Supporting Information

File S1: Visible Transitions

Figure SA shows an example of a visible transition when the coalescent topology remains the same and Figure SB shows an example of a visible transition when the coalescent topology changes (coalescence between the (a,(b,c)) branch and the (d,e) branch happens after coalescence of (f) and (g) on the left tree and before coalescence of (f) and (g) on the right tree.) Green lines mark the possible pruning locations that could have lead to the same visible transition; the red circle indicates the deleted node at coalescent time t_{del} and the blue circle indicates the new node created at coalescent time t_{new}.

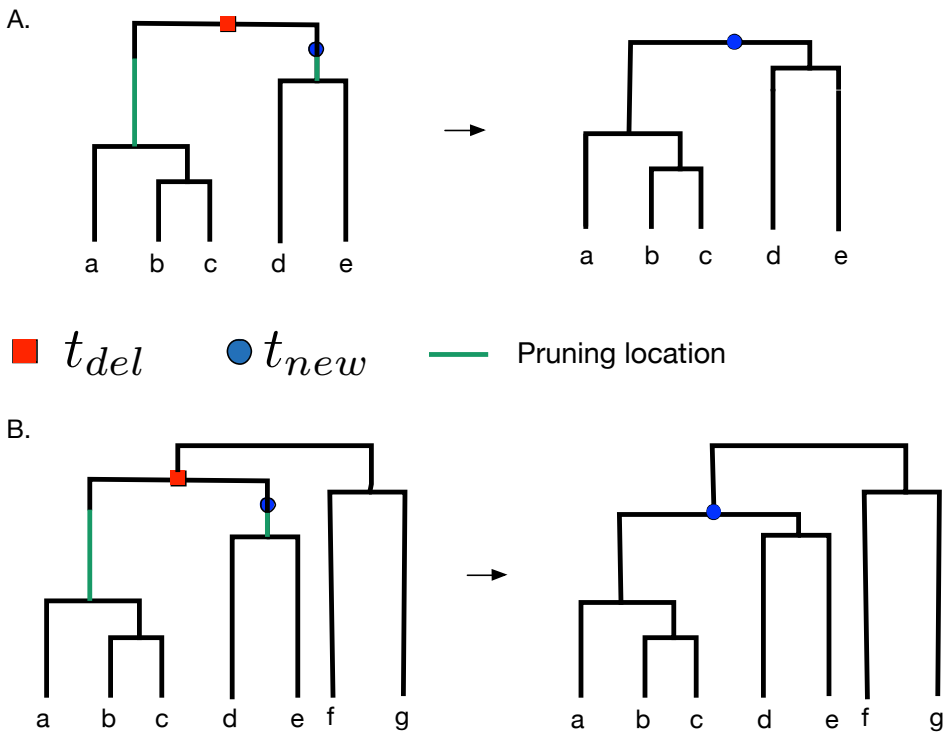

Figure S1: Examples of visible transitions when the pruning branch is uncertain. Red circle indicates deleted node at coalescent time t_{del}, blue circle indicates new node at coalescent time t_{new}. Green lines indicates possible pruning locations that could have resulted in such a visible transition. A. The topology remains the same. B. The topology changes.

File S2: Visible transitions between Tajima's genealogies

A Tajima's genealogy g^T corresponds to the pair of coalescent times and a ranked tree shape with n tips (i.e. with no labels but ranked coalescent events). In Figure S, we show four possible visible transitions. In the first case (Figure SA), when we compare the number of *children* of the blue circle node on the right tree at time t with the *children* of the red circle node on the left tree, we can conclude that only the green branch could have been selected for pruning. In Figure SB, comparing the *children* of the blue circle node on the right genealogy to the *children* of the red circle in the left genealogy, we conclude that the two *children* of the red circle are possible pruning locations. In Figures SC-D, $t_{new} < t_{del}$. This implies that the possible pruning locations will necessarily have heights up to t_{new}. Again, by comparing the *children* of the blue circle node on the right to the *children* of the red circle node on the left, we can asses the possible pruning locations.

