Supplemental Material

Transmission of human mtDNA heteroplasmy in the Genome of the Netherlands families: support for a variable size bottleneck

Mingkun Li, Rebecca Rothwell, Martijn Vermaat, Manja Wachsmuth, Roland Schröder, Jeroen F. J. Laros, Mannis van Oven, Paul I. W. de Bakker, Jasper A. Bovenberg, Cornelia M. van Duijn, Gert-Jan B. van Ommen, P. Eline Slagboom, Morris A. Swertz, Cisca Wijmenga, Genome of the Netherlands Consortium, Manfred Kayser, Dorret I. Boomsma, Sebastian Zöllner, Peter de Knijff, and Mark Stoneking
Processing mtDNA data

All reads were aligned with BWA (Li and Durbin 2009) to the hg19 human reference sequence, supplemented with the mtDNA revised Cambridge Reference Sequence (rCRS; GenBank accession number NC_012920) (Anderson et al. 1981; Andrews et al. 1999). Reads that mapped to nuclear DNA were removed, and the remaining reads were then remapped to the rCRS (with the first 500 bp of the rCRS copied to the end to account for the circularity of the mtDNA genome). A consensus sequence for each individual was called using the majority rule, and then the reads for that individual were remapped to this consensus sequence. The average coverage per site across the mtDNA genome was 1212X (117X-3559X).

Unexpectedly, the data generated at one of the four centers (Groningen) had systematically lower coverage of the mtDNA genome (Figure S1), despite having the same overall whole genome coverage as the other three centers. We have so far been unable to identify any difference in DNA extraction, processing of samples, sequencing, or downstream bioinformatics processing that could account for this difference in mtDNA coverage. However, the lower coverage did not result in any systematic differences with respect to number of heteroplasmies detected or average MAF (Table S1). Moreover, most of our analyses explicitly take coverage into account and hence are not influenced by the systematic difference in coverage between Groningen and the other centers. Where coverage could be an issue, we omitted the data from Groningen and repeated the analyses using just the data from the three centers with similar coverage levels; these analyses gave essentially the same results as the analyses that included the Groningen data.

After removing long mononucleotide/dinucleotide repeats (specifically, np 302-316, 513-526, 566-573, and 16181-16194), we used the following criteria to call heteroplasmies: a minimum minor allele frequency on each strand of 2% (a lower threshold increases false positives from sequencing error and NUMTs (Li et al. 2012; Li and Stoneking 2012)); at least three reads on each strand with the minor allele; a DREEP quality score (Li and Stoneking 2012) of 10 or more; coverage of at least 50X at that np in that individual; and coverage within 20-200% of the genome average. Note that the DREEP approach includes measures to control for false heteroplasmies caused by NUMTs (Li et al. 2012), and these criteria have been tested extensively with simulated data as well as empirical data from both rho-zero cells (which lack mtDNA and hence any mtDNA-associated reads are derived from NUMTs) and from artificial mixtures, and shown to accurately identify heteroplasmies with a false positive rate of <1% (Li et al. 2012; Li and Stoneking 2012). The local alignment around each inferred heteroplasmic position was further inspected manually to ensure that alignment issues with
potential nearby indels were not producing false inferences. As detecting heteroplasmies for indels requires a different approach, they are not considered here but will be the focus of another study.

**Droplet digital PCR**

A subset of the inferred heteroplasmies were selected for independent verification via droplet digital PCR (ddPCR), which was performed as described previously (Li et al. 2015). Six positions (Table S4) were analyzed in a total of 33 individuals, chosen to encompass a wide range of minor allele frequencies (Table S5). Briefly, standard PCR assays were prepared in a volume of 20 μL and containing two allele-specific probes labeled with different fluorescent dyes (Table S4); assays were then partitioned into ~20,000 emulsion droplets that each contained on average one template DNA molecule. After PCR the fluorescence was read and the minor allele frequency was estimated from droplets containing exactly one template molecule, as described previously (Li et al. 2015).

**Potential contamination**

Before we received the data, contamination was assessed in the whole genome sequence data and potential contaminants removed (Genome of the Netherlands 2014). Potential contamination in the mtDNA data was called if the minor alleles at five or more heteroplasmic positions in an individual could define an alternative haplogroup. While this procedure could miss contamination involving sequences that differ by less than five mutations, only about 0.7% of pairwise comparisons of the parental mtDNA genome sequences in the GoNL data differ by less than five mutations. In addition, a sample was regarded as contaminated if more than 80% of the heteroplasmies could be explained by contamination from another GoNL sample. Five samples (one father, three mothers, and one offspring) showed evidence of potential contamination. For the four parents showing evidence of contamination, the entire trio was removed, while the offspring showing evidence of contamination was a twin and hence only that sample was removed, thereby converting the twin quartet into a trio. In total 13 samples were removed, leaving 756 samples (228 trios, 8 DZ twin quartets, and 10 MZ twin quartets) for further analysis.

