

Figure S1: **Patterns of neutral diversity over time.** This figure is the analog to Figure 3 with lower population mutation rate ($N\mu = 0.5$). Correspondingly, the sweeps show characteristics of migrant-derived (rather than locally-derived) sweep patterns.

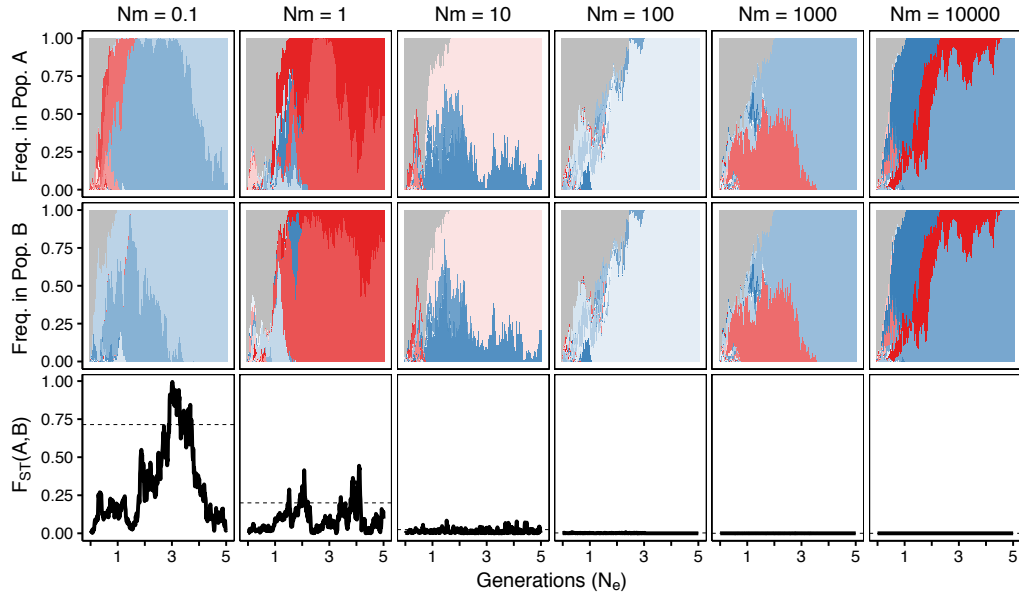
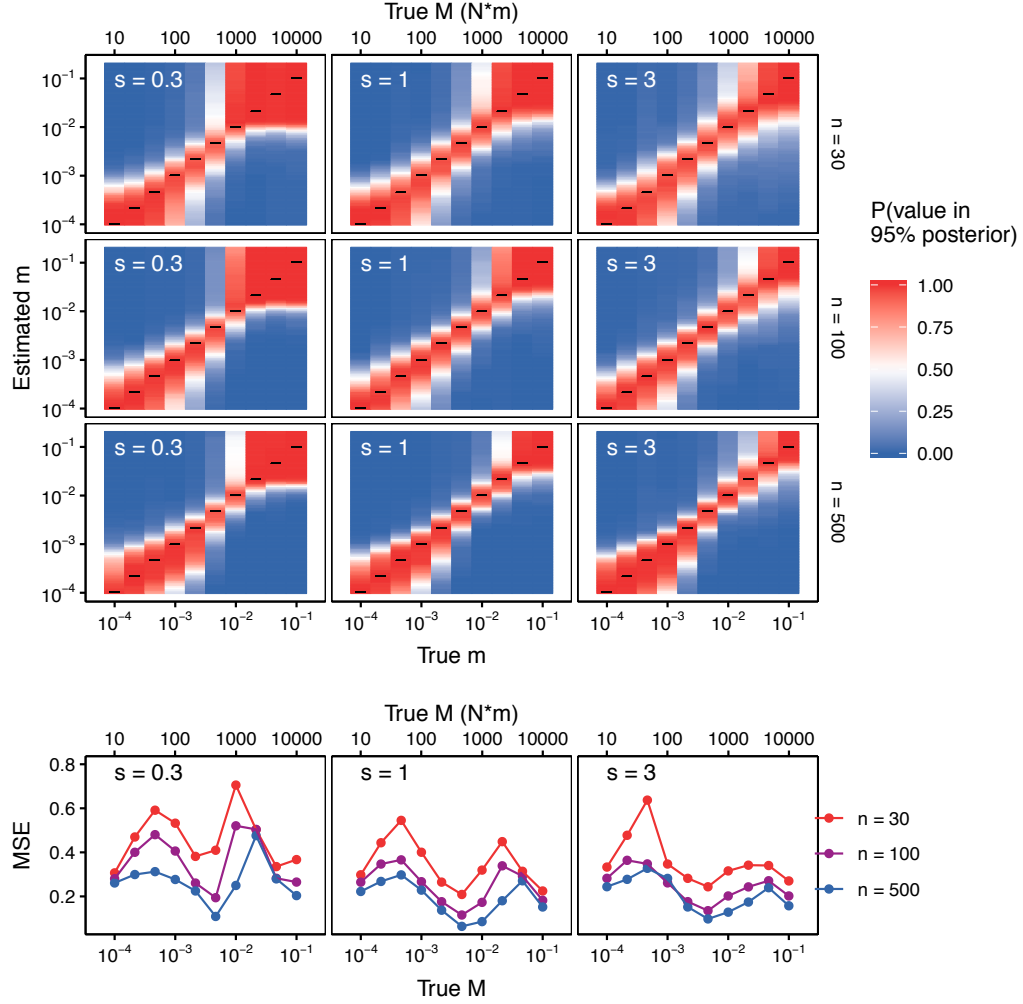


