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,*]
[1]Department 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., Pramoonjago, 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).

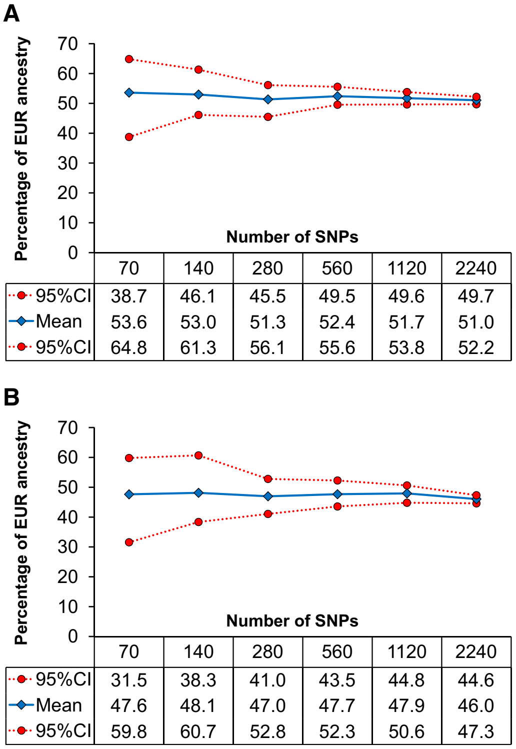
DOI 10.1016/j.ajhg.2009.10.025.

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 estimation "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.

In our study, we analyzed two Uyghur (UG) population samples, one from Yili, which is located in northern Xinjiang (hereafter referred to as N-UG), and the other from Hetian (or Hotan), located in southern Xinjiang (S-UG). Comparing the SNPs typed in Uyghur samples of Li et al. (Li-UG) and those in ours, 6 of 20 SNPs shared by Li-UG and S-UG and 3 of 6 SNPs shared by Li-UG and N-UG showed large allele frequency difference (>0.1). The average F_{ST}s between Li-UG and our Uyghur samples (0.0048 and 0.0094) are significantly larger than that between N-UG and S-UG (0.0009; see Table 1), suggesting that the Uyghur samples that Li et al. used could be rather different in genetic structure from ours. It is therefore not unexpected that the two studies give different estimations.

In addition, because Li et al. used only 68 random SNPs for ancestry estimation, we suspected that a substantial variation in estimations could be introduced by their small number of markers per se. Indeed, we noted in previous studies that the estimation of admixture proportion could vary in different marker panels even in the same population sample.[2,3] We conducted STRUCTURE analyses with a different number of SNPs that were randomly selected from genome-wide data in our previous study.[1] We found that as much as a 26% difference (the width of the 95% confidence interval) could be introduced in the genetic ancestry estimations when only 70 random SNPs were used (Figure 1), which is larger than the difference between the estimation of Li et al. (31.2%) and our estimation (47%–52%). It is clear that the variation of estimations decreases as the number of markers increases (Figure 1); for example, the difference was only 2.5% when 2240 SNPs were used, indicating that a considerable number of random markers are needed for a reliable estimation of genetic ancestry in an admixture population.



A

	70	140	280	560	1120	2240
● 95%CI	38.7	46.1	45.5	49.5	49.6	49.7
◆ Mean	53.6	53.0	51.3	52.4	51.7	51.0
● 95%CI	64.8	61.3	56.1	55.6	53.8	52.2

B

	70	140	280	560	1120	2240
● 95%CI	31.5	38.3	41.0	43.5	44.8	44.6
◆ Mean	47.6	48.1	47.0	47.7	47.9	46.0
● 95%CI	59.8	60.7	52.8	52.3	50.6	47.3

Figure 1. Estimation of the Percentage of European Ancestry in Uyghurs via Different Numbers of Markers

To estimate the European (EUR) ancestry in southern Uyghur (S-UG; A) and northern Uyghur (N-UG; B) populations, we used CEU and CHB (see text) as reference populations. For each given number of SNPs, 100 data sets with randomly selected markers were generated, and STRUCTURE analyses were performed with 20,000 iterations after a burn-in of length 30,000, with the admixture model and assuming that allele frequencies were correlated.

Furthermore, to exam the possible bias in estimation due to the selection of reference populations, we reanalyzed both the short tandem repeat (STR) and SNP data of a Human Genome Diversity Project panel that contained 940 unrelated individual samples collected from 52 worldwide populations. We ran STRUCTURE analysis for the 783 STRs[4] and 4600 randomly selected SNPs from Illumina HumanHap650K genome-wide data.[5] We found that the addition of more reference populations did not yield a significantly different estimation of European genetic ancestry in Uyghurs from those based on Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) and Han Chinese from Beijing (CHB) populations in the HapMap study (see Figure S1 available online). In addition, via haplotype sharing analysis, a new method that we proposed recently,[6] we estimated that the contribution of European genetic ancestry to S-UG is 56%, which is very close to our previous estimation based on STRUCTURE analysis and the same Uyghur samples as well as the same data.[2]

