Varied Mechanisms and Models for the Varying Mitochondrial Bottleneck

*Iain G. Johnston**

Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Bergen, Bergen, Norway

Mitochondrial DNA (mtDNA) molecules exist in populations within cells, and may carry mutations. Different cells within an organism, and organisms within a family, may have different proportions of mutant mtDNA in these cellular populations. This diversity is often thought of as arising from a "genetic bottleneck." This article surveys approaches to characterize and model the generation of this genetic diversity, aiming to provide an introduction to the range of concepts involved, and to highlight some recent advances in understanding. In particular, differences between the statistical "genetic bottleneck" (mutant proportion spread) and the physical mtDNA bottleneck and other cellular processes are highlighted. Particular attention is paid to the quantitative analysis of the "genetic bottleneck," estimation of its magnitude from observed data, and inference of its underlying mechanisms. Evidence that the "genetic bottleneck" (mutant proportion spread) varies with age, between individuals and species, and across mtDNA sequences, is described. The interpretation issues that arise from sampling errors, selection, and different quantitative definitions are also discussed.
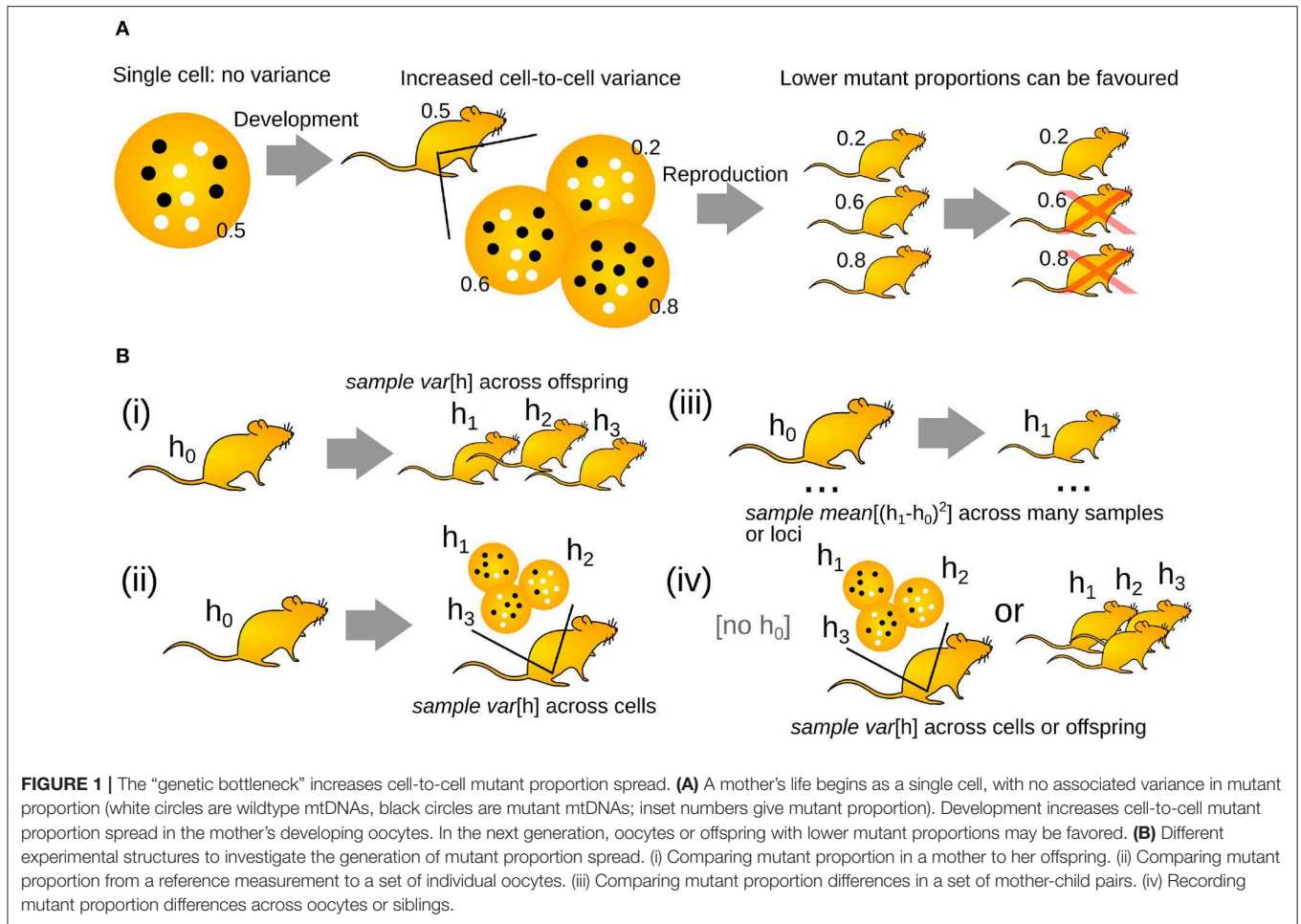
Keywords: mtDNA, bottleneck, development, inheritance, modeling, uncertainty, heterogeneity

1. INTRODUCTION

Mitochondria are vital energy-producing compartments in eukaryotic cells. As a result of their evolutionary history, they retain small genomes (mtDNA) which encode important respiratory machinery. In humans and other species, mtDNA molecules are inherited uniparentally, rarely recombine, and can acquire damaging mutations (Wallace and Chalkia, 2013). As hundreds or thousands of mtDNA molecules exist in the same cell, mutations may be present in some but not all molecules: we refer to the fraction of molecules in a cell with a given mutation as the "mutant proportion". MtDNA molecules within the same cell can harbor many different genetic variants at low proportions, a situation called microheteroplasmy (Guo et al., 2013). The mutant proportion associated with each single genetic variant is of scientific and translational interest, particularly as some variants (e.g., point mutations) have pathological consequences above a certain "threshold" proportion (Rossignol et al., 2003; Johnston and Burgstaller, 2019).

If mothers passed an identical mutant proportion onto each offspring, the buildup of mutations would eventually cause extinction (Muller, 1964). As a result, a developmental process has evolved to generate cell-to-cell variability in mutant proportion in animal germlines (Carling et al., 2011; Jokinen and Battersby, 2013; Stewart and Chinnery, 2015; Zhang et al., 2018)[1]. Thus, while some oocytes may receive higher mutant proportions, some will receive lower loads. Rather than all of a mother's oocytes having 50% mutant proportion, for example, they may range from 20 to 80% (**Figure 1A**). Oocytes with lower mutant proportions may then go on to become viable offspring, avoiding the buildup of mutation over generations. This increase in the oocyte-to-oocyte variance

---

[1]Other mechanisms, outside the scope of this article, exist to mitigate mtDNA mutation in other taxa (Johnston and Burgstaller, 2019).

FIGURE 1 | The "genetic bottleneck" increases cell-to-cell mutant proportion spread. **(A)** A mother's life begins as a single cell, with no associated variance in mutant proportion (white circles are wildtype mtDNAs, black circles are mutant mtDNAs; inset numbers give mutant proportion). Development increases cell-to-cell mutant proportion spread in the mother's developing oocytes. In the next generation, oocytes or offspring with lower mutant proportions may be favored. **(B)** Different experimental structures to investigate the generation of mutant proportion spread. (i) Comparing mutant proportion in a mother to her offspring. (ii) Comparing mutant proportion from a reference measurement to a set of individual oocytes. (iii) Comparing mutant proportion differences in a set of mother-child pairs. (iv) Recording mutant proportion differences across oocytes or siblings.

of mutant proportion is typically discussed as resulting from a "genetic bottleneck." Increasing cell-to-cell mtDNA variance has also been reported in somatic tissues, suggesting that the "genetic bottleneck" picture may also apply outside the germline (Sekiguchi et al., 2003; Wilton et al., 2018).

Oocyte-to-oocyte, and offspring-to-offspring, variance in mutant proportion is important in the fundamental biology of inheritance, and in human health and disease. While beneficial from an evolutionary perspective, this variance makes it hard to predict mtDNA inheritance patterns. As diseases result from high mutant proportions (Rossignol et al., 2003; Wallace and Chalkia, 2013), this unpredictability makes clinical planning difficult for families carrying dangerous mtDNA mutations (Poulton et al., 1998; Sallevelt et al., 2013). As such, substantial scientific effort is spent characterizing the processes that give rise to mtDNA variability.

The picture of the "genetic bottleneck" can be useful as a simple comparative statistic. However, experimental technology and mathematical theory has now advanced to the stage where we can ask (and begin to resolve) questions about the detailed physical mechanisms behind this genetic behavior. This article will attempt to compare the effective models and detailed mechanisms used to understand this important

process, and discuss how these vary through biology and in the scientific literature.

1.1. Terminology

The "genetic bottleneck" refers to a genetic quantity—an increase in cell-to-cell variability in mutant proportion. In humans and other animals, the genetic bottleneck is achieved in part (though likely not in full) by a "physical bottleneck" (described further below, and recently reviewed in Zhang et al., 2018). This "physical bottleneck" is a physical reduction in the copy number of mtDNA molecules per cell, which occurs during development. Because the word "bottleneck" appears in both terms, it is sometimes tempting to view the genetic and physical bottlenecks as equivalent. This is not generally the case. Unlike the physical bottleneck, the genetic bottleneck does not directly correspond to a observable number of molecules that can be directly measured by some experiment (Birky, 2001; Johnston and Jones, 2016). A genetic bottleneck of size 10, for example, does not mean that the physical copy number of mtDNAs per cell need ever be 10 at any point during development. As such, the term "mutant proportion spread," with less physical and more genetic implication, will be used here as a synonym for

Given a set of n heteroplasmy measurements $h_1, h_2, \ldots, h_n$, the *sample mean* $[h] = \frac{1}{n} \sum_{i=1}^{n} h_i$. Different ways exist to calculate sample variance. Typically, the "unbiased sample variance" is used, that is $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (h_i - sample\ mean\ [h])^2$. The "biased sample variance" is the mean squared difference from the mean $s_n^2 = \frac{1}{n} \sum_{i=1}^{n} (h_i - sample\ mean\ [h])^2$. The use of $n - 1$ rather than n, known as Bessel's correction, removes bias in the sample variance. As described in the text, studies calculate *sample var* $[h]$ using either s^2 (usually for **Figure 1B**i,ii,iv) or a mean squared difference approach more like s_n^2 (for **Figure 1B**iii).

In the literature, sample variances s^2 may also be found represented by $\mathbb{V}(h)$, $V(h)$, or σ^2 (but the latter is usually used for population variance). Sample means may be written $\bar{h}$, $\mathbb{E}(h)$, $\langle h \rangle$, μ (but the latter is usually used for population mean).

"genetic bottleneck." Note that a smaller "bottleneck" leads to more spread and vice versa. As described below, the "genetic bottleneck" (mutant proportion spread) may vary with species, individual, time, mtDNA sequence and other factors. The term "mutant proportion spread" perhaps captures this fluidity more than the more rigid "bottleneck."

We use "mutant proportion" rather than "heteroplasmy" because a heteroplasmy level over 50% is semantically difficult: the majority mtDNA type should then strictly be considered the reference type, and heteroplasmy redefined with respect to that type.

