An article template

The Author[1,*]

[1]Planet Earth
[*]email@address.edu

February 2, 2023

Abstract

A very simple template for an article class document.

1 Interpretation of results

1. Human generation time differences between populations are claimed to extend to more than 250,000 years. Extending the analysis beyond 10,000 generations shows that generation times do not converge until "78,000" generations ago (more than 2 million years). This is implausible.

2. Eurasian–West African divergence has been estimated to be around (e.g., Pagani et al., 2015)

3. Even papers that have inferred deep divergences of human populations in Africa would place the West Africa–Eurasian divergence around 100–150ka (e.g., Schlebusch et al., 2017).

4. It's worth noting that multiple studies have found evidence for deep population structure in the history of Bantu-speaking groups. However, it's unclear if such structure could account for such long-lasting differences in generation times, or if that structure is mis-inferred (as we argue in Ragsdale et al. (2022)).

2 The mutation spectrum over time

Issues:

1. GEVA-Atlas variant ages result in mutation spectra that change rapidly beyond 10,000 generates (which is likely why the weren't shown?)

2. Allele age estimates between three state-of-the-art methods [`GEVA` (Albers and McVean, 2020), Relate (Speidel et al., 2019), and `tsdate` (Wohns et al., 2022)] are only moderately correlated (for example, see Figure S20 in the Supplement of Wohns et al. (2022)).

3. Allele ages provided by each method results in distinct and unalike mutation spectrum histories (Figures 1–4).

4. In turn, these divergent histories provide estimates of generation time profiles that differ completely.

3 De novo mutations from trio data

1. *De novo* mutations from the Icelandic trio study have a different mutation profile than the most recent age bins (across all three of the variant dating methods).

2. In the most recent age bins (0 to 50 or 80 generations ago), the three methods estimate very similar spectra of young standing variation. They are closer to each other than any of them are to the spectrum from Jónsson et al. (2017).

3. Wang et al. (2023) acknowledge this:

   > We found that the mutation spectrum from the large pedigree study (14) consistently differed from the variant spectrum inferred from the 1000 Genomes Project data, possibly because we removed singletons from the polymorphism dataset to reduce errors."

4. However, they did not test (or at least did not show a test of) this hypothesis that the removal of singletons drives this signal. I don't think it does, but want to dive into it a bit more thoroughly. Comment: Update: GEVA does not provide ages for singletons. Relate and tsdate do, however, so we can test this hypothesis there.

5. Table 1 shows mutation spectrum from the most recent bin in each dataset.

6. I also don't think the approach they took is satisfying:

   > Therefore, to obtain absolute generation times for historical periods, we centered the observed spectra on the most recent bin, subtracting its difference with the average mutation spectrum estimated in (14) from each historical spectrum. This has the effect of assuming that parental ages in the pedigreed mutation dataset reflect generation times in the most recent historical bin.

   And I don't know what biases this introduces. It does have the effect of forcing recent bins to have roughly the same inferred average generation times for mothers and fathers as the iceland trio data (28.2 and 32, resp.). It's therefore not a *result* that recent time periods match other estimates. It's a built-in assumption of their model

4 Bigger picture issues and commentary

1. Allele age estimates are noisy, and probably shouldn't be used for such detailed inferences. You'll end up fitting the noise and bias of each method.

2. DNM estimates from trios have their own sets of problems. Do we know where the discrepancy between trio-estimated DNM spectrum and observations from pop-gen data come from? Probably needs to be sorted out.

