Assessing the maximum contribution from ancient populations–Supplementary Material

INTRODUCTION

The presentation here is intended to be mostly self contained and the notation differs slightly from that in the main text. We therefore repeat the set-up but with some small difference in notation.

A sample of size n_F from a particular locus in population F ("F" as in "focal") is traced back T units of coalescence time to population A ("A" as in "ancient"). Any population is here defined as the individuals living at a specific time point so that *e.g.* population F is no longer population F one generation back in time. We assume a biallelic marker with alleles u and "not u" ($\overline{u}$) and, without loss of generality, we define u to be the allele that has decreased in frequency from the ancient to the more recent sample (u can thus be either the ancestral or the derived allele). Following the lineages in population F backwards in time to population A, they are assigned to either u or $\overline{u}$ type according to a probability function generated by the observed (ancient DNA) sample drawn from population A. We denote the event of finding m_F alleles of type u and l_F alleles of type $\overline{u}$ in a sample of size $n_F = m_F + l_F$ from population F by $S_F = \{m_F, l_F\}$, the event of finding m_A alleles of type u and l_A alleles of type $\overline{u}$ in a sample of size $n_A = m_A + l_A$ from population A by $S_A = \{m_A, l_A\}$. The unknown population(s) will be designated by population X and the frequency of u in population X by U_X.

Software

The equations derived here are closed and arbitrarily exact, but cumbersome to calculate. A computer program package "MaxCon" that implements these formulas is available at

`www.ebc.uu.se/Research/IEG/evbiol/research/Jakobsson/software/MaxCon/.`

The programs are written in C++ and Python and takes as input m_F, n_F, m_A, n_A and a specified range of values for C and T for either when C is interpreted as the genetic contribution or when it is interpreted as the demographic contribution.

DERIVING THE PROBABILITY OF A RECENT SAMPLE GIVEN THE MODEL AND AN OLDER SAMPLE

(a) The probability of n genes having k ancestors t units of coalescent time ago ($g(k;n,t)$)

$g(k;n,t)$ can be found in Tavaré (1984):

$$g(k;n,t) = \begin{cases} 1 - \sum_{i=2}^{n} (-1)^i (2i-1) \frac{\binom{n}{i}}{\binom{n+i-1}{i}} e^{-\binom{i}{2}t} & \text{if } k = 1 \\ \sum_{i=k}^{n} (-1)^{i-k} \frac{2i-1}{k+i-1} \frac{\binom{n}{i}}{\binom{n+i-1}{i}} \binom{k+i-1}{k,k-1,i-k} e^{-\binom{i}{2}t} & \text{if } k \geq 2 \end{cases} \quad (1)$$

Here $0 < k \leq n$ and $t \geq 0$ is assumed. For $k = 0$ we define $g(0;0,t) = 1$ but $g(0;n,t) = 0$ when $n > 0$.

(b) The probability to draw a sample $\{\alpha, \beta\}$ conditioning on a previous sample $\{m, l\}$ ($s(\alpha, \beta; m, l)$)

Denoting the density for an allele X to have frequency x by $f_X(x)$,

$$s(\alpha, \beta; m, l) = \int_0^1 \binom{\alpha+\beta}{\alpha} x^\alpha (1-x)^\beta f_X(x|S = \{m, l\}) dx$$

Using Bayes' theorem, the posterior density for the frequency of an allele given that m copies of this allele in a sample of size $m + l$ is observed is

$$f_X(x|S = \{m, l\}) = P(S = \{m, l\}|X = x) \frac{f_X(x)}{P(S = \{m, l\})}$$

A beta prior for $f_X(x)$ with parameters a and b ($f_X(x) = x^{a-1}(1-x)^{b-1}\mathrm{B}(a, b)^{-1}$ where $\mathrm{B}(a, b)$ is the Beta function) implies that $f_X(x|S = \{m, l\})$ has a $Beta(m+a, l+b)$ distribution. $s(\alpha, \beta; m, l)$ is then a Beta-binomial distribution with parameters $m+b$ and $l+b$ and, in particular, assuming a uniform distribution for $f_X(x)$ (equivalent to setting $a = b = 1$):

$$s(\alpha, \beta; m, l) = \binom{\alpha+\beta}{\alpha} \frac{\mathrm{B}(\alpha + m + 1, \beta + l + 1)}{\mathrm{B}(m + 1, l + 1)} \tag{2}$$

If u is derived and assuming a pure drift model so that $f_X(x) = 1/x$ (see *e.g* Griffiths (2003) with $m(x)u_0(x) = 1/x$ in equation 28 in that article), then $f_X(x|S = \{m, l\})$ is $Beta(m, l+1)$ distributed. This leads to a Beta-binomial distribution with parameters m and $l + 1$:

$$s(\alpha, \beta; m, l) = \binom{\alpha+\beta}{\alpha} \frac{\mathrm{B}(\alpha + m, \beta + l + 1)}{\mathrm{B}(m, l + 1)} \tag{3}$$

Likewise, if u is ancestral we have $f_X(x|S = \{m, l\}) = x^m(1-x)^{l-1}\mathrm{B}(m + 1, l)^{-1}$ and

$$s(\alpha, \beta; m, l) = \binom{\alpha+\beta}{\alpha} \frac{\mathrm{B}(\alpha + m + 1, \beta + l)}{\mathrm{B}(m + 1, l)} \tag{4}$$

.

Comments: Note that $s(0, 0; m, l) = 1$ in all three cases and that equation (3) is assuming $m \geq 1$ while equation (4) is assuming $l \geq 1$.

(c) The probability of m u when there are $m + l$ lineages given that there were α u when there were $\alpha + \beta$ lineages ($h(m, l; \alpha, \beta)$)

In the standard coalescent process, the probability that there are m lineages of type u when there are $m+l$ lineages given that α lineages are of type u when there are $\alpha + \beta$ lineages is (Slatkin 1996)

$$h(m, l; \alpha, \beta) = \frac{\binom{m-1}{\alpha-1}\binom{l-1}{\beta-1}}{\binom{m+l-1}{\alpha+\beta-1}} \tag{5}$$

Comments: Note that this relies on $\binom{-1}{-1} \equiv 1$ and $\binom{k}{-1} \equiv 0$ for $k \geq 0$ so that (for instance) $h(0, l; \alpha, \beta)$ is larger than 0 only if $\alpha = 0$ and that $h(m, l; 0, \beta)$ is larger than 0 only if $m = 0$. Likewise, $\beta = 0$ implies $l = 0$.

MOTIVATING p_{con}

Here we present a more extensive motivation for p_{con} as a p-value, discuss an alternative p-value –p_{sup}– and compare some advantages and disadvantages of these two p-values. See also Bayarri and Berger (2000) for a more general discussion on the choice of p-value in the presence of a nuisance parameter.

