4-1-1995

Molecular Evolution of Numt, a Recent Transfer and Tandem Amplification of Mitochondrial DNA into the Nuclear Genome of the Domestic Cat (Felis catus)

Jose V. Lopez
Nova Southeastern University, joslo@nova.edu

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Molecular Evolution of *Numt*, a Recent Transfer and Tandem Amplification of Mitochondrial DNA into the Nuclear Genome of the Domestic Cat (*Felis catus*)

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University.

by

Jose Victor Lopez
Master of Science, The Florida State University
Bachelor of Science, Georgia Institute of Technology

Director: Dr. Alan H. Christensen, Ph.D
Department of Biology, George Mason University

Spring 1995
George Mason University
Fairfax, Virginia

MOLECULAR EVOLUTION OF *NUMT*, A RECENT TRANSFER AND TANDEM AMPLIFICATION OF MITOCHONDRIAL DNA INTO THE NUCLEAR GENOME OF THE DOMESTIC CAT (*FELIS CATUS*)

by

Jose Victor Lopez

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
the Requirements for the Degree
of
Doctor of Philosophy
Environmental Biology and Public Policy

Committee:

Alan H. Christensen, Director

Harold J. Morowitz

Jim Brown

Stephen Johnson

Jane M. Flinn

R. Christian Jones, Program Director

George Bradykowski, Department Chairperson

Hans Pick, Dean of the College of Arts and Sciences

Date: 3/22/95

Spring 1995
George Mason University
Fairfax, Virginia

Acknowlegements

This dissertation is dedicated to my grandparents -
Jose and Angela, Domingo and Victoria. Lahat mahal kita.

Many people have contributed to the success of this project and my education. Although I stated similar sentiments in the preface of my Master's thesis, becoming an evolutionary geneticist has made me even more aware of my indebtedness to all who have gone before me. First and foremost, I deeply thank Dr. Stephen J. O'Brien, chief of the Laboratory of Viral Carcinogenesis (LVC) for becoming my mentor, guiding this research and creating a professional environment which challenged and stimulated my growth as a scientist. This was an experience which I will carry with me always. I also gratefully acknowledge the contributions of Dr. Alan Christensen, chairman of my doctoral committee, for valuable, frank counsel on the technical and strategic molecular biological aspects of the project, as well as extending his biological interests into organisms with carnassial teeth. I thank Drs. Luther Brown, Harold Morowitz and Jane Flynn, other members of the doctoral committee, for lending me their time and wisdom. I feel privileged to have been able to interact closely with the talented and multi-faceted personnel of the LVC and NOAHS. The conscientious supervision and helpful advice from Dr. Naoya Yuhki *in situ* will always be remembered. I am grateful to Dr. William Modi, for applying his cytogenetics expertise towards solving the *Numt* problem, and Drs. Rus Hoelzel and Clay Stephens for supplementing and enhancing my

data with their erudite analyses and comments. I am also thankful for being able to share ideas and collaborations with diverse postdoctoral fellows: Drs. Ryuichi Masuda, Marilyn Raymond, Jill Pecon Slattery, Warren Johnson, Leslie Lyons, Abi Chidambarum, Lu Zhi, and Tom Laughlin. I will always cherish the fellowship and camaraderie associated with present and former graduate students of the LVC: Cheryl Winkler, Dennis Gilbert, Dianne Janczewski, Eric Brown, Sri Miththapala, Melanie Culver, Mike Cullen , Carlos Driscoll, Joelle Wentzell and Meredith Yeager. My productivity also relied on the smooth operation of the LVC, which flowed from the steadfast and cheerful presences of Janice Martenson, Mary Eichelberger, and Stanley Cevario. I am indebted to the friendly assistance of both PRI and LVC front offices comprised of Amy Huter-Imming, Theresa Covell, Laraine Main, Virginia Frye, Patricia Johnson, and the late Linda Ricketts. I am also grateful to Debbie Lomb, Mark Gunnell and Gary Smythers for invaluable help with computer programs. At the NCI-FCRDC graphics department, I am very grateful to Dru Claar-Thompson, Carolyn Whistler, Ellen Frazier, Jennifer Brown and other graphic artists for withstanding my endless figure revisions. Lastly, completion of this project would have been impossible without the care and support of my inner circle, which includes my beloved confidante, Megan James, many other Friends, my sister Stella Diffenderfer and her family, and most especially my dearest father and mother, Dr. Ernesto and Rosario Quedding Lopez. Their love and faith in my capabilities sustained me throughout the academic journey.

Table of Contents

Page

List of Tables

List of Figures

Figure

List of Abbreviations

bp - base pair
C-[value] - Complexity
cpm - counts per minute
CR/DL - control region or D[isplacement] loop
DNA - deoxyribonucleic acid
FADH - Flavin adenine dinucleotide, reduced
FISH - Fluorescent in situ hybridization
HSP- Heavy-strand origin of replication
indel - insertion or deletion mutation
kb - kilo base pair
λ - Lambda
LSP - light-strand promoter
ml - milliliters
MtDNA - Mitochondrial DNA
MYA - Million Years Ago
NADH - Nicotinamide adenine dinucleotide, reduced
Numt - Nuclear mitochondrial DNA
nt - nucleotide
OLR - Light strand origin of replication
ORF - open reading frame
PCR - polymerase chain reaction
PFGE - Pulsed field gel electrophoresis
pm - pico mole
rc - reverse complement
RFLP - Restriction Fragment Length Polymorphism
RPA - Ribonuclease Protection Assay
rRNA - ribosomal RNA
SCE - Sister chromatid exchange
SET - Serial Endosymbiosis Theory
SINES - short interspersed repeated sequences
TAS - termination-associated sequences
Ti - transition
Tv - transversion
μg/μl - microgram/microliter
UWGCG - University of Wisconsin Genetics Computer Group
VNTR - variable number of tandem repeats

Abstract

MOLECULAR EVOLUTION OF *NUMT*, A RECENT TRANSFER AND TANDEM AMPLIFICATION OF MITOCHONDRIAL DNA INTO THE NUCLEAR GENOME OF DOMESTIC CAT (*FELIS CATUS*)

Jose Victor Lopez, Ph.D.

George Mason University, 1995

Dissertation Director: Dr. Alan H. Christensen

Mitochondrial DNA (mtDNA) are functional cytoplasmic chromosomes, tracing origins to a symbiotic infection of eukaryotic cells by bacterial progenitors. As prescribed by the Serial Endosymbiosis Theory, symbionts have gradually transferred their genes to the nuclear genome that enable functional interaction. In this dissertation, a 7.9 kb transposition of a typically 17.0 kb mitochondrial genome to a specific chromosomal position in the domestic cat is reported. The integrated mtDNA has amplified about 38-76 times and now occurs as a "macrosatellite"-like tandem repeat with multiple length alleles resolved by pulse field gel electrophoresis (PFGE) segregating in cat populations. To examine the tempo and mode of evolution between different organelles, characterization of the complete 7946 bp nuclear mitochondrial DNA monomer, *Numt*, and cytoplasmic mtDNA (17,009 bp) sequences reveals about 95% similarity, which supports recent divergence within 1.8-2.0 MYA and the radiation

of four modern species in genus *Felis*. The motif, [ACACACGT], appears imperfectly repeated at the deletion junction of the control region (CR), and a likely target for recombination. Simple repeats are also implicated in indel generation. Most substitutions between the cat homologues are attributable to accelerated cytoplasmic mtDNA evolution, yet maintain a uniform rate of synonymous substitutions between different mitochondrial genes. Results of ribonuclease protection assays on cellular RNA verify the lack of *Numt*-specific transcription and the appraisal of *Numt* as a molecular "fossil". Despite an elevated number of transversions and no increase in dA/dT content over cytoplasmic mtDNA, *Numt* resembles archetypal pseudogene evolution.

To place the felid data in the context of functional mitochondrial genomes, pairwise similarity comparisons of all 37 mtDNA coding genes and the CR among eight complete mitochondrial genomes of five placental mammals were performed. In carnivores, the ND4L and ATPase 6 genes exhibit higher sequence conservation, while cyt B shows accelerated divergence. Lastly, the occurrence of *Numt*-like loci in other exotic felids deviates from current phylogenetic predictions. To confirm homology with the *F. catus Numt* locus, a series of experiments was conducted to isolate chromosomal sequences directly flanking *Numt*-like loci. These observations provide an empirical glimpse of historic genomic events that may parallel the accommodation of organelles in eukaryotes.

CHAPTER 1

ISOLATION AND MOLECULAR CHARACTERIZATION OF NUMT FROM THE DOMESTIC CAT NUCLEAR GENOME

BACKGROUND AND OBJECTIVES

"Every solution of a problem raises new unsolved problems; the more we learn about the world, the more conscious, specific, and articulate will be our knowledge of what we do not know, our knowledge of ignorance. For this, indeed, is the main source of our ignorance - the fact that our knowledge can be only finite, while our ignorance must necessarily be infinite." - Karl R. Popper, *Conjectures and Refutations* (1968)

In this dissertation, one of the largest transpositions of mitochondrial DNA (mtDNA) into the nuclear genome of a higher vertebrate, the domestic cat (*Felis catus*), is described. The cytoplasmic compartments, or organelles, possessing specialized functional roles have been recognized as a distinguishing trait of eukaryotic cells for many years (Wilson, 1959; Gilham, 1978; Alberts et al, 1989). As a eukaryotic organelle, mitochondria provide the site of energy production via the oxidation of carbon compounds and the coupling of chemiosmotic gradients with electron-transfer pathways to form phosphate bonds in ATP (adenosine triphosphate) (Hatefi, 1985). Through the process of oxidative phosphorylation, three vital and interrelated functions are served: reoxidation of NADH and FADH2, generation of ATP (energy), and regulation of body temperature by generating heat (Wallace, 1994).

Although mitochondrial genetics studies on yeast (Ephrussi, 1949) preceded the actual discovery of nucleic acids within mitochondria (Nass and Nass, 1963), the latter serendipitously ushered in a new era of molecular genetics research (Wallace, 1982; Attardi, 1985; Tzagloff and Myers, 1986). Along with the nearly universal distribution among eukaryotic organisms, several key features of mtDNA seized the attention of population and evolutionary biologists (Brown, 1983; Wilson et al, 1985; Moritz et al, 1987; Harrison, 1989; Avise, 1986; 1991; 1994; O'Brien et al, 1990). Firstly, mitochondrial gene sequences evolve on average about 2% per MY, which is about 5-10 times faster than most nuclear genes (Brown et al, 1979; 1982; Nei 1985; 1987). Secondly, in most eukaryotic organisms, mtDNAs undergo nearly uniparental transmission, typically through the matriline (Dawid and Blackler, 1972; Gyllensten et al, 1985). The resulting low frequency of recombination, preserves the integrity and faithful sorting of specific mtDNA haplotypes and increases its utility for population biology and systematics studies (Ballinger et al, 1992; Nei, 1987; Avise, 1994). Lastly, the compact size (ca. 16-17 kb) and relatively conserved gene organization of many animal mtDNAs also facilitated rapid and extensive studies across diverse phyla (Brown, 1983; 1985; Attardi, 1985; Harrison, 1989)(see Appendix C).

The subject of constrained or fluid size and gene content in many animal mitochondrial genomes sets a fundamental cornerstone for the rationale of characterizing mtDNA length variants. In spite of the great morphological and ecological diversity of vertebrate fauna in nature, endothermic vertebrate mtDNAs

appear restricted to sizes within 16.0 - 19.2 kb (Brown 1983; 1985; Roe et al, 1985; Arnason and Johnsson, 1992; Rand, 1993; 1994; Janke et al, 1994; Lee and Kocher, 1995). The double-stranded, circular mitochondrial genome typically codes for only 37 genes specifying 13 proteins, 22 tRNAs, and 2 rRNA subunits (Chomyn et al, 1985; Attardi, 1985; Wolstenholme, 1992). A compact mtDNA organization is further demonstrated by the lack of introns and paucity of short intervening sequences. In *Homo sapiens*, only 7% of the mitochondrial genome is noncoding (Attardi, 1985) and mitochondrial genes often directly abut one another, or overlap, as is the case for ND6 and ATPase 6 and between ND4 and ND4L, and several tRNA genes (Anderson et al, 1981; Bibb et al, 1981). The notion of a static vertebrate molecule with respect to size has barely changed with the continual expansion of the mtDNA database that began with the complete sequences of the human mitochondrial genome (Anderson et al, 1981). A general survey of the current mitochondrial genome database (GenBank release 86.0, 12/94) for vertebrate sequences finds nine mammals (not including cat), chicken, clawed toad (*Xenopus laevis*), and carp (*Cyprinus carpio*). Fig. 1 shows the structure of the domestic cat mitochondrial genome (fully described in chapter 2), which conforms to the structure of other placental mammalian mtDNAs.

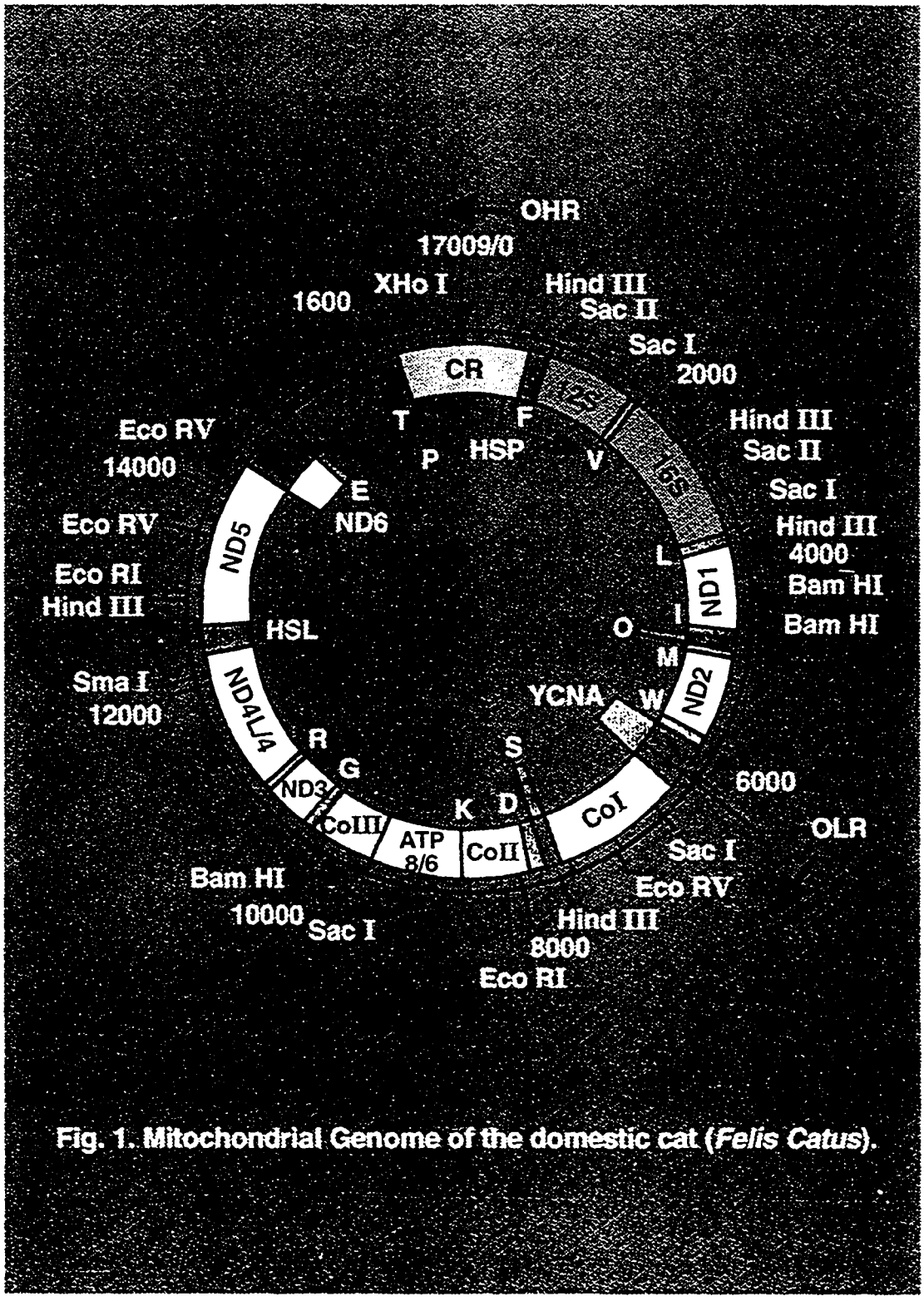


Fig. 1. Mitochondrial Genome of the domestic cat (*Felis Catus*).

Whether vertebrate or mammalian mtDNAs have become "frozen" at the current average of about 17.0 kb poses an intriguing question. For instance, invertebrate, fungal, and plant mitochondrial genomes exhibit a wider range of sizes from 14.3 kb in the nematode worm, *Ascaris suum* (Wolstenholme et al, 1987; Okimoto et al, 1992) to 115 kb in the ascomycetes fungus, *Cochliobolus heterostrophus*, and 2400 kb in the muskmelon, *Cucumis melo* (Gray, 1989a). Fluidity of mtDNA structure also finds support in the examples of length variations and rearrangements of mtDNA. These instances of *heteroplasmy*, a possession of more than one mtDNA genotype in an individual, can occur without a noticeable loss in organismal fitness (Harrrison and Rand, 1985; Hoeh et al, 1991; Wallace, 1992; Zouros et al, 1994). MtDNA length heteroplasmy may be correlated with basal metabolic rates, since it occurs more frequently in lower vertebrate and invertebrate genomes (Monnerot et al, 1984; Bermingham et al, 1986; Rand and Harrison, 1986; Hyman, 1988; Bentzen et al, 1988; Birky et al, 1991; Rand, 1993). For example, the parthenogenic lizard, *Cnemidophorus* exhibits mtDNA length variants caused by tandem duplications in the control region (CR) (Densmore et al, 1985; Moritz and Brown, 1987). More recently, Hoelzel et al (1993; 1994) demonstrated length heteroplasmy in the mitochondrial control region of two populations of elephant seals and various species of carnivores, while DNA deletions may encompass mtDNA coding regions (up to 7 kb) and cause heteroplasmy in human pathological conditions, such as myoclonic epilepsy and ragged red fibers (MERFF) and ocular myopathy, or simple aging (Schon et al, 1989; Wallace, 1992; 1994).

Since mtDNA holds only a minimal repertoire of genetic information, many structural proteins, enzymes and cofactors required for mitochondrial function must be actively transported into the organelle after cytoplasmic translation of nuclear transcripts (Gilham, 1978; Hurt and van Loon, 1986; Clayton, 1984; Hartl and Neupert, 1990; Rose et al, 1992). This paucity of genetic information has been interpreted in the context of the Serial Endosymbiosis Theory or SET (Margulis, 1981; 1993; Smith and Douglas, 1986; Ahmadhian, 1986; Gray, 1989a; 1989b), which proposed that mitochondria and plant chloroplasts (cp) arose from free-living prokaryotes, and organellar genomes are mere vestiges of a more complete ancestral genome. The SET is consistent with the dependence of mitochondrial biogenesis on nuclear instructions and strongly supported by molecular evidence showing a genetic similarity between animal mtDNA and the α subdivision of purple bacteria, which includes *Paracoccus denitrificans* (Yang et al, 1985; Spencer et al, 1984; Fenchel and Bernard, 1993). Primordial hosts to these endosymbionts may have been similar to the giant amoeba, *Pelomyxa palustris* or the mycoplasma-like prokaryote, *Thermoplasma acidophilum* (Smith and Douglas, 1986; Ahmadhian, 1986). An important corollary of the theory, however, is that during the evolution of each symbioses, symbiont (mitochondrial or chloroplast) gene sequences will have gradually transferred and integrated into the nuclear genome. Furthermore, integration of symbiont genes should be followed by their expression from nuclear loci to maintain function of the parent organelle.

Table 1. Examples of nuclear MtDNA previously characterized in non-felid organisms

Organism	MtDNA genes Integrated in Nucleus	Length (kb)	% DNA Diff.*	Time**	Nuc. Express	REP@	Est. Copy #	REA^	Postulated mechanisms of transfer	References
Yeast	Cytochrome b, var 1	1	15-20		Yes			Yes	petite mtDNA, Ty elements	Farrelly and Butow, 1983 Thorsness and Fox, 1990
Neurospora	ATPase subunit	0.183-0.225	35							van den Boogaart et al, 1982
Leguminous Plants	COII	ca. 1.2	15	60-200	Yes		ca.1-4	No	Reverse transcript of RNA intermed.	Nugent and Palmer, 1991
Locust	rRNA	0.8-3.5	n/a			Yes			Repetitive mobile elements	Gellison et al, 1983
Sea urchin	16S rRNA, COI	1	8.0-9.0	>30	low			Yes		Jacobs and Grimes, 1986
Rat	Control Region (CR), rRNA	0.5	20	6.5		Yes	1-?(low)		LINE/retroviral-like	Zullo et al., 1991 Hadler, 1983
Rodent	Cytochrome B	0.366	25-26					No		Smith et al, 1991
Human	ND4, ND5, tRNAs-(H,S,L) 16S rRNA	0.2-1.0	17.0-32.0	43		Yes	130		viral	Fukuda et al, 1985
Human	ND4, ND4L, 12S rRNA COI, D-loop	1.6	7.6-8.0	<9.0				Yes	transfected DNA autophagy	Kamimura et al, 1989
Human	12S rRNA, 16S rRNA	0.4 1.6	16-20	12.0-15.0				No		Nomiyama et al, 1984
Human	16S rRNA	5.0-8						Yes?	Kpn-I family repeat	Wakasugi et al, 1985
Domestic Cat***	CR --> COII	7.9^^ (240-600)	5.1	1.8 -3.0	No	Yes	38-76	Yes	Homologous or random recombination	This study

* - % D/I denotes overall percent difference between nuclear and cytoplasmic mtDNAs
- ** indicates estimated time of divergence between the nuclear and cytoplasmic homologues in million years
- *** Internal reference to this study was added for comparative purposes
^ - REA denotes whether mtDNA sequences were found rearranged in the nuclear genome
^^ - Ind cates that the 7.9 kb size represents only one monomer of the tandem array
@ - REP denotes whether nuclear mtDNA was found in multiple copies. The next column shows the number of copies determined in that study

Numerous studies (Table 1) have established that mtDNA (or cpDNA) can move or "transpose" into the nucleus or between heterologous organelles in various plant, invertebrate and vertebrate taxa (Van den Boogaart et al, 1982; Ellis, 1982, Tsuzuki et al, 1983; Quinn and White, 1987; Kamimura et al, 1989; Fukuda et al, 1985; Zullo et al, 1991; Nugent and Palmer, 1991; Smith et al, 1991). In at least two cases, long interspersed repetitive elements (LINES) or cryptic retroviral sequences were co-isolated with nuclear mtDNA, which suggests a possible vehicle for the transpositions (Wakasugi et al, 1985; Tsuzuki et al, 1983). 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.

During the course of developing the cat as an alternative animal model for genetics research (O'Brien, 1986; Yuhki and O'Brien, 1990; 1994; Gilbert et al, 1991; Brown et al, 1994; Lyons et al, 1994), extra mtDNA restriction fragments were observed by standard Southern hybridization analyses. Consequently, a set of experiments was initiated to characterize and demonstrate a nuclear location for this anomalous mtDNA in *F. catus*. Unlike previous studies of transferred mtDNA, the structure and organization of feline nuclear mtDNA, or *Numt*, appears to be an amplified tandem array of approximately one half of the mtDNA genome (7.9 kb) located on a specific feline chromosome, D2.

Based on the above rationale and discussion, the chapters of this dissertation address various aspects of the *Numt* phenomenon with the following primary objectives:

- Perform a comprehensive molecular characterization of the *Numt* phenomenon in *F. catus*, including the verification of its nuclear integration and location, Mendelian transmission, estimation of copy number and structural organization. (This chapter).

- Determine the complete DNA sequences of *Numt* and the homologous felid cytoplasmic mitochondrial genome to compare and contrast patterns of molecular evolution (e.g. existence of two different rates of evolution). (Chapter 2)

- Confirm whether *Numt* genes are transcriptionally silent, following the rationale of the SET. (Chapter 3)

- Verify and examine the possibility of similar mtDNA transpositions into the nuclear genome of other felid species in the context of Felidae phylogeny. Determine orthology of these events by isolating and characterizing the flanking chromosomal sequences directly adjacent to nuclear mtDNA loci. (Chapter 4)

- Discuss the implications of *Numt* with respect to evolutionary time frame, felid and mammalian genome evolution and possible mechanisms for *Numt* transposition.

MATERIALS AND METHODS

DNA Extraction and Southern Blot Analysis. Genomic DNAs were extracted from fresh lymphocytes or fibroblast 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 (Southern, 1975; Sambrook et al. 1989). Typically, 1.0 - 3.0 μg of DNA per lane were 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 5X Denhardt's solution [0.1% Ficoll, 0.1% polyvinyl-pyrolidone], and 0.1% bovine serum albumin fraction. Hybridization probes were radiolabeled with [^{32}P]-dCTP by random primer synthesis kits (Boehringer Mannheim) and added to filters at a specific activity of 5-10 X 10^8 cpm/μg. After washing with stringent conditions (once at 2X SSC, 0.5% SDS at 37°C, and twice with 1X 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, ten grams of frozen tissue were 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 KCl, 10 mM Tris-HCl, pH 7.5, 1.0 mM EDTA) and homogenized with a tight-fitting pestle in a Potter Elvejhem homogenizer and spun at 1000g for 10 min. and pellets (nuclear fractions) were used to isolate nuclear DNAs. After another centrifugation, supernatant was saved and spun at 20,000g 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, 1 mM EDTA, and EtBr and CsCl were added to bring up the concentration to 500 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 Cytoplasmic (Cyto) mtDNA molecular clones. The initial nuclear mtDNA clone, p*Numt*.1 was derived from a partial library of EcoRI restriction fragments in the 7.9 kb range that was cloned into the λ Zap 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 l 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 4.8-5.0 kb cytoplasmic clone, pCmt4.8, and the 12.0 kb clone, pCmt.12, were similarly isolated from separate EcoRI-digested, size-selected libraries derived from the same cat, FCA 65, and λ Zap and Fix vectors (Stratagene), respectively. Purified mtDNA restriction fragments from puma (*Felis concolor*) were employed as hybridization probes for screening the phage libraries.

Nucleotide Sequence Analyses. The p*Numt*.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 p*Numt*.1 DNA sequence or from literature sources for the following genes: tRNA-Q,Mf and NADH 2 -[5'GCATCCCACCTCAAACGT3'/5'GTGTAACTTCGGGCAC3']; CO I -[5'GATAGGATCTCCTCC3'/ 5'GGTGCCCCTGACATAG3']; and 12S rRNA (Kocher et al. 1989) -[5'AAAAAGCTTCAAACTGGGATTAGATACCCCACTAT 3'/ 5'TGACTGCAGAGGGTGACGGGCGGTGTGT3']; other cytoplasmic mtDNA gene sequences were obtained from sequencing M13 subclones. DNA sequencing reactions were performed by the dideoxy 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 (UWGCG) (Devereux et al. 1984; UWGCG Manual, 1994), Phylogeny Inference Package (PHYLIP) (Felenstein 1993), and Phylogenetic Analysis Using Parsimony (PAUP) (Swofford 1990).

Pulsed Field Gel Electrophoresis (PFGE). High molecular weight DNA's were prepared according to the procedures previously described (Sambrook et al. 1989). Typically, about 5-10 μg of genomic DNA were contained in each plug. Electrophoresis of restricted DNA plugs was carried out in a hexagonal CHEF PFGE apparatus at 6.5 V/cm in 0.5 X TBE (25 mM Tris-borate, 0.5 mM EDTA) for 36-48 h at 12°C. Switch times of 60-90 sec were set to provide optimal resolution in the 50 - 900 kb size range (Birren et al. 1988). Yeast (*Sacchromyces cerevisiae*) chromosomal DNAs and concatenated λ phage DNAs (Pharmacia LKB) were run as standard molecular weight markers. After the electrophoresis, the gel was treated for Southern transfer as previously described (Birren et al. 1988).

Fluorescent In Situ Hybridization (FISH) to Feline Chromosomes. The protocol for FISH has been previously described (Hoehe et al. 1991). The *Numt* DNA probe was labelled by nick translation with biotin-11-dUTP. Hybridization to metaphase chromosomes isolated from primary feline fibroblast cell cultures was carried out in a

10 μl volume of 50% formamide and 10% dextran sulphate in 2X SSC at pH 7.0 for 16 h at 37°C. Washes were performed once at 40°C in 50% formamide in 2X SSC, then three times in 2 X SSC only. Chromosomes were identified by QFH (Quinicrine Fluorescent Hoescht) banding. Photographs were taken with Kodak Ektachrome 400 ASA color slide film or EKTAR 1000 color print film.

Somatic Cell Hybrid Gene Mapping Panel. The generation and genetic characterization of the somatic hybrid panels, Chinese hamster X domestic cat and mouse X domestic cat, have been previously described (O'Brien and Nash 1982). These hybrid cells retain all rodent chromosomes but only some cat chromosomes in different combinations. DNAs from 41 hybrid cell lines used in the segregation analyses of *Numt* DNA were isolated and prepared for Southern hybridizations as described above.

PCR Amplification of Junction Regions between *Numt* Monomers. As a test for the tandem arrangement of *Numt* 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 *Numt* locus. The primer pair (J1: 5'- AACTGGGACGTGGGG-3' and J2: 5'-GCTCACGCACACAAG-3') was derived and oriented towards the extreme 5' and 3' termini of the p*Numt*.1 clone. The positions of each primer are indicated below the map in Fig. 5. Fragments in the expected size range of approximately 720 bp were then amplified by standard PCR protocols (Innis et al.

1991) using these primers and total genome DNA templates from several domestic and wild cat individuals.

Estimation of *Numt* copy number. A reconstruction experiment was performed using known amounts of the p*Numt*.1 clone and FCA 65 genomic DNA. First, the proportion (P) of *Numt* DNA in the felid haploid genome was calculated with the following equation,

$$n/G = P$$

where n is one copy of the 7.9 kb *Numt* repeat, and G is the size of the cat genome, estimated to be on the order of 3 X 10^9 bp. From this equation, 2.6 X 10^{-6} or 0.0003% of cat genomic DNA is represented by exactly one monomer of *Numt*. To find the relative weight of *Numt* in any given sample of DNA, P is multiplied by the weight of DNA (W). For the present case, 5.0 ug (5000 ng) genomic DNA X 0.0003% yields 0.015 ng as the weight of one copy of *Numt*. Weights for different copy number equivalents (0.1, 1, 10 , 20 and 50) of *Numt* were based on this W value and subsequently derived by appropriate serial dilutions of p*Numt*.1 DNA. The diluted p*Numt*.1 samples were combined with about 5.0 ug of EcoRI digested carrier *E.coli* genomic DNA and then loaded onto agarose gels for standard electrophoresis and Southern hybridization alongside 5.0 ug of cat genomic DNA digested with Eco RI

restriction enzyme. DNA concentrations were checked before restriction digests and electrophoresis with a TKO 100 microfluorometer (Hoefer) or DNA spectrophotometer (Beckman). The gel was blotted and hybridized to the complete p*Numt*.1 clone, following standard conditions (Sambrook et al, 1989). After development of the autoradiograph, bands corresponding to the 7.9 kb *Numt* were identified on the autoradiogram, and then cut from each lane representing a specific copy number equivalent or the genomic DNA sample on the original nylon filter. Radioisotopic signals (CPM) were quantified by liquid scintillation counting for each p*Numt*.1 dilution and used to produce a standard curve. Counts detected in the genomic fragment were interpolated to this curve to infer the copy number in the test cat.

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. 2A). 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 $\pm$ 200 bp has been found for nearly all mammalian mtDNAs including at least one felid 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 digests was evident in the nuclear and in total genome DNA preparations but absent in the supercoiled cytoplasmic DNA fraction (Fig. 2A). Extra fragments were also observed in digestions with other restriction enzymes (Fig. 2B) and in several cats (Fig. 2C). The restriction patterns of total genomic DNA were interpreted in a restriction map (Fig. 2D) that reflects the tandem repeat of *Numt* in nuclear DNA (see below). Because mitochondrial genomes are present in high copy number (ca. 10^2-10^4) 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 fragments were also present in multiple copies. Moreover, the ratio of the intensities between nuclear and

Fig. 2 (A) MtDNA restriction patterns observed in fractionated preparations of domestic cat DNAs. DNA fractions were total genomic DNA (T), nuclear (N) and cytoplasmic (C) as described in Experimental Procedures. Total mtDNA fragments were detected with a ^{32}P-labelled lambda clone, $\lambda^{3\text{-}2}$, containing a complete *F. catus* mtDNA genome (O'Brien et al., 1990). Band sizes are in kilobase pairs (kb). The (*) mark nucleus-specific mtDNA fragments. (B) Restriction patterns of *Numt* (labelled 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 p*Numt*.1. (C) Nuclear mtDNA bands can be visualized in several unrelated cats. 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 p*Numt*.1 (see text).

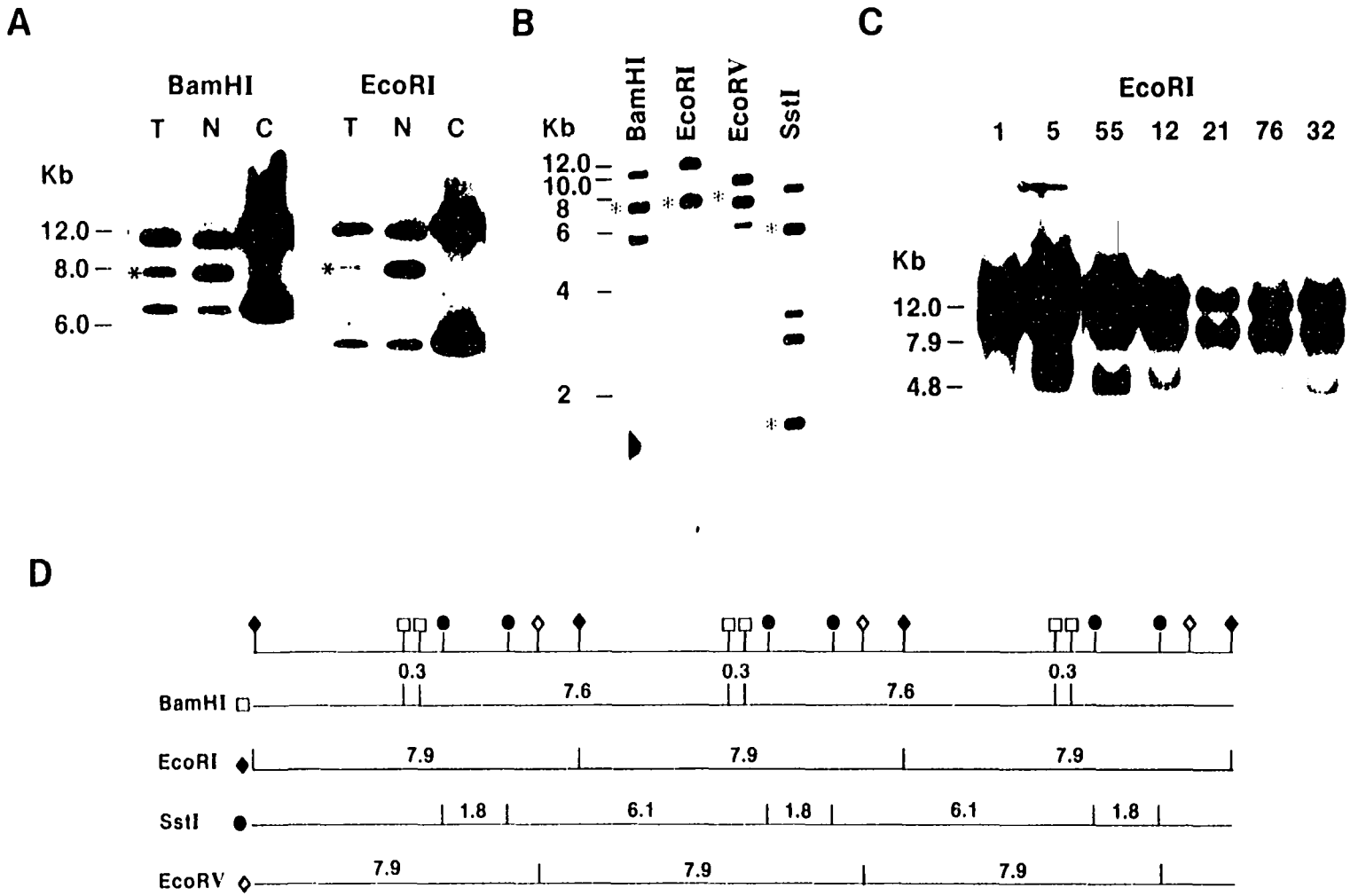


Figure 2. Mitochondrial restriction enzyme patterns observed in fractionated preparations of *F. catus* DNA.

cytoplasmic mtDNA bands varied in different individuals as did the intensity of different cytoplasmic mtDNA fragments (Fig. 2C). 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 a nuclear residence of the extra mtDNA sequences, we examined a panel of genetically characterized rodent X 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 nineteen feline chromosomes. Genomic DNA from 41 hybrids were scored for the occurrence of feline mtDNA fragments after digestions with SstI (Fig. 3A). 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%), with the presence of feline chromosome D2 (92%), and reciprocally discordant (26-53%) with each of the other cat chromosomes (Fig. 3B). These results affirm the nuclear location of *Numt* and implicate its position on chromosome D2. This conclusion was confirmed by florescence *in situ* hybridization (FISH) using a molecular clone of the nuclear mtDNA (p*Numt*.1; see below) to metaphase chromosome preparations from the domestic cat. The results of this analysis revealed a prominent hybridization signal near the centromeric region of chromosome D2 (Fig. 4).

Fig. 3. Segregation of *Numt* with specific cat chromosomes in a panel of rodent X cat somatic cell hybrids (O'Brien and Nash, 1982). (A) Pattern of mtDNA hybridization in SstI digest of genomic DNA from Chinese hamster X cat somatic cell hybrids probed with a feline mtDNA molecular clone. The 6.1- and 1.75-kb feline fragments represent *Numt* 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 *Numt* fragments with the 19 feline chromosomes segregating in the hybrid panels.

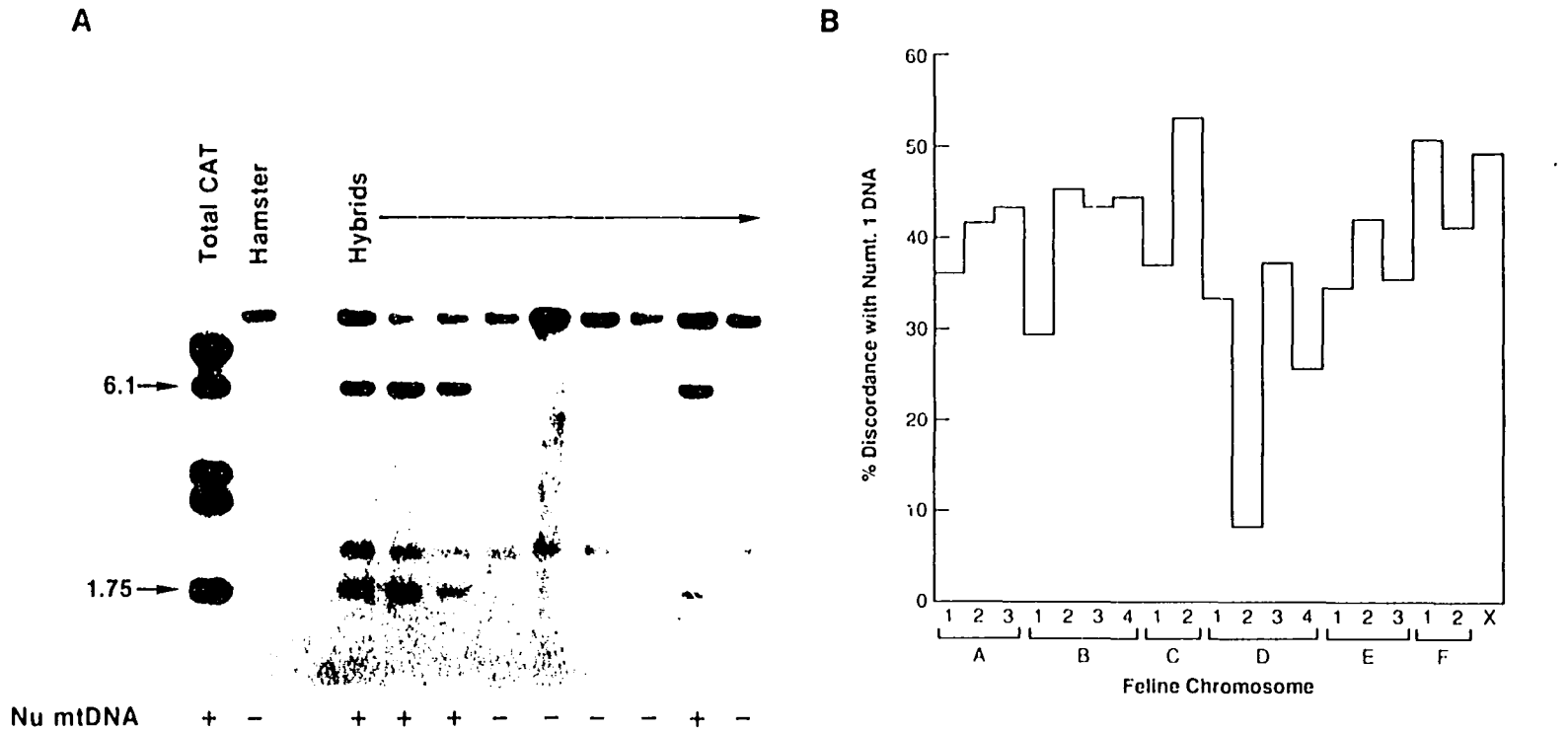


Figure 3. Segregation of *Numt* with specific cat chromosomes in a panel of somatic cell hybrids.

Fig. 4. 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. Slide preparation, and FISH results were kindly performed and provided by Dr. William S. Modi and Mary Eichelberger.

Molecular Cloning of Nuclear Mitochondrial DNA (Numt) and Cytoplasmic Mitochondrial DNA from F. catus

Seven clones were chosen from a size-selected EcoRI genomic library made with cat FCA 65 DNA and λ Zap II; one clone, designated p*Numt*.1, contained the predicted size for *Numt*, 7.9 kb (Fig. 2A) and was sequenced and characterized in detail (Fig. 5). In addition, two cytoplasmic mtDNA clones (pCmt.12 and pCmt4.8) were also recovered from FCA 65 and represent the two EcoRI digest fragments (12.0 and 4.8 kb respectively) shown in Figs. 2B and 5.

The gene content and order of the feline pCmt.12 clone and the nuclear mtDNA clone, p*Numt*.1, were determined by restriction mapping, by complete nucleotide sequencing of p*Numt*.1 and partial sequencing of pCmt.12 (Fig. 5). Restriction maps and sequence alignment of p*Numt*.1 and pCmt.12 were sufficiently conserved to identify all the mitochondrial gene segments present in the two constructs (Fig. 5). The p*Numt*.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, Norway rat and harbor and grey seals, and blue and fin whales (Anderson et al, 1981; Anderson et al, 1982; Bibb et al, 1981; Gadaleta et al, 1989; Arnason and Johnsson, 1989; Arnason et al, 1993; Arnason and Gullberg, 1993). The p*Numt*.1 clone represents a truncated and moderately divergent homologue of the mitochondrial genome that has

Fig. 5. (A) Comparison of restriction maps and gene content of *F. catus* nuclear (p*Numt*.1) and cytoplasmic mtDNA clones. Gene sizes are drawn to a scale based on the complete p*Numt*.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 p*Numt*.1 map represent regions of comparison which have been sequenced in both p*Numt*.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 p*Numt*.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 were verified by nucleotide sequencing. J1 and J2 indicate the primers used to amplify across the unique DL junction in *Numt*.

