Received: 31 December 2021 | Revised: 7 April 2022 | Accepted: 21 April 2022

DOI: 10.1002/humu.24390

RESEARCH ARTICLE

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

The transmission of human mitochondrial DNA in four-generation pedigrees

Qi Liu[1,2] | Muhammad Faaras Iqbal[3,4] | Tahir Yaqub[5] | Sehrish Firyal[4] | Yiqiang Zhao[1] | Mark Stoneking[6,7] | Mingkun Li[2,8,9]

[1]State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, China

[2]Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Center for Bioinformation, Beijing, China

[3]Department of Human Genetics and Molecular Biology, University of Health Sciences, Lahore, Pakistan

[4]Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore, Pakistan

[5]Institute of Microbiology, University of Veterinary and Animal Sciences, Lahore, Pakistan

[6]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[7]Laboratoire de Biométrie et Biologie Evolutive, CNRS, Université Lyon 1, Villeurbanne, France

[8]University of Chinese Academy of Sciences, Beijing, China

[9]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

Correspondence
Yiqiang Zhao, State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, China.
Email: yiqiangz@cau.edu.cn

Mark Stoneking, Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103, Leipzig, Germany.
Email: stonekg@eva.mpg.de

Mingkun Li, Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Center for Bioinformation, No. 1-104, Beichen West Rd, Chaoyang District, Beijing 100101, China.
Email: limk@big.ac.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 31871263; Max Planck Society

Abstract

Most of the pathogenic variants in mitochondrial DNA (mtDNA) exist in a heteroplasmic state (coexistence of mutant and wild-type mtDNA). Understanding how mtDNA is transmitted is crucial for predicting mitochondrial disease risk. Previous studies were based mainly on two-generation pedigree data, which are limited by the randomness in a single transmission. In this study, we analyzed the transmission of heteroplasmies in 16 four-generation families. First, we found that 57.8% of the variants in the great grandmother were transmitted to the fourth generation. The direction and magnitude of the frequency change during transmission appeared to be random. Moreover, no consistent correlation was identified between the frequency changes among the continuous transmissions, suggesting that most variants were functionally neutral or mildly deleterious and thus not subject to strong natural selection. Additionally, we found that the frequency of one nonsynonymous variant (m.15773G>A) showed a consistent increase in one family, suggesting that this variant may confer a fitness advantage to the mitochondrion/cell. We also estimated the effective bottleneck size during transmission to be 21–71. In summary, our study demonstrates the advantages of

Qi Liu and Muhammad Faaras Iqbal contributed equally to this study.

Yiqiang Zhao, Mark Stoneking, and Mingkun Li shared senior authorship.

multigeneration data for studying the transmission of mtDNA for shedding new light on the dynamics of the mutation frequency in successive generations.

KEYWORDS
heteroplasmy, inheritance, mtDNA, multigeneration pedigrees, transmission

1 | INTRODUCTION

As a vital organelle in eukaryotic cells, mitochondria supply the energy for cellular processes. The mitochondrion has its own genetic material, mitochondrial DNA (mtDNA), which is a 16,569 bp double-stranded circular molecule and encodes 13 protein genes, 22 transfer RNAs, and 2 ribosomal RNAs. The coexistence of wild-type mtDNA and mutant mtDNA in mitochondria or cells is known as mitochondrial heteroplasmy. Many pathogenic variants are in a heteroplasmic state and could compromise the function of the mitochondria (Stewart & Chinnery, 2015, 2021). The degree of heteroplasmy is related to the severity and onset of the disease; a threshold of 60%–95% is proposed for the exhibition of a clinical phenotype (Farrar et al., 2013; Stewart & Chinnery, 2015; Tatuch et al., 1992; Wallace et al., 1999).

The level of heteroplasmy dynamically changes throughout life, varies among different tissues, and is correlated with aging, the production of reactive oxygen species, and germline selection (Hahn & Zuryn, 2019; M. Li et al., 2015; Sondheimer et al., 2011; Wei et al., 2019). Moreover, the level of heteroplasmy could be significantly altered during transmission due to the so-called "bottleneck effect": only a subset of mtDNAs in the oocyte of the mother (MO) are transmitted to the offspring (OF) (Zhang et al., 2018). Understanding the inheritance of mtDNA is crucial for predicting the probability of a pathogenic variant being passed to the next generation and causing disease. Two important relevant scientific questions are how severe the bottleneck (Nb) is and whether the transmission process is neutral or under selection.

MtDNA Nb size can be estimated using a maximum likelihood algorithm to search for parameters that best fit the observed allele frequency change of shared variants in MO–oOF pairs. Based on the analysis of 39 MO–OF pairs, Rebolledo-Jaramillo et al. reported the Nb to be 26–35 (Rebolledo-Jaramillo et al., 2014). However, the average Nb was estimated to be nine using a much larger trio size of 228 in our previous study (M. Li et al., 2016). Zaidi et al. built a more sophisticated ontogenetic phylogenetic likelihood model that took genetic drift and variable mutation rates into consideration (Wilton et al., 2018; Zaidi et al., 2019), and obtained a similar Nb of 10–13.

Meanwhile, whether selection acts on mtDNA variants during transmission is still unclear (Burr et al., 2018). Rebolledo-Jaramillo et al. proposed purifying selection against transmission of nonsynonymous (NS) variants (Rebolledo-Jaramillo et al., 2014), whereas selection against novel/unknown variants in rRNA genes was found in other studies (M. Li et al., 2016; Wei et al., 2019). In contrast, Wilson et al. found no evidence of selection on five common heteroplasmic variants by investigating 577 MO–OF pairs (Wilson et al., 2016). These different findings may indicate a limited scale and intensity of selection during transmission; thus, a more sensitive method is needed.

Most previous studies of heteroplasmy transmission were conducted in two-generation pedigrees. These studies on single transmission events were unable to infer selection on specific variants unless the variant was shared by a large number of samples (Wilson et al., 2016), and could be heavily influenced by the random drift due to the severe Nb effect. In contrast, larger pedigrees, which include successive transmissions, can provide additional power to investigate selection and determine if specific variants are under continuous selection. Unfortunately, very few studies have been conducted on large pedigrees due to the difficulties in sampling. Zaidi et al. analyzed the inheritance of mtDNA variants in 7 three-generation and 2 four-generation pedigrees. However, the limited sample size and number of heteroplasmies found (12) prevented them from properly investigating selection (Zaidi et al., 2019). There are also two studies that compared single cell mtDNA in large pedigrees and before and after transplantation, but these studies only examined the control region, which may be subject to less stringent selection pressure (Yao, Childs, et al., 2007; Yao et al., 2013, 2015). In the present study, we explored the pattern of mtDNA transmission using 16 healthy four-generation pedigrees, including great grandmother (GGM), grandmother (GM), MO, and OF. We found that most heteroplasmies detected in the GGM were transmitted to the following generations, and selection was inferred for specific variants, although no broad selection signal was observed.

