Supporting Information

Constant rates yield unstable copy numbers for a model describing mtDNA genetic and network dynamics

We explored a simpler network system than the one presented in the Main Text, but found that it produced instability in mtDNA copy numbers, which we regard as biologically undesirable. Consider the following set of Poisson processes for singleton (s) and fused (f) species

$$s + s \xrightarrow{\gamma} f + f \tag{S1}$$
$$s + f \xrightarrow{\gamma} f + f \tag{S2}$$
$$f \xrightarrow{\beta} s \tag{S3}$$
$$s \xrightarrow{\alpha\rho} s + s \tag{S4}$$
$$f \xrightarrow{\alpha\rho} f + f \tag{S5}$$
$$s \xrightarrow{\eta\rho} \emptyset \tag{S6}$$
$$f \xrightarrow{\rho} \emptyset \tag{S7}$$

where Eq. (S1)-(S3) are analogous to Eq. (1)-(3) where mutant species are neglected. Eq. (S4) and (S5) are simple birth processes with a shared constant rate $\alpha\rho$. Eq. (S6) and (S7) are simple death processes with rates $\eta\rho$ and ρ respectively. The parameter ρ is shared amongst all of the birth and death reactions in Eqs. (S4)–(S7). ρ represents the intuitive assumption that, in order for a stable population size to exist, birth should balance death. However, for the network to have any effect at all, singletons should be at an increased risk of mitophagy relative to fused species. We represent the increased risk of singleton mitophagy with the parameter η. Since additional death is introduced into the system when $\eta > 1$, we include the parameter $\alpha > 1$ as an increased global biogenesis rate to balance the increased mitophagy of singletons.

We may write the above system as a set of ordinary differential equations

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\gamma s^2 - \gamma f s + \beta f + \alpha\rho s - \eta\rho s \tag{S8}$$
$$\frac{\mathrm{d}f}{\mathrm{d}t} = \gamma s^2 + \gamma f s - \beta f + \alpha\rho f - \rho f \tag{S9}$$

where we have enforced the stochastic reaction rate to be equivalent to the deterministic reaction rate, and hence the s^2 term is proportional to γ rather than 2γ (justification of this is presented below, see Eq. (S20)). In Figure S1 we see that the system displays a trivial steady state at $s = f = 0$ and a non-trivial steady state. Computing the eigenvalues of the Jacobian matrix at the non-trivial steady state indicates that it is a saddle node, and therefore unstable. Initial copy numbers which are too small tend towards extinction with time, and initial copy numbers which are too large tend towards a copy number explosion. This simple example suggests that a system of this form with constant reaction rates is unstable, and therefore biologically unlikely to exist under reasonable circumstances. We hence consider analogous biochemical reaction networks with a replication rate which is a function of state, to prevent extinction and divergence of the total population size.

Conversion of a chemical reaction network into ordinary differential equations

The following section outlines the steps in converting a set of chemical reactions into a set of ordinary differential equations (ODEs). In particular, we pay special attention to the fact that the rate of a chemical reaction with a stochastic treatment is not always equivalent to the rate in a deterministic treatment (Wilkinson 2011), as we will explain below. This subtlety is sometimes overlooked in the literature. This section draws on a number of standard texts (Gillespie 1976; Van Kampen 1992; Gillespie 2007; Wilkinson 2011) as well as (Grima 2010). We hope this harmonized treatment will be of help as a future reference.

Consider a general chemical system consisting of N distinct chemical species (X_i) interacting via R chemical reactions, where the j^{th} reaction is of the form

$$s_{1j}X_1 + \cdots + s_{Nj}X_N \xrightarrow{\hat{k}_j} r_{1j}X_1 + \cdots + r_{Nj}X_N \tag{S10}$$

where s_{ij} and r_{ij} are stoichiometric coefficients. We define $\hat{k}_j$ as the *microscopic* rate for this reaction. The dimensionality of this parameter will vary depending upon the stoichiometric coefficients s_{ij}. $\hat{k}_j$ may be loosely interpreted as setting the characteristic timescale (i.e. the cross section (Wilkinson 2011)) of reaction j.

The chemical master equation (CME) describes the dynamics of the joint distribution of the state of the system and time, moving forwards through time. Defining the state of the system as $\mathbf{x} = (x_1, \ldots, x_N)^T$, where x_i is the copy number of the i^{th} species, allows us to write the CME as (Grima 2010)

$$\frac{\partial P(\mathbf{x}, t|\mathbf{x}_0, t_0)}{\partial t} = \Omega \sum_{j=1}^{R} \left(\prod_{i=1}^{N} E_i^{-S_{ij}} - 1 \right) \hat{f}_j(\mathbf{x}, \Omega) P(\mathbf{x}, t|\mathbf{x}_0, t_0) \tag{S11}$$

where Ω is the volume of the compartment in which the reactions occur (also known as the system size), $S_{ij} = r_{ij} - s_{ij}$ is the stoichiometry matrix, and $E_i^{-S_{ij}}$ is referred to as the step operator and is defined through the relation $E_i^{-S_{ij}}(g(\mathbf{x})) = g(x_1, \ldots, x_i -$

$S_{ij}, \ldots, x_N)$, for any function of state $g(\mathbf{x})$. $\hat{f}_j(\mathbf{x}, \Omega)$ is the microscopic rate function of reaction j, which in general depends on both the state and the system size. A factor of Ω is explicitly included in this definition of the chemical master equation so that our treatment is compatible with Van Kampen's system size expansion (Van Kampen 1992). As a consequence of this, the probability that, given the current state $\mathbf{x}$, the j^{th} reaction occurs in the time interval $[t, t + \mathrm{d}t)$ somewhere in Ω (Gillespie 2007) is

$$\hat{a}_j(\mathbf{x}, \Omega)\mathrm{d}t := \Omega \hat{f}_j(\mathbf{x}, \Omega)\mathrm{d}t. \tag{S12}$$

$\hat{a}_j(\mathbf{x}, \Omega)$ is termed the propensity function (or "hazard") and is of particular relevance in the stochastic simulation algorithm (Gillespie 1976), since $\hat{a}_j(\mathbf{x}, \Omega) / \sum_j \hat{a}_j(\mathbf{x}, \Omega)$ determines the probability that the j^{th} reaction occurs next.

For the microscopic rate function, we may write

$$\hat{f}_j(\mathbf{x}, \Omega) = \hat{k}_j \prod_{i=1}^{N} \Omega^{-s_{ij}} \binom{x_i}{s_{ij}}. \tag{S13}$$

This equation counts the number of available combinations of reacting molecules (Gillespie 1976; Wilkinson 2011), whilst taking into account scaling with system size (Grima 2010).

We also introduce the deterministic rate equation (generally considered to be the macroscopic analogue of the CME) which is defined as (Van Kampen 1992; Grima 2010)

$$\frac{\mathrm{d}\phi_i}{\mathrm{d}t} = \sum_{j=1}^{R} S_{ij} \tilde{f}_j(\boldsymbol{\phi}) \tag{S14}$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)^T$ is the vector of macroscopic concentrations (of dimensions molecules per unit volume) and $\tilde{f}_j(\boldsymbol{\phi})$ is the macroscopic rate function satisfying

$$\tilde{f}_j(\boldsymbol{\phi}) = \tilde{k}_j \prod_{i=1}^{N} \phi_i^{s_{ij}} \tag{S15}$$

where $\tilde{k}_j$ is the *macroscopic* rate for the j^{th} reaction. We distinguish between $\hat{k}_j$ and $\tilde{k}_j$, respectively the rate constants for the discrete and continuous pictures, although this distinction is sometimes not emphasized in the literature (Grima 2010; Grima *et al.* 2011; Van Kampen 1992). The physical meaning of $\tilde{k}_j$ is not immediately obvious: we argue that this parameter only gains physical meaning through the following procedure.

As stated by Wilkinson (2011), if we intend for the microscopic description in Eq. (S11) to correspond to the macroscopic description in Eq. (S14), the rate of consumption/production of particles for every reaction must be the same in the deterministic limit of the stochastic system (the conditions for which we define below). Therefore, we apply the following constraint in the limit of large copy numbers

$$\lim_{x_i \to \infty} \hat{f}_j(\mathbf{x}, \Omega) = \tilde{f}_j(\boldsymbol{\phi}) \quad \forall\, i, j. \tag{S16}$$

In applying this constraint on all species i and all reactions j, we may derive a general relationship between $\hat{k}_j$ and $\tilde{k}_j$

$$\lim_{x_i \to \infty} \hat{k}_j \prod_{i=1}^{N} \Omega^{-s_{ij}} \frac{x_i!}{s_{ij}!(x_i - s_{ij})!} = \tilde{k}_j \prod_{i=1}^{N} \phi_i^{s_{ij}}. \tag{S17}$$

We can make two approximations to generate a more convenient relationship between the microscopic and macroscopic rates. Firstly, we assume that

$$x_i \approx \Omega \phi_i. \tag{S18}$$

This is a small noise approximation, since it is often assumed that $x_i = \Omega\phi_i + \Omega^{1/2}\xi_i$, where ξ_i is a noise term (Van Kampen 1992). If ξ_i is small then $x_i \approx \Omega\phi_i$ is a valid approximation. Secondly, we assume that

$$x_i(x_i - 1) \ldots (x_i - s_{ij} + 1) \approx x_i^{s_{ij}}. \tag{S19}$$

This is a large copy number approximation: in the case of e.g. a bimolecular reaction ($2X_i \to *$) with $s_{ij} = 2$, the approximation is of the form $x_i(x_i - 1) \approx x_i^2$ or $x_i \approx x_i - 1$. By applying Eq. (S19) to the factor of $x_i!$ in Eq. (S17), the factor of $(x_i - s_{ij})!$ cancels from the left-hand side. Simplifying using Eq. (S18), $\phi_i^{s_{ij}}$ cancels from both sides and we arrive at the important relationship

$$\tilde{k}_j \approx \hat{k}_j \prod_{i=1}^{N} \frac{1}{s_{ij}!}. \tag{S20}$$

With Eq. (S14), Eq. (S15) and Eq. (S20) one may therefore write down a set of ODEs for an arbitrary chemical reaction network, with constant reaction rates, in terms of the microscopic rates $\hat{k}_j$. This equation highlights that for reactions with $s_{ij} \geq 2$, $\tilde{k}_j \neq \hat{k}_j$, as is the case for bimolecular reactions of the form $2X_i \to *$ (see Eq. (1) and Eq. (S1)).

Importantly, if the microscopic rate function is a function of state then $\hat{k} = \hat{k}(\mathbf{x})$ and $\tilde{k} = \tilde{k}(\boldsymbol{\phi}) \approx \tilde{k}(\mathbf{x}/\Omega)$. In this case, Eq. (S20) still applies since the above argument assumed nothing about the particular forms of $\hat{k}$ and $\tilde{k}$. However, additional factors of Ω^{-1} are induced by applying Eq. (S18), which may carry through to the individual parameters of $\tilde{k}(\boldsymbol{\phi})$. A demonstration of this is given in the following section.

