

# Multiple sources of uncertainty confound inference of historical human generation times

Aaron P. Ragsdale<sup>1,\*</sup> and Kevin R. Thornton<sup>2</sup>

<sup>1</sup>University of Wisconsin–Madison, Wisconsin, USA

<sup>2</sup>University of California, Irvine, California, USA

\*apragdale@wisc.edu

May 25, 2023

## Abstract

WANG *et al.* (2023) recently proposed an approach to infer the history of human generation intervals from changes in mutation profiles over time. As the relative proportions of different mutation types depend on the ages of parents, binning variants by the time they arose allows for the inference of average paternal and maternal generation intervals over time. Applying this approach to published allele age estimates, WANG *et al.* (2023) 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 argue that the results and interpretations in WANG *et al.* (2023) are primarily driven by noise and biases in input data 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 downstream analyses may be strongly influenced by uncharacterized biases in their output.

Recent years have seen the rapid development of methods for reconstructing genetic genealogical structures of large unrelated 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) have the potential to transform population genetic inference. Biological and evolutionary processes determine the structure of these genealogies, including the distribution of mutations. In particular, the age of a variant can be estimated by mapping its mutation to the branch of the genealogy consistent with the observed data.

The past few years have also seen efforts to sequence large sets of pedigrees, providing increased resolution of important parameters of genome biology such as direct measurements of mutation rates and profiles. Through high-coverage sequencing of multiple generations of families, *de novo* mutations can be determined as maternally or paternally inherited. With a large enough sample size (meioses), both the number of mutations and proportions of mutation types (i.e., the *mutation spectrum* of A→C, A→G, etc. mutation types) can be correlated with parental age and sex (JÓNSSON *et al.*, 2017; HALLDORSSON *et al.*, 2019). These two sets of inferences, the estimated ages of mutations and the parental age- and sex-dependence of the mutation spectrum, can be combined to infer the history of average maternal and paternal generation intervals for human populations of diverse ancestries (MACIÀ *et al.*, 2021; WANG *et al.*, 2023). In order to avoid overfitting, this approach requires making a number of assumptions about the constancy of the mutational process over time and its similarity across populations (HARRIS, 2015; MATHIESON and REICH, 2017; HARRIS and PRITCHARD, 2017; DEWITT *et al.*, 2021), and of negligible effects of selection and biased gene conversion (LACHANCE and TISHKOFF, 2014; GLÉMIN *et al.*, 2015) on shaping the profile of segregating mutation types.

This approach has recently been criticized (GAO *et al.*, 2023). GAO *et al.* found that observed changes in the mutation spectrum over time cannot be explained by changes in maternal and paternal generation intervals alone. Male and female generation times are required to change in different directions within the same time interval in order to explain the data for different classes of mutation. They argue 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 reexamine the methodological assumptions and results in WANG *et al.* (2023). We first consider the reported inferred generation time histories and find that they are inconsistent with current understanding of human population history, in particular population structure within Africa. In exploring the source of this inconsistency, we show that allele age estimates are not just noisy, but age-stratified mutation spectra reconstructed using independent methods do not agree, with mutation profiles diverging in opposing directions. Thus, the results from WANG *et al.* (2023) do not reproduce when applying different estimates of allele ages from the same population samples. We further discuss the disagreement between the mutation rate profile found in pedigree studies and from young variants, the source of which is not well understood and complicates their comparison. In conclusion, we suggest that current methods of genealogical reconstruction and allele age estimation should continue to be evaluated for subtle biases, and that downstream analyses using estimated allele ages and mutation profiles should carefully validate their results.

## Long-lasting generation time differences are inconsistent with current human history models

Applying the proposed 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 their study. In discussing this inference the authors state, “[T]he 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 shaped present-day genetic variation (HAMMER *et al.*, 2011; HSIEH *et al.*, 2016; HEY *et al.*, 2018; RAGSDALE and GRAVEL, 2019; LORENTE-GALDOS *et al.*, 2019; DURVASULA and SANKARARAMAN, 2020).

However, in extending their analysis deeper into the past, WANG *et al.* (2023) find that ancestral generation intervals do not converge until many 10s of thousands of generations ago. Assuming an average generation time of 25–30 years, this corresponds to one to two million years in the past. 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 long-lasting isolation among ancestral populations is not impossible, it is not supported by genetic (e.g., RAGSDALE *et al.*, 2023) or archaeological (e.g., SCERRI *et al.*, 2018) evidence, which rather suggest at least periodic connectivity of ancestral human populations within Africa.

Genetic studies have estimated the Eurasian–West African divergence (i.e., the time of recent shared ancestry) at only  $\approx 75$ ka (thousand years ago) (e.g., PAGANI *et al.*, 2015; BERGSTRÖM *et al.*, 2020). While population genetic studies vary considerably in estimated split times, even those that infer deeper human divergences place the Eurasian–West African divergence at 100–150ka (e.g., SCHLEBUSCH *et al.*, 2017). If such estimated divergence times represent the majority of ancestry of the two groups (while allowing for a smaller portion to be due to long-lasting structure), then the shared portion of ancestry should be subject

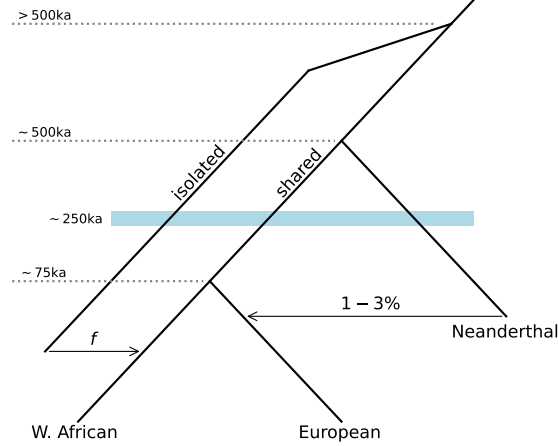


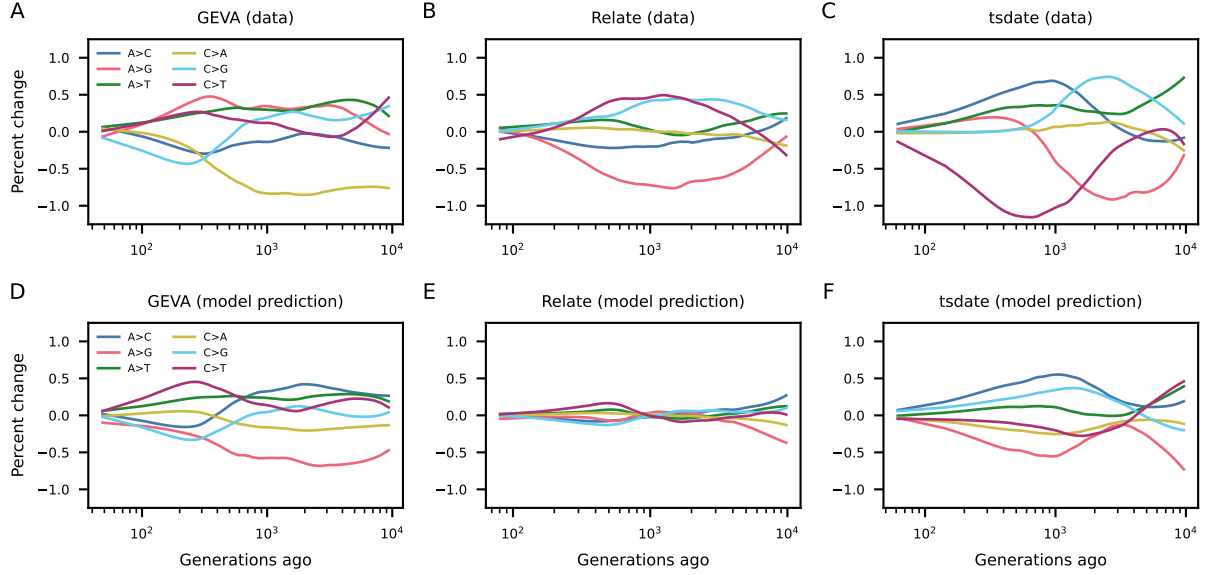
Figure 1: **A model for archaic admixture within Africa.** In such a model for “ghost” archaic introgression within Africa (DURVASULA and SANKARARAMAN, 2020), an isolated lineage diverged prior to the human-Neanderthal split and more recently contributed ancestry (with proportion  $f$ ) to West African, but not European populations. The mutation spectrum of West African populations among alleles dating to  $\sim 250 - 300\text{ka}$  is due to a combination of mutations from the “shared” and “isolated” lineages, while the European populations’ spectrum from this time is due to a combination from the “shared” and Neanderthal branches. WANG *et al.* (2023) showed that masking Neanderthal segments in Eurasian populations had a negligible effect on their inferred generation time histories, so we ignore this small contribution when comparing observed mutation spectra between West African and Eurasian groups.

to the same generation intervals prior to the divergence time. Any differences in the mutation spectrum from those epochs would be driven by differences in generation times affecting the minority of ancestry that remained isolated.