Since specifying an alternative model is difficult, we use a classical null-hypothesis testing set-up (in the case of an explicit alternative model, a Bayesian approach could have been used). We formulate a null-model or null-hypothesis and want to test the compatibility of the observed data with this model using a suitable statistic. The null is then rejected if the probability of obtaining the observed value of the statistic –or larger– is below some threshold given the model. The natural choice for such a statistic is, in our case, the allele frequency difference between the two samples as this is the observation that initially suggested that some alternative population may have contributed to population F. There

are two pieces of observed data, each corresponding to an alternative choice for setting up the conditioning in order to calculate a p-value: there is the ancient sample configuration $S_A = \{m_A, l_A\}$ and the more recent sample configuration $S_F = \{m_F, l_F\}$. Hence, besides conditioning on the null, one can condition on either S_A or S_F (or neither) to calculate a p-value. Conditioning on S_F implies that one should calculate the probability of obtaining $S_A = \{i, n_A - i\}$ for all i such that $0 \leq i \leq n_A$ and $|i/n_A - m_F/n_F| \geq \Delta$ where $\Delta \equiv m_A/n_A - m_F/n_F$ (≥ 0 by definition of u). This has the advantage of being based on the more recent sample which is typically larger and more trustworthy than the ancient sample. However, a disadvantage is that the mutation that separates allele u from $\overline{u}$ may occur in the time between the two samples and one should (strictly speaking) take the mutation rate into account. As long as the ancient sample is polymorphic, this problem is circumvented by conditioning on the ancient sample, S_A. A third alternative corresponds to calculating the probability of observing a frequency difference $\geq \Delta$ without conditioning on either sample but this implies a considerable loss of statistical power. Here we settle on conditioning on S_A.

In technical terms, we perform a one-tailed test of a change in allele frequency that is greater than or equal to Δ. Conditioning on $S_A = \{m_A, l_A\}$, the probability of obtaining $S_F = \{i, n_F - i\}$ for all i such that $0 \leq i \leq n_F$ and $|m_A/n_A - i/n_F| \geq \Delta$ is calculated. This later inequality has two solutions:

$$\frac{m_A}{n_A} - \frac{i}{n_F} \geq \Delta \quad \Leftrightarrow \quad \frac{i}{n_F} \leq \frac{m_F}{n_F} \tag{6}$$

and

$$\frac{m_A}{n_A} - \frac{i}{n_F} \leq -\Delta \quad \Leftrightarrow \quad \frac{i}{n_F} \geq \frac{m_A}{n_A} + \Delta \tag{7}$$

While $i = 0$ always fulfil inequality (6), the second inequality may not be satisfied by any i (*e.g.* if $m_A/n_A + \Delta > 1$). In order to help the presentation and comparison of the two p-values, define the set of sample configurations that satisfy (6) and (7) as

$$\begin{aligned}
\mathbf{D_1} &\equiv \{\{i, n_F - i\} : i \in \{0, \cdots, n_F\} \wedge \frac{i}{n_F} \leq \frac{m_F}{n_F}\} \\
\mathbf{D_2} &\equiv \{\{i, n_F - i\} : i \in \{0, \cdots, n_F\} \wedge \frac{i}{n_F} \geq \frac{m_A}{n_A} + \Delta\} \\
\mathbf{D_-} &\equiv \{\{i, n_F - i\} : i \in \{0, \cdots, n_F\} \wedge \frac{i}{n_F} \leq \frac{m_A}{n_A}\}
\end{aligned} \tag{8}$$

where $\mathbf{D_-}$ is the set of modern sample configurations with a frequency of u lower or equal to the frequency of u in the ancient sample ($\mathbf{D_-}$ contains $\mathbf{D_1}$ but not $\mathbf{D_2}$ as a subset since $m_F/nF < m_A/n_A$). Also define

$$P_{tot}(t, c, p_X, \mathbf{D}) \equiv \sum_{s \in \mathbf{D}} P(S_F = s | M(T = t, C = c, U_X = p_X) \wedge S_A = \{m_A, l_A\}) \tag{9}$$

where

$$P(S_F = \{m, l\} | M(T = t, C = c, U_X = p_X) \wedge S_A = \{m_A, l_A\})$$

is the probability of obtaining a modern sample with m u alleles and l $\overline{u}$ alleles given the model $M(T = t, C = c, U_X = p_X)$ and the ancient sample configuration with $\mathbf{D}$ some set of sample configurations.

To the extent that we do not want to specify the unknown population(s), U_X, the frequency of u in the unknown population(s), is a nuisance parameter that we would like to get rid of. One approach would be to maximize the probability (9) text over $0 \leq U_X \leq 1$ leading to

$$p_{sup} \equiv \sup_{0 \leq U_X \leq 1} P_{tot}(t, c, U_X, \mathbf{D_1} \cup \mathbf{D_2}) \tag{10}$$

The maximum of P_{tot} over $0 \leq U_X \leq 1$ is basically always at either $U_X = 0$ or $U_X = 1$ (see "Values of U_X that maximizes the probability of the observation" below). This set-up to calculate a p-value is perhaps more natural than p_{con} but is sometimes too conservative in that the value of U_X that maximizes (9) is not always supported by the observed data. To illustrate, imagine that $S_F = \{0, 10\}$ and $S_A = \{4, 6\}$. Then $i = 0$ satisfy inequality (6) while $i = 8, 9, 10$ satisfy inequality (7). Obviously, the sum of the latter values can dominate $i = 0$ in equation (9) so that the maximum is found at $U_X = 1$. This is problematic because the observation $S_F = \{0, 10\}$ and $S_A = \{4, 6\}$ certainly does not suggest that U_X is anywhere close 1 since the unknown population needs to provide $\overline{u}$ alleles, not u alleles, to make the model consistent with the data. A concern here is that any information of U_X when conditioning on $M(t, c, p_X) \wedge S_A = \{m_A, l_A\}$ is in the observation S_F which belongs to the sample space. Thus, in order to gain information of U_X, we need to condition on the observed S_F

but only use part of this information when calculating a p-value (the full use of this information would lead to that we would know the outcome and the probability (9) would be either 1 or 0 depending on $\mathbf{D}$). Intuitively, the error in doing so should however be a loss of statistical power and in cases when p_{sup} provides even less power this constitutes a superior alternative. Therefore, to avoid summing over $\mathbf{D_2}$ (and only over $\mathbf{D_1}$), we condition on that the sample frequency of u is lower in S_F than in S_A and we arrive at

$$p_{con} \equiv \max_{0 \leq U_X \leq 1} P_{tot}(t, c, U_X, \mathbf{D_1}|\mathbf{D}_-) = \max_{0 \leq U_X \leq 1} \frac{P_{tot}(t, c, U_X, \mathbf{D_1})}{P_{tot}(t, c, U_X, \mathbf{D}_-)} \tag{11}$$

which is identical to the definition in the main text.