Deriving an ODE description of the mitochondrial network system

In this section we show how to derive an ODE description of the network system described in Eq. (1)-Eq. (9) in the Main Text. In accordance with the notation in the previous section, we will redefine all of the rates in Eq. (1)-Eq. (10) with a hat notation ($\hat{a}$, for a general rate parameter a), to reflect that these are stochastic rates. Deterministic rates will be denoted with a tilde ($\tilde{a}$). Our aim will be to write a set of ODEs in terms of the stochastic rates, $\hat{a}$, for which we are able to estimate values.

We will begin by considering the fusion network equations Eq. (1) and Eq. (2). For clarity, we rewrite Eq. (1) to allow the reaction to proceed with some arbitrary rate $\hat{\rho}$:

$$X_S + X_S \xrightarrow{\hat{\rho}} X_F + X_F, \tag{S21}$$

where X denotes either a wild-type (W) or mutant (M). We will subsequently fix $\hat{\rho}$ to the rate of all other fusion reactions $\hat{\gamma}$. We do this because Eq. (1) is a bimolecular reaction involving one species: a fundamentally different reaction to bimolecular reactions involving two species, as we will now see.

Since $\hat{\rho}, \hat{\gamma} = \text{const}$, we may use Eq. (S20), resulting in the deterministic rates

$$\tilde{\rho} = \frac{\hat{\rho}}{2} \tag{S22}$$

$$\tilde{\gamma} = \hat{\gamma} \tag{S23}$$

for Eq. (S21) and Eq. (2) respectively. If we then enforce the microscopic rates to be equal for both of these fusion reactions, i.e. $\hat{\rho} = \hat{\gamma}$, then $\tilde{\rho} = \hat{\gamma}/2$. All other fusion reactions have $\tilde{\gamma} = \hat{\gamma}$ by application of Eq. (S20). Application of Eq. (S20) to the fission reaction in Eq. (3) shows that $\hat{\beta} = \tilde{\beta}$.

For Eq. (4), we have chosen a $\hat{\lambda}$ which is not a constant, but a function of the copy numbers of the chemical species ($\hat{\lambda} = \hat{\lambda}(\mathbf{x})$ where $\mathbf{x} = (w_s, w_f, m_s, m_f)$, see Eq. (10)). As pointed out in the previous section, care must be taken in writing down the deterministic analogue of $\hat{\lambda}$. Applying Eq. (S20), we have

$$\hat{\mu} + \hat{b}(\hat{\kappa} - (w_s + w_f + \hat{\delta} m_s + \hat{\delta} m_f)) = \tilde{\mu} + \tilde{b}(\tilde{\kappa} - (\phi_{w_s} + \phi_{w_f} + \tilde{\delta}\phi_{m_s} + \tilde{\delta}\phi_{m_f})). \tag{S24}$$

Applying Eq. (S18) and equating individual terms, we arrive at

$$\hat{\mu} = \tilde{\mu} \tag{S25}$$

$$\hat{b} w_s = \tilde{b}\frac{w_s}{\Omega} \implies \hat{b} = \tilde{b}/\Omega \tag{S26}$$

$$\hat{b}\hat{\kappa} = \tilde{b}\tilde{\kappa} \implies \hat{\kappa} = \tilde{\kappa}\Omega \tag{S27}$$

$$\hat{b}\hat{\delta} m_s = \tilde{b}\tilde{\delta}\phi_{m_s} \implies \hat{\delta} = \tilde{\delta}. \tag{S28}$$

In this study, we let $\Omega = 1$ so the above 4 parameters are identical to their deterministic counterparts. Hence, by application of Eq. (S14), we arrive at the following set of ODEs

$$\frac{\mathrm{d}\phi_{w_s}}{\mathrm{d}t} = -2\cdot\frac{\hat{\gamma}}{2}\phi_{w_s}^2 - \hat{\gamma}\phi_{w_s}\phi_{w_f} + \hat{\beta}\phi_{w_f} - (\hat{\mu} + \hat{b}(\hat{\kappa} - (\phi_{w_s} + \phi_{w_f} + \hat{\delta}\phi_{m_s} + \hat{\delta}\phi_{m_f})))\phi_{w_s} - \hat{\mu}\phi_{w_s} - \hat{\gamma}\phi_{m_f}\phi_{w_s} - \hat{\gamma}\phi_{w_s}\phi_{m_s} \tag{S29}$$

$$\frac{\mathrm{d}\phi_{m_s}}{\mathrm{d}t} = -2\cdot\frac{\hat{\gamma}}{2}\phi_{m_s}^2 - \hat{\gamma}\phi_{m_s}\phi_{m_f} + \hat{\beta}\phi_{m_f} - (\hat{\mu} + \hat{b}(\hat{\kappa} - (\phi_{w_s} + \phi_{w_f} + \hat{\delta}\phi_{m_s} + \hat{\delta}\phi_{m_f})))\phi_{m_s} - \hat{\mu}\phi_{m_s} - \hat{\gamma}\phi_{w_f}\phi_{m_s} - \hat{\gamma}\phi_{w_s}\phi_{m_s} \tag{S30}$$

$$\frac{\mathrm{d}\phi_{w_f}}{\mathrm{d}t} = 2\cdot\frac{\hat{\gamma}}{2}\phi_{w_s}^2 + \hat{\gamma}\phi_{w_s}\phi_{w_f} - \hat{\beta}\phi_{w_f} + (\hat{\mu} + \hat{b}(\hat{\kappa} - (\phi_{w_s} + \phi_{w_f} + \hat{\delta}\phi_{m_s} + \hat{\delta}\phi_{m_f})))(2\phi_{w_s} + \phi_{w_f}) + \hat{\gamma}\phi_{m_f}\phi_{w_s} + \hat{\gamma}\phi_{w_s}\phi_{m_s} \tag{S31}$$

$$\frac{\mathrm{d}\phi_{m_f}}{\mathrm{d}t} = 2\cdot\frac{\hat{\gamma}}{2}\phi_{m_s}^2 + \hat{\gamma}\phi_{m_s}\phi_{m_f} - \hat{\beta}\phi_{m_f} + (\hat{\mu} + \hat{b}(\hat{\kappa} - (\phi_{w_s} + \phi_{w_f} + \hat{\delta}\phi_{m_s} + \hat{\delta}\phi_{m_f})))(2\phi_{m_s} + \phi_{m_f}) + \hat{\gamma}\phi_{w_f}\phi_{m_s} + \hat{\gamma}\phi_{w_s}\phi_{m_s}. \tag{S32}$$

The steady state solution of this system of ODEs may be calculated, but its form is complex. For notational simplicity, we will drop the hat notation. Defining

$$x_1 = (b^2(\beta^2 + 2\beta(\gamma\kappa + 3\mu) + \gamma^2\kappa^2 + \mu^2 + 2\gamma\mu(\kappa + 2(\delta - 1)\phi_{m_s})) + 2b\gamma\mu(-\beta + \mu + \gamma(\kappa - 2\delta\phi_{m_s} + 2\phi_{m_s})) + \gamma^2\mu^2)^{1/2} \tag{S33}$$

the non-trivial, physically-realizable, component of the steady state may be parametrized in terms of ϕ_{m_s} and written as

$$\phi_{w_s} = -(\beta b^2\kappa + b^2\gamma\kappa^2 + b^2\kappa\mu + 2b^2\delta\mu\phi_{m_s} + \beta^2 b + \beta b\gamma\kappa + 5\beta b\mu + 3b\gamma\kappa\mu - \beta\gamma\mu + 2b\mu^2 - 2b\gamma\delta\mu\phi_{m_s} - 2b\gamma\mu\phi_{m_s} - x_1(b\kappa + \beta + 2\mu) + 2\gamma\mu^2 + 2\gamma^2\mu\phi_{m_s})/(2\mu(b - \gamma)^2) \tag{S34}$$

$$\phi_{w_f} = (b^3(\beta^2\kappa + \beta(\gamma\kappa(\kappa + (\delta-1)\phi_{m_s}) + \mu(3\kappa + \delta\phi_{m_s})) + \phi_{m_s}(\gamma^2(\delta-1)\kappa^2 - \delta\mu^2 + \gamma\mu((2\delta-3)\kappa + 2(\delta-1)\delta\phi_{m_s}))) + b^2(\beta^3 + \beta^2(6\mu + \gamma$$
$$(\kappa + (\delta-1)\phi_{m_s})) + \beta(5\mu^2 + \gamma^2\delta\kappa\phi_{m_s} + \gamma^2\kappa(-\phi_{m_s}) + 6\gamma\delta\mu\phi_{m_s} - 8\gamma\mu\phi_{m_s} - \kappa x_1) +$$
$$\phi_{m_s}(2\gamma^2\mu(\delta\kappa + \delta^2(-\phi_{m_s}) + \phi_{m_s}) + \gamma((6\delta-5)\mu^2 - (\delta-1)\kappa x_1) - \delta\mu x_1)) - b(\beta(\gamma^2\mu(-(\kappa + (4-3$$
$$\delta)\phi_{m_s})) + \gamma(2\mu^2 + (\delta-1)\phi_{m_s}x_1) + 3\mu x_1) + \gamma\mu\phi_{m_s}(\gamma(\delta-2)\mu + \gamma^2(\kappa - 2(\delta-1)\phi_{m_s}) + (\delta-3)x_1) +$$
$$\beta^2(2\gamma\mu + x_1)) + \gamma\mu(\beta - \gamma\phi_{m_s})(\gamma\mu + x_1))/(2b\mu(b-\gamma)^2(\beta + \gamma(\delta-1)\phi_{m_s})) \tag{S35}$$

$$\phi_{m_f} = (\phi_{m_s}(-b\beta + b\gamma\kappa + b\mu - 2b\gamma\delta\phi_{m_s} + 2b\gamma\phi_{m_s} + \gamma\mu + x_1))/(2b(\beta + \gamma(\delta-1)\phi_{m_s})). \tag{S36}$$

Since the steady state is parametrized by ϕ_{m_s}, the steady state is therefore a line.

Proof of heteroplasmy relation for linear feedback control

In this section we show that Eq. (13) holds for the system described by Eq. (1)-Eq. (9) given the replication rate in Eq. (10) using the Kramers-Moyal expansion under conditions of large copy number and fast network churn (to be defined below); the approach used here is similar to Constable *et al.* (2016). Here, we draw together elements from the literature to provide a coherent derivation; we therefore hope that the following exposition may provide clarity for a wider audience.

