

¹ Assessing the relationship of ancient and modern populations

² Joshua G. Schraiber

³ Started on October 22, 2016. Compiled on October 30, 2017

⁴ **Abstract**

⁵ Genetic material sequenced from ancient samples is revolutionizing our understand-
⁶ ing of the recent evolutionary past. However, ancient DNA is often degraded, resulting
⁷ in low coverage, error-prone sequencing. Several solutions exist to this problem, rang-
⁸ ing from simple approach such as selecting a read at random for each site to more
⁹ complicated approaches involving genotype likelihoods. In this work, we present a
¹⁰ novel method for assessing the relationship of an ancient sample with a modern popu-
¹¹ lation while accounting for sequencing error and post-mortem damage by analyzing raw
¹² read from multiple ancient individuals simultaneously. We show that when analyzing
¹³ SNP data, it is better to sequence more ancient samples to low coverage: two samples
¹⁴ sequenced to 0.5x coverage provide better resolution than a single sample sequenced
¹⁵ to 2x coverage. We also examined the power to detect whether an ancient sample is
¹⁶ directly ancestral to a modern population, finding that with even a few high cover-
¹⁷ age individuals, even ancient samples that are very slightly diverged from the modern
¹⁸ population can be detected with ease. When we applied our approach to European
¹⁹ samples, we found that no ancient samples represent direct ancestors of modern Euro-
²⁰ peans. We also found that, as shown previously, the most ancient Europeans appear
²¹ to have had the smallest effective population sizes, indicating a role for agriculture in
²² modern population growth.

23 1 Introduction

24 Ancient DNA (aDNA) is now ubiquitous in population genetics. Advances in DNA
25 isolation [Dabney et al., 2013], library preparation [Meyer et al., 2012], bone sampling
26 [Pinhasi et al., 2015], and sequence capture [Haak et al., 2015] make it possible to obtain
27 genome-wide data from hundreds of samples [Haak et al., 2015, Mathieson et al., 2015,
28 Allentoft et al., 2015, Fu et al., 2016]. Analysis of these data can provide new insight
29 into recent evolutionary processes which leave faint signatures in modern genomes,
30 including natural selection [Schraiber et al., 2016, Jewett et al., 2016] and population
31 replacement [Sjödin et al., 2014, Lazaridis et al., 2014].

32 One of the most powerful uses of ancient DNA is to assess the continuity of an-
33 cient and modern populations. In many cases, it is unclear whether populations that
34 occupied an area in the past are the direct ancestors of the current inhabitants of that
35 area. However, this can be next to impossible to assess using only modern genomes.
36 Questions of population continuity and replacement have particular relevance for the
37 spread of cultures and technology in humans [Lazaridis et al., 2016]. For instance, re-
38 cent work showed that modern South Americans are descended from people associated
39 with the Clovis culture that inhabited North America over 10,000 years ago, further enhancing
40 our understanding of the peopling of the Americas [Rasmussen et al., 2014].

41 Despite its utility in addressing difficult-to-answer questions in evolutionary biology,
42 aDNA also has several limitations. Most strikingly, DNA decays rapidly following
43 the death of an organism, resulting in highly fragmented, degraded starting material
44 when sequencing [Sawyer et al., 2012]. Thus, ancient data is frequently sequenced
45 to low coverage and has a significantly higher rate of misleadingly called nucleotides
46 than modern samples. When working with diploid data, as in aDNA extracted from
47 plants and animals, the low coverage prevents genotypes from being called with
48 confidence.

49 Several strategies are commonly used to address the low-coverage data. One of the
50 most common approaches is to sample a random read from each covered site and use

that as a haploid genotype call [Skoglund et al., 2012, Haak et al., 2015, Mathieson et al., 2015, Allentoft et al., 2015, Fu et al., 2016, Lazaridis et al., 2016]. Many common approaches to the analyses of ancient DNA, such as the usage of F-statistics [Green et al., 2010, Patterson et al., 2012], are designed with this kind of dataset in mind. F-statistics can be interpreted as linear combinations of simpler summary statistics and can often be understood in terms of testing a tree-like structure relating populations. Nonetheless, despite the simplicity and appeal of this approach, it has several drawbacks. Primarily, it throws away reads from sites that are covered more than once, resulting in a potential loss of information from expensive, difficult-to-acquire data. Moreover, as shown by Peter [2016], F-statistics are fundamentally based on heterozygosity, which is determined by samples of size 2, and thus limited in power. Finally, these approach are also strongly impacted by sequencing error, post-mortem damage, and contamination.

On the other hand, several approaches exist to either work with genotype likelihoods or the raw read data. Genotype likelihoods are the probabilities of the read data at a site given each of the three possible diploid genotypes at that site. They can be used in calculation of population genetic statistics or likelihood functions to average over uncertainty in the genotype [Korneliussen et al., 2014]. However, many such approaches assume that genotype likelihoods are fixed by the SNP calling algorithm (although they may be recalibrated to account for aDNA-specific errors, as in Jónsson et al. [2013]). However, with low coverage data, an increase in accuracy is expected if genotype likelihoods are co-estimated with other parameters of interest, due to the covariation between processes that influence read quality and genetic diversity, such as contamination.

A recent method that coestimates demographic parameters along with error and contamination rates by using genotype likelihoods showed that there can be significant power to assess the relationship of a single ancient sample to a modern population [Racimo et al., 2016]. Nonetheless, they found that for very low coverage data, inferences were not reliable. Thus, they were unable to apply their method to the large

80 number of extremely low coverage ($< 1x$) genomes that are available. Moreover, they
81 were unable to explore the tradeoffs that come with a limited budget: can we learn
82 more by sequencing fewer individuals to high coverage, or more individuals at lower
83 coverage?

84 Here, we develop a novel maximum likelihood approach for analyzing low coverage
85 ancient DNA in relation to a modern population. We work directly with raw read data
86 and explicitly model errors due to sequencing and post-mortem damage. Crucially,
87 our approach incorporates data from multiple individuals that belong to the same
88 ancient population, which we show substantially increases power and reduces error in
89 parameter estimates. We then apply our new methodology to ancient human data, and
90 show that we can perform accurate demographic inference even from very low coverage
91 samples by analyzing them jointly.

92 **2 Methods**

93 **2.1 Sampling alleles in ancient populations**

94 We assume a scenario in which allele frequencies are known with high accuracy in a
95 modern population. Suppose that an allele is known to be at frequency $x \in (0, 1)$ in
96 the modern population, and we wish to compute the probability of obtaining k copies
97 of that allele in a sample of n ($0 \leq k \leq n$) chromosomes from an ancient population.
98 As we show in the Appendix, conditioning on the frequency of the allele in the modern
99 population minimizes the impact of ascertainment, and allows this approach to be used
100 for SNP capture data.

To calculate the sampling probability, we assume a simple demographic model in which the ancient individual belongs to a population that split off from the modern population τ_1 generations ago, and subsequently existed as an isolated population for τ_2 generations. Further, we assume that the modern population has effective size $N_e^{(1)}$ and that the ancient population has effective size $N_e^{(2)}$, and measure time in diffusion units,

$t_i = \tau_i/(2N_e^{(i)})$. If we know the conditional probability that an allele is at frequency y in the ancient sample, given that it is at frequency x in the modern population, denoted $f(y; x, t_1, t_2)$, then the sampling probability is simply an integral,

$$\begin{aligned} P_{n,k}(x) &= \int_0^1 \binom{n}{k} y^k (1-y)^{n-k} f(y; x, t_1, t_2) dy \\ &= \binom{n}{k} \mathbb{E}_x (Y^k (1-Y)^{n-k}; t_1, t_2) \\ &\equiv \binom{n}{k} p_{n,k}(t_1, t_2) \end{aligned} \quad (1)$$

101 Thus, we must compute the binomial moments of the allele frequency distribution in
 102 the ancient population. In the Appendix, we show that this can be computed using
 103 matrix exponentiation,

$$p_{n,k}(t_1, t_2) = \left(e^{Qt_2} e^{Q^\downarrow t_1} \mathbf{h}_n \right)_i, \quad (2)$$

104 where $(\mathbf{v})_i$ indicates the i th element of the vector \mathbf{v} , $\mathbf{h}_n = ((1-x)^n, x(1-x)^{n-1}, \dots, x^n)^T$
 105 and Q and Q^\downarrow are the sparse matrices

$$Q_{ij} = \begin{cases} \frac{1}{2}i(i-1) & \text{if } j = i-1 \\ -i(n-i) & \text{if } j = i \\ \frac{1}{2}(n-i)(n-i-1) & \text{if } j = i+1 \\ 0 & \text{else} \end{cases}$$

106 and

$$Q_{ij}^\downarrow = \begin{cases} \frac{1}{2}i(i-1) & \text{if } j = i-1 \\ -i(n-i+1) & \text{if } j = i \\ \frac{1}{2}(n-i+1)(n-i) & \text{if } j = i+1 \\ 0 & \text{else.} \end{cases}$$

107 This result has an interesting interpretation: the matrix Q^\downarrow can be thought of as
 108 evolving the allele frequencies back in time from the modern population to the common

109 ancestor of the ancient and modern populations, while Q evolves the allele frequencies
 110 forward in time from the common ancestor to the ancient population (Fig 1).

111 [Figure 1 about here.]