Figure S2: **Patterns of neutral diversity over time.** This figure is the analog to Figure 3 under neutrality (i.e. selection strength $s = 0$). Correspondingly, the timescale for alleles to reach high frequencies is now in units of 10^5 (i.e., N_e) generations instead of single generations. The dashed lines mark the result $F_{ST} = 1/(4Nm + 1)$, and the figure caption is otherwise shared with Figure 3.



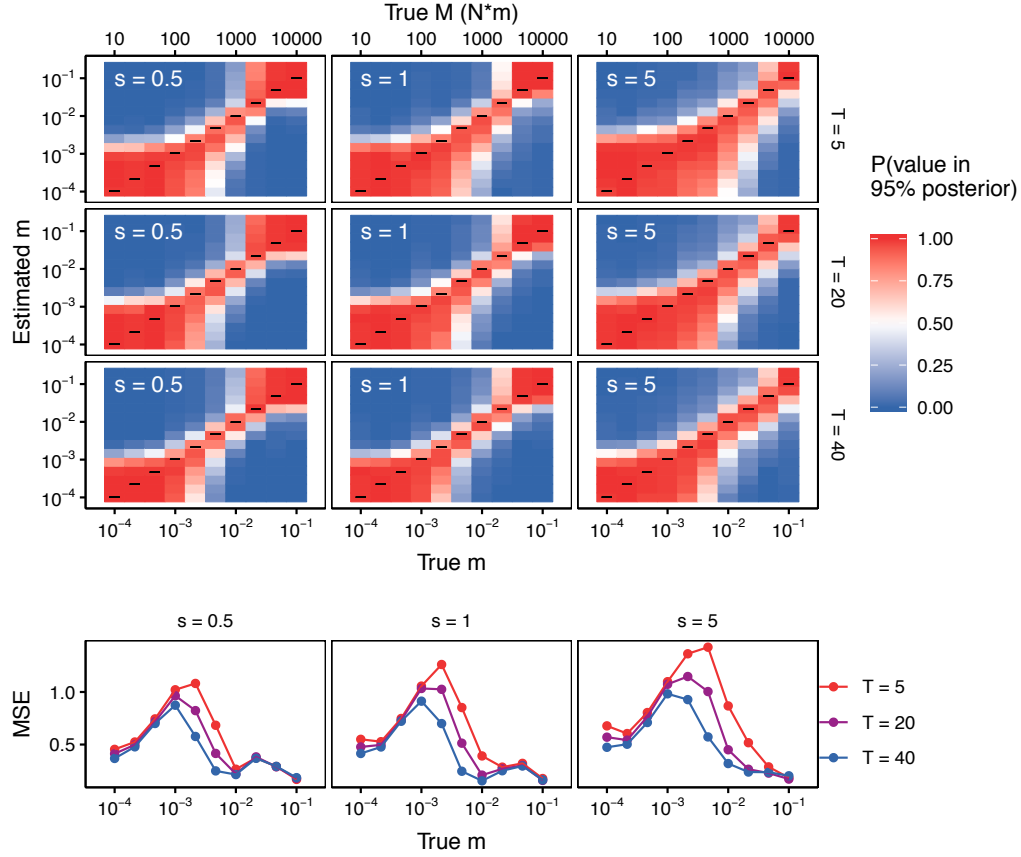


Figure S4: **Increasing T decreases the average distance between the posterior and the truth.** Similar to Figure 5, this figure displays the ability of the ABC procedure to recapture a test ‘Estimated m ’ under different values of s and M , with different sampling time configurations, as shown in rows. Sampling time points were taken at generations 