Recombination of Human Mitochondrial DNA

Yevgenya Kravtsberg,1 Marianne Schwartz,2 Timothy A. Brown,4 Konstantin Ebralidse,1 Wolfram S. Kunz,5 David A. Clayton,4 John Vissing,3 Konstantin Khrapko1*

Human mitochondrial DNA (mtDNA) is a 16.5-kb, circular genome essential for the maintenance of mitochondrial function and is present in multiple copies in most cell types. High sequence divergence and maternal inheritance make mtDNA useful in tracing human lineages. Whether recombination occurs between mitochondrial genomes is a long-standing question in mitochondrial biology, human evolution, and population studies (1). MtDNA recombination occurs in yeast, and recombinant mtDNA has been found in several animal species (2); however, the evidence for recombination between heterologous mtDNA in humans is controversial (1).

We searched for mtDNA recombinants in muscle tissue of an individual with paternal/maternally inherited mtDNA recombinants (3), where heterologous (paternal and maternal) mtDNAs are mixed and thus may have an opportunity to recombine. DNA was cleaved (4) by a paternal-specific restriction endonuclease at position 14,793 (fig. S1) to exclude a 10:1 excess of paternal mtDNA. It was then subjected to single-molecule polymerase chain reaction (PCR), which avoids in vitro recombination and has been successfully used to isolate nuclear DNA recombinants (5). In single-molecule PCR, each PCR product is a clone of identical molecules that originate from a single template. We recovered 450 PCR clones containing a maternal sequence at position 14,793 and screened them for paternal sequences at position 14,793, containing a maternal sequence at position 14,793 and screening them for paternal sequences at these two breakpoints (supporting online text).

Fig. 1. Structural map of recombinants. Each horizontal row represents a recombinant, with paternal and maternal segments in blue and red, respectively. Maternal to paternal nucleotide changes are shown above the map. Segments that join polymorphisms of different (paternal/maternal) descent, and thus contain sequence breakpoints, are white. Prominent stacks of white segments represent breakpoint hotspots, of which A, B, and C (but not the one at 15,218 to 16,189) are highly statistically significant (4). Nonmapped areas are gray. Vertical lines depict restriction sites used for recombinant screening (double black), the secondary origin of light-strand replication (lavender), and 3′ (green) and 5′ (orange) ends of the 7S DNA. Several recombinants contain more than two breakpoints (supporting online text).

References and Notes

4. Materials and methods are available as supporting material on Science Online.
9. We thank A. de Grey, M. Frank-Kamenetskii, B. Thilly, and anonymous reviewers for their input, and A. Pluzhnikov for help with statistical analysis. Supported in part by NIH grant nos. ES11343, AG19787, and AG18388 (K.K.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/304/5673/981/DC1
Materials and Methods
SOM Text
Fig. S1
References and Notes

2 February 2004; accepted 11 March 2004

1Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA. 2Department of Clinical Genetics. 3Department of Neurology and Copenhagen Muscle Research Center, National University Hospital Rigshospitalet, Copenhagen, Denmark. 4Howard Hughes Medical Institute, Chevy Chase, MD, USA. 5Department of Epileptology, University of Bonn Medical Center, Bonn, Germany.

*To whom correspondence should be addressed. E-mail: khrapko@hms.harvard.edu
Online Supporting Material.

Materials and Methods

**Single molecule PCR and mapping of recombinants.**
Total muscle DNA was digested with *BsrBI*, which cleaves at a paternal-specific restriction site at position 14,793 of mtDNA, and subjected to single molecule nested PCR as described elsewhere ([1](#)). In brief, DNA was aliquoted into microwell PCR plates at a concentration of about 0.3 amplifiable molecules per well (so that about one-third of the wells yielded PCR product) and subjected to a total of 50 cycles of nested PCR (27 cycles with outer pair and 23 with the inner pair of primers). To ascertain that the results were independent of the fragment size, two sets of nested primers were used that produced fragments of 7 kb and 9 kb (Fig. S1). In total, 100 9-kb and 350 7 kb PCR products were recovered and subjected to restriction analysis with *BsrBI* and *ApaLI* (paternal-specific site at position 73). Most of the products contained a pure maternal or (rarely) a pure paternal sequence (the latter originating from residual paternal templates uncut by *BsrBI*) at both polymorphic restriction sites. This proves that PCR products were clones each originating from a single template molecule, as expected in single-molecule PCR. In addition, *BsrBI*-negative and *ApaLI*-positive (i.e., carrying a *maternal* sequence at position 14,793 and a *paternal* sequence at position 73, respectively) PCR clones were identified. These clones were considered potential recombinants and were fully mapped using direct sequencing and/or restriction analysis. Every marker was sequenced and/or subject to restriction analysis at least two times. Several PCR products contained mixtures (mostly 50:50) of two sequences (a maternal and a recombinant or a maternal and a paternal). The frequency of mixed PCR reactions was consistent with the statistical expectation of the number of reaction wells with two or more templates. Recombinants were found at comparable frequency in DNA from each of two available biopsies taken from the right and the left *vastus lateralis* muscle, and using PCR fragments of both lengths (7 kb and 9 kb).

