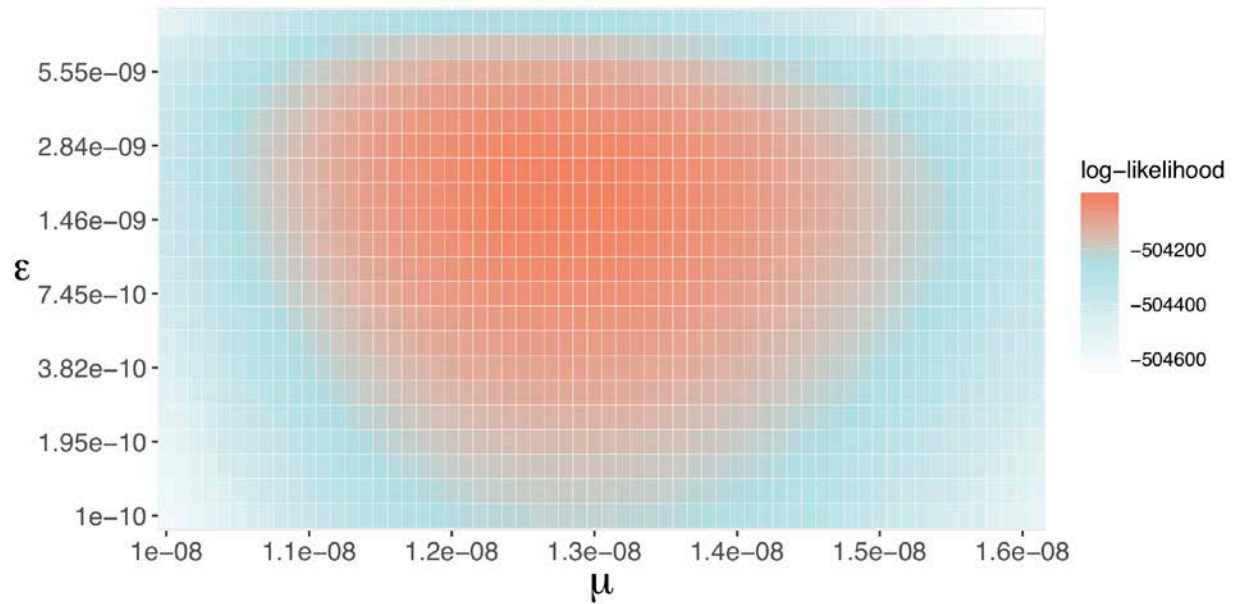


**The American Journal of Human Genetics, Volume 105**


**Supplemental Data**

**Estimating the Genome-wide Mutation Rate  
with Three-Way Identity by Descent**

**Xiaowen Tian, Brian L. Browning, and Sharon R. Browning**

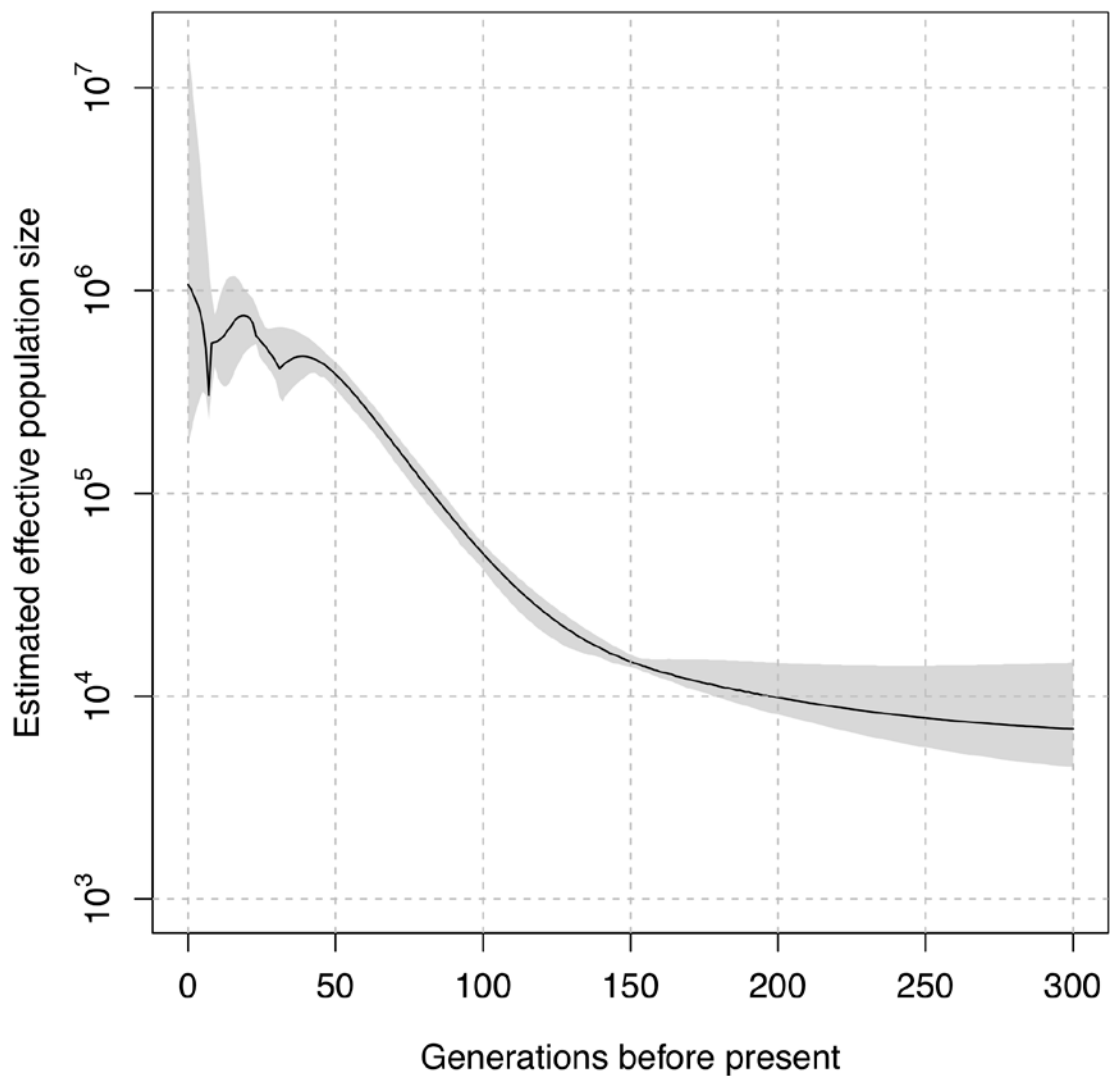


**Figure S1: An example of the likelihood contour for the mutation rate.** Data were simulated under the “super-exponential” model with errors simulated using the “unbiased” genotype error scheme (described in Methods). The simulated mutation rate is  $1.30 \times 10^{-8}$  per base pair per meiosis, and the error rate is 0.02%. The sample size is 2000 individuals. The likelihood is a function of two parameters: the mutation rate, and an error parameter  $\epsilon$  which is used to control for false apparent mutations cause by genotyping errors. The error parameter  $\epsilon$  is less than the genotype error rate because many genotype errors are removed by the requirement that two of the three IBD haplotypes carry the allele. We use an adaptive search grid to find the values of the parameters that maximize the likelihood.

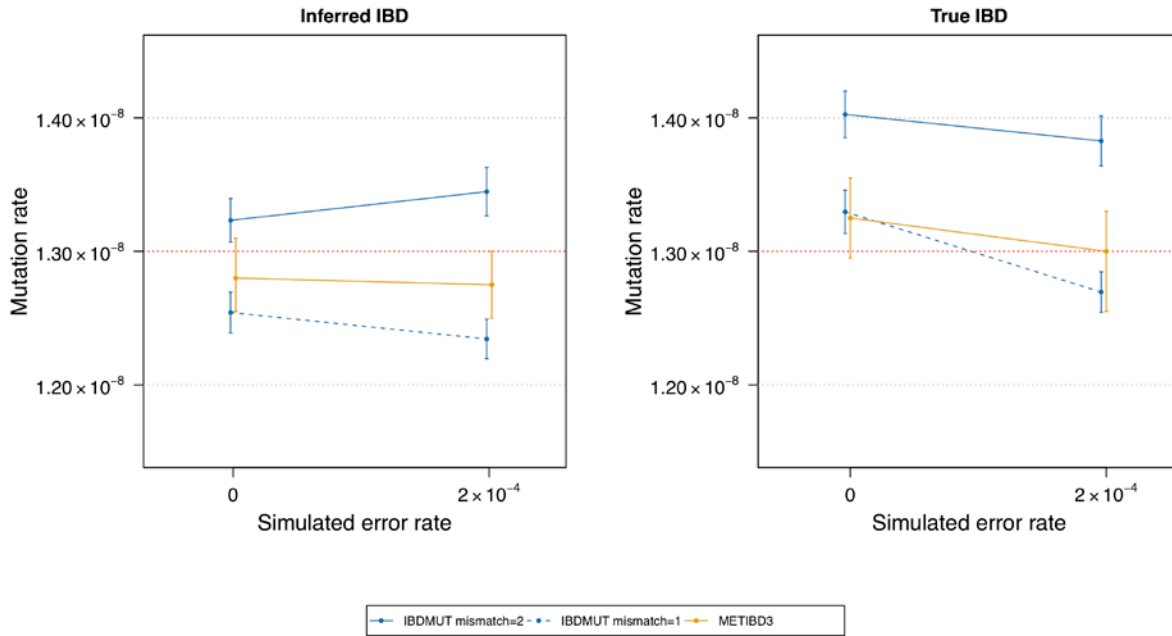
(1) 

(2) 

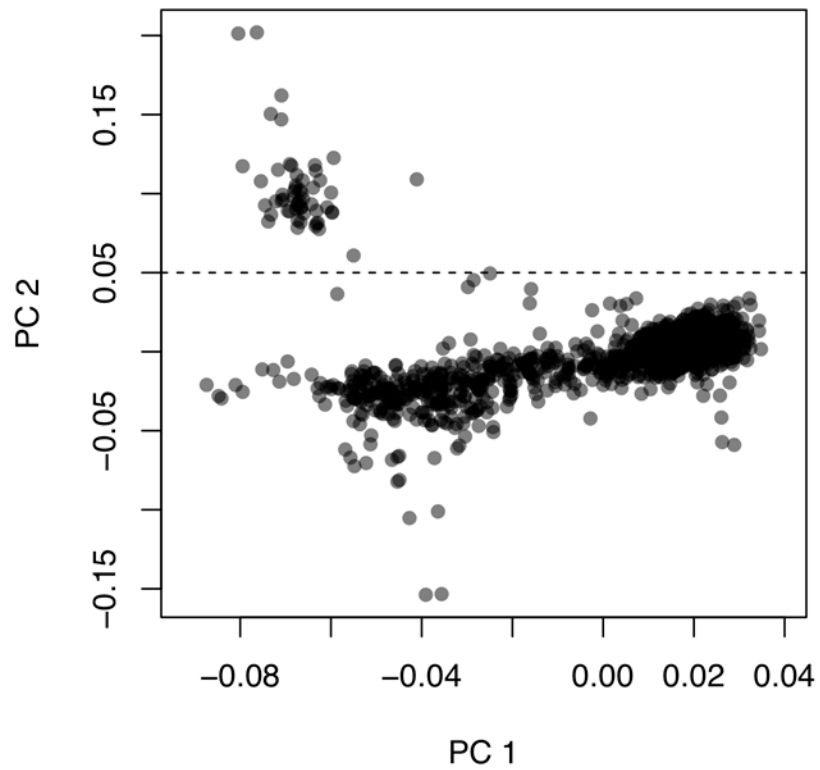
**Figure S2: An example of the gap-filling procedure.** The three detected Refined IBD segments for one pair of haplotypes are shown in (1) as *ab*, *cd*, and *ef*. If the gap *bc*, between the *ab* and *cd* segments has a maximum length of 0.5 cM and the maximum number of genotypes in *bc* that are inconsistent with IBD is 2, this gap will be filled, as shown in (2). Similar rules are applied to the gap *de* between the *cd* and *ef* segments.



