

Historical peopling of Siberia: the origin of Samoyedic-speaking populations

T. M. Karafet¹, L. P. Osipova^{2,3}, and M. F. Hammer¹

¹ University of Arizona, Tucson, USA; ² Institute of Cytology and Genetics, Novosibirsk, Russia;

³ Novosibirsk State University, Novosibirsk, Russia

Key words: Siberia, Samoyedic-speaking populations, genetic structure

Introduction.

The peoples of northwest Siberia speak languages belonging to the Finno-Ugric and Samoyedic branches of the Uralic linguistic family. Samoyedic people lead nomadic or semi-nomadic life in the Taymyr Peninsula and in the tundra or forest zone between the Ural Mountains and the Yenisey River. The Samoyedic languages fall into two main branches. The Northern Samoyedic branch includes languages of the Tundra and Forest Nentsi, Entsi, and Nganasans. The only Southern Samoyedic language with a considerable number of speakers is Selkup. Most Russian ethnographers and anthropologists believe that all contemporary Samoyedic-speaking peoples are descendants of Paleoasiatic (autochthonous) populations, which were assimilated by ancient Samoyeds who came from southern Siberia. Nevertheless, several questions remain open, such as “who were the ancestral autochthonous tribes” and “when did they settled these geographic areas.” A consensus has also not been achieved on the original homeland of Samoyeds. Our study examines autosomal genome-wide SNPs and Y-chromosomal variation to reconstruct the historical events that led to the peopling of Northwest Siberia.

Methods.

DNA samples were genotyped for 567,096 single nucleotide polymorphisms (SNPs) on Affymetrix platform using standard protocols. Analyses of linkage disequilibrium, IBD (identically by descent) and the distribution of runs of homozygosity were performed on full data set of autosomal SNPs. For several analyses we used 281,093 SNPs from the intersection of our data with publicly available ancient Siberian samples. We applied Principal Components, ADMIXTURE, and TreeMix program analyses on the merged autosomal data set, and F_{ST} , f_3 and D statistics were calculated.

Results.

Four Samoyedic-speaking populations were analyzed in this study: the Nganasans, Tundra Nentsi, Forest Nentsi, and Northern Selkups. We also included samples from an additional 23 populations from Siberia, Europe, Central Asia, East Asia, South Asia, and Near East. Our joint analyses based on autosomal and Y chromosome data show that Samoyedic-speaking populations do not cluster genetically strictly according to their linguistic affiliation. Only Forest and Tundra Nentsi demonstrate genetic similarity. Selkups are genetically close to Kets, while Nganasans are drawn toward Evenks and Yukagirs. Interestingly, nearly all Siberian populations exhibit a signal of admixture with Nganasans in f_3 and IBD analyses. When Siberian populations were modeled as a mixture of ancient Siberian cultures and present-day populations, the highly significant negative f_3 statistics were observed for a combination of ancient Mal'ta, Afanasievo, and Andronovo as the first source, and Nganasans or Evenks as the second mixing source population.

Conclusions.

The results of this study are consistent with the hypothesis that the Paleoasiatic ancestors were different for Nentsi (Forest and Tundra), Nganasans, and Selkups. Taking into account genetic variation and archaeological evidence we put forward a colonization model that encompasses early Paleolithic settlement processes, as well as events in historic times. We demonstrate that migration, distinctive marriage structure, and possibly selection shaped genetic differentiation among populations in Siberia.

Acknowledgements. This work was supported by National Science Foundation grants to T.M.K. and M.F.H. (PLR-1203874), and by the State Research Project (No 0324-2016-0002) to L.P.O.