As a simple test of such a scenario, we considered a model of archaic admixture within Africa, allowing for some proportion of admixture from a diverged lineage into West African populations (such as  $\approx 10\%$  as inferred by DURVASULA and SANKARARAMAN (2020), Figure 1). At 10,000 generations ago, average paternal and maternal generation intervals in the ancestors of Eurasians were inferred to both be  $\approx 20$  years, while the ancestral African generation intervals were at least 28 and 23 (see Figure S4 in WANG *et al.* (2023)). Using the same mutation model (JÓNSSON *et al.*, 2017), we can determine the generation intervals in the isolated lineage that is needed to result in a mutation spectrum matching that of the inferred generation times (see Supporting Information).

We assume that admixture proportions from the diverged and Eurasian-shared lineages ( $f$  and  $1 - f$ , respectively) result in surviving variation from those epochs having similar proportions of contributions (in effect, ignoring differences in total mutation rate and demographic effects that may distort the magnitude of the contributed mutation spectra, see Supporting Information). With 10% admixture into the ancestors of West Africans from a diverged lineage, the mean paternal age of conception would need to be 92 and the mean maternal age 48. These are unreasonably long generation times for *Homo* species. With 20% admixture from this diverged lineage (which is larger than has been proposed or inferred in previous genetic studies), mean ages would still need to be 58 and 34. For the paternal average generation time to be less than 40, this model would require  $f = 0.4$  or that the inferred generation times in the shared branch were considerably higher (Table S1).

Therefore, even assuming a model of long-lasting population structure with strict isolation within Africa, we find the reconstructed generation time intervals over the past 10,000 years from WANG *et al.* (2023) to be incompatible with plausible life histories of early humans. Given this, it is natural to ask what may be



**Figure 2: Time-stratified mutation spectra are inconsistent and poorly fit by historical generation time changes.** (A-C) Three methods for estimating allele ages provide incongruous mutation spectrum histories. Mutation spectra are the proportion of mutations falling within each of the six mutation classes, so “percent changes” refers to the change these proportions ( $\times 100$ ) over time relative to the most recent time bin. (D-F) Generation time histories were fit to each observed mutation spectrum profile (Figures S16–S18), and we calculated the mutation spectrum histories predicted by each of these inferred histories (shown here). None of the mutation spectrum histories observed in the data are recovered by the predicted from the inferred generation time histories. Figures S16–S18 show the mutation spectrum histories from data (top row) overlaid with predictions (bottom row). Trajectories are smoothed using LOESS regression.

causing such mis-inference. Below we show that multiple sources of uncertainty, namely noise and bias in allele age inference and inconsistencies in trio-base estimates of mutation profiles, confound inferences of generation times from age-stratified mutation spectra.

### Inconsistencies in inferred mutation spectra over time

WANG *et al.* (2023) used published allele ages from GEVA (ALBERS and McVEAN, 2020) to construct mutation spectra within time periods. GEVA estimates allele ages by considering the number of 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 and are not assigned an age. Partitioning variants by their estimated ages shows that the mutation spectrum has changed over time, assuming that the observed spectrum of segregating variation is not biased with respect to the spectrum of *de novo* mutations occurring during that time (Figure 2A and see Figure 1C in WANG *et al.* (2023)).

