

No evidence for intermolecular recombination in human fibroblast and blood mtDNA from individuals with biparental mtDNA transmission

Baoheng Gui^{1,2}, Zeyu Yang¹, Shiyu Luo^{1#}, Jesse Slone¹, Sushma Nagaraj³, Lisa Sadzewicz³, Luke J. Tallon³, and Taosheng Huang^{1*}

¹ Division of Human Genetics, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229, USA

² The Second Affiliated Hospital of Guangxi Medical University, 166 Daxue East Road, Nanning, Guangxi 530003, China

³ Institute for Genome Sciences, University of Maryland School of Medicine, 801 W. Baltimore Street, Baltimore, MD 21201, USA

Current affiliation: Division of Newborn Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA.

Correspondence:

*** Taosheng Huang, M.D., Ph.D.**

Division of Human Genetics

Cincinnati Children's Hospital Medical Center

3333 Burnet Avenue, Cincinnati, Ohio 45229, USA

Phone: 513-803-9260

Taosheng.Huang@cchmc.org

Abstract

Strictly maternal inheritance and lack of intermolecular recombination of the human mitochondrial genome (mtDNA) are the assumed preconditions for molecular evolution studies, phylogenetic reconstruction and population genetic analyses. This hypothesis, however, has been challenged by investigations providing evidence for genetic recombination of mtDNA, thus sparking controversy. Using single-molecule real-time (SMRT) sequencing technology, we sequenced the entire mtDNA from blood and fibroblast cells from five individuals with biparental mtDNA transmission in three separate, multiple-generation families. After phasing the single nucleotide polymorphism (SNP) genotypes of mtDNA, no intermolecular recombination between paternal and maternal mtDNA was found when the mtDNA was transmitted in either biparental or maternal mode. Our study provides support for the argument that intermolecular mtDNA recombination is absent or extremely rare in humans. As a consequence, these results support the feasibility of mtDNA-based molecular evolution studies and phylogenetic and population genetic analyses for humans, while also avoiding inaccurate phylogenetic inferences and incorrect rejection of the molecular clock.

The human mitochondrial genome (mtDNA) serves as a vital tool in molecular evolution studies, phylogenetic reconstruction and population genetic analyses (Ingman, et al. 2000), which mainly depend on the postulation that mtDNA is strictly maternally inherited. This, in turn, leads to the conclusion that mtDNA is maintained clonally, without the complexity generated from biparental recombination. This assumption, however, has been challenged by investigations that have sparked controversy by providing evidence for paternal inheritance and genetic recombination of mtDNA. The earliest evidence of mtDNA recombination is inferred from population genetic linkage disequilibrium (LD) analysis among pairs of mtDNA polymorphisms (Awadalla, et al. 1999) and phylogenetic analyses of excess homoplasies (homozygous mutations at the same genetic site) for phenotypically silent sites located in mtDNA coding regions (Eyre-Walker, et al. 1999). These analyses showed that LD decreased with distance between alleles and that frequency of homoplasies was much higher than expected on the basis of a single rate of synonymous mutations. Both of these traits are manifestations consistent with recombination. However, these results are not reproducible by using either the same measurement technique (Jorde and Bamshad 2000; Elson, et al. 2001) or an alternative methodology (Jorde and Bamshad 2000). Although reanalysis of mtDNA sequences from the formerly published data sets shows a negative correlation between LD and allele distance, most correlations were not statistically significant or even positive (Piganeau and Eyre-Walker 2004). Similarly, consistent and repeatable results of recombination in a human mtDNA data set cannot be obtained by using six well-established indirect tests of recombination (White and Gemmell 2009). In addition, rather than providing a conclusive argument for the existence of mtDNA recombination (Piganeau and Eyre-Walker 2004), high homoplasmy levels in phylogenetic trees could also be attributed to multiple mutations at hypervariable sites, which may be selected because of their phenotypic advantages. These results reveal a lack of unequivocal proof of mtDNA recombination in the published research using indirect evaluation such as LD analysis and homoplasmy evaluation, which may be incapable of determining the very infrequent occurrence of recombination in mtDNA.

Newer, more convincing cases of mtDNA recombination have been documented. Kraytsberg

et al. reported direct evidence for mtDNA recombination (Kraytsberg, et al. 2004). The authors searched for recombination events in the muscle tissues of an individual with biparental inheritance of the mitochondrial genome (Schwartz and Vissing 2002), which may provide an opportunity for intermolecular recombination between paternal and maternal mtDNA (**Figure 1**). Ultimately, they found 33 recombinants, accounting for ~0.7% of the total mtDNA molecules. Given the novelty of the result, they went to great lengths to ensure the *in vivo* origin of these recombinants, repeating the experiment with a reconstructed 10:1 mixture of paternal and maternal DNA to exclude the possibility of recombination produced in laboratory (either by PCR or some other errors). According to this result, the frequency of mtDNA recombination is possible but probably extraordinarily low, since the frequency of paternal leakage is rare.

Further direct evidence of mtDNA recombination between two human cytoplasmic hybrid cell lines supports the idea that recombination is possible in human mtDNA (D'Aurelio, et al. 2004) as evidenced in biparental leakage (Kraytsberg, et al. 2004). The authors fused two human cytoplasmic hybrid cell lines, each containing a distinct pathogenic mtDNA mutation and a specific set of genetic markers used for restriction fragment length polymorphism analysis and relative frequency measurement. Analysis of this hybrid cell model suggested the existence of recombination between the mtDNA haplogroups (~10% of total mtDNA) (D'Aurelio, et al. 2004). However, these artificially created heteroplasmic cells are in a suboptimal condition for studying recombination, as this condition rarely occurs naturally in humans. Therefore, despite this evidence of recombination, the debate remains.

Individuals with biparental mtDNA transmission will be ideal subjects to test possible intermolecular recombination between paternal and maternal mtDNA. We have identified three unrelated multiple-generation families with a high number and level of mtDNA heteroplasmy (Luo, et al. 2018) after amplifying entire mtDNA by a single long-range PCR and performing next-generation sequencing (NGS) analysis based on an established pipeline (Tang and Huang 2010; Huang 2011; Ma, et al. 2015). For instance, in one of the families (Family A, as shown in **Figure 2**), three immediate relative (lineal relative) members (grandfather-mother-son, or I-

I, II-1, and III-2 in the pedigree, respectively) were found to have high levels of mtDNA heteroplasmy (nearly 40% for the minor alleles and 60% for the major alleles). This heteroplasmy covers multiple mtDNA sites in the entire mitochondrial genome (31 sites in both II-1 and III-2, and 20 sites in I-1), which can serve as perfect genetic markers for recombination analysis. These human materials thus provide a unique opportunity to investigate mtDNA recombination.

In this study, we first investigated the possibility of mtDNA recombination by sequencing the entire mtDNA from blood cells of three immediate relative individuals (I-1, II-1, and III-2) with biparental mtDNA transmission in Family A, using single-molecule real-time (SMRT) sequencing technology. The SMRT sequencing technology allows us to revisit the controversial issue of mtDNA recombination accurately by determining the phase of all observed single nucleotide polymorphisms (SNPs) across all or nearly all of the entire 16.5-kbp mitochondrial genome at a single-molecule level, without cross-referencing between different molecules (as occurs in Sanger or NGS). After phasing the genotypes of mtDNA SNPs, no evidence of intermolecular recombination between paternal and maternal mtDNA was found, as shown in **Table 1**. For the immediate relative grandfather (I-1), two distinct haplogroups, R0a1 and H1a1, were identified after phasing of the 17 called genotypes with an average genotype frequency of ~40% and ~60%, respectively. The haplogroup R0a1 was transmitted to his daughter (II-1), who inherited another haplogroup, U5b1d1c, from her mother (I-10). No evidence of recombination between the R0a1 and H1a1 haplogroup was found in the grandfather (I-1) when transmitting the mtDNA in a biparental pattern. Similarly, when maternally transmitting the biparental mtDNA from the mother (II-1) to the son (III-2), no recombination appears to have occurred.

