

1 **A likelihood-based framework for demographic inference from genealogical trees**

2 Caoqi Fan^{1,2}, Jordan L. Cahoon^{1,3}, Bryan L. Dinh^{1,2}, Diego Ortega-Del Vecchyo⁴, Christian
3 Huber⁵, Michael D. Edge¹, Nicholas Mancuso^{1,2}, Charleston W.K. Chiang^{1,2}

4

5 ¹Department of Quantitative and Computational Biology, University of Southern California

6 ²Center for Genetic Epidemiology, Department of Population and Public Health Sciences, Keck
7 School of Medicine, University of Southern California

8 ³Department of Computer Science, University of Southern California

9 ⁴Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional
10 Autónoma de México, Juriquilla, Querétaro, México

11 ⁵Department of Biology, Penn State University, University Park, PA, USA.

12

13 **Abstract**

14 The demographic history of a population drives the pattern of genetic variation and is encoded
15 in the gene-genealogical trees of the sampled alleles. However, existing methods to infer
16 demographic history from genetic data tend to use relatively low-dimensional summaries of the
17 genealogy, such as allele frequency spectra. As a step toward capturing more of the information
18 encoded in the genome-wide sequence of genealogical trees, here we propose a novel
19 framework called the genealogical likelihood (gLike), which derives the full likelihood of a
20 genealogical tree under any hypothesized demographic history. Employing a graph-based
21 structure, gLike summarizes across independent trees the relationships among all lineages in a
22 tree with all possible trajectories of population memberships through time and efficiently
23 computes the exact marginal probability under a parameterized demographic model. Through
24 extensive simulations and empirical applications on populations that have experienced multiple
25 admixtures, we showed that gLike can accurately estimate dozens of demographic parameters
26 when the true genealogy is known, including ancestral population sizes, admixture timing, and
27 admixture proportions. Moreover, when using genealogical trees inferred from genetic data, we
28 showed that gLike outperformed conventional demographic inference methods that leverage
29 only the allele-frequency spectrum and yielded parameter estimates that align with established
30 historical knowledge of the past demographic histories for populations like Latino Americans
31 and Native Hawaiians. Furthermore, our framework can trace ancestral histories by analyzing a
32 sample from the admixed population without proxies for its source populations, removing the
33 need to sample ancestral populations that may no longer exist. Taken together, our proposed
34 gLike framework harnesses underutilized genealogical information to offer exceptional

35 sensitivity and accuracy in inferring complex demographies for humans and other species,
36 particularly as estimation of genome-wide genealogies improves.

37

38

39 **Introduction**

40 Accurately inferring the population history of humans has archaeological and historical
41 significance, and it also helps to properly account for population structure in association studies
42 and improve robustness in inferences about natural selection¹. Because of the complicated
43 interplay of random processes related to the underlying demography and observed genotypes –
44 including migration, coalescence, recombination, mutation, and genotyping error – demographic
45 inference is a challenging problem, often requiring simplifying assumptions or relatively coarse
46 data summaries. One popular way of estimating the size changes of a single population utilizes
47 a hidden Markov model (HMM) to describe the variation of haplotypes along the genome, where
48 the hidden states correspond to the underlying genealogical trees^{2–6}. As the number of potential
49 trees grows exponentially with sample size, these methods are computationally scalable by
50 tracking only a reduced representation of the underlying genealogy (e.g., SMC++⁵ and ASMC⁶
51 only track the coalescent times between a specific pair of haplotypes, while the remaining
52 samples assume auxiliary functions). These methods are typically constrained by small sample
53 sizes (usually <100) and the assumption of a single, homogeneous population, although they
54 are flexible with respect to the population size trajectories over time. To accommodate for larger
55 sample sizes that are more informative of the recent human history as well as more complex
56 demographic events such as splits, migrations, and admixture, alternative approaches to
57 demographic inference rely on a further reduced representation of the genealogy, the allele
58 frequency spectrum (AFS)^{7–11}. Although convenient to compute, the AFS may not contain
59 enough information to recover the history precisely^{12–14}.

60

61 HMM- and AFS-based methods are based on observed genotypes or haplotypes. However,
62 since neutral variation is related to demographic history entirely via the genealogical processes,
63 the (unknown) genealogy arguably has a more direct relationship with the underlying
64 demography than the downstream genotypes^{15–17}. Moreover, the complete genealogy of a
65 collection of samples, as represented by an ancestral recombination graph (ARG)^{18,19}, has
66 richer information than the AFS since it includes additional data not reflected in the allele
67 frequencies, such as the correlated coalescent histories between segments of a chromosome.
68 Therefore, a genealogy-based demographic inference method has the potential to leverage the

69 flexible topological structure of the ARG in distinguishing complex demographic histories,
70 especially those with multiple admixtures.

71

72 Here, we introduce a genealogical likelihood framework named gLike to compute the likelihood
73 of an observed genealogical tree under a parameterized demographic history. The intuition
74 behind gLike is that a genealogy in itself does not imply the assortment history of any of its
75 lineages (*i.e.* which set of discrete population memberships a particular lineage has traversed
76 over time), meaning that all possible cases have to be considered. Notably, this idea bears
77 similarity to the recently proposed “local ancestry path” problem by Pearson and Durbin²⁰, but
78 instead of inferring the population membership distribution of each individual node, gLike aims
79 to compute the total likelihood of all combinations. By defining a “state” as the population
80 memberships of all lineages existing at a specific time, possible movements between states
81 throughout the history can be summarized into a directed acyclic Graph of States (GOS). We
82 develop a full methodology for the GOS around three key problems: 1) constructing a minimal
83 GOS that contains all necessary states; 2) computing the conditional probabilities between
84 connected states with considerations of migrations, coalescences, and non-coalescences; and
85 3) propagating the marginal probabilities through the GOS to achieve the total likelihood of the
86 tree, which can then be combined across multiple independent trees across the genome. As a
87 general-purpose statistical framework and as a first step towards utilizing the information from
88 the entire ARG, gLike is applicable to a variety of demographic events – migrations, splits,
89 admixtures, and population size variations, providing tools for model selection and parameter
90 estimation.

91

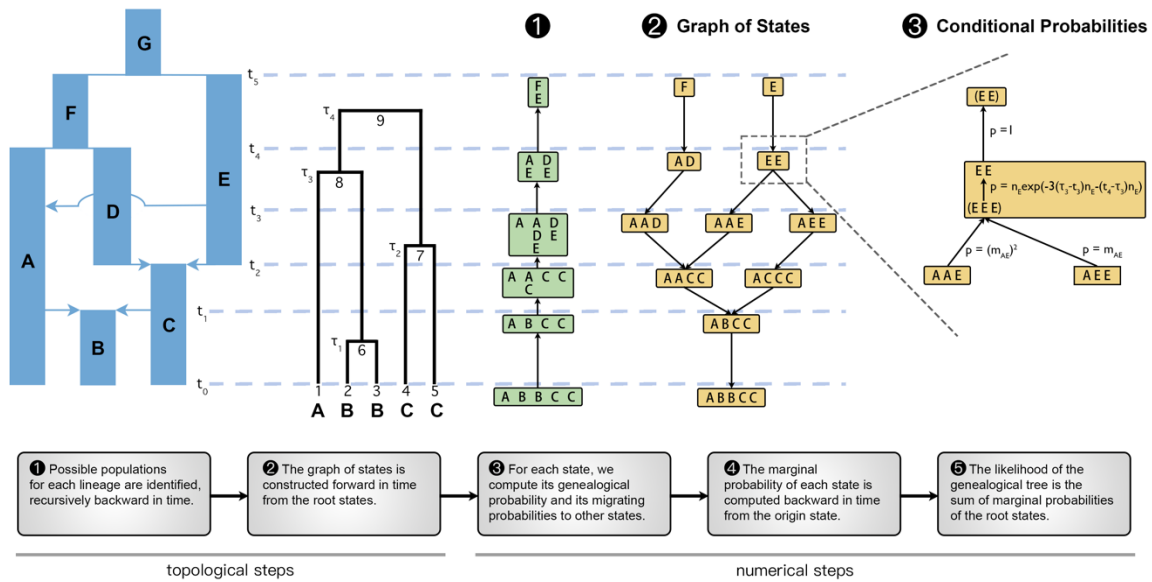
92 We demonstrate the advantage of genealogy-based demography inference by applying gLike to
93 simulated scenarios, with particular emphasis on complicated admixture histories such as three-
94 or four-way admixtures. gLike consistently outperforms existing AFS-based methods by
95 producing parameter estimates closer to the simulated truth. In analyses of genotyped samples
96 from Latino Americans and Native Hawaiians, the complex demography inferred by gLike is
97 consistent with the known history of both admixed populations and their ancestral populations –
98 Africans, Europeans, East Asians, Indigenous Americans, and Polynesians. Most notably, our
99 inference required no reference sample from the ancestral populations (such as samples from
100 Polynesians), nor explicit inference of local ancestries – information that is often not available or
101 is imprecisely estimated for understudied populations with complex history.