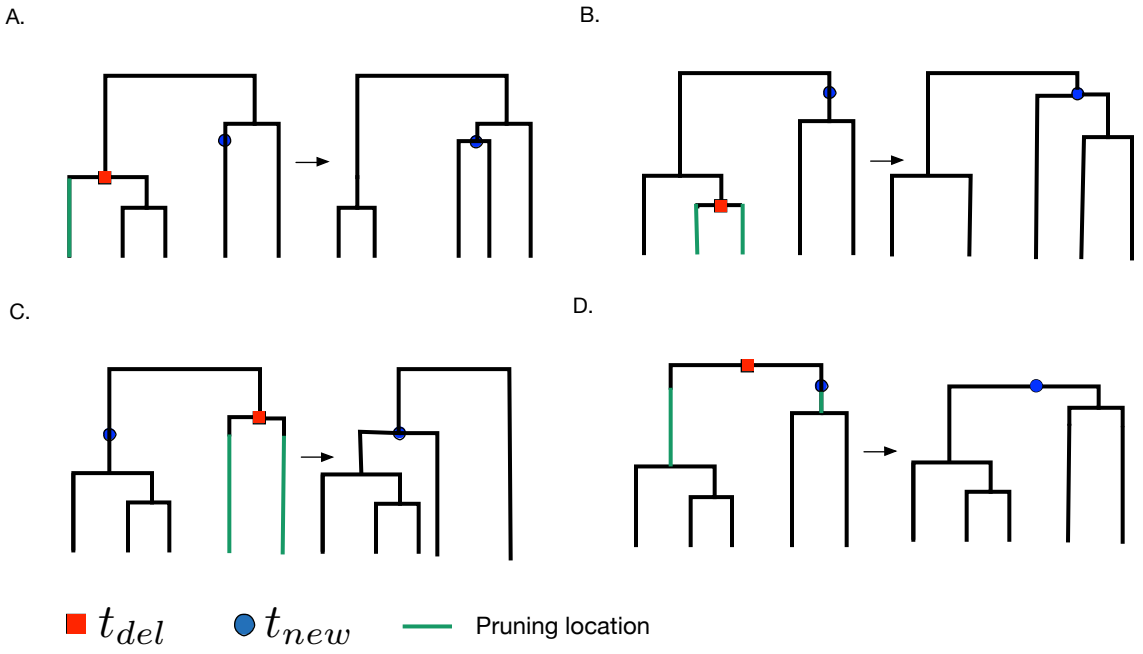

Figure S2: Examples of visible transitions between local Tajima's genealogies. Red circle indicates deleted node at coalescent time t_{del}, blue circle indicates new node at coalescent time t_{new}. Green lines indicates possible pruning locations that could have resulted in such a visible transition.

File S3: Simulations with `MaCS`

We use `MaCS` (Chen et al., 2009) for all our simulations with the following code lines:

Constant population size:

`./macs2` 300000 -t 1.0 -T -r .005 -h 1 (SEED: 1420480396)
`./macs20` 3000000 -t 1.0 -T -r .0002 -h 1 (SEED: 1399175725)
`./macs100` 3000000 -t 1.0 -T -r .0002 -h 1 (SEED: 1400528079)

Exponential growth and constant:

`./macs2` 300000 -t 1.0 -eG .1 10 -T -r .02 -h 1 (SEED: 1419985269)
`./macs20` 300000 -t 4.0 -eG .1 10 -T -r .002 -h 1 (SEED: 1420040333)
`./macs100` 300000 -t 1.0 -eG .1 10 -T -r .0002 -h 1 (SEED: 1401855826)

Bottleneck:

`./macs2` 300000 -t 4.0 -eN 0 1 -eN 0.3 0.1 -eN 0.5 1 -T -r .01 -h 1 (SEED: 1420824821)
`./macs20` 300000 -t 4.0 -eN 0 1 -eN 0.3 0.1 -eN 0.5 1 -T -r .002 -h 1 (SEED: 1420826310)
`./macs100` 300000 -t 4.0 -eN 0 1 -eN 0.3 0.1 -eN 0.5 1 -T -r .001 -h 1 (SEED: 1420826409)

File S4: EM sensitivity to parameter discretization

In Figure S3, we show EM estimates of a constant population size from 1000 local genealogies of 100 individuals. We show that different discretizations result in different estimates. We note that confidence intervals perform poorly in terms of coverage. The performance statistics corresponding to the three estimations displayed in Figure S3 are shown in Table S1.