In fitting generation time histories to data from the past 10,000 generations, we find that the inferred generation times provide a poor fit to the data (Figure 2D). The relative proportions of the predicted mutation spectrum trend in opposite directions for some mutation classes, confirming the concern from GAO *et al.* (2023) that a single generation time history cannot simultaneously explain each of the six mutation class frequency changes. For alleles older than 10,000 generations, the mutation spectrum fluctuates by large amounts (Figure S6). While estimated allele ages from the distant past have decreased accuracy (ALBERS and McVEAN, 2020; WANG *et al.*, 2023), the large differences in proportions suggests a bias in GEVA’s

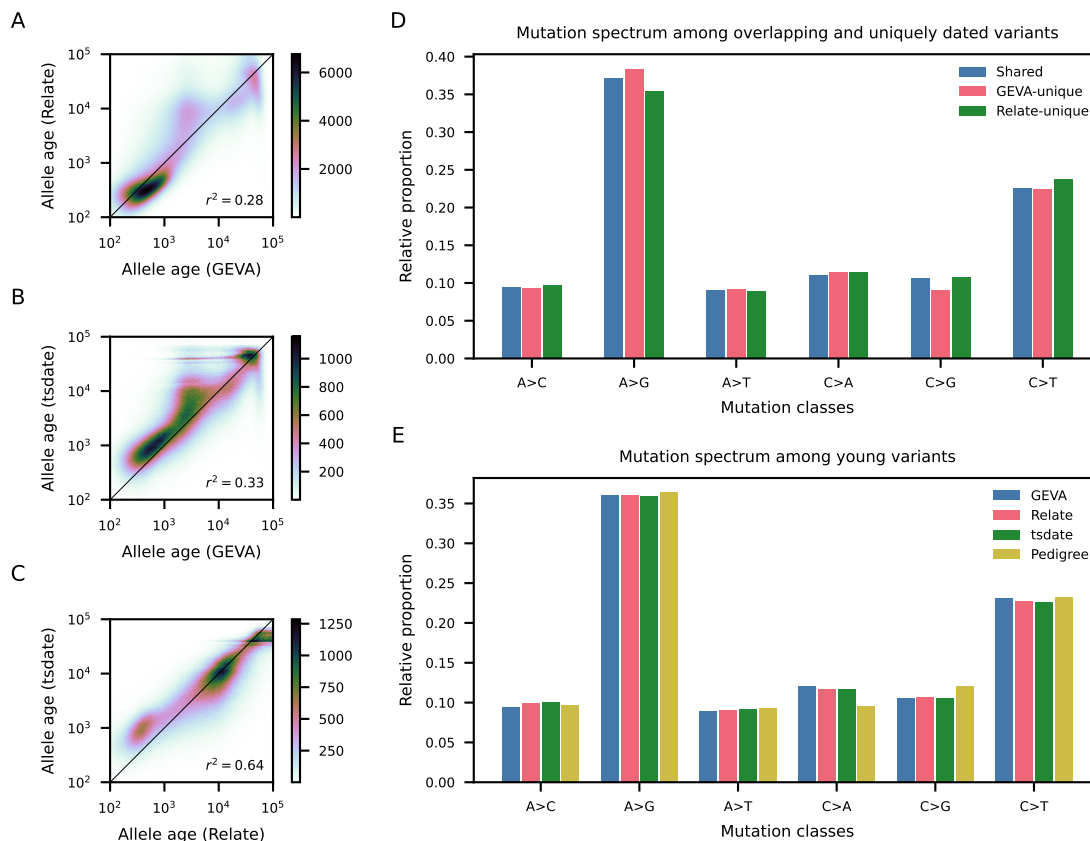


Figure 3: **Comparing allele ages and mutation spectra across data sources.** (A-C) Allele age estimates are only moderately correlated between methods. Shown here are allele age estimates for variants that were assigned ages by two or more of **GEVA**, **Relate** and **tsdate**. (D) The proportions of mutation types uniquely dated by **GEVA** and **Relate** differ from the variant proportions among mutations dated by both methods, indicating biases in the mutation types that are kept and discarded. Here, we compared **GEVA** and **Relate** (instead of **tsdate**) because allele ages were obtained from the same input data, namely phase three of the 1KGP. (E) While the mutation spectrum among young variants dated to  $< 100$  generations is similar between age estimation methods, the pedigree-based estimate of the *de novo* mutation spectrum (JÓNSSON *et al.*, 2017) differs, in particular for C→A and C→G mutations.

reported dates that correlates with mutation type.

