

Genetic Landscape of Eurasia and “Admixture” in Uyghurs

To the Editor: In the papers^{1,2} by Xu and Jin, the genetic structure of Uyghurs was described by 8150 ancestry-informative markers (AIMs). These markers estimated the admixture rate of the Uyghur population to be around 50% East Asian ancestry by comparing Uyghurs to East Asians and Europeans. However, we suspect that the estimate of Xu and Jin may be considerably biased by insufficient reference population coverage. In their study, Xu and Jin used the STRUCTURE program^{3,4} as the major method for estimating the admixture rate; however, results from the STRUCTURE program are strictly the probabilities of assignment to different estimated clusters and therefore are influenced by both the marker selection and the populations used to estimate allele frequencies in those other clusters. The results of Xu and Jin are unlikely to be biased due to marker selection because they used a large number of AIMs selected to distinguish Europeans from East Asians. However, the population coverage in their analysis was relatively sparse. They only analyzed Japanese, northern Chinese, and a very small sample ($n = 10$) of Mongols for the East Asian reference. For Europeans, they included some southern European populations, Russians, and Adygei. It is doubtful that this sparse population coverage can yield precise results. Although the cluster assignments can be interpreted as admixture under many circumstances, these reference populations may not serve as relevant ancestral populations for an admixture analysis of Uyghurs. We also note that, for populations in the middle of a clinal distribution, STRUCTURE cannot readily distinguish recent admixture from a historical cline established by an expanding population as in the original spread of modern humans out of Africa and across Eurasia.⁵

In our study, we typed 68 AIMs on 1766 individuals from 34 populations representing all subdivisions of Eurasia (Figure 1 and Table S1 available online; allele frequency data also in ALFRED). All samples were collected with informed consent by participants using protocols approved by all relevant institutional review boards. The SNPs were selected at random from a group of ~300 markers having a high global F_{ST} and already tested on many populations, including some reported here. However, the selection paid no attention to the pattern of variation across Eurasia. This “randomness” is illustrated by the Eurasian F_{ST} values for these 68 markers, which range from 0.028 to 0.681 in a highly skewed distribution with a median value of ~0.15 (Figure S1). Our set of populations contained several population samples for those regions around Uyghurs, such as Kazakhs, Mongols, Tibetan Khams, etc. to the east of Uyghurs and Hazaras, Khanty, etc. to the west. This comprehensive population coverage allowed more reliable estimates of relationships among the

populations. We used STRUCTURE 2.2 and SURFER 8.0 programs to illustrate the population similarities (Figure 1). The analyses showed a very clear west-east cline when $K = 2$. The median border estimated from these 68 SNPs divided Central Asia along the Ob River, the Kazakh highland, the western side of the Pamir Mountains, and the southwestern side of the Himalayas. Although the small number of markers examined cannot result in an accurate median border, the Uyghur population fell east of the median border with 31.2% assignment to the western cluster. Even the Hazaras and Khanty were to the east of the median border. This result matched the results of more anthropological studies on Eurasian population structures with the Hazaras around the median.⁶ Other anthropological studies also estimated Uyghurs to have ~30% western proportions,^{7,8} closer to our estimate.

The difference between the estimate of Xu and Jin (52%) and our estimate (31%) may stem from either the different population coverage or the sample size. We analyzed a different and larger sample of Uyghur individuals ($n = 48$) than that analyzed by Xu and Jin.² Their small sample size may have contributed to their overestimation of the European component to admixture (i.e., to cluster assignment). However, the insufficient population coverage may be more responsible for the difference than the sample size or the number of markers. Concerning the number of markers, it is known that a relatively small but specifically selected number of AIMs can accurately predict ethnicity proportion.⁹ As the two papers of Xu and Jin have demonstrated, the estimated admixture rates reported did not change much regardless of whether they were using chromosome 21 data only or the whole genome, and thus a large number of markers may not be necessary to estimate the “admixture” rate of Uyghurs. When we analyzed only the 12 markers with the highest F_{ST} values in our samples (Figure S1), the Uyghurs had a 30.2% assignment at $K = 2$ to the Europe and Western Asia cluster. This estimate was not significantly different from the above 31.2% when using all 68 markers. We consider it unlikely that a different set of appropriately chosen SNPs would give a markedly different answer based on unpublished data on some of these same populations.

