

Genetic Landscape of Eurasia Viewed from Large Allele Frequency Differences

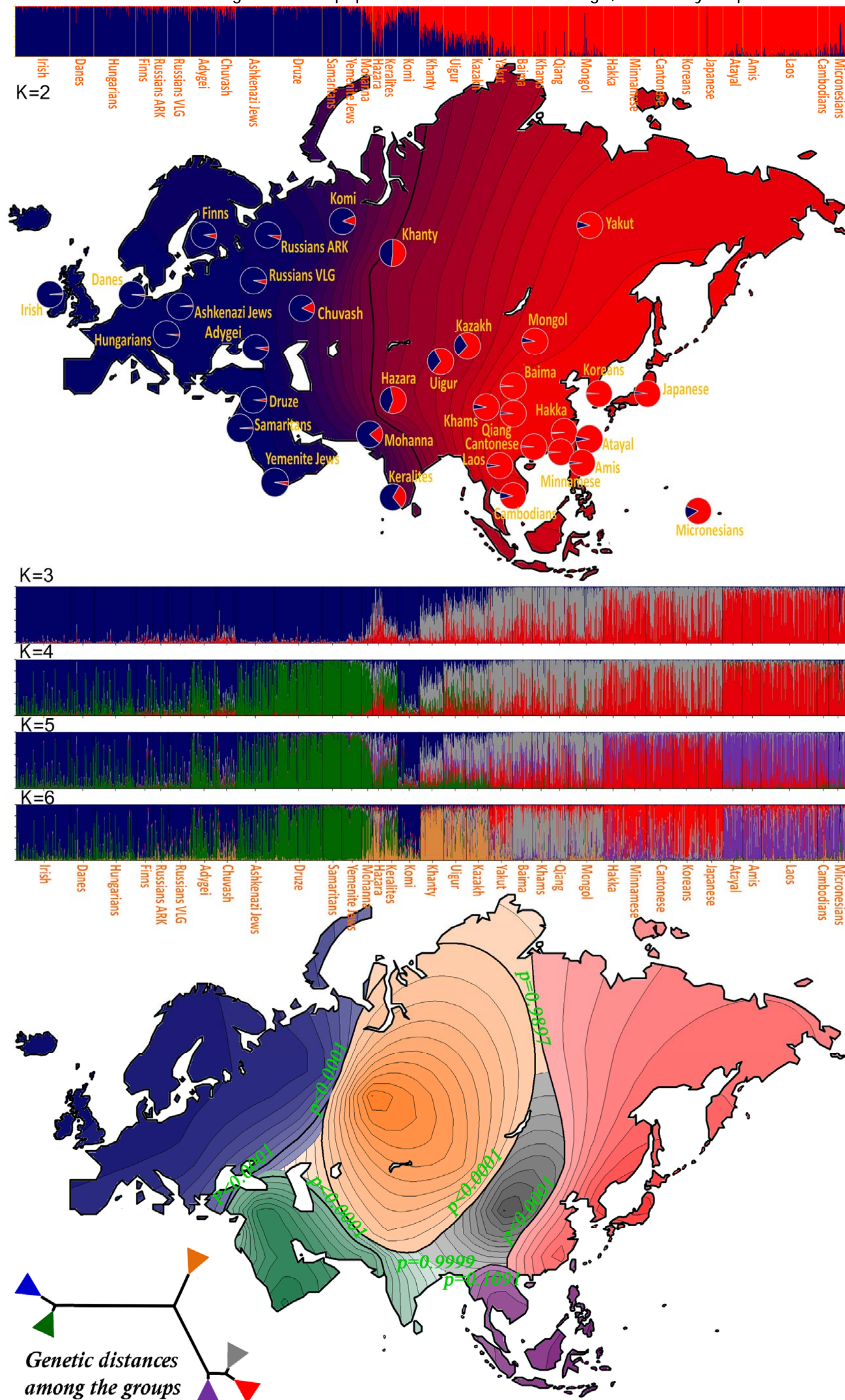
Hui Li*, Kelly Cho, Judith R. Kidd, Kenneth K. Kidd

Superlative Science,
Sensational Setting

Department of Genetics, School of Medicine, Yale University, New Haven, CT 06520 USA;

* State Key Lab of Genetics and MOE Key Lab of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433 China.