However, this initial analysis was lacking in certain regards. First, this initial dataset was generated when these families were first identified, and thus the data were generated using older PacBio technology with a much higher per base error rate and shorter average reads. The longest single molecules from this dataset were about 12 kbp, and the average read was around

5 kbp. There were relatively few of the ~12 kbp reads, and none of the individual single molecular reads fully covered the entire 16.5 kbp mitochondrial genome. These incomplete full mtDNA reads, while certainly better than the several hundred bases provided by traditional NGS sequencing methods, make this data less than ideal for detecting recombination, since there are much fewer variant sites on each continuous sequence. The high error rate and lower coverage also decrease the accuracy of base calls, and recombination hotspot analysis is likely to be heavily skewed due to very uneven coverage in regions far from the primer start site. There is also a more conceptual question related to the possibility of tissue-specific mtDNA recombination differences. It is possible that mtDNA recombination only occurs in particular cell lineages, and that some lineages experience high rates of mtDNA recombination while others experience no mtDNA recombination at all. Thus, it is important that we test DNA derived from tissues other than blood in order to maximize our chance of detecting mtDNA recombination events.

For these reasons, we obtained three new DNA samples for a new round of SMRT sequencing analysis based on new chemistry and technology (see Materials and Methods). The fibroblast DNA sample was included from individual I-1 of Family A. The other two samples were blood DNA samples obtained from individuals belonging to two other families showing biparental mtDNA transmission: Family B (Individual I-1, **Figure 3**) and Family C (Individual III-1, **Figure 4**). As before, all three DNA samples were subjected to long-range PCR amplification of their full-length mtDNA, and the resulting amplicons were submitted for SMRT sequencing. The resulting reads demonstrated this new approach has a much better average read length and accuracy. Most importantly, for each sample, we were able to obtain a minimum of 1000 full-length reads that span the full length (16,569 bp) of the mitochondrial genome (1712 reads for the fibroblast sample from Family A I-1, 2160 reads for Family B I-1, and 2719 reads for Family C III-1) (**Table 2**). This allowed for a much more refined analysis of recombination hotspots and recombination frequency.

As expected, our results demonstrated the existence of two full-length mtDNA species

corresponding to the predicted maternal and paternal haplogroups for each of the samples (**Figure 5**), further confirming the conclusions of our original study of biparental mtDNA transmission (Luo, et al. 2018). Curiously, our results also showed that a large proportion of reads contained at least one putative crossover event between the maternal and paternal haplotypes (**Figure 6**). However, a deeper analysis indicates that these recombination events are very likely spurious, for the following reasons. First, the distribution of breakpoints (i.e. putative “crossover” sites) in recombinants for all three samples was near linear, with no apparent recombination hotspots (**Figure 7**). This result would seem to suggest that recombination is random, which conflicts with literature showing breakpoint hotspots at sites with biological significance (Kraytsberg, et al. 2004). The number of recombinants found in our samples, 28-55% of total mtDNA, is also much higher than the rates of 0.7% in Kraytsberg et al. A likely explanation is that the recombinants were generated by PCR and not *in vivo* (Lahr and Katz 2009). Modeling of PCR chimera formation rates shows a similar distribution of recombinants and non-recombinants as observed in our actual data (**Figure 8**), further suggesting that these events are PCR artifacts rather than true, *in vivo* recombination events. It should also be noted that the ratio of purely paternal to purely maternal mtDNA sequences was different from the heteroplasmy levels obtained from short-read sequencing (**Table 2**). In the phased PacBio CCS data, the majority parental haplogroup (whether paternal or maternal) accounted for a greater portion of the full-length non-recombinant sequences than expected from previous variant calling. A possible explanation for this is that PCR-based recombination would shift the proportion of unrecombined reads in favor of the majority haplogroup. In the case where maternal sequences outnumber paternal sequences, both sequences have the same chance of failing to fully extend (the most common cause of PCR-based recombination). However, a partially extended sequence is more likely to anneal to a maternal template, since in this case, maternal sequences are more numerous. Under these conditions, when a maternal sequence undergoes PCR-based recombination, it is very likely to recombine to another maternal template, which would generate a purely maternal sequence. When the paternal sequence undergoes PCR-based recombination, however, it is also likely to recombine to a maternal sequence, generating a recombinant sequence. This shift can be seen in both cases

where the maternal sequence is the majority haplogroup (Family A and B) and in the case where the paternal sequence is the majority (Family C). Thus, on balance, the evidence obtained from our SMRT sequencing data indicates that these “recombination” events are PCR artifacts, and that true *in vivo* recombination is exceedingly rare, if not entirely non-existent.

By performing SMRT sequencing across multiple families and tissue types, our study further indicates the absence or lack of intermolecular mtDNA recombination in humans when the mtDNA is transmitted to offspring in either biparental or maternal mode. Our evidence comes from direct observation of recombination events in lineal relative members, rather than indirect population genetic inference. Furthermore, multiple informative SNPs from around the entire mtDNA were called for genotyping and phasing of the haplogroup, eliminating errors due to phasing mistakes and providing solid support for the argument. For a detectable mtDNA recombination event, it is necessary that at least two distinct non-allelic variants collocate in the same mtDNA sequence (**Figure 1**). Human mitochondrial genomes possess low-frequency rearrangements, i.e., partial duplication or deletion, generated by intramolecular recombination on break-point hotspot regions often defined by short direct repeats (Kajander, et al. 2000). However, little evidence has been found indicating the possibility of intermolecular recombination in human mtDNA. A plausible manner that mtDNA recombination could influence genetic diversity is by exchanging genetic information between biparental molecules at the stage of fertilization and pre-implantation embryo development, during which the paternal mtDNA may be eliminated gradually by active mechanisms of selective ubiquitin-dependent protein degradation and/or mitophagy (Song, et al. 2016). Even if paternal mtDNA fails to be effectively eliminated, resulting in biparental heteroplasmy (Schwartz and Vissing 2002; Luo, et al. 2018) and providing the potential for genomic exchanges, the occurrence of recombination between the paternal mtDNA and maternal mtDNA requires physical proximity of the two mitochondrial genomes. However, mammalian mitochondria typically only contain a single copy of mtDNA (Kukat, et al. 2011), which reduces the chance of recombination events between two molecules expected to be within close proximity. Additionally, no definitive evidence has been discovered to indicate whether or not paternal and maternal heterologous

mitochondria can fuse together to facilitate the proximity of the biparental mtDNA and exchange of their genetic material.

Intermolecular mtDNA recombination has been reported in some taxa, but is still controversial in others. There is increasing evidence confirming that mtDNA undergo recombination in plants (Stadler and Delph 2002; Bergthorsson, et al. 2003; Apitz, et al. 2013) and fungi (Saville, et al. 1998), and such events readily occur even in yeast (MacAlpine, et al. 1998; Fritsch, et al. 2014). However, it is still difficult to come to a definite conclusion for widespread animal mtDNA recombination (Greiner, et al. 2015). Direct evidence for recombination of mtDNA in species like the nematode *Meloidogyne javanica* (Lunt and Hyman 1997), the mussel *Mytilus galloprovincialis* (Ladoukakis and Zouros 2001a), and the flatfish *Platichthys flesus* (Hoarau, et al. 2002) has been well documented. It is worth noting that intermolecular recombination should be much easier to observe in species which do not follow strict uniparental inheritance of mitochondria. For example, the mussel *Mytilus galloprovincialis* possesses two lineages of mtDNA, each with a distinct transmission mode: a phenomenon known as doubly uniparental inheritance (DUI) (Zouros, Ball, et al. 1994; Zouros, Oberhauser Ball, et al. 1994). Several computational statistical analyses comparing the published animal mtDNA sequences also provided evidence for recombination even in organisms with standard maternal mtDNA inheritance (Ladoukakis and Zouros 2001b; Tsaousis, et al. 2005). However, when using different methods to test the potential recombination events of these mtDNA sequences, no or only weak evidence could be found in many of these species (Smith and Smith 2002). A recent study used an *in vivo* mouse model with germline heteroplasmy for a defined set of mtDNA mutations for more than 50 generations to detect potential recombination events (Hagstrom, et al. 2014). Based on cloning of single mtDNA molecules in the λ phage (without prior PCR amplification), followed by subsequent mutation analysis, no germline recombination was found after transmission of mtDNA under genetically and evolutionarily relevant conditions. These findings provide similar evidence as what was observed in our study; that is, the distinct populations of mtDNA carried by these mice did not show any observable evidence of recombination when transmitted maternally.