**Statistical analysis.**
To estimate the statistical significance of the observed recombination hotspots, we calculated, independently for each prospective hotspot, the probability that the hotspot has arisen by chance, assuming that breakpoints have an equal likelihood to arise at any base pair of the mitochondrial genome. This estimate is intended only as a rough conservative approximation of statistical significance. Note that all the recombinants isolated in this study were, by the criteria used, *maternal* at bp 14,793 and *paternal* at bp 73. Therefore in each recombinant, at least one breakpoint ought to be present within the 1,849 bp region from 14,793 to 73, and at least one - outside this region. Hence, the conditional probability of finding a breakpoint per base pair is different within and outside the 1,894 bp region. We thus divided the mapped part of the genome into two portions: one within and the other...
outside the region. The unmapped part of the genome was excluded from calculations. The average breakpoint densities per base pair per recombinant were calculated for the two portions of the genome by dividing the total number of breakpoints detected by the total number of nucleotides mapped within either portion.

For each of the 22 segments between neighboring paternal/maternal polymorphic sites we calculated $X_k$ (k=1,…,22, counting from left to right in Figure 1), the expected probability to find a breakpoint in the k-th segment per recombinant, by multiplying the appropriate breakpoint density by the length of k-th segment. Let $n_k$ be the observed number of breakpoints in k-th segment (e.g., for segment 5, also called hotspot A, $n_5=11$). For each segment, we estimated $P_k$, the probability that $n_k$ breakpoints occurred in the k-th segment by pure chance. $P_k$ was calculated as a binominal probability of $n_k$ or more successes to happen in a set of 33 (i.e., the number of recombinants) trials with the probability of success per trial equal to $X_k$. We considered a segment a highly significant hotspot if there were more breakpoints than expected ($n_k>33X_k$) and $P_k$ was less than 0.01. Only three segments (5, 20 and 21, denoted A, B, and C in Figure 1) proved to be highly statistically significant hotspots. Note that according to this criterion, segment 13 (15,218-16,189) is not a significant hotspot ($P_{13}=0.23$), although it does include a large number of breakpoints. This segment covers most of the region from 14,793 to 73, and thus is expected to contain many breakpoints (it actually contains merely 10% more breakpoints than expected under the equal likelihood assumption).

Supporting online text.

Proof of the in vivo origin of the recombinants.

While artificial recombinant molecules may be created by jumping PCR (2), single molecule PCR is not expected to be subject to this artifact (1). In this study, the following lines of evidence argue that the observed recombinants were generated in vivo. First, all but one recombinant fragment found in this study contain at least two breakpoints. Jumping PCR cannot explain this observation: two jumps in one molecule is an unlikely event given the low overall incidence of “jumps” in our experiments. In contrast, an in vivo recombinant must contain two breakpoints due to the circularity of mtDNA. Second, an association of breakpoint hotspots with functionally significant sites of the mitochondrial genome is unlikely to have been generated artificially, especially given that the hotspots are different in different classes of recombinants and do not coincide with the crucial BsrBI restriction site. Third, and most important, is the reconstruction experiment. Paternal and maternal DNA were mixed at a 10:1 ratio (as in the proband’s muscle) and subjected to BsrBI digestion and single molecule PCR. No recombinants were found upon screening of 150 PCR clones originating from this control DNA mixture, while screening of an equivalent number of clones from the proband’s DNA would have yielded more than 10 recombinant molecules. Thus, PCR artifacts do not represent a significant problem in our approach.

Recombinants with multiple breakpoints.

Several recombinant molecules contain more than two (up to six) breakpoints (Figure 1), while only two breakpoints are sufficient to satisfy the requirement imposed by the circularity of mtDNA molecule. One can only speculate about the origins of multiple breakpoints. In this context it is noteworthy that multiple breakpoints may result from recombinational repair of double stranded breaks (3). Alternatively, additional breakpoints might have arisen as a result of partial mismatch repair in paternal/maternal heteroduplexes, although mismatch repair has yet to be convincingly demonstrated in mitochondria. Repair of a single mismatch in a heteroduplex in most cases would create a marker surrounded by sequences of the opposite descent. Each such marker necessarily brings in what we interpret as a pair of "breakpoints", one from each side of the marker.

References:
