On the fraught inference of historical human generation times

Aaron P. Ragsdale[1,*] and Kevin Thornton[2]

[1]University of Wisconsin–Madison, Wisconsin, USA
[2]University of California, Irvine, California, USA
[*]email@address.edu

February 5, 2023

Abstract

Wang *et al.* (2023) recently proposed an approach to infer the history of human generation times from changes in mutation profiles over time. With the observation that the relative proportions of different mutation types depend on the ages of parents, stratifying variants by the age of mutation allows for the inference of average paternal and maternal generation times at past times. Applying this approach to published allele age estimates, Wang *et al.* inferred long-lasting sex differences in average generation times and surprisingly found that ancestral generation times of West African populations remained substantially higher than those of Eurasian populations extending tens of thousands of generations into the past. Here we show that the results and interpretations in Wang *et al.* (2023) are driven by noise and biases in input data, modeling and statistical choices, and a lack of validation using independent approaches for estimating allele ages. With the recent development of methods to reconstruct genome-wide gene genealogies, coalescence times, and allele ages, we caution that ...

Recent years have seen the rapid development of methods for reconstructing genealogical structures of large cohorts (Speidel *et al.*, 2019; Wohns *et al.*, 2022; Hubisz *et al.*, 2020), which are comprised of a series of gene genealogies (or trees) along the genome. Reconstructed genealogies (or informative summaries of them (Albers and McVean, 2020)) have the potential to transform evolutionary and population genetic inference, as biological and evolutionary processes impact the shape and correlation of gene trees and the distribution of variation that arises in the lineages they represent. Relevant to this study, the age of a variant can be estimated by mapping its mutation to the portion of the gene tree in which it is inferred to have occurred.

At the same time, the last few years have seen large sequencing efforts of families that provide ever-increasing resolution of genome biology, including direct measurements of mutation rates and profiles. Through high-coverage sequencing of multiple generations within known pedigrees, *de novo* mutations can be determined as maternally or paternally inherited, and with a large enough sample size both the number of mutations and proportions of mutation types (e.g., A→C, A→G, etc.) can be correlated with parental age and sex (Jónsson *et al.*, 2017; Halldorsson *et al.*, 2019). Wang *et al.* (2023) combined these two sets of inferences, the estimated ages of mutations and the parental age- and sex-dependence of the mutation spectrum, to infer the history of average maternal and paternal generation intervals for human populations of diverse ancestries. In order to avoid overfitting, this approach requires making a number of assumptions about the constancy of the mutational process over time **??**, its similarity across populations, APR: and others to point out? For now, let's accept these as stated...

This approach has recently been criticized (Gao *et al.*, 2022). Notably, Gao *et al.* show that observed changes in the mutation spectrum over time cannot be explained by changes in maternal and paternal generation times alone, as specific mutational signatures would require unique and divergent generation intervals to simultaneously explain them. Gao *et al.* also point out that pedigree-based estimates of the *de novo* mutation spectrum do not agree with the mutation spectrum among young variants in existing

population-level datasets, which we also discuss below. They argue instead that factors other than changes in generation intervals, including genetic modifiers and environmental exposure, must explain observed variation in mutation profiles.

In this note, we

1. Argue that the inferred generation time histories are inconsistent with the current understanding of human population history, in particular deep history within Africa.
2. Show that allele age estimates are not just noisy, but age-stratified mutation spectra reconstructed using independent methods disagree wildly, with mutation profiles diverging in opposing directions. Thus, the results from Wang *et al.* (2023) do not reproduce.
3. Discuss the problems with using mutation rates and profiles from pedigree-based studies to calibrate population genetic inferences.

In conclusion, we suggest that downstream analyses using estimated allele ages and mutation profiles should more carefully validate their results, which in turn should be interpreted with a heavy dose of skepticism APR: or something like that.

Long-lasting differences in population-specific generation intervals

In applying this inference approach to multiple populations of different continental ancestries, Wang *et al.* (2023) estimated that the ancestors of European, East Asian, and South Asian populations included in the 1000 Genomes Project Consortium *et al.* (2015) dataset (1KGP) have a history of significantly reduced average generation times compared to West African populations. These differences extend to over 10,000 generations, the time period highlighted in this study. In discussing this result the authors state, "the difference among populations beyond 2000 generations ago reflects population structure in humans before their dispersal out of Africa, a structure that is not fully captured by the 1000 Genomes AFR sample. This implies that the simple labels of 'African' and 'non-African' for these populations conceal differences in generation times that existed on our ancestral continent." Indeed, a number of recent genetic studies suggest that human population structure within Africa extending hundreds of thousands of years into the past has in part shaped modern-day genetic variation (Plagnol and Wall, 2006; Hammer *et al.*, 2011; Hsieh *et al.*, 2016; Hey *et al.*, 2018; Ragsdale and Gravel, 2019; Durvasula and Sankararaman, 2020; Lorente-Galdos *et al.*, 2019).