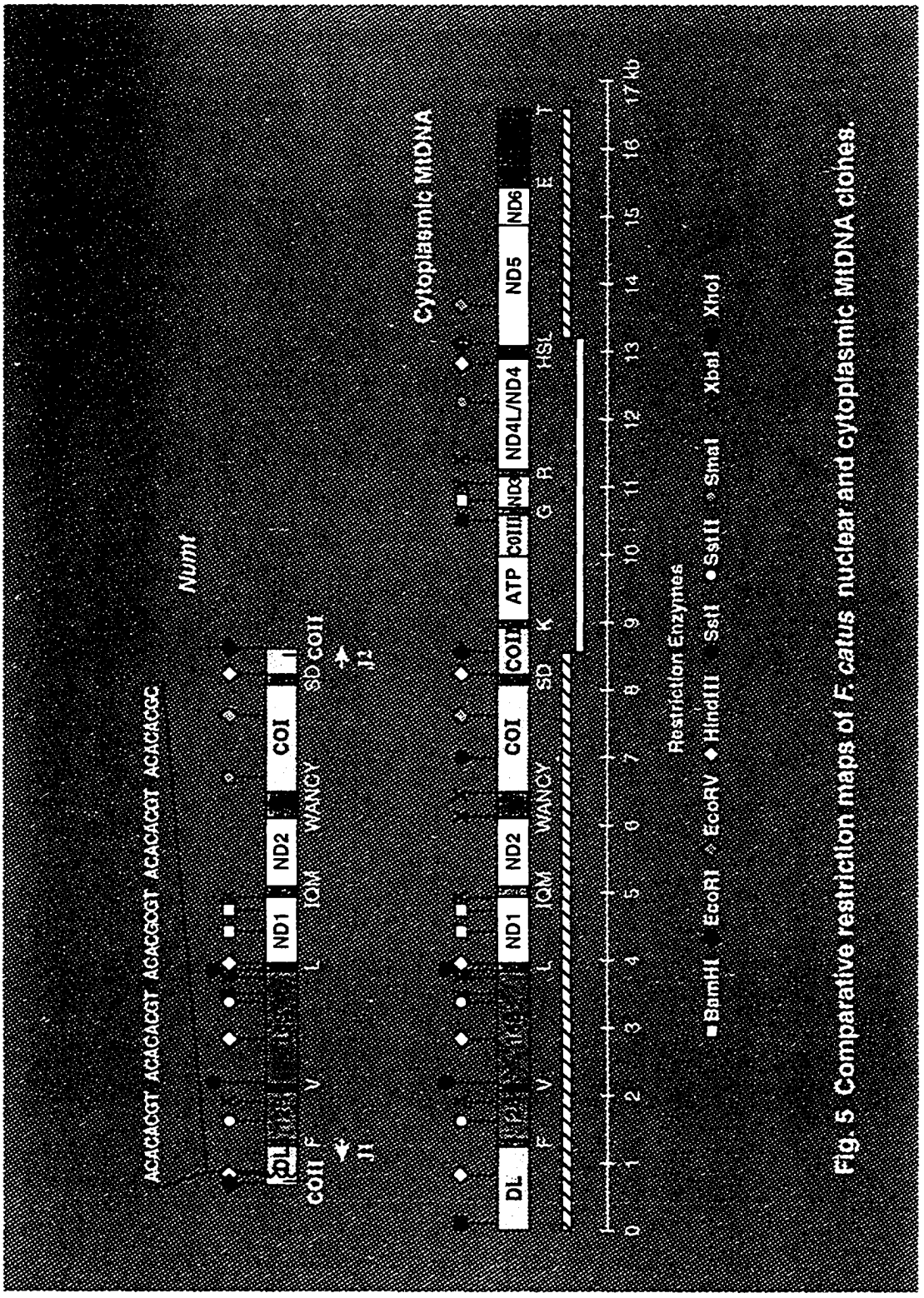


Fig. 5 Comparative restriction maps of *F. catus* nuclear and cytoplasmic MtDNA clones.

a COII gene interrupted downstream from the EcoRI site present in both cytoplasmic mtDNA and in *Numt* (Fig. 5). Contiguous regions of the mtDNA COII gene are found at either end of the EcoR1 site that defines the termini of the p*Numt*.1 molecular clone, suggesting a circular precursor of the 7.9-kb nuclear segment. The p*Numt*.1 COII sequence (552 bp) is foreshortened relative to cytoplasmic mtDNA COII genes in cat and other species (676 bp). COII in *Numt* is fused to about 200 bp of a truncated segment homologous to the mammalian mtDNA control or D-loop region (DL) in Fig. 5. 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 p*Numt*.1 with pCmt.12 and other DL regions revealed that p*Numt*.1 contained two segments that are very similar to regulatory conserved sequence blocks (CSBs) found in mammalian cytoplasmic mitochondrial DNA and shown in Fig. 6 (Chang and Clayton, 1984; Wallace et al, 1991). An origin of replication was not apparent in p*Numt*.1. There was a notable 8-bp motif, (ACACACGT), which was imperfectly repeated 5 times at the junction between COII and the D-loop control region (Fig. 5 and 6). 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; Hoelzel et al, 1994) and thus may be a candidate region for recombination between cytoplasmic and nuclear genomes.

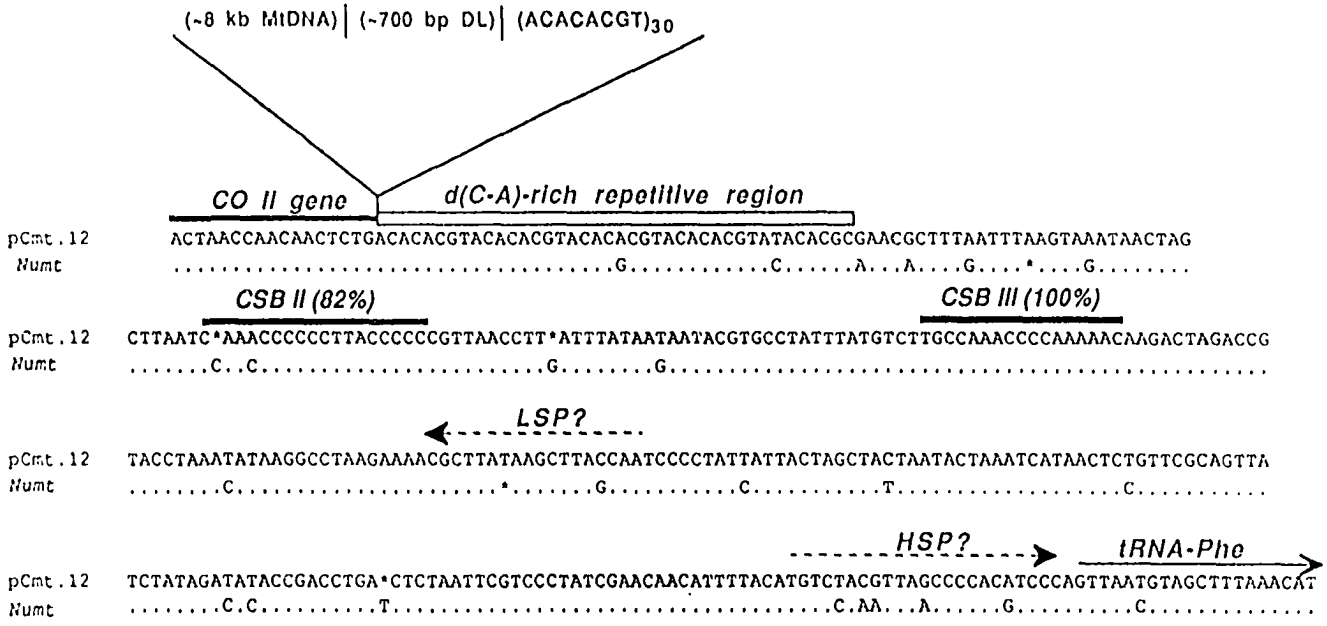


Fig. 6. Close-up view of internal *Numt* junction region after alignment of *Numt* and cytoplasmic mitochondrial D-loop (DL) sequences. The CR region in p*Numt*.1 is shortened relative to pCmt.12 by about 1 kb (700 + 240 bp) as indicated above the sequence alignment. The pCmt.12 clone contains a longer d(CA) repeat (240 bp) region, the remainder of the mitochondrial genome and 5' DL. Percent identify with human CSBs are 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 d(CA)-rich motif occurs at the deletion junction between COII and DL sequences in *Numt* (see text).

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

To account for the relatively high copy number of *Numt*, we postulate a tandemly repeated arrangement for p*Numt*.1 which includes the 7.9-kb fragment (Fig. 5) as the basic repeating unit. This hypothesis is supported by the hybrid panel and fluorescent in situ hybridization (FISH) results, which preclude episomal or chromosomally dispersed copies of *Numt* (Fig. 4) and the aforementioned splitting of COII gene sequences between the 5' and 3' termini of p*Numt*.1. The consistent size of *Numt* fragments (7.9 kb) with different restriction enzyme digestions plus our ability to produce a restriction map from genomic DNA (Fig. 2D) that was nearly identical to the p*Numt*.1 restriction map (Fig. 5) are also consistent with a tandem array of *Numt* sequences.

To confirm a tandem arrangement, we digested high molecular weight genomic DNA from several cats with six base-pair cutting restriction enzymes whose target sites were not present in the p*Numt*.1 DNA sequences. The restriction products were separated by pulse field gel electrophoresis (PFGE) and probed by Southern analysis for *Numt* DNA sequences. The genomic fragments were recovered and their estimated molecular size is illustrated in Fig. 7. The PFGE results revealed high molecular weight *Numt* specific fragments ranging in size from 300-600 kb, two orders of magnitude greater than the mean fragment size expected from a six base-pair cutting restriction enzyme (1 site per 4096 bp) and about twenty times the average mammalian

Fig. 7. PFGE analysis of restriction enzyme-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 seconds. The ^{32}P-labelled p*Nunu*.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 sec. 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 sec, respectively. The designation of nine proposed allelic states (A-I) is indicated alongside the band sizes.

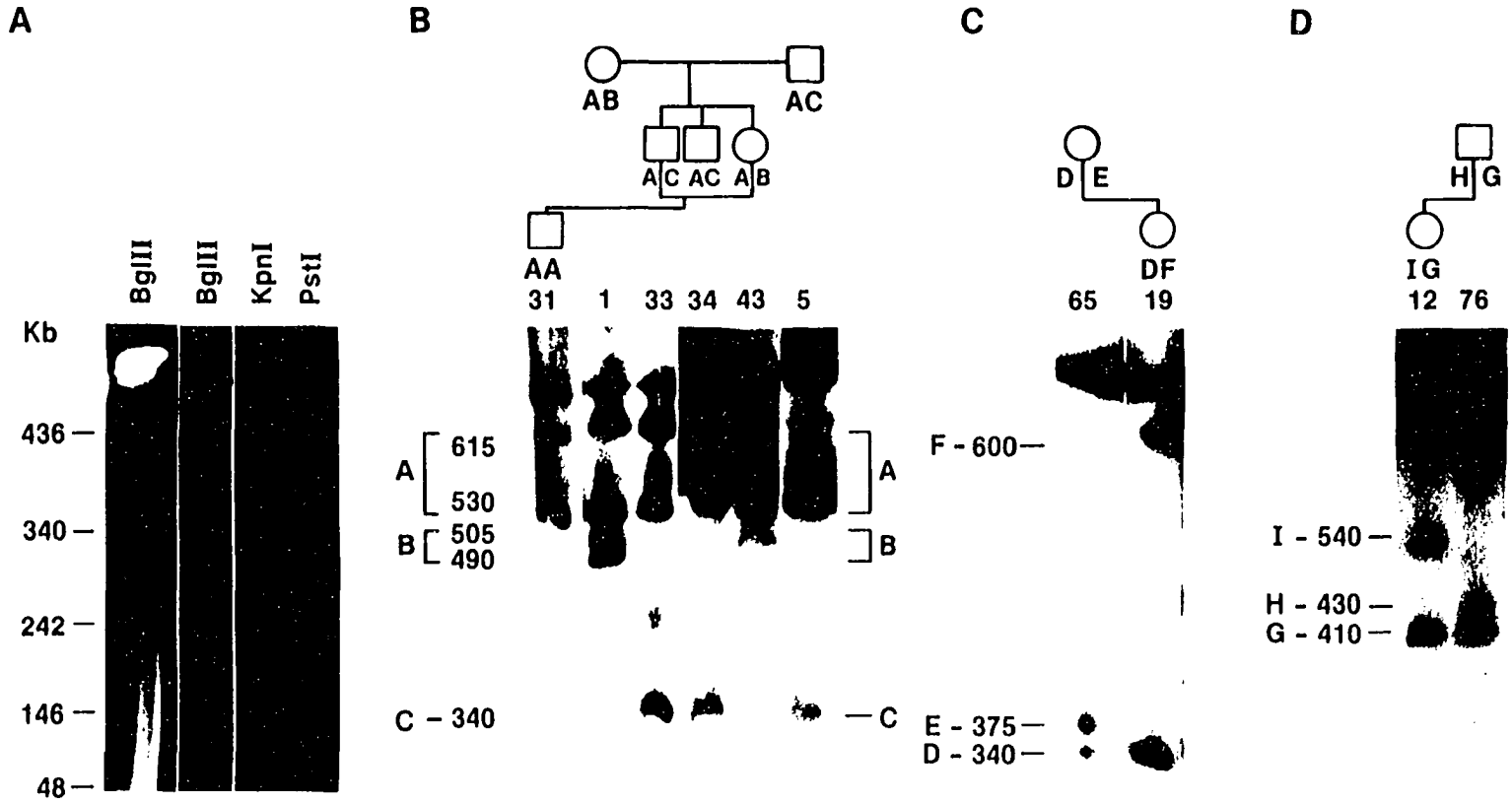


Figure 7. Pulsed Field Gel Electrophoretic analyses of *F. catus* genomic DNA.

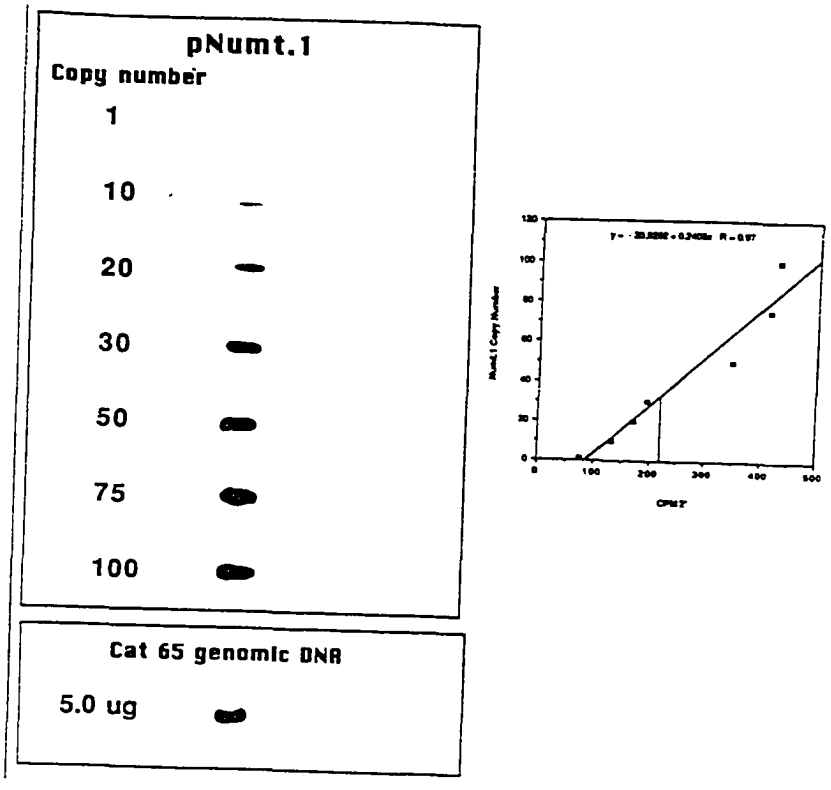


Fig. 8. Reconstruction experiment to estimate *Numt* copy number in FCA 65. Intensities of DNA bands of p*Numt*.1 copy number equivalents derived and cut from a Southern blot are shown on the left. Radioisotopic signals from each band were counted by liquid scintillation on a Beckman LS8100 counter and graphed at the right. Extrapolation of the hybridization signal (ca. 220 cpm) from 5.0 ug of genomic DNA indicated that the copy number of *Numt* in this cat was about 30.

mtDNA size. If the *Numt* locus were represented by a tandem array of 7.9-kb repeats like the p*Numt*.1 clone, the number of repeats would range from 38-76 copies.

To verify whether the multiple bands represent variant length alleles of *Numt* loci, pedigree analyses was also performed with PFGE. The high molecular weight *Numt* DNA fragments were polymorphic among the three families (Fig. 7B-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. 7B, at least three alleles (designated A, B, and C) are segregating in the three generations. Additional cat pedigrees (Fig. 7 C,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 38-76 copies. Furthermore, estimations of about 30 *Numt* copies from a reconstruction experiment using FCA 65 DNA also agree with the PFGE data (Fig. 8).

Comparative Sequence Analyses of Cytoplasmic and Nuclear mtDNAs

Nucleotide sequences were obtained from ten gene segments of p*Numt*.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 p*Numt*.1 and pCmt.12 sequenced regions shown in Fig. 9 (A-E) comprise the three different types of genes (rRNA, tRNA, and protein coding) found in most mitochondrial genomes, plus the control region (CR). For both p*Numt*.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. 9 and in Table 2.

Most (67/81) of the mutations between the p*Numt*.1 and pCmt.12 structural genes are synonymous and occur within non-conserved regions found in previous comparisons (Anderson et al, 1982). This indicates the presence of functional constraints against non-synonymous amino acid replacement substitutions (Kimura, 1983; Nei, 1987; Li et al, 1985). For instance, all three first position changes in the

Fig. 9. DNA sequence alignment of feline nuclear, p*Numt*.1, and cytoplasmic, pCmt.12, clones with homologous gene sequences from other mammal species. Non-felid 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 directly below the FCA pCmt.12 clone 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 sequences 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. libyca* (African wild cat); Fma-*F. margarita* (sand cat); Fni-*F. nigripes* (black-footed cat); Fch-*F. chaus* (jungle cat); Oma-*Otocolubus manul* (Pallas cat); Ogu-*Oncifelis guigna* (kodkod); Ple-*Panthera leo* (lion). Primer and template preparation, and phylogenetic analysis is given by Masuda et al., (in preparation). Other abbreviations and notation are as in (A). (C) tRNA sequences: the cytoplasmic sequences of Leu, Gln and Met_f-tRNAs were derived from asymmetric PCR products amplified from the pCmt.12 template. (D,E) ND1, ND2, COI and COII gene sequences. The deduced peptide sequences after translation of open reading frames in cytoplasmic pCmt.12 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).

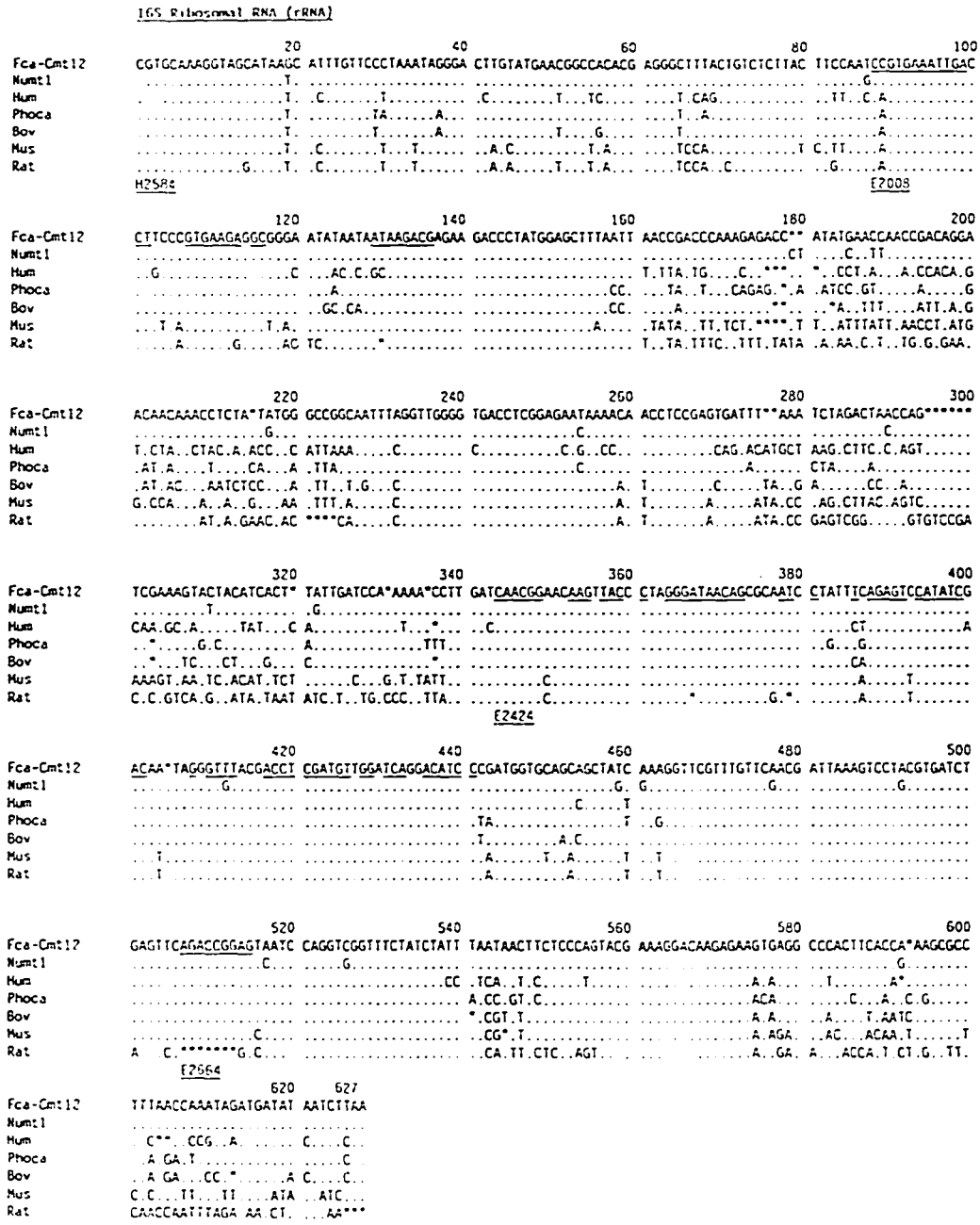


Figure 9. (Continued)

B.

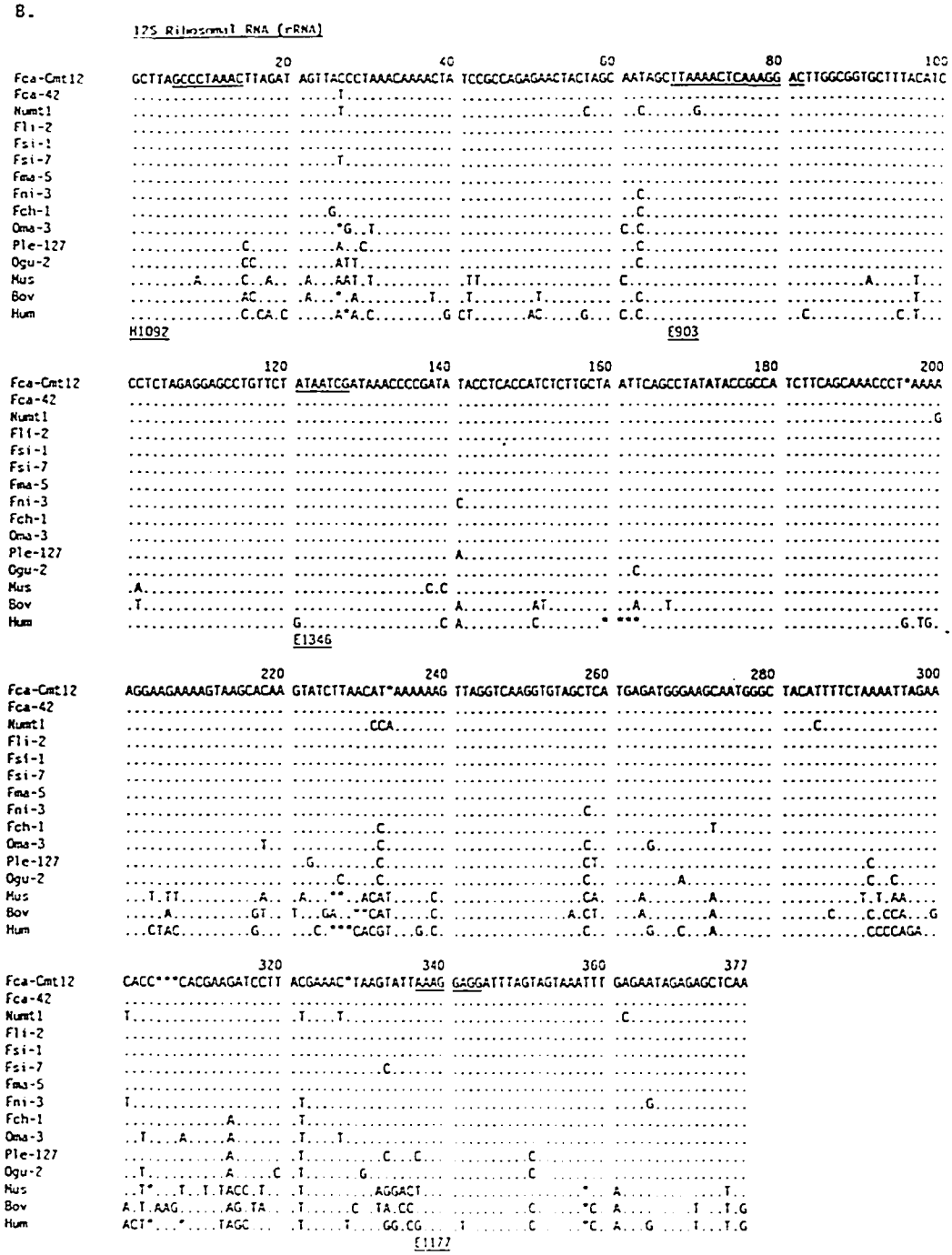


Figure 9. (Continued)

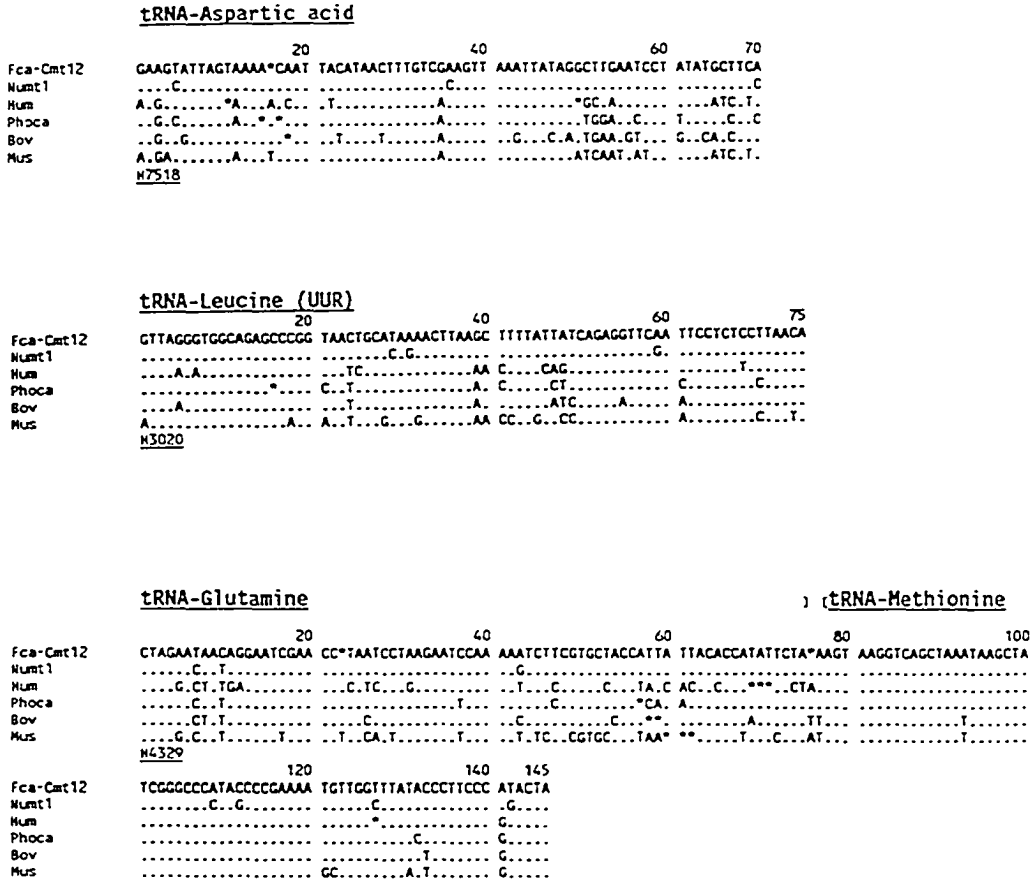


Figure 9. (Continued)

NADH Dehydrogenase 1 (ND1)

```
                    20                  40                  60                  80                 100
                    M  F  M  I  N  V  L   S  L  I  I  P  I  L   L  A  V  A  F  L   T  L  V  E  R  K  V   L  G  Y  M  Q  L  R
Fca-Cmt12           ATGTTTATAATTAATGTACT CTCACTAATTATTCCTATTC TCCTAGCTGTAGCCTTCCTA ACCCTAGTCGAACGAAAAGT GCTAGGCTATATGCAACTTC
Numt1               .....C...G..........  ..................G. ....C.............. ...T....T........... ..................C.
Hum                 ..ACCC..GGCC..CC.C.. ACTC..C...G.A..C.... .AA.C..AA.G..A..... .TG..TAC..........A. T...........A.....A.
Phoca               ..............A.CA.  ............C..A.... ....C..C.....T..... ..AT....A.....G..... A........C..A.....C.
Bov                 .....C........CA.CT. AAT.............C..C. .AT.G..C.....A.....T ..GT....G.......... T...........A.....C.
Mus                 G....CT.T......A.C.. AA....CC.CG.C..C.... .AA.C..CA.......... ..AT....A.....C...A. CT....G..C..A.....A.
                    H3307

                                  120                 140                 160                 180                 200
                      K  G  P  N  V  V   G  P  Y  G  L  L  Q   P  I  A  D  A  V  K   L  F  T  K  E  P   L  R  P  L  T  S  S
Fca-Cmt12           GCAAAGGACCAAATGTCGTA GGACCATACGGCCTACTTCA ACCTATCGCAGATGCTGTAA AACTCTTTACCAAAGAGCCT CTCCGACCCCTTACATCCTC
Numt1               .T.................. ............T.G..C.. ............C....... .......C........A... ...........C........
Hum                 .......C..C..C..G... ..C..C.....G.....A.. ...CT....T..C..CA... .......C...........C ..AAA....GCC.....TA.
Phoca               .A........C..CA..... .....C.....A..C..A.. ...C........C..C.... ....A..C...........C ..A.....A..A.....T..
Bov                 .A.....T............ ..T.....T........... ...C.....C.....AA.C. ....T..C.TT.....A..A ..A......GC......T..
Mus                 .A.....C..T..CA.T..T ..T.........A.TT.A.. ...AT.T.....C..CA... ..T.A....TA.....A..A A.A..C..TT.A...A....

                                  220                 240                 260                 280                 300
                      M  L  M  F  I  M  A   P  I  L  A  L  T   L  A  L  T  M  W  I   P  L  P  M  P  Y  P   L  I  N  M  N  L
Fca-Cmt12           CATATTAATATTCATCATAG CACCAATCCTAGCCCTCACA CTAGCCCTAACCATATGAAT CCCACTACCCATACCATACC CGCTCATTAACATAAACCTG
Numt1               ..............T.... ..................... ..................... ..................... .A.................A
Hum                 ...CACCC.C.A.....CC. .C..G.C.T....T.....C A.C..T..TCTAC......C ...C..C........CA... .C..GG.C...C.C.....A
Phoca               ..C.ACC........T..G. .C..T..T......T.AG.C ...........T......G. ...C..G.............. .C.....C...........A
Bov                 AGCC.C......T...C... ....T...A....TT.AGGC ......T.......G..... T..C.....A.....C..T. .T..T..C...........A
Mus                 T....CCT....T..T..T. ....T.C....T.A..... .....AT...GTC.....G. T..C.....A......C... .AT.A.....TT.......A

                                  320                 340  345
                      G  V  L  F  M  L  A   M  S  S  L  A  V  Y   S
Fca-Cmt12           GGAGTACTATTTATGCTAGC TATATCAAGCCTAGCTGTTT ATTCC
Numt1               ......T.......A..G.. ............G..C..C. .....
Hum                 ..CC.C.........T..... C.CC..T.........C.... .C..A
Phoca               .....GT....C..A..... A.................C. .C..A
Bov                 .....C........A..... C..............C..A. .C...
Mus                 ..GA.TT.......TT.... A.C...T......T.A.... .C...
```

Figure 9D. (Continued)

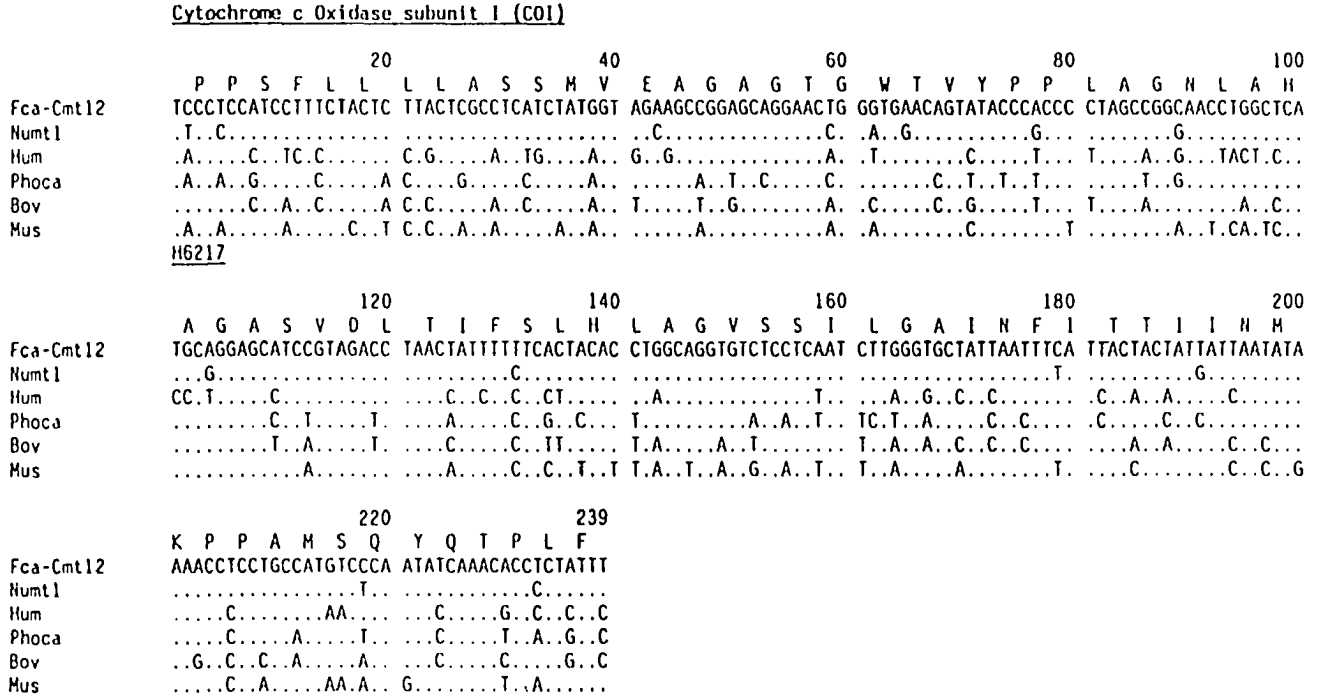


Figure 9E. (Continued)

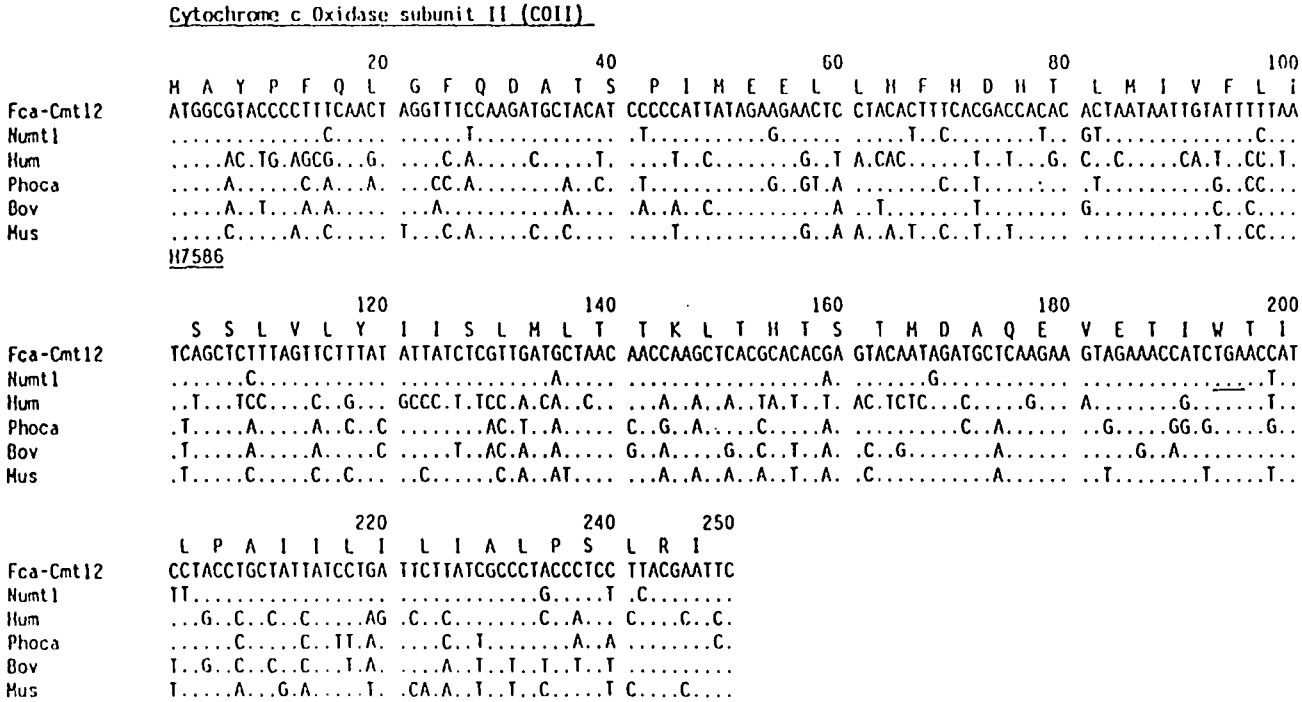


Figure 9. (Continued)

Table 2. Sequence comparison of feline cytoplasmic and nuclear mtDNA and between mammalian mtDNAs.

		Types of Changes btw. N and C				Codon Position*			Percent Difference (PD)***									
									Nucleotides (Nt)							Amino Acids		
Gene	Nt	Subst.	Ti	Tv	Gaps^	1	2	3	N/C	N/H	N/B	C/B	C/H	M/R	C/P	C/N**	C/B	C/H
Protein																		
CO-I	239	14	10	4	0	0	1(1)	13(2)	5.9	23.0	24.7	20.9	21.8	16.7	19.7	2.6	0.0	7.6
CO-II	250	20	20	0	0	3(0)	1(1)	16 (0)	8.0	35.2	21.9	19.1	32.4	15.0	20.0	1.2	1.0	27.0
ND-1	345	24	22	2	0	4(1)	0	20(1)	6.9	31.4	21.9	21.3	30.3	19.3	17.4	1.8	8.7	27.0
ND-2	326	23	18	5	2	5(4)	1(1)	17(3)	7.8	37.4	25.9	24.2	36.9	25.1	23.6	14.7	29.3	46.2
RNA																		
12S rRNA	371	12	10	2	1				3.5	21.0	16.4	15.6	20.0	6.9	14.3			
16S rRNA	611	18	10	8	1				3.1	21.6	15.7	14.1	20.8	23.2	12.0			
tRNA-Asp	69	3	1	2	0				4.3	21.7	30.4	26.1	17.4	11.9	17.6			
tRNA-Leu	75	3	3	0	0				4.0	18.7	14.7	10.7	14.7	4.2	12.0			
tRNA-Gln	74	3	3	0	0				4.0	27.8	18.2	11.0	30.6	5.6	10.8			
tRNA-Met	69	4	1	3	0				5.8	7.2	10.1	4.3	2.9	4.3	2.9			
Totals	2429	124	98	26	4				5.3^^									

Abbreviations: N-Numt.1; C-Cmt.12; B-bovine; P - phocine (seal); Ti-transitions; Tv-transversions

*Number of nonsynonomous substitutions (subst) are in parentheses

**Translation of Numt.1 sequence with the mitochondrial genetic code

***-Total PD counts gaps as single residue difference. Gaps are counted only in the FCA pNumt.1 and pCmt.12 DNA comparison.

^ - Nucleotide count is based only on the number read in FCA mtDNA.

^^The mouse (M) and rat (R) comparison encompasses regions identical to those compared in the cat.

COII gene of p*Numt*.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 2 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).

Phylogenetic Origin of Numt Transposition

We used 12S rRNA gene sequence data (Woese, 1987; Mindell and Honeycutt, 1990) for placing *Numt* DNA within an evolutionary framework among the Felidae. The six non-domestic cat species directly following the *F. catus* 12S rRNA sequences in Fig. 9B 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 phenetic (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. 9B. The results of these analyses, illustrated by the maximum likelihood phylogenetic tree in Fig. 10, 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. libyca*, *F. silvestris*, *F. margarita*, *F. nigripes*, *F. chaus* and *Otocolubus manul*) and indicated a hierarchical 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 *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*. 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.

Fig. 10 Phylogenetic analysis of the 12s rRNA gene (Fig. 9B) 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 ten species plus *Numt* using the DNAML algorithm of PHYLIP (Felsenstein, 1990; Felsenstein, 1981). The routine employed empirical base frequencies from the data set, a transition:transversion ratio of 5.0, and randomized input order. The log likelihood of the tree was -902.65. Each terminal node except the unresolved (*F. catus*, *F. sylvestris*, and *F. libyca*) group has limb lengths significantly ($p \leq 0.01$) greater than 0. Scale is percent nucleotide sequence differences between species; limb lengths are the same units. The same data was 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 (Saitou and Nei, 1987; Fitch and Margoliash, 1967; Felsenstein, 1993). 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 two criteria: 1) presence of mtDNA fragments in Southern blots of genomic DNA exceeding 22 kb; and 2) presence of PCR-generated fragments that used primers from *Numt* COII and D-loop sequences spanning the junction between repeat segments.

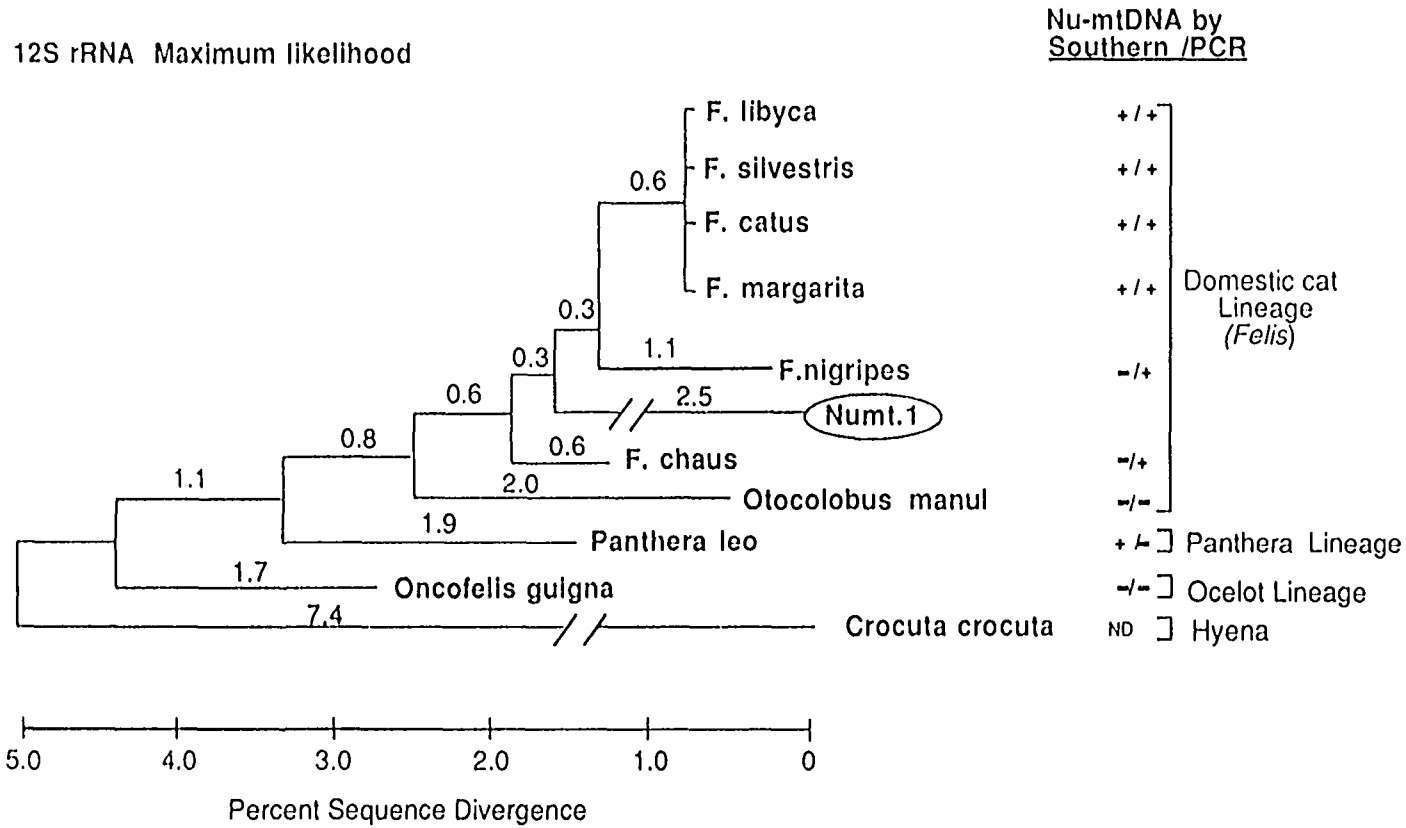


Figure 10. Phylogenetic analysis with 12S rRNA gene sequences.