102

103 **Results**

104 **Method overview: genealogical likelihood under multi-population demography**

105 A genealogical tree, despite being a complete record of the coalescent events of the sample
 106 haplotypes within a chromosomal interval, does not specify the migration history of lineages. In
 107 a typical genetic study, the samples (leaf nodes) are collected from known populations, which
 108 serves as the initial condition. The internal lineages could migrate, subject to the restriction that
 109 coalescences must happen within a population. Therefore, the probability of a given
 110 genealogical tree corresponds to the cumulative total of all migration scenarios that are
 111 compatible with this tree. Our proposed method, gLike, computes the likelihood of any given
 112 genealogical tree under a hypothesized demographic history (**Methods**). Operationally, it is
 113 broken into two topological steps to search for possible population memberships of lineages,
 114 followed by three numerical steps to compute the conditional and marginal probabilities (**Figure**
 115 **1**).
 116



117
 118 **Fig 1. A schematic of the major steps of the gLike algorithm with examples.** Starting from a parameterized
 119 demography and an observed genealogical tree with known sample populations, the fundamental data structure in
 120 gLike is the graph of states that summarizes all possible scenarios for all lineages to move through the populations
 121 across history. We denote the unique state at time zero that contains the observable population memberships of
 122 samples as the “origin state” (state “ABBCC” in this example), and the states about the root of the genealogical tree
 123 as the “root states” (states “F” and “E” in this example). The graph of states is constructed in Step 2, guaranteed by a
 124 preparatory Step 1 such that no redundant states will be generated, minimizing computational burden. Each column
 125 represents the population membership (in Step 2; e.g. “AD” means that lineage 8 is in population A and lineage 7 is

126 in population D, t_4 generations ago) or the set of possible memberships (in Step 1; e.g. at t_4 , lineage 8 may be in A
127 or E, and lineage 7 may be in D or E,) of a certain lineage. In Step 3, the conditional probabilities are computed for
128 all states in the GOS except the origin states, including the coalescence and non-coalescence probabilities implied in
129 each state and the migration probabilities between connected states. Conditional probabilities are exemplified within
130 the fourth epoch (between t_3 and t_4) around the state “EE”. Specifically, “EE” implies a unique hidden state “EEE”
131 near the t_3 end of the epoch because lineages 1 and 6 should both be in population E in order to coalesce into
132 lineage 8, which is in E given the state “EE.” The connection between “EE” and “EEE” is represented by the
133 “genealogical probability,” which consists of the probability that lineages 1, 6 and 7 did not coalesce before τ_3 (with
134 probability $\exp(-3(\tau_3 - t_3)n_E)$), that lineages 1 and 6 coalesced at τ_3 (with probability n_E), and that lineages 7 and
135 8 did not coalesce before t_4 (with probability $\exp(-(t_4 - \tau)n_E)$). The state “EE” has two child states, “AAE” and
136 “AEE,” according to Step 2, connected via the intermediate state “EEE”. The transition from “AAE” to “EEE”
137 requires two lineage migrations from “A” to “E,” which occurs with “migration probability” m_{AE}^2 . Similarly,
138 transition from “AEE” to “EEE” occurs with probability m_{AE} . In Step 4, the “marginal probability” of a state is
139 defined as the probability conditional on the origin state and is computed recursively. For state “EE”, $p(\text{state EE}) =$
140 $n_E \exp(-3(\tau - t_3)n_E - (t_4 - \tau)n_E) ((m_{AE})^2 p(\text{state AAE}) + m_{AE} p(\text{state AEE}))$. The marginal probabilities are
141 propagated backward in time until the root states, and the log likelihood of the genealogical tree (conditional on the
142 hypothesized demography) is, in step 5, the sum of all root states: $p(\text{tree}) = p(\text{state F}) + p(\text{state E})$.

143
144 We define a “state” as a specification of the population memberships of all lineages existing at a
145 specific time. All possible states before each historical event (occurring at t_1, t_2, \dots, t_5 in this
146 example) form a directed acyclic graph (step 2, **Figure 1**), which we call the “graph of states
147 (GOS)”, a complete representation of all possible migration scenarios. When a state specifies a
148 lineage in an impossible population, it becomes a dead-end state that does not connect to the
149 origin. For example, in step 2, if we imagine a state “AA” at t_4 as a child of “F”, it will not connect
150 to the origin state “ABBCC”, because the fourth and fifth samples cannot migrate from C to A
151 per the hypothesized demographic model (**Figure 1**; see also **Figure S1**). To reduce
152 computation time, we avoid generating any dead-end state by a preliminary step (step 1, **Figure**
153 **1**) that summarizes possible population memberships for each lineage. For example, in step 1
154 at t_4 , lineage 8 may be in “A” or “E”, and lineage 7 may be in “D” or “E”, thus “AA” is not a legal
155 state in step 2 (**Figure 1**). The graph of states is then constructed from the root states (“F” or “E”
156 in this example) forward in time, by searching for child states according to both the specified
157 migration events in the demography and the results in step 1. See **Figure S1** for intermediate
158 results and further operational details during these two steps.

159

160 After building the GOS, the relevant conditional probabilities are computed. Because lineages
161 are restricted to their respective population until a historical event, a state immediately before a
162 historical event t_s is sufficient to specify the population memberships of all lineages between
163 t_{s-1} and t_s . For example, the state “EE” implies that not only the two lineages, but also the
164 subtrees under both lineages are all in population E between t_3 and t_4 . Given memberships of
165 all lineages within the context of a state, we can compute the “genealogical probability” of the
166 state based on standard coalescent theory to describe the coalescence (or lack thereof) events
167 during the relevant interval on the tree. We also compute the “migration probability” between a
168 state and its child, which is the product of the migration probability of each lineage, according to
169 the migration matrix of the historical event (step 3, **Figure 1**). The “marginal probability” of a
170 state is then the probability conditional on the origin state and can be computed recursively
171 (step 4, **Figure 1**). Finally, we compute the likelihood of the genealogical tree as the sum of the
172 marginal probabilities of the root states (step 5, **Figure 1**). See **Figure 1** legend for more
173 explanation of genealogical, migration, marginal, and total probabilities related to the state “EE”
174 in steps 3-5.

175

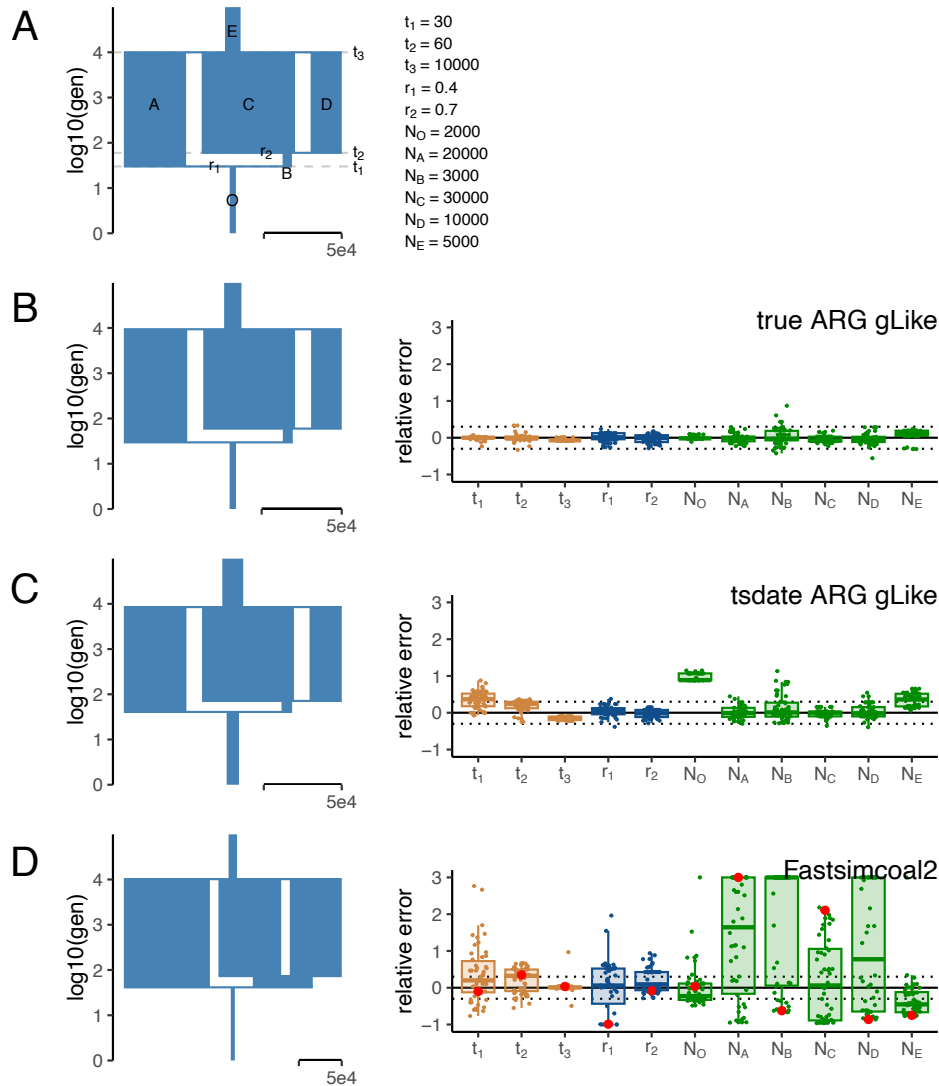
176 In practice, we apply gLike to a subsample of trees that are presumed independent, ideally from
177 evolutionarily neutral sites distantly spaced across the genome (usually 10-100, depending on
178 the computational resources), and the total likelihood is computed as the product over each
179 individual tree. The total likelihood as a function of the demographic parameters is then
180 optimized by simulated annealing. The final estimation of parameters is averaged over a
181 number of subsamples with replacement. The variance across subsamples serves as an
182 indicator of the uncertainty of the estimate.

183

184 **gLike accurately estimates all parameters in a three-way admixture demography**

185 Admixed populations, especially those with three ancestral components or more, pose
186 challenges to existing demographic inference methods. To showcase the performance of gLike
187 to analyze complex admixture, we simulated 1000 haplotypes on a 30Mb chromosome from a
188 population formed by two consecutive recent admixture events from three ancestral populations.
189 Such a demography is parameterized by 3 event times, 2 admixture proportions, and 7
190 population sizes, totaling up to 11 parameters (**Figure 2A**). When true genealogical trees were
191 available, the maximum likelihood estimates from gLike, averaged over 50 independent
192 simulations, for all 11 parameters achieved an overall 3.8% relative error (**Figure 2B**), while
193 gLike on the tsdate-reconstructed trees achieved an overall 23.3% relative error (**Figure 2C**).

194 We found that t_1 and N_O are the most overestimated parameters (by 35.6% and 97.3%,
 195 respectively) when using tsdate-reconstructed trees, likely due to tsdate's tendency to
 196 overestimate times of recent coalescences, prolonging the recent branches (**Figure S2**). Apart
 197 from t_1 and N_O , the other 9 parameters are estimated with 13.7% relative error.
 198



199
 200 **Fig 2. gLike accurately reconstructs three-way admixture without ancestral population samples.** (A) The true
 201 demography under which the genealogical trees and genotypes were simulated, with 6 populations involved:
 202 population O is admixed from A and B; B is the intermediate population admixed from C and D, where C is defined
 203 to be the major ancestor (proportion ≥ 0.5) without loss of generalizability; E is the ancestor of A, C and D. All
 204 population sizes are to scale. There are 11 parameters involved, including 6 population sizes and: t_1 , time of
 205 admixture of population O; t_2 , time of admixture of population B; t_3 , time of split from population E; r_1 , admixture
 206 proportion of A in O; r_2 , admixture proportion of C in B. The true value of each parameter is provided on the right.