Accurate dating of variant ages is central to the inference of generation intervals from time-stratified mutation spectra. Given the poor fit of the model to the **GEVA** data and the known uncertainty in age estimation for older variants, we attempted to reproduce the inferred generation time histories using allele age estimates from independent sources, **Relate** (SPEIDEL *et al.*, 2019) and **tsdate** (WOHNS *et al.*, 2022), both state-of-the-art genealogical reconstruction methods able to scale to modern sample sizes. The ages of variants dated by at least two of the methods are only moderately correlated (**GEVA** and **Relate**:  $r^2 \approx 0.28$ , **GEVA** and **tsdate**:  $r^2 \approx 0.33$ , **tsdate** and **Relate**:  $r^2 \approx 0.64$ ; Figure 3A-C and see Figure S20 from WOHNS *et al.* (2022)). Imperfect correlations should be expected, as genealogical reconstruction methods have been shown to vary in the accuracy of estimated  $T_{MRCA}$  (BRANDT *et al.*, 2022).

Despite the imperfect correlation, it is still possible that differences in estimate allele ages are unbiased with respect to mutation type. However, we find that ages provided by each method disagree at the level of

individual mutation spectrum histories (Figures 2A-C and S1-S3), with changes in mutation proportions often trending in opposite directions over the same time periods. In turn, these divergent spectrum histories provide qualitatively different inferred generation time histories (Figures S13-S15). None of them provide a reasonable fit to the data (Figure 2D-F).

Published allele ages from both ALBERS and McVEAN (2020) (**GEVA**) and SPEIDEL *et al.* (2019) (**Relate**) used the same input dataset: the phase 3 release of 1KGP. However, they differ in the total number of variants that were assigned ages (roughly 30 million and 48 million, respectively). **GEVA** and **Relate** provide allele age estimates for data that is consistent with certain assumptions, such as allowing only a single mutation at a given site. For multi-allelic mutations or those that are inconsistent with the underlying tree topology, **Relate** and **tsdate** retain this data by allowing multiple mutations at a site, although such sites are discarded from downstream analyses. Of variants with a single allele age estimate, 25 million were dated by both **GEVA** and **Relate**, with the remaining uniquely assigned an age by one or the other method. When comparing mutation spectrum histories restricted to variants dated by both **GEVA** and **Relate**, the two methods still predict different mutation profile trajectories (Figure S12).

In comparing the proportions of mutation types assigned ages by both methods or uniquely by one, we find that those proportions differ considerably (Figure 3D). For example, **GEVA** provides age estimates for relatively more A→G mutations and fewer C→G mutations, while **Relate** provides age estimates for relatively more C→T and fewer A→G mutations. This owes to methodological differences in evaluating whether a particular variant can be reliably dated, which may have both bioinformatic (e.g., genotyping, phasing and polarization errors) and biological (e.g., recurrent mutations) causes. Therefore, when comparing mutation spectra that are conditioned on variants being dated by a particular method, it is not apparent which method, if any, provides unbiased estimates of age-stratified mutation spectra.

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

Independent of biases in allele age estimate, there are further problems 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 from the spectrum of young segregating variation (e.g., variants estimated to be less than 100 generations old, Figure 3E and Table S2). 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. This is the equivalent of 10s of thousands of SNPs in the most recent bin, with the exact number depending on the variant-dating method. While estimated mutation spectra from **GEVA**, **tsdate** and **Relate** differ at older times, they closely agree for mutations inferred to be less than 100 generations old (Table S2).

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.” While **GEVA** does not provide estimates of allele ages for singletons, both **tsdate** and **Relate** do report singleton allele ages, and their inclusion does not strongly affect the mutation spectrum for very young variants (Table S2), though it does impact the mutation profiles in older time periods (Figures S7, S11). 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 accuracy of the mutation spectrum among young variants is unlikely to be driven by differences in coverage.

To account for the differences between the *de novo* spectrum from pedigree studies (JÓNSSON *et al.*, 2017) and the spectrum among young variants, WANG *et al.* (2023) subtracted this difference from each historical mutation spectrum. “[This choice] has the effect of assuming that parental ages in the pedigreed mutation dataset reflect generation times in the most recent historical bin” (WANG *et al.*, 2023). However, the average generation time among present-day Icelanders may not reflect average generation times in world-wide populations over the past 3-5 thousand years, represented by the most recent time bin in their analysis.

Still more concerning is that we do not know the source of the discrepancy between the Iceland and 1KGP mutation spectra. Without knowing this, it is unclear that simply subtracting the differences between them properly accounts for it. Instead, mutation class-specific genotyping differences may distort the underlying mutation model, potentially driving deviations in predicted mutation profiles in unexpected ways.

What could be driving the large disagreement between the spectrum of *de novo* mutations from Icelandic pedigrees and that of young variants in the 1KGP dataset? First, there could be true differences in the mutation spectrum between the Iceland cohort and 1KGP populations, although this seems unlikely, as populations of different ancestries in 1KGP have more similar recent mutation spectra to each other than to the Iceland *de novo* spectrum (including the EUR populations, Table S3). If the differences are real, then the Icelandic pedigree data is an inappropriate calibration for the mutation spectrum in other populations. Second, there could be differences in selective pressures between mutations of different classes, although selection would need to be very strong for some classes compared to others in order to see the observed difference among young variants. Third, the signal may be driven by genotyping error or bioinformatics choices, though the agreement between high- and low-coverage 1KGP datasets suggests that genotyping error does not have a strong effect in the 1KGP data. Instead, filtering and bioinformatics choices in the pedigree approach are the likely culprit (BERGERON *et al.*, 2022). Indeed, GAO *et al.* (2023) show that two studies of Icelandic trios that partially overlap in families that were included (JÓNSSON *et al.*, 2017; HALLDORSSON *et al.*, 2019) estimate significantly different *de novo* mutation spectra due to filtering choices (see Figure 3, supplement 2 in GAO *et al.* (2023)). Until the source of these differences are understood and properly accounted for, we caution that population-genetic inferences should avoid calibration using mutation rates and profiles from pedigree studies.

## Summary of previous simulation studies

BRANDT *et al.* (2022) looked at argweaver, relate and tsinfer/tsdate. argweaver is not relevant here. They focus on the mean squared errors (MSE) of estimated coalescence times. Sadly, the SI material genetics seems broken so we go to the preprint at <https://www.biorxiv.org/content/10.1101/2021.11.15.468686v4>.

Their results:

1. Figure S1: relate and tsinfer/tsdate tend to underestimate pairwise coalescence times. (The latter looks especially bad.)
2. Figure S2: The MSE of the estimate coalescence time increases with increasing true coalescence time.
3. Figure S2: All programs OVER estimate recent times and UNDER estimate older times.
4. Figure S2: The sims have mutation rate equal to the recombination rate.

From this: the MSE is very low (looks zero-ish) for recent times, but large for older times. From their results, downstream analyses that either take point estimates at face value or bin point estimates will have inaccurate point estimates or estimates moving into the wrong bins for older nodes/mutations.

All of these simulations are highly idealized—the methods say, “all sims done under standard coalescent w/recombination” and cite Hudson 1983.

From the GEVA paper (ALBERS and MCVEAN, 2020):

1. Fig 2A shows results based on simulations comparable to BRANDT *et al.* (2022), but with much larger sample size. GEVA works very well for this case.
2. For “complex demography” and perfect data, Fig 2B shows a tendency to over-estimate recent times and under estimate older times.
3. These trends get more pronounced for imperfect data (Fig 2B bottom panels.)

4. The SI figs converge on a story (by SI Fig 3) that complex demography + data (genotyping) error gives the same types of biases as described by Brandt et al. (Phasing seems to make things a touch worse – SI Fig 4.)

Complex demography means the Gutenkunst model. Simulating data errors means introducing genotyping errors to mimic what they think is reasonable in light of real data. They also add additional error for some results by adding a phasing step.

I would argue that none of this explains what is going on in the real analysis! If older mutations usually end up in wrong time bins, that should be equally likely for all mutation classes. But that is NOT what we see! If we were to simulate data where we mess around with the transition matrix in different time epochs (and mess with the mutation rate to fake changes in generation time), I doubt that we'd mimic mutation class A showing a different pattern from B. It would seem that what we see much be some *technical* issue that we still don't understand?

## Prose re: simulation results

APR: Seems we could have a short discussion-y section on this, as well as some of the assumptions that go into the methods (such as constancy of mutation rates, and that information about the mutation types are unseen and unused by the tree sequence reconstruction approach). This would also satisfy some of reviewer 1's requests for discussion.