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. 10). The tandem expansions of *Numt* are apparently still occurring through a process reminiscent of minisatellite expansion/contraction as multiple length alleles (Fig. 7) 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 X 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 by the tandemly repeated

arrangement of *Numt*, its very recent origin, and 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 due to organelle fusion (Hayashi et al, 1985), intramolecular recombination is more common and has been implicated in the origin of human mitochondrial pathologies (Schon et al, 1989; Wallace, 1992; 1994). An alluring hypothesis (Fig. 11) would parallel the origin of *rho*$^-$ or "petite" mtDNA mutations in the yeast, *Sacchromyces cerevisae* (Locker et al, 1979; Dujon and Belcour, 1989). The mtDNAs in these respiratory-deficient organelles experienced *intra*molecular recombination, causing large deletions of the mitochondrial genome. The remnant mtDNA becomes amplified extrachromosomally and persists as an episome. If the ancestral *Numt* underwent 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 extra-nuclear concatonization occurs in transgenesis experiments (Capecchi, 1989).

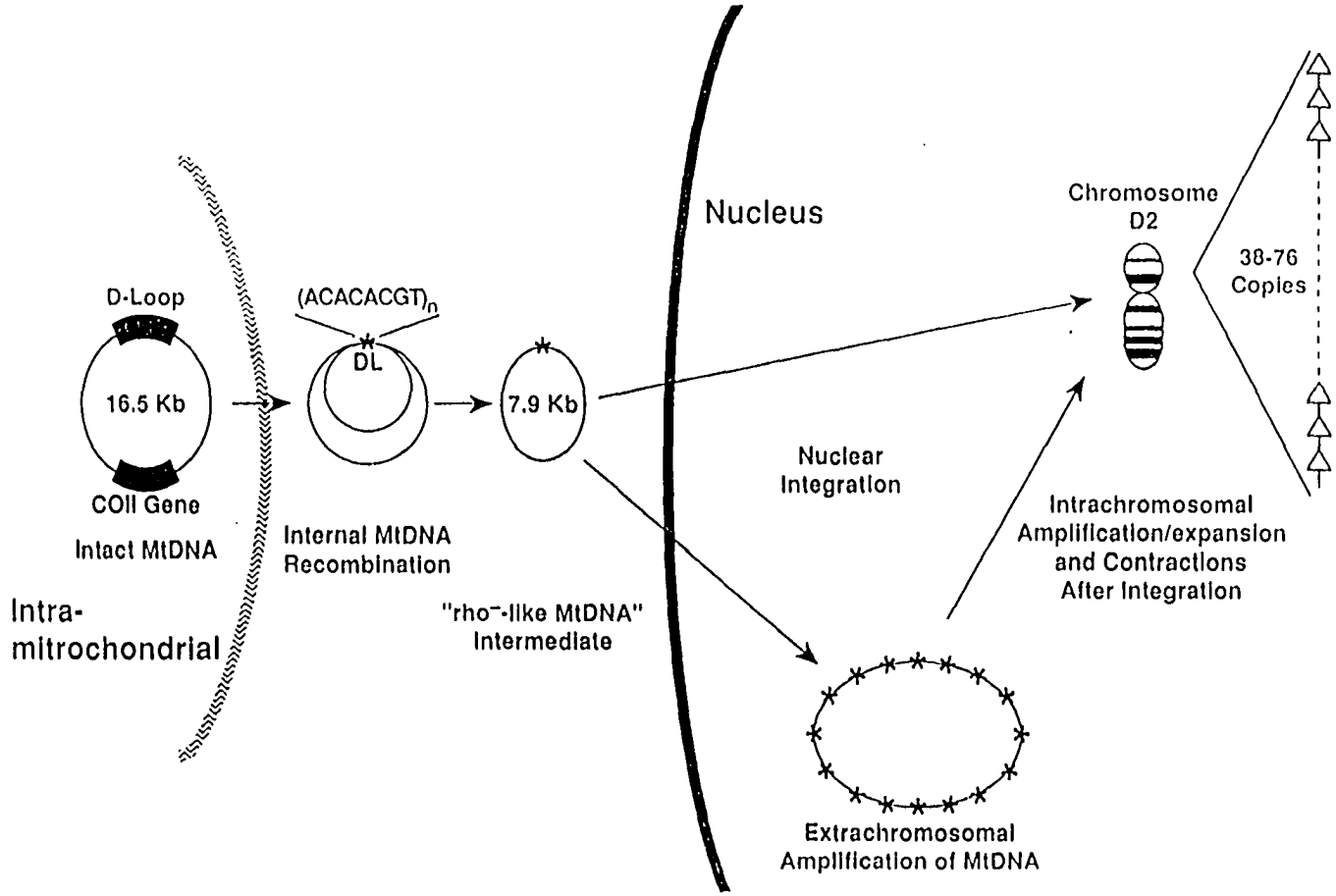


Fig. 11 - Model for the origin of Numt based on analogous recombination and extrachromosomal amplification of petite mutations in yeast (see text). Sequence analysis of Numt suggests a simple break in the D-loop (or CR) and in the COII gene fused to a circle and transposed from the 17.0 kb mitochondrial genome. Amplification and chromosomal integration followed but temporal order is difficult to ascertain. Further, it is possible that transposition, integration and amplification events may be either simultaneous or discordant in evolutionary time.

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; Gantt et al, 1991) The presence of normally untranscribed control region regulatory elements in *Numt* (Fig. 5, 6) 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-like" repeats are found in several carnivore mitochondrial genomes (Arnason and Johnsson, 1992; Hoelzel et al, 1993; Hoelzel et al, 1994) as well as being widely dispersed in the mammalian nuclear genome (Weber, 1990; Love et al, 1990; Dieterich 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* (Chandley and Mitchell, 1988; Wahls et la, 1990).

Length polymorphism in a tandem array 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. 7) supporting the interpretation that the PFGE fragments were alleles that differed 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 (Stephan, 1989; Jeffreys et al, 1991; Levinson and Gutman, 1987; Dover, 1982;

Zuckerkandl, 1992; Charlesworth et al, 1994). Other interpretations for individual heterogeneity such as short intervening genomic (non-mitochondrial) 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. 7B).

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 duplication and divergence has been the working hypothesis for the origin of nuclear genes with mitochondrial functions [e.g., nuclear-encoded members of the citric acid cycle, cytochrome chain and oxidative phosphorylation pathways] in the context of the Serial Endosymbiosis Theory (Margulis, 1970; Gray, 1989; Smith and Douglas, 1986; Ahmadhian, 1986).

Examination of the pattern of the *Numt* sequence in genes present in Fig 6 indicates several aspects that would preclude these sequences from producing functional gene products. First, for the protein coding genes of *Numt*, there are termination codons or frameshift mutations in all possible open reading frames which are compounded by the existence of two different genetic codes between the nucleus and

mitochondria (Brown 1985; Anderson et al. 1981; Breitenberger and RajBhandary 1985). For example in the *Numt* ND2 gene, a single base insertion causes chain termination codons in both genetic codes as illustrated in Fig. 9E. Second, the *Numt* control region is truncated relative to the cytoplasmic D-loop and may lack upstream regulatory elements. Third, the two identified regulatory elements (CSBII and III in Fig. 6) are involved in transcriptional promotion catalyzed by mitochondrial RNA polymerase and trans-activating factors that are not known to operate on nuclear genes (Schinkel and Tabak 1989).

Phylogenetic Interpretations

Alignment of *Numt* sequences to cytoplasmic mtDNA homologues in cat species demonstrated a high degree of homology (Table 2) 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 2) to the estimation of Li et al. (1981) whereby the fraction of sequence divergence is:

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

where μ_1 and μ_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 *Numt* transposition and amplification were contemporaneous events or occurred at different times in the evolution of the felid genomes. The pattern of *Numt* mutational divergence (Table 2) is relevant to this question since the observed changes appear to reflect selective constraints of functional genes. Most of the changes between *Numt* and cytoplasmic mtDNA were transitions (99/125=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 (*circa* 33%) for synonymous mutations indicating that non-synonymous (codon altering) mutations in *Numt* ancestors had been eliminated by natural selection, which could have occured while evolving as mitochondrial genomes (Kimura 1983; Ohta 1992; Hughes and Nei 1988; Yuhki and O'Brien 1990). Further, the occurrence of *Numt*-like sequences in other cat species outside of the *Felis* genus (e.g., lions; see Fig. 10) but not in all species more closely related to domestic cat is enigmatic and may reflect a discordance in evolutionary time between *Numt* transposition and *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 saltatory 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 *Numt* in distantly related Felidae species.

CHAPTER 2

COMPARATIVE ANALYSIS OF THE COMPLETE NUCLEOTIDE SEQUENCES OF NUMT AND THE DOMESTIC CAT MITOCHONDRIAL GENOME

BACKGROUND

"Cheshire-Puss ...would you please tell me which way I ought to go from here?", asked Alice.
"That depends a good deal on where you want to get to," said the Cat.
"I don't much care where ---,"
"--- so long as I get *somewhere*," Alice added as an explanation.
"Oh, you're sure to do that," said the Cat, "if you only walk long enough."
- from *Alice's Adventure's in Wonderland* (1865) by Lewis Carrol

Mitochondrial DNA, the genetic blueprint for a small set of proteins involved in oxidative phosphorylation and electron transfer on the mitochondrial membrane (Hatefi, 1985), has proved to be a pivotal molecule for population, evolutionary and phylogenetic studies (Avise, 1994; Brown, 1985; Wilson et al, 1985; Moritz et al, 1987; O'Brien, 1994b). Yet, continuing discovery of idiosyncracies in mtDNA gene evolutionary patterns reveal important exceptions to simple interpretations. For example, the extremely rapid rate of nucleotide substitution combined with a typically slower amino acid replacement frequency for some mitochondrial genes have recently

raised questions about mtDNA's reliability in phylogenetic inference (Graybeal, 1994; Meyer, 1994). Also, relatively few analyses have examined the pattern and consequences of mutation during the evolutionary divergence of mtDNA genes, i.e. the correlation of specific replacement mutations with secondary or higher order structure of proteins, enzymology or the effects on electron transport processes per se (Irwin et al, 1991; Adkins and Honeycutt, 1994). Pathological, somatic mtDNA rearrangements or deletions occur more frequently within specific regions of mitochondrial genomes and with increasing age (Larsson et al, 1990; Wallace, 1992). Furthermore, recent demonstration of an imperfect mode of maternal inheritance due to the leakage of paternal genotypes (Zouros et al, 1994) and mtDNA heteroplasmy (Buroker et al, 1990; Rand and Harrison, 1989; Biju-Duval et al, 1991; Hoelzel et al, 1993; Hoelzel et al, 1994) have also unsettled the paradigm of mtDNA's structural constancy or "economy" (Attardi, 1985). Lastly, several examples of genetic transfer between mitochondria and chromosomes have been described, including the *macrosatellite*-like locus, *Numt*, a transposition and tandem amplification (38-76X) of 7.9 kb of mtDNA into the nuclear genome of *Felis catus* and several closely related species of the genus *Felis* (Fukuda et al, 1985; Zullo et al, 1991, Smith et al, 1991; Lopez et al, 1994).

Substitution rates in mtDNA genes are known to vary between phyletic lineages, between organelles in the same cell, and between genes and gene regions within the same mitochondrial genome (Brown et al, 1982; Miyata et al, 1982; Britten, 1986; Vawter and Brown 1986; Hasegawa and Kishino, 1989; Kondo et al, 1993; Bulmer et

al, 1991; Li, 1993; Rand, 1993; 1994; Hillis and Huelsenbeck, 1992; Lynch and Jarrell, 1993). Factors that affect lineage-specific substitution rates include metabolic rates and body size, generation time, or nucleotide composition (Wolfe et al, 1989; Martin and Palumbi, 1993; Li, 1993). The rate of change between nucleotide and amino acid levels appears incongruent even within the same gene (Graybeal, 1994). Therefore, to avoid spurious conclusions, allowances for rate heterogeneity should be incorporated into phylogenetic studies which employ standard assays such as RFLP or DNA sequence analysis of mitochondrial genes. For example, the UPGMA method for phylogenetic reconstruction can produce erroneous topologies due to the inability to account for varying mutation rates (Huelsenbeck and Hillis, 1993).

Nine complete mammalian mitochondrial genome sequences (human, mouse, cow, rat, fin whale, blue whale, harbor seal, grey seal and American opossum) have been published to date (Anderson et al, 1981; 1982; Bibb et al, 1981; Gadeleta et al, 1989; Arnason et al, 1991; Arnason and Johnsson, 1992 ; Arnason and Gullberg 1993; Arnason et al, 1993; Janke et al, 1994). Feline mtDNA represents the first sequence of a terrestrial carnivore mitochondrial genome in sequence databases. The sequences of two distinct forms of feline mtDNA - one cytoplasmic, the other nuclear - will be useful from several evolutionary perspectives including a) description of carnivore divergence at the genetic level, b) facilitation of population and evolutionary biology studies of free-ranging endangered felid species (Yuhki and O'Brien, 1994; Pecon-Slattery et al, 1994; O'Brien, 1994a; 1994b; Masuda et al, in prep.), c) developing the

cat as a genetic model for human disease (O'Brien, 1986; Brown et al, 1994), and d) characterization of substitution rate heterogeneity and divergence of mitochondrial genes within and between different organellar compartments.

The phenomenon of *Numt* transposition represents an unusual opportunity to analyze a relatively intact paralogous duplication of mammalian mtDNA (Hardison and Gelinas, 1986; Goodman, 1981). It is most likely that *Numt* became a pseudogene immediately after transfer to the nucleus in light of organellar differences in genetic code. Evolutionary divergence between *Numt* and the rapidly evolving cytoplasmic genome is also probably uncoupled, owing to an environment of increased superoxide radicals, and less-efficient DNA repair in the mitochondria (Boulikas 1992; Miquel, 1992; Wallace, 1992). The unprecedented large scale of *Numt* transposition allows a comprehensive study of mutational processes across several contiguous genes in the cat mitochondrial genome, related to the "transparent substitutional sieve" described by Gillespie (1991). By comparing sequence divergence of *Numt* with feline and other mammalian cytoplasmic mtDNAs, we can directly observe the polarity, type, and varying rate of mutations occurring between pseudogene and functional genomes and different organelles.

MATERIALS AND METHODS

Cloning and Sequencing. Total nuclear and cytoplasmic nucleic acids (DNA and RNA) were extracted from fresh lymphocytes of a domestic cat (FCA 65) according to standard procedures (Sambrook et al, 1989; Lopez et al, 1994). All of the nuclear and cytoplasmic mtDNA sequences shown were encompassed in three Eco RI fragments (ca. 12.0, 7.9 and 5.0 kb), which were purified from preparative agarose gels. Isolation and characterization of the original 7.9 kb nuclear (pNumt.1) and two cytoplasmic clones (pCmt.12 and pCmt4.8) containing all of the sequences presented in this paper were described by Lopez et al (1994). Sequences of both Heavy and Light-strands were determined by either a) subcloning 1.0 -2.0 kb fragments into M13 mp18/mp19 single-stranded phage vectors (Sambrook et al, 1989) or b) walking in both 5' and 3' directions along the original intact lambda phage or pBlueScript phagemid (Stratagene) clones of pNumt.1, pCmt.12, or pCmt4.8. For walking, forward and reverse primers were designed at approximately 300 bp intervals and synthesized on an (Applied Biosystems Inc.-ABI) automated 394 DNA/RNA synthesizer. The three clones served as templates for cycle sequencing reactions run on an automated DNA sequencer 373A (ABI) using a fluorescence-labeled dideoxynucleotide termination method (dye-terminator). Some regions (about 30%) of *Numt* DNA were read manually by polyacrylamide gel electrophoresis using [^{35}S]-dATP in Sequenase reactions (U.S.

Biochemical). Verification of sequence data was evaluated with ABI analysis software, Sequencher 2.1 program (Gene Codes Corporation, 1994), or by visual inspection of chromatograms. Genetic regions which contained any ambiguous or unreadable nucleotides were sequenced again.

Sequence Analysis. Feline sequences were analyzed by programs of the University of Wisconsin Genetics Computer Group (GCG) (1994). Phylogenetic analysis was performed with PHYLIP 35c (Felsenstein, 1993), and Phylogenetic Analysis Using Parsimony (PAUP version 3.1.1) (Swofford 1993) on VMS VAX mainframe and Macintosh computers. Secondary structures were predicted with FOLD by Zuker and Steigler (1981) on GCG. The CMATRIX program developed at the LVC/NCI-FCRDC was used to calculate total percent similarities in nucleotide (nt) or amino acid sequences of mtDNAs. CMATRIX imposed a penalty of 1.0 for each gap encountered, and did not evaluate varying degrees of chemical similarity between DNA or amino acid residues. Therefore, our usage of the term, *similarity*, will be commensurate with the common use of sequence "identity" in the literature (Deveraux et al, 1984; GCG Manual, 1994). Multiple sequence alignments were created by either PILEUP or PRETTY in UWGCG. Most of the alignments to determine homology and gene boundaries within cat mtDNA were made with either cow or the harbor seal, *Phoca vitulina*, which encompass the phylogenetically nearest mtDNA sequences to cat available (Arnason and Johnsson, 1992; Li et al, 1990). The numbering system used

for cat cytoplasmic mtDNA follows harbor seal conventions (Arnason and Johnsson, 1992).

To determine relative DNA and amino acid substitution rates among all mitochondrial genes, all possible pairwise comparisons of percent sequence similarity of the following mammalian mitochondrial genomes retrieved from Genbank (release 86, 12/1994) and EMBL (release 39, 6/1994), were performed - harbor seal (*Phoca vitulina*), grey seal (*Halichoerus grypus*), fin whale (*Balaenoptera physalus*), blue whale (*Balaenoptera musculus*), human (*Homo sapiens*), cow (*Bos taurus*), mouse (*Mus musculus*), and rat (*Rattus norvegicus*). The American opossum (*Didelphis virginia*) genome was primarily used as an outgroup taxon. Values for the mean, standard deviation and range of DNA and amino acid percent similarity for each mitochondrial gene comparison were calculated and listed. The *Numt* sequence was submitted to GenBank in the form of the *in vivo* pNumt.1 clone, isolated by Lopez et al (1994). GenBank accession numbers for the cat cytoplasmic mtDNA and *Numt* are U20753 and U20754 respectively.

RESULTS

Composition of Cytoplasmic MtDNA Sequences

The *F. catus* mitochondrial genome is comprised of 13 structural ORFs, 22 tRNA genes, both large and small subunit rRNA genes and a regulatory control region (CR) (Fig. 12, 13; see also Fig. 1). The feline mtDNA sequence possesses several features found in other species. First, all of the ORFS are oriented in the same direction as homologous ORFs found in the other eight mammalian mitochondrial genomes, with no major rearrangements (Fig. 13A). Second, ND1, ND2, COIII, ND3, ND4 genes lack complete termination codons (Table 3). However, complete stop codons may be read within the tRNA-Trp gene directly downstream of the ND2 gene. Presuming that polyadenylation of processed transcripts occurs in cat mtDNA according to the model prescribed for humans (Anderson et al 1981), most stop codons in the cat mtDNA appear to be TAA. Third, disregarding the ORFs which have stop codons in their downstream tRNA genes (e.g. ND1 and ND2), Coding sequences overlap between the ATPase 8 and ATPase 6, ND4 with ND4L, and ND5 and ND6 genes. Fourth, the cat Light (L)-strand origin of replication (ORL) is less d(C-G)-rich (53%) than the harbor seal sequence, but the cat ORL can still fold into a stable stem-loop structure (Fig. 13B). Neither of the two rRNA genes nor any of the major structural genes show large length differences relative to other mammalian mtDNAs.

Table 3 - Characteristics of the Domestic Cat Cytoplasmic Mitochondrial Genome

Gene	From	To	Size*		Start codon	Stop codon	5' intervening spacer
tRNA-Phe	866	935	70				
12S rRNA	936	1895	960				
tRNA-Val	1896	1963	68				
16S rRNA	1964	3537	1574				
tRNA-Leu (UUR)	3538	3612	75				
NADH dehydrogenase subunit 1(ND1)	3615	4571	957		ATG	TAa^^	AC
tRNA-Ile	4571	4639	69				
tRNA-Gln	4637	4710	74	L			
tRNA-Met	4712	4780	69				A
NADH dehyrogenase subunit 2 (ND2)	4781	5822	1042		ATC	Taa^^	
tRNA-Trp	5823	5890	68				
tRNA-Ala	5907	5975	69	L			CACATCTAAACCATTC
tRNA-Asn	5977	6049	73	L			A
Origin of L-strand replication (OLR)	6050	6081	32				
tRNA-Cys	6082	6147	66	L			
tRNA-Tyr	6148	6214	67	L			
Cytochrome c oxidase subunit I (COI)	6216	7760	1545		ATG	TAA	T
tRNA-Ser (UCN)	7759	7828	70	L			
tRNA-Asp	7833	7901	69				TTAA
Cytochrome c oxidase subunit II (COII)	7902	8585	684		ATG	TAA	
tRNA-Lys	8589	8656	68				ATT
ATPase 8	8658	8861	204		ATG	TAA	C
ATPase 6	8819	9499	681		ATG	TAA	
Cytochrome c oxidase subunit III (COIII)	9499	10282	784		ATG	Taa^^	
tRNA-Gly	10283	10351	69				
NADH dehyrogenase subunit 3 (ND3)	10352	10698	347		ATA	TAa^^	
tRNA-Arg	10699	10767	69				
NADH dehyrogenase subunit 4L (ND4L)	10768	11064	297		ATG	TAA	
NADH dehyrogenase subunit 4 (ND4)	11058	12435	1378		ATG	Taa^^	
tRNA-His	12436	12504	69				
tRNA-Ser (AGY)	12505	12563	59				
tRNA-Leu (CUN)	12564	12633	70				
NADH dehyrogenase subunit 5 (ND5)	12634	14454	1821		ATA	TAA	
NADH dehyrogenase subunit 6 (ND6)	14438	14965	528	L	ATG	TAA	
tRNA-Glu	14966	15034	69	L			
Cytochrome B (Cyt B)	15038	16177	1140		ATG	AGA	TTA
tRNA-Thr	16178	16247	70				TT
tRNA-Pro	16248	16314	67	L			
Control region (CR)	16315	865	1559				

* ORFs end at the last base of putative stop codon.
L- Designates Light-strand transcript.
^^- Signifies an incomplete termination codon as shown in Fig 1, lower case denotes predicted codon after polyadenylation
-In the comparison with Numt a) there is a 335 bp overlap of control region sequences with Numt
b) The 16 bp spacer region between tRNA-Trp and tRNA-Ala contains 1 gap, c) a total of 2 mutations occur in other spacer regions

Table 4- Total codon usage in the cat cytoplasmic MtDNA

TTT (Phe)	108(68)	TCT (Ser)	33(35)	TAT (Tyr)	66(50)	TGT (Cys)	11(8)
TTC	117(163)	TCC	73(74)	TAC	76(84)	TGC	14(16)
TTA (Leu)	109(68)	TCA	112(106)	TAA (Ter)	7(7)	TGA (Trp)	92(91)
TTG	17(20)	TCG	8(10)	TAG (Ter)	0(0)	TGG	12(13)
CTT (Leu)	59(57)	CCT (Pro)	57(81)	CAT (His)	31(36)	CGT (Arg)	6(5)
CTC	97(95)	CCC	59(42)	CAC	63(64)	CGC	9(13)
CTA	273(297)	CCA	71(68)	CAA (Gln)	84(72)	CGA	45(43)
CTG	42(58)	CCG	7(7)	CAG	7(14)	CGG	7(4)
ATT (Ile)	151(131)	ACT (Thr)	72(51)	AAT (Asn)	56(40)	AGT (Ser)	18(17)
ATC	179(203)	ACC	101(107)	AAC	94(107	AGC	34(42)
ATA	182(192)	ACA	124(137)	AAA (Lys)	88(89)	AGA (Ter)	1(1)
ATG (Met)	68(57)	ACG	14(22)	AAG	14(12)	AGG (Ter)	0(0)
GTT (Val)	33(26)	GCT (Ala)	51(46)	GAT Asp)	29(26)	GGT (Gly)	33(48)
GTC	38(44)	GCC	116(98)	GAC	39(46)	GGC	57(36)
GTA	95(90)	GCA	84(100)	GAA (Glu)	76(74)	GGA	97(109)
GTG	21(28)	GCG	8(8)	GAG	23(25)	GGG	31(22)

Numbers in parentheses indicate total codon usage in harbor seal mtDNA sequence.
(Arnason and Johnsson, 1992)

Fig. 12 - A. The complete nucleotide sequences of domestic cat (*F. catus*) cytoplasmic mtDNA (top) and one aligned *Numt* DNA repeat unit (bottom). The L-strand sequence is shown. Alignment of both sequences in the 7946 bp region of homology was performed with GAP (GCG, 1994), using a gap weight of 5.0 and a gap length weight of 0.3. Indels are marked by (.) and highlighted in gray shading. The translated aa sequences for each structural gene are given above the nucleotide sequences using standard nomenclature and translated with the mitochondrial genetic code. Repetitive regions in the control region are marked as either RS2 or RS3 following the nomenclature of Hoelzel (1993). At RS2, boundaries of each monomer are marked by an (▼) and the total region is delineated by a (∇). RS2 palindromes are marked by arrows below the sequence with the 5'-->3' arrows showing strong conservation to TAS consensus sequences (Foran et al, 1988). All nucleotide numbering in the text refers to the cytoplasmic mtDNA designations of the cat and follows homologous human mtDNA conventions (Anderson et al, 1982). Abbreviations for all mitochondrial genes was based on human nomenclature (Wallace, 1992), except for the tRNA genes which followed Arnason and Johnsson (1992). Due to typeset editing, sequence line 33 beginning with nucleotide 3177 is 1 bp short, and line 7 beginning with 601 is 1 bp longer than normal due to an insertion.

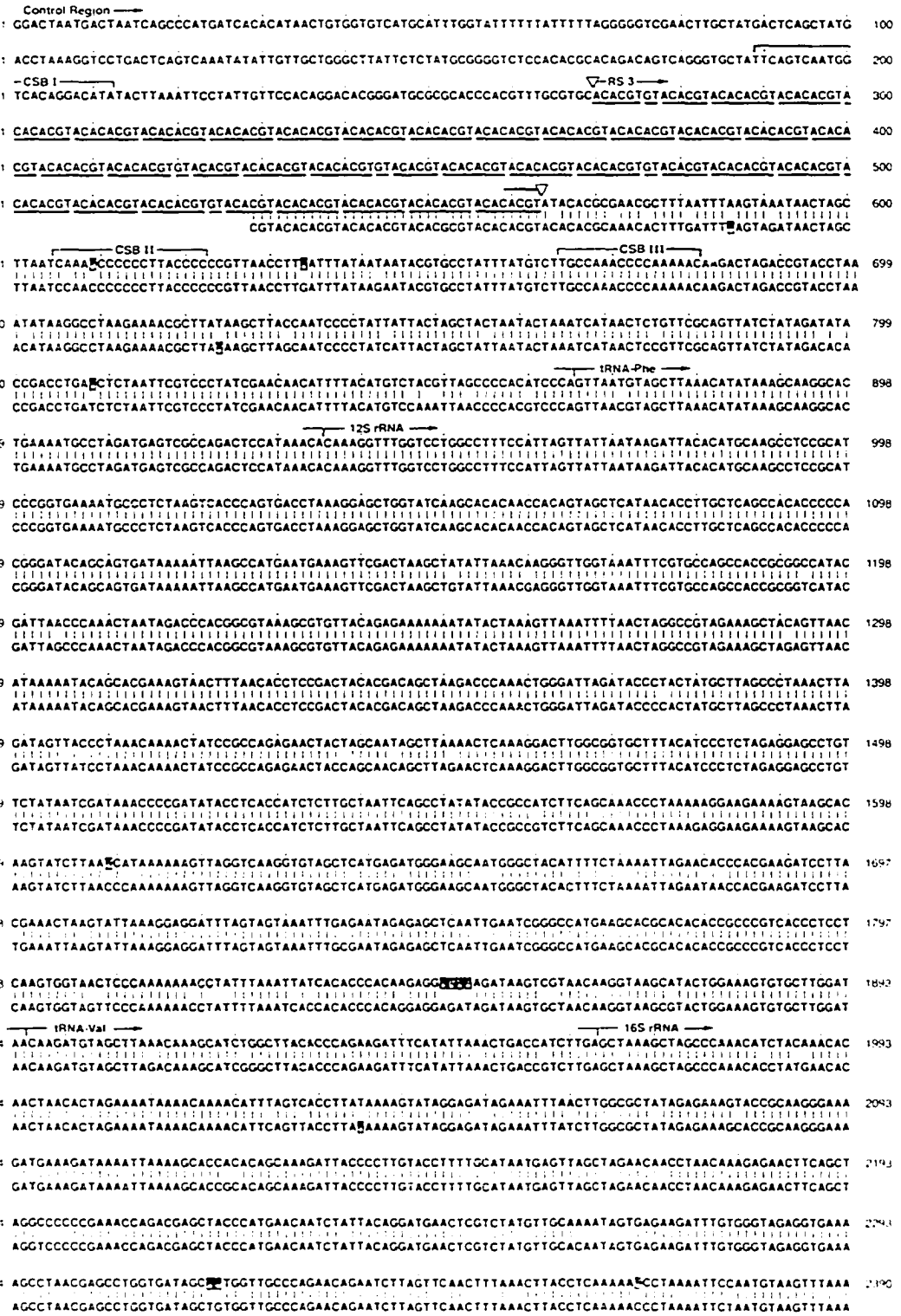


Figure 12. (Continued)

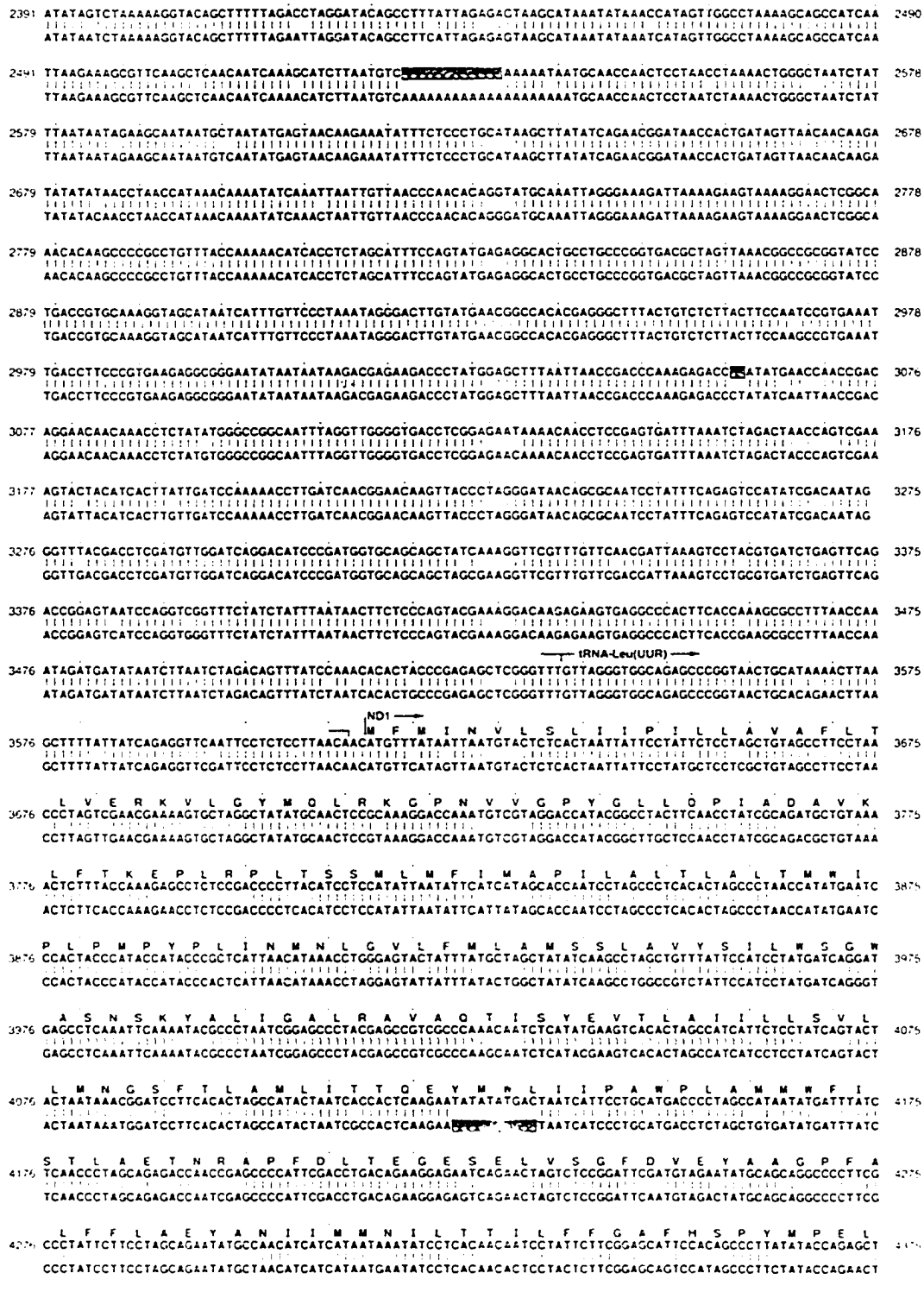


Figure 12. (Continued)

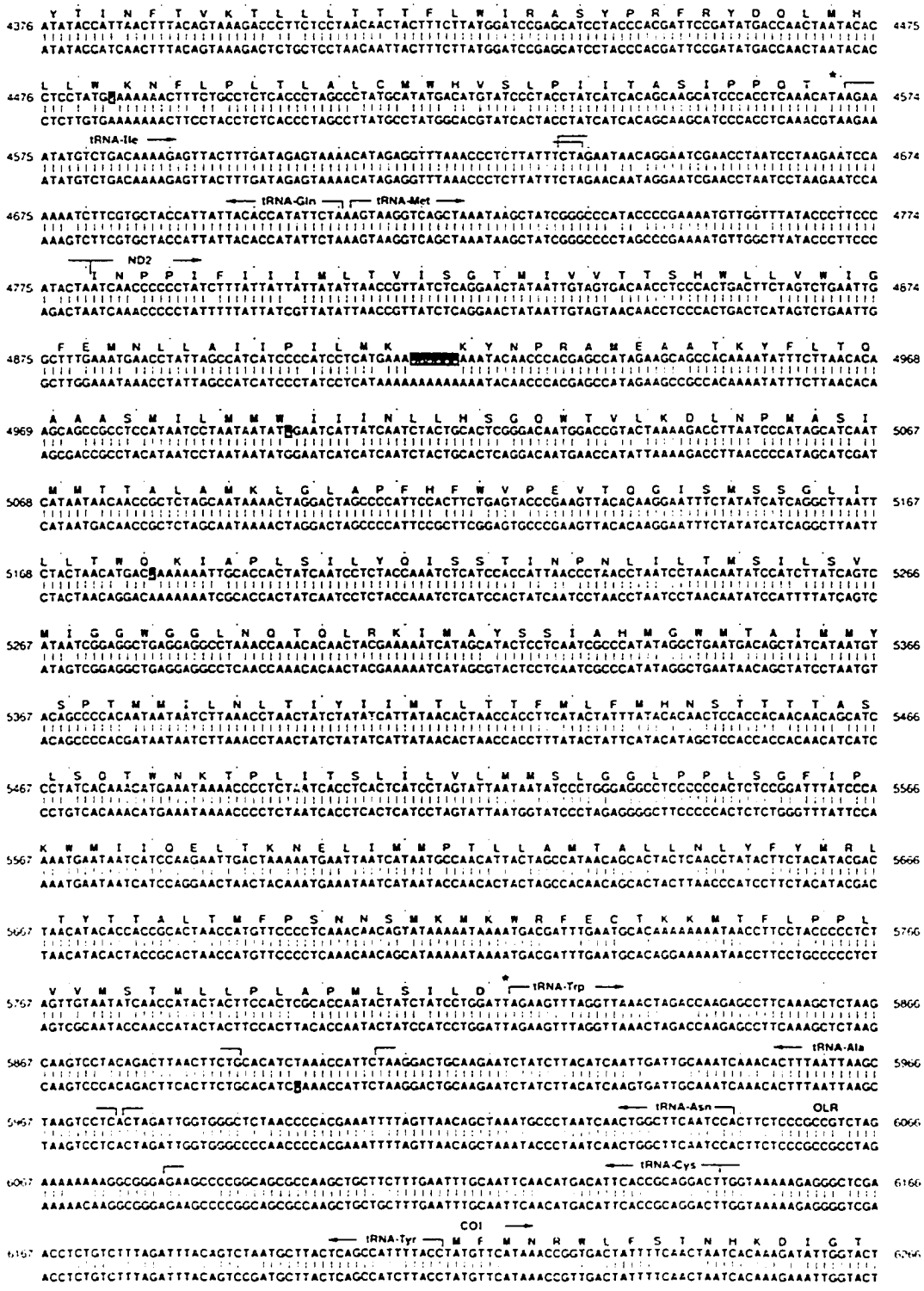


Figure 12. (Continued)

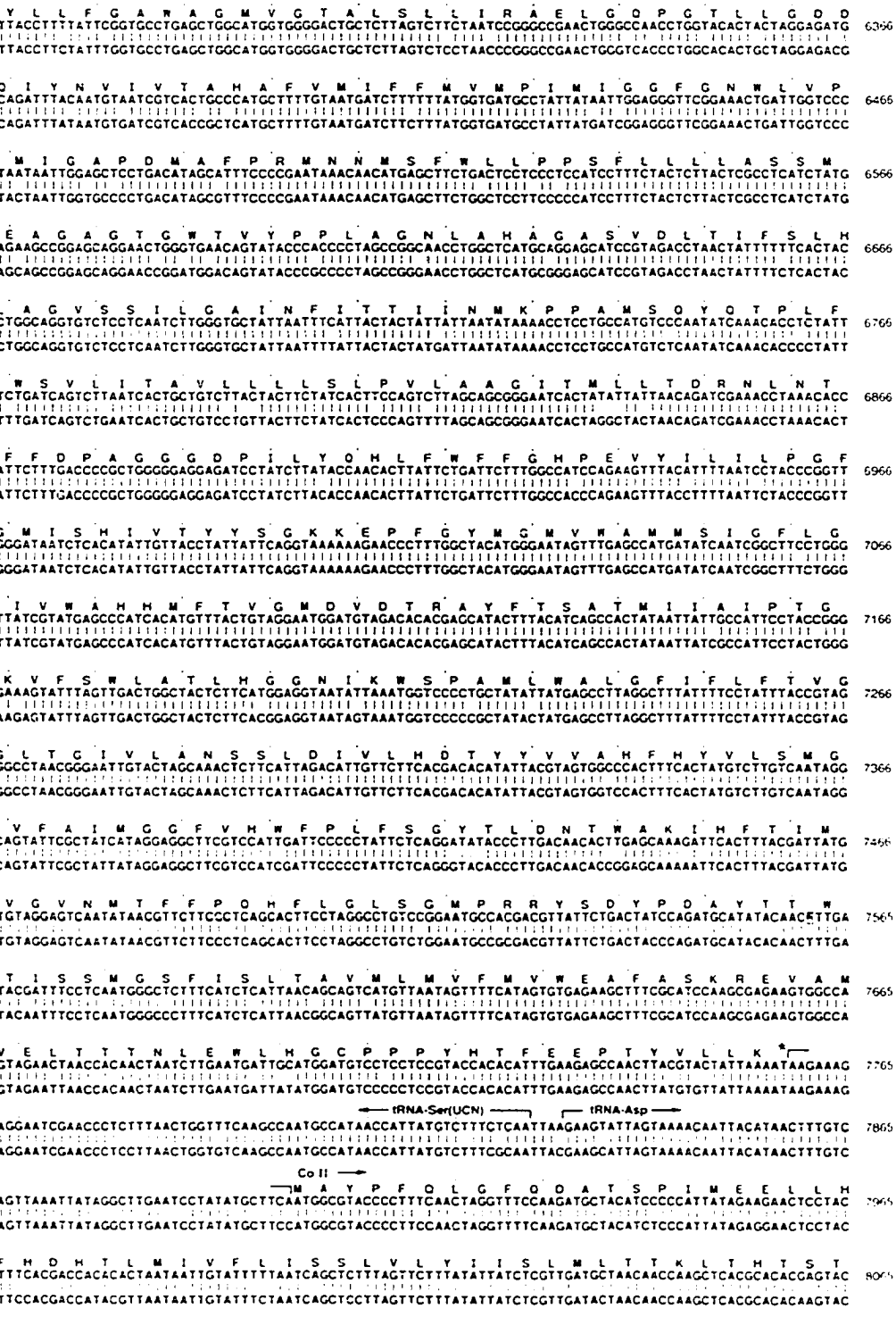


Figure 12. (Continued)

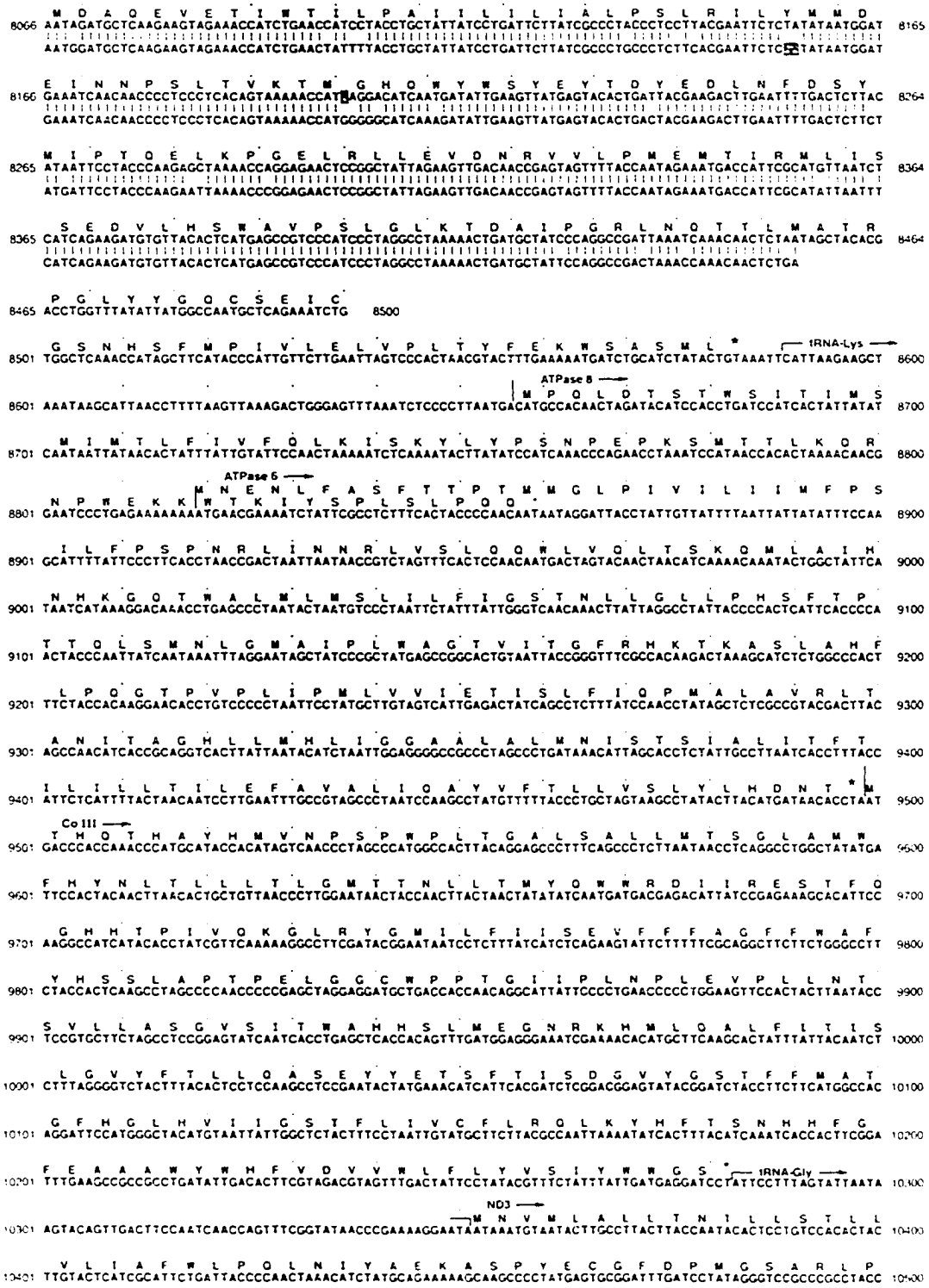


Figure 12. (Continued)