112 Because of the fragmentation and degradation of DNA that is inherent in obtaining
 113 sequence data from ancient individuals, it is difficult to obtain the high coverage data
 114 necessary to make high quality genotype calls from ancient individuals. To address this,
 115 we instead work directly with raw read data, and average over all the possible genotypes
 116 weighted by their probability of producing the data. Specifically, we follow Nielsen et al.
 117 [2012] in modeling the probability of the read data in the ancient population, given the
 118 allele frequency at site l as

$$\mathbb{P}(R_l|k) = \sum_{g_{1,l}=0}^2 \dots \sum_{g_{n,l}=0}^2 \mathbb{I}\left(\sum_{i=1}^m g_{i,l} = k\right) \prod_{i=1}^n \binom{2}{g_{i,l}} \mathbb{P}(R_{i,l}|g_{i,l}),$$

119 where $R_{i,l} = (a_{i,l}, d_{i,l})$ are the counts of ancestral and derived reads in individual i at
 120 site l , $g_{i,l} \in \{0, 1, 2\}$ indicates the possible genotype of individual i at site l (i.e. 0 =
 121 homozygous ancestral, 1 = heterozygous, 2 = homozygous derived), and $\mathbb{P}(R_{i,l}|g_{i,l})$ is
 122 the probability of the read data at site l for individual i , assuming that the individual
 123 truly has genotype $g_{i,l}$. We use a binomial sampling with error model, in which the
 124 probability that a truly derived site appears ancestral (and vice versa) is given by
 125 ϵ . We emphasize that the parameter ϵ will capture both sequencing error as well as
 126 post-mortem damage (c.f. Racimo et al. [2016] who found that adding an additional
 127 parameter to specifically model post-mortem damage does not improve inferences).
 128 Thus,

$$\mathbb{P}(R|g) = \binom{a+d}{d} p_g^d (1-p_g)^a$$

with

$$\begin{aligned} p_0 &= \epsilon \\ p_1 &= \frac{1}{2} \\ p_2 &= 1 - \epsilon \end{aligned}$$

129 .

130 Combining these two aspects together by summing over possible allele frequencies
131 weighted by their probabilities, we obtain our likelihood of the ancient data,

$$L(D) = \prod_{l=1}^L \sum_{k=0}^n \mathbb{P}(R_l|k) p_{n,k}(x_l). \quad (3)$$

132

3 Results

133

3.1 Impact of coverage and number of samples on inferences

134

To explore the tradeoff of sequencing more individuals at lower depth compared to fewer
135 individuals at higher coverage, we performed simulations using `msprime` [Kelleher et al.,
136 2016] combined with custom scripts to simulate error and low coverage data. Briefly, we
137 assumed a Poisson distribution of reads at every site with mean given by the coverage,
138 and then simulated reads by drawing from the binomial distribution described in the
139 Methods.

140

First, we examined the impact of coverage and number of samples on the ability
141 to recover the drift times in the modern and the ancient populations. Figure 2 shows
142 results for data simulated with $t_1 = 0.02$ and $t_2 = 0.05$, corresponding to an ancient
143 individual who died 300 generations ago from population of effective size 1000. The
144 populations split 400 generations ago, and the modern population has an effective
size of 10000. We simulated approximately 180000 SNPs by simulating 100000 500

base pair fragments. Inferences of t_1 can be relatively accurate even with only one low coverage ancient sample (Figure 2A). However, inferences of t_2 benefit much more from increasing the number of ancient samples, as opposed to coverage (Figure 2B). Supplementary Table 1 shows that there is very little change in the average estimated parameter, indicating that most of the change in RMSE is due to decreased sampling variance. Thus, two individuals sequenced to 0.5x coverage have a much lower error than a single individual sequenced to 2x coverage, even though there is very little bias in either case. To explore this effect further, we derived the sampling probability of alleles covered by exactly one sequencing read (see Appendix). We found that sites covered only once have no information about t_2 , suggesting that evidence of heterozygosity is very important for inferences about t_2 . Finally, though we showed through simulation that there is sufficient power to disentangle t_1 from t_2 , estimates of these parameters are negatively correlated, due to the necessity of fitting the total drift time $t_1 + t_2$ (Supplementary Figure 1).

[Figure 2 about here.]

We next examined the impact of coverage and sampling on the power to reject the hypothesis that the ancient individuals came from a population that is directly ancestral to the modern population. We analyzed both low coverage (0.5x) and higher coverage (4x) datasets consisting of 1 (for both low and high coverage samples) or 5 individuals (only for low coverage). We simulated data with parameters identical to the previous experiment, except we now examined the impact of varying the age of the ancient sample from 0 generations ago through to the split time with the modern population. We then performed a likelihood ratio test comparing the null model of continuity, in which $t_2 = 0$, to a model in which the ancient population is not continuous. Figure 3 shows the power of the likelihood ratio test. For a single individual sequenced to low coverage, we see that the test only has power for very recently sampled ancient individuals (i.e. samples that are highly diverged from the modern population). However, the power increases dramatically as the number of individuals or the coverage per

174 individual is increased; sequencing 5 individuals to 0.5x coverage results in essentially
175 perfect power to reject continuity. Nonetheless, for samples that are very close to the
176 divergence time, it will be difficult to determine if they are ancestral to the modern
177 population or not, because differentiation is incomplete.

178 [Figure 3 about here.]

179 **3.2 Impact of admixture**

180 We examined two possible ways that admixture can result in violations of the model to
181 assess their impact on inference. In many situations, there may have been secondary
182 contact between the population from which the ancient sample is derived and the
183 modern population used as a reference. We performed simulations of this situation by
184 modifying the simulation corresponding to Figure 2 (300 generation old ancient sample
185 from population of size 1000 split from a population of size 10000 400 generations ago)
186 to include subsequent admixture from the ancient population to the modern popula-
187 tion 200 generations ago (NB: this admixture occurred *more recently* than the ancient
188 sample). In Figure 4, we show the results for admixture proportions ranging from 0
189 to 50%. Counterintuitively, estimates of t_1 initially *decrease* before again increasing.
190 This is likely a result of the increased heterozygosity caused by admixture, which acts
191 to artificially inflate the effective size of the modern population, and thus decrease t_1 .
192 As expected, t_2 is estimated to be smaller the more admixture there is; indeed, for an
193 admixture rate of 100%, the modern and ancient samples are continuous. The impact
194 on t_2 appears to be linear, and is well approximated by $(1 - f)t_2$ if the admixture
195 fraction is f .

196 [Figure 4 about here.]

197 In other situations, there may be admixture from an unsampled “ghost” population
198 into the modern population. If the ghost admixture is of a high enough proportion, it
199 is likely to cause a sample that is in fact a member of a directly ancestral population to

not appear to be ancestral. We explored this situation by augmenting our simulations in which the ancient sample is continuous with an outgroup population diverged from the modern population 0.04 time units ago (corresponding to 800 generations ago) and contributed genes to the modern population 0.01 time units ago (corresponding to 200 generations ago). We then assessed the impact on rejecting continuity using the likelihood ratio test (Figure 5). As expected, we see that low-power sampling strategies (such as a single individual sequenced to low coverage) are very minimally impacted by ghost admixture. However, for more powerful sampling strategies, moderate rates of ghost admixture ($\sim 10\%$) result in rejection of continuity.

[Figure 5 about here.]

3.3 Impact of contamination

We also explored the impact of foreign DNA contamination on inferences made using this approach. Briefly, we modified the simulations to include a chance c of a read being from a modern sample instead of the ancient sample when simulating reads. We again simulated data corresponding to Figure 2, with a 300 generation old ancient sample from population of size 1000 split from a population of size 10000 400 generations ago. In Figure 6, we see that relatively modest amounts of contamination can result in estimating zero or near-zero drift times. Interestingly, for the same contamination fraction, higher coverage samples are impacted slightly less. Together, this suggests that contamination will result in samples to be falsely inferred to be directly continuous with the modern population.

[Figure 6 about here.]

3.4 Application to ancient humans

We applied our approach to ancient human data from Mathieson et al. [2015], which is primarily derived from a SNP capture approach that targeted 1.2 million SNPs.

225 Based on sampling location and associated archeological materials, the individuals
226 were grouped into *a priori* panels, which we used to specify population membership
227 when analyzing individuals together. We analyzed all samples for their relationship to
228 the CEU individuals from the 1000 Genomes Project [Consortium, 2015]. Based on
229 our results that suggested that extremely low coverage samples would yield unreliable
230 estimates, we excluded panels that are composed of only a single individual sequenced
231 to less than 2x coverage.

232 We computed maximum likelihood estimates of t_1 and t_2 for individuals as grouped
233 into populations (Figure 7A; Table 1). We observe that t_2 is significantly greater
234 than 0 for all populations according to the likelihood ratio test. Thus, none of these
235 populations are consistent with directly making up a large proportion of the ancestry of
236 modern CEU individuals. Strikingly, we see that $t_2 \gg t_1$, despite the fact these samples
237 died in the past, and thus they belonged to a lineage that must have existed for fewer
238 generations since the population split than the modern samples. This suggests that
239 all of the ancient populations are characterized by extremely small effective population
240 sizes.

241 [Table 1 about here.]

242 [Figure 7 about here.]

