Numt, a Recent Transfer and Tandem Amplification of Mitochondrial DNA to the Nuclear Genome of the Domestic Cat

Jose V. Lopez,1 Naoya Yuhki,2 Ryuichi Masuda,2* William Modi,1 Stephen J. O’Brien2

1 Biological Carcinogenesis and Development Program, Program Resources, Inc./DynCorp, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702, USA; Department of Biology, George Mason University, Fairfax, VA 22030, USA
2 Laboratory of Viral Carcinogenesis, National Cancer Institute-FCRDC, Frederick, MD 21702, USA

Received: 1 September 1993 / Revised: 18 December 1993 / Accepted: 27 December 1993

Abstract. The mitochondrial DNA of plant and animal cells is transcriptionally active genome that traces its origins to a symbiotic infection of eucaryotic cells by bacterial progenitors. As described by the Serial Endosymbiosis Theory, symbiotic organelles have gradually transferred their genes to the eucaryotic genome, producing a functional interaction of nuclear and mitochondrial genes in organelle function. We report here a recent remarkable transposition of 7.9 kb of a typically 17.0-kb mitochondrial genome to a specific nuclear chromosomal position in the domestic cat. The integrated segment has subsequently become amplified 38–76 times and now occurs as a tandem repeat macrosatellite with multiple-length alleles resolved by pulse-field gel electrophoresis (PFGE) segregating in cat populations. Sequence determination of the nuclear mitochondrial DNA segment, Numt, revealed a d(CA)-rich 8-bp motif [ACACACGT] repeated imperfectly five times at the deletion junction that is a likely target for recombination. The extent and pattern of sequence divergence of Numt genes from the cytoplasmic mtDNA homologues plus the occurrence of Numt in other species of the family Felidae allowed an estimate for the origins of Numt at 1.8–2.0 million years ago in an ancestor of four modern species in the genus Felis. Numt genes do not function in cats; rather, the locus combines properties of nuclear minisatellites and pseudogenes. These observations provide an empirical glimpse of historic genomic events that may parallel the accommodation of organelles in eucaryotes.

Key words: Mitochondrial DNA — Transposition — Amplification — d(CA)-rich repeats — Pseudogenes

Introduction

Mitochondrial DNA (mtDNA) has proven to be a useful tool in the molecular evolutionary genetics of animals due to an accelerated nucleotide substitution rate compared with the nuclear genome, near clonal and maternal transmission, low recombination between different mtDNA genotypes, and a conserved gene order in higher animals (Attardi 1985; Brown 1985; Wilson et al. 1985; Moritz et al. 1987; Avise 1991). Since mtDNA holds only a minimal repertoire of genetic information, many structural proteins, enzymes, and cofactors required for viable organelle function must be actively imported after cytoplasmic translation of nuclear genes (Hartl and Neupert 1990). These requirements have been interpreted in the context of the Serial Endosymbiosis Theory, or SET (Margulis 1970), which has proposed that mitochondria and plant chloroplasts arose from free-living procaryotic ancestors. A necessary assumption of the SET is that throughout the evolution of the eucaryotes, mtDNA genetic information has been gradually transferred and integrated into the nuclear genome. The SET is supported by molecular evidence showing a genetic similarity between animal
mtDNA and the α subdivision of purple bacteria (Yang et al. 1985; Gray 1989).

Several studies have demonstrated that mtDNA has been transferred between membrane-bound organelles in plant, invertebrate, and vertebrate taxa (Van den Boogaart et al. 1982; Tetszu et al. 1983; Kaminura et al. 1989; Fukuda et al. 1985; Zullo et al. 1991; Smith et al. 1991; Ellis 1982; Gelissen and Michaelis 1987). In at least two cases, long, interspersed, repetitive elements (LINES) or cryptic retroviral sequences were co-isolated with the nuclear mtDNA, which suggests a possible vehicle for the transpositions (Tsuzuki et al. 1983; Waksasugi et al. 1985). Amplification of the nuclear mtDNA has occurred in some but not all of the documented cases. These observations indicate the periodic occurrence of genetic exchanges between organelles and nuclear genomes but fail to identify a unifying mechanism or common DNA sequences which may be directly involved in the transpositions.

In the present paper, one of the largest transpositions of mtDNA into the nucleus of a higher vertebrate, the domestic cat (Felis catus), is described. During the course of developing the cat as an animal model for genetics research (O’Brien 1986), extra mtDNA restriction fragments were observed by standard Southern hybridization analyses. We initiated a set of experiments that demonstrated a nuclear location for this anomalous mtDNA in F. catus. Unlike previous studies of transferrred mtDNA, the structure and organization of feline nuclear mtDNA appear to be an amplified tandem array of approximately one-half of the mtDNA genome (7.9 kb) located on a specific feline chromosome, D2. The evolutionary implications, time frame, and possible mechanisms for the transposition and nuclear integration are discussed.