207 (B-D) The reconstructed demography using parameter estimates averaged over 50 independent simulations (left) and
208 boxplots of relative errors $((\text{estimated}-\text{true})/\text{true})$ in each simulation (right). Boxplots are capped at 300% relative
209 error for ease of visualization. Trees and genotypes of 1000 haplotypes drawn from population O were simulated on
210 a 30 Mb chromosome. The demographic parameters were estimated by gLike on the true trees (B), by gLike on the
211 tsinfer+tsdate reconstructed trees from the true genotypes (C), and by Fastsimcoal2 on the allele frequency spectra
212 derived from true genotypes (D). For Fastsimcoal2 results, the parameter estimates for the single run with the
213 highest likelihood out of 50 independent runs, a practice commonly adopted by Fastsimcoal2, are labeled in red. A
214 reference for the width of the population sizes equivalent to 50,000 is given in each panel.

215

216 We also tested Fastsimcoal2 (ref.¹¹), which is capable of flexibly inferring complex demography
217 using allele frequency spectra. Based on true genotypes and the same three-way admixture
218 model, Fastsimcoal2 estimates had a relative error of 51.4%, which led to a visually distorted
219 demography (**Figure 2D**). This is in sharp contrast to Fastsimcoal2 showing comparable
220 accuracy to gLike on a three-population split demography (**Figure S3**). gLike also outperformed
221 a Generative Adversarial Network (GAN)-based deep learning approach, pg-gan²¹, which was
222 designed to overcome the limitations of relying on summary statistics such as the frequency
223 spectrum. In our benchmarking, pg-gan performed well for a two-population split demography
224 but was less accurate compared with Fastsimcoal2 and gLike on the three-population split and
225 admixture demographies (**Figure S4** and data not shown). We thus did not test pg-gan further in
226 this study. Nevertheless, our experiments with pg-gan were conducted without specialized
227 neural network hardware and do not dismiss GANs' potential as an emerging approach. Further
228 training and improved procedures may enhance GAN-based demographic inference²².

229

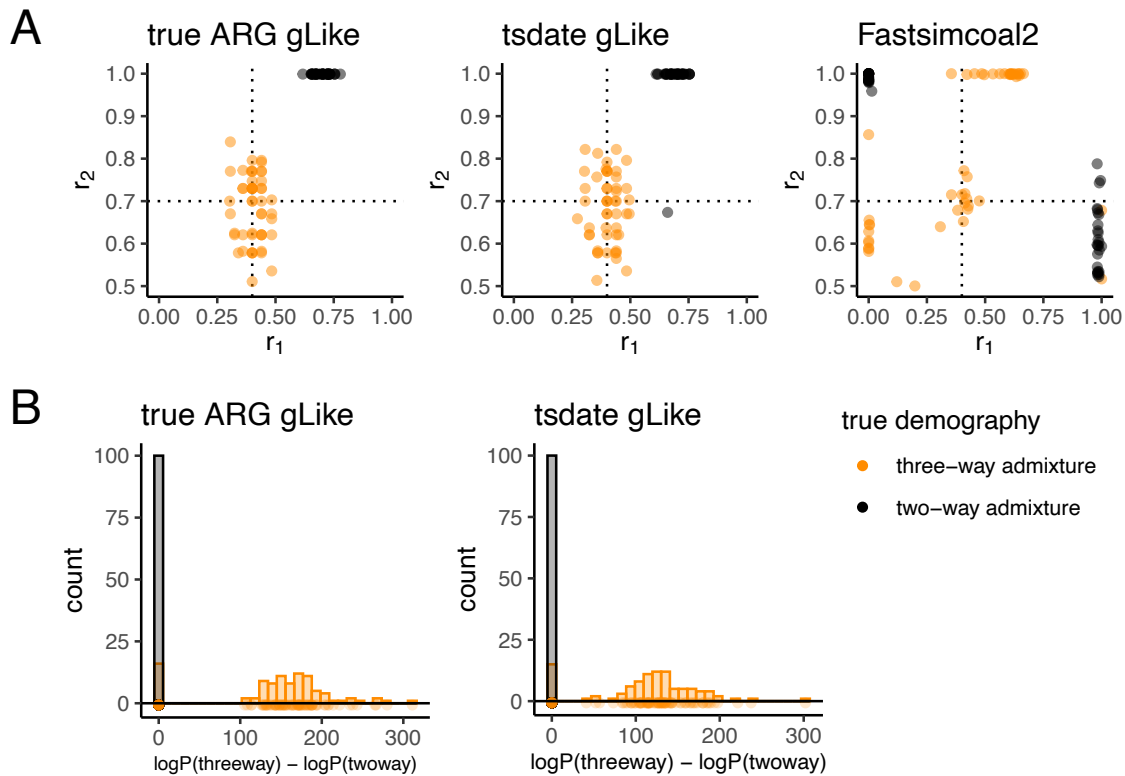
230 We find that in our application with gLike for the demographies we have studied, analyses using
231 tsinfer+tsdate-estimated genealogical trees produced more accurate estimated demographies
232 those using trees estimated by Relate. The difference in performance may trace to the fact that
233 recent coalescence times are overestimated by Relate to a greater extent than by tsdate,
234 causing a 20~50% depletion of coalescences within the recent dozens of generations (**Figure**
235 **S2A**), thereby leading to mis-estimations in the gLike framework. As a result, gLike on Relate-
236 reconstructed trees was not tested further in this study. Notably, Relate is more accurate in
237 estimating the ancient part of the ARG, including the tree-wise times to the most recent common
238 ancestor (tMRCAs) than tsinfer+tsdate (**Figure S2B**), which explains why in other applications
239 utilizing the genealogical trees, such as inferring the genome-wide expected relationship
240 matrix¹⁷ (eGRM), Relate may outperform tsdate.

241

242 **gLike detects components of admixture with high confidence**

243 We examined the ability of gLike to distinguish two-way from three-way admixtures. We expect
244 that the estimated parameters should reduce a complex model into a simpler one if the simpler
245 model is closer to the true underlying model. Conversely, the likelihood should increase
246 substantially when switching from a simple model to a complex one if the complex model is
247 closer to the true underlying model. We first applied gLike under a hypothesized three-way
248 admixture model to simulated trees and observed the estimated admixture proportions, r_1 and r_2
249 (**Figure 3A**, left and middle panels). Across 50 replicate simulations, when the true demography
250 was a three-way admixture, the estimated admixture proportion for the third ancestry
251 component, r_2 , centered around the true value (0.7) and was always far from the boundaries
252 (0.5 and 1.0). When the true demography was a two-way admixture, the estimated r_2 was
253 almost always 1.0, with only one exception (**Figure 3A**). This indicates that gLike correctly
254 reduced a three-way admixture model into a two-way model if the truth were indeed two-way
255 admixed. In contrast, both r_1 and r_2 were estimated to be the boundary values around half of the
256 time by Fastsimcoal2, regardless of the true demography (**Figure 3A**, right panel).

257
258 We next evaluated the maximum likelihood achieved under a two-way admixture model and a
259 three-way admixture model (**Methods**). AIC model selection was applied on the log-likelihood
260 differences between two models to select the more likely model between the two-way and three-
261 way admixtures. Across 100 independent simulations, the three-way admixture model was
262 never preferred when the true admixture was two-way, and the three-way admixture model was
263 preferred over two-way when it was the true model ~85% of the time with both true ARGs and
264 tsdate-reconstructed ARGs, resulting in a ~92% accuracy of classification.
265



266

267 **Fig 3. gLike distinguishes three-way admixture from two-way admixture.** True (left) and tsinfer+tsdate
268 reconstructed (middle) trees were obtained from simulated three-way (orange, same model as **Figure 2**) and two-
269 way (grey, r_2 was set to 1, removing contribution from population D) admixed populations. (A) gLike was applied
270 assuming a three-way admixture model. The estimated r_1 and r_2 values in each of 50 independent simulations are
271 shown, dashed lines denote true values of r_1 and r_2 in three-way admixture simulations. (B) gLike was first applied
272 under a two-way admixture model, then the model is expanded into a three-way admixture and gLike likelihood is
273 optimized while fixing shared parameters between two models (see **Methods** for technical details). The distributions
274 of log likelihood improvement after model expansion are shown as histogram. Model selection through the Akaike
275 information criterion (AIC) resulted in a classification accuracy of 92%.

276

277 **gLike reproduces complex demographic histories from stdpopsim**

278 Having established that gLike sensitively detects components of admixtures and estimates
279 parameters with high accuracy, we further evaluate its ability to reconstruct two additional
280 demographic models with increasing complexity, as published in stdpopsim²³ – the American
281 Admixture (stdpopsim model 4B11; **Figure 4**) and the Ancient Europe (stdpopsim model 2A21;
282 **Figure 5**) demographies.