Kramers-Moyal expansion of the chemical master equation for large copy numbers Customarily, the Kramers-Moyal expansion is formed using a continuous-space notation (Gardiner 1985), so we will initially proceed in this way. Following the treatment by Gardiner (1985), we begin by re-writing the chemical master equation Eq. (S11) (CME) as

$$\frac{\partial P(\mathbf{x},t)}{\partial t} = \int_{-\infty}^{\infty} \mathrm{d}\mathbf{x}' \left[T(\mathbf{x}|\mathbf{x}')P(\mathbf{x}',t) - T(\mathbf{x}'|\mathbf{x})P(\mathbf{x},t)\right] \tag{S37}$$

where we have set $\Omega = 1$. $T(\mathbf{x}|\mathbf{x}')$ is the transition rate from state $\mathbf{x}' \to \mathbf{x}$, and the dependence upon the initial condition has been suppressed for notational convenience. We now proceed by expanding the CME. The multivariate Kramers-Moyal expansion may be written as

$$\frac{\partial P(\mathbf{x},t)}{\partial t} \approx \int_{-\infty}^{\infty} \left(-\nabla\left(T(\mathbf{x}'|\mathbf{x})P(\mathbf{x})\right)^T \cdot (\mathbf{x}'-\mathbf{x}) + \frac{1}{2}(\mathbf{x}'-\mathbf{x})^T \cdot \mathbf{H} \cdot (\mathbf{x}'-\mathbf{x})\right) \mathrm{d}\mathbf{x}' \tag{S38}$$

where $\mathbf{H}(\mathbf{x})$ is the Hessian matrix of $T(\mathbf{x}'|\mathbf{x})P(\mathbf{x})$

$$\mathbf{H} := \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} & \cdots & \frac{\partial^2}{\partial x_1 \partial x_N} \\ \vdots & & \vdots \\ \frac{\partial^2}{\partial x_N \partial x_1} & \cdots & \frac{\partial^2}{\partial x_N^2} \end{pmatrix} T(\mathbf{x}'|\mathbf{x})P(\mathbf{x}) \tag{S39}$$

(see (Gardiner 1985) for a proof of this in the univariate case).

A transition to each possible neighbouring state $\mathbf{x}'$ corresponds to some reaction j which moves the state from $\mathbf{x} \to \mathbf{x}'$. Since we know the influence of each reaction on state $\mathbf{x}$ through the constant stoichiometry matrix S_{ij}, and that the propensity of a reaction does not depend upon $\mathbf{x}'$ itself (see Eq. (S13)), we may transition from a notation involving $\mathbf{x}$ and $\mathbf{x}'$ into a notation involving $\mathbf{x}$ and j. We may therefore define $T_j(\mathbf{x}) := T(\mathbf{x}'|\mathbf{x}) \equiv \hat{f}_j(\mathbf{x})$ (see Eq. (S13)), and let $\mathbf{H}(\mathbf{x}) \to \mathbf{H}_j(\mathbf{x})$.

We now make a large copy number assumption in order to simplify $T_j(\mathbf{x})$. To take a large copy number limit, we assume that $x_i! \approx x_i^{s_{ij}}(x_i - s_{ij})!$ resulting in

$$T_j(\mathbf{x}) \approx \hat{k}_j \prod_{i=1}^{N} \frac{x_i^{s_{ij}}}{s_{ij}!}. \tag{S40}$$

This approximation is exact when $s_{ij} = 0, 1$, but inexact when $s_{ij} \geq 2$. For example, if we consider the second-order bimolecular reaction in Eq. (1), Eq. (S40) is equivalent to assuming $w_s^2 \approx w_s(w_s - 1)$; consequently, a factor of $1/(s_{ij}!) = 1/2$ arises in $T_j(\mathbf{x})$ as a combinatorial factor from stochastic considerations.

Fokker-Planck equation for chemical reaction networks We now wish to re-write Eq. (S38) as a Fokker-Planck equation. Since the integral in Eq. (S38) is over $\mathbf{x}'$, and every $\mathbf{x}'$ corresponds to a reaction j, we may interpret the integral in Eq. (S38) as a sum over all reactions, i.e. $\int \mathrm{d}\mathbf{x}' \to \sum_{j=1}^{R}$. Hence, for the j^{th} reaction, $[(\mathbf{x}' - \mathbf{x})]_i = S_{ij}$. With these observations, we may write the first integral of Eq. (S38) as

$$\begin{aligned} \int_{-\infty}^{\infty} -\nabla(T(\mathbf{x}'|\mathbf{x})P(\mathbf{x}))^T \cdot (\mathbf{x}'-\mathbf{x})\,\mathrm{d}\mathbf{x}' &= \int_{-\infty}^{\infty} -\nabla(T_j(\mathbf{x})P(\mathbf{x},t))^T \cdot (\mathbf{x}'-\mathbf{x})\,\mathrm{d}\mathbf{x}' \\ &= -\sum_{j=1}^{R}\sum_{i=1}^{N} \frac{\partial}{\partial x_i}(T_j(\mathbf{x})P(\mathbf{x},t))S_{ij} \\ &= -\sum_{i=1}^{N} \frac{\partial}{\partial x_i} A_i P(\mathbf{x},t) \end{aligned} \tag{S41}$$

where

$$\mathbf{A} := \mathbf{S} \cdot \mathbf{T}. \tag{S42}$$

$\mathbf{A}$ is a vector of length N, $[\mathbf{S}]_{ij} := r_{ij} - s_{ij}$ is the $N \times R$ stoichiometry matrix Eq. (S10), and $\mathbf{T}$ is the vector of transition rates, of length R (for which we have taken a large copy number approximation in Eq. (S40)). To re-write the second integral of Eq. (S38), we write an element of the Hessian $\mathbf{H}_j$ in Eq. (S39) as

$$H_{jlm} = \frac{\partial^2}{\partial x_l \partial x_m} T_j(\mathbf{x}) P(\mathbf{x}, t) \tag{S43}$$

where $j = 1, \ldots, R$ and $l, m = 1, \ldots, N$. Thus, we may write

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{1}{2} (\mathbf{x}' - \mathbf{x})^T \cdot \mathbf{H}_j \cdot (\mathbf{x}' - \mathbf{x}) \, \mathrm{d}\mathbf{x}' &= \frac{1}{2} \sum_{j=1}^{R} \sum_{l=1}^{N} \sum_{m=1}^{N} S_{lj} H_{jlm} S_{mj} \\
&= \frac{1}{2} \sum_{j=1}^{R} \sum_{l=1}^{N} \sum_{m=1}^{N} S_{lj} \frac{\partial^2}{\partial x_l \partial x_m} T_j P(\mathbf{x}, t) S_{mj} \\
&= \frac{1}{2} \sum_{l=1}^{N} \sum_{m=1}^{N} \frac{\partial^2}{\partial x_l \partial x_m} \left(\sum_{j=1}^{R} S_{lj} T_j S_{mj} \right) P(\mathbf{x}, t) \\
&= \frac{1}{2} \sum_{i,m=1}^{N} \frac{\partial^2}{\partial x_i \partial x_m} B_{im} P(\mathbf{x}, t)
\end{aligned} \tag{S44}$$

where

$$\mathbf{B} := \mathbf{S} \cdot \mathrm{Diag}(\mathbf{T}) \cdot \mathbf{S}^T. \tag{S45}$$

$\mathbf{B}$ is an $N \times N$ matrix, and $\mathrm{Diag}(\mathbf{Y})$ is a diagonal matrix whose main diagonal is the vector $\mathbf{Y}$. We may therefore re-write Eq. (S38) as a Fokker-Planck equation for the state vector $\mathbf{x}$ of the form

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} \approx -\sum_{i=1}^{N} \frac{\partial}{\partial x_i} [A_i(\mathbf{x}) P(\mathbf{x}, t)] + \frac{1}{2} \sum_{i,m=1}^{N} \frac{\partial^2}{\partial x_i \partial x_m} [B_{im}(\mathbf{x}) P(\mathbf{x})]. \tag{S46}$$

Fokker-Planck equation for an arbitrary function of state We now wish to make a change of variables in Eq. (S46) to write down a Fokker-Planck equation for an arbitrary scalar function of state $\mathbf{x}$ (which we will later set to be heteroplasmy). To do this, we wish to make use of Itô's formula, which allows a change of variables for an SDE. In general, the Fokker-Planck equation in Eq. (S46) is equivalent (Jacobs 2010) to the following Itô stochastic differential equation (SDE)

$$\mathrm{d}\mathbf{x} = \mathbf{A} \, \mathrm{d}t + \mathbf{G} \, \mathrm{d}\mathbf{W} \tag{S47}$$

where $\mathbf{G}\mathbf{G}^T \equiv \mathbf{B}$ (where $\mathbf{G}$ is an $N \times R$ matrix) and $\mathrm{d}\mathbf{W}$ is a vector of independent Wiener increments of length R, and a Wiener increment $\mathrm{d}W$ satisfies

$$\int_0^t \mathrm{d}W := W(t), \; P(W, t) \equiv \frac{1}{\sqrt{2\pi t}} e^{-W^2/(2t)}. \tag{S48}$$

Itô's formula states that, for an arbitrary function $h(\mathbf{x}, t)$ where $\mathbf{x}$ satisfies Eq. (S47), we may write the following SDE

$$\mathrm{d}h(\mathbf{x}, t) = \left\{ \frac{\partial h}{\partial t} + (\nabla h)^T \mathbf{A} + \frac{1}{2} \mathrm{Tr} \left[\mathbf{G}^T \mathbf{H}_h(\mathbf{x}) \mathbf{G} \right] \right\} \mathrm{d}t + (\nabla h)^T \mathbf{G} \, \mathrm{d}\mathbf{W}, \tag{S49}$$

where $\mathbf{H}_h(\mathbf{x})$ is the Hessian matrix of $h(\mathbf{x}, t)$ (see Eq. (S39), where $T(\mathbf{x}'|\mathbf{x})P(\mathbf{x})$ should be replaced with $h(\mathbf{x}, t)$). Given the form of $\mathbf{B}$ in Eq. (S45) we let

$$\mathbf{G} = \mathbf{S} \cdot \mathrm{Diag}(\sqrt{\mathbf{T}}), \tag{S50}$$

which satisfies $\mathbf{G}\mathbf{G}^T \equiv \mathbf{B}$.

For convenience, we may also perform the transformation purely at the level of Fokker-Planck equations. Let $h(\mathbf{x}, t)$ satisfy the general Fokker-Planck equation

$$\frac{\partial P(h, t)}{\partial t} = -\frac{\partial}{\partial h} [\tilde{A}(h, t) P(h, t)] + \frac{1}{2} \frac{\partial^2}{\partial h^2} [\tilde{B}(h, t) P(h, t)] \tag{S51}$$

for scalar functions $\tilde{A}(h, t)$ and $\tilde{B}(h, t)$. Using the cyclic property of the trace in Eq. (S49), we may identify

$$\tilde{A} = \frac{\partial h}{\partial t} + (\nabla h)^T \mathbf{A} + \frac{1}{2} \mathrm{Tr} \left[\mathbf{B} H_h(\mathbf{x}) \right] \tag{S52}$$

where Tr is the trace operator. Also, from Eq. (S49),

$$\tilde{B} = [(\nabla h)^T \mathbf{G}][(\nabla h)^T \mathbf{G}]^T = (\nabla h)^T \mathbf{B} (\nabla h). \tag{S53}$$

Hence, using Eq. (S51), Eq. (S52) and Eq. (S53), we may write down a Fokker-Planck equation for an arbitrary function of state in terms of $\mathbf{A}$ and $\mathbf{B}$.

