Ancient mtDNA sequences in the human nuclear genome: A potential source of errors in identifying pathogenic mutations

(Alzheimer's disease/pseudogenes/mutation/mitochondrial disease)

DOUGLAS C. WALLACE*, CAROL STUGARD, DEBORAH MURDOCK, THEODORE SCHURR, AND MICHAEL D. BROWN

Center for Molecular Medicine, Emory University, Atlanta, GA 30322

ABSTRACT Nuclear-localized mtDNA pseudogenes might explain a recent report describing a heteroplasmic mtDNA molecule containing five linked missense mutations dispersed over the contiguous mtDNA CO1 and CO2 genes in Alzheimer's disease (AD) patients. To test this hypothesis, we have used the PCR primers utilized in the original report to amplify CO1 and CO2 sequences from two independent ρ° (mtDNA-less) cell lines. CO1 and CO2 sequences amplified from both of the ρ° cells, demonstrating that these sequences are also present in the human nuclear DNA. The nuclear pseudogene CO1 and CO2 sequences were then tested for each of the five "AD" missense mutations by restriction endonuclease site variant assays. All five mutations were found in the nuclear CO1 and CO2 PCR products from ρ° cells, but none were found in the PCR products obtained from cells with normal mtDNA. Moreover, when the overlapping nuclear CO1 and CO2 PCR products were cloned and sequenced, all five missense mutations were found, as well as a linked synonymous mutation. Unlike the findings in the original report, an additional 32 base substitutions were found, including two in adjacent tRNAs and a two base pair deletion in the CO2 gene. Phylogenetic analysis of the nuclear CO1 and CO2 sequences revealed that they diverged from modern human mtDNAs early in hominid evolution about 770,000 years before present. These data would be consistent with the interpretation that the missense mutations proposed to cause AD may be the product of ancient mtDNA variants preserved as nuclear pseudogenes.

In the 9 years since the discovery of the first mtDNA missense mutation associated with an inherited human disease (1), there has been rapidly growing interest in the possibility that mtDNA mutations may be involved in a wide variety of degenerative diseases (2–4). As increasingly complex diseases have been investigated, the sensitivity of the methods for detecting low levels of heteroplasmic mutations has increased substantially (5, 6). However, the use of such techniques for identifying mtDNA variants using total cellular DNA has a potential risk. The human nuclear DNA harbors mtDNA sequences that have been transferred from the cytoplasm over the course of mammalian evolution, and, as a consequence, these sequences may be detected and misinterpreted as pathologically relevant.

The first human mtDNA sequences localized in the nucleus were found by screening a λ phage genomic library with a mtDNA 16S rRNA probe. The clones isolated in the screens contained both mtDNA sequences and flanking nuclear sequences (7). Subsequently, additional human nuclear sequences have been shown to be homologous to mtDNA 16S rRNA sequences (8–10); ND4 (URF4) and ND5 (URF5) sequences

(11); contiguous CO1, 12S rRNA, and ND4L/ND4 sequences (12); tRNA^{Gly} and ND3 sequences (13); and the 7S DNA of the control region (14). Human nuclear-localized mtDNA fragments have now been isolated both by cloning from human genomic libraries and by PCR amplification from cultured human cells cured of their cytoplasmic mtDNAs (ρ° cells) by growth in ethidium bromide (10, 13).

Nuclear copies of mtDNA sequences appear to be a common phenomenon, having been documented in sea urchin (15), birds (16), rodents (17, 18), and nonhuman primates (19). This transfer over evolutionary time undoubtedly accounts for the current nuclear location of most of the genes that encode mitochondrial-specific polypeptides, including those required for oxidative phosphorylation (OXPHOS) and mitochondrial biogenesis (20). This genetic transfer is a continuing process, as shown by the recent integration of mtDNA CO3 gene sequences into one of the two nuclear c-myc gene alleles in HeLa cells. This modified gene is transcribed into a chimeric RNA with the CO3 sequences in the opposite orientation to the c-myc gene sequence (21).

The presence of mtDNA sequences in animal cell nuclei has proven to be an important source of contamination for studies on ancient DNA. In one study, consensus sequence primers were used to PCR amplify a 174-nucleotide pair (np) mtDNA cytochrome b fragment from 80 million-year-old dinosaur bone fragments. A cytochrome b sequence was detected in 9 out of 494 reactions (1.8%). Each successful reaction was sequenced directly, with seven falling into one consensus group and two into another, both of which differed from bird, reptile, and mammalian sequences (22). However, subsequent phylogenic analysis revealed that these cytochrome b sequences were more closely related to mammals, and particularly to primates, than to birds or amphibians (19, 23–25). Moreover, a similar fragment was amplified from the mtDNA-free nuclear DNA obtained from differentially lysed human spermatozoa, and two of the spermatozoan nuclear cytochrome b sequences clustered with the sequences derived from the ancient bone, all of which are more homologous to human mtDNA sequences than to bird (25). Hence, it appeared that human nuclear DNA contamination was the source of the DNA sequences derived from the dinosaur bones.

More recently, it has been reported that Alzheimer's disease (AD) patients are heteroplasmic for several missense mutations in the mtDNA genes for cytochrome *c* oxidase (COX) subunits CO1 and CO2 (5). The CO1 missense mutations observed were a G-to-A transition at np 6366 (Val to Ile) and an A-to-G transition at np 7146 (Thr to Ala). The CO2 missense mutations were a C-to-T transition at np 7650 (Thr to Ile), a C-to-T

transition at np 7868 (Leu to Phe), and an A-to-G transition at np 8021 (Ile to Val). The CO1 gene was also reported to harbor a silent substitution, C to T at np 6483 (5). Interestingly, these same nucleotide differences relative to the reference human sequence (26) are found in the mtDNA sequences of both chimpanzee (*Pan troglodytes*) and gorilla (*Gorilla gorilla*) (5).