Abstract: The diversification leading to modern human populations in Eurasia is one of the most important topics in the study of human expansions out of Africa. Most studies of Eurasian populations have either used limited markers or involved insufficient population coverage. We chose 68 markers based on large allele frequency differences among several Eurasian populations and then typed them on 1766 individuals from 34 populations representing all subdivisions of Eurasia. Analyses using the STRUCTURE program showed a clinal east-west division when K=2, with a median border dividing Central Asia along the Ob River, the Kazakh highland, the western side of Pamir Mountains, and the southwestern side of the Himalayas. We fit curves to the STRUCTURE loadings using distances of the population coordinates from the median border. The genetic structure changed dramatically only within 2000km on each side of the border. At higher values of K, the western populations of East Asia are the first to be distinguished (at K=3); Mongols, Tibetans, Qiang, and Baima, are most distinct from the more eastern populations. At K=4, Southwest and South Asians are distinguished from the Europeans; Southeast Asians and Central Asians are successively distinguished from eastern East Asians at K=5 and 6, respectively. Several more isolated populations such as Samaritans, Atayals, or Micronesians were distinguished in different independent runs when K=7 providing no clear anthropological information. South Asians were always clustered with Southwest Asians with pronounced similarity to Central Asians. The failure to distinguish South Asians maybe due to the selection of the markers with large allele frequency differences specifically between Europeans and East Asians. We also tested for statistical differences in the allele frequencies for all pairs of clusters when K=6. The results showed significant borders (P<0.0001) including those between western East Asians and eastern East Asians or Central Asians; however, insignificant borders were observed between Southwest Asians and Southeast Asians or western East Asians and between Central Asians and eastern East Asians. This indicates substantial gene flow in North Asia between eastern East Asians and Central Asians, and in South Asia between South Asians and Southeast Asians. Using increased population and marker coverage, this study helps to understand the details of genetic diversity and landscape of Eurasians.



Introduction: The genetic landscape of Eurasia, or the genetic structure of Eurasian populations, is one of the most intriguing topics in the field of human genetics. However, a comprehensive study with sufficient population coverage and genomic markers is yet to be done. Here we use large allele frequency difference data of various population samples to view the genetic landscape.

Materials & Methods: A total of 1766 individuals from 34 populations (Fig.1) were typed on 68 markers with high global Fst. The genetic similarities among the populations were estimated using STRUCTURE 2.2 program, and the results are displayed onto a map in Figure 1 using SURFER 8.0. Correlation between genetic and geographic distances was estimated by the curve fit tool in SPSS 13.0. The geographic distance between two locations A and B was calculated according to the following formula: $Dis=R*ACOS(1-(SQRT((R*COS(latA/180*\pi))^2+(R*COS(latB/180*\pi))^2+(R*(SIN(latB/180*\pi)-SIN(latA/180*\pi))^2-2*R^2*COS(latA/180*\pi)*COS(latB/180*\pi)*COS(ABS(lngA-lngB)/180*\pi))^2/(2*R^2)))$. For the six clusters, the significance of borders was examined using t-tests (p values marked in Fig.1)

Results: A clear west-east cline was observed when K=2 (Fig.1) with a median line between Tibet and Iran. The genetic structure changes gradually away from the median line and reaches the extremum within 2000km (Fig.2). The whole Eurasia can be divided into six parts genetically. Beside the well-known European and West Asian clusters, East Asia was divided into three, the South, the West, and the East. The Central Asian cluster was indistinguishable until when K=6, indicating that Central Asia is an admixture of the east and the west.

These results suggest a further discussion in the fields of anthropology, epidemiology and population genetics.