An SDE for heteroplasmy forced onto the steady state line in the high-churn limit It has been demonstrated that SDE descriptions of stochastic systems which possess a globally-attracting line of steady states may be formed in the long-time limit by forcing the state variables onto the steady state line (Constable *et al.* 2016; Parsons and Rogers 2017). Such descriptions may be formed in terms of a parameter which traces out the position on the steady state line, hence reducing a high-dimensional problem into a single dimension (Constable *et al.* 2016; Parsons and Rogers 2017). In our case, heteroplasmy is a suitable parameter to trace out the position on the steady state line. We seek to use similar reasoning to verify Eq. (13). In what follows, we will assume that $\mathbf{x}(t=0) = \mathbf{x}_{\text{ss}}$, where $\mathbf{x}_{\text{ss}}$ is the state which is the solution of $\mathbf{A} = \mathbf{0}$ (which is equivalent to finding the steady state solution of the deterministic rate equation in Eq. (S14) due to our assumption of large copy numbers and $\Omega = 1$), so that we may neglect any deterministic transient dynamics.

Inspection of the steady state of the ODE description of our system reveals that the set of steady state solutions forms a line (see Eqs. (S34)–(S36)). Inspection of the steady state solution reveals that the steady state depends on the fusion (γ) and fission (β) rates. Mitochondrial network dynamics occur on a much faster timescale than the replication and degradation of mtDNA: the former occurring on the timescale of minutes (Twig *et al.* 2008) whereas the latter is hours or days (Johnston and Jones 2016). We seek to use this separation of timescales to arrive at a simple form for $\mathbb{V}(h)$. We redefine the fusion and fission rates such that

$$\begin{aligned} \gamma &\to M\gamma \\ \beta &\to M\beta \end{aligned} \tag{S54}$$

where M is a constant which determines the magnitude of the fusion and fission rates, which we call the "network churn".

We now wish to use heteroplasmy

$$h(\mathbf{x}, t) = h(\mathbf{x}) := (m_s + m_f)/(w_s + w_f + m_s + m_f), \tag{S55}$$

as our choice for the function of state in the Fokker-Planck equation in Eq. (S51). We will first compute the diffusion term $\tilde{B}$ for heteroplasmy using Eq. (S53). If we constrain the state $\mathbf{x}$ to be forced onto the steady state line $\mathbf{x}_{\text{ss}}$ (as per (Constable *et al.* 2016; Parsons and Rogers 2017)) in the high-churn limit, then upon defining

$$\theta := \sqrt{b^2(\beta + \gamma\kappa)^2 - 2b\gamma\mu\,(\beta - \gamma\kappa + 2\gamma(\delta - 1)m_s) + \gamma^2\mu^2} \tag{S56}$$

we have

$$\begin{aligned} \lim_{M\to\infty}\left(\tilde{B}\big|_{\mathbf{x}=\mathbf{x}_{\text{ss}}}\right) = (16b^2\gamma^2\mu m_s(\beta + \gamma(\delta-1)m_s)^2(b^2(\beta^3 + 2\beta^2\gamma(\delta-1)m_s + \beta\gamma^2(\kappa^2 + (1-2\delta)\kappa m_s + (\delta-1)(2\delta-1)m_s^2) \\ +\gamma^3\kappa m_s((\delta-1)m_s - \kappa)) + b(-\beta^2(2\gamma\mu + \theta) + \beta\gamma(\kappa(2\gamma\mu + \theta) + m_s(\gamma(5-4\delta)\mu - 2(\delta-1)\theta)) + \\ \gamma^2 m_s((\delta-1)m_s(3\gamma\mu + \theta) - \kappa(2\gamma\mu + \theta))) + \gamma\mu(\gamma\mu + \theta)(\beta - \gamma m_s)))/(\beta^3(b(\gamma(\kappa - 2\delta m_s + 2m_s) - \beta) + \gamma\mu + \theta)^4). \end{aligned} \tag{S57}$$

Eq. (S57) is difficult to understand. In order to perform further simplification, we make an ansatz for the form of $\tilde{B}$ ($\tilde{B}_{\text{An}}$) and seek to determine whether our ansatz is equivalent to the derived form of $\tilde{B}$ under the constraints defined on the left-hand side of Eq. (S57). Our ansatz takes the form

$$\tilde{B}_{\text{An}} := \lim_{M\to\infty}\left(\frac{2\mu h(1-h)}{n(\mathbf{x})}\cdot f_s(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}_{\text{ss}}}\right) \tag{S58}$$

where $f_s(\mathbf{x}) := (w_s + m_s)/(w_s + w_f + m_s + m_f)$ and $n(\mathbf{x}) := w_s + w_f + m_s + m_f$. Notice that this ansatz is more general than Eq. (S57), since it has no explicit dependence upon the parameters of the control law assumed in Eq. (10), and only explicitly depends upon functions of state $\mathbf{x}$.

Upon substituting the steady state $\mathbf{x}_{\text{ss}}$ into the ansatz in Eq. (S58) and taking the high-churn limit, we find that

$$\begin{aligned} \tilde{B}_{\text{An}} = -(m_s(\beta + \gamma(\delta-1)m_s)(b(\beta + \gamma\kappa) - \gamma\mu + \theta)(b(\beta + \gamma\kappa) + \gamma\mu - \theta)(b(m_s(2\beta\delta - \beta + \gamma\kappa) - 2\beta\kappa) + m_s(\theta - \gamma\mu))) \\ /(2b\beta^3(\kappa - \delta m_s + m_s)^2(b(\gamma(\kappa - 2\delta m_s + 2m_s) - \beta) + \gamma\mu + \theta)). \end{aligned} \tag{S59}$$

After some algebra (see GitHub repository for Mathematica notebook), it can be shown that Eq. (S57) and Eq. (S59) are equivalent, i.e.

$$\tilde{B}_{\text{An}} \equiv \lim_{M\to\infty}\left(\tilde{B}\big|_{\mathbf{x}=\mathbf{x}_{\text{ss}}}\right). \tag{S60}$$

As such, we may use $\tilde{B}$ and $\tilde{B}_{\text{An}}$ interchangeably in the limit of high network churn. Furthermore, it can be shown after some algebra that the drift of heteroplasmy when forced onto the steady state line is 0, i.e.

$$\tilde{A}\big|_{\mathbf{x}=\mathbf{x}_{\text{ss}}} \equiv 0. \tag{S61}$$

A similar result is shown in (Constable *et al.* 2016) ((Equation S59) therein). Substituting $h(\mathbf{x}, t) = h$, $\tilde{A}$ and $\tilde{B}$ into the Fokker-Planck equation for an arbitrary function of state Eq. (S51), we have

$$\frac{\partial P(h,t)}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial h^2}\left[\left(\frac{2\mu h(1-h)}{n(\mathbf{x})}\cdot f_s(\mathbf{x})\right)\Big|_{\mathbf{x}=\mathbf{x}_{\text{ss}}(h)} P(h,t)\right] \tag{S62}$$

which is equivalent to the following SDE for heteroplasmy

$$\mathrm{d}h = \sqrt{\frac{2\mu h(1-h)f_s(\mathbf{x})}{n(\mathbf{x})}}\bigg|_{\mathbf{x}=\mathbf{x}_{\mathrm{ss}}(h)} \mathrm{d}W \tag{S63}$$

in the limit of large network churn, large copy numbers, and a second-order truncation of the Kramers-Moyal expansion. Although the state has been forced onto the steady state, stochastic fluctuations mean that trajectories may move along the line of steady states, so the diffusion coefficient is not constant in general. We may calculate the new value of $\mathbf{x}_{\mathrm{ss}}(h)$ for every displacement due to Wiener noise in h, and substitute into $f_s(\mathbf{x})$ and $n(\mathbf{x})$ to determine the diffusion coefficient at the next time step.

However, for sufficiently short times, and large copy numbers (i.e. low diffusivity of h), we may assume that the diffusion coefficient in Eq. (S63) may be approximated as constant. Since the general solution of the SDE

$$\mathrm{d}y = \sqrt{B}\,\mathrm{d}W \tag{S64}$$

for $B = \text{const}$ is

$$y \sim \mathcal{N}(y|y_0, Bt) \tag{S65}$$

where $\mathcal{N}(y|y_0, \sigma^2)$ is a Gaussian distribution on y with mean y_0 and variance σ^2, and $y_0 = y(t=0)$. Since we have assumed that the state is initialised at $\mathbf{x}(t=0) = \mathbf{x}_{\mathrm{ss}}$, there are no deterministic transient dynamics, so we may write

$$\mathbb{V}(h) \approx \frac{2\mu t}{n(\mathbf{x})} h(\mathbf{x})(1-h(\mathbf{x}))f_s(\mathbf{x})\bigg|_{\mathbf{x}=\mathbf{x}_{\mathrm{ss}}}, \tag{S66}$$

where $\mathbb{V}$ returns variance of a random variable. In this equation, we take $\mathbf{x} = \mathbf{x}_{\mathrm{ss}} = \text{const}$, since we have assumed a low-diffusion limit. We observe that this equation is of precisely the same form as (Equation 12) of Johnston and Jones (2016), except with an additional proportionality factor of f_s induced by the inclusion of a mitochondrial network.

Heteroplasmy variance relations for alternative model structures and modes of genetic mtDNA control

Here we explore the implications of alternative model structures upon Eq. (S63). Firstly, we may consider replacing Eq. (4) with

$$X_S \xrightarrow{\lambda} X_S + X_S. \tag{S67}$$

This corresponds to the case where replication coincides with fission, see (Lewis *et al.* 2016). Repeating the calculation in the previous section also results in Eq. (S63), so the result is robust to the particular choice of mtDNA replication reaction (see GitHub repository for Mathematica notebook).

Secondly, we may explore the impact of allowing non-zero mtDNA degradation of fused species. This could correspond to autophagy-independent degradation of mtDNA, for example via the exonuclease activity of POLG (Medeiros *et al.* 2018). To encode this, we may add the following additional reaction

$$X_F \xrightarrow{\xi\mu} \emptyset \tag{S68}$$

where $0 \leq \xi \leq 1$. We were not able to make analogous analytical progress in this instance. However, numerical investigation (Figure S3E) revealed that the following ansatz was able to predict heteroplasmy variance dynamics

$$\mathbb{V}(h) \approx \frac{2\mu t}{n(\mathbf{x})} h(\mathbf{x})(1-h(\mathbf{x}))(f_s(\mathbf{x}) + \xi(1-f_s(\mathbf{x})))\bigg|_{\mathbf{x}=\mathbf{x}_{\mathrm{ss}}}. \tag{S69}$$

In other words, allowing degradation of fused species results in a linear correction to our heteroplasmy variance formula in Eq. (13). If fused species are susceptible to degradation at the same rate as unfused species ($\xi = 1$), then $\mathbb{V}(h)$ loses f_s dependence entirely and the mitochondrial network has no influence over heteroplasmy dynamics.

We also explored various different forms of $\lambda(\mathbf{x})$ and $\mu(\mathbf{x})$, which we label A-G after (Johnston and Jones 2016), and X-Z for several newly-considered functional forms, see Table S3 and Figure S4A-I. Control D of (Johnston and Jones 2016) involves no feedback, which we do not explore – see Figure S1, and the discussion surrounding Eq. (S1). The argument presented in the previous section requires the steady state solution of the system to be solvable, since we require the explicit form of $\mathbf{x}_{\mathrm{ss}}$ in Eq. (S57), Eq. (S59) and Eq. (S61). For controls B, C, E, F, G, Y and Z in Table S3, the steady states are solvable and similar arguments to the above can be applied (see the GitHub repository for Mathematica notebooks). Controls B, C, E, F all satisfy Eq. (S63); this can be shown numerically for controls A and X. However, controls G, Y and Z satisfy

$$\mathrm{d}h = \sqrt{\frac{2\mu h(1-h)}{n(\mathbf{x})}}\bigg|_{\mathbf{x}=\mathbf{x}_{\mathrm{ss}}(h)} \mathrm{d}W. \tag{S70}$$

Notably, Eq. (S70) does not depend on f_s, unlike Eq. (S63) (see GitHub repository for Mathematica notebooks). This is because control of copy number occurs in the degradation rate, rather than the replication rate, for controls G, Y and Z. A modified version of a Moran process (presented below) can provide intuition for why the diffusion rate of heteroplasmy variance depends on the network state when the population is controlled through replication, and does not depend on network state when the population is controlled through degradation.

Choice of nominal parametrization

In this section we discuss our choice of nominal parametrization for the network system in Eq. (1)-Eq. (9), given the replication rate in Eq. (10). We will first discuss our choice of network parameters.

Cagalinec *et al.* (2013) found that the average fission rate in cortical neurons is 0.023 ± 0.003 fissions/mitochondria/min. Assuming that this value is representative of the fission rate in general, and converting this to units of per day, we may write the mitochondrial fission rate as $\beta = 33.12\ \text{day}^{-1}$.

The dimensions of β are day^{-1} and not $\text{mitochondrion}^{-1}\ \text{day}^{-1}$. This is because if the propensity (see Eq. (S12), where $\Omega = 1$) of e.g. Eq. (3) is $\hat{a}_{\text{fis},w} = \beta w_f$ then the mean time to the next event is $1/(\beta w_f)$; therefore the dimension of β is per unit time and copy numbers are pure numbers, i.e. dimensionless. Similar reasoning constrains the dimension of the fusion rate, see below.

Evaluation of the fusion rate is more involved, since fusion involves two different chemical species coming together to react whereas fission may be considered as spontaneous. Furthermore, there are 7 different fusion reactions whereas there are only 2 fission reactions. For simplicity, assume that all species have a steady-state copy number of $x_i = 250$ (resulting in a total copy number of 1000, heteroplasmy of 0.5 and 50% of mitochondria existing in the fused state). Neglecting subtleties relating to bimolecular reactions involving one species (see Eq. (S20)), each fusion reaction proceeds at rate $\hat{a}_{\text{fus},j} \approx \gamma x_i^2$. Since there are 7 fusion reactions (Eq. (1), Eq. (2), Eq. (7)-Eq. (9)), the total fusion propensity is $\hat{a}_{\text{fus}} \approx 7\gamma x_i^2$. Similarly, the total fission propensity is $\hat{a}_{\text{fis}} = \beta(w_f + m_f) = 2\beta x_i$. Since we expect macroscopic proportions of both fused and fissioned species in many physiological settings, we may equate the fusion and fission propensities, $\hat{a}_{\text{fus}} = \hat{a}_{\text{fis}}$, and rearrange for the fusion rate γ to yield $\gamma = 2\beta x_i/(7x_i^2) \approx 3.8 \times 10^{-2}\ \text{day}^{-1}$. The orders of magnitude difference between β and γ stems from the observation that fusion propensity depends on the square of copy number whereas the fission propensity depends on copy number linearly.

Given the network parameters, we then explored appropriate parametrizations for the genetic parameters: the mitophagy rate (μ) and the parameters of the linear feedback control (κ, b and δ, see Eq. (10)). mtDNA half-life is observed to be highly variable: in mice this can be between 10-100 days (Burgstaller *et al.* 2014a). For consistency with another recent study investigating the relationship between network dynamics and heteroplasmy, we use an mtDNA half-life of 30 days (Tam *et al.* 2015).

The parameter δ in the replication feedback control (see Eq. (10)) may be interpreted as the "strength of sensing of mutant mtDNA" in the feedback control (Hoitzing *et al.* 2017). Assuming that fluctuations in copy number of mutants and wild-type molecules are sensed identically (as may be the case for e.g. non-coding mtDNA mutations) we may reasonably assume a model of $\delta = 1$ as the *simplest* case of a neutral mutation (although $\delta \neq 1$ still defines a neutral model, since both mutant and wild-type alleles experience the same replication and degradation rates per molecule, see Eqs.(4)–(6)).

We are finally left with setting the parameters κ and b in the linear feedback control Eq. (10). In the absence of a network state and mutants, κ is precisely equal to the steady state copy number, since the degradation rate equals the replication rate when $\kappa = w$. However, the presence of a network means that a subpopulation of mtDNAs (namely the fused species) are immune to death, resulting in κ no longer being equivalent to the steady state copy number. The parameter b may be interpreted as the feedback control strength, which determines the extent to which the replication rate changes given a unit change in copy number.

Given a particular value of b, we may search for a κ which gives a total steady state copy number (n) which is closest to some target value (e.g. 1000 as a typical total mtDNA copy number per cell in human fibroblasts (Kukat *et al.* 2011)). We swept a range of different values of b and found that, for values of b smaller than a critical value (b^*), a κ could not be found whose deterministic steady state was sufficiently close to $n = 1000$. This result is intuitive because in the limit of $b \to 0$, $\lambda = \text{const}$. From the analysis above we have shown that constant genetic rates (μ, λ) result in unstable copy numbers, and therefore a sufficiently small value of b is not expected to yield a stable non-trivial steady state solution. We chose $b \approx b^*$, and the corresponding κ, such that the steady state copy number is controlled as weakly as possible given the model structure.

Rate renormalization

In Eqs. (1)–(10) we have neglected reactions such as

$$X_F + X_F \to X_F + X_F \tag{S71}$$

because they do not change the number of molecules in our state vector $\mathbf{x} = (w_s, w_f, m_s, m_f)$. One may ask whether neglecting such reactions means that it is necessary to renormalize the fission-fusion rates which were estimated in the preceding section. In estimating the nominal parametrization above, we began by using a literature value for the mitochondrial fission rate, and then matched the fusion rate such that the summed hazard of a fusion event approximately balanced the fission rate. This matching procedure is reasonable, since we observe a mixture of fused and fissioned mitochondria under physiological conditions: choosing a fusion rate which is vastly different results in either a hyperfused or fragmented network. We must therefore only justify the fission rate. Eq. (3) assumes that a fission reaction always results in a singleton, and a singleton is by definition a molecule which is susceptible to mitophagy (see Eq. (6)). Therefore, if fission reactions always result in mitochondria containing single mtDNAs which are susceptible to mitophagy, then we expect our model to match well to true physiological rates. If, on the other hand, fission reactions between large components of the network which are too large to be degraded are common, then renormalization of β by the fraction of fission events which result in a sufficiently small mitochondrion would be necessary, which would in turn renormalize γ through our matching procedure. We are not aware of experimental measurements of the fraction of fission events which result in mitochondria which contain a particular number of mtDNAs. Such an experiment, combined with the distribution of mitochondrial sizes which are susceptible to mitophagy, would allow us to validate our approach. Despite this, the robustness of our results over approximately 4 orders of magnitude for the fission-fusion rate (Figure S4A-I) provides some indication that our results are likely to hold in physiological regions of parameter space.

A modified Moran process may account for the alternative forms of heteroplasmy variance dynamics under different models of genetic mtDNA control

We sought to gain insight into why control of population size through the replication rate ($\lambda = \lambda(\mathbf{x})$, $\mu = \text{const}$) results in heteroplasmy variance depending on the fraction of unfused mitochondria (see Eq. (13)), whereas control of population size through the degradation rate ($\mu = \mu(\mathbf{x})$, $\lambda = \text{const}$) results in heteroplasmy variance becoming independent of network state, where

$$\mathbb{V}(h) \approx \frac{2\lambda t}{n(\mathbf{x})} h(\mathbf{x})(1-h(\mathbf{x}))\Big|_{\mathbf{x}=\mathbf{x}_{\text{ss}}}. \tag{S72}$$

We will proceed by considering an analogous Moran process to the set of reactions presented in Eqs. (1) (9).

First, consider a haploid biallelic Moran process consisting of wild-types and mutants, in a population of fixed size n. At each step in discrete time, a member of the population is chosen for birth, and another for death. Let m_t denote the copies of mutants at time t. Then,

$$P(m_{t+1} = j | m_t = i) = \begin{cases} i(n-i)/n^2 & \text{if } j = i \pm 1 \\ i^2/n^2 + (n-i)^2/n^2 & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases} \tag{S73}$$

It follows that

$$\mathbb{E}(m_{t+1} | m_t) = m_t. \tag{S74}$$

Defining $h_t := m_t/n$ then from Eq. (S73)

$$\mathbb{V}(m_{t+1} | m_t) = 2h_t(1-h_t) \tag{S75}$$

and therefore

$$\mathbb{V}\left(\frac{m_{t+1}}{n}\middle| m_t\right) = \mathbb{V}(h_{t+1}|m_t) = \frac{2}{n^2} h_t(1-h_t). \tag{S76}$$

Suppose that, instead of the process occurring with discrete time, instead the process occurs with continuous time, where each event is a simultaneous birth and death, and is modelled as a Poisson process. Suppose that events occur at a rate μ per capita. The waiting time between successive events (τ) is an exponential random variable with rate μN. Hence the expected waiting time between successive events is

$$\mathbb{E}(\tau) = \frac{1}{\mu n}. \tag{S77}$$

If we take the ratio of Eq. (S76) and Eq. (S77), we have

$$\frac{\mathbb{V}(h_{t+\tau}|m_t)}{\mathbb{E}(\tau)} = \frac{2\mu h_t(1-h_t)}{n}. \tag{S78}$$

Heuristically, one could interpret Eq. (S78) as a ratio of differentials as follows. If we were to suppose that n were large enough such that $\mathbb{E}(\tau)$ is very small, and h_t is approximately constant (h_0) after a small number of events, then

$$\frac{\Delta\mathbb{V}(h)}{\Delta\tau} \approx \frac{\mathbb{V}(h_{t+\tau}|m_t)}{\mathbb{E}(\tau)} \implies \frac{\mathrm{d}\mathbb{V}(h)}{\mathrm{d}t} \approx \frac{2\mu h_0(1-h_0)}{n} \implies \mathbb{V}(h,t) \approx \frac{2\mu h_0(1-h_0)}{n} t \tag{S79}$$

where we have replaced the inter-event time τ with physical time t. This result is analogous to Eq. (S72) and Eq. (12) of Johnston and Jones (2016), and agrees with simulation (Figure S6A).