5, fixation, and fixation + T , with $T = (5, 20, 40)$ Increasing T can result in estimation more focused around the $x = y$ line, indicating precise and accurate estimation. This performance is displayed in the bottom row as decreased log-MSE among the highest T simulations, across values of s and m . This effect is most important for particular values of T ($N\mu = 1$).

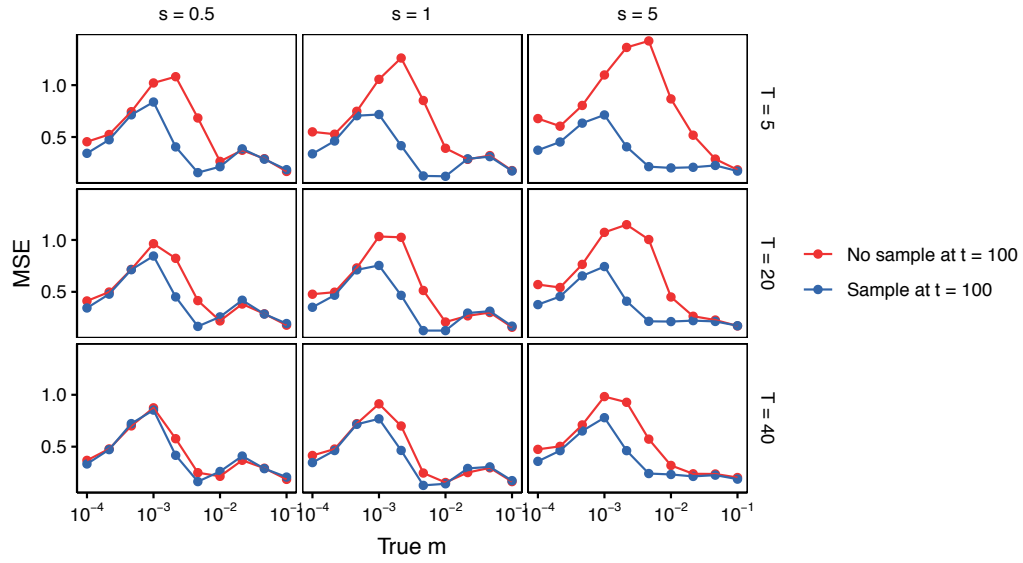


Figure S5: **Adding one sample time point at generation 100 substantially improves fit, but significantly more when T is small.** For different values of T , s and M , adding an additional time point at generation 100 decreases the log-MSE of the population. When $T = 40$, there is relatively little improvement, because the third sampling time point is likely already near $t = 100$ ($N\mu = 1$).