5 Figures

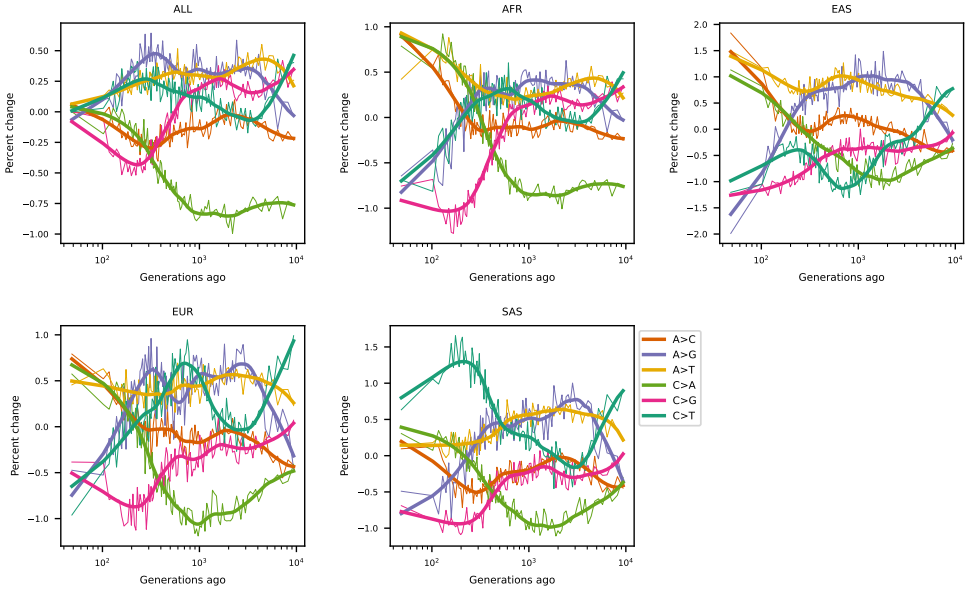

Figure 1: **`GEVA`-inferred mutation spectrum history.**

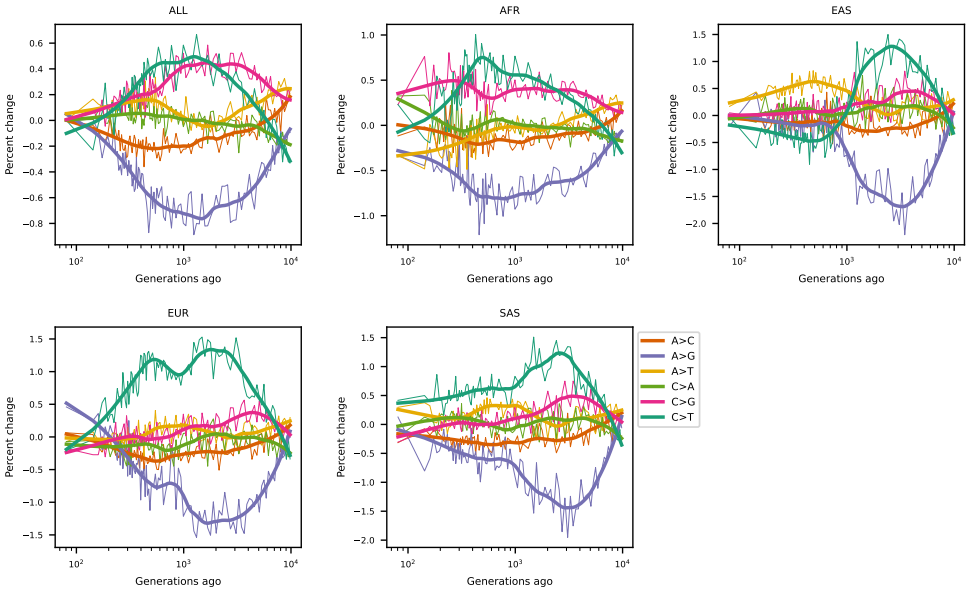

Figure 2: **`Relate`-inferred mutation spectrum history.**

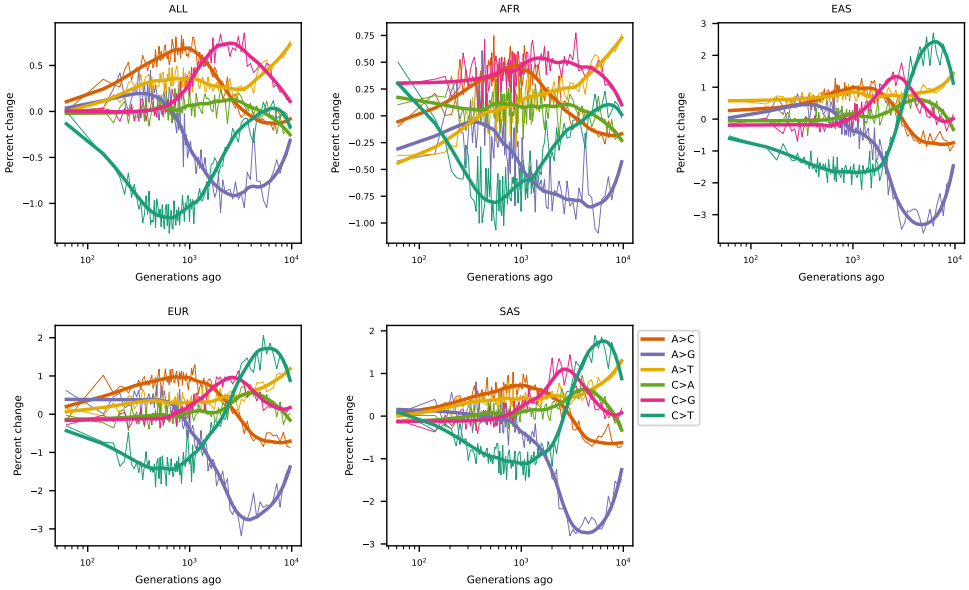

Figure 3: `tsdate`**-inferred mutation spectrum history.**

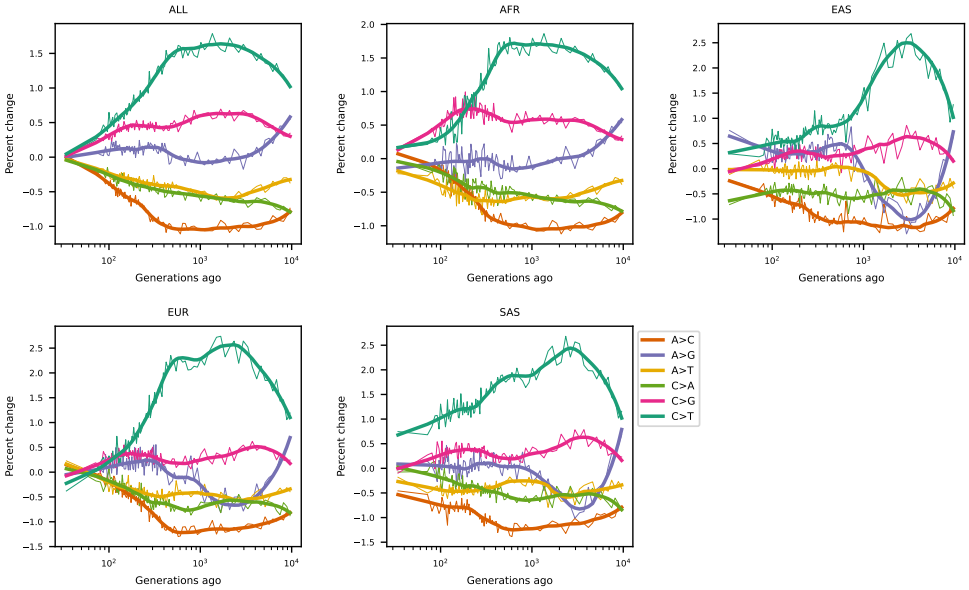

Figure 4: `Relate`**-inferred mutation spectrum history, including singletons.**

