns during the 13th century, the paternal ancestors of the Yakuts were decimated or reduced to very small numbers and as a result significantly impacting the phylogeographic patterns of Y chromosome diversity observed in Siberia today. For example, in a study of Asian Y-STR haplotypes by Zerjal et al. (2003), a widespread sublineage of haplogroup C* revealed Mongolian ancestry based on comparative considerations and produced a coalescent date of approximately 1,000 years BP, which strongly suggest that it represents a genetic marker of Genghis Khan's vast empire and long-lasting male dynasty. Whether or not the Yakut ancestors were impacted by this protracted period of Mongol warmongering or some other demographic upheaval is not known, although the archaeological record for the Kurykan people, the purported ancestors of Yakuts, presents a precipitous disappearance from the Lake Baikal region. The exact origins of the Yakut Y chromosomes therefore remain a mystery and further research is needed to fully elucidate Yakut evolutionary history.

ACKNOWLEDGMENTS

We are grateful to all the participants from Yakutia for providing blood samples and to the people who contributed to their collection. We thank Chris Anderson for his help in constructing input files for the SERIAL SIMCOAL program and Dennis O'Rourke and two anonymous reviewers for their valuable comments on this article.

LITERATURE CITED

- Alekseev AN. 1996. Ancient Yakutia: The Iron Age and the Medieval Epoch (in Russian). Novosibirsk: Izdatel'stvo Instituta Arkheologii i Etnografii.
- Anderson CNK, Ramakrishnan U, Chan YL, Hadley EA. 2005. Serial Simcoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* 21:1733–1734.
- Belle EMS, Ramakrishnan U, Mountain JL, Barbujani G. 2006. Serial coalescent simulations suggest a weak genealogical relationship between Etruscans and modern Tuscans. *Proc Natl Acad Sci USA* 103:8012–8017.
- Comrie B. 1981. *The languages of the Soviet Union*. Cambridge: Cambridge University Press.
- Dolgikh BO. 1960. *The tribal composition of the peoples of Siberia in the 17th century (in Russian)*. Moscow: Izdatel'stvo Akademii Nauk SSSR.
- Ewens WJ. 1972. The sampling theory of selectively neutral alleles. *Theor Pop Biol* 3:87–112.
- Federova SA, Bermisheva MA, VILLEMS R, Maksimova NR, Khusnutdinova EK. 2003. Analysis of mitochondrial DNA lineages in Yakuts. *Mol Biol* 37:544–553.
- Forsyth J. 1992. *A history of the peoples of Siberia*. Cambridge: Cambridge University Press.
- Goebel T. 1999. Pleistocene human colonization of Siberia and peopling of the Americas: An ecological approach. *Evol Anthropol* 8:208–227.
- Gogolev AI. 1993. *The Yakuts. Problems of their ethnogenesis and formation of their culture (in Russian)*. Yakutsk: Izdatel'stvo JaGU.
- Hagelberg E. 2003. Recombination or mutation rate heterogeneity? Implications for mitochondrial Eve. *Trends Genet* 19:84–90.
- Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC. 1993. Toward a more accurate time scale for the human mitochondrial DNA tree. *J Mol Evol* 37:347–354.
- Heyer E, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, Labuda D. 2001. Phylogenetic and familial estimates of mitochondrial substitution rates: Study of control region mutations in deep-rooting pedigrees. *Am J Hum Genet* 69:1113–1126.
- Ho SYW, Larson G. 2006. Molecular clocks: When times are a-changin'. *Trends Genet* 22:79–83.
- Howell C, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Hernstadt C. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: There is a difference between phylogenetic and pedigree rates. *Am J Hum Genet* 72:659–670.
- Janhunen J. 1996. *Manchuria. An ethnic history*. Helsinki: The Finno-Ugrian Society.
- Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger K, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Sajantila A. 2000. Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* 66:1580–1588.
- Konstantinov IV. 1975. *The origins of the Yakut people and their culture (in Russian)*. Yakutsk: Yakutskiy Filial SO AN SSSR. p 106–173. Publications of the Prilenskaya Archaeological Expedition.
- Macaulay VA, Richards MB, Forster P, Bendall KE, Watson E, Sykes B, Bandelt HJ. 1997. mtDNA mutation rates—No need to panic. *Am J Hum Genet* 61:983–990.
- Meyer S, Weiss G, Haeseler A. 1999. Pattern of nucleotide substitution and rate heterogeneity in hypervariable regions I and II of human mtDNA. *Genetics* 152:1103–1110.
- Okladnikov AP. 1956. *Ancient population of Siberia and its culture (in Russian)*. In: Levin MG, Potapov LP, editors. *Narody Sibiri*. Moscow: Russian Academy of Science. p 13–98.
- Okladnikov AP. 1970. *Yakutia: Before its incorporation into the Russian state*. Montreal: McGill-Queen's University Press.
- Pakendorf B, Novgorodov IN, Osakovskij VL, Danilova AP, Protopopov AP, Stoneking M. 2006. Investigating the effects of prehistoric migrations in Siberia: Genetic variation and the origins of Yakuts. *Hum Genet* 120:334–353.
- Pakendorf B, Stoneking M. 2005. Mitochondrial DNA and human evolution. *Ann Rev Genom Hum Genet* 6:165–183.
- Pakendorf B, Wiebe V, Tarskaia LA, Spitsyn VA, Soodyall H, Rodewald A, Stoneking M. 2003. Mitochondrial DNA evidence for admixed origins of Central Siberian populations. *Am J Phys Anthropol* 120:211–224.