These variants could only be identified from DNA extracted from frozen platelet-enriched pellets when prepared in a specific way. The samples were thawed, centrifuged, and washed in Dulbecco's PBS; the pellets were suspended in water and incubated in a boiling water bath for 10 min; and the solution was extracted by SDS-proteinase K digestion and phenol-chloroform extraction. The lysis by boiling was reported to be a critical step in detecting these mutant sequences. The CO1 and CO2 genes were PCR amplified from these blood cell extracts by using pairs of flanking primers, cloned into the pCRII vector with the TA-cloning kit (Invitrogen), and sequenced. Ten clones were sequenced for each subunit. Four of the CO1 clones had the above-mentioned base substitutions at np 6366, 6483, and 7146. One clone was reported to have the missense substitutions at np 6366 and 7146 but not the synonymous mutation at np 6483. The remaining five CO1 clones had the nucleotides reported in the reference human sequence (26). Similarly, three of the CO2 clones had all three missense mutations at np 7650, 7868, and 8021, whereas the remaining seven clones had the reference human sequence. For rapid screening of a large number of samples, Davis and coworkers (5) used a competitive primer extension assay to analyze these base changes in DNA extracts of blood samples from AD patients and controls. This assay "revealed that the six mutations were apparent at low levels in most individuals but that the frequency of these mutant alleles are elevated in most AD cases" (5). Based on these data, the authors suggested that the observed point mutations in the CO1 and CO2 genes cause the cytochrome oxidase defect in AD (5).

These base substitutions have several novel features relative to the numerous and well studied pathogenic mtDNA mutations that result in human disease (2, 3). First, the six heteroplasmic mtDNA variants co-occur on the same molecule, even though they extend over a region of 1655 np. By contrast, virtually all other heteroplasmic pathogenic mutations studied to date involve only a single altered nucleotide (2, 3). Second, the variant sequences could be detected only in boiled samples, whereas previously studied homoplasmic and heteroplasmic mtDNA variants are commonly and readily detected by using direct SDS-proteinase K digestion with organic extractions or other standard DNA extraction techniques. Third, the variant sequences were found in virtually all human samples, whereas known pathogenic mutations are generally confined to patients with known mitochondrial diseases (3). Finally, in a footnote to their paper, the authors mentioned that a comparison of the AD CO1 and CO2 variants revealed similarities with the great apes (5).

The prevalence and diversity of mtDNA sequences in the human nucleus, the apparent link of the AD DNA mutations with those present in the chimpanzee and gorilla, and the existence of nuclear-mtDNA sequence contamination in ancient samples suggested to us that the putative, heteroplasmic AD genotypes might represent ancient nuclear pseudogenes rather than new mtDNA mutations. To test this possibility, we attempted to PCR amplify these "AD" CO1 and CO2 sequences from two different human ρ° cell lines that lack cytoplasmic mtDNA, using the primers reported by Davis and coworkers (5). Amplified sequences were obtained, cloned, and sequenced, and these nuclear-derived CO1 and CO2 sequences were found to contain the same missense mutations as did the putative AD mtDNAs. In addition, the sequence of our nuclear mtDNA clones revealed multiple additional silent substitutions. A phylogenetic comparison of our nuclear CO1 and CO2 sequences suggested that they were transferred to the nucleus early in hominid evolution.

MATERIALS AND METHODS

Cells and Cell Cultures. The human lymphoblastoid cell line WAL2A and the osteosarcoma cell line 143B-TK$^-$, and their mtDNA-free (ρ°) derivatives, have been described (27). Standard growth conditions were used (27).

DNA Analyses. DNA was extracted from 5×10^6 cells of each cell line by using the Puregene DNA extraction kit (Gentra Systems, Inc., Minneapolis). To confirm the presence of cytoplasmic mtDNAs in the parental cell lines and the absence of mtDNA in the ρ° cells, each DNA extract was tested for mtDNA amplification by using mtDNA primers that are known to react with mtDNA but not with nuclear DNA. These mtDNA-specific primers encompassed np 14430–15461 (14430FOR/15461REV), np 15238–15865 (15238FOR/15865REV), np 3007–3370, and np 9151–10047.

The CO1 and CO2 mtDNA and nuclear DNA sequences were amplified by using the primers described by Davis and colleagues (5). For amplification of the CO1 gene, the forward primer encompassed np 5803–5825 (CO1FOR) and reverse primer np 7570–7548 (CO1REV). For amplification of the CO2 gene, the forward primer encompassed np 7483–7503 (CO2FOR) and reverse primer np 8383–8361 (C02REV). Finally, for the CO3 gene, the forward primer encompassed np 9106–9130 (CO3FOR) and the reverse primer np 10110–10088 (CO3REV). For amplification of the combined CO1 and CO2 sequences, primer pair CO1FOR and CO2REV was used. All PCRs were performed with 0.3 μM of each primer, Perkin–Elmer AmpliTaq DNA polymerase and buffer, 1.5 mM MgCl$_2$, and 200 μM of each dNTP (Pharmacia). All reactions were performed by using hot-start PCR, with an annealing temperature of 55°C and 34 cycles of amplification.

