

REPLY TO LAZARIDIS AND REICH:

Robust model-based inference of male-biased admixture during Bronze Age migration from the Pontic-Caspian Steppe

Amy Goldberg^a, Torsten Günther^b, Noah A. Rosenberg^a, and Mattias Jakobsson^{b,c,1}

By comparing the sex-specifically inherited X chromosome to the autosomes in ancient genetic samples, we (1) studied sex-specific admixture for two prehistoric migrations. For each migration we used several admixture estimation procedures—including ADMIXTURE model-based clustering (2)—to compare X-chromosomal and autosomal ancestry in contemporaneous Central Europeans and we interpreted greater admixture from the migrating population on the autosomes as male-biased migration. For migration into late Neolithic/Bronze Age Central Europeans (BA) from the Pontic-Caspian Steppe (SP) we inferred male-biased admixture at 5–14 males per migrating female.

Lazaridis and Reich (3) contest this male-biased migration claim. For simulated individuals, they claim that ADMIXTURE provides biased X-chromosomal ancestry estimates. They argue that if the bias is taken into account, then X-chromosomal steppe ancestry is similar to our autosomal ancestry estimate, and that hence steppe male and female contributions are similar.

Many factors affect ancestry inferences from ADMIXTURE and related programs (2, 4–8). To understand ADMIXTURE inferences for X-chromosomal ancient DNA, we performed simulations examining the effects of multiple variables. First, we used “reference” individuals in ref. 1 to simulate analogs of the BA population.

Fig. 1 plots estimated X-chromosomal ancestry for simulated BA individuals (Fig. 1 A and B), showing that for high true ancestry levels ADMIXTURE overestimates steppe ancestry, whereas for low levels it underestimates it. For the intermediate ancestry in ref. 1 (0.366), however, ADMIXTURE is accurate, and our estimate is robust to bias.

Because our interest in ref. 1 was the X/autosomal comparison, we next simulated autosomes, finding bias similar to that of the X chromosome (Fig. 1 C and D). Bias-corrected X/autosomal ancestry estimates translate in a constant-admixture model (1) to four to seven migrating steppe males per female. Thus, accounting for

ADMIXTURE bias, substantial male excess during the steppe migration remains supported.

We next tested whether specific data features—haploid ancient genotypes, high missing-data rates, and small reference samples—might underlie previously unseen ADMIXTURE biases. We performed analogous simulations using modern HapMap samples without these features. This analysis traces the bias to the small reference samples available in haploid ancient data (1) (Fig. 2).

The greater bias in ADMIXTURE in ref. 3 than here thus likely arises from two sources. First, ancestry values underlying the simulation in ref. 3 trend toward parameter values that generate higher bias than with our even spacing. Second, ADMIXTURE inference in ref. 3 discards one individual per source population, potentially enlarging bias from small reference samples.

We note that the authors of ref. 3 also consider a second program, qpAdm (9); their wide confidence intervals for this summary-statistic method relying on f_4 calculations permit multiple interpretations (male bias, female bias, or no bias). Direct f_4 calculation (10), however, trends toward male-biased migration: BA share more alleles with SP than with early Neolithic Central Europeans (CE) on autosomes [$f_4(\text{chimp}, \text{BA}; \text{CE}, \text{SP}) = 0.0014$; $Z = 6.78$, $P < 0.0001$], but have more CE X-chromosomal sharing ($f_4 = -0.0068$; $Z = -0.561$).

We conclude that our inference of male-biased Pontic-Caspian Steppe migration, seen using ADMIXTURE, STRUCTURE, mechanistic simulations, and X/autosomal F_{ST} , is robust. Our analysis further illuminates the impact of small haploid reference samples on ADMIXTURE; we look forward to refining sex-specific migration estimates as larger, higher-coverage ancient samples become available.

Acknowledgments

We thank Jonathan Kang and Joshua Schraiber for helpful discussions.

^aDepartment of Biology, Stanford University, Stanford, CA 94305; ^bDepartment of Organismal Biology, Uppsala University, 75236 Uppsala, Sweden; and ^cScience for Life Laboratory, Uppsala University, 75123 Uppsala, Sweden

Author contributions: A.G., T.G., N.A.R., and M.J. designed research; A.G., T.G., N.A.R., and M.J. performed research; A.G. and T.G. analyzed data; and A.G., T.G., N.A.R., and M.J. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence should be addressed. Email: mattias.jakobsson@ebc.uu.se.

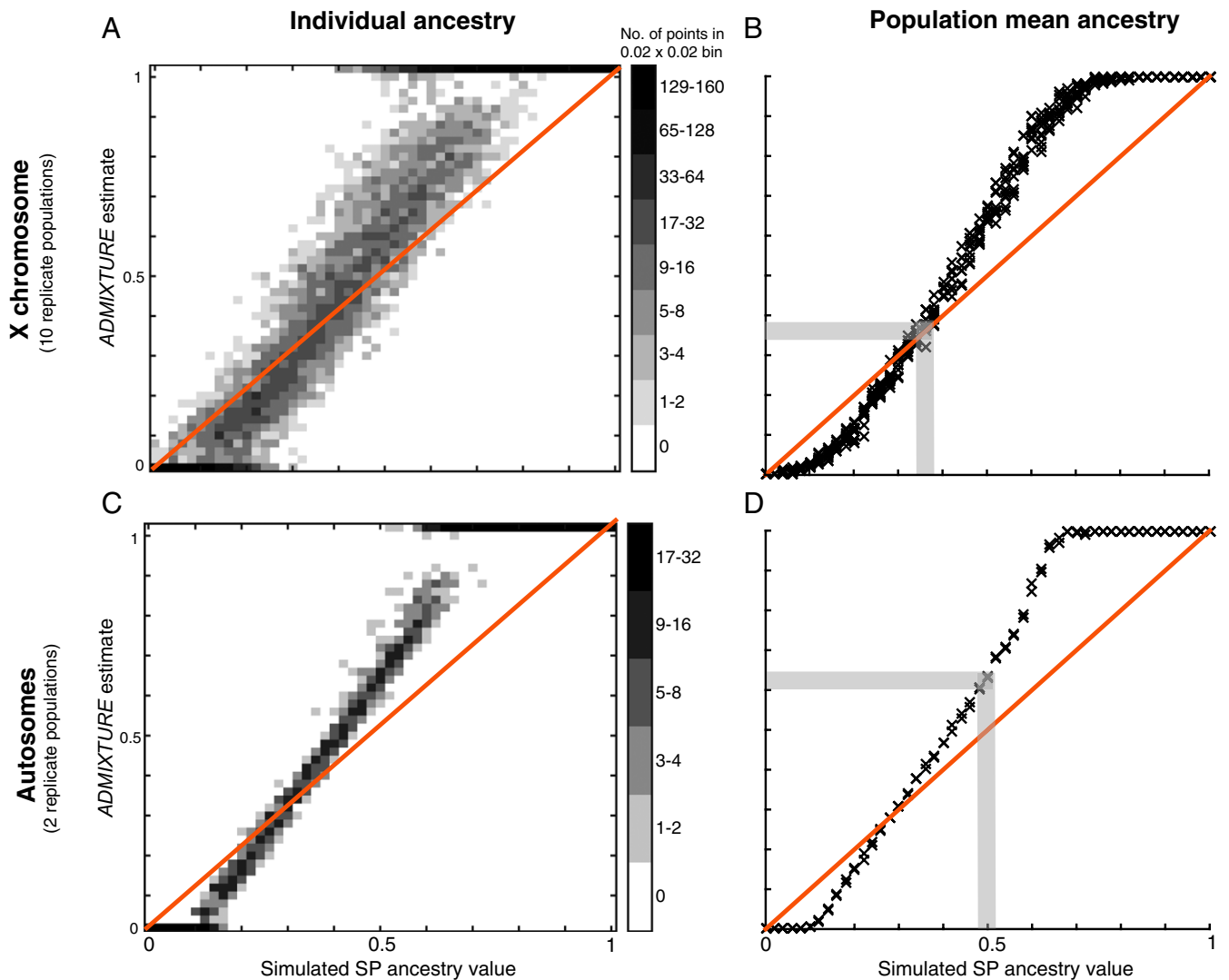


Fig. 1. ADMIXTURE inference in simulated ancient genomes. We simulated admixed BA individuals for steppe-related ancestry values in 0.02 increments fixing the ratio of AF (Anatolian farmers) to HG (European hunter-gatherers) ancestry as 0.85 to 0.15. Using linkage disequilibrium-pruned SNP sets from ref. 1, independently at each SNP, we drew a reference population and chose the allele randomly from among individuals in that population (including missing data). We used a 10-seed average with supervised ADMIXTURE (1), considering replicate 16-individual populations for each SP ancestry value (A and C). Shaded bars (B and D) show the range of simulated ancestries that corresponds to estimated ancestries in ref. 1: [0.34, 0.38] for the X, corresponding to the same range in simulated values, and [0.60, 0.64] for autosomes, corresponding to ~ 0.500 . The updated autosomal ancestry value generates X/autosomal ancestry ratio $0.366/0.500 = 0.732$, compared with 0.592 in ref. 1; this ratio generates an inference of four to seven migrating males per female by the mechanistic model of ref. 1.