**Estimating the size of the bottleneck during mtDNA transmission**

We aim to estimate the size and nature of the bottleneck during the inheritance of mitochondria based on the change in minor allele frequency of heteroplasmic mtDNA sites transmitted from mother to offspring. We considered four models: a constant size bottleneck model, in which each mtDNA genome is a segregating unit and the bottleneck size does not vary between individuals; a
variable size bottleneck model, in which each mtDNA genome is a segregating unit and the bottleneck size is allowed to vary between individuals; a constant size nucleoid model, in which a nucleoid containing a variable number of identical mtDNA genomes (with mean = 7.5 genomes per nucleoid) is the segregating unit and the bottleneck size does not vary between individuals; and a variable size nucleoid model, in which a nucleoid containing a variable number of identical mtDNA genomes (with mean = 7.5 genomes per nucleoid) is the segregating unit and the bottleneck size is allowed to vary between individuals.

We first describe the most basic model: a constant size bottleneck with the transmission of individual mitochondria. Let \( n \) be the size of the bottleneck, \( m_{\text{obs}} \) the number of copies of the minor allele in the mother, \( m_N \) the total number of reads in the mother, \( c_{\text{obs}} \) the number of copies of the minor allele in the offspring, and \( c_N \) the total number of reads in the offspring. We aim to maximize \( L(n|m_{\text{obs}}, c_{\text{obs}}, m_N, c_N) \). To this end, we model the bottleneck as sampling \( n \) mtDNA genomes with \( x \) copies of the minor allele where each transmitted mtDNA genome is sampled independently from a large number of maternal mtDNA genomes. We calculate the probability of observing \( c_{\text{obs}} \) given \( m_{\text{obs}} \) when \( n \) mtDNA genomes are transmitted:

\[
L(n|m_{\text{obs}}, c_{\text{obs}}, m_N, c_N) \propto P(c_{\text{obs}}|n, m_{\text{obs}}, m_N, c_N) = \sum_{x=0}^{n} \frac{P(c_{\text{obs}}|x, n, m_{\text{obs}}, m_N, c_N)P(x|n, m_{\text{obs}}, m_N, c_N)}{B} \frac{P(x|n, m_{\text{obs}}, m_N)}{A}
\]

(1)

This simplification arises because by conditioning on \( x \): \( c_{\text{obs}} \) given \( x \) and \( n \), \( c_N \) is independent of \( m_{\text{obs}} \) and \( m_N \), and \( x \) depends only on \( m_{\text{obs}} \) and \( m_N \). Therefore, this likelihood consists of two expressions: (A) the probability of transmitting \( x \) minor alleles in a bottleneck of size \( n \); and (B) the probability of observing \( c_{\text{obs}} \) minor alleles in a mature offspring, conditional on \( x \) and \( n \). To calculate (A), we consider the probability of sampling \( m_{\text{obs}} \) minor alleles in the mother by integration over the unknown maternal minor allele frequency, \( m_f \):

\[
P(x|n, m_{\text{obs}}, m_N)
= \int_{0}^{1} P(x|m_{\text{obs}}, m_N, n, m_f)P(m_f|m_{\text{obs}}, n, m_N)dm_f
= \int_{0}^{1} \frac{P(x|m_f, n)P(m_f|m_{\text{obs}}, n, m_N)}{A_1} \frac{P(m_f|m_{\text{obs}}, n, m_N)}{A_2} dm_f
\]

(2)
In the first expression of this equation, \( A_1 \), \( x \) given \( m_f \) is independent of \( m_{obs}, m_N \). \( A_1 \) is binomial, giving the probability of observing \( x \) minor alleles given \( n \) transmitted mtDNA genomes with probability \( m_f \). The second expression of this equation (\( A_2 \)) is the probability that the underlying maternal minor allele frequency is \( m_f \) given the observed maternal minor allele count, \( m_{obs} \), which is independent of \( n \). In this expression, consider the possibility of genotyping errors among the maternal reads. Then \( m_t \), the true unknown number of maternal alleles in the sample carrying the derived allele, may differ from the observed number \( m_{obs} \). We then calculate \( A_2 \) as:

\[
P(m_f|m_{obs}, m_N) = \sum_{m_t=0}^{m_N} \left[ \frac{p(m_t|m_f, m_N)P(m_f|m_N)}{\int_0^1 p(m_t|m_f, m_N)P(m_f|m_N)dm_f} \right] P(m_t|m_{obs}, m_N) \tag{3}
\]

\[
P(m_t|m_{obs}, m_N) = \sum_{m_t=0}^{m_N} \left[ \frac{p(m_t|m_f, m_N)P(m_f|m_N)}{\int_0^1 p(m_t|m_f, m_N)P(m_f|m_N)dm_f} \right] P(m_t|m_{obs}, m_N) \tag{4}
\]