When taking biological observations and comparing them to models, *population* and *sample* statistics must be considered. Population statistics are summaries of a quantity—like the mean and variance—over the entire population of interest— for example, all oocytes in an organism. Sample measurements of statistics like mean and variance are those derived from a limited number of samples of a larger population. Experimental limitations usually mean that we must consider sample statistics—for example, a set of 20 oocytes from an organism. By contrast, quantitative models typically phrase their predictions in terms of population statistics. Accidents of sampling may lead to differences between sample measurements and population statistics.

When considering these statistics, different studies often use different symbols for the same quantity (**Box 1**). Here, we will attempt to make equations as verbally "readable" as possible. We write *sample var* $[h]$ for sample variances, *sample mean* $[h]$ for sample means, *var* $[h]$ for population variances and *mean* $[h]$ for population means. The sample quantities are computed as described in **Box 1**.

2. OBSERVATIONS

The fundamental observation that implies the existence of a "genetic bottleneck" (mutant proportion spread) is that offspring have different mutant proportions to their parents (**Figure 1A**). mutant proportions also differ from offspring to offspring. Therefore, at some point(s) between generations, variability in mutant proportion is induced. Parent-to-offspring differences

in mtDNA mutant proportion were first reported in cattle (Hauswirth and Laipis, 1982; Ashley et al., 1989; Koehler et al., 1991). Following this, experimental evidence for a "genetic bottleneck" (mutant proportion spread) has been found in animals from flies (Solignac et al., 1984), crickets (Rand and Harrison, 1986), mice (Wai et al., 2008; Burgstaller et al., 2018), salmon (Wolff et al., 2011), and penguins (Millar et al., 2008) to humans (Marchington et al., 1997; Rebolledo-Jaramillo et al., 2014; Li et al., 2016). Some examples of the variety of experimental bottleneck studies are compiled in **Figure 2**.

A mother starts her life as a single fertilized oocyte. As this is a single cell, there is no cell-to-cell variability in mutant proportion; there is only a single value. The oocytes that later develop in that mother, however, may vary substantially in mutant proportion. This suggests that the reason for offspring differences may be the induction of cell-to-cell mtDNA variability in germline development.

To compute the size of the "genetic bottleneck" (mutant proportion spread), we need a set of "before and after" measurements (**Figure 1B**). Often, the "before" measurement is taken from a mother. Different studies have different "after" observation structures. In animal models and some human experiments, sets of "after" observations are obtained: for example, measurements across a set of offspring (**Figure 1B**i), or a set of single-cell oocyte measurements (**Figure 1B**ii). Developmental studies aimed at identifying mechanisms rather than "bottleneck size" may take samples of oocytes or their precursors at different stages of development. In other experiments, particularly in human population genetics, a single "after" observation is taken: for example, a single offspring (**Figure 1B**iii). Many before-after pairs are then used to characterize the population. When a "before" observation is not available, mutant proportion spread may be characterized from "after" measurements and some estimate of the "before" state is constructed (**Figure 1B**iv). This estimate is often the sample mean of the "after" measurements, thus assuming that no selective shift has occurred.

The mutant proportion variability for a system is typically reported as the sampled variance across a set of "after" observations *sample var* $[h]$ (**Figure 1B**). Most models describing mtDNA statistics (see below) predict that the population variance will follow the form:

$$var\ [h] = h_0(1 - h_0) \times \ldots, \qquad (1)$$

where $\ldots$ is some expression that may vary according to the model, and h_0 is the mutant proportion in the initial "before" population from which sampling takes place (not the new "after" population). In other words, most models predict cell-to-cell mutant proportion variance to depend on initial mutant proportion h_0, and specifically to be proportional to $h_0(1 - h_0)$.

Because most models have the above form, we often work with a quantity which we here call "mutant proportion spread" but which is usually called "normalized heteroplasmy variance":

$$sample\ var'\ [h] = \frac{sample\ var\ [h]}{h_0(1 - h_0)}. \qquad (2)$$

Reference	Organism	Intergenerational / developmental?	Tissue	Calculation	sample var' [h] before technical uncertainty
Solignac et al. 1984	Fly	Intergenerational	Oocytes	From V(h)	0.054 to 0.17
Rand et al. 1986	Cricket	Intergenerational	Organism	From V(h)	0.025 to 0.11
Ashley et al. 1989	Cattle	Intergenerational	Brain/liver	From V(h)	0.41
Howell et al. 1992	Human (14560)	Intergenerational	Blood	From V(h)	0.11
Jenuth et al. 1996	Mouse	Intergenerational	Tail	From V(h)	0.033 to 0.18
Bendall et al. 1996	Human	Intergenerational	Blood	Bayesian fit with binomial model	Variable by individual, interpreted as 0.071 to 0.33
Millar et al. 2008	Penguin	Intergenerational	Blood	MSD	0.032
Wolff et al. 2011	Salmon	Intergenerational	Fin	From V(h)	0.0059 to 0.018
Monnot et al. 2011	Human (3243)	Intergenerational	Blood (also PGD samples)	From V(h)	0.27
Guo et al. 2013	Human	Intergenerational	Blood	Correlation approach	Interpreted as 0.005
Rebolledo-Jaramillo et al. 2014	Human	Intergenerational	Blood, cheek	MSD	0.029
Pallotti et al. 2014	Human (3243)	Intergenerational	Several	From V(h)	Interpreted as 0.24 to 0.25
Li et al. 2016	Human	Intergenerational	Blood	MSD	Variable by locus, around 0.11
Wilson et al. 2016	Human	Intergenerational	Blood meta-analysis	Beta	Variable by locus, incl modes around 0.4 for 8993 and 0.1 for 3243
Otten et al. 2018	Human	Intergenerational	PGD samples	Kimura fit	Variable by locus, incl 0.79 for 8993, 0.18 for 3243
Jenuth et al. 1996	Mouse	Developmental	Single oocytes	From V(h)	Variable by stage, 2.4e-7 to 0.0019 PGCs; 0.017 to 0.053 primary oocytes; 0.013 to 0.21 mature oocytes
Marchington et al. 1997 / Poulton et al. 1998	Human	Developmental	Single oocytes	From V(h)	Interpreted as 0.11
Brown et al. 2001	Human	Developmental	Single oocytes	V(h) w/o h0	0.13
Wai et al. 2008	Mouse	Developmental	Single oocytes	V(h) w/o h0	Variable by time, 0.0015 to 0.0022 4 dpc; 0.013 to 0.043 8+ dpc
Wolff et al. 2011	Salmon	Developmental	Single oocytes	From V(h)	0.007 to 0.014
Lee et al. 2012	Monkey	Developmental	Blastomeres	V(h) w/o exact h0	Variable by division, 0.031 2-cell, 0.065 4-cell, 0.095 8-cell
Burgstaller et al. 2018	Mouse	Developmental	Oocytes (also offspring)	From V(h)	Variable by age, around 0.01 for young mice and 0.12 for old mice

FIGURE 2 | Mutant proportion spreads observed in different systems. Some examples of the diverse mutant proportion spread *sample var'* [h] observed experimentally. Loci in brackets refer to specific human mtDNA mutations; PGC, primordial germ cell.

The reason for working with *sample var'* [h] is that its normalized value does not typically depend on the specific initial mutant proportion values h_0 from one particular experiment. The results from different experiments, with different values of h_0, can then be more naturally compared.

These variability observations are typically studied from two different perspectives. First, at the "statistical" level: what is the distribution of mutant proportions that will arise from a given mother? This perspective often uses "genetic bottleneck size" as a single number that reflects the observed sample-to-sample mutant proportion spread. Second, at the "mechanistic" level: what physical mechanisms give rise to this distribution of mutant proportions? This perspective attempts to link the coarse-grained outcome of the "genetic bottleneck" to specific, measurable physical rates and properties. In this article, we will first discuss concepts related to this first perspective, before surveying recent progress on the second.

3. THE "GENETIC BOTTLENECK" ABSTRACTED AS SAMPLING EVENTS OR DRIFT

3.1. Abstracting the "Genetic Bottleneck" as a Single Sampling Event

For convenience, studies often describe the "genetic bottleneck" (mutant proportion spread) as the result of a single abrupt event that creates many new individuals, with different mutant

FIGURE 3 | Constructing new populations from random sampling of an initial population. **(A)** Sampling an initial population to construct many new populations of size N. Smaller N generates more variability between the new populations. **(B)** Reamplification of sampled populations to the original size can be deterministic (left, preserving mutant proportion) or stochastic (right, changing mutant proportion). **(C)** Structures of several distributions related to the study of the "genetic bottleneck" (mutant proportion spread), parameterized by effective "bottleneck size" N.

proportions, from an initial individual (**Figure 3A**). In this case, the resulting "bottleneck size" is simply a readout of mutant proportion spread, and does not directly correspond to any physical observable. In particular, it is not generally equal to the minimum copy number of mtDNA molecules (the "physical bottleneck") (Birky, 2001; Jokinen and Battersby, 2013; Johnston and Jones, 2016; Zhang et al., 2018). This is because the "genetic bottleneck" folds together all mechanisms that can influence mutant proportion spread—the physical bottleneck, cell divisions, random mtDNA dynamics, and so on. The specific number associated with "genetic bottleneck size" may therefore be substantially lower than the physical bottleneck during development.

The goal in this perspective is typically to characterize the "genetic bottleneck" (mutant proportion spread) under different conditions. These may involve, for example, different genetic features, different populations, and different species. Knowledge of the value associated with the "genetic bottleneck" (mutant proportion spread) in these cases can inform fundamental biology and clinical planning (Sallevelt et al., 2013).

The concept underlying this approach is a model of "random sampling." Here, we start with an initial population of mtDNA, with mutant proportion h_0. To create one instance of a final population—for example, the population in one oocyte in the next generation—we randomly sample that initial population. Specifically, we pick at random one member of the initial population and put an mtDNA molecule of that type in our final population. If we are sampling "with replacement," we retain the picked member in the original population. The alternative is sampling "without replacement," which involves removing the picked member from the source population so it cannot be picked again.

If we use N picks with replacement to construct one new population, and another N picks with replacement to construct a second new population, and so on, the new populations will likely differ (**Figure 3A**). This is because we are likely to choose

different numbers of each mtDNA type when we are constructing the new populations.

Quite how different the new populations will be depends on N, the number of picks. If we just pick $N = 1$ mtDNA from our source population for each new population, different new populations may differ substantially: each will contain only one mtDNA type, so some populations will have a mutant proportion of 0 and some a mutant proportion of 1. By contrast, if N is high, we draw many samples from our initial population, and are likely to end up with new populations that look rather like the initial one (with mutant proportions close to h_0). We can immediately see that our mutant proportion spread (genetic bottleneck) will decrease as N decreases (**Figure 3A**).

This process is called binomial sampling. The actual variance between our new populations is well-known from theory, and is

$$var\left[h\right] = \frac{h_0(1 - h_0)}{N}. \qquad (3)$$

A common picture of the "genetic bottleneck" is exactly this N. That is, if we observe a certain mutant proportion spread across cells or samples, we work out how large or small N would have to be to generate that amount of spread through this binomial sampling, and call this number the "genetic bottleneck."

