Supplementary material for "A population-phylogenetic view of mitochondrial heteroplasmy".

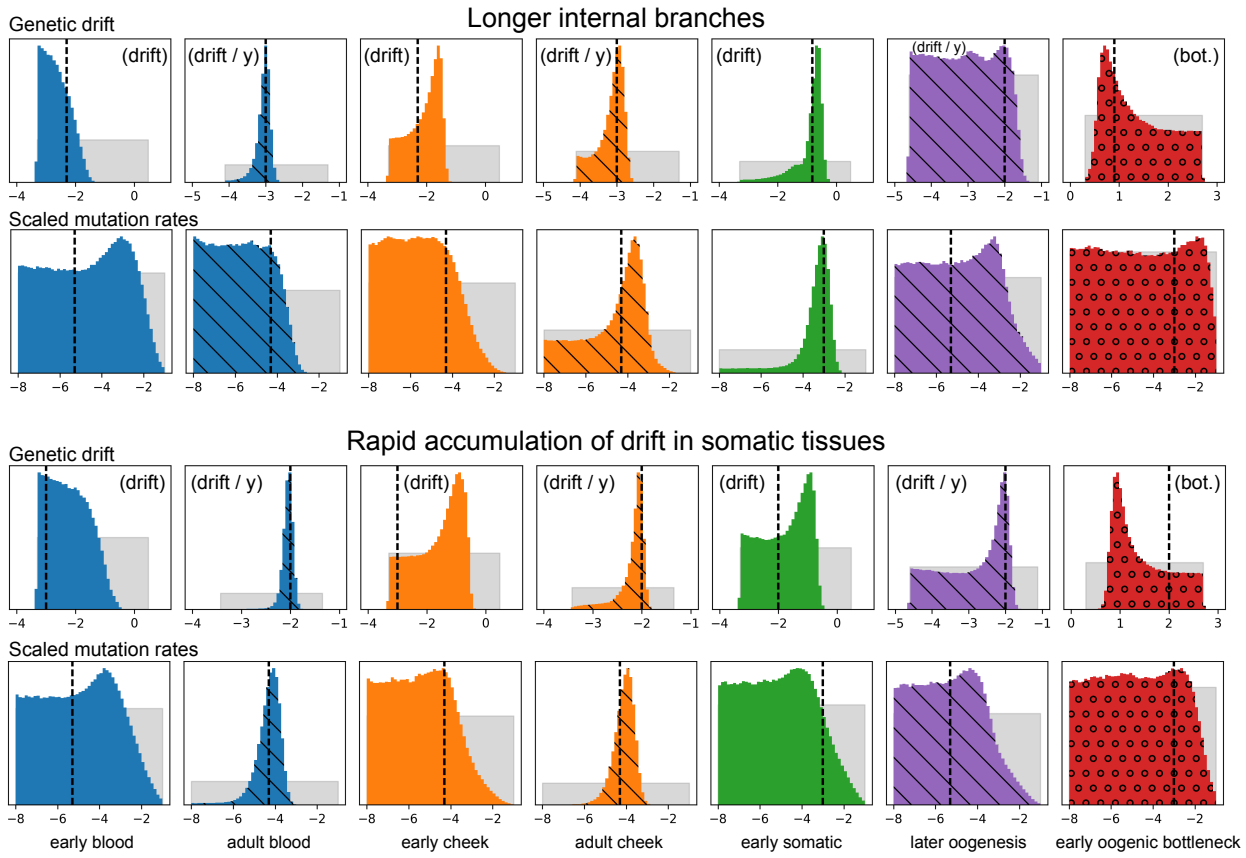

Figure S1: Inference results from additional simulations under the model assumed by our inference procedure. The first two rows show posterior distributions of parameters estimated under simulations in which the internal branches are relatively long compared to the simulations presented in the main text. These parameters were inferred from the frequencies of 103 heteroplasmic loci amongst 500 independently sites in 40 simulated families. The second pair of rows shows posterior distributions for simulations in which the rates of accumulation of genetic drift in the somatic tissues is increased compared to the simulations in the main text. In these simulations, there were 109 heteroplasmic loci amongst 400 independently segregating loci simulated in 80 families. Posterior distributions are shown with colored histograms, prior distributions are shown with gray histograms, and true parameter values are shown with dashed vertical lines. Colors match the corresponding developmental processes in Figure **??