


Avoiding misleading estimates using mtDNA heteroplasmy statistics to study bottleneck size and selection

Konstantinos Giannakis,¹ Amanda K. Broz,² Daniel B. Sloan,² Iain G. Johnston ^{1,3,*}

¹Department of Mathematics, University of Bergen, 5007 Bergen, Norway

²Department of Biology, Colorado State University, Fort Collins, CO 80523, USA

³Computational Biology Unit, University of Bergen, 5008 Bergen, Norway

*Corresponding author: Email: iain.johnston@uib.no

Abstract

Mitochondrial DNA heteroplasmy samples can shed light on vital developmental and genetic processes shaping mitochondrial DNA populations. The sample means and sample variance of a set of heteroplasmy observations are typically used both to estimate bottleneck sizes and to perform fits to the theoretical “Kimura” distribution in seeking evidence for mitochondrial DNA selection. However, each of these applications raises problems. Sample statistics do not generally provide optimal fits to the Kimura distribution and so can give misleading results in hypothesis testing, including false positive signals of selection. Using sample variance can give misleading results for bottleneck size estimates, particularly for small samples. These issues can and do lead to false positive results for mitochondrial DNA mechanisms—all published experimental datasets we re-analyzed, reported as displaying departures from the Kimura model, do not in fact give evidence for such departures. Here we outline a maximum likelihood approach that is simple to implement computationally and addresses all of these issues. We advocate the use of maximum likelihood fits and explicit hypothesis tests, not fits and Kolmogorov–Smirnov tests via summary statistics, for ongoing work with mitochondrial DNA heteroplasmy.

Keywords: mtDNA, heteroplasmy, selection, bottleneck, Kimura distribution, statistical genetics

Introduction

Mitochondrial DNA (mtDNA) encodes vital energetic apparatus. Populations of tens to thousands of mtDNA molecules exist in eukaryotic cells, and some of these molecules may carry mutations that cause disease. Because of the high ploidy of mtDNA, a cell may contain a coexisting “heteroplasmic” mixture of mutant and wildtype mtDNA molecules. Cells may support a certain heteroplasmy—that is, a certain proportion of mutant mtDNA—before suffering any consequences, but if that threshold is exceeded, pathologies result (Wallace and Chalkia 2013; Stewart and Chinnery 2015).

Because germ cells also contain highly polyploid populations of mtDNA, its inheritance differs from that of nuclear DNA. MtDNA is often exclusively maternally inherited. If a mother carries an mtDNA mutation, her oocytes will in general have a range of different heteroplasmies for that mutation. This variability is generated during development by a collection of random processes referred to as the genetic “bottleneck”, where a reduced effective population size increases cell-to-cell variability in the oocyte population (Stewart and Chinnery 2015; Johnston 2019). The variability generated by the bottleneck allows some offspring to inherit heteroplasmy levels lower than their mother, helping to ensure viable offspring can be generated. This segregation of mutational damage across offspring is observed across eukaryotic kingdoms (Edwards et al. 2021).

The amount of heteroplasmy variance generated between oocytes and offspring is of central importance both in mitochondrial

biology and in clinical efforts to understand and plan for families carrying mtDNA mutations. The magnitude of heteroplasmy variance is often quantified by considering a “bottleneck size” n_b —the sample size that would generate the same magnitude of heteroplasmy variance if a single binomial sampling event was responsible for generating all variability. Inferring bottleneck size with samples of observed genetic data is of interest and importance from fundamental biology to clinical planning. It is common to estimate bottleneck size using the sample mean $\bar{h} = (1/n) \sum h_i$ and sample variance $s^2 = (1/(n-1)) \sum (h_i - \bar{h})^2$ of a set of heteroplasmy measurements $\{h_i\}$. Specifically, bottleneck size is typically estimated using

$$\hat{n}_b = p(1-p)/V,$$

where p and V are sample statistics, calculated from observations in different ways in different projects. Most commonly, given a sample of heteroplasmy values, the sample mean \bar{h} is used to estimate p (thus assuming an absence of selection (Johnston 2019)), and the sample variance s^2 for V . In other cases, an earlier or reference heteroplasmy measurement may be used for p ; the population variance expression is sometimes used for V . All of these quantities are estimators, based on a sample, for the population relationship $n_b = p_0(1-p_0)/\sigma^2$, which in turn comes from a binomial model where the population variance is given in terms of the parameters $\sigma^2 = n_b p_0(1-p_0)$.

All the above assumes that no systematic pressures favor mutant or wildtype mtDNA during development and inheritance. The

Received: October 18, 2022. Accepted: March 10, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of The Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

question of whether selective differences do act between mtDNA types in the germline is hotly debated in different systems, including humans (Johnston 2019; Wei et al. 2019). One approach that has been used to explore mtDNA selection involves the comparison of heteroplasmy measurements to a theoretically predicted heteroplasmy distribution under neutral drift (Kimura 1955). This comparison has been used in several recent studies (Wonnapijit et al. 2008, 2010; Monnot et al. 2011; Freyer et al. 2012; Jokinen et al. 2016; Otten et al. 2018; Zhang et al. 2021) and to model mtDNA inheritance (Samuels et al. 2013). Within the mtDNA field, this neutral-drift distribution is commonly referred to as the “Kimura” distribution, although Kimura actually derived several other distributions describing the role of selective differences and other features (Kimura 1954; see below). The Kimura distribution of interest here is the theoretical distribution of allele frequencies after neutral drift has acted on an initial population for some given time. The theory is built up under assumptions of discrete generations, no selection, migration, or mutation, and fixed population size. The application to heteroplasmy combines the time and population size parameters into a single “drift” parameter so that the distribution is parameterized by an initial frequency (heteroplasmy) and an amount of drift, which is related to the similarly effective bottleneck size above. Depending on the parameterization, the Kimura distribution can take the form of a tight peak around an initial value or a wider normal-like bell curve, or (after more drift) can have substantial density at the point values $h=0$ and $h=1$, reflecting the fixation or extinction of an allele (some examples are shown in Fig. 1).

Fits of observed heteroplasmy data to the Kimura distribution, like bottleneck estimates, are also typically performed using the sample mean and sample variance for a set of measurements, in a method outlined by Wonnapijit et al. (2008; 2010) and referred to as the WCS-K approach after those authors and Kimura (Wallace and Chalkia 2013). The interpretation here is that if the Kimura distribution—which assumes no selective differences—does not fit the data well, there may be some support for selective differences in the system of interest. The Kimura distribution as used by the WCS-K approach takes two parameters p (initial heteroplasmy) and b (describing the extent of drift), both between 0 and 1. A value of b close to 1 corresponds to little drift and variance; lower values correspond to more drift and higher variance (b is connected to bottleneck size n_b via $n_b = 1/(1-b)$). The population mean and variance are

$$\begin{aligned}\mu &= p \\ \sigma^2 &= p(1-p)(1-b).\end{aligned}$$

Typically, a fit to observed heteroplasmy samples is performed by calculating the sample mean and sample variance of the sample, asserting that these are equal to the population mean μ and population variance σ^2 above, and using these relationships to provide estimates for the parameters p and b (Wonnapijit et al. 2008; 2010). If the equality assertion is instead recognized as an estimation, this is the so-called “method-of-(central)-moments” for fitting a distribution. Having fitted this distribution, the WCS-K approach uses a Kolmogorov–Smirnov (KS) test to compare the empirical distribution of observed heteroplasmy values to a large set of samples drawn from the fitted distribution (although this workflow has statistical flaws, described below and by Crutcher (1975)). The KS test considers the cumulative distribution functions (CDFs) of two sampled distributions and seeks the maximum absolute difference between these functions (illustrated in

Fig. 1b). If this maximum distance exceeds a threshold, computed under the null hypothesis that both sample sets come from the same distribution, a significant difference is claimed between the fitted theory and observed samples. In the WCS-K approach, this is inferred to provide evidence for selection, as the theory is developed under assumptions of neutrality.

The problems with these fitting approaches can be illustrated with 3 related problems.