How we estimate N depends on the structure of our experiment (**Figure 1B**). First consider the case where we have a single "before" measurement and a set of "after" measurements (for example, a mother mutant proportion and a set of offspring (**Figure 1B**i) or oocyte (**Figure 1B**ii) mutant proportions). Take the sample variance $sample\ var\left[h\right]$ of the "after" measurements. Call the "before" measurement h_0. Then the definition of "bottleneck size" is often taken to be

$$N = \frac{h_0(1 - h_0)}{sample\ var\left[h\right]} = \frac{1}{sample\ var'\left[h\right]} \qquad (4)$$

based on this binomial sampling picture. If h_0 is not known, as in **Figure 1B**iv, it is sometimes estimated to be equal to

FIGURE 4 | Estimating the "genetic bottleneck" from different data sources. **(A)** Simulated data, comparing estimates of "bottleneck size" N using either 16 "after" measurements ("Variance") or 16 sets of single "after" measurements ("MSD"). Estimates are performed in the presence of different levels of selection ($s = 1$, no selection; decreasing s is increasing selective pressure). **(B)** Uncertainty in a "bottleneck size" estimate using $mean\,[h] = 0.5$, $var\,[h] = 0.01$, and different numbers of samples n. Particularly for $n < 10$, bottleneck size estimates can have large uncertainty.

sample mean $[h]$. That is, the assumption is made that no shift in mutant proportion has taken place due to selection or accidents of sampling.

The idea here is to convert a less intuitive quantity *sample var'* $[h]$ into a more intuitive one (an effective number of segregating units). However, this binomial sampling picture has some issues. First, it does not correspond to a plausible biological mechanism. Development does not involve a single, abrupt sampling event. How reamplification of mtDNA back to its original level takes place is rarely considered (**Figure 3B**), although models for reamplification do exist (see below). Second, and related, a binomial sampling regime predicts a binomial distribution for final mutant proportion (**Figure 3C**). For a small value of N, this means that mutant proportion can only take one of a restricted set of values. For example, if N is 4, we would only expect mutant proportions of 0, 25, 50, 75, and 100% after sampling. Other models have been proposed to address these shortcomings (see below).

Next, consider the case where we have a set of paired "before" and "after" observations (for example, the mother-single offspring pairs in **Figure 1B**iii). The prevailing approach to calculate a "bottleneck size" here is via an expression derived in references (Millar et al., 2008; Hendy et al., 2009) based on
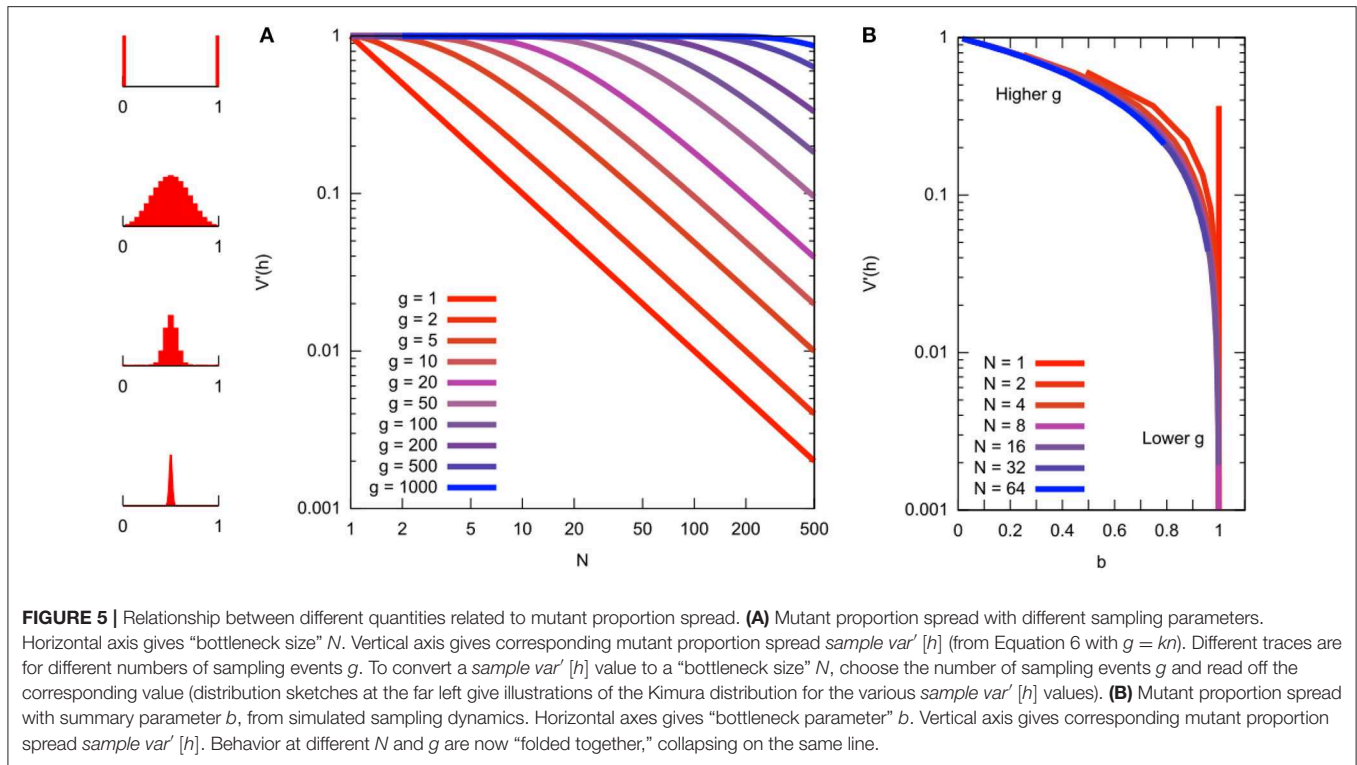
$$N = \frac{h_0(1 - h_0)}{sample\,mean\,\left[(h - h_0)^2\right]}. \tag{5}$$

Here, the spread of "after" measurements *sample var* $[h]$ $=$ $\frac{1}{n-1}\sum_i(h_i - sample\,mean\,[h])^2$ has been replaced by the mean square difference between the "before" and "after" measurements

$\frac{1}{n}\sum_i(h_i - h_0)^2$. That is, the approach assumes that the average "after" measurement *sample mean* $[h]$ is equal to the "before" measurement h_0—in other words, that no shifts in mean mutant proportion act between generations. If selection is in fact present, Equations (4) and (5) give different results, and Equation (5) quickly fails to capture the true bottleneck size even in abstracted systems (**Figure 4**). Because Equation (5) deals with squared differences, selective shifts in different directions do not "cancel out" but rather reinforce the resultant discrepancy.

3.2. The "Genetic Bottleneck" Abstracted as Several Sampling Events or Drift

A single binomial sampling event does not represent a real biological mechanism. To improve this picture, some studies consider the "genetic bottleneck" (mutant proportion spread) as arising from a series of sampling events, modeling cell divisions that randomly partition mtDNA molecules between cells. Early work on mtDNA inheritance (Solignac et al., 1984; Rand and Harrison, 1986; Ashley et al., 1989; Howell et al., 1992) drew on a classical result from Sewall Wright (Wright, 1942, 1984) to this end. This result describes the spread of allele frequencies due to "accidents of sampling" in repeated generations, where the individuals in one generation are a random sample from the previous generation. For mtDNA, this "Wright equation" predicts

$$var\,[h] = h_0(1 - h_0)\left(1 - \left(1 - \frac{1}{N}\right)^{kn}\right), \tag{6}$$

FIGURE 5 | Relationship between different quantities related to mutant proportion spread. **(A)** Mutant proportion spread with different sampling parameters. Horizontal axis gives "bottleneck size" N. Vertical axis gives corresponding mutant proportion spread *sample var'* $[h]$ (from Equation 6 with $g = kn$). Different traces are for different numbers of sampling events g. To convert a *sample var'* $[h]$ value to a "bottleneck size" N, choose the number of sampling events g and read off the corresponding value (distribution sketches at the far left give illustrations of the Kimura distribution for the various *sample var'* $[h]$ values). **(B)** Mutant proportion spread with summary parameter b, from simulated sampling dynamics. Horizontal axes gives "bottleneck parameter" b. Vertical axis gives corresponding mutant proportion spread *sample var'* $[h]$. Behavior at different N and g are now "folded together," collapsing on the same line.

where k is the number of random samplings per generation (for example, the number of cell divisions in germ line formation) and n is the number of generations. The reader will notice that if $kn = 1$, describing a single sampling event as above, Equation (4) is recovered.

For convenience, some studies have since defined new "bottleneck parameters" to simplify this expression. One choice is to set $\alpha = (1 - 1/N)^k$. Another more recent alternative is to define $b = \exp(-g/N)$. Here $g = kn$ represents an amalgamated number of samplings, and the exponential form is used for algebraic convenience because $\exp(-g/N) \simeq (1-1/N)^g$. In these cases, Equation 6 becomes:

$$var\left[h\right] = h_0(1 - h_0)\left(1 - \alpha^n\right) \simeq h_0(1 - h_0)\left(1 - b\right). \quad (7)$$

The advantage of using these "bottleneck parameters" is that they fold together two unknown quantities: the number of generations and the effective population size. Under Equation (6), readouts of "bottleneck size" (mutant proportion spread) using N are contingent on a particular choice of g, the number of generations for which the bottleneck applies (**Figure 5A**). Readouts using b (and α) absorb this dependency, providing a simple readout of mutant proportion spread that makes no assumptions about the number of sampling events (**Figure 5B**). Given a number of generations, N can be recovered from b via $N = -g/\ln b$.

An approximation of the cell-to-cell distribution of mutant proportion under these repeated-sampling models is the so-called Kimura distribution (Wonnapinij et al., 2008) (**Figure 3C**). Strictly, the assumptions involved in deriving this approximation rely on N being large (Kimura, 1955). However, the Kimura distribution reproduces intuitive behavior for the distribution of mutant proportion under drift, and several studies use a fit to the Kimura distribution to estimate b (Wonnapinij et al., 2008; Otten et al., 2018).

The relationship between these quantities N, b, g, and *sample var'* $\left[h\right]$ is illustrated in **Figure 5**, which may serve as a reference for comparison of reported "genetic bottleneck" (mutant proportion spread) statistics in different studies.

3.3. Drift Manifest Through Random Replication

One issue with a simple sampling picture is that it predicts a set of cellular mtDNA populations consisting of N molecules. In most circumstances, this N value is much lower than the size of typical cellular populations. For example, in animal germline development, the number of mtDNA molecules per cell is amplified several orders of magnitude from a minimum copy number back to a functional level (Cree et al., 2008; Wai et al., 2008; Cotterill et al., 2013; Zhang et al., 2018). If this reamplification happened perfectly deterministically, no further change in $var\left[h\right]$ would occur. However, cell biology is rarely deterministic, and there is good reason to believe that this reamplification process involves a random component (Birky, 1994; Chinnery and Samuels, 1999; Capps et al., 2003; Johnston et al., 2015).

Several recent studies have considered models for this reamplification (Johnston and Jones, 2015; Wilson et al., 2016). Most are based on the idea of random mtDNA replication. That is, an mtDNA molecule is randomly chosen from the current

population, and replicated. Then a new mtDNA molecule is randomly chosen and replicated, and so on until the desired population size is achieved. This is a modified Moran model (Moran, 1958) (the usual Moran model involves removing one molecule per replication, so that overall population size remains constant), also known as a Pólya urn model (Eggenberger and Pólya, 1923; Johnson and Kotz, 1977).