However, in extending their analysis farther into the past, Wang *et al.* find that ancestral generation intervals do not converge until many 10s of thousands of generations age. With an average generation time of 25–30 years, this corresponds to well over one million years ago. This observation would require some portion of the ancestries of Eurasian and West African populations to have remained isolated for many hundreds of thousands of years, for those structured ancestral populations to have had large differences in average generation times over the course of this history, and for those groups to have contributed substantively to different contemporary human populations. While such a scenario of very long-lasting isolation among ancestral popualtions is not impossible, it is not supported by genetic (Ragsdale *et al.*, 2022; **?**) or archaeological (Scerri *et al.*, 2018; **?**) evidence, which rather suggest at least periodic connectivity of ancestral human populations within Africa.

Rather than long-lasting isolation between a large portion of ancestry of West African and Eurasian populations (that represented large differences in generation intervals), Eurasian–West African divergence has been estimated at only $\approx$ 75ka (thousand years ago) (e.g., Pagani *et al.*, 2015; **?**). Even studies that have inferred deeper divergences of human populations within African would place the Eurasian–West African divergence at around 100–150ka (Schlebusch *et al.*, 2017).

APR: If these estimated divergences represents the majority of ancestry of the two populations, can we do a back of the envelope calculation to see what difference in generation time would be needed between

the hypothesized minority of deeply structured ancestry contributing to these pops in order to explain the estimated average differences in generation times? I bet it would be huge.

Given the implausibility of the stated result, it is natural to ask what may be causing such mis-inferences. Below we show that multiple sources of uncertainty, namely noise and bias in allele age inference and incosistencies in trio-base estimates of mutation profiles, confound inferences of generation times from time series of mutation spectra.

Inconsistencies in inferred mutation spectra over time

Central to the inference of generation intervals from time-stratified mutation spectra is the dating of variant ages. Wang *et al.* (2023) used published allele ages from Albers and McVean (2020) (using the software `GEVA`). `GEVA` estimates allele ages by considering the number of additional mutations that have accumulated on the ancestral haplotype carrying the focal variant, as well as the effect of recombination in reducing the size of that ancestral haplotype. Singletons are excluded from analysis by `GEVA` and are not assigned an age. Partitioning variants by their estimated ages shows that the mutation spectrum (i.e., the distribution of six mutation types) has changed over time, assuming that the observed spectrum of segregating variation is not biased with respect to the spectrum of *de novo* mutations occuring during that time (Figure **??** and see Figure 1C in Wang *et al.*). APR: Would require a selection (or genotyping error) argument.

Focusing on the `GEVA` data,

- Beyond 10,000 generations, `GEVA`-ages spectra fluctuate by a very large amount (although Wang *et al.* (2023) "note that age estimates of mutations in the very distant past have decreased accuracy.")
- The fit is poor between data and model predictions, with model spectra trending in opposite directions from the data for some mutation classes (Gao *et al.*, 2022)

Given the poor fit of the model to the data and the known uncertainty in age estimation for older variants (Albers and McVean, 2020), we attempted to reproduce the inferred generation interval histories using allele age estimates from independent methods, names `Relate` (Speidel *et al.*, 2019) and `tsdate` (Wohns *et al.*, 2022), two state-of-the-art genealogical reconstruction methods.

- Allele age estimates between the three methods are only moderately correlated (as shown in Figure S20 in the Supplement of Wohns *et al.* (2022)).
- I estimated this correlation from our parsed data APR: `GEVA` and `Relate`: $r^2 \approx 0.28$, `GEVA` and `tsdate`: $r^2 \approx 0.34$, `tsdate` and `Relate`: $r^2 \approx 0.64$
- Despite this low to moderate correlation, we would hope that differences are unbiased with respect to the age-stratified mutation spectra. However, allele ages provided by each method result in distinct and unalike mutation spectrum histories (Figures 1–2), with mutation spectrum changes often trending in opposite directions over the same epochs.
- Even between `tsdate` and `Relate`, which have higher corretion in inferred allele ages, we do not see agreement of mutation spectrum history.
- In turn, these divergent histories provide estimates of generation time profiles that qualitatively differ.

Conclusions from this section:

- Mutation spectrum histories stratified by estimated allele ages are unreliable, as methods disagree even for fairly young mutations, and it's not clear whether *any* of the methods get it right (relevant to Gao *et al.* (2022)).
- It is not obvious where the discrepancies are coming from (need to look into Brandt *et al.* (2022))

Mutation spectra differ between *de novo* mutations and young alleles

The large disagreements in mutation spectrum histories between multiple variant age-estimation methods should cause skepticism of down-stream inferences that rely on them. But if we were to accept one of the mutation spectrum histories as accurate, there is a further cause for concern in comparing age-stratified mutation spectra to those estimated from pedigree studies (Jónsson *et al.*, 2017; Halldorsson *et al.*, 2019). As Wang *et al.* (2023) acknowledge, the spectrum of *de novo* mutations identified in Icelandic trios (Jónsson *et al.*, 2017) differs considerably from the spectrum of young segregating variation (e.g., variants estimated to be less than 100 generations old, Table 1). Gao *et al.* (2022) argue that these differences are unlikely to be driven by biological processes.