Materials and Methods

DNA Extraction and Southern Blot Analysis. Genomic DNAs were extracted from fresh lymphocytes or fibrolast cell cultures of domestic cats according to the procedures previously described (Sambrook et al. 1989). The purified DNAs were digested with restriction endonucleases according to the manufacturer’s instructions (BRL, Gaithersburg, MD). Electrophoresis and Southern transfer were also performed according to standard procedures (Sambrook et al. 1989). Typically, 1.0–3.0 µg of DNA per lane was loaded on the gels. BioTrace RP (Gelman Sciences) or Hybond-N (Amersham) nylon membranes were hybridized in a solution containing 50% formamide, 1 M NaCl, 50 mM PIPES (pH 6.8), 200 µg/ml salmon testis DNA, 0.1% Sarkosyl, 10 mM EDTA, and 5 × Denhardt’s solution (0.1% Ficoll, 0.1% polyvinyl-pyrolidone), and 0.1% bovine serum albumin fraction. Hybridization probes were radiolabeled with [32P]-dCTP by random primer synthesis kits (Boehringer Mannheim) and added to filters at a specific activity of 5 × 10^8 cpm/µg. After washing with stringent conditions (0.1× SSC, 0.1% SDS at 50°C), filters were exposed to Kodak XRP-5 X-ray film for at least 16 h and developed.

Isolation of Mitochondrial DNAs. Mitochondrial DNAs were purified by the method of Drouin (1980). Briefly, 10 g of frozen tissue was pulverized with pestle in liquid nitrogen. After evaporation of the liquid nitrogen, powdered tissue was resuspended in homogenization buffer (0.25 M sucrose, 0.15 M KCI, 10 mM Tris-HCl, pH 7.5, 1.0 mM EDTA) and homogenized with a tight-fitting pestle in a Potter Elvexem homogenizer and spun at 1,000 g for 10 min, and pellets (nuclear fractions) were used to isolate nuclear DNAs. After another centrifugation, supernatant was saved and spun at 20,000 g for 10 min. A part of the pellets was used to isolate cytoplasmic DNAs. Remaining pellets were resuspended in a solution containing 0.25 M sucrose, 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA, and EtBr and CsCl were added to bring the concentration to 300 µg/ml and a refractive index of 1.39, respectively. Pellets were then centrifuged at 55,000 rpm in a Beckman Ti60 rotor at 20°C. Supercoiled DNAs were isolated from a lower band and used as purified mitochondrial DNAs.

Isolation of Nuclear and Cyttoplasmic mtDNA Clones. The nuclear mtDNA clone, pNutm.1, was derived from a partial library of EcoRI restriction fragments in the 7.5 kb range that was cloned into the λ Zap II vector (Stratagene). Genomic DNA from a domestic cat designated FCA 65 was digested with EcoRI restriction endonuclease, separated by agarose gel electrophoresis, gel purified with Geneclean elution kits (Bio 101), and ligated with the λ vector DNA. The DNAs were then packaged in vitro with Gigapack Gold II reagent (Stratagene). The isolation and preparation of recombinant phage clones were performed according to standard procedures (Sambrook et al. 1989). The 12.0 kb cytoplasmic mtDNA clone, pCmt.12, was similarly isolated from a second size-selected library cloned into the λ Fix vector from the same cat, FCA 65. Purified mtDNA restriction fragments from puma (Felis concolor) were employed as hybridization probes for screening the phage libraries.

Nucleotide (nt) Sequence Analyses. The pNutm.1 DNA sequence was determined in both directions by subcloning restriction fragments into M13 mp18 and mp19 (BRL) single-stranded phage vectors (Sambrook et al. 1989). The oligonucleotide primer pairs used to amplify asymmetric PCR templates for DNA sequencing (Gyllensten and Erlich 1988) of homologous cytoplasmic mtDNA genes in pCmt.12 were derived from the pNutm.1 DNA sequence or from literature sources for the following genes: rRNA-Q,M and NADH-2—5’GATCCCACCTCCTAACCAGT3’/5’GTGTAACCTTCGGGCA3’; COI—5’GATAGTTAGCTCCCTCCG/5’GTTGGGCTCCATAG3’; and 12S rRNA (Koehler et al. 1989)—5’AAAACCTTCAAAGTTGGA/5’TGACTCGGAGGGTGAACGCGCGGTTGTG3’. Other cytoplasmic mtDNA gene sequences were obtained from sequencing M13 subclones. DNA sequencing reactions were performed by the dye chain termination method using either commercial Sequenase version 2.0 kits (U.S. Biochemical Corporation) or Taq dye primer/terminator cycle sequencing kits required for the automated DNA Sequencer (Applied Biosystems model 373A). Sequences were analyzed by programs of the University of Wisconsin Genetics Computer Group (UWGGC) (Devereux et al. 1984), by the Phylogeny Inference Package (PHYLIP) (Felsenstein 1993), and by Phylogenetic Analysis Using Parsimony (PAUP) (Swofford 1990).

PCR Amplification of Junction Regions between Nurm Monomers. As a test for the tandem arrangement of Nurm in genomic DNA, PCR primers were designed to amplify the DNA region spanning the unique junction between COII and D-loop gene sequences found at the genomic Nurm locus. The primer pair (J1: 5’-AATGCGGGACGTGGGGAGG3’ and J2: 5’-GCTCACGACACAG-3’) was derived and oriented toward the extreme 5’ and 3’ termini of the pNutm.1 clone. The positions of each primer are indicated below the map in Fig. 4. Fragments in the expected size range of approximately 720 bp were then amplified by standard PCR protocols (Innis et al. 1990) us-
Fig. 1. A MiDNA restriction patterns observed in fractionated preparations of domestic cat DNAs. DNA fractions were total genomic DNA (T), nuclear DNA (N), and cytoplasmic DNA (C), as described in Experimental Procedures. Total miDNA fragments were detected with a λP-labeled lambda clone, λattB-λA, containing a complete F. catus miDNA genome (O'Brien et al. 1990). Band sizes are in kilobase pairs (kb). The (*) marks the nucleus-specific mtDNA fragments. B Restriction patterns of Nuni (labeled fragments) in cat FCA 65 are identical to those in A and sum to approximately 7.9 kb. Hybridization conditions were also identical to A except that the hybridization probe was pNumt.1. C Nuclear mtDNA bands can be visualized in several unrelated cats (1–32). Genomic DNAs were digested with EcoRI, blotted, and probed with the complete mtDNA clone used in A. D Interpretation of restriction digest patterns of four enzymes shown in B conforms to a tandemly arrayed model for chromosomal pNumt.1. (See text.)

Results

Nuclear Chromosome Location of mtDNA Sequences in Cats

Nuclear mitochondrial DNA sequences (Numt, pronounced “new might”) were first detected in Southern hybridization analyses of F. catus DNA probed with a full-length mtDNA molecular clone (Fig. 1A). In both EcoRI and BamHI digestions the total molecular sizes of all mtDNA fragments were greater than 20.0 kb, although a mean size of 16.5 kb ± 200 bp has been found for nearly all mammalian mtDNAs including at least one feline species (Attardi 1985; O'Brien et al. 1990). An extra mtDNA fragment of approximately 7.9 kb in EcoRI digests or 7.6 plus 0.3 kb (visible only after longer exposure times) in BamHI digestion was evident in the nuclear and in total genome DNA preparations but absent in the supercoiled cytoplasmic DNA fraction (Fig. 1A). Extra fragments were also observed in digestions with other restriction enzymes (Fig. 1B) and in several unrelated domestic cats (Fig. 1C). The restriction patterns of total genomic DNA were interpreted in a restriction map (Fig. 1D) that reflects the tandem repeat of Numt in nuclear DNA. (See below.) Because mitochondrial genomes are present in high copy number (ca. 10^2–10^5) in most mammalian somatic cells (Birky 1978), the high intensities of suspected nuclear DNA fragments (comparable in intensity to authentic cytoplasmic mtDNA fragments) suggested that the nuclear frag-
ments were also present in multiple copies. Moreover, the ratio of the intensities between nuclear and cytoplasmic mtDNA bands varied in different individuals similar to the intensity of different cytoplasmic mtDNA fragments (Fig. 1C). This dose fluctuation was evident even when equivalent amounts of DNA were controlled by feline HOX3A gene hybridizations (Masuda et al. 1991) indicating varying copy numbers of both the extra and the cytoplasmic mtDNA among cats.

As additional evidence for nuclear residence of the extra mtDNA sequences, we examined a panel of genetically characterized rodent × cat somatic cell hybrids (O'Brien and Nash 1982; Gilbert et al. 1988). These hybrids retain the full complement of rodent chromosomes plus different combinations of each of the 19 feline chromosomes. Genomic DNA from 41 hybrids was scored for the occurrence of feline mtDNA fragments after digestions with SstI (Fig. 2A). Three feline cytoplasmic mtDNA fragments (9.0, 3.4, and 2.9 kb) were absent in all hybrids, indicating loss of feline mitochondria in the hybrids, while the two “extra” or nuclear fragments (6.1 and 1.75 kb) appeared in some hybrids but not in others. The appearance of the extra fragments was highly concordant with each other (100%) and with the presence of feline chromosome D2 (92%) and reciprocally discordant (26–53%) with each of the other cat chromosomes (Fig. 2B). These results affirm the nuclear location of Nunt and implicate its position on chromosome D2. This conclusion was confirmed and extended by fluorescence in situ hybridization (FISH) using a molecular clone of the nuclear mtDNA (pNunt.1; see below) to metaphase chromosome preparations from the domestic cat. The results of this analy-

Fig. 2. Segregation of Nunt with specific cat chromosomes in a panel of rodent × cat somatic cell hybrids (O'Brien and Nash 1982). A Pattern of mtDNA hybridization in SstI digest of genomic DNA from Chinese hamster × cat somatic cell hybrids probed with a feline mtDNA molecular clone. The 6.1- and 1.75-kb feline fragments represent Nunt while the 9.6- and 2.2-kb bands in the hybrid lanes stem from cross-hybridization with hamster cytoplasmic mtDNA. B Discordancy plot of the Nunt fragments with the 19 feline chromosomes segregating in the hybrid panels.