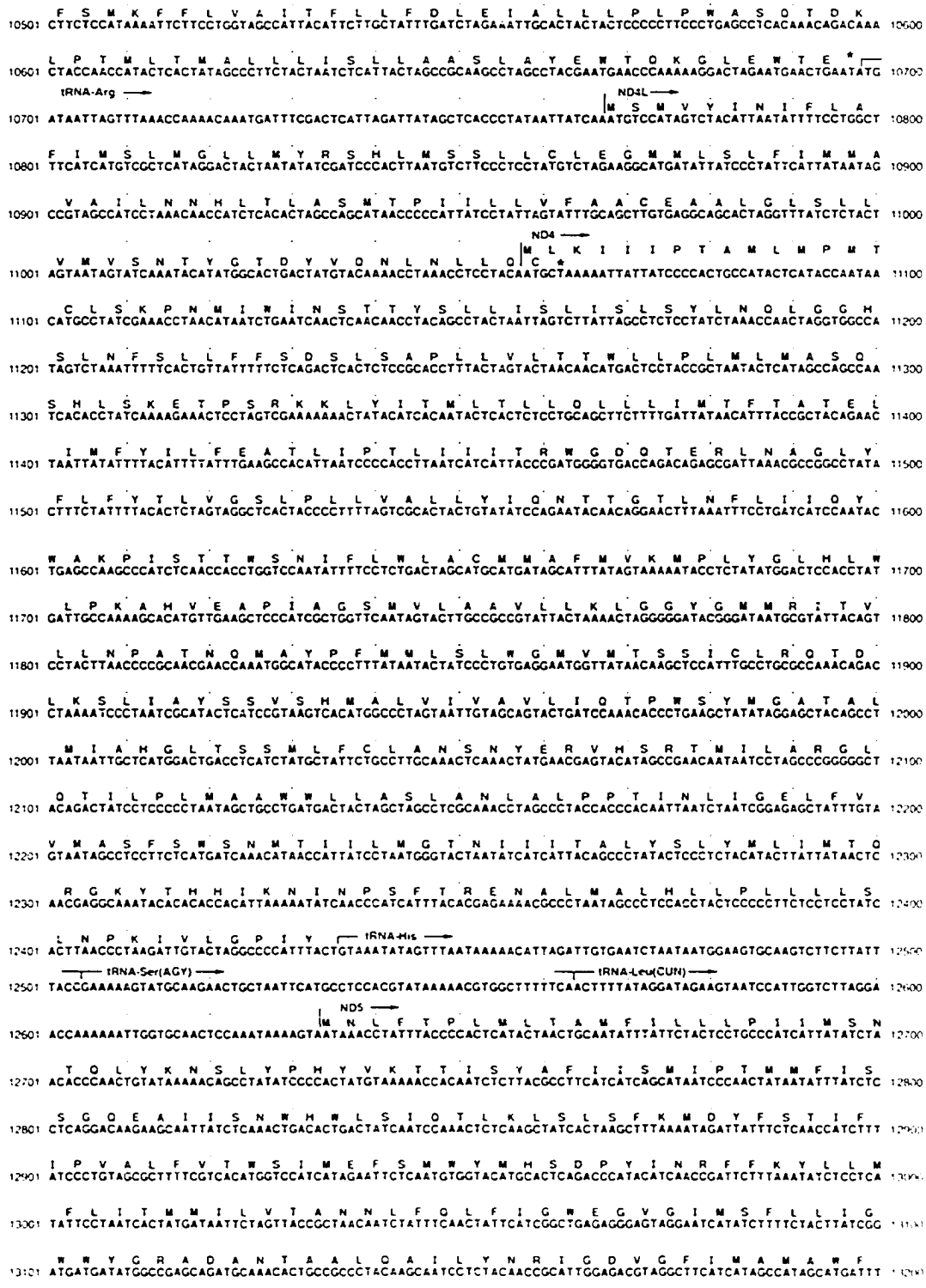


Figure 12. (Continued)

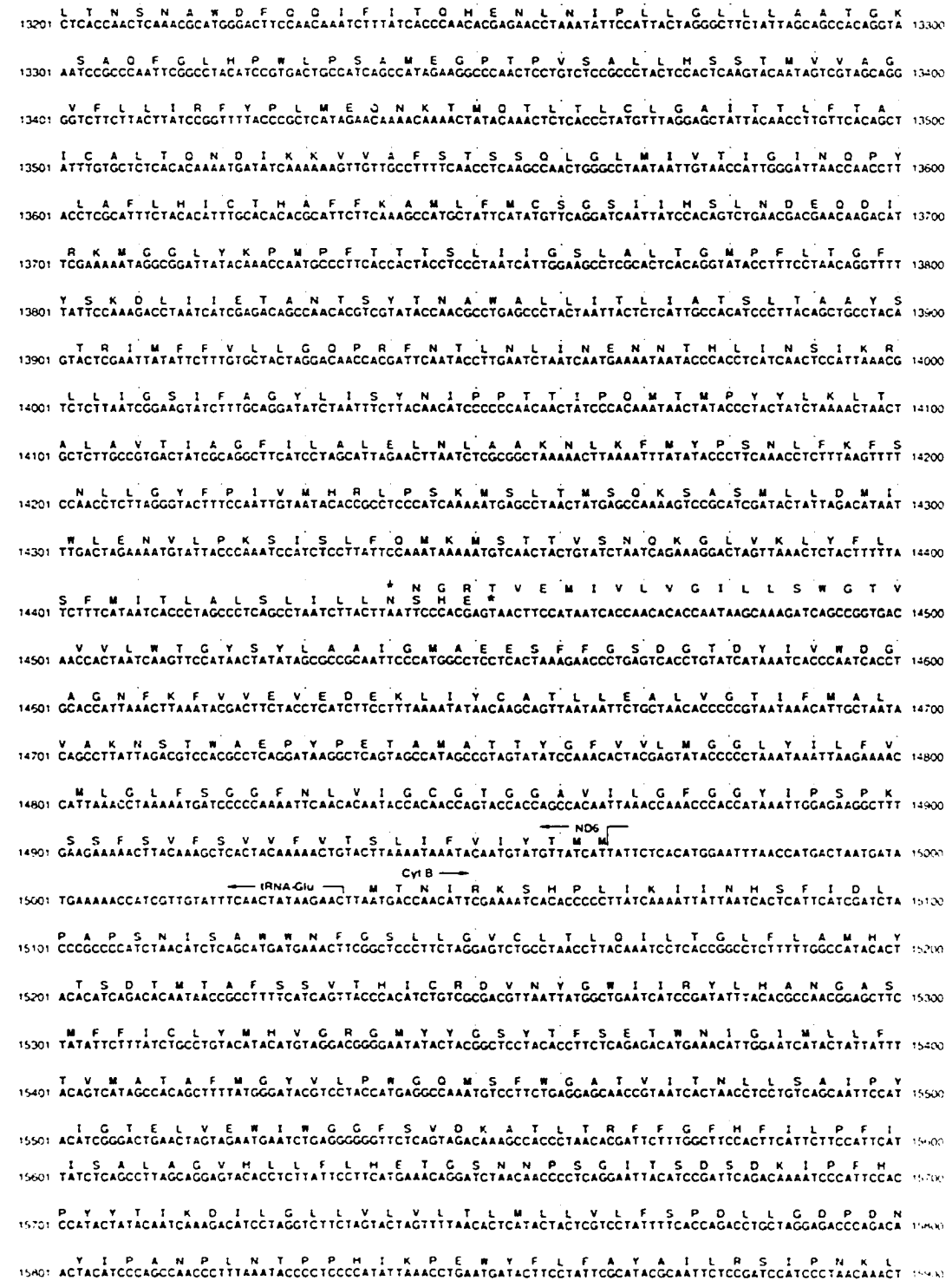


Figure 12. (Continued)

Figure 12. (Continued)

The total length of the cat sequence is 17009 bp. Total base composition of the mitochondrial genome is 32.5% A , 26.2% C, 14.2% G, and 27.1% T. Similar to seal mtDNAs, the cat mitochondrial genome exhibits a high overall dG content among mammals but retains the strong bias against G at the third codon position in structural genes. The pattern of codon usage in cat mtDNA (Table 4) follows the preference patterns observed in other mammalian mtDNA sequences, with the possible exception of Ile and Phe codons. Compared to harbor seal codon usage, cat mtDNA shows an increase of TTT, probably at the expense of TTC codons. Also, the cat initiates the ND2 gene with a non-AUG codon, ATC (Ile).

In contrast to other mammalian genomes, the ATPase 8 gene is extended by the duplication of one Q residue at the 3' end which may have resulted from slippage during DNA replication. Also, a total of 34 bp comprise the untranslated spacer nucleotides between mitochondrial genes.

Analysis of tRNA Genes

The canonical secondary structure features common to most vertebrate mitochondrial tRNA molecules (Cedergren et al, 1981), such as the anticodon (AC) stem-loops, and T-phi-C and amino acid-acceptor (A-A) arms, are also observed in some feline mitochondrial tRNAs (Fig. 13C). When compared with cow or harbor seal tRNA sequences, most differences occur in either the T-phi-C arm, dihydrouridine-loop, or the "variable" loop region between the AC stem and the T-phi-C arm, and

Fig. 13. - A, Physical map of coding genes and major restriction sites within the cat cytoplasmic mtDNA. Genes on the inner circle are transcribed from the L-strand. Locations of the tRNA genes (shaded boxes) conform to the canonical placental mammalian arrangement and have been previously drawn (Lopez et al, 1994). Listed enzymes recognize 6-base bp sites and cut less than four times, except HindIII which has five sites. The following abbreviations were used: HSP - putative heavy-strand promoter, OHR - origin of heavy-strand replication; OLR - origin of light-strand replication. B, Predicted secondary structure of the OLR (L-strand origin of replication) (energy = -14.0). Sequences are shown in H-strand orientation; C, Representative tRNA cloverleaf secondary structure of tRNA-Lys in cat mtDNA. Diagram was produced with FOLD program of Zuker and Steigler (1981) in GCG (energy =-11.0).

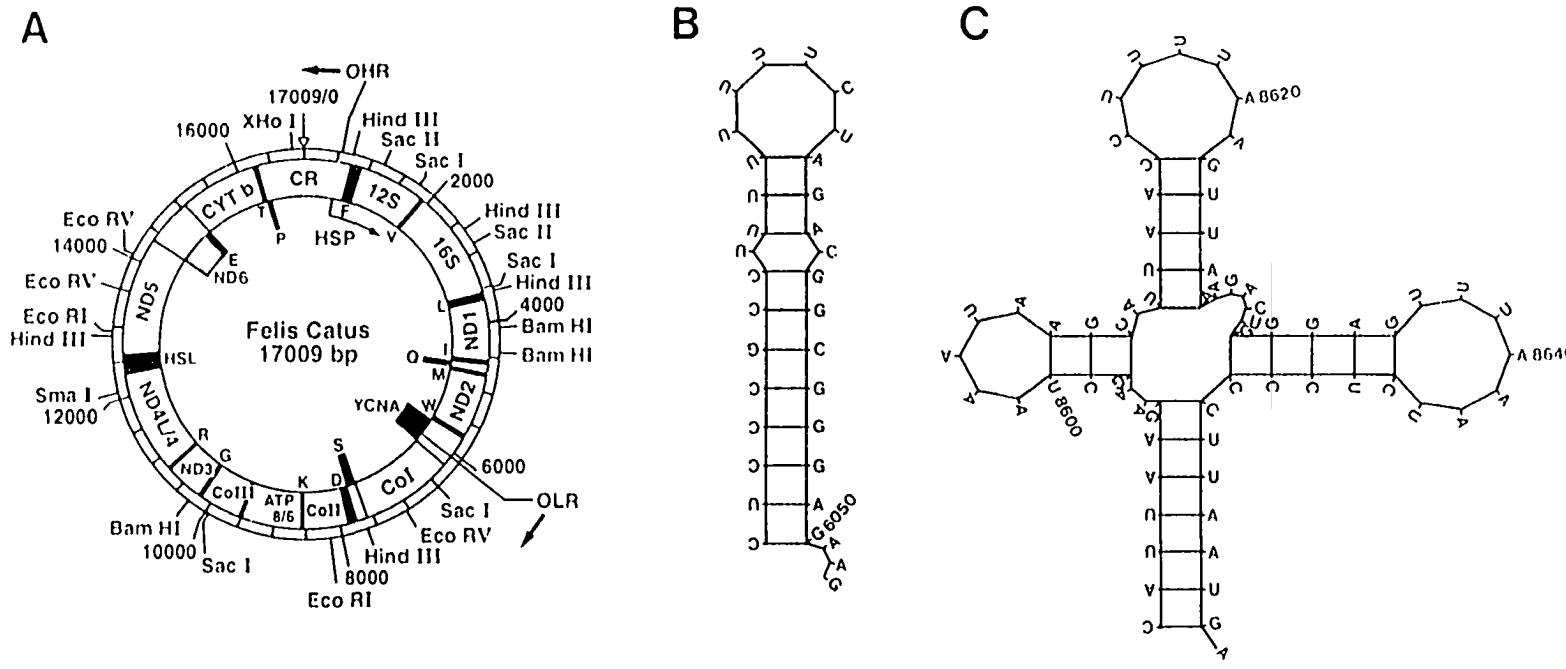


Figure 13. Physical map and secondary structures in the *F. catus* mitochondrial genome.

many substitutions in one stem strand were compensated in the corresponding stem strand. Insertion or deletion mutations (indels) are observed in the cat dihydrouridine-loops of His, Gln, Phe, Pro, Tyr, Leu and Asp-tRNAs relative to bovine mtDNA genes. For example, the tRNA-Phe gene is 3 bp longer in cat than in cow mtDNA. The AC loop region is the most conserved tRNA region among all comparisons. Furthermore, the cat tRNA-Leu (CUN) gene is longer relative to the harbor seal sequence, which is best explained by an arbitrary placement of the ND5 5' gene boundary in *P. vitulina* mtDNA, causing it to be three residues (M-K-V) longer and discordant with all other mammalian ND5 sequences. Feline tRNA sequences for Ala, Val and Met contain a large number of mutations in the AC stem when compared with bovine mtDNA, although most changes are compensated.

Mitochondrial DNA Control Region of the Cat

The feline control region (CR) spans about 1560 bp (Fig. 14A). An unusual characteristic of the cat mtDNA CR is the presence of two distinct sites of repetitive sequences (RS2 and RS3) on opposite sides of the highly conserved core of the control region, which together extend the CR 447 bp longer than the human sequence. The locations of these repeats appear to be highly conserved when compared to other mammalian CR repeats (Wilkinson and Chapman, 1991; Ghivizzani et al, 1993; Arnason and Johnsson, 1992). RS3, a 294 bp d(C-A)-rich repeat, which resembles nuclear microsatellites (Love et al, 1990) as well as other carnivore CR repeats

Fig. 14. - A, Schematic diagram of the feline control region (drawn to scale). Numbers correspond to cytoplasmic mtDNA nucleotide positions shown in Fig. 1. Hatched boxes represent repetitive sequence sites, RS2 or RS3, following terminology of Hoelzel (1993). CSB I-III designate closest matches to previously identified "conserved sequence blocks" (Saccone et al, 1991; Lopez et al, 1994). B, Multiple alignment of three complete RS2-type repeats (80, 80, 82 bp respectively) in the cat cytoplasmic CR. The RS2 region spans nt positions 16504 to 16779 in the cat mtDNA. Evening bat and masked shrews sequence are also listed for comparison. C, Secondary structure of two of three 80 bp repeats at the RS2 site (pos. 16538) in the CR (energy = -25.3) produced by the FOLD. Black arrows mark the location of palindromic sequences shown in Fig. 1A, and white arrows indicate the dG substitution observed in two repeats.

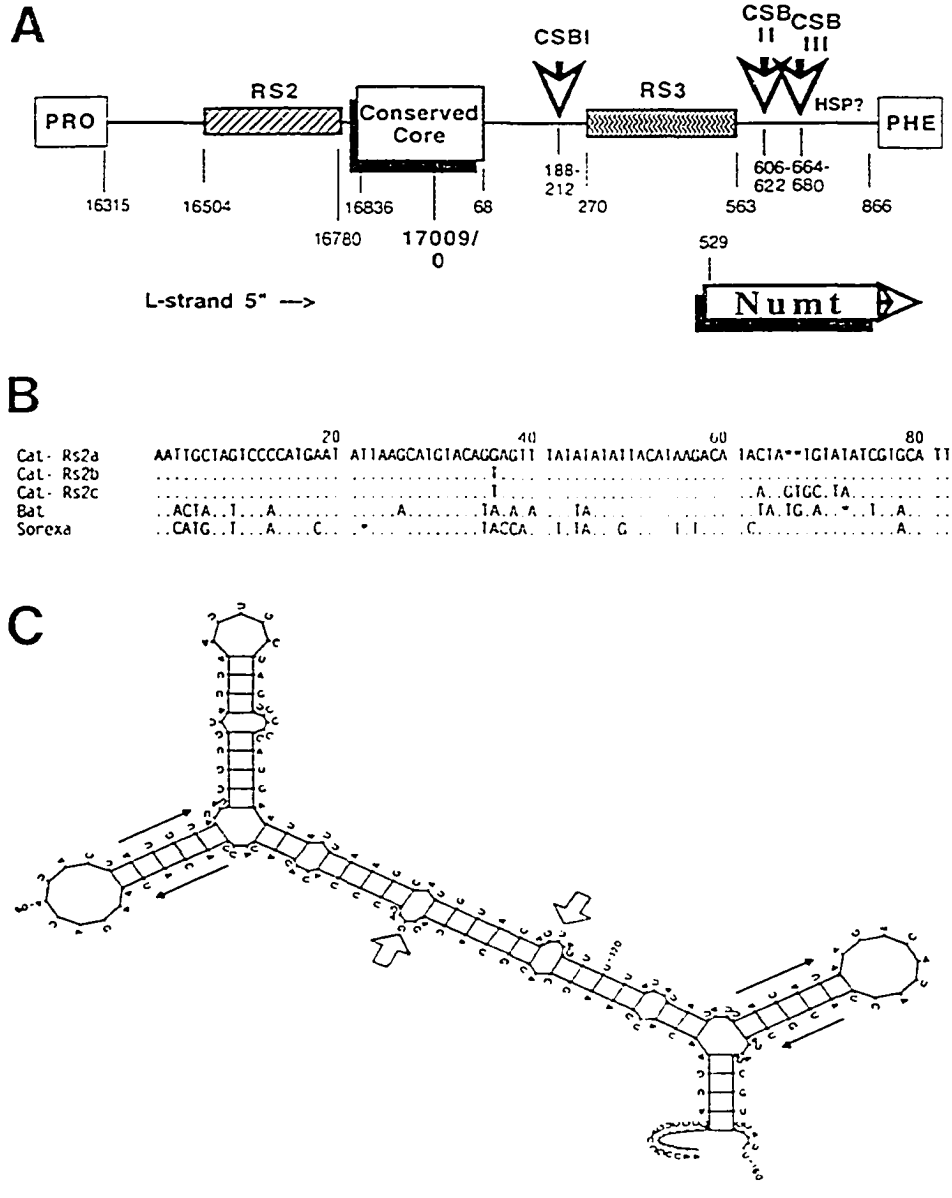


Figure 14. Schematic diagram of the feline mitochondrial control region.

(Hoelzel et al. 1994), is found at the L-strand 3' end (beginning at nt. pos. 270) of feline mtDNA (Fig. 12). RS3 contains a 6-8 bp core unit, ACACACGT, imperfectly repeated 37 times int he mtDNA sequence. The RS2 element at the 5' (left) end of the CR L-strand consists of three complete 80-82 bp monomers (a-c), which are highly conserved among each other (91-98% similarity) (Fig. 14B). The cat RS2 repeats also show 72-75% and 67-74% sequence similarity to homologous mtDNA CR repeats observed in the evening bat and masked shrew, respectively (Wilkinson and Chapman, 1991; Stewart and Baker, 1994). The 3'-most cat repeat (RS2c) showed the greatest divergence at its own 3'end, while the most 5' repeat unit (pos. 16504), possessing 94% similarity and one deletion relative to the consensus, is truncated after only 34 bp in cat mtDNA. The cat RS2 sequences contain several palindromic motifs (5' TACAT-----ATGTA 3') beginning at pos. 16507 which could potentially form secondary structures and possibly function as terminal associated sequences (TAS) involved in D-loop replication (Foran et al. 1988; Saccone et al. 1991; Madsen et al. 1993; Brown, 1986). Array secondary structures appeared more stable with two or three RS2 repeats versus one repeat unit. These secondary structures may have facilitated the expansion/contraction of repeat numbers following the model of Buroker et al (1990). The non-integral repeat number may reflect mutational decay or misaligned slippage during rounds of duplication. Preliminary data indicate that the RS2 region is hyper-mutable and highly heteroplasmic within individuals of other species of Felidae (M. Culver, unpublished observations).

Comparison of Cytoplasmic MtDNA and Numt Homologous Region

Homologous *Numt* DNA sequences extend 7946 bp and were derived from a single monomer of the tandemly repeated chromosomal array estimated to range from about 38-76 copies (Lopez et al, 1994). The overlapping homologous region between cytoplasmic mtDNA and *Numt* begins at cytoplasmic nt. pos. 529 within the RS3 repeat and includes about 80% of the COII gene to nucleotide 8454 (Fig. 12). The last 304 bp of the *Numt* COII gene in Fig. 12 occur upstream of the CR RS3 repeat sequences in the original pNumt.1 clone, corroborating the observations of an *in situ* excision, circularization, nuclear integration and tandem amplification at the chromosomal *Numt* locus (Lopez et al, 1994).

Tables 5 and 6 summarize the pattern of mutational divergence between *Numt* and cytoplasmic mtDNA genes of cat and other species. A total uncorrected nucleotide sequence difference of about 5.1% exists between the two feline complements. The overall base composition of *Numt* did not increase in dA and dT content relative to cytoplasmic mtDNA: Cytoplasmic- (33.3%) A, (26.4%) T, (24.8%) C, (15.4%) G vs *Numt*- (32.6%) A, (25.8%) T, (25.2%) C, and (16.3%) G. When changes in the cat homologous region are grouped according to codon position, 72% of the base substitutions occur at the 3rd position, and 23/51 (45%) 1st codon substitutions are at synonymous Leu codons (Table 5). These data suggest that more changes than

Table 5. Nucleotide Substitutions in Protein Coding Genes of the Homologous Region between Cat Cytoplasmic mtDNA and Numt by each Codon Position

Gene	# Nts.	Codon position 1**	2	3	Total Subst.
ND 1	957	14 (23%) / 3	3 (5%) / 1	44 (72%) / 5	61/9
ND 2	1042	16 (21%) / 6	9 (12%) / 3	52 (68%) / 6	77/15
CO I	1544	15 (17%) / 3	7 (8%) / 4	64 (73%) / 6	86/13
CO II*	250	6(17%) / 1	2 (6%) / 1	28 (77%) / 1	36/3
Totals	3793	51 (20%) / 13	21 (8%) / 9	188 (72%) / 18	260/40

-Entry values: [Total # Substitutions (% Substitutions) / # Tv]
* Includes only the homologous mtDNA region in Numt.
**- 23 substitutions at Leu codons are synonomous; the remainder are nonsynonomous (NS)

expected by random chance were synonymous, and further indicate that most of the divergence between *Numt* and cytoplasmic mtDNA occurred in the cytoplasmic genome, where selection pressure constrained random mutational drift.

A comparison of the pattern of mutational divergence between feline cytoplasmic mtDNA and *Numt*, relative to mtDNA divergence between related species from other mammalian families (e.g. fin/blue whales, harbor/grey seals) is presented in Table 6. All three comparisons involve recently diverged mtDNA sequences (seals: 2-2.5 MYA - Million Years Ago; whales: <5.0 MYA; and *Numt*/cytoplasmic ca. 2.0 MYA) (Arnason et al, 1993; Arnason and Gullberg, 1993). The percent similarity values for each gene appear comparable among the three datasets and likely reflect gene-specific evolutionary rates. Thus, CO subunit and rRNA genes are the most conserved, followed by the ND1 and ND2 genes in all three comparisons. Moreover, the varying Ti:Tv ratios observed between the different gene classes (protein, tRNA and rRNA), probably reflect the different selective constraints acting on each gene class. For all three groups, rRNA genes consistently show the lowest Ti:Tv ratios. Despite having the longest estimated divergence times among the three data sets, the whale data set retains a high overall Ti:Tv ratio of 9:1, similar to the seal ratio of 12:1. In contrast, overall (3.5:1) and individual feline gene Ti:Tv ratios greatly deviate from both seal and whale data sets, while corresponding sequence similarities remain relatively uniform (Table 6). Furthermore, between the two seals, only 5 percent of third codon position changes were transversions, while twice as many transversions accrue at the same position in

Table 6. Nucleotide Substitution Patterns from Pairwise Comparisons of Closely Related Mammalian Taxa

	Num/Domestic Cat						Harbor Seal/Grey Seal						Fin Whale/Blue Whale					
Gene	Subst.	Gap	Ti	Tv	Ti:Tv	% Similar	Subst.	Gaps	Ti	Tv	Ti:Tv	% Similar	Subst.	Gaps	Ti	Tv	Ti:Tv	% Similar
Control region**	21	5	16	5	3:1	92	11	1	10	1	10:1	97	12	1	7	5	1.4:1	96
t-PHE	1	0	1	0	1:0	98	1	0	1	0	1:0	99	8	0	7	1	7:1	89
12S rRNA	32	2	22	10	2:1	96	20	1	18	2	3:1	98	44	6	36	8	5:1	95
t-VAL	3	0	2	1	2:1	96	2	0	2	0	2:0	97	2	0	1	1	1:1	97
16S rRNA	42	5	30	12	2:1	97	35	6	28	7	4:1	97	80	3	67	13	5:1	95
t-Leu (UUR)	3	0	3	0	3:0	95	0	0	0	0		100	2	2	2	0	2:0	95
ND 1	61	2	52	9	6:1	93	29	0	29	0	29:0	97	72	0	70	2	35:1	92
t-ILE	0	0	0	0		100	1	2	0	1	0:1	97	1	0	1	0	0	98
t-GLN	3	0	3	0	3:0	96	1	1	1	0	1:0	99	0	0	0	0		100
t-MET	4	0	1	3	1:3	94	1	1	1	0	1:0	98	1	0	1	0	1:0	98
ND 2	77	3	62	15	4:1	93	46	0	43	3	14:1	96	101	0	89	12	7:1	90
t-TRP	2	0	1	1	1:1	97	1	1	1	0	1:0	99	1	0	1	0	1:0	98
t-ALA	1	0	0	1	0:1	99	0	0	0	0		100	2	0	2	0	2:0	97
t-ASN	3	0	3	0	3:0	96	0	0	0	0		100	0	0	0	0		100
OLR	2	0	1	1	1:1	94	0	0	0	0		100	1	0	1	0	1:0	97
t-CYS	1	0	0	1	0:1	98	0	1	0	0		100	2	0	2	0	2:0	97
t-TYR	4	0	3	1	3:1	94	5	0	5	0	5:0	93	0	0	0	0		100
CO I	86	1	73	13	6:1	94	56	0	52	4	13:1	96	115	0	107	8	13:1	92
t-SER (UCN)	3	0	1	2	1:2	96	1	0	1	0	1:0	99	4	0	4	0	4:0	94
t-ASP	3	0	1	2	1:2	96	1	0	1	0	1:0	99	0	0	0	0		100
CO II*	36	2	33	3	11:1	94	14	0	14	0	14:1	94	21	0	19	2	10:1	92
TOTALS	388	20	308	80	3.8:1	95	225	14	207	18	12:1	97	469	12	417	52	9:1	94
(Gene Class Totals)																		
tRNA	31	0	*19*	*12*	*1.6:1*	97	14	6	*13*	*1*	*13:1*	98	23	2	*21*	*2*	*10:1*	97
rRNA	74	7	*52*	*22*	*2.4:1*	97	55	7	*46*	*9*	*5:1*	98	124	9	*103*	*21*	*5:1*	95
Protein	260	8	*220*	*40*	*6:1*	93	145	0	*138*	*7*	*20:1*	96	309	0	*285*	*24*	*12:1*	92

'- Mean similarities in structural genes were calculated with a penalty of 1 substitution for each gap

-Ti:Tv ratios were not listed for genes with 100% similarity

*-Comparison involves only the first 250 bp which are homologous between cat mtDNA sequences

**- Comparisons encompass the extreme 3' end of the CR: 336 bp in the cat (see Fig. 1), 346 bp between the seals, and 350 bp in whales.

- Boxed values designate overall Ti:Tv ratios for each mammalian group.

- DNA sequences and estimated divergence times for the seal (2.0-2.5 MYA) and whale (5-7 MYA) pairs were derived from Arnason and Johnsson, (1992), Arnason et al (1991), Arnason and Gullberg (1993). The 2.0 MYA divergence time for the two cat mtDNAs was based on their divergence and reference mutation rates for nuclear pseudogenes (Lopez et al, 1994), and also conforms with previously estimated divergence times for other species of genus Felis (Collier and O'Brien, 1985) known to carry nuclear mtDNA.

the cat mtDNAs. The greatest flux of cat Ti:Tv values appear in the tRNA class, which had the lowest ratios (1.6:1). In several studies, the ratio of transitions to transversions has been shown to exceed 20:1 in recently diverged mtDNA sequences (Brown et al, 1982; Ruvulo et al, 1993), and generally, transversions cause more amino acid replacements, which accumulate with increasing divergence time (Aquadro and Greenberg, 1983; Jukes, 1987).

The feline mtDNA homologous region contains a total of 21 gaps, representing indels that ranged from 1-10 bp in length (Fig. 12); of these, 12 were single nucleotide indels. Fourteen of the gaps (66%) are insertions in the *Numt* sequence, which lengthen *Numt* by at least 20 bp over the homologous cytoplasmic mtDNA. Five gaps occur in the CR, 7 in the rRNA genes, 5 in the ND subunits and only 3 in the CO genes. A large proportion (25%) of gap mutations are found in the variable 3' terminus of the CR, reconfirming the relaxed mutational constraints in this region. These mutations may derive from DNA polymerase slippage during DNA replication, since at least eight indels occurred at sites which are "simple", homopolymeric, or with one alternating nucleotide motif (Tautz et al, 1986). For example, two gaps involving >1 bp occur at sites with alternating residues or direct repeats (nt pos. 1848, 4124). Other long (12 bp, 6 bp) insertions of poly-A sequences occur in the 16S rRNA gene (pos. 2533) and in the ND2 gene (pos. 4918), respectively. These observations plus preliminary measures of cryptic simplicity suggest the influence of stochastic DNA turnover mechanisms with respect to indel mutations and other sequence changes (Dover, 1982;

Tautz et al, 1986; Hoelzel et al, 1993). Lastly, because two-thirds of the indels would disrupt ORFs in mitochondrial structural proteins, the lineage of these mutations likely derive from the *Numt* sequence.

Variable Evolutionary Rates of Individual Mitochondrial Genes among Mammals

To quantify relative divergence rates of mitochondrial genes for molecular evolutionary studies, mean percent similarities in DNA and amino acid sequences for all 37 mtDNA genes and the CR within five mammalian orders (rodents, primates, carnivores, artiodactyls, and cetaceans) were calculated (Tables 7-9). In addition, the average divergence of mammalian mtDNA genes from an outgroup species, chicken (*Gallus gallus*) was also computed. The most consistently conserved mitochondrial genes were COI, COIII, and 12S rRNA sequences, while the most rapidly evolving mtDNA sequences were both 5' and 3' CR termini, the ATPase 8, ND2 and ND6 genes (Table 7, column I). At the DNA level, the 5' L-strand (left) CR was the most rapidly evolving region between closely-related pairs of taxa (i.e. both seals, both whales, seal-cat), while the 3' CR became almost equally divergent in more distant comparisons. Sequence divergence from the outgroup, chicken mtDNA (Table 7, column II), did not always produce the same ranking as mammalian-only estimates (Table 7, column I). For example, the 12S rRNA gene and the conserved core of the control region showed greater divergence from chicken mtDNA, while the ND4 and ND5 genes exhibited higher overall conservation. The 16S rRNA gene also had highly

Table 7 Pairwise comparisons of nucleic acid sequence similarity in Mitochondrial genes and the Control region (CR).*

		I				II		III		IV		V	VI
		mammal^		Std.		mammal - chicken []		cat -		mouse - rat		harbor seal -	fin whale -
region	size	average	range	Dev.	order #	average	order #	harbor seal	order #		order #	grey seal	blue whale
CR-3' +	1014	47	28	9	1	42	1	73	2	70	1	96	97
CR-5' +	576	48	33	13	2	44	2	52	1	79	4	89	87
ATP8	204	67	17	5	3	51	3	76	3	80	5	97	93
ND2	1044	67	15	5	4	58	6	78	5	75	2	96	90
ND6	549	69	18	6	5	54	5	80	9	81	10	94	91
ND5	1851	70	11	4	6	63	11	76	4	77	3	96	91
ND3	350	72	12	5	7	60	8	80	10	80	6	98	91
ND4	1378	72	12	3	8	63	12	78	6	80	7	95	90
ND4L	297	73	14	5	9	58	7	79	7	81	9	98	91
ATP6	684	73	7	2	10	61	9	80	8	85	15	97	91
16S	1669	75	14	5	11	62	10	86	17	80	8	97	95
ND1	978	76	10	3	12	70	15	84	15	81	11	97	93
CR-C +	264	76	20	6	13	52	4	86	18	87	17	98	99
CO2	684	76	13	5	14	69	14	82	12	85	16	96	92
CYB	1146	77	7	3	15	73	17	81	11	84	14	95	92
12S	1010	78	8	2	16	68	13	84	16	92	18	98	95
CO3	786	78	6	2	17	72	16	82	14	82	12	96	93
CO1	1554	79	3	1	18	77	18	82	13	83	13	96	93

*-A UPGMA alignment algorithm with gap weight=5 and gap length weight=0.3 was used except for the CR where stringency was reduced to a gap weight=1 and gap length weight=0.05. For similarity calculations a gap penalty of 1 substitution was used.

^-Average percent similarity of pairwise comparisons between a representative species from 5 mammalian orders (human, domestic cat, fin whale, cow & mouse)

[]-Average percent similarity between chicken and a representative species from 5 mammalian orders (human, domestic cat, fin whale, cow, and mouse).

+-The CR was split into three regions, the highly conserved region (CR-C) and the two variable regions located 5' and 3' to the conserved region (CR-5' nd CR-3' respectively), relative to light strand replication.

#-The order was determined by using four decimal places to order the similarity values. An order number of 1 was assigned to the least similar comparison and 18 was assigned to the most similar comparison. Ties in order number occurred in the cat-harbor seal and mouse-rat columns, in both cases the mammal order number was used to break the tie.

___ -A cluster analysis was performed on the mammal comparisons and the clusters with significant differences occur above and below the line.

Table 8. Pairwise comparisons of nucleic acid sequence similarity in Mitochondrial tRNA genes.*

		I				II		III		IV		V	VI
		mammal^		Std.		mammal - chicken []		cat -		mouse - rat		harbor seal -	fin whale -
region	size	average	range	Dev	order #	average	order #	harbor seal	order #		order #	grey seal	blue whale
tTHR	72	68	24	9	1	48	1	83	7	84	3	96	93
tCYS	69	68	33	15	2	59	5	91	17	87	7	99	97
tVAL	73	68	23	7	3	58	4	85	9	87	8	99	97
tGLY	70	70	28	11	4	63	12	73	1	88	9	100	97
tSER (AGY)	67	72	30	11	5	54	3	97	20	72	1	98	95
tPRO	70	72	17	6	6	60	8	82	5	91	15	100	94
tGLN	84	73	24	8	7	63	13	89	14	85	5	99	100
tTRP	76	75	34	11	8	59	7	81	4	91	14	97	99
tLYS	72	76	20	6	9	60	9	88	11	85	6	96	94
tPHE	73	77	19	5	10	59	6	87	10	82	2	99	86
tASP	70	78	17	5	11	67	14	82	6	84	4	100	100
tHIS	70	79	14	5	12	73	19	80	3	90	12	99	93
tARG	69	80	19	6	13	52	2	88	12	93	17	100	97
tASN	75	80	10	4	14	69	16	85	8	95	20	100	100
tSER (UCN)	76	82	18	5	15	69	15	92	18	90	13	99	94
tLEU (UUR)	75	82	13	5	16	74	20	89	15	97	22	100	97
tALA	69	84	16	2	17	73	18	90	16	88	10	100	97
tGLU	69	84	18	6	18	62	10	88	13	94	18	99	100
tTYR	71	90	4	2	19	72	17	79	2	90	11	93	100
tLEU (CUN)	72	90	7	3	20	84	21	99	21	96	21	100	99
tILE	72	92	7	2	21	62	11	96	19	91	16	99	99
tMET	69	95	6	2	22	91	22	97	22	94	19	99	99

*-A UPGMA alignment algorithm with gap weight=5 and gap length weight=0.3 was used except for the CR where stringency was reduced to a gap weight=1 and gap length weight=0.05. For similarity calculations a gap penalty of 1 substitution was used.

^-Average percent similarity of pairwise comparisons between a representative species from 5 mammalian orders (human, domestic cat, fin whale, cow, and mouse).

[]-Average percent similarity between chicken and a representative species from 5 mammalian orders (human, domestic cat, fin whale, cow, and mouse).

#-The order was determined by using four decimal places to order the similarity values. An order number of 1 was assigned to the least similar comparison and 18 was assigned to the most similar comparison. Ties in the order number occurred in the mammal, cat-harbor seal and mouse-rat columns. The mammal-chicken order number was used to break the mammal tie, the mammal order number was used to break the other ties.

variable rates between different lineages (e.g 80% conservation in the mouse-rat pair vs. 95% between whales and 97% seals). The tRNA genes showed the greatest fluctuation in sequence similarity between lineages (Table 8). Also, the lowest Ti:Tv ratios were seen in the CR, and highest ratios were observed in the tRNA genes (avg. ratio of 0.6 in the CR, 0.91 in protein; 0.92 in rRNA and 1.55 in the tRNA genes).

In the amino acid comparison (Table 9), the ATPase subunit 8 gene was the most variable mitochondrial gene with the lowest mean similarity of 55% (range = 34% pts) among all five mammalian species. The ND6 and ND2 genes followed with 63% mean similarity. As seen for nucleotide sequences, the CO subunit amino acid sequences were ranked high in overall sequence similarity with the COI gene (93% similarity) as the most consistently conserved gene followed by the COIII gene (88%). The cyt b gene was also relatively slowly evolving with 83% mean similarity, although a slight rate increase may have occurred in the carnivores (Table 9, column III). On the other hand, the ND4L gene was highly conserved between seals and cats relative to all mammals ($p < 0.01$).

Phylogenetic analysis with feline mtDNA and *Numt* 16S rRNA sequences was conducted to show its relationship with the other mammalian mtDNA sequences (Fig. 15). A maximum parsimony topology extends the conclusion of Janke et al (1994), with additional cat sequences, and recapitulates phylogenetic relationships produced with other algorithms (e.g neighbor-joining, maximum likelihood) (Felsenstein, 1993). We used the 16S rRNA gene, since missense mutations in *Numt* usually obliterated any

identity (but not homology) after alignment of the translated amino acid sequences. Nevertheless, other conserved mitochondrial genes, such as COI and COIII, produced branching hierarchies similar to the 16S rRNA results, which show the affinity of the two cat mtDNAs with each other and the seal sequences, as well as the recapitulation of an artiodactyl-carnivore grouping (Li et al, 1990).

Table 9. Pairwise Percent Similarity (PS) in Amino Acid Sequences of Mt Genes among Five Mammalian Orders.

			I			II		III		IV		V	VI
Gene	Size (# res.)	Mammals (All Orders)*	Range^^ of PS	Std Dev.	Order #	Chicken vs. Mammals**	Order #	Cat vs. H. Seal	Order #	Rat vs. Mouse	Order #	H. Seal G. Seal	F. whale B. whale
ATPase 8	67	55	29	9.4	1	34	1	72	1	80	4	96	92
ND 6	175	63	30	11.1	2	37	2	85	4	79	3	97	94
ND 2	347	63	24	8.5	3	45	3	76	2	75	1	98	95
ND 5	606	70	18	6.1	4	56	6	82	3	78	2	98	96
ND 4L	98	73	20	7.8	5	46	4	94	10	85	6	100	93
ND 3	115	75	24	9	6	57	7	89	6	86	8	99	100
ND 4	459	76	20	6.8	7	59	8	88	5	87	9	98	97
ATPase 6	227	78	12	3.6	8	55	5	92	8	85	7	97	95
ND 1	318	81	13	5.1	9	72	10	92	9	81	5	99	99
Cyt B	379	83	10	3.9	10	73	11	89	7	94	10	97	97
CO II	227	85	26	10.6	11	67	9	96	12	99	13	100	99
CO III	261	88	7	2.7	12	75	12	94	11	96	11	99	99
CO I	514	93	6	2	13	87	13	98	13	97	12	99	99

- Listings are ordered from least (#1) to highest percent similarity (#13) among mt genes.
*- All statistics (mean, range, std. dev.) are based on values derived from all pairwise comparisons between human, mouse, fin whale, cat, and cow mt genes.
- Gap penalties equivalent to one substitution were imposed on all similarity calculations by CMATRIX.
**- Chicken (Gallus gallus) sequences derived from (Desjardins and Morais, 1990)
- ^^The range statistic may be viewed as a measure of consistency, proportional to the amount of rate variation among the five lineages.
- Ties were resolved by carrying out numbers to the fourth decimal place.

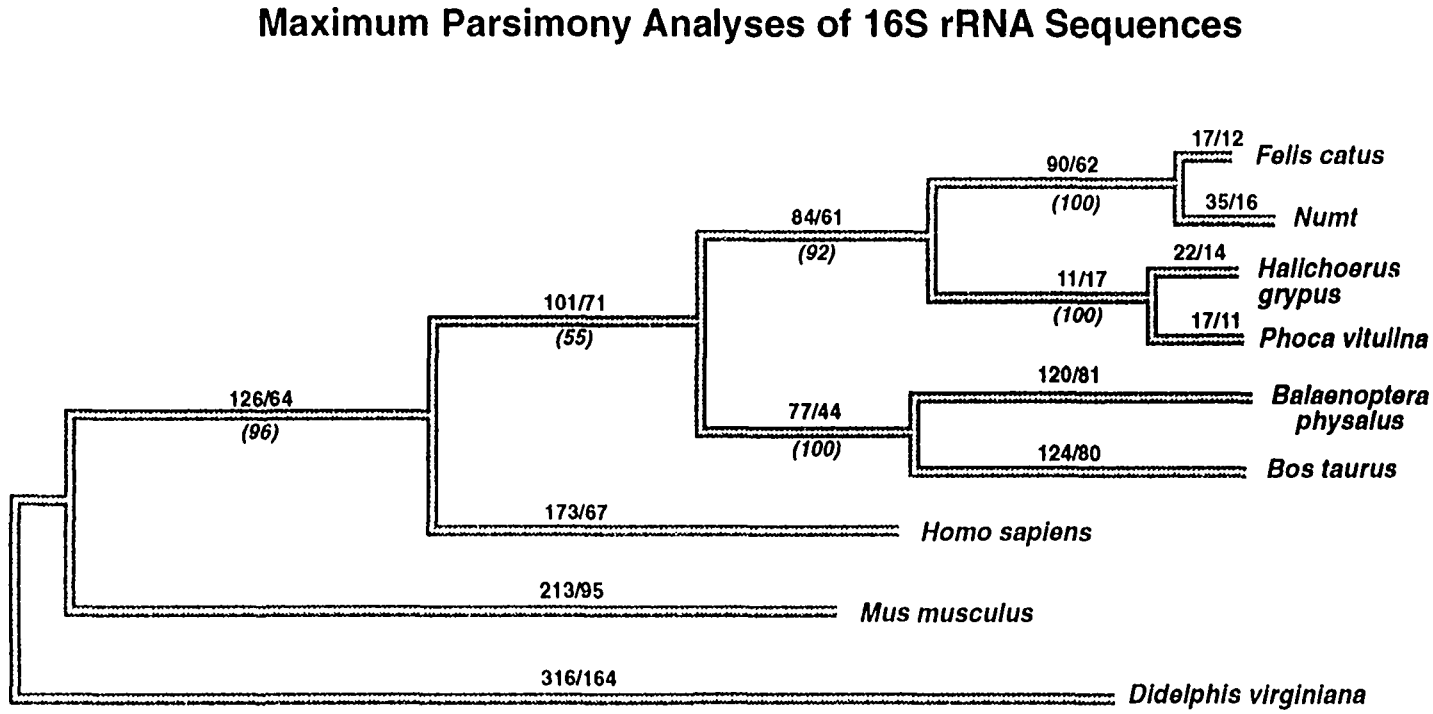


Fig. 16. Phylogenetic reconstruction with total 16S rRNA gene sequences. The 50% majority-rule consensus tree was created with PAUP 3.1.1 employing unweighted maximum parsimony criteria and branch-swapping options (Swofford, 1993). Alignment of the complete gene (ca. 1600 bp) from the respective taxa was done with default parameters (gap weight = 3.0) of PILEUP in GCG. The tree length equals 1615 steps, with consistency index of 0.755. Numbers above the branches designate the number of total changes/homoplasies. Bootstrap percentages in support of each node from 100 iterations are shown below the branches in italics and parentheses for each node.

DISCUSSION

Domestic cat mtDNA can be distinguished from other mammalian mtDNA sequences by its possession of a large, 7.9 kb tandemly repeated, homologue in the nuclear genome, termed *Numt* (Lopez et al, 1994). The remarkable transposition and amplification of *Numt* provides the opportunity to examine evolutionary leitmotifs, such as inter-organellar substitution rate differences, mutational polarity and frequency and synonymous substitution rates within both mitochondrial and nuclear compartments. Evidence that mtDNA genes can transpose to the nucleus, creates the alluring prospect of directly analyzing intracellular (paralogous) duplication events (Goodman, 1981; Hardison and Gelinas, 1986; Fukuda et al, 1985; Smith et al, 1991; Zullo et al., 1991).

Structural aspects of cytoplasmic mtDNA organization

In the cytoplasmic mitochondrial genome, the control region is longer than average due to two repetitive motifs, RS2 and RS3, at opposite ends, but its length does not exceed the 1838 bp CR of lagomorphs (Mignotte et al 1990; Biju-Duval, 1991). The compact vertebrate mitochondrial genome structure as defined by Attardi (1985) is probably maintained by selective pressures and therefore may limit the accrual of novel features such as CR simple repeats. Furthermore, length variation at homopolymer tracts has recently been associated with molecular drive mechanisms in

the *Drosophila* gene, *mastermind* (Dover, 1982; Newfeld et al, 1994), and DNA repeats in non-CR regions have been implicated in various mtDNA deletions associated with human disease states (Wallace, 1992). Other evolutionary explanations for mtDNA repeats are described elsewhere (Hoelzel, 1993; Hoelzel et al, 1994; Ghivizzani et al, 1993; Buroker et al, 1990; Rand, 1993). With respect to codon usage, base composition, gene order and gene size, and site of mutations, the cytoplasmic mtDNA sequence of *F. catus* conforms to most placental mammalian mtDNA genomes (Gadeleta et al, 1989; Anderson et al 1982; Kumazawa et al, 1994).