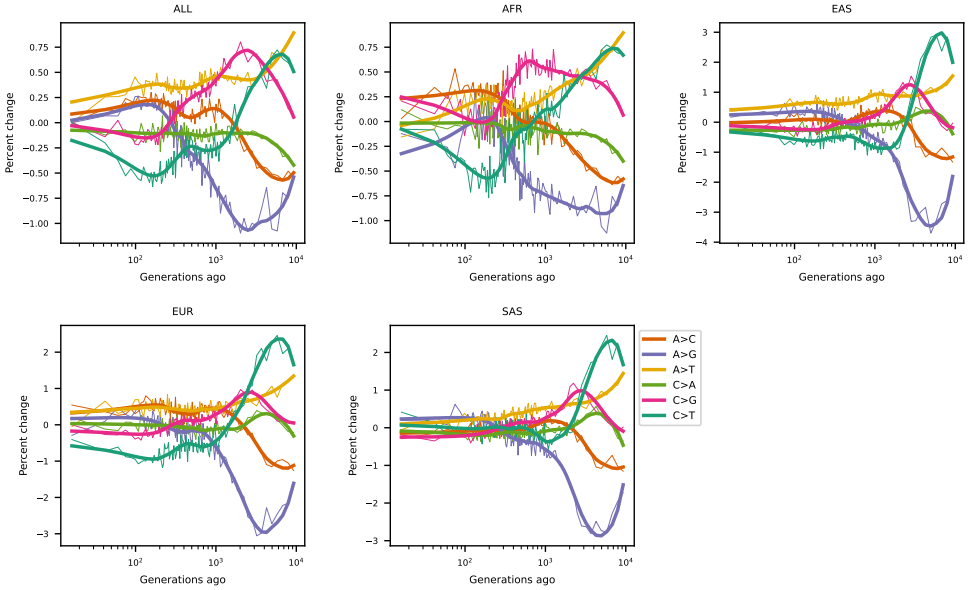

Figure 5: **`tsdate`-inferred mutation spectrum history, including singletons.**

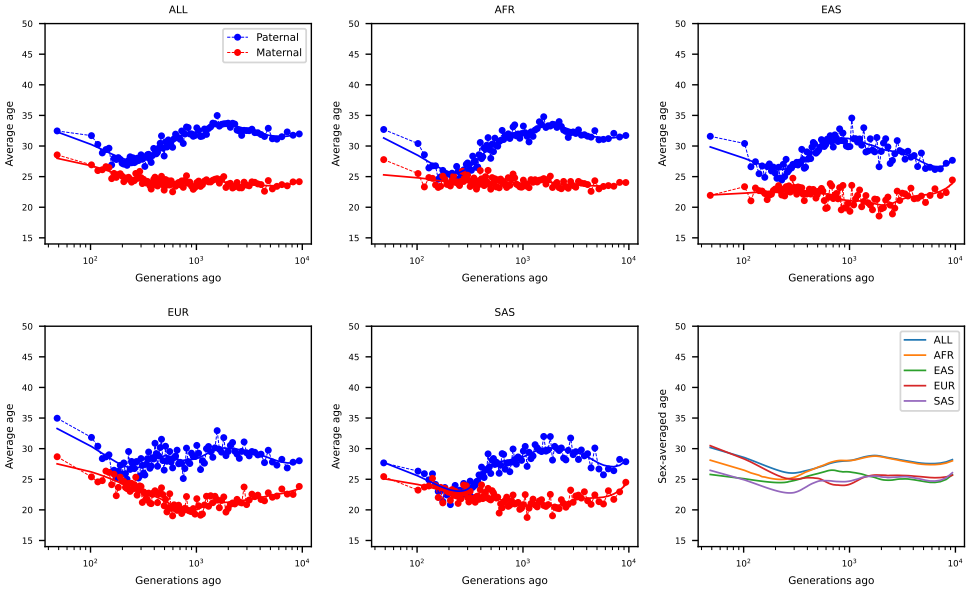

Figure 6: **`GEVA`-inferred generation time histories.**

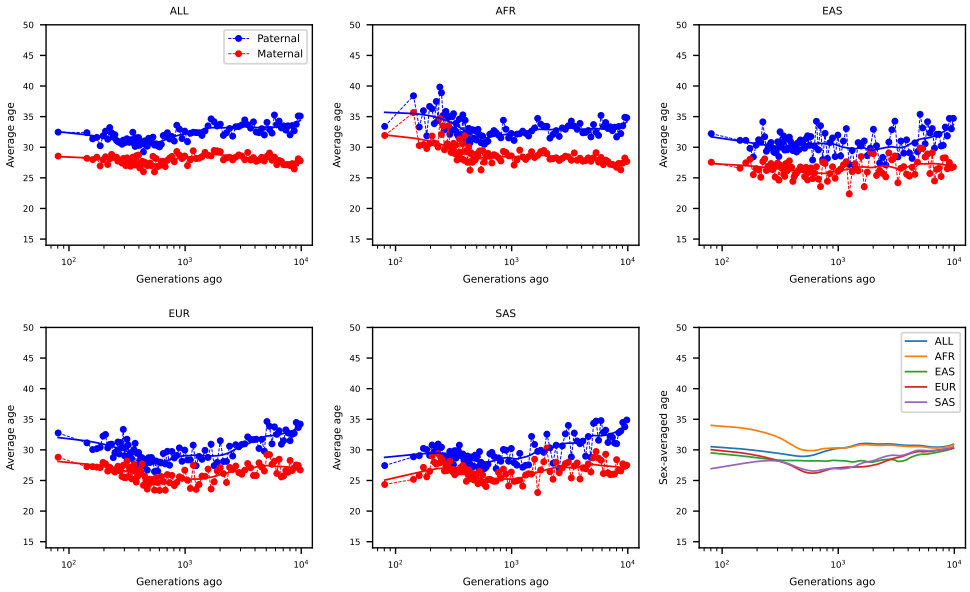