CO1 Variants. The np 6366 G-to-A transition was detected by loss of a *Bsm*AI site at np 6371 in the CO1FOR/CO1REV PCR product. Digestion of the PCR product from normal mtDNA yields 996-, 569-, and 203-bp fragments, whereas the AD variant sequence produces 996- and 772-bp fragments. The np 6483 C-to-T transition was synonymous and, thus, not tested. The np 7146 A-to-G transition did not change a pre-existing restriction site. Therefore, the mutation was detected with a mismatch primer prepared such that the 3' end was adjacent to the variant nucleotide and the primer contained a mismatch that in conjunction with the variant nucleotide created a *Nru*I site. The sequence of the forward primer began at np 6968 (6968FOR) and was 5'-CATTGTATTAGCAAACTCATCAC-3', and the reverse primer began at np 7170 (7170REV) and was 5'-GATTTACGC-CGATGAATATGATcG-3', in which the mismatched base is in lowercase type. The resulting 202-bp fragment remains uncut when amplified from the normal human mtDNA sequence but is cut into 178- and 24-np fragments when amplified from AD sequences.

CO2 Variants. The np 7650 T-to-C variant was detected by digestion of the CO2FOR/CO2REV PCR product with *Hph*I. The np 7650 polymorphism results in the loss of the *Hph*I site at np 7639. The variant sequence also has an additional base change, which removes a second *Hph*I site at np 8128. *Hph*I digestion of normal sequence DNA results in 489-, 255-, and 157-bp fragments, whereas the AD nuclear variant sequence gives a 901-bp fragment.

The np 7868 C-to-T transition was also tested by using a mismatch primer that creates an *Mse*I site if the PCR product contains the AD variant. The forward primer began at np 7680 (7680FOR) and was 5'-TCCTTATCTGCTTCCTAGTCC-3', and the reverse primer began at np 7893 (7893REV) and was 5'-TGGTGGCCAATTGATTTGATGGTtA'-3, with the lowercase letter representing the mismatched base. For the normal human sequence, the 214-np PCR product remains uncut, whereas the AD variant is cut into 189- and 25-bp fragments.

Finally, the np 8021 A-to-G transition creates a *Hpa*II site. The region is amplified by a forward primer starting at np 7900

(7900FOR, 5′-CTGAACCTACGAGTACACCG-3′) and a reverse primer at np 8225 (8225 REV, 5′-TTAATTCTAGGAC-GATGGGC-3′). When digested with *Hpa*II, the normal human sequence produces 213-, 75-, and 38-bp fragments. The AD np 8021 polymorphism creates a new *Hpa*II site dividing the 213-bp fragment into 119- and 94-bp fragments. The nuclear fragments from the ρ° WAL2A and 143B-TK⁻ cell lines also contain an additional nucleotide change at np 8152 resulting in a site loss that fuses the 75- and 38-bp fragments into a 113-bp fragment.

All restriction endonucleases were obtained from New England Biolabs, and all restriction fragments were resolved on 1.0% Seakem plus 2.5% NuSieve GTG agarose (FMC BioProducts) in 1× Tris-borate buffer.

Cloning of CO1 and CO2 Fragments. The 1768-bp CO1 PCR product and the 901-bp CO2 PCR product were amplified from WAL2A-EB2 ρ° nuclear DNA by using the above primers and 32 cycles of amplification. One microliter of the CO1 or CO2 PCRs was ligated into plasmid pCR 2.1 by using the Invitrogen TA-cloning kit. White LacZ-deficient colonies were screened for inserts with the M13–43/T7 primers for amplification. Five plasmids with appropriately sized inserts were sequenced for both the CO1 and CO2 clones by using various internal mtDNA primers. Both strands of each plasmid were sequenced by using cycle sequencing with Prism dideoxy terminator dyes, and the sequences were analyzed with an ABI 373 automated sequencer.

The nuclear WAL2A-EB2 ρ° sequence (GeneBank accession no. AF035429) was compared with the standard human (V00662 or J01415), chimpanzee (D38113), gorilla (D38114), orangutan (D38115), and gibbon (X99256) sequences by using PHYL1P3.57c (28).

The phylogenetic relationship of the sequence data was analyzed by using the Neighbor-Joining (NJ) method (29). Genetic distances were first estimated from the sequence data by using three different models in DNADIST (28), including the Kimura two-parameter (30), Jukes and Cantor (31), and DNAML (32) methods, and a transition-to-transversion ratio of 30:1. In addition, distances were estimated with the parsimony method available in DNAPARS (28). The resulting distance matrices were then analyzed with the NJ method to generate unrooted phylogenies. To test the reliability of the branching order of the NJ trees, the sequence data were bootstrapped over 100 replicates in SEQBOOT (28), and the bootstrapped data sets used to generate genetic distances in DNADIST. These sets of distances were again analyzed with the NJ method, and consensus trees were generated from the resulting NJ trees with CONSENSE (28), in which the percent support for the branches represents the bootstrap value for the nodes of the trees.

RESULTS

Nuclear Location of CO1 and CO2 Sequences. To determine whether CO1- and CO2-like sequences were located in the human cell nucleus, we used primer pairs homologous to different regions of the mtDNA to amplify the intervening sequences from total cellular DNA extracted from two pairs of cell lines—143B-TK⁻ containing mtDNA (Fig. 1, lanes 3, 8, 13, and 18) and 143B-87, which lacks mtDNA (ρ°) (Fig. 1, lanes 4, 9, 14, and 19)—plus WAL2A containing mtDNA (Fig. 1, lanes 5, 10, 15, and 20) and WAL2A-EB2 without mtDNA (ρ°) (Fig. 1, lanes 6, 11, 16, and 21). When primers in the ND6-cytochrome b region were used for amplification (14430FOR/15461REV and 15238FOR/15866REV), mtDNA sequences were amplified from only the parental cell lines containing normal mtDNA (143B-TK⁻, lanes 3 and 8, and WAL2A, lanes 5 and 10). No fragments were amplified from the corresponding ρ° cells (143B-87, lanes 4 and 9, and WAL2A-EB2, lanes 6 and 11), confirming the total absence of normal mtDNA molecules in these cell lines (27). Additional mtDNA primer pairs also failed to amplify mtDNA-like sequences from WAL2A-EB2 ρ° DNA. These in-
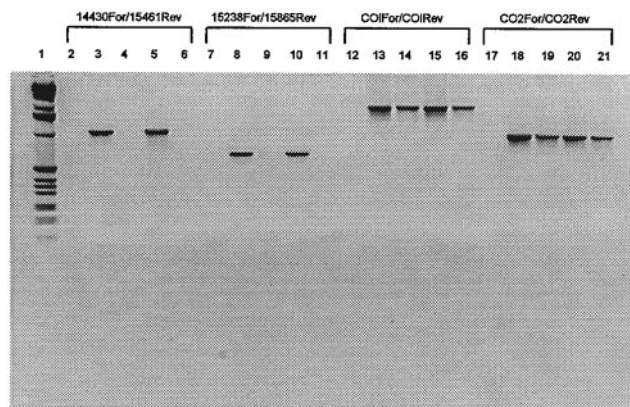