For some mutation classes, the relative proportion of *de novo* mutations in the trio-based study differs from the young-variant spectrum by up to 0.02, which would imply a large over- or under-count of different mutation types. `GEVA`, `tsdate`, and `Relate`, while they differ for mutations that are inferred to be older, very closely agree for mutations inferred to be less than 100 generations old (Table 1). In discussing this discrepancy, Wang *et al.* (2023) state, "We found that the mutation spectrum from the large pedigree study 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." Rather, `GEVA` does not provide estimates of allele ages for singletons, so this suggested source of discrepancy cannot be checked with their published allele ages. Both `tsdate` and `Relate` do report allele ages for singletons, and their inclusion does not strongly affect the mutation spectrum in the most recent time period (Table 1), though it does impact the mutation profiles in older time periods (Figures 4, 5). Of note, reported ages from `GEVA` and `Relate` both used the low-coverage phase 3 1KGP data while `tsdate` used the more recent independently sequenced high-coverage 1KGP data (Byrska-Bishop *et al.*, 2022), so the similarity of mutation profiles among young variants is unlikely to be driven by differences in coverage.

What could be driving the large disagreement between the spectrum of *de novo* mutations from pedigree-based approaches and that of young variants in the 1KGP dataset?

- True differences in mutation spectrum between the Iceland population and 1KGP populations APR: not likely – populations of different ancestries in 1KGP are consistent, and the EUR populations differ from Iceland
- Extremely recent large-scale changes in the *de novo* mutation spectrum APR: also not likely to occur at this scale, but if it were true, we should not be using the Iceland trio data to calibrate population genetics models at all
- Differences in selective pressures between mutations of different classes APR: selection would need to be very different, and affect many variants genome-wide. How strong would selection need to be to decrease certain mutation classes by a given amount? We could use *moments* for this...
- Genotyping error or bioinformatics choices APR: the agreement between high and low coverage data suggests that genotyping error does not have a strong effect in the 1KGP data. APR: instead, filtering and bioinformatics choices in the pedigree approach are the likely culprit. (Bergeron *et al.*, 2022)

Finally, a paragraph on model choices:

- 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

Conclusions

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.
3. Finally, Wang *et al.* (2023) gives us an excellent exmaple of the need for validation in population genetics studies, especially when inferences are built upon previous inferences that are known to be noisy and that need additional validations in their own right.

Figures

Many of these to the supplement.

