Supplementary information

Multiple distinct evolutionary mechanisms govern the dynamics of selfish mitochondrial genomes in *Caenorhabditis elegans*

Bryan L. Gitschlag[1,2]*, Claudia V. Pereira[1], James P. Held[1], David M. McCandlish[2], & Maulik R. Patel[1,3,4,5]*

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA

[2]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724

[3]Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, TN, USA

[4]Diabetes Research and Training Center, Vanderbilt University School of Medicine, Nashville, TN, USA

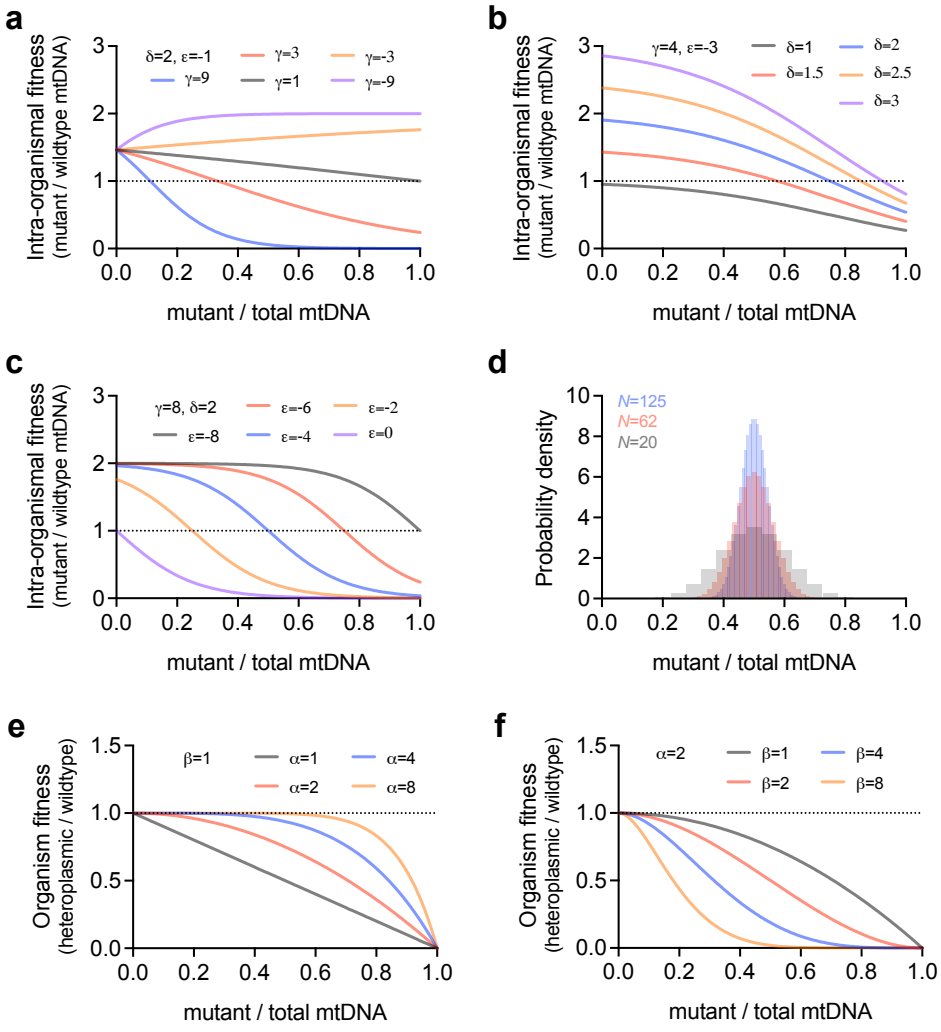
[5]Evolutionary Studies, Vanderbilt University, VU Box #34-1634, Nashville, TN, USA

*To whom correspondence should be addressed: gitschl@cshl.edu (BLG); maulik.r.patel@vanderbilt.edu (MRP)

This PDF file includes:

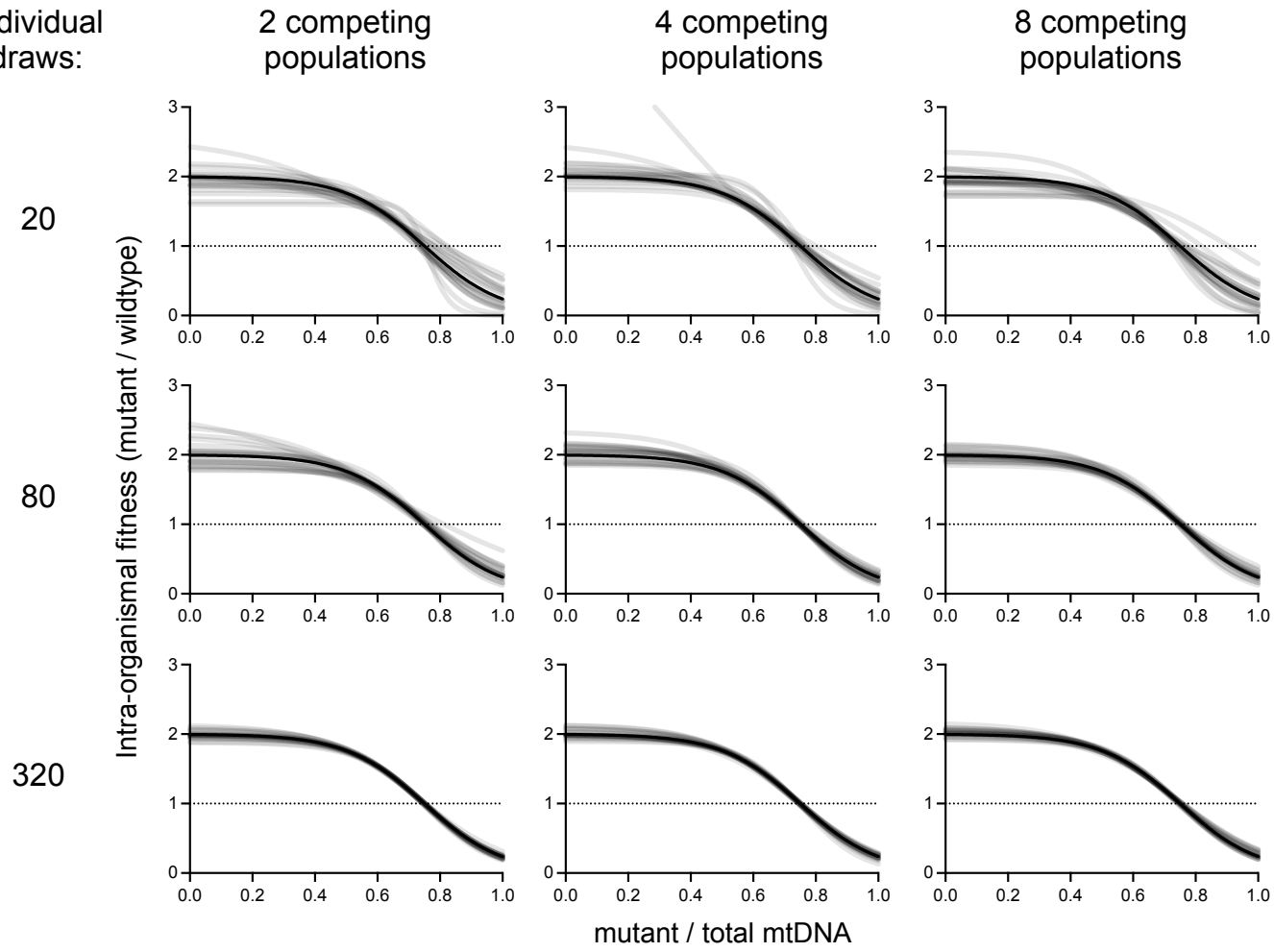
Supplementary Figures 1-9

Supplementary Tables 1-2



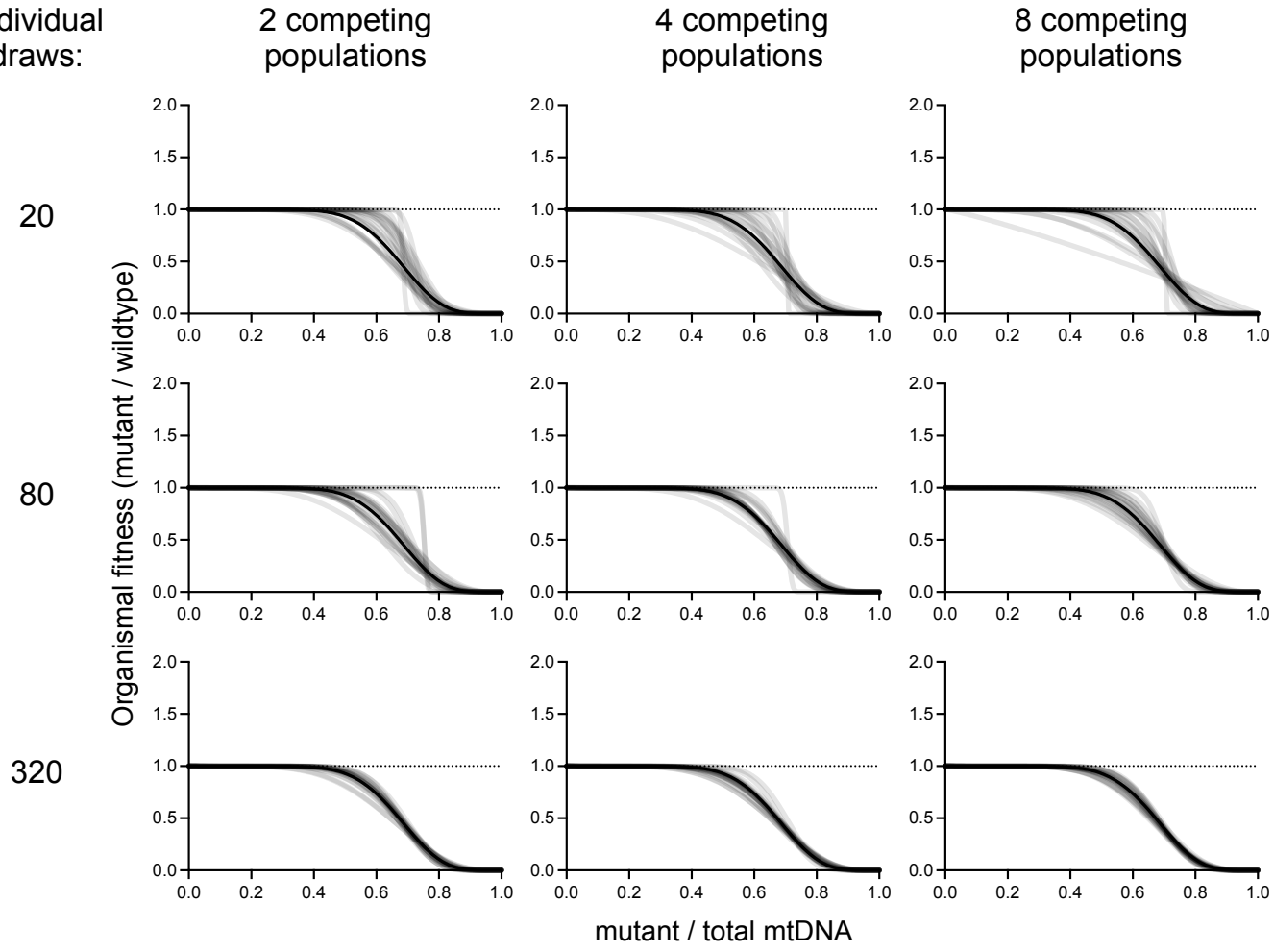
Supplementary Figure 1: Effects of variation in each model parameter specifying selection and drift.