**. Distributions hashed with diagonal lines correspond to processes with drift parameterized by rates of accumulation of genetic drift with age, and circles in the red posterior distributions indicate that this process is modeled by an explicit bottleneck.

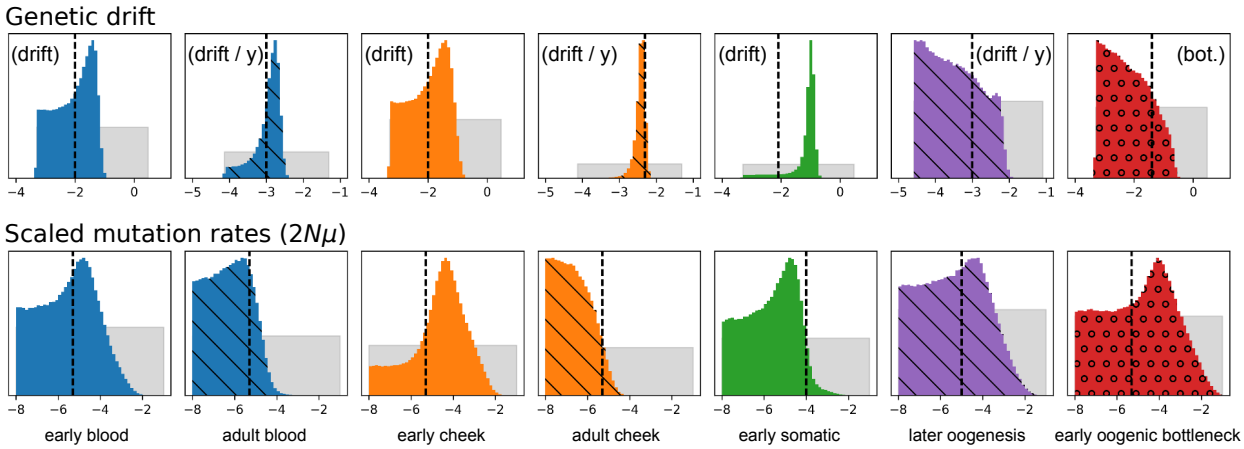

Figure S2: Inference results from simulations with no recombination between heteroplasmic loci segregating within a single family. The first row shows posterior distributions (color histograms), prior distributions (gray distributions) and simulated parameter values (dashed vertical lines) for genetic drift parameters. The second row shows the same for scaled mutation rate parameters. In order to simplify simulations, the period of genetic drift in early oogenesis was modeled as a period of genetic drift in a fixed population size rather than as a single-generation bottleneck.