243 We further explored the relationship between the dates of the ancient samples and
244 the parameters of the model by plotting t_1 and t_2 against the mean sample date of
245 all samples in that population (Figure 7B, C). We expected to find that t_1 correlated
246 with sample age, under the assumption that samples were members of relatively short-
247 lived populations that diverged from the “main-stem” of CEU ancestry. Instead, we
248 see no correlation between t_1 and sample time, suggesting that the relationship of
249 these populations to the CEU is complicated and not summarized well by the age
250 of the samples. On the other hand, we see a strong positive correlation between
251 t_2 and sampling time ($p < 1 \times 10^{-4}$). Because t_2 is a compound parameter, it is
252 difficult to directly interpret this relationship. However, it is consistent with the most

ancient samples belonging to populations with the smallest effective sizes, consistent with previous observations [Skoglund et al., 2014].

Finally, we examined the impact of grouping individuals into populations in real data. We see that estimates of t_1 for low coverage samples are typically lower when analyzed individually than when pooled with other individuals of the same panel (Figure 8A); because Supplementary Table 1 shows that there is no downward bias in t_1 for low coverage, this suggests that there may be some heterogeneity in these panels. On the other hand, there is substantial bias toward overestimating t_2 when analyzing samples individually, particularly for very low coverage samples (Figure 8B). This again shows that for estimates that rely on heterozygosity in ancient populations, pooling many low coverage individuals can significantly improve estimates.

[Figure 8 about here.]

4 Discussion

Ancient DNA (aDNA) presents unique opportunities to enhance our understanding of demography and selection in recent history. However, it also comes equipped with several challenges, due to postmortem DNA damage [Sawyer et al., 2012]. Several strategies have been developed to deal with the low quality of aDNA data, from relatively simple options like sampling a read at random at every site [Green et al., 2010] to more complicated methods making use of genotype likelihoods [Racimo et al., 2016]. Here, we presented a novel maximum likelihood approach for making inferences about how ancient populations are related to modern populations by analyzing read counts from multiple ancient individuals and explicitly modeling relationship between the two populations. We explicitly condition on the allele frequency in a modern population; as we showed in the appendix, this renders our method robust to ascertainment in modern samples. Thus, it can be used with SNP capture data. Moreover, confidence intervals can be calculated using a nonparametric bootstrap. Using this approach, we

279 examined some aspects of sampling strategy for aDNA analysis and we applied our
280 approach to ancient humans.

281 We found that sequencing many individuals from an ancient population to low
282 coverage (.5-1x) can be a significantly more cost effective strategy than sequencing
283 fewer individuals to relatively high coverage. For instance, we saw from simulations
284 that far more accurate estimates of the drift time in an ancient population can be
285 obtained by pooling 2 individuals at 0.5x coverage than by sequencing a single indi-
286 vidual to 2x coverage (Figure 2). We saw this replicated in our analysis of the real
287 data: low coverage individuals showed a significant amount of variation and bias in
288 estimating the model parameters that was substantially reduced when individuals were
289 analyzed jointly in a population (Figure 8). To explore this further, we showed that
290 sites sequenced to 1x coverage in a single individual retain no information about the
291 drift time in the ancient population. This can be intuitively understood because the
292 drift time in the ancient population is strongly related the amount of heterozygosity
293 in the ancient population: an ancient population with a longer drift time will have
294 lower heterozygosity at sites shared with a modern population. When a site is only
295 sequenced once in a single individual, there is no information about the heterozygosity
296 of that site. We also observed a pronounced upward bias in estimates of the drift time
297 in the ancient population from low coverage samples. We speculate that this is due to
298 the presence of few sites covered more than once being likely to be homozygous, thus
299 deflating the estimate of heterozygosity in the ancient population. Thus, for analysis
300 of SNP data, we recommend that aDNA sampling be conducted to maximize the num-
301 ber of individuals from each ancient population that can be sequenced to ~1x, rather
302 than attempting to sequence fewer individuals to high coverage. This suggestion can
303 be complicated when samples have vastly different levels of endogenous DNA, where it
304 may be cost effective to sequence high quality samples to higher coverage. In that case,
305 we recommend sequencing samples to at least 3-4x coverage; as evidenced by Figures
306 2 and 3, single samples at <4x coverage provide extremely limited information about
307 the drift time in the ancient population and thus little power to reject continuity.

When we looked at the impact of model misspecification, we saw several important patterns. First, the influence of admixture from the ancient population on inferences of t_2 is approximately linear, suggesting that if there are estimates of the amount of admixture between the modern and ancient population, a bias-corrected estimate of t_2 could be produced (Figure 4B). The impact on inference of t_1 is more complicated: admixture actually *reduces* estimates of t_1 (Figure 4A). This is likely because admixture increases the heterozygosity in the modern population, thus causing the amount of drift time to seem reduced. In both cases, the bias is not impacted by details of sampling strategy, although the variance of estimates is highly in a way consistent with Figure 2.

Of particular interest in many studies of ancient populations is the question of direct ancestry: are the ancient samples members of a population that contributed substantially to a modern population? We emphasize that this does not mean that the particular samples were direct ancestors of any modern individuals; indeed, this is exceedingly unlikely for old samples [Rohde et al., 2004, Chang, 1999, Baird et al., 2003, Donnelly, 1983]. Instead, we are asking whether an ancient sample was a member of a population that is directly continuous with a modern population. Several methods have been proposed to test this question, but thus far they have been limited to many individuals sequenced at a single locus [Sjödin et al., 2014] or to a single individual with genome-wide data [Rasmussen et al., 2014]. Our approach provides a rigorous, maximum likelihood framework for testing questions of population continuity using multiple low coverage ancient samples. We saw from simulations (Figure 3) that data from single, low coverage individuals result in very little power to reject the null hypothesis of continuity unless the ancient sample is very recent (i.e. it has been diverged from the modern population for a long time). Nonetheless, when low coverage individuals are pooled together, or a single high coverage individual is used, there is substantial power to reject continuity for all but the most ancient samples (i.e. samples dating from very near the population split time).

Because many modern populations may have experienced admixture from unsam-

337 plied “ghost” populations, we also performed simulations to test the impact of ghost
338 admixture on the probability of falsely rejecting continuity. We find that single an-
339 cient samples do not provide sufficient power to reject continuity even for high levels of
340 ghost admixture, while increasingly powerful sampling schemes, adding more individ-
341 uals or higher coverage per individual, reject continuity at higher rates. However, in
342 these situations, whether we regard rejection of continuity as a false or true discovery
343 is somewhat subjective: how much admixture from an outside population is required
344 before considering a population to not be directly ancestral? In future work it will be
345 extremely important to estimate the “maximum contribution” of the population an
346 ancient sample comes from (c.f Sjödin et al. [2014]).

347 To gain new insights from empirical data, we applied our approach to ancient
348 samples throughout Europe. Notably, we rejected continuity for all populations that we
349 analyzed. This is unsurprising, given that European history is extremely complicated
350 and has been shaped by many periods of admixture [Lazaridis et al., 2014, Haak et al.,
351 2015, Lazaridis et al., 2016]. Thus, modern Europeans have experienced many periods
352 of “ghost” admixture (relative to any particular ancient sample). Nonetheless, our
353 results show that none of these populations are even particularly close to directly
354 ancestral, as our simulations have shown that rejection of continuity will not occur
355 with low levels of ghost admixture.

356 Secondly, we observed that the drift time in the ancient population was much larger
357 than the drift time in the modern population. Assuming that the ancient sample were
358 a contemporary sample, the ratio t_1/t_2 is an estimator of the ratio $N_e^{(2)}/N_e^{(1)}$; in fact,
359 because the ancient sample existed for fewer generations since the common ancestor
360 of the ancient and modern populations, t_1/t_2 acts as an upper bound on $N_e^{(2)}/N_e^{(1)}$.
361 Moreover, this is unlikely to be due to unmodeled error in the ancient samples: error
362 would be expected increase the heterozygosity in the ancient sample, and thus *decrease*
363 our estimates of t_2 . Another potential complication is the fact that modern Europeans
364 are a mixture of multiple ancestral populations [Lazaridis et al., 2014, Haak et al.,
365 2015]. As shown through simulation, admixture increases heterozygosity in the modern

366 population and thus decreases estimates of t_1 . However, even very large amounts of
367 ghost admixture did not result in the order-of-magnitude differences we see in the
368 real data, suggesting that ghost admixture cannot account for all the discrepancy
369 between modern and ancient N_e . Thus, we find strong support for the observation that
370 ancient Europeans were often members of small, isolated populations [Skoglund et al.,
371 2014]. We interpret these two results together as suggestive that many ancient
372 samples found thus far in Europe were members of small populations that ultimately
373 went locally extinct. Nonetheless, there may be many samples that belonged to larger
374 metapopulations, and further work is necessary to specifically examine those cases.

375 We further examined the effective sizes of ancient populations through time by
376 looking for a correlation between the age of the ancient populations and the drift
377 time leading to them (Figure 7C). We saw a strong positive correlation, and although
378 this drift time is a compound parameter, which complicates interpretations, it appears
379 that the oldest Europeans were members of the smallest populations, and that effective
380 population size has grown through time as agriculture spread through Europe.

381 We anticipate the further development of methods that explicitly account for dif-
382 ferential drift times in ancient and modern samples will become important as aDNA
383 research becomes even more integrating into population genomics. This is because
384 many common summary methods, such as the use of Structure [Pritchard et al., 2000]
385 and Admixture [Alexander et al., 2009], are sensitive to differential amounts of drift
386 between populations [Falush et al., 2016]. As we've shown in ancient Europeans, an-
387 cient samples tend to come from isolated subpopulations with a large amount of drift,
388 thus confounding such summary approaches. Moreover, standard population genetics
389 theory shows that allele frequencies are expected to be deterministically lower in an-
390 cient samples, even if they are direct ancestors of a modern population. Intuitively,
391 this arises because the alleles must have arisen at some point from new mutations, and
392 thus were at lower frequencies in the past. A potentially fruitful avenue to combine
393 these approaches moving forward may be to separate regions of the genome based on
394 ancestry components, and assess the ancestry of ancient samples relative to specific

395 ancestry components, rather than to genomes as a whole.