Problem 1. Misleading bottleneck size estimates. For our first example, we consider a dataset on mtDNA heteroplasmy recently published in Broz et al. (2022). A mother plant with heteroplasmy $h_0=0.81$ produced 24 offspring. 15 had $h=0$, 3 had $h=1$, and the remaining h values were (0.91, 0.91, 0.92, 0.94, 0.95, 0.98). As we have an initial “reference” heteroplasmy (the mother’s h_0), we can estimate the bottleneck size with

$$\begin{aligned}\hat{n}_b &= h_0(1-h_0)/s^2 \\ &= 0.81 \times 0.19/0.225 \\ &= 0.684,\end{aligned}$$

Leaving us with a bottleneck size of 0.68 segregating units. Such a value is nonsensical because we cannot segregate less than a single unit of information through a bottleneck (even a theoretical, binomial one). This problem can readily be demonstrated with a simpler example of two heteroplasmy measurements (0, 1) (see Supplementary File 1), and is not a rare case: other mtDNA (and plastid DNA) measurements from plants also yield such bi-homoplasmic observations.

Problem 2. False positive signals of selection (or other processes) via Kimura fit. We next consider a dataset from Freyer et al. (2012), plotted in Fig. 1a. Using the WCS-K approach, we fit the Kimura distribution using summary statistics $p = \bar{h} = 0.605$ and $b = 1 - s^2/(\bar{h}(1-\bar{h})) = 0.901$. A Monte Carlo implementation of the KS test (Wonnapijit et al. 2008) gives a P -value of 0.028 (close to that reported in the original publication), suggesting a significant departure from the theoretical model. Such a departure could be interpreted as a signature of selection.

But now consider a different Kimura distribution, with $p = 0.628$ and $b = 0.935$ (the provenance of these parameters will be explained below). Now the KS test gives a P -value of 0.556. So there is no reason to reject the null hypothesis that these observations were drawn from a Kimura distribution—removing our reason to suggest that selection may be acting. Because we previously used an (inappropriate) fit based on summary statistics, we chose the wrong Kimura distribution with which to compare our data and obtained a false positive. This is before accounting for the more general fact that the application of the KS test in this situation—comparing data to samples from a distribution whose parameters are fitted using the same data—is statistically incorrect as described below (Crutcher 1975).

This issue can be illustrated more dramatically with a synthetic example, involving the dataset $\underline{h} = (0, 0, 0, 0, 0.9, 0.9, 0.9, 0.9)$ (plotted in Fig. 1b). Here, the distributions corresponding to different fit methods (see below) have more dramatically different structures and the different KS distances in each case can be more easily seen.

We have chosen a particular real and synthetic dataset to simply illustrate this point, and a synthetic example that makes the quantities involved clear. But the issue is systemic: upon re-analyzing a range of real datasets for which significant departures were reported (see below), we have not identified any cases that in fact departed significantly from the Kimura distribution at a threshold of $P < 0.05$.

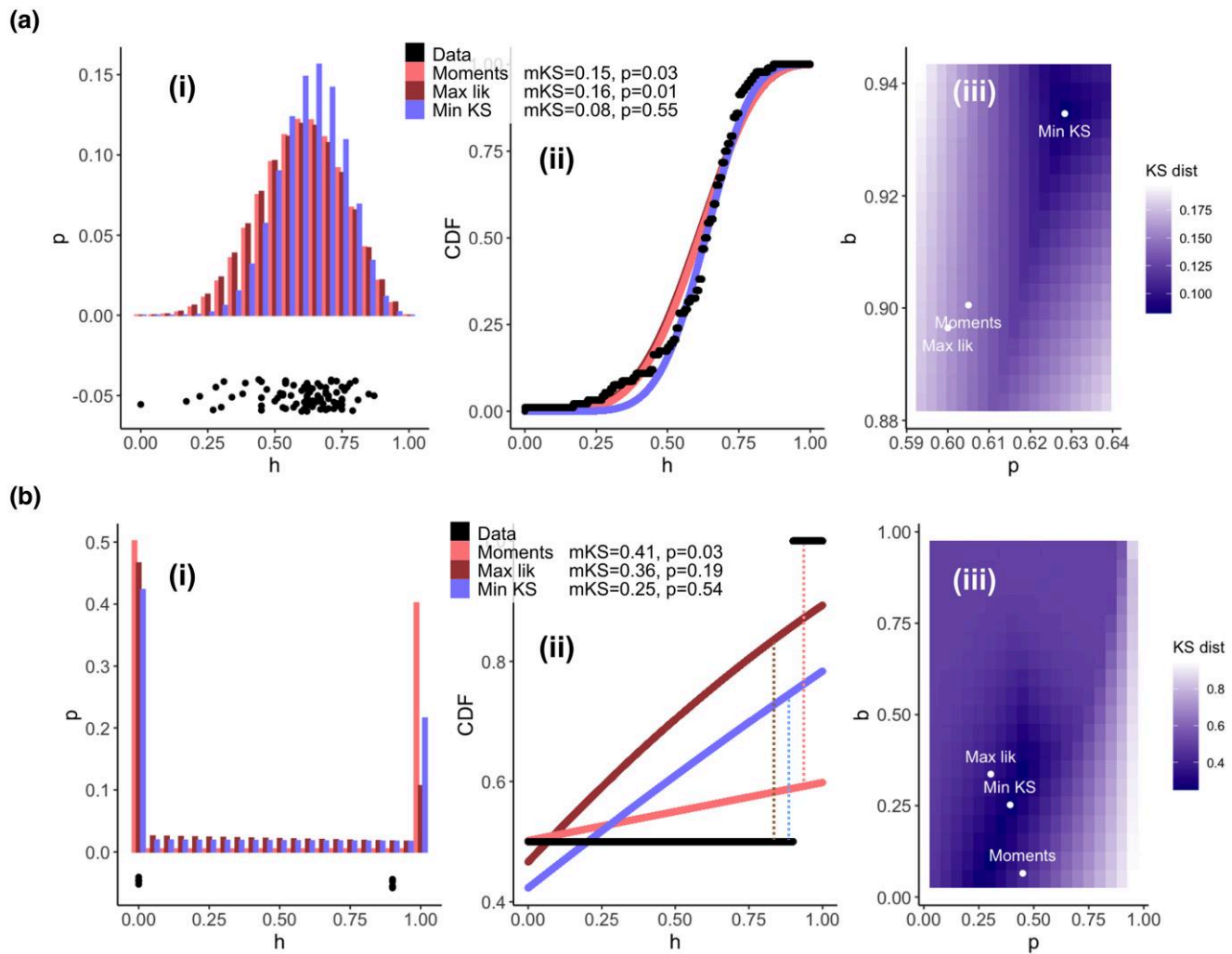


Fig. 1. Different estimators for the Kimura distribution give dramatic differences in hypothesis testing. Fits via the (commonly implemented) method-of-moments, maximum likelihood, and minimum KS distance to example datasets. Each method's KS distance (mKS) and p-value from the Monte Carlo test described in the text is given in the inset. a) Data from Freyer et al. (2012) (specifically, the concatenated set of observations for mother heteroplasmies ≤ 0.6); b) synthetic data as described in the text. (i) Probability distributions for each fit; (ii) comparison of CDFs from the data and for each fit; (iii) the KS distance between empirical observations and theoretical distribution with different choices of parameter p and b of the Kimura distribution (parameterizations from specific fits marked as points). The maximum distance between CDFs of the data and a fitted distribution gives the KS distance (illustrated by dotted lines in (b) (ii)); the method-of-moments (and maximum likelihood) give KS distances rather higher than the minimum KS distance approaches, with correspondingly low (and false positive) P-values in hypothesis testing. Even the subtle differences in parameterization for (a) give dramatically different P-values in the KS test.

Problem 3. Hypothesis testing with bottleneck sizes. Sample variances are not drawn from a normal distribution. Even when the population from which samples are taken is normal, the sample variance follows a chi-squared distribution; when the population distribution is not normal (as for heteroplasmy measurements), the sample variance distribution is in general not known. We cannot then use tests that invoke a normality assumption (like the t-test) to compare sample variances or the bottleneck size estimates that are derived from them. Must we use low-powered non-parametric approaches instead?