A fundamental problem with this estimate at $K = 2$ is the high improbability of the Uyghur population being admixed of two widely separated populations such as Europeans and eastern East Asians. Therefore, we also tested for additional subdivisions with STRUCTURE to see which populations were more closely related to Uyghurs. We observed that Central Asian populations including Uyghurs, Kazakhs, and Khanty did not form their own cluster until $K = 6$, indicating that the Central Asian cluster was not a completely distinct population group. From $K = 3$ to $K = 5$, the western part of East Asia (Kham, Baima, Qiang, Mongol) was distinguished from the eastern part of East Asia (Japanese, Korean, Hakka, Minnamese, Cantonese). Examining the “admixture” pattern of Uyghurs, it is clear that the proportion of western East

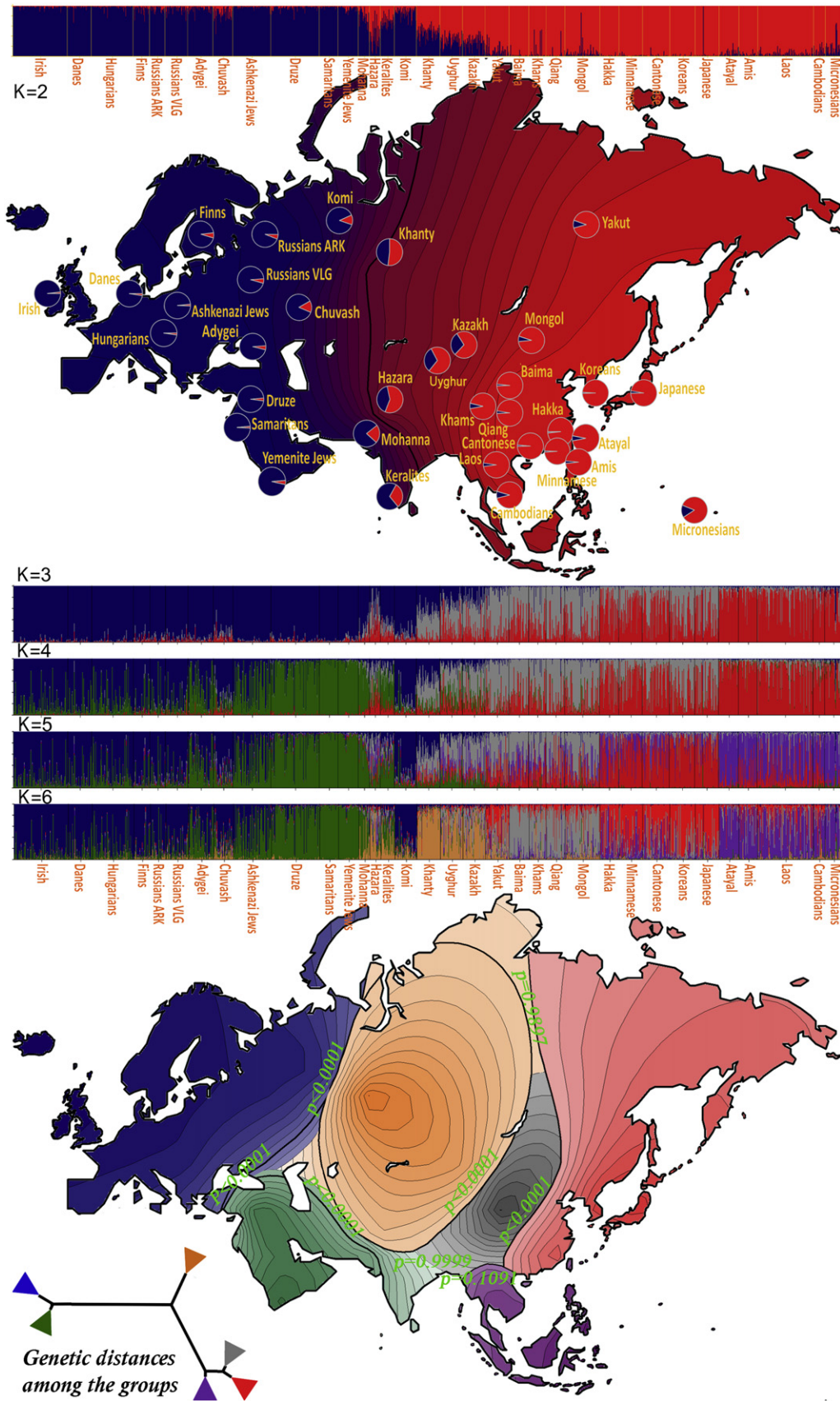


Figure 1. STRUCTURE Bar Views for K = 2–6 and Contour Plots for K = 2 and K = 6
 The contour plots and genetic distance tree are color coded to correspond to the STRUCTURE plots. The significance levels for pairwise comparisons of clusters are given on the K = 6 contour plot (bottom) at the relevant borders.

Asia is much higher than that of eastern East Asia, especially when $K = 5$ (0.418 versus 0.128, respectively). Moreover, the South Asians and West Asians also contributed more than the Europeans to Uyghurs (0.180 versus 0.100, respectively). Including the western East Asians, South Asians, and West Asians when estimating the admixture rate of Uyghurs illustrates the difficulty with the concept of admixture for such intermediate populations—the estimate depends on the populations hypothesized to be admixed in the target population. Considering only the Europeans and the eastern East Asians may seem to indicate equal “contributions” to Uyghurs, which was the case in the study of Xu and Jin.