2 | MATERIALS AND METHODS

2.1 | Sample collection and library preparation

Forty-nine four-generation families were visited in different districts of Punjab Province of Pakistan. After taking verbal consent from participants of the study or legal guardians, blood (BL) samples or buccal swabs (BS) were collected from participants. BL samples were collected primarily from adults, while the BS were primarily collected from children. We took the specimen from only one person per generation. No clinical or phenotypic data was collected.

DNA was extracted from samples using the QIAamp DNA mini kit (Qiagen). Extracted DNA was diluted to a concentration of 10 ng/μl and subjected to DNA fragmentation using a Diagenode Bioruptor (Diagenode). The mtDNA was enriched by hybridization with polymerase chain reaction-generated mtDNA baits, and the sequencing libraries were prepared following a multiplex sequencing protocol (Maricic et al., 2010). Dual indices were used. The libraries were then

sequenced on an Illumina HiSeq. 2000 (Illumina) in paired-end mode with a read length of 125 bp. Finally, 63.2G data were obtained.

2.2 | Quality control

The library construction of seven samples failed, resulting in 189 samples being included in the subsequent analysis. Most samples from the OF were BS (35 out of 46), and samples from other family members were mostly from BL (137 out of 143).

MitoMutCall was used to assemble the consensus sequence (allele frequency ≥0.7, read count supporting the allele on each strand ≥3, sequencing depth ≥5) (Liu et al., 2019). To screen for possible mislabeling or incorrect kinship, we compared the consensus sequences among family members; a problem was suspected if there were more than four discrepant nucleotides between one sample and the remaining family members, which identified 15 problematic samples from 14 families. Strikingly, the minimum number of discrepancies was 13 in these samples, indicating that they were undoubtedly not from the same maternal lineage as other family members. Moreover, the discrepant sequences did not match any other families, suggesting possible adoption; thus, these samples were excluded from the subsequent analysis.

2.3 | Detection of heteroplasmies

DREEP was used to call heteroplasmies (minor allele frequency (MAF) ≥ 0.05, MAF on each strand ≥0.02, read count supporting either allele on each strand ≥3, sequencing depth ≥200, DREEP score ≥10 using a Poisson or the empirical distribution) (M. Li & Stoneking, 2012). No indels were allowed in the 3 bp of the flanking sequence in each direction. We also excluded the low-complexity and ambiguity regions (302–316, 513–526, 566–573, 3106–3107, and 16,181–16,194). At positions with at least one family member with heteroplasmy satisfying the above criteria, an inherited heteroplasmy was called if both alleles were observed with a frequency higher than 2% in the individual's MO.

In particular, the nuclear mitochondrial DNA (NUMTs) may be enriched during the process of hybridization capture, which could lead to false-positive heteroplasmies (Husami et al., 2020; Yao et al., 2008). In our previous study (M. Li et al., 2012), we have demonstrated that the mapping strategy and criteria for calling heteroplasmy employed in this study (i.e., reads were simultaneously mapped to the nuclear DNA and mtDNA, with a minimum frequency higher than 2%) could eliminate all NUMTs-induced false-positive heteroplasmies.

2.4 | Identification of potential contamination

Two scenarios of contamination were considered in this study (M. Li et al., 2015). The first scenario was contamination from other experimental samples. Contamination was suspected if more than five heteroplasmies could be explained by contamination from another sample in the same sequencing library, and more than 80% of the nucleotides from the potential contamination donor at the donor–recipient discrepant positions were detectable in the sample (MAF ≥ 1%, MAF on each strand ≥1%, read count on each strand ≥3). The second scenario was contamination from samples not included in the sequencing library. Contamination was suspected if more than five heteroplasmies could be explained by contamination from a specific haplogroup, and more than 80% of the variants defining this haplogroup were detectable in the sample. According to these criteria, 32 samples were suspected to be contaminated (9 and 23 samples, respectively, for the two scenarios). An additional seven samples were suspected to be problematic, as the distribution of MAF at all heteroplasmic positions was significantly different from that in other family members (adjusted $p < 0.05$).

Overall, 135 out of 189 samples passed all of the above filters, resulting in data from 16 four-generation families, 16 three-generation families, 9 two-generation families, and 5 one-generation families.

2.5 | Annotation of the heteroplasmies

Variants were annotated by ANNOVAR (K. Wang et al., 2010). The mitochondrial haplogroup was determined by HaploGrep2 (Weissensteiner et al., 2016) and Phylotree (van Oven & Kayser, 2009). Pathogenicity was obtained via MITOMAP (Lott et al., 2013) and MitImpact (Castellana et al., 2015) using the default parameters.

2.6 | Estimation of the Nb size

The maximum-likelihood estimation (MLE) method described in our previous study was used to estimate the Nb size (M. Li et al., 2016). Briefly, we constructed a constant-size model for the transmission of the heteroplasmies which considered the influence of sequencing error and the stochastic effect. We simulated different Nb sizes to find the parameters that best fit the observed frequency change of the heteroplasmies during the transmission from MO to OF. The detailed formulas and procedures are described in the supplementary file of the original article (M. Li et al., 2016). Because it is unclear whether the heteroplasmies identified in the same individual were on the same molecule or different molecules, the Nb size was estimated for each of the two scenarios (one heteroplasmy was randomly selected from each MO–OF pair in the former scenario).

