
Copyright:

© The Author 2016. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

DOI link to article:

http://dx.doi.org/10.1093/hmg/ddv626

Date deposited:

12/05/2016

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 Unported License
Mitochondrial DNA sequence characteristics modulate the size of the genetic bottleneck

Ian J. Wilson1,†, Phillipa J. Carling1,2,†, Charlotte L. Alston2,3, Vasileios I. Floros4,5, Angela Pyle1,2, Gavin Hudson1,2, Suzanne C.E.H. Sallevelt6, Costanza Lamperti7, Valerio Carelli8,9, Laurence A. Bindoff10,11, David C. Samuels12, Passorn Wonnapinij13, Massimo Zeviani4,7, Robert W. Taylor2,3, Hubert J.M. Smeets6, Rita Horvath1,2 and Patrick F Chinnery2,4,5,*

1Institute of Genetic Medicine, 2Wellcome Trust Centre for Mitochondrial Research and 3Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK, 4Medical Research Council Mitochondrial Biology Unit, Cambridge, UK, 5Department of Clinical Neurosciences, School of Clinical Medicine, University of Cambridge, Cambridge, UK, 6Department of Clinical Genetics, Research Schools GROW/CARIM, Maastricht University Medical Center, Maastricht, Netherlands, 7Division of Molecular Neurogenetics, National Neurological Institute ‘C. Besta’, Milano, Italy, 8IRCCS Institute of Neurological Sciences of Bologna, Bellaria Hospital, Bologna, Italy, 9Unit of Neurology, Department of Biomedical and NeuroMotor Sciences (DIBINEM), University of Bologna, Bologna, Italy, 10Department of Neurology, Haukeland University Hospital, Bergen, Norway, 11Department of Clinical Medicine (K1), University of Bergen, Bergen, Norway, 12Vanderbilt Genetics Institute, Department of Molecular Physiology and Biophysics, Vanderbilt School of Medicine, Nashville, TN, USA and 13Department of Genetics, Faculty of Science, Kasetsart University, Bangkok, Thailand

*To whom correspondence should be addressed at: Department of Clinical Neurosciences, University of Cambridge, Addenbrookes Biomedical Campus, Cambridge CB2 0QQ, UK. Tel: +44 1223217091; Fax: +44 122333694; Email: pf25@medschl.cam.ac.uk

Abstract

With a combined carrier frequency of 1:200, heteroplasmic mitochondrial DNA (mtDNA) mutations cause human disease in ∼1:5000 of the population. Rapid shifts in the level of heteroplasmy seen within a single generation contribute to the wide range in the severity of clinical phenotypes seen in families transmitting mtDNA disease, consistent with a genetic bottleneck during transmission. Although preliminary evidence from human pedigrees points towards a random drift process underlying the shifting heteroplasmy, some reports describe differences in segregation pattern between different mtDNA mutations. However, based on limited observations and with no direct comparisons, it is not clear whether these observations simply reflect pedigree ascertainment and publication bias. To address this issue, we studied 577 mother–child pairs transmitting the m.11778G>A, m.3460G>A, m.8344A>G, m.8993T>G/C and m.3243A>G mtDNA mutations. Our analysis controlled for inter-assay differences, inter-laboratory variation and ascertainment bias. We found no evidence of selection during transmission but show that different mtDNA mutations segregate at different rates in human pedigrees. m.8993T>G/C

†I.J.W. and P.J.C. contributed equally to the study.

Received: September 2, 2015. Revised: November 26, 2015. Accepted: December 22, 2015

© The Author 2016. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
segregated significantly faster than m.11778G>A, m.8344A>G and m.3243A>G, consistent with a tighter mtDNA genetic bottleneck in m.8993T>G/C pedigrees. Our observations support the existence of different genetic bottlenecks primarily determined by the underlying mtDNA mutation, explaining the different inheritance patterns observed in human pedigrees transmitting pathogenic mtDNA mutations.

Introduction

First described in 1988, point mutations of mitochondrial DNA (mtDNA) have emerged as a major cause of maternally inherited human disease (1,2). Pathogenic mtDNA mutations causing a severe multisystem phenotype are usually heteroplasmic, with a mixture of mutated and wild-type mtDNA present in the same individual. The percentage level of mutated mtDNA largely determines whether a biochemical defect is expressed on a cellular level, and the inherited level of heteroplasmy correlates more or less with the severity of the clinical phenotype (3).