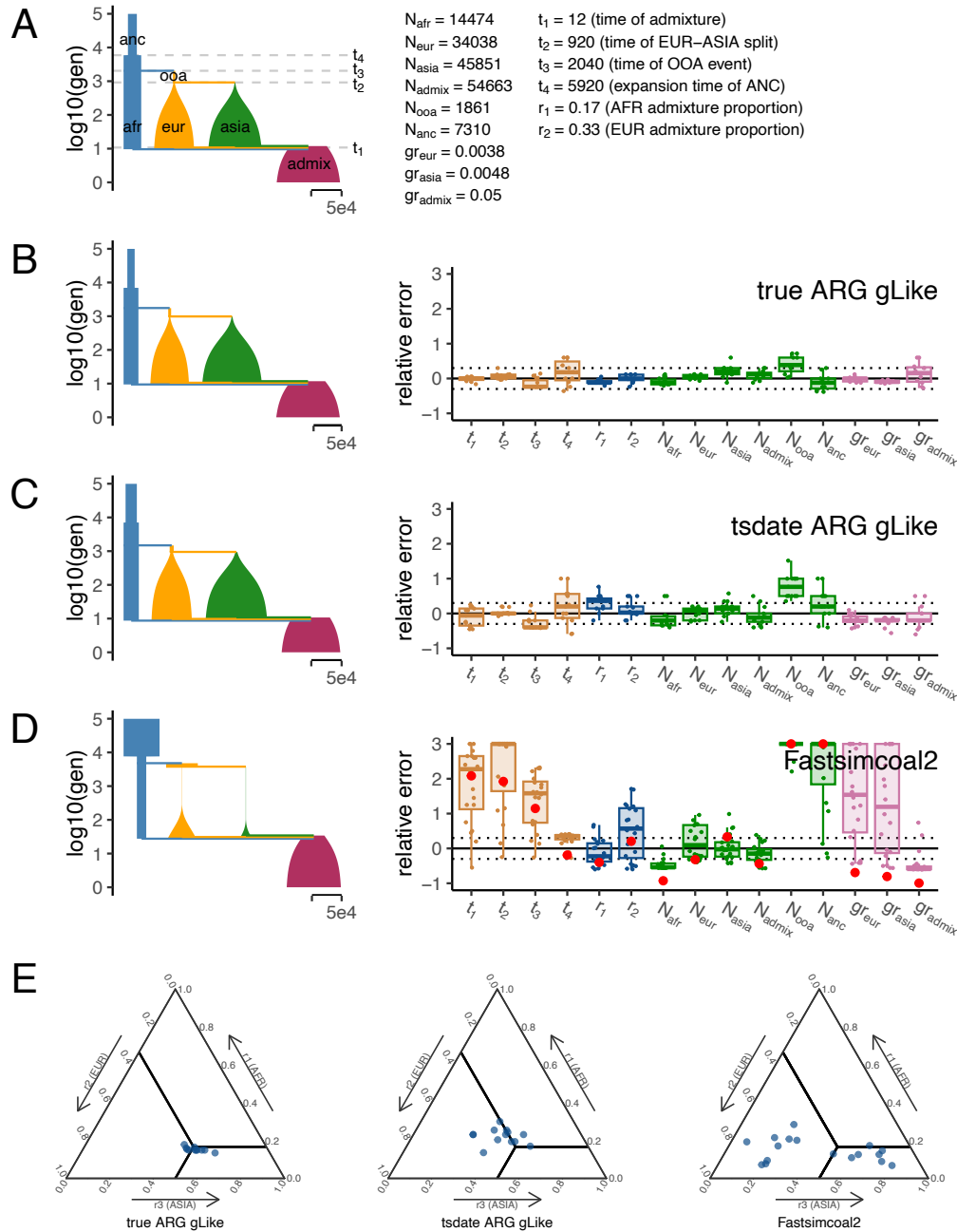
283

284 The American Admixture model consists of four populations (AFR, EUR, ASIA and ADMIX;
285 **Figure 4A**), where ADMIX is formed by a very recent admixture from the other three
286 populations. This model has 15 parameters, including 4 event times, 2 admixture proportions, 6
287 population sizes and 3 exponential growth rates. We simulate 1000 haplotypes from population
288 ADMIX on a 30Mb chromosome. gLike on the true trees inferred all 15 parameters with overall
289 11.3% relative error (**Figure 4B**). The majority of the error was in N_{ooa} , the size of the out-of-
290 Africa predecessor of the European population, which was overestimated by 38.5%. gLike on
291 the tsdate-reconstructed trees inferred parameters with overall 23.5% relative error (**Figure 4C**).
292 Except from the overestimation of N_{ooa} by 77.8%, the error concentrated on the African branch.
293 For example, r_1 (the African admixture proportion) was overestimated by 30.2%, and N_{anc} was
294 overestimated by 27.1%. Fastsimcoal2, in comparison, estimated the same set of parameters
295 with 258.7% relative error (**Figure 4D**). Fastsimcoal2 estimated the African proportion fairly
296 accurately, but appears unable to distinguish between the European and Asian proportions
297 (**Figure 4E**).

298

299 As AFS-based methods presumably have better performance in the presence of a multi-
300 dimensional allele frequency spectrum, we compared gLike and Fastsimcoal2 in additional
301 simulations where 500 haplotypes from each ancestral population were sampled to supplement
302 the 1000 admixed samples (**Figure S5**). Presence of ancestry reference samples improved the
303 accuracy and consistency of Fastsimcoal2's estimation of almost all parameters (an average of
304 213.1% relative error), especially the admixture proportions. But gLike based on the true and
305 inferred trees (5.8% and 16.7% relative errors, respectively) was still more accurate in capturing
306 the histories of these populations (**Figure S5**).

307



308

309 **Fig 4. gLike reconstructs the American admixture demography**

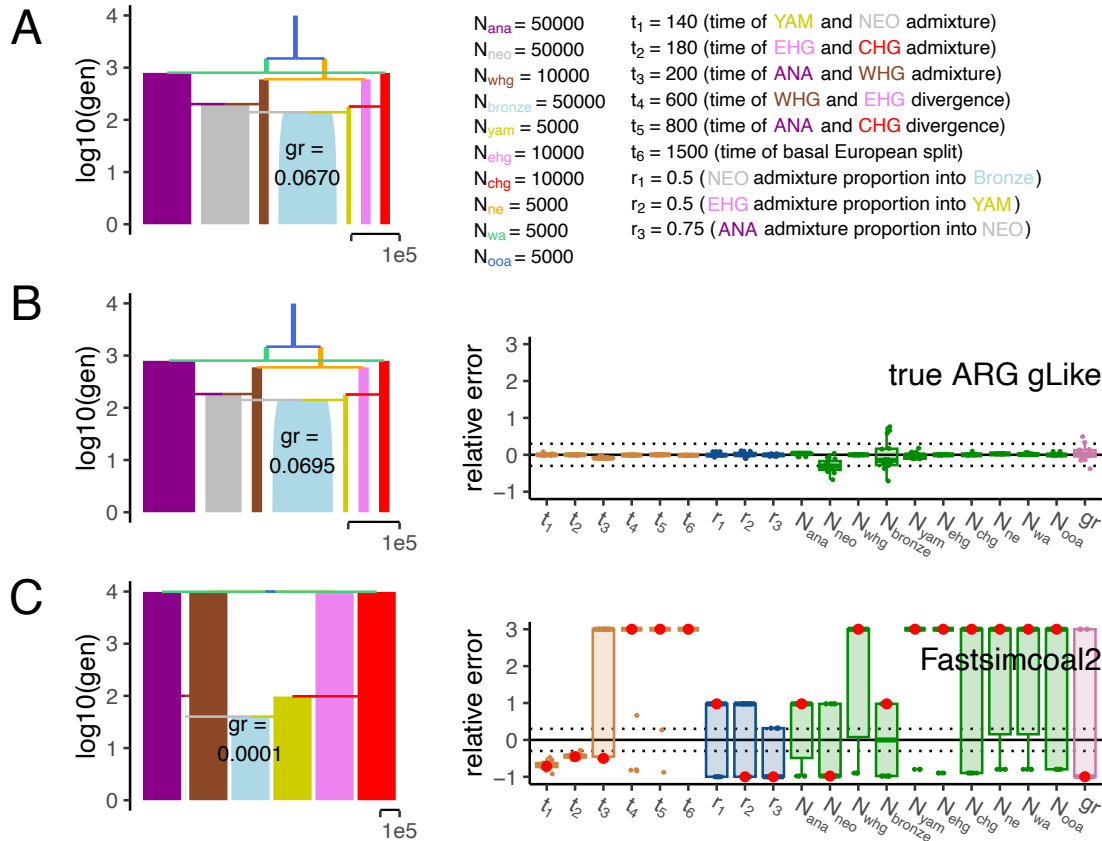
310 (A) American admixture demography with parameters from stdpopsim model 4B11. All population sizes are drawn
 311 to scale. (B-D) The reconstructed demography using estimations averaged over 50 replicate simulations (left) and
 312 boxplots of relative errors in each simulation (right). Trees and genotypes of 1,000 haplotype from the admixed
 313 population were simulated on a 30 Mb chromosome, the demographic parameters were estimated by gLike on the
 314 true trees (B) or the tsinfer+tsdate reconstructed trees (C), and by Fastsimcoal2 on the allele frequency spectra
 315 derived from true genotypes (D). Boxplots are capped at 300% relative error for ease of visualization. For
 316 Fastsimcoal2 results, the parameter estimates for the single run with the highest likelihood out of 50 independent

317 runs are labeled in red. A reference for the width of the population sizes equivalent to 50,000 is given in each panel.
318 (E) Ternary plots showing admixture proportions estimated by gLike on the true trees (left), by gLike on the
319 tsinfer+tsdate reconstructed trees (middle) or by Fastsimcoal2 on the allele frequency spectra of the true genotypes
320 (right), with slide lines indicating true parameters.

321

322 To test gLike's performance on intra-continental admixtures, we also evaluated the Ancient
323 Europe model from stdpopsim (2A21). This model is a four-way admixture model where the two
324 intermediate ancestors of Bronze Age population are each in turn admixed from two ancestors
325 (**Figure 5A**). We simulated 1000 haplotypes from the present-day population that descended
326 from the Bronze Age, and 200 from each of the ancient populations, according to the times
327 specified by stdpopsim. Applying gLike to the true trees resulted in estimates of the 20
328 parameters with overall 3.0% relative error (**Figure 5B**). The main misestimated parameter was
329 the 29.6% underestimation of N_{neo} , an ancient population that only existed for 20 generations
330 (180-200gen) when its samples were collected. Fastsimcoal2 estimated all parameters with an
331 average relative error of 132.3% (**Figure 5C**). The estimates of several population sizes reside
332 near the preset borders -- a behavior that has been suggested to be an intrinsic pitfall of AFS-
333 based methods²⁴. We did not test tsdate in this experiment because its ARG inference method
334 does not currently make full use of the ancient samples (instead, they are inserted as "proxy
335 sample ancestors" onto the existing ARG). Given our evaluation above, however, we would
336 expect that gLike will substantially improve over Fastsimcoal2 in accuracy of parameter
337 estimates if inferred ARGs can accurately incorporate ancient samples, and that gLike can
338 generally handle intra-continental admixtures when ancestral populations may be relatively
339 closely related.