Materials and methods

Statistics and code

The statistical analyses we employ here are described in the text and were implemented in R within a new package *heteroplasmy* (<https://github.com/kostasgian21/heteroplasmy>). The existing R packages used are *kimura* (<https://github.com/lbozhilova/>

kimura), *ggplot2* (Wickham 2016) for plotting, *foreach* (Weston and Microsoft Corporation 2020) (<https://CRAN.R-project.org/package=foreach>), and optionally *doParallel* (Weston and Microsoft Corporation 2022) (<https://CRAN.R-project.org/package=doParallel>) for parallel execution of the code, and *devtools* (Wickham et al. 2020) (<https://CRAN.R-project.org/package=devtools>) to download and install the *kimura* and *heteroplasmy* packages. The code is freely available at <https://github.com/StochasticBiology/heteroplasmy-analysis>.

Heteroplasmy data

The LE and HB mouse oocyte heteroplasmy data were taken from Burgstaller et al. (2018). In our paper, only indicative samples of the dataset in Burgstaller et al. (2018) were used. LE and HB correspond to two genetically distinct heteroplasmic mouse lines; HB and LE stand for Hohenberg and Lehsten respectively, the regions from which the founders of the mouse models were isolated (Burgstaller et al. 2014). For the purposes of this work,

only samples from oocytes were used. Heteroplasmy measurements were taken as found in [Supplementary File 1](#).

Data from progenitor germ cells (PGCs), oocytes, and offspring were taken from [Freyer et al. \(2012\)](#). Grouping of individual measurements was undertaken to match the grouping within the original publication, according to the ranges of the mother's reference heteroplasmy for PGCs and offspring as classified in the paper (i.e. [0.0–0.6], [0.6–0.7], [0.7–1.0]). For the oocyte data, all 5 samples were individually tested (again, as in the original paper). Because of ambiguity with the order of reporting of the associated significance results in the original paper for the oocyte data, we mapped each of our results to the numerically closest ones from the original paper.

Human, fly (*Drosophila mauritiana* and *D. simulans*), and mouse heteroplasmy measurements from [Wonnapijit et al. \(2008\)](#) were also used. Originally, the human data came from [Brown et al. \(2001\)](#), fly data from [de Stordeur et al. \(1989\)](#), and mouse data from [Jenuith et al. \(1996\)](#). The raw data are not available in either publication, so we use the binned data, as depicted in relevant histograms in [Wonnapijit et al. \(2008\)](#). To interpret the binned data as estimated individual measurements we tested several methods, including taking the mean of each bin, taking the mean and then adding small noise disturbances to cut the ties, and others. The method that most faithfully reproduced the original results was resampling the data within each bin by taking uniformly random values within the bin interval. As noted also by [Wallace and Chalkia \(2013\)](#), caution is needed when binned data are used to test for selection by fitting a Kimura distribution. Therefore, we reported here ranges of P-values when using the method-of-moments approach to replicate the results from ([Wonnapijit et al. 2008](#)), instead of single values (in [Table 1](#)).

Finally, organellar heteroplasmy data were taken from [Broz et al. \(2022\)](#), including between-generation measurements of mitochondrial heteroplasmy for the *mt91017* and *mt334038* SNPs from *msh1* plants, and the *mt334038* SNP from wildtype plants.

All data, if not presented as such, were normalized to proportions (on the [0,1] interval) rather than percentages for analysis.

For the synthetic data, the normal samples were generated using the *morm* command in base R. For the data generated by a Kimura distribution, the *rkimura* function from the *kimura* R package (<https://github.com/lbozhilova/kimura>) was used.

Results

A parametric solution—maximum likelihood fitting

How can we estimate a meaningful bottleneck size and derive sensible confidence intervals in the challenging cases above? Fitting heteroplasmy measurements to the Kimura distribution can actually answer all of these questions. However, the above examples make it clear that we cannot in general perform a simple matching of distribution parameters to summary statistics (choosing the parameters, p and b , that give a distribution with the same mean and variance as the sample). This method-of-moments will only find the parameters that are most compatible with the observations in the case of infinite sample size, where the sample mean and variance converge to the population mean and variance. For many purposes, the method-of-moments estimates may be close enough to these most-compatible values to provide useful information. However, as we show here, the differences between them can have substantial misleading effects on hypothesis testing when studying heteroplasmy.

When individual heteroplasmy measurements are available (sometimes only summary statistics are reported), and there is reason to believe the Kimura model, a more appropriate approach is instead to identify the maximum likelihood parameters given the full set of measurements (as used previously in at least one study ([Otten et al. 2018](#))). The maximum likelihood parameters for a statistical model are those that give the highest joint probability of observing our measurements under that model. Although these do correspond to the sample mean and sample variance in the case of a normal distribution model, for other distributions (including the Kimura distribution) this is not generally the case (as in the example above). Instead, we have to find the parameters with the highest associated likelihood. The Kimura distribution Kimura ($h | p, b$) gives the probability density for a given heteroplasmy observation h . We write

$$L = \prod_i \text{Kimura}(h_i | p, b),$$

and seek the p, b combination that maximizes L for a given h . If we wish to enforce a particular p —for example, if we have a reliable initial heteroplasmy measurement—we can instead perform the search only over b . Both p and b are confined in [0,1] here, respecting the constraints of the system.

This maximum likelihood process will identify the b (and p , if required) that is most likely given the set of observations. We can also derive confidence intervals on this estimate using Fisher information or bootstrapping (see [Supplementary File 1](#)), obtaining, for example, an estimate for b and its 95% confidence intervals. These can then be interpreted as bottleneck size n_b via $n_b = 1/(1 - b)$.

To reiterate, this process is not the same as fitting a Kimura distribution based on summary statistics, as is often used. In that case, we are losing information about the distribution of heteroplasmy samples and will not in general identify the maximum likelihood parameterization.

How does this approach perform on the example problem datasets above? First, consider the heteroplasmy measurements from [Broz et al. \(2022\)](#) which gave a bottleneck size estimate under 1. Using a maximum likelihood fit gives an estimate for b of 0.204 (95% c.i. 0.107–0.354), corresponding to an estimate for n_b of 1.26 (95% c.i. 1.12–1.55), all readily interpretable parameters of the model. More generally, the maximum likelihood approach readily identifies the most likely model parameters given observations ([Fig. 2](#)). In all these cases the confidence intervals on n can readily be computed using likelihood derivatives, immediately giving an interpretable uncertainty on bottleneck size.

Testing fits to the Kimura distribution

[Wonnapijit et al. \(2008\)](#) propose using a Monte Carlo method based on the KS statistic between the empirical cumulative distribution function of a heteroplasmy sample and an ensemble of samples generated from a fitted Kimura distribution. However, this fitting is typically carried out using the method-of-moments, which is not guaranteed to give a distribution that will generate samples with the lowest KS distance from the data. The maximum likelihood approach above does not solve this problem: the maximum likelihood parameterization of the Kimura distribution is also not guaranteed to minimize KS distance from the data. In general, fitting by moments, likelihood, and goodness-of-fit (including KS distance) will give different estimators for finite samples ([Figs. 1 and 2](#)).