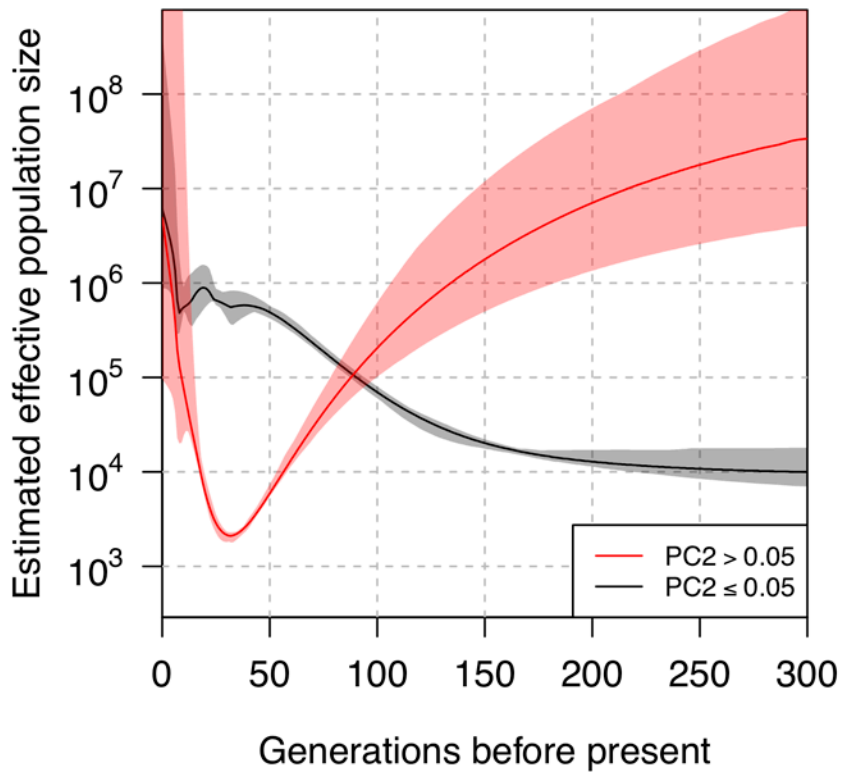
**Figure S3: Estimated recent effective population size for the Framingham sample.** Generations before present are shown on the x-axis, and estimated effective population size is shown on the y-axis. The black line gives the estimated size while the gray region gives the 95% bootstrap confidence intervals. The reduction in estimated effective size approximately 10 generations ago likely reflects the bottleneck effect of European migration to the US. A similar estimated bottleneck was seen in analysis of European-ancestry individuals sampled from Memphis, Tennessee.<sup>1</sup>



