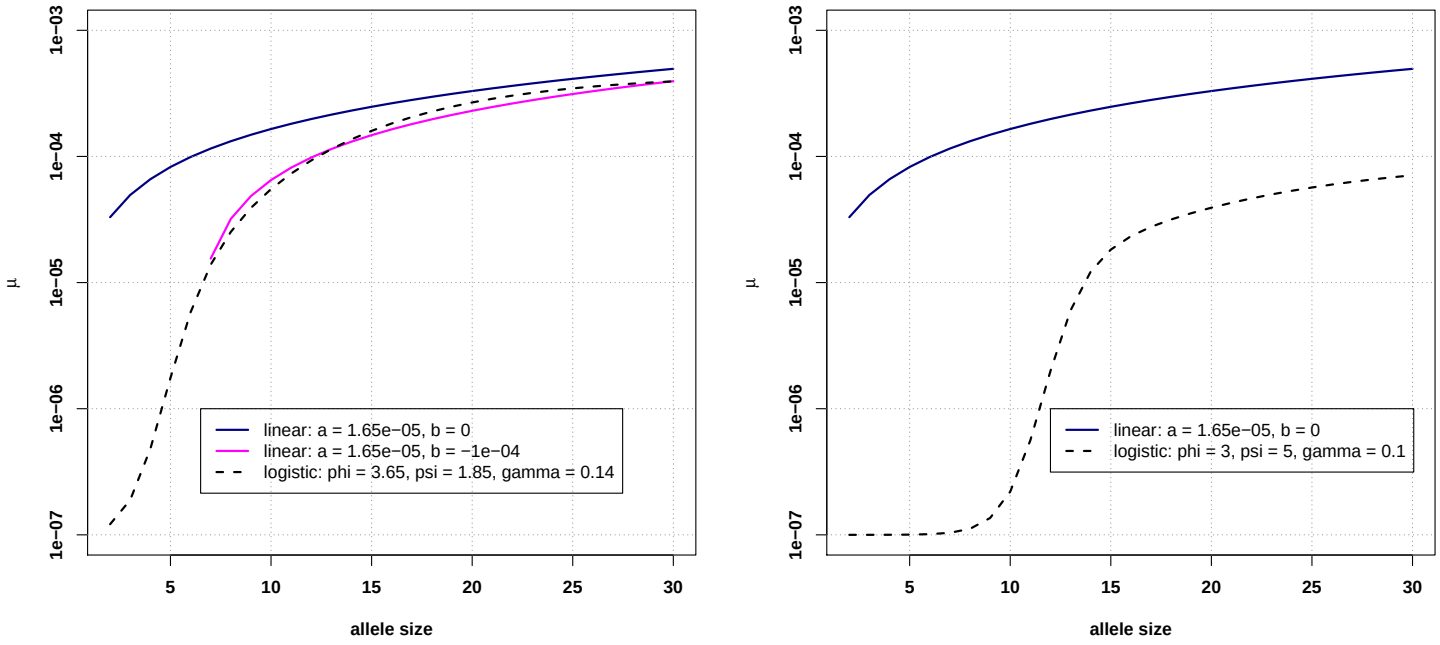
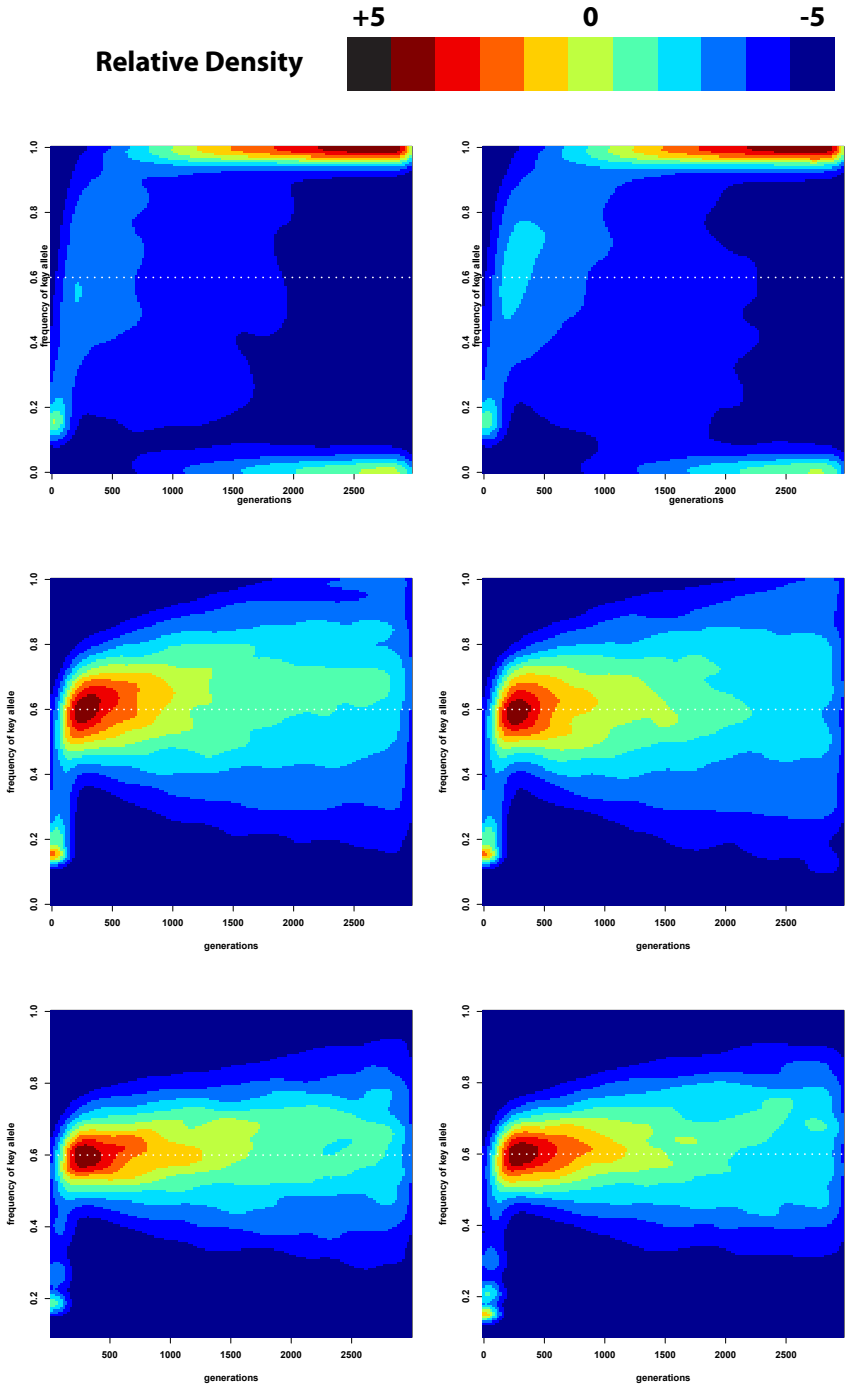
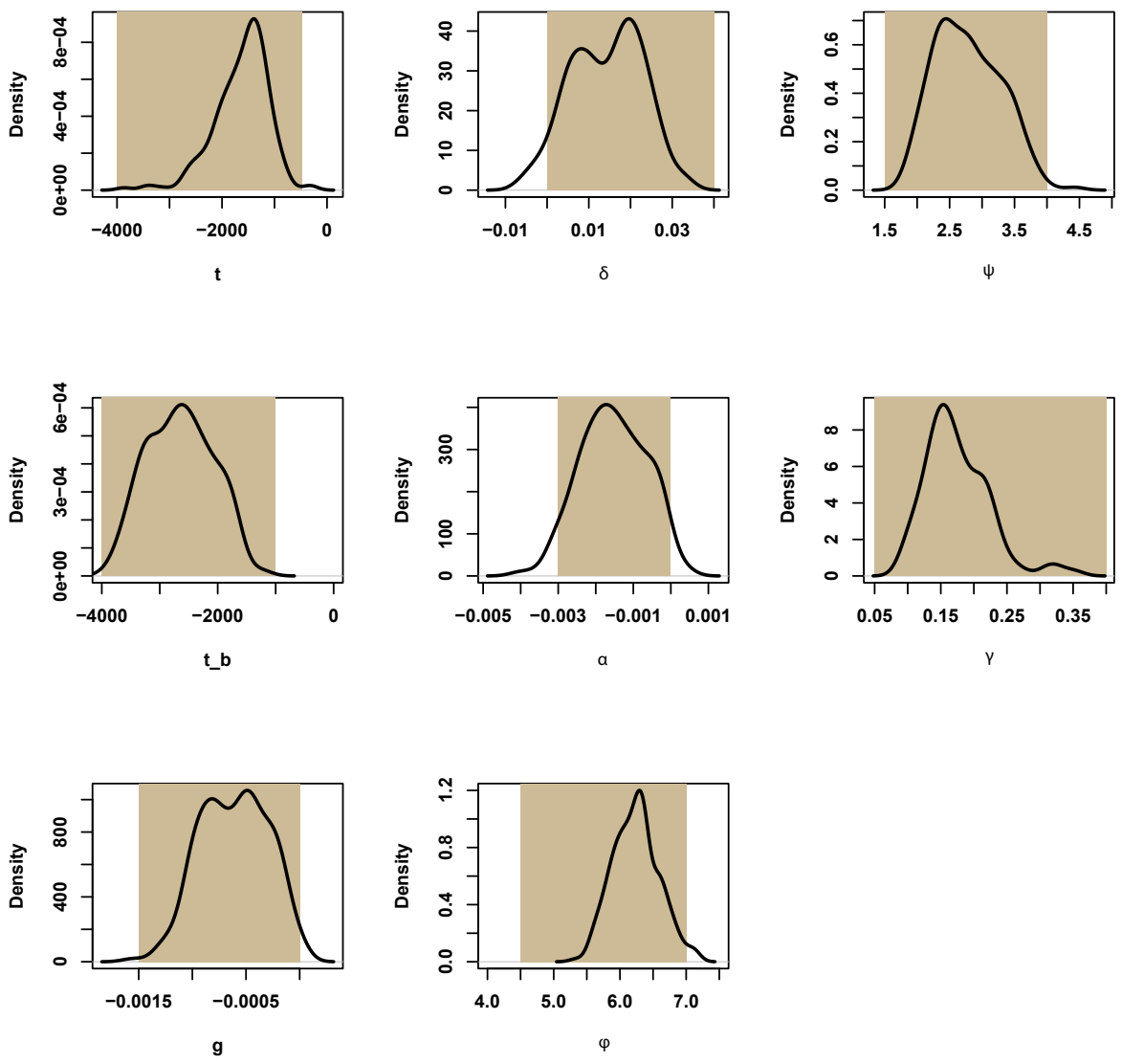


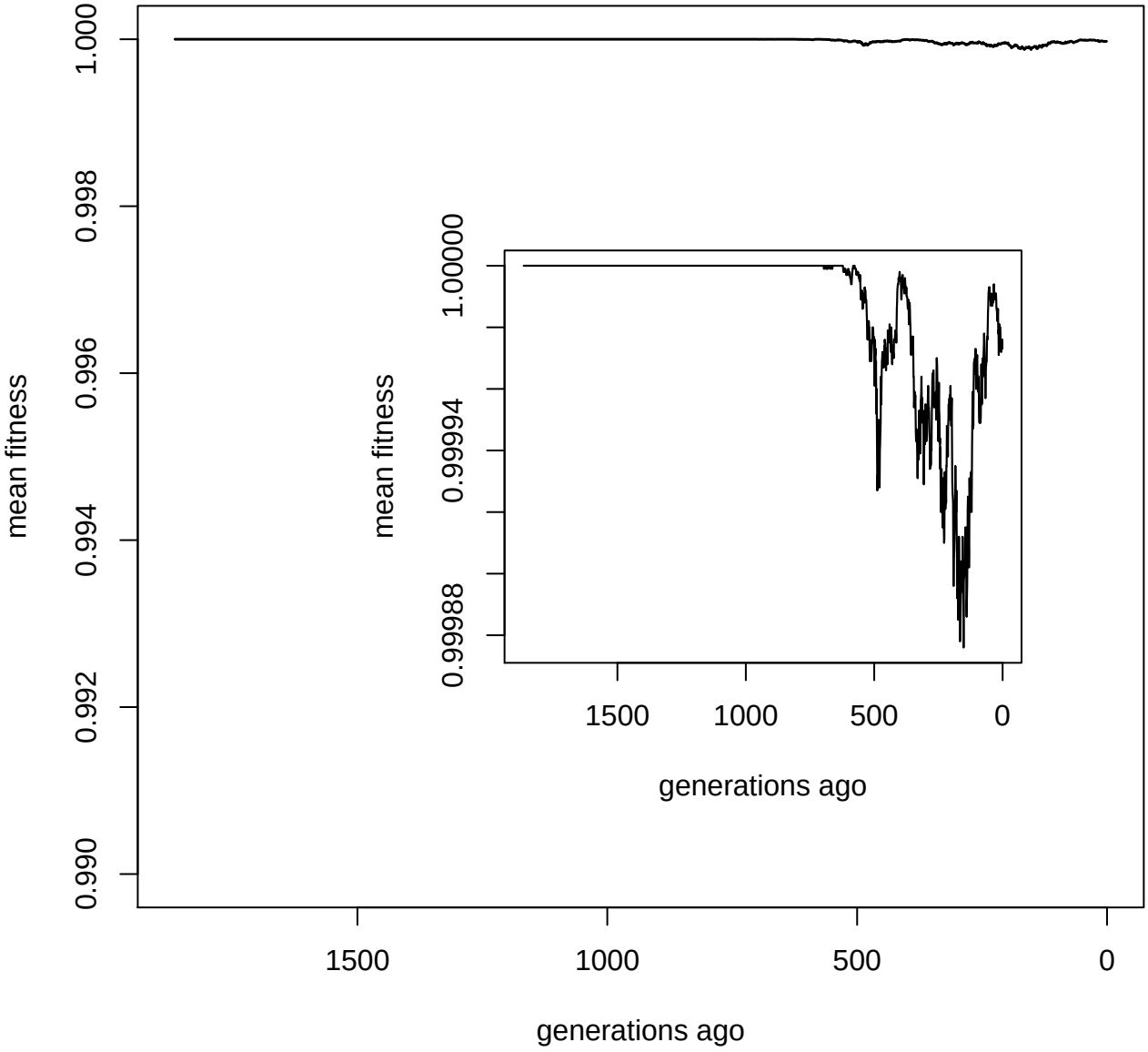
Supplementary Figure 1 Comparing logistic (dashed lines) and linear (solid lines) models of allele-specific mutation rate. The logistic model (dashed line) allows low but appreciable mutation rates for the smallest allele sizes, while the linear model can lead to $\mu = 0$ for certain negative mutation rates (left, magenta line). Also note that for larger allele sizes, the logistic model can largely