

1 Supplementary Information

2 Appendix A.

3 Initializing the ancestral haplotype for the MCMC.

4 To decrease run times for the MCMC, we initialize the starting sequence for the ancestral
 5 haplotype using a heuristic algorithm which exploits the decrease in polymorphism near the
 6 selected site. Let A^0 denote the initial ancestral haplotype to be estimated, and let the indicator
 7 variable I_{ij} denote whether chromosome i is part of the ancestral haplotype at site j :

$$I_{ij} = \begin{cases} 1 & \text{if } X_{ij} = A_j^0; \\ 0 & \text{if } X_{ij} \neq A_j^0 \end{cases} \quad (1)$$

8 The algorithm proceeds as follows:

- 9 1. At $j = 1$ all chromosomes with the beneficial allele are specified to be on the ancestral
 10 haplotype at the selected site, i.e. $\sum_{i=1}^n I_{i1} = n$ and $A_j^0 = 1$.
- 11 2. Moving to the next adjacent SNP, we calculate the allele frequency, F_j , among chromo-
 12 somes on the ancestral haplotype at the previous site:

$$F_j = \frac{\sum_{i=1}^n X_{ij} I_{i(j-1)}}{\sum_{i=1}^n I_{i(j-1)}} \quad (2)$$

- 13 3. The major allele among advantageous allele carriers is assumed to be the putative ancestral
 14 allele and minor alleles are assumed to be the result of a putative recombination event off
 15 of the ancestral haplotype in the previous SNP interval. For $j > 0$,

$$A_j^0 = \begin{cases} 1 & \text{if } F_j > 0.5; \\ 0 & \text{if } F_j < 0.5 \end{cases} \quad (3)$$

16 Because we expect there to be some rare or singleton variants on the ancestral haplotype,
 17 singletons are removed before step 1 in an effort to improve estimates of the ancestral haplotype
 18 at more distant sites. In addition, major and minor alleles can't be identified at sites with alleles

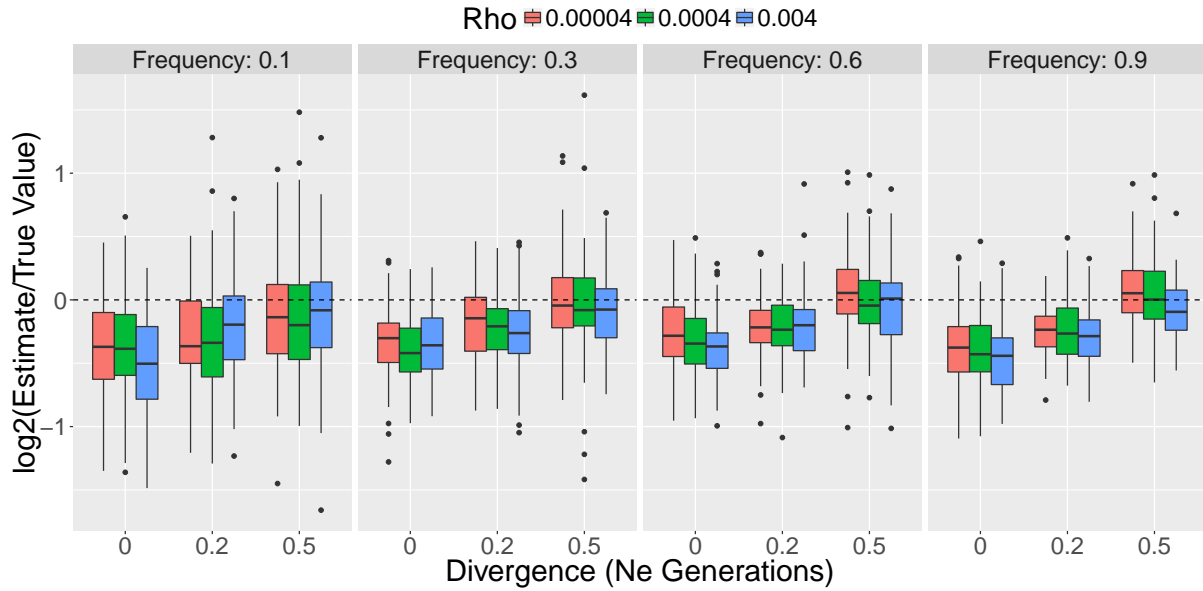
at 0.5 frequency and are also removed initially. Steps 2 and 3 are computed iteratively until reaching the end of the locus ($j = L$) on both sides flanking the selected site. The sites that were removed ($F_j = 0.5$ and singletons) are then added back in and take values of I_{ij} from I_{ij+1} . A_j^0 for the added sites are computed using equations 12 and 13. At sites for which $\sum_{i=1}^n I_{i1} = 0$, $A_j^0 = \text{Binomial}(1, P_j)$, where $P_j = \frac{1}{n} \sum_{i=1}^n X_{ij}$. After getting the initial estimate A_j^0 , the MCMC is run and evaluated for convergence by visual inspection of trace plots.

Appendix B.

Modelling singletons and invariant sites on background haplotypes within the panel of carriers.

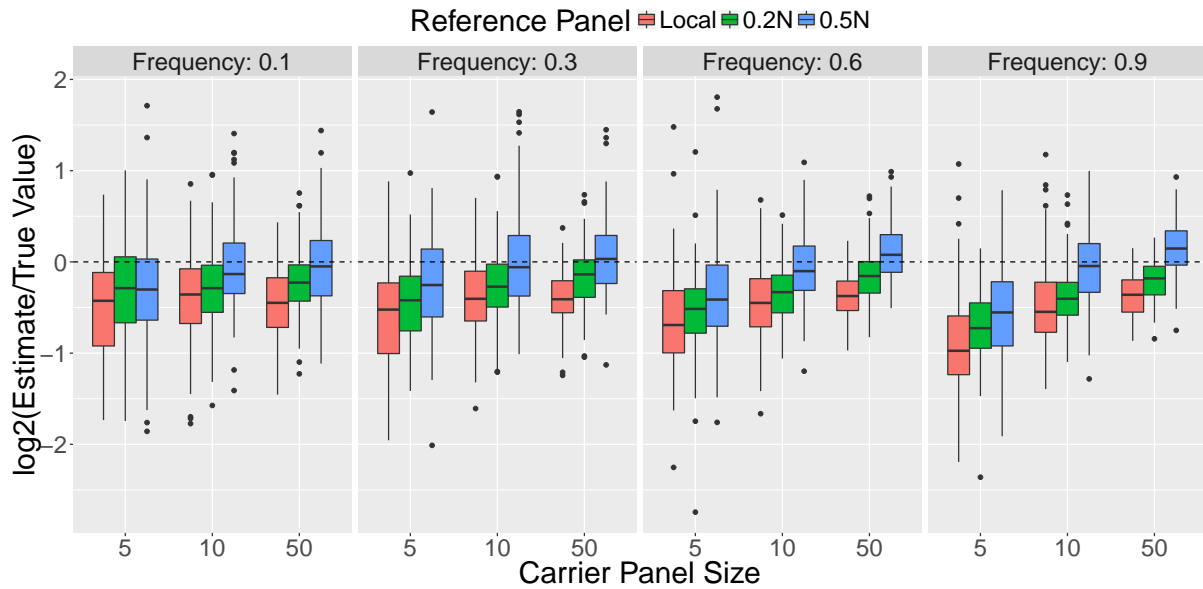
We implemented an approach for modeling invariant sites on the background haplotypes among carriers of the beneficial allele, the accuracy of which are summarized in Supplementary Table 6 below. This portion of the likelihood is denoted β_{iw} in the model description, and corresponds to the Li and Stephens [2003] haplotype copying model. The original formulation of this model ignores invariant sites, which is equivalent to assuming the probability of observing them is 1. We modified our likelihood by first noting that, under the star genealogy assumption, all variants which are found in the carrier panel but absent in the reference panel should be singletons. Thus we can estimate the probability of a invariant site by considering the rate of singletons in the reference panel. Specifically, we estimate the probability as $1 - (S/nL)$, where S is the number of singletons in the reference panel, n is the number of reference panel haplotypes and L is the number of basepairs at the locus. For a given haplotype i and SNP w in β_{iw} , the probability of observing d invariant sites is $(1 - (S/nL))^d$ (Model B in Supplementary Table 6).