In summary, the results in our study largely deny the occurrence of intermolecular mtDNA recombination in both biparental and maternal transmission modes, thus supporting the feasibility of mtDNA-based molecular evolution studies, phylogenetic inference and population genetic analyses for humans, while also avoiding spurious phylogenetic inferences and incorrect rejection of the molecular clock. Furthermore, these findings also broaden our knowledge of human mtDNA-associated diseases. However, the question of intermolecular mtDNA recombination remains far from resolved, and more evidence and critical discussion are warranted to come to a definite conclusion and to develop our understanding of human and mitochondrial evolution.

Material and Methods

Subjects

Three unrelated multi-generation families with a high number and level of mtDNA heteroplasmy were previously identified (Luo, et al. 2018). For this study, heteroplasmic individuals from each of these families were chosen for further analysis. For Family A, three immediate related individuals (lineal relatives) across multiple generations were chosen: a grandfather (I-1), mother (II-1), and son (III-2), as shown in the pedigree in **Figure 2**. For Family B and C, a single individual was chosen from each family for analysis: I-1 for Family B (**Figure 3**), and III-1 for Family C (**Figure 4**). Family A and Family B were originally evaluated at the MitoClinic and referred to the Mitochondrial Diagnostic Laboratory at the Human Genetics Division of Cincinnati Children's Hospital Medical Center (CCHMC), and Family C was evaluated at Mayo Clinic and genetic testing was done at the Diagnostic Laboratory at Baylor College of Medicine. For all families, mtDNA analysis was performed based on long-range PCR and next-generation sequencing (Tang and Huang 2010; Huang 2011; Ma, et al. 2015; Luo, et al. 2018). To investigate the possibility of recombination events in these individuals' mtDNA, written informed consent was obtained from all participants. This study was approved by the Institutional Review Board of CCHMC (approval study ID: 2013-

7868).

Whole mitochondrial DNA amplification

Genomic DNA was isolated from the blood and fibroblast samples using Genra DNA extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Entire mtDNA was amplified as a single, long-range PCR amplicon as previously described (Tang and Huang 2010; Huang 2011; Ma, et al. 2015). One hundred nanograms of total genomic DNA were used as template in a 50- μ L PCR system. The primers specifically recognize genuine mtDNA: F-2120 (GGACACTAGGAAAAACCTTG TAGAGAGAG) and R-2119 (AAAGAGCTGTTCTCTTTGGACTAACA). PCR amplifications were performed using TaKaRa LA Taq Hot Start polymerase (TaKaRa Biotechnology, Kyoto, Japan). PCR conditions were: 94 °C for 1 min; 98 °C for 10 sec and 68 °C for 16 min, 30 cycles; 72 °C for 10 min and hold at 4 °C.

Single-molecule real-time (SMRT) sequencing and analysis

The long-range PCR products of entire mtDNA molecules were used for SMRT sequencing at the Institute of Genome Science, University of Maryland for the blood DNA samples for I-I, II-1, and III-2 of Family A, and at Novogene (Beijing, China) for the fibroblast DNA from I-1 of Family A and the blood DNA for I-1 in Family B and III-1 in Family C.

For the data generated at the University of Maryland, DNA libraries were constructed and prepared for sequencing using the SMRTbell Template Prep Kit 2.0 and the DNA/Polymerase Binding Kit 2.0 (Pacific Biosciences, Menlo Park, CA) according to the manufacturer's instructions. 16.5-kbp fragments were size-selected with the BluePippin device (Sage Sciences, Beverly, MA), and were sequenced with the DNA Sequencing Kit 2.0 (Pacific Biosciences, Menlo Park, CA) using a 4-hour movie collection. The raw subread data was analyzed using PacBio's Long Amplicon Analysis module (Smrtanalysis 2.3.0) to find phased consensus sequences, using PacBio subreads that were longer than 13 kbp. The resulting phased sequences were aligned to the mitochondrial reference genome and variant calling was

performed for each. The variants were then used to assign a genotype to each phased sequence by comparing them with variants found in the parental/ancestral mitochondrial genomes (Kurtz, et al. 2004). Circular consensus sequence (CCS) were also generated from the amplicon data and used to calculate allele/haplogroup frequencies (Chaisson and Tesler 2012).

For the data generated at Novogene, DNA libraries were constructed and prepared for sequencing using the SMRTbell Template Prep Kit 2.0 and the Sequel II Binding Kit 2.0 (Pacific Biosciences, Menlo Park, CA) according to the manufacturer's instructions. The PCR amplicon were sequenced without size-selection using PacBio's Sequel System (Pacific Biosciences, Menlo Park, CA), using Chemistry 3.0, Software version 6.0, and a 10-hour movie collection time. The raw subread data was used to generate consensus sequences using pbccs, a command-line tool provided by PacBio. All three data sets were run with minimum subread length at 13 kbp, filtering out shorter reads. Due to the position of the primers, full-length reads started at base position 2120 of the reference NC_012920 genome and ended at 2119. To facilitate accurate alignment of these long reads, the reference mtDNA sequence was modified by moving bases 1-2119 to the end of the reference sequence. The consensus reads were then aligned to the modified sequence using pbalgn, also part of the PacBio command-line tools. Using a quality filter of $RQ \geq .995$ and $minLength = 16000$, the aligned consensus reads were reduced to only high-quality reads that are full-length or nearly full-length. A custom script was then used to separate the aligned reads into parental haplogroups and recombinant groups. We used the SNP sites that differed between the two parental haplogroups to count the number of paternal and maternal variants on each read. Using the pattern of variant haplotypes, we sorted the reads into their respective groups. The position and haplotype of the sites were based on the original analysis of the short-read data for these families (Luo, et al. 2018). Each CCS sequence was assigned to either the paternal or maternal haplogroup based on the first heteroplasmic variant site. Each site where the paternal and maternal haplogroups disagreed was checked to see if it was paternal or maternal. If variants switched haplogroup and the new haplogroup was sustained for at least two variant sites, then the CCS sequence was marked as

recombinant and a crossover at the site recorded. The reads were visualized and checked in Integrative Genomics Viewer (IGV) software (Robinson, et al. 2011). A map of the breakpoint sites was generated to check for potential recombination hotspots.

The model of PCR recombinant distribution was generated based on the assumption that for each round of PCR, a certain percentage of sequences would not fully extend. The template to which the partially extended DNA strand anneals is determined by the distribution of sequences after the previous round of PCR. The final distribution of parental and recombinant sequences after 30 rounds of PCR was compared to the observed distribution of sequences in the three samples. A predicted distribution was generated for all combinations of heteroplasmy level and recombination rates, and the results used to generate chi-square statistics to find the combination of best fit.

For all datasets, the mitochondrial haplogroups were identified based on defining SNP genotypes provided by SNPedia (https://www.snpedia.com/index.php/MtDNA_Haplogroup) and MITOMAP (<http://www.mitomap.org/foswiki/bin/view/MITOMAP/HaplogroupMarkers>).

Declaration of interests

None.

Acknowledgements and funding

The authors would like to thank the subjects for participating in our study. We also thank the Institute of Genome Science, University of Maryland, Novogene, and the DNA Core facility of CCHMC, where the SMRT sequencing and next-generation sequencing were performed. This work was supported by the Cincinnati Children's Hospital Research Foundation and the Eunice Kennedy Shriver National Institute of Child Health & Human Development at the National Institute of Health (1R01HD092989-01A1) to T.H.

References

Apitz J, Weihe A, Pohlheim F, Borner T. 2013. Biparental inheritance of organelles in *Pelargonium*: evidence for intergenomic recombination of mitochondrial DNA. *Planta* 237:509-515.

Awadalla P, Eyre-Walker A, Smith JM. 1999. Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* 286:2524-2525.

Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424:197-201.

Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* 13:238.

D'Aurelio M, Gajewski CD, Lin MT, Mauck WM, Shao LZ, Lenaz G, Moraes CT, Manfredi G. 2004. Heterologous mitochondrial DNA recombination in human cells. *Hum Mol Genet* 13:3171-3179.

Elson JL, Andrews RM, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 2001. Analysis of European mtDNAs for recombination. *Am J Hum Genet* 68:145-153.