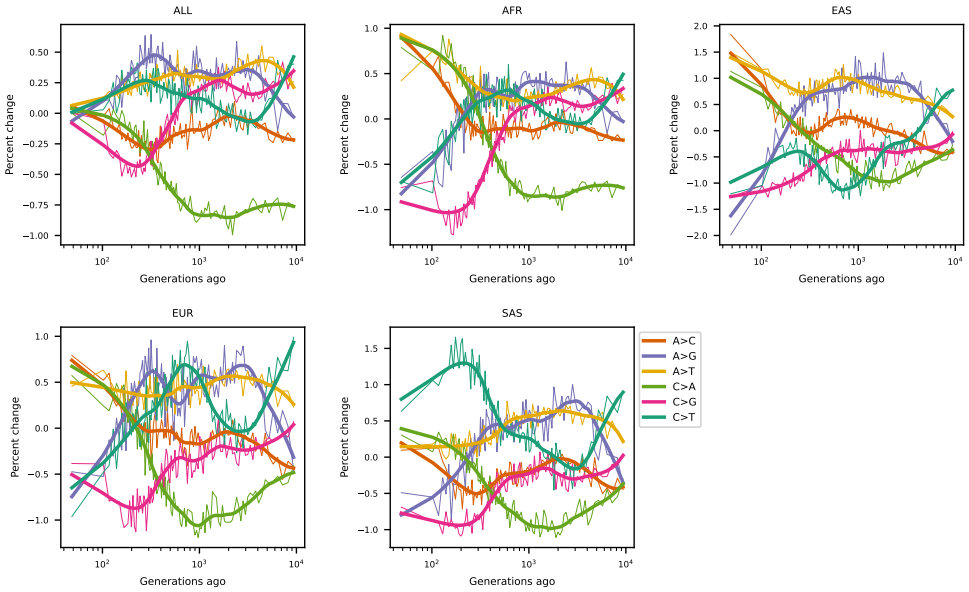

Figure 1: **`GEVA`-inferred mutation spectrum history.**

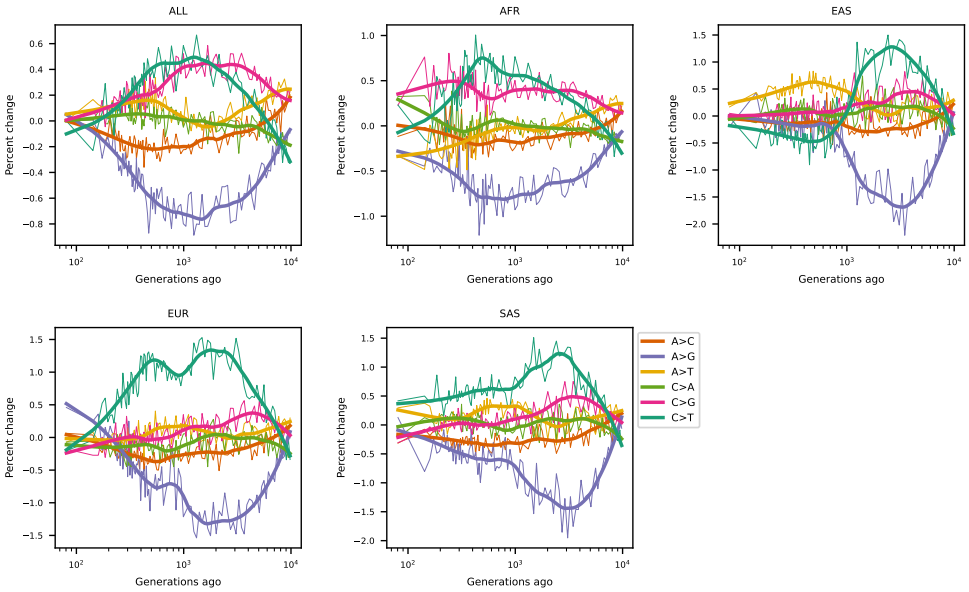

Figure 2: **`Relate`-inferred mutation spectrum history.**

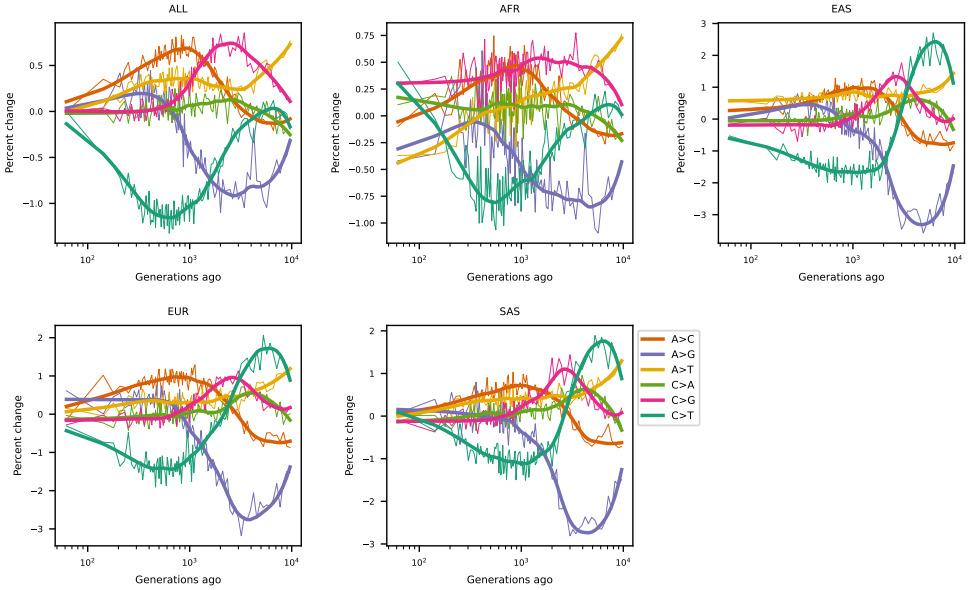

Figure 3: `tsdate`**-inferred mutation spectrum history.**

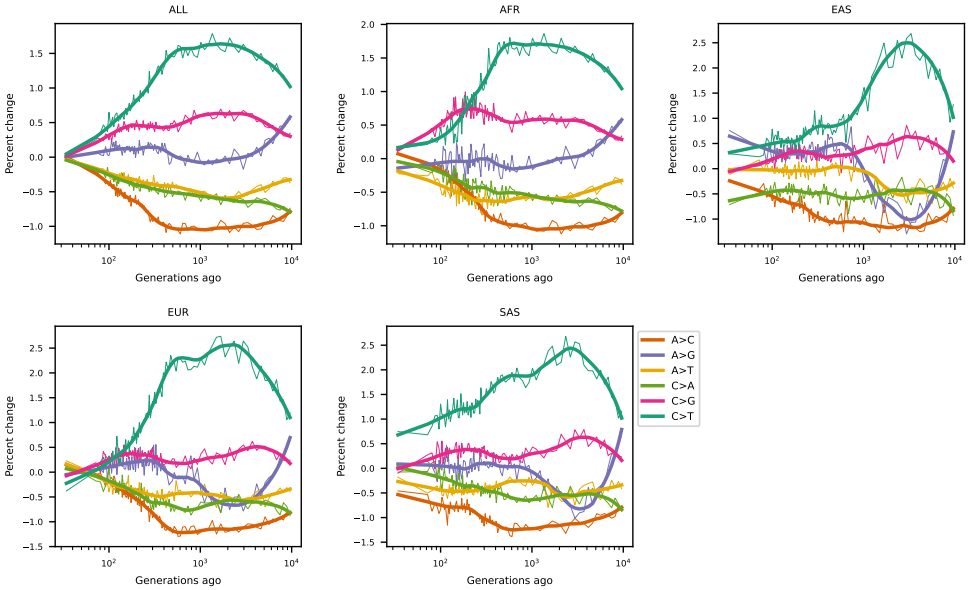