Despite this attempt to more accurately model the background haplotypes in the carrier panel, we did not find any consistent improvement in bias or accuracy.

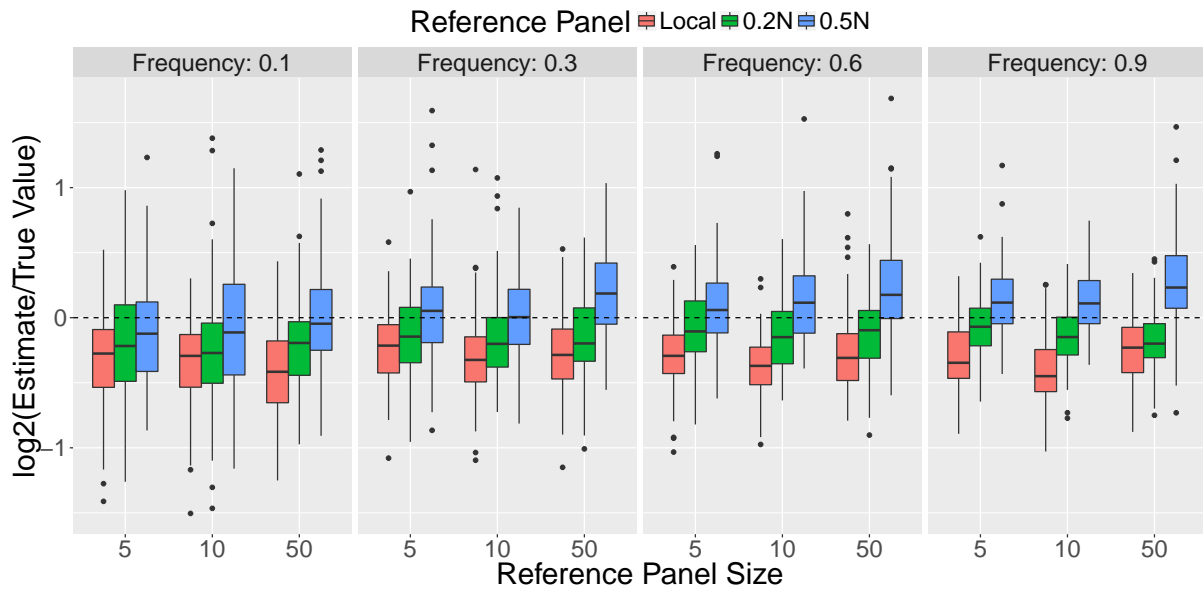


42

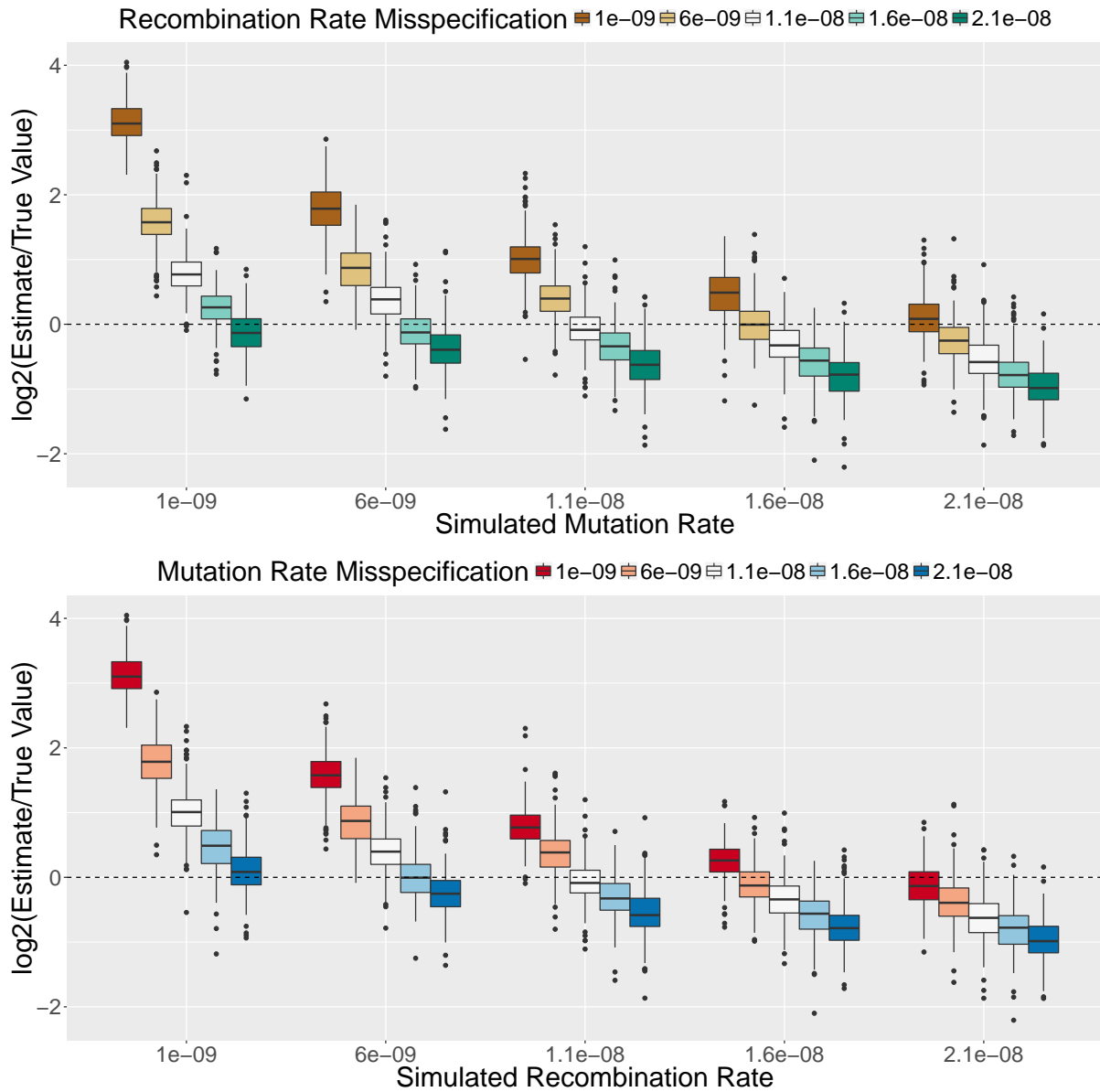
43 **Supplementary Figure 1. Effect of misspecifying ρ .** Accuracy results for 3 different
 44 values of ρ used in the Li and Stephens [2003] copying model for background haplotypes. All
 45 other parameter values are identical to Figure 2 in the main text. The divergence value of 0
 46 refers to a local reference panel. Allele frequency refers to the end frequency of the beneficial
 47 allele trajectory.