In the limit of infinite reamplification, the model gives rise to a beta distribution for mutant proportion spread (**Figure 3C**). Infinite reamplification may not seem realistic, but actually the structure of this distribution is quickly stabilized after a relatively small number of replications, so the simple infinite limit is similar to more reasonable cases. However, results also exist for intermediate cases, and their exploration may be a fruitful area of future research. The beta distribution takes two parameters, α and β, intuitively corresponding to the number of mutant and wildtype molecules in the cell before any replication. If we $\alpha = h_0 N$ and $\beta = (1 - h_0)N$, the mean of the beta distribution is $mean\left[h\right] = h_0$ as expected, and the variance of the beta distribution is

$$var\left[h\right] = \frac{h_0(1 - h_0)}{N + 1}. \tag{8}$$

3.4. Uncertainty

The "genetic bottleneck" (mutant proportion spread) is a readout of variance. Revealing trends in cell-to-cell variance is more challenging than revealing trends in average behavior, and requires more data. Wonnapinij et al. (2010) have drawn attention to the challenging nature of obtaining reliable estimates of mutant proportion spread. Uncertainty in estimated mutant proportion spread is often large, challenging precise estimates of the "bottleneck size" and leading to variability in these estimates. Even in the case of no technical error (see below), sampling errors can lead to large variability in estimates of "bottleneck size," particularly if fewer than 10 samples are used (**Figure 4B**).

Uncertainty in readouts of variance can be an unintuitive quantity. We are perhaps more used to thinking about mean values as the quantity of interest, with variance around a mean value corresponding to uncertainty. However, we can—and should—also describe and estimate the uncertainty associated with an observation of variance.

One way to estimate uncertainty in $sample\ var\left[h\right]$ involves assuming that mutant proportion samples are drawn from a normal distribution. This is not generally the case (as seen in **Figure 3C**), but is a simple illustration that may be applied when spread is low. Confusingly, there are two expressions in circulation for the sampling error in this case. Which of these values gets used depends on how the variance was computed. If the variance is calculated using an estimate of the mean taken from the same dataset (employing Bessel's correction, as with many "after" measurements), Wonnapinij et al. (2010) cite:

$$SE\left[sample\ var\left[h\right]\right] = var\left[h\right] \times \sqrt{\frac{2}{n - 1}}, \tag{9}$$

for the standard error in $sample\ var\left[h\right]$, where n is the number of samples used to characterize $sample\ var\left[h\right]$. If the mean is estimated from a different source (omitting Bessel's correction, as with mean-squared-difference calculations using a single "after" measurement), an estimate of the variance of the sample variance is $(var\left[h\right])^2(n-1)/n^2$, as quoted in reference (Millar et al., 2008), corresponding to a standard error of $var\left[h\right] \times \sqrt{(n-1)/n^2}$. The standard error associated with a variance measurement can then be estimated by using $var\left[h\right] \simeq sample\ var\left[h\right]$ in these expressions.

However, for wide spreads or means close to 0 or 1, mutant proportion distributions do not have normal structure. In this case (Wonnapinij et al., 2010), cite a more general result:

$$SE\left[sample\ var\left[h\right]\right] = \sqrt{\frac{1}{n}\left(D_4 - (var\left[h\right])^2 \times \left(\frac{n-3}{n-1}\right)\right)} \tag{10}$$

where $D_4 = (n-1)/n^3 \times ((n^2 - 3n + 3)\mu_4 + 3(2n - 3)\mu_2^2)$, and $\mu_2 = 1/n\sum_{i=1}^{n}(h_i - h_0)^2$, $\mu_4 = 1/n\sum_{i=1}^{n}(h_i - h_0)^4$. While more complicated in structure, all these quantities can readily be worked out from the set of observed mutant proportion measurements.

All these expressions have the standard error of $sample\ var\left[h\right]$ scale roughly with the observed value divided by $\sqrt{n}$. Thus, unless a large number n of samples are used to characterize mutant proportion spread, the associated uncertainty in $sample\ var\left[h\right]$ can be rather high. As "bottleneck size" estimates depend on $1/sample\ var\left[h\right]$, the corresponding uncertainty can be enormous for low sample sizes (**Figure 4B**).

These expressions are based on the statistics of sampled variances, and assume that the sample mutant proportion values themselves have no associated uncertainty (in other words, there is no technical error associated with the genetic measurement). Technical error should also be included in the uncertainty associated with these estimates. Several studies include considerations of technical error in their estimates of mtDNA statistics (Bendall et al., 1996; Millar et al., 2008; Li et al., 2016; Wilson et al., 2016). This is typically achieved through simple uncertainty propagation, that is, considering an observed variance to be a combination of natural variance and technical variance. The technical variance may either be quantified through experimental calibration (Millar et al., 2008) or as part of a statistical inference process (Bendall et al., 1996; Li et al., 2016; Wilson et al., 2016).

3.5. Results

Early reports of the size of the genetic bottleneck (mutant proportion spread) varied substantially across organisms. Contributing to this variability was the fact that different studies used different values of kn in Equation (6). These different values reflected, for example, estimates of the number of cell divisions involved in germline development in different species. More recently, it has become more common to set $kn = 1$ and assume a single binomial sampling event, or to use a "bottleneck parameter," usually b, to summarize mutant proportion spread as above. **Figure 2** summarizes the mutant proportion spreads observed in several key experimental studies across species.

The rapid intergenerational shifts observed in cattle (Hauswirth and Laipis, 1982; Koehler et al., 1991) have given rise to the highest mutant proportion spread values so far observed. Insects appear to have lower mutant proportion spreads (Solignac et al., 1984; Rand and Harrison, 1986). In mice, several experiments have observed the increase of mutant proportion spread through germline development (Jenuth et al., 1996; Wai et al., 2008). Fish show similar behavior (Wolff et al., 2011).

Mutant proportion spread in humans was observed some time ago (Bendall et al., 1996; Marchington et al., 1997), but its magnitude remains debated. Variability in the behavior of mutant proportion spread was quickly apparent. Blok et al. found dramatic skew toward extreme mutant proportions in transmission of the 8993 mutation (Blok et al., 1997). Lutz et al. (2000) found evidence for variable mutant proportion spread in a human family; while they did not provide quantitative estimates they noted that the different spreads they observed suggest a varying "bottleneck size" which could be very small. Bendall et al. (1996) used a Bayesian approach to show that it was unlikely that their study families had the same "bottleneck size." More recently, two large-scale population-genetic studies suggest rather different "bottleneck sizes" (Rebolledo-Jaramillo et al., 2014; Li et al., 2016). Pathogenic mutations seem to involve more mutant proportion spread, particularly the 8993 mutation (Blok et al., 1997; Monnot et al., 2011; Wilson et al., 2016; Otten et al., 2018). Ongoing preimplantation genetic diagnoses approaches continue to provide data on mutant proportion spread at different developmental stages (Monnot et al., 2011; Treff et al., 2012; Sallevelt et al., 2013). Pallotti et al. (2014) performed a meta-analysis of 3243 bottlenecks along with their own experiments and found reasonable consistency in mutant proportion spread. Notably, different studies still use different protocols for reporting a "bottleneck size," sometimes setting $g(= kn) = 24$ or $g(= kn) = 1$ in Equation (6).

While not a focus of this article, we note that genetic bottlenecks (increasing mutant proportion spread) (Sekiguchi et al., 2003; Wilton et al., 2018) and physical bottlenecks (Cao et al., 2007; Otten et al., 2016; Floros et al., 2018) have also been reported in somatic tissues.

4. THE "GENETIC BOTTLENECK" AS A SET OF PHYSICAL PROCESSES

In parallel with statistical characterization of the "genetic bottleneck" (mutant proportion spread), related research attempts to understand the physical processes that give rise to an observed "genetic bottleneck" (mutant proportion spread) in a given system. The goal here is typically to identify biological mechanisms and potential targets for intervention.

A plausible physical mechanism for the "genetic bottleneck" (mutant proportion spread) must account for both physical and genetic observations over time during development. The physical observations involve mtDNA copy number per cell and the occurrence of cell divisions; the genetic observations involve cell-to-cell variability in mutant proportion. An example from a meta-analysis of mouse observations is shown in **Figures 6A,B**. The joint prediction of these physical and genetic observations is very important because it constrains the mechanisms that are possible—for example, the size of the physical bottleneck, the timing of cell divisions, and the rate of reamplification all influence the resulting genetic statistics of mtDNA populations.

While not a focus of this article, the specific genetic players behind the physical processes below are increasingly being revealed, and have been reviewed in, for example, references (Carling et al., 2011; Jokinen and Battersby, 2013).
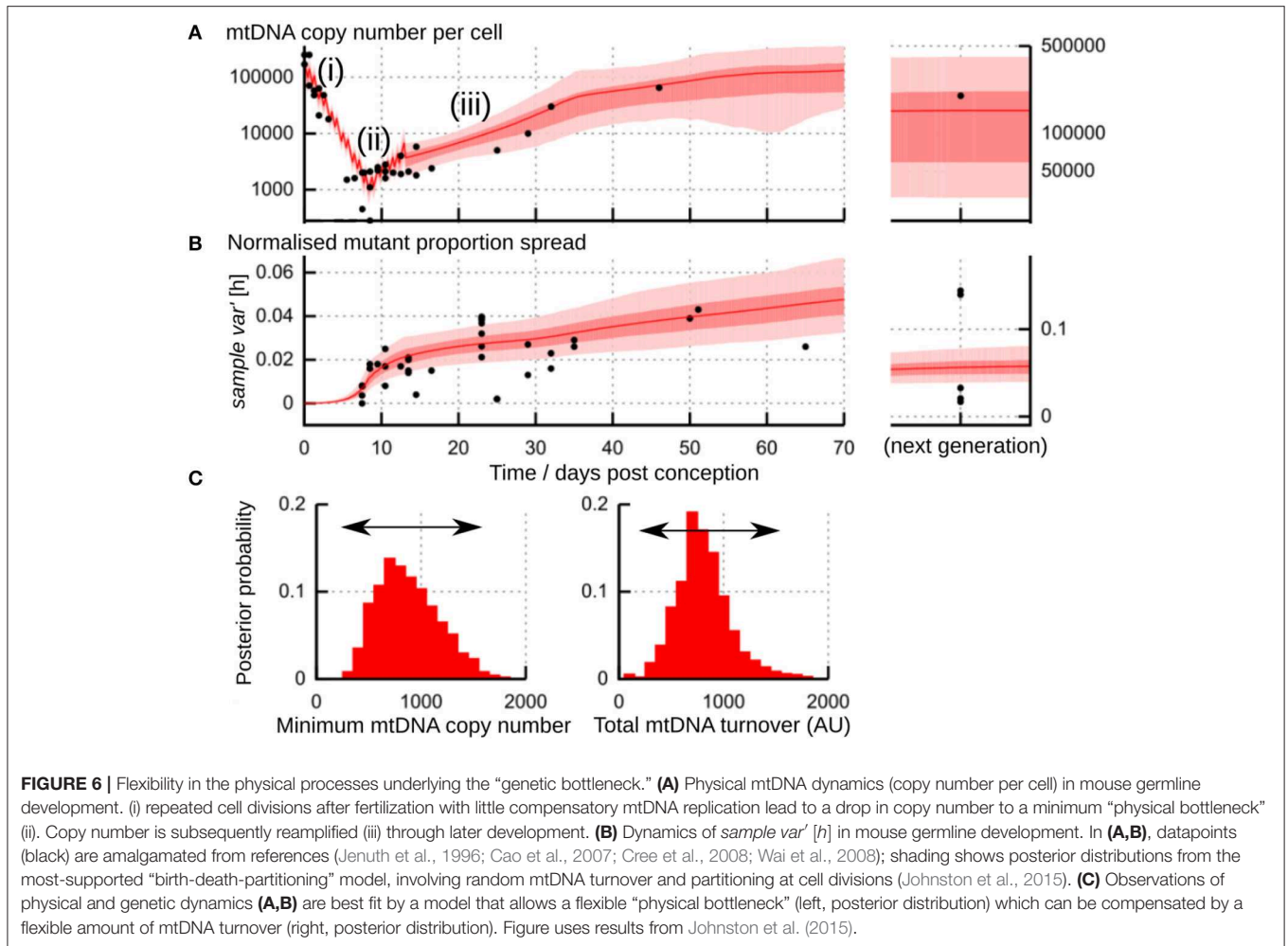
4.1. The Physical Bottleneck During Development