Fig. 1. Amplification of PCR products from total cellular DNAs extracted from two different human cell lines and their mtDNA-deficient ρ° counterparts using various pairs of primers homologous to the normal human mtDNA sequence (26). The cell lines are 143B-TK⁻ (lanes 3, 8, 13, and 18), 143B-87 (ρ°) (lanes 4, 9, 14, and 19), WAL2A (lanes 5, 10, 15, and 20), and WAL2A-EB2 (ρ°) (lanes 6, 11, 16, and 21). The primer pairs used are 14430FOR/15461REV (lanes 2–6) and 15238FOR/15865REV (lanes 7–11), both encompassing positions of cytochrome b; 5803COIFOR/7570 CO1REV encompassing CO1 (lanes 12–16) and 7483CO2FOR/8383CO2REV encompassing CO2 (lanes 17–21). Negative controls involving amplification without template are shown in lanes 2, 7, 12, and 17. A size standard is shown in lane 1.

cluded primers encompassing part of the 16S rRNA gene, the tRNA$^{Leu(UUR)}$ gene, and part of the ND1 gene (np 3007–3370) as well as primers encompassing the CO3 gene (np 9106–10110). However, a primer pair encompassing the 3′ end of ATPase 6 and most of CO3 (np 8829–9859) did amplify from the ρ° DNA (data not shown).

By contrast, when the CO1 (mtDNA np 5803–7570) and the CO2 (mtDNA np 7483–8383) primers of Davis and coworkers (5) were used, appropriate-length fragments were amplified from all four cell lines, whether or not they contained cytoplasmic mtDNA. However, the amount of product was somewhat greater for the cell lines with mtDNA (143B-TK⁻, lanes 13 and 18, and WAL2A, lanes 15 and 20) than for the cell lines without mtDNA (143B-87, lanes 14 and 19, and WAL2A-EB2, lanes 16 and 21). Thus, the ρ° 143B-87 and WAL2A-EB2 cells lack any detectable mtDNA but still harbor sequences homologous to CO1 and CO2. The only reasonable conclusion is that these and presumably all human cells harbor nuclear sequences homologous to CO1 and CO2.

To determine if the CO1 and CO2 sequences from the nucleus were continuous, as they are in the mtDNA, WAL2A-EB2 ρ° cell DNA was used as template to amplify the entire CO1 and CO2 region. When the encompassing primers CO1FOR and CO2REV were used, the expected 2581-bp fragment was obtained. Hence, the nucleus contains a continuous section of the mtDNA encompassing both genes (data not shown; see the sequence below).

The Nuclear CO1 and CO2 Genes Contain the AD Missense Mutations. To confirm that the CO1 and CO2 PCR products from the ρ° cells contained the missense variants described by Davis and coworkers (5), we devised specific restriction endonuclease tests for each of the five mutations. The PCR products from both the 143B-TK⁻ cell line and its mtDNA-deficient (ρ°) derivative, 143B-87, were then tested for each mutation.

In every case, the CO1 or CO2 PCR product from the 143B-TK⁻ cells containing normal mtDNA had the restriction sites expected for the human reference mtDNA sequence (26), whereas the PCR products from the 143B-87 ρ° cell line had the restriction site patterns consistent with the reported AD sequences (Fig. 2). Specifically, the np 6366 G-to-A variant was detected by the loss of a *Bsm*AI site. Accordingly, the CO1 PCR

Neurobiology: Wallace *et al.*

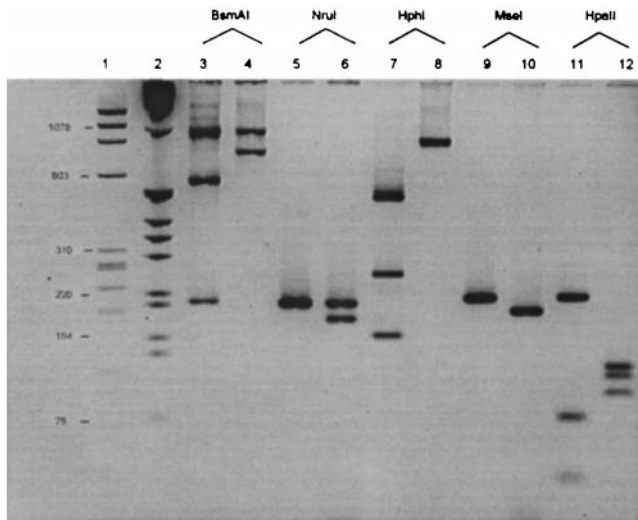*Proc. Natl. Acad. Sci. USA* 94 (1997)   14903