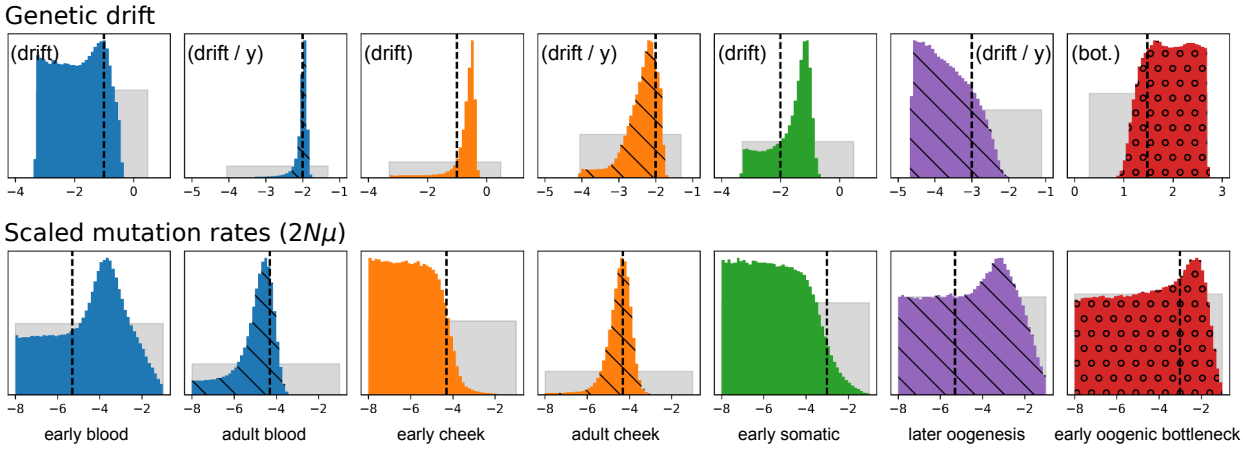

Figure S3: Inference results from simulations with noisy heteroplasmy detection. False negatives were simulated with probability 0.4 for each truly heteroplasmic locus with frequency between 0.1% and 2%. False negatives were produced at a rate of 3×10^{-5} per bp, so that 14 false negatives and 5 false positives were produced in a dataset of 101 heteroplasmic loci. As in other figures presented here, the first row shows posterior distributions (color histograms), prior distributions (gray distributions) and simulated parameter values (dashed vertical lines) for genetic drift parameters. The second row shows the same for scaled mutation rate parameters.

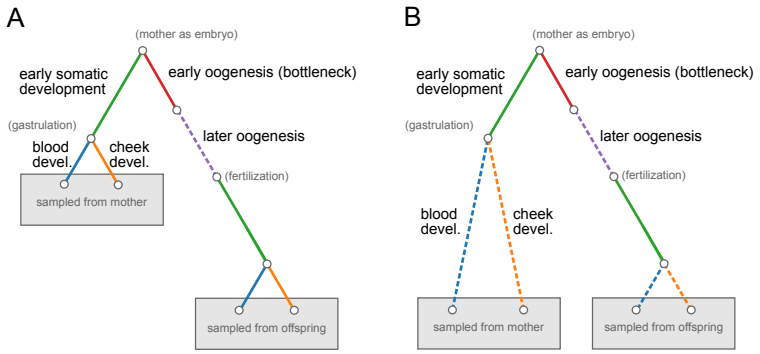

Figure S4: Two additional ontogenetic phylogenies for which we calculated the total Bayesian evidence. The two models differ in how they model genetic drift and mutation in the somatic tissues. The "fixed" model (Panel **A**) assumes that all genetic drift and mutation in the somatic tissues occurs early in development, and the the "linear" model (Panel **B**) assumes that genetic drift and mutation in somatic tissues accumulate linearly with the age of the individual. Compare these to the model in Figure **??** (termed "both"), which assumes that genetic drift and mutation in somatic tissues occurs both in a fixed amount during early development and in adulthood, accumulating linearly with age.

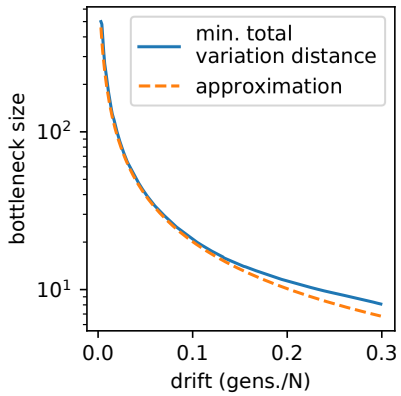

Figure S5: Translation of genetic drift into effective bottleneck sizes. The blue line shows, for different drift durations, the effective bottleneck size minimizing the total variation distance to the allele frequency transition distribution parameterized by generations per effective population size. The dashed orange line shows Equation (**??**), our approximate translation between the two parameterizations.