a-c, intra-organismal fitness functions showing the result of varying parameters gamma (**a**), delta (**b**), and epsilon (**c**), each while holding the other parameters constant. **d**, mutant frequency variation among progeny resulting from a single parent at z=0.5, given by different values of the effective mtDNA bottleneck size N. **e-f**, organismal fitness functions showing the result of varying parameters alpha (**e**) and beta (**f**), each while holding the other parameter constant.



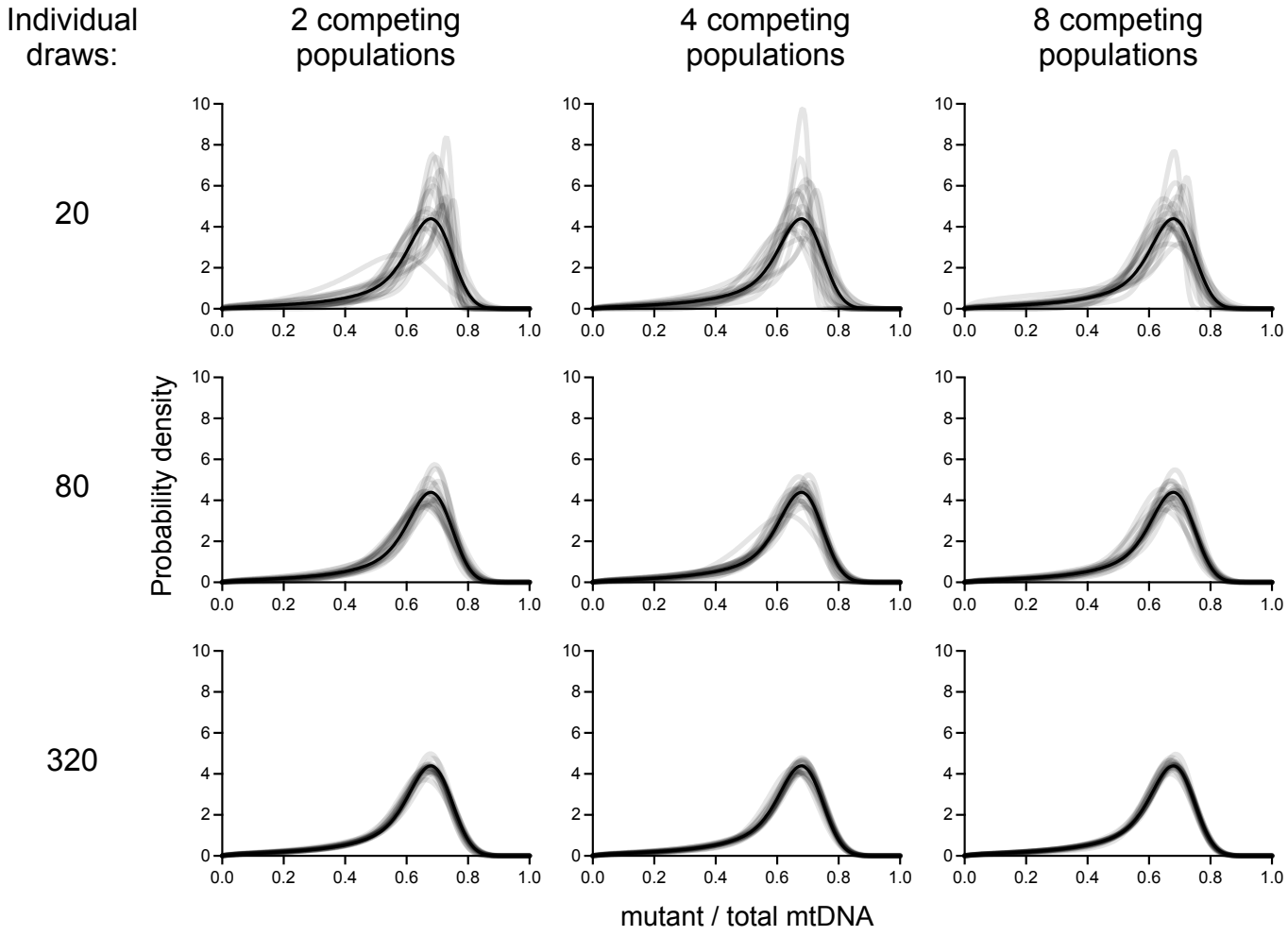
Supplementary Figure 2: Intra-organismal fitness functions for simulated data of different sample sizes.