Rapid intergenerational shifts in the level of mtDNA heteroplasmy levels were first observed in Holstein cows (4,5) and subsequently documented in human pedigrees transmitting pathogenic mtDNA mutations. A restriction in the number of mitochondrial genomes repopulating the female germ line (the mtDNA bottleneck) is thought to explain this phenomenon, supported by observations in mice (6,7). Although the initial analysis of human pedigrees implied differences in the mtDNA bottleneck between different families (8), the analysis of aggregate data did not support earlier reports (9). Limited data analysis pointed towards a random genetic drift mechanism acting on all heteroplasmic mtDNA mutations (9), but in some instances, there appeared to be evidence of selection in favour of mutated genomes (10), even after minimising the effects of ascertainment bias. These conflicting data lead to two fundamental unanswered questions: do all mtDNA mutations behave the same during inheritance, and is there selection for or against different levels of heteroplasmy? Clarifying the underlying trends is key to providing reliable recurrence risks for patients with mtDNA diseases, and understanding the underlying mechanisms involved may open new avenues for preventative treatment in the future (11).

To address this issue, we have performed the largest analysis of inherited mtDNA heteroplasmic mutations in humans to date. Our findings show that differences in the behaviour of the mtDNA bottleneck between specific pathogenic mtDNA mutations, explaining the variability in clinical inheritance pattern observed in human pedigrees transmitting different mtDNA mutations.

Results

Determining the potential impact of ascertainment bias

Given previous concerns about ascertainment bias when studying the inheritance of heteroplasmy in human pedigrees (9), first we performed a simulation experiment to determine the possible consequences of identifying pedigrees through a clinically affected child. We then determined whether the standard approach of omitting the affected proband minimizes any bias to an acceptable level.

The simulations were based on an established model for the mtDNA genetic bottleneck using measurements of heteroplasmy made in human oocytes for neutral alleles (i.e. with no selection) (12,13). We studied the difference in heteroplasmy level between a mother and child (ΔM-O) in simulated pedigrees in silico, with three possible strengths of the mtDNA genetic bottleneck (where bottleneck parameter, b = 0.9, is a weak bottleneck; b = 0.7 is an intermediate strength bottleneck; and b = 0.2 is a strong bottleneck). We then modelled the effects of ascertaining the whole pedigrees (Fig. 1) using three forms of ascertainment: (1) through a mother (A Muk), (2) through an affected child (A OC) or (3) through the other unaffected child, but disregarding data contributed by the affected child (A OC)—the latter mimicking the conventional approach used to minimize ascertainment bias by omitting to analyse data from an affected proband. For a full description, see the Materials and Methods and Figure 1.

The simulations confirmed previous assumptions (9) that ascertainment through an affected child (A OC) leads to a skewed distribution of ΔM-O, creating the false impression that there is selection bias in favour of the mtDNA mutation (Fig. 2, blue line compared with the grey histogram). This was most prominent when the bottleneck was narrow (i.e. b was low, a strong bottleneck). However, when families were ascertained through the mother (A Muk), or the data from the affected child (A OC) were not included, the ΔM-O distribution resembled the entire data set before sampling (Fig. 2, red and green lines compared with the grey histogram), showing that omission of the proband (A OC) minimized the ascertainment bias.

Analysis of human pedigree data

To generate the largest data set possible, we studied pedigrees derived from a meta-analysis of published data (n = 532 mother–child pairs transmitting five common heteroplasmic mtDNA mutations: m.11778G>A (n = 117), m.3460G>A (n = 74), m.8344A>G (n = 96), m.8993T>G/C (n = 117) and m.3423A>G (n = 128), citations shown in the Material and Methods); and 45 new unpublished mother–child pairs [m.8344A>G (n = 9), m.8993T>G/C (n = 2) and m.3423A>G (n = 34)] measured in two centres. Given that there was no obvious difference between the two data sets (Fig. 3), we merged all of the data and minimized ascertainment bias by analysing separately data with and without the clinically affected probands. The final data set included 467 mother–child pairs for the uncorrected data. m.3243A>G was analysed before and after correcting for the known decrease in leucocyte heteroplasmy levels for this specific mutation using the published approach (14). For this correction, only individuals with a heteroplasmy level of <95% were included to avoid pairs where the mother or offspring values corrected to >100%. This reduced the sample size of m.3243A>G from 137 to 99 pairs. The entire data set was used to determine the likely size of the bottleneck parameter, b, using the model described previously (12,13), incorporating the laboratory assay and laboratory site as covariates. Bayesian statistical analyses were performed using JAGS (15).