3 | RESULTS

3.1 | Summary of the sequencing data

BL and buccal swab samples from 49 four-generation families were subjected to mtDNA enrichment and next-generation sequencing (see details Section 2). Seven samples failed in library construction. For the remaining samples, the average sequencing depth was 2424 (interquartile

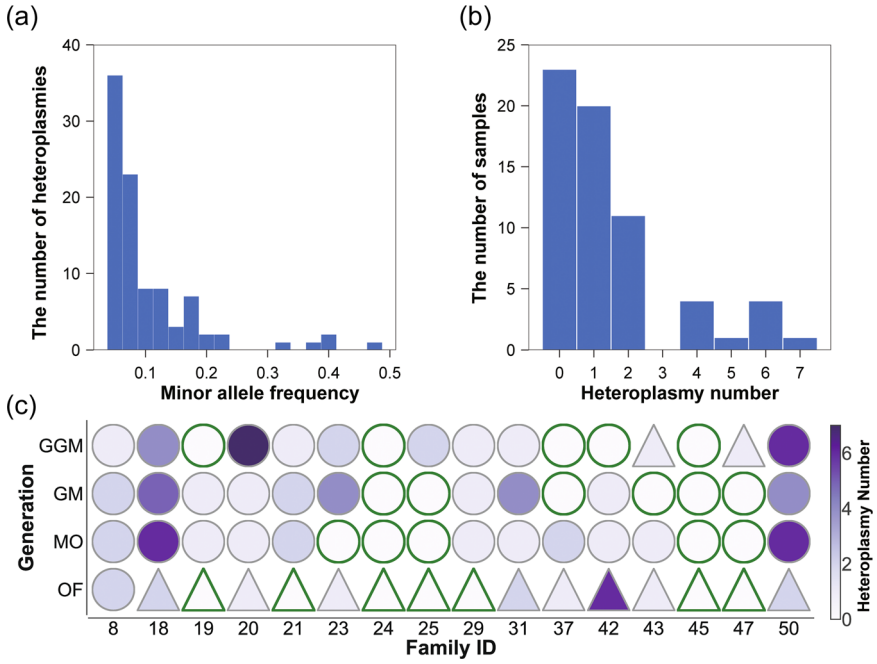

FIGURE 1 Heteroplasmies in the 16 four-generation pedigrees. (a) The distribution of minor allele frequency of all heteroplasmies. (b) The distribution of heteroplasmy number per sample. (c) The number of heteroplasmies for each sample. Circles and triangles denote the BL and BS tissues, respectively. The density of the color indicates the number of heteroplasmies. The green border indicates that no heteroplasmy was identified in the sample. BL, blood; BS, buccal swabs; GGM, great grandmother; GM, grandmother; MO, mother; OF, offspring.

range: 1332–3186). Fifty-four samples were suspected to be mislabeled or contaminated through quality analysis (see details Section 2). Finally, 135 samples from 46 families were used in the following analysis, including 16 four-generation families, 16 three-generation families, nine two-generation families, and five one-generation families.

3.2 | Summary of the heteroplasmies

In total, we detected 194 heteroplasmies with MAF at least 5% in 83 individuals, which were located at 118 positions. Heteroplasmies were overrepresented in the D-loop region, followed by the *CYTB* and *ND2* genes (Supporting Information: Figure S1). To further investigate the transmission pattern of mtDNA heteroplasmies, we focused on the 16 four-generation families. For these, 94 heteroplasmies were identified at 60 positions in 14 families, and 64.9% of the heteroplasmies had an MAF less than 10% (Figure 1a). The number of heteroplasmies per sample ranged from 0 to 7, which followed a Poisson distribution ($p = 0.26$, χ^2 test, Figure 1b,c and Supporting Information: Figure S2).

3.3 | Transmission of the heteroplasmies in the four-generation pedigrees

A heteroplasmy was classified as inherited if the same alleles were observed in an individual and her/his MO with MAF at least 2%; otherwise, a de novo variant was called. Among the 45 heteroplasmies observed in the first generation (GGM), 33 (73.3%) were transmitted to the second generation (GM), 29 (64.4%) were transmitted to the third generation (MO), and 26 (57.8%) were transmitted to the fourth generation (OF) (Table 1). The number of de novo variants and disappearing variants were comparable (25 vs. 31), showing a dynamic balance of the variant number during transmission. We noted that more inherited heteroplasmies would be called if looser criteria were applied. For example, if an MAF of 0.5% in the MO was required to define an inheritance event, the heritable proportion of variants passed to different generations in the GGM increased to 83.6%, 78.2%, and 72.7%, respectively. However, inherited heteroplasmies might be overestimated with this low frequency threshold due to the occurrence of independent somatic variants at the same position in different generations. Thus, an MAF threshold of 2% was used in the following analyses. Moreover, positions with a high substitution rate (defined as more than 10 independent occurrences in the mtDNA phylogenetic tree) were also excluded.

The MAF of the heteroplasmies was significantly correlated between successive generations ($r = 0.87$, $p = 5.43e{-}15$ for GM–MO;

TABLE 1 The inheritance of the heteroplasmic variants in four-generation families.

Generation	All	De novo	Inherited	GMM[a]	GM[b]	MO[c]	OF[d]
GGM	45	-	-	12	4	3	26
GM	43	10	33		12	5	26
MO	34	3	31			7	27
OF	39	12	27	-	-	-	-

Abbreviations: GGM, great grandmother; GM, grandmother; MO, mother; OF, offspring.

[a]GMM: the number of variants that can only be observed in GGM.

[b]GM: the number of variants that were transmitted to GM, but not MO, or OF.

[c]MO: the number of variants that were transmitted to GM and MO, but not OF.

[d]OF: the number of variants that were transmitted to GM, MO, and OF.

$r = 0.89$, $p = 3.12e{-}14$ for MO–OF), except for GGM–GM ($r = 0.18$, $p = 0.19$) (Supporting Information: Figure S3). The lack of correlation between GGM and GM might be caused by the fact that GGMs are older; hence, there was more time for heteroplasmy frequencies to change since transmission, which has been observed in previous studies (Arbeithuber et al., 2020; Burgstaller et al., 2018; Zaidi et al., 2019). Consistent with this explanation, we found that the variance of the MAF change in GGM–GM was greater than that in GM–MO and MO–OF ($p = 0.005$, and $p = 0.001$, F test).

The MAF changes of heteroplasmies during transmission seemed to be random and followed a normal distribution with a mean of zero ($p < 0.001$, Shapiro–Wilk normality test, Figure 2a,b). Meanwhile, the direction and magnitude of the frequency change during the three generations of transmission were not consistently correlated (Figure 2b,c), suggesting that the frequency change was driven by random drift rather than persistent selection in one direction. Thus, we speculated that most variants were functionally neutral or at most mildly deleterious. In addition, we observed that the variance of frequency change of the variant in one generation (C1), two generations (C2), and three generations (C3) were significantly increased with the number of transmissions (0.0055, 0.0098, and 0.0132, respectively, for C1, C2, and C3; C1 vs. C2, $p = 2.93e{-}4$, C1 vs. C3, $p = 6.68e{-}6$, C2 vs. C3, $p = 0.087$, F test) (Figure 2a), indicating that the frequency of variants could change significantly even without the involvement of selection.