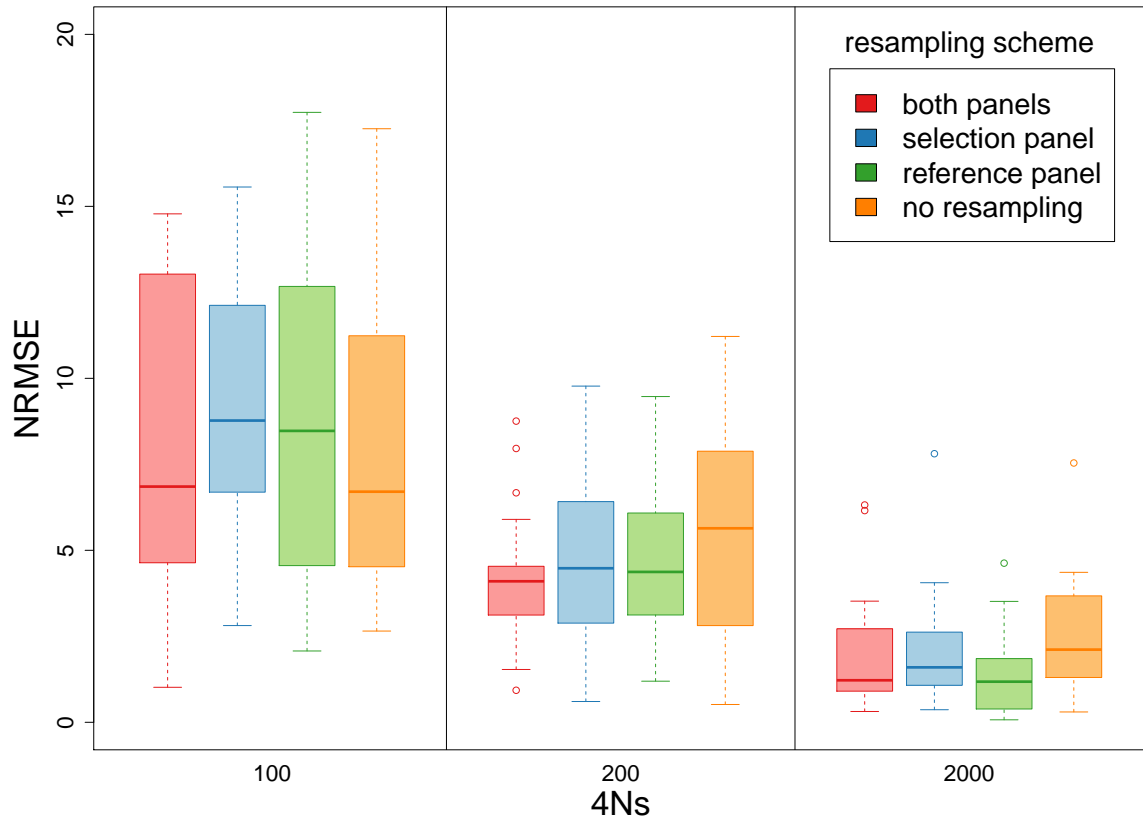
Supplementary Figure 2. Effect of beneficial allele carrier sample size. Accuracy results for 3 different sample sizes for the panel of haplotypes carrying the beneficial allele. The selection strength for all simulations was set to 0.01. All other parameter values are identical to Figure 2 in the main text. Allele frequency refers to the end frequency of the beneficial allele trajectory.



Supplementary Figure 3. Effect of reference panel sample size. Accuracy results for 3 different sample sizes for the reference panel of haplotypes without the selected allele. The selection strength for all simulations was set to 0.01. All other parameter values are identical to Figure 2 in the main text. Allele frequency refers to the end frequency of the beneficial allele trajectory.

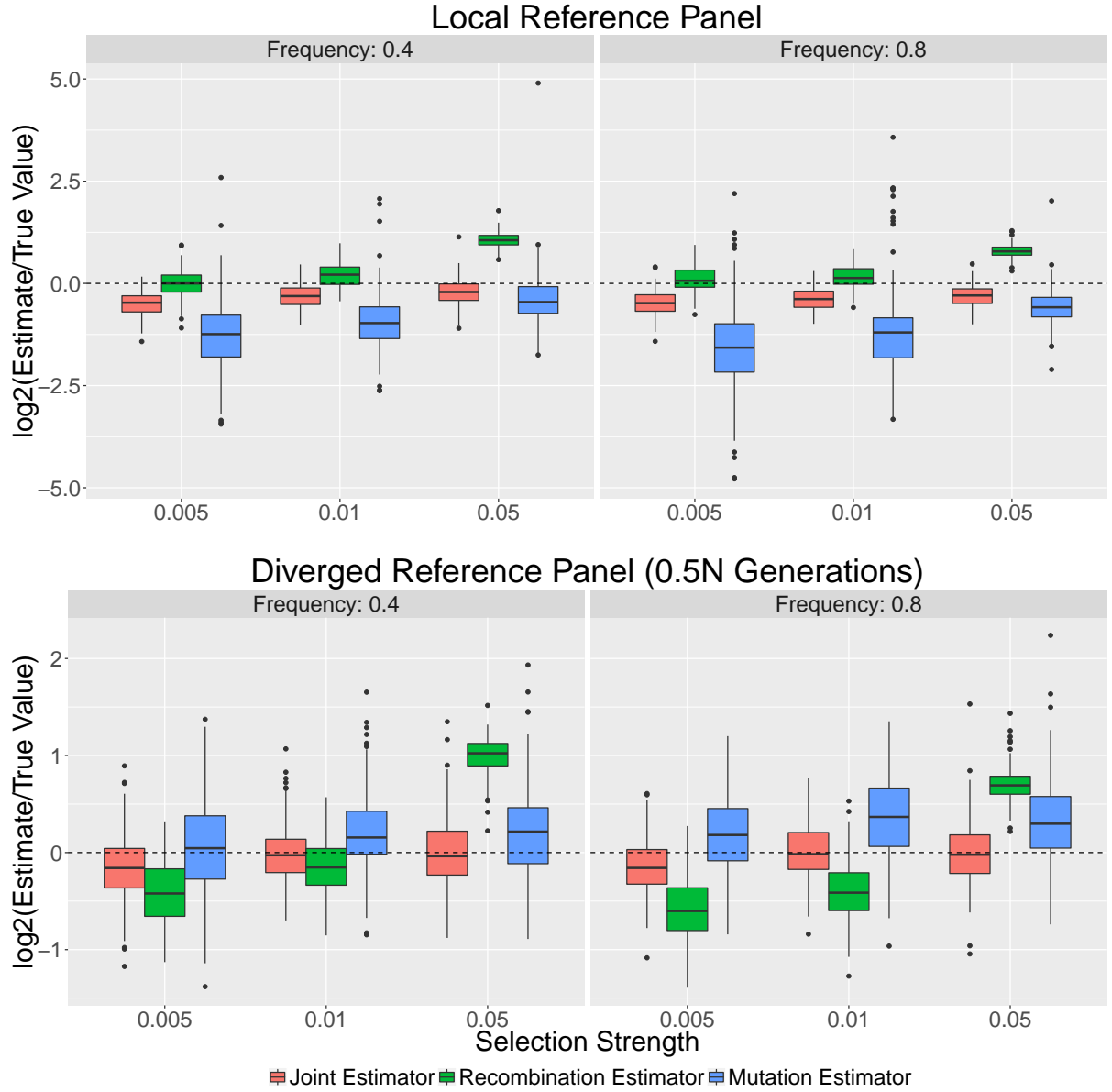


Supplementary Figure 4. Effect of misspecifying the mutation and recombination rates. Accuracy results for varying degrees of mutation and recombination rate misspecification. In both panels, the parameter values on the x-axis were used both for simulation and inference. For the colored boxplots, the true values are in white ($1.1e-8$) and the colors refer to different degrees of misspecification used for inference. Simulations were performed with a local reference panel and a selection strength of 0.01. All other parameter values are identical to Figure 2 in the main text.