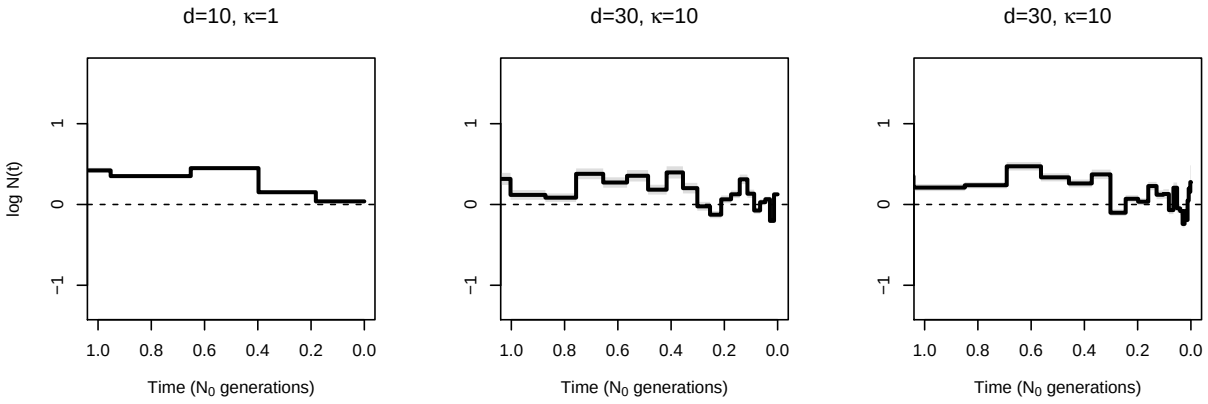

Figure S3: **EM sensitivity to parameter discretization.** Comparison of population size trajectories estimated from 1000 simulated genealogies ($m = 1000$) of 100 individuals with a constant population size. EM inference with different discretizations varying the parameters in Equation 15.

Table S1: *Summary of simulation results depicted in Figure S3. SRE is the sum of relative errors (Equation 24), MRW is the mean relative width of the 95% BCI (Equation 25), and ENV (Equation 26).*

	SRE	MRW	ENV
EM $d = 10$, $\kappa = 10$	43.41	0.99	48.6%
EM $d = 30$, $\kappa = 10$	34.25	0.76	42.6%
EM $d = 30$, $\kappa = 100$	43.96	0.99	46.0%

File S5: Analysis of Human data

We use *ARGweaver* (Rasmussen et al., 2014) with the following code lines:

European population:

```
arg-sample -s data1000/CEU_10.sites
   -N 11534 -r 1.6e-8 -m 1.26e-8
   --ntimes 200 --maxtime 200e3 -c 1 -n 10
   -o data1000/CEU.sample/out
```

Yoruban population:

```
arg-sample -s data1000/YRI_10.sites
   -N 11534 -r 1.6e-8 -m 1.26e-8
   --ntimes 200 --maxtime 200e3 -c 1 -n 10
   -o data1000/YRI.sample/out
```

ARGweaver time is measured in units of generations, so in order to generate Figure 8, we multiplied time by $1/(2 \times 11,534)$. To obtain $\log N(t)$ displayed in Figure S4, we multiplied our estimates by $1/(8 \times 11,532)$ and converted them in logarithmic scale.

We note that *ARGweaver* assumes the SMC and not the SMC$'$ model so our estimates of $N(t)$ are biased. One source of such a bias is that the $A^i(t)$ functions that indicate the number of lineages present at time t in the SMC$'$ are replaced by $A^i(t) - 1$ if the pruning branch is present at time t in the SMC. Another source of bias is the lack of invisible recombination events in *ARGweaver* realizations. To approximate the effect of this difference we re-run our algorithm replacing $A^i(t)$ by $A^i(t) - 1$. Figure S4 shows that the main conclusions about inferred recent past population sizes remain valid; in our analysis of human data (Figure 8) we only focus on the recent past.