Table 1. Pairwise F_{ST} between Uyghur Population Samples

	Li-UG	N-UG
N-UG	0.0048 ± 0.0047	
S-UG	0.0094 ± 0.0078	0.0009 ± 0.0004

Li-UG represents Uyghur samples used by Li et al.; N-UG represents Uyghur samples collected from northern Xinjiang (Yili); S-UG represents Uyghur samples collected from southern Xinjiang (Hotan). F_{ST} between Li-UG and N-UG was calculated based on allele frequencies of 6 overlapped SNPs; F_{ST} between Li-UG and S-UG was calculated based on allele frequencies of 20 overlapped SNPs; F_{ST} between N-UG and S-UG was obtained from Xu and Jin.[1] Standard deviation is shown for each pairwise F_{ST}.

In summary, our extended analyses suggest that the difference between the two estimations could be attributable to either the selection of different Uyghur population samples by Li et al. versus those that we used in our previous studies, the small number of markers that they used, which could introduce considerable variance in the estimation, or both. Regardless of the true cause of the discrepancy, the different estimations should prompt a close examination of the presence of a possibly substantial variation of admixture among Uyghur populations, a phenomenon similar to that observed in African Americans and Hispanics.

Shuhua Xu[1,2] and Li Jin[1,2,3,4,*]
[1]Chinese Academy of Sciences and Max Planck Society (CAS-MPG) Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; [2]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Chinese Academy of Sciences, Shanghai 200031, China; [3]State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433, China; [4]China Medical City Institute of Health Sciences, Taizhou, Jiangsu 225300, China
*Correspondence: ljin007@gmail.com

Supplemental Data

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

References

1. 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.
2. 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.
3. Xu, S., Huang, W., Wang, H., He, Y., Wang, Y., Wang, Y., Qian, J., Xiong, M., and Jin, L. (2007). Dissecting linkage disequilibrium in African-American genomes: Roles of markers and individuals. Mol. Biol. Evol. *24*, 2049–2058.
4. Rosenberg, N.A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J.K., and Feldman, M.W. (2005). Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet. *1*, e70.
5. 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.
6. Xu, S., Jin, W., and Jin, L. (2009). Haplotype-sharing analysis showing Uyghurs are unlikely genetic donors. Mol. Biol. Evol. *26*, 2197–2206.

DOI 10.1016/j.ajhg.2009.10.025. ©2009 by The American Society of Human Genetics. All rights reserved.

Haplotype Background, Repeat Length Evolution, and Huntington's Disease

To the Editor: Warby et al.[1] present fascinating data on the haplotype background of chromosomes carrying the Huntington's disease (HD [MIM 143100]) mutation and the length distribution of the CAG repeat for different haplotypes within the general population. One of their conclusions is that *cis*-elements are likely to represent a major predisposing element in HD expansion. Here, I use evolutionary modeling of the CAG repeat length distribution within populations to argue that the distribution of CAG repeat length and disease incidence in different haplotypes can be explained by founder events, each of which involved expansion of repeats to lengths that are classified as normal by HD investigators (<28 repeats). There is therefore no need to invoke *cis*-element polymorphism within the human population.

Mutation of the HD CAG repeat is both upwardly biased (increases in repeat length are more frequent than decreases) and length dependent (longer repeats mutate more frequently than short ones). Based on sperm typing data, Falush et al.[2] estimated that the mutation rate was proportional to the number of repeats to the power of eight, so that, for example, alleles with 23 copies of the repeat would be approximately 10 times more mutable than alleles with 17 repeats, and alleles with 32 repeats would be approximately 100 times more mutable. The strong length dependence of the mutation rate means that CAG length in itself is a powerful factor in determining the stability of the repeat. Additionally, beyond approximately 55 repeats, the HD mutation causes juvenile HD, which makes further transmission impossible. In fact, the data argue that in modern populations, selection acts strongly against repeat lengths of 44 or more.[2]

I simulated the repeat length distribution in an infinite population based on the mutational model in Falush et al. In order to simulate the effect of natural selection, I removed all repeats of length 50 or more from the population. Simulations show that the assumptions made in modeling selection against disease alleles of different lengths have a negligible effect on the repeat length distribution among normal chromosomes (data not shown). In small populations, e.g., the early settlers of Europe, particular haplotypes can drift to high frequency, also increasing the frequency of the CAG repeat that they carry. In order to investigate the effect of founder events, the population was initially started with three haplotypes each at 1/3 frequency and with initial repeat lengths of 17, 23, and 32 (Figure 1).

A repeat of length 17 has a <0.2% chance of mutating in each generation, so that after 100 generations, most repeats of this length remained unchanged. A repeat of length 32 has a 20% chance within each generation. After 100 generations, most of the repeats of length 32 have mutated at least once, with a majority expanding to length 50 and being removed by natural selection. Consequently,