68

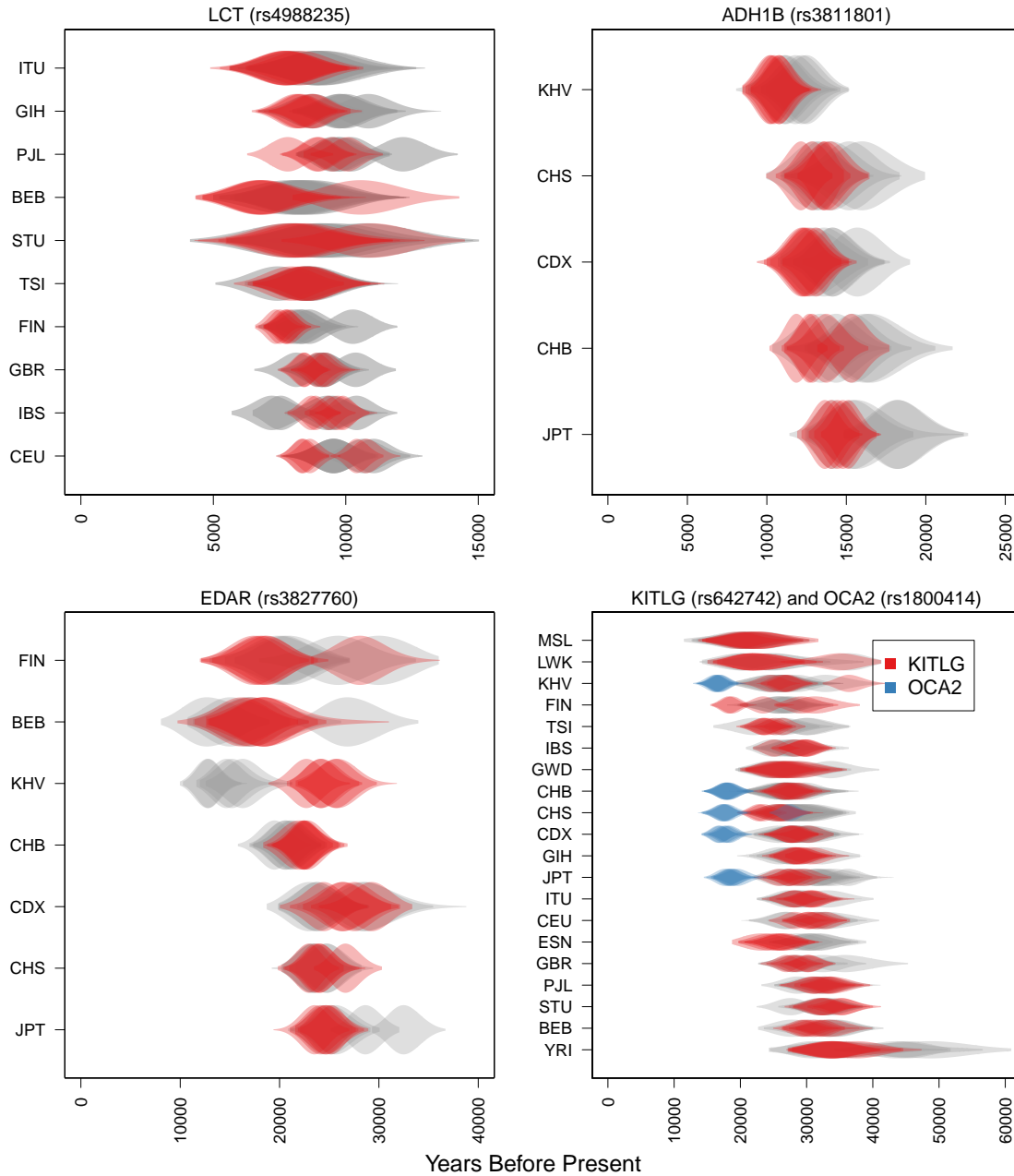
Supplementary Figure 5. Effect of resampling subsets of complete data. Estimated accuracy and among independent MCMC runs for different resampling schemes. Frequency trajectories were simulated to an end frequency of 0.1. Under each 2Ns value and resampling scheme indicated in the legend, 20 data sets were simulated and inference was performed on the 5 replicate MCMCs. In each simulation, the full dataset includes sample sizes of 100 for the selected and reference panels. Inference for each replicate was then performed on 50 selected haplotypes and 20 reference haplotypes according to the sampling scheme in the legend. Normalized RMSE values are calculated using the estimates and true TMRCA value, while the standard deviations are calculated using the estimates and their mean.



78

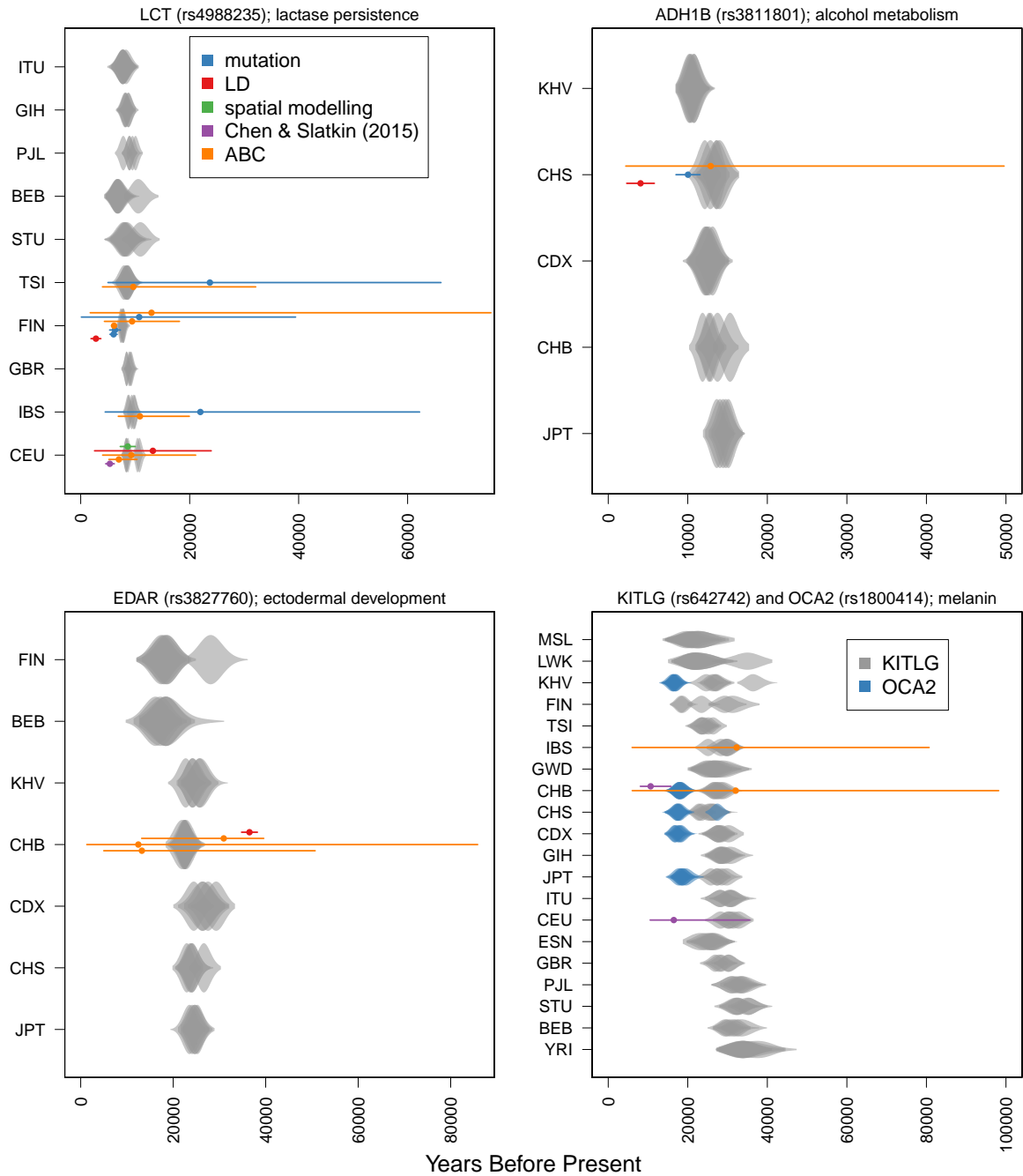
79 **Supplementary Figure 6. Comparison to heuristic estimates.** We compared our TM-
 80 RCA estimator (joint estimator) to an estimate which uses the mean length of haplotype lengths
 81 and another estimate which uses number of derived mutations on the ancestral haplotype. In
 82 all simulations a selection strength of 0.01 was used. All other parameter values are identical
 83 to Figure 2 in the main text. Frequency refers to the end frequency of the beneficial allele
 84 trajectory.