To model the genotyping errors, given by the probability \( P(m_t|m_{obs}, m_N) \) in equation (3), we incorporated a position-specific error rate \( \varepsilon \), which was estimated as the average minor allele frequency at that position across all individuals (irrespective of the specific minor allele observed, so \( \varepsilon \) is position-specific but not allele-specific); the average \( \varepsilon \) for the data used in the bottleneck size estimation was 0.0008. Then \( P(m_t|m_{obs}, m_N) \) can be calculated using Bayes’ rule (equation (5) and reduces to equation (6)), the probability of observing the maternal minor allele count based on the sequencing error \( \varepsilon \). Equation (6) is made up of two additional binomials that model: the probability that \( i \) minor alleles were correctly called minor (with probability \( (1-\varepsilon) \)); and the probability that the remaining \( m_{obs} - i \) alleles were incorrectly called minor (with probability \( \varepsilon \)).

\[
P(m_t|m_{obs}, m_N) = \frac{p(m_{obs}|m_t, m_N)P(m_t|m_N)}{p(m_{obs}|m_{obs}, m_N)} = P(m_{obs}|m_t, m_N) \tag{5}
\]

\[
P(m_t|m_{obs}, m_N) = \sum_{i=0}^{\min(m_t, m_{obs})} \binom{m_t}{i} (1-\varepsilon)^i \varepsilon^{m_t-i} \left( \frac{m_{obs}-m_t}{m_{obs}-i} \right) (\varepsilon)^{m_{obs}-i} (1-\varepsilon)^{(m_{obs}-m_t)-(m_{obs}-i)} \tag{6}
\]

Note that in equation (5), \( m_t \) is independent of \( n \) given \( m_{obs} \) and \( m_N \). Furthermore, we are assuming a uniform prior on \( (m_t|m_N) \), so \( P(m_t|m_N)P(m_{obs}|m_N) \) is constant with \( m_t \).

In equation (3), \( m_f \) given \( m_t \) is independent of \( m_{obs} \), such that \( P(m_f|m_t, n, m_{obs}, m_N) = P(m_f|m_t, n) \) is the probability of the true underlying maternal minor allele frequency \( m_f \) given that our sample of size \( m_N \) contains \( m_t \) true minor alleles. We calculate this using Bayes’ rule, as shown in equation (4), in which \( P(m_t|m_f) \) is the binomial probability of \( m_t \) minor alleles in a sample of size \( m_N \), drawn from the underlying frequency \( m_f \). We assume a uniform prior on \( m_f \).
We focus now on the probability of the observed minor allele count in the child, given \( x \) transmitted minor alleles and \( n \) transmitted mtDNA genomes (probability \( B \) in equation (1)). We model the three processes that occur after transmission: (1) replication within the child to achieve the final minor allele count \( (x_f) \) and final total allele count \( (n_f) \) in the child from the bottleneck size of \( n \) transmitted mtDNA genomes (drift); (2) sampling from this final population; and (3) genotyping error in our sample. This probability is obtained by summing over all possible values of the true offspring minor allele count \( c_t \), where \( c_N \) is the total coverage in the child:

\[
P(c_{\text{obs}}|x,n,c_N) = \sum_{c_t}^N P(c_{\text{obs}}|c_t, x, n, c_N) P(c_t|x, n, c_N) = \\
\sum_{c_t}^N \left[ \frac{P(c_{\text{obs}}|c_t, c_N)}{B_1} \sum_{x_f} \frac{P(c_t|x_f, n_f, c_N) P(x_f|x, n, n_f)}{B_2} \right] \\
\sum_{c_t}^N \left( \sum_{x_f} \frac{P(c_t|x_f, n_f, c_N) P(x_f|x, n, n_f)}{B_2} \right) \\
\sum_{c_t}^N \left( \sum_{x_f} \frac{P(c_t|x_f, n_f, c_N) P(x_f|x, n, n_f)}{B_2} \right) \\
\sum_{c_t}^N \left( \sum_{x_f} \frac{P(c_t|x_f, n_f, c_N) P(x_f|x, n, n_f)}{B_2} \right)
\]

The first expression in equation (7), \( B_1 \), arises because \( c_{\text{obs}} \) is independent of \( x \) and \( n \) given \( c_t \) and \( c_N \). Then \( B_1 \) is the sequencing error probability, calculated as in equation (6). The second term, \( B_2 \), arises by conditioning on \( x_f, n_f \), and \( c_N \), such that \( (c_t|x_f, n_f, c_N) \) is independent of \( x \) and \( n \). Then, \( B_2 \) is a binomial that corresponds to observing \( c_t \) minor alleles after sampling \( c_N \) mitochondria from the adult offspring mtDNA population where minor alleles are sampled with probability \( x_f/n_f \).

The last portion, \( B_3 \), models the replication process to the full size offspring population from the bottleneck size at transmission. In \( B_3 \), \( x_f \) is independent of \( c_N \) given \( x, n \). Expanded in equation (8), \( B_3 \) is calculated using a modified Moran model without replacement (Moran 1958). Briefly, we assume that in the initial population, there are \( n \) mtDNA genomes with \( x \) (\( n > x > 0 \)) carrying minor alleles. At each replication event, one genome is chosen at random. This mtDNA is replicated and added to the population, increasing the population size by one. We repeat this process until we reach the assumed final population size of \( n_f=1000 \), based on the known copy number of mtDNA of \( 10^3-10^4 \) (Shoubridge 2000; Lan et al. 2008). Based on this model the probability of the final minor allele count can be calculated (for details see section below on modeling the replication process):

\[
P(x_f|x, n, n_f) = \binom{n_f-n}{x_f-x} \binom{xf-1}{x-1} \binom{n_f-x_f-1}{n-x-1} \binom{xn-x-1}{n_f-1}
\]

Combining these equations, we return to the overall summation in equation (1). Maximizing equation (1) gives the maximum-likelihood estimate of \( n \) for an individual site. The estimate across all sites is the joint likelihood, obtained by multiplying the individual likelihoods across all sites.

Building on the most basic model of a constant size bottleneck, we construct three more complex models. The variable size bottleneck model differs from the constant size bottleneck model by modeling \( n \), the number of mtDNA genomes transmitted to the child, as a Poisson distributed
The estimate of \( \lambda \) can be obtained by maximizing the likelihood of \( \lambda \) while summing over the unknown values of \( x \) and \( n \):

\[
L(\lambda|m_{obs}, c_{obs}) = P(c_{obs}|\lambda, m_{obs}, m_N, c_N) \propto \sum_{n=0}^{\infty} P(c_{obs}|n, m_{obs}, m_N, c_N)P(n|\lambda) \quad (9)
\]

Because the upper limit of \( n \) is infinite for a Poisson distribution, we calculate this sum until \( P(n|\lambda) \) reaches a lower limit (arbitrarily set at \( 10^{-10} \)).

The third model, the constant size bottleneck with nucleoids, differs from the first two models in that the estimate of \( n \) now represents the number of nucleoids transmitted to the child, with each nucleoid containing only identical copies of either the major allele or the minor allele. We assume each nucleoid \( i \) has a random size \( g_i \), \( i = 0 \ldots n \) modeled as a Poisson-distributed random variable with mean \( \lambda = 7.5 \) (based on empirical studies that find that each nucleoid has 5-10 mtDNA genomes (Jacobs et al. 2000; Cao et al. 2007; Khrapko 2008)). Without loss of generality, the first \( x \) groups contain the minor allele. This gives \( \sum_i g_i \) as the total number of transmitted mitochondria and \( \sum_i x g_i \) as the total number of copies of the minor allele. Under this nucleoid model, we adjust \( B_3 \) in equation (7) which models the replication process to the full size offspring population from the bottleneck size at transmission. Using the same model of replication, we now assume that in the initial population, there are \( \sum g_i \) mtDNA genomes with \( \sum x g_i \) carrying minor alleles. Because we lack a closed form equation for all possibilities of the Poisson-distributed random sizes of \( g_i \), we use a Monte-Carlo approximation to calculate this term. The other terms of equation (7), \( B_1 \) and \( B_3 \), are again made up of the sequencing error probability and the probability of having \( c_t \) minor alleles in our sample of size \( c_N \) given \( x_f \) and \( n_f \), the final minor allele count and final total allele count in the child after replication. The remainder of the maximum-likelihood estimation was calculated as for the constant size bottleneck model.

Finally, we consider the variable size bottleneck with nucleoids. Similar to the variable size bottleneck model in which each mtDNA genome is a segregating unit, this model differs from the constant size bottleneck model with nucleoids in that we now estimate \( \lambda \), the mean of a Poisson distributed random variable that represents the mean number of nucleoids transmitted to the child. The estimate of \( \lambda \) can be obtained by maximizing the likelihood of \( \lambda \) while summing over the unknown values of \( x \) and \( n \), as in equation (9), with \( \lambda \) now representing the mean number of nucleoids transmitted to the child.