Intra-organismal fitness functions obtained by simulating data of different sample sizes (n=20 simulations per plot) from a single set of unique input model parameters. The true fitness function for the model parameters is shown as a solid black line on each plot. Model parameters were re-estimated by joint maximum likelihood for each simulated data set, resulting in a unique combination of a new intra-organismal fitness function (gray lines), organismal fitness function (Supplementary Fig. 2), and most evolutionarily stable mutant frequency distribution (Supplementary Fig. 3). The data points drawn from the mutant frequency distribution vary by row (increasing from top to bottom). These are used as samples of the mutant frequency distribution and as the parents for modeling intra-organismal selection. The number of replicate populations, simulated for the purposes of modeling organismal selection, vary by column (increasing from left to right).



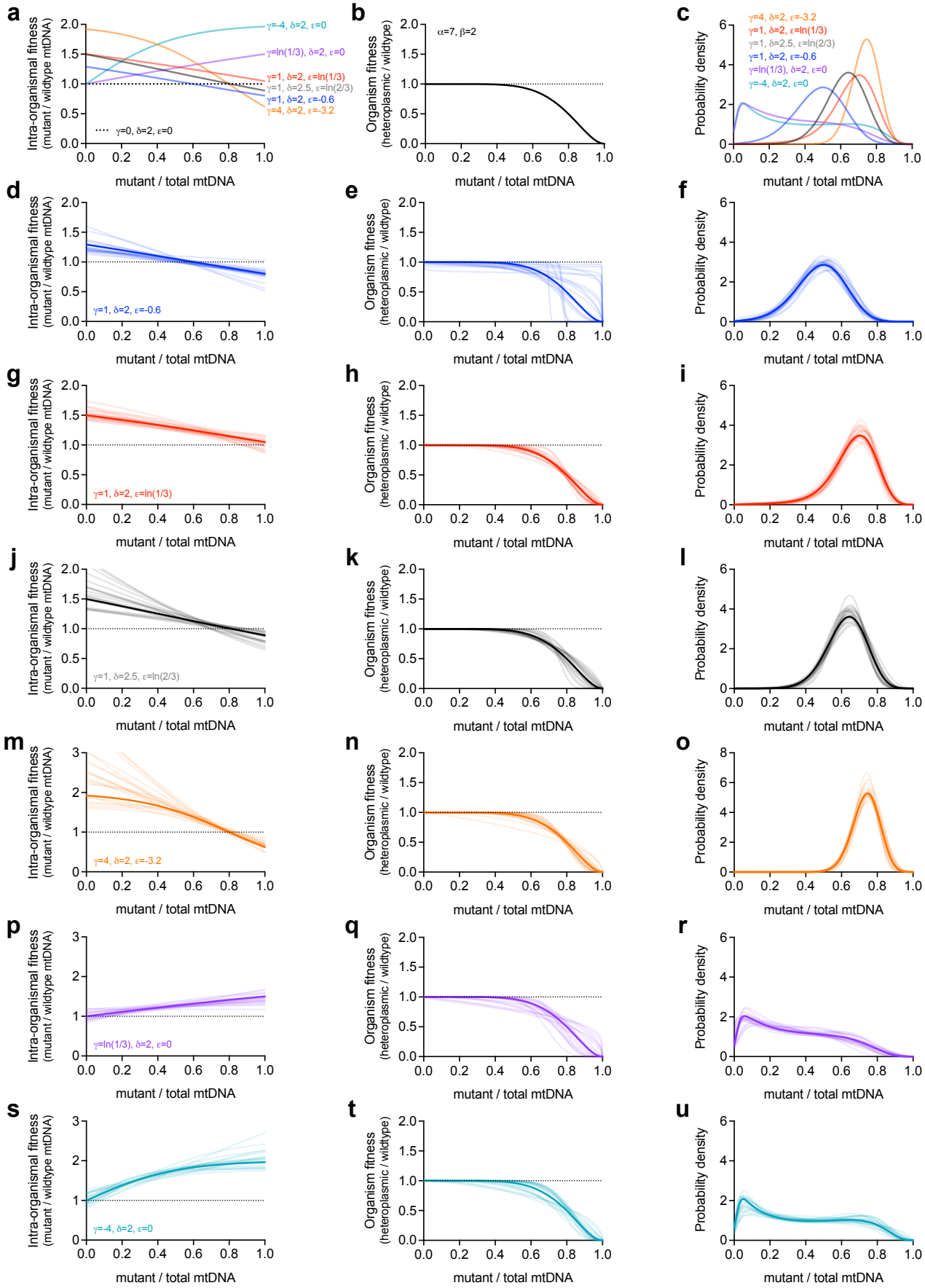
Supplementary Figure 3: Organismal fitness functions for simulated data of different sample sizes.