Fig. 3. Metaphase chromosomes of the domestic cat following fluorescent in situ hybridization (FISH). The hybridization signals (yellow spots) are localized at the pericentric region of chromosome D2. Molecular Cloning of Nuclear Mitochondrial DNA (Nunt) and Cytoplasmic Mitochondrial DNA (cmt) from F. catus

Seven clones were isolated from a size-selected EcoRI genomic library made with cat FCA 65 DNA and λ Zap II; one clone, designated pNunt.1, contained the predicted size for Nunt, 7.9 kb (Fig. 1A), and was sequenced and characterized in detail (Fig. 4). In addition, two cytoplasmic mtDNA clones (pCmt.12 and pCmt.4.8) were also recovered from FCA 65 and represent the two EcoRI digest fragments (12.0 and 4.8 kb, respectively) shown in Figs. 1 and 4.
Fig. 4. A Comparison of restriction maps and gene content of F. catus nuclear (pNumt.1) and cytoplasmic mtDNA clones. Gene sizes are drawn to a scale based on the complete pNumt.1 DNA sequence and the mean sizes of previously characterized human, cow, and mouse mtDNAs. Cytoplasmic mtDNA is represented by two cloned EcoRI fragments, pCmt.12 and pCmt.4.8, and indicated by hatched and open bars, respectively. Black bars below the pNumt.1 map represent regions of comparison which have been sequenced in both pNumt.1 and pCmt.12. Dark boxes represent mitochondrial tRNA genes based on the mammalian consensus and written in standard amino acid notation. The precise size or order of the genes in the regions that do not overlap with pNumt.1 (e.g., pCmt.4.8) are approximate and have not been completely verified by DNA sequencing. In pCmt.4.8 (open bar), the presence of ATP synthase subunit 6 and COIII genes was verified by nt sequencing. J1 and J2 indicate the primers used to amplify across the unique DL junction in Numt. B Alignment of Numt and cytoplasmic mitochondrial D-loop (DL) sequences. The DL region in pNumt.1 is shortened relative to pCmt.12 by about 1 kb (700 + 240 bp) as indicated above the sequence alignment. The pCmt.12 clone contains a longer dCA repeat (240 bp) region, the remainder of the mitochondrial genome, and 5’ DL. Percent identity with human CSBs is in parentheses. The putative positions of heavy and light strand promoters (HSP and LSP, respectively) are inferred from human and other mammalian DL sequences (Saccone et al. 1991). The imperfectly repeated 8-bp dCA-rich motif occurs at the deletion junction between COII and DL sequences in Numt. (See text.)

The gene content and order of the feline pCmt.12 clone and the nuclear mtDNA clone, pNumt.1, were determined by restriction mapping, by complete nucleotide sequencing of pNumt.1, and by partial sequencing of pCmt.12 (Fig. 4). Restriction maps and sequence alignment of pNumt.1 and pCmt.12 were sufficiently conserved to identify all the mitochondrial gene segments present in the two constructs (Fig. 4). The pNumt.1 DNA sequence represented a 7.9-kb fragment with strong sequence (and contiguous gene order) homology to the feline mtDNA (pCmt.12) and to previously reported sequences of mitochondrial DNA genomes of human, cow, mouse, rat, and harbor seal (Anderson et al. 1981, 1982; Bibb et al. 1981; Gadaleta et al. 1989; Arnesson and Johnsson 1992).

The pNumt.1 clone represents a truncated and moderately divergent homologue of the mitochondrial genome that has a COII gene interrupted downstream from the EcoRI site present in both cytoplasmic mtDNA and in Numt (Fig. 4). Contiguous regions of the mtDNA COII gene are found at either end of the EcoRI site that defines the termini of the pNumt.1 molecular
Fig. 5. PFGE analysis of restriction digested genomic DNA. A DNAs shown are from cat FCA 65. Restriction enzymes used in the digestion are indicated above the sample lanes. Pulse time was 65 s. The 32P-labeled pNutm.1 clone was used to probe all of the PFGE blots. Undigested genomic DNA also hybridized to the probe and appears as the highest band on all of the autoradiograms. B–D Mendelian transmission determined by PFGE analysis of high-molecular-weight cat DNA from unrelated pedigrees. All samples were cut with the restriction endonuclease BglII. Each panel represents separate families run under specific PFGE conditions. Numbers above the autoradiogram identify individuals. Pedigree 1 (B) is composed of two different gels run with a 60-s switch time; cats 34 and 43 were run at a slower velocity (4.5 V/cm). Pulse times for pedigrees 2–3 (C and D, respectively) were 60 and 90 s, respectively. The designation of nine proposed allelic states (A–I) is indicated alongside the band sizes.

c and i

c

clone, suggesting a circular precursor of the 7.9-kb nuclear segment. The pNutm.1 COII sequence (552 bp) is foreshortened relative to cytoplasmic mtDNA COII genes in cat and other species (676 bp). COII in Nunt is fused to about 200 bp of a truncated segment homologous to the mammalian mtDNA control or D-loop region (DL) in Fig. 4. Taken together these data support a simple deletion of an ancient cytoplasmic mtDNA molecule followed by circularization, joining a truncated COII gene with a deleted DL control region.

Sequence comparison of the DL segment of pNutm.1 with pCmt.12 and other DL regions revealed that pNutm.1 contained two segments that are very similar to regulatory conserved sequence blocks (CSBs) found in mammalian cytoplasmic mitochondrial DNA and shown in Fig. 4B (Chang and Clayton 1984; Wallace et al. 1991). An origin of replication was not apparent in pNutm.1. There was a notable 8-bp motif (ACA-CACGT), which was imperfectly repeated five times at the junction between COII and the D-loop control region (Fig. 4B). A similar d(CA)-rich repeat is also present as a longer version in pCmt.12 and in other carnivore mtDNA (Hoelzel et al. 1993) and thus may be a candidate region for recombination between mitochondrial and nuclear genomes.

