DOI 10.18699/SBB-2022-230

Detailed phylogeny of the Y-chromosome haplogroup N1a2 in the populations of Siberia and Eastern Europe

Kharkov V.*, Stepanov V.

Research Institute of Medical Genetics, TNMRC, Tomsk, Russia

** vladimir-kharkov@medgenetics.ru*

Key words: human population genetics, haplogroups, phylogeny and phylogeography of Y-chromosome lineages

Motivation and Aim: The modern level of solving the fundamental problems of general human genetics and population genetics – the study of genetic diversity and the structure of the gene pools of human populations, the mechanisms of its formation, the reconstruction of migrations and human settlement, is most thoroughly implemented in the aspect of analysis of data from sequencing of complete genomes and phylogeographic and phylogenetic analysis of Y-chromosome haplogroups, using the latest advances in the discovery and genotyping of new SNP markers by massive parallel sequencing.

Methods and Algorithms: We carried out a detailed analysis of phylogeny of the Y-chromosome haplogroup N1a2 in various populations of Siberia and the Volga-Ural region using bioinformatics processing of whole genome sequencing data. The samples represent all indigenous peoples living in these territories and belong to different subclades previously identified by YSTR haplotypes. Samples were genotyped from Siberian (Khanty, Nenets, Siberian Tatars, Southern Altaians, Kumandins, Tubalars, Chelkans, Khakases (Sagais and Kachins), Shors, Chulym Turks, Tuvans, Buryats (eastern and western), Evenks, Yakuts) and Eastern European populations (Komi, Udmurts, Mari, Bessermen, Mordovians, Kazan Tatars, Bashkirs). After analyzing genome-wide data, hundreds of terminal SNPs were selected for the most detailed separation of subhaplogroups. Some of these SNPs were completely genotyped on the entire array of population samples.

Results: According to the results of genotyping, the age of separation of various branches of all subclades was calculated, their component contribution to the gene pools of all the studied peoples and their genetic continuity between different populations were assessed. The phylogeny of all ethnospecific sublines of the haplogroup N1a2 is described in detail. Median networks of YSTR haplotypes were constructed, reflecting the expansion of the number of ancestral populations, which coincide very well with the data on SNP markers. The initial territory for the emergence of all basic subclades N1a2 is the Sayan-Altai region. Further, their separation occurred during the migrations of various ancestral groups in different time periods. The main connection between the distribution of different N1a2 subclades can be traced with representatives of the Samoyedic and Finno-Ugric languages. The frequency of different sublines of this haplogroup in different populations correlates well with the proportions of different genetic components in their gene pools, calculated based on the analysis of whole genome sequencing data. Almost all sublines show significant population specificity. The results of a comparative analysis of the spectrum of haplotypes convincingly indicate the specificity of Y-chromosome sublines and clusters of haplotypes not only at

the level of ethnic groups and populations but also at the level of phratries, tribes, and individual family groups.

Conclusion: Genotyping of a wide range of selected terminal YSNPs and YSTRs on the entire array of population samples allowed for the most accurate and detailed separation of all sublines between populations. This further clarifies the relationship of subhaplogroups with migration processes, founder effects in different populations, genetic components, demographic growth, and the continuity of the genetic contribution between different populations.

Acknowledgements: The study is supported by The Research Institute of Medical Genetics, Tomsk National Research Medical Center (122020200083-8).