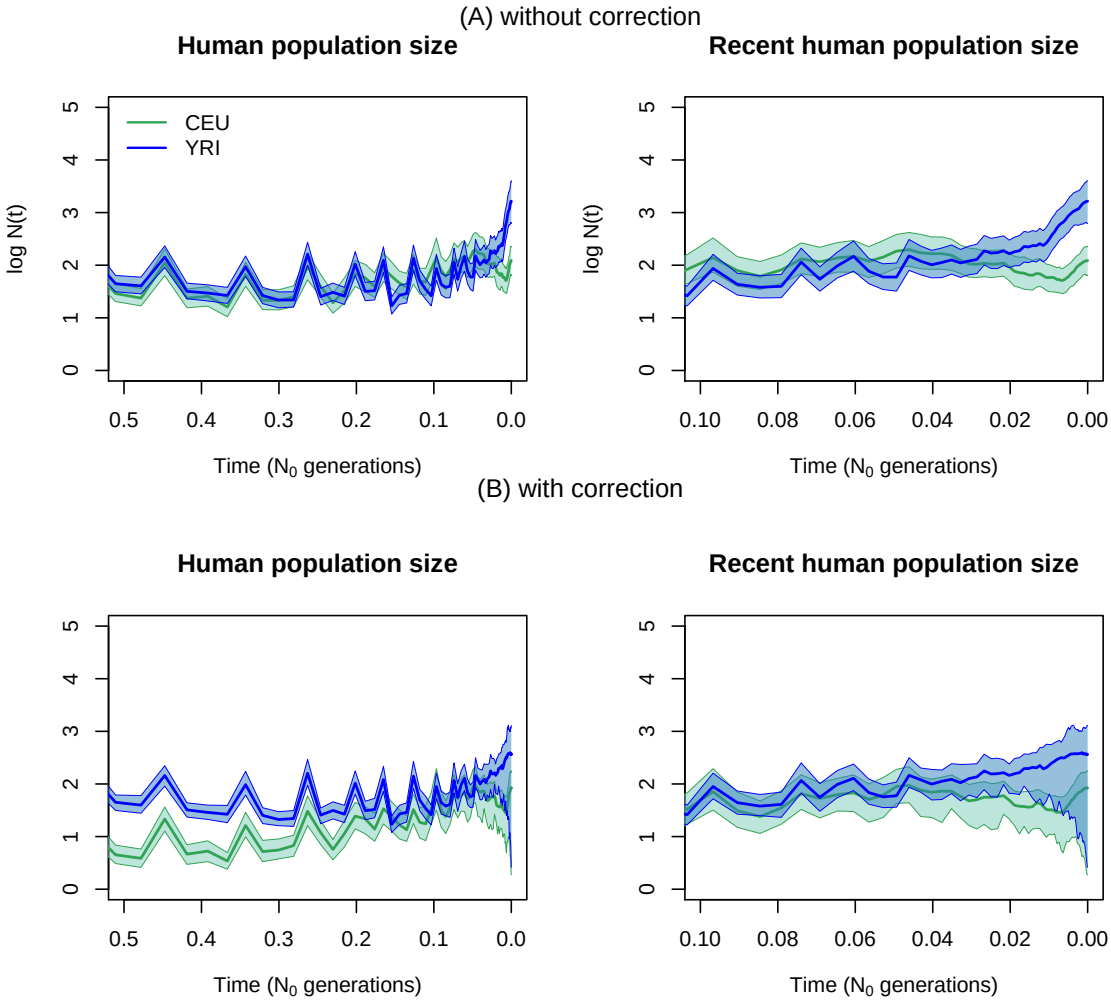

Figure S4: **Inference of human population size trajectories** $N(t)$ **for** $n = 10$**.** Green solid line and green shaded areas represent the posterior median and 95% BCI for European population (CEU) and blue solid line and blue shaded areas represent the posterior median and 95% BCI for Yoruban population. **(A) Without correction.** We ignore the fact that our genealogies were generated assuming the SMC process instead of SMC′. **(B) With correction.** We corrected the function of the number of lineages to approximate the SMC likelihood. Figures on the right show the same results as in the left side for the recent past $(0,0.1N_0)$.

File S6: Sampling more individuals on simulations

In Figure S5 we re-arrange our results on simulations shown in Figures 5−7 to compare our estimations when increasing the number of samples. We find that increasing n does not necessarily improve estimation from 1000 local genealogies.