Organismal fitness functions obtained by simulating data of different sample sizes (n=20 simulations per plot) from the same set of unique input model parameters as used in Supplementary Fig. 1. The true fitness function for the model parameters is shown as a solid black line on each plot. Model parameters were re-estimated by joint maximum likelihood for each simulated data set, resulting in a unique combination of a new intra-organismal fitness function (Supplementary Fig. 1), organismal fitness function (gray lines), and most evolutionarily stable mutant frequency distribution (Supplementary Fig. 3). The data points drawn from the mutant frequency distribution vary by row (increasing from top to bottom). These are used as samples of the mutant frequency distribution and as the parents for modeling intra-organismal selection. The number of replicate populations, simulated for the purposes of modeling organismal selection, vary by column (increasing from left to right).



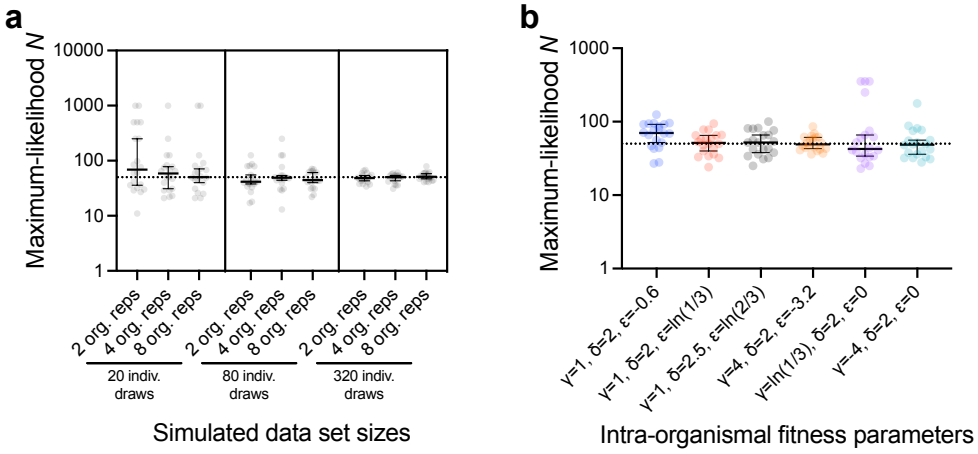
Supplementary Figure 4: Stationary distributions for simulated data of different sample sizes.

Stationary distributions obtained by simulating data of different sample sizes (n=20 simulations per plot) from the same set of unique input model parameters as used in Supplementary Fig. 4 and 5. The true fitness function for the model parameters is shown as a solid black line on each plot. Model parameters were re-estimated by joint maximum likelihood for each simulated data set, resulting in a unique combination of a new intra-organismal fitness function (Supplementary Fig. 1), organismal fitness function (Supplementary Fig. 2), and the most evolutionarily stable mutant frequency distribution (shown here in gray). The data points drawn from the mutant frequency distribution vary by row (increasing from top to bottom). These are used as samples of the mutant frequency distribution and as the parents for modeling intra-organismal selection. The number of replicate populations, simulated for the purposes of modeling organismal selection, vary by column (increasing from left to right).