**Modeling the replication process**
Assume that in the initial population, there are \( n \) individuals with \( x \) \((n > x > 1)\) carrying minor alleles. One individual is chosen at random, copied, and added to the next generation, increasing the population size by one. Then for the first generation of \( n + 1 \) individuals, the probability of \( k_1 \) individuals carrying minor alleles is binomial:

\[
P(k_1 | x) = \begin{cases} 
\frac{n-x}{n} & \text{if } k_1 = x \\
\frac{x}{n} & \text{if } k_1 = x + 1 \\
0 & \text{otherwise}
\end{cases}
\]

Similarly, for the second generation of \( n + 2 \) individuals, the probability of \( k_2 \) individuals carrying minor alleles is based on two binomial samplings:

\[
P(k_2 | x) = \begin{cases} 
\frac{n-x}{n} \left(\frac{n-x+1}{n+1}\right) & \text{if } k_2 = x \\
2 \frac{x(n-x)}{n(n+1)} & \text{if } k_2 = x + 1 \\
\frac{x}{n} \left(\frac{x+1}{n+1}\right) & \text{if } k_2 = x + 2 \\
0 & \text{otherwise}
\end{cases}
\]

Therefore, the closed form for the probability of observing \( x + z \) minor alleles in generation \( j \) is:

\[
P(k_j = x + z | x) = \binom{j}{z} \frac{(x + z - 1)!}{(x - 1)!} \frac{((n - x) + (j - z - 1))!}{(n - x - 1)!} \frac{(n - 1)!}{(n + (j - 1))!}
\]

We apply this approach to the replication within the child to achieve the final minor allele count \( x_f \) and final total allele count \( n_f \) in the child from the bottleneck size of \( n \) transmitted mtDNA genomes (drift) with \( x \) carrying minor alleles. The probability of observing \( x_f \) after \( n_f - n \) generations:
\begin{align*}
P(x_f | x, n, n_f) &= P(k_{n_f-n} = x + (x_f - x) | k_n = x) \\
&= \frac{(n_f - n)}{(x_f - x)} \frac{(x + (x_f - x) - 1)!}{(x - 1)!} \frac{(n - x) + (n_f - n) - (x_f - x) - 1) \right) !}{(n - x - 1)!} \frac{(n - 1)!}{(n + ((n_f - n) - 1)!)} \\
&= \left( \frac{(n_f - n)}{(x_f - x)} \right) \frac{(x_f - 1)!}{(x - 1)!} \frac{(n_f - x_f - 1)!}{(n - x - 1)!} \frac{(n - 1)!}{(n_f - 1)!}
\end{align*}
Figure S1. Average coverage for the mtDNA genome in the GoNL data. Top, average coverage across the mtDNA genome for data generated from the four centers. Bottom, box plots of the coverage. The mtDNA coverage was systematically lower for the samples processed in Groningen, for unknown reasons.
Figure S2. Comparison of the MAF estimated via sequencing to that estimated via ddPCR for a subset of the data.
Figure S3. The number of heteroplasmies per individual follows a Poisson distribution. The plot shows the observed number of heteroplasmies (open circles) in each of the 492 mothers and fathers, and the expected number (vertical lines) based on the Poisson distribution.
Figure S4. Coverage at heteroplasmic positions. Box plots are shown for heteroplasmies that were transmitted or not transmitted from mothers, and for heteroplasmies that were either received or not received (i.e., arose *de novo*) in the offspring. There are no significant differences in coverage between either transmitted and non-transmitted heteroplasmies ($P=0.401$, Mann-Whitney $U$ test) or between inherited and non-inherited heteroplasmies ($P=0.391$, Mann-Whitney $U$ test).
Figure S5. Number of synonymous and nonsynonymous heteroplasmies with different minor allele frequencies.
Figure S6. Proportion of nonsynonymous mutations in different functional impact categories in high level (MAF≥0.05) vs. low level heteroplasmy (MAF<0.05). Nonsynonymous mutations were categorized in terms of likely functional impact on the protein as high risk, medium risk, low risk, or neutral.
Figure S7. Distribution of the number of heteroplasmies for polymorphic alleles that were present in mothers but not present in offspring (disappeared heteroplasmies). To control for the effect of frequency differences between novel alleles and polymorphic alleles, for each novel allele one polymorphic allele which had a similar allele frequency (difference ≤ 0.03) was randomly retrieved. Among 100000 resamplings, none of the polymorphic alleles had the number of disappearing events equal to or higher than the observed number of disappearing events for novel alleles (empirical P-value < 0.00001).
Table S1. Number of individuals sequenced, mtDNA and whole genome coverage, number of heteroplasmies, and average minor allele frequency (MAF) for each of the four centers contributing to the GoNL project. The number of heteroplasmies per sample is significantly higher in the Rotterdam population than in the other populations (Mann-Whitney U tests: Rotterdam vs. Groningen, p=0.0042; Rotterdam vs. Leiden, p=0.035; Rotterdam vs. Amsterdam, p=0.00331). This may reflect the higher coverage for the samples from Rotterdam. There were no significant differences among populations with respect to the distribution of average MAF (after Bonferroni correction for the number of tests). Only unrelated individuals (fathers and mothers) were used.