For each mutation, the average change in heteroplasmy was not significantly different from zero, consistent with no selection for or against the mtDNA mutations during transmission. The posterior differences in bottleneck strength, b, are shown in Figure 4. As predicted from the simulations, the bottleneck strength was overestimated (i.e. b is low, a narrower bottleneck) when we included pairs ascertained through a clinically affected proband. However, after the exclusion of affected probands, we observed a difference in the strength of the bottleneck parameter, b, estimated for different mtDNA mutations. m.8993T>G/C showed the largest difference,
closely followed by m.3460A>G. A two tailed Bayesian hypothesis test was used to determine whether the bottleneck parameter, b, was significantly stronger for m.8993T>G/C than for m.11778G>A, m.8344A>G and m.3460G>A. This test gave estimated posterior probabilities of 0.031, 0.044 and 0.54. The bottleneck parameter distribution for the uncorrected m.3243A>G data was distorted, likely due to the known confounding effect of changing heteroplasmy levels with age (Fig. 4, far right hand panel) (14). We therefore used age-corrected data in the Bayesian hypothesis test comparing the bottleneck parameter, b, for m.8993T>G/C to m.3243A>G, which revealed an estimated posterior probability of 0.001. No other tests reached statistical significance at the 5% level.

Discussion

Our analysis accounted for several sources of potential variability, including different analytical techniques to measure heteroplasmy levels, different laboratories, differences between individual pedigrees themselves and differences in maternal age (16). Having minimised the likely effects of ascertainment bias, we observed different rates of heteroplasmy segregation during maternal transmission between mtDNA mutations. Patients harbouring the m.8993T>G/C mutation showed more rapid segregation of heteroplasmy levels than any of the other mutations. Although this trend has been suspected for some time (17), the analysis we present here provides the first direct comparison and a formal demonstration that m.8993T>G/C behaves differently to other pathogenic mtDNA mutations. Given that the size of the mitochondrial genetic bottleneck is the primary factor determining the rate of segregation during transmission, our findings indicate differences in the behaviour of the mtDNA bottleneck based on the underlying mtDNA genotype.

The more rapid segregation observed in m.8993T>G/C pedigrees is consistent with earlier reports of major shifts in heteroplasmy observed in pedigrees transmitting this mutation (17,18). This explains why severely affected children are often the only affected individuals in families transmitting this mutation, and why their mothers are usually asymptomatic. Conversely, the less dramatic rates of segregation seen for other mtDNA mutations explain why a wide range of heteroplasmy values are seen in different family members, with oligo-symptomatic individuals transmitting the mutation in larger pedigrees, with multiple moderate and severely affected family members (19,20). Our findings also provide a potential explanation for the relative frequency of different pathogenic mutations in epidemiological studies (for example, m.3243A>G being much more common than m.8993T>G/C), despite the same background frequency of healthy carriers in the population with low heteroplasmy levels (21,22). It should be noted, however, that this discussion relates to the observed statistical trends, and although uncommon, rapidly segregating families with m.3243A>G (23) and slowly segregating families with m.8993T>G/C have been described (24,25) and are also consistent with our findings.

Studies in mice have shown a dramatic reduction in the amount of mtDNA within single cells at an early stage in mammalian germ cell development (6,7). This reduction is sufficient to explain the segregation of mtDNA heteroplasmy in mice (6). Mathematical models predict that subtle differences in the size of the mtDNA genetic bottleneck at this critical period will have a dramatic impact on the rate of segregation of mtDNA heteroplasmy (26). Given emerging evidence that the amount of mtDNA within cells can be influenced by the mtDNA sequence (27,28), it is plausible that the different mtDNA mutations we have studied cause differences in the amount of mtDNA within the developing germ line (29), either through a replication advantage or as a compensation for the lower rates of oxidative phosphorylation and higher levels of reactive oxygen species production (28,30). The differences in mtDNA level would lead to differences in the rate of segregation through the genetic bottleneck. Alternatively, selection either for or against a particular mutation would lead to differences in the rate of segregation (31). First described by population genetic theory (13,32), the effective bottleneck size, Ne, can be defined like the effective population size during genetic drift. Studies in isolated reduced populations have shown that selective pressures can influence Ne, without directly influencing the true population size. The apparent difference in bottleneck size between the different mtDNA mutations could thus reflect the selection pressure and need not necessarily be caused by a difference in the actual amount of mtDNA during germ cell development. Although it will be fascinating to determine the underlying molecular mechanisms, this will not alter our conclusions, or the relevance of our findings for women transmitting the mtDNA mutations we have studied.