EFFICIENT CALCULATION

Based on the observation that the probability of there being k ancestral lineages left out of n at time t (function $g(k; n, t)$ derived in (Tavaré 1984)) is very close to zero for most values of k, (*e.g.* Jakobsson and Rosenberg 2007; Maruvka et al. 2011) we implemented a way to speed up calculations. This can be essential in cases where n_F (the size of the more recent sample) is large. For a given "error" ϵ, the implementation identifies a set of k-values $\mathbf{K} = \{k : K_{min} \leq k \leq K_{max}\}$ such that

$$\sum_{k \in \mathbf{K}} g(k; n, t) \geq 1 - \epsilon$$

Here K_{min} and K_{max} are found by the following algorithm:

1. Calculate $g(k; n, t)$ for $k = 1, \cdots, n$
2. Identify k^*, the k that maximizes $g(k; n, t)$, and set $p_{accumulated} = g(k^*; n, t)$ and $K_{min} = K_{max} = k^*$
3. While not $p_{accumulated} \geq 1 - \epsilon$,
   (a) if $K_{min} > 1$: put $K_{min} = K_{min} - 1$ and $p_{accumulated} = p_{accumulated} + g(K_{min}; n, t)$
   (a) if $K_{max} < n$: put $K_{max} = K_{max} + 1$ and $p_{accumulated} = p_{accumulated} + g(K_{max}; n, t)$

Error bounds for demographic contribution

In the case when C is interpreted as demographic contribution, the error made by ignoring factors that include $g(k; n, t)$ with $k \notin \mathbf{K}$ can be seen by noting that equation (8) in the main text can be written as

$$\sum_{k=1}^{n_F} g(k; n_F, t) \sum_{\alpha=0}^{k} h(m_F, l_F; \alpha, k - \alpha) \zeta(\alpha, k - \alpha; m_A, l_A, c, p_X)$$

noting that $h(m_F, l_F; \alpha, k - \alpha)$ will be zero for $\alpha > m_F$ and $k - \alpha > l_F$. This can be identified as

$$\sum_{k=1}^{n_F} g(k; n_F, t) Prob(S_F = \{m_F, l_F\}|\theta(k))$$

where $\theta(k)$ is the event that there were $n_F - k$ coalescent events among during the time t as well as $C = c$, $S_A = \{m_A, l_A\}$ and $U_X = p_X$. Equation (9) can then be written as

$$P_{tot}(t, c, p_X, \mathbf{D}) = \sum_{k=1}^{n_F} g(k; n_F, t) Prob(S_F \in \mathbf{D}|\theta(k))$$

$$= \sum_{k \in \mathbf{K}} g(k; n_F, t) Prob(S_F \in \mathbf{D}|\theta(k)) + \sum_{k \notin \mathbf{K}} g(k; n_F, t) Prob(S_F \in \mathbf{D}|\theta(k))$$

$$\leq P^*_{tot}(t, c, p_X, \mathbf{D}) + \sum_{k \notin \mathbf{K}} g(k; n_F, t) \leq P^*_{tot}(t, c, p_X, \mathbf{D}) + \epsilon$$

where $P^*_{tot}(t, c, p_X, \mathbf{D}) \equiv \sum_{k \in \mathbf{K}} g(k; n_F, t) Prob(S_F \in \mathbf{D}|\theta(k))$ is $P_{tot}(t, c, p_X, \mathbf{D})$ when ignoring factors with $g(k; n_F, t)$ and $k \notin \mathbf{K}$ which is always smaller than $P_{tot}(t, c, p_X, \mathbf{D})$. Thus

$$P_{tot}(t, c, p_X, \mathbf{D}) - \epsilon \leq P^*_{tot}(t, c, p_X, \mathbf{D}) \leq P_{tot}(t, c, p_X, \mathbf{D}) \tag{12}$$

Error bounds for genetic contribution

This case is slightly more complicated as the equations involve $g(k;r,t)$ terms with $1 \leq k \leq r$ and $1 \leq r \leq n$. Instead, define $\mathbf{K}(r)$ so that $\sum_{k \notin \mathbf{K}(r)} g(k;r,t) \leq \epsilon$. Noting that equation (5) in the main text can be written as

$$\Phi(\alpha,\beta;m_A,l_A,t) = \sum_{k=0}^{\alpha+\beta} g(k;\alpha+\beta,t) \sum_{i=0}^{min(k,\alpha)} h(\alpha,\beta;i,k-i)s(i,k-i;m_A,l_A)$$

so that equation (6) in the main text becomes

$$\sum_{r=0}^{n_F} \sum_{\alpha=0}^{min(r,m_F)} Bin(r;n_F,c)Bin(m_F-\alpha;n_F-r,p_X) \sum_{k=0}^{r} g(k;r,t) \sum_{i=0}^{min(k,\alpha)} h(\alpha,r-\alpha;i,k-i)s(i,k-i;m_A,l_A)$$

$$= \sum_{r=0}^{n_F} Bin(r;n_F,c) \sum_{k=0}^{r} g(k;r,t) \sum_{\alpha=0}^{min(r,m_F)} Bin(m_F-\alpha;n_F-r,p_X) \sum_{i=0}^{min(k,\alpha)} h(\alpha,r-\alpha;i,k-i)s(i,k-i;m_A,l_A)$$

Defining

$$f^*(\{m_F,l_F\},r,p_X,m_A,l_A) \equiv \sum_{\alpha=0}^{min(r,m_F)} Bin(m_F-\alpha;n_F-r,p_X) \sum_{i=0}^{min(k,\alpha)} h(\alpha,r-\alpha;i,k-i)s(i,k-i;m_A,l_A)$$

which is the probability of obtaining $S_F = \{m_F,l_F\}$ given that there are r (out of n_F) lineages tracing back to population A and that there were $r-k$ coalescent events among these r lineages during the time t so that $0 \leq f^*(\{m_F,l_F\},r,p_X,m_A,l_A) \leq 1$. Equation (9) can then be written as

$$P_{tot}(t,c,p_X,\mathbf{D}) = \sum_{s \in \mathbf{D}} \sum_{r=0}^{n_F} Bin(r;n_F,c) \sum_{k=0}^{r} g(k;r,t) f^*(s,r,p_X,m_A,l_A)$$

$$= \sum_{r=0}^{n_F} Bin(r;n_F,c) \sum_{k=0}^{r} g(k;r,t) \sum_{s \in \mathbf{D}} f^*(s,r,p_X,m_A,l_A)$$

