TEXT S8   ADMIXTURE linked simulations analysis

For the linked simulations we have compared our results with the program AD-MIXTURE (Alexander, Novembre, and Lange 2009). ADMIXTURE computes the same likelihood as STRUCTURE but performs maximum-likelihood analysis, i.e. it does not perform MCMC sampling and does not apply a prior. This makes is significantly faster and avoids many mixing problems, and is easily applicable to the HGDP dataset. To perform this analysis, we took the same phased haplotype data used as input for ChromoPainter and converted it to PLINK format (Purcell, Neale, Todd-Brown, Thomas, Ferreira, Bender, Maller, Sklar, de Bakker, Daly, and Sham 2007) (PLINK version 1.07, downloaded from `http://pngu.mgh.harvard.edu/~purcell/plink/`) using which we extracted the SNPs with minor frequency > 0.01. The filesizes were too large for manipulation within PLINK with 200 regions and therefore we used a minor frequency cutoff of 0.02 in this case. We then ran ADMIXTURE for various numbers of populations K and for varying number of regions.

We here show the details of the ADMIXTURE results since direct comparison between the methods is not possible, fineSTRUCTURE being an MCMC based no-admixture model and ADMIXTURE reporting only maximum-likelihood admixture results. The correlation reported in the paper is created by forcing the admixture solution to choose the most likely population for each individual; however, performing the correlation on the admixed solution does not change the results qualitively.

References

ALEXANDER, D. H., J. NOVEMBRE, and K. LANGE, 2009 Fast model-based estimation of ancestry in unrelated individuals. Genome Research **19:** 1655–1664.

PURCELL, S., B. NEALE, K. TODD-BROWN, L. THOMAS, M. FERREIRA, D. BENDER, J. MALLER, P. SKLAR, P. DE BAKKER, M. DALY, and P. SHAM, 2007 PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics **81:** 559–75.