Natural selection might nonetheless be involved in the transmission of some heteroplasmies. We found that NS variants showed a higher chance of being lost during transmission than other variants, despite having similar MAFs ($p = 0.0198$, Fisher's exact test). In

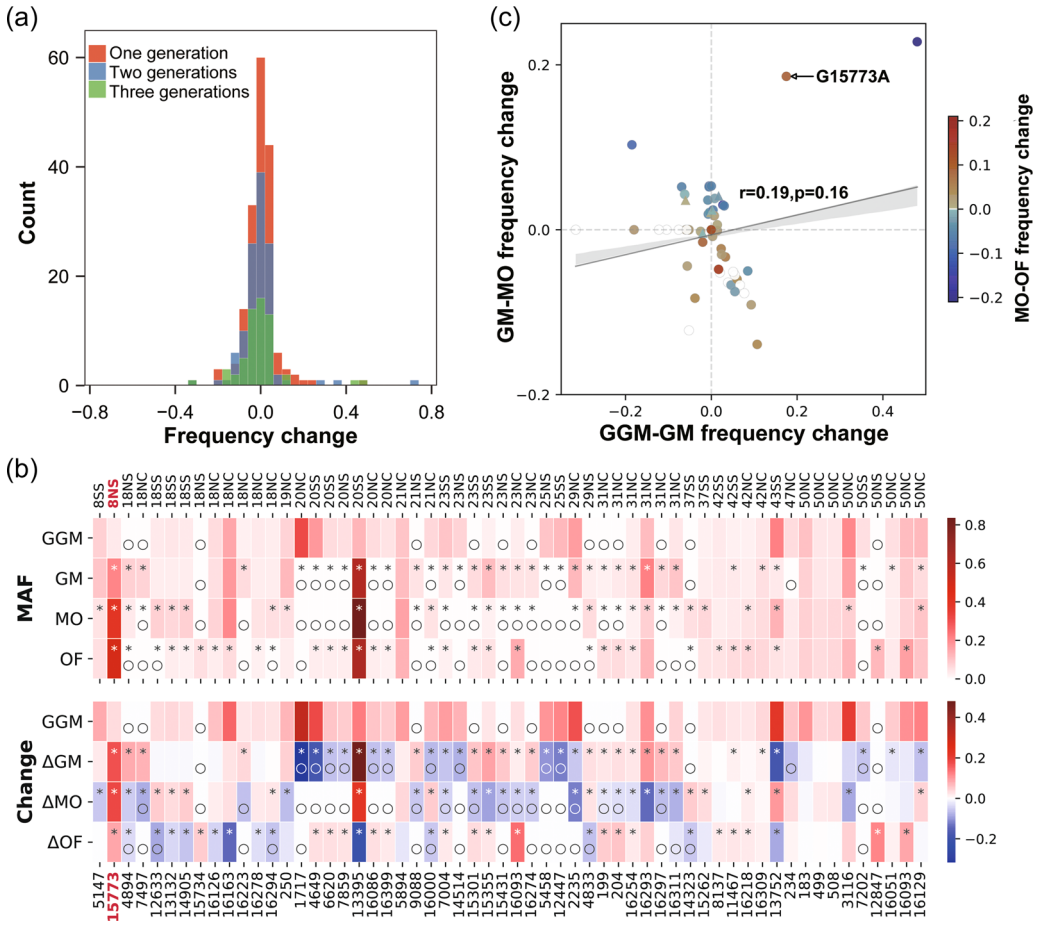

FIGURE 2 Heteroplasmy minor allele frequency (MAF) change during transmission. (a) The distributions of MAF changes in one generation, two generations, and three generations. (b) MAF changes during transmission. The top panel shows the MAF for each sample. The bottom panel shows the MAF change during transmission. Circles indicate that the heteroplasmy disappeared in the sample, while an asterisk indicates significant changes in MAF between generations. The labels at the top denote the family ID and mutation type (NC, nocoding; NS, nonsynonymous; SS, synonymous). The labels at the bottom denote the nucleotide position, with the position showing a continuous MAF change in the same direction indicated in red. (c) The correlation of MAF change between different transmissions. There was no significant correlation in the frequency change between GGM–GM and GM–MO, whereas a negative correlation was observed between GM–MO and MO–OF ($r = -0.44$, $p = 4.92e{-}4$). The only variant showing continuous MAF change in the same direction is indicated by an arrow. Triangles indicate the five variants whose MAF did not significantly change across all three transmissions. GGM, great grandmother; GM, grandmother; MO, mother; OF, offspring.

addition, we found one heteroplasmy whose frequency showed a consistent increase across generations (Figure 2b). Specifically, the frequency of the mutant A allele at position 15,773 in family 8 increased from 3.6% in GGM to 21.1%, 39.7%, and 48.5% in GM, MO, and OF, respectively. This heteroplasmy was located in the *MT-CYB* gene, and the variant was reported to be possibly associated with LHON (Lott et al., 2013).

Two additional pathogenic heteroplasmies reported in MITOMAP (Lott et al., 2013) were observed in the data set. Variant m.7497G>A was reported to be associated with exercise intolerance, muscle pain, and lactic acidaemia (Grafakou et al., 2003). This variant was observed in the GM in family 18 with a frequency of 7.7% and disappeared in MO. The NS variant m.7859G>A in the *MT-CO2* gene, associated with progressive encephalomyopathy (Uusimaa et al., 2004), was observed in all members of family 20 with similar frequencies (GGM 94.6%, GM 99.5%, MO 99.4%, OF 97.1%), suggesting no significant selection acting on this variant.

3.4 | Estimation of the Nb size during transmission

A MLE method was applied to calculate the Nb size during transmission (M. Li et al., 2016). In the first model, we assumed that heteroplasmies in the same individual were randomly distributed on different molecules, so heteroplasmies were transmitted independently. Thirty-six MO–OF pairs involving 91 heteroplasmies with MAF greater than 2% were used to calculate the Nb, which resulted in an Nb of 48 (Figure 3a). Nbs estimated from different transmissions were 38, 53, and 71 for GGM–GM, GM–MO, and MO–OF, respectively (Supporting Information: Figure S4). Correspondingly, we found that the three greatest frequency changes were observed in GGM–GM transmissions, even though the overall distribution was not significantly different between different transmissions ($p = 0.24$, Kruskal–Wallis test).