340



341

342

343 Fig 5. gLike reconstructs the ancient Europe demography

344 (A) Ancient Europe demography with parameters from stdpopsim model 4A21. The Bronze Age population is
 345 plotted with initial size true to scale, but the growth rate is shown as text to avoid a disproportionate figure. All other
 346 population sizes are constant size and drawn to scale. (B, C) The reconstructed demography using estimates
 347 averaged over 50 replicate simulations (left) and boxplots of percentage errors in each simulation (right). Trees and
 348 genotypes were simulated on a 30 Mb chromosome. A total of 2200 haplotype samples (1000 contemporary samples
 349 descended directly from the Bronze Age population and 200 ancient samples each from the six ancient populations)
 350 were drawn at collection times as described by stdpopsim. The demographic parameters were estimated by gLike on
 351 the true trees (B) or by Fastsimcoal2 on the allele frequency spectra of the true genotypes (C). Boxplots are capped
 352 at 300% relative error for ease of visualization. For Fastsimcoal2 results, the parameter estimates for the single run
 353 with the highest likelihood out of 50 independent runs are labeled in red. A reference for the width of the population
 354 sizes equivalent to 10,000 is given in each panel.

355

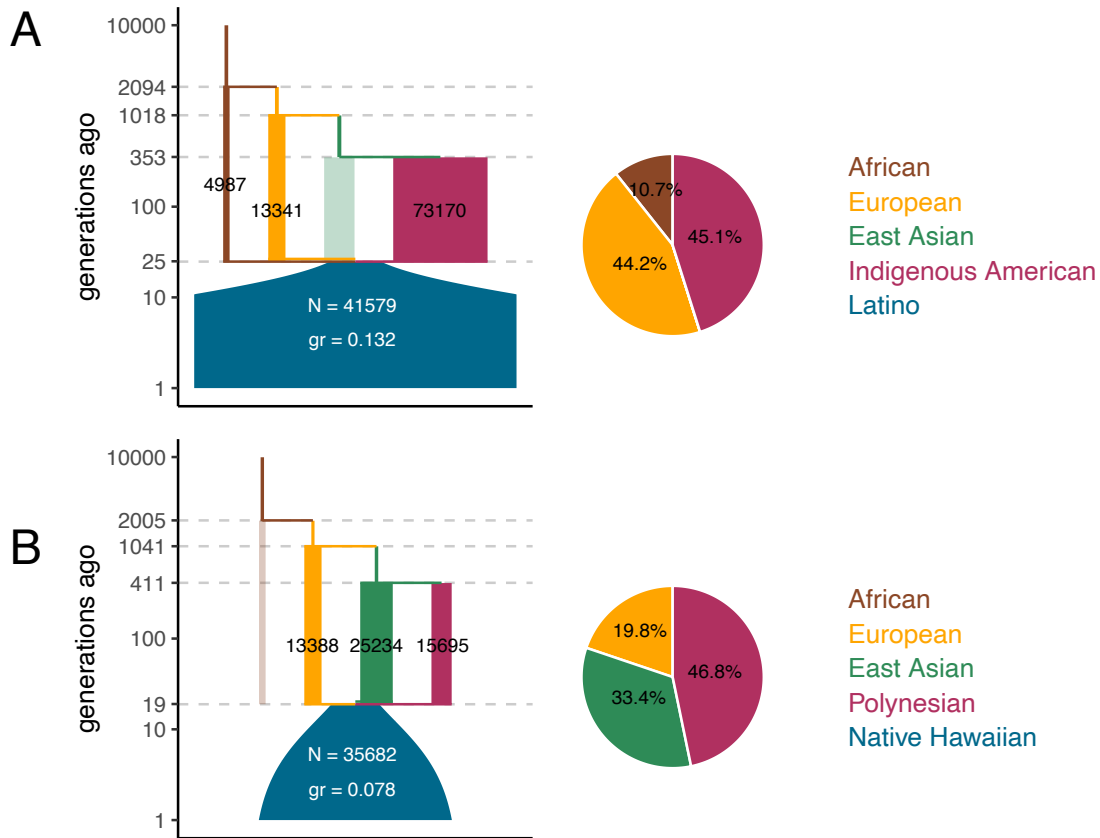
356 Inferring admixture history of Latinos and Native Hawaiians using genome-wide array

357 data

358 We applied gLike to investigate populations with complex demographic history using genome-
359 wide genotyping data from Latinos and Native Hawaiians, each with 500 subsampled diploid
360 individuals. We parameterized a four-way admixture model consisting of Africans, Europeans,
361 East Asians and a fourth ancestral population, which is used to model the Indigenous
362 Americans (for Latinos) or the Polynesians (for Native Hawaiians). We estimated genealogical
363 trees from the genotyping data using tsdate and estimated a total of 16 parameters using gLike
364 (**Figure 6; Supplemental Table 1**). We estimated the Latino lineages to be 10.7% from
365 Africans, 44.2% from Europeans, 45.1% from Indigenous Americans, and 0% (across all 20
366 independent threads) from East Asians, while the Native Hawaiian lineages were 19.8% from
367 Europeans, 33.4% from East Asians, 46.8% from Polynesians, and 0% (across all 20
368 independent threads) from Africans (**Figure 6**). As expected, we estimated the Native
369 Hawaiians to be more recently admixed than the Latinos (19 compared to 25 generations ago).
370 Also, the Native Hawaiians had a slightly smaller initial population size than the Latinos
371 ($35,682 \pm 10,656$ compared to $41,579 \pm 16,851$; but both are likely overestimated. See
372 **Discussion**) and grew at a slower rate (0.078 ± 0.009 compared to 0.132 ± 0.012) since the
373 admixture.

374
375 The European ancestries participated in both admixtures. As expected, we found the estimates
376 of its population size ($13,388 \pm 2,388$ and $13,341 \pm 4,702$) and of time of divergence with the East
377 Asians ($1,018 \pm 172$ and $1,041 \pm 87$ generations ago) to be highly concordant between two data
378 sets, suggesting the same underlying population that colonized the Americas and Polynesia.
379 Note that this ancestry should be more appropriately interpreted as the ancestral population
380 responsible for the colonization, which is less genetically diverse than the entire European
381 continent currently or at the time. The Indigenous Americans and Polynesians, though
382 represented as the same component in the model, were estimated to have different sizes
383 ($73,170 \pm 28,939$ compared to $15,695 \pm 7,393$), which may reflect greater population sizes or more
384 extensive structure in the ancestors to the Latino samples than to the Native Hawaiian samples.
385 Considering the potential errors during the ARG-reconstruction process (as have been seen in
386 **Figures 2, 4 and S2**) and biases due to the lack of high-quality sequencing data for these two
387 admixed cohorts (**Table S2**), these estimates of the demographic parameters for both
388 populations should be taken with caution. Nevertheless, our results suggest that gLike is able to
389 qualitatively capture known features of the demographic history of Latinos and Native Hawaiians
390 without reference data from their ancestral populations, and the results stand to improve as
391 ARG-reconstruction approaches advance.

392



393

394 **Fig 6. parameter estimations for the demographic histories of Latinos and Native Hawaiians.**

395 gLike was applied under a potential four-way admixture model reminiscent of stdpopsim model 4B11 for both the
396 Latino (A) and Native Hawaiian (B) data. The four potential ancestral populations are African, European, East
397 Asian, and Indigenous American (for Latinos) and Polynesian (for Native Hawaiians). The reconstructed
398 demographic diagrams are to scale, marked with relevant parameters. N, size of the admixed population in diploids
399 at time of admixture; gr, growth rate of the admixed population. Ancestral populations estimated to have 0%
400 admixture proportion are shown as translucent, because their sizes cannot be estimated. Pie charts show the
401 estimated admixture proportions of ancestral populations.

402

403 **Discussion**

404 With the fast development of scalable ARG inference over the past few years, development of
405 population-genetic approaches that explicitly use the ARG or its marginal trees is an exciting
406 area of active research. With this in mind, our current study introduced a framework that
407 explains the stochastic formation of the genealogical trees in a multi-population context, and
408 computes the full likelihood of each demographic scenario. Our results revealed that the history
409 of at least three ancestral populations can be clearly decoded from the genealogical trees of a
410 single admixed sample without knowledge of the ancestral populations. For many understudied
411 diverse populations across the world, it is often unclear whether they are admixed, and if so,
412 what the ancestral populations were. Even if the ancestral populations are known or can be
413 hypothesized, they likely no longer exist or are difficult to sample. For these populations,
414 demographic inference using allele frequencies is difficult, since distinct demographic scenarios
415 can give similar AFSs¹². gLike has the potential to provide new insights into studies of these
416 understudied or ancient populations, as well as the demographic history of other species.

417

418 It is worth clarifying that the admixture proportions in the demographic context (such as those
419 estimated by Fastsimcoal2¹¹ and gLike here) have a slightly different meaning from that in the
420 genomic context (such as those estimated by STRUCTURE²⁵ and ADMIXTURE²⁶). As a
421 demographic parameter, the admixture proportion describes the probability of a lineage to
422 migrate (backward in time) from one population to another, while in the genomic context, this
423 proportion describes how much of the genome one population shares with another. The two
424 concepts can deviate primarily in two cases: 1. There is considerable genetic drift after the
425 admixture, especially when the population size is small; 2. The admixed population, O, may
426 have a genetic component from, say, population A, not because A participated in the formation
427 of O, but because of more ancient migrations from A to other ancestries of O. In gLike results,
428 all admixture proportions should be interpreted in demographic context. In practice, admixture
429 proportions could be estimated through other means in the genomic context, and then be used
430 as the initial values for gLike to improve optimization speed and stability, while allowing gLike to
431 make further adjustments as needed.

432

433 We also note that currently gLike is not utilizing the full information encoded in an ARG, but
434 rather is relying on sets of presumed independent trees. In many ways, gLike was inspired by
435 HMM-based demographic inference²⁻⁶, where genealogical trees are implicitly utilized.
436 However, these methods are computationally intensive and have limited scalability, primarily

437 due to the intricate handling of recombination events. We reasoned that while recombination
438 events are essential for ARG inference, they are less informative for ARG-based demographic
439 inference. Once the ARG (and thus the genealogical trees within the ARG) has been accurately
440 inferred from the genotypes, reliance on recombination events for insights into demography
441 becomes less important. Recombination can be modeled as a random breakpoint in the
442 genealogical tree re-coalesced onto the rest of the tree – the random break is independent of
443 demography, and the re-coalescence holds minimal information compared to the numerous
444 coalescences already on the tree. In light of this, gLike currently focuses on rigorously modeling
445 lineage assortments and coalescent events within individually independent trees, rather than the
446 variability between neighboring trees, to achieve greater scalability (in order to handle
447 thousands of samples and multiple populations). Future enhancements of gLike may then
448 model recombination to incorporate the remaining information encoded in the ARG.
449 Furthermore, gLike has some commonality with approaches to species-tree inference based on
450 gene trees, where gene trees can be used to estimate the topology and branch lengths of a
451 phylogenetic tree²⁷. Whereas such methods estimate the whole topology, we pre-specify the
452 demographic history and estimate parameters related to it, including processes like admixture
453 that do not feature as prominently in species-tree inference. In cases where the demographic
454 history is sketchy, it may be possible to develop approaches akin to the species-tree inference
455 to estimate parts of the topology.

456
457 One current limitation of gLike is that certain parameters are not individually identifiable, but
458 could only be optimized in combination. For example, the effects of population size and growth
459 rate are hard to separate if a population exists for only a short time (**Figure S6**). Any
460 combination of the two parameters that produces the same average coalescence rate will have
461 a similar likelihood, making it difficult to identify the global optimum. Such entangled parameters
462 are in fact a limitation in many demographic inference methods and often result in similar
463 likelihoods for many combinations of parameters. When applying gLike with hill-climbing-based
464 optimization methods, the estimates of entangled parameters could be path dependent. Thus, a
465 grid search on specific entangled parameters after a general optimization routine may be
466 beneficial to an unbiased estimation of the demography.

467
468 In addition, continuous migration is not currently supported by gLike, because it drastically
469 increases the number of states. In the American Admixture simulations (**Figure 4**), we omitted
470 the weak migrations (10^{-5} - 10^{-4} per generation) between continental populations as originally

471 specified by the stdpopsim model. Omitting the continuous migrations have no visible impact on
472 estimating the remaining parameters unless they are ~100 times more intense than that
473 currently specified in the stdpopsim model and presumed to be typical between continental
474 human populations (**Figure S7**). However, such frequent migrations (10^{-3} - 10^{-2} per generation)
475 may exist between intra-continental populations where geographical separations are minimal.
476 Estimating the migration rate itself is also of interest in ecological studies of other species, and a
477 future focus will be extending gLike to incorporate continuous migration. One obvious solution is
478 to discretize the continuous migration into a number of pulse migrations, which results in many
479 layers each containing a large number of states. An effective discretization strategy, as well as
480 an efficient random sampling technique on the states, seems necessary to address this
481 challenge.

482
483 Current ARG inference methods have achieved remarkable scalability and accuracy, but their
484 biases and errors still deserve attention in genetic applications. We have showcased the varying
485 performance of tsinfer+tsdate and Relate at different time scales in admixed populations
486 (**Figure S2**). The overestimation of branch lengths at recent times appears to be a common
487 problem for both methods, but is more severe in Relate-inferred trees, to the degree that
488 meaningful GOSs are difficult to construct. Tsinfer and tsdate are also faster because they use
489 heuristic algorithms to avoid the $O(n^2)$ pair-wise comparisons. However, the bottom-up
490 approach of tsdate is somewhat less accurate for ancient coalescences, whereas Relate's
491 hierarchical clustering-based method infers the deep part of the genealogies with higher
492 accuracy (especially beyond 1000 generations ago), and thus captures global relatedness more
493 robustly¹⁷. There may be techniques to adjust one's result with the other, thus combining both of
494 their advantages. With scalable and accurate ARG inference across broader scales, we expect
495 the reliability and accuracy of gLike demographic inference to be further improved.

496
497 Finally, we acknowledge that human migrations and admixtures exist on a continuum. In the
498 current framework we opted to model discrete populations and components of ancestries, as is
499 customary when modeling the histories of recently admixed populations such as the Latinos.
500 But one of the advantages of an ARG-based view of human history may be to remove the
501 notion of distinct populations. Enabling continuous rather than pulse-like migrations between
502 populations to enhance gLike may be another step forward, but future developments of ARG-
503 based demographic inference may emphasize on the paradigm shift to represent human
504 histories and structure on a continuum.

505

506 **Methods**

507 **Formalization of the problem: Probability of a genealogical tree under a demography**

508 The demographic history of K populations can be represented by the interplay between two
509 stochastic processes affecting the lineages – coalescence and movement among populations.

510 The coalescence rate $n_a(t)$ of each population a as a function of time t is

$$511 \quad n_a(t) = \frac{1}{kN_a(t)}, \quad a \in \{1, \dots, K\}, \quad t \in (0, \infty),$$

512 where N_a is the effective population size, and k is ploidy. And the migration probability matrix m
513 at each of the S historical events is

$$514 \quad m_{ab}(t_s), \quad a, b \in \{1, \dots, K\}, \quad s \in \{1, \dots, S\},$$

515 where t_s is the time of the s -th historical event, and $m_{ab}(t_s)$ is the instantaneous probability for a
516 lineage to move (backward in time) from population a to b .

517

518 The demography is thus defined as

$$519 \quad \mathcal{D} = (n, m) = (\{n_a\}, \{m_{ab}\}),$$

520 a size- K vector of coalescence rates defined on continuous time, and a $K \times K$ matrix of
521 migration probabilities defined on a discrete set of times. While gLike currently does not
522 explicitly incorporate continuous migration, it can potentially be represented as a series of
523 historical events through discretization.

524

525 A genealogical tree with N nodes can be defined by the time and children of each node

$$526 \quad \mathcal{G} = \{(\tau_i, \pi_i) | i \in \{1, \dots, N\}\},$$

527 where τ_i is the time of the node i (or equivalently, the emergence of lineage i), and π_i is the set
528 of its child nodes (which is empty if i is a leaf node). The end time ω_i of lineage i can be
529 calculated as time of its parent node (that is, $\omega_i = \tau_j$ if $i \in \pi_j$) or ∞ if it has no parent. Our goal is
530 to compute $\mathbb{P}(\mathcal{G}|\mathcal{D})$ for arbitrary \mathcal{G} and \mathcal{D} , and we will omit thereafter the “conditional on \mathcal{D} ”
531 notation, which is always implied.

532

533 It is helpful to define the set of lineages existing at time t as

$$534 \quad L(t) = \{i | \tau_i \leq t < \omega_i\},$$

535 and the lineages emerging between t and t' as

$$536 \quad L(t, t') = \{i | t < \tau_i, \omega_i < t'\}.$$

537

538 Migration trajectory and states

539 The population identity of a lineage i during its existence,

$$540 \quad x_i(t), \quad t \in [\tau_i, \infty)$$

541 is a time-dependent variable taking values from $\{1, \dots, K\}$ that describes how this lineage, or its
542 ancestor lineage when $t > \omega_i$, migrates in history. For convenience, the value of $x_i(t)$ at exactly
543 the time of a historical event is defined as the left limit $x_i(t_s) = \lim_{t \rightarrow t_s^-} x(t)$, so that $x(t)$ is left-
544 continuous.

545

546 The population identity of all lineages existing at any time throughout the history is

$$547 \quad x(t) = \{x_i(t) | i \in L(t)\}, \quad t \in [0, \infty),$$

548 which gives a complete migration trajectory of the genealogical tree. The genealogical tree itself
549 does not dictate x , and the probability of it should be computed as the sum over all possible
550 trajectories,

$$551 \quad \mathbb{P}(\mathcal{G}) = \sum_x \mathbb{P}(\mathcal{G} \cap x).$$

552 In order to compute $\mathbb{P}(\mathcal{G})$ recursively over time, we define $\mathcal{G}(0, t)$ as the genealogical history in \mathcal{G}
553 until time t , and define a “state” as

$$554 \quad \mathcal{G}(0, t) \cap x(t).$$

555 For example, the state “ABCC” in **Figure 1** at t_1 contains $\mathcal{G}(0, t_1)$, which indicates that lineages
556 2 and 3 coalesced at τ_1 but all other possible coalesces has not happened at t_1 , and $x(t_1) =$
557 ABCC, which indicates that the remaining four lineages (1,6,4 and 5) are in populations A,B,C
558 and C, respectively, at t_1 .

559

560 Now $\mathbb{P}(\mathcal{G})$ can be expressed as the sum of probability of root states

$$561 \quad \mathbb{P}(\mathcal{G}) = \mathbb{P}(\mathcal{G}(0, \infty)) = \sum_{x(\infty)} \mathbb{P}(\mathcal{G}(0, \infty) \cap x(\infty))$$

562

563 Conditional probability between states

564 The conditional probability between states

$$565 \quad \mathbb{P}(\mathcal{G}(0, t_{s+1}) \cap x(t_{s+1}) | \mathcal{G}(0, t_s) \cap x(t_s)) \\ 566 \quad = \mathbb{P}(\mathcal{G}(0, t_s) \cap x(t_{s+1}) | \mathcal{G}(0, t_s) \cap x(t_s)) \mathbb{P}(\mathcal{G}(0, t_{s+1}) \cap x(t_{s+1}) | \mathcal{G}(0, t_s) \cap x(t_{s+1}))$$

567 consists of a migration probability and a genealogical probability.

568

569 The migration probability

570
$$\mathbb{P}(\mathcal{G}(0, t_s) \cap x(t_{s+1}) | \mathcal{G}(0, t_s) \cap x(t_s)) = \prod_{i \in L(t_s)} m_{x_i(t_s)x_i(t_{s+1})}(t_s)$$

571 describes the migration of each lineage i from $x_i(t_s)$ to $x_i(t_{s+1})$ at time t_s .

572

573 The genealogical probability $\mathbb{P}(\mathcal{G}(0, t_{s+1}) \cap x(t_{s+1}) | \mathcal{G}(0, t_s) \cap x(t_{s+1}))$ describes how likely the
 574 genealogical tree grows according to \mathcal{G} backward in time from t_s to t_{s+1} , given population
 575 identities $x(t_{s+1})$. This requires that every coalescence in \mathcal{G} happened exactly at its time in \mathcal{G}
 576 (which we call the coalescence probability) and that any other possible coalescence did not
 577 happen (which we call the non-coalescence probability).

578

579 The coalescence probability is

580
$$\prod_{i \in L(t_s, t_{s+1})} [n_{x_i(t_{s+1})}(\tau_i)]^{\max(0, |\pi_i| - 1)}$$

581 where $n_{x_i(t_{s+1})}(\tau_i)$ is the coalescence rate of lineage i 's population when it emerges. Note that
 582 the lack of migration between τ_i and t_{s+1} guarantees $x_i(\tau_i) = x_i(t_{s+1})$. And $\max(0, |\pi_i| - 1)$ is
 583 the number of coalescences at the emergence of i (for example, a binary node is formed with
 584 one coalescence, a ternary node can be viewed as two coalescences at the same moment, and
 585 a leaf node or unary node does not have coalescence).