Substitution patterns between homologous felid mtDNAs

Several salient reasons indicate the transcriptional silence of *Numt* genes, including the disruption of ORFs by indels, genetic code differences between organelles, and the resemblance of the *Numt* locus to satellite DNA and likely proximity to heterochromatin-enriched centromeric regions (Lopez et al, 1994; Charlesworth et al, 1994). Therefore, *Numt* is expected to resemble pseudogene evolution through a stochastic pattern of substitution, presumably after its "release" from functional constraints (Kimura, 1983). Nevertheless, *Numt* can proceed under several possible modes of evolution relative to cytoplasmic mtDNA:

a) *Numt* and cytoplasmic mtDNA evolve at the same rate;

b) *Numt* evolution is virtually "frozen" compared with the rapid (up to 10X) substitution rate of cytoplasmic mtDNA (Brown et al, 1982, Nei, 1987);

c) *Numt* as a pseudogene evolves faster than the mitochondrial genome.

To determine which of the above models best describes the evolution of *Numt* and cytoplasmic mtDNA in the cat, we performed relative rate tests (Wu and Li, 1985) with both cat mtDNA homologues using gene sequences from ocelot or panthera lineage species (ocelots, lions etc.) as outgroups (Table 10). For 12S rRNA sequences, the relative rate tests indicated that *Numt* sequences were evolving faster (not statistically significant at $p = 1.0$) than cytoplasmic mtDNA, supporting previous conclusions and the topology showing unequal branch lengths leading to *Numt* (Lopez et al, 1994). Using the more rapidly evolving 16S rRNA (Table 7), however, *Numt* and cytoplasmic mtDNA substitution rates appeared nearly equivalent in comparison to various outgroup felids (Table 10).

The relative rate results, together with the analysis of variation among different mitochondrial genes (Tables 7-9, discussed below), provide an important corollary to the assumptions of rate variation between nuclear and cytoplasmic mtDNAs: namely, each model can realistically reflect the situation between nuclear and cytoplasmic mtDNAs, but ultimately depends on the specific mitochondrial gene being compared. Moreover, by default, the more rapidly evolving mtDNA homologue will more

Table 10. Comparison of Relative Evolutionary Rates between Numt and Cytoplasmic MtDNA in Two Different Mitochondrial Genes and with Diverse Outgroup Taxa.

	MAXIMUM PARSIMONY ^					
	12S rRNA			16S rRNA		
	% DNA Similarity			% DNA Similarity		
Outgroup species	Numt	FCA-Cyto	Ratio*	Numt	FCA-Cyto	Ratio*
Pallas cat (OMA)	16.0	3.0	5.3	19.0	6.0	3.2
Lion (PLE)	16.0	3.0	5.3	14.0	12.0	1.2
Tiger (PTI)	22.0	3.0	7.3	13.0	13.0	1.0
Puma (PCO)	16.0	3.0	5.3	19.0	6.0	3.2
Leopard cat (PBE)	16.0	3.0	5.3	18.0	7.0	2.6
Ocelot (LPA)	16.0	3.0	5.3	18.0	7.0	2.6
Kodkod (OGU)	15.0	4.0	3.8	19.0	6.0	3.2
Combined Taxa	16.0	3.0	5.3	13.0	13.0	1.0

* - Represents the ratio of Numt : Cytoplasmic mtDNA branch lengths.
- Numerical values represent the branch lengths in character changes (parsimony analysis) leading to either Numt or to the corresponding cytoplasmic mtDNA sequence. Parsimony analysis more effectively "filtered out" the variation on the internode branch leading from the outgroup to the bifurcating feline sequences.
- Although different individuals may have been used for the different genes, comparisons between Numt and cytoplasmic mtDNA were determined within one individual of the species.
^ - Weighted parsimony analysis was performed with each outgroup, using PAUP 3.1.1 (Swofford, 1993). Transitions were typically weighted 5X more than transversions.

profoundly affect the overall pattern of the gene being compared. Also, higher overall mutation frequencies (ca. 70%) observed at third, versus first or second, codon positions in structural mitochondrial genes re-emphasize the presence of selected, functional constraints on cytoplasmic mtDNA substitutions (Table 5; Brown et al, 1982; Miyata et al, 1982). For gene sequences (e.g. 12S rRNA, COI) which evolve slowly in the mitochondria, the nuclear pseudogene rate of *Numt* will appear rapid relative to the cytoplasmic rate.

These interpretations of rate variation in mitochondrial genes and between organelles can also be viewed in the context of recent arguments in support of equivalent substitution rates between the nucleus and mitochondrial (mode a) (Lynch and Jarrell, 1993) or unequal rates (modes b or c) (Slade et al, 1994). Besides the cat data, however, we are unaware of examples demonstrating the third mode of faster evolution at nuclear loci in mammals.

As a secondary observation in the above analysis, similar proportions of mutations (ca. 70%) at the mostly silent third codon position in four different genes (Table 5) provide indirect evidence for constant synonymous substitution rates within felid mitochondrial genome. The constancy of synonymous substitution rates among coding genes continues to spark investigation and discourse (Langley and Fitch, 1974; Gojobori et al, 1982; Li et al, 1985; Ticher and Graur, 1989; Wolfe et al, 1989; Bernardi et al, 1993; Kondo et al., 1993).

Dating the birth of Numt in the cat nuclear genome

Choice of the most appropriate model (a-c) has a bearing on estimates for the precise date of *Numt* integration into the cat nuclear genome. Using model (a) with equivalent rates between *Numt* and cytoplasmic mtDNA probably facilitates the simplest method for dating *Numt* integration. For example, with preliminary data indicating 4.4% mean 16S rRNA sequence divergence among several of the large cats of Panthera genus (lion, leopard, jaguar, clouded leopard, snow leopard), whose divergence times are known from fossils at 1.8 - 3.0 MYA, a date of only 1.0 - 1.6 MY is calculated for the birth of *Numt* (Wayne et al, 1989; 1991; Janczewski et al, in press; Johnson et al, in preparation). Alternatively, a previously calculated value of about 1.8 MY for the birth of *Numt* was based on the second assumption of a faster cytoplasmic rate and 5.3% total divergence between *Numt* and cytoplasmic mtDNA (Lopez et al, 1994). With a final value of 5.1% total divergence presented in this study, the date of integration decreases to 1.7 MY, using the same calculation. These values define a fairly robust time frame from 1.0 - 1.7 MY since the original *Numt* integration within an ancestral Felis species.

A curiosity in the comparison of nucleotide substitution patterns between the three closely related mammalian pairs is the finding of lower bias against transversions and net decreases in dA and dT content in *Numt* relative to felid cytoplasmic mtDNA (Table 6), which challenges expectations for pseudogenes and non-coding sequence

evolution (Gojobori 1982; Li et al, 1984, 1985). To help explain the disparity, elevated Ti:Tv ratios are usually more pronounced in mtDNA rather than nuclear sequence comparisons (DeSalle et al, 1987). The disparities in mutational spectra probably also relate to the exceptionally dissimilar cellular environments within the mitochondria and the nucleus, which encompass differences in amounts of oxidative damage to DNA, the presence or absence of different enzymes and DNA polymerases involved in repair processes, and the physical structure of the double helix *in vivo* (Clayton, 1991; McBride et al, 1992; Miquel, 1992; Wallace, 1992). For example, Feig and Loeb (1993) observed a relationship between hypermutable nucleotide hotspots and the "pausing" of mammalian beta-polymerase at specific DNA secondary structures. Furthermore, since DNA repair of the nuclear genes may be directly influenced by the presence/absence of methylated residues (Hare and Taylor, 1985), the integration of *Numt* with a novel or non-existent pattern of methylation could dramatically affect repair mechanisms. However, the paucity of comprehensive studies directly quantifying methylation in mitochondrial genomes (Pollack et al, 1984; Mazin et al, 1988) restrict the capability of deriving general or reliable interpretations of these mutation patterns between *Numt* and cytoplasmic mtDNA. Although overall DNA content is expected to become dA/dT-rich, due to the action of common mutagenic agents, it is possible that an insufficient amount of time has elapsed since the relatively recent divergence of *Numt* from cytoplasmic mtDNA for the expected accumulations. Lastly, the ancestral

population of cytoplasmic mtDNAs that donated the original *Numt* fragment, may have been polymorphic and possessed an a priori bias in base composition.

Mutation Rates of Individual Mitochondrial Genes within Functional Genomes

Tight linkage between genes on the same mitochondrial genome does not predispose them to the same rate. Although similar conclusions have been previously derived using more rigorous statistical methods (Adachi et al, 1993; Lynch and Jarrell, 1993), the present analysis of relative sequence similarities and mutational dynamics in functional mitochondrial genes, was broad in scope and verifies rate heterogeneity between i) different mammalian phyletic lineages ii) different organelles (previous section) and iii) different mitochondrial genes within the same genome, especially with respect to nonsynonymous substitutions (Tables 7-9). Since it could impose a strong influence on the interpretation of rate heterogeneity between different mitochondrial genes from divergent lineages (Li et al, 1985), base composition and codon usage in feline mtDNA was assessed but did not show any major deviations compared to harbor seal (Table 4) or other mammals. In general, primates (e.g. human mtDNA) represent the only lineage which significantly differs in both codon usage and base composition among five mammalian orders (Janke et al, 1994).

Within single genomes, understanding the variation in evolutionary rates between distinct mitochondrial genes is similarly meaningful for phylogenetic studies, since genes with higher substitution rates will better distinguish closely related taxa, while

slower genes can resolve older bifurcations (Mindell and Honeycutt, 1990; Wainright et al, 1993; Ruvulo et al 1993; Lynch and Jarrell, 1993; Graybeal, 1994; Slade et al, 1994). Among five mammalian orders, mitochondrial genes can be divided into three groups with either very fast (CR), fast (ATPase 8, ND2, ND6 etc.) and slow (12S rRNA, COI etc.) substitution rates. Although others have demonstrated that amino acid substitutions accumulate more linearly compared to the DNA level for distantly related phyla (Brown et al, 1982; Brown, 1985; Lynch and Jarrell, 1994; Adachi et al, 1993), pairwise comparisons of amino acid similarities for many mitochondrial genes have higher variances and ranges (Table 9) compared to the same statistics at the DNA level (Tables 7, 8). This likely reflects the expected mutational saturation for many mtDNA sequences in the time frame of mammalian radiations spanning about 80 MY (Carroll, 1988; Novacek, 1992; Graybeal, 1994).

Each gene carries its own peculiar biases in base composition (Wolfe et al, 1989; Kondo et al, 1993; Bernardi et al, 1993; Adachi et al, 1993), codon preferences (Li et al, 1985; Moriyama and Hartl, 1993), hypermutable or invariant sites (Palumbi, 1989), response to selective or mutation pressures etc. (Kimura, 1983; Gillespie, 1986; 1991; Hasegawa and Kishino, 1989), which can all significantly affect substitution rates. Overall, these data support the contention that molecular evolution or phylogenetic studies view each mitochondrial gene as distinct units with various levels of rate heterogeneity.

CHAPTER 3

TRANSCRIPTIONAL INACTIVITY OF THE NUMT LOCUS

BACKGROUND

Non sum qualis eram - "I am not what I used to be"

Several compelling reasons support the transcriptional silence of *Numt* locus genes, such as the differences in genetic code between organelles, (Barrell, 1980; Attardi, 1985), indel mutations, and a computer searches of the *Numt* sequence revealing a paucity of potential ORFs (Fig. 16). Nonetheless, the empirical verification of *Numt* transcriptional inactivity was a major objective of this dissertation, because bestowing pseudogene status on *Numt* genes can profoundly affect the implications and interpretations of evolutionary patterns, such as *Numt*'s potential utility as a neutral reference for calibrating felid-specific mtDNA mutation rates, or as a novel source of mitochondrial proteins. The last point bears directly on a primary (but rarely addressed) assumption of SET, which is that the transfer and successful integration of genetic information between symbiotic organelles should be immediately followed by their activation and expression at the new locus (Gray 1989; Margulis, 1981; Margulis, 1993). Also, Li (1985) has postulated that "with the non-functionality of pseudogenes,

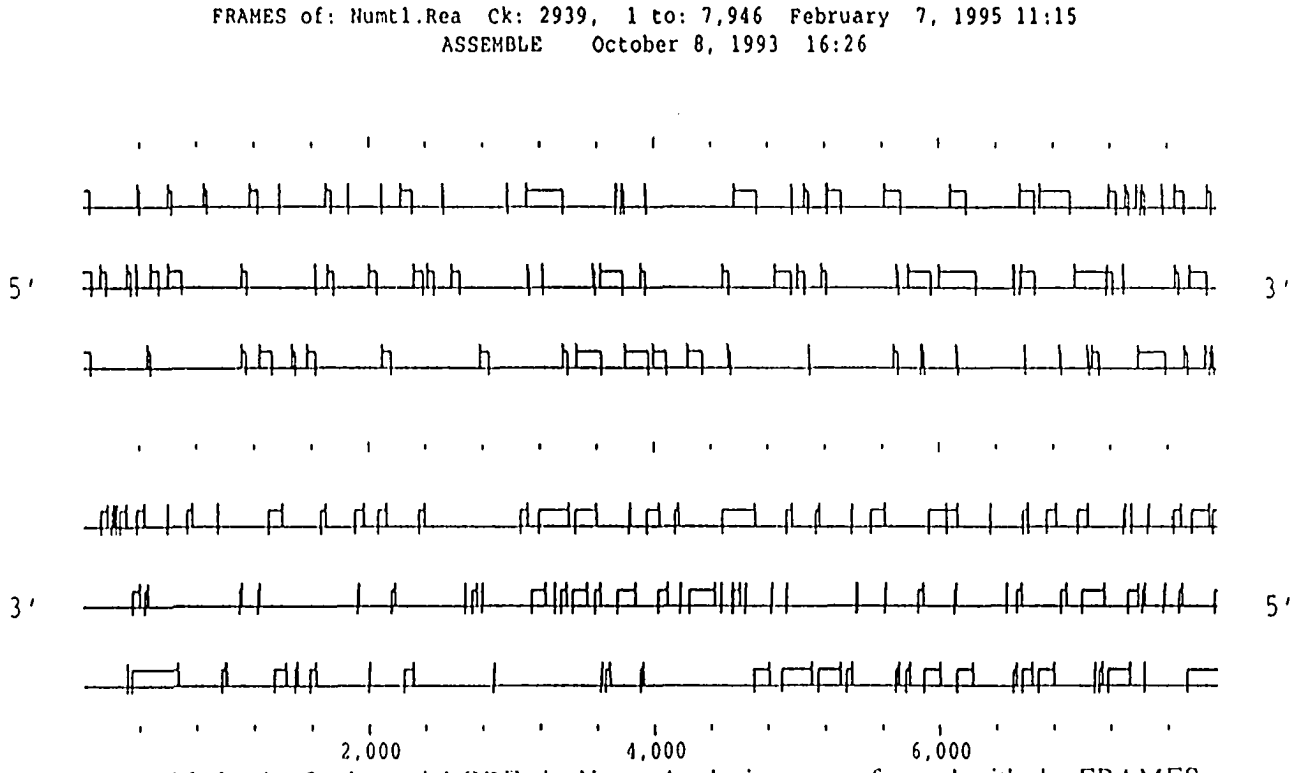


Fig. 16. Lack of substantial ORFs in *Numt*. Analysis was performed with the FRAMES program of UWGCG, using the nuclear genetic code, all six possible frames and the complete 7.9 kb *Numt* sequence in Fig. 12. Potential start codons are indicated by lines extending above the reading frame box and stop codons are denoted by lines extending below the box. Numbering corresponds to the mtDNA sequence in Fig. 12. The longest frames observed were < 400 bp and did not correspond to expected canonical mtDNA ORFs.

all mutations in them will be selectively neutral, and therefore become fixed in the population with equal probability. Thus the pattern of nucleotide substitution in pseudogenes should reflect the pattern of spontaneous point mutation." Lastly, since only one unit (p*Numt*.1) of the multimeric *Numt* tandem array was effectively characterized in previous chapters, the possibility remains that other repeat units at the *Numt* "macrosatellite" locus could be transcribed and translated by mutations which eradicate apparent nonsense differences in p*Numt*.1. In fact, evidence exists that apparently "functionless" satellite DNA can be transcribed if they are transposed and read through by illegitimate promoters or visa versa (Miklos, 1985; Sealy et al, 1981).

The high sequence conservation between *Numt* and cytoplasmic mtDNA sequences (ca. 95%), however, makes the differentiation of mtDNA gene expression from either source non-trivial. For example, the precise origin of RNA transcripts immobilized on Northern ("filter") blots (Sambrook et al, 1989) would be difficult to discern, since probes could theoretically cross-hybridize to both mtDNA complements under standard conditions.

Therefore, I chose to employ the ribonuclease protection assay (RPA) to identify specific mRNA transcripts (Fig. 17; Myers et al, 1985; Murakawa et al, 1988). Briefly, the RPA involves "solution" hybridization of RNA probe molecules to cellular mRNA, which is preferable to filter methods due to more predictable hybridization kinetics, higher sensitivity and increased capacity for quantifying numbers of hybrid RNA molecules. *Numt* and cytoplasmic mtDNA-specific RNA probes can be

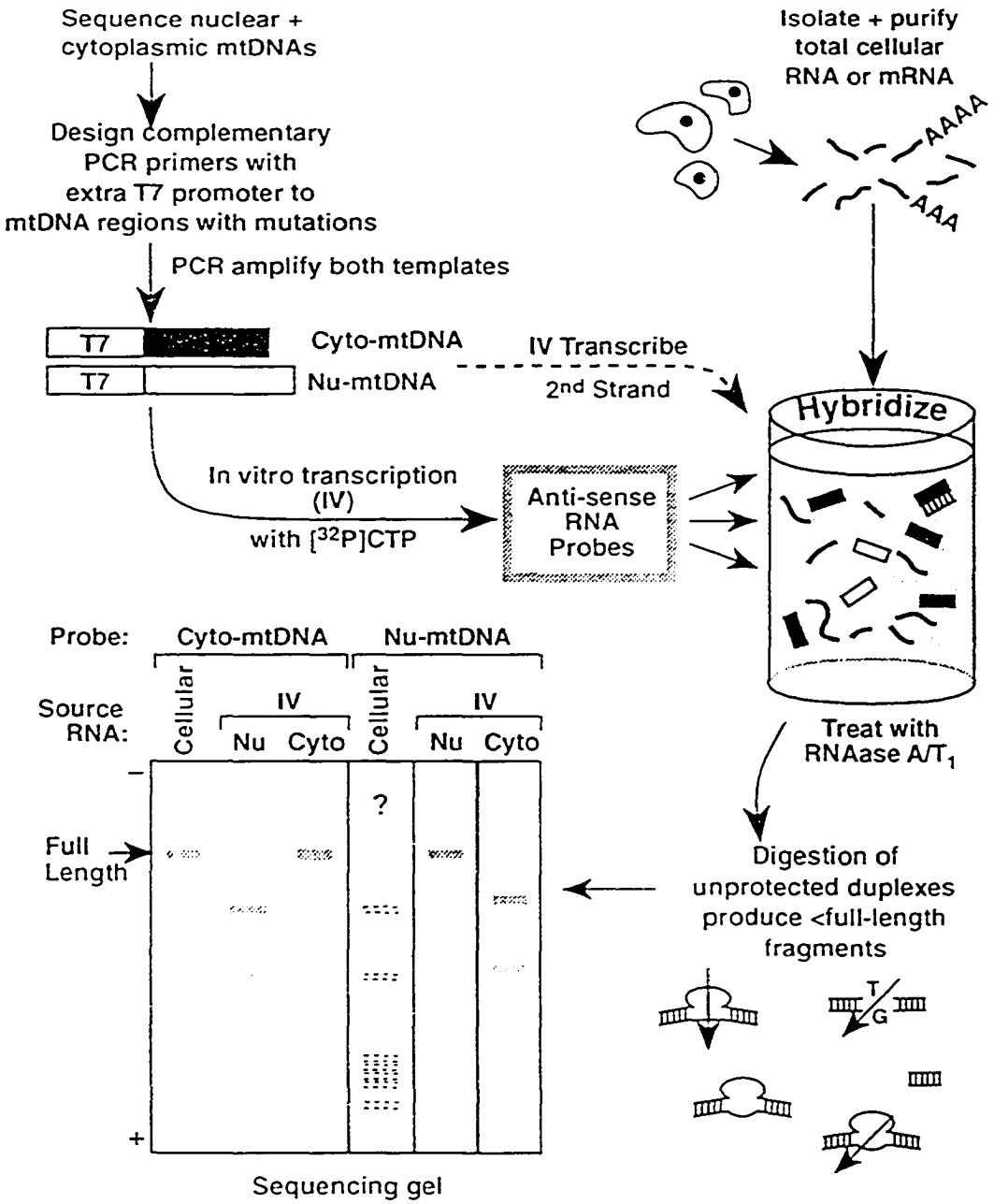


Fig. 17 RNAase Protection Assay designed to differentiate cytoplasmic and Numt mtDNA transcripts.

```
      1362
CYTO  GGGATTAGATACCCCACTAT GCTTAGCCCTAAACTTAGAT AGTTACCCTAAACAAAACTA
NUMT  .................... .................... .....T..............

      1422
CYTO  TCCGCCAGAGAACTACTAGC AATAGCTTAAAACTCAAAGG ACTTGGCGGTGCTTTACATC
NUMT  ................C... ..C......G.......... ....................

      1482
CYTO  CCTCTAGAGGAGCCTGTTCT ATAATCGATAAACCCCGATA TACCTCACCATCTCTTGCTA
NUMT  .................... .................... ....................

      1542
CYTO  ATTCAGCCTATATACCGCCA TCTTCAGCAAACCCTAAAAA GGAAGAAAAGTAAGCACAAG
NUMT  ................G... ................G... ....................

      1602
CYTO  TATCTTAA*CATAAAAAAGT TAGGTCTAGGTGTAGCTCAT GAGATGGGAAGCAATGGGCT
NUMT  ........C.CA........ .................... ....................

      1661
CYTO  ACATTTTCTAAAATTAGAAC ACCCACGAAGATCCTTACGA AACTAAGTATTAAAGGAGGA
NUMT  ...C..............T .A..............T.. ..T.................

      1728
TTTAGTAG  <-- T7 PRO
........
```

Fig. 18. Alignment of homologous cytoplasmic (top) and Numt (bottom) 12S rRNA gene regions selected for RNAse protection assays. Symbols and terminology are identical to those in Fig. 19.

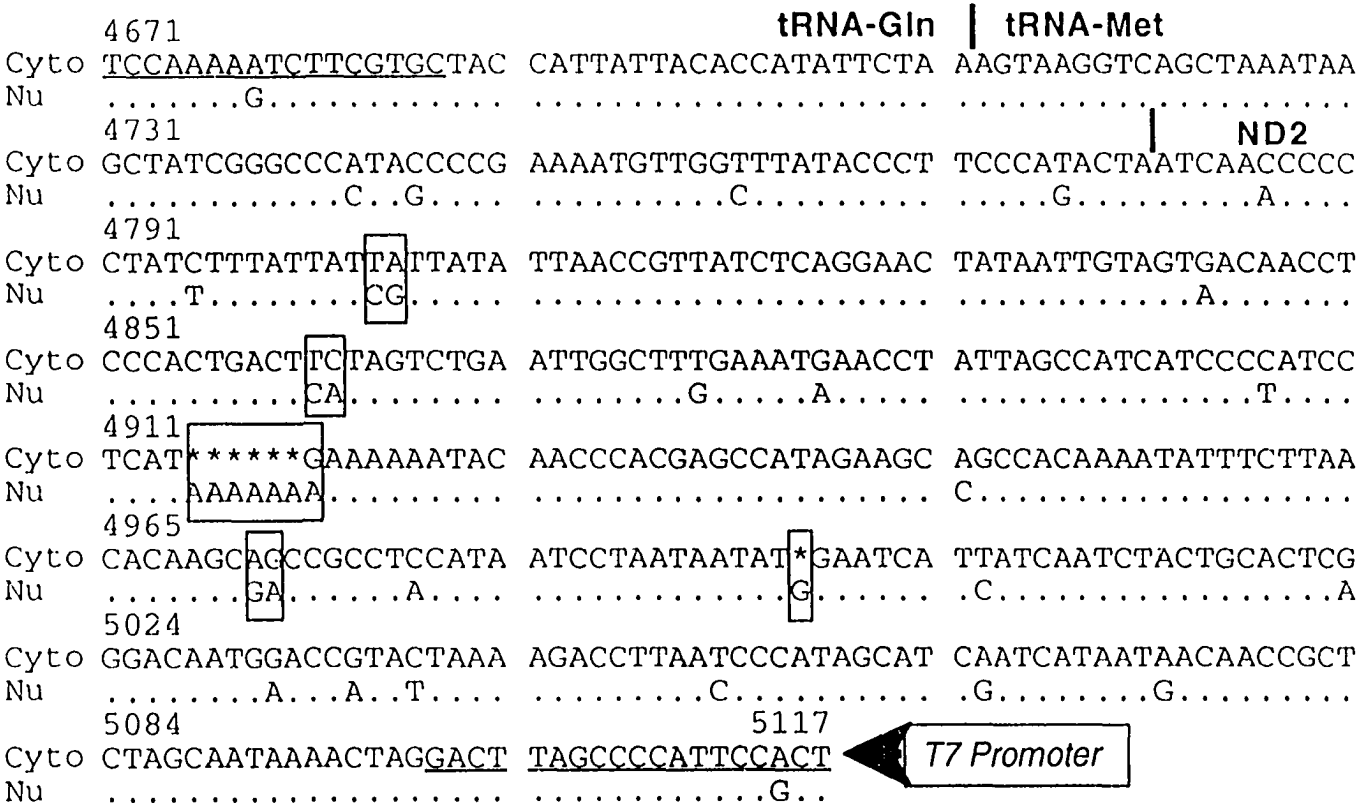


Fig. 19. Alignment of homologous NADH dehydrogenase subunit 2 (ND2), tRNA-Gln and tRNA-Met gene regions in cytoplasmic (cyto) and Numt (Nu) mtDNA. Primer sequences used for PCR amplification of templates for anti-sense RNA probes, were originally derived from pNumt.1 sequences and are underlined. The single bp mismatch in the primers did not inhibit amplification from cytoplasmic pCmt.12, when employing standard reaction conditions (Innis et al, 1990) and 50^{o}C annealing temperatures. The box at the 3' terminus marks the attachment of T7 promoter sequences necessary for *in vitro* transcription of the PCR product. Internal boxes highlight gap mutations. Bars indicate gene boundaries and nucleotide positions correspond to the cat mtDNA sequence (Fig. 12).

synthesized directly from templates derived from either p*Numt*.1 or pCmt.12 clones by PCR (Stoflet et al, 1988; Murakawa et al, 1988; Innis et al, 1990). For the present experiments, three mitochondrial gene regions (12S rRNA, ND1 and ND2) were chosen for analysis, since they contain a relatively high number of mutations (especially indels) between *Numt* and cytoplasmic mtDNA. By the attachment of a T7 promoter sequence to the appropriate PCR primers, amplification products can be transcribed *in vitro* to produce "anti-sense" RNA probes which are complementary to cellular mRNAs predicted from the sequence alignments (Fig. 18, 19). After radioisotopic labeling, these synthetic probes are hybridized to either poly-(A+) mRNA, total cellular RNA, or control *in vitro* RNA transcripts. Hybrid RNA duplexes are then subjected to over digestion with ribonucleases T1 and A to detect mismatches. Fragment sizes should correspond to completely protected duplexes or the distances between cleaved mismatches. For example, full-length fragments would result from complete protection of *Numt* or cytoplasmic mtDNA probe by cellular transcripts (Fig. 17). More complicated patterns of protected fragments are possible when probes span tRNA gene boundaries involved in processing polycistronic mtDNA mRNAs (Clayton, 1984; 1991).

The experimental regimen of RPA aims to demonstrate the plausibility of detecting either cytoplasmic mtDNA or *Numt*-specific RNA messages. Absence of the latter in the context of the cytoplasmic controls, will permit the rejection of *Numt* transcription (null hypothesis), and consequently support the alternative hypothesis that *Numt* is an unexpressed pseudogene.

MATERIALS AND METHODS

RNA extraction and preparation. Total cellular RNA was extracted from fresh or snap-frozen kidneys or liver of a sacrificed cat (FCA 65) using the standard guanidine salt method and the following modifications (Sambrook et al, 1989). Tissue was ground with mortar and pestle in liquid nitrogen, added to 20 ml of 8 M guanidine-HCL and homogenized (one pass) and then spun at 12,000 RPM for ten minutes. The supernatant was filtered through 25 micron cloth to remove lipid and precipitated with 0.1 vol of 3M sodium acetate and 0.5 volume cold 95% ethanol at -20°C for > 2 hours. The pellet was resuspended in 10 mls 4 M guanidine HCL and precipitated two more times and then dissolved in 0.02 mM EDTA for one round of chloroform extraction. NaCl was then added to 150 mM (in ten mls), and the solution was heated to 56°C for 10 minutes and cooled on ice. After spinning at 12,000 RPM for 10 minutes the supernatant was saved and stored by precipitating with ethanol overnight. Poly-A mRNA was either derived from total RNA samples or from fresh tissue using protocols of commercial isolation kits (Invitrogen). The integrity of purified RNAs was evaluated by formaldehyde gel electrophoresis and Northern blotting (Sambrook et al, 1989).

Ribonuclease Protection Assays (RPA). The mitochondrial 12S rRNA, ND2 and ND1 subunit gene regions were chosen for assays after determining that a sufficient number of mutations between p*Numt*1 and cytoplasmic pCmt.12 clones existed (Fig. 12; Table 2). The *Numt* and cytoplasmic mtDNA sequences are aligned to show the expected mismatched nucleotides, which would be cleaved by RNAase (Figs 18 - 19). To offset the possibility of less than 100% cleavage at base substitutions by RNAse, the selected ND2 gene region includes two gaps, such as the 6 bp insertion of A residues at position 4914, while the 12S rRNA sequence has one gap at position 1609. Digestion of RNA heteroduplexes at these sites would prohibit the observation of fragments larger than 244 bp in either RPA.

The following oligonucleotide primers were derived from the above gene sequences showing the highest conservation between the *Numt* and cytoplasmic mtDNA. For each pair, the 25 bp sequence of the T7 promoter [T7] - 5'ACCTAATACGACTCACTATAGGGAG 3' was added to the 5' end of the appropriate RC primer to permit the *in vitro* transcription of anti-sense RNA molecules from the original PCR product (Stoflet et al, 1988; Krieg et al, 1991). Consequently, *in vitro* transcripts would be complementary to *in vivo* H-strand transcripts and provide protection from RNAse.

Optimal PCR reaction conditions were determined empirically for each primer pair, but essentially followed guidelines previously described (Ehrlich et al, 1991; Innis et al, 1990). Amplifications were visualized on ethidium-bromide stained agarose gels

and were considered acceptable when the PCR product equaled the predicted size plus the extra [T7]-promoter DNA. After spinning amplification products through Centricon-100 filter units to remove salts and unincorporated nucleotides and washing with sterile DEPC-treated water, labeled probes were prepared by *in vitro* transcription of the PCR templates with kits from Ambion Inc. (Austin, Texas) or Stratagene. Reactions included 4.84 μM ^{32}P-rCTP 5' triphosphate, 80 uM of cold CTP, 400 μM for each ATP, GTP and UTP, 1 μl DTT, RNAse inhibitor, and T7 RNA polymerase in about 25 μl. After labeling, probes were run and cut from 6% polyacrylamide gels and then eluted into 0.5 M NH_4OAc, 10 mM Mg acetate, 1 mM EDTA, and 0.1% SDS. Approximately 3 X 10^5 - 3 X 10^6 cpm of probe were recovered, which was then diluted to 1 X 10^5 cpm for each hybridization reaction.

Solution hybridizations of labeled probes with total RNA, poly-A mRNA or *in vitro* RNA transcripts were carried out in 20 ul of Ambion hybridization buffer (80% deionized formamide, 100 mM sodium citrate (pH 6.4), 300 mM sodium acetate (pH 6.4), and 1 mM EDTA) at 47-50°C overnight, after heating each reaction to 90°C for 3-4 min. A control reaction containing yeast tRNA was included in every hybridization experiment to test RNAse activity.

RNAse digestions were carried out in Ambion digestion buffer or 5 mM EDTA, 10mM Tris [pH = 7.5], and 300 mM NaCl. Concentrations of RNAse T1/A enzyme mixtures (250 Kunitz units/ml RNAse A + 10,000 units/ml cloned RNAse T1) were empirically determined or followed the manufacturer's guidelines (e.g 1:100 dilution

of Ambion's enzyme mixture into hybridization buffer). The following general guidelines were also used to assess digestions: RNAse A preferentially cuts at pyrimidines [C:A, C:C, C:T, U:T] (Myers et al, 1985), while T1 cuts at the 3' end of GpN runs. Also, the stability of hybrid duplexes generally decreases from RNA:RNA > RNA:DNA > DNA:DNA. Digestions were incubated at 37°C for 30 min. and stopped with inactivation/precipitation mixture (Ambion). After resuspension in loading buffer, samples were run on 6% polyacrylamide sequencing gels for about 2 hrs and processed according to standard sequencing gel protocols.

To provide positive controls, similar procedures were performed on transcripts derived from the opposite PCR strand of both p*Numt*.1 and cytoplasmic pCmt.12 templates. These PCR-derived "sense" transcripts were hybridized with complementary test anti-sense RNA probes in parallel with cellular RNA to test whether the test probe could "protect itself".

RESULTS AND DISCUSSION

Table 11 summarizes the expected and observed protected fragment sizes for each gene tested in the RPA. The expected sizes are based on the location of mutations shown in the sequence alignments (Fig. 18-19).

In the ND2 RPA (Fig. 20), the cytoplasmic RNA probe clearly exhibited greater protection of transcripts from total cellular (and poly-A) RNA than the *Numt* probe. Although showing lighter band intensities due to lower concentrations, poly-A mRNA band patterns usually paralleled those in total RNA lanes. The protected 320 bp fragment probably corresponds to the processed ND2 transcript. Smaller processed transcripts of 69 and 41 bp, also predicted from the alignment of Fig. 19, would code for tRNA-Met and 3' portion of tRNA-Gln, respectively, but have run off the gel shown in Fig 20. Nonetheless, an unprocessed, 447 bp full-size transcript is also protected in lane 3, and probably derived from H-strand polycistronic messages. The cytoplasmic band at 197 bp was unexpected and could only be explained by a random cleavage or duplex formation with an incomplete or anonymous RNA message.

In contrast, RPA results with the *Numt*-specific ND2 probe (Fig. 20; lanes 7-11) greatly differ from the cytoplasmic probe hybridizations by revealing virtually no protection of predicted full-size cytoplasmic transcripts with *in vitro* reactions, total

Table 10. Parameters and Results of Ribonuclease Protection Assays between Feline Numt and Cytoplasmic MtDNA

RPA No	MtDNA Gene region	Assay Type*	Primers pairs to produce in vitro RNA Template	DNA Sequence (5'→3')	Cat MtDNA Position	Size*** of PCR Template	Probe Size (bp)	RESULTS Size of Protected Fragments after Hybridization with - pNumt 1 Transcripts@	pCmt 12 Transcripts@	In vivo Total (or poly-A+) RNA
1	ND2 & tRNA-Q, M	A Cyto	1 625RC	(T7) AGCGGAATGGGGCTAGTCCTAG	5102	447	447	Exp. - None	Exp -345^,69^, 41^^	Exp. - 447, 345^, 69^, 41^
			2 71/203	CCAAAAGTCTTTCGTGC	4672			Obs - None	Obs. - smear, possible 447	Obs - ca 440, 320, 197, 107
		B Numt	" "	" " "		454	454	Exp - 454	Exp. - < 40 bp	Exp - < 40 bp
			" "	" " "				Obs. - NA	Obs. - ca 107	Obs - < 40 bp
2	12S rRNA	A Cyto	1 322 RC	(T7) CTACT AAATC CTCCT TTAAT AC	1707	392	375	Exp** - None	Exp - 375	Exp. - 375
			2 12S-L	AAGCTTCAAACTGGGATTAGATACCCCAC	1382			Obs. - None	Obs - smear ca 350	Obs - ca 350
		B Numt	" "	" " "		393	374	Exp -374	Exp - < 100	Exp - < 100
			" "	" " "				Obs - smear, ca 3	Obs - 230, 90	Obs - 230, 90
3	ND1	A Cyto	1 290 RC	(T7) GTTAATGAGCGGGTATGGTAT	3890	500	471	Exp - None	Exp - 471, ca 291, 57^^	Exp - ca 291, 57^^
			2 16S/551	CCC AGT ACG AAA GGA CAA GAG	3419			Obs - heavy smear	Obs - NA	Obs. - 285, 175, < 100 bp
		B Numt	" "	" " "			471	Exp. - 471	Exp - < 291 bp	Exp. - < 291 bp
			" "	" " "				Obs - smear	Obs - NA	Obs - < 130 bp

* - Each assay (#1-3) should be viewed in two parts, dependent on the type of probe applied in hybridizations. Therefore, results for each probe (cytoplasmic mtDNA or Numt) are listed by rows
- Columns indicate type of complementary RNA used in hybridizations
- The same primer pairs were utilized for amplifying PCR templates from either pNumt 1 or cytoplasmic pCmt 12 clones
** - Exp. refers to the expected size of protected fragments based on sequence alignments (Fig 19,20###) and mutations; Obs = observed
- [T7] indicates the attachment of T7 promoter sequences; RC denotes reverse complement of the actual L-strand sequence shown in Fig. 12; NA (data not available)
-***All size values are in base pairs (bp). Generally, sizes of PCR templates were larger than probe due to the additional T7 promoter sequence
^ Processed transcript predicted from sequence; @ Derived from in vitro transcription of PCR generated templates
- ^^ Represents either the 3' portions of processed tRNA-Gln and tRNA-Leu (UUR) genes

Fig. 20. RPA results for the ND2, tRNA-Q, tRNA-M region. The dried polyacrylamide gel begins with the tRNA positive control (lane 1), which appears completely digested and confirms RNAse activity. Although probes and hybridizations were immediately used in hybridization reactions, some degradation of both are evident in the observation of smearing and non-specific bands. Additional streaking may have been caused by a combination of overloading probes and film overexposure. As a size standard, the sequences of M13mp18 single-stranded phage vector initiated from the universal primer are shown alongside RPA reactions.

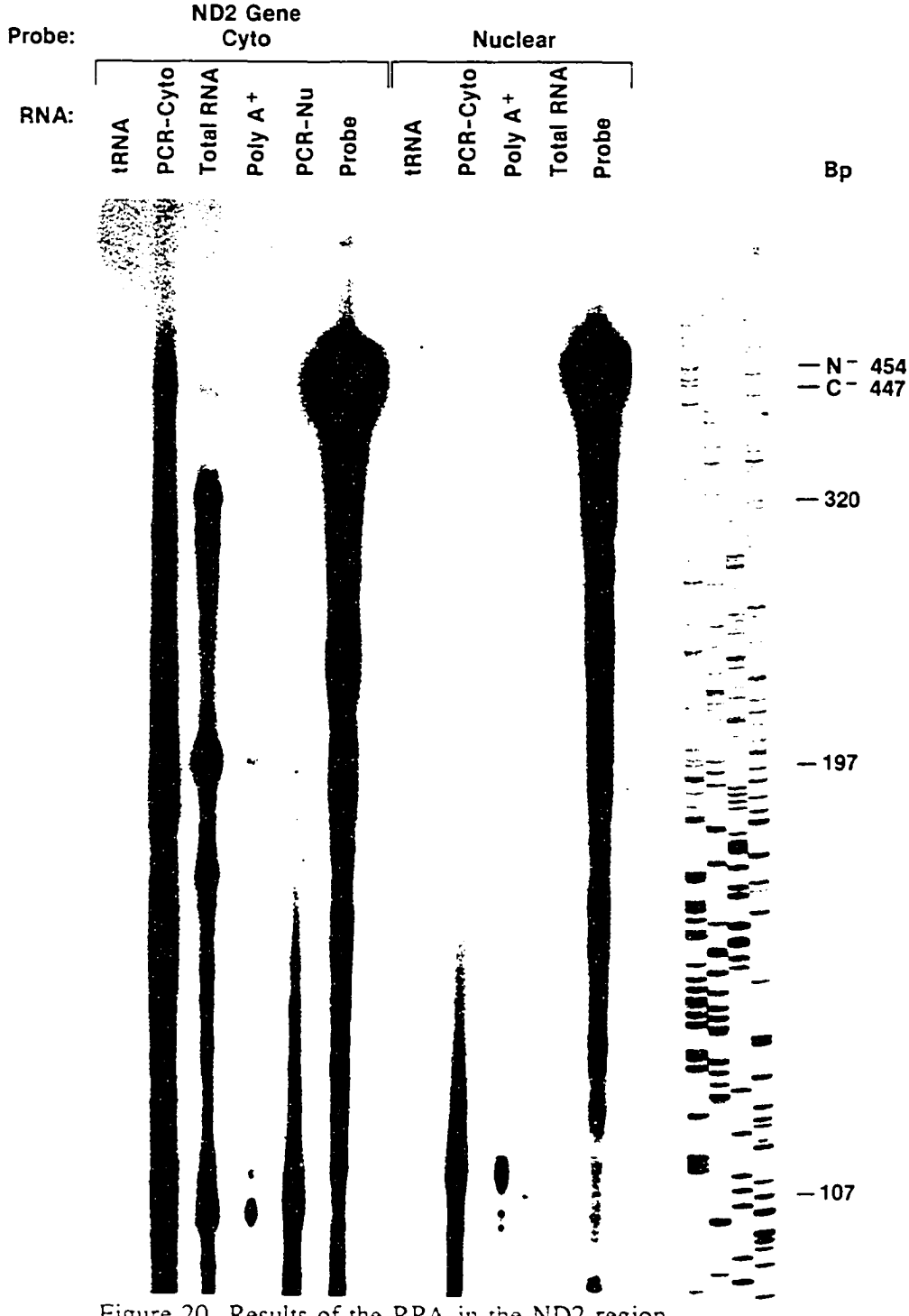


Figure 20. Results of the RPA in the ND2 region.

cellular or poly-A+ RNA. Also, the aberrant 197 bp band was less visible in *Numt* probe hybridizations of the with *in vivo* cytoplasmic transcripts.

An example of the positive control reaction's ability to detect non-complementary regions in hybrid duplexes is shown in lane 5, where the cytoplasmic probe and PCR-derived p*Numt*.1 transcripts are severely degraded. Although the negative control (lane 2) with the pure cytoplasmic mtDNA hybrids exhibited a high degree of smearing, a full-size transcript comparable to the protected probe and the ND2 fragment in the total RNA lane (3) may be present. Generally, control hybridizations with *in vitro* transcripts derived from PCR products complementary to either cytoplasmic (lane 2) or *Numt* sequences were less clear than cellular RNAs, due to high backgrounds or insufficient digestion. Also the hybridization of the *Numt* probe to its *in vitro* sense strand was lost due to mishandling.

Ribonuclease Protection Assay of 12S rRNA gene sequences

Since the 12S rRNA gene sequences chosen for RPA do not overlap with normal processing sites (Attardi, 1985), the cytoplasmic probe was expected to protect a transcript of equivalent size (Table 11). As with the ND2 and tRNA-Met and tRNA-Gln gene results, protection of nearly full-size cytoplasmic transcripts of the 12S rRNA cytoplasmic probe was evident after hybridization with total RNA (lane 4) and possibly the PCR negative control (lane 2) (Fig. 21). A slight decrease in size of the *in vivo* cytoplasmic transcript in the total RNA control lane was apparent, but probably due to

Fig. 21. RPA results for the 12S rRNA gene region. The largest band in the cytoplasmic probe lane (5) is 375 bp. Although it cannot be precisely determined due to overexposure of the signal, the cytoplasmic probe appears to be at least one bp shorter than the *Numt* probe, as predicted from the primary sequence. Smaller fragments in the probe lanes again probably represent non-enzymatic degradation products. Amounts of total RNA used in hybridizations with the *Numt* probe are indicated above the lanes at the far right. The M13 phage was again used as the size standard.