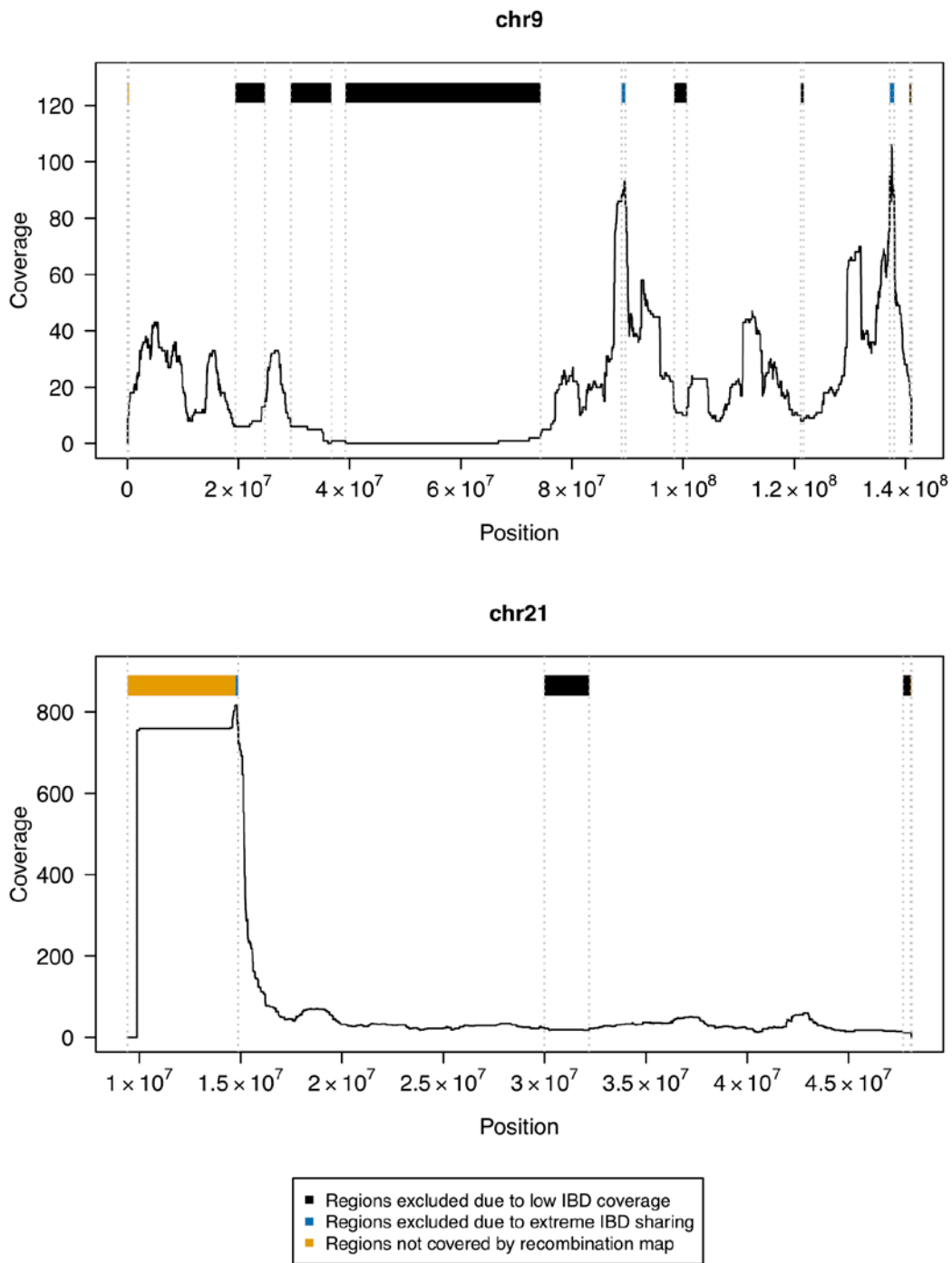
**Figure S4: Estimated mutation rates from IBDMUT.** Data were simulated under the “super-exponential” model with errors simulated using the “unbiased” genotype error scheme (described in Methods). The simulated mutation rate is  $1.30 \times 10^{-8}$  per base pair per meiosis and is indicated with the red dotted line. The sample size is 2000 individuals. We used GERMLINE for IBD segment detection with different values for the number of allowed mismatch sites. Genotype errors were added before IBD detection and thus influence the accuracy of IBD inference. We show results when using the inferred effective population size from IBDNe based on the Refined IBD segments (left panel), and also show results when using the true effective population size from the simulation model (right panel). Point estimates (dots) and 95% confidence intervals (bars) are shown. Results from our method (METIBD3) on the same data are shown for comparison.



**Figure S5: The first two principal components for genetic data from 1362 founders in Framingham Heart Study.** The dashed line indicates  $PC2=0.05$ , which separates the two clusters of individuals.