These findings have important implications for our understanding for the recurrence risks of heteroplasmic mtDNA diseases, for the prevention of mtDNA disease using conventional techniques such as prenatal and pre-implantation diagnosis, and also suggest that therapies aimed at manipulating a germ cell mtDNA content could influence the underlying rate of segregation of pathogenic mtDNA mutations in humans. On the one hand, increasing the mtDNA content in germ cells could slow down segregation and thereby prevent the expression of severe disease in subsequent generations. On the other hand, factors known to reduce mtDNA content could lead to more rapid segregation of heteroplasmic alleles increasing the probability of generating germ cells with very low heteroplasm levels (as well as very high heteroplasm levels, which may not be viable at an early stage in pregnancy). Finally, environmental factors leading to selection pressure could also alter the behaviour of the genetic bottleneck, as described earlier. Given the recent observation of widespread low-level mtDNA heteroplasm (33), these reductions in mtDNA content could lead to the emergence of pre-existing pathogenic variants, previously present at a very low level in the female germline. With recent evidence that common genetic polymorphisms of mtDNA can influence mtDNA levels (28), it is conceivable that geographic, ethnic and even inter-familial differences in segregation occur, influenced by non-synonymous single base-pair substitutions on the background mtDNA haplotype. This could explain differences in the rates of segregation of the same pathogenic mtDNA mutation, which have been described in different families, and also influence the segregation of low-level heteroplasm, providing a potential mechanism for the association of mtDNA haplogroups with common late-onset human diseases (34).

Materials and Methods

Heteroplasmic levels in mothers and offspring

We studied the transmission of mtDNA heteroplasm in both published and new unpublished pedigrees.

Ascertained of published data

Published pedigrees transmitting m.11778G>A, m.3460G>A, m.8344A>G, m.8993T>G/C and m.3423A>G were identified through a systematic review of the literature. We included all papers where the age, clinical status and laboratory methods were clearly recorded, and there was at least one heteroplasmic individual. These parameters, along with the laboratory location,
1. Simulate a population of heteroplasmic mothers (red), each with two offspring (blue) using an established bottleneck model (see text).

2. Determine the likelihood that a mother is clinically affected and thus ascertained based on her heteroplasm level.

3. Identify mother-offspring pairs through the affected mother, $A_{AM}$.

4. Remove the affected mothers from the original distribution. Bold colors = ascertained data, light colors = non-ascertained data.

5. Determine the likelihood that a child is clinically affected and thus ascertained based on the heteroplasm level.

6. Identify mother-offspring pairs through the affected child, $A_{AC}$.

7. Identify mother-offspring pairs through the unaffected other child, $A_{OC}$.
were incorporated in subsequent analyses. We identified 77 publications, containing blood DNA heteroplasm levels from 532 mother–child pairs transmitting 5 common heteroplasmic mitochondrial DNA mutations: m.11778G>A (n = 117 mother–child pairs) (35–45), m.3460G>A (n = 74) (35,37,46–51), m.8344A>G (n = 96) (52–62), m.8993T>G/C (n = 117) (24,62–84) and m.3423A>G (n = 128) (20,54,85,86,87,88,89,90,91,92,93,94,95–105).

**Unpublished data**

Unpublished pedigrees transmitting the m.8344A>G (n = 9 mother–child pairs), m.8993T>G/C (n = 2) and m.3423A>G (n = 34) mutations were identified from the following accredited diagnostic laboratories: Newcastle, UK; Maastricht, NL; Milan and Bologna, Italy; Bergen, Norway and Munich, Germany. Genomics DNA was analysed in two laboratories: Newcastle, UK; and Maastricht, NL.