Now consider the modified Moran process in Figure S6B, which we refer to as a "protected" Moran process. Let $0 < f_s \leq 1$ be the fraction of individuals susceptible to death, which is a constant. nf_sh_t and $nf_s(1-h_t)$ mutants and wild-types are randomly chosen to be susceptible to death, respectively, where n is large. In this continuous-time model, the inter-event time is $\tau \sim \text{Exponential}(\Gamma)$ where Γ will be defined below. Then an individual from the susceptible population is chosen for death, and any individual is allowed to be born. The birth and death events occur simultaneously in time.

Again, using t as an integer counter of events, we have

$$P(m_{t+1} = j | m_t = i) = \begin{cases} \frac{nf_sh_t}{nf_s}\frac{n(1-h_t)}{N} & \text{if } j = i-1 \\ \frac{nf_s(1-h_t)}{nf_s}\frac{nh_t}{N} & \text{if } j = i+1 \\ \frac{nf_sh_t}{nf_s}\frac{nh_t}{N} + \frac{nf_s(1-h_t)}{nf_s}\frac{n(1-h_t)}{n} & \text{if } j = i \\ 0 & \text{otherwise} \end{cases} \tag{S80}$$

$$= \begin{cases} h_t(1-h_t) & \text{if } j = i \pm 1 \\ h_t^2 + (1-h_t)^2 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases} \tag{S81}$$

which is equivalent to the definition of a Moran process in Eq. (S73), meaning that Eq. (S76) applies to the protected Moran process as well.

We consider two heuristic arguments for choosing the inter-event rate Γ, where the inter-event time $\tau \sim \text{Exponential}(\Gamma)$. Firstly, if the death rate per capita is constant (μ), then the rate at which a death event occurs in the system (Γ_{death}) is proportional to the number of individuals which are susceptible to death: $\Gamma_{\text{death}} = \mu n f_s$. If we assume that the overall birth rate is matched to the overall death rate so that population size is maintained, as is the case when $\lambda = \lambda(\mathbf{x})$ in the network system, then the overall birth rate (Γ_{birth}) must also be $\Gamma_{\text{birth}} = \mu n f_s$. Hence,

$$\Gamma = \Gamma_{\text{birth}} + \Gamma_{\text{death}} = 2\mu n f_s \tag{S82}$$

where μ is a proportionality constant. Since, for a Moran event to occur, both a birth and a death event must occur, time effectively runs twice as fast in a Moran model relative to a comparable chemical reaction network model. We therefore rescale time by taking $\mu \to \mu/2$, and thus

$$\Gamma = \mu n f_s. \tag{S83}$$

As a result, $\mathbb{E}(\tau) = 1/(\mu n f_s)$ and therefore, using Eq. (S76) and the reasoning in Eq. (S79),

$$\frac{\Delta \mathbb{V}(h)}{\Delta \tau} \approx \frac{\mathbb{V}(h_{t+\tau}|m_t)}{\mathbb{E}(\tau)} \implies \frac{\mathrm{d}\mathbb{V}(h)}{\mathrm{d}t} \approx \frac{2\mu f_s h_0(1-h_0)}{n} \implies \mathbb{V}(h) \approx \frac{2\mu f_s h_0(1-h_0)}{n} t. \tag{S84}$$

This is analogous to when $\lambda = \lambda(\mathbf{x})$ and $\mu = \text{const}$ in the network system. *Hence, when $\lambda = \lambda(\mathbf{x})$ and $\mu = const$, the absence of death in the fused subpopulation means the timescale of the system (being the time to the next death event) is proportional to f_s.* This argument is only a heuristic, since the Moran process is defined such that birth and death events occur simultaneously and therefore do not possess separate propensities (Γ_{birth} and Γ_{death}).

The second case we consider is when each individual has a constant rate of birth, hence $\Gamma_{\text{birth}} \propto n$. Then the death rate is chosen such that $\Gamma_{\text{birth}} = \Gamma_{\text{death}}$. In this case $\Gamma = \lambda n$, where λ is a proportionality constant. The same argument from Eq. (S77) to Eq. (S79) may be applied, with an appropriate rescaling of time, and we arrive at Eq. (S72). This is analogous to when $\mu = \mu(\mathbf{x})$ and $\lambda = \text{const}$ in the network system. *Hence, when $\mu = \mu(\mathbf{x})$ and $\lambda = const$, the presence of a constant birth rate in the entire population means the timescale of the system (being the time to the next birth event) is independent of f_s.*

References

Burgstaller, J. P., I. G. Johnston, N. S. Jones, J. Albrechtová, T. Kolbe, *et al.*, 2014 MtDNA segregation in heteroplasmic tissues is common in vivo and modulated by haplotype differences and developmental stage. Cell Rep. **7**: 2031–2041.

Cagalinec, M., D. Safiulina, M. Liiv, J. Liiv, V. Choubey, *et al.*, 2013 Principles of the mitochondrial fusion and fission cycle in neurons. J. Cell Sci. **126**: 2187–2197.

Constable, G. W., T. Rogers, A. J. McKane, and C. E. Tarnita, 2016 Demographic noise can reverse the direction of deterministic selection. Proc. Natl. Acad. Sci. USA p. 201603693.

Gardiner, C., 1985 *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer Series in Synergetics, Springer Berlin Heidelberg.

Gillespie, D. T., 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comp. Phys. **22**: 403–434.

Gillespie, D. T., 2007 Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem. **58**: 35–55.

Grima, R., 2010 An effective rate equation approach to reaction kinetics in small volumes: Theory and application to biochemical reactions in nonequilibrium steady-state conditions. J. Chem. Phys. **133**: 07B604.

Grima, R., P. Thomas, and A. V. Straube, 2011 How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations? J. Chem. Phys. **135**: 084103.

Hoitzing, H., P. A. Gammage, M. Minczuk, I. G. Johnston, and N. S. Jones, 2017 Energetic costs of cellular and therapeutic control of stochastic mtDNA populations. bioRxiv p. 145292.

Jacobs, K., 2010 *Stochastic processes for physicists: understanding noisy systems*. Cambridge University Press.

Johnston, I. G. and N. S. Jones, 2016 Evolution of cell-to-cell variability in stochastic, controlled, heteroplasmic mtDNA populations. Am. J. Hum. Genet. **99**: 1150–1162.

Kukat, C., C. A. Wurm, H. Spåhr, M. Falkenberg, N.-G. Larsson, *et al.*, 2011 Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. Proc. Natl. Acad. Sci. USA **108**: 13534–13539.

Lewis, S. C., L. F. Uchiyama, and J. Nunnari, 2016 ER-mitochondria contacts couple mtDNA synthesis with mitochondrial division in human cells. Science **353**: aaf5549.

Medeiros, T. C., R. L. Thomas, R. Ghillebert, and M. Graef, 2018 Autophagy balances mtDNA synthesis and degradation by DNA polymerase POLG during starvation. J. Cell Biol. pp. jcb–201801168.

Mouli, P. K., G. Twig, and O. S. Shirihai, 2009 Frequency and selectivity of mitochondrial fusion are key to its quality maintenance function. Biophys. J. **96**: 3509–3518.

Parsons, T. L. and T. Rogers, 2017 Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology. J. Phys. A **50**: 415601.

Tam, Z. Y., J. Gruber, B. Halliwell, and R. Gunawan, 2013 Mathematical modeling of the role of mitochondrial fusion and fission in mitochondrial DNA maintenance. PloS ONE **8**: e76230.

Tam, Z. Y., J. Gruber, B. Halliwell, and R. Gunawan, 2015 Context-dependent role of mitochondrial fusion-fission in clonal expansion of mtDNA mutations. PLoS Comp. Biol. **11**: e1004183.

Twig, G., A. Elorza, A. J. Molina, H. Mohamed, J. D. Wikstrom, *et al.*, 2008 Fission and selective fusion govern mitochondrial segregation and elimination by autophagy. EMBO J. **27**: 433–446.
Van Kampen, N. G., 1992 *Stochastic processes in physics and chemistry*, volume 1. Elsevier.
Wilkinson, D. J., 2011 *Stochastic modelling for systems biology*. CRC press.

Table S1 Key predictions from our mathematical models.

The following results hold for our neutral genetic model of a post-mitotic cell, with a simple model of mitochondrial network dynamics:

1. The rate of increase of heteroplasmy variance is proportional to the fraction of unfused mitochondria, but independent of the absolute magnitude of fission-fusion rates, due to a rescaling of time by the mitochondrial network (Eq.(13), Eq.(S63), Figure 2E-H).
2. The rate of accumulation of *de novo* mutations increases as the fraction of unfused mitochondria increases, due to a rescaling of time by the mitochondrial network (Figure 3B-D)
3. When fusion is selective, intermediate fusion-fission ratios are optimal for reducing mean heteroplasmy (Figure 4A)
4. When mitophagy is selective, complete fission is optimal for reducing mean heteroplasmy (Figure 4B)

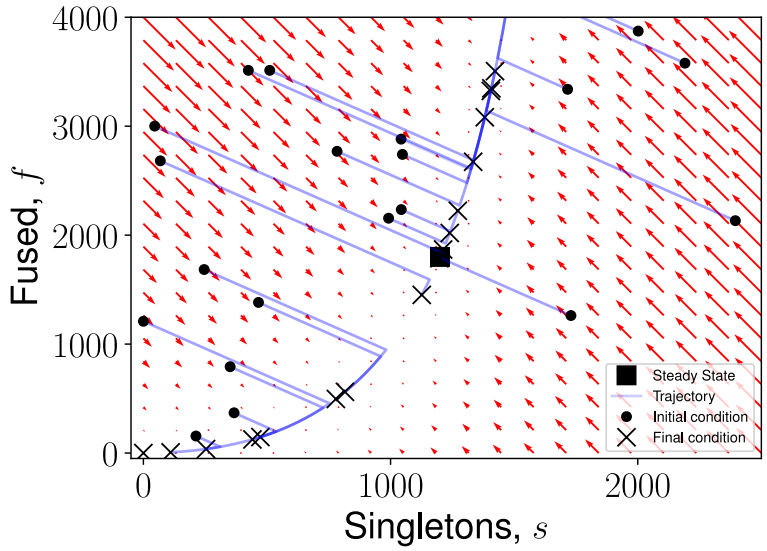

Figure S1 Phase portrait for an ODE representation of a network system with constant rates. The system displays two steady states: a trivial steady state at $s + f = 0$, and a non-trivial steady state. At a steady state, both time derivatives in Eq. (S8) and Eq. (S9) vanish. Trajectories (blue) show the evolution of the system up to $t = 1000$. The direction and magnitude of the derivative at points in space are shown by red arrows. Trajectories can be seen either decaying to $s = f = 0$ or tending to infinity. $\gamma = 0.01656$, $\beta = 33.12$, $\rho = 0.023$, $\eta = 1.1$, $\alpha = 1.04$.