In addition, our study showed that Central Asia clearly formed a cluster with significant borders when $K = 6$. The border of the Central Asian cluster went along Lake Baikal, the A-erh-chin Mountains, the Kunlun Mountains, the Hindu Kush, the Caspian Sea, and the Ural Mountains, matching the traditional anthropological definition of Central Asia. We performed pairwise *t* tests for the six clusters to estimate the significance of the borders (*p* values are marked on the map at bottom in Figure 1). The only insignificant gap around the Central Asian cluster was in northern Siberia, where people led a nomadic hunting lifestyle. Another insignificant border was the southern border of the western East Asian cluster, where people also led a nomadic hunting lifestyle. That area is believed to have been a migration pathway into East Asia for early modern humans.^{10,11} Such a nomadic hunting lifestyle might have been the easiest way to blur borders arising among populations through admixture during more recent human history. Though nomadic lifestyles may have blurred distinctions, those nonsignificant cluster comparisons might also simply be due to a lack of power for this specific set of SNPs to distinguish differences. Other borders among all pairs of the six clusters were significantly distinct. It is reasonable that the Caucasus, the Anatolian plateau, and the Himalayas became the borders of the clusters by minimizing gene flow and allowing allele frequency differences to accumulate by drift.

Our analyses significantly divided East Asia into eastern and western parts, agreeing with the hypothesis that early modern humans entered East Asia in the south along two routes, the western route from Myanmar to Yunnan and the eastern route from Vietnam to Guangdong.^{10,12,13} The descendants of the migrants along these two routes would have accumulated significant genetic differences and subsequently would have had different effects on the gene pool of Uyghurs. Here we have demonstrated that the western East Asians are more closely related to Uyghurs than the eastern East Asians.

To confirm the relative genetic affinity between Uyghurs and the other Eurasian populations, we performed principal component analysis (PCA) with SPSS 13.0. The results of principal components 1 versus 2 are plotted in Figure 2. It is obvious that two clusters exist: a European-West Asian and a far East Asian. The South Asian and Central Asian

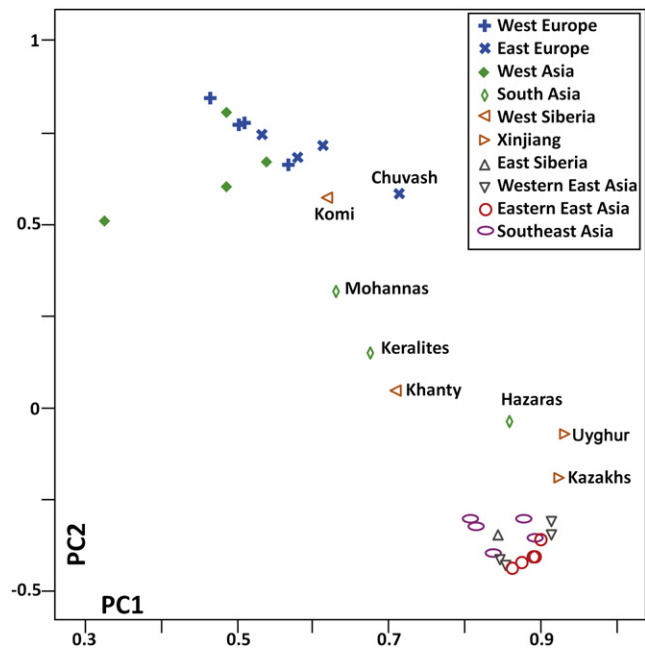


Figure 2. Principal Component Plot of Eurasian Populations Just over 80% of the variance among these populations based on the 68 SNPs in this study can be explained by the first two components.

populations are scattered between the clusters of Europeans and East Asians. Notably, Uyghurs are much closer to the East Asians than to the Europeans. The clinal pattern seen in the PCA is also suggested by the sequential subdividing of the groups of populations in the series of increasing K for the STRUCTURE analyses. The first two PCA components account for just over 80% of the total variance, probably because some of the SNPs have very large allele frequency differences between Europe and far East Asia (Table S1).

STRUCTURE cannot distinguish recent admixture from a cline of other origin, and these analyses cannot prove admixture in the Uyghurs; however, historical records indicate that the present Uyghurs were formed by admixture between Tocharians from the west and Orkhon Uyghurs (Wugusi-Huihu, according to present Chinese pronunciation) from the east in the 8th century CE.¹⁴ The Uyghur Empire was originally located in Mongolia and conquered the Tocharian tribes in Xinjiang. Tocharians such as Kroran have been shown by archaeological findings to appear phenotypically similar to northern Europeans,¹⁵ whereas the Orkhon Uyghur people were clearly Mongolians. The two groups of people subsequently mixed in Xinjiang to become one population, the present Uyghurs. We do not know the genetic constitution of the Tocharians, but if they were similar to western Siberians, such as the Khanty, admixture would already be biased toward similarity with East Asian populations.

In conclusion, we argue that the Uyghurs’ genetic structure is more similar to East Asians than to Europeans, in contrast to the reports by Xu and Jin, whose work may

have been affected by their sparse population coverage. The median line of the Eurasian genetic landscape appears to lie to the west of the Xinjiang Uyghur Autonomous Region of China. When we have collected more data on these 34 populations, we should be able to refine these estimates.

Hui Li,¹ Kelly Cho,¹ Judith R. Kidd,¹
and Kenneth K. Kidd^{1,*}

¹Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA

*Correspondence: kenneth.kidd@yale.edu

Supplemental Data

Supplemental Data include one figure and one table and can be found with this article online at <http://www.cell.com/AJHG>.

Acknowledgments

This study was funded in part by National Institutes of Health grant P01 GM057672 (J.R.K. and K.K.K.) and NIJ grants 2004-DN-BX-K025 and 2007-DN-BX-K197 (K.K.K.) awarded by the National Institute of Justice, Office of Justice Programs, United States Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the Department of Justice. We thank all of the colleagues who helped us assemble the population samples, including the National Laboratory for the Genetics of Israeli Populations at Tel Aviv University. We give special thanks to the many hundreds of individuals from the relevant populations who volunteered to give blood samples for studies such as this.