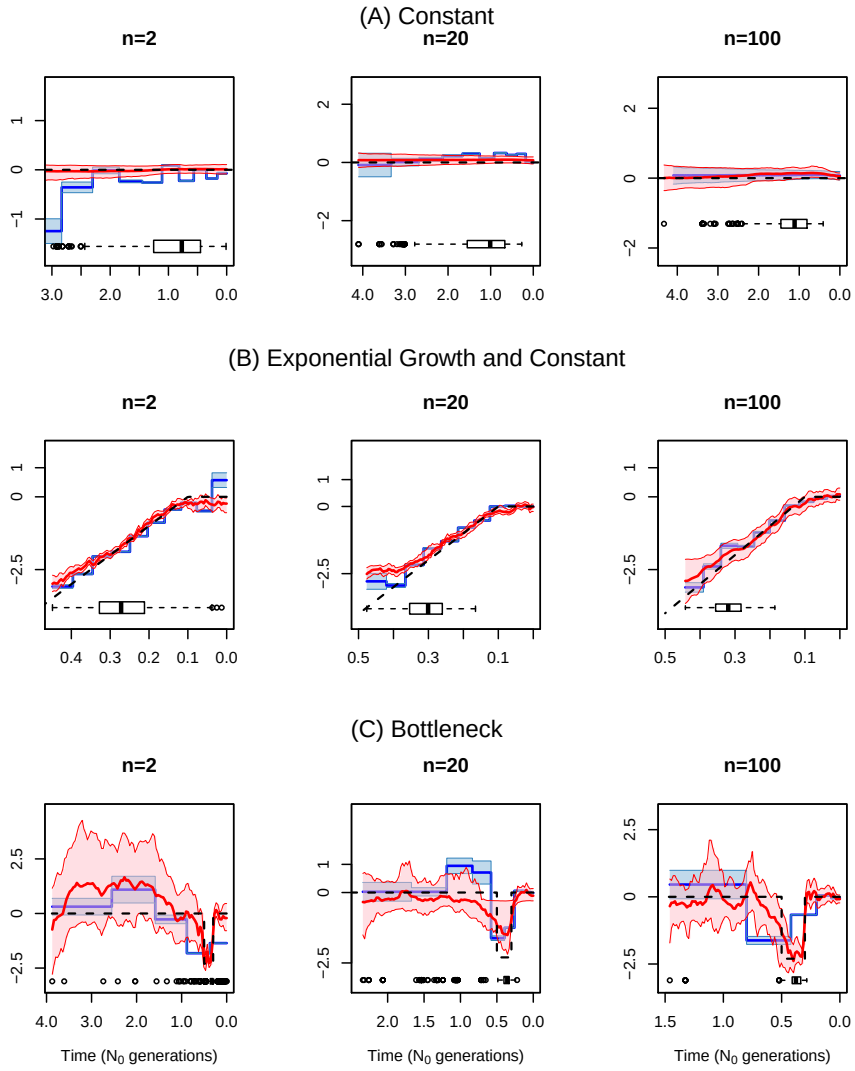

Figure S5: **Comparison of population size trajectories $N(t)$ inferred varying the number of samples from** 1000 **simulated local genealogies** (A) Simulated data under constant population size, (B) exponential and constant trajectory, and (C) a bottleneck. We rearrange the plots displayed in Figures 5−7 corresponding to 1000 simulated genealogies. Red curves are pink areas correspond to our Bayesian GP estimates and blue curves and areas correspond to the EM estimates.

File S7: Fisher Information Calculation

The calculation of the Fisher information needed to estimate confidence intervals of a piecewise constant trajectory of population sizes, requires the following expected values:

$$\mathrm{E}[z_j^i z_k^l \mid \mathbf{Y}] = \begin{cases} \mathrm{E}[z_j^i \mid \mathbf{Y}] & j = k, i = l \\ 0 & j \neq k, i = l \\ \mathrm{E}[z_j^i \mid \mathbf{Y}]\mathrm{E}[z_k^l \mid \mathbf{Y}] & i \neq l \end{cases}$$

$$\mathrm{E}[\Delta_j^i \Delta_k^l \mid \mathbf{Y}] = \begin{cases} \Delta_{j,k}^i & i = l \\ \mathrm{E}[\Delta_j^i \mid \mathbf{Y}]\mathrm{E}[\Delta_k^l \mid \mathbf{Y}] & i \neq l \end{cases}$$

$$\mathrm{E}[z_j^i \Delta_k^l \mid \mathbf{Y}] = \begin{cases} \mathrm{E}[z_j^i \mid \mathbf{Y}]\mathrm{E}[\Delta_k^l \mid \mathbf{Y}] & i \neq l \\ \mathrm{E}[z_j^i \Delta_j^i \mid \mathbf{Y}] & i = l, j = k \\ 0 & i = l, j < k \\ (z\Delta)_{jk}^i & k < j \end{cases}$$

For $k < j$ and $i \in \mathcal{I}$

$$(z\Delta)_{jk}^i = (x_{k+1}^i - x_k^i)\frac{\sum_{l=1}^{k-1} \hat{F}_{l,j}^i \hat{P}_{l,j}^i}{\sum_{j=1}^{D_i-1}\sum_{k=j}^{D_i-1} \hat{F}_{k,j}^i \hat{P}_{k,j}^i}$$

$$+ \int_{x_k^i}^{x_{k+1}^i} (x_{k+1}^i - u) \exp\left\{-\frac{(x_{k+1}^i - u)A^i(x_{k+1}^i)}{N(x_{k+1}^i)}\right\} du \frac{\hat{F}_{k,j}^i \frac{\hat{P}_{k,j}^i}{\hat{Q}_k^i}}{\sum_{j=1}^{D_i-1}\sum_{k=j}^{D_i-1} \hat{F}_{k,j}^i \hat{P}_{k,j}^i}$$

and for $k < j$ and $i \in \mathcal{I}^c$

$$(z\Delta)_{jk}^i = z_j^i E[\Delta_k^i \mid \mathbf{Y}]$$

For $j < k$ and $i \in \mathcal{I}$

$$\Delta_{j,k}^i = (x_{j+1}^i - x_j^i)(x_{k+1}^i - x_k^i)\frac{\sum_{l=1}^{j-1}\sum_{m=k+1}^{D_i-1} \hat{F}_{l,m}^i \hat{P}_{l,m}^i}{\sum_{j=1}^{D_i-1}\sum_{k=j}^{D_i-1} \hat{F}_{k,j}^i \hat{P}_{k,j}^i}$$