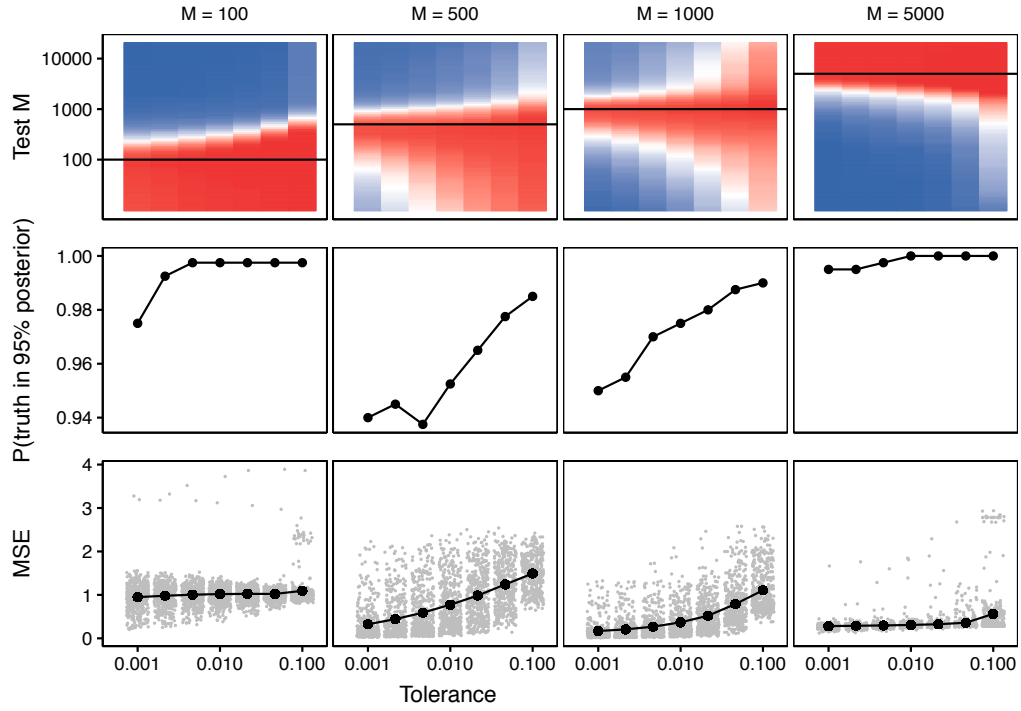


Figure S6: **Decreasing the tolerance decreases the probability of finding the actual truth, but decreases the average distance between the posterior and the truth** The performance of the ABC procedure is evaluated in three different ways. In the top row, we show the probability of including different values of M within posteriors computed with different tolerances. The true value of M is shown with the black line. The blue to red probability spectrum indicates that lower tolerances include fewer false values of M within their corresponding 95% posteriors. In the second row, the proportion of trials with a given tolerance that successfully capture the true M decreases with the tolerance. In the third row, the log-MSE decreases as the tolerance is decreased. Each grey dot shows an individual simulation with a given M and the black dots display the median ($N\mu = 1, s = 1$).

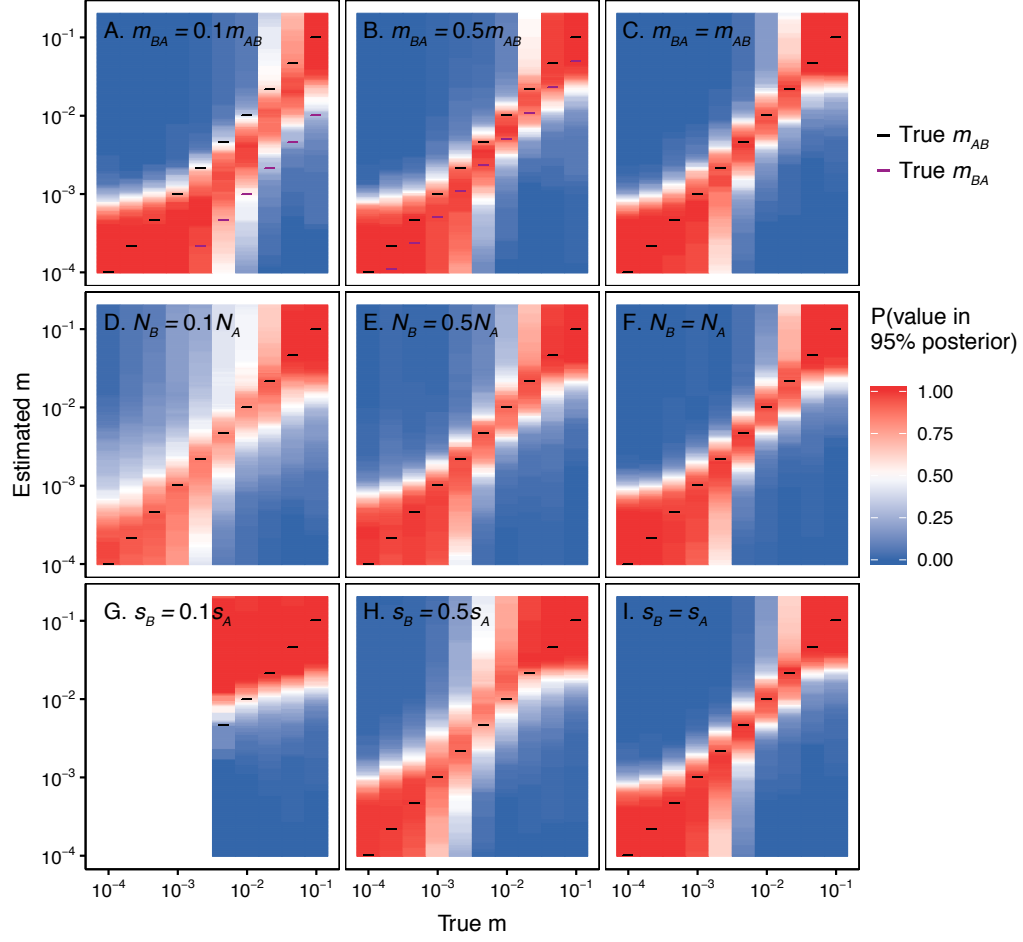


Figure S7: **Impact of asymmetry of model parameters on estimation procedure.** 100 simulations are run for each true value of m on the x -axis and 100 posteriors are produced. We plot the proportion of 95% posteriors for a given true m (shown on the x -axis) that include the various values of estimated m (shown on the y -axis). Values that are more red indicate that the posteriors successfully capture the test 'Estimated m ' value on the y -axis whereas blue values indicate that few posteriors include the y -axis value. Black horizontal lines indicate where the true value matches the y -axis 'Estimated m .' Simulations are generated across a gradient of values for m ($m \in (10^{-4} - 10^{-1})$). Unless otherwise noted, parameter choices are given as follows: $s_A = s_B = (1)$, sampling at generations $(5, T = \text{fixation}, T + 30, 100)$, $N_A = N_B = 10^5$, $n = 100$). In addition, in panels **A**, **B** and **C**, $m_{BA} = pm_{AB}$, $p \in (0.1, 0.5, 1)$, respectively. In panels **D**, **E** and **F**, $N_B = pN_A$, $p \in (0.1, 0.5, 1)$, respectively. In panels **G**, **H** and **I**, $s_B = ps_A$, $p \in (0.1, 0.5, 1)$, respectively.