| Populations | Number of samples | Average mtDNA coverage | Number of samples with ≥1 heteroplasmy | Total number of heteroplasmies | Number of heteroplasmies per sample | Average MAF |
|-------------|------------------|------------------------|----------------------------------------|-------------------------------|-------------------------------------|------------|
| Rotterdam   | 122              | 1464                   | 74                                     | 111                           | 0.91                                | 0.101      |
| Groningen   | 106              | 304                    | 47                                     | 63                            | 0.59                                | 0.120      |
| Leiden      | 48               | 1315                   | 21                                     | 29                            | 0.60                                | 0.129      |
| Amsterdam   | 216              | 1384                   | 101                                    | 139                           | 0.64                                | 0.108      |
Table S2. Heteroplasmies identified in the MZ twin quartets.

| Trio | Position | Major allele | Minor allele | Mother | MZ1 | MZ2 | P-value |
|------|----------|--------------|--------------|--------|-----|-----|---------|
| A105 | 379      | A            | G            | 1      | 0.975* | 0.973* | 1       |
| A148 | 195      | T            | C            | 0.793  | 0.558*** | 0.584*** | 0.448   |
| A148 | 11080    | T            | C            | 0.938  | 0.914 | 0.925 | 0.384   |
| A148 | 16093    | C            | T            | 0.973  | 0.983 | 0.984 | 1       |
| A163 | 16104    | T            | C            | 0.953  | 0.984*** | 0.983** | 0.899   |
| A164 | 16124    | T            | C            | 0.994  | 0.969*** | 0.985   | 0.0009  |
| A173 | 16234    | C            | T            | 0.629  | 0.325*** | 0.317*** | 0.749   |

Asterisks indicate P-values (Fisher’s exact test) for differences in major allele frequency between each twin (MZ1 or MZ2) and the mother: *, P<0.001; **, P<0.00001; ***, P<0.0000001

1Major allele frequency
2Fisher’s exact test of the null hypothesis: no difference in major allele frequencies in MZ1 vs. MZ2
| Trio | Position | Major allele | Minor allele | Mother$^1$ | DZ1$^1$ | DZ2$^1$ | P-value$^2$ |
|------|----------|--------------|--------------|------------|----------|----------|------------|
| A124 | 195      | T            | C            | 0.994      | 0.883*** | 1        | 8.8x10$^{-22}$ |
| A124 | 385      | A            | G            | 0.964      | 1***     | 1**      | 1          |
| A124 | 15848    | A            | G            | 0.974      | 1***     | 1***     | 1          |
| A125 | 16220    | A            | G            | 0.991      | 0.970*   | 1*       | 3.8x10$^{-13}$ |
| A127 | 709      | G            | A            | 1          | 0.974**  | 0.998    | 9.6x10$^{-6}$ |
| A127 | 2600     | A            | G            | 1          | 1        | 0.952*** | 3.9x10$^{-26}$ |
| A128 | 16256    | C            | T            | 0.976      | 1***     | 0.999*** | 0.21       |
| A177 | 14470    | T            | C            | 1          | 1        | 0.861*** | 4.2x10$^{-66}$ |
| A178 | 7980     | A            | G            | 0.633      | 0.332*** | 0.145*** | 3.3x10$^{-42}$ |

Asterisks indicate P-values (Fisher’s exact test) for differences in major allele frequency between each twin (DZ1 or DZ2) and the mother: *, P<0.001; **, P<0.00001; ***, P<0.0000001

$^1$Major allele frequency

$^2$Fisher’s exact test of the null hypothesis: no difference in major allele frequencies in DZ1 vs. DZ2.
Table S4. Primer and probe sequences used in the ddPCR verification of a subset of the heteroplasmies. Numbers refer to nucleotide positions; F and R refer to the forward and reverse primers used to amplify the sequence surrounding each position; the designation “Probe” followed by a small letter indicates the probe sequence used to detect that allele (with the variable position indicated by small letters in the sequence.