*I don't know quite where to put this yet.*

Previous work have evaluated the performance of **GEVA**, **Relate**, and of **tsdate** using coalescent simulation (BRANDT *et al.*, 2022; ALBERS and MCVEAN, 2020). For data simulated under neutral models with constant population size, and assuming perfect data, **Relate** and **tsdate** tend to overestimate recent coalescent times and underestimate older times (BRANDT *et al.*, 2022). For this same case, **GEVA** will tend to overestimate "intermediate" times and underestimate older times (see ALBERS and MCVEAN, 2020, Fig S1). Under demographic models inferred for human populations, **GEVA** will begins to show the same behavior as **Relate** and **tsdate**, over- and under- estimating recent and older times, respectively (ALBERS and MCVEAN, 2020, Fig S2). Genotyping and phasing errors exacerbate these biases (ALBERS and MCVEAN, 2020, Figs S3 and S4). BRANDT *et al.* (2022) did not perform simulations under such conditions. For **Relate** and **tsdate**, the mean squared error (MSE) of node ages increases for more ancient nodes BRANDT *et al.* (2022). ALBERS and MCVEAN (2020) did not provide a comparable evaluation for **GEVA**.

The accuracy of mutation age estimates is determined by the accuracy of the coalescent time estimates (node ages). We therefore expect the ages of recent mutations to be overestimated and the ages of older variants to be underestimated. Analyses that "bin" mutations into time windows will tend to have older mutations placed into incorrect bins (due to the higher MSE of the age estimates).

While estimation errors may contribute to the poor fit of the mutation spectra to the data (2), it is not clear why different mutation types (A to T, etc.) should be affected different nor why they should give different inferences about generation time (all see GAO *et al.*, 2023). (Mumble mumble something else msut be happening.)

## Conclusions

Large-scale reconstruction of gene genealogies, dating mutations, and  $T_{\text{MRCA}}$  inference are all exciting developments, but these methods are quite new and will likely advance considerably in the coming years. Currently, existing methods show different accuracies even under overly-idealized conditions (BRANDT *et al.*, 2022). We should therefore be cautious of unexpected and subtle biases that can impact downstream analyses. In the case of generation time inference (WANG *et al.*, 2023), we showed that patterns of variation that were attributed to biological processes (variation in generation times) are predominantly driven by nonbiological artifacts in the input data. As the field continues to evaluate the accuracy of allele age estimation



and genealogical reconstruction methods, we recommend that analyses that rely on their output thoroughly validate their results using independent approaches.

## Acknowledgements

We thank Ziyue Gao and Priya Moorjani for helpful discussions.

## 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.
- BERGSTRÖM, A., S. A. MCCARTHY, R. HUI, M. A. ALMARRI, Q. AYUB, *et al.*, 2020 Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**: eaay5012.
- 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.
- DEWITT, W. S., K. D. HARRIS, A. P. RAGSDALE, and K. HARRIS, 2021 Nonparametric coalescent inference of mutation spectrum history and demography. *Proceedings of the National Academy of Sciences* **118**: e2013798118.
- 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, 2023 Limited role of generation time changes in driving the evolution of mutation spectrum in humans. *eLife* **12**: e81188.
- GLÉMIN, S., P. F. ARNDT, P. W. MESSER, D. PETROV, N. GALTIER, *et al.*, 2015 Quantification of gc-biased gene conversion in the human genome. *Genome Research* **25**: 1215–1228.
- 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.
- HARRIS, K., 2015 Evidence for recent, population-specific evolution of the human mutation rate. *Proceedings of the National Academy of Sciences* **112**: 3439–3444.
- HARRIS, K., and J. K. PRITCHARD, 2017 Rapid evolution of the human mutation spectrum. *eLife* **6**: e24284.
- 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.
- LACHANCE, J., and S. A. TISHKOFF, 2014 Biased gene conversion skews allele frequencies in human populations, increasing the disease burden of recessive alleles. *The American Journal of Human Genetics* **95**: 408–420.
- 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.
- MACIÀ, M. C., L. SKOV, B. M. PETER, and M. H. SCHIERUP, 2021 Different historical generation intervals in human populations inferred from neanderthal fragment lengths and mutation signatures. *Nature Communications* **12**: 5317.
- MATHIESON, I., and D. REICH, 2017 Differences in the rare variant spectrum among human populations. *PLoS Genetics* **13**: e1006581.
- 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.
- 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. G. HOAL, M. MÖLLER, *et al.*, 2023 A weakly structured stem for human origins in africa. *Nature* : 1–9.
- 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.