$$+(x_{k+1}^i - x_k^i) \int_{x_j^i}^{x_{j+1}^i} (x_{j+1}^i - u) \exp\left\{-\frac{(x_{j+1}^i - u)A^i(x_{j+1}^i)}{N(x_{j+1}^i)}\right\} du \frac{\sum_{l=k+1}^{D_i-1} \hat{F}_{j,l}^i \frac{\hat{P}_{j,l}^i}{\hat{Q}_j^i}}{\sum_{j=1}^{D_i-1}\sum_{k=j}^{D_i-1} \hat{F}_{k,j}^i \hat{P}_{k,j}^i}$$

$$+ \int_{x_j^i}^{x_{j+1}^i} \int_{x_k^i}^{x_{k+1}^i} (x_{j+1}^i - u)(t - x_k^i) \exp\left\{-\frac{(x_{j+1}^i - u)A^i(x_{j+1}^i)}{N(x_{j+1}^i)} - \frac{(t - x_k^i)A^i(x_{k+1}^i)}{N(x_{k+1}^i)}\right\} du dt \frac{\hat{F}_{j,k}^i \frac{\hat{P}_{j,k}^i}{\hat{Q}_j^i(1-q_k^i)}}{\sum_{j=1}^{D_i-1}\sum_{k=j}^{D_i-1} \hat{F}_{k,j}^i \hat{P}_{k,j}^i}$$

and

$$\Delta_{j,j}^i = (x_{j+1}^i - x_j^i)^2 \frac{\sum_{k=1}^{j-1}\sum_{l=j+1}^{D_i-1} \hat{F}_{k,l}^i \hat{P}_{k,l}^i}{\sum_{j=1}^{D_i-1}\sum_{k=j}^{D_i-1} \hat{F}_{k,j}^i \hat{P}_{k,j}^i}$$

$$+ \int_{x_j^i}^{x_{j+1}^i} (x_{j+1}^i - u)^2 \exp\left\{-\frac{(x_{j+1}^i - u)A^i(x_{j+1}^i)}{N(x_{j+1}^i)}\right\} du \frac{\sum_{k=j+1}^{D_i-1} \hat{F}_{j,k}^i \frac{\hat{P}_{j,k}^i}{\hat{Q}_j^i}}{\sum_{j=1}^{D_i-1}\sum_{k=j}^{D_i-1} \hat{F}_{k,j}^i \hat{P}_{k,j}^i}$$

$$+\int_{x_j^i}^{x_{j+1}^i}\int_u^{x_{j+1}^i}(t-u)^2\frac{1}{N(x_{j+1}^i)}\exp\left\{-\frac{(t-u)A^i(x_{j+1}^i)}{N(x_{j+1}^i)}\right\}dtdu\frac{\hat{F}_{j,j}}{\sum_{j=1}^{D_i-1}\sum_{k=j}^{D_i-1}\hat{F}_{k,j}^i\hat{P}_{k,j}^i}$$

$$+\int_{x_j^i}^{x_{j+1}^i}(t-x_j^i)^2\exp\left\{-\frac{(t-x_j^i)A^i(x_{j+1}^i)}{N(x_{j+1}^i)}\right\}dt\frac{\sum_{k=1}^{j-1}\hat{F}_{k,j}^i\frac{\hat{P}_{k,j}^i}{1-\hat{q}_j^i}}{\sum_{j=1}^{D_i-1}\sum_{k=j}^{D_i-1}\hat{F}_{k,j}^i\hat{P}_{k,j}^i}$$

For $i\in\mathcal{I}^c$ and $j<k$

$$\Delta_{j,k}^i=\begin{cases}0 & \sum_{l=1}^j I^i(x_{l+1}^i)=0 \text{ or } y_j^i=0\\ (x_{j+1}^i-x_j^i)(x_{k+1}^i-x_k^i) & I^i(x_{j+1}^i)=0,\sum_{l=1}^j I^i(x_{l+1}^i)>0,y_j^i=1,y_k^i=1\\ (x_{k+1}^i-x_k^i)\delta_j^i & I^i(x_{j+1}^i)=1,\sum_{l=1}^j I^i(x_{l+1}^i)>0,y_k^i=1\end{cases},$$

where δ_j^i is as defined in Equation 29, and

$$\Delta_{j,j}^i=\begin{cases}0 & \sum_{l=1}^j I^i(x_{l+1}^i)=0 \text{ or } y_j^i=0\\ (x_{j+1}^i-x_j^i)^2 & I^i(x_{j+1}^i)=0,\sum_{l=1}^j I^i(x_{l+1}^i)>0,y_j^i=1\\ \delta_{j,j}^i & I^i(x_{j+1}^i)=1,\sum_{l=1}^j I^i(x_{l+1}^i)>0,y_j^i>0\end{cases},$$