**Quantification of heteroplasmy: Newcastle**

Heteroplasmy was measured by quantitative pyrosequencing. Pyromark Assay Design Software v.2.0 (Qiagen) was used to design primers for template generation, one of which contains a biotinylated tag (*BIO), and the pyrosequencing reaction (Seq). For m.3243A>G: F-*BIO-TAAAGCCCCCTACTTACACAAAGCG, R-CCGATTAGAATGGTACATTGAG, Seq-ATGCCATTACGGGGG; for m.8344A>G: F-*BIO-CATGGCCCATCGTACGGAAT, R-CTTTAATGCGCTTTGCTTGAGG, Seq-TAGAGCTTAAAGATTAAGAGA; and for m.8993T>G/C F-AGGCAACCTACACCCCTTA, R-*TGTGAAAACGTAGGCTTGGAT, Seq-TAAAGTTAAAGATTAAGAGA; and for m.8993T>G/C F-AGGCAACCTACACCCCTTA, R-*TGTGAAAACGTAGGCTTGGAT, Seq-TAAAGTTAAAGATTAAGAGA. Product templates were generated with a GoTaq® DNA Polymerase (Promega) reaction according to manufacturer’s protocol. Pyrosequencing was performed using the Pyromark Q24 platform according to the manufacturer’s protocol, using the designed pyrosequencing primers for each mutation. Pyromark Q24 software was used to quantify the heteroplasmy levels of each mutation through comparison of the relevant peak heights of both wild-type and mutant mtDNA. The accuracy of the pyrosequencing assay was determined by generating wild-type and mutant clones, which when mixed at the correct proportions mimicked a range of heteroplasmy levels between 0 and 100% for each mutation. Each mixed sample was assessed for mutation

---

**Figure 1.** Overview of the inheritance model. Simulations used to determine the potential effects of ascertainment bias through a mother, and affected child, or an unaffected child. Mothers = red; offspring = blue; unaffected offspring = green. Darker distributions represent ascertained pedigrees; lighter distributions represent non-ascertained pedigrees. The probability of a family being recruited depends on an individual developing the disease, which in turn is correlated with the heteroplasmy level. The probability that an individual is recruited, $p_i$, based on a distribution that is zero below a level of heteroplasmy $d_i$, and one above heteroplasmy $d_i$, with a linear increase between these points. The probability of being recruited with a heteroplasmy of $\Delta M = d$ is the range of heteroplasmies where the probability of disease increases linearly from 0 (lower level) to 1 (upper level) (see legend to Fig. 1 for details of the simulation model).

**Figure 2.** Modelling the effects of ascertainment bias on simulated pedigrees transmitting mtDNA heteroplasmy. Frequency distribution histograms for the difference between maternal and offspring heteroplasm levels ($\Delta M$) for three different values of the bottleneck strength (bottleneck parameter, $b = 0.9$, a weak bottleneck; $b = 0.7$, an intermediate bottleneck; and $b = 0.2$, a strong bottleneck), and for three different methods of ascertainment: sampling the families where there was an affected mother—(ascertainment = affected mother, AAM, red); sampling the families where there was an affected child—(ascertainment = affected child, AAC, blue); and sampling the families through an affected child, only but including the transmission from the mother to the other child—(ascertainment = other child, ACO, green). The grey histogram gives the simulated distribution before any ascertainment, thus corresponding to data without any ascertainment bias. Each row reflects different thresholds heteroplasmy values required to cause disease. The parameter $d$ models the threshold for disease resulting from the pathogenic mutation. The range given for $d$ is the range of heteroplasmies where the probability of disease increases linearly from 0 (lower level) to 1 (upper level) (see legend to Fig. 1 for details of the simulation model).
Figure 3. Relationship between the level of mtDNA heteroplasmy in mothers and offspring for five pathogenic mtDNA mutations. Maternal and offspring heteroplasmy levels are displayed as a proportion. Red symbols = affected proband, green symbols = mother of the proband, blue = other relatives. Circles = meta-analysis data, triangles = new pedigree data measured at Centre 1, squares = new pedigree data measured at Centre 2. m.3243A>G corrected refers to heteroplasmy values after a correction for the known decline in heteroplasmy levels with age (14).

Figure 4. Differences in the size of the mitochondrial DNA genetic bottleneck of five pathogenic mtDNA mutations. Violin plots show the probability density for a given bottleneck strength parameter, $b$, calculated from actual heteroplasmy measurements in human pedigrees shown in Figure 3. All = all of the mother–child pairs. No probands = mother–child pairs remaining after the exclusion of an affected proband, thus minimising ascertainment bias. ‘m.3243A>G corrected’ refers to heteroplasmy values after a correction for the known decline in heteroplasmy levels with age (14).
load using pyrosequencing and used to generate a standard curve (Supplementary Material, Fig. S1).