In the second model, we assumed that variants observed within an individual might be located on the same molecule, and thus, they were always transmitted together and should be regarded as a single transmission event. This hypothesis was supported by the observation that the frequency changes of two transmitted heteroplasmies in the same individual were more similar than those of two transmitted heteroplasmies in different individuals ($p = 1.93e{-}13$, Wilcoxon rank-sum test, Supporting Information: Figure S5). Then, one heteroplasmy was randomly selected from each MO–OF pair to calculate the Nb, and the process was repeated 1000 times. The optimum Nb estimated from all MO–OF pairs was 30 (Figure 3b), while the Nbs estimated from the transmission of GGM–GM, GM–MO, and MO–OF were 21, 36, and 56, respectively (Supporting Information: Figure S6).

4 | DISCUSSION

mtDNA possesses a higher mutation rate than nuclear DNA(Brown et al., 1979). Whether these variants affect the function of the mitochondrion and how deleterious variants are removed from the mtDNA genome are important fundamental scientific questions. Comparative analysis of mtDNA in different species and geographic locations has revealed that some variants are under either positive or negative selection at the population level (Cavadas et al., 2015; Pereira et al., 2011; Ruiz-Pesini et al., 2004; Soares et al., 2013), suggesting that they are functionally important. Recently, selection at the individual level has been investigated via the analysis of the frequency change of heteroplasmic variants during transmission (Giuliani et al., 2014; M. Li et al., 2016; Rebolledo-Jaramillo et al., 2014; Wei et al., 2019; Zaidi et al., 2019). A great challenge in these studies is that the MAF change during transmission is affected not only by selection but also by the severe Nb that occurs during oogenesis, which results in random frequency changes during transmission (Floros et al., 2018; Zaidi et al., 2019). Thus, instead of looking for variants/sites/genes under selection, previous studies mostly focused on selection at a genome-wide level, which may not be very helpful for downstream functional analysis.

In this study, we collected a large sample of four-generation pedigrees to investigate the inheritance of mtDNA heteroplasmies. This multigeneration data enabled us to investigate the selection on

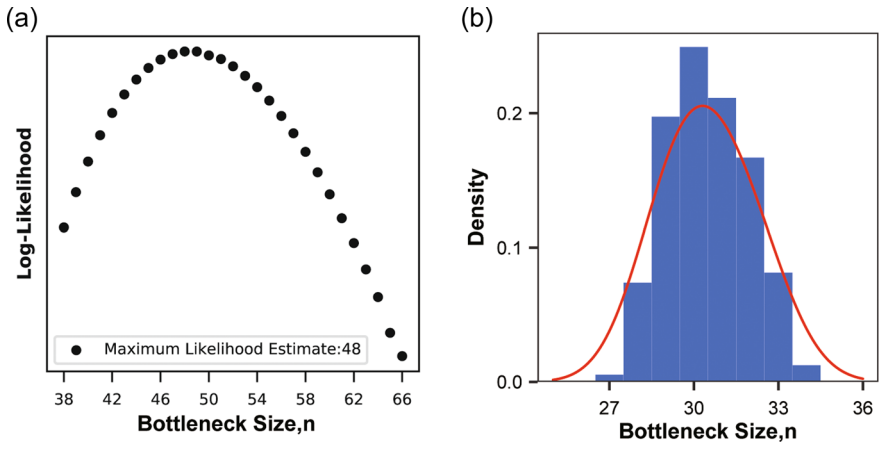

FIGURE 3 The estimated mitochondrial DNA bottleneck (Nb) size. (a) Likelihood curve of the Nb size when using all heteroplasmies with a frequency greater than 2%. (b) Distribution of the optimum Nb sizes when using only one heteroplasmy from each sample. A heteroplasmy was randomly selected when there was more than one heteroplasmy in the sample, and the process was repeated 1000 times. The red line represents the best fitting curve.

specific variants. Among 45 heteroplasmic variants identified in the GGM, we found only one variant that showed a persistent increase in frequency during three transmissions (with MAFs of 3.6%, 21.1%, 39.7%, and 48.5%). The increase may be due to the possibility that this NS variant in the *MT-CYB* gene increased the fitness of the mitochondrion/cell over the wild type, or it could just be a chance event, further experimental investigation is needed. The most significant frequency change was observed at position 13,395 (a synonymous variant in the *MT-ND5* gene) in family 20, whose MAF increased from 12.9% to 60.9% after the first transmission from GGM to GM and further increased to 83.7%, which is high enough to have an impact on mitochondrial function. Interestingly, the frequency decreased to 62.9% after the third transmission, consistent with the speculation that the function of the mutant mitochondria might be impaired and thus selected against when it is at high frequency. The second greatest frequency change was observed at position 1717 in the *MT-RNR2* gene in the same family, whose MAF was reduced from 31.6% to 0% after the first transmission and was not observed in the subsequent generations. Although selection on this position is possible, it is difficult to distinguish selection from genetic drift in this situation. Of note, the MAF of another two heteroplasmies (m.2235C>T and m.3116C>T) in the same gene also significantly decreased during transmission (from 17.4% to 0 and from 20.4% to 12.8%), suggesting that this region may be under purifying selection, consistent with the findings in previous two-generation studies (M. Li et al., 2016; Wei et al., 2019). Additionally, we found that NS variants were more prone to be lost during transmission, suggesting that they may be subject to purifying selection during transmission, which has also been found in other studies using pedigree data (Floros et al., 2018; Rebolledo-Jaramillo et al., 2014; Zaidi et al., 2019). Meanwhile, there were also studies on population data that identified the signal of purifying selection on heteroplasmic mutations, which may be associated with the pathogenic potential of these mutations (H. Li et al., 2017; Y. Wang et al., 2021; Ye et al., 2014). Of note, the selection on mutations might be context-dependent (i.e., haplogroup and background mutations), which was not investigated due to the lack of mutations that were concurrently observed in different pedigrees (Yao, Ellison, et al., 2007; Yao et al., 2013).