One process that occurs during germline development in animals is a physical reduction in the number of mtDNA molecules per cell (Zhang et al., 2018). This reduction is observed in animals including mice (Cao et al., 2007; Cree et al., 2008; Wai et al., 2008), fish (Wolff et al., 2011; Otten et al., 2016), sheep (Cotterill et al., 2013), and humans (Floros et al., 2018). For some time after fertilization, cell divisions repeatedly halve the cellular mtDNA population, with little compensatory replication. This halving leads to a pronounced drop in mtDNA copy number per cell (**Figure 6A**i). A fertilized oocyte typically contains many mtDNA molecules [hundreds of thousands in mice (Cree et al., 2008; Wai et al., 2008); around a million in humans (Floros et al., 2018)]. The size of the physical bottleneck—that is, the lowest copy number of mtDNA per cell during development—remains debated, but is often orders of magnitude lower; Zhang et al. (2018) have recently provided a survey of mtDNA reduction in different species. In mice, the lowest copy number may lie between 200 and 1,000 (Cao et al., 2007, 2009; Cree et al., 2008; Wai et al., 2008; Johnston et al., 2015) (**Figure 6A**ii). In humans, mean copy numbers around 1400 are observed in progenitor germ cells (Floros et al., 2018). In zebrafish, decreases from tens of millions to hundreds of mtDNAs per cell are observed (Otten et al., 2016). The copy number of mtDNA during development seems to depend on genetic characteristics of the mtDNA (Monnot et al., 2013), potentially making the physical bottleneck sequence-dependent.

Pictured as drawing a random selection of mtDNA molecules from a larger population, copy number reduction provides a way to generate variability between cells. Additionally, the magnitude of variability generated through other random sampling processes is amplified by low copy numbers.

4.2. Random Replication of a Subset of mtDNA Molecules

In this mechanism, at some point(s) in germline development, a random subset of a cell's mtDNA population is allowed to replicate, while all others are eventually subject to degradation or loss (Wai et al., 2008). This subset may be, for example, those mtDNAs within a certain distance of the nucleus (Wallace, 2018). As the random subset chosen will differ in different cells, this process imposes a natural sampling inducing variance between cells. A smaller subset of molecules will lead to more mutant proportion spread.

FIGURE 6 | Flexibility in the physical processes underlying the "genetic bottleneck." **(A)** Physical mtDNA dynamics (copy number per cell) in mouse germline development. (i) repeated cell divisions after fertilization with little compensatory mtDNA replication lead to a drop in copy number to a minimum "physical bottleneck" (ii). Copy number is subsequently reamplified (iii) through later development. **(B)** Dynamics of *sample var'* [*h*] in mouse germline development. In **(A,B)**, datapoints (black) are amalgamated from references (Jenuth et al., 1996; Cao et al., 2007; Cree et al., 2008; Wai et al., 2008); shading shows posterior distributions from the most-supported "birth-death-partitioning" model, involving random mtDNA turnover and partitioning at cell divisions (Johnston et al., 2015). **(C)** Observations of physical and genetic dynamics **(A,B)** are best fit by a model that allows a flexible "physical bottleneck" (left, posterior distribution) which can be compensated by a flexible amount of mtDNA turnover (right, posterior distribution). Figure uses results from Johnston et al. (2015).

Wai et al. observed a sharp increase in mutant proportion spread in mice aged between 4 and 8 days (as in **Figure 6B**) (Wai et al., 2008; Samuels et al., 2010). Using microscopy, they showed that only a subset of mtDNA molecules was involved in replication at a given time. They propose this subset replication model as the mechanism by which variability is generated at this development stage (folliculogenesis, **Figure 6A**iii). Johnston et al. (2015) suggest that this observation is also compatible with random mtDNA turnover (see below), where a non-fixed subset of mtDNAs is expected to be involved in replication at any given time.

The random replication model described above, connected to the beta distribution, can describe the dynamics of the subset-replication model. Care must be taken here to ensure that physical copy number dynamics are reproduced: for example, the small amount of replicating mtDNAs must balance the large number of degrading mtDNAs as copy number is amplified (**Figure 6A**iii).

4.3. Random Partitioning of mtDNA Molecules at Cell Divisions

This mechanism is possible during specific times when cells are undergoing divisions. At division, a "parent" cell distributes its population of mtDNA to its two "daughter" cells. The assignment of each mtDNA molecule to one or the other daughter may follow a random process (Birky, 2001; Huh and Paulsson, 2011; Johnston et al., 2012). In this case, each division will increase the cell-to-cell mutant proportion variability between daughter cells. If mtDNA molecules are partitioned in clusters, this increase will be faster (Cao et al., 2007). Larger clusters will lead to more mutant proportion spread.

Whether the "unit of inheritance" of mtDNA is a single molecule or a cluster is a debated question. MtDNA within mitochondria is packaged into complexes called nucleoids. These were thought to contain around 5–10 mtDNA genomes (Jacobs et al., 2000; Cao et al., 2007; Khrapko, 2008), suggesting that clusters of mtDNA may be the natural state. However, evidence from microscopy suggests that nucleoids may only contain around 1 mtDNA genome (Kukat et al., 2011). Model selection for mouse germline development (Johnston et al., 2015) and human transmission (Li et al., 2016) both suggest that single mtDNA molecules are the unit of inheritance.

Observations in rhesus monkeys (Lee et al., 2012) showed a dramatic induction of variance by the 8-cell stage, presumably due to random partitioning of mtDNAs over the first three cell divisions. In this study, mtDNA admixtures in oocytes were

created by fusing two cytoplasm halves from different oocytes. Cell divisions then immediately followed the construction of these admixed oocytes. It is thus not inconceivable that physical heterogeneity in the distribution of mtDNA molecules, remaining from the cytoplasm fusion, may contribute to this high mutant proportion spread. For example, if the fusion process created a "north hemisphere" containing exclusively one mtDNA type and a "south hemisphere" containing exclusively the other, and the first cell division occurred along the "equator," the resultant cells would then immediately have maximum mutant proportion differences. Natural systems may be expected to have more physically mixed mtDNA populations, and so potentially show less extreme mutant proportion spreads in these early stages.

The "repeated sampling" approaches above attempt to model cell divisions during development as a series of random binomial samples. Partitioning dynamics can also be embedded in stochastic models of mtDNA replication and degradation (Johnston and Jones, 2015, 2016). Much of this work assumes binomial partitioning; however, recent work in yeast has suggested that partitioning of mtDNA is tighter than binomial sampling (Jajoo et al., 2016). Mathematical results do exist for more controlled partitioning, or the partitioning of clusters of mtDNA (Johnston and Jones, 2015) but are often complicated, so simulation is often used to make quantitative predictions in these cases (Johnston et al., 2015; Li et al., 2016).

4.4. Random Turnover of mtDNA Molecules

MtDNA replicates and degrades quasi-independently of the cell cycle. The noisy environment of the cell means that these processes have a random component (Birky, 1994; Chinnery and Samuels, 1999; Capps et al., 2003; Johnston et al., 2015). The ongoing action of this random turnover creates cell-to-cell mutant proportion variability. For example, two cells that start with identical mtDNA populations will diverge over time, as different molecules undergo replication and degradation. Faster turnover, or turnover of clusters, will lead to more mutant proportion spread.

To account for the full set of processes that an individual mtDNA molecule may undergo, several stochastic modeling approaches have been developed (reviewed in Hoitzing et al., 2017). These approaches model every individual mtDNA molecule in a cell and subjects them to the physical processes that we may expect to occur during development. Typically, these processes will have a random component, so that if the model is simulated twice, the precise outcomes will differ. These differences can be used to characterize the variability supported by different mechanisms.

A well-known model involves "relaxed replication," that is, replication of mtDNA independent of the cell cycle (Birky, 1994). Models of this process typically involve mtDNA molecules degrading with a fixed rate, and replicating randomly with a rate that depends on population size (Chinnery and Samuels, 1999; Capps et al., 2003). This model generates variability over time because of these random dynamics. Cree et al. propose this mechanism, amplified by the physical bottleneck,

to generate mutant proportion spread in mouse development (Cree et al., 2008).

More recently, the different ways that the cell could control this replication rate have recently been explored in detail using "birth-death" models (Johnston et al., 2015; Johnston and Jones, 2016; Hoitzing et al., 2019). Strikingly, this work showed that no matter how the cell controls mtDNA replication, if there is some mutant proportion, the variance of this mutant proportion will increase linearly over time.

Specifically, in a population of N mtDNAs, random turnover of molecules with rate β over time t gives rise to the behavior

$$sample\ var'\left[h\right] = \frac{2f\beta t}{N},\tag{11}$$

so that, for example, a year of mtDNA turnover, with average rate one degradation event per week, in a cell with 1,000 mtDNA molecules would give a mutant proportion spread of $(2 \times 52)/1,000 = 0.104$. This would be interpreted as a "bottleneck size" around 9.6. In followup theoretical developments (Aryaman et al., 2019), the factor f in Equation (11) has been shown to be the fraction of unfused mitochondria, that is, mitochondria containing mtDNAs subject to mitophagy (Youle and Narendra, 2011; Diot et al., 2016). Mitochondrial quality control, linked to fission-fusion dynamics, contributes to the turnover of mitochondria in the cell (Twig et al., 2008) and provides one way that mitochondrial dynamics may influence both mean and variance dynamics of mtDNA populations (Hoitzing et al., 2015; Johnston, 2018; Latorre-Pellicer et al., 2019). Higher rates of quality control related turnover can result in higher cell-to-cell mutant proportion variance (Johnston et al., 2015) [and, if mitochondria associated with one mtDNA type are preferentially degraded, this selective pressure will also influence mean mutant proportions (Twig et al., 2008; Hoitzing et al., 2015)]. Equation (11) provides a coupling between the physical fission-fusion dynamics of mitochondria and the time behavior of mtDNA mutant proportion spread (Hoitzing et al., 2015; Johnston, 2018; Aryaman et al., 2019).

