Population genetic structure in the Balkans

Margaritis Tsifintaris[1], Fotis Tsetsos[1], Iordanis Karagiannidis[1],
Dimitrios Mantzaris[1], Petros Drineas[2], Peristera Paschou[1]

[1]Department of Molecular Biology & Genetics, Democritus University of Thrace, Alexandroupoli, Dragana, Greece
[2]Department of Computer Science, Rensselaer Polytechnic Institute Troy NY,United States of America
*e-mail: margtsf@mbg.duth.gr

Populations that have lived in the broader region of Balkans have played a crucial role for the European and world-wide history, acting as a gateway that shaped European genetic variation. In an effort to shed some light on the genetic diversity of the populations of the Balkan Peninsula we analyzed data from 11 populations that were extracted from the Estonian Biocentre dataset. In order to process the above-mentioned datasets, a principal component analysis (PCA) approach was followed. Data were analyzed using Eigensoft 6.0.1 and PLINK 1.07. The EIGENSOFT package combines functionality from population genetics methods (Patterson et al. 2006) and also includes EIGENSTRAT, a method for stratification correction (Price et al. 2006). In particular, a total of 133 individuals (86 men and 47 women) originating from the 11 populations were analyzed. A total of 554,417 single nucleotide polymorphisms (SNPs) were extracted using self-developed scripts and used for further analysis. Despite their geographical proximity, we uncover great diversity and structure among the neighboring populations in the Balkans. Results presented in the PCA, reflect the geographical distribution of the studied populations and especially the variation from the south to north.

1. Patterson N, Price AL, Reich D (2006) Population Structure and Eigenanalysis. PLoS Genet 2(12): e190. doi:10.1371/journal.pgen.0020190 .
2. Alkes L Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick & David Reich (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38, 904–909.