Table S2 Nominal parametrization for network system. Dimensions of parameters are derived using individual terms in Eqs. (S29)–(S32); copy numbers of particular species are pure numbers and therefore dimensionless. Since we have chosen a system size of $\Omega = 1$ throughout, we set the dimension of volume to 1. See Choice of nominal parametrization for further details of parameter justification.

Parameter	Description	Value	Dimensions	Remarks
β	Fission rate	33.12	day^{-1}	For cortical neurons, see (Cagalinec *et al.* 2013)
γ	Fusion rate	3.79×10^{-2}	day^{-1}	Approximately balances fission
μ	Mitophagy rate	0.023	day^{-1}	From (Tam *et al.* 2015)
δ	Mutant feedback sensitivity	1	dimensionless	Potentially appropriate for e.g. non-coding mtDNA mutations
b	Feedback control strength	1.24×10^{-5}	day^{-1}	Chosen as the weakest control strength which has a non-trivial steady state and total copy number of 1000
κ	Steady state copy number parameter	11.7	dimensionless	See remark for b

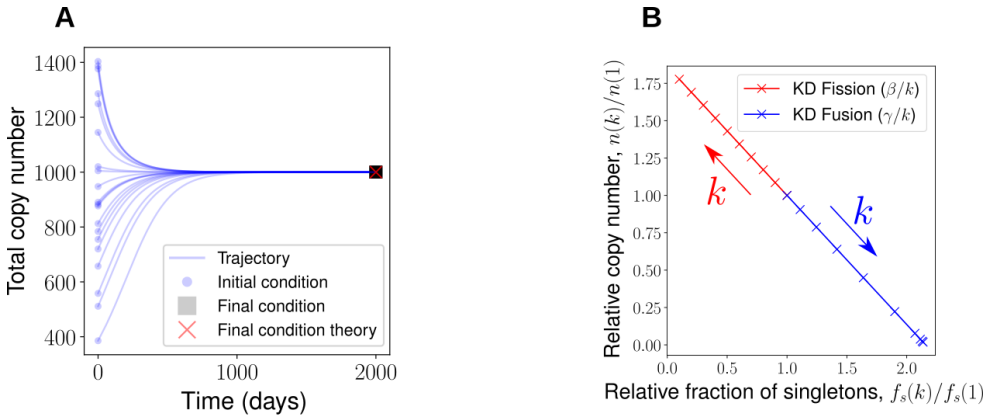

Figure S2 Deterministic treatment of network system. (**A**) Deterministic dynamics of total copy number under linear feedback control, which is controlled to a particular steady state value (see Figure 2A&B). (**B**) Defining a knock-down (KD) factor ($k^{-1} = 0.1, 0.2 \ldots, 1.0$), the fission rate was rescaled to $\beta \to \beta/k$ (red) and the fusion rate to $\gamma \to \gamma/k$ (blue), causing a linear increase and decrease in total copy number respectively under a deterministic treatment (see Figure 2A&B).

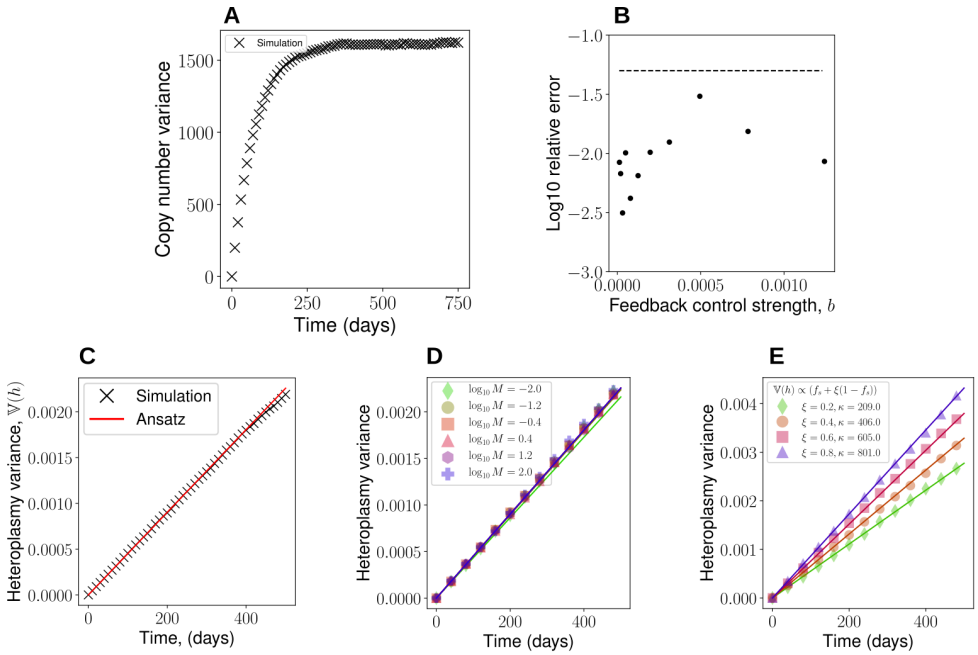

Figure S3 Stochastic treatment of network system. (**A**) Copy number variance for stochastic simulations initially increases, since all stochastic simulations begin with the same initial condition, but then plateaus since the steady state line is globally attracting (see Figure 2C&D). (**B**) Error in Eq. (13) in a sweep over the feedback control strength, b. Dotted line denotes a 5% error according to Eq. (12). (**C**) $\mathbb{V}(h)$ profile for the parametrization with the largest error in (**B**). (**D**) Sweeps of the network rate magnitude (see Figure 2H). Heteroplasmy variance is approximately independent of absolute network rates over a broad range of network magnitudes. (**E**) Allowing fused species to be degraded with relative rate ξ (Eq. (S68)), stochastic simulations for heteroplasmy variance (markers) and Eq. (S69) (lines) are shown. Fused species degradation induces a linear correction to the heteroplasmy variance formula.

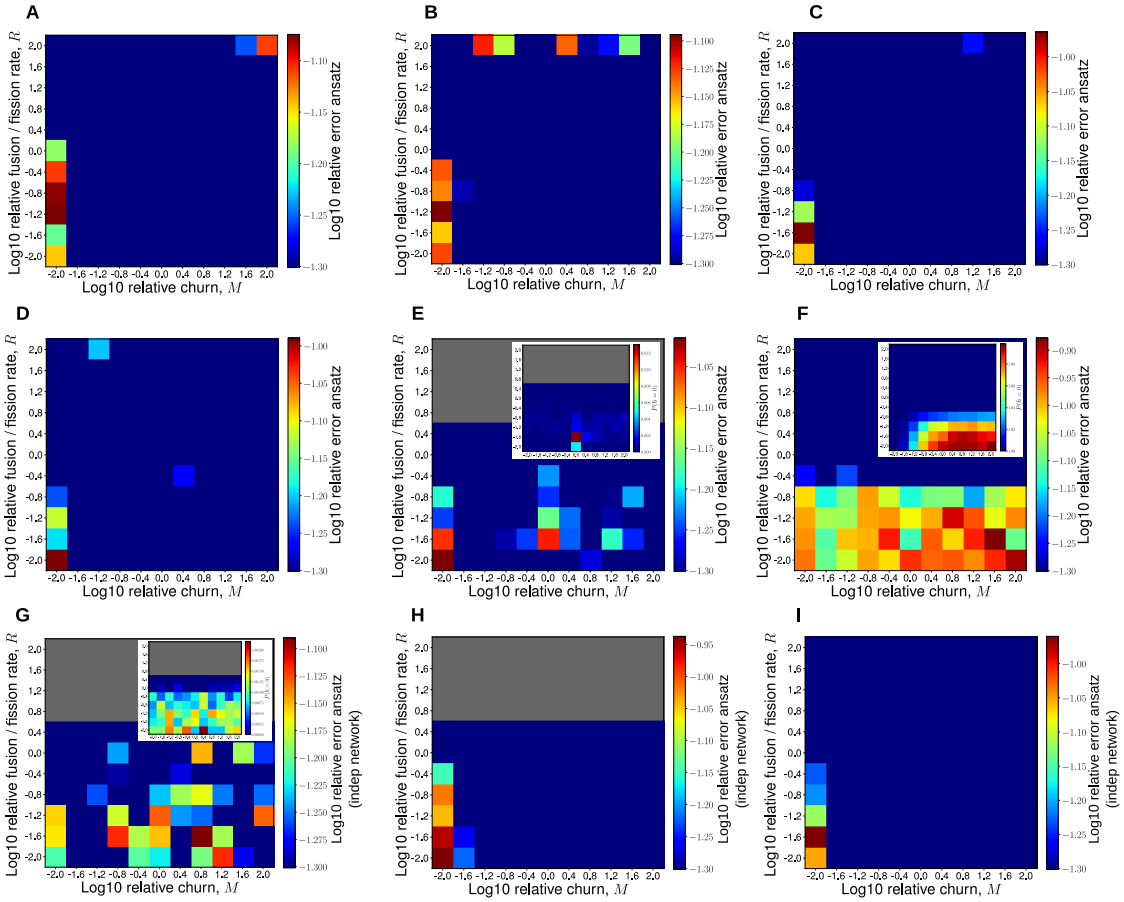

Figure S4 Parameter sweeps of network fission-fusion rates for replication-based and degradation-based control modes show robustness of two heteroplasmy variance formulae respectively. (**A-I**) Error in Eq. (13) (ansatz) and Eq. (S72) (ansatz indep network) for sweeps of the fusion and fission rates for the corresponding feedback control functions in Table S3. Equations are accurate to at least 5% (blue regions) across large regions of parameter space, for many control laws. Fusion and fission rates are redefined as $\gamma \to \gamma_0 MR$ and $\beta \to \beta_0 M$ where M and R denote the magnitude and ratio of the network rates, and γ_0, β_0 denote the nominal parametrizations of the fusion and fission rates respectively. Summary statistics for 10^4 realizations, with initial condition $h = 0.3$ and evaluated at $t = 500$ days. Errors in $\mathbb{V}(h)$ (see Eq. (12)) smaller than 5% are truncated and are shown as blue. Parametrizations where a deterministic steady state could not be found for an initial condition of $h = 0.3$ are shown in grey. Inset figures, where present, display the probability of fixation at $h = 0$. Where insets are not present, the probability of fixation is negligible. (**A-F**) When $\lambda = \lambda(x)$ and $P(h = 0)$ is low, Eq. (13) performs well in the high-churn limit. (**G-I**) When $\mu = \mu(x)$ and $P(h = 0)$ is low, Eq. (S72) performs well in the high-churn limit.

Table S3 Nominal parametrizations for the alternative feedback control functions explored. Nominal parametrizations for the feedback controls in Fig. S4. In all cases, the nominal fission and fusion rates were $\beta = 33.12$, $\gamma = 0.038$ respectively.