which is identified as

$$= \sum_{r=0}^{n_F} Bin(r;n_F,c) \sum_{k=0}^{r} g(k;r,t) P(S_F \in \mathbf{D}|\theta(r,k))$$

where $\theta(r,k)$ is the event that there are r (out of n_F) lineages tracing back to population A and that there were $r-k$ coalescent events among these initial r lineages during the time t (as well as $C=c$, $S_A=\{m_A,l_A\}$ and $U_X=p_X$). Furthermore,

$$\sum_{r=0}^{n_F} Bin(r;n_F,c) \sum_{k \notin \mathbf{K}(r)} g(k;r,t) P(S_F \in \mathbf{D}|\theta(r,k)) \leq \sum_{r=0}^{n_F} Bin(r;n_F,c)\epsilon = \epsilon \sum_{r=0}^{n_F} Bin(r;n_F,c) = \epsilon$$

so that

$$P_{tot}(t,c,p_X,\mathbf{D}) - \epsilon \leq P^{**}_{tot}(t,c,p_X,\mathbf{D}) \leq P_{tot}(t,c,p_X,\mathbf{D}) \tag{13}$$

where $P^{**}_{tot}(t,c,p_X,\mathbf{D})$ is $P_{tot}(t,c,p_X,\mathbf{D})$ when only summing over $k \in \mathbf{K}(r)$.

Error bounds for p_{con}

The error when calculating p_{con} can be delimited by noting that p_{con} has the form (see definitions (11) and (8))

$$\frac{A+a}{A+a+B+b}$$

where

$$A = \sum_{k \in \mathbf{K}} g(k;n_F,t) Prob(S_F \in \mathbf{D_1} | \text{there were } k \text{ ancestral lineages at } t)$$

$$a = \sum_{k \notin \mathbf{K}} g(k; n_F, t) Prob(S_F \in \mathbf{D_1} | \text{there were } k \text{ ancestral lineages at } t) \leq \epsilon$$

$$B = \sum_{k \in \mathbf{K}} g(k; n_F, t) Prob(S_F \in \mathbf{D}_- \setminus \mathbf{D_1} | \text{there were } k \text{ ancestral lineages at } t)$$

$$b = \sum_{k \notin \mathbf{K}} g(k; n_F, t) Prob(S_F \in \mathbf{D}_- \setminus \mathbf{D_1} | \text{there were } k \text{ ancestral lineages at } t) \leq \epsilon$$

and (using that $d((A+x)/(A+B+x))/dx \geq 0$)

$$\frac{A}{A+B+\epsilon} \leq \frac{A}{A+B+b} \leq p_{con} \leq \frac{A+a}{A+B+a} \leq \frac{A+\epsilon}{A+B+\epsilon} \tag{14}$$

USING SIMULATIONS

To simulate under the demographic null model given $T = t$, $C = c$, $U_X = p_X$, and $S_A = \{m_A, l_A\}$

1. Simulate the number of ancestors to the sample from population F at $T = t$ by recursively drawing X_i from the exponential distribution with rate parameter $i(i-1)/2$ for $i = n_F, n_F - 1, \cdots$ until $X_{n_F} + X_{n_F-1} + \cdots + X_k > t$. Then k is the simulated number of ancestors.

2. The number among these k ancestors that are from population A, k_A, is drawn from $Bin(k, c)$, which gives the number of ancestors from population X as $k - k_A$.

3. The number of u alleles among the k ancestors, m_0, is drawn from $Bin(k-k_A, p_X) + BetaBinomial(m_A+1, l_A+1, k_A)$ (alternatively $Bin(k-k_A, p_X) + BetaBinomial(m_A, l_A+1, k_A)$ or $Bin(k-k_A, p_X) + BetaBinomial(m_A+1, l_A, k_A)$, see above).

4. Simulate m_F^{sim}, the sample from population F, by repeating (starting with $\alpha = k$ and $\beta = m_0$):

   - Draw X from $Bernoulli(\beta/\alpha)$
   - If $X = 1$, set $\beta = \beta + 1$
   - Set $\alpha = \alpha + 1$

   $n_F - k$ times. Set $m_F^{sim} = \beta$.

To simulate under the genetic null model given $T = t$, $C = c$, $U_X = p_X$, and $S_A = \{m_A, l_A\}$.

1. The number of lineages in the sample that trace back to population A, r is drawn from $Bin(n_F, c)$.

2. Simulate the number of ancestors to these r lineages at $T = t$ by recursively drawing X_i from the exponential distribution with rate parameter $i(i-1)/2$ for $i = r, r-1, \cdots$ until $X_r + X_{r-1} + \cdots + X_k > t$. Then k is the simulated number of ancestors.

3. The number of u alleles among the k ancestors, m_0, is drawn from $BetaBinomial(m_A + 1, l_A + 1, k)$ (alternatively $BetaBinomial(m_A, l_A + 1, k)$ or $BetaBinomial(m_A + 1, l_A, k)$, see appendix I above).

4. Starting with $\alpha = k$ and $\beta = m_0$:

   - Draw X from $Bernoulli(\beta/\alpha)$
   - If $X = 1$, set $\beta = \beta + 1$
   - Set $\alpha = \alpha + 1$

   $r - k$ times.

5. Draw m_F^{sim} from $\beta + Bin(n_F - r, p_X)$.

p_{con} given $U_X = p_X$ can then be estimated as the proportion of simulations for which $m_F^{sim} \leq m_F$ among those simulations where $m_F^{sim} \leq n_F \frac{m_A}{n_A}$.

p_{con} AT DIFFERENT LIMITS

Here we study what happens at different limits, especially when T is very large or very small. To simplify the presentation let $\{i, n_F - i\}$ denote a sample of size n_F from population F (the modern population) with i u-alleles and let $P_{\{i,n_F-i\}}$ denote the probability of obtaining such a sample (the notation for the conditioning is suppressed). Moreover, n and m will be used instead of n_F and m_F below.