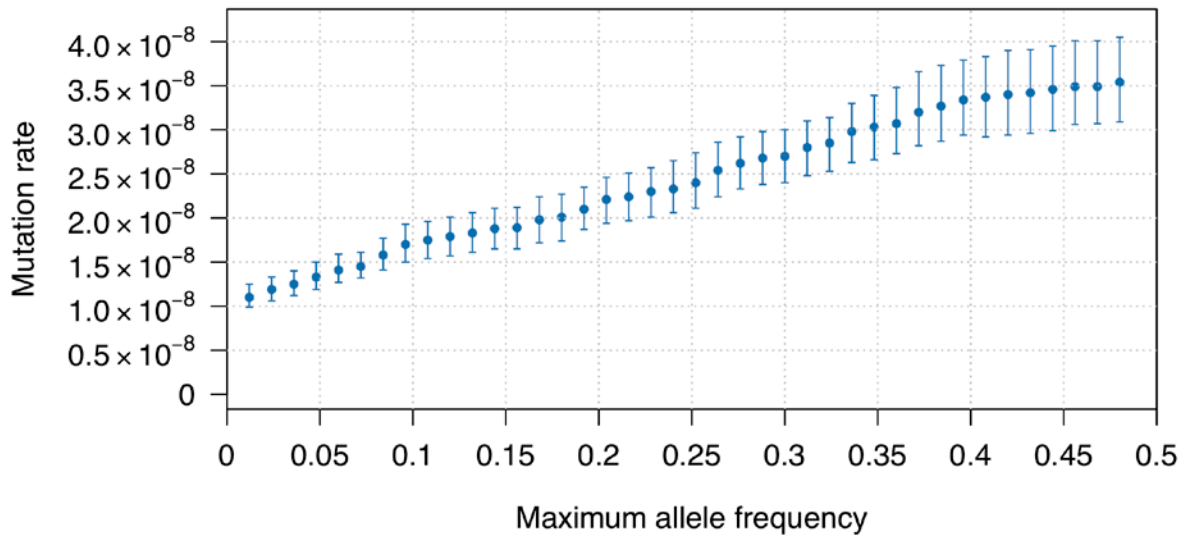


**Figure S6: Estimated effective population size of two clusters of individuals from the Framingham Heart Study.** Generations before present are shown on the x-axis, and estimated effective population size is shown on the y-axis. The solid line in black represents the estimated size from samples with PC2 less than or equal to 0.05, while the solid line in red represents the estimated size from samples with PC2 greater than 0.05. The shaded region gives the 95% bootstrap confidence intervals.



**Figure S7: Examples of 3-way IBD coverage along the genome.** Levels of three-way IBD are shown in windows of 500 base pairs for two representative chromosomes. Black bars represent regions with zero 3-way IBD coverage after removing IBD segments of length greater than 6 cM. Blue bars represent regions excluded from the analysis due to extremely high levels of apparent IBD sharing. Orange bars represent regions not covered by the Rutgers recombination map.





**Figure S8: Estimated mutation rate from the Framingham Heart Study data as a function of maximum allowed allele frequency.** Point estimates (dots) and 95% confidence intervals (bars) are shown.

Three-way IBD sharing (IBD3)	$P(\text{IBD3}   \text{tree3})$
	$P_1$
	$P_2$
	$P_3$
	$P_3$
	$P_4$

**Table S1: Probabilities for all possible three-way IBD segment configurations.** The right column gives the probability of the IBD segment configuration in the left column, conditional on the 3-haplotype coalescent tree, *tree3*, being the tree shown in Figure 1 (haplotypes A and B coalesce  $g_1$  generations ago and their common ancestor and haplotype C coalesce  $g_1 + g_2$  generations ago). The corresponding probability is given by one of the equations represented by  $P_1, P_2, P_3, P_4$  below. The positions  $x_1, x_2, x_3, x_4$  of changes in IBD status are measured in Morgans.

Let  $G_1(x; \lambda) = \lambda e^{-\lambda x}$  denote an exponential distribution and  $G_2(x) = (3g_1 + 2g_2)^2 x e^{-(3g_1 + 2g_2)x}$  denote the gamma distributions with shape parameter 2 and rate parameter  $3g_1 + 2g_2$ . Then

$$P_1 = \frac{(g_1 + 2g_2)g_1}{(3g_1 + 2g_2)^2} G_2(x_3 - x_2) G_1(x_2 - x_1; 2g_1) G_1(x_4 - x_3; 2g_1 + 2g_2)$$

$$P_2 = \frac{(g_1 + 2g_2)g_1}{(3g_1 + 2g_2)^2} G_2(x_3 - x_2) G_1(x_2 - x_1; 2g_1 + 2g_2) G_1(x_4 - x_3; 2g_1)$$

$$P_3 = \frac{g_1^2}{(3g_1 + 2g_2)^2} G_2(x_3 - x_2) G_1(x_2 - x_1; 2g_1 + 2g_2) G_1(x_4 - x_3; 2g_1 + 2g_2)$$

$$P_4 = \frac{(g_1 + 2g_2)^2}{(3g_1 + 2g_2)^2} G_2(x_3 - x_2) G_1(x_2 - x_1; 2g_1) G_1(x_4 - x_3; 2g_1)$$

The probabilities corresponding to any three-way coalescent tree can be found by appropriate relabeling of the haplotypes.

## Supplemental References

1. Browning, S.R., Browning, B.L., Daviglus, M.L., Durazo-Arvizu, R., Schneiderman, N., Kaplan, R., and Laurie, C.C. (2018). Ancestry-specific recent effective population size in the Americas. *PLoS Genet* 14, e1007385.