586

587 The non-coalescence probability is

588
$$\prod_{a \in \{1, \dots, K\}} \exp\left(-\int_{t_s}^{t_{s+1}} \binom{l_a(t)}{2} \cdot n(t) dt\right)$$

589 where

590
$$l_a(t) = |\{i | i \in L(t), x_i(t_{s+1}) = a\}|$$

591 is the number of lineages in population a at time t (if population identities are specified by
 592 $x_i(t_{s+1})$), which is a step function that jumps when lineages emerge or coalesce; $\binom{l_a(t)}{2} =$

593 $\frac{l_a(t)(l_a(t)-1)}{2}$ is the number of lineage pairs in a that are possible to coalesce; and the exponential

594 term is the probability that none of them actually coalesced during (t_s, t_{s+1}) , which is derived

595 from a nonhomogeneous Poisson process with rate $\lambda(t) = \binom{l_a(t)}{2} \cdot n(t)$. Note that $n(t)$ can be

596 any integrable function, enabling flexibility to the population size variation in the demographic

597 model.

598

599 We conclude that the conditional probability between states is

$$\begin{aligned}
 & \mathbb{P}(\mathcal{G}(0, t_{s+1}) \cap x(t_{s+1}) | \mathcal{G}(0, t_s) \cap x(t_s)) \\
 &= \left(\prod_{i \in L(t_s)} m_{x_i(t_s)x_i(t_{s+1})}(t_s) \right) \cdot \left(\prod_{i \in L(t_s, t_{s+1})} [n_{x_i(t_{s+1})}(\tau_i)]^{\max(0, |\pi_i| - 1)} \right) \\
 & \quad \cdot \left(\prod_{a \in \{1, \dots, K\}} \exp\left(-\int_{t_s}^{t_{s+1}} \binom{l_a(t)}{2} \cdot n(t) dt\right) \right) \\
 &= (\text{migration probability}) \cdot (\text{coalescence probability}) \cdot (\text{noncoalescence probability}) \\
 &= (\text{migration probability}) \cdot (\text{genealogical probability})
 \end{aligned}$$

603
604
605
606 Practically, the migration probability has to be computed between any parent-child state pair,
607 but the genealogical probability is independent from the child state and needs to be calculated
608 only once for every state. As a boundary condition, the origin state at the bottom (*i.e.* leaves) of
609 the tree has probability one

$$610 \quad \mathbb{P}(\mathcal{G}(0, 0) \cap x(0)) = \mathbb{P}(x(0)) = 1,$$

611 where $x(0)$ specifies the population identities of each individual in the study samples.

612

613 **The minimal graph of states**

614 All possible states at all times of all historical events t_1, t_2, \dots, t_s form a directed acyclic graph,
615 named as the graph of states (GOS), where states in adjacent layers (one at t_s and the other at
616 t_{s+1}) are connected with their conditional probability as introduced above. A state with zero
617 marginal probability will not contribute to the marginal probability of its parent state and is
618 redundant in the graph. A GOS without redundant states is called a minimal GOS.

619

620 The coalescence probability and non-coalescence probability are always above zero, because
621 population sizes cannot be zero or infinity. This means that, to judge if a state is possible or not,
622 we only have to check the migration probabilities, which are decomposable into migrations of
623 each individual lineage. In other words, a state is possible if every lineage is in a possible
624 population. To put it mathematically, we have

$$625 \quad \mathbb{P}(\mathcal{G}(0, t_s) \cap x(t_s)) > 0 \Leftrightarrow \left[I(x_i(0)) \prod_{1 \leq r \leq s} m(t_r) \right]_{x_i(t_s)} > 0, \quad \forall i \in L(0)$$

626 where $I(x_i(0))$ is a size- K indicator vector with value 1 at the population $x_i(0)$ where sample i
627 was collected, and all other elements zero; $\prod_{1 \leq r \leq s} m(t_r)$ is the transition matrix summarizing the

628 first s historical events; and $[I(x_i(0)) \prod_{1 \leq r \leq s} m(t_r)]_{x_i(t_s)}$ is the probability that lineage i migrated
629 from $x_i(0)$ to $x_i(t_s)$. **Figure 1** step 1 can be understood as the non-zero elements in
630 $I(x_i(0)) \prod_{1 \leq r \leq s} m(t_r)$ for every s .

631

632 **Implementation details and optimization**

633 With the above-mentioned theory to calculate $\mathbb{P}(\mathcal{G}|\mathcal{D}_\theta)$ on a demographic model \mathcal{D}_θ
634 parameterized by θ , the estimated parameters that best explains the observed \mathcal{G} is

$$635 \theta^* = \operatorname{argmax}_{\theta} \mathbb{P}(\mathcal{G}|\mathcal{D}_\theta)$$

636 gLike encapsulates the likelihood computation and a simulated annealing-based optimization
637 into an open-source Python package, alongside a C extension to accelerate Cartesian product
638 operations when searching for child states (GitHub page: <https://github.com/Ephraim-usc/glike>).
639 All probabilities are implemented in log scale, and sums of probabilities are calculated with the
640 `scipy logsumexp` function. If the number of states at a layer exceeds the preset limit (10^5 by
641 default), a random subsample of states is generated to approximate the likelihood. When
642 multiple, presumed independent and neutrally evolving, trees are provided, the final log
643 likelihood is the sum of log likelihoods of each tree. We presume independence of trees as the
644 total likelihood would assume more complicated forms if trees were nearby and not
645 independent. We also presume neutrality as coalescence probabilities would deviate from the
646 inverse of population sizes when there are variants under natural selection. We set a user-
647 defined parameter to drop some proportion (default: 50%) the lowest likelihood trees during
648 optimization, as we found in practice that this filtering improves robustness against errors in tree
649 reconstruction (such as erroneous coalescences) and migrations that are neglected in the
650 demographic model.

651

652 **Demographic inference in simulations**

653 All simulations were performed on a 30 Mb chromosome with both recombination and mutation
654 rates set to 10^{-8} per generation per base pair, with a sample size of 1,000 haplotypes from the
655 admixed population. The demographic parameters are annotated in the corresponding figures,
656 or cloned from `stdpopsim`²³ models 4B11 (American Admixture) and 4A21 (Ancient Europe). In
657 American Admixture simulations, we ignored the continuous migrations in our simulations and
658 estimations. The extent to which hidden migrations potentially undermines gLike results was
659 tested on additional simulations with 1-, 10- and 100-times continuous migrations as reported by
660 `stdpopsim` 4B11. In the Ancient Europe simulation, we additionally sampled 200 haplotypes

661 from each ancestral population according to the collection times reported by stdpopsim, in order
662 to mimic genetic studies with ancient DNA.

663

664 To evaluate gLike, ARGs and genotypes were simulated by msprime²⁸. ARG reconstructions by
665 tsinfer+tsdate^{29,30} or Relate³¹ were performed with all default parameters as suggested in the
666 user manual. One hundred evenly spaced trees across the chromosome were selected for
667 gLike inference. The precision of gLike parameter estimation (i.e., the minimal step size during
668 optimization by simulated annealing, relative to the current estimate) was set to 2%. The
669 absolute difference between the average estimate and the truth, divided by truth, is defined as
670 the relative error. The average estimates across 50 replicate simulations were used as the final
671 pictorial representation of the reconstructed demography, with boxplots of the relative errors
672 across 50 replicates also shown. The standard deviation across 50 replicate simulations serves
673 as an indicator of the parameter uncertainties as listed in **Tables S1** and **S2**.

674

675 To compare gLike to Fastsimcoal2 (ref¹¹), derived allele frequency spectra were computed on
676 all simulated SNPs (including singletons), and parameter estimation was performed with
677 100,000 simulations and 40 ECM (expectation/conditional-maximization) loops, using the
678 commands “-n 1 -s0 -d -k 1000000” for AFS simulation and “-n 100000 -s0 -d -M -L 40” for
679 parameter estimation. The estimate with the highest likelihood obtained among 50 independent
680 runs was used as the final pictorial representation of the reconstructed demography (following
681 the same practice recommended by the authors of Fastsimcoal2³²), with estimates from all 50
682 shown in the accompanying boxplots. We also compared gLike performance to pg-gan²¹, a
683 deep learning demographic parameter inference method that uses generative adversarial
684 networks to create realistic simulated training data. Genotypes from simulated ARGs of the
685 same demographic model were used as training data, run for up to 300 training iterations with
686 default training parameters. We also used the same range for each demographic parameter to
687 be consistent with the Fastsimcoal2 comparisons. Since pg-gan gives multiple sets of
688 parameter proposals at end of training, the set of inferred demographic parameters with the
689 lowest relative error compared to the true parameters was selected as the final estimate of this
690 run. A total of 50 independent runs were conducted.

691

692 To characterize the impact of ARG reconstruction using array data instead of sequencing data,
693 we performed additional simulation experiment in which SNPs were retained with the probability

694

$$p(\text{MAF}) = C_{\text{ref}}(\text{MAF})/C_{\text{sim}}(\text{MAF}),$$

695 where MAF is the minor allele frequency of the simulated SNP, $C_{\text{ref}}(\text{MAF})$ is the number of
696 occurrences of MAF in the Latinos array data, and C_{sim} is the number of occurrences of MAF in
697 a simulated genome (3,000Mb). As expected, it was found that C_{sim} is greater than C_{ref} across
698 all values of $\text{MAF} \in [0, 0.5]$, which ensures p is always less than one. We then inferred the ARG
699 using tsinfer+tsdate using the simulated array data.

700

701 **Model selection in simulations**

702 To test for the existence of an additional ancestral component, gLike was applied under a two-
703 way admixture model and a three-way admixture model, and the maximum likelihoods achieved
704 under both models were compared. Specifically, the two-way admixture model structurally
705 mimicked the three-way admixture as in **Figure 2A**, but without population D, so that all
706 lineages from population B entered population C. As such, the two-way admixture model had
707 two fewer parameters – r_2 (admixture proportion from D) and N_D (population size of D). gLike
708 was then applied in a two-step manner. First, the parameters were estimated under the two-way
709 admixture model with the default hill-climbing optimization. Next, we applied gLike under the
710 three-way admixture model and perform a grid search on r_2 , N_C and N_D , while fixing other
711 parameters at their two-way admixture estimates. Finally, the difference between the maximum
712 log likelihoods achieved under two models was used for AIC model selection (with 2 degrees of
713 freedom, to account for the two extra parameters in the three-way admixture model), and the
714 model with the higher AIC value was selected.