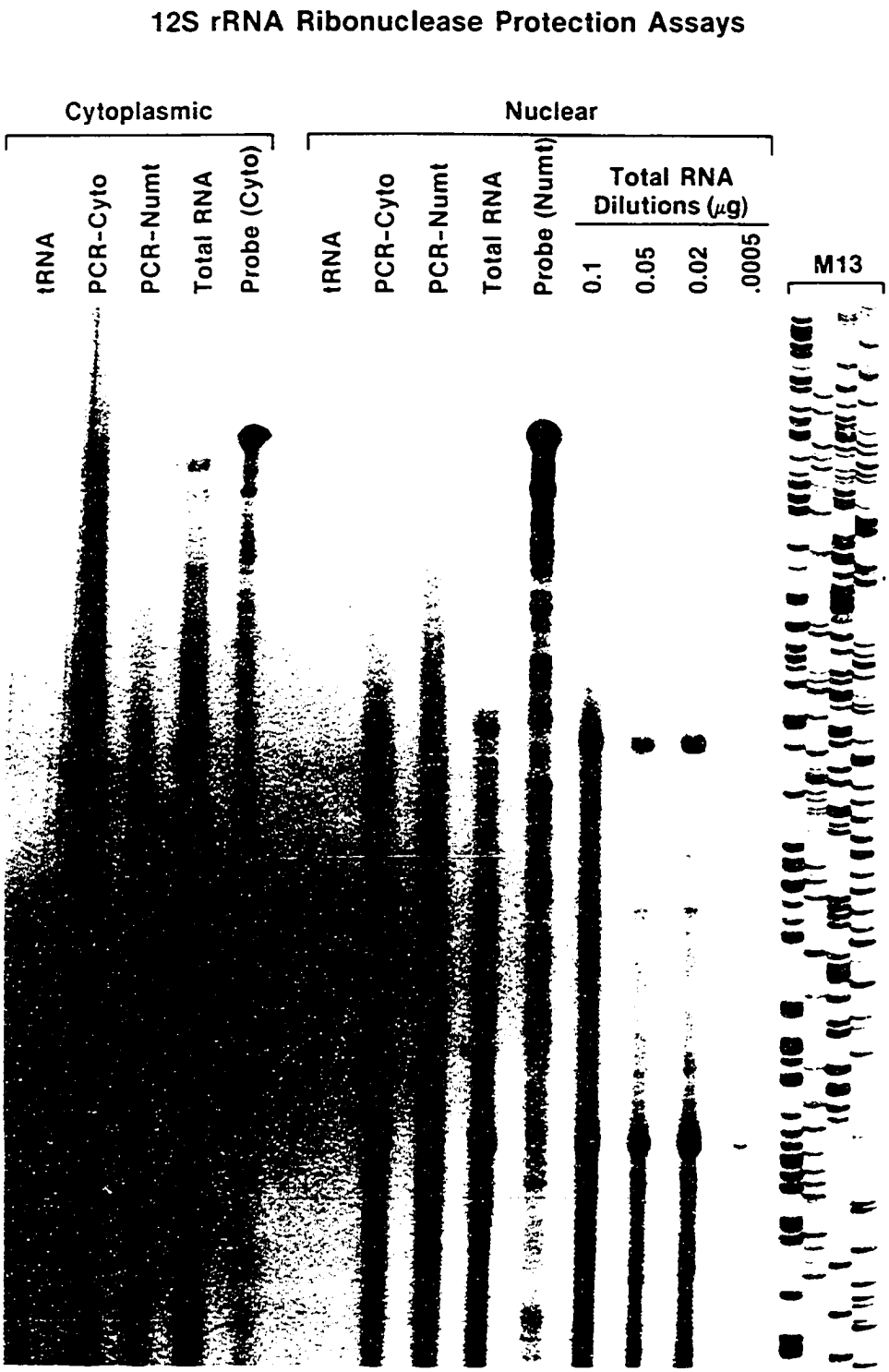


Figure 21. Results of the RPA in the 12S rRNA region.

the digestion of the T7 promoter tail from the original primer after annealing. This could not be seen in Fig 20 due to overexposure of the ND2 probes. The reciprocal hybridization of the cytoplasmic probe with *Numt* sense RNA (lane 3) did not yield a full size transcript.

In hybridizations with the *Numt*-specific probe, some protection was afforded to two prominent digestion products of approximately 230 and 133 bp in the total cellular RNA hybridizations (Fig. 21; lanes 9, 11-14). These fragments are larger than sizes predicted from the sequence alignment, and may be explained by less than 100% or preferential cleavage at mismatches between *Numt* and cytoplasmic mtDNA. These two bands can be dismissed as artifactual, since they i) are present in the hybridizations with cytoplasmic control (lane 7) yet ii) do not appear in the negative control reaction (lane 8). Furthermore, the lack of fragments over 250 bp with the *Numt*-specific anti-sense RNA is consistent with the predictions based on the highly susceptible indel at mtDNA position 1609 (Table 11). These results reiterate the previous ND2 findings in the ND2 gene region, of the failure of the *Numt* probes to fully protect *in vivo*, cellular transcripts.

Unlike the ND2 assay, PCR-derived 12S rRNA sense transcripts in the control hybridizations better protected both cytoplasmic and *Numt* probes, generating nearly full-size protected fragments (lanes 4 and 8, respectively). For example, the fragment at about 375 bp produced in the *Numt* negative control hybridization (lane 8), probably

results from the same phenomenon of T7 linker degradation observed with the cytoplasmic 12S rRNA probe in lane 4.

Although irrelevant due to the negative results with the *Numt* probe, serial dilutions of total cellular RNA were also hybridized to the *Numt* probe, to evaluate the sensitivity of the RPA. Despite the pipetting error between the 0.05 and 0.02 dilutions, transcripts appear to be detectable down to a concentration of about 0.0005 μg. Also, gel results for the ND1 RPA are not shown, due to excessive smearing in probe and control lanes. However, similar conclusions could be drawn with this gene (Table 11): namely, an inability of probes derived from p*Numt*.1 DNA templates to protect full-sized *in vivo* transcripts.

Based on predictions from the primary sequence, multiple bands observed in several of the RPA gels probably result from *in vivo* processing of the multi-cistronic mtNDA transcript. Unexpected fragments (e.g 197 bp, Fig. 21) in cytoplasmic RNA control lanes may stem from heteroplasmic mtDNAs, which are cleaved at substitutions. Alternatively, production of these bands could be directly related to specific reaction conditions, and a consequence of incomplete or non-specific cleavage by RNAse. Nevertheless, these artifacts do not alter the primary, striking observation that *Numt*-specific transcripts appear to be non-existent (even in total RNA preparations) or at least not expressed at normal levels comparable to cytoplasmic mtDNA RNA messages. Conversely, full-length or processed cytoplasmic *in vivo* mtDNA transcripts were protected only with the anti-sense probes derived from cytoplasmic mtDNA sequences.

While these RPA results eliminate the likelihood of *Numt* expression in the cat, the paradox of gene expression implicit in the SET remains unsettled. If the concept of genetic transfer between organelles is to remain a viable mode of integration, diversification and general evolution of eukaryotic genomes (Gray, 1989; Margulis, 1993), inauguration of functional gene expression in new organelles or chromosomal environments after indiscriminate genetic transfers should be accounted for. For example, plants possess the ability to edit RNA messages, which explained the transcription of nuclear mtDNA fragments in the cowpea (*Vigna unguiculata*) and other legumes (Nugent and Palmer, 1991). Mammals, however, lack this capability, and therefore, transcription of mitochondrial genes displaced to the nucleus must overcome the organellar differences in genetic code, potential inhibiting effects of any surrounding satellite DNA (e.g. scarcity of promoter sequences) and also re-route the targeting and active transport of nuclear-encoded proteins (e.g. gamma-RNA polymerase) and transcription factors required for mitochondrial metabolism (Clayton, 1984; 1991; Hurt and van Loon, 1986; Fisher et al, 1991; Rose et al, 1992). At this stage of evolution, mustering the necessary means and flexibility to overcome these obstacles may be too difficult for present-day vertebrate genomes.

CHAPTER 4

FLANKING SEQUENCES AND IMPLICATIONS OF NUMT IN EXOTIC SPECIES OF FELIDAE

BACKGROUND

"In the dime stores and bus stations/People talk of situations,
Read books, repeat quotations/Draw conclusions on the wall.
Some speak of the future/My love she speaks softly,
She knows there's no success like failure
And that failure's no success at all."
-Love Minus Zero/No Limit, Bob Dylan

Although the objectives of this dissertation have heretofore primarily encompassed the characterization of nuclear mtDNA fragments in the one felid species (*F. catus*) actively targeted as an alternative model for human disease and genetics research (O'Brien, 1986; Lyons et al, 1994), continuing characterization of nuclear mtDNA fragments in exotic species of cats impacts ongoing evolutionary studies in our laboratory. For several years, mtDNA data (DNA sequences, RFLP) has been routinely employed to elucidate population substructure, quantitate standing genetic variation and infer phylogenetic relationships within and among diverse groups of mammals (O'Brien et al, 1990; Miththapala, doctoral dissertation, 1992; Janczewski et al, in press; Johnson et al, in press; Masuda et al, submitted; O'Brien, 1994; Baker et al, 1990; Hoelzel et al, 1993; Avise, 1994). Among felids, the genetic status and differentiation

of various subspecies of puma (*P. concolor*) and the increasingly threatened tiger (*P. tigris*) are being investigated with mtDNA gene sequences.

Clarification of a cat family (Felidae) phylogeny, involving most of the 38 recognized species, has been a goal of this laboratory for several years (Collier and O'Brien, 1985; O'Brien et al, 1987), and to expedite the endeavor, many types of molecular data have been employed. For example, techniques such as microcomplement fixation to determine serum albumin immunological distances, differential segregation of endogenous retroviruses, comparative cytogenetics (Modi and O'Brien, 1988) and protein electrophoresis (O'Brien et al, 1987; Pecon-Slattery et al, 1994) have established at least three major cat lineages: domestic cat, pantherine and ocelot (Fig 22A; Janczewski et al, in press). These studies also complement non-molecular (morphological, behavioral, reproductive) phylogenetic interpretations (Salles, 1992). Overall, gene trees can provide a reliable portrayal of species radiations within mammalian families and orders (Nei, 1987; Irwin et al, 1991; Miyamoto and Cracraft, 1991; Meyer et al; Slade et al, 1994; Avise, 1994).

To illustrate the vital role of mtDNA data for phylogenetic inferences in the Felidae, a current view of the cat family tree recently derived with mitochondrial cyt b sequences is depicted (Fig. 22B) (Masuda et al, in prep). This topology involves a greater number of felid taxa, does not include *Numt* sequences, but remains similar to the phylogeny shown in Fig. 10 (Chapter 1), by recapitulating a monophyletic *Felis* genus. Nonetheless, resolution of felid phylogeny with mtDNA characters remains

Fig. 22A. Phylogenetic relationships based upon diverse molecular methods (see text). (a) indicates likely integration point of endogenous retrovirus RD-114; (b) and (c) indicate the differentiation of karyotypes in either Panthera group (2N = 36 chromosomes) or South American ocelot lineages (2N = 38). This figure was borrowed with permission from D. Janczewski and the Laboratory of Viral Carcinogenesis. (Rationale and explicit methods of phylogenetic analysis used to derive this consensus are described in articles by Janczewski et al, in press).

B. Neighbor-joining tree derived with mtDNA cyt b DNA sequences. The topology was constructed with the MEGA 1.01 program (Kumar et al, 1993), applying a Ti:Tv ratio of 6:1 and all three codon positions. Bootstrap percentage values derived from 100 replications of the cyt b data with replacement, are shown below the branches in parentheses (Swofford and Olsen, 1990; Hillis and Bull, 1993). Branch lengths are shown as genetic distances estimated with Kimura's 2-parameter model of evolution (Kimura, 1980). Sequence data and analysis are described in more detail in the article by Masuda et al (submitted) in which the candidate collaborated. (+) indicates the detection of *Numt*-like fragments by either extra restriction fragments on Southern blots or amplification of unique deletion junctions in wild cats.

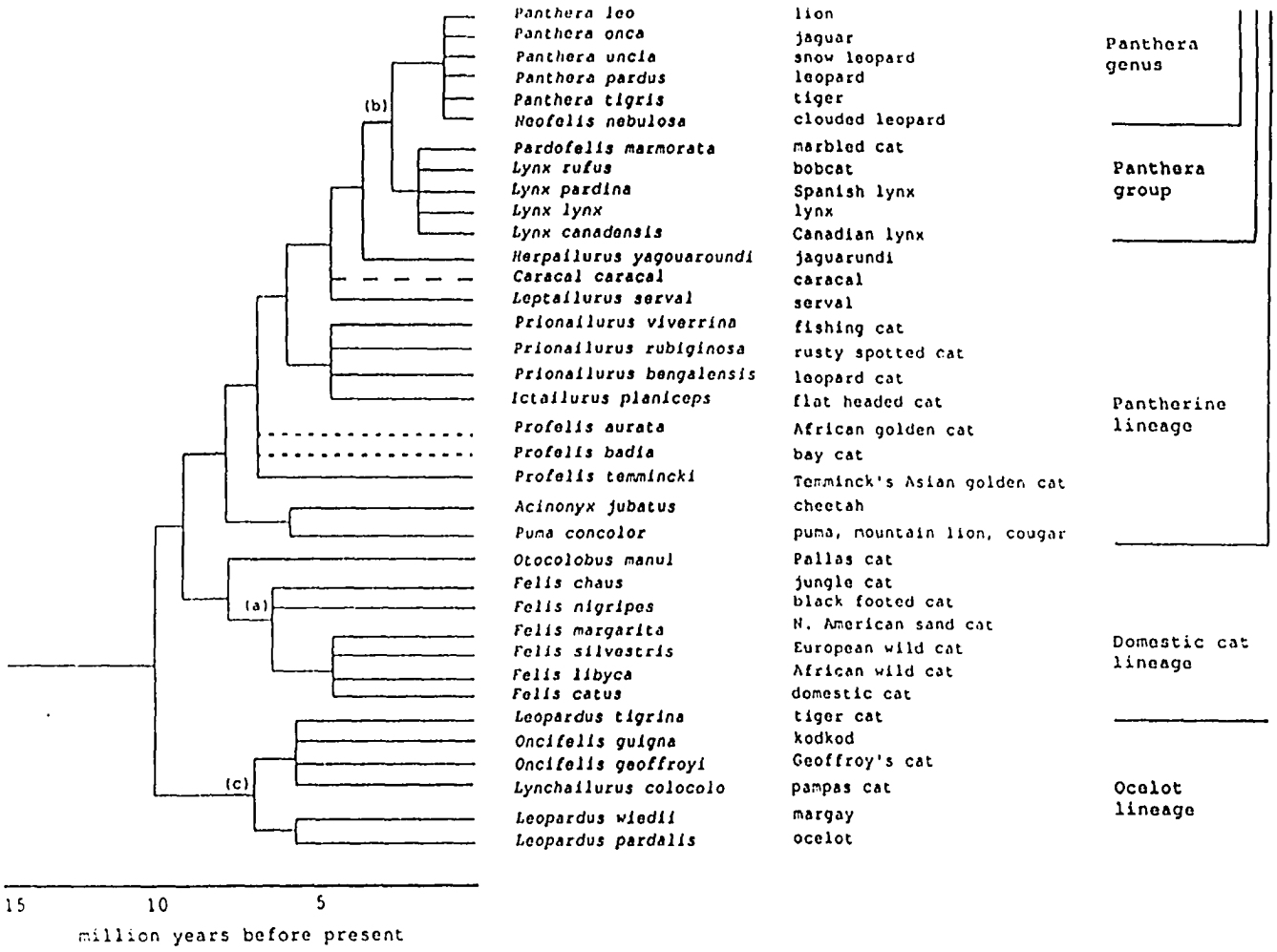
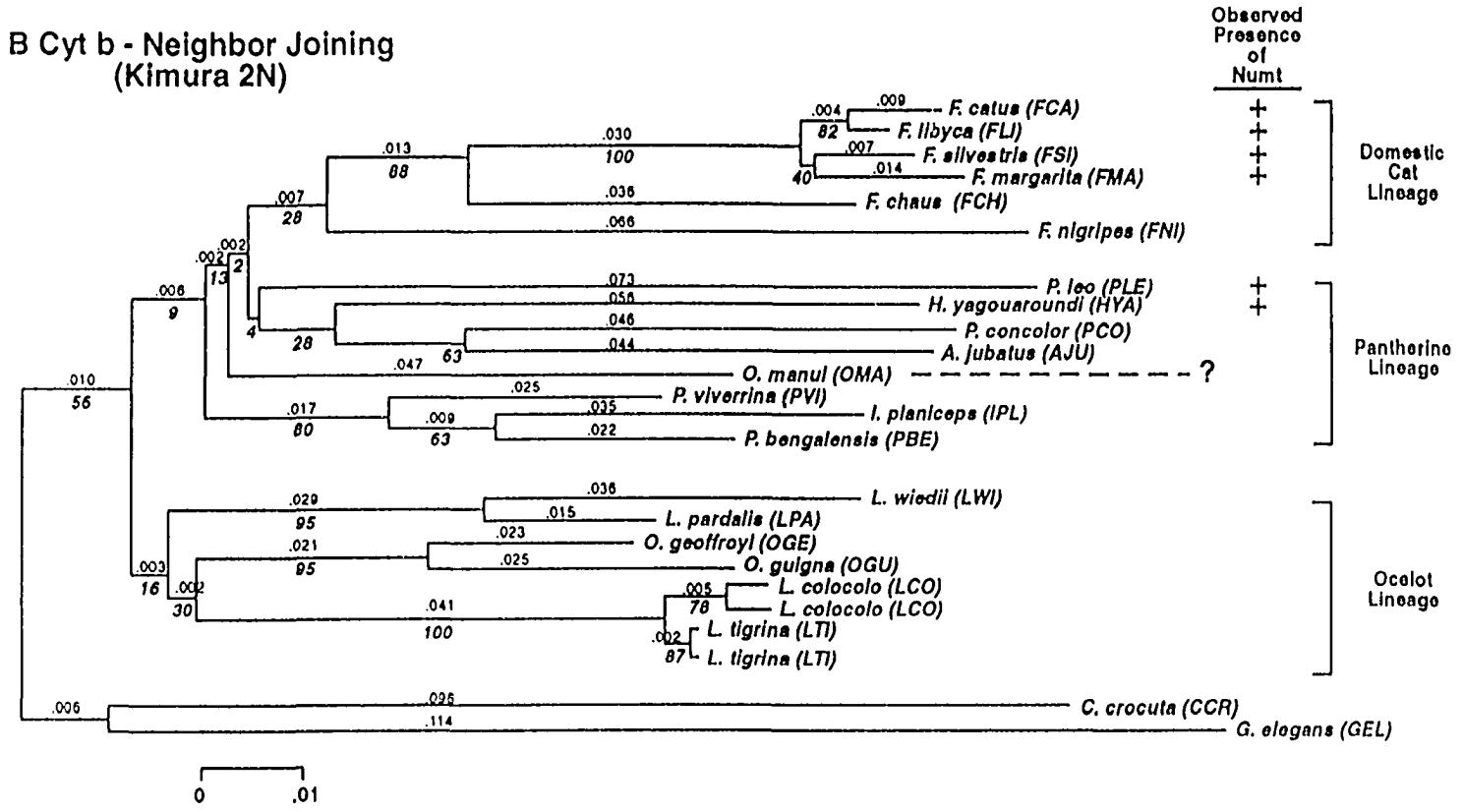


Figure 22. (A) Consensus phylogeny of the Felidae.



equivocal, as the NJ tree indicates that the highest bootstrap values mostly support terminal (more recent) nodes and clades, while internal branching order is less robust.

Considering the ambiguities among some felid associations, identification of homologous *Numt* sequences in one or more of exotic felid taxa could improve the power of phylogenetic reconstruction with mtDNA. This was partially demonstrated in Chapter 1, as *Numt* served as a reliable phylogenetic marker among the *Felis* species most closely related to *F. catus*.

Before these observations can be fully exploited to consolidate phylogenetic relationships, however, it is imperative to determine whether the different *Numt* loci among Felidae are truly *orthologous* (divergent due to speciation rather than another duplication event) or at least homologous (Schlegel, 1994). In other words, are the putative nuclear bands observed in lion and other exotic species derived from the same integration event which produced *Numt* of the domestic cat? According to the phylogeny of Fig. 10 and this chapter, and preliminary RFLP data (Johnson et al, in prep), the occurrence of *Numt* in non-*Felis* species appears enigmatic. The tree suggests that the distribution of *Numt* fragments among felid taxa is not congruent with phylogeny but rather sporadic. Thus, multiple integrations of mtDNA into the nucleus during Felidae radiation remain plausible, which can conflict with idea of *Numt* homology. An alternative explanation is that a single *Numt* integration occurred in an older common ancestor to both pantherine and domestic cat lineages, followed by random *losses* of nuclear mtDNA in certain lineages or species (serval, caracal, puma

etc.) during Felidae radiation. This second explanation is not supported by the relatively high DNA sequence conservation between cytoplasmic mtDNA and *Numt* in *F. catus*, however. Also, no evidence for mtDNA in the nuclear genome in any ocelot lineage species currently exists.

Due to the high sequence conservation between nuclear and cytoplasmic mtDNAs, paralogous mtDNA fragments could be confused with bona fide orthologous mtDNA. A more definitive means of establishing the orthology/homology of *Numt* loci in divergent felid taxa may be achieved by characterizing chromosomal DNA which is directly adjacent to the *Numt* locus (Goodman, 1981; Hardison and Gelinas, 1986). After isolation of these genomic flanking DNAs from the domestic cat, they could be used as hybridization probes or enable the design of oligonucleotide primers for PCR to determine the status (presence/absence) and general characteristics of other putative *Numt* loci in exotic felid genomes. Secondly, flanking sequences could elucidate or refute some of the hypothesized mechanisms of *Numt* integration into the nuclear genome. For instance, the involvement of stochastic processes such as recombination at hotspots or turnover at microsatellite (repetitive) loci or transposons or "retroid"-like elements that could act as shuttling vectors between organelles (Wakasugi et al, 1985; Tsuzuki et al, 1983; Li and Graur, 1991; Charlesworth et al, 1994) may be distinguished with molecular characterization of flanking DNA.

Several key factors were considered in designing experiments (e.g. cloning, PCR) to isolate *Numt* genomic flanking sequences:

i) As demonstrated in chapter 1, *Numt* is a tandem repeat (up to 75 X in some individual cats), with sequences clearly homologous to cytoplasmic mtDNA. Since overall conservation reaches to about 95% sequence identity, precautions had to be incorporated which would eliminate cloning artifactual fragments stemming from either a) multiple *Numt* repeats or b) numerous cytoplasmic mtDNA copies. The second problem could be precluded by using restriction enzymes which did not recognize any cytoplasmic sequences, prohibiting their uptake into a recombinant library. However, since no cloning vector, other than yeast artificial chromosomes (YACs) (Burke et al, 1987; Anand et al, 1990; Franke and Zimmer, 1994), can accommodate a full size *Numt* locus (> 240 kb) plus its flanking genomic DNA, one enzyme had to be chosen which cut at least once within *Numt* (*Numt*+) sequences, thus lowering the size range of clonable fragments. Candidates for each type of enzyme are shown in Fig. 23A.

ii) If *Numt* actually exists as a single-copy locus in the domestic cat, then a minimum of 4 unique *Numt*/genomic DNA junctions will be present per diploid cat genome. In contrast, the high copy number of mtDNA genomes per cell (10^2-10^4) (Birky, 1978) will generate a proportionately high number of clonable restriction fragment termini in spite of the use of single-cut restriction enzymes, thereby increasing the probability of cytoplasmic mtDNA or internal *Numt* inserts.

iii) Determination of the physical distance between the end of the *Numt* locus and the nearest cloning restriction site was not possible with preliminary data available. Due to the aforementioned high copy number of *Numt* and cytoplasmic mtDNAs, mtDNA fragments on standard Southern blots produce signals of high intensity on autoradiograms (Fig 2 and 24; see also O'Brien et al, 1990). Increasing X-ray film exposure times to detect minor bands (and determine their distance from *Numt*) would only exacerbate this condition, due to the expected overexposure from canonical mtDNA fragments during autoradiography. This problem will also be encountered with DNA enriched for nuclear sequences, since *Numt* is also present in high copy number.

iv) Although the complete *Numt* sequence has been determined (Chapter 2), the actual identity of *Numt* sequences at the junction with chromosomal DNA was not evident in the original p*Numt*1 clone. Consequently, an informed decision in choosing *Numt*+ single-site enzymes that could lessen the distance to chromosomal restriction sites, was not possible.

In Chapter 1, it was alluded that nuclear mtDNA fragments could be detected in other members of Felidae by Southern hybridization and PCR. In this chapter, the results of those pilot experiments are formalized and quantitated in greater detail using the same techniques. Furthermore, major emphasis is directed towards the description of a multi-pronged strategy to isolate chromosomal sequences directly flanking *Numt* in *F. catus*.

MATERIALS AND METHODS

Determination of heretical nuclear mtDNA in other felid species. Total felid genomic DNA was purified from lymphocytes or fibroblasts and utilized in Southern hybridizations according to standard protocols previously described (Sambrook et al, 1989; see also Chapter 1). The following felid species (with their tree letter species codes) were analyzed and represent a cross-section of all three major cat lineages (Collier and O'Brien, 1985): lion (PLE- *Panthera leo*), leopard (PPA - *Panthera pardus*), jaguar (PON- *P. onca*), snow leopard (PUN- *P. uncia*), tiger (PTI- *P. tigris*), marbled cat (PMA-*Pardofelis marmorata*), Canadian lynx (LCA - *Lynx canadensis*), lynx (LLY - *L. lynx*), LRU (LRU - *L. rufus*), jaguarundi (HYA - *Herpailurus yagouaroundi*), cheetah (AJU- *Acinonyx jubatus*), caracal (CCA- *Caracal caracal*), serval (LSE- *Leptailurus serval*), flat-headed cat (IPL- *Ictailurus planiceps*), fishing cat (PVI- *Prionailurus viverrina*), leopard cat (PBE - *Prionailurus bengalensis*), clouded leopard (NNE- *Neofelis nebulosa*), African golden cat (PAU- *Profelis aurata*), Temminck's Asian golden cat (PTE- *Profelis temmincki*), puma (PCO- *Puma concolor*), ocelot (LPA- *Leopardus pardalus*), margay (LWI-*L. wiedii*), tigrina (LTI- *L. tigrina*), Geoffroy's cat (OGE - *Oncifelis geoffroyi*), kodkod (OGU- *O. guigna*) and pampas cat (LCO- *Lynchailurus colocolo*). Felid species in the domestic cat lineage or genus *Felis* are described in Chapter 1.

Purified DNAs were digested with various six bp-recognizing restriction enzymes. Total mt genome sizes were estimated in each species by adding all fragments detected with the complete mtDNA probe in each lane and computing the average.

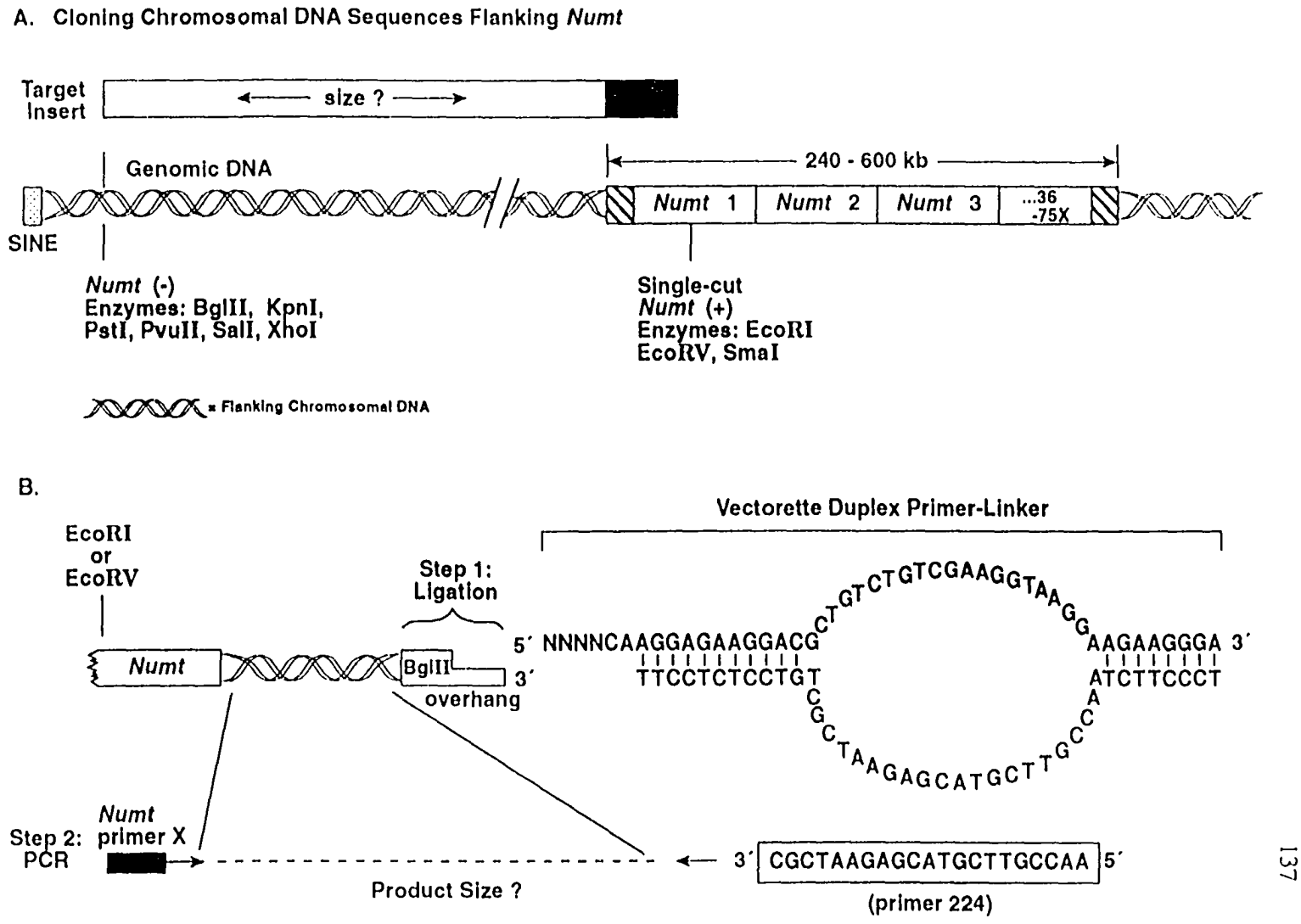
PCR amplification of *Numt* junction. The unique junction presumably stemming from the deletion and/or recombination which juxtaposes COII and CR DNA sequences in *Numt* can be amplified with the following primer pairs, and which were described in Chapter 1 (Fig. 5): 1. [J2] 5' GCTCACGCACACACAAG 3'(8048) (or more downstream [KS/230] - 5' TACACTCATGAGCCGTCCC 3' (8380)) and 2. [J1]RC - 5' AACTGGGACGTGGGG 3'(854). Primer names are in [ ], while their exact nucleotide position in *F. catus* cytoplasmic mtDNA (Fig. 12, Chapter 2) is shown in (). Primer 2 is written as the reverse complement (RC) of control region sequences. Both of these primers (50 pm) were used with various exotic felid genomic DNA templates in standard 100 ul PCR reactions: 94°C denaturation for 1 min, 50°C annealing for 1 min, 72°C polymerization for 1 min for 30 cycles in a Perkin Elmer 480 ThermoCycler.

Isolation of flanking chromosomal sequences.

The following strategies (numbered 1 - 4) were implemented to isolate chromosomal DNA which flanks *F. catus Numt*.

1. Asymmetric cloning. To facilitate cloning nuclear sequences and eliminate contaminating cytoplasmic mtDNA, genomic DNA from cat FCA 65 was enriched for nuclear DNA by isolating nuclei (Hewish and Burgoyne, 1973; Wu, 1980; Lopez, Master's Thesis, 1988). One strategy was to clone directionally, by generating asymmetric restriction fragments, where both 5' and 3' termini of a potential insert are derived from two different restriction digestions (Sambrook et al, 1989). A restriction enzyme which does not cut within the *Numt* repeat (*Numt*-), but presumably has a site in genomic DNA at some unknown physical distance from the *Numt* locus, generates the chromosomal terminus of the flanking fragment (see Fig. 23A). The opposite end of the fragment is generated by a restriction enzyme which cleaves within *Numt* (*Numt*+) preferably only once to simplify the range of restriction fragments, and guarantee the presence of *Numt* sequences in a positive clone which could then be detected with a homologous *Numt* probe. Under these criteria, suitable enzymes were limited to a small pool of candidates enzymes (Fig. 23A). For example, Bgl II was chosen as an appropriate non-cut (*Numt*-) enzyme, since it produced high molecular weight fragments in the PFGE experiments (Fig. 7). EcoRI or EcoRV was typically used as the single-site *Numt*+ enzyme. Overall, this strategy would produce a fragment

Fig. 23A. Diagram of directional cloning strategy for cloning genomic DNA flanking *Numt*. Restriction enzymes are listed from the pool of possible candidates meeting the criteria discussed in Methods. Hatched box indicate the actual junction abutting both *Numt* and chromosomal sequences, whose sequences are currently anonymous. Also, all adjacent *Numt* sequences are unknown. Blackened region in the target represents DNA overlapping with Numt, which is essential for recognition by homologous probes.
B. Diagram of the Vectorette linker primer system utilized for identifying anonymous DNA sequences (after Riley et al, 1990). Instead of YAC vector sequences, for which the procedure was originally developed, *Numt* sequences provide the second anchor primer (black bar) which allow synthesis of the complementary template essential for primer 224 binding.



with one end homologous to *Numt* sequences, while the other end would consist of uncharacterized genomic flanking DNA. (The exact proportions of either would remain unknown until sequencing could be performed).

The choice of cloning vectors was dictated by the criteria outlined above. pBluescript plasmid vectors (Stratagene) served as the most convenient and effective vectors for directional cloning, since they can accommodate large recombinant inserts (5-20 kb) and typically possess asymmetrical polycloning sites. Cosmid vectors were not used, since a) even the smallest continuous *Numt* fragment (ca. 240 kb) could not be accommodated within a single cosmid, which typically has a size limit for inserts of about 50 kb, and b) most cosmid vectors have symmetrical polycloning sites which would preclude their use in asymmetric cloning (Sambrook et al, 1989).

Vectors and purified DNA from nuclei were double-digested with the appropriate enzymes and ligated. Probes for screening recombinant libraries were derived either from specific PCR amplification products or gel-purified fragments derived from the p*Numt*.1 clone. Bacterial transformations, probe labeling, plaque and colony hybridizations were performed according to standard procedures (Sambrook et al, 1989). Electroporation at 2.0 - 2.3 kV was the method of choice for introducing recombinant plasmid into bacterial hosts (e.g *E. coli* NM554, Stratagene).

2. Symmetric cloning of partial digests or the "shadow" Sst I mtDNA band. A second cloning strategy involved the use of symmetric genomic DNA fragments derived from

partial digestions with one restriction enzyme. The possibility of cloning multimers of *Numt* repeats would be diminished by constraints on insert size imposed by specific cloning vectors. For example, rapid plaque screening would be facilitated in Lambda (λ) vectors such as the λ DASH phage (Stratagene), which holds inserts of 9-23 kb (less than 3 Numt repeats). Furthermore, putative flanking sequences could be identified by any deviation from 7.9 kb Numt monomers, generated by single site Numt+ enzymes.

In a second experiment, the low-intensity Sst I restriction fragment which migrates around 4.5 kb in Southern blots of domestic cat DNA was targeted for cloning (see Fig. 2, Chapter 1). This procedure would also involve symmetrical cloning. This fragment was hypothesized to contain genomic flanking DNA, since its size did not conform to map predictions and its signal intensity was much reduced.

Sst I fragments in the 4.5 kb size range were purified with GeneClean (Bio 101) and ligated to a pBluescript II KS(+) bluescript plasmid vector, also cut with Sst I. This library was screened with radio-labeled *Numt* probes as in strategy 1.

3. Vectorette or "Bubble" PCR. This PCR-based method was developed by Riley et al (1990) to amplify terminal sequences from a defined starting point in high-molecular weight yeast artificial chromosomes (YAC) (Burke et al, 1987; Roberts et al, 1992; Franke and Zimmer, 1994). The procedure hinges upon synthetic duplex oligonucleotide linkers, termed "vectorettes", which possess 4 bp overhangs

complementary to appropriate restriction enzyme sticky ends. The linkers also have a region or bubble of non-complementarity nucleotides (Fig. 23B). Other oligonucleotides could be generated for different restriction enzymes with 3' overhangs or blunt ends. Ligation of these duplexes to genomic DNA digests creates vectorette libraries with the specified terminal sequences of the linker.

After creation of the library, PCR amplification is performed using a) a primer (oligonucleotide 224) which is identical to the bottom primer of the bubble and b) a *Numt*-specific (*Numt*+)primer. The crucial aspect of this step is that the complementary strand, essential for primer 224 binding will be generated only after an initial round of DNA polymerization has successfully occurred from the *Numt*-specific primer. In this way, PCR products can be generated containing both *Numt* and anonymous DNA sequences, which may be subsequently identified by direct DNA sequencing. In a minor modification prior to the PCR step, genomic DNA was digested with a single *Numt*+ restriction enzyme to decrease the length of potential PCR amplification products from felid DNA.

As in the cloning protocols, the *Numt*+ enzymes, Eco RV and EcoRI, were used for digestion within *Numt*, while the non-cutting *Numt*- enzyme, Bgl II, was chosen as the target chromosomal site for linker annealing. Also many different *Numt*+ oligonucleotide primers spanning most of the 8.0 kb clone were available after sequencing the entire p*Numt*.1 clone (Chapter 2), and utilized in various combinations of PCR as the second primer with primer 224.

Prior to PCR, oligonucleotides were synthesized on an Applied Biosystems Inc. (ABI) automated DNA/RNA synthesizer (model 394). Equimolar concentrations [1.0 ug] of top and bottom oligonucleotides comprising the vectorette linker were heated to 60°C and then allowed to anneal at room temperature. Duplexed vectorette linker primers (ca. 2.0 ug) were then kinased with 10 U T4 polynucleotide kinase (Bethesda Research Labs), purified with Centricon-50 or -100 (Amersham) spin filters, and ligated to BglII-digested genomic DNA.

As a positive control, p*Numt*.1 or cat genomic DNA was digested with the *Numt*+ enzyme, Bam HI (which has sticky ends compatible with the Bgl II linker), ligated with kinased vectorette Bgl II linkers, and amplified with the 224 vectorette and *Numt*+ primers in parallel with genomic PCR reactions. Specific amplification product sizes could be predicted based on the second *Numt*+ primer used in the reaction and Bam HI sites mapped in *Numt*.

4. Alu- PCR. Another PCR-based strategy was developed to take advantage of any "short interspersed repeated" (SINES) or Alu-like (Deininger and Batzer, 1993) sequences in the cat genome. SINES comprise a family of short (ca. 300 -500 bp) DNA repeats, numbering about 10^5 in several mammalian genomes (Li and Graur, 1991; Weiner et al, 1986). For example, the 3-6% representation of SINES in the human genome suggests an occurrence of an Alu repeat about every 10 kb on average. SINE DNA sequences contain species-specific conserved regions, which have facilitated

the design of primers to obtain chromosome-specific probes for gene mapping in the human genome project via Alu-PCR (Nelson et al, 1989).

With this technology, SINE sequences can be used to anchor one end of an anonymous PCR reaction. DNA sequences for carnivore-specific Alu-PCR in the cat were obtained from H.J. van der Vlugt and A. Lenstra (unpublished data):

1. CANSINE1 - 5' TAACCCACTGAGCCACCCAG 3'

2. CANSINE2 - 5' CCTTGGGCTCAGGTCATGATC 3'

Together with various combinations of *Numt*+ primers, standard and long PCR reactions were performed on cat genomic DNA derived from lymphocytes or somatic cell hybrids which segregate known portions of the cat genome in a rodent background (see chapter 1; O'Brien and Nash, 1992). If it was necessary to identify potential *Numt* SINE products from within a smear, aliquots of completed PCR reactions could be Southern blotted and hybridized with *Numt*-specific probes.

For the above PCR methods of 3 and 4, it was assumed that the distance between the *Numt*-chromosomal junction and Alu or vectorette linker sites would be longer than the typical size range of products generated by standard PCR reactions (Innis et al, 1991). Therefore, various PCR conditions (reaction buffers, annealing and polymerization times) were tested by the candidate that could best generate "long-PCR" products (over 7-10 kb) (Innis et al, 1990). In general, reaction buffers with 0.5-2.5 mM Mg^{++}(Boehringer Mannheim), 2.5 U Taq polymerase, temperatures and times for denaturation (94°C, 1 min) and annealing (50-60°C for 1 min, depending on the primer

pair) followed standard recommendations. Conditions which varied included increase of polymerization (72°C) time up to 7 minutes and lower concentrations (< 50 ng/reaction) of template DNA (Maga and Richardson, 1991). Using the λ clone pCmt.12 as a template, it was possible to generate fragments >10 kb (Fig. 26B).

RESULTS AND DISCUSSION

Confirmation of Numt Loci in Exotic Felid Species

A representative Southern blot of felid mtDNA fragments is shown in Fig 24. Besides the domestic cat (FCA), felid species which show extra mtDNA fragments in the Bam HI digest, making their total mtDNA complement significantly greater than 17.0 kb, are the sand cat (*Felis margarita* - FMA) and the European wildcat (*Felis silvestris* - FSI). These results were repeated with five different restriction enzymes, and distinguish felid species possessing a total mtDNA complement greater than the 17.0 kb expected for felid or other mammalian cytoplasmic mt genomes (Table 12). For all digestions, ethidium-bromide staining of the gels prior to blotting was performed and verified sufficient digestion of total DNA. Mean values of total mtDNA size were based on enzymes which produced the best profiles, i.e unequivocal autoradiograms. For example, Xho I digestions were not scored since it seldom cleaved mtDNA or was generally inefficient in cutting genomic DNA. Large variances associated with mtDNA sizes of some taxa (e.g. PAU, HYA) should be treated with caution, since they were often attributable to faint or ambiguous scoring of bands on autoradiograms. The reciprocal error of overestimating mt genome size is more difficult, however, when at least two different enzymes were used for scoring, and many of the larger genomes were scored with at least two enzymes. This data shows that besides the above *Felis*

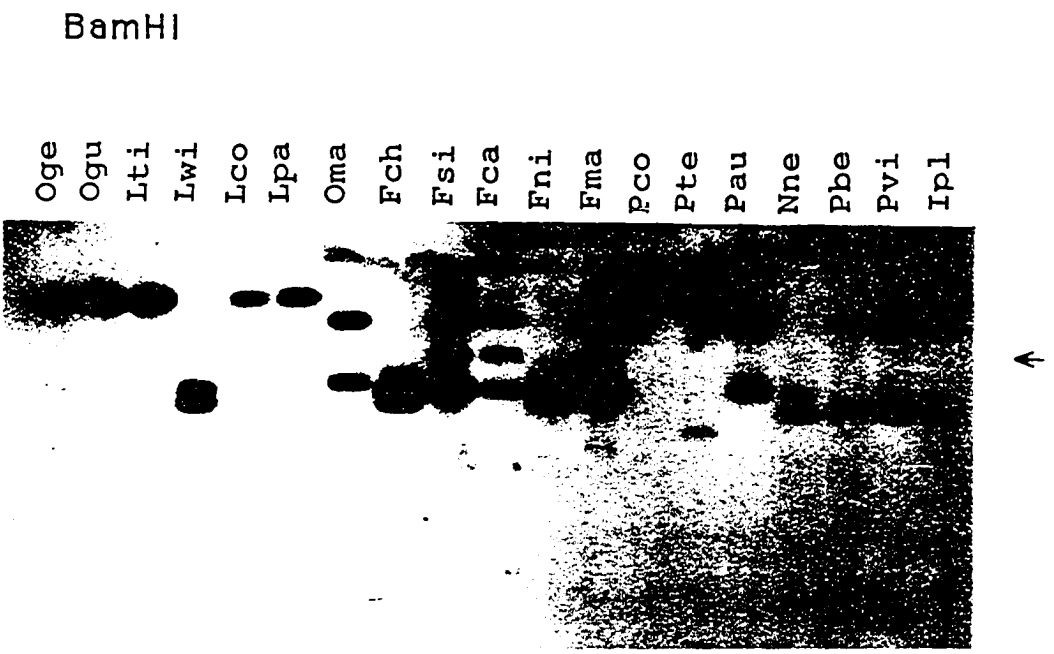


Fig. 24. Autoradiogram with the total mtDNA profile of genomic DNAs in various exotic felids. Approximately 1.0 ug of Bam HI-digested genomic DNAs, were loaded and used in electrophoresis. The clone λ3-2 (O'Brien et al, 1990) containing a complete felid mt genome was used as a hybridization probe for detecting all possible mt fragments. Bands which migrate at the same size range and therefore most likely analogous to the 7.9 kb *Numt* of *F. catus* are indicated by the arrow.