Eyre-Walker A, Smith NH, Smith JM. 1999. How clonal are human mitochondria? *Proc Biol Sci* 266:477-483.

Fritsch ES, Chabbert CD, Klaus B, Steinmetz LM. 2014. A genome-wide map of mitochondrial DNA recombination in yeast. *Genetics* 198:755-771.

Greiner S, Sobanski J, Bock R. 2015. Why are most organelle genomes transmitted maternally? *Bioessays* 37:80-94.

Hagstrom E, Freyer C, Battersby BJ, Stewart JB, Larsson NG. 2014. No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline. *Nucleic Acids Res* 42:1111-1116.

Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL. 2002. Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. *Mol Biol Evol* 19:2261-2264.

Huang T. 2011. Next generation sequencing to characterize mitochondrial genomic DNA heteroplasmy. *Curr Protoc Hum Genet* Chapter 19:Unit19 18.

Ingman M, Kaessmann H, Paabo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708-713.

Jorde LB, Bamshad M. 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* 288:1931.

Kajander OA, Rovio AT, Majamaa K, Poulton J, Spelbrink JN, Holt IJ, Karhunen PJ, Jacobs HT. 2000. Human mtDNA

sublimons resemble rearranged mitochondrial genomes found in pathological states. *Hum Mol Genet* 9:2821-2835.

Kraytsberg Y, Schwartz M, Brown TA, Ebralidse K, Kunz WS, Clayton DA, Vissing J, Khrapko K. 2004. Recombination of human mitochondrial DNA. *Science* 304:981.

Kukat C, Wurm CA, Spahr H, Falkenberg M, Larsson NG, Jakobs S. 2011. Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *Proc Natl Acad Sci U S A* 108:13534-13539.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* 5:R12.

Ladoukakis ED, Zouros E. 2001a. Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Mol Biol Evol* 18:1168-1175.

Ladoukakis ED, Zouros E. 2001b. Recombination in animal mitochondrial DNA: evidence from published sequences. *Mol Biol Evol* 18:2127-2131.

Lahr DJ, Katz LA. 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *Biotechniques* 47:857-866.

Lunt DH, Hyman BC. 1997. Animal mitochondrial DNA recombination. *Nature* 387:247.

Luo S, Valencia CA, Zhang J, Lee NC, Slone J, Gui B, Wang X, Li Z, Dell S, Brown J, et al. 2018. Biparental Inheritance of Mitochondrial DNA in Humans. *Proc Natl Acad Sci U S A* 115:13039-13044.

Ma H, Folmes CD, Wu J, Morey R, Mora-Castilla S, Ocampo A, Ma L, Poulton J, Wang X, Ahmed R, et al. 2015. Metabolic rescue in pluripotent cells from patients with mtDNA disease. *Nature* 524:234-238.

MacAlpine DM, Perlman PS, Butow RA. 1998. The high mobility group protein Abf2p influences the level of yeast mitochondrial DNA recombination intermediates in vivo. *Proc Natl Acad Sci U S A* 95:6739-6743.

Piganeau G, Eyre-Walker A. 2004. A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. *Heredity (Edinb)* 92:282-288.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* 29:24-26.

Saville BJ, Kohli Y, Anderson JB. 1998. mtDNA recombination in a natural population. *Proc Natl Acad Sci U S A* 95:1331-1335.

Schwartz M, Vissing J. 2002. Paternal inheritance of mitochondrial DNA. *N Engl J Med* 347:576-580.

Smith JM, Smith NH. 2002. Recombination in animal mitochondrial DNA. *Mol Biol Evol* 19:2330-2332.

Song WH, Yi YJ, Sutovsky M, Meyers S, Sutovsky P. 2016. Autophagy and ubiquitin-proteasome system contribute to sperm mitophagy after mammalian fertilization. *Proc Natl Acad Sci U S A* 113:E5261-5270.

Stadler T, Delph LF. 2002. Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *Proc Natl Acad Sci U S A* 99:11730-11735.

Tang S, Huang T. 2010. Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system. *Biotechniques* 48:287-296.

Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E. 2005. Widespread recombination in published animal mtDNA sequences. *Mol Biol Evol* 22:925-933.

White DJ, Gemmell NJ. 2009. Can indirect tests detect a known recombination event in human mtDNA? *Mol Biol Evol* 26:1435-1439.

Zouros E, Ball AO, Saavedra C, Freeman KR. 1994. Mitochondrial DNA inheritance. *Nature* 368:818.

Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR. 1994. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proc Natl Acad Sci U S A* 91:7463-7467.

Figures and Figure Legends

Figure 1. Model of hypothetical recombination between mtDNA molecules. Pink and blue circles represent maternal and paternal mtDNA molecules, respectively. (A) Hypothetical genomic crossover between homoplasmic mtDNA molecules transmitted maternally, which would be expected to rarely generate new haplotypes. (B) Hypothetical recombination between maternal and paternal mtDNA molecules in a biparental inheritance mode. For a detectable recombination event to occur, at least two distinct non-allelic variants must collocate in the mtDNA sequence. In this example, the genotypes C and A are collocated in the maternal mtDNA sequence, while the genotypes T and G are collocated in the paternal sequence. If a crossover event happens (at the locus indicated in the figure), the maternal haplotype turns into C combined with G, while the paternal haplotype changes to T combined with A. After sequencing, different reads covering different regions (as indicated in the figure) are used for genotyping and phasing of the haplotypes.