Figure 4: `Relate`**-inferred mutation spectrum history, including singletons.**

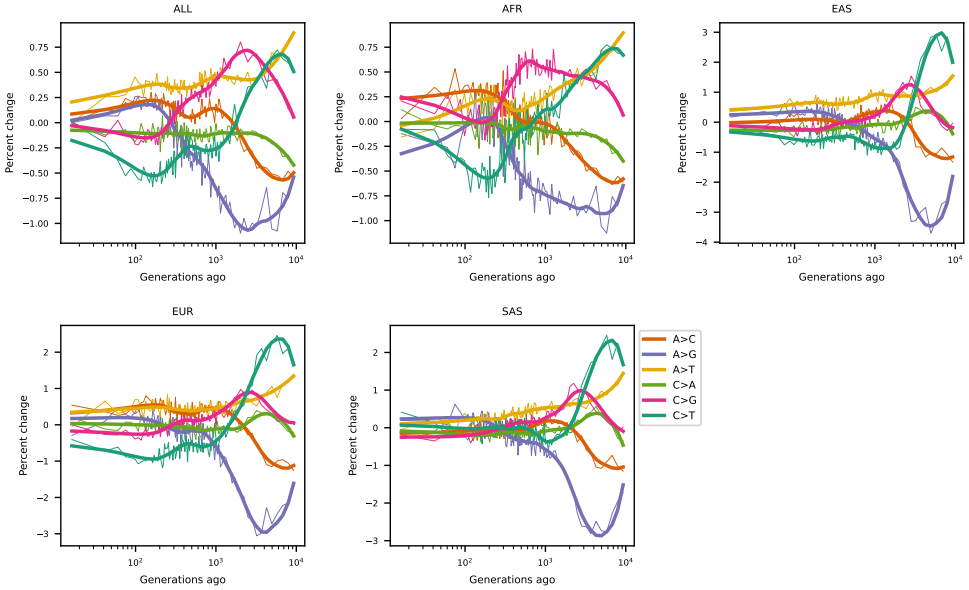

Figure 5: **`tsdate`-inferred mutation spectrum history, including singletons.**

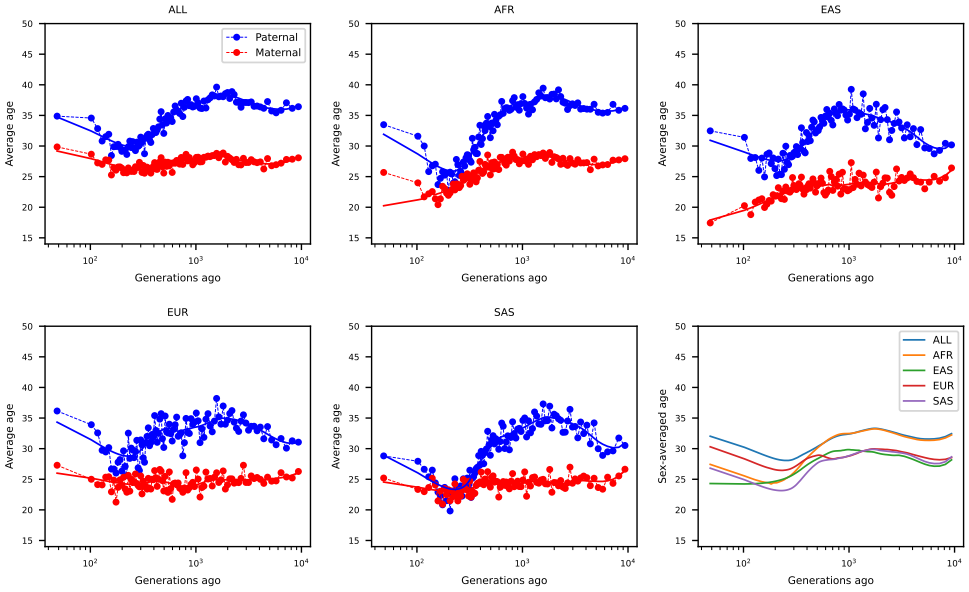

Figure 6: **`GEVA`-inferred generation time histories.**