recapitulate the mutational dynamics of the linear model. (right) Another illustration of the fact that the logistic model provides a larger dynamic range along the axis of allele size.



Supplementary Figure 2 Individual-based (left column) and recursion (right column) simulations produce near identical distributions of key allele frequency and time under identical starting conditions. Mutation, selection, and drift were modeled. Each contour map shows the joint distribution (across 1000 independent simulations) of the frequency of the key allele (also the most-fit allele) and the number of generations since selection began. Top row: $N_e = 1000$, middle row: $N_e = 5000$, bottom row: $N_e = 10000$. In the absence of drift the frequency of the key allele at mutation-selection equilibrium was 0.6008.



Supplementary Figure 3 Prior and posterior distributions for estimated parameters surrounding evolution of the microsatellite that causes Friedreich's ataxia. Tan boxes indicate the range of uniform priors for each parameter. Solid black lines are the regression-adjusted posterior densities. Because posterior distributions are regression adjusted, some parameters show small densities outside the prior range.



Supplementary Figure 4 Genetic load associated with emergence of E class alleles is minimal. Mean fitness as a function of time since emergence of the founding LN allele is shown. Minor deflections are noticeable on the main graph, but it is necessary to magnify the y-axis in order to observe these tiny deflections of mean fitness (inset). These data are drawn from a single simulation using the median posterior estimates of the relevant parameters (Table 1).