A new linear combination method of haplogroup distribution central vectors to model population admixtures

Tibor Török, Kitti Maár, István Gergely Varga, University of Szeged
Zoltán Juhász Hun-REN Centre for Energy Research

In this study, we examined the genetic composition of female members of 101 ancient and 71 modern populations. We characterized the populations by the frequency distribution vectors of 74 appropriately selected mitochondrial haplogroups.

Our first goal was to find groups populations having similar haplogroup distributions. We identified 35 such groups using the "self-organising cloud" algorithm developed for this purpose. The algorithm described these clusters by the mean vector of the haplogroup distributions of the populations belonging to the given cluster.

It is an obvious assumption that we attributed 35 hypothetical ancient populations to the 35 mean haplogroup distributions determined by the algorithm. This hypothesis was verified by certain real ancient populations that proved to be significantly and convincingly similar to one of the 35 mean haplogtoup distributions.

After that, we posed the question how we can characterize the real populations as a mixture of these hypothetical ancient populations. For this purpose, we developed the linear combination algorithm. This algorithm approximated the real population as the weighted sum of 35 the mean haplogroup distributions. Real and modeled haplogroup distributions of Copper Age and 9-11th century Hungary, as well as modern Hungarians are shown in Figure 1.

Three populations were scrutinized in the article: the populations of the Carpathian Basin in Copper Age, in the 9-11th century and of the Modern Age. According to the results the female haplogroups of the Copper Age populations of the Carpathian Basin remained deterministic components both in the 9th century and Modern Age.
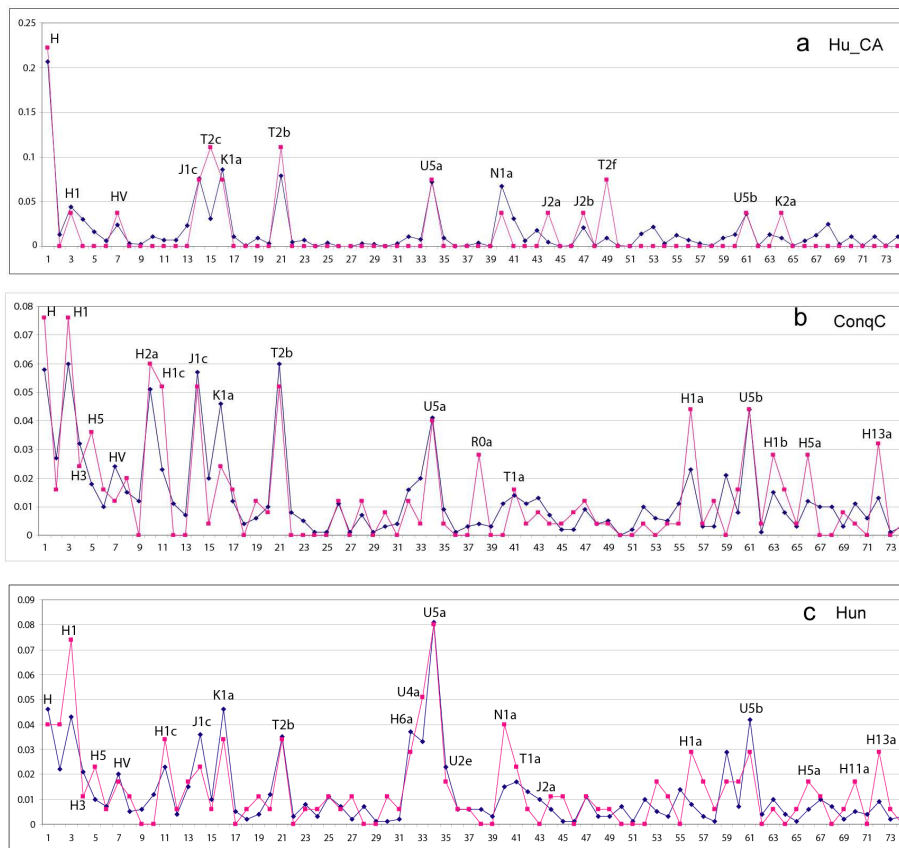
Figure 1. Hg distributions of a: Hungarian Copper Age (Hu_CA), b: Conqueror commoner (9-11th century ConqC) and c: modern Hungarian (Hun) populations. Red lines indicate original Hg distributions, while blue lines show modeled distributions by the linear combination algorithm. The similarity of the haplogroup distributions of 9-11th century Conqeror (ConqC) and modern Hungarian haplogroup distributions is particularly striking.

Reference:

Török T, Maár K, Varga IG, Juhász Z. A new linear combination method of haplogroup distribution central vectors to model population admixtures. Mol Genet Genomics. 2022 May;297(3):889-901. doi: 10.1007/s00438-022-01888-0. Epub 2022 Apr 11. PMID: 35411488; PMCID: PMC9130205.

Online source:

https://link.springer.com/article/10.1007/s00438-022-01888-0