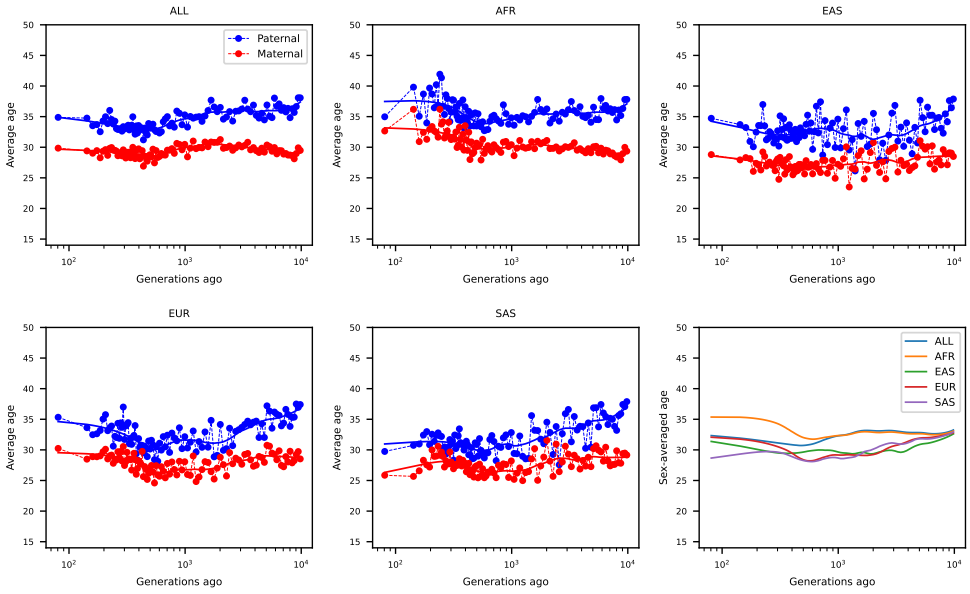

Figure 7: `Relate`**-inferred generation time histories.**

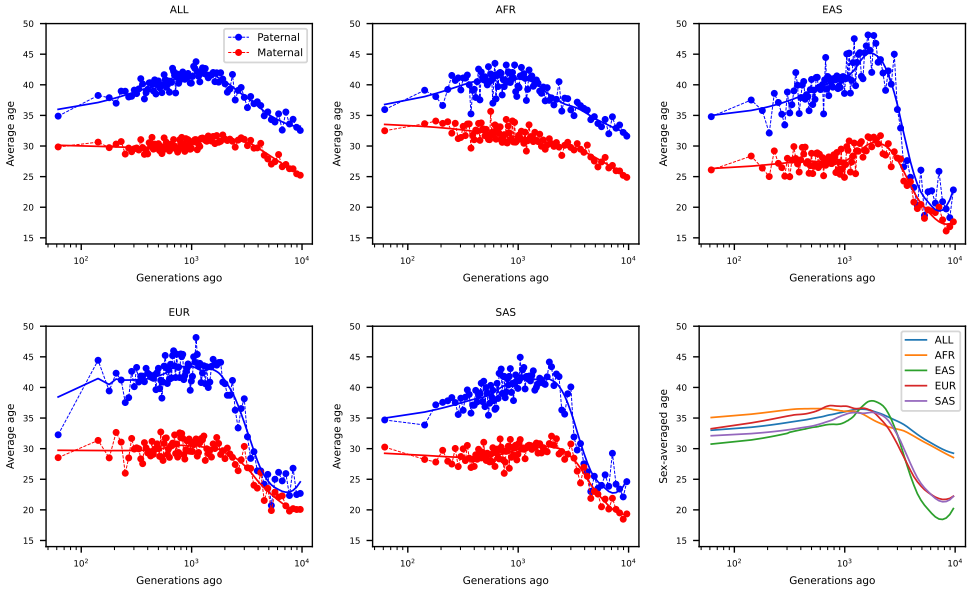

Figure 8: `tsdate`**-inferred generation time histories.**

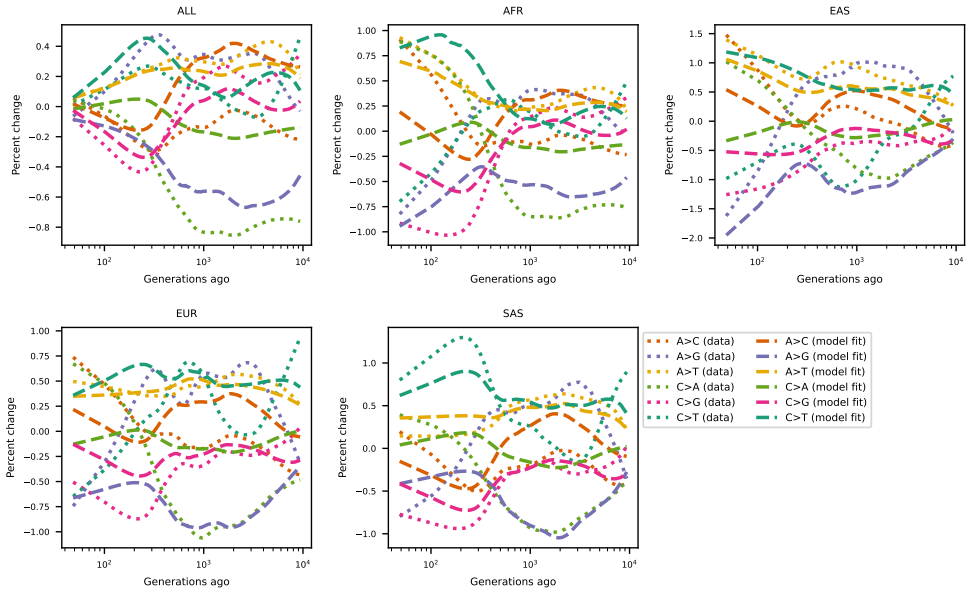

Figure 9: **Prediction of mutation spectrum history from `GEVA`-inferred generation times.**