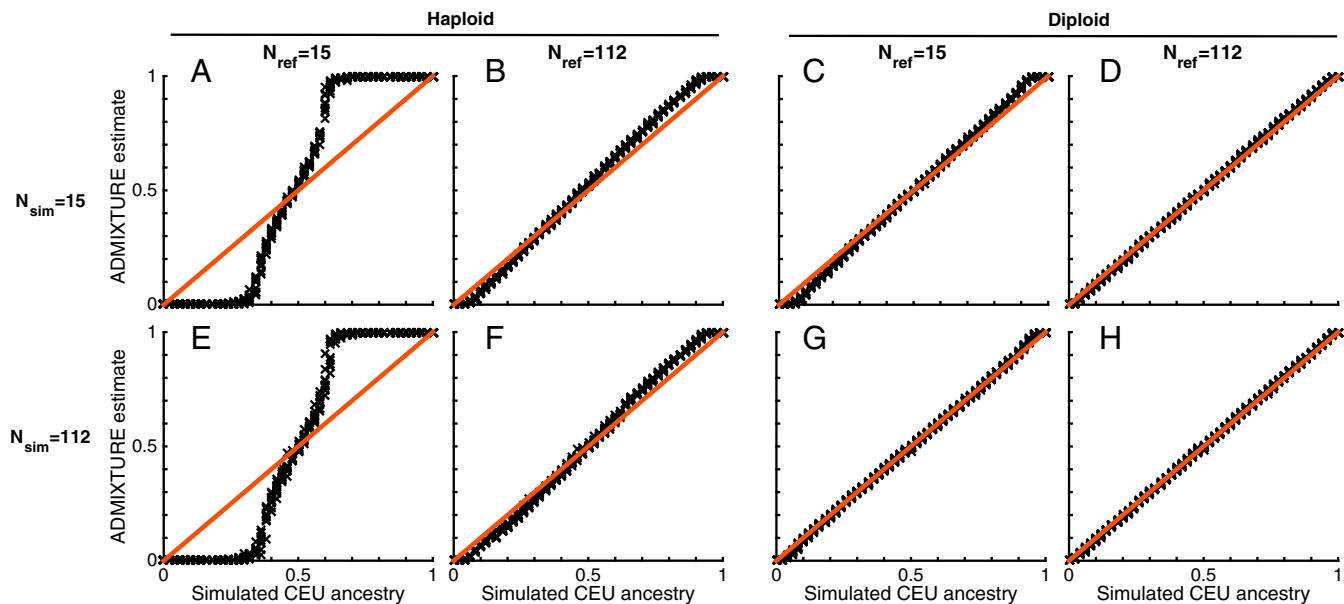


Fig. 2. ADMIXTURE sample-size effects. Following the procedure in Fig. 1, we estimated ancestry in simulated admixed haploid and diploid genomes. Considering 4,605 randomly sampled autosomal SNPs, we varied the HapMap YRI and CEU sample sizes used per source population when simulating genomes (N_{sim}) and as reference samples in ADMIXTURE (N_{ref}). With N_{ref} comparable to ref. 1, ADMIXTURE bias in haploid simulations matches that observed for ancient data (A). Therefore, the bias is likely not due to missingness. As N_{ref} increases (B and F), bias decreases. Using differing N_{sim} and N_{ref} , we attribute the bias to small reference samples during inference rather than simulation (B vs. E). Diploid simulations have minimal bias (C, D, G, and H), suggesting that haploidy of ancient data combines with small reference sizes to generate the observed bias in ADMIXTURE.

- 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 Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664.
- 3 Lazaridis I, Reich D (2017) Failure to replicate a genetic signal for sex bias in the steppe migration into central Europe. *Proc Natl Acad Sci USA* 114:E3873–E3874.
- 4 Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.
- 5 Rosenberg NA, et al. (2005) Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet* 1:e70.
- 6 Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7:574–578.
- 7 Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12:246.
- 8 Allentoft ME, et al. (2015) Population genomics of Bronze Age Eurasia. *Nature* 522:167–172.
- 9 Haak W, et al. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522:207–211.
- 10 Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192:1065–1093.