Further recall that $\mathbf{D_1}$ includes all (modern) sample configurations $\{i, n-i\}$ with $i/n \leq m/n$ (always including $i = 0$) where $\{m, n-m\}$ is the observed modern sample configuration. We condition on the u-allele being less frequent in the modern sample than in the ancient sample and $\mathbf{D}_-$ are all (modern) sample configurations $\{i, n-i\}$ for which $i/n < m_A/n_A$ where $\{m_A, n_A - m_A\}$ is the observed ancient sample configuration (note that $i/n = m_A/n_A$ is not allowed). Define R to be the largest i such that $i/n < m_A/n_A$. Then an equivalent way of expressing p_{con} is

$$\max_{0 \leq U_X \leq 1} \frac{\sum_{i=0}^{m} P_{\{i,n-i\}}}{\sum_{i=0}^{R} P_{\{i,n-i\}}}$$

and we will study how

$$\frac{\sum_{i=0}^{m} P_{\{i,n-i\}}}{\sum_{i=0}^{R} P_{\{i,n-i\}}} \tag{15}$$

depends on U_X (the frequency of the u-allele in population X) and C (the contribution). For this purpose we assume that the frequency of the u-allele in the sampled ancient population A is fixed at U_A. Although we put a density on U_A conditional on the observed ancient sample configuration and integrate over this density in the calculation of p_{con}, whatever is true for all values of U_A is also true for p_{con}.

We show that at $T = 0$, equation (15) is maximized at $U_X = 0$ and that p_{con} is non-increasing in C for both demographic and genetic contribution (these are the same when $T = 0$). Next, we show that this is also true for demographic contribution when T is large. We then show that this is not necessarily the case for genetic contribution for $C > 0$ and T large: the value of U_X that maximizes equation (15) is not always at 1 for large T and $C \approx 1$ and equation (15) is increasing in C for large values of C if U_X is set to 0. We finish with a thorough search for cases where the value of U_X that maximizes equation (15) is not at $U_X = 0$ for small sample sizes and find no instances when this is the case for demographic contribution. For genetic contribution, the proportion of such cases levels off at a bit more than 10% as T grows but is a minor problem for small T.

Contribution at $T \approx 0$

At $T = 0$, there is no genetic drift, demographic and genetic contribution are the same and

$$\sum_{i=0}^{b} P_{\{i,n-i\}} = \sum_{i=0}^{b} \sum_{k=1}^{n} g(k,n,0) \sum_{j=0}^{k} Bin(j,k,x) h(i,n,i,j,k-j) = \sum_{i=0}^{b} Bin(i,n,x) \equiv F(b,n,x)$$

where $Bin(k,n,p) \equiv \binom{n}{k} p^k (1-p)^{n-k}$ and $x = CU_A + (1-C)U_X$ is the frequency of u. Furthermore

$$\frac{d}{dx} F(b,n,x) = -\frac{n-b}{1-x} Bin(b,n,x)$$

so that

$$\frac{d}{dx} \frac{\sum_{i=0}^{m} P_{\{i,n-i\}}}{\sum_{i=0}^{R} P_{\{i,n-i\}}} = \frac{d}{dx} \frac{F(m,n,x)}{F(R,n,x)} =$$

$$\frac{(n-R)Bin(R,n,x)F(m,n,x) - (n-m)Bin(m,n,x)F(R,n,x)}{(1-x)(F(R,n,x))^2}$$

Obviously the denominator is always positive and the sign is the same as the sign of the nominator. Since

$$(n-R)Bin(R,n,x)F(m,x) - (n-m)Bin(m,n,x)F(R,x)$$

$$= \sum_{i=0}^{m} x^{R+i}(1-x)^{2n-R-i} \left((n-R)\binom{n}{R}\binom{n}{i} - (n-m)\binom{n}{m}\binom{n}{R-m+i} \right) - F(R-m-1)$$

it is enough to show that

$$(n-m-D)\binom{n}{m+D}\binom{n}{i}-(n-m)\binom{n}{m}\binom{n}{D+i}\leq 0$$

where $D \equiv R-m \geq 0$. Below we show this for $0 \leq i \leq m$ which is clearly enough for proving that the nominator is negative and it follows that

$$\frac{d}{dx}\frac{F(m,n,x)}{F(R,n,x)} \leq 0$$

implying that $F(m,n,x)/F(R,n,x)$ is non-increasing in x. Thus, $U_X = 0$ minimizes $x = CU_A + (1-C)U_X$ so that

$$p_{con} = \frac{\sum_{i=0}^{m} Bin(i,n,CU_A)}{\sum_{i=0}^{R} Bin(i,n,CU_A)}$$

Furthermore, since the derivative of this expression *w.r.t.* C is obviously also non-positive, p_{con} is non-increasing in C. Since $n-m-D < n-m$, it is sufficient to show

$$\binom{n}{m+D}\binom{n}{i}-\binom{n}{m}\binom{n}{D+i}\leq 0$$

to prove that

$$(n-m-D)\binom{n}{m+D}\binom{n}{i}-(n-m)\binom{n}{m}\binom{n}{D+i}\leq 0$$

If this this were not the case there are combinations of values that fulfill

$$\binom{n}{m+D}\binom{n}{i} > \binom{n}{m}\binom{n}{D+i}$$

Logging both sides, this inequality can be written as

$$\sum_{j=n-m-D+1}^{n-m} log(j) + \sum_{j=i+1}^{D+i} log(j) > \sum_{j=m+1}^{m+D} log(j) + \sum_{j=n-i-D+1}^{n-i} log(j)$$

$$\Leftrightarrow$$

$$\sum_{j=0}^{D-1} log\frac{n-m-j}{n-i-j} > \sum_{j=0}^{D-1} log\frac{D+m-j}{D+i+j}$$

$$\Leftrightarrow$$

$$\prod_{j=0}^{D-1} \frac{n-m-j}{n-i-j}\frac{D+i-j}{D+m-j} > 1$$

This is not possible since

$$\frac{n-m-j}{n-i-j} \leq 1$$

and

$$\frac{D+i-j}{D+m-j} \leq 1$$

for $0 \leq i \leq m$.

Demographic contribution when T is large

There will only be a single ancestral lineage to the sample at $T \to \infty$ implying that $P_{\{i,n-i\}}$ will be non-zero only for $i = 0$ and $i = n$. In this case,

$$p_{con} = \max_{0 \le U_X \le 1} \frac{P_{\{0,n\}}}{P_{\{0,n\}}} = 1$$

since $R < n$ by definition.
The case when $T = t$ is such that the probability of there being more than 2 lineages left at T can be ignored can also be studied. For this purpose, designate the probability of there being 1 lineage left by v and the probability of there being 2 lineages left by $1 - v$. Letting x be the frequency of u in the merged population ($x = CU_A + (1 - C)U_X$), then

$$\sum_{i=0}^{b} P_{\{i,n-i\}} = \sum_{k=1}^{n} g(k,n,t) \sum_{i=0}^{k} Bin(i,k,x) \sum_{\alpha=0}^{b} h(\alpha, n-\alpha, i, k-i)$$

using the notation above. Since $g(1,n,t) = v$, $g(2,n,t) = 1 - v$ by assumption and because $h(\alpha, n-\alpha, 0, j) = 1$ for $\alpha = 0$ but 0 otherwise and since factors with $h(\alpha, n-\alpha, 2, 0)$ disappear since these are 0 for $n - \alpha > 0$ which is always the case for $b < n$ (note that this holds for $b = 0$):

$$\sum_{i=0}^{b} P_{\{i,n-i\}} = vBin(0,1,x) + (1-v)(Bin(0,2,x) + Bin(1,2,x) \sum_{\alpha=0}^{b} h(\alpha, n-\alpha, 1, 1))$$

$$= (1-x)(v + (1-v)(1 - x + 2x \sum_{\alpha=0}^{b} h(\alpha, n-\alpha, 1, 1))) = (1-x)(1 - x(1-v)(1 - 2\frac{b}{n-1}))$$