| Position and primer/probe | Sequence ( 5´-3´) | 5’Modification |
|---------------------------|-------------------|----------------|
| 16093_F                   | GTTCTTTTCATGGGGAAGCAG |               |
| 16093_R                   | GGGGTTTTTGATGTGGATT |               |
| 16093_Probe_c             | AACCGCTATGTATcTCGTACATTACTG | 6FAM |
| 16093_Probe_t             | AACCGCTATGTATITCGTACATTACTG | HEX |
| 195_F                     | TGTTTTGATTCCCTGCCTCA |               |
| 195_R                     | GCTGTGCGACATTTCAATTGT |               |
| 195_Probe_c               | CGAACATACTTACTAAAGTGTGTAAATTT | 6FAM |
| 195_Probe_t               | CGAACATACCCTACTAAAGTGTGTAAATTT | HEX |
| 8705_F                    | CGACTAATCACCACACCAACA |               |
| 8705_R                    | TCCGAGGAGGTTAGTGTGTG |               |
| 8705_Probe_t              | ATAAACCAtACACCAACTAAAGGACGA | 6FAM |
| 8705_Probe_c              | ATAAACCACACACCAACCAACACAAAGGACGA | HEX |
| 15191_F                   | ACATCGGCATTATCCTCCTG |               |
| 15191_R                   | GTGTGAGGCTGGGACTCTT |               |
| 15191_Probe_t             | AGTAATTACAAACTTACTATCCGGCATC | 6FAM |
| 15191_Probe_c             | AGTAATTACAAACCCTACTATCCGGCATC | HEX |
| 7980_F                    | ACGATCCTCCTCTTACACATC |               |
| 7980_R                    | TTATACGAATGGGGGCTTCA |               |
| 7980_Probe_a              | AACCAGGGGaCCTGCGA | 6FAM |
| 7980_Probe_g              | AACCAGGCGgCCTGCGA | HEX |
| 15152_F                   | AAACCTGAAACATCGGCCATT |               |
| 15152_R                   | AATGTATGGGATGGCGGATA |               |
| 15152_Probe_a             | CTCCCGTGAAgGCAAATATC | 6FAM |
| 15152_Probe_g             | CTCCCGTGAgGCAAATATC | HEX |
Table S5. Comparison of heteroplasmy MAF estimated by ddPCR and by sequencing.

| Position | Sample | Major allele | Minor allele | ddPCR | Sequencing |
|----------|--------|--------------|--------------|-------|------------|
| 195      | A124d  | T            | C            | 0.007 | 0.000      |
| 195      | A148d  | T            | C            | 0.457 | 0.416      |
| 195      | A124b  | T            | C            | 0.011 | 0.006      |
| 195      | A124c  | T            | C            | 0.116 | 0.117      |
| 195      | R5b    | T            | C            | 0.027 | 0.002      |
| 195      | R5c    | T            | C            | 0.044 | 0.049      |
| 195      | R18b   | T            | C            | 0.169 | 0.157      |
| 195      | R18c   | T            | C            | 0.528 | 0.567      |
| 195      | A148b  | T            | C            | 0.165 | 0.207      |
| 195      | A148c  | T            | C            | 0.454 | 0.442      |
| 7980     | A178b  | A            | G            | 0.397 | 0.367      |
| 7980     | A178c  | A            | G            | 0.654 | 0.668      |
| 7980     | A178d  | A            | G            | 0.856 | 0.855      |
| 8705     | A143B  | T            | C            | 0.360 | 0.371      |
| 8705     | A143C  | T            | C            | 0.425 | 0.428      |
| 8705     | A156B  | T            | C            | 0.095 | 0.096      |
| 8705     | A156C  | T            | C            | 0.207 | 0.212      |
| 15152    | A170b  | A            | G            | 0.775 | 0.787      |
| 15152    | A170c  | A            | G            | 0.001 | 0.000      |
| 15191    | A20B   | T            | C            | 0.275 | 0.263      |
| 15191    | A20C   | T            | C            | 0.679 | 0.666      |
| 15191    | A28B   | T            | C            | 0.412 | 0.408      |
| 15191    | A28C   | T            | C            | 0.322 | 0.326      |
| 16093    | A23B   | C            | T            | 0.043 | 0.046      |
| 16093    | A23C   | C            | T            | 0.031 | 0.028      |
| 16093    | G55B   | C            | T            | 0.057 | 0.084      |
| 16093    | G55C   | C            | T            | 0.017 | 0.010      |
| 16093    | A110B  | C            | T            | 0.069 | 0.064      |
| 16093    | A110C  | C            | T            | 0.021 | 0.022      |
| 16093    | A152B  | C            | T            | 0.028 | 0.029      |
| 16093    | A152C  | C            | T            | 0.012 | 0.012      |
| 16093    | A169B  | C            | T            | 0.034 | 0.009      |
| 16093    | A169C  | C            | T            | 0.010 | 0.064      |
Table S6. Discrepant positions between mother-offspring pairs.