It is possible to find the parameterization of the Kimura distribution that is guaranteed to give the minimum KS distance from a

Table 1. Appropriate parameter fits to the Kimura distribution remove statistically significant signals of selection.

| Reference | Specific dataset | P-values: | | Kimura parameter estimates: | | | | |
|------------------------------------------------------|--------------------------------------------------------|-------------------|-------------------------------|----------------------------------|------------------------|------------------------|---------------------------|---------------------------|
| | | Original reported | This study MoM estimate range | This study min KS estimate range | Estimated p from MoM | Estimated b from MoM | Estimated p from min KS | Estimated b from min KS |
| Freyer et al. (2012) | PGC ($0.4 < h_0 \leq 0.6$) | 0.5 | [0.42–0.52] | [0.94–0.95] | 0.537 | 0.935 | 0.531 | 0.949 |
| | PGC ($0.6 < h_0 \leq 0.7$) | 0.5 | [0.37–0.51] | [0.99–1] | 0.637 | 0.930 | 0.638 | 0.945 |
| | PGC (>0.7) | 0.2 | [0.23–0.3] | [0.46–0.48] | 0.760 | 0.880 | 0.773 | 0.899 |
| | Offspring ($h_0 \leq 0.6$) | 0.04 | [0.026–0.045] | [0.54–0.56] | 0.605 | 0.901 | 0.628 | 0.935 |
| | Offspring ($0.6 < h_0 \leq 0.7$) | NA | [0.34–0.55] | [0.87–0.88] | 0.645 | 0.956 | 0.652 | 0.964 |
| | Offspring (>0.7) | 0.03 | [0.017–0.083] | [0.34–0.36] | 0.664 | 0.958 | 0.674 | 0.967 |
| | Oocytes set 1 ^a | 0.31 | [0.24–0.30] | [0.94–0.95] | 0.656 | 0.986 | 0.660 | 0.991 |
| | Oocytes set 2 ^a | 0.51 | [0.44–0.73] | [0.96–0.97] | 0.628 | 0.981 | 0.634 | 0.987 |
| | Oocytes set 3 ^a | 0.08 | [0.035–0.080] | [0.21–0.23] | 0.701 | 0.989 | 0.703 | 0.986 |
| | Oocytes set 4 ^a | 0.42 | [0.37–0.57] | [0.89–0.90] | 0.532 | 0.986 | 0.538 | 0.988 |
| Oocytes set 5 ^a | 0.84 | [0.78–0.88] | [0.93–0.94] | 0.764 | 0.992 | 0.766 | 0.992 | |
| Wonnapijit et al. (2008) (via other primary sources) | Human primary oocytes | 0.827 | [0.37–0.88] | [0.96–0.99] | 0.132 | 0.873 | 0.127 | 0.892 |
| | Mice 515 tail biopsy | 0.646 | [0.52–0.94] | [0.96–0.99] | 0.051 | 0.948 | 0.055 | 0.938 |
| | Mice 515 mature oocytes | 0.049 | [0.02–0.06] | [0.86–0.95] | 0.041 | 0.860 | 0.034 | 0.923 |
| | Mice 517 tail biopsy | 0.75 | [0.44–0.95] | [0.96–0.99] | 0.075 | 0.935 | 0.077 | 0.923 |
| | Mice 517 mature oocytes | 0.834 | [0.66–0.98] | [0.96–0.99] | 0.095 | 0.861 | 0.103 | 0.870 |
| | Mice 603A mature oocytes | 0.681 | [0.56–0.98] | [0.98–1] | 0.009 | 0.979 | 0.009 | 0.975 |
| | Mice 603A primary oocytes | 0.037 | [0.099–0.36] | [0.49–0.74] | 0.013 | 0.978 | 0.012 | 0.984 |
| | Mice 603B mature oocytes | 0.435 | [0.32–0.68] | [0.88–0.98] | 0.031 | 0.969 | 0.033 | 0.973 |
| | Mice 603B primary oocytes | 0.223 | [0.38–0.58] | [0.73–0.88] | 0.025 | 0.973 | 0.026 | 0.969 |
| | <i>Drosophila mauritiana</i> H1 unfertilized eggs | 0.004 | [1e-5–0.12] | [0.049–0.15] | 0.412 | 0.277 | 0.444 | 0.492 |
| | <i>Drosophila mauritiana</i> G20-5 unfertilized eggs | 0.922 | [0.38–0.99] | [0.80–0.99] | 0.319 | 0.926 | 0.324 | 0.929 |
| | <i>Drosophila mauritiana</i> G71-12 unfertilized eggs | 0.542 | [0.60–0.99] | [0.85–0.99] | 0.600 | 0.939 | 0.605 | 0.922 |
| | <i>Drosophila mauritiana</i> H1-31 M unfertilized eggs | 0.587 | [0.12–0.97] | [0.53–0.99] | 0.180 | 0.932 | 0.173 | 0.950 |
| | <i>Drosophila mauritiana</i> H1-18D unfertilized eggs | 0.993 | [0.74–0.99] | [0.92–0.99] | 0.478 | 0.918 | 0.477 | 0.907 |
| | <i>Drosophila mauritiana</i> H1-12B unfertilized eggs | 0.865 | [0.62–0.88] | [0.68–0.94] | 0.812 | 0.828 | 0.824 | 0.851 |
| | <i>Drosophila simulans</i> 6YF16 unfertilized eggs | 0.79 | [0.51–0.98] | [0.85–0.99] | 0.129 | 0.907 | 0.130 | 0.902 |

Comparison of reported P-values (from method-of-moments fits) with the range of P-values from method-of-moment fits in this study, and P-value ranges from minimum KS distance fits. Estimated parameters of the Kimura distribution p and b are also reported in each case. Data are from several original sources where significant departures from the Kimura distribution were reported (Wonnapijit et al., 2008; Freyer et al., 2012). In some cases, our method-of-moments P-value estimate range departs from that originally reported, illustrating the sensitivity of KS P-values on the specific values used (Supplementary Fig. a in Supplementary File 1; see Discussion). See Methods for details of how data were collected and parsed. PGC, primordial germ cell; NA, not available.

^a The ordering of these experiments was not specified in the original publication; we have ordered them to align as much as possible with our estimated MoM P-values.

given dataset, by minimizing the KS distance across parameter values (Fig. 1; see Supplementary File 1). This parameterization can yield much lower KS distances than the two other estimators. In all cases we studied, this means that the P -value computed from the Monte Carlo approach does not pass a significance threshold, even when the other parameterizations (less suited to minimizing KS distance) suggested a significant departure. Some examples are shown in Fig. 1; this is the approach used to identify the alternative parameterization in the motivating example problem above.

In other words, the fact that the Monte Carlo approach gives a low P -value when used with a distribution fitted via moments does not mean that the Kimura distribution is incompatible with observations. It is very possible that a different fit, minimizing KS distance instead of matching moments, will give a P -value that does not pass a significance threshold—so the hypothesis that a Kimura distribution (just not the moment-fitted one) generated the data cannot be discarded. We found this to be the case in every dataset we investigated, including several where a significant departure from the Kimura distribution was previously reported (Wonnapijit et al. 2008; Freyer et al. 2012) (Table 1). Other studies using the Kimura fit did not report significant departures from the Kimura distribution (Monnot et al. 2011; Jokinen et al. 2016; Zhang et al. 2021); in these cases, the minimum KS distance fit will simply yield even higher P -values (as in Fig. 1 and Table 1).

There are some other important points to consider here. First, the KS test as applied in the WCS-K approach relies on sampling and, therefore, does not produce a single, fixed P -value as is typically reported. Instead, a range of P -values arises from the approach, and this range can be quite wide depending on the sampling method used, and whether observations are rounded before testing (Table 1). Second, the P -value range is very sensitive to even small perturbations in the specific heteroplasmy values analyzed. In Supplementary Fig. a in Supplementary File 1 we artificially add noise levels <1% to heteroplasmy observations and find consistent and, in some cases, dramatic increases in the range of P -values reported. Third, and more broadly, it is a well-known statistical result that the KS test cannot be used in the usual way to compare data with a distribution that has been fitted using that same data (Crutcher 1975). As in the WCS-K approach, when a parametric distribution is fitted using data and those data are then compared to it, the nonparametric KS test approach must be corrected.

The fact that judicious parameterizations of the Kimura distribution can fit this wide range of heteroplasmy distributions underlines that the Kimura distribution is remarkably flexible and capable of capturing a wide range of heteroplasmy structures (including all those observed to date, to our knowledge). Despite this flexibility, it is possible to construct a dataset that cannot be well fit (in the KS sense) by any parameterization of the Kimura distribution. An example is given by expanding the above (0, 0.9) example so that there are 50 observations of each value ($P=0.01$ from Monte Carlo KS test using minimum KS distance fit, Supplementary Fig. b in Supplementary File 1). This structure is challenging to fit because, on one hand, the many zeroes suggest either a low p or very high segregation, and the many 0.9s suggest neither can be true. However, this and comparable examples are far removed from any real-world observations of which the authors are aware.

