

# Comment

---

Supplementary information to:

## Counter the weaponization of genetics research by extremists

A Comment published in *Nature* **610**, 444–447 (2022)

<https://doi.org/10.1038/d41586-022-03252-z>

---

Jedidiah Carlson, Brenna M. Henn, Dana R. Al-Hindi & Sohini Ramachandran

---

This Supplementary information comprises:

1. Methodology for the graphic 'Racists use geneticists' infographics'

## Supplementary information

### Methodology for the graphic ‘Racists use geneticists’ infographics’

In the original Li et al. analysis [Li et al. 2008], the authors infer ancestry or “admixture” proportions for each individual in their dataset using genome-wide genotypes. Their figure was generated using 938 samples from the Human Genome Diversity Project and 642,690 genetic variants ascertained in a few global populations. The number of ancestries is set by the user-defined parameter “K”; each K color represents the estimated ancestral allele frequencies.

In our new “African-centric” figure (panels A and C), we use a similar approach but with a different breakdown of African (21) and non-African (5) population samples. We included all populations from the African Genome Variation Project (AGVP) [Gurdasani et al. 2015] with more than 20 individuals, along with Nama from South Africa [van Eeden et al. 2022]. The Ugandan population from AGVP was down-sampled to random set of 100 individuals. Samples from the AGVP were collected by local researchers across several countries within Africa. We selected five populations from the 1000 Genomes Project [The 1000 Genomes Project Consortium 2015] from Europe, Asia and North America. Low-coverage genomes were multi-sample called together to prevent batch effects. Our analysis included 1918 individual genomes, of which 1450 were of African-descent, and 9,189,395 genetic variants after extracting chromosomes 1 through 10 and applying LD pruning and minor allele frequency filters.

In our figure, we mimic Li and colleagues’ (2008) clustering of the data set into K=7 ancestries. Here, we used the program ADMIXTURE [Alexander 2009] to estimate individuals’ ancestry proportions and partition their genomes into distinct ancestries, which correspond to a unique color in the graph. The *frappe* algorithm in Li et al. [see Tang 2005] and ADMIXTURE are conceptually very similar, differing primarily in the form of optimization used to analyze thousands of genetic markers.

Funding for the AGVP sample collections was in part by the Wellcome Trust (grant number WT077383/Z/05/Z), The Wellcome Trust Sanger Institute (grant number WT098051), the Bill and Melinda Gates Foundation, the Foundation for the National Institutes of Health (grant number 566), and the UK Medical Research Council (grant numbers G0901213-92157, G0801566, G0600718 and MR/K013491/1).

### Supplemental references

Li, J. Z. et al., 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319(5866): pp.1100-1104.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A. and Ritchie, G.R., 2015. The African genome variation project shapes medical genetics in Africa. *Nature* 517(7534), pp.327-332.

van Eeden, G., Uren, C., Pless, E., Mastoras, M., van der Spuy, G.D., Tromp, G., Henn, B.M. and Möller, M., 2022. The recombination landscape of the Khoe-San likely represents the upper limits of recombination divergence in humans. *Genome Biology* 23(1), pp.1-14.

The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. *Nature* 526, pp.68–74.

Alexander, D.H., Novembre, J. and Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19(9), pp.1655-1664.

Tang, H., Peng, J., Wang, P. and Risch, N.J., 2005. Estimation of individual admixture: analytical and study design considerations. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society* 28(4), pp.289-301.