The Kalash Genetic Isolate? The Evidence for Recent Admixture

To the editor: The recent paper "The Kalash Genetic Isolate: Ancient Divergence, Drift, and Selection," by Ayub et al.[1] suggests that the Kalash people of present-day Pakistan experienced "no detectable gene flow from their geographic neighbors in Pakistan or from other extant Eurasian populations" since their split from those populations over 8,000 years ago. They note that this finding of apparent genetic isolation contradicts the results of Hellenthal et al.,[2] who inferred DNA introgression dated to 910–220 BCE in an overlapping sample of Kalash individuals. Hellenthal et al.[2] inferred the (unknown) source of this DNA to have genetic similarities to a wide range of modern-day groups from West Asia and Europe, including Germany-Austria and Turkey, for example.

Ayub et al. apply methods using two fundamental sources of information, which differ in the types of admixture they are designed to detect. The first source of information, leveraged by ALDER, a method developed in work involving several of us,[3] is unique patterns of linkage disequilibrium generated by admixture. Ayub et al. claim that ALDER "showed no evidence of gene flow into the Kalash."[1] However, this is an error, as the ALDER results reported in their Table S4 show highly significant evidence (p value $< 10^{-10}$) in the Kalash when using Armenia and Chamar as surrogates. Eight other pairings of surrogates give p values $< 10^{-5}$. In all cases, the surrogate pairs include one group from South Asia (Chamar, Kol) and the other from West Eurasia (Armenia, Adygei, Brahui, Hungarians, Palestinians, Tuscans), consistent with admixture from a West Eurasian source.[2] The admixture date point estimates range from 92 to 125 generations ago (with SE < 20), consistent to that inferred in Hellenthal et al. using GLOBETROTTER (95% CI: 76–101 generations).[2] Indeed, the original ALDER paper also found a highly significant signal of mixture in the Kalash (Table 1 of that paper).[3]

The second source of information is based on tests for population mixture, as implemented in the ADMIXTOOLS[4] and TREEMIX[5] software, that model allele frequency correlation patterns among populations but not correlation patterns along the genome. The authors fail to detect a signal using these methods, although previous work has shown that signals of population mixture using such methods can be masked by the effects of genetic drift[3,6] or model mis-specification.[3] In detail, the negative f_3-statistic test implemented in ADMIXTOOLS measures the correlation in allele frequency differences between a sampled group C and two other groups A and B. A significantly negative f_3(C;A,B)—indicating that the frequencies of alleles in population C tend to be intermediate between A and B—can only arise if population C descends from a mixture of populations related (anciently) to A and B.[4] However, as the developers of the negative f_3-statistic test note, "a history of admixture does not always result in a negative f_3(C; A,B) statistic. If population C has experienced a high degree of population-specific drift (perhaps due to founder events after admixture), it can mask the signal so that f_3(C; A,B) might not be negative."[4,6] In other words, allele frequencies in population C will have drifted enough that they will no longer tend to be intermediate between those of A and B. As Ayub et al. showed and has been reported previously, the Kalash experienced strong drift effects—among the highest of Eurasian populations studied to date.[1,3,7–9] Thus, the failure to observe a negative f_3-statistic does not provide meaningful evidence against admixture.

TREEMIX[5] also does not provide evidence of admixture in the Kalash according to the analyses reported in Ayub et al. However, this might be explained by the enormous search space necessary to explain all potential population merges, split times, and migrations among the 30 sampled populations considered in Figure S3 of Ayub et al.[1] In the original TREEMIX paper, those authors speculate that "in graphs with complex structure … several different histories will be compatible with the data."[5] As an example, that paper notes how the well-documented admixture from Neanderthals into non-African populations[10] is missed when applying TREEMIX to data from Neanderthals, Denisovans, and the 53 world-wide populations of the Human Genome Diversity Panel (HGDP).[5]

These observations indicate that, contrary to the claim of Ayub et al. that the ancestors of the Kalash have been isolated from the ancestors of other extant populations for over 8,000 years, there is in fact strong evidence that they have *not* been isolated over this time frame. Hellenthal et al.[2] also inferred a similar signal of ancient admixture from a West Eurasian source into several populations from neighboring regions to the Kalash (the Balochi, Brahui, Makrani, Pathan, Sindhi), suggesting this introgression might be shared as part of a broader signal (though they note the Kalash event appeared to involve a more European-like source relative to the other groups). Whether or not this admixture event involved the armies of Alexander the Great is an unresolved question. One promising direction for future insight is ancient DNA analysis of skeletal remains from northern Pakistan. Further studies of this unique group[11] are important, and we hope that future studies will shed light on the contributing populations.

Garrett Hellenthal, Daniel Falush, Simon Myers,
David Reich, George B.J. Busby, Mark Lipson,
Cristian Capelli, and Nick Patterson

References

1. Ayub, Q., Mezzavilla, M., Pagani, L., Haber, M., Mohyuddin, A., Khaliq, S., Mehdi, S.Q., and Tyler-Smith, C. Am. J. Hum. Genet. *96*, 775–783.
2. Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S. (2014). Science *343*, 747–751.
3. Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Genetics *193*, 1233–1254.
4. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Genetics *192*, 1065–1093.
5. Pickrell, J.K., and Pritchard, J.K. (2012). PLoS Genet. *8*, e1002967.
6. Reich, D., Thangaraj, K., Patterson, N., Price, A.L., and Singh, L. (2009). Nature *461*, 489–494.
7. Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., Zerjal, T., Tyler-Smith, C., and Mehdi, S.Q. (2002). Am. J. Hum. Genet. *70*, 1107–1124.
8. Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A.S., et al. (2004). Am. J. Hum. Genet. *74*, 827–845.
9. 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). Science *319*, 1100–1104.
10. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., et al. (2010). Science *328*, 710–722.
11. Lines, M. (1999) (Peshawar, Pakistan: Emjay Books International).

http://dx.doi.org/10.1016/j.ajhg.2015.12.025.