85



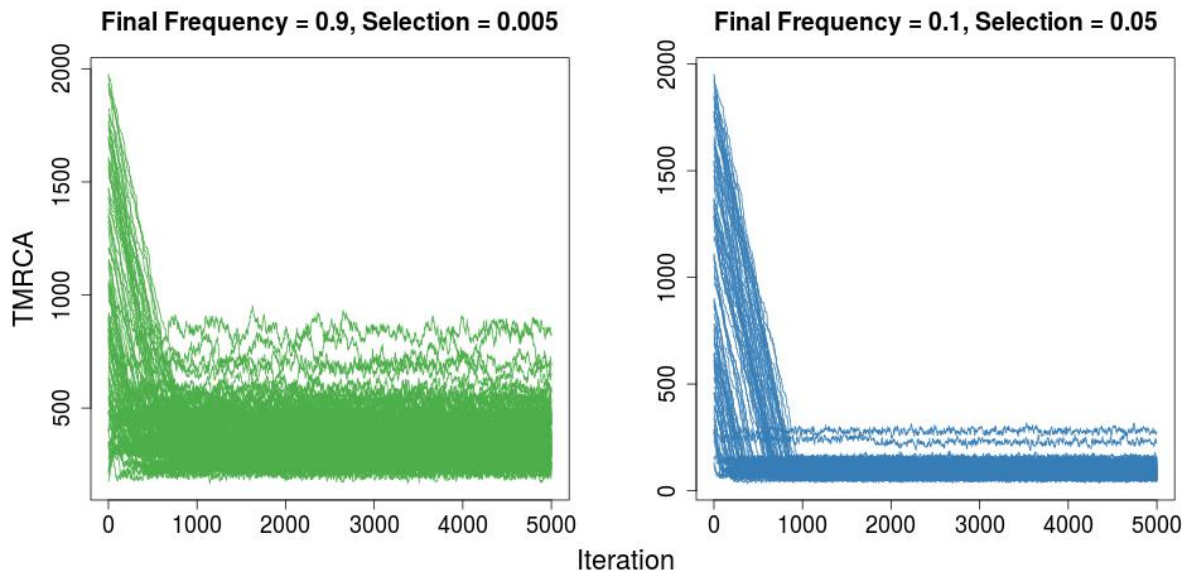
86

87 **Supplementary Figure 7. Comparison of fine-scale and Mbp-scale recombination**
 88 **maps.** A comparison between estimates made using the fine-scale Decode recombination map
 89 (grey) and a uniform recombination rate (red and blue). The uniform recombination rate used
 90 for each gene is the mean rate for the 1Mb region around each variant indicated by the rs
 91 number. Five replicate MCMCs were performed for each variant and population by resampling
 92 the selected and reference panels with replacement.



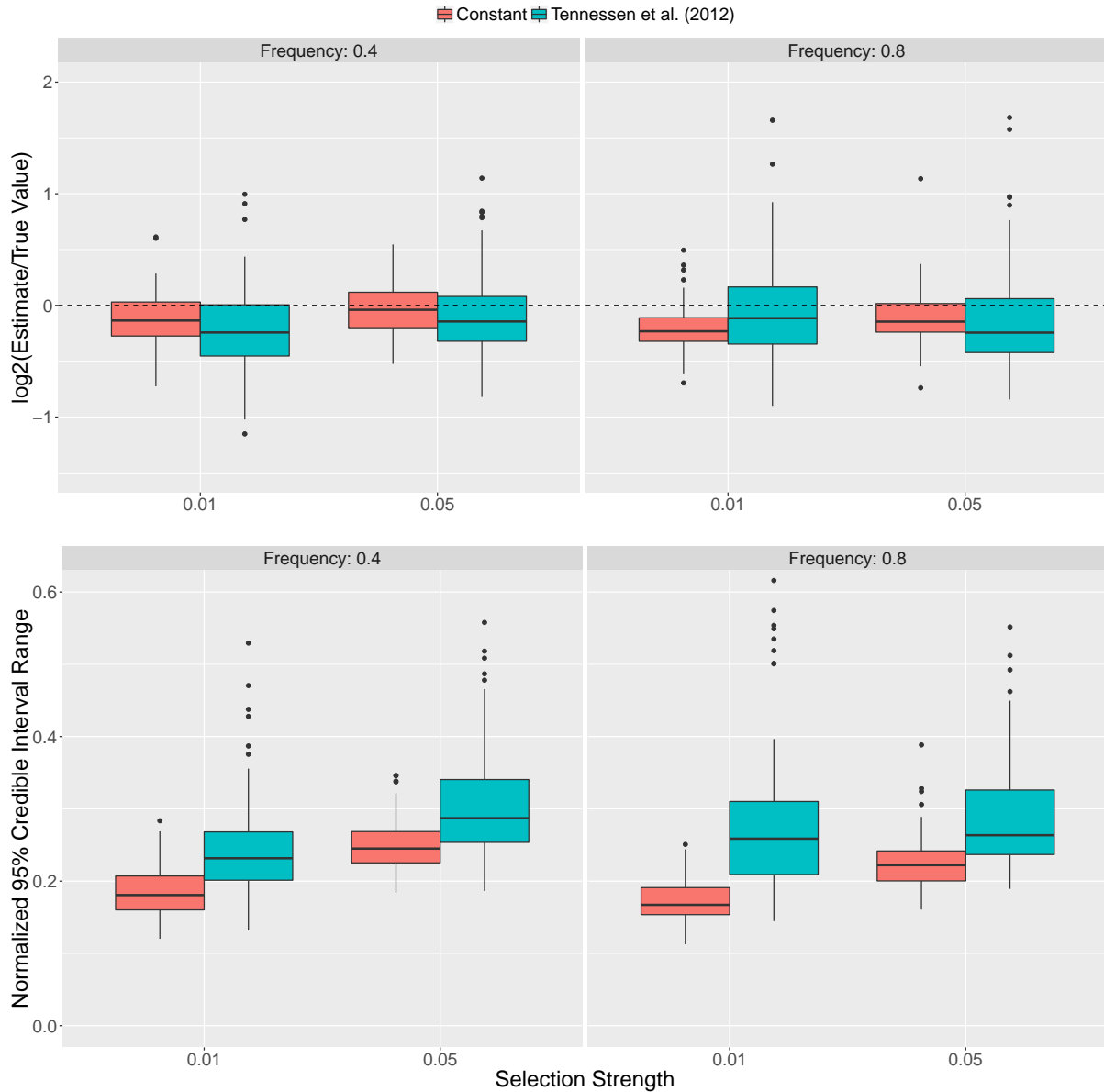
93

94 **Supplementary Figure 8. Comparison of TMRCA estimates and previous estimate**
 95 **approaches.** Results from Fig 4 (main text) sorted into different plots for different variants.
 96 Previous estimates are colored by an abbreviated description of the type of information used in
 97 the data. The blue violin plots in the KITLG/OCA2 plot are estimates for the OCA2 variant.
 98 The purple and orange previous estimates for CHB in the KITLG/OCA2 plot refer to OCA2
 99 and KITLG, respectively.



100

101 **Supplementary Figure 9. Traces of MCMC results from simulated data.** Results
 102 from Figure 2 (main text) for data simulated in a single population using a local reference panel.
 103 Each plot is the result of MCMC runs performed on 100 simulated data sets. The simulated
 104 parameter values in the left plot represent the oldest TMRCAs and those in the right are the
 105 youngest.



106

107 **Supplementary Figure 10. Effects of non-equilibrium demographic history on es-**
 108 **timate accuracy.** A comparison of estimate accuracy and credible interval ranges using data
 109 simulated under the European demographic history inferred by Tennesen et al. [2012] and a
 110 constant population size model. To decrease computation time, we used a present day popula-
 111 tion size of 150000 rather than 500000. All relative changes in growth rate and bottleneck sizes
 112 are identical to those inferred by Tennesen et al. [2012]. We used a local reference panel for
 113 both demographic histories, and other parameter values are identical to those used for Figure 2
 114 in the main text.

| Selection Strength | Frequency | TMRCAs |
|--------------------|-----------|--------|
| 0.005 | 0.1 | 525 |
| 0.005 | 0.3 | 789 |
| 0.005 | 0.6 | 1030 |
| 0.005 | 0.9 | 1411 |
| 0.01 | 0.1 | 322 |
| 0.01 | 0.3 | 461 |
| 0.01 | 0.6 | 596 |
| 0.01 | 0.9 | 772 |
| 0.05 | 0.1 | 94 |
| 0.05 | 0.3 | 120 |
| 0.05 | 0.6 | 144 |
| 0.05 | 0.9 | 179 |

Supplementary Table 1. Simulated TMRCAs values (mean generations). These are mean TMRCAs values from simulations using 3 selection strengths and 4 ending frequencies for the beneficial allele. Each mean TMRCAs is computed with 300 simulations.

| Abbreviation | Sample |
|--------------|--|
| CHB | Han Chinese in Beijing, China |
| JPT | Japanese in Tokyo, Japan |
| CHS | Southern Han Chinese |
| CDX | Chinese Dai in Xishuangbanna, China |
| KHV | Kinh in Ho Chi Minh City, Vietnam |
| CEU | Utah Residents with Northern and Western European Ancestry |
| TSI | Toscans in Italy |
| FIN | Finnish in Finland |
| GBR | British in England and Scotland |
| IBS | Iberian Population in Spain |
| YRI | Yoruba in Ibadan, Nigeria |
| LWK | Luhya in Webuye, Kenya |
| GWD | Gambian in Western Divisions in the Gambia |
| MSL | Mende in Sierra Leone |
| ESN | Esan in Nigeria |
| GIH | Gujarati Indian from Houston, Texas |
| PJL | Punjabi from Lahore, Pakistan |
| BEB | Bengali from Bangladesh |
| STU | Sri Lankan Tamil from the UK |
| ITU | Indian Telugu from the UK |

Supplementary Table 2. Sample abbreviations for the 1000 Genomes Project panel.

| Gene | Population | Mbp-Scale Recombination Map | | Fine-Scale Recombination Map | |
|-------|------------|-----------------------------|-----------------------|------------------------------|-----------------------|
| | | t (Years) | 95% Credible Interval | t (Years) | 95% Credible Interval |
| KITLG | FIN | 18733.11 | 16675.03 - 20816.16 | 26343.99 | 21185.68 - 31439.77 |
| KITLG | MSL | 22339.79 | 15723.06 - 28949.85 | 20244 | 15042.13 - 26063.71 |
| KITLG | LWK | 22783.5 | 17921.93 - 28011.88 | 24200.86 | 16839.43 30730.11 |
| KITLG | KHV | 24544.92 | 21643.47 - 27192.74 | 26697.31 | 22249.2 31526.1 |
| KITLG | ESN | 26254.27 | 22854.93 - 29657.6 | 31791.02 | 26440.75 36543.8 |
| KITLG | TSI | 26427.6 | 24109.65 - 28905.17 | 22776.53 | 18379.65 27588.96 |
| KITLG | CHS | 26535.39 | 23456.37 - 29651.06 | 28396.03 | 23284.08 33294.5 |
| KITLG | CHB | 26772.95 | 24297.24 - 30141.32 | 28968.9 | 24451.06 34718.66 |
| KITLG | GBR | 26785.87 | 23841.96 - 29252.6 | 36132.87 | 31410.16 41697.2 |
| KITLG | GWD | 27669.05 | 21900.96 - 33664.19 | 25833.85 | 20134.3 32433.62 |
| KITLG | ITU | 28093.72 | 24607.36 - 31040.55 | 28090.16 | 24212.08 32250.46 |
| KITLG | CDX | 28362.37 | 25128.88 - 31245.57 | 29010.65 | 24457.71 33869.7 |
| KITLG | JPT | 28636.9 | 26351.84 - 31139.23 | 31634.52 | 27551.51 36471.26 |
| KITLG | GIH | 29029.82 | 25862.18 - 32439.4 | 28935.4 | 24599.77 33752.46 |
| KITLG | IBS | 29730.96 | 26169.86 - 32812.62 | 25373.45 | 21393.79 30031.51 |
| KITLG | CEU | 31287.49 | 27866.12 - 34512.88 | 34009.13 | 29818.32 38072.96 |
| KITLG | STU | 32021.3 2 | 8243.9 - 36318.9 | 27693.36 | 23968.81 32516.82 |
| KITLG | BEB | 32030.37 | 29000.59 - 34975.94 | 34375.32 | 29254.91 39578.63 |
| KITLG | PJL | 33719.74 | 30137.36 - 37310.9 | 31384.02 | 26814.24 36005.58 |
| KITLG | YRI | 33947.53 | 28861.11 - 39098.89 | 44437.11 | 36047.45 54074.11 |
| EDAR | FIN | 17386.19 | 13887.25 - 20794.2 | 20176.21 | 15053.08 25838.39 |
| EDAR | BEB | 18370.17 | 14325.16 - 22871.72 | 18418.06 | 13680.78 25409.82 |
| EDAR | CHB | 22192.42 | 19682.73 - 25735.6 | 19262.27 | 16921.98 21521.19 |
| EDAR | JPT | 23508.87 | 21595.24 - 25644.81 | 25730.04 | 23096.63 28826.86 |
| EDAR | CHS | 24058.94 | 22005.79 - 26678.85 | 24813.16 | 22493.15 27204.94 |
| EDAR | CDX | 24360.34 | 21572.05 - 27044.11 | 24346.84 | 21214.9 28019.59 |
| EDAR | KHV | 25683.33 | 23169.98 - 28379.79 | 12686.77 | 11001.3 14645.05 |
| OCA2 | KHV | 16370.39 | 14439.12 - 18102.08 | 26904.93 | 22093.63 32402.26 |
| OCA2 | CHS | 17316.96 | 14913.26 - 19799.16 | 26377.66 | 21217.62 31921.16 |
| OCA2 | CHB | 17838.58 | 15336.98 - 20174.82 | 25159.82 | 20764 29688.86 |
| OCA2 | CDX | 18083.2 1 | 6231.36 - 20253.74 | 28644.18 | 24241.91 33819.6 |
| OCA2 | JPT | 18598.62 | 16110.22 - 20785.6 | 31582.69 | 27875.01 35522.35 |
| ADH1B | KHV | 10841.65 | 9720.032 - 12147.5 | 11186.97 | 9503.454 12862.22 |
| ADH1B | CHS | 12101.84 | 10668.909 - 13479.33 | 15352.24 | 12969.029 17974.85 |
| ADH1B | CDX | 12176.61 | 10678.377 - 13699.32 | 13568.88 | 11183.01 15941.4 |
| ADH1B | JPT | 13996.17 | 12670.869 - 15278.67 | 18317.67 | 15995.495 20911.5 |
| ADH1B | CHB | 15377.36 | 13763.712 - 17281.5 | 13526.93 | 11280.86 16210.89 |
| LCT | BEB | 6869.385 | 5143.203 - 8808.557 | 7971.853 | 5893.793 10443.94 |
| LCT | FIN | 7545.399 | 6982.857 - 8112.515 | 10332.821 | 9349.834 11427.629 |
| LCT | ITU | 7795.401 | 6199.996 - 9419.64 | 8972.475 | 7043.311 11015.566 |
| LCT | TSI | 7936.011 | 6616.676 - 9435.192 | 8630.238 | 7084.033 10230.152 |
| LCT | STU | 8197.625 | 6167.62 - 10338.243 | 7671.266 | 5205.261 10364.956 |
| LCT | GBR | 8412.48 | 7754.023 - 9084.704 | 8185.932 | 7111.164 9226.64 |
| LCT | CEU | 8662.519 | 8064.022 - 9340.064 | 10701.2 | 9579.387 11839.975 |
| LCT | GIH | 8732.25 | 7724.106 - 9921.599 | 9926.97 | 8596.736 11379.234 |
| LCT | IBS | 9341.408 | 8687.717 - 9988.713 | 7593.055 | 6602.516 8681.566 |
| LCT | PJL | 9514.453 | 8596.386 - 10382.874 | 9500.563 | 8511.207 10618.241 |

Supplementary Table 3. TMRCA estimates from 1000 Genomes Project data. TMRCA estimates from the 1000 Genomes Project panel using the Mbp and fine-scale recombination rate. These results represent the distributions with the highest posterior probability among the 5 replicates shown with transparency in Figures 3 and 4. All estimates are scaled to a generation time of 29 years.

| Gene | Population | Years Before Present | Time Estimated | Approach | Method | Reference |
|-------|----------------|-----------------------|------------------|-------------------------------|-----------------------------|---------------------------|
| LCT | CEU | 8560 (7328 - 9861) | t_1 | LD and Allele Freq. | HMM | Chen et al. [2015] |
| LCT | CEU | 7466 (5516 - 11019) | t_1 | LD, Mutation and Allele Freq. | ABC | Nakagome et al. [2016] |
| LCT | CEU | 9277 (4021 - 21102) | t_1 | LD, Mutation and Allele Freq. | ABC | Tishkoff et al. [2007] |
| LCT | CEU | 13246 (2538 - 23954) | t_{ca} | LD | Reich and Goldstein [1999] | Bersaglieri et al. [2004] |
| LCT | Finland | 2791 (1885 - 3698) | t_{ca} | LD | Reich and Goldstein [1999] | Bersaglieri et al. [2004] |
| LCT | Finland (west) | 5921 (5266 - 6576) | t_{ca} | Mutation | Bandelt et al. [1999] | Enattah et al. [2008] |
| LCT | Finland (east) | 6177 (5209 - 7145) | t_{ca} | Mutation | Bandelt et al. [1999] | Enattah et al. [2008] |
| LCT | Finland (west) | 6119 (5655 - 6542) | t_{ca} | LD | Serre et al. [1990] | Enattah et al. [2007] |
| LCT | Finland | 9425 (4350 - 18125) | t_{ca} | LD | Seixas et al. [2001] | Coelho et al. [2005] |
| LCT | Finland (east) | 7155 (77 - 26293) | t_{ca} | Mutation | Stumpf and Goldstein [2001] | Enattah et al. [2007] |
| LCT | Finland | 10730 (0 - 39440) | t_{ca} | Mutation | Stumpf and Goldstein [2001] | Coelho et al. [2005] |
| LCT | Italy | 9645 (3990 - 32120) | t_{ca} | LD | Seixas et al. [2001] | Coelho et al. [2005] |
| LCT | Italy | 23710 (5000 - 66120) | t_{ca} | Mutation | Stumpf and Goldstein [2001] | Coelho et al. [2005] |
| LCT | Portugal | 10869 (6890 - 19940) | t_{ca} | LD | Seixas et al. [2001] | Coelho et al. [2005] |
| LCT | Portugal | 21959 (4489 - 62199) | t_{ca} | Mutation | Stumpf and Goldstein [2001] | Coelho et al. [2005] |
| LCT | Finland | 12992 (1740 - 75284) | t_1 | LD, Mutation and Allele Freq. | ABC | Peter et al. [2012] |
| LCT | European | 8632 (7257 - 10020) | t_1 | Spatial Modeling | ABC | Itan et al. [2009] |
| KITLG | Portugal | 32277 (6003 - 80683) | t_1 | LD, Mutation and Allele Freq. | ABC | Beleza et al. [2013b] |
| KITLG | Han Chinese | 32045 (6032 - 98165) | t_1 | LD, Mutation and Allele Freq. | ABC | Beleza et al. [2013b] |
| KITLG | CEU | 26386 (16846 - 56928) | t_1 | LD and Allele Freq. | HMM | Chen et al. [2015] |
| OCA2 | Han Chinese | 17056 (12912 - 25246) | t_1 | LD and Allele Freq. | HMM | Chen et al. [2015] |
| ADH1B | East Asians | 4060 (2320 - 5800) | t_{ca} | Mutation | Su et al. [1999] | Li et al. [2011] |
| ADH1B | East Asians | 10026 (8512 - 11540) | t_{ca} | LD | Serre et al. [1990] | Peng et al. [2010] |
| ADH1B | Han Chinese | 8957 (1533 - 34618) | t_1 | LD, Mutation and Allele Freq. | ABC | Peter et al. [2012] |
| EDAR | East Asians | 8666 (914 - 59711) | t^{fix} | LD and Mutation | ABC | Bryk et al. [2008] |
| EDAR | Han Chinese | 8463 (3192 - 32443) | t_1 | LD, Mutation and Allele Freq. | ABC | Peter et al. [2012] |
| EDAR | Han Chinese | 35873 (15283 - 45907) | t_1 | Allele Freq. Spectrum | ABC | Kamberov et al. [2013] |
| EDAR | Han Chinese | 42328 (40339 - 44317) | t_1 | Allele Freq. Spectrum | MLE | Kamberov et al. [2013] |

Supplementary Table 4. Previous allele age point estimates and 95% confidence intervals for the loci considered in this study. All estimates are scaled to a generation time of 29 years and, where possible for SNP data, scaled to a mutation rate of 1.6×10^{-8} . For the times estimated in each case, t_1 refers to the time of mutation, t_{ca} is time to the common ancestor and t^{fix} is time since fixation [Przeworski, 2003].

| | | startmrca | | | | Chen et al. [2015] | | ForSim | | IS-Age | |
|------------|-----------|------------------|---------------|---------------------|---------------|--------------------|---------------|---------|--------|---------|--------|
| | | local ref. panel | | diverged ref. panel | | | | | | | |
| | | Mean | RMSE | Mean | RMSE | Mean | RMSE | Mean | RMSE | Mean | RMSE |
| Freq = 80% | | | | | | | | | | | |
| | s = 0.005 | -0.4786 | 0.5638 | -0.1365 | 0.3186 | -0.0410 | 0.3637 | -1.2113 | 1.2313 | 0.0330 | 0.7255 |
| | s = 0.01 | -0.3830 | 0.4657 | 0.0132 | 0.3014 | -0.0190 | 0.3053 | -0.2830 | 0.4125 | -0.0984 | 0.6384 |
| | s = 0.05 | -0.3074 | 0.4044 | 0.0027 | 0.3289 | 0.1469 | 0.3253 | 1.7866 | 1.8039 | 0.2719 | 0.4296 |
| Freq = 40% | | | | | | | | | | | |
| | s = 0.005 | -0.5070 | 0.5882 | -0.1607 | 0.3801 | -0.0450 | 0.7080 | -0.4826 | 0.6284 | -0.3158 | 0.6331 |
| | s = 0.01 | -0.3198 | 0.4294 | -0.0132 | 0.2897 | -0.0621 | 0.5949 | 0.2720 | 0.4275 | -0.1586 | 0.8631 |
| | s = 0.05 | -0.2009 | 0.3731 | -0.0024 | 0.3497 | 0.2884 | 0.6554 | 2.1822 | 2.1822 | 0.4936 | 0.6391 |

Supplementary Table 5. Method comparison. Comparison of accuracy and bias results between our estimator “startmrca” and previously reported results from Chen et al. [2015] under different end frequencies (Freq) and selection strengths (s). Root mean squared errors (RMSE) and means were computed using $\log_2(\text{Estimated}/\text{True})$ TMRCA values. Results in bold indicate the method with lower RMSE values than the others. Simulations were matched to include a sample size of 200 haplotypes of length 1Mbp with a mutation and recombination rate of 1×10^{-8} . The diverged reference panel is sampled from a population that split with the beneficial allele carrier population .5N generations in the past. ForSim is the forward simulation method by Beleza et al. [2013a]; and IS-Age is the importance sampling-based method by Chen and Slatkin [2013].

| Reference Panel | Selection | Frequency | Model A | | Model B | |
|-----------------|-----------|-----------|----------|----------------|----------|----------------|
| | | | Mean | RMSE | Mean | RMSE |
| Diverged | 0.005 | 0.4 | -0.16070 | 0.38010 | -0.17842 | 0.38143 |
| Diverged | 0.01 | 0.4 | -0.01320 | 0.28970 | -0.00003 | 0.34665 |
| Diverged | 0.05 | 0.4 | -0.00240 | 0.34970 | 0.03975 | 0.40088 |
| Diverged | 0.005 | 0.8 | -0.13650 | 0.31860 | -0.16190 | 0.33011 |
| Diverged | 0.01 | 0.8 | 0.01320 | 0.30140 | 0.05729 | 0.28370 |
| Diverged | 0.05 | 0.8 | 0.00270 | 0.32890 | 0.04515 | 0.36350 |
| Local | 0.005 | 0.4 | -0.50700 | 0.58820 | -0.52444 | 0.61450 |
| Local | 0.01 | 0.4 | -0.31980 | 0.42940 | -0.35912 | 0.44024 |
| Local | 0.05 | 0.4 | -0.20090 | 0.37310 | -0.23170 | 0.37473 |
| Local | 0.005 | 0.8 | -0.47860 | 0.56380 | -0.52219 | 0.60239 |
| Local | 0.01 | 0.8 | -0.38300 | 0.46570 | -0.44445 | 0.50871 |
| Local | 0.05 | 0.8 | -0.30740 | 0.40440 | -0.30716 | 0.40357 |