Here we used that $h(0,n,1,1) = 0$ and $h(\alpha, n-\alpha, 1, 1) = 1/(n-1)$ for $\alpha > 0$. Note that this holds also for $b = 0$. Thus

$$\frac{\sum_{i=0}^{m} P_{\{i,n-i\}}}{\sum_{i=0}^{R} P_{\{i,n-i\}}} = \frac{1 - x(1-v)(1 - 2\frac{m}{n-1})}{1 - x(1-v)(1 - 2\frac{R}{n-1})}$$

and

$$\frac{d}{dx} \frac{\sum_{i=0}^{m} P_{\{i,n-i\}}}{\sum_{i=0}^{R} P_{\{i,n-i\}}} = \frac{2(1-v)(m-R)}{(n-1)(1 - x(1-v)(1 - 2\frac{R}{n-1}))^2}$$

which is never positive since $m \le R$.
As above, this shows that for $0 \le m \le R$, the maximum of $\sum_{i=0}^{m} P_{\{i,n-i\}} / \sum_{i=0}^{R} P_{\{i,n-i\}}$ is obtained at $x = CU_A + (1 - C)0 = CU_A$ and that p_{con} is a non-increasing function in C.

Genetic contribution at $C \approx 0$

Recall that the genes in the sample from the modern population F are first partitioned into those that eventually reach population A and those that do not. A single gene in the sample traces its ancestry back to population A with probability C (the genetic contribution). When $C = 0$ no lineages in the sample from population F has an ancestry in population A and

$$P_{\{\alpha,n-\alpha\}} = Bin(\alpha, n, U_X)$$

This is analogous to the case $T = 0$ above so that $U_X = 0$ maximizes equation (15).

Genetic contribution at $C \approx 1$ and $T \to \infty$

Here we study the case when C is very close to 1 such that the event of more than 1 gene *not* having ancestry in population A is negligible. The probability of all genes in the sample tracing back to population A is denoted by v and the probability of all but one gene having ancestry in population A by $1 - v$. Note that $P_{\{i,n-i\}}$ will be zero for $2 \le i \le n-2$ because, for these values of i, more than 2 lineages need to enter population X as all lineages that trace their ancestry to population A will be of the same allelic type. Thus, the only values of i to consider are $i = 0, 1, n-1$ (n need not be considered as $m \le R < n$) and

$$P_{\{0,n\}} = P_{\{0,n\}^A \wedge \{0,0\}^X} + P_{\{0,n-1\}^A \wedge \{0,1\}^X} = v(1-U_A) + (1-v)(1-U_A)(1-U_X)$$

$$P_{\{1,n-1\}} = P_{\{0,n-1\}^A \wedge \{1,0\}^X} = (1-v)(1-U_A)U_X$$
$$P_{\{n-1,1\}} = P_{\{n-1,0\}^A \wedge \{0,1\}^X} = (1-v)U_A(1-U_X)$$

where $P_{\{i,k-i\}^A \wedge \{j,n-k-j\}^X}$ designates the probability of k genes (out of the n) tracing their ancestry back to population A and that, among these k genes, i are of allelic type u while there are j genes of allelic type u among the $n-k$ genes that do not have an ancestry in population A. Further note that $P_{\{0,n\}} + P_{\{1,n-1\}} = 1 - U_A$ and $P_{\{n-1,1\}} + P_{\{n,0\}} = U_A$. For $0 < m < R$:

$$\sum_{i=0}^{m} P_{\{i,n-i\}} = P_{\{0,n\}} + P_{\{1,n-1\}} = 1 - U_A$$

and

$$\sum_{i=0}^{R} P_{\{i,n-i\}} = \begin{cases} 1-U_A & R < n-1 \\ 1-U_A+(1-v)U_A(1-U_X) & R = n-1 \end{cases}$$

so that $p_{con} = 1$ for $R < n-1$.
If $R = n-1$,

$$p_{con} = \max_{0 \le U_X \le 1} \frac{1-U_A}{1-U_A+(1-v)U_A(1-U_X)} = 1$$

where the maximum is found at $U_X = 1$. Hence $p_{con} = 1$ for $0 < m < R$.
For $m = 0$ and $R < n-1$,

$$p_{con} = \max_{0 \le U_X \le 1} \frac{v(1-U_A)+(1-v)(1-U_A)(1-U_X)}{1-U_A} = \max_{0 \le U_X \le 1} 1 - U_x(1-v) = 1$$

where the maximum is found at $U_X = 0$. The case $m = 0$ and $R = n-1$ is slightly more complicated but p_{con} is close to 1 for v close to 1. The derivative of the ratio $\sum_{i=0}^{m} P_{\{i,n-i\}} / \sum_{i=0}^{R} P_{\{i,n-i\}}$ *w.r.t.* U_X is negative if $v < (1-U_A)/U_A$ and positive for $v > (1-U_A)/U_A$ so that the maximizing value of U_X is at 0 if $v < (1-U_A)/U_A$. Since v is typically close to 1 and since $R = n-1$ should imply a large U_A (and small $(1-U_A)/U_A$), this suggests that $U_X = 1$ is the maximizing value of U_X.
The case of $m = R$ gives (trivially) $p_{con} = 1$.
Thus $p_{con} = 1$ at $T \to \infty$ and C close to 1 but the maximizing value of U_X is at 1 for $m > 0$ and depends on R and the combination of C and U_A for $m = 0$.