Nuclear Mitochondrial Sequences Occur as a Tandem Repeat in the Cat Genome

To account for the relatively high copy number of Nunt, we postulate a tandemly repeated arrangement for pNutm.1 which includes the 7.9-kb fragment (Fig. 4A) as the basic repeating unit. This hypothesis is supported by the hybrid panel and fluorescent in situ hybridization (FISH) results, which preclude epigenomal or chromosomally dispersed copies of Nunt (Fig. 3) and the aforementioned splitting of COII gene sequences between the 5' and 3' termini of pNutm.1. The consistent size of Nunt fragments (7.9 kb) with different restriction enzyme digestions plus our ability to produce a restriction map from genomic DNA (Fig. 1D) that was nearly identical to the pNutm.1 restriction map (Fig. 4A) are also consistent with a tandem array of Nunt sequences.

To confirm a tandem arrangement, we digested high-molecular-weight genomic DNA from several cats with 6-bp cutting restriction enzymes whose target sites were not present in the pNutm.1 DNA sequences. The restriction products were separated by pulse-field gel electrophoresis (PFGE) and probed by Southern analysis for Nunt DNA sequences. The genomic fragments were recovered and their estimated molecular size is illustrated in Fig. 5. The PFGE results revealed high-molecular-weight Nunt-specific fragments ranging in size from 300 to 600 kb, two orders of magnitude greater than the mean fragment size expected from a 6-bp cutting restriction enzyme (1 site per 4,096 bp) and about 20 times the average mammalian mtDNA size. If the Nunt locus were represented by a tandem array of 7.9-kb repeats like the pNutm.1 clone, the number of repeats would range from 38 to 76 copies.
Fig. 6. DNA sequence alignment of feline nuclear, pNum1.1, and cytoplasmic, pCnt1.2, clones with homologous gene sequences from other mammalian species. Non-feld mammalian DNA sequences (cited in text) are abbreviated as follows: Bov, Bos taurus (cow); Phoca, Phoca vitulina (harbor seal); Hum, Homo sapiens (human); Rat, Rattus rattus (rat); and Mus, Mus musculus (mouse). Dots symbols directly below the FCA-Cmnt1.2 sequence indicate conserved nucleotides. Asterisks (*) denote gaps in the sequence. Underlined nucleotides designate the highly conserved sequences shared with prokaryotes (Eperon et al. 1980). The underlined number below the sequence refers to the corresponding nt position in the published human (H) or E. coli (E) sequences, while all other numbering refers to the feline nt positions in this study. A 16S rRNA gene; B comparison of 12S rRNA sequences. A second individual cytoplasmic FCA sequence (FCA 42) is shown. The remainder of the sequences include Fsi—F. silvestris (European wild cat); Fli—F. lybica (African wild cat); Fma—F. margarita (sand cat); Fni—F. nigripes (black-footed cat); Fch—

F. chaus (jungle cat); Oma—Onocolobus manul (Pallas cat); Og—Oncley's guigna (kodok); Ple—Panthera leo (lion). Primer and template preparation and phylogenetic analysis are given by Masuda et al. (in preparation). Other abbreviations and notation are as in A. C tRNA sequences: the cytoplasmic sequences of I, Q, and M1 tRNAs were derived from asymmetric PCR products amplified from the pCnt1.2 template. D,E ND1, ND2, COI, and COII gene sequences. The deduced peptide sequences after translation of open reading frames in cytoplasmic pCnt1.2 mtDNA are written above the nucleotide sequence in standard amino acid nomenclature, utilizing the mitochondrial code. Each amino acid letter marks the first position of a codon. Underlined codons in the mtDNA sequences denote a termination codon recognized in the nuclear genetic code. Double underlines denote termination codons recognized exclusively in the mitochondrial genetic code. Other symbols and species abbreviations remain as in A.
To verify whether the multiple bands represent variant length alleles of *Numt* loci, pedigree analyses were also performed with PFGE. The high-molecular-weight *Numt* DNA fragments were polymorphic among the three families (Fig. 5B–D).

Although the patterns were complex in some cases (i.e., some alleles were represented by more than one PFGE fragment), in every case the transmission of fragments conformed to Mendelian segregation expectations. For example, in Fig. 5B, at least three alleles (designated A, B, and C) are segregating in the three generations. Additional cat pedigrees (Fig. 5C,D) also demonstrate Mendelian transmission of at least one fragment from parent to offspring. The PFGE data
strongly affirm the interpretation that the Numt locus on cat chromosome D2 consists of a tandem array of multiple 7.9-kb segments of mitochondrial DNA repeated differently in cat chromosomes from 38 to 76 copies.

Comparative Sequence Analyses of Cytoplasmic and Nuclear mtDNAs

Nucleotide sequences were obtained from 10 gene segments of pNumt.1 and pCmt.12 to address the following points: (1) confirmation of homology, gene content, and Numt gene-coding capability; (2) characterization of the types of sequence changes between Numt and cytoplasmic mtDNA (e.g., substitutions, deletions, insertions, etc.); and (3) measurement of rates of DNA sequence divergence that occurred between the two distinct mtDNA sequences. The pNumt.1 and pCmt.12 sequenced regions shown in Fig. 6A–E comprise the three different types of genes (tRNA, rRNA, and protein coding) found in most mitochondrial genomes, plus the D-loop region (Brown 1985). For both pNumt.1 and pCmt.12, the previously characterized mammalian mtDNAs were used to identify operational open reading frames. The high degree of amino acid conservation between other mammal sequences and pCmt.12 indicates that the cat mtDNA structural genes encode functional mitochondrial proteins. A summary of the pattern and extent of nucleotide and amino acid differences for segments of 10 mitochondrial genes is illustrated in Fig. 6 and in Table 1.