The estimated Nb size in this study ranged between 21 and 71. This result is close to that reported by Rebolledo-Jaramillo et al., 2014 but larger than that estimated by Zaidi et al. and our previous study (M. Li et al., 2016; Zaidi et al., 2019). Since the estimation is based on the distribution of the frequency change in two successive generations, the discrepancy may stem from the differences in the sample size, tissues, age, sex, haplogroup, or health status of the participants (Arbeithuber et al., 2020; Johnston et al., 2015; Nandakumar et al., 2021; Zaidi et al., 2019), which can potentially lead to an altered MAF. Moreover, the Nb size estimated from the younger generation data tended to be larger than that estimated using the older generation data. A possible explanation is that heteroplasmy frequencies may change during aging (Arbeithuber et al., 2020; Johnston & Jones, 2016). A significant correlation was observed between age and the frequency of mtDNA variants (Elson et al., 2001; Rajasimha et al., 2008), as expected since mutation frequencies do change over time. This can result in a frequency change distribution with a larger variance, which could lead to an underestimation of the Nb size.

Our study has several limitations. First, two types of specimens were used, the specimens taken from the OF (most of whom were children) were mostly BS, whereas the specimens from other generations were BL. As the level of heteroplasmy varies between tissues (M. Li et al., 2015), the comparison results between two tissues should be interpreted with caution. Nevertheless, we noted that the two tissues were similar in terms of the number and frequency of the heteroplasmies (Supporting Information: Figure S7), and recent studies found a high correlation between the buccal swab and BL in heteroplasmy level ($r = 0.97$) (Zaidi et al., 2019), especially in younger individuals ($r^2 = 0.92$) (Rebolledo-Jaramillo et al., 2014). Therefore, the difference in sample types should have a limited impact on the investigation of mtDNA transmission in this study. Second, the specimens investigated in our study, as well as previous studies on the inheritance of mtDNA, are tissues that have undergone a long period of time after the transmission, which may give rise to new somatic mutations and alter the frequency of the mutation. We call heteroplasmies using a relatively higher frequency threshold, so that more than 90% of the mutations had a MAF greater than 10% in at least one family member, a frequency that somatic mutation rarely reached (M. Li et al., 2015). Thus, our results on the mtDNA transmission should not be heavily influenced by somatic mutations.

In summary, we investigated the dynamics of the frequency changes in mtDNA heteroplasmies over three successive generations in humans. Although the frequency change of variants seems to be random, the frequency of a few variants showed consistent changes over the course of multiple generations, which may result in significant changes in mitochondrial function. Selection is suspected to be involved in the transmission of some variants, in particular selection against NS variants, and discovering the underlying mechanism and to what extent it could help to avoid the accumulation of harmful variants merit further investigation.

ACKNOWLEDGMENTS

We thank the high-performance computing platform of the State Key Laboratory of Agrobiotechnology for their support. We also thank Roland Schröder from the Max Planck Institute for Evolutionary Anthropology for helping generate the sequences. This study was supported by the National Natural Science Foundation of China (31871263) and the Max Planck Society. Open Access funding enabled and organized by Projekt DEAL.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The raw sequencing data reported in this article have been deposited in the Genome Warehouse in the National Genomics Data Center

(under accession number HRA001581). The variant data have been sumitted to LOVD database (https://databases.lovd.nl/shared/references/DOI:10.1002/humu.24390).

ORCID

Mingkun Li http://orcid.org/0000-0003-1041-1172

REFERENCES

Arbeithuber, B., Hester, J., Cremona, M. A., Stoler, N., Zaidi, A., Higgins, B., Anthony, K., Chiaromonte, F., Diaz, F. J., & Makova, K. D. (2020). Age-related accumulation of de novo mitochondrial mutations in mammalian oocytes and somatic tissues. *PLoS Biology*, *18*(7), e3000745. https://doi.org/10.1371/journal.pbio.3000745

Brown, W. M., George, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, *76*(4):1967. https://doi.org/10.1073/pnas.76.4.1967

Burgstaller, J. P., Kolbe, T., Havlicek, V., Hembach, S., Poulton, J., Piálek, J., Steinborn, R., Rülicke, T., Brem, G., Jones, N. S., & Johnston, I. G. (2018). Large-scale genetic analysis reveals mammalian mtDNA heteroplasmy dynamics and variance increase through lifetimes and generations. *Nature Communications*, *9*(1), 2488. https://doi.org/10.1038/s41467-018-04797-2

Burr, S. P., Pezet, M., & Chinnery, P. F. (2018). Mitochondrial DNA heteroplasmy and purifying selection in the mammalian female germ line. *Development, Growth & Differentiation*, *60*(1), 21–32. https://doi.org/10.1111/dgd.12420

Castellana, S., Rónai, J., & Mazza, T. (2015). MitImpact: An exhaustive collection of pre-computed pathogenicity predictions of human mitochondrial non-synonymous variants. *Human Mutation*, *36*(2), E2413–E2422. https://doi.org/10.1002/humu.22720

Cavadas, B., Soares, P., Camacho, R., Brandão, A., Costa, M. D., Fernandes, V., Pereira, J. B., Rito, T., Samuels, D. C., & Pereira, L. (2015). Fine time scaling of purifying selection on human nonsynonymous mtDNA mutations based on the worldwide population tree and mother–child pairs. *Human Mutation*, *36*(11), 1100–1111. https://doi.org/10.1002/humu.22849

Elson, J. L., Samuels, D. C., Turnbull, D. M., & Chinnery, P. F. (2001). Random intracellular drift explains the clonal expansion of mitochondrial DNA mutations with age. *American Journal of Human Genetics*, *68*(3), 802–806. https://doi.org/10.1086/318801

Farrar, G. J., Chadderton, N., Kenna, P. F., & Millington-Ward, S. (2013). Mitochondrial disorders: Aetiologies, models systems, and candidate therapies. *Trends in Genetics*, *29*(8), 488–497. https://doi.org/10.1016/j.tig.2013.05.005

Floros, V. I., Pyle, A., Dietmann, S., Wei, W., Tang, W. C. W., Irie, N., Payne, B., Capalbo, A., Noli, L., Coxhead, J., Hudson, G., Crosier, M., Strahl, H., Khalaf, Y., Saitou, M., Ilic, D., Surani, M. A., & Chinnery, P. F. (2018). Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos. *Nature Cell Biology*, *20*(2), 144–151. https://doi.org/10.1038/s41556-017-0017-8

Giuliani, C., Barbieri, C., Li, M., Bucci, L., Monti, D., Passarino, G., Luiselli, D., Franceschi, C., Stoneking, M., & Garagnani, P. (2014). Transmission from centenarians to their offspring of mtDNA heteroplasmy revealed by ultra-deep sequencing. *Aging*, *6*(6), 454–467. https://doi.org/10.18632/aging.100661