For clarity, it helps to specify the hypotheses that these various approaches test. The original approach tests the hypothesis that a specific Kimura distribution, parameterized by the method-of-moments, commonly generates samples with a higher KS distance from the

distribution than the data's KS distance. Because that particular parameterization is not guaranteed to reflect the true distribution, this test is hard to interpret. The proposed approach tests the hypothesis that any Kimura distribution commonly generates samples with a higher KS distance from the distribution than the data's KS distance. This test is more (but not fully) aligned with the scientific question of whether the data may be generated given the assumptions underlying the Kimura model.

The question of whether samples are unlikely to be drawn from a given family of distributions is complicated. While results like variants of the Anderson–Darling test exist for many distributions (Stephens 2017), to our knowledge these results do not exist for the Kimura distribution. We suggest that such tests are not yet developed enough to provide scientific insight, and instead advocate the likelihood-based testing of alternative hypotheses as in Otten et al. (2018).

Nonparametric solutions I— h -statistics

In some cases, this maximum likelihood approach may be impossible (if we do not have access to individual measurements) or undesirable (if we do not believe that the Kimura distribution, or any other parametric model, is a good description of the system). In such cases, we may be forced to use a nonparametric approach to estimate bottleneck size. Here, there is no way of avoiding some of the issues above, as without a model we cannot naturally enforce scientific constraints on the values involved. With this caveat, sample statistics can be used to provide an estimator of the uncertainty in sample variance (Wonnapijit et al. 2010). However, as shown in our introductory problems, several issues can arise with this approach and require careful interpretation; we also believe that the expressions in Wonnapijit et al. (2010) need some adjustment to be generally applicable.

The variance of the sample variance s^2 is

$$V(s^2) = (1/n)(\mu_4 - (n-3)/(n-1)\sigma^4),$$

which requires two population quantities, the variance σ^2 and the 4th central moment μ_4 , to be estimated from a sample of data.

Wonnapijit et al. (2010) quote a result for a quantity D_4 , which is proposed as an unbiased estimator of the 4th central moment of a distribution

$$D_4 = ((n-1)(n^2 - 3n + 3)/n^3)\mu_4 + (3(2n-3)(n-1)/n^3)\mu_2^2.$$

However, we cannot find a justification for this estimator. In the cited source (Dodge and Rousson 1999), the left-hand side of this equation (Wonnapijit's D_4) is not presented as an estimator of μ_4 , but is the expected value of the sample moment m_4 . The reference states that the expected value of that sample moment is the expression on the right-hand side, not that this is an estimator for the population μ_4 . Indeed, μ_4 (the quantity to be estimated) itself appears on the right-hand side. In tandem, several other key expressions in the WCS-K approach, including for the normal and Kimura special cases, involve population quantities σ^2 , p , b , and even μ_4 itself, which are not directly accessible from a sample.

Happily, all this is resolved by the existence of a unique unbiased symmetric estimator for μ_4 in terms of sample moments, which is the corresponding h -statistic (Dwyer 1937; Halmos 1946; Rose and Smith 2002; the h label here and h for heteroplasmy are a coincidence):

$$h_4 = (3(3-2n)n^2 m_2^2 + n^2(n^2 - 2n + 3)m_4)/((n-3)(n-2)(n-1)n),$$

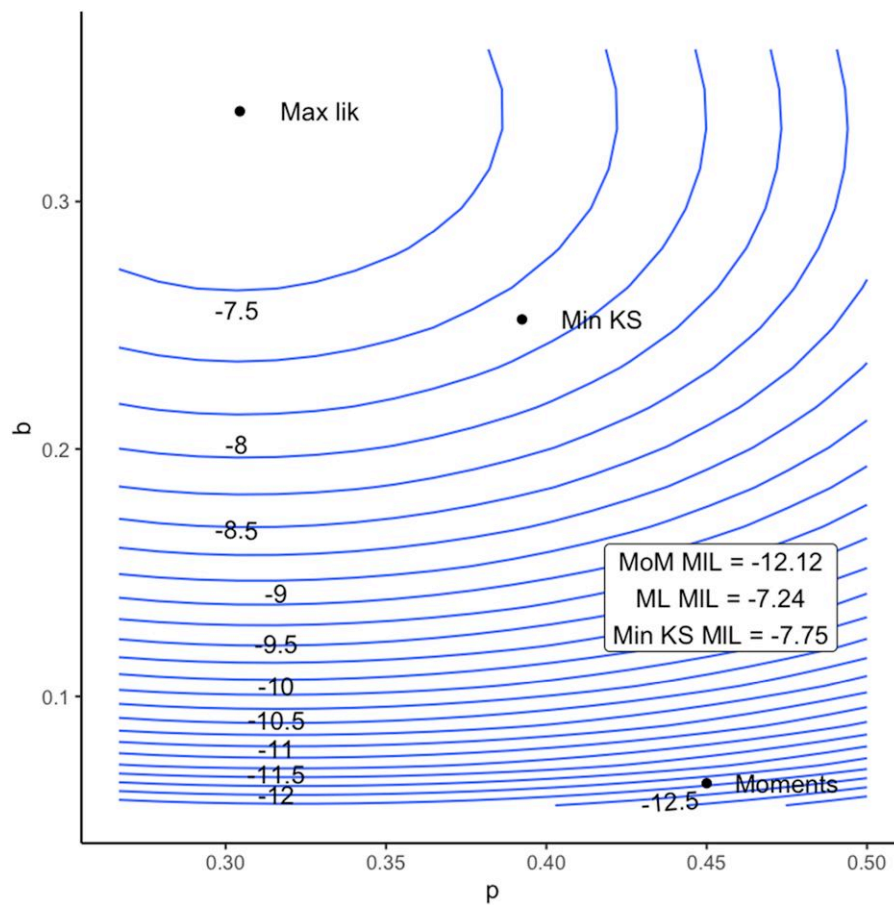


Fig. 2. Likelihoods for different model fits. Likelihood surface for the synthetic dataset in Fig. 1b with different values of the Kimura distribution parameters p and b . The parameter estimates from different approaches (MoM, method-of-moments; Min KS, minimum KS distance; Max lik, maximum likelihood) are given as individual points. The maximum likelihood fit by construction identifies the parameterization with the highest likelihood; the method-of-moments and minimum KS distance fits identify parameterizations some distance from this peak.

where

$$\begin{aligned} m_2 &= 1/n \sum (h - \bar{h})^2 \\ m_4 &= 1/n \sum (h - \bar{h})^4 \\ \bar{h} &= 1/n \sum h \end{aligned}$$

(note that these expressions are all functions of the data sample alone, not population quantities as in the WCS-K expressions). Here we can immediately take our n measurements of h , compute the sample mean \bar{h} and sample moments m_2 and m_4 , and obtain our h_4 estimate of μ_4 for further use.

The expression in Wonnapijit et al. (2010) for μ_4 in the Kimura distribution is correct (the algebra required is in Supplementary File 1) but, as before, this is a population quantity and we cannot in general simply plug in a set of estimators and obtain an unbiased μ_4 estimate. Using the h -statistic immediately resolves this issue.

We, therefore, propose the following estimator for the variance of the sample variance, based directly on sample statistics from the data

$$\hat{V}(s^2) = (1/n)(h_4 - (n-3)/(n-1)s^4).$$

Once this variance estimate has been obtained, its interpretation as confidence intervals requires a parametric choice (for example,

writing ± 1.96 s.e. invokes a normal assumption). To avoid these issues we can use a resampling approach.

Nonparametric solutions II—resampling to estimate variance uncertainty

An alternative approach is possible without relying on a parametric model, particular estimators, and without requiring a parametric choice at any point in the analysis. The bootstrap and jackknife are two algorithms from applied statistics that allow a very general estimation of uncertainty on any statistic of interest, computed by “resampling” the data set (Efron and Tibshirani 1994). This process involves generating a set of new samples from the original sample, related but different, and computing the statistic of interest for each new sample. This set of computed values estimates the true distribution of the statistic of interest. In the bootstrap, B new samples of size n are constructed by sampling with replacement from the original sample. In the jackknife, n new samples of size $n-1$ are constructed by omitting each element of the original sample in turn. We focus on the bootstrap here for simplicity.

Bootstrap estimates for heteroplasmy variance are then constructed by creating B new samples and working out the heteroplasmy variance for each, with the uncertainty in the overall estimate given by the spread of values across this resampled set. There is an important technical point here: resampling with

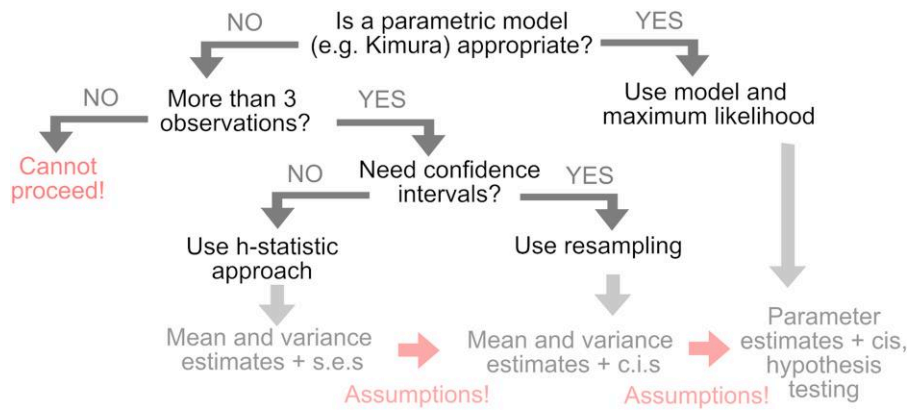


Fig. 3. Choosing approaches to analyze heteroplasmy measurements. A decision tree outlining our proposed use of different statistical approaches to analyze heteroplasmy. The “assumptions” involved are (i) a parametric way of interpreting standard errors as confidence intervals (for example, a normal model of ± 1.96 s.e.) and (ii) a link between the estimated summary statistics of a dataset and the parameters of a generative model. s.e., standard error; c.i., confidence interval.

replacement leads to bias in summaries of dispersion (like variance) because the same observation will often be repeated in a re-sampled set (Wiley 2001). Several methods have been proposed to correct this bias, described and benchmarked in Nguyen (2018). We adopt the simple approach in Wiley (2001) (generalized in Brennan (2007), and supported by the benchmarking across a wide range of cases)—using a correcting factor of $n/(n-1)$ to compensate for the expected bias. The estimate of the standard error of heteroplasmy is then

$$\hat{V}(s^2) = n/(n-1)1/B \sum_i V(\hat{h}_{(i)}),$$

where the sum is over B bootstrap resamples, each giving a re-sampled dataset $\hat{h}_{(i)}$.

A roadmap for heteroplasmy analysis

Taken together, these approaches give us a set of options for the analysis of heteroplasmy data. These options are outlined in the form of a decision tree in Fig. 3.

In Fig. 4, the different approaches for estimating uncertainty in heteroplasmy variance are illustrated for a selection of the (highly segregated) samples in the plant organelle dataset from Broz et al. (2022). In Supplementary Fig. c in Supplementary File 1, these different approaches are illustrated for a wider range of synthetic and real datasets. Across examples, nonparametric (h-statistic and bootstrap) estimates are typically consistent, and parametric fits to the Kimura distribution give more conservative (larger) estimates for $V(h)$ variance.

Bonus results I—comparing bottleneck sizes

Consider the problem of how to compare bottleneck sizes in different systems. We cannot use a t-test—bottleneck size is the reciprocal of a sample variance and cannot possibly be normally distributed. Nonparametric approaches will lose statistical power. Can we use the above idea to perform well-powered testing of hypotheses about bottleneck size?

Yes, readily. Given the ability to perform maximum likelihood estimates for our parametric models, we can use a likelihood ratio test to compare two models: first, where a different bottleneck applies to different observations, and second, where the same bottleneck describes all observations—following Broz et al. (2022). Consider the case where we have two groups, each consisting of several sets of heteroplasmy observations. We are

interested in whether the bottleneck size is the same in the two groups. We consider two model structures. First, each set in each is assumed to be drawn from a Kimura distribution, where p is unique for each set and b is unique for each group. Second, each set is assumed to be drawn from a Kimura distribution, where p is unique for each set and the same b value applies to both groups. Hence

$$L_1 = \prod_j \prod_i \text{Kimura}(h_i | p_i, b_j)$$

$$L_2 = \prod_j \prod_i \text{Kimura}(h_i | p_i, b).$$

We then maximize L_1 and L_2 over p_i and b_j (for model 1) or b (model 2). We can then use the likelihood ratio test to investigate support for model 1 (different bottleneck sizes in the 2 groups) against model 2 (the same bottleneck size in both groups):

$$\Lambda = -2(L_2 - L_1).$$

Comparing Λ to a χ^2 distribution with 1 degree of freedom (capturing the one parameter difference between the two models) will then give a P -value against the null hypothesis of no difference in bottleneck size between the groups.

For example, consider two groups: Group 1 with $h_{11} = (0.2, 0.4)$, $h_{12} = (0.6, 0.7)$ and Group 2 with $h_{21} = (0.1, 0.7)$ and $h_{22} = (0.3, 1)$. The maximum likelihood process above gives estimates for the bottleneck size of 34 (9.4–130) and 2.3 (1.3–5.8), respectively, and the likelihood ratio test gives a P -value of 0.005 against the null hypothesis of equal bottlenecks.

This P -value may seem surprisingly small given the low sample size in our example. But this is the strength of an appropriately chosen parametric approach. It is very unlikely that a single Kimura distribution capable of generating the high-variance pairs would also generate the low-variance ones and vice versa. To draw a parallel with the (perhaps more familiar) t-test, if we have two very distant pairs of internally close observations, it is very unlikely that the same normal distribution would generate them all, and we can obtain an arbitrarily low P -value against this null hypothesis as the pair spacing increases (e.g. (0, 0.01) and (0.99, 1) gives $P < 0.0001$).

Bonus results II—the case of nonzero selection

Paralleling his work on allele distributions under neutral drift (Kimura 1955), Kimura also derived distributions for the case of selective pressure favoring one allele (and for the case where

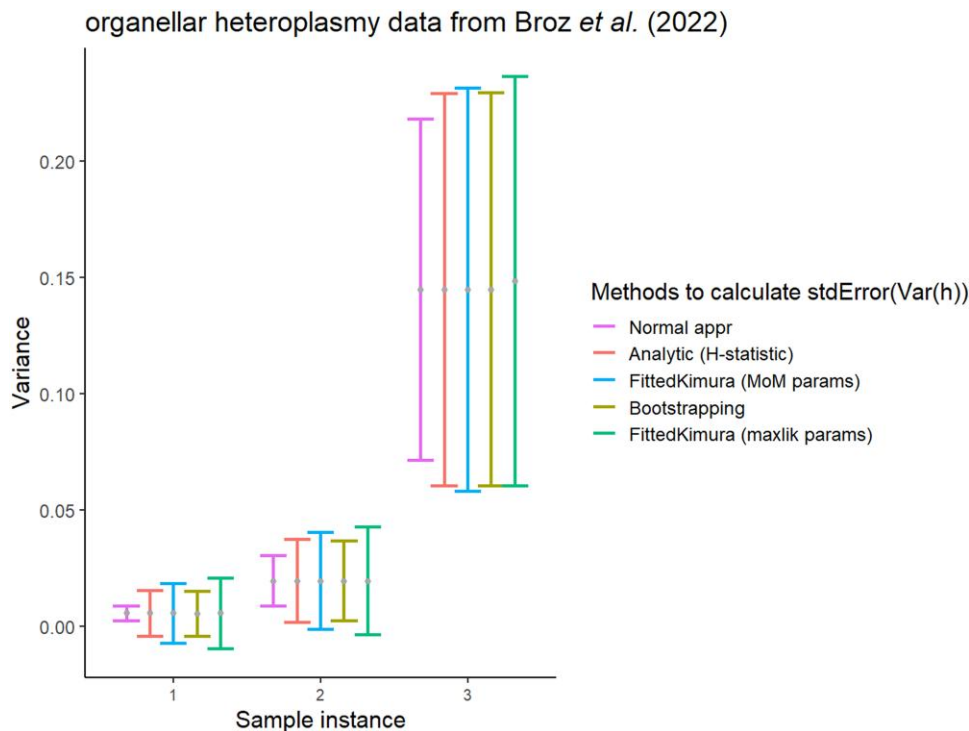


Fig. 4. Different ways to estimate the uncertainty of variance in heteroplasmy samples. The variance of a subset of organellar heteroplasmy samples from Broz et al. (2022). For comparison, error bars are set to twice the estimated standard error of variance. Nonparametric (h -statistic and bootstrap) approaches give generally consistent estimates; parametric fits to the Kimura distribution are more conservative. Method-of-moments and maximum likelihood fits of Kimura parameters can give different estimates of uncertainty, particularly pronounced in the final set. See [Supplementary File 1](#) for examples from other experimental studies.