where

$$\delta_{j,j}^i=(x_{j+1}^i-x_j^i)^2\left[\frac{\sum_{k=1}^{j-1}I^i(x_{k+1}^i)\hat{Q}_k^i\prod_{l=k+1}^{D_i-1}[\hat{q}_l^i]^{y_l^i}}{\sum_{k=1}^{D_i-1}I^i(x_{k+1}^i)\hat{Q}_k^i\prod_{l=k+1}^{D_i-1}[\hat{q}_l^i]^{y_l^i}}\right]$$

$$+\int_{x_j^i}^{x_{j+1}^i}(x_{j+1}^i-u)^2\exp\left\{-\frac{(x_{j+1}^i-u)A^i(x_{j+1})}{N(x_{j+1}^i)}\right\}du\left[\frac{I^i(x_{j+1}^i)\prod_{l=j+1}^{D_i-1}[\hat{q}_l^i]^{y_l^i}}{\sum_{k=1}^{D_i-1}I^i(x_{k+1}^i)\hat{Q}_k^i\prod_{l=k+1}^{D_i-1}[\hat{q}_l^i]^{y_l^i}}\right]$$

and

For $i\in\mathcal{I}$

$$\mathrm{E}[z_j^i\Delta_j^i\mid\mathbf{Y}]=\int_{x_j^i}^{x_{j+1}^i}\int_u^{x_{j+1}^i}(t-u)\frac{1}{N(x_{j+1}^i)}\exp\left\{-\frac{(t-u)A^i(x_{j+1}^i)}{N(x_{j+1}^i)}\right\}dtdu\frac{\hat{F}_{j,j}}{\sum_{j=1}^{D_i-1}\sum_{k=j}^{D_i-1}\hat{F}_{k,j}^i\hat{P}_{k,j}^i}$$

$$+\int_{x_j^i}^{x_{j+1}^i}(t-x_j^i)\exp\left\{-\frac{(t-x_j^i)A^i(x_{j+1}^i)}{N(x_{j+1}^i)}\right\}dt\frac{\sum_{k=1}^{j-1}\hat{F}_{k,j}^i\frac{\hat{P}_{k,j}^i}{1-\hat{q}_j^i}}{\sum_{j=1}^{D_i-1}\sum_{k=j}^{D_i-1}\hat{F}_{k,j}^i\hat{P}_{k,j}^i}.$$

and for $i\in\mathcal{I}^c$

$$\mathrm{E}[z_j^i\Delta_j^i\mid\mathbf{Y}]=\delta_j^i z_j^i$$

The gradient vector of the complete data log-likelihood has lth element

$$\frac{\partial}{\partial\log N_l}\mathcal{L}_c(\mathbf{Y}_c;\hat{\mathbf{N}})=B_l-A_l+C_l-Z_l \tag{1}$$

With

$$A_l=\sum_{j=1}^{D}a_j^0 1_{l,j}^0$$

$$B_l=\sum_{j=1}^{D}0.5A^0(x_{j+1}^0)[A^0(x_{j+1}^0)-1](x_{j+1}^0-x_j^0)1_{l,j}^0\exp[-\log N_l],$$

$$C_l = \sum_{i=0}^{m-2} \sum_{j=1}^{D} A^i(x^i_{j+1}) \Delta^i_j 1^i_{l,j} \exp[-\log N_l],$$

and

$$Z_l = \sum_{i=0}^{m-2} \sum_{j=1}^{D} z^i_j 1^i_{l,j}.$$

Next, differentiating Equation 1 in this file, we have $\frac{\partial^2 l_c(\mathbf{Y}_c;\hat{\mathbf{N}})}{\partial \log N_l \partial \log N_m} = 0$ for all $l \neq m$, so the Hessian is a diagonal matrix with (l,l)th element

$$\frac{\partial^2}{\partial \log N_l^2} \mathcal{L}_c(\mathbf{Y}_c;\hat{\mathbf{N}}) = -B_l - C_l$$

and

$$\mathrm{E}\left[\left(\frac{\partial l_c(\mathbf{Y}_c;\hat{\mathbf{N}})}{N_l}\right)^2 \mid \mathbf{Y}\right] = (B_l - A_l)^2 + 2(B_l - A_l)\mathrm{E}[C_l - Z_l \mid \mathbf{Y}] + \mathrm{E}[(C_l - Z_l)^2 \mid \mathbf{Y}]$$

where

$$\mathrm{E}[C_l^2 \mid \mathbf{Y}] = \exp[-2logN_l] \sum_{i=0}^{m-2} \sum_{j=1}^{D_i-1} \left[\left\{A^i(x^i_{j+1})\right\}^2 \Delta^i_{j,j} 1^i_{l,j} + 2 \sum_{k=j+1}^{D_i-1} A^i(x^i_{j+1}) A^i(x^i_{k+1}) \Delta^i_{j,k} 1^i_{l,j} 1^i_{l,k} \right]$$