FIG. 2. Detection of the five missense mutations in CO1 and CO2 genes amplified from DNA of 143B-TK⁻ and 143B-87 ρ° cells by restriction endonuclease digestion. PCR products obtained from 143B-TK⁻ DNA are shown in lanes 3 and 5 for CO1 and lanes 7, 9, and 11 for CO2. PCR products obtained from 143B-87 ρ° DNA are shown in lanes 4 and 6 for CO1 and lanes 8, 10, and 12 for CO2. Detection of the np 6366 G-to-A mutation in CO1 by *Bsm*AI digestion is presented in lanes 3 and 4. Detection of the np 7146 A-to-G mutation in CO1 by *Nru*I digestion is shown in lanes 5 and 6. The double band in lane 6 is most likely the result of partial digestion. Detection of the np 7650 C-to-T mutation by *Hph*I digestion is shown in lanes 7 and 8. Detection of the np 7868 C-to-T mutation by *Mse*I digestion is shown in lanes 9 and 10. Detection of the np 8021 A-to-G mutations by *Hpa*II digestion is shown in lanes 11 and 12. Lanes 1 and 2 are size standards: lane 1 is a ∅X174 *Hae*III digest and lane 2 is a 1-kb DNA ladder.

product from 143B-TK⁻ cells gave the expected 996-, 569-, and 203-np fragments (Fig. 2, lane 3), whereas the PCR product from the ρ° cells lacked the site associated with the 6366-base change and gave 996- and 772-np fragments (Fig. 2, lane 4). The CO1 np 7146 A-to-G variant was detected by using a mismatch primer, which in the presence of the AD variant creates a *Nru*I site. The PCR product from the 143B-TK⁻ cells lacked this site, giving a 202-np product (Fig. 2, lane 5), whereas the PCR product from the ρ° cells had the site and gave 178- and 24-bp products (Fig. 2, lane 6). The CO2 variant at np 7650 (C to T) results in the loss of the *Hph*I site at np 7639. The nuclear CO2 PCR product also contains a second base change at np 8140 that eliminates a second *Hph*I site at np 8128 (Table 1). Consistent with the normal human sequence, the 143B-TK⁻ PCR product gave 489-, 255-, and 157-np fragments (Fig. 2, lane 7), whereas the PCR product from the ρ° cell gave a single fragment of 901 np (Fig. 2, lane 8). For the CO2 np 7868 C-to-T variant, a mismatch primer that creates a *Mse*I site was synthesized. Consistent with the normal mtDNA sequence, the CO2 PCR product from the 143B-TK⁻ cells gave an uncut 214-np product (Fig. 2, lane 9), whereas the PCR product from the ρ° cell had the *Mse*I site and gave the expected 189- and 25-np products (Fig. 2, lane 10). Finally, the CO2 np 8021 A-to-G variant creates a *Hpa*II site. The CO2 PCR product from the 143B-TK⁻ cell line gave the expected normal mtDNA restriction pattern of 213-, 75-, and 38-np fragments (Fig. 2, lane 11), whereas the PCR product from the ρ° cell line had the 213-np fragment split into 119- and 94-np fragments, consistent with the presence of the AD variant. The ρ° PCR product also lacks an additional *Hpa*II restriction site that fuses the 75- and 38-np fragments from the normal sequence into a 113-np fragment. Hence, the *Hpa*II digest of the ρ° cell PCR product gives the expected 119-, 113-, and 94-np fragments (Fig. 2, lane 12).

This analysis thus makes a clear distinction. The CO1 and CO2 PCR products from cells containing normal human mtDNAs do not contain the mutant "AD" nucleotides at the variant positions

Table 1. Nuclear and cytoplasmic CO1 and CO2 sequences

np position	Human mtDNA	ρ° cell DNA	Chimp mtDNA	Gorilla mtDNA
tRNA^Tyr				
5840	C	T	T	C
COI				
6023	G	A	A	A
6221	T	C	C	C
6242	C	T	T	C
6266	A	C	C	C
6299	A	G	A	A
6366	**G**	**A**	**A**	**A**
6383	G	A	A	G
6410	C	T	T	C
6452	C	T	C	T
6483	**C**	**T**	**T**	**T**
6512	T	C	T	T
6542	C	T	T	T
6569	C	A	A	A
6641	T	C	T	A
6935	C	T	T	C
6938	C	T	C	T
7146	**A**	**G**	**G**	**G**
7232	C	T	C	C
7256	C	T	T	T
7316	C	A	G	A
tRNA^Asp				
7521	G	A	A	A
COII				
7616	G	A	G	G
7650	**C**	**T**	**T**	**T**
7705	T	C	C	C
7810	C	T	T	G
7868	**C**	**T**	**T**	**T**
7891	C	T	T	C
7912	G	A	A	A
8021	**A**	**G**	**G**	**G**
8065	G	A	A	A
8140	C	T	C	C
8152	G	A	A	A
8167	T	C	C	C
8197	C	Δ	C	C
8198	A	Δ	A	A
8203	C	T	T	T
8254	C	T	C	C

but instead harbor the "normal" nucleotide found in the human reference sequence (26). By contrast, the CO1 and CO2 PCR products from ρ° human cells, which lack all normal mtDNA, contain only the CO1 and CO2 variants reported for the "AD" mtDNAs. Hence, the nuclear CO1 and CO2 sequences of human cells harbor all five of the missense mutations attributed to the AD-specific mtDNA by Davis and coworkers (5), whereas the normal mtDNA CO1 and CO2 sequences do not.