| Child ID | Position | Reference allele | Mother’s alleles | Major allele frequency | Offspring’s alleles | Major allele frequency | Gene Annotation | Functional effect |
|----------|----------|------------------|------------------|-----------------------|--------------------|-----------------------|-----------------|-------------------|
| A173c⁴  | 16234    | C                | C/T              | 0.629                 | T/C                | 0.675                 | CR              |                   |
| A173d⁴  | 16234    | C                | C/T              | 0.629                 | T/C                | 0.683                 | CR              |                   |
| L96c     | 1250     | C                | A/C              | 0.795                 | C/A                | 0.823                 | MT-RNR1         |                   |
| G52c     | 13824    | A                | G/A              | 0.994                 | G/A                | 0.554                 | SS(MT-ND5)      | CR                |
| A151c    | 9275     | A                | G/A              | 0.621                 | A/G                | 0.663                 | SS(MT-COX3)     |                   |
| G55c     | 10365    | G                | A/G              | 0.724                 | G/A                | 0.662                 | NS(MT-ND3)      | Medium            |
| G55c     | 16312    | A                | A/G              | 0.83                  | G/A                | 0.63                  | CR              |                   |
| A33c     | 8405     | A                | A/C              | 0.999                 | G/A                | 0.695                 | NS(MT-ATP8)     | not annotated    |
| R55c     | 8654     | T                | T/C              | 0.756                 | C/T                | 0.614                 | NS(MT-ATP6)     | Neutral           |
| A170c    | 8902     | G                | A/G              | 0.567                 | G/A                | 0.999                 | NS(MT-ATP6)     | Medium            |
| A170c    | 15152    | G                | G/A              | 0.787                 | A/G                | 0.9996                | NS(MT-CYTB)     | High              |
| A178c⁵   | 7980     | A                | A/G              | 0.633                 | G/A                | 0.668                 | NS(MT-COX2)     | Neutral           |
| A178d⁵   | 7980     | A                | A/G              | 0.633                 | G/A                | 0.855                 | NS(MT-COX2)     | Neutral           |
| R18c     | 195      | T                | T/C              | 0.843                 | C/T                | 0.567                 | CR              | associated with bipolar disorder |
| A157c    | 16292    | C                | T/C              | 0.832                 | C/T                | 0.646                 | CR              |                   |
| L87c     | 16311    | T                | C/T              | 0.56                  | T/C                | 0.696                 | CR              |                   |
| L106c    | 789      | T                | C/T              | 0.78                  | T/C                | 0.676                 | MT-RNR1         |                   |
| A20c     | 15191    | T                | T/C              | 0.737                 | C/T                | 0.666                 | SS(MT-CYTB)     |                   |

¹rCRS allele
²first allele is major allele, second is minor allele
³as predicted by Mutationassessor or from Mitomap (http://www.mitomap.org)
⁴A173c and A173d are MZ twins
⁵A178c and A178d are DZ twins
Table S7. Maximum-likelihood estimates (MLE) and Akaike Information Criteria (AIC) values for the four bottleneck models.

| Model          | MLE | Log-Likelihood at MLE | AIC   |
|----------------|-----|-----------------------|-------|
| Simple         | 8   | -560.96               | 1123.92 |
| Variable       | 9   | -558.58               | 1119.16 |
| Nucleoid       | 7   | -594.19               | 1190.38 |
| Variable Nucleoid | 9   | -582.50               | 1167.00 |
References

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290**: 457-465.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**: 147.

Cao L, Shitara H, Horii T, Nagao Y, Imai H, Abe K, Hara T, Hayashi J, Yonekawa H. 2007. The mitochondrial bottleneck occurs without reduction of mtDNA content in female mouse germ cells. *Nat Genet* **39**: 386-390.

Genome of the Netherlands Consortium. 2014. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* **46**: 818-825.

Jacobs HT, Lehtinen SK, Spelbrink JN. 2000. No sex please, we're mitochondria: a hypothesis on the somatic unit of inheritance of mammalian mtDNA. *Bioessays* **22**: 564-572.

Khrapko K. 2008. Two ways to make an mtDNA bottleneck. *Nat Genet* **40**: 134-135.

Lan Q, Lim U, Liu CS, Weinstein SJ, Chanock S, Bonner MR, Virtamo J, Albanes D, Rothman N. 2008. A prospective study of mitochondrial DNA copy number and risk of non-Hodgkin lymphoma. *Blood* **112**: 4247-4249.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Li M, Schroder R, Ni S, Madea B, Stoneking M. 2015. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proc Natl Acad Sci U S A* **112**: 2491-2496.

Li M, Schroeder R, Ko A, Stoneking M. 2012. Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic Acids Res* **40**: e137.

Li M, Stoneking M. 2012. A new approach for detecting low-level mutations in next-generation sequence data. *Genome Biol* **13**: R34.

Moran PAP. 1958. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society* **54**: 60-71.
Shoubridge EA. 2000. Mitochondrial DNA segregation in the developing embryo. *Hum Reprod* **15 Suppl 2**: 229-234.