Most (67/81) of the mutations between the pNumt.1 and pCmt.12 structural genes are synonymous and occur within nonconserved regions found in previous comparisons (Anderson et al. 1982). This indicates the presence of functional constraints against nonsynonymous amino acid replacement substitutions (Nei 1987). For instance, all three first-position changes in the COII gene of pNumt.1 and pCmt.12 occur at leucine codons, which are synonymous. In general, the percent sequence substitutions between the two cat mtDNAs shown in Table 1 are much lower than the values reported in previous studies comparing nuclear mtDNA and cytoplasmic DNA in other species (Fukuda et al. 1985; Smith et al. 1991).

Mutational residue differences between pNumt.1 and pCmt.12 were directly observed by polymerase chain reaction (PCR) sequence reactions (Fig. 7). When unFractionated genomic DNAs were used as templates to generate asymmetric PCR products for sequencing, the sequencing gels showed fragments in two lanes as if the species were heterozygous or heteroplasmic. In contrast, using individual cloned pNumt.1 or pCmt.12 fragments as templates yields unequivocal sequences. Fig-
Figure 7 illustrates the types of mutations detected in the 12S rRNA gene including base substitutions and the single-d(A) insertion; the latter causes a shifting of bands in the total DNA gels.

Phylogenetic Origin of Nnum Transposition

We used 12S rRNA gene sequence data (Woese 1987; Mindell and Honeycutt 1990) to place Nnum DNA with-
in an evolutionary framework among the Felidae. The six non-domestic cat species directly following the F. catus 12S rRNA sequences in Fig. 6B have been classified within the genus Felis (domestic cat lineage) by morphological and molecular techniques (Wurster-Hill and Centerwall 1982; Collier and O’Brien 1985; O’Brien et al. 1987; Nowak 1991). The Panthera leo (lion) and Oncifelis guigna (kodkod) DNA sequences represent the other two major branches in the cat family—the Panthera and ocelot groups, respectively.

The aligned 12S rRNA sequence data were analyzed using three distinct phylogenetic approaches: (1) a phylogenetic (distance matrix) analysis of overall pairwise sequence divergence between species; (2) a maximum parsimony (cladistic) analysis of character changes; and (3) a maximum likelihood evaluation of all possible phylogenetic trees relating the sequences in Fig. 6B. The results of these analyses, illustrated by the maximum likelihood phylogenetic tree in Fig. 8, were topologically consistent in that they reaffirmed the distinction of the three lineages of Felidae (ocelot lineage, O. guigna; Panthera lineage, P. leo; and domestic cat lineage, F. catus, F. libycus, F. silvestris, F. margarita, F. nigripes, F. chaus, and Otocolobus manul) and indicated a hier-

Fig. 6. Continued.
Table 1. Sequence comparison of feline cytoplasmic (pCmt.12) and nuclear (pNumt.1) mitochondrial DNA and between mammalian mitochondrial DNA^a

<table>
<thead>
<tr>
<th>Gene</th>
<th>Nt^b</th>
<th>Subst.</th>
<th>Trt</th>
<th>Trv</th>
<th>Gaps</th>
<th>Codon position^c</th>
<th>Nucleotides</th>
<th>Amino Acid^d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO-I</td>
<td>239</td>
<td>14</td>
<td>10</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1(1)</td>
<td>13(2)</td>
</tr>
<tr>
<td>CO-II</td>
<td>250</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>3(0)</td>
<td>1(1)</td>
<td>16(0)</td>
</tr>
<tr>
<td>ND-1</td>
<td>345</td>
<td>24</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>4(1)</td>
<td>0</td>
<td>20(1)</td>
</tr>
<tr>
<td>ND-2</td>
<td>326</td>
<td>23</td>
<td>18</td>
<td>5</td>
<td>2</td>
<td>5(4)</td>
<td>1(1)</td>
<td>17(3)</td>
</tr>
<tr>
<td>RNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12s rRNA</td>
<td>371</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16s rRNA</td>
<td>611</td>
<td>18</td>
<td>10</td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Asp</td>
<td>69</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Leu</td>
<td>75</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Gln</td>
<td>74</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Met</td>
<td>69</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>2429</td>
<td>124</td>
<td>98</td>
<td>26</td>
<td>4</td>
<td>5.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^a Abbreviations: N—pNumt.1; C—pCmt.12; H—human; B—bovine; P—porcine (seal); Trt—trinitis; Trv—transversions
^b Total PD counts gaps as single residue difference. Gaps are counted only in the FCA pNumt.1 and pCmt.12 DNA comparison
^c Numbers of nonsynonymous substitutions (subst.) are in parentheses
^d Translation is based on the mitochondrial genetic code

Architectural divergence pattern within the domestic cat (Felis) lineage. The derived phylogeny places the Numt origin within the radiation of the species of Felis at approximately the time that the ancestors of F. nigripes diverged from the common ancestor of F. catus, F. silvestris, F. libyca, and F. margarita. The result is consistent with the disposition of large Numt nuclear fragments in these species as Numt nuclear fragments are present in F. catus, F. silvestris, F. libyca, and F. margarita, but absent in F. nigripes, F. chaus, and O. manul (Fig. 8). Taken together the results suggest the transfer of Numt from cytoplasmic mtDNA to the nuclear genome occurred prior to the species divergence of the domestic cat and its three closest relatives but subsequent to the split of F. nigripes and F. chaus from the Felis lineage.