4.5. Combinations of Mechanisms

Several of these processes are conceptually linked. For example, when a cell divides, it loses around half of its mtDNA content, immediately restricting the subset of mtDNAs that are available for replication. If mtDNA molecules are involved in ongoing random turnover, only a subset of molecules will be replicating at any given time (Johnston et al., 2015).

In each of these cases, a smaller mtDNA population acts to amplify increases in mutant load spread, because the influence of random events is less "smoothed out" in small populations. Therefore, we can end up with the same amount of spread by either (i) generating a smaller amount and amplifying it more through small population size; or (ii) generating a larger amount and amplifying it less. Indeed, analysis of mouse data suggests that the same amount of spread can be achieved with a small physical bottleneck and less mtDNA turnover (less generation, more amplification) or a large physical bottleneck and more mtDNA turnover (more generation, less amplification) (Johnston

et al., 2015) (**Figure 6C**). This flexibility may help reconcile differing reports on the size of the physical bottleneck (Cao et al., 2007, 2009; Cree et al., 2008; Johnston et al., 2015). It is not inconceivable that some mechanism may allow the cell to sense and control this choice, so that, for example, embryos with slightly lower mtDNA turnover have their mtDNA populations depleted more to compensate.

To consider these mechanisms together, the birth-death framework above was coupled to a description of cell divisions to provide a detailed stochastic model of germline development in mice (Johnston and Jones, 2015; Johnston et al., 2015). Compared to other detailed models, this birth-death-partitioning model provided the best fit to a meta-analysis of existing physical and genetic data. The best model for cell-to-cell spread of mutant proportion had two components: a contribution from partitioning at cell divisions and a contribution from ongoing drift due to mtDNA turnover.

The birth-death-partitioning model provides closed-form, though complicated, expressions for full distributional details of mutant proportion at all times through development, which well-predicted independent experimental observations of mutant proportion distributions in oocytes (Johnston et al., 2015). The combined birth-death-partitioning model was also used to provide an update to the Wright equation (Equation 6) to include random mtDNA turnover (Johnston and Jones, 2016), predicting:

$$sample\ var'\left[h\right] = 1 - \left(1 - \frac{1}{N}\right)^g + \frac{4t}{3N\tau}, \qquad (12)$$

where N is now a physical mtDNA copy number, g a physical number of cell divisions, t is time and τ is the timescale of mtDNA degradation. Append:

The final term in Equation (12) estimates the ongoing increase in mutant load spread due to mtDNA turnover, increasingly linearly with time t.

5. RECENT TOPICS

5.1. Model Selection and Predictions

We have discussed a range of different proposed mechanisms for the "genetic bottleneck" (mutant proportion spread). A comparatively recent set of studies has attempted to identify the mechanisms that are most supported by data. This has been attempted through the use of model selection (Kirk et al., 2013), a process that compares the statistical support for different mechanisms while guarding against overfitting. Li et al. used likelihood-based model selection with a human dataset to provide support for a "genetic bottleneck" (mutant proportion spread) that varies for different sequences and involves individual mtDNAs (rather than clusters) as segregating units (Li et al., 2016). Johnston et al. used likelihood-free model selection for mouse data to identify the mechanism(s) most supported by data. They found little support for partitioning of clustered mtDNA, and most support for the birth-death-partitioning model above, which was further supported by followup experiments (Johnston et al., 2015). A theoretical comparison of different models for

mtDNA control (Johnston and Jones, 2016) revealed the above principles of increasing variance that hold regardless of which specific mechanism is true. More recently, large-scale inter-generational data from mice was used in a statistical framework to identify which processes influence mtDNA statistics during development and aging (Burgstaller et al., 2018).

These detailed mathematical models present the opportunity to refine the prediction of mutant proportion distributions. The birth-death-partitioning model predicted distributional details of oocyte mutant proportion in developing mice (Johnston et al., 2015). Based on the picture of increasing mutant proportion spread in aging oocytes, a simple model involving a variation of a logit-normal distribution for mutant proportion predicted distributional details of mutant proportion in mouse litters (Burgstaller et al., 2018).

5.2. Sequence-Specific Behavior in Mutant Proportion Spread

Substantial recent attention has been focussed on whether the genetic bottleneck (mutant proportion spread) is sequence-specific. Evidence for this hypothesis includes observations from different pathological mtDNA mutations (Monnot et al., 2011; Wilson et al., 2016; Otten et al., 2018). Consideration of different human variants in a population genetic context also suggests that the magnitude of the genetic bottleneck (mutant proportion spread) depends on the specific variant under investigation (Li et al., 2016). A particularly striking difference appears to exist between the 3243 and 8993 mutations (Monnot et al., 2011; Wilson et al., 2016; Otten et al., 2018). The aforementioned population study (Li et al., 2016) also found a variable-size bottleneck to be most statistically supported for non-pathological mutations.

As discussed throughout, sequence-specific proliferative advantages of one mtDNA type over another can confound attempts to analyse the genetic bottleneck (mutant proportion spread). A sequence-specific increase in mutant proportion spread can arise without a proliferative difference between sequences: for example, if one sequence experiences both higher replication and degradation rates, increasing random turnover without an overall selective advantage. Conversely, under some experimental designs, sequence-specific differences in the behavior of mean mutant proportion (i.e., proliferative differences) could be interpreted instead as differences in mutant proportion variance if it is assumed that no proliferative differences exist (as in **Figure 4A**). Further theoretical work unpicking the behavior of mtDNA statistics as mean and variance change together will be useful in interpreting these observations.

5.3. Ongoing Increase of Mutant Proportion Spread During Aging

Recent large-scale intergenerational data in mice has shown an ongoing increase in mutant proportion spread in oocytes over time in adult mice (**Figure 7**). This increasing oocyte-to-oocyte spread of mutant proportion with age has been directly observed in mouse oocytes (Burgstaller et al., 2018), and has been shown to be more statistically supported than a constant-spread model in
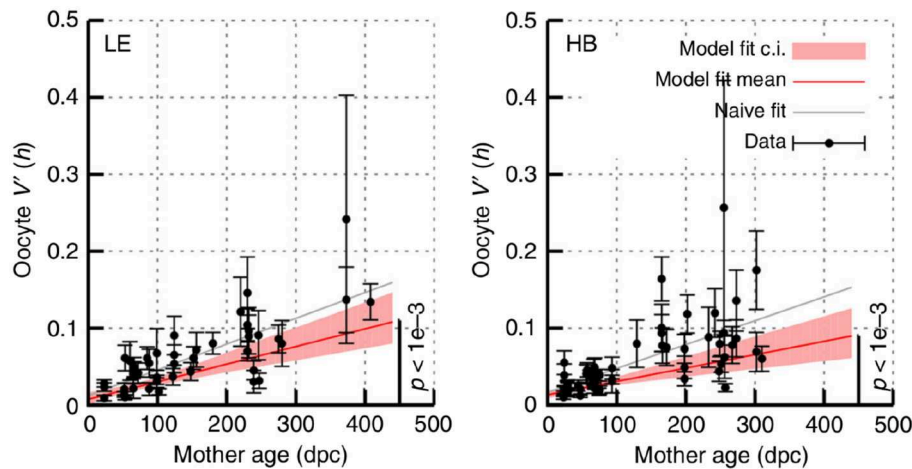
FIGURE 7 | Increasing mutant proportion spread in oocytes with mouse age. Data from Burgstaller et al. (2018), reporting *sample var'* [*h*] in sets of individual oocytes from mice of different ages. HB and LE label two genetic models, involving admixtures of wild-derived haplotypes HB and LE, respectively with haplotype C57Bl/6N. Error bars are derived using Equation (10); "model fit" accounts for this uncertainty and "naive fit" simply fits the bare observations. In all cases a significant linear increase in *sample var'* [*h*] with time, following Equation (11), is observed.

independent observations in flies, mice, and humans (Johnston and Jones, 2016).

The mechanism(s) behind the ongoing shrinking of the genetic bottleneck (increasing mutant proportion spread) remains unclear (Johnston et al., 2015; Zhang et al., 2018). However, random turnover of mtDNA may be a reasonable candidate mechanism (Johnston et al., 2015; Johnston and Jones, 2016; Burgstaller et al., 2018). The cumulative action of stochastic replication (and degradation) is to generate cell-to-cell spread in mitochondrial statistics, including in mutant proportion. Other processes like diversifying selection, physical clustering, and even mutagenesis could all contribute to the observed increase in spread.

These results are from systems involving cellular admixtures of two main haplotypes. Other results suggest a consistent picture, for example, showing an increasing number of heteroplasmic sites in children from older mothers, which the authors suggest is likely attributable to oocyte aging (Rebolledo-Jaramillo et al., 2014). Another study found non-uniform changes in heteroplasmy with age in humans (Sondheimer et al., 2011).

In light of this observation, this article would advocate an additional careful analysis of the contribution of maternal age to observed mutant proportion patterns. As we expect the genetic bottleneck (mutant proportion spread) to decrease with age, any systematic differences in age between these compared variances could confound other relationships. Conversely and more positively, appropriate accounting for age would help increase the statistical power of these comparisons.

6. THE PROBLEM OF SELECTION

Throughout the above, we have alluded to the problems that systematic selection for one or more mtDNA types can cause in these analyses. Theory describing the influence of selection has been established, but is complicated (Johnston et al., 2015). In particular, if approaches that assume the absence of selection

are used when it is in fact present, errors can arise in estimates of genetic properties and physical mechanisms. As pointed out above, these issues may lead to dramatic underestimation of "bottleneck size," and cannot be assumed to "cancel out."

Several of the results above are valid only in the absence of selection: when no mtDNA type experiences an advantage over any other. This is known to be false for many mtDNA pairings in many somatic tissues, where selection for one mtDNA type over another is often observed (reviewed in Burgstaller et al., 2014). Selection in the germline has been more debated, but evidence is increasing. In several studies, the transmission of pathological mutations seems to be subject to selective pressure. The maximum level of transmission for the 3243 mutation in humans has appeared to be limited (Monnot et al., 2011; Otten et al., 2018), and selection against severe mtDNA mutations has been observed in mice (Fan et al., 2008). Recent observations in mice (Burgstaller et al., 2018; Latorre-Pellicer et al., 2019) and humans (Wei et al., 2019) have indeed observed selection at different loci. Burgstaller et al. (2018) suggest that selection may act in different directions at different developmental stages (very recently supported by Latorre-Pellicer et al., 2019), and that these directions may either cancel out or provide a net selective shift. Mathematical theory for the behavior of mutant proportion spread when selection is present remains less well-developed and represents an important future theoretical target. The birth-death-partitioning approach in references (Johnston et al., 2015; Johnston and Jones, 2016) can account for selection but are mathematically complicated. Otten et al. (2018) have proposed a truncated Kimura distribution to describe a selective regime where mutant proportions above a certain value are prohibited, and found that it is supported by observations of the 3243 mutation.