Grafakou, O., Hol, F. A., Otfried Schwab, K., Siers, M. H., Ter Laak, H., Trijbels, F., Ensenauer, R., Boelen, C., & Smeitink, J. (2003). Exercise intolerance, muscle pain and lactic acidaemia associated with a 7497G>A mutation in the tRNASer(UCN) gene. *Journal of Inherited Metabolic Disease*, *26*(6), 593–600. https://doi.org/10.1023/A:1025960300710

Hahn, A., & Zuryn, S. (2019). Mitochondrial genome (mtDNA) mutations that generate reactive oxygen species. *Antioxidants*, *8*(9), 392. https://www.mdpi.com/2076-3921/8/9/392

Husami, A., Slone, J., Brown, J., Bromwell, M., Valencia, C. A., & Huang, T. (2020). Clinical utility of whole genome sequencing for the detection of mitochondrial genome mutations. *Journal of Genetics and Genomics*, *47*(3), 167–169. https://doi.org/10.1016/j.jgg.2020.03.001

Johnston, I. G., Burgstaller, J. P., Havlicek, V., Kolbe, T., Rülicke, T., Brem, G., Poulton, J., & Jones, N. S. (2015). Stochastic modelling, Bayesian inference, and new in vivo measurements elucidate the debated mtDNA bottleneck mechanism. *eLife*, *4*, e07464. https://doi.org/10.7554/eLife.07464

Johnston, I. G., & Jones, N. S. (2016). Evolution of cell-to-cell variability in stochastic, controlled, heteroplasmic mtDNA populations. *The American Journal of Human Genetics*, *99*(5), 1150–1162. https://doi.org/10.1016/j.ajhg.2016.09.016

Li, H., Bi, R., Fan, Y., Wu, Y., Tang, Y., Li, Z., He, Y., Zhou, J., Tang, J., Chen, X., & Yao, Y. -G. (2017). mtDNA heteroplasmy in monozygotic twins discordant for schizophrenia. *Molecular Neurobiology*, *54*(6), 4343–4352. https://doi.org/10.1007/s12035-016-9996-x

Li, M., Rothwell, R., Vermaat, M., Wachsmuth, M., Schröder, R., Laros, J. F. J., van Oven, M., de Bakker, P. I. W., Bovenberg, J. A., van Duijn, C. M., van Ommen, G.-J. B., Slagboom, P. E., Swertz, M. A., Wijmenga, C., Genome of Netherlands, C., Kayser, M., Boomsma, D. I., Zöllner, S., de Knijff, P., & Stoneking, M. (2016). Transmission of human mtDNA heteroplasmy in the genome of the Netherlands families: Support for a variable-size bottleneck. *Genome Research*, *26*(4), 417–426. https://doi.org/10.1101/gr.203216.115

Li, M., Schröder, R., Ni, S., Madea, B., & Stoneking, M. (2015). Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proceedings of the National Academy of Sciences*, *112*(8), 2491–2496. https://doi.org/10.1073/pnas.1419651112

Li, M., Schroeder, R., Ko, A., & Stoneking, M. (2012). Fidelity of capture-enrichment for mtDNA genome sequencing: Influence of NUMTs. *Nucleic Acids Research*, *40*(18), e137. https://doi.org/10.1093/nar/gks499

Li, M., & Stoneking, M. (2012). A new approach for detecting low-level mutations in next-generation sequence data. *Genome Biology*, *13*(5), R34. https://doi.org/10.1186/gb-2012-13-5-r34

Liu, Q., Lin, D., Li, M., Gu, Z., & Zhao, Y. (2019). Evidence of neutral evolution of mitochondrial DNA in human hepatocellular carcinoma. *Genome Biology and Evolution*, *11*(10), 2909–2916. https://doi.org/10.1093/gbe/evz214

Lott, M. T., Leipzig, J. N., Derbeneva, O., Xie, H. M., Chalkia, D., Sarmady, M., Procaccio, V., & Wallace, D. C. (2013). mtDNA variation and analysis using mitomap and mitomaster. *Current Protocols in Bioinformatics*, *44*(1), 1.23.21–21.23.26. https://doi.org/10.1002/0471250953.bi0123s44

Maricic, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*, *5*(11), e14004. https://doi.org/10.1371/journal.pone.0014004

Nandakumar, P., Tian, C., O'Connell, J., Hinds, D., Paterson, A. D., & Sondheimer, N. (2021). Nuclear genome-wide associations with mitochondrial heteroplasmy. *Science Advances*, *7*(12), eabe7520. https://doi.org/10.1126/sciadv.abe7520

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, *30*(2), E386–E394. https://doi.org/10.1002/humu.20921

Pereira, L., Soares, P., Radivojac, P., Li, B., & SamuelsDavid C. (2011). Comparing phylogeny and the predicted pathogenicity of protein variations reveals equal purifying selection across the global human mtDNA diversity. *The American Journal of Human Genetics*, *88*(4), 433–439. https://doi.org/10.1016/j.ajhg.2011.03.006

Rajasimha, H. K., Chinnery, P. F., & Samuels, D. C. (2008). Selection against pathogenic mtDNA mutations in a stem cell population leads to the loss of the 3243A>G mutation in blood. *American Journal of Human Genetics*, *82*(2), 333–343. https://doi.org/10.1016/j.ajhg.2007.10.007

Rebolledo-Jaramillo, B., Su, M. S. -W., Stoler, N., McElhoe, J. A., Dickins, B., Blankenberg, D., Korneliussen, T. S., Chiaromonte, F., Nielsen, R., Holland, M. M., Paul, I. M., Nekrutenko, A., & Makova, K. D. (2014). Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences*, *111*(43), 15474. https://doi.org/10.1073/pnas.1409328111

Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., & Wallace, D. C. (2004). Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*, *303*(5655), 223–226. https://doi.org/10.1126/science.1088434

Soares, P., Abrantes, D., Rito, T., Thomson, N., Radivojac, P., Li, B., Macaulay, V., Samuels, D. C., & Pereira, L. (2013). Evaluating purifying selection in the mitochondrial DNA of various mammalian species. *PLoS One*, *8*(3), e58993. https://doi.org/10.1371/journal.pone.0058993

Sondheimer, N., Glatz, C. E., Tirone, J. E., Deardorff, M. A., Krieger, A. M., & Hakonarson, H. (2011). Neutral mitochondrial heteroplasmy and the influence of aging. *Human Molecular Genetics*, *20*(8), 1653–1659. https://doi.org/10.1093/hmg/ddr043