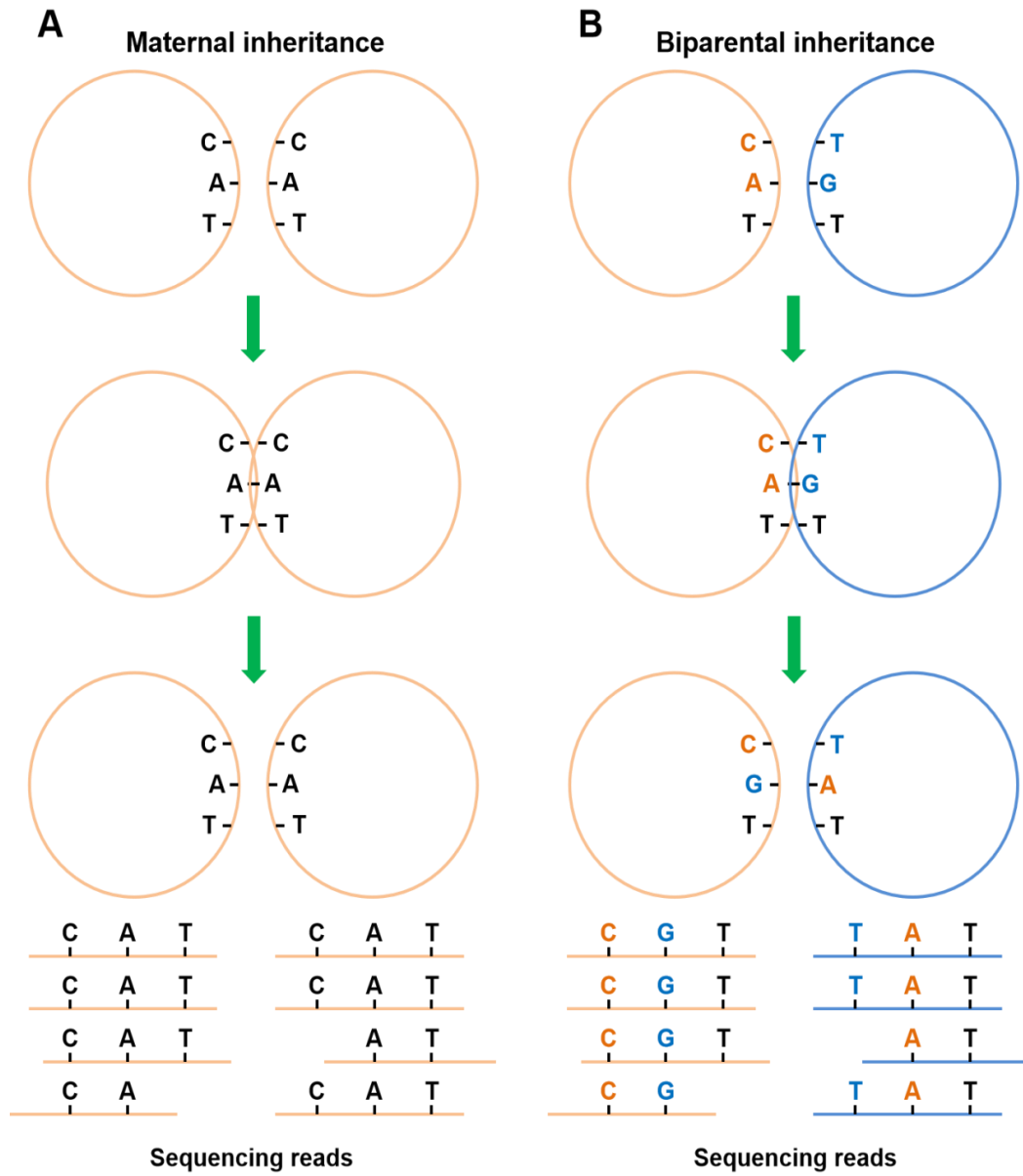


Figure 2. Pedigree of a multiple-generation family (Family A) with a high number and level of mtDNA heteroplasmy. The arrow indicates the proband for the family. Dark squares and circles show individuals with a high level of mtDNA heteroplasmy (nearly 40% for the minor alleles and 60% for the major alleles) at multiple mtDNA sites (31 sites in II-1, III-1, III-2 and III-3, and 20 sites in I-1). Their haplogroups and the corresponding average heteroplasmy ratio are indicated. For example, the haplogroup R0a1 and H1a1 account for ~40% and ~60% of the mtDNA, respectively, in I-1. For SMRT sequencing, whole mitochondrial DNA was PCR-amplified from the blood samples of I-1, II-2, and III-2 (**Table 1**), as well as from the fibroblasts of I-1 (**Figures 7-8**).

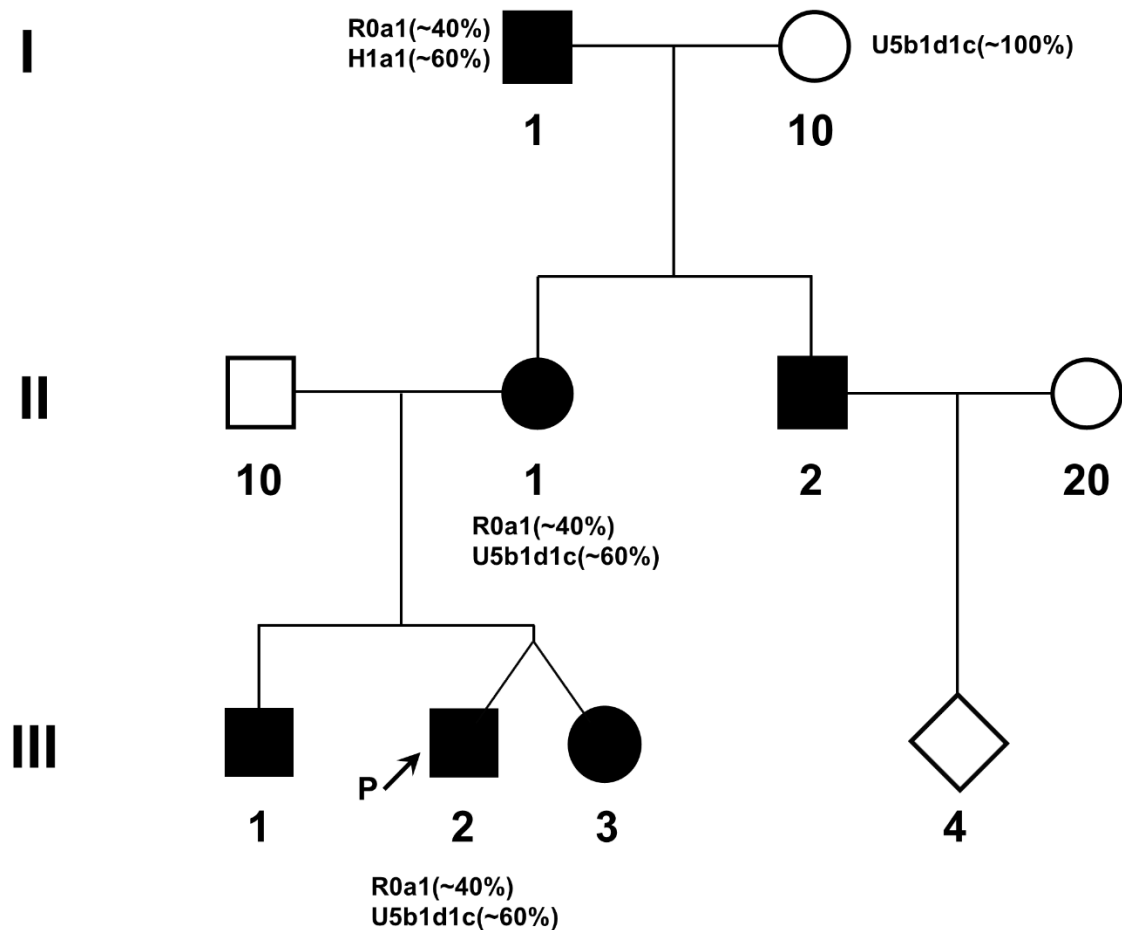


Figure 3. Pedigree of a multiple-generation family (Family B) with a high number and level of mtDNA heteroplasmy. The arrow indicates the proband for Family B. Dark squares and circles show individuals with a high level of mtDNA heteroplasmy at multiple mtDNA sites (29 sites in I-1 and 44 sites in II-2). Their haplogroups and the corresponding average heteroplasmy ratio are indicated. For SMRT sequencing, whole mitochondrial DNA was PCR-amplified from the blood sample of I-1, the proband's father (**Figures 7-8**).

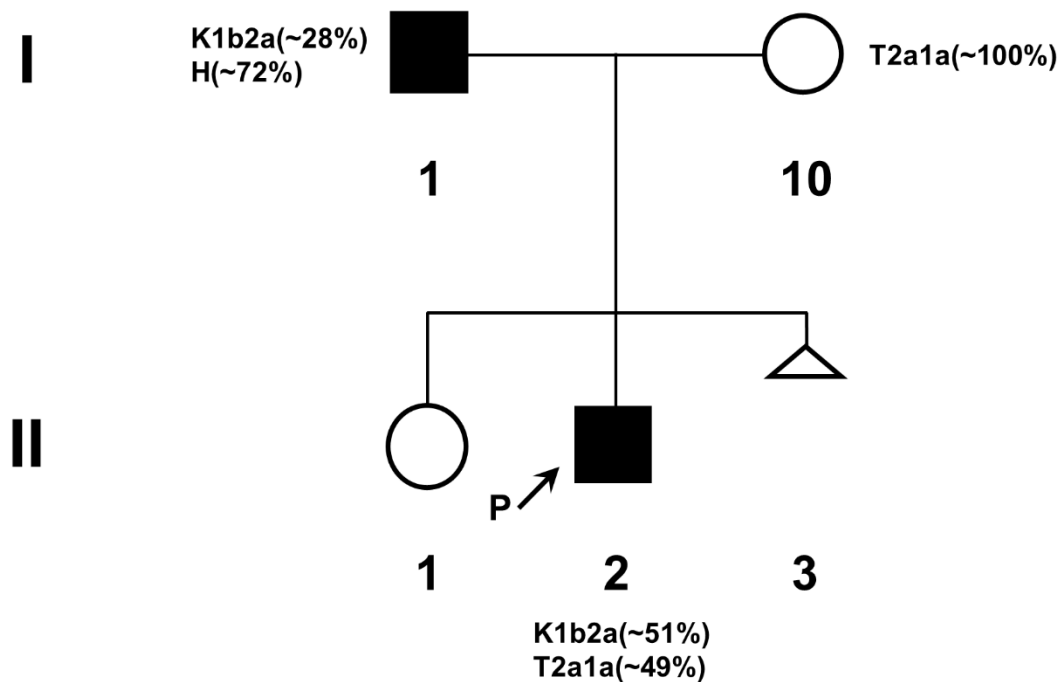


Figure 4. Pedigree of a multiple-generation family (Family C) with a high number and level of mtDNA heteroplasmy. The arrow indicates the proband for Family C. Dark squares and circles show individuals with a high level of mtDNA heteroplasmy at multiple mtDNA sites (51 sites in I-1 and 13 sites in II-1, II-2, III-1, and III-2). Their haplogroups and the corresponding average heteroplasmy ratio are indicated. For SMRT sequencing, whole mitochondrial DNA was PCR-amplified from the blood sample of III-1, the proband (**Figures 5-8**).