Comparison to the Kimura distribution is often used to argue for an absence of selection. However, this approach must be interpreted with caution. Depending on the mechanism of selection, Kimura-distributed samples may be observed even

when selection has occurred. In particular, as approaches using the Kimura distribution sometimes use several "after" but no "before" measurements, it is possible that an early shift in mutant proportion will not be detected.

7. CONCLUSIONS

7.1. The Variable "Genetic Bottleneck"

This article has attempted to review the various models and mechanisms that have been considered for the "genetic bottleneck" (mutant proportion spread). Some diversity in reported mtDNA behavior comes from the choice of analysis protocol: the use of bottleneck parameters, rather than bottleneck sizes that allow a choice of "generation number," can help avoid this. The reporting of $sample\ var'\ [h]$, the fundamental observation from which these statistics are derived, and its associated uncertainty, will also help interpretability and comparison.

Ongoing research has provided evidence that the "genetic bottleneck" (mutant proportion spread) varies with age, species, individual, and genetic features. Intriguingly, the coupling of physical and genetic behavior of mitochondria (Equation 11; Tam et al., 2013; Aryaman et al., 2019) suggests that heterogeneity in mitochondrial dynamics may induce heterogeneity in mutant proportion.

A diverse range of studies on the mtDNA bottleneck continues to provide a wealth of insight into this important process. However, the very diversity of this research risks confusion arising, particularly around aspects of the prevailing terminology. This article has attempted to clarify some of the concepts involved, to serve as a reference for the increasingly interdisciplinary community working in this field.

Some takehome messages for reference include:

- The "genetic bottleneck" is a readout of mutant proportion spread that is generally not an observable physical quantity, and is measured reported in diverse ways through the literature;
- Observations of mutant proportion spread can have substantial uncertainty both from sampling and technical error, particularly if under 10 samples are used (when the standard error can approach half the observation);
- The physical mechanisms underlying the "genetic bottleneck" (mutant proportion spread) include a combination of copy number reduction (a physical bottleneck), random replication and degradation of mtDNA molecules, and random partitioning at cell divisions;
- The magnitude of the physical bottleneck appears to be flexible, as flexibility in mtDNA turnover can compensate to produce the same effects on mutant proportion spread;
- The presence of mtDNA selection complicates estimates of mutant proportion spread, and different experimental designs report different statistics in this case;
- The "genetic bottleneck" (mutant proportion spread) likely varies by species, individual, age, and mtDNA sequence.

DATA AVAILABILITY STATEMENT

This study did not generate new experimental data. Code for the simulations and visualizations involved is publically available at https://github.com/StochasticBiology/bottleneck-review.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Aryaman, J., Bowles, C., Jones, N. S., and Johnston, I. G. (2019). Mitochondrial network state scales mtDNA genetic dynamics. *Genetics* 212, 1429–1443. doi: 10.1534/genetics.119.302423

Ashley, M., Laipis, P., and Hauswirth, W. (1989). Rapid segregation of heteroplasmic bovine mitodiondria. *Nucleic Acids Res.* 17, 7325–7331.

Bendall, K. E., Macaulay, V. A., Baker, J. R., and Sykes, B. C. (1996). Heteroplasmic point mutations in the human mtdna control region. *Am. J. Hum. Genet.* 59:1276.

Birky Jr., C. (1994). Relaxed and stringent genomes: why cytoplasmic genes don't obey mendel's laws. *J. Heredity* 85, 355–369.

Birky Jr., C. W. (2001). The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu. Rev. Genet.* 35, 125–148. doi: 10.1146/annurev.genet.35.102401.090231

Blok, R. B., Gook, D. A., Thorburn, D. R., and Dahl, H.-H. M. (1997). Skewed segregation of the mtDNA N 8993 (TRG) mutation in human oocytes. *Am. J. Hum. Genet.* 60, 1495–1501.

Burgstaller, J. P., Johnston, I. G., and Poulton, J. (2014). Mitochondrial DNA disease and developmental implications for reproductive strategies. *Mol. Hum. Reprod.* 21, 11–22. doi: 10.1093/molehr/gau090

Burgstaller, J. P., Kolbe, T., Havlicek, V., Hembach, S., Poulton, J., Piálek, J., et al. (2018). Large-scale genetic analysis reveals mammalian mtdna heteroplasmy dynamics and variance increase through lifetimes and generations. *Nat. Commun.* 9:2488. doi: 10.1038/s41467-018-04797-2

Cao, L., Shitara, H., Horii, T., Nagao, Y., Imai, H., Abe, K., et al. (2007). The mitochondrial bottleneck occurs without reduction of mtDNA content in female mouse germ cells. *Nat. Genet.* 39:386. doi: 10.1038/ng1970

Cao, L., Shitara, H., Sugimoto, M., Hayashi, J.-I., Abe, K., and Yonekawa, H. (2009). New evidence confirms that the mitochondrial bottleneck is generated without

reduction of mitochondrial dna content in early primordial germ cells of mice. *PLoS Genet.* 5:e1000756. doi: 10.1371/journal.pgen.1000756

Capps, G. J., Samuels, D. C., and Chinnery, P. F. (2003). A model of the nuclear control of mitochondrial DNA replication. *J. Theor. Biol.* 221, 565–583. doi: 10.1006/jtbi.2003.3207

Carling, P. J., Cree, L. M., and Chinnery, P. F. (2011). The implications of mitochondrial dna copy number regulation during embryogenesis. *Mitochondrion* 11, 686–692. doi: 10.1016/j.mito.2011.05.004

Chinnery, P. F., and Samuels, D. C. (1999). Relaxed replication of mtDNA: a model with implications for the expression of disease. *Am. J. Hum. Genet.* 64, 1158–1165.

Cotterill, M., Harris, S. E., Collado Fernandez, E., Lu, J., Huntriss, J. D., Campbell, B. K., et al. (2013). The activity and copy number of mitochondrial dna in ovine oocytes throughout oogenesis *in vivo* and during oocyte maturation *in vitro*. *Mol. Hum. Reprod.* 19, 444–450. doi: 10.1093/molehr/gat013

Cree, L. M., Samuels, D. C., de Sousa Lopes, S. C., Rajasimha, H. K., Wonnapinij, P., Mann, J. R., et al. (2008). A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nat. Genet.* 40:249. doi: 10.1038/ng.2007.63

Diot, A., Morten, K., and Poulton, J. (2016). Mitophagy plays a central role in mitochondrial ageing. *Mammal. Genome* 27, 381–395. doi: 10.1007/s00335-016-9651-x

Eggenberger, F., and Pólya, G. (1923). Über die statistik verketteter vorgänge. *ZAMM J. Appl. Math. Mech.* 3, 279–289.

Fan, W., Waymire, K. G., Narula, N., Li, P., Rocher, C., Coskun, P. E., et al. (2008). A mouse model of mitochondrial disease reveals germline selection against severe mtdna mutations. *Science* 319, 958–962. doi: 10.1126/science.1147786

Floros, V. I., Pyle, A., Dietmann, S., Wei, W., Tang, W. W., Irie, N., et al. (2018). Segregation of mitochondrial dna heteroplasmy through a developmental genetic bottleneck in human embryos. *Nat. Cell Biol.* 20:144. doi: 10.1038/s41556-017-0017-8

Guo, Y., Li, C.-I., Sheng, Q., Winther, J. F., Cai, Q., Boice, J. D., et al. (2013). Very low-level heteroplasmy mtDNA variations are inherited in humans. *J. Genet. Genom.* 40, 607–615. doi: 10.1016/j.jgg.2013.10.003

Hauswirth, W. W., and Laipis, P. J. (1982). Mitochondrial dna polymorphism in a maternal lineage of holstein cows. *Proc. Natl. Acad. Sci. U.S.A.* 79, 4686–4690.

Hendy, M. D., Woodhams, M. D., and Dodd, A. (2009). Modelling mitochondrial site polymorphisms to infer the number of segregating units and mutation rate. *Biol. Lett.* 5, 397–400. doi: 10.1098/rsbl.2009.0104

Hoitzing, H., Gammage, P. A., Minczuk, M., Johnston, I. G., and Jones, N. S. (2019). Energetic costs of cellular and therapeutic control of stochastic mtdna populations. *PLoS Comput. Biol.* 15:e1007023. doi: 10.1371/journal.pcbi.1007023

Hoitzing, H., Johnston, I. G., and Jones, N. S. (2015). What is the function of mitochondrial networks? A theoretical assessment of hypotheses and proposal for future research. *Bioessays* 37, 687–700. doi: 10.1002/bies.201400188

Hoitzing, H., Johnston, I. G., and Jones, N. S. (2017). "Stochastic models for evolving cellular populations of mitochondria: disease, development, and ageing," in *Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology*, ed D. Holcman (Cham: Springer), 287–314.

Howell, N., Halvorson, S., Kubacka, I., McCullough, D., Bindoff, L., and Turnbull, D. (1992). Mitochondrial gene segregation in mammals: is the bottleneck always narrow? *Hum. Genet.* 90, 117–120.

Huh, D., and Paulsson, J. (2011). Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat. Genet.* 43:95. doi: 10.1038/ng.729

Jacobs, H. T., Lehtinen, S. K., and Spelbrink, J. N. (2000). No sex please, we're mitochondria: a hypothesis on the somatic unit of inheritance of mammalian mtdna. *Bioessays* 22, 564–572. doi: 10.1002/(SICI)1521-1878(200006)22:6<564::AID-BIES9>3.3.CO;2-W

Jajoo, R., Jung, Y., Huh, D., Viana, M. P., Rafelski, S. M., Springer, M., et al. (2016). Accurate concentration control of mitochondria and nucleoids. *Science* 351, 169–172. doi: 10.1126/science.aaa8714

Jenuth, J. P., Peterson, A. C., Fu, K., and Shoubridge, E. A. (1996). Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nat. Genet.* 14:146.

Johnson, N. L., and Kotz, S. (1977). *Urn Models and Their Application; An Approach to Modern Discrete Probability Theory*. New York, NY; London: Wiley.

Johnston, I. G. (2018). Tension and resolution: dynamic, evolving populations of organelle genomes within plant cells. *Mol. Plant* 12, 764–783. doi: 10.1016/j.molp.2018.11.002

Johnston, I. G., and Burgstaller, J. P. (2019). Evolving mtdna populations within cells. *Biochem. Soc. Trans.* 47, 1367–1382. doi: 10.1042/BST20190238

Johnston, I. G., Burgstaller, J. P., Havlicek, V., Kolbe, T., Rülicke, T., Brem, G., et al. (2015). Stochastic modelling, Bayesian inference, and new *in vivo* measurements elucidate the debated mtDNA bottleneck mechanism. *Elife* 4:e7464. doi: 10.7554/eLife.07464

Johnston, I. G., Gaal, B., das Neves, R. P., Enver, T., Iborra, F. J., and Jones, N. S. (2012). Mitochondrial variability as a source of extrinsic cellular noise. *PLoS Comput. Biol.* 8:e1002416. doi: 10.1371/journal.pcbi.1002416

Johnston, I. G., and Jones, N. S. (2015). Closed-form stochastic solutions for non-equilibrium dynamics and inheritance of cellular components over many cell divisions. *Proc. R. Soc. A* 471:20150050. doi: 10.1098/rspa.2015.0050

Johnston, I. G., and Jones, N. S. (2016). Evolution of cell-to-cell variability in stochastic, controlled, heteroplasmic mtDNA populations. *Am. J. Hum. Genet.* 99, 1150–1162. doi: 10.1016/j.ajhg.2016.09.016

Jokinen, R., and Battersby, B. J. (2013). Insight into mammalian mitochondrial dna segregation. *Ann. Med.* 45, 149–155. doi: 10.3109/07853890.2012.693190

Khrapko, K. (2008). Two ways to make an mtdna bottleneck. *Nat. Genet.* 40:134. doi: 10.1038/ng0208-134

Kimura, M. (1955). Solution of a process of random genetic drift with a continuous model. *Proc. Natl. Acad. Sci. U.S.A.* 41, 144–150.