Web Resources

The URL for data presented herein is as follows:

ALFRED, <http://alfred.med.yale.edu>

References

1. Xu, S., Huang, W., Qian, J., and Jin, L. (2008). Analysis of genomic admixture in Uyghur and its implication in mapping strategy. *Am. J. Hum. Genet.* 82, 883–894.
2. Xu, S., and Jin, L. (2008). A genome-wide analysis of admixture in Uyghurs and a high-density admixture map for disease-gene discovery. *Am. J. Hum. Genet.* 83, 322–336.
3. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
4. Falush, D., Stephens, M., and Pritchard, J.K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7, 574–578.
5. Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., and Feldman, M.W. (2002). Genetic structure of human populations. *Science* 298, 2381–2385.
6. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., and Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319, 1100–1104.
7. Du, R. (1997). Human population genetics studies in China. *Bulletin of Biology* 32, 9–12.
8. Xiao, F.X., Yang, J.F., Cassiman, J.J., and Decorte, R. (2002). Diversity at eight polymorphic Alu insertion loci in Chinese populations shows evidence for European admixture in an ethnic minority population from northwest China. *Hum. Biol.* 74, 555–568.
9. Sampson, J.N., Kidd, K.K., Kidd, J.R., and Zhao, H. (2008). Selecting SNPs to correctly predict ethnicity [Platform Session 20: Statistical Genetics I, No. 60]. Presented at the annual meeting of The American Society of Human Genetics, November 11–15, 2008, Philadelphia. Available at the following URL: <http://www.ashg.org/2008meeting/abstracts/fulltext/>.
10. Li, D., Li, H., Ou, C., Lu, Y., Sun, Y., Yang, B., Qin, Z., Zhou, Z., Li, S., and Jin, L. (2008). Paternal genetic structure of Hainan aborigines isolated at the entrance to East Asia. *PLoS ONE* 3, e2168.
11. Li, H., Cai, X., Winograd-Cort, E.R., Wen, B., Cheng, X., Qin, Z., Liu, W., Liu, Y., Pan, S., Qian, J., et al. (2007). Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am. J. Phys. Anthropol.* 134, 481–488.
12. Li, H., Huang, Y., Mustavich, L.F., Zhang, F., Tan, J.Z., Wang, L.E., Qian, J., Gao, M.H., and Jin, L. (2007). Y chromosomes of prehistoric people along the Yangtze River. *Hum. Genet.* 122, 383–388.
13. Li, H., Wen, B., Chen, S.J., Su, B., Pramoongjago, P., Liu, Y., Pan, S., Qin, Z., Liu, W., Cheng, X., et al. (2008). Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evol. Biol.* 8, 146.
14. Mackerras, C. (1972). *The Uighur Empire According to the T'ang Dynastic Histories* (Canberra, Australia: Australian National University Press).
15. Mallory, J.P., and Mair, V.H. (2000). *The Tarim Mummies: Ancient China and the Mystery of the Earliest Peoples from the West* (London: Thames & Hudson).

DOI 10.1016/j.ajhg.2009.10.024. ©2009 by The American Society of Human Genetics. All rights reserved.

Response to Li et al.

To the Editor: Li et al. analyzed 68 SNPs genotyped on 1766 individuals from 34 populations and provided a lower estimation (31.2%) than ours (47%–52%)¹ of European genetic ancestry in Uyghurs. They argue that our estima-

tion “may have been affected by sparse population coverage.” The study of Li et al. is very interesting and provides some new insights into the genetic landscape of Eurasia. Here we show that the discrepancy between the two estimations could be attributable to either the difference in Uyghur samples per se, the variation of the estimation using a small number of markers by Li et al., or both.