Figure 7: `Relate`**-inferred generation time histories.**

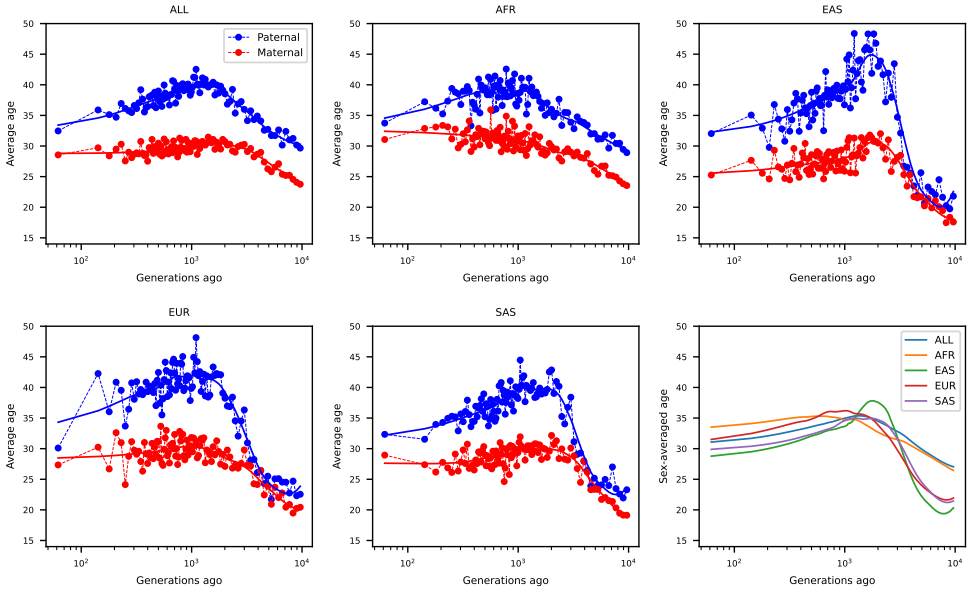

Figure 8: `tsdate`**-inferred generation time histories.**

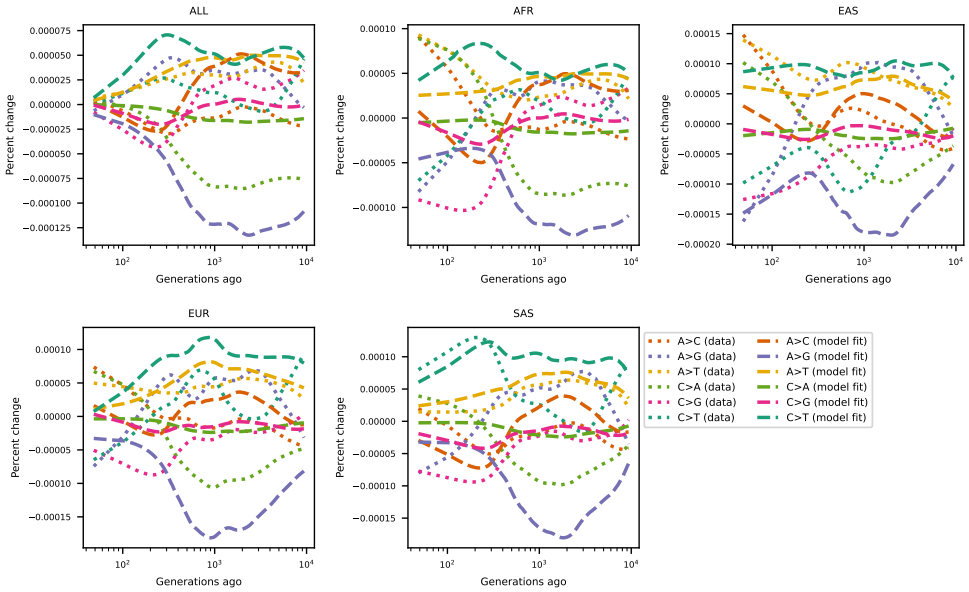