Genetic contribution at $U_X = 0$ and $T \to \infty$

When $U_X = 0$ and $T \to \infty$, no genes from population X can be of type u while all genes with an ancestry in population A are of the same allelic type. Thus, using the notation above

$$p_{con} = \frac{\sum_{i=0}^{m} P_{\{i,n-i\}}}{\sum_{i=0}^{R} P_{\{i,n-i\}}}$$

and

$$P_{\{i,n-i\}} = \begin{cases} \sum_{j=0}^{n} P_{\{0,j\}^A \wedge \{0,n-j\}^X} & i = 0 \\ P_{\{i,0\}^A \wedge \{0,n-i\}^X} & i > 0 \end{cases}$$
$$= \begin{cases} (1-U_A)\sum_{j=1}^{n} Bin(j,n,C) + Bin(0,n,C) & i = 0 \\ U_A Bin(i,n,C) & i > 0 \end{cases}$$
$$= \begin{cases} 1-U_A+U_A Bin(0,n,C) & i = 0 \\ U_A Bin(i,n,C) & i > 0 \end{cases}$$

so that

$$\sum_{i=0}^{b} P_{\{i,n-i\}} = 1 - U_A + U_A \sum_{i=0}^{b} Bin(i,n,C)$$

Furthermore,

$$\frac{d\sum_{i=0}^{b} P_{\{i,n-i\}}}{dC} = -U_A \frac{n-b}{1-C} Bin(b,n,C)$$

so that

$$\frac{d}{dC}\frac{\sum_{i=0}^{m} P_{\{i,n-i\}}}{\sum_{i=0}^{R} P_{\{i,n-i\}}} = \frac{U_A}{1-C}\frac{(n-R)Bin(R,n,C)\sum_{i=0}^{m} P_{\{i,n-i\}} - (n-m)Bin(m,n,C)\sum_{i=0}^{R} P_{\{i,n-i\}}}{(\sum_{i=0}^{R} P_{\{i,n-i\}})^2}$$

Here the denominator is never negative so the nominator determines the sign. First we find a bound such that if C is smaller than this the derivative is negative:

$$(n-R)Bin(R,n,C)\sum_{i=0}^{m} P_{\{i,n-i\}} - (n-m)Bin(m,n,C)\sum_{i=0}^{R} P_{\{i,n-i\}}$$

$$\leq (n-R)Bin(R,n,C)\sum_{i=0}^{R} P_{\{i,n-i\}} - (n-m)Bin(m,n,C)\sum_{i=0}^{R} P_{\{i,n-i\}}$$

which is negative if

$$(n-R)Bin(R,n,C) < (n-m)Bin(m,n,C)$$

$$\Leftrightarrow$$

$$\left(\frac{C}{1-C}\right)^{R-m} < A$$

$$\Leftrightarrow$$

$$C < \frac{A^{\frac{1}{R-m}}}{1+A^{\frac{1}{R-m}}}$$

where

$$A \equiv \frac{(n-m)\binom{n}{m}}{(n-R)\binom{n}{R}}$$

Next we find a bound such that if C is larger than this the derivative is positive:

$$(n-R)Bin(R,n,C)\sum_{i=0}^{m} P_{\{i,n-i\}} - (n-m)Bin(m,n,C)\sum_{i=0}^{R} P_{\{i,n-i\}}$$

$$\geq (n-R)Bin(R,n,C)(1-U_A) - (n-m)Bin(m,n,C)\sum_{i=0}^{R} P_{\{i,n-i\}}$$

$$\geq (n-R)Bin(R,n,C)(1-U_A) - (n-m)Bin(m,n,C)\sum_{i=0}^{n} P_{\{i,n-i\}}$$

$$= (n-R)Bin(R,n,C)(1-U_A) - (n-m)Bin(m,n,C)$$

which is positive if

$$(n-R)Bin(R,n,C)(1-U_A) > (n-m)Bin(m,n,C)$$

$$\Leftrightarrow$$

$$\left(\frac{C}{1-C}\right)^{R-m} > B$$

$$\Leftrightarrow$$

$$C > \frac{B^{\frac{1}{R-m}}}{1+B^{\frac{1}{R-m}}}$$

where

$$B \equiv \frac{(n-m)\binom{n}{m}}{(1-U_A)(n-R)\binom{n}{R}} = A/(1-U_A)$$

To illustrate, in the Woolly mammoth example $n = 61$, $m = 0$, $R = 14$ and the distribution of U_A is centered around $11/47$ so that the derivative is negative if C is less than approximately 0.324 but positive if C is larger than approximately 0.325.

Searching for the maximizing value of U_X

We searched for the value of U_X among $0, 0.05, \cdots, 0.95, 1$ that maximizes the expression (15). We tried all possible combinations of $C \in \{0.1, 0.2, \cdots, 0.9\}$ ($C = 0$ and $C = 1$ are trivial cases), and $\{n_F, n_A\} \in \{\{5,5\},\{4,4\},\{4,3\},\{3,4\},\{3,3\}\}$ with $0 \leq m < n_F$ and $0 < m_A \leq n_A$ for which $m/n_F < m_A/n_A$ (u is defined to be more common in population A). The three considered priors for the frequency of u in population A were investigated: i) a uniform prior (441 cases), ii) assuming that u is the derived variant using (3) (270 cases since $m_A = 0$ is not allowed), and iii) assuming that u was the ancestral variant using (4) (441 cases). This amounts to a total of 1152 number of cases.

Both interpretation of C were investigated: C interpreted as demographic contribution, and C interpreted as genetic contribution. For C interpreted as demographic contribution we investigated the times $T = 0, 0.2, \cdots, 1.8, 2.0$ and found no case when $U_X > 0$ maximized the value of (15). For C interpreted as genetic contribution we investigated the times $T = 0, 0.01, \cdots, 0.19, 0.2$ as well as $T = 0.2, 0.3, \cdots, 2.9, 3.0$ and $T = 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0$. We found no cases of when $U_X > 0$ maximized the value of (15) for $T = 0, 0.01, \cdots 0.17$ but from that point on the number of cases increased until it reached a steady level of 155 (out of 1152) at $T \approx 3.0$ (figures 1 and 2).

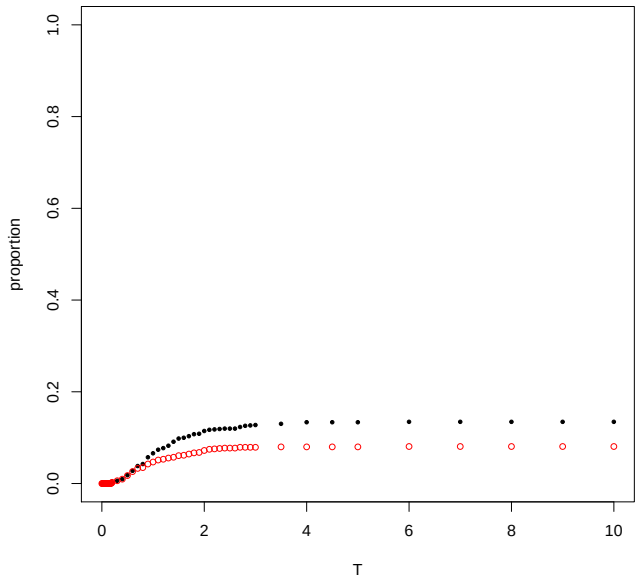

Figure 1: The proportion of cases for genetic contribution when the value of U_X that maximizes equation (15) is not at $U_X = 0$ (black dots) or neither at $U_X = 0$ nor $U_X = 1$ (red circles).