Stewart, J. B., & Chinnery, P. F. (2015). The dynamics of mitochondrial DNA heteroplasmy: Implications for human health and disease. *Nature Reviews Genetics*, *16*(9), 530–542. https://doi.org/10.1038/nrg3966

Stewart, J. B., & Chinnery, P. F. (2021). Extreme heterogeneity of human mitochondrial DNA from organelles to populations. *Nature Reviews Genetics*, *22*(2), 106–118. https://doi.org/10.1038/s41576-020-00284-x

Tatuch, Y., Christodoulou, J., Feigenbaum, A., Clarke, J. T., Wherret, J., Smith, C., Rudd, N., Petrova-Benedict, R., & Robinson, B. H. (1992). Heteroplasmic mtDNA mutation (T–G) at 8993 can cause Leigh disease when the percentage of abnormal mtDNA is high. *American Journal of Human Genetics*, *50*(4), 852–858. https://pubmed.ncbi.nlm.nih.gov/1550128

Uusimaa, J., Finnilä, S., Remes, A. M., Rantala, H., Vainionpää, L., Hassinen, I. E., & Majamaa, K. (2004). Molecular epidemiology of childhood mitochondrial encephalomyopathies in a finnish population: Sequence analysis of entire mtDNA of 17 children reveals heteroplasmic mutations in tRNAArg, tRNAGlu, and tRNALeu(UUR) genes. *Pediatrics*, *114*(2), 443–450. https://doi.org/10.1542/peds.114.2.443

Wallace, D. C., Brown, M. D., & Lott, M. T. (1999). Mitochondrial DNA variation in human evolution and disease. *Gene*, *238*(1), 211–230. https://doi.org/10.1016/S0378-1119(99)00295-4

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), e164. https://doi.org/10.1093/nar/gkq603

Wang, Y., Guo, X., Ye, K., Orth, M., & Gu, Z. (2021). Accelerated expansion of pathogenic mitochondrial DNA heteroplasmies in Huntington's disease. *Proceedings of the National Academy of Sciences*, *118*(30), e2014610118. https://doi.org/10.1073/pnas.2014610118

Wei, W., Tuna, S., Keogh, M. J., Smith, K. R., Aitman, T. J., Beales, P. L., Bennett, D. L., Gale, D. P., Bitner-Glindzicz, M. A. K., Black, G. C., Brennan, P., Elliott, P., Flinter, F. A., Floto, R. A., Houlden, H., Irving, M., Koziell, A., Maher, E. R., Markus, H. S., & Wood, Y. (2019). Germline selection shapes human mitochondrial DNA diversity. *Science*, *364*(6442), eaau6520. https://doi.org/10.1126/science.aau6520

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., & Schönherr, S. (2016). HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, *44*(W1), W58–W63. https://doi.org/10.1093/nar/gkw233

Wilson, I. J., Carling, P. J., Alston, C. L., Floros, V. I., Pyle, A., Hudson, G., Sallevelt, S. C., Lamperti, C., Carelli, V., Bindoff, L. A., Samuels, D. C., Wonnapinij, P., Zeviani, M., Taylor, R. W., Smeets, H. J., Horvath, R., & Chinnery, P. F. (2016). Mitochondrial DNA sequence characteristics modulate the size of the genetic bottleneck. *Human Molecular Genetics*, *25*(5), 1031–1041. https://doi.org/10.1093/hmg/ddv626

Wilton, P. R., Zaidi, A., Makova, K., & Nielsen, R. (2018). A population phylogenetic view of mitochondrial heteroplasmy. *Genetics*, *208*(3), 1261. https://doi.org/10.1534/genetics.118.300711

Yao, Y.-G., Childs, R. W., Kajigaya, S., McCoy, J. P., Jr., & Young, N. S. (2007). Mitochondrial DNA sequence heterogeneity of single CD34+ cells after nonmyeloablative allogeneic stem cell transplantation. *Stem Cells*, *25*(10), 2670–2676. https://doi.org/10.1634/stemcells.2007-0269

Yao, Y.-G., Ellison, F. M., McCoy, J. P., Chen, J., & Young, N. S. (2007). Age-dependent accumulation of mtDNA mutations in murine hematopoietic stem cells is modulated by the nuclear genetic background. *Human Molecular Genetics*, *16*(3), 286–294. https://doi.org/10.1093/hmg/ddl457

Yao, Y.-G., Kajigaya, S., Feng, X., Samsel, L., McCoy, J. P., Torelli, G., & Young, N. S. (2013). Accumulation of mtDNA variations in human single CD34+ cells from maternally related individuals: Effects of aging and family genetic background. *Stem Cell Research*, *10*(3), 361–370. https://doi.org/10.1016/j.scr.2013.01.006

Yao, Y.-G., Kajigaya, S., & Young, N. S. (2015). Mitochondrial DNA mutations in single human blood cells. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *779*, 68–77. https://doi.org/10.1016/j.mrfmmm.2015.06.009

Yao, Y.-G., Kong, Q. P., Salas, A., & Bandelt, H. J. (2008). Pseudomitochondrial genome haunts disease studies. *Journal of Medical Genetics*, *45*(12), 769. https://doi.org/10.1136/jmg.2008.059782

Ye, K., Lu, J., Ma, F., Keinan, A., & Gu, Z. (2014). Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proceedings of the National Academy of Sciences*, *111*(29), 10654–10659. https://doi.org/10.1073/pnas.1403521111

Zaidi, A. A., Wilton, P. R., Su, M. S., Paul, I. M., Arbeithuber, B., Anthony, K., Nekrutenko, A., Nielsen, R., & Makova, K. D. (2019). Bottleneck and selection in the germline and maternal age influence transmission of mitochondrial DNA in human pedigrees. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(50), 25172–25178. https://doi.org/10.1073/pnas.1906331116

Zhang, H., Burr, Stephen, P., & Chinnery, Patrick, F. (2018). The mitochondrial DNA genetic bottleneck: Inheritance and beyond. *Essays in Biochemistry*, *62*(3), 225–234. https://doi.org/10.1042/ebc20170096

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Liu, Q., Iqbal, M. F., Yaqub, T., Firyal, S., Zhao, Y., Stoneking, M., & Li, M. (2022). The transmission of human mitochondrial DNA in four-generation pedigrees. *Human Mutation*, 1–9. https://doi.org/10.1002/humu.24390