this selection fluctuates) (Kimura 1954). Wright and Kerr (1954) also made quantitative progress on this question; Kimura (1954) links the two approaches. Interpreting these allele frequency distributions as heteroplasmy distributions, we have a probability density function for heteroplasmy measurements given an effective population size and selection coefficient. Once again, a maximum likelihood approach can be used to estimate these parameters and associated uncertainty. Further, a likelihood ratio test can be used to seek statistical support for the presence of selection at the intracellular level. In Otten et al. (2018), a similar philosophy is used for the intercellular level, with a likelihood ratio test against an alternative hypothesis of a “truncated Kimura” distribution—a Kimura distribution with probability mass removed at high heteroplasmy levels, to model an absence (due to removal or death) of cells with such high heteroplasmy levels. In both cases, a parametric model for the action of selection (within or between cells) is invoked to provide a well-posed statistical test.

Discussion

The work of Wonnapijit et al. (2008, 2010) was groundbreaking in applying statistical methods and stochastic models from population-genetic theory to modern mtDNA observations. The Kimura distribution is a convenient and powerful model for allele frequencies, though as critiqued in Wallace and Chalkia (2013), it is not without issues. Here we suggest that its use within a maximum likelihood setting, rather than using method-of-moments fitting, resolves several issues that have arisen in its application. In particular, we caution against the KS testing of the Kimura distribution fitted by moments. As we have shown, the moment fit does not generally give the parameters that are most compatible

(in the KS sense) with the data, making it very likely that false positive errors occur.

This is fundamentally a statistical story about how different estimators can give different results, and how any results must be interpreted with the estimator in mind. There is nothing intrinsically good or bad about the different estimators (method-of-moments, maximum likelihood, minimum KS distance) that we use here. However, approaches for testing hypotheses with parameters require those parameters to be inferred in a compatible way. When the hypothesis is, for example, any Kimura distribution has a large KS distance from the empirical data, the estimator (minimum KS distance) that minimizes this distance and therefore more strictly tests the hypothesis should be chosen. For estimates of uncertainty in sample variance in real datasets, the difference between different estimators rarely provides a substantial effect. However, for fitting and testing distribution structure, the appropriate estimator is a much more important issue.

Wallace and Chalkia (2013) previously outlined some shortcomings of using this approach, noting that the Kimura model itself relies on several assumptions which are not met in real mtDNA situations and that departures from these assumptions may challenge the model. They also discuss that the use of the KS test to compare the theoretical distribution with empirical data is not without several other issues—including the applicability of its underlying population-genetic assumptions, limited sensitivity, and ability to capture the non-Mendelian dynamics of mtDNA inheritance. Another pitfall with using the KS test arises when one tries to compare data with a distribution whose parameters were estimated by the same data (Crutcher 1975). The issue we highlight here is a different one and stands in parallel with these important points—even if the Kimura model and KS test approach are

accepted, the way that this model is typically fitted with experimental observations makes further testing uninterpretable.

How should we detect selection? In the case of a mechanism that fundamentally changes the family of distributions from which heteroplasmy is drawn, the likelihood ratio test approach of Otten *et al.* (2018) is well suited. Here, support is compared for a truncated Kimura distribution (modeling cell-level selection) and a Kimura distribution (modeling neutrality). The case of intracellular selection is more challenging. Kimura (1954) shows that even in the presence of selection, distributions can closely resemble those from a neutral model. The best approach here is to use longitudinal data (an early reference measurement or a time course) and to fit a model that allows for selection and generates all observations. It is important to note that any early reference measurement will itself be a sample and cannot be regarded as ground truth (i.e. as a population parameter).

The various nonlinearities involved in these expressions and distributions mean that even small differences in individual heteroplasmy measurements—and certainly in estimated parameters—can have dramatic differences in the *P*-values from the analysis (as in Fig. 1b, and several examples in Table 1). Because of this, rounding and binning heteroplasmy values before the WCS-K analysis can have strong effects on the consequent findings (as in some examples in Table 1). The other approaches outlined here are more robust to such small deviations (which will always arise due to measurement noise).

More generally, mtDNA (and plastid DNA) heteroplasmy is a remarkably awkward quantity. If we use the near-universal definition $h = m/(w + m)$ where *m* is the mutant copy number and *w* wildtype copy number, *h* can strictly only take values where *w* and *m* are integers, and where *w + m* is the cellular copy number. So steps smaller than 0.01 are not permitted for a cell with 100 mtDNAs: we can have $h = 0.05$ or $h = 0.06$ but not $h = 0.054$. As the ratio of two random variables, the variance (and mean) of *h* does not have a convenient closed-form representation, even if we have a perfect theory for how *w* and *m* change. The WCS-K model is one attempt to work with *h* by employing a particular set of simplifying assumptions (which have been critically discussed, for example, in Wallace and Chalkia (2013)). Other approaches include attempting to develop predictions for *w* and *m* and then using our approximating for a complicated sum over all possibilities (Johnston *et al.* 2015) or a Taylor expansion approximation for *h* (Johnston *et al.* 2015; Aryaman *et al.* 2019; Hoitzing *et al.* 2019; Edwards *et al.* 2021; Insalata *et al.* 2022). These approaches have merit (and shortcomings (Glastad and Johnston 2023)) when attempting to build a bottom-up theory from microscopic dynamics; the Kimura model is convenient for top-down, data-driven analysis. For this reason, it is a valuable approximation of use in heteroplasmy analysis—providing the estimator used is appropriate for the statistical task at hand.

Data availability

This study did not generate new primary research data. All code is freely available at <https://github.com/StochasticBiology/heteroplasmy-analysis>.

Supplemental material available at G3 online.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 805046

(EvoConBio) to IGJ). DBS and AKB are supported by National Institutes of Health (NIH) Grant R01 GM118046.

Conflicts of interest

The author(s) declare no conflict of interest.