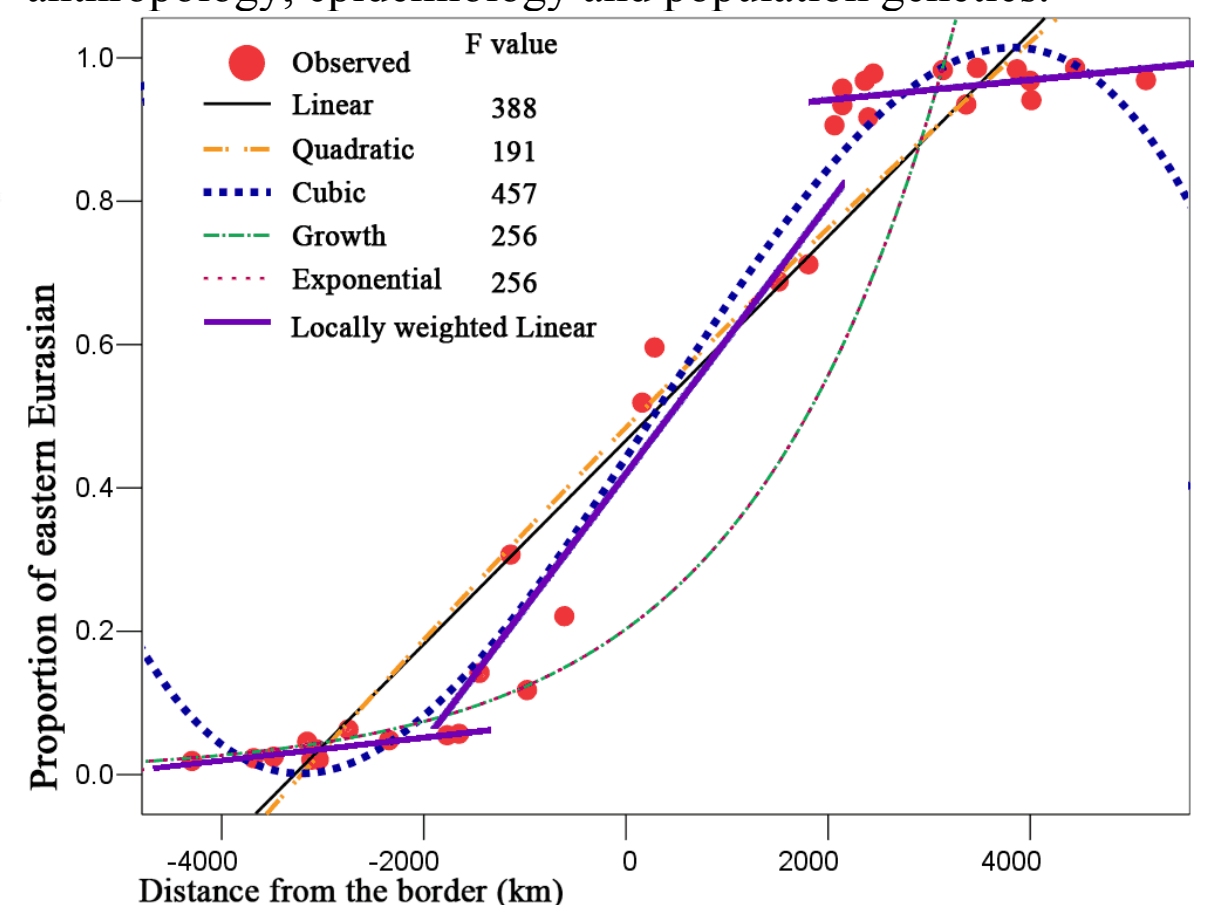


Figure 1. Bar views and contour map views of the genetic structure of Eurasian populations.

Figure 2. Relationship between genetic and geographic distances when K=2.

This study was funded, in part, by National Institute of Health grant P01 GM057672 (KKK and JRK) and NIH grants 2004-DN-BX-K025 and 2007-DN-BX-K197 (KKK) awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. HL was also supported by funding from Fudan University.