Figure 9: **Prediction of mutation spectrum history from `GEVA`-inferred generation times.**

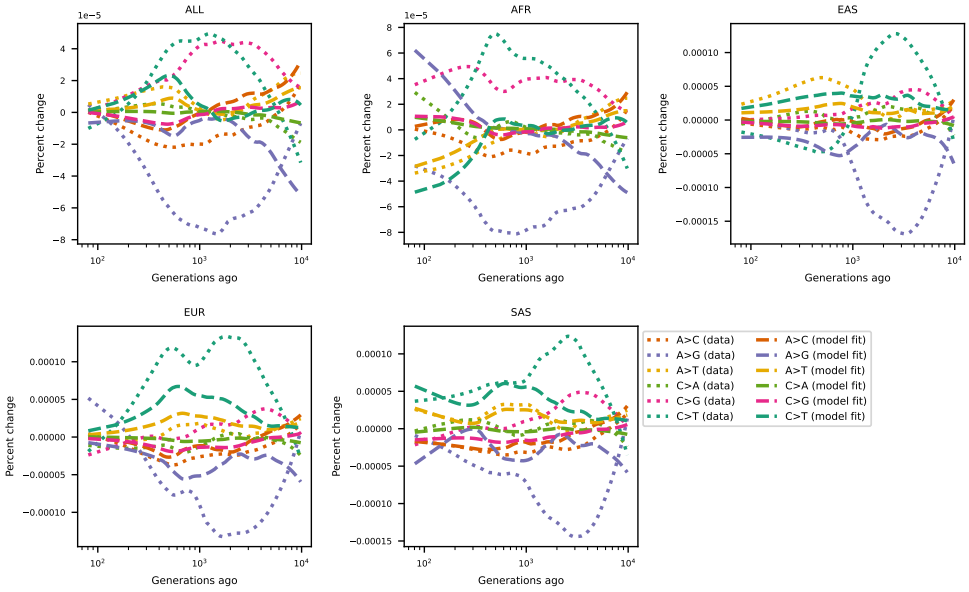

Figure 10: **Prediction of mutation spectrum history from `Relate`-inferred generation times.**