Discussion

The data summarized here provide evidence for the ancestral transposition of approximately one-half of the cytoplasmic mitochondrial DNA genome to a chromosomal region of the nuclear genome in domestic cats, F. catus. The segment was subsequently amplified in tandem and occurs in modern cats as a 300–600-kb segment of nuclear mitochondrial (Numt) DNA that segregates as a repeat length polymorphism in domestic cat populations. A phylogenetic analysis of 12S RNA sequences from Numt, cat, and related feline species' cytoplasmic mtDNA indicated that the transposition was an historic event, occurring at a precise time point during the evolutionary divergence of species of the genus Felis (Fig. 8). The tandem expansions of Numt are apparently still occurring through a process reminiscent of minisatellite expansion/contraction as multiple length alleles (Fig. 5) segregate in cat populations.

The evidence for the nuclear location and concatameric mtDNA sequences included: (1) occurrence of 7.9 kb of "extra" mtDNA restriction fragments in nuclear DNA from cats; (2) segregation of these fragments concordant with feline chromosome D2 in a rodent × cat hybrid panel; (3) localization of Numt to chromosome D2 by fluorescent in situ hybridization; and (4) Mendelian transmission of large PFGE chromosomal Numt fragments in domestic cat pedigrees. The Numt transposition differs from nuclear mtDNAs described in other vertebrate species in the tandemly repeated arrangement of Numt, in its very recent origin, and in the unparalleled magnitude of the feline mtDNA transposition.

Mechanisms for the Generation of Numt

Hypotheses concerning the origin of the Numt locus must incorporate at least two primary molecular processes: recombination and gene amplification. Although intermolecular recombination between discrete mammalian mtDNAs occurs infrequently by organelle fusion (Hayashi et al. 1985), intramolecular recombina-
a similar process, a chromosomal integration (reminiscent of modern transgenesis) would place a large segment in the chromosomal targets. The absence of interspersed nuclear DNA sequences in the *Numt* repeat motif would support (but not prove) the extrachromosomal tandem amplification of *Numt* prior to its original integration. Similar extranuclear concatenation occurs in transgenesis experiments (Capecchi 1989).

Previously described transpositions of mtDNA to nuclear genomes have involved transfer of both edited RNA precursors (Nugent and Palmer 1991) and non-transcribed DNA segments (Quigley et al. 1988; Ganitt et al. 1991). The presence of normally untranscribed D-loop regulatory elements in *Numt* (Fig. 4) would suggest that the precursor of *Numt* was a DNA fragment that contained untranscribed signal sequences plus a portion of the d(CA)-rich repeat. Similar “microsatellite” repeats are found in several mammalian mitochondrial genomes including cats (Fig. 4) (Anderson et al. 1982; Arnason and Johnsson 1992; Hoelzel et al. 1993) as well as being widely dispersed in the mammalian nuclear genome (Weber 1990; Dietrich et al. 1992; Serikawa et al. 1992). These repeats may facilitate exchange between DNA segments as they appear to enhance the rate of homologous recombination in vitro (Wahls et al. 1990).

**Array Length Polymorphism of *Numt***

The segments of *Numt* found in the cat genome likely represent a tandem array of 7.9 kb of mitochondrial DNA detected in PFGE fragments of 300–600 kb (or 38–76 copies). Different size fragments found in different individuals segregated in a Mendelian fashion (Fig. 5), supporting the interpretation that the PFGE fragments are alleles that differ in repeat number. In parallel with studies of minisatellite length repeats, these alleles may be generated by DNA replication slippage, unequal crossing over, or both (Jeffreys et al. 1991; Levinson and Gutman 1987). Other interpretations for individual heterogeneity—such as short intervening genomic (nonmitochondrial) spacer DNA, or the occurrence of novel restriction sites in a fraction of the repeat members—are possible explanations for our observation of more than one fragment in a postulated allele (e.g., alleles A and B in Fig. 5B).