The Sequences of the Nuclear-Encoded CO1 and CO2 Genes. To further characterize the nuclear-encoded CO1 and CO2 sequences, we amplified the CO1 (np 5803–7570) and CO2 (np 7483–8383) sequences from WAL2A-EB2 ρ° cells and cloned and sequenced five clones from each PCR. Because these sequences are contiguous in the nuclear genome and the PCR products overlap, alignment of the sequences gave a continuous sequence that extends from the beginning of the CO1 to the end of the CO2 gene. Of the five CO1 cloned sequences, four were identical and one was different. The one deviant sequence was more highly divergent from the normal mtDNA than the other four clones. For the CO2 PCR clones, all five were essentially identical and overlapped with the four homologous CO1 sequences. Alignment of the four CO1 and five CO2 clones provided a continuous sequence extending from np 5804 to np 8384.

Comparison of the consensus nuclear CO1 and CO2 sequences with the reference human sequence (26) (Table 1) revealed that the nuclear sequence harbored all of the sequence variants attributed to the AD mtDNA by Davis and coworkers (5). These included the missense mutations at np 6366 (G to A), np 7146 (A to G), np 7650 (C to T), np 7868 (C to T), and np 8021 (A to G). The nuclear sequence also contained the synonymous substitution at np 6483 (C to T). Surprisingly, the nuclear CO1 and CO2 sequences also contained an additional 32 base differences relative to the reference human sequence. Two of these were in tRNAs, one in tRNA^{Tyr} at np 5840 and the other in tRNA^{Asp} at np 7521; 17 were base substitutions in CO1; 11 were base substitutions in CO2; and 2 were deletions in CO2 at np 8197 and 8198 (Table 1). The two deleted bases would create a frameshift near the C-terminal end of the CO2 protein, which confirms that this sequence is a pseudogene (Table 1).

Alignment of the nuclear-encoded CO1 and CO2 sequences with the mtDNA sequences of chimpanzee, gorilla, and human revealed that the nuclear sequence has features in common with all three of the higher primates. In fact, all six of the base substitutions attributed to the AD mtDNA, including the five missense mutations, are found in the nuclear-encoded CO1 and CO2 sequences and are identical to both the chimpanzee and the gorilla mtDNA sequences but different from the human mtDNA sequence. Because the cytoplasmic mtDNA evolves much more rapidly than do nuclear mtDNA pseudogenes, the primate nuclear mtDNA pseudogenes can be assumed to be essentially evolutionarily static (16). Under that assumption, we calculated the maximum likelihood sequence divergence between the human nuclear-encoded CO1 and CO2 sequence and the normal human, chimpanzee, and gorilla mtDNA sequences. This comparison revealed a 1.4% sequence divergence between the human nuclear sequence and human mtDNA sequences, an 8.4% divergence between the human nuclear sequence and the chimpanzee mtDNA sequence, and an 11.7% divergence between the human nuclear sequence and the gorilla mtDNA sequence. Because the sequence divergence between the human and chimpanzee mtDNA CO1 and CO2 sequences was 9.1% and because these species are thought to have diverged about 5 million years before present, this divergence would suggest that the nuclear CO1 and CO2 DNA segment was transferred from the cytoplasm to the

nucleus about 770,000 years before present. This relationship is confirmed by phylogenetic analysis that shows that the nuclear CO1 and CO2 sequences branch from the other primates closer to human than to chimpanzee or gorilla (Fig. 3).

The fifth CO1 sequence is significantly more divergent than the contiguous CO1 and CO2 sequences. Of the 751 np of sequence that could be obtained by using human mtDNA primers from the reference mtDNA sequence (26), the fifth CO1 sequence proved to be more similar to that of gibbon mtDNA than to human.

DISCUSSION

There is a rapidly growing interest in identifying new mtDNA mutations that are associated with both rare and common degenerative diseases (2–4). For late-onset degenerative diseases such as AD, it might be argued that contributory mtDNA mutations could be both mild and common (33). Indeed, one mtDNA mutation in the tRNA^{Gln} gene at np 4336, which has been proposed to contribute to late-onset AD, is a relatively mild sequence variant associated with the most common European mtDNA lineage, haplogroup H. This variant has been reported to be present in approximately 5% of AD patients, but only about 0.7% of random normal controls (34, 35). However, as increasingly sensitive analytical techniques are applied to detecting hypothetical low frequency and heteroplasmic mtDNA mutations, there is a rising risk that the sequences detected will be low copy number mtDNA sequences that are integrated into the nuclear DNA. The potential for such a scenario is underscored by the estimation that there are approximately 1000 mtDNA-derived sequences integrated into the human nuclear genome (12). Because most of these mtDNA sequences have been integrated into the nuclear genome before the origin of anatomically modern humans, these sequences will contain phyogenetically ancient nucleotide sequence variants that may be inadvertently interpreted as pathogenic variants.

Based on these considerations, it is most likely that the five linked CO1 and CO2 missense mutations attributed to AD mtDNAs by Davis and coworkers (5) are actually from nuclear-encoded sequences. There are multiple arguments to support this contention. First, the method used to extract the DNA from the platelet-enriched pellet was unorthodox and may have enriched the samples for nuclear DNA sequences. Freezing–thawing and washing the platelets in PBS may have lysed the platelets, resulting in the loss of some mitochondria and mtDNAs. Subsequent boiling of the suspension of remaining material would lyse nuclei and release sheared nuclear DNA. The boiling may also have released novel nuclear DNA molecules, such as small polydispersed circular nuclear DNAs (36, 37). Apparent heteroplasmy might then result from a mixture of both the cytoplasmic mtDNA and enriched nuclear-localized mtDNA sequences.

Second, the probability of the same five linked mtDNA missense mutations co-occurring repeatedly over two genes and 1655 np of sequence is unprecedented. In the past, relatively severe pathogenic mtDNA mutations have arisen spontaneously, have been heteroplasmic, and occurred on a variety of different background mtDNA haplotypes. Alternatively, milder pathogenic mutations may have occurred in the more distant past on a particular mtDNA haplotype and are typically homoplasmic (38).

Third, contiguous CO1 and CO2 genes containing the five putative AD missense mutations as well as the single synonymous mutation have been isolated from ρ° cell DNAs containing only nuclear DNA by PCR amplification with the same primers reported by Davis and coworkers (5). By contrast, these missense mutations were not detected in the PCR products when using total cellular DNA extracted from cells containing a normal complement of mitochondria and mtDNAs. Moreover, previous analyses of hundreds of mtDNAs from throughout the world have yet to detect these variants (38).
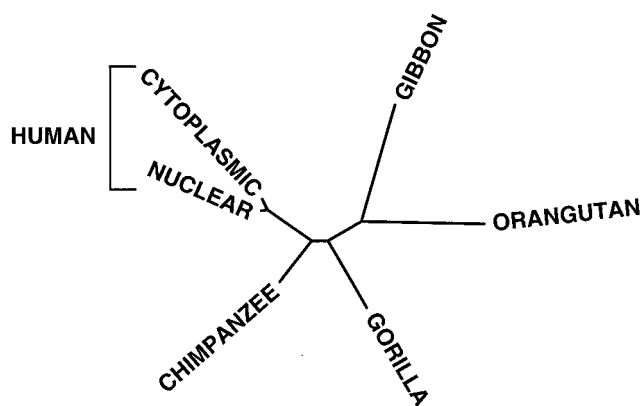

FIG. 3.   Unrooted NJ tree relating the human nuclear CO1 and CO2 sequences to the homologous sequences from the human, chimpanzee, gorilla, gibbon, and orangutan mtDNAs. This tree is based on genetic distances calculated using the maximum likelihood model (DNAML) (28). Boostrap analysis from 100 independent trees generated the diagrammed result in 100% of the comparisons between the human nuclear and cytoplasmic sequences and between the gibbon and orangutan sequences, in 99% of the comparisons between the two human sequences and the chimpanzee sequence, and in 82% of the comparisons between the chimpanzee and gorilla sequences and between the gorilla and the collective gibbon and orangutan sequences. Phylogenies with identical branching orders were obtained by using genetic distances calculated by the parsimony, Jukes-Cantor, and Kimura two-parameter methods.

Fourth, a complete CO3 gene could not be amplified from ρ° cell DNA by using the primers of Davis and coworkers (5), although CO3 was readily amplified from DNA extracted from cells containing normal mtDNA. This is relevant because Davis and colleagues (5) reported that they did not find any sequence variants in CO3, although multiple changes were found in CO1 and CO2.

Fifth, interspecies comparisons revealed that the putative AD CO1 and CO2 missense mutations and the associated synonymous mutation were identical to those found in the same positions in normal chimpanzee and gorilla mtDNAs. This result suggests an ancient rather than recent origin for these variants, a conclusion which was strongly supported by phylogenetic analysis of the human nuclear, human cytoplasmic, chimpanzee, and gorilla CO1 and CO2 sequences.

Sixth, although the Davis and coworkers (5) paper reports COX deficiency in cybrids prepared by fusion of platelets from AD patients as compared with platelets from controls, they did not provide data or comment on whether or not the cybrids contained mtDNAs with the CO1 or CO2 "AD" mutations. By contrast, previous assignments of OXPHOS enzyme defects to the mtDNA by cybrid transfer have routinely shown that the transfer of the biochemical defect is always linked to the transfer of mutant mtDNA (39, 40).

There is one fact that may argue that the nuclear CO1 and CO2 genes reported in the current paper are not the same as the "AD" mutations reported by the Davis and colleagues (5). In our nuclear DNA fragments, we found that many additional silent base substitutions were linked to the five missense mutations reported by Davis and coworkers (5). Interestingly, the one synonymous substitution reported by Davis and colleagues (5) was among those found in our nuclear CO1 and CO2 pseudogenes. The multiple additional synonymous variants found in our nuclear CO1 and CO2 sequence further support its intermediate position between human and chimpanzee and gorilla. Hence, the absence of these additional synonymous mutations in the Davis and coworkers (5) report indicates either that these authors are observing a different and novel genetic phenomenon that we have not adequately replicated or that they overlooked the additional sequence variants observed in our sequence analysis.

Regardless of the origin of the putative AD missense mutation mtDNAs, the nuclear CO1 and CO2 sequences reported in this study are interesting in their own respect. They were transferred from the mtDNA to the nucleus long after the hominid lineage separated from the chimpanzee and gorilla lineages. Because the time of insertion of the sequence into the nucleus is estimated to be about 770,000 years before present, the transfer of these sequences might have occurred in archaic *Homo* (41).

In conclusion, as increasingly more sensitive methods are used to detect low frequency, heteroplasmic mtDNA mutations in clinical materials, increasing caution must be exercised to avoid the inadvertent detection of mtDNA pseudogenes integrated into the nuclear genome. Such sequences might be identified by application of phylogenic comparisons to detect interspecific homologies and by testing for the potential amplification of the mutant mtDNA sequences using DNA extracted from ρ° cells.