715

716 **Latinos and Native Hawaiians data processing**

717 A total of 5,382 self-identified Native Hawaiians and 3,659 self-identified Latinos from the
718 Multiethnic Cohort (MEC) were genotyped on two separate GWAS arrays: Illumina MEGA and
719 Illumina Global Diversity Array (GDA). After taking the intersection of SNPs found on both
720 arrays, the genotyping data were lifted to hg38 using *triple-liftover*³³ to ensure alleles in inverted
721 sequences between reference genome builds were properly lifted. We removed variants that
722 were genotyped in fewer than 95% of individuals, variants out of Hardy-Weinberg Equilibrium (p
723 $< 10^{-6}$), and individuals with greater than 2% missing genotypes (though no one was removed
724 with this threshold). After quality check, the Native Hawaiian and Latino datasets contained
725 990,549 and 1,093,693 SNPs, respectively. The data were phased without a reference using
726 EAGLE³⁴ and its default hg38 genetic map. We randomly subsampled 1,000 haploids and
727 removed monomorphic SNPs, resulting in 879,040 and 927,254 SNPs in the Native Hawaiian
728 and Latinos datasets, respectively. The ancestral alleles were called by a comparison with the

729 human ancestor GRCh38 e107 genome (URL: <ftp.ensembl.org/pub/release->
730 [86/fasta/ancestral_alleles/](ftp.ensembl.org/pub/release-86/fasta/ancestral_alleles/)). Tsinfer and tsdate were used with all default parameters as
731 suggested in the user manual to reconstruct the ARG. The human neutralome³⁵ (*i.e.*, the
732 regions of the human genome identified as likely selectively neutral) was converted into hg38
733 coordinates, and 319 neutral regions that are at least 5Mb from each other were selected for
734 gLike analysis. Ten trees were sampled in each gLike optimization thread, and 20 threads were
735 run in parallel. The estimates of demographic parameters were averaged over 20 threads. The
736 precision of gLike parameter estimation was set to 5%, higher than 2% used in simulations. This
737 choice is due to the broader span of the likelihood curve's plateau, which generally extends
738 beyond 5%, wider than observed in simulations. Therefore, using smaller step sizes would
739 increase computational costs with little gain in performance.

740

741 **Data Availability**

742 The individual level genetic data for Native Hawaiian and Latino datasets were derived from the
743 Multiethnic Cohort (MEC), and are available on dbGaP (accession numbers: phs000220.v2.p2
744 and phs002183.v1.p1). The gLike package is available on its github page
745 (<https://github.com/Ephraim-usc/glike2>).

746

747 **Acknowledgement**

748 We would like to thank Iain Mathieson and Sara Mathieson for discussions and advice.
749 Research reported in this publication was supported by National Institute of Health under award
750 number R35GM142783 and R01HG12605 to C.W.K.C. Computation for this work was
751 supported by USC's Center for Advanced Research Computing (<https://carc.usc.edu>).

752

753 **Author's Contributions**

754 C.W.K.C., D.O.D.V., and C.D.H. conceived of the study. C.F. and C.W.K.C. designed the study.
755 C.F. and J.L.C. performed the analysis. B.L.D. curated the data. C.F., M.D.E., N.A.M., and
756 C.W.K.C. interpreted the data. C.F., J.L.C., M.D.E. and C.W.K.C. wrote the manuscript with
757 input from all co-authors.

758

759 **Competing Interests**

760 The authors declare no competing interests

761 **References**

- 762 1. Mathieson, I. & McVean, G. Differential confounding of rare and common variants in spatially
763 structured populations. *Nat Genet* **44**, 243–6 (2012).
- 764 2. Li, H. & Durbin, R. Inference of human population history from individual whole-genome
765 sequences. *Nature* **475**, 493–6 (2011).
- 766 3. Sheehan, S., Harris, K. & Song, Y. S. Estimating Variable Effective Population Sizes from
767 Multiple Genomes: A Sequentially Markov Conditional Sampling Distribution Approach.
768 *Genetics* **194**, 647–662 (2013).
- 769 4. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple
770 genome sequences. *Nat Genet* **46**, 919–25 (2014).
- 771 5. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history
772 from hundreds of unphased whole genomes. *Nat Genet* **49**, 303–309 (2017).
- 773 6. Palamara, P. F., Terhorst, J., Song, Y. S. & Price, A. L. High-throughput inference of pairwise
774 coalescence times identifies signals of selection and enriched disease heritability. *Nat*
775 *Genet* **50**, 1311–1317 (2018).
- 776 7. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the
777 Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency
778 Data. *PLoS Genet* **5**, e1000695 (2009).
- 779 8. Bhaskar, A., Wang, Y. X. R. & Song, Y. S. Efficient inference of population size histories and
780 locus-specific mutation rates from large-sample genomic variation data. *Genome Res.* **25**,
781 268–279 (2015).
- 782 9. Jouganous, J., Long, W., Ragsdale, A. P. & Gravel, S. Inferring the Joint Demographic
783 History of Multiple Populations: Beyond the Diffusion Approximation. *Genetics* **206**, 1549–
784 1567 (2017).

- 785 10. Kamm, J., Terhorst, J., Durbin, R. & Song, Y. S. Efficiently Inferring the Demographic
786 History of Many Populations With Allele Count Data. *Journal of the American Statistical*
787 *Association* **115**, 1472–1487 (2020).
- 788 11. Excoffier, L. *et al.* fastsimcoal2: demographic inference under complex evolutionary
789 scenarios. *Bioinformatics* **37**, 4882–4885 (2021).
- 790 12. Myers, S., Fefferman, C. & Patterson, N. Can one learn history from the allelic spectrum?
791 *Theor Popul Biol* **73**, 342–348 (2008).
- 792 13. Terhorst, J. & Song, Y. S. Fundamental limits on the accuracy of demographic inference
793 based on the sample frequency spectrum. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7677–7682
794 (2015).
- 795 14. Lapierre, M., Lambert, A. & Achaz, G. Accuracy of Demographic Inferences from the Site
796 Frequency Spectrum: The Case of the Yoruba Population. *Genetics* **206**, 439–449 (2017).
- 797 15. McVean, G. A Genealogical Interpretation of Principal Components Analysis. *PLoS Genet* **5**,
798 e1000686 (2009).
- 799 16. Opgen-Rhein, R., Fahrmeir, L. & Strimmer, K. Inference of demographic history from
800 genealogical trees using reversible jump Markov chain Monte Carlo. *BMC Evol Biol* **5**, 6
801 (2005).
- 802 17. Fan, C., Mancuso, N. & Chiang, C. W. K. A genealogical estimate of genetic relationships.
803 in (2020).
- 804 18. Hudson, R. R. *Gene genealogies and the coalescent process*. (Oxford surveys in
805 evolutionary biology, 1990).
- 806 19. Griffiths, R. C. & Marjoram, P. Ancestral Inference from Samples of DNA Sequences with
807 Recombination. *Journal of Computational Biology* **3**, 479–502 (1996).
- 808 20. Pearson, A. & Durbin, R. *Local Ancestry Inference for Complex Population Histories*.
809 <http://biorxiv.org/lookup/doi/10.1101/2023.03.06.529121> (2023)
810 doi:10.1101/2023.03.06.529121.

- 811 21. Wang, Z. *et al.* Automatic inference of demographic parameters using generative
812 adversarial networks. *Mol Ecol Resour* **21**, 2689–2705 (2021).
- 813 22. Gower, G., Picazo, P. I., Lindgren, F. & Racimo, F. *Inference of population genetics*
814 *parameters using discriminator neural networks: an adversarial Monte Carlo approach.*
815 <http://biorxiv.org/lookup/doi/10.1101/2023.04.27.538386> (2023)
816 doi:10.1101/2023.04.27.538386.
- 817 23. Adrion, J. R. *et al.* A community-maintained standard library of population genetic models.
818 *eLife* **9**, e54967 (2020).
- 819 24. Rosen, Z., Bhaskar, A., Roch, S. & Song, Y. S. Geometry of the Sample Frequency
820 Spectrum and the Perils of Demographic Inference. *Genetics* **210**, 665–682 (2018).
- 821 25. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of Population Structure Using
822 Multilocus Genotype Data. *Genetics* **155**, 945–959 (2000).
- 823 26. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in
824 unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- 825 27. Kubatko, L. S., Carstens, B. C. & Knowles, L. L. STEM: species tree estimation using
826 maximum likelihood for gene trees under coalescence. *Bioinformatics* **25**, 971–973 (2009).
- 827 28. Kelleher, J., Etheridge, A. M. & McVean, G. Efficient Coalescent Simulation and
828 Genealogical Analysis for Large Sample Sizes. *PLoS Comput Biol* **12**, e1004842 (2016).
- 829 29. Kelleher, J. *et al.* Inferring whole-genome histories in large population datasets. *Nat Genet*
830 **51**, 1330–1338 (2019).
- 831 30. Wohns, A. W. *et al.* *A unified genealogy of modern and ancient genomes.*
832 <http://biorxiv.org/lookup/doi/10.1101/2021.02.16.431497> (2021)
833 doi:10.1101/2021.02.16.431497.
- 834 31. Speidel, L., Forest, M., Shi, S. & Myers, S. R. A method for genome-wide genealogy
835 estimation for thousands of samples. *Nat Genet* **51**, 1321–1329 (2019).

- 836 32. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust
837 Demographic Inference from Genomic and SNP Data. *PLoS Genet* **9**, e1003905 (2013).
- 838 33. Sheng, X. *et al.* Inverted genomic regions between reference genome builds in humans
839 impact imputation accuracy and decrease the power of association testing. *Human Genetics*
840 *and Genomics Advances* **4**, 100159 (2023).
- 841 34. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium
842 panel. *Nat. Genet.* **48**, 1443–1448 (2016).
- 843 35. Woerner, A. E., Veeramah, K. R., Watkins, J. C. & Hammer, M. F. The Role of
844 Phylogenetically Conserved Elements in Shaping Patterns of Human Genomic Diversity.
845 *Molecular Biology and Evolution* **35**, 2284–2295 (2018).
- 846
- 847