$$+2 \exp[-2logN_l] \sum_{i=0}^{m-2} \sum_{j=1}^{D_i-1} \left[A^i(x^i_{j+1}) \mathrm{E}[\Delta^i_j \mid \mathbf{Y}] 1^i_{l,j} \sum_{p=i+1}^{m-2} \sum_{k=1}^{D_p-1} A^p(x^p_{k+1}) \mathrm{E}[\Delta^p_k \mid \mathbf{Y}] 1^p_{l,k} \right],$$

$$\mathrm{E}[Z_l^2 \mid \mathbf{Y}] = \sum_{i=0}^{m-2} \sum_{j=1}^{D_i-1} \left[\mathrm{E}[z^i_j \mid \mathbf{Y}] 1^i_{l,j} + 2\mathrm{E}[z^i_j \mid \mathbf{Y}] 1^i_{l,j} \sum_{p=i+1}^{m-2} \sum_{k=1}^{D_p-1} \mathrm{E}[z^p_k \mid \mathbf{Y}] 1^p_{l,k} \right]$$

and

$$\mathrm{E}[C_l Z_l \mid \mathbf{Y}] = \frac{1}{N_l} \sum_{i=0}^{m-2} \sum_{j=1}^{D} A^i(x^i_{j+1}) \mathrm{E}[z^i_j \Delta^i_j \mid \mathbf{Y}] 1^i_{l,j} + \sum_{i=0}^{m-2} \sum_{j=1}^{D} A^i(x^i_{j+1}) \sum_{k=j+1}^{D} (z\Delta)^i_{k,j} 1^i_{l,k} 1^i_{l,j}$$

$$+ \sum_{i=0}^{m-2} \sum_{j=1}^{D} A^i(x^i_{j+1}) \sum_{p=1, p\neq i}^{m-2} \sum_{k=1}^{D_p-1} \mathrm{E}[\Delta^i_j \mid \mathbf{Y}] \mathrm{E}[z^p_k \mid \mathbf{Y}] 1^i_{l,j} 1^p_{l,k}$$

Also,

$$\mathrm{E}\left[\left(\frac{\partial l_c(\mathbf{Y}_c;\hat{\mathbf{N}})}{N_l}\right)\left(\frac{\partial l_c(\mathbf{Y}_c;\hat{\mathbf{N}})}{N_k}\right) \mid \mathbf{Y}\right] = (B_l - A_l)(B_k - A_k) + (B_l - A_l)\mathrm{E}[C_k - Z_k \mid \mathbf{Y}]$$

$$+(B_k - A_k)\mathrm{E}[C_l - Z_l \mid \mathbf{Y}] + \mathrm{E}[(C_l - Z_l)(C_k - Z_k) \mid \mathbf{Y}]$$

where

$$\mathrm{E}[C_l C_k] = \exp[-logN_l - logN_k] \sum_{i=0}^{m-2} \sum_{j=1}^{D_i-1} A^i(x^i_{j+1}) 1^i_{l,j} \sum_{o=1}^{m-2} \sum_{p=1}^{D_o-1} A^o(x^o_{p+1}) 1^o_{k,p} \mathrm{E}[\Delta^i_j \Delta^o_p \mid \mathbf{Y}],$$

$$\mathrm{E}[Z_l Z_k] = \sum_{i=0}^{m-2} \sum_{j=1}^{D_i-1} \sum_{o \neq i}^{m-2} \sum_{p=1}^{D_o-1} 1^i_{l,j} 1^o_{k,p} \mathrm{E}[z^i_j z^o_p \mid \mathbf{Y}]$$

and for $l < o$

$$\mathrm{E}[C_l Z_o \mid \mathbf{Y}] = \frac{1}{N_l} \sum_{i=0}^{m-2} \sum_{j=1}^{D} A^i(x^i_{j+1}) 1^i_{l,j} \left\{ \sum_{k=j+1}^{D} (z\Delta)^i_{k,j} 1^i_{o,k} + \sum_{p=1, p\neq i}^{m-2} \sum_{k=1}^{D} \mathrm{E}[\Delta^i_j \mid \mathbf{Y}] \mathrm{E}[z^p_k \mid \mathbf{Y}] 1^p_{o,k} \right\}$$

References

Chen, G. K., Marjoram, P., and Wall, J. D. (2009). Fast and flexible simulation of DNA sequence data. *Genome Research*, 19(1):136–142.

Rasmussen, M. D., Hubisz, M. J., Gronau, I., and Siepel, A. (2014). Genome-wide inference of ancestral recombination graphs. *PLoS Genet*, 10(5):e1004342.