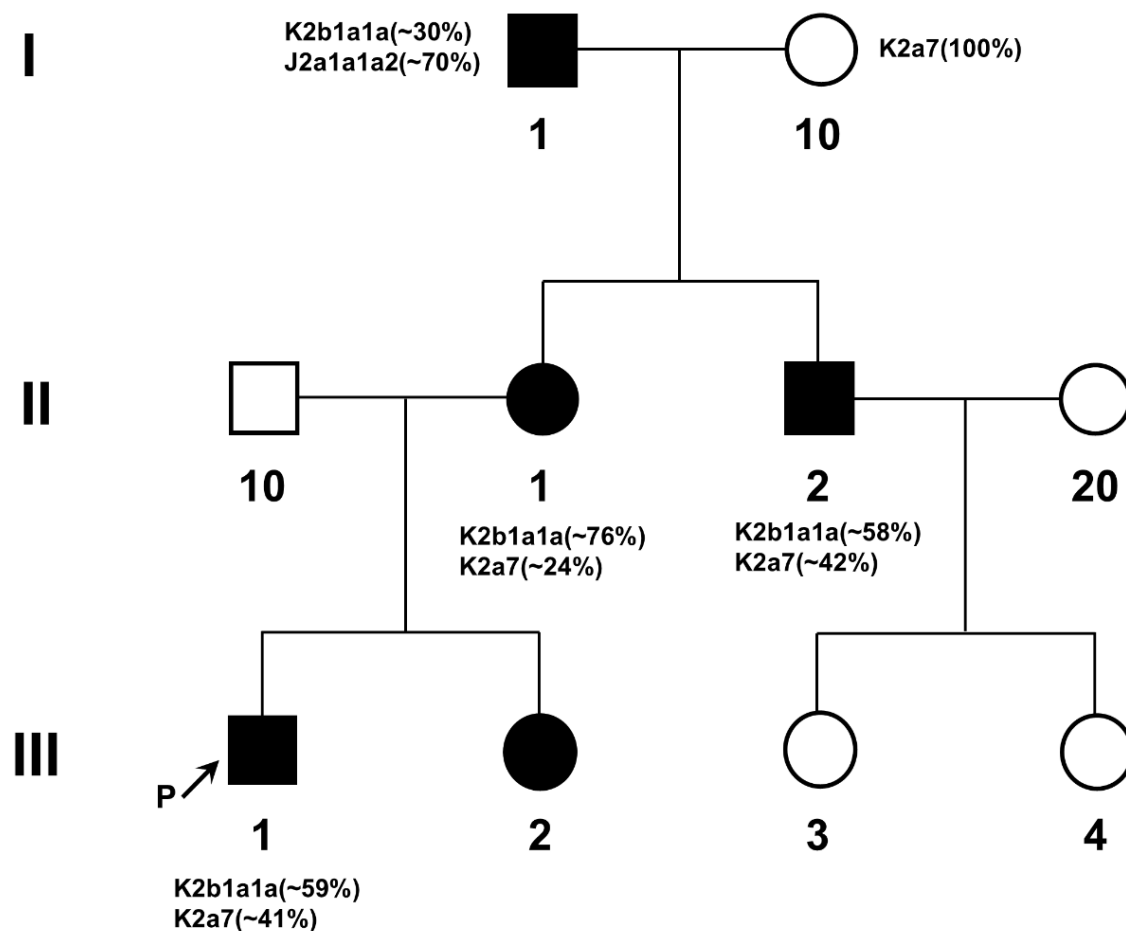


Figure 5. Comparison of sequence reads for parental haplogroups of III-1 of Family C.

This diagram shows the non-recombinant reads from the original parental haplogroups for III-1 from Family C. The vertical bars with different colors represent different base substitutions when compared to the rCRS reference genome (NC_012920). The maternal haplogroup is K2a7, and the paternal haplogroup is K2b1a1a. The CCS reads are aligned to the modified reference file, filtered for quality and length, then visualized in Integrative Genomics Viewer (IGV) software.

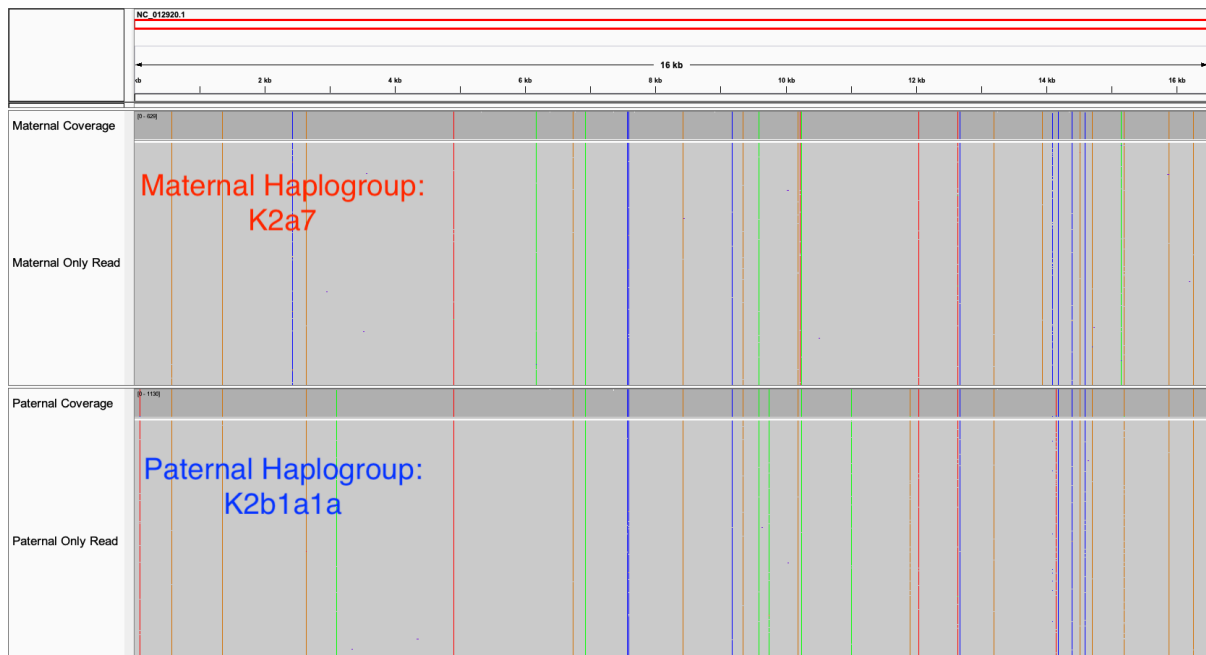


Figure 6. Recombinant sequences with one breakpoint from III-1 in Family C. This diagram shows all of the single-breakpoint reads from III-1 in Family C. The yellow band indicates the location where the sequences switch from one haplogroup to the other. The vertical bars with different colors represent different base substitutions when compared to the rCRS reference genome (NC_012920). For comparison, non-recombinant reads for the maternal haplogroup are shown at the top of the diagram, and non-recombinant reads for the paternal haplogroup are shown at the bottom. As in **Figure 5**, the original maternal haplogroup is K2a7, and the paternal haplogroup is K2b1a1a.