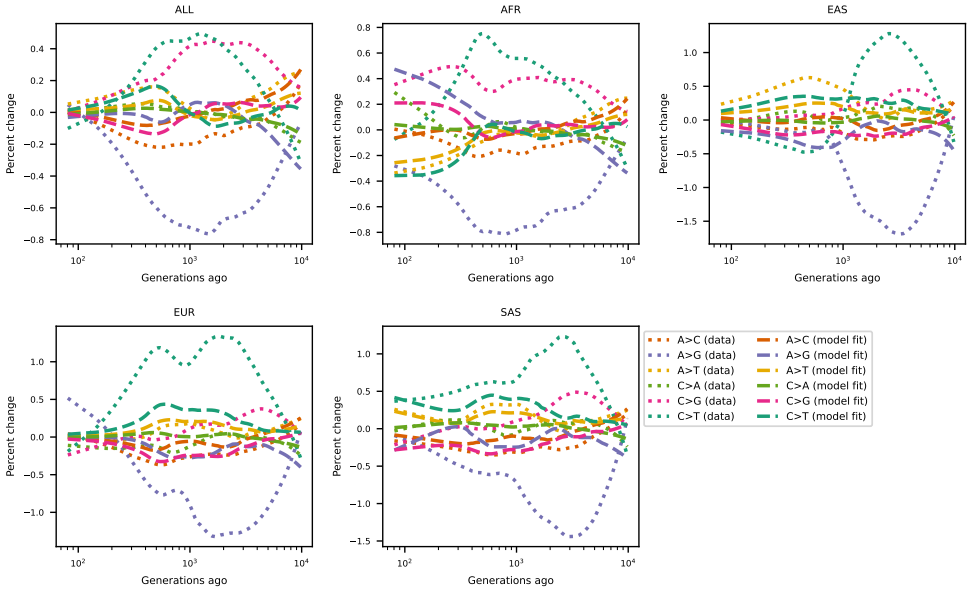

Figure 10: **Prediction of mutation spectrum history from `Relate`-inferred generation times.**

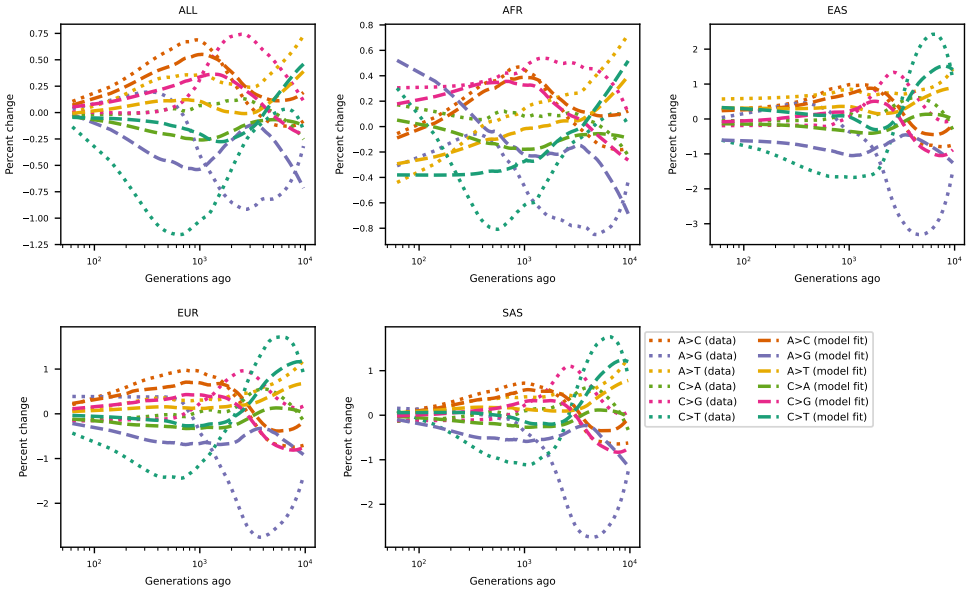

Figure 11: **Prediction of mutation spectrum history from `tsdate`-inferred generation times.**

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 unchanged. Note that `GEVA` does not report ages for singletons. While the three methods provide similar spectra from recent mutations, the spectrum from the Iceland pedigrees differs, in particular for the C→A and C→G classes. These differences are up to 2% of the proportion among all mutations, which corresponds to an under- or over-count of up to $\sim 20\%$ of C→A and C→G mutations, respectively. This difference remains whether the spectrum is estimated from only mutations that were phased in Jónsson *et al.* (2017) or from all mutations (phased and unphased).