**Numt as a Pseudogene: Functional Implications***

Although *Numt* shares several aspects of moderately repetitive DNA families, an important difference is that its ancestors were coding genes, albeit mitochondrial ones. This history makes its interpretation reminiscent of a pseudogene with an unusual tandem amplification in its origins. Transfer to the nucleus followed by du-
Fig. 8. Phylogenetic analysis of the 12s rRNA gene (Fig. 6B) in homologous Numt and cytoplasmic mtDNA from F. catus, from other species within the genus Felis, plus more divergent P. leo (lion) of the Panthera lineage and O. guigna from the ocelot lineage (after Masuda et al., in preparation). The tree presented is a maximum likelihood analysis of 10 species plus Numt using the DNAML algorithm of PHYLIP (Felsenstein 1981, 1993). The routine employed empirical base frequencies from the data set, a transversion/translation ratio of 5.0, and a randomized input order. The log likelihood of the tree was -902.65. Each terminal node except the unresolved (F. catus, F. silvestris, and F. libycus) group has limb lengths significantly ($P < 0.01$) greater than 0. Scale is percent nucleotide sequence differences between species; limb lengths are the same units. The same data were used to produce a phenetic distance matrix that was analyzed using the neighbor-joining method and the least squares (FITCH) procedure incorporated in the PHYLIP computer package (Devereux et al. 1984; Saitou and Nei 1987; Fitch and Margoliash 1967). In addition, a parsimony or cladistic analysis employed the PAUP package (Swofford 1990). Each of the phylogenetic analyses produced trees that were topologically equivalent but slightly different in limb length.

The presence of Numt was investigated in other feline species based on the presence of mtDNA fragments in Southern blots of genomic DNA exceeding 22 kb. A positive Southern result was also observed in P. leo which was unexpected from the phylogenetic tree. In addition, PCR-generated fragments that used the J1 + J2 primers from Numt COI and D-loop sequences spanning the junction between repeat segments were detected in F. catus, F. silvestris, F. margarita, and F. libycus but not in O. manul.

**Phylogenetic Interpretation**

Alignment of Numt sequences to cytoplasmic mtDNA homologues in cat species demonstrated a high degree of homology (Table 1) phylogenetically consistent with a divergence of Numt within the radiation of the genus Felis. As this evolutionary divergence likely occurred toward the late Pliocene, within the last 3.0 million years (MY) (Masuda et al. in preparation; Collier and O'Brien 1985; Kurten 1968), the phylogenetic analyses support the divergence of Numt at roughly 1.8 MY BP (before present). This date is estimated from the application of the overall genetic distance of Numt vs cytoplasmic mitochondrial DNA of 5.3% (Table 1) to the estimation of Li et al. (1981) whereby the fraction of sequence divergence is:

$$
\delta = (\mu_1 + \mu_2)t
$$

where $\mu_1$ and $\mu_2$ are the mutation rates of diverging sequences and $t$ is the time elapsed. Substituting $\mu_1 = 2.5 \times 10^{-8}$ substitutions/site/year for cytoplasmic mitochondrial DNA (Hasegawa et al. 1985) and $\mu_2 = 4.7 \times 10^{-9}$ substitutions/site/year for nuclear pseudogene divergence (Li et al. 1981), we compute 1.78 MY as the time elapsed since Numt and cat cytoplasmic mtDNA diverged.

We cannot discern from the available data whether
the \textit{Numt} transposition and amplification were contemporaneous events or occurred at different times in the evolution of the felid genomes. The pattern of \textit{Numt} mutational divergence (Table 1) is relevant to this question since the observed changes appear to reflect selective constraints of functional genes. Most of the changes between \textit{Numt} and cytoplasmic mtDNA were transitions (98/124 = 79\%) and the majority of mutations in coding genes (COI, COII, ND1, and ND2) (67/81 = 83\%) were synonymous substitutions. A selectively neutral model of nuclear pseudogene divergence would predict a much lower incidence (ca. 33\%) for synonymous mutations, indicating that non-synonymous (codon altering) mutations in \textit{Numt} ancestors had been eliminated by natural selection, which could have occurred while evolving as mitochondrial genomes (Kimura 1983; Ohta 1992; Hughes and Nei 1988; Yuuki and O’Brien 1990). Further, the occurrence of \textit{Numt}-like sequences in other cat species outside of the \textit{Felis} genus (e.g., lions; see Fig. 8) but not in all species more closely related to domestic cat is enigmatic and may reflect a discordance in evolutionary time between \textit{Numt} transposition and \textit{Numt} amplification. That is, an ancient transposition earlier in the Felidae or Carnivore radiation that persisted as a single or low copy number until periodic salatory amplifications occurred is possible since our methods would only detect amplified copies in divergent species. Resolution of these questions would offer considerable insight into the evolutionary patterns of genome organization and must await molecular genetic analysis of the homologous chromosomal target of \textit{Numt} in distantly related Felidae species.

Acknowledgments. Portions of this paper fulfill the PhD requirements for JVL at George Mason University. The authors wish to acknowledge the excellent technical assistance of Mary Eichelberger, Alyce Burke, and Stanley Cevario, plus the early empirical observations of Drs. M. George and L. Forman. We are also grateful to Drs. W.F. Martin, J.C. Stephens, Michael Dean, Mary Carrington, Rus Hoelzel, Leslie Lyons, Jill Slattery, Marilyn Raymond, Mike Clegg, and Ms. Janice Martenson for critical reading of the manuscript. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. Serum and tissue samples of endangered species were collected in full compliance with specific federal permits (CITES, Endangered and Threatened Species; Captive Breeding) issued to the National Cancer Institute, National Institutes of Health, principal officer S.J. O’Brien, issued by the U.S. Fish and Wildlife Service of the Department of Interior.

References


Capecchi MR (1989) Altering the genome by homologous recombination. Science 244:1288–1292


Swofford DL (1990) Phylogenetic analysis using parsimony (PAUP), version 3.0. Illinois Natural History Survey, Champaign, IL