Literature cited

- Aryaman J, Bowles C, Jones NS, Johnston IG. Mitochondrial network state scales mtDNA genetic dynamics. *Genetics*. 2019;212(4):1429–1443. doi:10.1534/genetics.119.302423.
- Brennan RL. Unbiased estimates of variance components with bootstrap procedures. *Educ Psychol Meas*. 2007;67(5):784–803. doi:10.1177/0013164407301534.
- Brown DT, Samuels DC, Michael EM, Turnbull DM, Chinnery PF. Random genetic drift determines the level of mutant mtDNA in human primary oocytes. *Am J Hum Genet*. 2001;68(2):533–536. doi:10.1086/318190.
- Broz AK, Keene A, Gyorfy MF, Hodous M, Johnston IG, Sloan DB. Sorting of mitochondrial and plastid heteroplasmy in Arabidopsis is extremely rapid and depends on MSH1 activity. *Proc Natl Acad Sci U S A*. 2022;119(34):e2206973119. doi:10.1073/pnas.2206973119.
- Burgstaller JP, Johnston IG, Jones NS, Albrechtova J, Kolbe T, Vogl C, Futschik A, Mayrhofer C, Klein D, Sabitzer S, *et al.* MtDNA segregation in heteroplasmic tissues is common in vivo and modulated by haplotype differences and developmental stage. *Cell Rep*. 2014;7(6):2031–2041. doi:10.1016/j.celrep.2014.05.020.
- Burgstaller JP, Kolbe T, Havlicek V, Hembach S, Poulton J, Piálek J, Steinborn R, Rüllicke T, Brem G, Jones NS, *et al.* Large-scale genetic analysis reveals mammalian mtDNA heteroplasmy dynamics and variance increase through lifetimes and generations. *Nat Commun*. 2018;9(1):2488. doi:10.1038/s41467-018-04797-2.
- Crutcher HL. A note on the possible misuse of the Kolmogorov-Smirnov test. *J Appl Meteorol* (1962–1982). 1975;14(8):1600–1603. doi:10.1175/1520-0450(1975)014<1600:ANOTPM>2.0.CO;2
- de Stordeur E, Solignac M, Monnerot M, Mounolou JC. The generation of transplasmic *Drosophila simulans* by cytoplasmic injection effects of segregation and selection on the perpetuation of mitochondrial DNA heteroplasmy. *Mol General Genet MGG*. 1989;220(1):127–132. doi:10.1007/BF00260866.
- Dodge Y, Rousson V. The complications of the fourth central moment. *Am Stat*. 1999;53(3):267–269. doi:10.1080/00031305.1999.10474471
- Dwyer PS. Moments of any rational integral isobaric sample moment function. *Ann Math Stat*. 1937;8(1):21–65. doi:10.1214/aoms/1177732451.
- Edwards DM, Røyrvik EC, Chustecki JM, Giannakis K, Glastad RC, Radzvilavicius AL, Johnston IG. Avoiding organelle mutational meltdown across eukaryotes with or without a germline bottleneck. *PLoS Biol*. 2021;19(4):e3001153. doi:10.1371/journal.pbio.3001153.
- Efron B, Tibshirani RJ. An introduction to the Bootstrap. Boca Raton (FL): CRC press; 1994.
- Freyer C, Cree LM, Mourier A, Stewart JB, Koolmeister C, Milenkovic D, Wai T, Floros VI, Hagström E, Chatzidaki EE, *et al.* Variation in germline mtDNA heteroplasmy is determined prenatally but modified during subsequent transmission. *Nat Genet*. 2012;44(11):1282–1285. doi:10.1038/ng.2427.
- Glastad RC, Johnston I. Mitochondrial network structure controls cell-to-cell mtDNA variability generated by cell divisions. *PLoS*

- Comput Biol. 2023;19(3):e1010953. doi: 10.1371/journal.pcbi.1010953
- Halmos PR. The theory of unbiased estimation. *Ann Math Stat.* 1946; 17(1):34–43. doi:10.1214/aoms/1177731020.
- Hoitzing H, Gammage PA, Haute LV, Minczuk M, Johnston IG, Jones NS. Energetic costs of cellular and therapeutic control of stochastic mitochondrial DNA populations. *PLoS Comput Biol.* 2019; 15(6):e1007023. doi:10.1371/journal.pcbi.1007023.
- Insalata F, Hoitzing H, Aryaman J, Jones NS. Stochastic survival of the densest and mitochondrial DNA clonal expansion in ageing. *Proc Natl Acad Sci U S A.* 2022;119(49):e2122073119. doi:10.1073/pnas.2122073119.
- Jenuth JP, Peterson AC, Fu K, Shoubridge EA. Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nat Genet.* 1996;14(2):146–151. doi:10.1038/ng1096-146.
- Johnston IG. Varied mechanisms and models for the varying mitochondrial bottleneck. *Front Cell Dev Biol.* 2019;7:294. doi:10.3389/fcell.2019.00294.
- Johnston IG, Burgstaller JP, Havlicek V, Kolbe T, Rülcke T, Brem G, Poulton J, Jones NS. Stochastic modelling, Bayesian inference, and new in vivo measurements elucidate the debated mtDNA bottleneck mechanism. *Elife.* 2015;4:e07464. doi:10.7554/eLife.07464.
- Jokinen R, Marttinen P, Stewart JB, Neil Dear T, Battersby BJ. Tissue-specific modulation of mitochondrial DNA segregation by a defect in mitochondrial division. *Hum Mol Genet.* 2016; 25(4):706–714. doi:10.1093/hmg/ddv508.
- Kimura M. Stochastic processes and distribution of gene frequencies under natural selection. In: *Cold Spring Harbor Symposia on Quantitative Biology (Vol. 20)*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1954. p. 33–53.
- Kimura M. Solution of a process of random genetic drift with a continuous model. *Proc Natl Acad Sci U S A.* 1955;41(3):144–150. doi: 10.1073/pnas.41.3.144.
- Monnot S, Gigarel N, Samuels DC, Burlet P, Hesters L, Frydman N, Frydman R, Kerbrat V, Funalot B, Martinovic J, et al. Segregation of mtDNA throughout human embryofetal development: m. 3243A>G as a model system. *Hum Mutat.* 2011;32(1): 116–125. doi:10.1002/humu.21417.
- Nguyen NT. Evaluation of using the bootstrap procedure to estimate the population variance [thesis]. Steven F Austin State University; 2018. Retrieved from <https://scholarworks.sfasu.edu/etds/157>.
- Otten AB, Sallevelt SC, Carling PJ, Dreesen JC, Drüsedau M, Spierts S, Paulussen AD, de Die-Smulders CE, Herbert M, Chinnery PF, et al. Mutation-specific effects in germline transmission of pathogenic mtDNA variants. *Hum Reprod.* 2018;33(7):1331–1341. doi:10.1093/humrep/dey114.
- Rose C, Smith MD. *Mathstata: mathematical statistics with mathematics*. In: *Compstat*. Heidelberg: Physica; 2002. p. 437–442.
- Samuels DC, Wonnapijit P, Chinnery PF. Preventing the transmission of pathogenic mitochondrial DNA mutations: can we achieve long-term benefits from germ-line gene transfer? *Hum Reprod.* 2013;28(3):554–559. doi:10.1093/humrep/des439.
- Stephens MA. Tests based on EDF statistics. In: *Goodness-of-fit Techniques*. New York (NY): Routledge; 2017. p. 97–194.
- Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet.* 2015;16(9):530–542. doi:10.1038/nrg3966.
- Wallace DC, Chalkia D. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harb Perspect Biol.* 2013;5(11):a021220. doi:10.1101/cshperspect.a021220.
- Wei W, Tuna S, Keogh MJ, Smith KR, Aitman TJ, Beales PL, Bennett DL, Gale DP, Bitner-Glindzicz MA, Black GC, et al. Germline selection shapes human mitochondrial DNA diversity. *Science.* 2019; 364(6442):eaau6520. doi:10.1126/science.aau6520.
- Weston S; Microsoft Corporation. *foreach: Provides Foreach Looping Construct*; 2020.
- Weston S; Microsoft Corporation. *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*; 2022.
- Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016.
- Wickham H, Hester J, Chang W. *devtools: Tools to Make Developing R Packages Easier*. New York: Springer-Verlag; 2020.
- Wiley EW. Bootstrap strategies for variance component estimation: theoretical and empirical results [thesis]. Stanford University; 2001. Retrieved from <https://www.proquest.com/docview/304729620>.
- Wonnapijit P, Chinnery PF, Samuels DC. The distribution of mitochondrial DNA heteroplasmy due to random genetic drift. *Am J Hum Genet.* 2008;83(5):582–593. doi:10.1016/j.ajhg.2008.10.007.
- Wonnapijit P, Chinnery PF, Samuels DC. 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.* 2010;86(4):540–550. doi:10.1016/j.ajhg.2010.02.023.
- Wright S, Kerr WE. Experimental studies of the distribution of gene frequencies in very small populations of *Drosophila melanogaster*. II. *Bar*. *Evolution.* 1954;8(2):775–781. doi:10.2307/2405441
- Zhang H, Esposito M, Pezet MG, Aryaman J, Wei W, Klimm F, Calabrese C, Burr SP, Macabelli CH, Viscomi C, et al. Mitochondrial DNA heteroplasmy is modulated during oocyte development propagating mutation transmission. *Sci Adv.* 2021;7(50):eabi5657. doi:10.1126/sciadv.abi5657.

Editor: J. Comeron