**Quantification of heteroplasmy: Maastricht**

Heteroplasmy was measured by semiquantitative restriction fragment length polymorphism (RFLP) analysis using mutation-specific restriction enzymes. Primers used for m.3243A>G were as follows: CAACCTTATGATTACCCACACAACTTAGATTATA CCCACAC (forward) and TTGTCTGGGTTAAACCCT (reverse); for m.3243A>G: TCCTCTTGCTTAATCCTGC (forward) and GTG TATTAGTTGGGCAATTTACCTGTAAGCCGGTGTTG (reverse); and for m.8993T>C/G: CACACTCACCCTTTATCCC (forward) and TCATTATTGTGTGTGTGTTCAG (reverse). For each forward primer, an unlabelled and FAM-labelled primer was available, polymerase chain reaction (PCR) was performed on the GeneAmp PCR System 9700 (Perkin–Elmer Applied Biosystems) in a total volume of 50 µl, containing 1× PCR buffer (Invitrogen) and 1 U of Taq DNA polymerase (Invitrogen), 2 mm MgCl2 (Invitrogen) and 0.1 mm dNTP (Pharmacia). Unlabelled forward (15 pmol) and reverse primer (3 pmol) sets were used in the first round. First round PCR started with 5 min denaturation at 94°C followed by 32 cycles of 1-min denaturation at 94°C, 45 s annealing at 53°C and 45-s elongation at 72°C, followed by a final elongation step of 7 min at 72°C. 15 ml of first round PCR product was adjusted to 50 µl with second round PCR mix, containing 1× PCR buffer (Invitrogen) 1 U of DNA polymerase, 2 mm MgCl2 (Invitrogen) and a fluorescently labelled forward primer (15 pmol). One final amplification cycle was performed of 5 min denaturation at 94°C, 1-min annealing at 53°C and 7-min elongation at 72°C. The labelled second round PCR product (15 ml) was digested in the appropriate digestion buffer in a total volume of 50 µl containing 10 U HaeIII (10 U/µl; Roche) for the m.3243A>G mutation, 10 U BglI (10 U/µl) for the m.3244A>G mutation and 10 U HhaII (10 U/µl; Roche) for the m.8993T>C/G mutation. The m.3243A>G-amplicon contains an additional HaeIII restriction site as an internal control for restriction enzyme digestion completion. For each mutation, a sample with known mutation load was included as reference. After digestion, samples were purified using a QIAquick PCR purification kit (Qiagen). Samples were analysed by capillary electrophoresis on an ABI Prism 3730 Genetic Analyser followed by GeneScan analysis. Each PCR product was divided by the sum of the peak area of the wild-type and mutant peak. The ratio was calculated using the peak area peak area of the wild-type and mutant peak. The ratio was calculated using the peak area of the wild-type and mutant peak. The ratio was calculated using the peak area of the wild-type and mutant peak. The ratio was calculated using the peak area.

**Simulation of mtDNA heteroplasmy inheritance**

An overview of the simulation is shown in Figure 1. First, we generated a population of simulated mothers with different heteroplasmy levels. We then modelled the inheritance of heteroplasmy from each of these mothers to two offspring (see below). Next we determined the likelihood of the individuals being clinically affected. We then sampled the simulated families based on the following ascertainment criteria, in the following order:

1. Sampling the families where there was an affected mother—(A_m). Here, we included the transmission from the mother to both offspring.

2. Sampling the families where there was an affected child—(A_c), but the mother was not affected. Here, we also included the transmission from the mother to both offspring.

3. Sampling the families through an affected child, only but including the transmission from the mother to the other child—(A_o).

This sampling algorithm mimicked the different ways that mother–child pairs could be identified in a study of real human pedigrees, and the resulting difference in distributions for the three different pairs (Supplementary Material, Fig. S2) resembles those from the real data from the human pedigrees shown in Figure 3.

We then compared the sampled mother–child pairs with the original simulated pedigrees to determine whether the ascertainment method influenced the observed distribution of heteroplasmy transmissions.

**Simulating the transmission of mtDNA heteroplasmy (the mtDNA genetic bottleneck)**

We modelled the mtDNA genetic bottleneck using established model derived from measurements of heteroplasmy made in human oocytes (12,13). Here, the bottleneck is modelled as an infinite population of mtDNA molecules that was instantaneously reduced to the minimum bottleneck size N for g generations, before expanding back to a large size. We simulated the transmission of mtDNA heteroplasmy of a neutral allele from the mother to two offspring through the bottleneck by approximating with a beta distribution with parameters α = p b/(1−b) and β = (1−p)b/(1−b), where b is the drift parameter b = exp(−g/N). A large b indicates a weak (or wide) genetic bottleneck. When g is large compared with N, b approaches 0, corresponding to a very strong (or narrow) bottleneck. Using this model, we generated a series of simulated heteroplasmy transmissions from mothers starting with a range of heteroplasmy values.