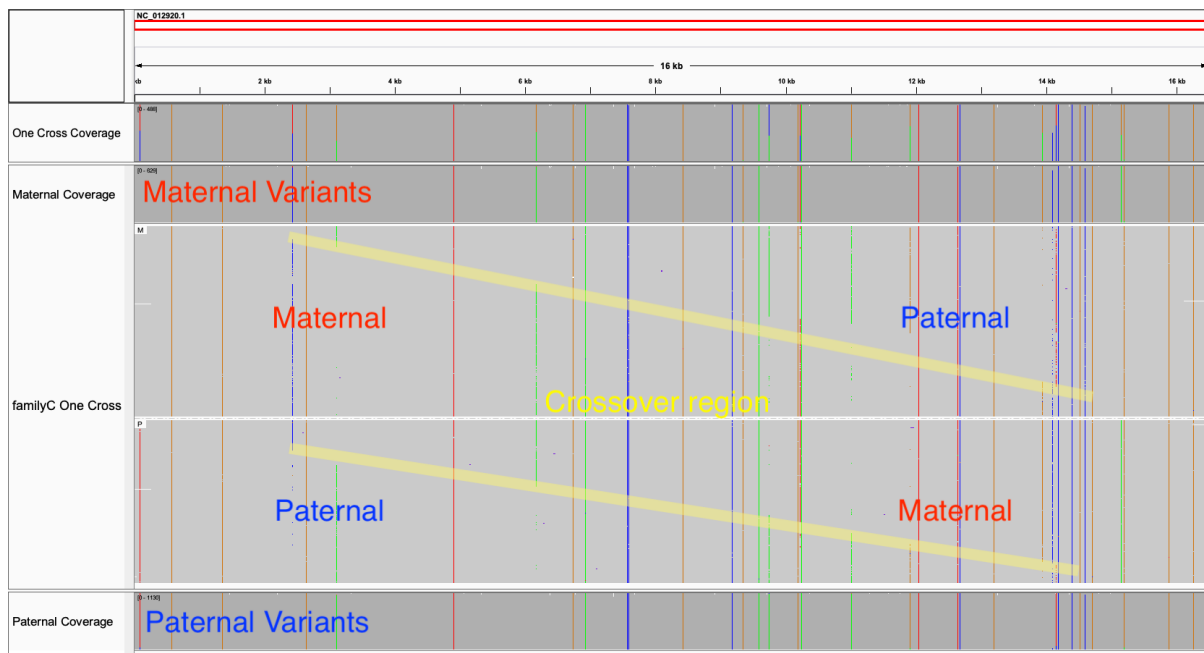


Figure 7. Map of breakpoint sites and frequency. These graphs show the cumulative percentage of reads with a breakpoint at each site in the mitochondrial genome. A custom script was used to count the number of strands with a breakpoint at each variant site. The rates are cumulative to normalize for the distance between variant sites. Position “0” below corresponds to position “2120” in the rCRS, which is the annealing site of the forward primary used to generate the PCR amplicon used for sequencing. In the same way, position “16569” below corresponds to “2119” in the rCRS, which is the annealing site of the reverse primer used to generate the amplicon.

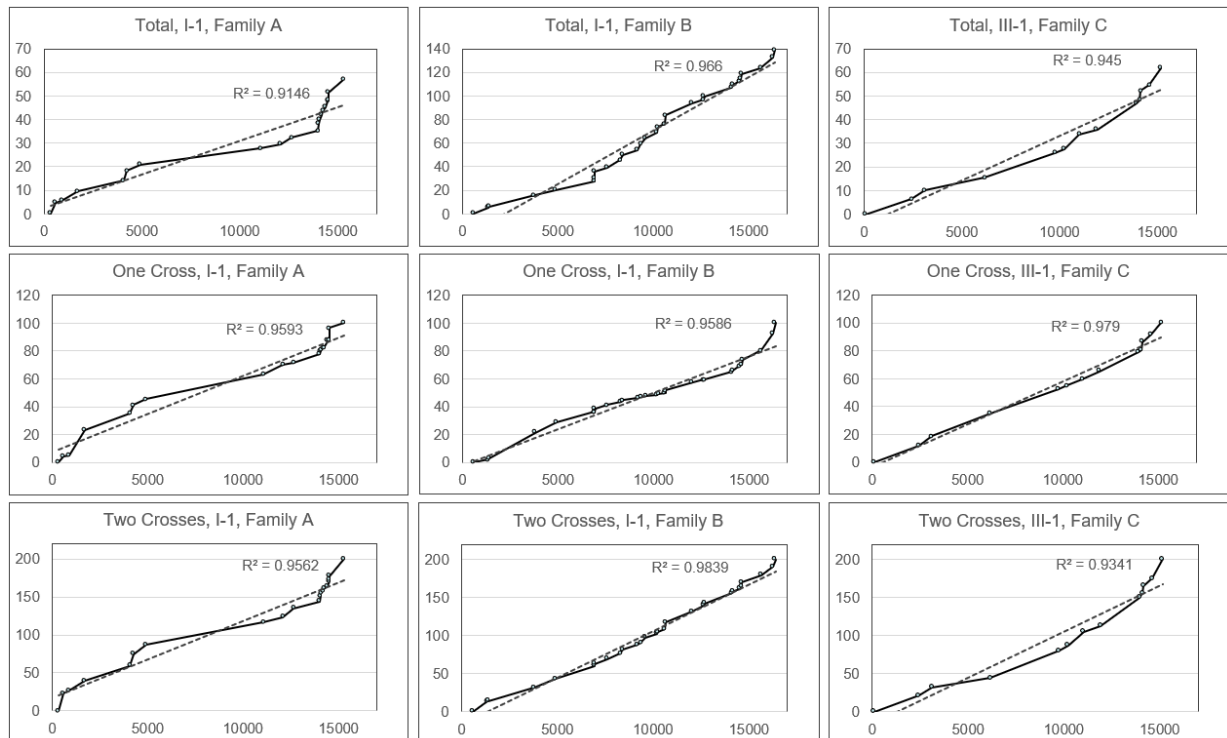
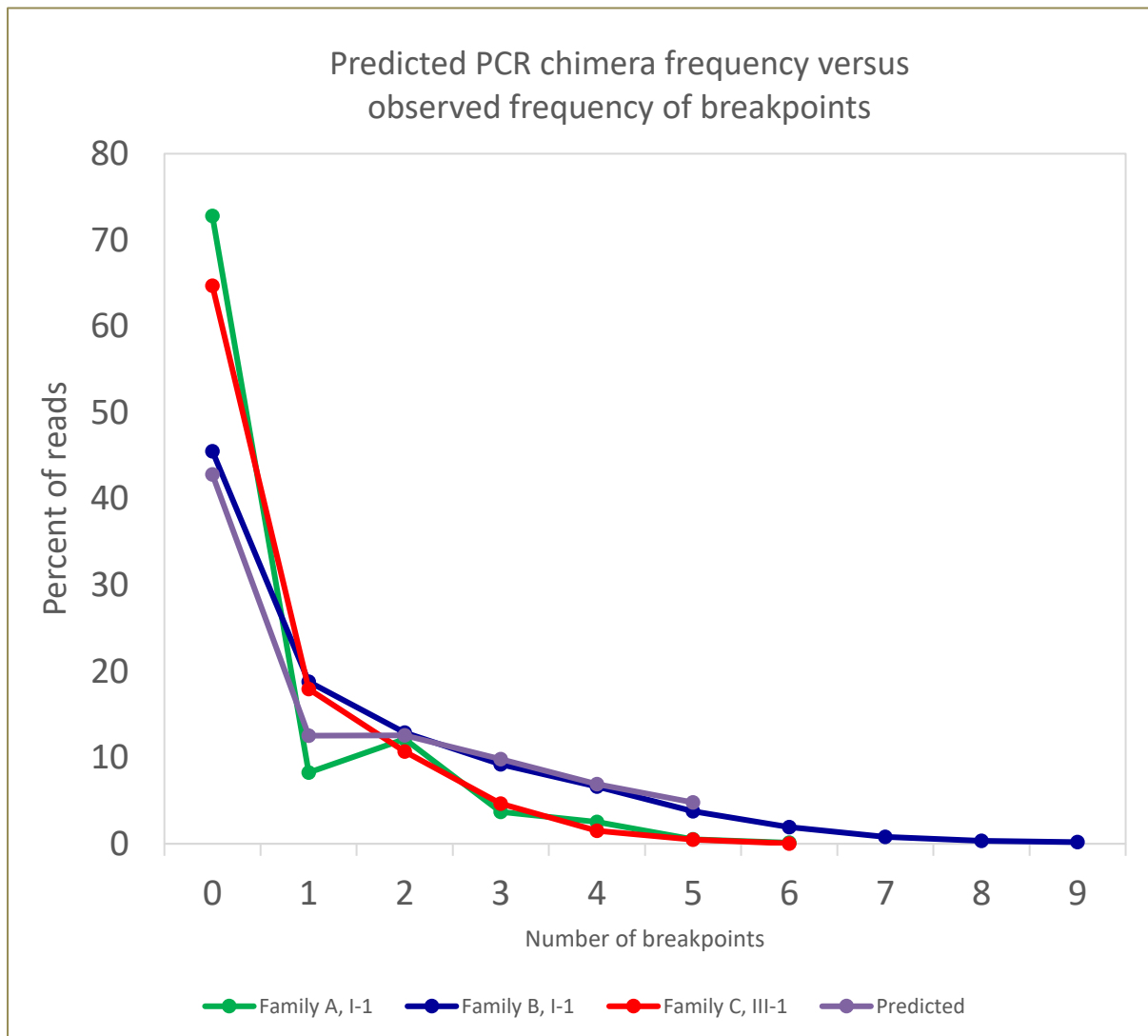