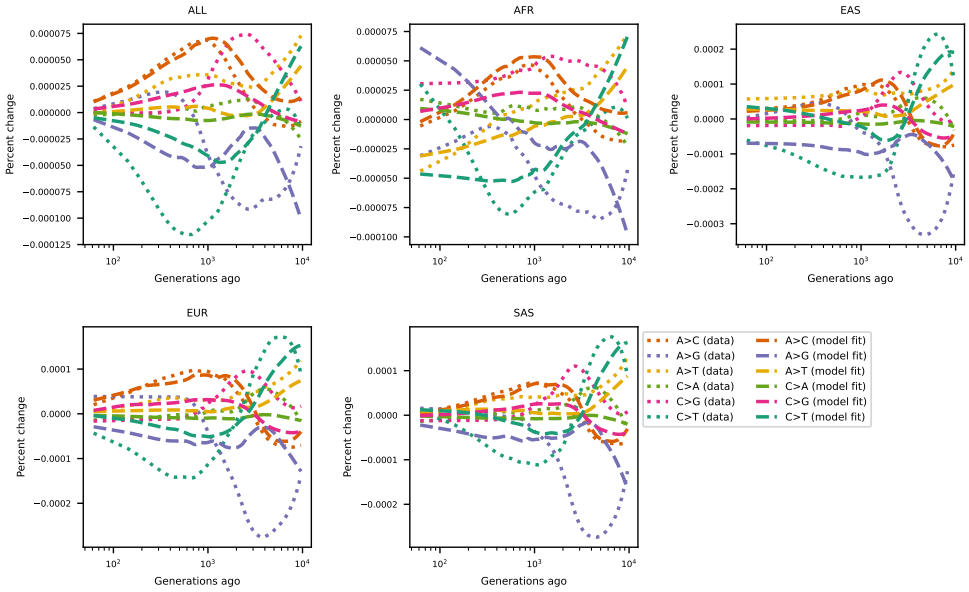

Figure 11: **Prediction of mutation spectrum history from `tsdate`-inferred generation times.**

6 Tables

Table 1: **Mutation profiles from the past 100 generations, compared to Iceland trios.** The most recent time bin for each method included the past ≈ 150 generations. When singletons were included (when using data from `tsdate` and `Relate`), the spectra of estimated recent standing variation were largely similar. Note that `GEVA` does not report ages for singletons. While the three methods provide similar spectra from recent mutations, the spectrum from the Iceland trios differs substantially. The similarity in the spectrum between mutations that were phased in Jónsson et al. (2017) and all mutations (phased and unphased) are very similar to each other.

Dataset	A→C	A→G	A→T	C→A	C→G	C→T
`GEVA`	0.0946	0.3600	0.0886	0.1201	0.1057	0.2310
`tsdate`	0.0931	0.3579	0.0899	0.1146	0.1061	0.2384
`tsdate` (w/singletons)	0.0989	0.3598	0.0908	0.1168	0.1062	0.2275
`Relate`	0.0991	0.3610	0.0863	0.1124	0.1038	0.2374
`Relate` (w/singletons)	0.1002	0.3590	0.0921	0.1164	0.1060	0.2263
Trios (phased)	0.1071	0.4100	0.0881	0.0481	0.0847	0.2620
Trios (all mutations)	0.1080	0.4086	0.0903	0.0483	0.0839	0.2610

References

Patrick K Albers and Gil McVean. Dating genomic variants and shared ancestry in population-scale sequencing data. *PLoS biology*, 18(1):e3000586, 2020.

Hákon Jónsson, Patrick Sulem, Birte Kehr, Snaedis Kristmundsdottir, Florian Zink, Eirikur Hjartarson, Marteinn T Hardarson, Kristjan E Hjorleifsson, Hannes P Eggertsson, Sigurjon Axel Gudjonsson, et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature*, 549(7673): 519–522, 2017.

Luca Pagani, Stephan Schiffels, Deepti Gurdasani, Petr Danecek, Aylwyn Scally, Yuan Chen, Yali Xue, Marc Haber, Rosemary Ekong, Tamiru Oljira, et al. Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *The American Journal of Human Genetics*, 96(6):986–991, 2015.

Aaron P Ragsdale, Timothy D Weaver, Elizabeth G Atkinson, Eileen Hoal, Marlo Möller, Brenna M Henn, and Simon Gravel. A weakly structured stem for human origins in africa. *bioRxiv*, pages 2022–03, 2022.

Carina M Schlebusch, Helena Malmström, Torsten Günther, Per Sjödin, Alexandra Coutinho, Hanna Edlund, Arielle R Munters, Mário Vicente, Maryna Steyn, Himla Soodyall, et al. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science*, 358(6363):652–655, 2017.

Leo Speidel, Marie Forest, Sinan Shi, and Simon R Myers. A method for genome-wide genealogy estimation for thousands of samples. *Nature genetics*, 51(9):1321–1329, 2019.

Richard J Wang, Samer I Al-Saffar, Jeffrey Rogers, and Matthew W Hahn. Human generation times across the past 250,000 years. *Science Advances*, 9(1):eabm7047, 2023.

Anthony Wilder Wohns, Yan Wong, Ben Jeffery, Ali Akbari, Swapan Mallick, Ron Pinhasi, Nick Patterson, David Reich, Jerome Kelleher, and Gil McVean. A unified genealogy of modern and ancient genomes. *Science*, 375(6583):eabi8264, 2022.