Kirk, P., Thorne, T., and Stumpf, M. P. (2013). Model selection in systems and synthetic biology. *Curr. Opin. Biotechnol.* 24, 767–774. doi: 10.1016/j.copbio.2013.03.012

Koehler, C. M., Lindberg, G., Brown, D., Beitz, D., Freeman, A. E., Mayfield, J., et al. (1991). Replacement of bovine mitochondrial dna by a sequence variant within one generation. *Genetics* 129, 247–255.

Kukat, C., Wurm, C. A., Spåhr, H., Falkenberg, M., Larsson, N.-G., and Jakobs, S. (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. U.S.A.* 108, 13534–13539. doi: 10.1073/pnas.1109263108

Latorre-Pellicer, A., Lechuga-Vieco, A. V., Johnston, I. G., Hämäläinen, R. H., Pellico, J., Justo-Méndez, R., et al. (2019). Regulation of mother-to-offspring transmission of mtDNA heteroplasmy. *Cell Metab.* 30, 1–11. doi: 10.1016/j.cmet.2019.09.007

Lee, H.-S., Ma, H., Juanes, R. C., Tachibana, M., Sparman, M., Woodward, J., et al. (2012). Rapid mitochondrial dna segregation in primate preimplantation embryos precedes somatic and germline bottleneck. *Cell Rep.* 1, 506–515. doi: 10.1016/j.celrep.2012.03.011

Li, M., Rothwell, R., Vermaat, M., Wachsmuth, M., Schröder, R., Laros, J. F., et al. (2016). Transmission of human mtDNA heteroplasmy in the genome of the netherlands families: support for a variable-size bottleneck. *Genome Res.* 26, 417–426. doi: 10.1101/gr.203216.115

Lutz, S., Weisser, H.-J., Heizmann, J., and Pollak, S. (2000). Mitochondrial heteroplasmy among maternally related individuals. *Int. J. Legal Med.* 113, 155–161. doi: 10.1007/s004140050288

Marchington, D., Hartshorne, G., Barlow, D., and Poulton, J. (1997). Homopolymeric tract heteroplasmy in mtDNA from tissues and single oocytes: support for a genetic bottleneck. *Am. J. Hum. Genet.* 60:408.

Millar, C. D., Dodd, A., Anderson, J., Gibb, G. C., Ritchie, P. A., Baroni, C., et al. (2008). Mutation and evolutionary rates in adélie penguins from the antarctic. *PLoS Genet.* 4:e1000209. doi: 10.1371/journal.pgen.1000209

Monnot, S., Gigarel, N., Samuels, D. C., Burlet, P., Hesters, L., Frydman, N., et al. (2011). Segregation of mtdna throughout human embryofetal development: m.3243A>G as a model system. *Hum. Mutat.* 32, 116–125. doi: 10.1002/humu.21417

Monnot, S., Samuels, D. C., Hesters, L., Frydman, N., Gigarel, N., Burlet, P., et al. (2013). Mutation dependance of the mitochondrial dna copy number in the first stages of human embryogenesis. *Hum. Mol. Genet.* 22, 1867–1872. doi: 10.1093/hmg/ddt040

Moran, P. A. P. (1958). "Random processes in genetics," in *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 54 (Cambridge, UK: Cambridge University Press), 60–71.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutat. Res.* 1, 2–9.

Otten, A. B., Sallevelt, S. C., Carling, P. J., Dreesen, J. C., Drüsedau, M., Spierts, S., et al. (2018). Mutation-specific effects in germline transmission of pathogenic mtdna variants. *Hum. Reprod.* 33, 1331–1341. doi: 10.1093/humrep/dey114

Otten, A. B., Theunissen, T. E., Derhaag, J. G., Lambrichs, E. H., Boesten, I. B., Winandy, M., et al. (2016). Differences in strength and timing of the mtdna bottleneck between zebrafish germline and non-germline cells. *Cell Rep.* 16, 622–630. doi: 10.1016/j.celrep.2016.06.023

Pallotti, F., Binelli, G., Fabbri, R., Valentino, M. L., Vicenti, R., Macciocca, M., et al. (2014). A wide range of 3243A>G/tRNALeu (UUR) (MELAS) mutation loads may segregate in offspring through the female germline bottleneck. *PLoS ONE* 9:e96663. doi: 10.1371/journal.pone.0096663

Poulton, J., Macaulay, V., and Marchington, D. (1998). Mitochondrial genetics' 98 is the bottleneck cracked? *Am. J. Hum. Genet.* 62:752.

Rand, D. M., and Harrison, R. G. (1986). Mitochondrial DNA transmission genetics in crickets. *Genetics* 114, 955–970.

Rebolledo-Jaramillo, B., Su, M. S.-W., Stoler, N., McElhoe, J. A., Dickins, B., Blankenberg, D., et al. (2014). Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15474–15479. doi: 10.1073/pnas.1409328111

Rossignol, R., Faustin, B., Rocher, C., Malgat, M., Mazat, J.-P., and Letellier, T. (2003). Mitochondrial threshold effects. *Biochem. J.* 370, 751–762. doi: 10.1042/bj20021594

Sallevelt, S. C., Dreesen, J. C., Drüsedau, M., Spierts, S., Coonen, E., van Tienen, F. H., et al. (2013). Preimplantation genetic diagnosis in mitochondrial DNA disorders: challenge and success. *J. Med. Genet.* 50, 125–132. doi: 10.1136/jmedgenet-2012-101172

Samuels, D. C., Wonnapinij, P., Cree, L. M., and Chinnery, P. F. (2010). Reassessing evidence for a postnatal mitochondrial genetic bottleneck. *Nat. Genet.* 42, 471–472.

Sekiguchi, K., Imaizumi, K., Matsuda, H., Mizuno, N., Yoshida, K., Senju, H., et al. (2003). mtDNA sequence analysis using capillary electrophoresis and its application to the analysis of mtdna in hair. *Jpn. J. Sci. Technol. Identif.* 7, 123–130. doi: 10.3408/jasti.7.123

Solignac, M., Génermont, J., Monnerot, M., and Mounolou, J.-C. (1984). Genetics of mitochondria in drosophila: mtDNA inheritance in heteroplasmic strains of d. mauritiana. *Mol. Gen. Genet.* 197, 183–188.

Sondheimer, N., Glatz, C. E., Tirone, J. E., Deardorff, M. A., Krieger, A. M., and Hakonarson, H. (2011). Neutral mitochondrial heteroplasmy and the influence of aging. *Hum. Mol. Genet.* 20, 1653–1659. doi: 10.1093/hmg/ddr043

Stewart, J. B., and Chinnery, P. F. (2015). The dynamics of mitochondrial dna heteroplasmy: implications for human health and disease. *Nat. Rev. Genet.* 16:530. doi: 10.1038/nrg3966

Tam, Z. Y., Gruber, J., Halliwell, B., and Gunawan, R. (2013). Mathematical modeling of the role of mitochondrial fusion and fission in mitochondrial DNA maintenance. *PLoS ONE* 8:e76230. doi: 10.1371/journal.pone.0076230

Treff, N. R., Campos, J., Tao, X., Levy, B., Ferry, K. M., and Scott Jr., R. T. (2012). Blastocyst preimplantation genetic diagnosis (PGD) of a mitochondrial dna disorder. *Fertil. Steril.* 98, 1236–1240. doi: 10.1016/j.fertnstert.2012.07.1119

Twig, G., Elorza, A., Molina, A. J., Mohamed, H., Wikstrom, J. D., Walzer, G., et al. (2008). Fission and selective fusion govern mitochondrial segregation and elimination by autophagy. *EMBO J.* 27, 433–446. doi: 10.1038/sj.emboj.7601963

Wai, T., Teoli, D., and Shoubridge, E. A. (2008). The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. *Nat. Genet.* 40:1484. doi: 10.1038/ng.258

Wallace, D. C. (2018). Mitochondrial genetic medicine. *Nat. Genet.* 50, 1642–1649. doi: 10.1038/s41588-018-0264-z

Wallace, D. C., and Chalkia, D. (2013). Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harbor Perspect. Biol.* 5:a021220. doi: 10.1101/cshperspect.a021220

Wei, W., Tuna, S., Keogh, M. J., Smith, K. R., Aitman, T. J., Beales, P. L., et al. (2019). Germline selection shapes human mitochondrial DNA diversity. *Science* 364:eaau6520. doi: 10.1126/science.aau6520

Wilson, I. J., Carling, P. J., Alston, C. L., Floros, V. I., Pyle, A., Hudson, G., et al. (2016). Mitochondrial DNA sequence characteristics modulate the size of the genetic bottleneck. *Hum. Mol. Genet.* 25, 1031–1041. doi: 10.1093/hmg/ddv626

Wilton, P. R., Zaidi, A., Makova, K., and Nielsen, R. (2018). A population phylogenetic view of mitochondrial heteroplasmy. *Genetics* 208, 1261–1274. doi: 10.1534/genetics.118.300711

Wolff, J. N., White, D. J., Woodhams, M., White, H. E., and Gemmell, N. J. (2011). The strength and timing of the mitochondrial bottleneck in salmon suggests a conserved mechanism in vertebrates. *PLoS ONE* 6:e20522. doi: 10.1371/journal.pone.0020522

Wonnapinij, P., Chinnery, P. F., and Samuels, D. C. (2008). The distribution of mitochondrial DNA heteroplasmy due to random genetic drift. *Am. J. Hum. Genet.* 83, 582–593. doi: 10.1016/j.ajhg.2008.10.007

Wonnapinij, P., Chinnery, P. F., and Samuels, D. C. (2010). Previous estimates of mitochondrial DNA mutation level variance did not account for sampling error: comparing the mtDNA genetic bottleneck in mice and humans. *Am. J. Hum. Genet.* 86, 540–550. doi: 10.1016/j.ajhg.2010.02.023

Wright, S. (1942). Statistical genetics and evolution. *Bull. Am. Math. Soc.* 48, 223–246.

Wright, S. (1984). *Evolution and the Genetics of Populations, Volume 2: Theory of Gene Frequencies*, Vol. 2. Chicago, IL; London: University of Chicago Press.

Youle, R. J., and Narendra, D. P. (2011). Mechanisms of mitophagy. *Nat. Rev. Mol. Cell Biol.* 12:9. doi: 10.1038/nrm3028

Zhang, H., Burr, S. P., and Chinnery, P. F. (2018). The mitochondrial DNA genetic bottleneck: inheritance and beyond. *Essays Biochem.* 62, 225–234. doi: 10.1042/EBC20170096