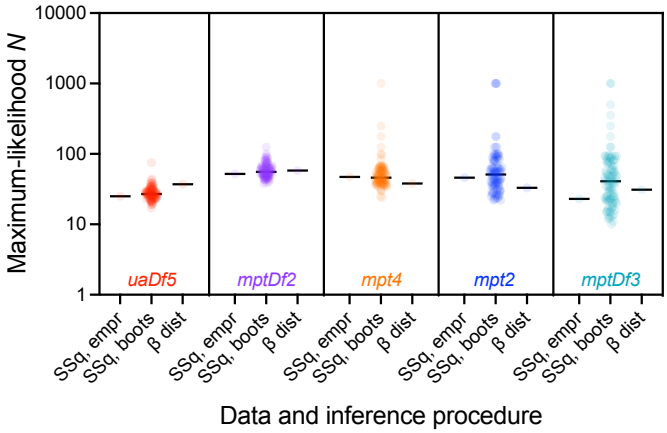
Supplementary Figure 5: Variation in intra-organismal selection dynamics and their effects on mutant frequency.

a-c, intra-organismal fitness functions corresponding to different mechanisms of selfish mitochondrial genome dynamics, including both positive and negative frequency-dependent selection (**a**), which combine with a given organismal fitness effect (**b**) to produce different stationary heteroplasmic frequency distributions (**c**). **d-u**, ground-truth functions (opaque lines) and maximum-likelihood functions for simulated data sets (semi-transparent lines, n=20) for each selection parameter set in **a** (*N*=50), with maximum-likelihood inferences initialized at α=1, β=1, γ=0, δ=2, ε=0 (corresponding to the line at neutral intra-organismal fitness).



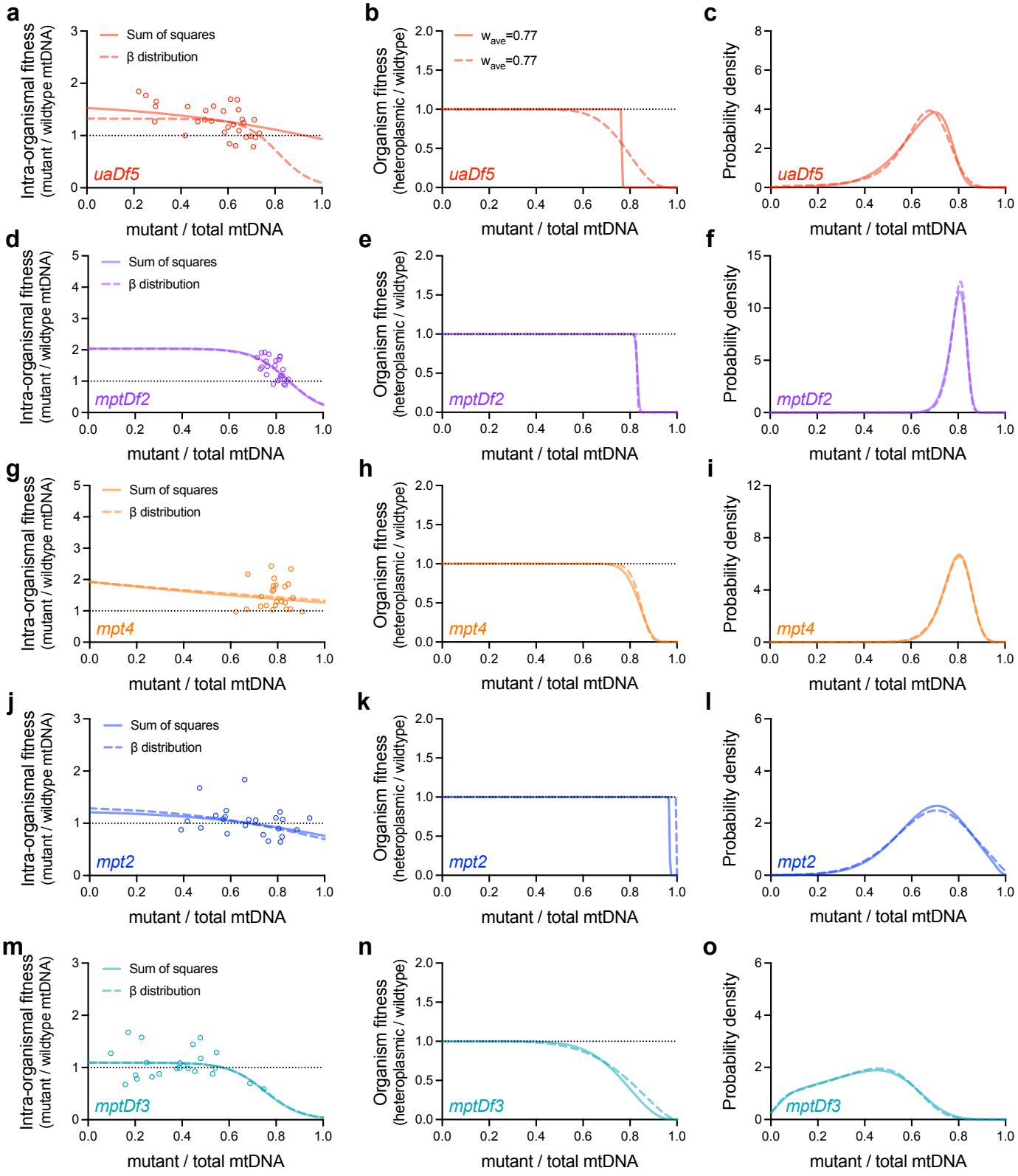
Supplementary Figure 6: Estimation of drift parameter *N* from simulated data.

a, maximum-likelihood estimates of *N* for the same simulated data sets represented in Supplementary Fig. 2-4. **b**, maximum-likelihood estimates of *N* for the same simulated data sets represented in Supplementary Fig. 5d-u. Dotted line represents the ground-truth (*N*=50) that was used to simulate the data. Error bars represent median and 95% confidence interval for 20 simulated data sets (represented by individual data points).