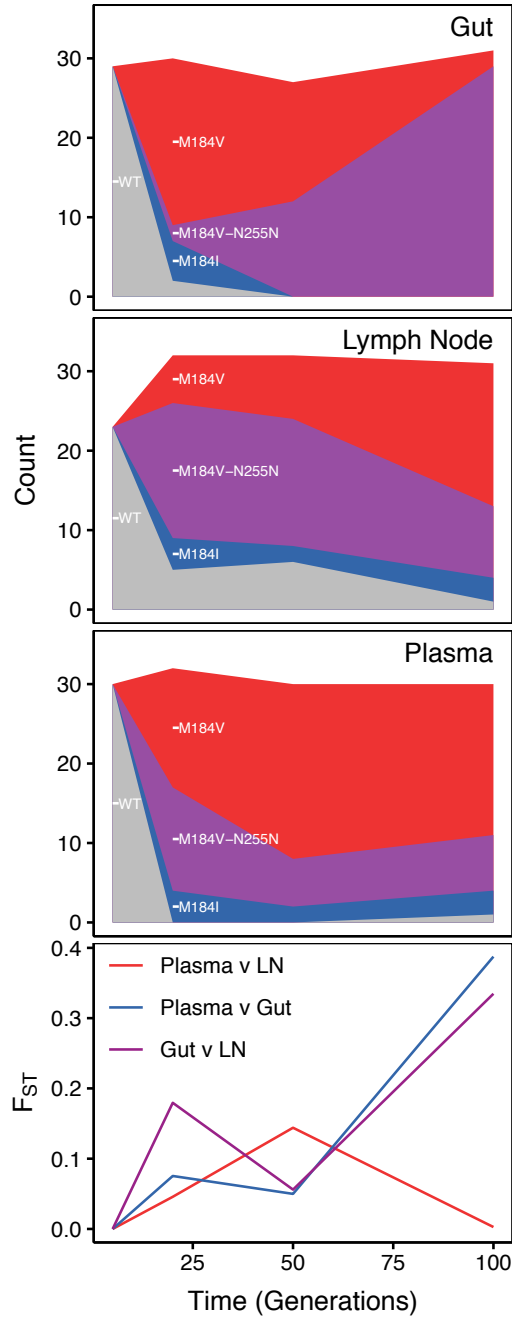


Figure S9: **Dynamics of drug resistance fixation across space and time in a treated Simian-HIV population with strict haplotype definitions** ($c = 5$). The top row shows diagrams of drug resistant haplotypes spreading in different sampling locations over time sampled at generations 7, 21, 49 and 98 after the onset of selection via the drug FTC in the gut, lymph node and blood plasma. Each color represents a distinct lineage separated by at least one mutation. Mutations are included if they are observed in at least five copies at or before a sweep is at least 70% complete (see Materials and Methods for full description of how haplotypes were defined). The bottommost panel plots pairwise F_{ST} between pairs of sampling locations.

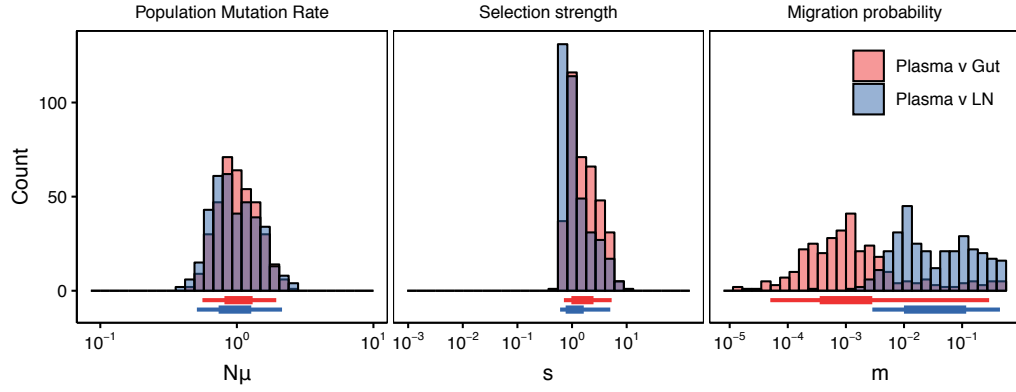


Figure S10: **Estimation of population parameter rates from intra-patient Simian-HIV data sampled from different subpopulations with lenient haplotype definitions ($c = 1$).** The top row shows diagrams of drug resistant haplotypes spreading in different subpopulations over time sampled at generations 7, 21, 49 and 98 in the gut, lymph node and blood plasma. Each color represents a distinct lineage separated by at least one mutation. The resulting posteriors are given for the ABC procedure for comparisons between the plasma and gut (red) and plasma and lymph node (blue). $tol = 2.5 \times 10^{-4}$. 95% posteriors for these distributions are shown in Table S1.