Table 12. Molecular Weights (Kb) of Total Felid and Carnivore MtDNA Fragments in Five Different Restriction Digestions

Species	Enzymes: HpaI	Sst I	EcoRI	BamHI	Hind III	Mean	Std. Dev	VAR
IPL	18.2	15.1	16.0	17.7	n/a	**16.8**	**1.4**	2.1
PVI	14.4	17.0	14.2	18.1	n/a	**15.9**	**1.9**	3.7
PBE	14.5	18.2	18.8	18.1	n/a	**17.4**	**2.0**	3.8
NNE	14.7	18.0	16.3	15.8	n/a	**16.2**	**1.4**	1.9
PAU	14.1	32.0	29.3	18.7	n/a	**23.5**	**8.4**	71.2
PTE	17.1	17.5	13.3	19.2	n/a	**16.8**	**2.5**	6.3
PCO	17.1	17.5	17.4	17.3	n/a	**17.3**	**0.2**	0.0
FMA	20.3	23.6	16.4	25.6	n/a	**21.5**	**7.4**	54.8
FNI	14.4	17.8	17.5	17.1	n/a	**16.7**	**1.5**	2.3
FCA	21.2	31.0	25.8	25.6	n/a	**25.9**	**4.0**	16.0
FSI	21.2	31.0	25.8	25.6	n/a	**25.9**	**4.0**	16.0
FCH	18.3	18.1	17.5	17.1	n/a	**17.8**	**0.6**	0.4
OMA	17.5	18.0	18.1	16.9	n/a	**17.6**	**0.6**	0.4
LPA	17.3	18.0	17.8	17.3	n/a	**17.6**	**0.4**	0.2
LCO	18.6	18.0	15.3	17.3	n/a	**17.3**	**1.4**	2.0
LWI	16.8	18.0	17.1	15.2	n/a	**16.8**	**1.2**	1.4
LTI	19	18.0	17.6	15.9	n/a	**17.6**	**1.3**	1.7
OGU	19	18.0	17.6	15.9	n/a	**17.6**	**1.3**	1.7
OGE	19	18.0	16.3	15.9	n/a	**17.3**	**1.4**	2.0
LSE	n/a	n/a	16.3	14.3	19.7	**16.8**	**2.7**	7.3
CCA	n/a	n/a	17.5	15.8	19.0	**17.4**	**1.6**	2.6
AJU	n/a	n/a	18.2	14.8	12.3	**15.1**	**3.0**	9.0
HYA	n/a	n/a	31.5	16.1	18.8	**22.1**	**8.2**	67.2
LRU	n/a	n/a	36.9	15.5	22.4	**24.9**	**10.9**	118.8
LLY	n/a	n/a	17.5	14.5	20.7	**17.6**	**3.1**	9.6
LCA	n/a	n/a	17.5	15.1	14.8	**15.8**	**1.5**	2.3
PMA	n/a	n/a	16.7	15.5	12.8	**15.0**	**2.0**	4.0
PTI	n/a	n/a	22.4	23.2	18.3	**21.3**	**2.6**	6.8
PUN	n/a	n/a	18.6	17.5	25.1	**20.4**	**4.1**	16.8
PON	n/a	n/a	24.0	26.0	n/a	**25.0**	**1.4**	2.0
PPA	n/a	n/a	27.9	n/a	22.9	**25.4**	**3.5**	12.3
PLE	n/a	n/a	27.9	19.1	22.7	**23.2**	**4.4**	19.4
Spotted Hyena	n/a	n/a	21.0	13.2	18.0	**17.4**	**4.0**	16.0
Mongoose	n/a	n/a	21.0	28.5	18.2	**22.6**	**5.3**	28.1
Siberian Polecat	n/a	n/a	7.7	14.9	15.9	**12.8**	**4.4**	19.4
Racoon	n/a	n/a	21.0	13.6	14.3	**16.3**	**4.0**	16.0
Bear	n/a	n/a	21.0	7.9	18.0	**15.6**	**6.8**	46.2
Dog	n/a	n/a	17.6	18.8	17.0	**17.8**	**0.9**	0.8

* - Derived from a single high MW band; n/a - data not available

^ Faint bands or low amount of DNA observed on gel.

- Although all species were digested with the five listed enzymes, some blots produced equivocal results (e.g. faint or monomorphic banding patterns), and therefore were not included.

species, most of the *Panthera* species (especially lion and leopard), lynx, jaguarundi, African golden cat, and the non-felid mongoose exhibit mtDNAs greater 20 kb, which is also more than 9 standard deviation units (±430 bp) from the average genome size among five orders of mammals, including Carnivora ($p < .001$) (Appendix C). Although preliminary Southern blot results indicate the presence of extra cyt b sequences in the nuclear mtDNA of large cats, which differs from *F. catus Numt* (N. Yuhki, personal communication), the presence of nuclear mtDNA fragments can reconstitute the *Panthera* genus as a monophyletic clade.

To verify the presence of *Numt* in wildcats closely related to *F. catus*, and confirm Southern hybridization results, PCR amplifications using primers flanking the unique *Numt* junction were performed (Fig. 25). The results indicate positive PCR products similar to sizes predicted from the domestic cat *Numt* sequence. It was interesting that jungle cat (FCH) and black-footed cat (FNI) were also positive for this junction, since neither showed significantly larger mitochondrial genomes by Southern blotting (Table 12). Subsequent DNA sequence determination of these products confirms their similarity to the *F. catus* deletion junction (data not shown). This suggests that analogous *Numt*-like bands in these species are present but not amplified. Conversely, the same primers proved negative on cats of the ocelot species, lion, lynx and African golden cat, which can be interpreted as either sufficient DNA sequence divergence at the respective primer-binding sites that inhibit PCR with *F. catus Numt* primers, or the existence of different *Numt*-like junctions in these non-*Felis* species.

Fig. 25. PCR amplification of the internal *Numt*-specific junction region in several wild *Felis* species. The primers are located in COII and control region sequences which flank the unique junction shown in Fig. 5 (Chapter 1). Species which were positive for *Numt* junction fragments by Southern analysis but negative by PCR amplification with J1 and KS/230 primers included PLE, PTI, AJU, PAU, HYA LRU and OMA.

PCR Amplification of Internal *Numt* Deletion Junction in Diverse Felids

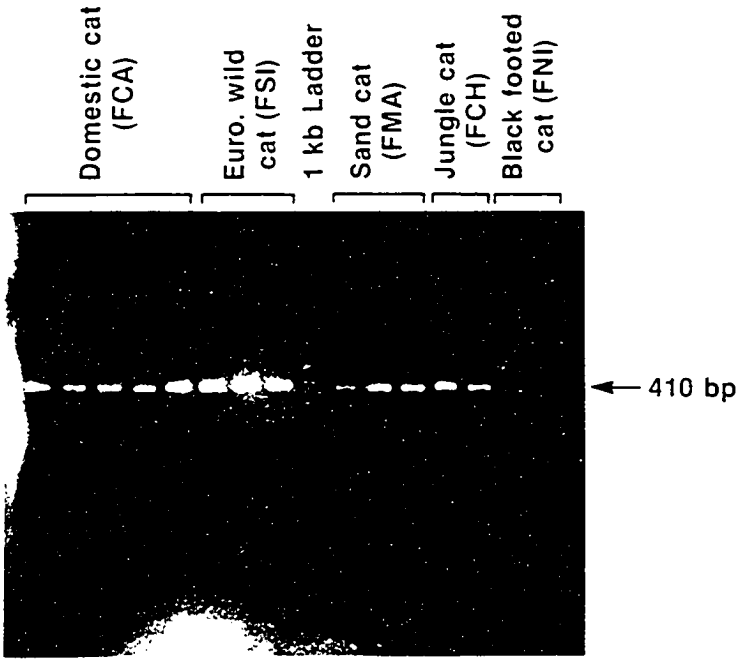


Figure 25. PCR amplification of the internal *Numt*-specific junction region in several wild *Felis* species.

Overall, these preliminary results suggest the possibility of multiple integrations of mtDNA during the evolution of the Felidae. Effective testing of this hypothesis will require more in-depth characterization of nuclear mtDNA loci in exotic felids and interpretation in the context of the phylogenetic divergence of the Felidae (Janczewski et al, in press; Masuda et al, submitted).

Cloning of chromosomal DNA flanking Numt

Four separate libraries, one λ-based library and three separate plasmid libaries, were constructed and screened according to strategies 1 and 2. Each of these libraries typically yielded few positive hybridization signals (e.g. 10-20 in number for each library), representing putative flanking clones. The hybridization signals were also of low intensity and difficult to detect. Overall, screening two genomic libraries following strategy 1 in plasmid vectors failed to yield candidate clones containing chromosomal DNA flanking the *Numt* locus.

The nuclear preparation procedure did not completely eliminate the possibility of cytoplasmic contamination, although there was a reduction of cytoplasmic mtDNA molecules by about 75%, as gauged in Southern hybridizations. Increasing the number of washes of the nuclear pellet could have lowered contamination, but also would have increased the risk of damaging intact nuclei. Nonetheless, these results indicated the efficacy of using nuclear preparations and the directional cloning procedure to reduce potentially contaminating cytoplasmic mtDNA.

Four positive clones from the λ DASH partial digestion library were isolated with strategy 2. After purification, digestion with either Eco RI revealed the presence of non-canonical *Numt* restriction patterns, which was expected from cloning novel genomic fragments (Fig. 26A). Southern hybridization of this gel with *Numt* probes confirmed the presence of *Numt* DNA sequences within the novel set of fragments. However, sequencing the insert in clone L151.O1 indicated that the junction between *Numt* sequences and anonymous genomic sequences was precisely demarcated by an Eco RI cloning site, suggesting they were ligated by the cloning procedure. To verify this conclusion, the non-*Numt* sequences derived from the clone were used to synthesize oligonucleotides for PCR with appropriate *Numt* primers. Results of these subsequent PCR experiments using *F. catus* genomic DNA as a template did not yield products of the predicted size.

Similar negative results were obtained after screening the symmetric Sst I library generated with strategy 2. In general, difficulty was encountered differentiating the likely faint signal of any positive recombinant clones from the high background of false positive plaques or colonies.

PCR strategies for obtaining flanking chromosomal DNA

For both PCR strategies, the primary objective was to obtain discrete amplification products within, at most, 1 - 4 fragments, each potentially representing a unique junction of the single copy *Numt* locus in the diploid genome of FCA 65.

Fig. 26A. Restriction profile of four putative clones identified in one λ DASH library (L151) cut with Eco RI enzyme used in cloning. The two largest bands represent phage arms. Several of the smaller bands hybridized to *Numt* probes in Southern blots, which prompted further characterization by DNA sequencing (see text).
B. Agarose gel showing amplification of a long PCR product. Using the cat primers [13330] - 5' CCAACACGAGAACCTAAATATTCC 3'(nt pos. 13245) and [7280 RC] - 5' GGGAAGAACGTTATATTGACTCC 3' (7473) and the cloned mt genome in pCmt.12 as a template, the predicted 11,259 bp PCR product was generated. Reaction conditions varied from standard conditions primarily by increasing the 72°C extension to 7 minutes, and using 20 cycles. Only 10% of the reaction was loaded on the gel.

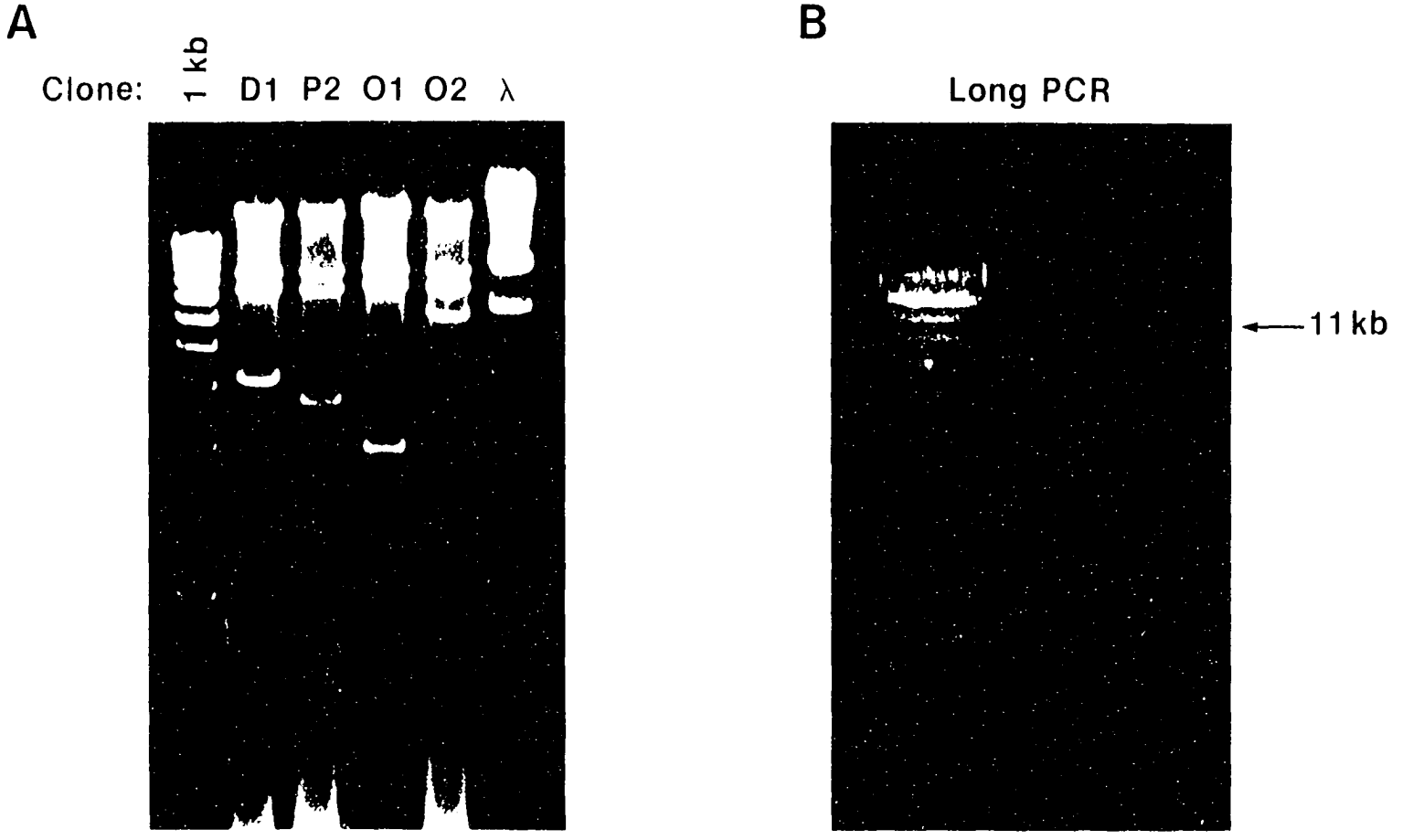


Figure 26. (A) Restriction profile of four putative flanking clones identified in one Lambda library. (B) Amplification of a long PCR product.

Various combinations (>20) of different *Numt*+ primers, canvassing the entire length of the p*Numt*.1 sequence, with 224 vectorette or carnivore SINE primers (CANSINEs) did not generate desired amplification products resembling flanking chromosomal DNA in either type of PCR reaction. In Alu-PCR pilot experiments using either CANSINE primers, as well as either total cat genomic DNA or somatic cell hybrid DNA identified as chromosome D2+ as templates, only a smear of amplification products was produced, suggesting non-specific or excessive priming from SINEs.

These results occurred in spite of affirmative control experiments. For example, after vectorette linkers were ligated to cat genomic DNA digested with *Numt*(+) BamHI, instead of *Numt*(-) BglII, PCR reactions with a *Numt*+ primer (nt position 3556) and 224 primer yielded the expected 530 bp product. Likewise, long-PCR products up to 11 kb in size were generated in control reactions (Fig 26B).

Obstruction of the vectorette strategy could have occurred at any one of the many steps (annealing, ligation of linker to genomic DNA, kinasing or long PCR) of the protocol. Also optimization of the long PCR procedure was conducted on the DNA of clone pCmt.12 (Lopez et al, 1994). Since Cheng et al (1994) subsequently demonstrated that genomic DNA templates were more difficult to amplify than cloned DNA and also produced lower total yields of amplified products, the conditions I used for generating large PCR products may not have been optimal for the potentially large distances involved in traversing genomic DNA flanking *Numt*.

In ALU-PCR experiments, heavy smearing patterns suggested an excess of non-specific amplification products. These could have resulted from too many SINE elements being recognized, in spite of the inclusion of a specific *Numt*+ primer in each reaction. Failure of the Alu-PCR method could also have been due to non-specific or mis-binding of the primer to its target site under the PCR conditions used. In retrospect, different ratios of concentrations for each respective primer pair, as well as annealing temperatures, could have been tested more thoroughly. However, the carnivore ALU-like sequences, CANSINE 1 and 2, were derived from unpublished data, and subsequent testing or corrections by the original authors (J.A. Lenstra and H.H.J. van der Vlugt) may have revealed errors in the DNA sequence unknown to the candidate. Also similar to vectorette experiments, size constraints for the Alu-PCR product could have been a major limiting factor, if the physical distance between the closest CANSINE element and the *Numt*-genomic DNA junction was beyond the range of conditions utilized for PCR. Lastly, although useful in the human genome project, the objectives for using Alu-PCR in mapping the human genome was to generate > 1 DNA fragments as probes for specific chromosomes, and therefore multiple, non-specific PCR products which exhibit a fair amount of smearing in PCR reactions was not considered a hindrance in that context (Nelson et al, 1989; Li et al, 1994). However, the specific single-locus target of *Numt* flanking DNA sequences may have been overwhelmed by the potentially manifold representation of SINE elements in the cat genome.

CHAPTER 5

SYNTHESIS AND CONCLUSIONS

"Similarity is observation. Homology is conclusion." - Walter Fitch

Applying modern, versatile biotechnologies and experimental strategies, this dissertation has answered many of the pressing questions related to heretical mtDNA fragments in the nuclear genome of cats. The *Numt* characterization followed a typical paradigm of genetics studies, whereby the "mutant" is identified or thoroughly described before the wild-type form. The major conclusions of the research follow.

1. Transposition of mtDNA to the nucleus has occurred in *F. catus* and its close relatives in the domestic cat lineage, and now segregates with nuclear genomic DNA.

2. Nuclear mtDNA, or *Numt*, is homologous to half of the cytoplasmic mitochondrial genome in the domestic cat.

3. The 7.9 kb *Numt* monomer is arranged as a head-to-tail tandem repeat array *in situ*, ranging in size from 240-600+ kb at a single locus on cat chromosome D2.

4. Overall sequence divergence between cat cytoplasmic and *Numt* mtDNAs is relatively low (5.1%). In rapidly evolving protein genes most substitutions can be attributed to neutral synonymous mutations stemming from the cytoplasmic genome, which supports the notion of more rapid evolution in the mitochondria versus nuclear sequences. In rRNA genes, however, the rate of *Numt* substitutions appears faster or equivalent to the cytoplasmic mtDNA.

5. Although *Numt* is considered a pseudogene sequence, it had a lower than expected transversion:transition ratio and no increase in dA/dT content compared with cytoplasmic mtDNA.

6. The central core of the control region, 12S rRNA, COI and COIII subunit genes were the most conserved mitochondrial gene sequences, while the ATPase 8, ND2 and ND6 genes were the most rapidly evolving among five mammalian orders.

7. In the carnivore lineage, the ND4L (at the protein level) and ND6 (at the DNA level) genes may have experienced a slow down, while the cyt b gene appeared to be evolving faster at both levels.

8. *Numt* appears transcriptionally silent.

9. Similar *Numt*-like loci appear in the nuclear genomes of other species of Felidae.

Numt closely resembles satellite DNA

For various reasons which are discussed below, viewing *Numt* as a nascent satellite DNA (Miklos, 1985; Grieg et al, 1993; Modi, 1992; Warburton et al, 1993) can shed further light on the mechanisms surrounding its origin, amplification and evolution. Although not a typical satellite per se, the data presented thus far, strongly indicates that *Numt* closely matches the definition of this ubiquitous, highly repeated DNA of eukaryotes by its a) large physical expanse, b) the telescoping of smaller repeat motifs within the a single *Numt* monomer and the total tandem array, and c) possible proximity to centromeric regions (Fig. 4, 11; see also Miklos, 1985; Charlesworth et al, 1994). However, the 7.9 kb length of one *Numt* unit would qualify the locus as one of the largest satellite DNAs ever described.

Not entirely synonymous with one another, constitutive heterochromatin and highly repeated satellite DNA sequences have been studied for many years, but their precise functions remain conspicuously enigmatic (Brutlag, 1977; Miklos, 1985). Structurally, heterochromatin can consist of simple tandem repeats from 10 bp to 171 bp (found in alpha satellites of primates) to over 2.0 kb in some rodents (Charlesworth et al, 1994; Modi, 1992). Despite the relative simplicity of these tandem arrays in heterochromatic regions (Lohe et al, 1993), the entire span of highly repetitive loci can

reach up to 5 MB (Charlesworth et al, 1994), which is similar to *Numt*. Most eukaryotic chromosomal centromeres possess different families of repetitive satellite DNA (Willard, 1990; Miklos, 1985).

Cats have relatively stable karyotypes (few centric rearrangements) but only a small amounts of C-banding, which define most heterochromatic regions (Matthews et al, 1980; Pathak and Wurster-Hill, 1977). Previous cytogenetic studies have indicated that amounts of C-band heterochromatin are more variable among different felid species than in other mammal or carnivore groups (Fanning et al, 1988; Modi et al, 1988). Also, felid satellite DNA described by Fanning et al (1988) was localized at telomeric rather than centromeric regions of felid chromosomes. In general, many species exhibit specific centromeric or kinetochore variants of satellite DNA, but their possible roles in speciation mechanisms, such as postzygotic chromosome mispairing, remain largely unsubstantiated and are often dismissed (Miklos, 1985).

Although the cytogenetic data (Fig. 4) did not unequivocally localize *Numt* sequences within the primary constriction region of chromosome D2 (Willard, 1990), the possible proximity of *Numt* to centromeric regions has implications on cell function and genome evolution. As a parallel, mouse major satellite DNA lies beyond centromeric regions, yet is probably involved in sister chromatid contact and kinetochore apposition (Radic et al, 1987; Willard, 1990). Furthermore, the initial *Numt* integration presumably occurred on a single haploid chromosome, creating D2 chromosomes of unequal size (if pre-amplified) in an individual of a *Felis* species

(possibly *F. nigripes* or *F. chaus*; see Fig. 4 and 10) ancestral to the domestic cat, which must have survived and propagated through that ancestral population and species. If *Numt* had deleterious effects on chromosome pairing and fitness, natural selection would be expected to purge it soon after its appearance. Consistent with the finding that autosomal pairing does not depend on heterochromatin (Yamamoto, 1979), *Numt* does not appear to interfere with normal pairing of D2 chromosomes in cats. For example, recent studies involving interspecies crosses between modern domestic cats and leopard cats (PBE- *Prionailurus bengalensis*), which do not possess *Numt*-like loci, produce viable F1 hybrids which are available for subsequent backcross matings (Lyons et al, 1994).

Although the sequential timing of *Numt* amplification, either before or after integration in the nuclear genome, was not precisely determined (Fig. 11), continuing generation of *Numt* repeats would be consistent with models that simulate the expansion of repetitive arrays in regions associated with low recombination (Stephan, 1989; Charlesworth et al, 1994). According to these models, lower recombination rates coupled with weak selection against tandem arrays will increase the tendency to longer and more complex repeat units (Stephan, 1989). These models are not totally inconsistent with the invocation of unequal sister chromatid exchange (SCE) (vs. homologous chromosome exchange) in satellite DNA to explain the "concertina"-like behavior - expansion/contraction - of tandemly repeated monomers (Cabot et al, 1993; Wu and Hammer, 1991; Begun and Aquadro, 1992). SCE has been observed within

centromeric heterochromatin, which has low recombination frequencies between homologues. For example, in the *Drosophila* tandem array, *Responder*, Cabot et al (1993) found evidence of meiotic drive and low interchromosomal differences primarily due to sister chromatid exchange, which is necessary to produce the copy number variability often associated with tandem arrays.

A similar type of mechanism may be operating at the *Numt* locus to produce the pattern of polymorphic PFGE fragments observed among different individual cats (Fig. 7; Warburton et al, 1993). The presence of short repeat motifs (at RS3; Fig. 6) in the CR of felid mtDNA and *Numt* may facilitate unequal crossing-over by homologous recombination (Ayares et al. 1986), which may be necessary for array elongation. The extreme monomer length of *Numt* precludes strand slippage as a major alternative mechanism for array amplification. Moreover, the detection of an *F. catus Numt* deletion junction in *F. chaus* and possible *F. nigripes* by PCR (Fig. 6, 26), sans amplification of *Numt* bands, suggests that the above mechanisms may be absent, modified, or inactive in these exotic species.

Lastly, various examples from *Xenopus* to cattle demonstrate that amplification of satellite DNA can continue after the insertion of single-copy, foreign non-homologous DNA within an array, which subsequently becomes repeated (Lam and Carrol, 1983; Miklos, 1985). This raises the possibility that *Numt* is also embedded in an unrelated stretch of highly repetitive DNA, as a result of random integration into the nuclear genome. For example, the moderately repeated nuclear mtDNA found in

Locusta migratoria (Table 1; Gellissen et al, 1983) appears to be directly flanked by simple repetitive DNA. By increasing the distance to nearby restriction sites and biasing the primary DNA code, long stretches of simple repeats adjacent to *Numt* could explain the failure of cloning *Numt* flanking chromosomal sequences in Chapter 4. In spite of these close parallels with satellite DNA, the *Numt* phenomenon unfortunately does not resolve the concomitant mysteries associated with this ubiquitous element of eukaryotic genomes.

Numt exemplifies a fossilization process at the molecular level

With the realization of its cellular origin and previous capacity of coding for essential mitochondrial proteins and translational elements, *Numt* also matches the description of a molecular "fossil". Likewise, processed pseudogenes are RNA transcripts copied by reverse transcriptase into DNA that integrates back into the genome, and represent another type of molecular fossil which abounds in mammalian genomes (Weiner et al, 1986; Deininger and Batzer, 1993). Similar to *Numt*, pseudogenes have lost their original functions - tRNA, rRNA, snRNA or 7SL RNA (Li and Graur, 1991). When other types of "retro-elements", which encompass retroviruses to transposable elements carrying sequences for reverse transcriptase, lose the ability of transposition from a chromosomal locus, they can effectively become fossilized. The heterochromatin of *Drosophila* has been described as a graveyard for many types of previously-mobile elements (Li and Graur, 1991). Also, the *Alu* family of short

interspersed repeats discussed briefly in Chapter 5, matches these characteristics of being simultaneously a transposable and non-functional DNA sequence. The cumulative evidence in this dissertation did not support an RNA intermediate for *Numt* origin, however.

This discussion leads to further questions of the contribution of nongenic/noncoding elements to the growth of genome size and the C-value paradox, which refers to the incongruence of genetic information (or organismic complexity) and the absolute DNA content among many living organisms. For example, several unicellular protists have larger genomes than *Homo sapiens* (Cavalier-Smith, 1985), while organisms of similar phenotypic class vary widely in genome content or C-value. Most of this variation has now been attributed to the presence of non-genic or "junk or selfish" DNA, which exhibits virtually no phenotypic effect on the organism (Ohno, 1972; Dawkins, 1976; Orgel and Crick, 1980; Doolittle and Sapienza, 1980). In spite of these comprehensive and thoughtful theories, evidence concerning specific factors or mechanisms that maintain nongenic DNA has been scarce. It is highly unlikely that selection drives *Numt* amplification in the fashion of "homogeneous staining regions" (HSR) or double minutes (Schimke, 1980).

By demonstrating that *Numt* genes are not expressed *in vivo* and therefore equivalent to pseudogenes, the working hypothesis of the SET could be rejected in the cat, but describing *Numt* as junk or selfish DNA may be premature and should be done with caution. For example, the *Responder* tandem repeat locus of *Drosophila* mentioned

above has been shown to affect organismal fitness, when present at higher copy numbers (Wu et al, 1989). Another intriging hypothesis for the maintenance of junk DNA is its inhibitory effect on rates of development, which may be advantageous for some organisms (e.g. plants) (Pagel and Johnstone, 1992). Also, the role of satellite DNA in kinetochore anchoring and binding of specific centromeric proteins can be recalled in this context (Willard, 1990). Since the histone-less or "naked" DNA form of mtDNA would not be expected to persist on *Numt* in the nuclear genome, the effects of *Numt* integration on the phasing of nucleosomes or chromatin condensation on chromosome D2 are largely unknown, yet highly dependent on primary DNA sequences. The possibility therefore remains open that specific protein-binding capabilities or other ancillary genomic/cellular functions, apart from a strict RNA-coding capability, could subsequently evolve for *Numt* (Zuckerkandl, 1992).

Future research directions

The data provided by this research should facilitate and assist the interpretations of ongoing and future molecular genetics studies on the domestic cat, and its felid relatives who possess *Numt*-like bands. The possible advantages to having inert mtDNA sequences in the felid genome are worth analyzing and emphasizing in the context of directions for future research in population genetics, felid systematics, and molecular evolution. As elaborated in the previous two sections, the ramifications of meiotic drive (Cabot et al, 1993) and putative fitness components of tandem arrays may be applicable

to *Numt* in a populations genetics context. The competing hypotheses of multiple integration and orthology/homology of *Numt* loci among the Felidae, and perhaps other carnivores (e.g. herpestids; Table 12), could also be pursued more rigorously and with better techniques. The concordance of *Numt* appearances in wild felids may be useful for consolidating Felidae phylogenies.

With respect to molecular evolution, isolation of other *Numt* monomers from the same array could test the tenets of the molecular drive hypothesis (Dover, 1982; 1986) and enable a more detailed study of homogenizing mechanisms, while a *Numt*-like repeat from an exotic felid species would allow a better assessment and comparison of rate constancy and/or variation in both pseudogenes and functional mtDNA sequences within a single species.

References

References

Adachi, J., Cao, Y., Hasegawa, M. (1993) Tempo and mode of mitochondrial DNA evolution in vertebrates at the amino acid sequence level: Rapid evolution in warm-blooded vertebrates. J Mol Evol 36:270-281

Adkins, R.M., Honeycutt, R.L. (1994) Evolution of the primate cytochrome c oxidase subunit II gene. J. Mol. Evol. 38:215-231.

Ahmadhian, V. (1986) *Symbiosis: An Introduction to Biological Associations*. Univ. Press of England, Hanover.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D. (1989) *Molecular Biology of the Cell*. Garland, New York.

Anand, R., Rilley, J.H., Butler, R., Smith, J.C., Markham, A.F. (1990) A 3.5 genome equivalent multi-access YAC library: Construction, characterization, screening and storage. Nucl. Acids Res. 18:1951-1956.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L, Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F.A., Schreier, P.H., Smith, A.J.H., Staden, R., Young, I.G. (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457-465.

Anderson, S., De Bruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F., Young, I.G. (1982) Complete sequence of bovine mitochondrial DNA (Conserved features of the mammalian mitochondrial genome). J. Mol. Biol. 156:683-717.

Aquadro, C.F., Greenberg, B.D. (1983) Human mitochondrial DNA variation and evolution: Analysis of nucleotide sequences from seven individuals. Genetics 103:287-312.

Arnason, U., Gullberg, A., Widegren, B. (1991) The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. J. Mol. Evol. 33:556-568.

Arnason, U., Johnsson, E. (1992) The complete mitochondrial DNA sequence of the harbor seal, *Phoca vitulina*. J. Mol. Evol. 34:493-505

Arnason, U., Gullberg, A. (1993) Comparison between the complete mtDNA sequences of the blue and fin whale, two species that can hybridize in nature. J. Mol. Evol. 37:312-322.

Arnason, U., Gullberg, A., Johnsson, E., Ledje, C. (1993) The nucleotide sequence of the mitochondrial DNA molecule of the grey seal, *Halichoerus grypus*, and comparison with mitochondrial sequences of other true seals. J. Mol. Evol. 37:323-330.

Attardi, G. (1985) Animal mitochondrial DNA: An extreme example of genetic economy. Int. Rev. Cyto. 93:93-145.

Avise, J.C. (1986) Mitochondrial DNA and the evolutionary genetics of higher animals. Phil. Trans R. Soc. London. B312:325-342.

Avise, J.C. (1991) Ten unorthodox perspectives on evolution prompted by comparative population genetic finding on mitochondrial DNA. Ann Rev Genet 25:45-69

Avise, J.C. (1994) *Molecular Markers, Natural History and Evolution*. Chapman and Hall, New York.

Ayares, D., Chekuri, L., Song, K.-Y., Kucherlapati, R. (1986) Sequence homology requirements for intermolecular recombination in mammalian cells. Proc. Natl. Acad. Sci. USA 83:5199-5203.

Baker, C.S., Palumbi, S.R., Lambertsen, R.H., Weinrich, M.T., Calambokidis, J., O'Brien, S.J. (1990) The influence of seasonal migration on geographic distribution of mitochondrial DNA haplotypes in humpback whales. Nature 344:238-240.

Ballinger, S.W., Shuur, T.G., Torroni, A., Gan, Y.Y., Hodge, J.A., Hassan, K., Chen, K.-H., Wallace, D.C. (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. Genetics 130:139-152.

Barell, B.G., Anderson, S., Bankier, A.T., de Bruijn, M.H.L., Chen, E., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Shreier, P.H., Smith, A.J.H., Staden, R., Young, I.G. (1980) Different patterns of codon recognition by mammalian mitochondrial tRNAs. Proc. Natl. Acad. Sci. USA 77:3164-3166.

Begun, D.J., Aquadro, C.F. (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356:519-520.

Bentzen, P., Leggett, W.C., Brown, G.G. (1988) Length and restriction heteroplasmy in the mitochondrial DNA of American shad (*Alosa sapidissima*). Genetics 118:509-518.

Bermingham, E., Lamb, T., Avise, J.C (1986) Size polymorphism and heteroplasmy in the mtDNA of lower vertebrates. J. Hered 77:249-252.

Bernardi, G., Mouchiroud, D., Gautier, C. (1993) Silent substitutions in mammalian genomes and their evolutionary implications. J. Mol. Evol 37:583-589.

Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W., Clayton, D.A. (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167-180.

Biju-Duval, C., Enafaa, H., Dennebouy, N., Monnerot, M., Mignotte, F., Soriguer, R.C., Gaaied, A.E., Hili, A.E., Mounolou, J.-C. (1991) Mitochondrial DNA evolution in Lagomorphs: Origin of Systematic heteroplasmy and organization of diversity in European rabbits. J. Mol. Evol. 33:92-102.

Birky, C.W., Jr., Fuerst, P., Maruyama, T. (1989) Organelle gene diversity undermigration , mutation, and drift; equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics 121:613-627.

Birky, C.W., Jr. (1978) Transmission genetics of mitochondria and chloroplasts. Ann. Rev. Genet. 12:471-512.

Birren, W.B., Wai, E., Hook, L., Simon, M.I. (1988) Optimized conditions for pulsed field gel electrophoresis separations of DNA. Nucleic Acids Res 16:7563-7582.

Boulikas, T. (1992) Evolutionary consequences of nonrandom damage and repair of chromatin domains. J. Mol. Evol 35:156-180.

Breitenberger, C., Rajbhandary, U.L. (1985) Some highlights of mitochondrial research based on analyses of *Neurospora crassa* mitochondrial DNA. Trends Biochem Sci 10:478-483

Britten, R.J. (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 165:349-357.

Brown, W.M., George Jr., M., Wilson, A.C.(1979) Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad Sci 76:1967-1971.

Brown, W.M. Prager E.M, Wang, A., Wilson, A.C. (1982) Mitochondrial DNA sequences of primates: Tempo and Mode of evolution. J. Mol. Evol. 18:225-239.

Brown, W.M. (1983) Evolution of animal mitochondrial DNA (1983) *In Evolution of Genes and Proteins*. (Eds. M. Nei, R.K. Koehn) Sinauer, Sunderland, Mass. pp 62-88.

Brown, W.M. (1985) The mitochondrial genome of animals. In *Molecular Evolutionary Genetics* (Ed. R. J. MacIntyre), Plenum, New York. pp. 95-130.

Brown, G.G. (1986) Structural conservation and variation in the D-loop containing region of vertebrate mitochondrial DNA. J. Mol Biol 192:503-511.

Brown, E., Yuhki, N., Packer, C., O'Brien, S.J. (1994) Three phylogenetic clusters of feline immunodeficiency virus in free-ranging African lions. J. Virol. 68:5953-5968.

Bruford, M.W., Wayne, R.K. (1993) Microsatellites and their application to population genetic studies. Curr. Op. Genet. Devel. 3:939-943.

Bulmer, M., Wolfe, K.H., Sharp, P.M. (1991) Synonomous nucleotide substitution rates in mammalian genes: Implications for the molecular clock and the relationship of mammalian orders. Proc. Natl. Acad Sci 88:5974-5978.

Burke, D.T., Carle, G.F., Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236:806-811.

Buroker, N.E., Brown, J.R., Gilbert, T.A., O'Hara, P.J., Beckenbach, A.T, Thomas, W.K., Smith, M.J. (1990) Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics 124:157-163.

Cabot, E.L., Doshi, P., Wu, M.-L., Wu, C.-I. (1993) Population genetics of tandem repeats in centromeric heterochromatin: Unequal crossing over and chromosomal divergence at the *Responder* locus of *Drosophila melanogaster*. Genetics 135:477-487.

Capecchi M.R. (1989) Altering the genome by homologous recombination. Science 244:1288-1292.

Carroll, R.L. (1988) *Vertebrate Paleontology and Evolution*. Freeman, New York.

Cavalier-Smith, T. (1985) *The Evolution of Genome Size*. Wiley, New York.

Cedergren, R.J., Sankoff, D, Larue, B, Grosjean H. (1981) The evolving tRNA molecule. CRC Crit. Rev. Bioch. 11:35-104.

Chandley A.C., Mitchell A.R. (1988) Hypervariable minisatellite regions are sites for crossing-over at meiosis in man. Cytogenet Cell Genet 48:152-155.

Chang D.D., Clayton D.A. (1984) Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. Cell 36:635-643.

Charlesworth, B., Sniegowski, P, Stephan, W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215-220.

Cheng, Z., Fockler, C., Barnes, W.M., Higuchi, R. (1994) Effective amplification of long targets from cloned inserts and human genomic DNA. Proc. Natl. Acad. Sci. 91:5695-5699.

Clayton, D.A. (1982) Replication of animal mitochondrial DNA. Cell 28:693-705.

Clayton, D.A. (1984) Transcription of the mammalian mitochondrial genome. Ann. Rev. Biochem. 53:573-594.

Clayton, D.A. (1991) Replication and transcription of vertebrate mitochondrial DNA. Annu. Rev. Cell Biol. 7:453-478.

Collier, G.E., O'Brien, S.J. (1985) A molecular phylogeny of the Felidae: Immunological distance. Evolution 39:473-487.

Dawid, I.B., Blackler, A.W. (1972) Maternal and cytoplasmic inheritance of mitochondrial DNA in *Xenopus*. Dev. Biol. 29:152-161.

Dawkins, R. (1976) *The Selfish Gene*. Oxford Univ. Press, New York.

Deininger, P.L., Batzer, M.A. (1993) Evolution of Retroposons. *Evol Biol* (Ed. M.K. Hecht) Vol. 27. Plenum Press, New York.

Densmore, L.D., Wright, J.W., Brown, W.M. (1985) Length variation and heteroplasmy are frequent in mitochondrial DNA from parthenogenic and bisexual lizard (Genus *Cnemidophorus*). Genetics 110:687-707.

De Salle, R., Freedman, T., Prager, E.M., Wilson, A.C. (1987) Tempo and mode of sequence evolution in mitochondrial DNA in Hawaiian *Drosophila*. J. Mol. Evol. 26:157-164.

Desjardins P., Morais, R. (1990) Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. J. Mol. Biol 212:599-635.

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387-395

Dietrich W., Katz H., Lincoln S.E., Shin H.-S., Friedman J., Dracopoli N.C., Lander E.S. (1992) A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131:423-447.

Doolittle, W.F., Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601-603.

Dover, G.A. (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111-117.

Dover (1986) Molecular drive: a cohesive mode of species evolution. Nature 299:111-117.

Drouin J. (1980) Cloning of human mitochondrial DNA in *Escherichia coli*. J Mol Biol 140:15-34

Dujon B, Belcour L (1989) Mitochondrial DNA instabilities. In Berg DE, Howe MM (eds) Mobile DNA. American Society Microbiology, Washington, DC, pp 861-878

Ehrlich, H.A., Gelfand, D., Sninsky, J.J. (1991) Recent advances in the polymerase chain reaction. Science 252:1643.

Ellis J. (1982) Promiscuous DNA - chloroplast genes inside plant mitochondria. Nature 299:678-680

Eperon I.C., Anderson S., Nierlich D.P. (1980) Distinctive sequence of human mitochondrial ribosomal RNA genes. Nature 286:460-467

Ephrussi, B., Hottinguer, H., Chimenes, A.M. (1949) Ann. Inst. Pasteur, Paris 76:351-364.

Fanning, T.G. (1987) Origin and evolution of a major feline satellite DNA. J. Mol. Biol. 197:627-634.

Fanning, T.G., Modi, W.S., Wayne, R.K., O'Brien, S.J. (1988) Evolution of heterochromatin-associated satellite DNA loci in felids and canids (Carnivora) Cytogenet Cell Genet. 48:214-219.

Farrelly, F., Butow, R.A. (1983) Rearranged mitochondrial genes in the yeast nuclear genome. Nature 301:296-301.

Feig, D.I., Loeb, L.A. (1993) Mechanisms of mutation by oxidative DNA damage: reduced fidelity of mammalian DNA polymerase beta. Biochemistry 32:4466-4473.

Felsenstein J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol Evol 17:368-376.

Felsenstein, J. (1993) PHYLIP: Phylogenetic Inference Package, version 3.5c. University of Washington, Seattle.

Fenchel, T., Finlay, B.J. (1991) Ciliates evolve from anaerobic lifestyle from aerobic ancestors. Eur. J. Protist 26:210-215.

Fenchel, T., Bernard, C. (1993) A purple protist. Nature 362:300.

Fisher, R.P., Parisi, M.A., Clayton, D.A. (1989) Flexible recognition of rapidly evolving promoter sequences by mitochondrial transcription factor 1. Genes and Development 3:2202-2217

Fitch W.M., Margoliash E. (1967) Construction of phylogenetic trees. Science 135:279-284

Foran, D.R., Hixson, J.E., Brown, W.M. (1988) Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucl. Acids Res. 16:5841-5861.

Fox, T.D. (1983) Mitochondrial genes in the nucleus. Nature 301:371-372.

Franke, Y., Zimmer, M. (1994) Construction of a human MluI Yac library. Genomics 21:58-62.

Fukuda, M., Wakasugi, S., Tsuzuki, T., Nomiyama, H., Shimada, K. (1985) Mitochondrial DNA-like sequences in the human nuclear genome. J. Mol. Biol. 186:257-266.

Gadeleta, G., Pepe, G., De Candia, G., Quagliariello, C., Sibisa, E., Saccone, C. (1989) The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J. Mol. Evol. 28:497-516.

Gall, J.G. (1981) Chromosome structure and the C-value paradox. J. Cell Biol. 91:3s-14s.

Gantt, J.S., Baldauf, S.L., Calie, P.J., Weeden, N.F., Palmer, J.D. (1991) Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J. 10(10):3073-3078.

Gellissen, G., Bradfield, J.Y., White, B.N., Wyatt, G.R. (1983) Mitochondrial DNA sequences in the nuclear genome of the locust. Nature 301:631-634.

Gellissen, G., Michaelis, G. (1987) Gene transfer: Mitochondria to nucleus. In *Endocytobiology*. (Eds J.J. Lee, J.F. Frederick.) Ann. N.Y. Acad. Sci. 503:391.

Genetics Computer Group (UWGCG) (1994) Program Manual for Wisconsin Package, Version 8. Madison, WI.

Ghivizzani, S.C., Mackay, S.L.D., Madsen, C.S., Laipis, P.J., Hauswirth, W.W. (1993) Transcribed heteroplasmic repeated sequences in the porcine mitochondrial DNA D-loop region. J. Mol. Evol. 37:36-47.

Gilbert, D.A., O'Brien, J.S., O'Brien S.J. (1988) Chromosomal mapping of lysosomal enzyme structural genes in the domestic cat. Genomics 2:329-336

Gilham, N.W. (1978) *Organelle Heredity*. Raven, New York.

Gillespie, J.H. (1991) *The Causes of Molecular Evolution*. Oxford University Press, New York. p. 76.

Gojobori, T, Li, WH, Graur, D. (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. J. Mol. Evol. 18:360-369.

Goodman, M. (1981) Decoding the pattern of protein evolution. Prog. Biophys Mol Biol 38:105-164.

Graybeal, A. (1994) Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. Syst Biol 43(2):174-193.

Gray, M.W. (1989a) Origin and evolution of mitochondrial DNA. Ann. Rev. Cell Biol. 5:25-50.

Gray M.W. (1989b) The evolutionary origins of organelles. Trends Genet 5:294-299

Gyllensten U.B. Wharton, D., Wilson, A.C. (1985) Maternal inheritance of mitchondrial DNA during backcrossing of two species of mice. J. Hered. 76:321-324

Gyllensten UB, Erlich HA (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. Proc Natl Acad Sci USA 85:7652-7656

Gyllensten, U., Wharton, D., Josefsson, A., Wilson, A.C. (1991) Paternal inheritance of mitochondrial DNA in mice. Nature 352:255

Hare, J.T., Taylor, J.H. (1985) One role for DNA methylation in vertebrate cells is strand discrimination in mismatch repair. Proc. Natl. Acad. Sci. USA 82:7350-7354.

Hardison, R.C., Gelinas, R. (1986) Assignment of orthologous relationships among mammalian alpha-globin genes by examining flanking regions reveals a rapid rate of evolution. Mol. Biol. Evol. 3:243-261.

Harrison, R.G. (1989) Animal Mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends Ecol. Evol. 4:6.

Hartl F.-U., Neupert W. (1990) Protein sorting to mitochondria: evolutionary conservations of folding and assembly. Science 247:930-939.

Hasegawa M., Kishino H., Yano T. (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160-174.

Hasegawa, M., Kishino, H. (1989) Heterogeneity of tempo and mode of mitochondrial DNA evolution among mammalian orders. Jpn J. Genet. 64:243-258.

Hatefi, Y. (1985) The mitochondrial electron transport and oxidative phosphorylation system. Ann. Rev. Biochem. 54:1015-69.

Hayashi J.I., Tagashira Y., Yoshida M.C. (1985) Absence of extensive recombination between inter- and intraspecies mitochondrial DNA in mammalian cells. Exp. Cell Res. 160:387-395.

Hecht, N.B., Liem, H., Kleene, K.C., Distel, R.J., Ho, S.-M. (1984) Maternal inheritance of the mouse mitochondrial genome is not mediated by a loss or gross alteration of the paternal mitochondrial DNA or by methylation of the oocytoe mitochondrial DNA. Devel. Biol. 102:452-461.