- Puzyrev VP, Stepanov VA, Golubenko MV, Puzyrev KV, Maximova NR, Kharkov VN, Spiridonova MG, Nogovitsina AN. 2003. mtDNA and Y-chromosomal lineages in the Yakut population. *Russ J Genet* 39:816–822.
- Ray N, Currat M, Excoffier L. 2003. Intra-deme molecular diversity in spatially expanding populations. *Mol Biol Evol* 20:76–86.
- Ricaud F-X, Kolodesnikov S, Keyser-Tracqui C, Alekseev AN, Crubézy E, Ludes B. 2004. Genetic analysis of human remains found in two eighteenth century Yakut graves at At-Dabaan. *Int J Legal Med* 118:24–31.
- Ricaud F-X, Kolodesnikov S, Keyser-Tracqui C, Alekseev AN, Crubézy E, Ludes B. 2006. Molecular genetic analysis of 400-year-old human remains found in two Yakut burial sites. *Am J Phys Anthropol* 129:55–63.
- Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9:552–569.
- Ruhlen M. 1987. *A guide to the world's languages*. Stanford: Stanford University Press.
- Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M. 1994. Mismatch distributions of mtDNA reveal recent human population expansions. *Hum Biol* 66:761–775.
- Shirokogoroff SM. 1966. *Social organization of the northern Tungus*. Netherlands: Anthropological Publications.
- Siguroardóttir S, Helgason A, Gulcher JR, Stefánsson K, Donnelly P. 2000. The mutation rate in the human mtDNA control region. *Am J Hum Genet* 66:1599–1609.
- Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512–526.
- Tokarev SA, Gurchich IS. 1956. The Yakuts (in Russian). In: Levin MG, Potapov LP, editors. *Narody Sibiri*. Moscow: Russian Academy of Science. p 267–328.
- Vasilevich GM. 1969. *Evenki. Istoriko-etnograficheskie ocherki*. Leningrad: Izdatel'stvo, Nauka' Leningradskoe otdelenie.
- Ward RH, Frazier BL, Dew-Jager K, Pääbo S. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci USA* 88:8720–8724.
- Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurler ME, Robinson E, Gerelsaikhan T, Dashnyam B, Mehdi SQ, Tyler-Smith C. 2003. The genetic legacy of the Mongols. *Am J Hum Genet* 72:717–721.
- Zlojutro M, Tarskaia LA, Sorensen M, Snodgrass JJ, Leonard WR, Crawford MH. 2008. The origins of the Yakut people: Evidence from mitochondrial DNA diversity. *Int J Hum Genet* 8:119–130.