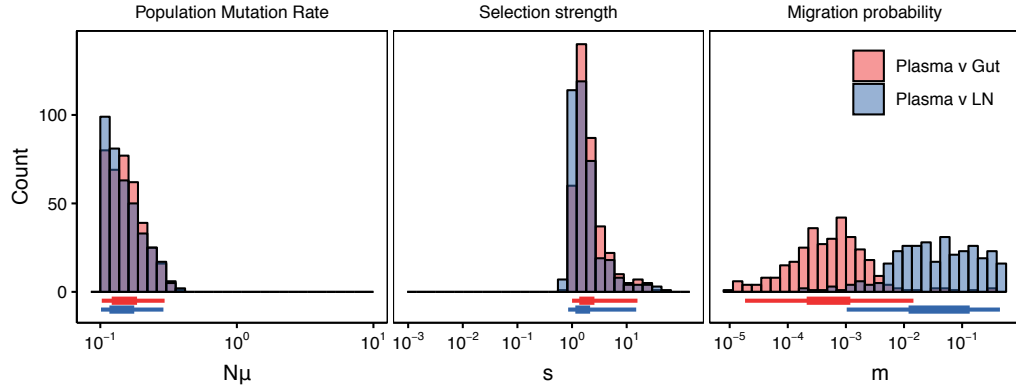


Figure S11: **Estimation of population parameter rates from intra-patient Simian-HIV data sampled from different subpopulations with strict haplotype definitions ($c = 5$).** The top row shows diagrams of drug resistant haplotypes spreading in different subpopulations over time sampled at generations 7, 21, 49 and 98 in the gut, lymph node and blood plasma. Each color represents a distinct lineage separated by at least one mutation. The resulting posteriors are given for the ABC procedure for comparisons between the plasma and gut (red) and plasma and lymph node (blue). $tol = 2.5 \times 10^{-4}$. 95% posteriors for these distributions are shown in Table S1.

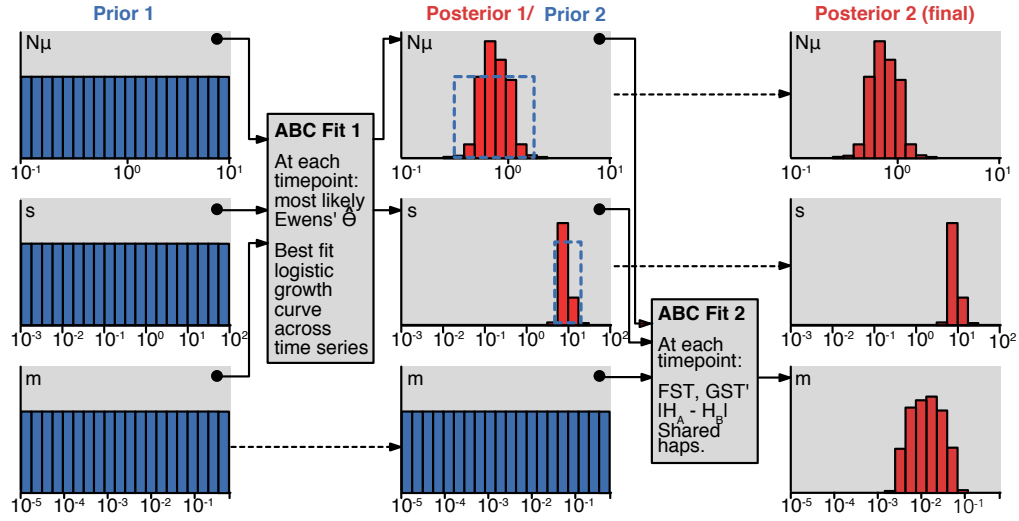


Figure S12: **Schematic of 2-step approximate Bayesian computation (ABC) procedure** Priors for all three parameters are initially \log_{10} uniform with $N\mu \in (10^{-1}, 10^1)$, $s \in (10^{-3}, 10^2)$, $m \in (10^{-5}, .5)$, and these priors are used to estimate posteriors over $N\mu$ and s with summary statistics tailored for mutation and selection as shown in “ABC Fit 1.” The posteriors from the first ABC fit are used as priors for the second round ABC, while the posterior for m remains \log_{10} uniform with $m \in (10^{-5}, .5)$. The summary statistics in the second fit are tailored for estimating migration. The final posteriors are m posterior from ABC 2 and the s and $N\mu$ posteriors from ABC 1. For more details, see Materials and Methods, Approximate Bayesian Computation.