Interpretation	Replication rate	Degradation rate	Note
Linear feedback (see Figure 2, S2, S3, S4A)	$\mu + b(\kappa - w_T - \delta m_T)$; see Table S2	μ; see Table S2	Control E in (Johnston and Jones, 2016) and GitHub repository
Relaxed replication (see Figure S4B)	$\alpha\mu(w_{opt} - w_T - \delta m_T)/(w_T + m_T)$; $\alpha = 1$, $w_{opt} = 1000$, $\delta = 1$	μ; $\mu = 0.023$	Control A in (Johnston and Jones, 2016) and GitHub repository
Differential control for target population (see Figure S4C)	$\alpha(w_{opt} - w_T)$	μ; $\mu = 0.023$	Control B in (Johnston and Jones, 2016) and GitHub repository
Ratiometric control for target population (see Figure S4D)	$\alpha(w_{opt}/w_T - 1)$; $\alpha = 1$, $w_{opt} = 1000$	μ; $\mu = 0.023$	Control C in (Johnston and Jones, 2016) and GitHub repository
Production independent of wild-type (see Figure S4E)	α/w_T; $\alpha = 5$	μ; $\mu = 0.023$	Control F in (Johnston and Jones, 2016) and GitHub repository
General linear feedback (see Figure S4F)	$\mu + b(\kappa - \delta_1 w_s - \delta_2 w_f - \delta_3 m_s - \delta_4 m_f)$; see Table S2, $\delta_1 = 0.8$, $\delta_2 = 1.0$, $\delta_3 = 0.2$, $\delta_4 = 0.3$	μ; see Table S2	Control X in GitHub repository
Ratiometric control through degradation (see Figure S4G)	λ; $\lambda = 0.023$	$\mu w_T/w_{opt}$; $\mu = 0.023$; $w_{opt} = 200$	Control G in (Johnston and Jones, 2016) and GitHub repository
Linear feedback in degradation (see Figure S4H)	λ; $\lambda = 0.023$	$\mu + b(w_T + \delta m_T - \kappa)$; $\mu = 0.023$, $b = 10^{-4}$, $\kappa = 1000$, $\delta = 1$	Control Y in GitHub repository
Differential control for target population in degradation (see Figure S4I)	λ; $\lambda = 0.023$	$\alpha(w_T - w_{opt})$; $\alpha = 1$, $w_{opt} = 1000$	Control Z in GitHub repository

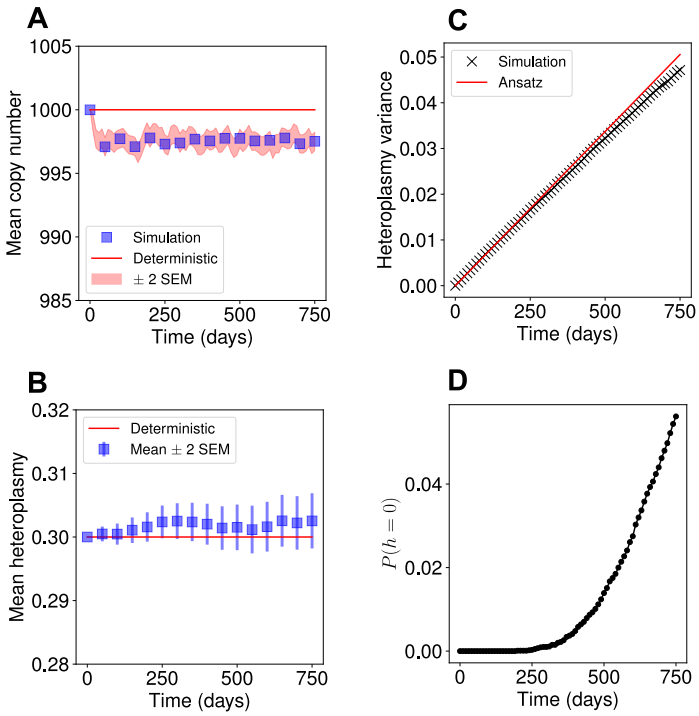

Figure S5 Ansatz predicts heteroplasmy variance for linear feedback control in a fast mtDNA turnover regime when fixation probability is low. Stochastic simulations of the linear feedback control network system with an mtDNA half-life of 2 days, corresponding to $\mu = \ln(2)/2$. (**A**) The mean copy number, and (**B**) the mean heteroplasmy show qualitatively similar dynamics to the nominal parametrisation presented (see Fig. 2C&D). (**C**) Eq.(13) predicts $\mathbb{V}(h)$ accurately up to approximately 250 days, where the Eq.(13) begins to overestimate the variance. (**D**) The over-estimation of heteroplasmy variance coincides with an increase in the probability of fixation at $h = 0$. Parameters apart from μ were chosen according to the protocol outlined in Choice of nominal parametrization, with 10^4 repeats. $\kappa = 101.6$, $b = 2.07 \times 10^{-4}$, for all other parameters see Table S2.

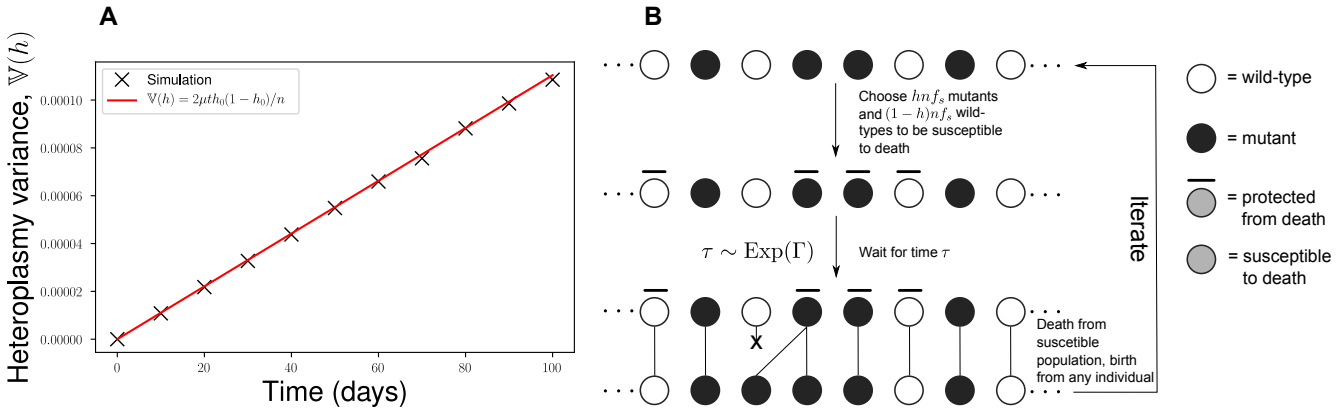

Figure S6 Exploration of analogous Moran processes. (**A**) The original biallelic Moran process satisfies Eq. (S72), where h_0 is the initial heteroplasmy, which is equivalent to $\mathbb{E}(h)$. (**B**) The "protected" Moran process. The population size is constrained to be fixed to some large constant, n. There exist two alleles, mutants (black circle) and wild-types (white circles). hnf_s mutant and $(1-h)nf_s$ wild-type molecules are susceptible to death, the rest are protected from death (denoted by a bar). An exponential random variable is drawn as the waiting time to the next event (see A modified Moran process may account for the alternative forms of heteroplasmy variance dynamics under different models of genetic mtDNA control for discussion of the form of the rate Γ). Time is incremented by the waiting time, then a death event occurs from the susceptible population and a birth event from any individual simultaneously. The same individual is allowed to be chosen for both birth and death. The process is then repeated iteratively.

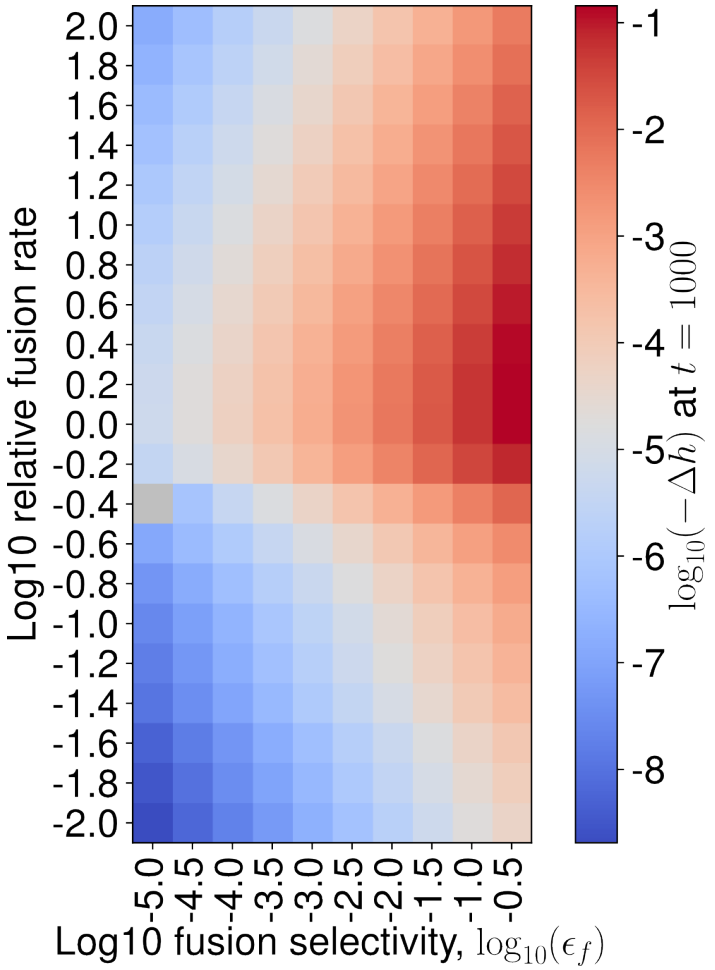

Figure S7 A deterministic parameter sweep of fusion selectivity and the relative fusion rate for mitochondrial quality control. An ODE treatment allows smaller heteroplasmy changes to be probed without the need to resort to an infeasible number of stochastic simulations. Displaying the relative change in heteroplasmy (Δh) after $t = 1000$ days. We observe that a reduction in heteroplasmy is achieved at intermediate fusion rates at all non-zero fusion selectivities investigated. Grey denotes a change which is smaller than floating point precision.

Table S4 Comparison of previous models with our model of mitochondrial genetic/network dynamics. Each of the key differences with our model is enumerated, and has a corresponding comment; see Discussion.

Model	Key assumptions	Key differences	Comments
(Tam *et al.* 2013, 2015)	• Cell consists of 16 subcompartments • Fission/fusion induces migration between subcompartments	1. Slower fission-fusion dynamics result in larger rate of increase in $\mathbb{V}(h)$ 2. Fast fission-fusion rates cause $\mathbb{E}(h)$ to increase	1. Limiting to regimes where fission-fusion is fast, likely results in loss of rate magnitude sensitivity 2. Could be explained by spatial effects which we neglect here
(Mouli *et al.* 2009)	• Fission follows fusion • Mitochondria consist of multiple units • Sigmoidal relationship between number of functional units per mitochondrion and activity	1. When fusion is selective, higher fusion rates are optimal 2. When mitophagy is selective, intermediate fusion rates are optimal	1. The optimal fusion rate is an order of magnitude lower than the highest fusion rate considered 2. Non-linearity between function, intra-mitochondrial heteroplasmy, and network state, may result in intermediate fusion optimality