396 Our current approach leaves several avenues for improvement. We use a relatively
397 simple error model that wraps up both post-mortem damage and sequencing error
398 into a single parameter. While Racimo et al. [2016] shows that adding an additional
399 parameter for PMD-related error does not significantly change results, the recent work
400 of Kousathanas et al. [2017] shows that building robust error models is challenging and
401 essential to estimating heterozygosity properly. Although our method is robust to non-
402 constant demography because we consider only alleles that are segregating in both the
403 modern and the ancient population, we are losing information by not modeling new
404 mutations that arise in the ancient population. Similarly, we only consider a single
405 ancient population at a time, albeit with multiple samples. Ideally, ancient samples
406 would be embedded in complex demographic models that include admixture, detailing
407 their relationships to each other and to modern populations [Patterson et al., 2012,
408 Lipson and Reich, 2017]. However, inference of such complex models is difficult, and
409 though there has been some progress in simplified cases [Lipson et al., 2014, Pickrell
410 and Pritchard, 2012], it remains an open problem due to the difficult of simultaneously
411 inferring a non-tree-like topology along with demographic parameters. Software such as
412 `momi` [Kamm et al., 2016] that can compute the likelihood of SNP data in an admixture
413 graph may be able to be used to integrate over genotype uncertainty in larger settings
414 than considered here.

415 5 Appendix

416 5.1 Computing allele frequency moments in the ancient popu- 417 lation

418 We wish to compute moments of the form

$$\mathbb{E}_x(g(Y); t_1, t_2) = \int_0^1 g(y)f(y; x, t_1, t_2)dy. \quad (4)$$

419 To do so, we make use of several results from diffusion theory. To ensure that this
 420 paper is self contained, we briefly review those results here. The interested reader may
 421 find much of this material covered in Ewens [2012], Karlin and Taylor [1981]. Several
 422 similar calculations can be found in Griffiths [2003].

423 Let the probability of an allele going from frequency x to frequency y in τ genera-
 424 tions in a population of size N_e be $f(x, y; t)$, where $t = \tau/(2N_e)$. Under a wide variety
 425 of models, the change in allele frequencies through time is well approximated by the
 426 Wright-Fisher diffusion, which is characterized by its generator,

$$\mathcal{L} = \frac{1}{2}x(1-x)\frac{d^2}{dx^2}.$$

427 The generator of a diffusion process is useful, because it can be used to define a differ-
 428 ential equation for the moments of that process,

$$\frac{d}{dt}\mathbb{E}_x(g(X_t)) = \mathbb{E}_x(\mathcal{L}g(X_t)). \quad (5)$$

429 We will require the *speed measure* of the Wright-Fisher diffusion, $m(x) = x^{-1}(1 -$
 430 $x)^{-1}$, which essentially describes how slow a diffusion at position x is “moving” com-
 431 pared to a Brownian motion at position x . Note that all diffusions are reversible with
 432 respect to their speed measures, i.e.

$$m(x)f(x, y; t) = m(y)f(y, x; t).$$

433 We additionally require the probability of loss, i.e. the probability that the allele
 434 currently at frequency x is ultimately lost from the population. This is

$$u_0(x) = 1 - x.$$

435 Note that it is possible to condition the Wright-Fisher diffusion to eventually be lost.

436 The transition density can be computed as

$$f^\downarrow(x, y; t) = f(x, y; t) \frac{u_0(y)}{u_0(x)}$$

437 by using Bayes theorem. The diffusion conditioned on loss is characterized by its
438 generator,

$$\mathcal{L}^\downarrow = -x \frac{d}{dx} + \frac{1}{2}x(1-x) \frac{d^2}{dx^2}.$$

439 In an infinite sites model, in which mutations occur at the times of a Poisson
440 process with rate $\theta/2$ and then each drift according to the Wright-Fisher diffusion, a
441 quasi-equilibrium distribution will be reached, known as the frequency spectrum. The
442 frequency spectrum, $\phi(x)$, predicts the number of sites at frequency x , and can be
443 written in terms of the speed measure and the probability of loss,

$$\phi(x) = \theta m(x) u_0(x).$$

444 To proceed with calculating (4), note that the conditional probability of an allele
445 being at frequency y in the ancient population given that it's at frequency x in the
446 modern population can be calculated

$$f(y; x, t_1, t_2) = \frac{f(x, y; t_1, t_2)}{\phi(x)}$$

447 where $f(x, y; t_1, t_2)$ is the joint probability of the allele frequencies in the modern and
448 ancient populations and $\phi(x)$ is the frequency spectrum in the modern population.

449 Assuming that the ancestral population of the modern and ancient samples was at
450 equilibrium, the joint distribution of allele frequencies can be computed by sampling
451 alleles from the frequency spectrum of the ancestor and evolving them forward in time
452 via the Wright-Fisher diffusion. This can be written mathematically as

$$f(x, y; t_1, t_2) = \int_0^1 f(z, x; t_1) f(z, y; t_2) \phi(z) dz.$$

We now expand the frequency spectrum in terms of the speed measure and the probability of loss and use reversibility with respect to the speed measure to rewrite the equation,

$$\begin{aligned}
\int_0^1 f(z, x; t_1) f(z, y; t_2) \phi(z) dz &= \theta \int_0^1 f(z, x; t_1) f(z, y; t_2) m(z) u_0(z) dz \\
&= \theta \int_0^1 \frac{m(x)}{m(z)} f(x, z; t_1) f(z, y; t_2) m(z) u_0(z) dz \\
&= \theta m(x) u_0(x) \int_0^1 f(x, z; t_1) \frac{u_0(z)}{u_0(x)} f(z, y; t_2) dz \\
&= \phi(x) \int_0^1 f^\downarrow(x, z; t_1) f(z, y; t_2) dz.
\end{aligned}$$

453 The third line follows by multiplying by $u_0(x)/u_0(x) = 1$. This equation has the interpretation of sampling an allele from the frequency spectrum in the modern population,
 454 then evolving it *backward* in time to the common ancestor, before evolving it *forward*
 455 in time to the ancient population. The interpretation of the diffusion conditioned on
 456 loss as evolving backward in time arises by considering the fact that alleles arose from
 457 unique mutations at some point in the past; hence, looking backward, alleles must
 458 eventually be lost at some point in the past.
 459

To compute the expectation, we substitute this form for the joint probability into (4),

$$\begin{aligned}
\int_0^1 g(y) f(y; x, t_1, t_2) dy &= \int_0^1 g(y) \left(\int_0^1 f^\downarrow(x, z; t_1) f(z, y; t_2) dz \right) dy \\
&= \int_0^1 \left(\int_0^1 g(y) f(z, y; t_2) dy \right) f^\downarrow(x, z; t_1) dz,
\end{aligned}$$

where the second line follows by rearranging terms and exchanging the order of integration. Note that this formula takes the form of nested expectations. Specifically,

$$\begin{aligned}
\int_0^1 g(y) f(z, y; t_2) dy &= \mathbb{E}_z(g(Y_{t_2})) \\
&\equiv h(z)
\end{aligned}$$

and

$$\begin{aligned} \int_0^1 h(z) f^\downarrow(x, z; t_1) dz &= \mathbb{E}_x^\downarrow(h(Z_{t_1})) \\ &= \mathbb{E}_x(g(Y); t_1, t_2). \end{aligned}$$

460 We now use (5) to note that

$$\frac{d}{dt} p_{n,k} = \frac{k(k-1)}{2} p_{n,k-1} - k(n-k)p_{n,k} + \frac{(n-k)(n-k-1)}{2} p_{n,k+1}$$

461 and

$$\frac{d}{dt} p_{n,k}^\downarrow = \frac{k(k-1)}{2} p_{n,k-1}^\downarrow - k(n-k+1)p_{n,k}^\downarrow + \frac{(n-k+1)(n-k)}{2} p_{n,k+1}^\downarrow$$

462 with obvious boundary conditions $p_{n,k}(0; z) = z^k(1-z)^{n-k}$ and $p_{n,k}^\downarrow(0; x) = x^k(1-x)^{n-k}$.
463

464 These systems of differential equations can be rewritten as matrix differential equa-
465 tions with coefficient matrices Q and Q^\downarrow respectively. Because they are linear, first
466 order equations, they can be solved by matrix exponentiation. Because the expectation
467 of a polynomial in the Wright-Fisher diffusion remains a polynomial, the nested expec-
468 tations can be computed via matrix multiplication of the solutions to these differential
469 equations, yielding the formula (2).

470 5.2 Robustness to ascertainment in the modern population

471 By conditioning on the allele frequency in the modern population, we gain the power
472 to make inferences that are robust to ascertainment in the modern population. To see
473 this, note from Equation 3 in Nielsen and Signorovitch [2003] that

$$f(x|A) = \frac{f(x, A)}{f(A)}$$

474 where A indicates the event that the allele was ascertained in the modern population.
 475 A simple generalization of this shows that

$$f(x, y|A) = \frac{f(x, y, A)}{f(A)}.$$

So,

$$\begin{aligned} f(y|x, A) &= \frac{f(x, y|A)}{f(x|A)} \\ &= \frac{f(x, y, A)}{f(x, A)} \\ &= \frac{f(A|x, y)f(x, y)}{f(A|x)f(x)} \\ &= \frac{f(x, y)}{f(x)} \end{aligned}$$

476 where the final line follows by recognizing that $f(A|x, y) = f(A|x)$ since the allele was
 477 ascertained in the modern population. Thus, we see that the ascertainment is removed
 478 by conditioning and we recover the original formula. Note that the robustness to
 479 ascertainment is only exact if the allele is ascertained in the modern population, but is
 480 expected to be very close to true so long as the allele is ascertained in a population
 481 closer to the modern population than to the ancient population.

482 5.3 Sites covered exactly once have no information about drift 483 in the ancient population

484 Consider a simplified model in which each site has exactly one read. When we have
 485 sequence from only a single individual, we have a set l_a of sites where the single read is
 486 an ancestral allele and a set l_d of sites where the single read is a derived allele. Thus,
 487 we can rewrite (3) as

$$L(D) = \prod_{l \in l_a} \left((1 - \epsilon)P_{2,0}(x_l) + \frac{1}{2}P_{2,1}(x_l) + \epsilon P_{2,2}(x_l) \right) \prod_{l \in l_d} \left(\epsilon P_{2,0}(x_l) + \frac{1}{2}P_{2,1}(x_l) + (1 - \epsilon)P_{2,2}(x_l) \right).$$

We can use formulas from Racimo et al. [2016] to compute $P_{2,k}(x_l)$ for $k \in \{0, 1, 2\}$,

$$\begin{aligned} P_{2,0}(x_l) &= 1 - x_l e^{-t_1} - \frac{1}{2} x_l e^{-(t_1+t_2)} + x_l \left(x_l - \frac{1}{2} \right) e^{-(3t_1+t_2)} \\ P_{2,1}(x_l) &= x_l e^{-(t_1+t_2)} + x_l (1 - 2x_l) e^{-(3t_1+t_2)} \\ P_{2,2}(x_l) &= x_l e^{-t_1} - \frac{1}{2} x_l e^{-(t_1+t_2)} + x_l \left(x_l - \frac{1}{2} \right) e^{-(3t_1-t_2)}. \end{aligned}$$

488 Note then that

$$(1 - \epsilon)P_{2,0}(x_l) + \frac{1}{2}P_{2,1}(x_l) + \epsilon P_{2,2}(x_l) = 1 - \epsilon - x(1 - 2\epsilon)e^{-t_1}$$

489 and

$$\epsilon P_{2,0}(x_l) + \frac{1}{2}P_{2,1}(x_l) + (1 - \epsilon)P_{2,2}(x_l) = \epsilon + x(1 - 2\epsilon)e^{-t_1}.$$

490 Neither of these formulas depend on t_2 ; hence, there is no information about the drift
491 time in the ancient population from data that is exactly 1x coverage.

492 6 Software Availability

493 Python implementations of the described methods are available at www.github.com/schraiber/continuity/

494 7 Acknowledgments

495 We wish to thank Melinda Yang, Iain Mathieson, Pontus Skoglund, and Fernando
496 Racimo for several discussions during the conception of this work that greatly improved
497 its scope and rigor. We are grateful to Fernando Racimo and Benjamin Vernot for
498 comments on early versions of this work that significantly improved its quality. We
499 also appreciate comments on the preprint from Alex Kim and Aylwyn Scally. We
500 would also like to thank Tamara Broderick for several stimulating discussions about
501 clustering that ultimately didn't result in any interesting application to ancient DNA.

502 References

- 503 David H Alexander, John Novembre, and Kenneth Lange. Fast model-based estimation
504 of ancestry in unrelated individuals. *Genome research*, 19(9):1655–1664, 2009.
- 505 Morten E Allentoft, Martin Sikora, Karl-Göran Sjögren, Simon Rasmussen, Morten
506 Rasmussen, Jesper Stenderup, Peter B Damgaard, Hannes Schroeder, Torbjörn
507 Ahlström, Lasse Vinner, et al. Population genomics of bronze age eurasia. *Nature*,
508 522(7555):167–172, 2015.
- 509 SJE Baird, NH Barton, and AM Etheridge. The distribution of surviving blocks of an
510 ancestral genome. *Theoretical population biology*, 64(4):451–471, 2003.
- 511 Joseph T Chang. Recent common ancestors of all present-day individuals. *Advances*
512 *in Applied Probability*, 31(4):1002–1026, 1999.
- 513 1000 Genomes Project Consortium. A global reference for human genetic variation.
514 *Nature*, 526(7571):68–74, 2015.
- 515 Jesse Dabney, Matthias Meyer, and Svante Pääbo. Ancient dna damage. *Cold Spring*
516 *Harbor perspectives in biology*, 5(7):a012567, 2013.
- 517 Kevin P Donnelly. The probability that related individuals share some section of
518 genome identical by descent. *Theoretical population biology*, 23(1):34–63, 1983.
- 519 Warren J Ewens. *Mathematical population genetics 1: theoretical introduction*, vol-
520 ume 27. Springer Science & Business Media, 2012.
- 521 Daniel Falush, Lucy van Dorp, and Daniel Lawson. A tutorial on how (not) to over-
522 interpret structure/admixture bar plots. *bioRxiv*, page 066431, 2016.
- 523 Qiaomei Fu, Cosimo Posth, Mateja Hajdinjak, Martin Petr, Swapan Mallick, Daniel
524 Fernandes, Anja Furtwängler, Wolfgang Haak, Matthias Meyer, Alissa Mittnik, et al.
525 The genetic history of ice age europe. *Nature*, 2016.

- 526 Richard E Green, Johannes Krause, Adrian W Briggs, Tomislav Maricic, Udo Stenzel,
527 Martin Kircher, Nick Patterson, Heng Li, Weiwei Zhai, Markus Hsi-Yang Fritz, et al.
528 A draft sequence of the neandertal genome. *science*, 328(5979):710–722, 2010.
- 529 RC Griffiths. The frequency spectrum of a mutation, and its age, in a general diffusion
530 model. *Theoretical population biology*, 64(2):241–251, 2003.
- 531 Wolfgang Haak, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick,
532 Bastien Llamas, Guido Brandt, Susanne Nordenfelt, Eadaoin Harney, Kristin Stew-
533 ardson, et al. Massive migration from the steppe was a source for indo-european
534 languages in europe. *Nature*, 522(7555):207–211, 2015.
- 535 Ethan M Jewett, Matthias Steinrücken, and Yun S Song. The effects of population
536 size histories on estimates of selection coefficients from time-series genetic data.
537 *Molecular Biology and Evolution*, page msw173, 2016.
- 538 Hákon Jónsson, Aurélien Ginolhac, Mikkel Schubert, Philip LF Johnson, and Ludovic
539 Orlando. mapdamage2. 0: fast approximate bayesian estimates of ancient dna dam-
540 age parameters. *Bioinformatics*, 29(13):1682–1684, 2013.
- 541 John A Kamm, Jonathan Terhorst, and Yun S Song. Efficient computation of the joint
542 sample frequency spectra for multiple populations. *Journal of Computational and
543 Graphical Statistics*, (just-accepted):1–37, 2016.
- 544 Samuel Karlin and Howard E Taylor. *A second course in stochastic processes*. Elsevier,
545 1981.
- 546 Jerome Kelleher, Alison M Etheridge, and Gilean McVean. Efficient coalescent sim-
547 ulation and genealogical analysis for large sample sizes. *PLoS Comput Biol*, 12(5):
548 e1004842, 2016.
- 549 Thorfinn Sand Korneliussen, Anders Albrechtsen, and Rasmus Nielsen. Angsd: analysis
550 of next generation sequencing data. *BMC bioinformatics*, 15(1):356, 2014.

- 551 Athanasios Kousathanas, Christoph Leuenberger, Vivian Link, Christian Sell, Joachim
552 Burger, and Daniel Wegmann. Inferring heterozygosity from ancient and low cover-
553 age genomes. *Genetics*, 205(1):317–332, 2017.
- 554 Iosif Lazaridis, Nick Patterson, Alissa Mitnik, Gabriel Renaud, Swapan Mallick,
555 Karola Kirksnow, Peter H Sudmant, Joshua G Schraiber, Sergi Castellano, Mark
556 Lipson, et al. Ancient human genomes suggest three ancestral populations for
557 present-day europeans. *Nature*, 513(7518):409–413, 2014.
- 558 Iosif Lazaridis, Dani Nadel, Gary Rollefson, Deborah C Merrett, Nadin Rohland, Swa-
559 pan Mallick, Daniel Fernandes, Mario Novak, Beatriz Gamarra, Kendra Sirak, et al.
560 Genomic insights into the origin of farming in the ancient near east. *Nature*, 536
561 (7617):419–424, 2016.
- 562 Mark Lipson and David Reich. Working model of the deep relationships of diverse
563 modern human genetic lineages outside of africa. *Molecular Biology and Evolution*,
564 page msw293, 2017.
- 565 Mark Lipson, Po-Ru Loh, Nick Patterson, Priya Moorjani, Ying-Chin Ko, Mark
566 Stoneking, Bonnie Berger, and David Reich. Reconstructing austromesian popu-
567 lation history in island southeast asia. *Nature communications*, 5, 2014.
- 568 Iain Mathieson, Iosif Lazaridis, Nadin Rohland, Swapan Mallick, Nick Patterson,
569 Songül Alpaslan Roodenberg, Eadaoin Harney, Kristin Stewardson, Daniel Fernan-
570 des, Mario Novak, et al. Genome-wide patterns of selection in 230 ancient eurasians.
571 *Nature*, 528(7583):499–503, 2015.
- 572 Matthias Meyer, Martin Kircher, Marie-Theres Gansauge, Heng Li, Fernando Racimo,
573 Swapan Mallick, Joshua G Schraiber, Flora Jay, Kay Prüfer, Cesare De Filippo, et al.
574 A high-coverage genome sequence from an archaic denisovan individual. *Science*, 338
575 (6104):222–226, 2012.