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)	0.0953	0.3649	0.0890	0.0960	0.1216	0.2332
Trios (all mutations)	0.0962	0.3638	0.0923	0.0951	0.1202	0.2324

References

1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, *et al.*, 2015 A global reference for human genetic variation. Nature **526**: 68–74.

Albers, P. K., and G. McVean, 2020 Dating genomic variants and shared ancestry in population-scale sequencing data. PLoS biology **18**: e3000586.

Bergeron, L. A., S. Besenbacher, T. Turner, C. J. Versoza, R. J. Wang, *et al.*, 2022 The mutationathon highlights the importance of reaching standardization in estimates of pedigree-based germline mutation rates. Elife **11**: e73577.

Brandt, D. Y., X. Wei, Y. Deng, A. H. Vaughn, and R. Nielsen, 2022 Evaluation of methods for estimating coalescence times using ancestral recombination graphs. Genetics **221**: iyac044.

Byrska-Bishop, M., U. S. Evani, X. Zhao, A. O. Basile, H. J. Abel, *et al.*, 2022 High-coverage whole-genome sequencing of the expanded 1000 genomes project cohort including 602 trios. Cell **185**: 3426–3440.

Durvasula, A., and S. Sankararaman, 2020 Recovering signals of ghost archaic introgression in african populations. Science Advances **6**: eaax5097.

Gao, Z., Y. Zhang, N. Cramer, M. Przeworski, and P. Moorjani, 2022 Limited role of generation time changes in driving the evolution of mutation spectrum in humans. bioRxiv : 2022–06.

Halldorsson, B. V., G. Palsson, O. A. Stefansson, H. Jonsson, M. T. Hardarson, *et al.*, 2019 Characterizing mutagenic effects of recombination through a sequence-level genetic map. Science **363**: eaau1043.

Hammer, M. F., A. E. Woerner, F. L. Mendez, J. C. Watkins, and J. D. Wall, 2011 Genetic evidence for archaic admixture in africa. Proceedings of the National Academy of Sciences **108**: 15123–15128.

Hey, J., Y. Chung, A. Sethuraman, J. Lachance, S. Tishkoff, *et al.*, 2018 Phylogeny estimation by integration over isolation with migration models. Molecular biology and evolution **35**: 2805–2818.

Hsieh, P., A. E. Woerner, J. D. Wall, J. Lachance, S. A. Tishkoff, *et al.*, 2016 Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in central african pygmies. Genome research **26**: 291–300.

Hubisz, M. J., A. L. Williams, and A. Siepel, 2020 Mapping gene flow between ancient hominins through demography-aware inference of the ancestral recombination graph. PLoS genetics **16**: e1008895.

Jónsson, H., P. Sulem, B. Kehr, S. Kristmundsdottir, F. Zink, *et al.*, 2017 Parental influence on human germline de novo mutations in 1,548 trios from Iceland. Nature **549**: 519–522.

Lorente-Galdos, B., O. Lao, G. Serra-Vidal, G. Santpere, L. F. Kuderna, *et al.*, 2019 Whole-genome sequence analysis of a pan african set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-saharan populations. Genome biology **20**: 1–15.

Pagani, L., S. Schiffels, D. Gurdasani, P. Danecek, A. Scally, *et al.*, 2015 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**: 986–991.

Plagnol, V., and J. D. Wall, 2006 Possible ancestral structure in human populations. PLoS genetics **2**: e105.

Ragsdale, A. P., and S. Gravel, 2019 Models of archaic admixture and recent history from two-locus statistics. PLoS genetics **15**: e1008204.

Ragsdale, A. P., T. D. Weaver, E. G. Atkinson, E. Hoal, M. Möller, *et al.*, 2022 A weakly structured stem for human origins in africa. bioRxiv : 2022–03.

Scerri, E. M., M. G. Thomas, A. Manica, P. Gunz, J. T. Stock, *et al.*, 2018 Did our species evolve in subdivided populations across africa, and why does it matter? Trends in ecology & evolution **33**: 582–594.

Schlebusch, C. M., H. Malmström, T. Günther, P. Sjödin, A. Coutinho, *et al.*, 2017 Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. Science **358**: 652–655.

Speidel, L., M. Forest, S. Shi, and S. R. Myers, 2019 A method for genome-wide genealogy estimation for thousands of samples. Nature genetics **51**: 1321–1329.

Wang, R. J., S. I. Al-Saffar, J. Rogers, and M. W. Hahn, 2023 Human generation times across the past 250,000 years. Science Advances **9**: eabm7047.

Wohns, A. W., Y. Wong, B. Jeffery, A. Akbari, S. Mallick, *et al.*, 2022 A unified genealogy of modern and ancient genomes. Science **375**: eabi8264.