Figure 8. Modeling of predicted PCR chimera rates relative to observed breakpoint frequencies. This graph shows the distribution of reads according to the number of breakpoints observed in the SMRT sequencing data for the fibroblasts from I-1 of Family A, as well as the blood samples from I-1 of Family B and III-1 of Family C. The predicted distribution for PCR-generated chimeras is plotted for comparison.



Tables

Table 1. Phasing of the haplogroups in the three immediate relative individuals from Family A. The query genotypes found in specific individuals represent genotypes different from the reference sequence (rCRS, or “Revised Cambridge Reference Sequence”). The rCRS is the corrected human mtDNA reference sequence, which is deposited in the GenBank NCBI database under accession number NC_012920. The percentage of specific genotypes for different mtDNA markers varied mainly because they are covered by different sequencing reads, resulting in different read depth for specific loci.

Position	rCRS	Grandfather (I-1)									Mother (II-1)									Son (III-2)								
		A	C	G	T	Query	R0a1	H1a1	% R0a1 genotype	% H1a1 genotype	A	C	G	T	Query	R0a1	U5b1d1c	%R0a1 genotype	%U5b1d1c genotype	A	C	G	T	Query	R0a1	U5b1d1c	%R0a1 genotype	%U5b1d1c genotype
58	T	0	33	0	53	C	C	T	38%	62%	0	16	0	25	C	C	T	39%	61%	0	33	0	67	C	C	T	33%	67%
64	C	0	52	0	22	T	T	C	30%	70%	0	26	0	12	T	T	C	32%	68%	0	63	0	19	T	T	C	23%	77%
73	A	32	0	54	0	G	A	G	37%	63%	14	0	27	0	G	A	G	34%	66%	37	0	63	0	G	A	G	37%	63%
150	C										0	19	0	21	T	C	T	48%	53%	0	32	0	63	T	C	T	34%	66%
827	A	51	0	27	0	G	G	A	35%	65%	31	0	14	0	G	G	A	31%	69%	61	0	40	0	G	G	A	40%	60%
2179	A										19	0	25	0	G	A	G	43%	57%	37	0	56	0	G	A	G	40%	60%
2442	T	0	23	0	49	C	C	T	32%	68%	0	15	0	26	C	C	T	37%	63%	0	29	0	58	C	C	T	33%	67%
2706	A	67	0	34	0	G	G	A	34%	66%																		
3010	G	62	0	34	0	A	G	A	35%	65%																		
3197	T										0	20	0	16	C	T	C	44%	56%	0	94	0	55	C	T	C	37%	63%
3847	T	0	25	0	72	C	C	T	26%	74%	0	12	0	20	C	C	T	38%	63%	0	38	0	105	C	C	T	27%	73%

5437	C										0	11	0	21	T	C	T	34%	66%	0	44	0	88	T	C	T	33%	67%	
5656	A										16	0	26	0	G	A	G	38%	62%	61	0	92	0	G	A	G	40%	60%	
6182	G	33	0	60	0	A	A	G	35%	65%	17	0	26	0	A	A	G	40%	60%	56	0	97	0	A	A	G	37%	63%	
6365	T	0	55	0	34	C	T	C	38%	62%																			
7028	C	0	51	0	30	T	T	C	37%	63%																			
7768	A										10	0	15	0	G	A	G	40%	60%	30	1	61	0	G	A	G	33%	66%	
7912	G										13	0	10	0	A	G	A	43%	57%	52	0	30	0	A	G	A	37%	63%	
9477	G										6	0	3	0	A	G	A	33%	67%	12	0	5	0	A	G	A	29%	71%	
11467	A										1	0	5	0	G	A	G	17%	83%	5	0	12	0	G	A	G	29%	71%	
11719	G										4	0	1	0	A	G	A	20%	80%	14	0	2	0	A	G	A	13%	88%	
12308	A										2	0	1	0	G	A	G	67%	33%	3	0	15	0	G	A	G	17%	83%	
12372	G										1	0	3	0	A	G	A	75%	25%	14	0	5	0	A	G	A	26%	74%	
13188	C	0	8	0	5	T	T	C	38%	62%	0	4	0	4	T	T	C	50%	50%	0	16	0	8	T	T	C	33%	67%	
13617	T										0	6	0	5	C	T	C	45%	55%	0	22	0	16	C	T	C	42%	58%	
14182	T										0	11	0	10	C	T	C	48%	52%	0	35	0	17	C	T	C	33%	67%	
14198	G	21	0	12	0	A	G	A	36%	64%																			
14766	C	0	31	0	23	T	T	C	43%	57%																			
15631	A										16	0	25	0	G	A	G	39%	61%	32	0	67	0	G	A	G	32%	68%	
15721	T										0	25	0	15	C	T	C	38%	63%	0	63	0	33	C	T	C	34%	66%	
16126	T	0	34	0	49	C	C	T	41%	59%	0	16	0	25	C	C	T	39%	61%	0	33	0	64	C	C	T	34%	66%	
16162	A	28	0	53	0	G	A	G	35%	65%																			
16192	C										0	14	0	27	T	C	T	34%	66%	0	34	0	65	T	C	T	34%	66%	
16209	T	0	52	0	32	C	T	C	38%	62%																			
16218	C										0	16	0	25	T	C	T	39%	61%	0	36	0	63	T	C	T	36%	64%	

16270	C									0	16	0	25	T	C	T	39%	61%	0	31	0	66	T	C	T	32%	68%	
16291	C	0	55	0	30	T	T	C	35%	65%	0	26	0	16	T	T	C	38%	62%	0	65	0	34	T	T	C	34%	66%
16320	C										0	15	0	24	T	C	T	38%	62%	0	29	0	64	T	C	T	31%	69%
16362	T	0	18	0	48	C	C	T	27%	73%	0	9	0	25	C	C	T	26%	74%	1	13	0	64	C	C	T	17%	82%
16519	T	0	45	0	34	C	T	C	43%	57%																		

Table 2. Number of parental and recombinant sequences for each sample. A custom script was used to look at all of the variant positions for each sample and phase sequences into their mtDNA haplogroups. For this table, a “breakpoint” is defined as a switch from one parental haplogroup to the other that is sustained for at least two sites.

	Family A, I-1		Family B, I-1		Family C, III-1	
	# of reads	Percent %	# of reads	Percent %	# of reads	Percent %
Total	1712	100	2160	100	2719	100
Paternal	12	0.70093458	75	3.47222222	1130	41.5593968
Maternal	1234	72.0794393	908	42.037037	629	23.133505
0 Breakpoints	1246	72.7803738	983	45.5092593	1759	64.6929018
1 Breakpoint	141	8.23598131	406	18.7962963	488	17.9477749
2 Breakpoints	208	12.1495327	278	12.8703704	291	10.7024641
3 Breakpoints	63	3.67990654	199	9.21296296	126	4.63405664
4 Breakpoints	43	2.51168224	143	6.62037037	41	1.50790732
5 Breakpoints	9	0.52570093	81	3.75	13	0.47811695
6 Breakpoints	2	0.11682243	42	1.94444444	1	0.03677823
7 Breakpoints			17	0.78703704		
8 Breakpoints			7	0.32407407		
9 Breakpoints			4	0.18518519		