Supplementary Table 6. Model Comparison. Comparison of accuracy and bias results between different approaches for modelling invariant sites among background haplotypes in the carrier panel (β_{iw}). Model A refers to the original Li and Stephens [2003] model. Model B uses the singleton rate in the reference panel (see Appendix C). As in Supplementary Table 6 above, root mean squared errors (RMSE) and means were computed using $\log_2(\text{Estimated}/\text{True})$ TMRCA values. Results in bold indicate the model with lowest RMSE value. Frequency refers to the end frequency of the beneficial allele trajectory.

| | Posterior | | Bootstraps | |
|------------|-----------|-------------|------------|-------------|
| | Mean | 95% C.I. | Mean | 95% C.I. |
| Freq = 80% | | | | |
| s = 0.01 | 728 | (643 - 813) | 721 | (623 - 838) |
| s = 0.1 | 76 | (62 - 90) | 77 | (67 - 88) |
| Freq = 40% | | | | |
| s = 0.01 | 388 | (345 - 434) | 413 | (360 - 485) |
| s = 0.1 | 52 | (43 - 62) | 53 | (44 - 64) |

Supplementary Table 7. Bootstrap Estimate Comparisons. Comparison of TMRCA estimates from posterior results of simulated data versus estimates from 100 bootstrap replicates of those same datasets. For each dataset, we simulated 100 beneficial allele carriers and 20 non-carriers for the reference panel. Bootstrap replicates were generated by resampling among the beneficial allele carriers. We used mutation and recombination rates of 1×10^{-8} and a population size of 10000.

References

- Hans-Jurgen Bandelt, Peter Forster, and Arne Röhl. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1):37–48, 1999.
- Sandra Beleza, Nicholas A Johnson, Sophie I Candille, Devin M Absher, Marc A Coram, Jailson Lopes, Joana Campos, Isabel Inês Araújo, Tovi M Anderson, Bjarni J Vilhjálmsson, et al. Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genetics*, 9(3):e1003372, 2013a.
- Sandra Beleza, António M Santos, Brian McEvoy, Isabel Alves, Cláudia Martinho, Emily Cameron, Mark D Shriver, Esteban J Parra, and Jorge Rocha. The timing of pigmentation lightening in Europeans. *Molecular Biology and Evolution*, 30(1):24–35, 2013b.
- Todd Bersaglieri, Pardis C Sabeti, Nick Patterson, Trisha Vanderploeg, Steve F Schaffner, Jared A Drake, Matthew Rhodes, David E Reich, and Joel N Hirschhorn. Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics*, 74(6):1111–1120, 2004.
- Jarosław Bryk, Emilie Hardouin, Irina Pugach, David Hughes, Rainer Strotmann, Mark Stoneking, and Sean Myles. Positive selection in East Asians for an EDAR allele that enhances NF- κ B activation. *PLoS One*, 3(5):e2209–e2209, 2008.
- Hua Chen and Montgomery Slatkin. Inferring selection intensity and allele age from multilocus haplotype structure. *G3: Genes— Genomes— Genetics*, 3(8):1429–1442, 2013.
- Hua Chen, Jody Hey, and Montgomery Slatkin. A hidden Markov model for investigating recent positive selection through haplotype structure. *Theoretical Population Piology*, 99:18–30, 2015.
- Margarida Coelho, Donata Luiselli, Giorgio Bertorelle, Ana Isabel Lopes, Susana Seixas, Giovanni Destro-Bisol, and Jorge Rocha. Microsatellite variation and evolution of human lactase persistence. *Human Genetics*, 117(4):329–339, 2005.
- Nabil Sabri Enattah, Aimee Trudeau, Ville Pimenoff, Luigi Maiuri, Salvatore Auricchio, Luigi Greco, Mauro Rossi, Michael Lentze, JK Seo, Soheila Rahgozar, et al. Evidence of still-

- ongoing convergence evolution of the lactase persistence T-13910 alleles in humans. *The American Journal of Human Genetics*, 81(3):615–625, 2007.
- Nabil Sabri Enattah, Tine GK Jensen, Mette Nielsen, Rikke Lewinski, Mikko Kuokkanen, Heli Rasinpera, Hatem El-Shanti, Jeong Kee Seo, Michael Alifrangis, Insaf F Khalil, et al. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *The American Journal of Human Genetics*, 82(1):57–72, 2008.
- Yuval Itan, Adam Powell, Mark A Beaumont, Joachim Burger, Mark G Thomas, et al. The origins of lactase persistence in Europe. *PLoS Computational Biology*, 5(8):e1000491–e1000491, 2009.
- Yana G Kamberov, Sijia Wang, Jingze Tan, Pascale Gerbault, Abigail Wark, Longzhi Tan, Yajun Yang, Shilin Li, Kun Tang, Hua Chen, et al. Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell*, 152(4):691–702, 2013.
- Hui Li, Sheng Gu, Yi Han, Zhi Xu, Andrew J Pakstis, Li Jin, Judith R Kidd, and Kenneth K Kidd. Diversification of the ADH1B gene during expansion of modern humans. *Annals of Human Genetics*, 75(4):497–507, 2011.
- Na Li and Matthew Stephens. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, 165(4):2213–2233, 2003.
- Shigeki Nakagome, Gorka Alkorta-Aranburu, Roberto Amato, Bryan Howie, Benjamin M Peter, Richard R Hudson, and Anna Di Rienzo. Estimating the ages of selection signals from different epochs in human history. *Molecular Biology and Evolution*, 33(3):657–669, 2016.
- Yi Peng, Hong Shi, Xue-bin Qi, Chun-jie Xiao, Hua Zhong, Z Ma Run-lin, and Bing Su. The ADH1B Arg47His polymorphism in East Asian populations and expansion of rice domestication in history. *BMC Evolutionary Biology*, 10(1):15, 2010.
- Benjamin M Peter, Emilia Huerta-Sanchez, and Rasmus Nielsen. Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS Genetics*, 8(10):e1003011, 2012.

- 168 Molly Przeworski. Estimating the time since the fixation of a beneficial allele. *Genetics*, 164(4):
169 1667–1676, 2003.
- 170 DE Reich and DB Goldstein. Estimating the age of mutations using variation at linked markers.
171 *Microsatellites: Evolution and Applications*. Oxford University Press, Oxford, pages 129–138,
172 1999.
- 173 Susana Seixas, Oscar Garcia, Jesus M Trovada, Teresa M Santos, António Amorim, and Jorge
174 Rocha. Patterns of haplotype diversity within the serpin gene cluster at 14q32. 1: insights
175 into the natural history of the α 1-antitrypsin polymorphism. *Human Genetics*, 108(1):20–30,
176 2001.
- 177 Jean-Louis Serre, Brigitte Simon-Bouy, Etienne Mornet, Bartolomé Jaume-Roig, Angeliki Bal-
178 assopoulou, Marienne Schwartz, Agnès Taillandier, Joëlle Boue, and André Boué. Studies of
179 RFLP closely linked to the cystic fibrosis locus throughout Europe lead to new considerations
180 in populations genetics. *Human Genetics*, 84(5):449–454, 1990.
- 181 Michael PH Stumpf and David B Goldstein. Genealogical and evolutionary inference with the
182 human Y chromosome. *Science*, 291(5509):1738–1742, 2001.
- 183 Bing Su, Junhua Xiao, Peter Underhill, Ranjan Deka, Weiling Zhang, Joshua Akey, Wei Huang,
184 Di Shen, Daru Lu, Jingchun Luo, et al. Y-chromosome evidence for a northward migration of
185 modern humans into Eastern Asia during the last Ice Age. *The American Journal of Human*
186 *Genetics*, 65(6):1718–1724, 1999.
- 187 Sarah A Tishkoff, Floyd A Reed, Alessia Ranciaro, Benjamin F Voight, Courtney C Babbitt,
188 Jesse S Silverman, Kweli Powell, Holly M Mortensen, Jibril B Hirbo, Maha Osman, et al.
189 Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*,
190 39(1):31–40, 2007.