**Statistical methods**

Changes in heteroplasmy between mother and child were modelled using the same beta distribution as for the simulation, using a hierarchical model with a different bottleneck parameter b for each mtDNA mutation. Measurements of heteroplasmy were assumed to be normally distributed about the true values, with censored at 0 and 1, and with an error rate ϑ for publication i, to allow for different measurement technologies. Writing mi for the observed maternal heteroplasmy for mother i, and m_j for the heteroplasmy for child j of mother i, where j = 1, . . . , n, and n_i is the number of children for Mother i,

\[
m_i \sim N(\mu_i, \sigma_i^2), \quad \alpha_j \sim N(\eta_j, \sigma_j^2), \quad \text{and} \quad n_j \sim \text{Beta}(\mu_j b_{m_i}, (1−\mu_j)(1−b_{m_i}))
\]

where \(b_h = \exp(−(g/N_{ij}))\), m_i gives the mutation for Mother i and r_i gives the publication for mother i. To test for systematic increases or decreases in the level of heteroplasmy (i.e. selection), an additional model was built. This allowed for changes in the expected heteroplasmy in the offspring, by multiplying the expected child heteroplasmy, \(\mu_j\), by a random factor, \(s_i\), which models selection by increasing or decreasing the average heteroplasmy value. The JAGS statistical package (15) was used to make inferences about the strength of the bottleneck using the statistical model above. A Bayesian framework was used with uniform priors between 0
and 1 for the $b_k$, and broad priors for the error parameters $\sigma_k^2$. The model and all JAGS and R scripts are available through the following URL: github.com/ijwilson/mtBottleneck.

**Bayesian comparison of bottleneck sizes**

Maternal heteroplasmy values were corrected for age as described (14). Bayesian hypothesis tests of $b_j$ versus $b_k$ were calculated using the empirical joint posterior distributions of $b_j$ and $b_k$ from JAGS (15) and using $2 \max(P(b_j > b_k), P(b_k > b_j))$. The JAGS script is available through the following URL: github.com/ijwilson/mtBottleneck.

**Supplementary Material**

Supplementary Material is available at HMG online.

**Acknowledgements**

P.F.C. is a Wellcome Trust Senior Fellow in Clinical Science (101876/Z/13/2), and a UK NIHR Senior Investigator, who receives support from the Medical Research Council Mitochondrial Biology Unit (MC_UP_15012), the Wellcome Trust Centre for Mitochondrial Research (096919/Z/11/Z), the Medical Research Council (UK) Centre for Translational Muscle Disease research (G0601943), EU FP7 TIRCON, and the National Institute for Health Research (NIHR) Biomedical Research Centre based at Cambridge University Hospitals NHS Foundation Trust and the University of Cambridge. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. R.H. is supported by the European Research Council (309548). P.F.C. and R.H. are supported by the Mitochondrial European Educational Training, ITN Marie Curie People (317433). R.W.T. (309548). P.F.C. and R.H. are supported by the Mitochondrial European Educational Training, ITN Marie Curie People (317433). R.W.T. also receives support from The Lily Foundation and the UK NHS Specialist Commissioners, which funds the ‘Rare Mitochondrial Disorders of Adults and Children’ Diagnostic Service in Newcastle upon Tyne. C.L.A. is the recipient of an NIHR doctoral fellowship (NIHR-HCS-D12-03-04). We acknowledge the ‘Cell lines and DNA Bank of Paediatric Movement Disorders and Neurodegenerative Diseases’ of the Telethon Network of Genetic Biobanks (grant GTB12001J) and the EurobiobanK Network. This work was supported by the Medical Research Council, the Pierfranco and Luisa Mariani Foundation, Telethon Grant GGP11011 and ERC Advanced Grant FP7-322424. The publication is the work of the authors, and Patrick Chinnery will serve as guarantor for the contents of this paper.

Conflict of Interest statement. None declared.

**Funding**

Funding to pay the Open Access publication charges for this article was provided by the Wellcome Trust.

**References**