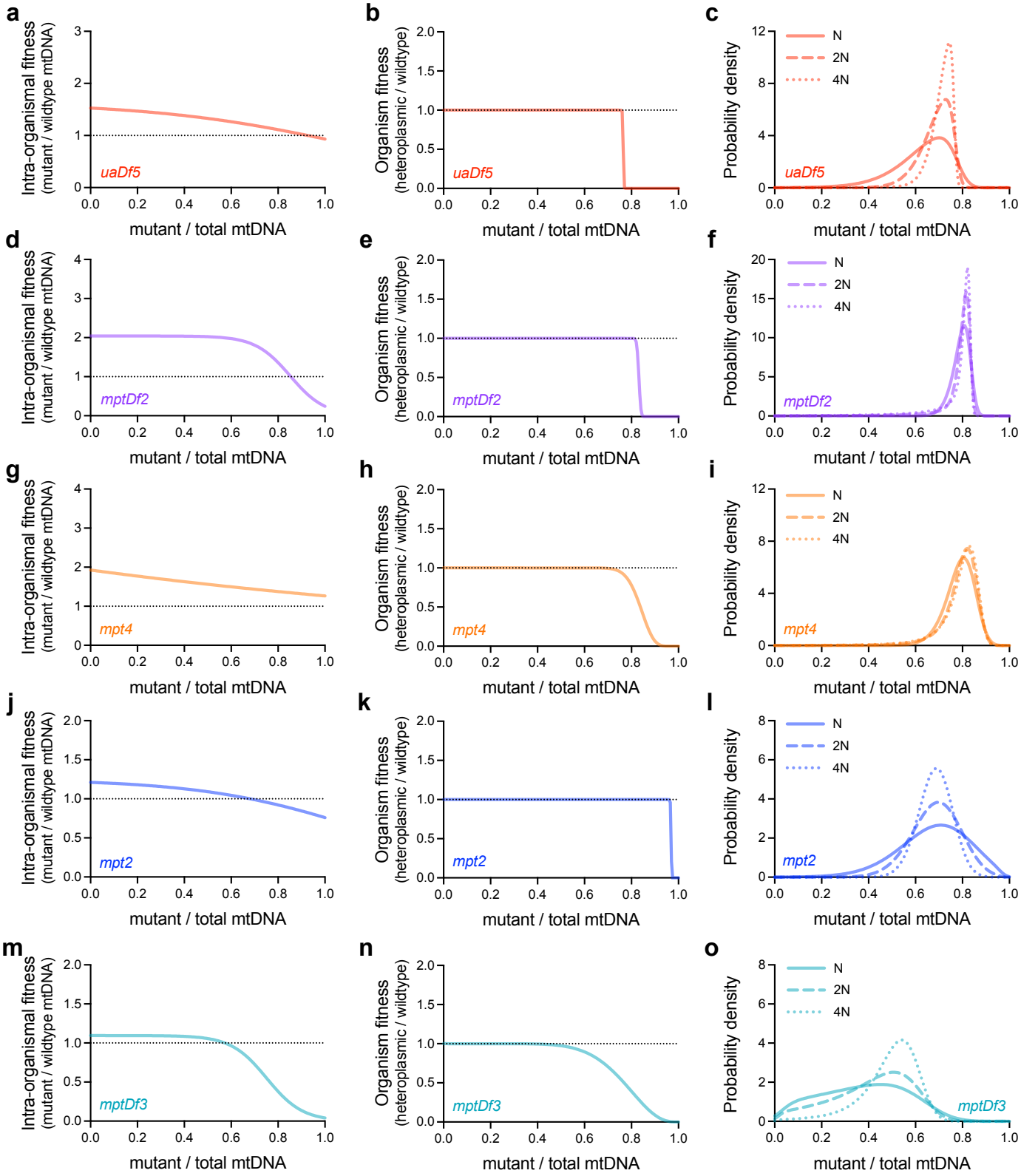
Supplementary Figure 7: Estimation of drift parameter *N* based on alternative methods for estimating the likelihood fo intra-organismal data.

Maximum-likelihood *N* was estimated for each of the five mitochondrial genotypes (Fig. 1). The likelihood for the intra-organismal data is estimated using the sum-of-squares method (*SSq, empr* and *SSq, boots* for empirical data and 100 parametric bootstrap data sets, respectively), and also using a beta distribution that is a function of *N* for the empirical data set (see Methods, equation 9). Apart from the approach to evaluate the likelihood with respect to the intra-organismal data, the maximum-likelihood procedure is left unchanged.



