LETTER

Failure to replicate a genetic signal for sex bias in the steppe migration into central Europe

Iosif Lazaridis[a,b,1] and David Reich[a,b,c,1]

Goldberg et al. (1) used genome-wide ancient DNA data (2) from central European Bronze Age (BA) populations and their three ancestral sources of steppe pastoralists (SP), Anatolian farmers (AF), and European hunter-gatherers (HG) to investigate whether the SP migration into central Europe after 5,000 years ago (3, 4) was sex-biased. By estimating a lower proportion of SP ancestry on the X chromosome (36.6%), which is primarily carried in females, than on the autosomes (61.8%), they suggest that the migration involved a ratio of 5–14 SP males for every female.

We attempted to replicate this finding using qpAdm (3), which leverages allele frequency correlations between the admixed (BA) and source (SP, AF, and HG) populations with distant outgroups to eliminate potential biases due to genetic drift between the true source populations and the ones used as surrogates for them. Our outgroups are Mota (5), Ust_Ishim (6), Kostenki14 (7), GoyetQ116-1 (7), Vestonice16 (7), MA1 (8), AfontovaGora3 (7), and Levantine Neolithic farmers (9). We ran qpAdm with allsnps: YES and Mota as the basis population. The model fits the data ($P = 0.072$ autosomes and $P = 0.747$ chromosome X). For the BA population, we estimate 61.4 ± 2.9% SP, 31.0 ± 1.2% AF, and 7.6 ± 2.9% HG ancestry using all autosomal SNPs and 67.5 ± 17% SP, 26.5 ± 6.9% AF, and 6.0 ± 16.4% HG using all X-chromosome SNPs; thus, we do not find less SP ancestry on the X chromosome.

To diagnose why we do not replicate the finding of ref. 1 we simulated pseudo-BA individuals with known admixture proportions and then used supervised ADMIXTURE (10)—the method used in ref. 1 on the same 4,605 X-chromosome SNPs. We started our simulations using all possible combinations of (20 AF) × (9 HG) × (8 SP) individuals. For each of these 1,440 triples we simulated 16 individuals, randomly sampling alleles from the source individuals according to the qpAdm-estimated proportions of the 16 real BA individuals and removing simulated individuals with fewer than 1,000 SNPs as in ref. 1 (results were little changed with no filtering or a 2,000-SNP threshold). We estimated ancestry for each individual by performing 10 random-seeded supervised ADMIXTURE replicates and averaging the results (we used the remaining 19 AF, 8 HG, and 7 SP individuals as source populations) (Fig. 1*A*). Supervised ADMIXTURE predicts real ancestry poorly in this setting (Fig. 1*B*). The estimation error (estimated – real ancestry) is strongly correlated ($r = 0.91$) with the estimated SP ancestry, allowing us to predict it by a regression (Fig. 1*C*), which indicates upward bias for high SP ancestry estimates and downward bias for low ones. For the pool of 16 individuals that the authors of ref. 1 estimated had 36.6% ancestry the error is predicted to be −19.5%, largely eliminating the discrepancy between the X-chromosome estimate of ref. 1 and the autosomal ancestry estimates of ADMIXTURE (1) and qpAdm. These results show that bias in the estimation of admixture proportions, rather than sex bias in the steppe migration, drives the findings of ref. 1.

Acknowledgments

We thank members of the D.R. laboratory for critical comments and the authors of ref. 1 for sharing data and program output with us. This work was supported by NIH Grant GM100233 and by NSF HOMINID Grant BCS-1032255. D.R. is an investigator of the Howard Hughes Medical Institute.

[a]Department of Genetics, Harvard Medical School, Boston, MA 02115; [b]Broad Institute of MIT and Harvard, Cambridge, MA 02142; and [c]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115

Author contributions: I.L. and D.R. designed research; I.L. and D.R. performed research; I.L. analyzed data; and I.L. and D.R. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: lazaridis@genetics.med.harvard.edu or reich@genetics.med.harvard.edu.

A

Individual	SP	AF	HG
I0059	39±8%	35±3%	26±8%
I1532	71±8%	25±4%	4±8%
I0049	78±9%	19±4%	3±9%
I0099	41±7%	40±3%	19±7%
I0115	48±9%	34±4%	17±9%
I0117	50±8%	35±3%	15±7%
I0803	65±11%	31±5%	4±10%
I0047	59±8%	34±3%	8±8%
I0164	54±8%	33±3%	12±8%
RISE00	75±8%	26±4%	0±8%
RISE150	71±8%	28±3%	1±8%
RISE577	71±8%	38±3%	-9±8%
I0103	84±8%	19±3%	-3±7%
I0104	72±8%	22±3%	6±8%
I0116	71±9%	33±4%	-3±8%
I0118	42±8%	38±3%	20±7%
Mean	62%	31%	7%

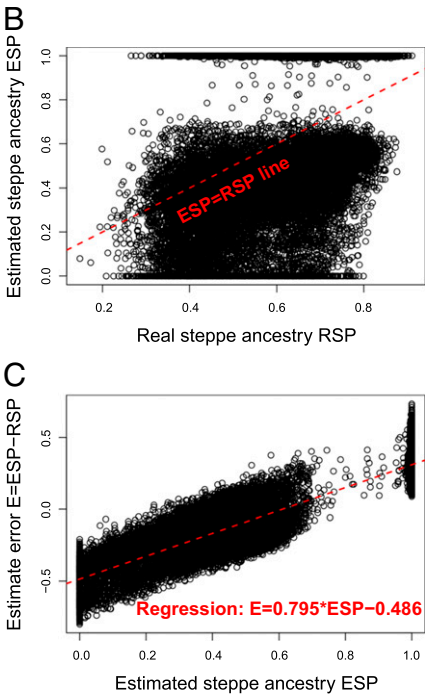

Fig. 1. (*A*) qpAdm estimates of autosomal ancestry of 16 BA individuals. (*B*) Real X-chromosome steppe ancestry (RSP) is not well predicted by supervised ADMIXTURE-estimated X-chromosome steppe ancestry (ESP). (*C*) Error (ESP-RSP) is correlated with ESP, allowing us to infer a 19.5% downward bias at the 36.6% ESP ancestry level.

1 Goldberg A, Günther T, Rosenberg NA, Jakobsson M (2017) Ancient X chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. *Proc Natl Acad Sci USA* 114:2657–2662.
2 Mathieson I, et al. (2015) Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528:499–503.
3 Haak W, et al. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522:207–211.
4 Allentoft ME, et al. (2015) Population genomics of Bronze Age Eurasia. *Nature* 522:167–172.
5 Gallego Llorente M, et al. (2015) Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* 350:820–822.
6 Fu Q, et al. (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514:445–449.
7 Fu Q, et al. (2016) The genetic history of Ice Age Europe. *Nature* 534:200–205.
8 Raghavan M, et al. (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505:87–91.
9 Lazaridis I, et al. (2016) Genomic insights into the origin of farming in the ancient Near East. *Nature* 536:419–424.
10 Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12:246.