- 576 Rasmus Nielsen and James Signorovitch. Correcting for ascertainment biases when
577 analyzing snp data: applications to the estimation of linkage disequilibrium. *Theo-*
578 *retical population biology*, 63(3):245–255, 2003.
- 579 Rasmus Nielsen, Thorfinn Korneliussen, Anders Albrechtsen, Yingrui Li, and Jun
580 Wang. Snp calling, genotype calling, and sample allele frequency estimation from
581 new-generation sequencing data. *PloS one*, 7(7):e37558, 2012.
- 582 Nick Patterson, Priya Moorjani, Yontao Luo, Swapan Mallick, Nadin Rohland, Yiping
583 Zhan, Teri Genschoreck, Teresa Webster, and David Reich. Ancient admixture in
584 human history. *Genetics*, 192(3):1065–1093, 2012.
- 585 Benjamin M Peter. Admixture, population structure, and f-statistics. *Genetics*, 202
586 (4):1485–1501, 2016.
- 587 Joseph K Pickrell and Jonathan K Pritchard. Inference of population splits and mix-
588 tures from genome-wide allele frequency data. *PLoS Genet*, 8(11):e1002967, 2012.
- 589 Ron Pinhasi, Daniel Fernandes, Kendra Sirak, Mario Novak, Sarah Connell, Songül
590 Alpaslan-Roodenberg, Fokke Gerritsen, Vyacheslav Moiseyev, Andrey Gromov, Pál
591 Raczyk, et al. Optimal ancient dna yields from the inner ear part of the human
592 petrous bone. *PloS one*, 10(6):e0129102, 2015.
- 593 Jonathan K Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population
594 structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.
- 595 Fernando Racimo, Gabriel Renaud, and Montgomery Slatkin. Joint estimation of
596 contamination, error and demography for nuclear dna from ancient humans. *PLoS*
597 *Genet*, 12(4):e1005972, 2016.
- 598 Morten Rasmussen, Sarah L Anzick, Michael R Waters, Pontus Skoglund, Michael De-
599 Giorgio, Thomas W Stafford Jr, Simon Rasmussen, Ida Moltke, Anders Albrechtsen,
600 Shane M Doyle, et al. The genome of a late pleistocene human from a clovis burial
601 site in western montana. *Nature*, 506(7487):225–229, 2014.

- 602 Douglas LT Rohde, Steve Olson, and Joseph T Chang. Modelling the recent common
603 ancestry of all living humans. *Nature*, 431(7008):562, 2004.
- 604 Susanna Sawyer, Johannes Krause, Katerina Guschanski, Vincent Savolainen, and
605 Svante Pääbo. Temporal patterns of nucleotide misincorporations and dna frag-
606 mentation in ancient dna. *PloS one*, 7(3):e34131, 2012.
- 607 Joshua G Schraiber, Steven N Evans, and Montgomery Slatkin. Bayesian inference of
608 natural selection from allele frequency time series. *Genetics*, 203(1):493–511, 2016.
- 609 Per Sjödin, Pontus Skoglund, and Mattias Jakobsson. Assessing the maximum con-
610 tribution from ancient populations. *Molecular biology and evolution*, page msu059,
611 2014.
- 612 Pontus Skoglund, Helena Malmström, Maanasa Raghavan, Jan Storå, Per Hall, Eske
613 Willerslev, M Thomas P Gilbert, Anders Götherström, and Mattias Jakobsson. Ori-
614 gins and genetic legacy of neolithic farmers and hunter-gatherers in europe. *Science*,
615 336(6080):466–469, 2012.
- 616 Pontus Skoglund, Helena Malmström, Ayça Omrak, Maanasa Raghavan, Cristina
617 Valdiosera, Torsten Günther, Per Hall, Kristiina Tambets, Jüri Parik, Karl-Göran
618 Sjögren, et al. Genomic diversity and admixture differs for stone-age scandinavian
619 foragers and farmers. *Science*, 344(6185):747–750, 2014.

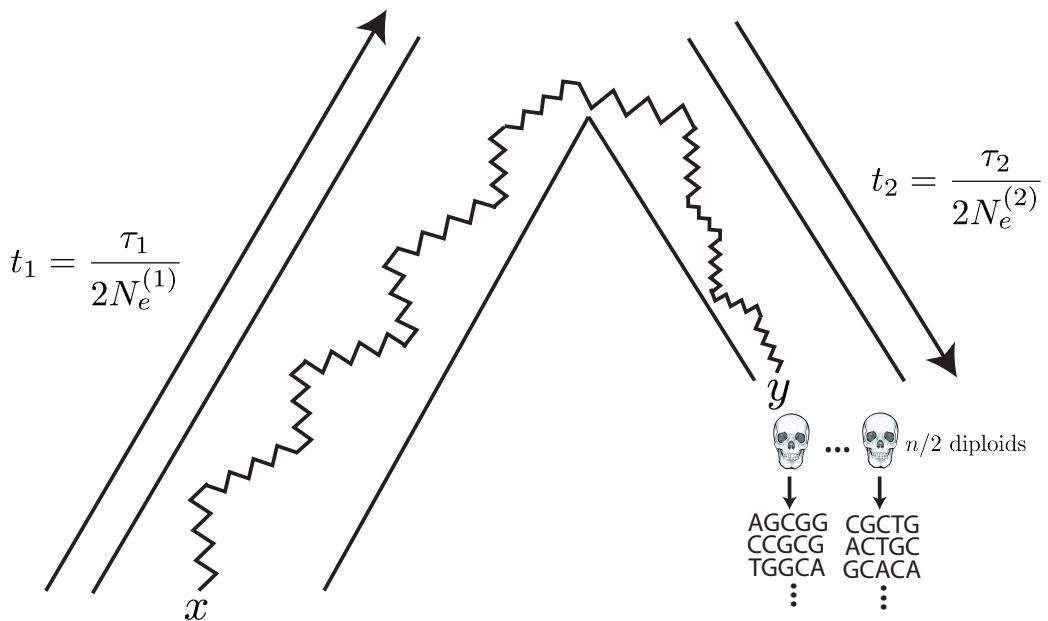


Figure 1: The generative model. Alleles are found at frequency x in the modern population and are at frequency y in the ancient population. The modern population has effective size $N_e^{(1)}$ and has evolved for τ_1 generations since the common ancestor of the modern and ancient populations, while the ancient population is of size $N_e^{(2)}$ and has evolved for τ_2 generations. Ancient diploid samples are taken and sequenced to possibly low coverage, with errors. Arrows indicate that the sampling probability can be calculated by evolving alleles *backward* in time from the modern population and then forward in time to the ancient population.

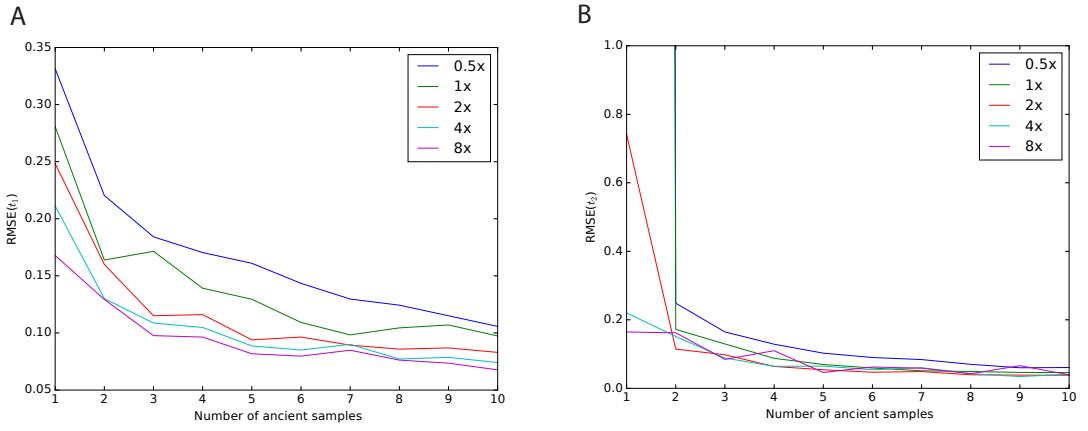


Figure 2: Impact of sampling scheme on parameter estimation error. In each panel, the x axis represents the number of simulated ancient samples, while the y axis shows the relative root mean square error for each parameter. Each different line corresponds to individuals sequenced to different depth of coverage. Panel A shows results for t_1 while panel B shows results for t_2 . Simulated parameters are $t_1 = 0.02$ and $t_2 = 0.05$.

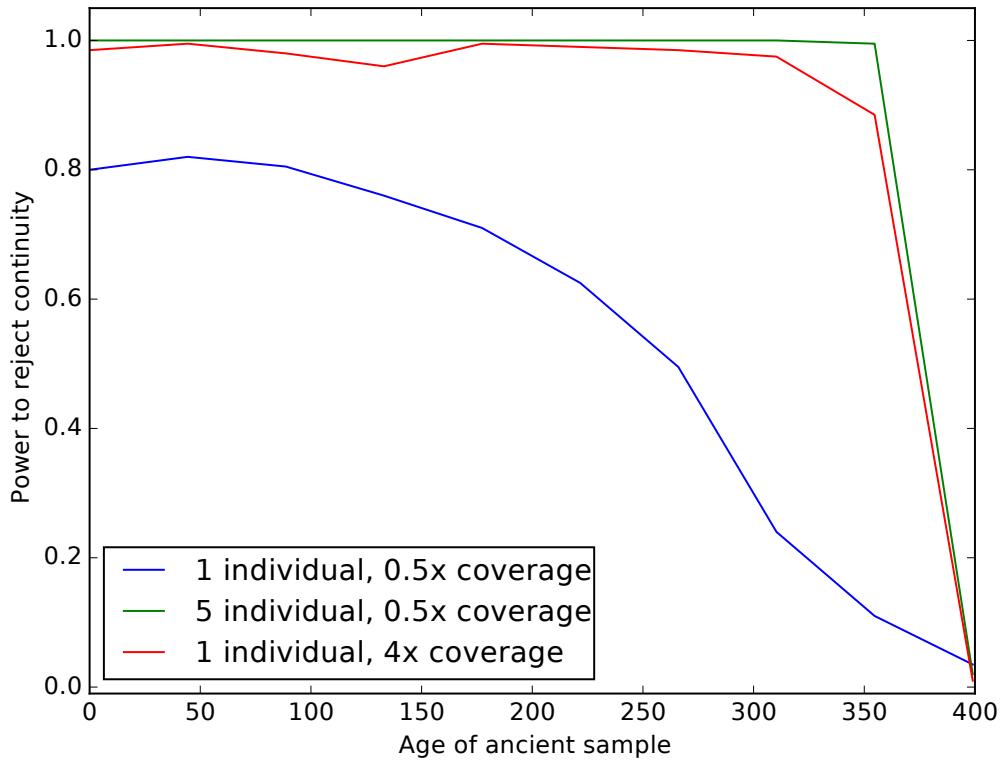


Figure 3: Impact of sampling scheme on rejecting population continuity. The x axis represents the age of the ancient sample in generations, with 0 indicating a modern sample and 400 indicating a sample from exactly at the split time 400 generations ago. The y axis shows the proportion of simulations in which we rejected the null hypothesis of population continuity. Each line shows different sampling schemes, as explained in the legend.

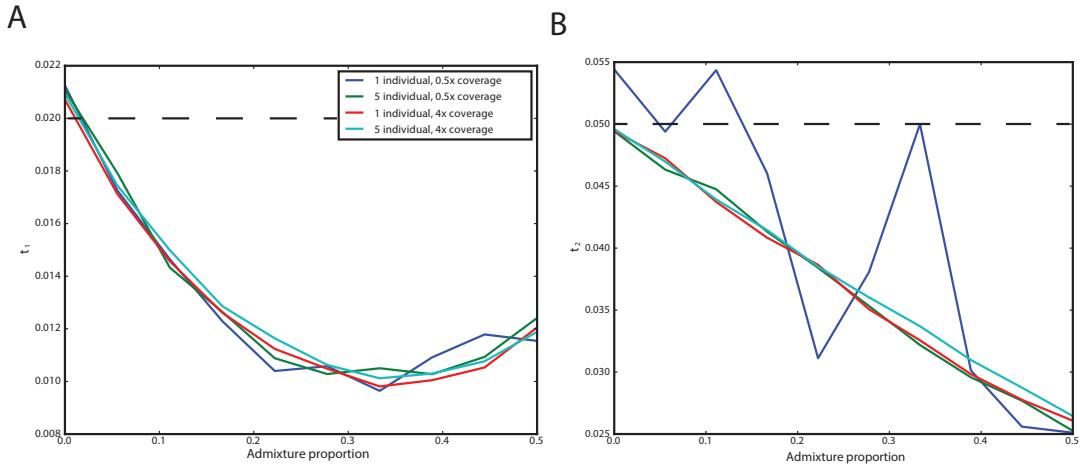


Figure 4: Impact of admixture from the ancient population on inferred parameters. The x axis shows the admixture proportion and the y axis shows the average parameter estimate across simulations. Each line corresponds to a different sampling strategy, as indicated in the legend. Panel A shows results for t_1 and Panel B shows results for t_2 . The true values of $t_1 = 0.02$ and $t_2 = 0.05$ are indicated by dashed lines.

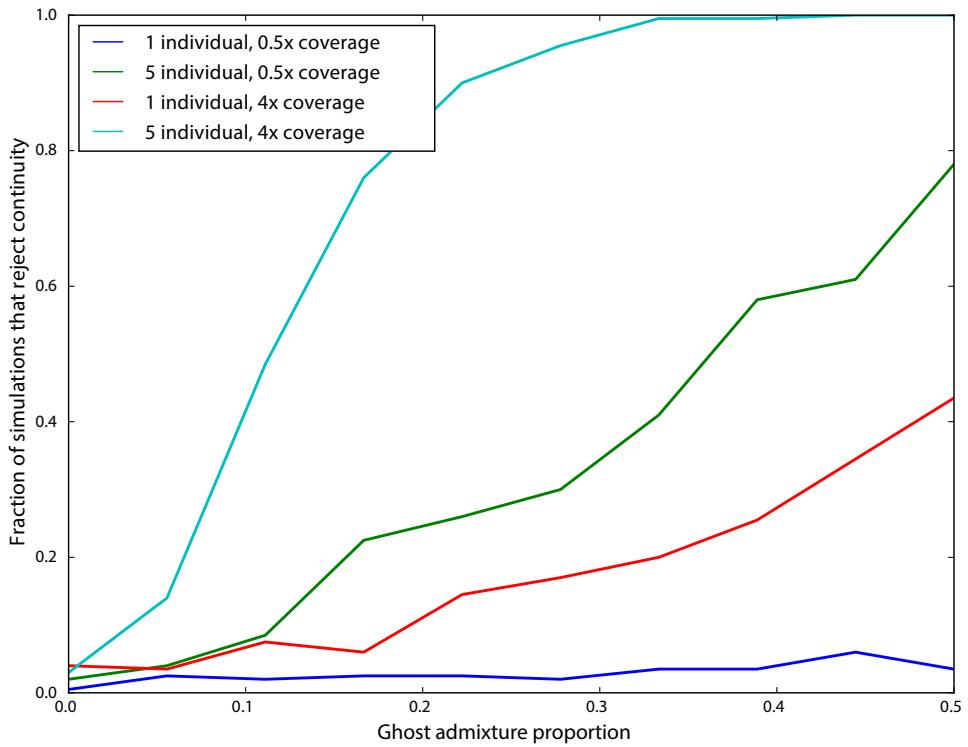


Figure 5: Impact of ghost admixture on rejecting continuity. The x axis shows the admixture proportion from the ghost population, and the y axis shows the fraction of simulations in which continuity was rejected. Each line corresponds to a different sampling strategy, as indicated in the legend.

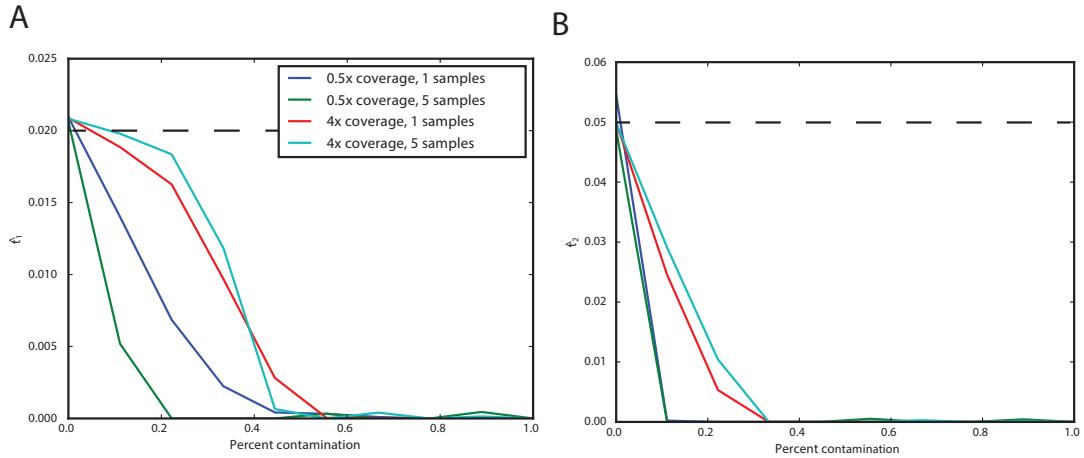


Figure 6: Impact of contamination on parameter inference. The x axis shows the contamination fraction, and the y axis shows the average parameter estimate from simulations. Each line corresponds to a different sampling strategy, as indicated in the legend. Panel A shows t_1 , and Panel B shows t_2 . Dashed lines indicate the true values of $t_1 = 0.02$ and $t_2 = 0.05$

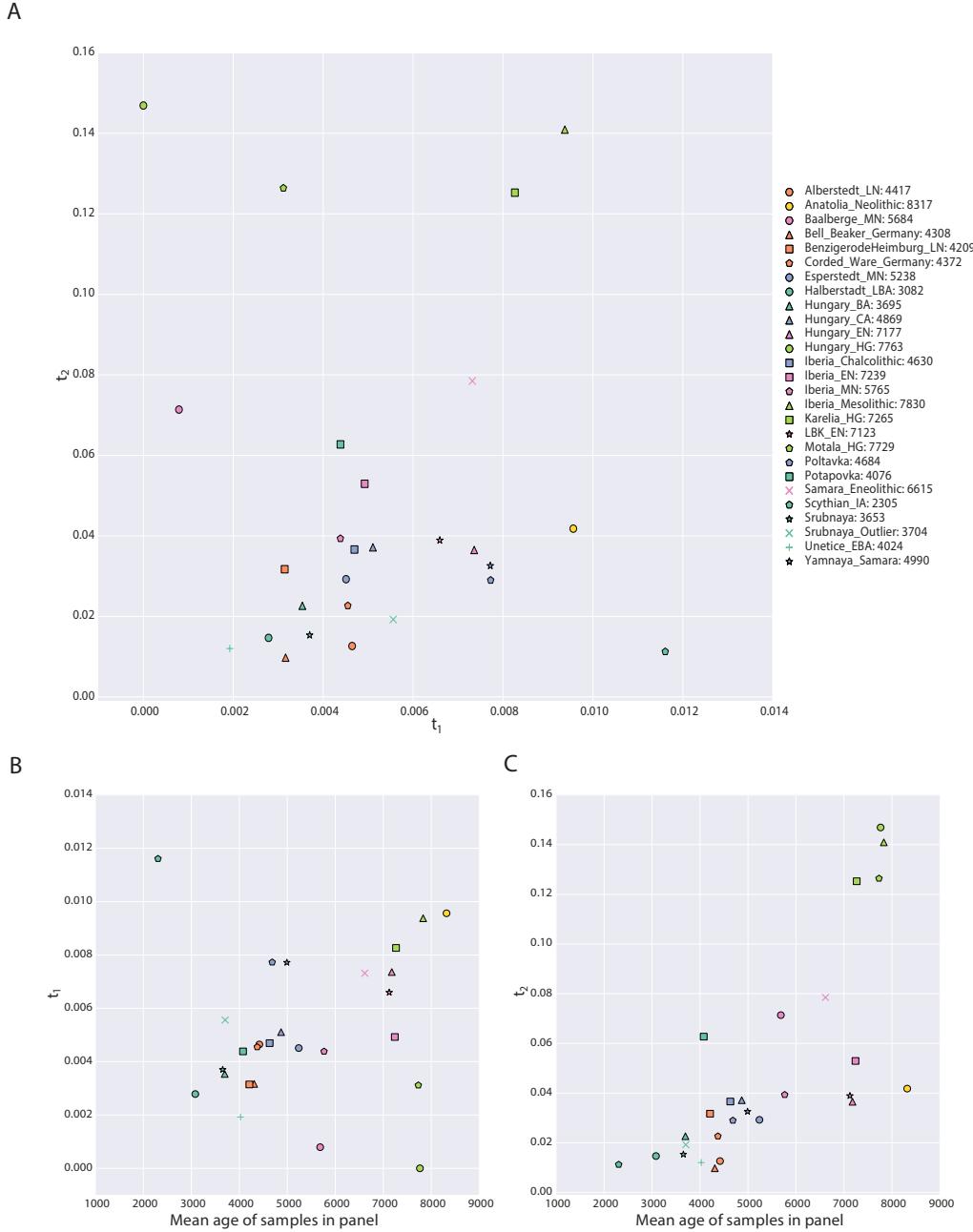


Figure 7: Parameters of the model inferred from ancient West Eurasian samples. Panel A shows t_1 on the x-axis and t_2 on the y-axis, with each point corresponding to a population as indicated in the legend. Numbers in the legend correspond to the mean date of all samples in the population. Panels B and C show scatterplots of the mean age of the samples in the population (x-axis) against t_1 and t_2 , respectively. Points are described by the same legend as Panel A.

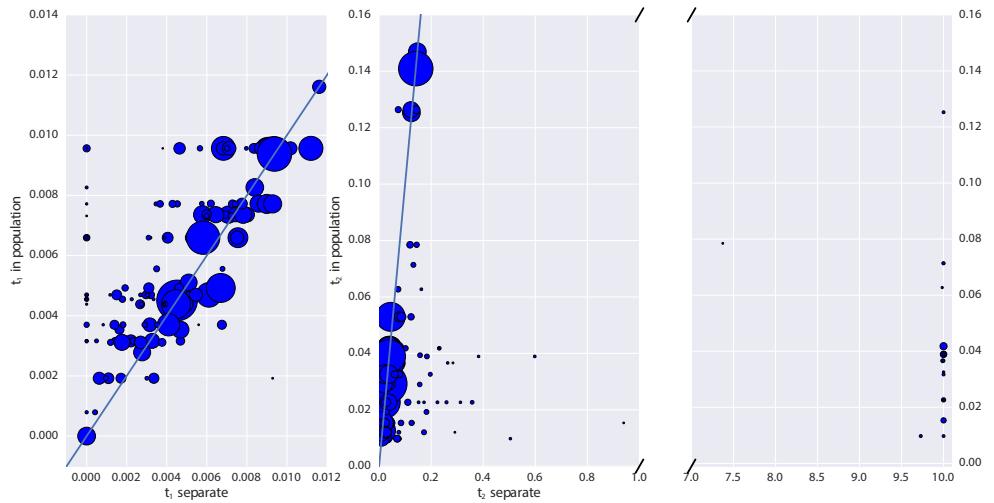


Figure 8: Impact of pooling individuals into populations when estimating model parameters from real data. In both panels, the x-axis indicates the parameter estimate when individuals are analyzed separately, while the y-axis indicates the parameter estimate when individuals are grouped into populations. Size of points is proportional to the coverage of each individual. Panel A reports the impact on estimation of t_1 , while Panel B reports the impact on t_2 . Note that Panel B has a broken x-axis. Solid lines in each figure indicate $y = x$.

pop	cov	date	t_1	t_2	lnL	t_1 (cont)	lnL (cont)
Alberstedt_LN	12.606	4417.000	0.005	0.013	-779411.494	0.006	-779440.143
Anatolia_Neolithic	3.551	8317.500	0.010	0.042	-9096440.714	0.044	-9106156.877
Baalberge_MN	0.244	5684.333	0.001	0.071	-201575.306	0.007	-201750.419
Bell_Beaker_Germany	1.161	4308.444	0.003	0.010	-1834486.744	0.008	-1834652.858
BenzigerodeHeimburg_LN	0.798	4209.750	0.003	0.032	-346061.545	0.007	-346134.356
Corded_Ware_Germany	2.250	4372.833	0.005	0.023	-2139002.723	0.017	-2139858.192
Esperstedt_MN	30.410	5238.000	0.005	0.029	-975890.329	0.009	-976047.889
Halberstadt_LBA	5.322	3082.000	0.003	0.015	-558966.522	0.004	-558993.078
Hungary_BA	3.401	3695.750	0.004	0.023	-789754.969	0.010	-789939.889
Hungary_CA	5.169	4869.500	0.005	0.037	-504413.094	0.010	-504549.603
Hungary_EN	4.033	7177.000	0.007	0.036	-3478429.262	0.033	-3481855.461
Hungary_HG	5.807	7763.000	0.000	0.147	-469887.471	0.015	-471652.083
Iberia_Chalcolithic	1.686	4630.625	0.005	0.037	-2351769.869	0.028	-2354249.543
Iberia_EN	4.875	7239.500	0.005	0.053	-1483274.628	0.030	-1485675.934
Iberia_MN	5.458	5765.000	0.004	0.039	-1491407.962	0.023	-1492793.179
Iberia_Mesolithic	21.838	7830.000	0.009	0.141	-720759.133	0.030	-723091.935
Karelia_HG	2.953	7265.000	0.008	0.125	-652952.676	0.033	-655352.439
LBK_EN	2.894	7123.429	0.007	0.039	-3656617.954	0.033	-3660838.639
Motala_HG	2.207	7729.500	0.003	0.126	-1477338.076	0.068	-1489573.895
Poltavka	2.211	4684.500	0.008	0.029	-1334662.071	0.020	-1335358.630
Potapovka	0.267	4076.500	0.004	0.063	-220112.816	0.011	-220251.379
Samara_Eneolithic	0.463	6615.000	0.007	0.078	-362161.674	0.020	-362689.209
Scythian_IA	3.217	2305.000	0.012	0.011	-492961.306	0.013	-492973.694
Srubnaya	1.662	3653.273	0.004	0.015	-2578065.957	0.013	-2578645.731
Srubnaya_Outlier	0.542	3704.500	0.006	0.019	-285828.766	0.008	-285851.523
Unetice_EBA	1.320	4024.786	0.002	0.012	-1676798.610	0.008	-1677026.310
Yamnaya_Samara	1.937	4990.500	0.008	0.033	-2440183.354	0.028	-2442192.801

Table 1: Details of populations included in analysis. “pop” is population name, “cov” is mean coverage of individuals in the population, “date” is mean date of individuals in the population, “ t_1 ” is the maximum likelihood estimate of t_1 in the full model, “ t_2 ” is the maximum likelihood estimate of t_2 in the full model, “LnL” is the maximum likelihood value in the full model, “ t_1 (cont)” is the maximum likelihood estimate of t_1 in the model where $t_2 = 0$, “LnL” is the maximum likelihood value in the model where $t_2 = 0$.