Supplementary Figure 8: Fitness functions and mutant frequency distributions based on alternative methods for estimating the likelihood fo intra-organismal data.

a-c, maximum-likelihood intra-organismal (**a**) and inter-organismal (**b**) fitness functions and the stable mutant frequency distribution (**c**) for *uaDf5*, each as a function of heteroplasmic mutant mtDNA frequency, similar to Fig. 3d-f but excluding bootstraps and instead comparing two methods of estimating the likelihood for the intra-organismal data: the sum-of-squares method (solid line) versus a beta distribution that is a function of *N* (dashed line), similar to the comparison for estimating *N* (Supplementary Fig. 7). Note that despite different slopes for the *uaDf5* organism-level fitness effects (**b**), the estimated mean fitness (0.77) is essentially identical under each model; these slopes are qualitatively similar to the apparent multimodal inter-organismal selection models in the *uaDf5* bootstraps (see Fig. 3e). **d-o**, similar to **a-c** but for *mptDf2* (**d-f**), *mpt4* (**g-i**), *mpt2* (**j-l**), and *mptDf3* (**m-o**). Apart from the approach to evaluate the likelihood with respect to the intra-organismal data (left column), the maximum-likelihood procedure is left unchanged. Individual data points (left column) show the mutant frequency and fitness of individual parent-progeny lineages in Fig. 3a (**a**) and in the left column of Fig. 4 (**d**,**g**,**j**,**m**).



Supplementary Figure 9: Effect of variation in magnitude of drift on the mutant mtDNA frequency distribution under the maximum-likelihood selection parameters.

a-c, maximum-likelihood intra-organismal (**a**) and inter-organismal (**b**) fitness functions and the stable mutant frequency distribution (**c**) for *uaDf5*, each as a function of heteroplasmic mutant mtDNA frequency, similar to Fig. 5 but excluding bootstraps and showing the stable mutant frequency distribution (**c**) under maximum-likelihood selection parameters and the drift parameter *N*, in addition to two- and four-fold changes in magnitude of drift (2*N* and 4*N*, respectively). **d-o**, similar to **a-c** but for *mptDf2* (**d-f**), *mpt4* (**g-i**), *mpt2* (**j-l**), and *mptDf3* (**m-o**).

Genotype	Generations until probability of invasion ≥0.5			
	N_e=10	N_e=100	N_e=1,000	N_e=10,000
uaDf5	791	79	8	1
	146–6.5*10^4	15–6,487	1–649	<1–65
mptDf2	1.2*10^9	1.2*10^8	1.2*10^7	1.2*10^6
	2,975–1.6*10^{21}	298–1.6*10^{20}	30–1.6*10^{19}	3–1.6*10^{18}
mpt4	7.0*10^6	7.0*10^5	7.0*10^4	7,016
	4,943–7.0*10^{21}	494–7.0*10^{20}	49–7.0*10^{19}	5–7.0*10^{18}
mpt2	9.4*10^4	9,444	944	94
	466–3.3*10^{10}	47-3.3*10^9	5–3.3*10^8	<1–3.3*10^7
mptDf3	24	2	<1	<1
	4–1,043	<1–104	<1–10	<1–1

Supplementary Table 1: Expected persistence times of the heteroplasmic state.

Persistence or robustness of the heteroplasmic state, defined by the expected number of generations until the invasion by a de novo homoplasmic lineage (introduced by the loss of the selfish genome through intra-organismal drift) becomes more probable than no invasion. Since the loss of the mutant genome is beneficial to host fitness, the probability of invasion is treated as the fixation probability for a beneficial genotype, described in ref. 48. Briefly, the per-generation probability of invasion is $1\text{-}e^{-2spNe}$, where the selection coefficient s is equal to w-1, where w is the relative fitness of homoplasmic organisms, $1/w_{org}$. Here, p is the per-generation probability of de novo introduction of a homoplasmic lineage (equal to the probability of mutant loss, provided in Table 1 column titled probability of loss), and Ne is the effective organismal population size. Since the per-generation probability of no invasion is equal to e^{-2spNe}, the persistence time as defined is thus equal to the minimum value of generations g such that $(e^{-2spNe})^g \leq 0.5$. Each entry: persistence time given maximum-likelihood model parameters (first line), in addition to 95% bootstrap confidence interval (second line).

Genotype	Average *z*	*z**
uaDf5	0.63	0.93
	0.60–0.65	>0.73
mptDf2	0.79	0.85
	0.78–0.80	>0.83
mpt4	0.79	>1
	0.76–0.82	>0.94
mpt2	0.67	0.68
	0.61–0.72	0.62–0.89
mptDf3	0.38	0.57
	0.32–0.46	0–0.75

Supplementary Table 2: Comparison of average heteroplasmic mutant mtDNA frequency and heteroplasmic frequency predicted by intra-organismal balancing selection (z*).

Each entry: maximum likelihood estimate (first line), 95% bootstrap confidence interval (second line).