1. Wallace, D. C., Singh, G., Lott, M. T., Hodge, J. A., Schurr, T. G., Lezza, A. M. S., Elsas, L. J., II, & Nikoskelainen, E. K. (1988) *Science* **242,** 1427–1430.
2. Wallace, D. C., Lott, M. T. & Brown, M. D. (1996) in *Human Gene Mapping 1995*, ed. Cuticchia, J. (Johns Hopkins Univ. Press, Baltimore), pp. 1280–1331.
3. Wallace, D. C., Brown, M. D. & Lott, M. T. (1996) in *Emery & Rimoin's Principle & Practice of Medical Genetics*, eds. Rimoin, D. L., Connor, J. M. & Pyeritz, R. E. (Churchill Livingstone, New York), 3rd Ed., pp. 277–332.
4. Wallace, D. C. (1997) in *The Molecular & Genetic Basis of Neurological Disease*, eds. Rosenberg, R. N., Prusiner, S. B., DiMauro, S. & Barchi, R. L. (Butterworth-Heinemann, Boston), pp. 237–269.
5. Davis, R. E., Miller, S., Herrnstadt, C., Ghosh, S. S., Fahy, E., Shinobu, L. A., Galasko, D., Thal, L. J., Beal, M. F., Howell, N. & Parker, W. D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 4526–4531.
6. Fahy, E., Nazarbarghi, R., Zomorrodi, M., Herrnstadt, C., Parker, W. D., Davis, R. E. & Ghosh, S. S. (1997) *Nucleic Acids Res.* **15,** 3102–3109.
7. Tsuzuki, T., Nomiyama, H., Setoyama, C., Maeda, S. & Shimada, K. (1983) *Gene* **25,** 223–229.
8. Nomiyama, H., Fukuda, M., Wakasugi, S., Tsuzuki, T., Shimada, K. & Miyata, T. (1985) *Nucleic Acids Res.* **13,** 1649–1658.
9. Nomiyama, H., Tsuzuki, T., Wakasugi, S., Fududa, M. & Shimada, K. (1984) *Nucleic Acids Res.* **12,** 5225–5234.
10. Hu, G. & Thilly, W. G. (1994) *Gene* **147,** 197–204.
11. Fukuda, M., Wakasugi, S., Tsuzuki, T., Nomiyama, H. & Shimada, K. (1985) *J. Mol. Biol.* **186,** 257–266.
12. Kamimura, N., Ishii, S., Ma, L. D. & Shay, J. W. (1989) *J. Mol. Biol.* **210,** 703–707.
13. Hu, G. & Thilly, W. G. (1995) *Curr. Genet.* **28,** 410–414.
14. Zischler, H., Gelsert, H., von Haeseler, A. & Paabo, S. (1995) *Nature (London)* **378,** 489–492.
15. Jacobs, H. T., Posakony, J. W., Grula, J. W., Roberts, J. W., Xin, J. H., Britten, R. J. & Davidson, E. H. (1983) *J. Mol. Biol.* **165,** 609–632.
16. Arctander, P. (1995) *Proc. R. Soc. London Biol. Sci.* **262,** 13–19.
17. Smith, M. F., Thomas, W. K. & Patton, J. L. (1992) *Mol. Biol. Evol.* **9,** 204–215.
18. Zullo, S. (1991) *J. Mol. Evol.* **221,** 1223–1235.
19. Collura, R. V. & Stewart, C. B. (1995) *Nature (London)* **378,** 485–489.
20. Wallace, D. C. (1982) *Microbiol. Rev.* **46,** 208–240.
21. Shay, J. W., Baba, T., Zhan, Q., Kamimura, N. & Cuthbert, J. A. (1991) *Oncogene* **6,** 1869–1874.
22. Woodward, S. R., Weyand, N. J. & Bunnell, M. (1994) *Science* **266,** 1229–1194.
23. Hedges, S. B. & Schweitzer, M. H. (1995) *Science* **268,** 1191–1192.
24. Allard, M. W., Young,, D. & Huyen, Y. (1995) *Science* **268,** 1192.
25. Zischler, H., Hoss, M., Handt, O., von Haeseler, A., van der Kuyl, A. C., Goudsmit, J. & Pääbo, S. (1995) *Science* **268,** 1192–1193.
26. Anderson, S., Bankier, A. T., Barrell, B. G., deBruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., *et al.* (1981) *Nature (London)* **290,** 457–465.
27. Trounce, I. A., Kim, Y. L. Jun, A. S. & Wallace, D. C. (1996) *Methods Enzymol.* **264,** 484–509.
28. Felsenstein, J. (1995) PHYLIP, Phylogeny Inference Package (University of Washington, Seattle), Version 3.57.
29. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
30. Kimura, M. (1980) *J. Mol. Evol.* **16,** 111–120.
31. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York), pp. 21–132.
32. Felsenstein, J. (1981) *J. Mol. Evol.* **17,** 368–376.
33. Brown, M. D., Shoffner, J. M., Kim, Y. L., Jun, A. S., Graham, B. H., Cabell, M. F., Gurley, D. & Wallace, D. C. (1996) *AJMG* **61,** 283–289.
34. Shoffner, J. M., Brown, M. D., Torroni, A., Lott, M. T., Cabell, M. F., Mirra, S. S., Beal, M. F., Salvo, R., Yang, C.-C., Gearing, M., Watts, R. L., Juncos, J. L., Hansen, L. A., Crain, B. J., Fayad, M., Reckord, C. L. & Wallace, D. C. (1993) *Genomics* **17,** 171–184.
35. Hutchin, T. & Cortopassi, G. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6892–6895.
36. Kunisada, T., Yamagishi, H. & Sekguchi, T. (1983) *Plasmid* **10,** 242–250.
37. Kunisada, T. & Yamagishi, H. (1984) *Gene* **31,** 213–223.
38. Wallace, D. C. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 8739–8746.
39. Trounce, I. A., Neill, S. & Wallace, D. C. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 8334–8338.
40. Jun, A. S., Trounce, I. A., Brown, M. D., Shoffner, J. M. & Wallace, D. C. (1996) *Mol. Cell. Biol.* **16,** 771–777.
41. Wood, B. (1996) *BioEssays* **18,** 945–954.