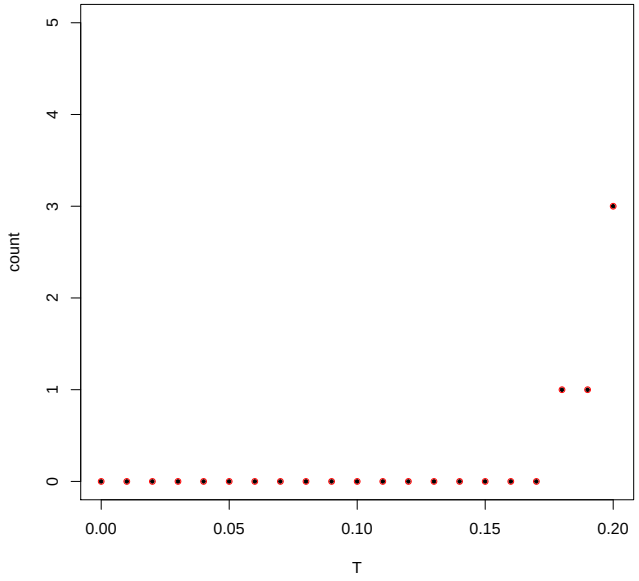

Figure 2: The number of cases for genetic contribution when the value of U_X that maximizes equation (15) is not at $U_X = 0$ (black dots) or neither at $U_X = 0$ nor $U_X = 1$ (red circles).

VALUES OF U_X THAT MAXIMIZES THE PROBABILITY OF THE OBSERVATION

This is relevant for p_{sup} defined above. We searched for the value of U_X among $0, 0.05, \cdots, 0.95, 1$ that maximizes the expression (4) in the main text. We tried all possible combinations of $T \in \{0, 0.0001, 0.01, 1.0\}$, $C \in \{0.1, 0.2, \cdots, 0.9\}$ ($C = 0$ and $C = 1$ are trivial cases), and $\{n_F, n_A\} \in \{\{5,5\},\{4,4\},\{4,3\},\{3,4\},\{3,3\}\}$ with $0 \leq m_F \leq n_F$ and $0 \leq m_A \leq n_A$ for which $m_F/n_F \leq m_A/n_A$ (u is defined to be more common in population A). This investigation was performed for the two cases; C interpreted as the demographic contribution, and C interpreted as the genetic contribution. The same investigation was performed using a uniform prior for the frequency of u, assuming that u was the derived variant using (3), and assuming that u was the ancestral variant using (4).

When C is interpreted as demographic contribution, the values of U_X that maximizes equation (4) in the main text were exclusively either $U_X = 0$ or $U_X = 1$. However there were cases when the maximum U_X varied with C. For instance, for a uniform prior and $m_F = 0$, $n_F = 3$, $m_A = 1$, $n_A = 4$, and $T = 0, 0.0001, 0.01$ (but not $T = 1.0$), the value of U_X that maximize the probability is $U_X = 1$ for $C \leq 0.7$ but $U_X = 0$ for $C \geq 0.8$. We investigated this transition point in more detail and found that the maximum probability is found at $U_X = 1$ for $C \leq 0.7777$ but at $U_X = 0$ for $C \geq 0.7778$ when $T = 0$; $U_X = 1$ for $C \leq 0.7776$ but at $U_X = 0$ for $C \geq 0.7777$ when $T = 0.0001$; $U_X = 1$ for $C \leq 0.7699$ but at $U_X = 0$ for $C \geq 0.7700$ when $T = 0.01$. We found no transition point for $T = 1$ and even for $C = 10^{-8}$, the maximum probability was found at $U_X = 0$ (for $C = 0$, equation (4) in the main text is equal to 1 for both $U_X = 0$ and $U_X = 1$ when $m_F = 0$, $n_F = 3$, $m_A = 1$, and $n_A = 4$). When $T = 0.5$, the transition point was found between $C = 0.1095$ and $C = 0.1096$. In summary, although the transition point is dependent on time, in no case was there a value of $0 < U_X < 1$ that maximized the probability in equation (4) in the main text.

Under the genetic contribution model there were exceptions to this rule if C, see Table 1. However, only 19 out of 6300 exceptions ($< 0.31\%$) were found and they were all found for large values of T suggesting that either $U_X = 0$ or $U_X = 1$ is a reasonable assumption.

Table 1: The values of U_X (among $U_X = 0, 0.05, \cdots, 0.95, 1$) that maximizes the value of equation (4) in the main text over the whole range of C for the choices of m_F, n_F, m_A, n_A and T where exceptions were found. In all cases, C is interpreted as the genetic contribution.

						C								
prior for U_x	m_F	n_F	m_A	n_A	T	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
uniform	1	5	2	5	1.0	1	1	1	1	1	0	0.05	0.05	0.05
uniform	2	5	3	5	1.0	0	0	0	0	0	1	0.95	0.95	0.95
uniform	1	4	2	4	1.0	1	1	1	1	1	1	0.5	0.5	0.5
U derived	2	5	3	5	1.0	0	0	0	0	0	0	0.85	0.85	0.9
U derived	1	4	2	4	1.0	0	0	0	0	0	0	0	0.15	0.25
U ancestral	1	5	2	5	1.0	1	1	1	1	1	1	0.15	0.15	0.1
U ancestral	1	4	2	4	1.0	1	1	1	1	1	1	1	0.85	0.75

REFERENCES

Bayarri MJ, Berger JO. 2000. P Values for Composite Null Models. *Journal of the American Statistical Association.* 95:1127–1142.

Griffiths RC. 2003. The frequency spectrum of a mutation, and its age, in a general diffusion model. *Theor. Popul. Biol.* 64:241–251.

Jakobsson M, Rosenberg NA. 2007. The probability distribution under a population divergence model of the number of genetic founding lineages of a population or species. *Theor. Popul. Biol.* 71:502–523.

Maruvka YE, Shnerb NM, Bar-Yam Y, Wakeley J. 2011. Recovering Population Parameters from a Single Gene Genealogy: An Unbiased Estimator of the Growth Rate. *Mol. Biol. Evol.* 28:1617–1631.

Slatkin M. 1996. Gene genealogies within mutant allelic classes. *Genetics.* 143:579–587.

Tavaré S. 1984. Line-of-Descent and Genealogical Processes, and Their Apllications in Population Genetics Models. *Theor. Popul. Biol.* 26:119–164.