Hewish, D.R., Burgoyne, L.A. (1973) Calcium dependent endonuclease activity of isolated nuclear preparations - Relationships between its occurrence and occurrence of the classes of enzymes. Bioch. Biophys. Res. 52:475-481.

Hillis, D.M., Huelsenbeck, J.P. (1992) Signal, noise and reliability in molecular evolutionary analyses. J. Hered 83:189-195.

Hoehe M.R., Caenazzo L., Martinez M.M., Hsieh WT, Modi W.S., Gershon E.S., Bonner T.I. (1991) Genetic and physical mapping of the human cannabinoid receptor gene to chromosome 6q14-q15. New Biol. 3:880-885.

Hoelzel, A.R., Hancock, J.M, Dover, G.A. (1993) Generation of VNTRs and heteroplasmy by sequence turnover in the mitochondrial control region of two elephant seal species. J. Mol. Evol 37:190-197.

Hoelzel, A.R., (1993) Evolution by DNA turnover in the control region of vertebrate mitochondrial DNA. Curr. Op in Genet. Devel. 3:891-895.

Hoelzel, A.R., Lopez, J.V., Dover, G.A., O'Brien, S.J. (1994) Rapid evolution of a heteroplasmic repetitive sequence in the mitochondrial DNA control region of carnivores. J. Mol. Evol. 39:191-199.

Holland, M.J., Innis, M.A. (1990) In vitro transcription of PCR templates. (Eds. M.A.Innis, D.H. Gelfand, J.J Sninsky, T.J. White) *PCR Protocols: A Guide to Methods and Applications*. Academic Press Inc., San Diego. pp 169-176.

Horak, I., Coon, H.G., Dawid, I.B. (1974) Interspecific recombination of mitochondrial DNA molecules in hybrid somatic cells. Proc. Natl. Acad. Sci. USA 71:1828.

Howell, N. (1989) Evolutionary conservation of protein regions in the protonmotive cytochrome b and their possible roles in redox catalysis. J. Mol. Evol. 29:157-169.

Huelsenbeck, J.P., Hillis, D.M. (1993) Success of phylogenetic methods in the four-taxon case. Syst. Biol. 42:247-264.

Hughes A.L., Nei M. (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167-170.

Hurt, E.C., Van Loon, A.P.G.M (1986) How proteins find mitochondria and intramitochondrial compartments. Trends in Bioch. Sci. 11:204-207.

Hyman, B.C., Beck, J.L., Weiss, K.C. (1988) Sequence amplification and gene rearrangement in parasitic nematode mitochondrial DNA. Genetics 120:707-712.

Innis M.A., Gelfand D.H., Srinsky J.J., White T.J. (Eds.) (1990) *PCR Protocols: A Guide to Methods and Applications*. Academic Press, San Diego, CA.

Jacobs, H.T., Grimes, B. (1986) Complete nucleotide sequence of the nuclear pseudogenes for cytochrome subunit I and large mitochondrial rRNA in sea urchin *Stronglycentrotus purpuratus*. J. Mol. Biol. 187:509.

Janczewski, D.N., Modi, W.S., Stephens, J.C., O'Brien, S.J. (1995) Molecular evolution of mitochonrial 12S rRNA and cytochrome b sequences in the pantherine lineage of Felidae. Mol. Biol. Evol. in press.

Janke, A., Fuchs-Feldmaier, G., Kelly Thomas, W., von Haeseler, A., Paabo, S. (1994) The marsupial mitochondrial genome and the evolution of placental mammals. Genetics 137:243-256.

Jeffreys A.J., MacLead A., Tamaki K., Neil D.L., Moncleton D.G. (1991) Minisatellite repeat coding as a digital approach to DNA typing. Nature 354:204-209

Johnson, W.J., Dratch, P.A., Martenson, J.S., O'Brien, S.J. (in preparation) Resolution of recent radiations within three contrasting felid lineages using analysis of mtDNA RFLP variation.

Jukes, T.H. (1987) Transitions, transversions, and molecular evolutionary clock. J. Mol. Evol. 26:87-98.

Kamimura N., Ishii S., Linadong M., Shay J.W. (1989) Three separate mtDNA sequences are contiguous in human genomic DNA. J Mol Biol 210:703-707

Kimura M. (1983) *The Neutral Theory of Molecular Evolution.* Cambridge University Press. Cambridge, England.

Kocher T.D., Thomas W.K., Meyer A., Edwards S.V., Paabo S., Villablanca F.X., Wilson A.C. (1989) Dynamics of mitochondrial DNA animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci USA 86:6196-6200.

Kondo R., Satoshi H., Satta Y., Takahata N. (1993) Evolution of hominoid mitochondrial DNA with special reference to the silent substitution rate over the genome. J Mol Evol 36:517-531.

Krieg, P.A. (1991) Synthesis of RNA probes using SP6, T7, and T3 RNA polymerase. In *Methods in Gene Technology*. Vol. 1 (Eds. J.W. Dale and P.G. Sanders). J.A.I. Press.

Kumar, S., Tamura, K., Nei, M. (1993) MEGA: Molecular Evolutionary Genetics Analysis, version 1.01. The Pennsylvania State University, University Park, PA 16802.

Kumazawa, Y., Nishida, M. (1993) Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenies. J. Mol. Evol. 37:380-398.

Kurten B. (1968) *Pleistocene Mammals of Europe.* Aldine Press, Chicago, IL.

Lam, B.S., Carroll, D. (1983) Tandemly repeated DNA sequences from *Xenopus laevis*. I. Studies on sequence organization and variation in satellite I DNA (741 bp repeat). J. Mol. Biol. 165:567-585.

Langley, C., Fitch, W. M (1974) An examination of the constancy of the rate of molecular evolution J. Mol Evol. 3:161-177.

Larsson, N.G., Holme, E., Kristianson, B., Oldfors, A., Tulinius, M. (1990) Progressive increase of the mutated mitochondrial DNA fraction in Kearns-Sayre syndrome. Pediatr. Res. 28:131-136.

Lee, W-J., Kocher, T.D. (1995) Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. Genetics 139:873-887.

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203-221

Li, Z., Wise, C.A., Le Paslier, D., Hawkins, A.L., Griffin, C.A., Pittler, S.J., Lovett, M., Jabs, E.W. (1994) A YAC contig of approximately 3 Mb from human chromosome 5q31-> q33. Genomics 19:470-477.

Li W.-H., Gojobori T., Nei M. (1981) Pseudogenes as a paradigm of neutral evolution. Nature 292:237-239.

Li, W-H, Wu, C.I., Luo, C.C. (1984) Nonrandom point mutation as reflected in nucleotide substitutions and its evolutionary implications. J. Mol. Evol 21:58-71.

Li, W-H, Luo, C.C., Wu, C.I. (1985) Evolution of DNA sequences. In *Molecular Evolutionary Genetics*. (Ed. R.J. MacIntyre) Plenum, New York. pp 1-94.

Li, W-H, Gouy, M., Sharp, P.M., O'Huigin, C., Yang, Y-W (1990) Molecular phylogeny of rodentia, lagomorpha, primates, artiodactyla, and carnivora and molecular clocks. Proc. Natl. Acad. Sci. 87:6703-6707.

Li, W-H., Graur, D. (1991) *Fundamentals of Molecular Evolution*. Sinauer Assoc., Sunderland MA.

Li, W-H. (1993) So, what about the molecular clock hypothesis? Curr. Op. Gen. Devel. 3:896-901.

Lopez, J.V. (1988) DNAse I hypersensitive sites and their correlation tot he differential expression of an exogenous thymidine kinase gene. Master of Science Thesis. Florida State University, Tallahasee, Florida.

Lopez, J.V., Yuhki, N., Masuda, R., Modi, W. S., O'Brien, S.J. (1994) *Numt*, a transposition and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J. Mol. Evol 39:174-190.

Lopez, J.V., Culver, M.C., Cevario, S. O'Brien, S.J. Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a nuclear mtDNA tandem repeat: Analysis of heterogeneous evolutionary rates. J. Mol. Biol. in preparation.

Locker J., Lewin A., Rabinowitz M. (1979) The structure and organization of mitochondrial DNA from petite yeast. Plasmid 2:155-181.

Love, J.M., Knight, A.M., McAleer, M.A., Todd, J.A. (1990) Towards construction of a high resolution map of the mouse genome using PCR analyzed microsatellites. Nucl. Acids Res. 14:4123-4130.

Lynch, M., Jarrell, P.E. (1993) A method for calibrating molecular clocks and its application to animal mitochondrial DNA. Genetics 135:1197-1208.

Lyons, L.A., Raymond, M.M., O'Brien, S.J. (1994) Comparative genomics: The next generation. Animal Biotechnology 5:103-112.

Madsen, C.S., Ghivizzani, S.C., Hauswirth, W.W. (1993) Protein binding to a single termination-associated sequence in the mitochondrial DNA D-loop region. Mol. Cell. Biol. 13:2162-2171.

Maga, E.A., Richardson, T. (1991) Amplification of a 9.0 kb fragment using PCR. Biotechniques 11:185-186.

Margulis L. (1970) *Origin of Eukaryotic Cells*. Yale University Press, New Haven, CT.

Margulis, L. (1993) *Symbiosis in Cell Evolution*. Freeman and Co., New York.

Martin, A.P, Palumbi, S.R. (1993) Body size, metabolic rate, generation time and the molecular clock. Proc. Natl. Acad. Sci 90:4087-4091.

Masuda, R.M., O'Brien, S.J., Pecon-Slattery, J., Yuhki, N., Lopez, J.V. (1995) Partial resolution of the cat family tree with mitochondrial 12S rRNA and cytochrome b gene sequences. In preparation.

Masuda R, Yuhki N, O'Brien SJ (1991) Molecular cloning, chromosomal assignment and nucleotide sequences of the feline homeobox HOX3A. Genomics 11:1007-1013.

Matthews, H.R., Pearson, M.D., MacLean, N. (1980) Cat satellite DNA. Isolation using netropisin with CsCl gradients. Biochim. Biophys. Acta 606:228-235.

Mazin, A.L., Boiko, L.M., Ogarkova, O.A., Vanhyushin, B.F. (1988) Loss of CpG dinucleotides from DNA. VI. Methylation of mitochondrial and chloroplast genes. Molekulyarnaya Biologiya 22:1688-1696.

McBride, T.J., Preston, B.D., Loeb, L.A. (1991) Mutagenic spectrum resulting from DNA damage by oxygen radicals. Biochemistry 30:207-213.

Meyer, A. (1994) Shortcomings of the cytochrome B gene as a molecular marker. Trends Ecol. Evol. 9:278-280.

Mignotte, F., Champagne, A.M., Gueride, M., Mounolou, J.C. (1990) Direct repeats in the noncoding region of rabbit mitochondrial DNA: involvement in the generation of intra- and inter-individual heterogeneity. Eur. J. Bioch 194:561.

Miklos, G.L.G. (1985) Localized highly repetitive DNA sequences in vertebrate and invertebrate genomes. In *Molecular Evolutionary Genetics*. (Ed. R.J. MacIntyre) pp. 241-313.

Mindell, D.P., Honeycutt, R.L. (1990) Ribosomal RNA in vertebrates: Evolution and phylogenetic applications. Annu. Rev. Ecol. Syst. 21:541-566.

Miquel, J. (1992) An update on the mitochondrial-DNA mutation hypothesis of cell-aging. Mutation Res. 275:209-216.

Miththapala, S., Seidensticker, J., Phillips, L.G., Goodrowe, K.L., Fernando, S.B.U., Forman, L., O'Brien, S.J. (1991) Genetic variation in Sri Lankan leopards. Zoo Biol. 10:139-146.

Miththapala, S. (1992) Genetic and morphological variation in the leopard (*Panthera pardus*): A geographically widespread species. Ph.D thesis. Univ. of Florida. 249 pp.

Miyamoto, M.M., Cracraft, J.(Eds.) (1991) Phylogenetic inference, DNA sequence analysis, and the future of molecular systematics. In *Phylogenetic Analysis of DNA Sequences*. Oxford University Press, New York.

Miyata, T., Hayashida, H, Kikuno, R., Hasegawa, M., Kobayashi, M. (1982) Molecular clock of silent substitutions: At least six-fold preponderance of silent substitutions in mitochondrial genes over those in nuclear genes. J. Mol. Evol. 19:28-35.

Modi, W.S., Fanning, T.G., Wayne, R.K., O'Brien, S.J. (1988) Chromosomal localization of satellite DNA sequences among 22 species of felids and canids (Carnivora) Cytogenet. Cell Genet. 48:208-213.

Modi, W.S. (1992) Nucleotide sequence and genomic organization of a tandem satellite array from the rock vole *Microtus chrotorrhinus* (Rodentia). Mammalian Genome 3:226-232.

Monnerot, M., J-C., Mounolou, Solignac, M. (1984) Intra-individual length hereogeneity of Rana esculenta mitochondrial DNA. Biol. Cell 52:213-218.

Moritz, C., Dowling, T. E., Brown, W.M. (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann. Rev. Ecol. Syst. 18:269-292.

Moriyama, E.N., Hartl, D.L. (1993) Codon usage bias and base composition of nuclear genes in Drosophila. Genetics 134:847-858.

Mullis, K.B., Faloona, F. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. 155:335-350.

Murakawa, G.J., Zaia, J.A., Spallone, P.A., Stephens, D.A., Kaplan, B.E., Wallace, R.B., Rossi, J.J. (1988) Laboratory methods: Direct detection of HIV-1 RNA from AIDS and ARC patient samples. DNA 7:287-295.

Myers, R.M., Larin, Z., and Maniatis, T. (1985) Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. Science 230:1242-1246.

Nass, M.M.K., Nass, S. (1963) Intramitochondrial fibers with DNA characteristics. I. Fixation and electron staining reactions. J.Cell Biol. 19:593-611.

Nei M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York

Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Soliz, Webster, T.D., Ledbetter, D.H., Caskey, C.T. (1989) Alu polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. Proc. Natl. Acad. Sci. 86:6686-6690.

Newfeld, S.J., Tachida, H., Yedvobnick, B. (1994) Drive-selection equilibrium: Homopolymer evolution in the *Drosophila* gene *mastermind*. J. Mol. Evol. 38:637-641.

Novacek, M.J. (1992) Mammalian phylogeny: shaking the tree. Nature 356:121-125.

Nowak R.M. (1991) *Walker's Mammals of the World.* 5th ed. The Johns Hopkins University Press Baltimore, MD.

Nugent J.M., Palmer J.D. (1991) RNA-mediated transfer of the gene COXII from the mitochondrion to the nucleus during flowering plant evolution. Cell 66:473-481.

O'Brien S.J., Nash W.G. (1982) Genetic mapping in mammals: chromosome map of domestic cat. Science 216:257-265.

O'Brien, S.J. (1986) Molecular genetics in the domestic cat and its relatives. Trends in Genet. 2:137-143.

O'Brien S.J., Collier G.E., Benveniste R.E., Nash W.G., Newman A.K., Simonson J.M., Eichelberger M.A., Seal U.S., Janssen D., Bush M., Wildt D.E. (1987) Setting the molecular clock in Felidae: the great cats, *Panthera.* In *Tigers of the World.* (Eds. Tilson R.L., Seal U.S.) Noyes Publications, Park Ridge, NJ, pp. 10-27.

O'Brien S.J., Roelke M.E., Yuhki N., Richards K.W., Johnson W.E., Franklin W.L., Anderson A.E., Bass O.L. Jr, Belden R.C., Martenson J.S. (1990) Genetic introgression within the Florida panther *Felis concolor coryi.* Natl Geo Res 6:485-494.

O'Brien, S.J. (1994) A role for molecular genetics in biological conservation. Proc. Natl. Acad. Sci. 91:5748-5755.

O'Brien, S.J. (1994) Genetic and phylogenetic analyses of endangered species. Annu. Rev. Genet. 28:467-489.

Ohta, T., Kimura, M. (1971) On the constancy of the evolutionary rate of cistrons. J. Mol. Evol. 1:18-25.

Ohta, T. (1992) The nearly neutral theory of molecular evolution. Ann. Rev. Ecol. Syst. 23:263-286.

Okimoto, R., Macfarlane, J.L., Clary, D.O., Wolstenholme, D.R. (1992) The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum.* Genetics 130:471-498.

Orgel, L.E., Crick, F.H.C. (1980) Selfish DNA: The ultimate parasite. Nature 284:604-607.

Pagel, M., Johnstone, R.A. (1992) Variation across species in the size of the nuclear genome supports the junk-DNA explanation fo the C-value paradox. Proc. R. Soc. London B 249:119-124.

Palumbi, S.R. (1989) Rates of molecular evolution and the fraction of nucleotide positions free to vary. J. Moi. Evol. 29:180-187.

Pathak, S., Wurster-Hill, D.H. (1977) Distribution of constitutive heterochromatin in carnivores. Cytogenet. Cell. Genet. 18:245-254.

Pecon Slattery, J., Johnson, W.E., Goldman, D., O'Brien, S.J. (1994) Phylogenetic reconstruction of South American felids defined by protein electrophoresis. J. Mol. Evol. 39:296-305.

Pepe, G., Holtrop M., Gadaleta G., Kroon, A.M., Cantatore P., Gallerani R., de Benedetto C., Quagliariello C., Sbisa', E., Saccone C. (1983) Non-random patterns of nucleotide substitutions and codon strategy in the mammalian mitochondrial genes coding for identified and unidentified reading frames. Biochem Int 6:553-563.

Pollack, Y., Kasir, J., Shemer, R., Shulamit, M., Szyf, M. (1984) Methylation pattern of mouse mitochondrial DNA. Nucl. Acids Res. 12:4811-4824.

Quigley F., Martin W.F., Ceriff R. (1988) Intron conservation across the prokaryotic-eukaryotic boundary: structure of the nuclear gene for chloroplast glyceraldehyde 3-phosphate dehydrogenase from maize. Proc Natl Acad Sci USA 85:2672-2676.

Radic, M.Z., Lundgren, K., Hamkalo, B.A. (1987) Curvature of mouse satellite DNA and condensation of heterochromatin. Cell 50:1101-1108.

Rand, D.M. (1993) Endotherms, ectotherms and mitochondrial genome-size variation. J. Mol. Evol. 37:281-295.

Rand, D.M. (1994) Thermal habit, metabolic rate and the evolution of mitochondrial DNA. Trends Ecol. Evol. 9(4):125-131.

Richter, C., Park, J.-W., Ames, B.N. (1988) Normal oxidative damage to mitochondrial and nuclear DNA is extensive. Proc. Natl. Acad. Sci. USA 85:6465-6467.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J.C., Markham, A.F. (1990) A novel, rapid method for the isolation of terminal

sequences from yeast artifical chromosome (YAC) clones. Nucl. Acids Res. 18(10):2887-2890.

Roberts, R.G., Coffey, A.J., Martin Bobrow, M., Bentley, D.R. (1992) Determination of the exon structure of the distal portion of the dystrophin gene by vectorette PCR. Genomics 13:942-950.

Roe, B.A., Ma, D.P., Wilson, R.K., Wong, J.F.H. (1985) The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J.Biol. Chem. 260:9759-9774.

Rose, A.M., Joyce, P.B.M., Hopper, A.K., Martin, N.C. (1992) Separate information required for nuclear and subnuclear localization: Additional complexity in localizing an enzyme shared by mitochondria and nuclei. Mol. Cell. Biol 12:5652-5658.

Ruvulo, M., Zehr, S., Dornum, M V, Pan, D., Chang, B., Lin, J. (1993) Mitochondrial COII sequences and modern human origins. Mol. Biol. Evol. 10:1115-1135.

Saccone C., Pesole G., Sbisa E. (1991) The main regulatory region of mammalian mitochondrial DNA: Structure-function model and evolutionary pattern. J Mol Evol 33:83-91.

Saitou N., Nei M. (1987) The neighbor joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sanger, F., Nicklen, S., Coulsen, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Sealy, L., Hartley, R., Donelson, J., Chalkley, R., Hutchison, N., Hamkalo, B. (1981) Characterization of a highly repetitive sequence family in rat. J. Mol. Biol. 145:291-318.

Schimke, R.T. (1980) Gene amplification and drug resistance. Sci. Amer. 243:60-69.

Schinkel A.H., Tabak H.F. (1989) Mitochondrial RNA polymerase: dual role in transcription and replication. Trends Genet 5:149-154.

Schlegel, M. (1994) Molecular phylogeny of eukaryotes. Trends in Ecol. Evol. 9:330-335.

Schon E.A., Rizzuto R., Moraes C.T., Nakase H., Zeriani M., Dimauro S (1989) A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA. Science 24:346-349

Serikawa T., Kuramoto T., Hilbert P., Mori M., Yamada J., Dubay C.J., Lindpainter K., Ganten D., Guenet J.L., Lathrop G.M., Beckmann J.S. (1992) Rat gene mapping using PCR-analyzed microsatellites. Genetics 131:701-702

Slade, R.W., Moritz, C., Heideman, A. (1994) Multiple nuclear-gene phylogenies: Application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny. Mol. Biol. Evol. 11:341-356.

Smith, D.C., Douglas, A. (1989) *The Biology of Symbiosis*. E.A. Arnold, London.

Smith, M.F., Thomas, W.K., Patton, J.L. (1991) Mitochondrial DNA-like sequence in the nuclear genome of an akodintine rodent. Mol. Biol. Evol. 9:204-215.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

Stephan, W. (1989) Tandem-repetitive noncoding DNA: Forms and forces. Mol. Biol. Evol. 198-212.

Stewart, D.T., Baker, A.J. (1994) Evolution of mtDNA D-loop sequences and their use in phylogenetic studies of shrews in the subgenus *Otisorex* (Sorex:Soricidae:Insectivora). Mol. Phylo. Evol. 3(1):38-46.

Stoflet, E.S., Koegerl, D.D., Sarkar, G., Sommer, S.S. (1988) Genomic amplification with transcript sequencing. Science 239:491-494.

Stohl, L.L., Clayton, D.A. (1992) *Saccharomyces cerevisiae* contains an RNase MRP that cleaves at a conserved mitochondrial RNA sequence implicated in replication priming. Mol. Cell Biol. 12(6):2561-2569.

Swofford, D.L. (1993) Phylogenetic anlysis using parsimony (PAUP), version 3.1.1. Smithsonian Institution, Washington D.C.

Tautz, D., Trick, M., Dover, G.A. (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322:652-656.

Ticher, A., Graur, D. (1989) Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein coding genes. J Mol Evol 28:286-298.

Tsuzuki, T., Nomiyama, H., Setoyaja, C., Maeda, S, Shimada, K. (1983) Presence of mitochondrial-DNA-like sequences in the human nuclear DNA. Gene 25:223-229.

Tzagoloff, A., Myers, A.M. (1986) Genetics of mitochondrial biogenesis. Ann. Rev. Biochem. 55:249-285.

Van den Boogaart P., Samalio J., Agsteribbe E. (1982) Similar genes for a mitochondrial ATPase subunit in the nuclear and mitochondrial genomes of *Neurospora crassa*. Nature 298:187-189

Van der Vlugt, H.H.J., Lenstra, J.A. (1994) SINE elements in Carnivores. Submitted.

Vawter, L., Brown, W.M. (1986) Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in their molecular clock. Science 234:194-196.

Wainright, P.O., Hinkle, G., Sogin, M.L., Stickel, S.K. (1993) Monophyletic origins of the metazoa: an evolutionary link with fungi. Science 260:340-342.

Wahls W.P., Wallace L.J., Moore P.D. (1990) Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. Cell 60:95-103.

Wakasugi, S., Nomiyama, H., Fukuda, M., Tsuzuki, T., Shimada K. (1985) Insertion of a long Kpn I family member within a mitochondrial-DNA-like sequence present in the human nuclear genome. Gene 36:281-288.

Wallace, D.C. (1982) Structure and evolution of organelle genomes. Microbiol. Rev. 46:208-240.

Wallace D.C., Lott M.T., Torroni A., Shoffner J.M. (1991) Report of the committee on human mitochondrial DNA. Cytogenet Cell Genet 58:1103-1123.

Wallace, D.C. (1992) Diseases of the mitochondrial DNA. Annu. Rev. Biochem 61:1175-1212.

Wallace, D.C., Lott, M.T., Torroni, A., Brown, M.D. (1993) Report of the committee on human mitochondrial DNA. Genome Priority Reports 1:727-757.

Warburton, P.E., Waye, J.S., Willard, H.F. (1993) Nonrandom localization of recombination events in human alpha satellite repeat unit variants: Implications for higher-order structural characteristics within centromeric heterchromatin. Mol. Cell. Biol 13:6520-6529.

Wayne, R.K., Benveniste, R.E., Janczewski, D.N., O'Brien, S.J. (1989) Molecular and biochemical evolution of the Carnivora. In *Carnivore Behavior, Ecology and Evolution*. (Gittleman, J.L., Ed.) Cornell University Press, Ithaca, NY. pp. 465-494.

Wayne, R.K, Valkenburgh, B.V., O'Brien, S.J. (1991) Molecular distance and divergence time in carnivores and primates. Mol. Biol. Evol. 8:297-319.

Weber J.L. (1990) Human DNA polymorphisms based on length variations in simple-sequence tandem repeats. Genome Analysis 1:159-181

Weiner, A.M., Deininger, R.L., Efstratisdis, A. (1986) Nonviral retrposons: Genes, pseudogenes and transposablew elements generated by reverse flow of genetic information. Annu. Rev. Biochem. 55:633-61.

Wilkinson, G.S., Chapman, A.M. (1991) Length and sequence variation in evening bat D-loop mtDNA. Genetics 128:607-617.

Willard, H.F. (1990) Centromeres of mammalian chromosomes. Trends in Genetics 6(12):410-416.

Wilson, A.C., Cann, R.L., Carr, S.M., George, M., Gyllensten, U.B., Helm-Bychowski, K.M., Higuchi, R.G., Palumbi, S.R., Prager, E.M., Sage, R.D., Stoneking, M. (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linnean Soc. 26:375-400.

Wilson, E.B. (1959) *The Cell in Development and Heredity*. Macmillan, New York.

Woese, C.R. (1987) Bacterial evolution. Micro Rev 51:221-271

Wolfe, K.H., Li, W.-H., Sharp, P.M. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84:9054-9058.

Wolfe, K.H., Sharp, P.M., Li, W.-H. (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283-285.

Wolstenholme, D.R. (1992) Genetic novelties in mitochondrial genomes of multicelluar animals. Curr. Op. Gen. & Devei. 2:918-925.

Wu, C. (1980) The 5' ends of *Drosophila* heat-shock genes in chromatin are sensitive to DNAse I. Nature 286:854-860.

Wu, C-I., Li, W.-H. (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA 1741-1745.

Wu, C.-I., True, J.R., Johnson, N. (1989) Fitness reduction associated with the deletion of a satellite DNA array. Nature 341:248-251.

Wu, C.-I, Hammer, M.F. (1991) Molecular evolution of ultraselfish genes of meiotic drive systems. In *Evolution at the Molecular Level* (Eds. R.K. Selander, A.G. Clark, T.S. Whittam). Sinauer, Sunderland, Mass.

Wurster-Hill, D.H., Centerwall, W.R. (1982) The interrelationships of chromosome patterns in canids, mustelids, hyena and felids. Cytogenet Cell Genet 34:178-192

Yamamoto, M. (1979) Cytological studies of heterochromatin function in *D. melanogaster* males: autosomal meiotic pairing. Chromosoma 72:293-328.

Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G.J., Woese, C.R. (1985) Mitochondrial origins. Proc Natl Acad Sci USA 82:4443-4447.

Yuhki, N., O'Brien, S.J. (1990) DNA recombination and natural selection pressure sustain genetic sequence diversity of the feline MHC class I genes. J. Exp. Med. 172:621-630.

Yuhki, N., O'Brien, S.J. (1994) Exchanges of short polymorphic DNA segments predating speciation in feline major histocompatibilty complex class I genes. J. Mol. Evol. 39:22-33.

Zouros, E., Ball, A.O., Saavedra, C., Freeman, K.R. (1994) An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. Proc. Natl. Acad. Sci. USA 91:7463-7467.

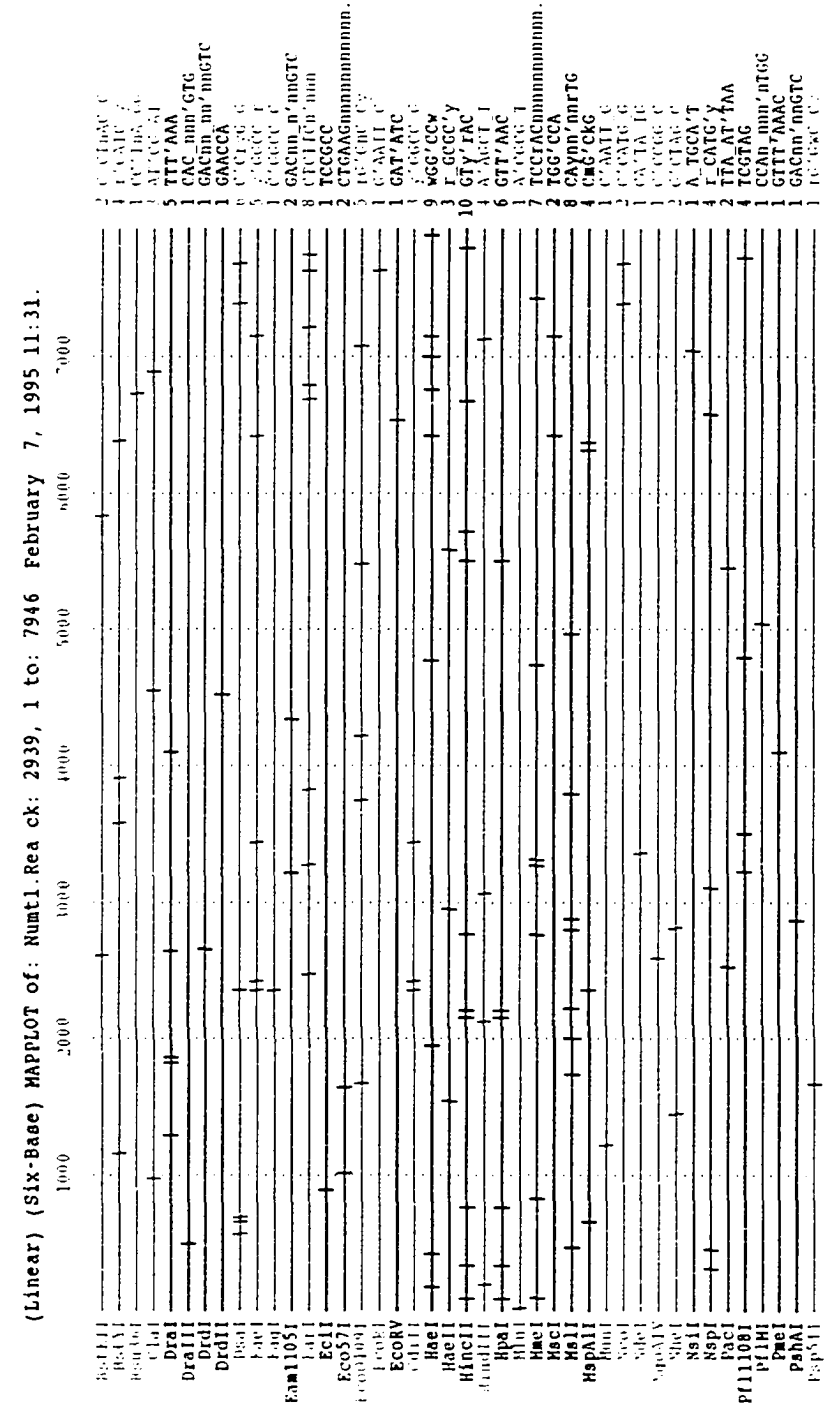
Zuckerkandl, E. (1992) Revisiting junk DNA. J. Mol. Evol. 34:259-271.

Zuker, M., Steigler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucl. Acids Res. 9:133-148.

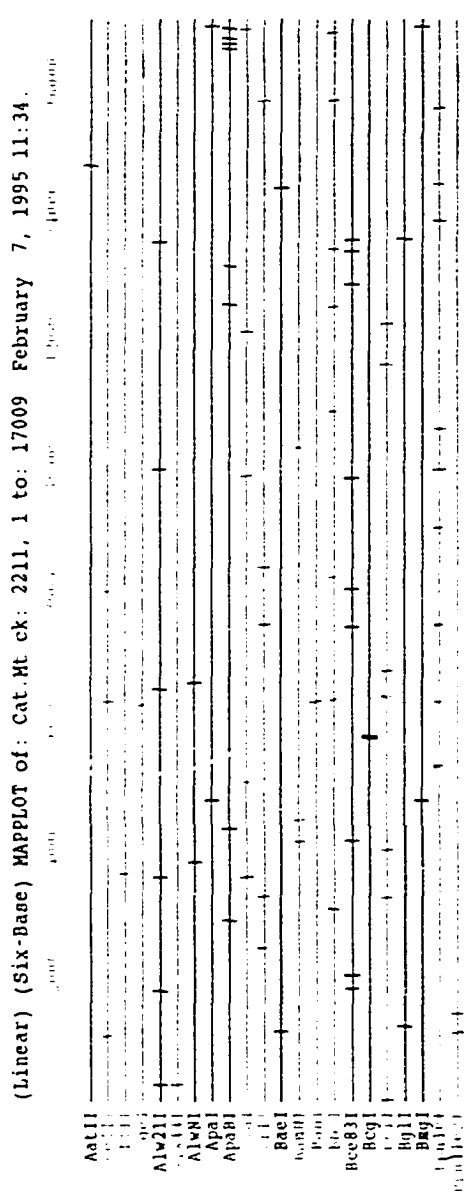
Zullo, S., Sieu, L.C., Slightom, J.L., Hadler, H.I., Eisenstadt, J. M. (1991) Mitochondrial D-loop sequences are integrated in the rat nuclear genome. J. Mol. Biol 221:1223-1235.

Appendices

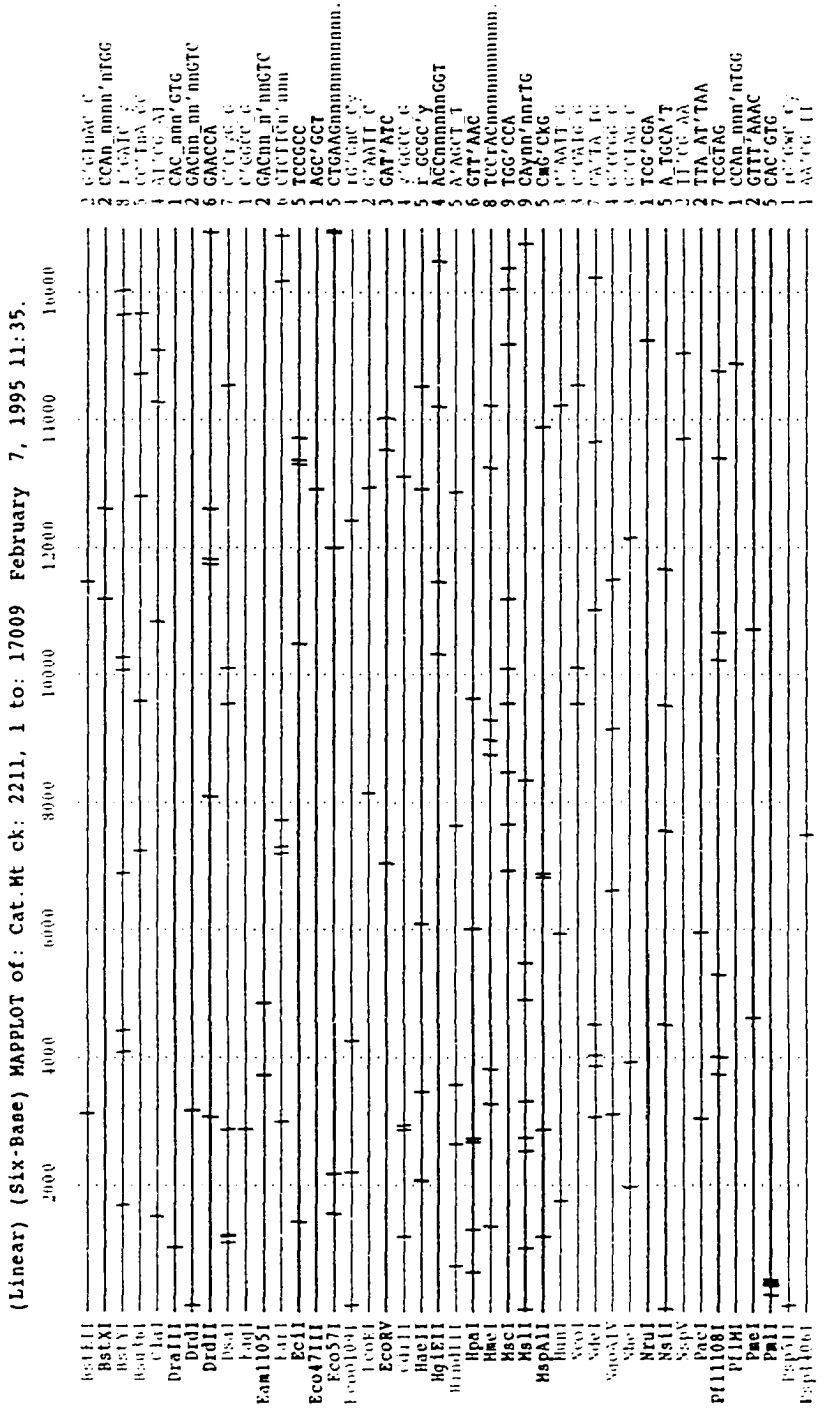
Appendix A (Continued)



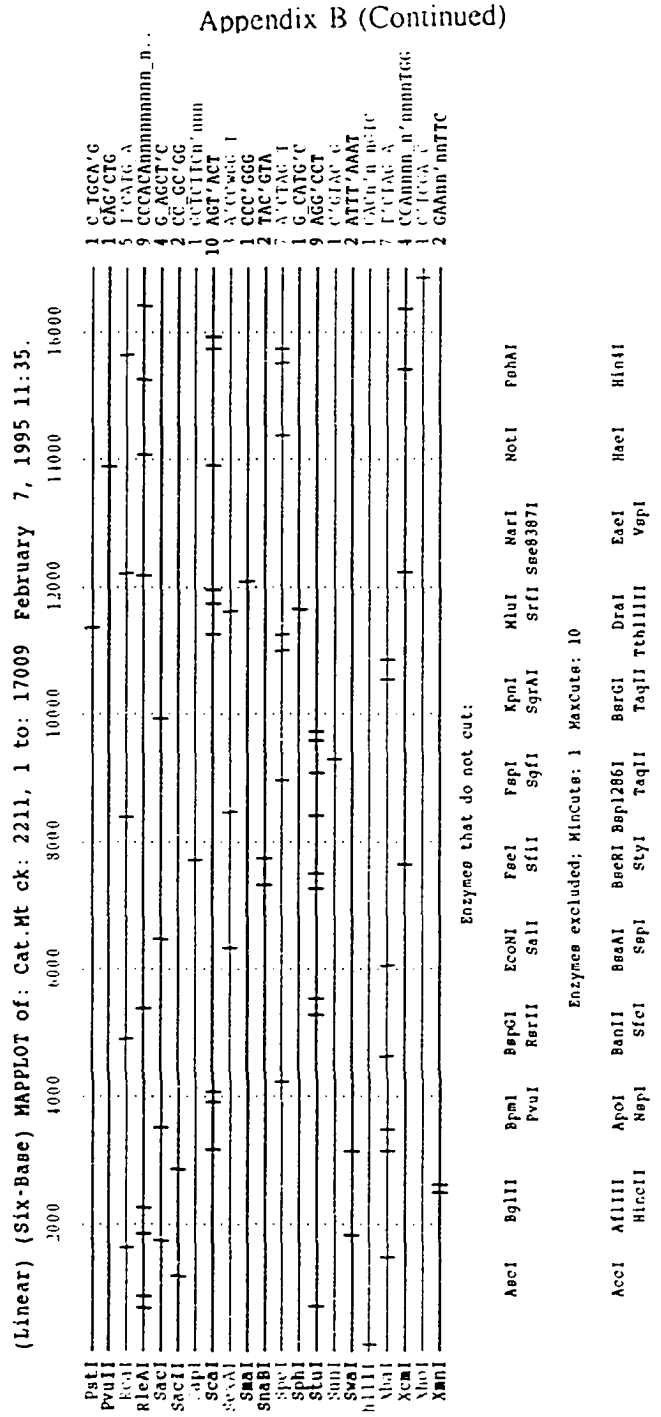
Appendix B



Appendix B (Continued)



Appendix B (Continued)



Appendix C. Total Sizes of Various Vertebrate Mt Genomes

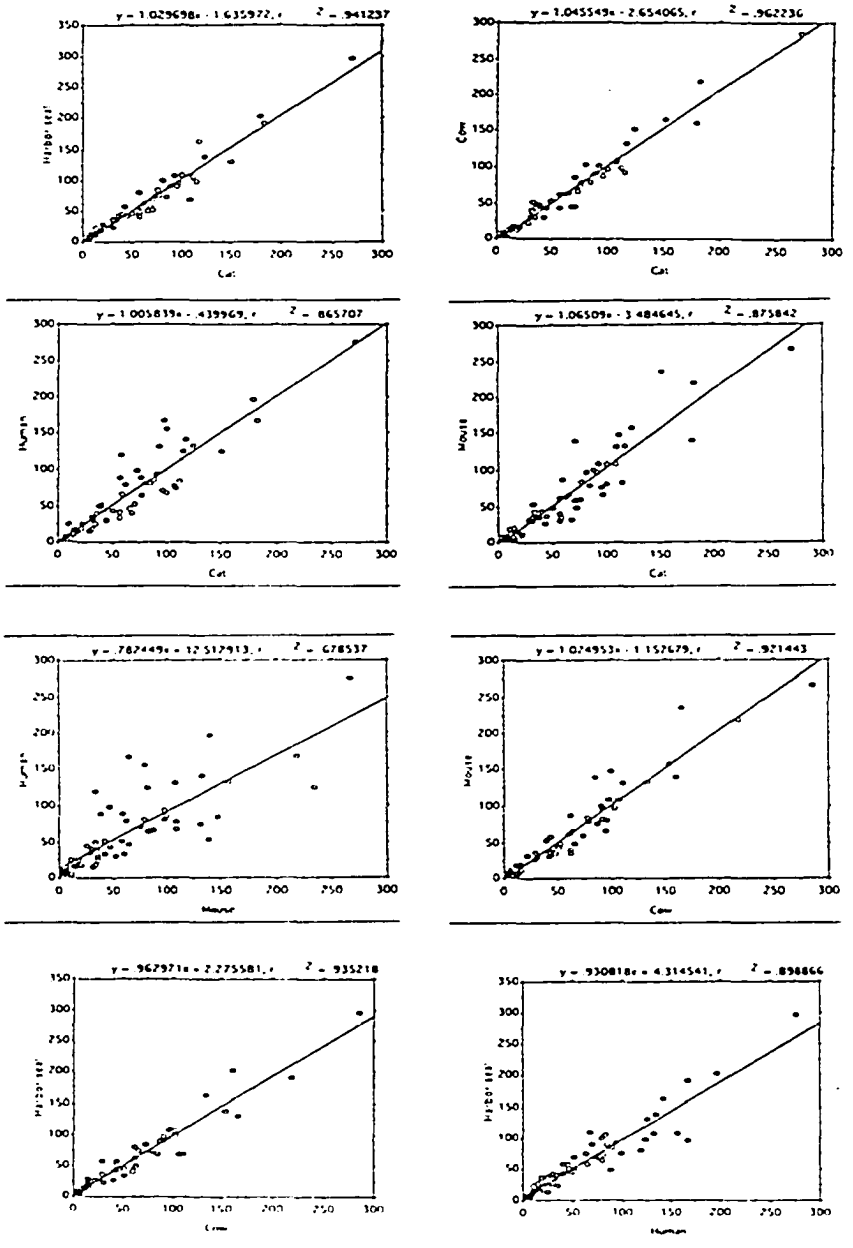
Species				
Common Name	Binomial	Size (bp)	Std. Dev	Reference
Domestic cat	*Felis catus*	**17009**		Lopez et al, in prep
Cow	*Bos taurus*	**16338**		Anderson et al, 1982
Mouse	*Mus musculus*	**16295**		Bibb et al, 1981
Norway rat	*Rattus norvegicus*	**16298**		Gadeleta et al, 1989
Human	*Homo sapiens*	**16569**		Anderson et al, 1981
Harbor seal	*Phoca vitulina*	**16826**		Arnason and Johnsson, 1992
Grey seal	*Halichoerus grypus*	**16797**		Arnason et al, 1993
Fin whale	*Baleaenoptera physalus*	**16398**		Arnason et al, 1991
Blue whale	*Baleaenoptera musculus*	**16402**		Arnason and Gullberg, 1993
Chicken	*Gallus gallus*	**16775**		Desjardins and Morais, 1990
Carp	*Cyprinus carpio*	**16575**		Chang et al, 1994
Sea lamprey	*Petromyzon marinus*	**16201**		Lee and Kocher, 1995
African toad	*Xenopus laevis*	**17700**		Roe et al, 1985
Domestic rabbit^	*Oryctolagus cuniculus*	**17300**		Brown, 1983
	Avg. =	**16677**	+430	

- Data was also available from GenBank release 86.0 (12.94)
^ - Inferred from techniques other than direct DNA sequencing.

Appendix D. Pairwise comparisons of codon usage for all structural genes in the mitochondrial genomes of four mammalian orders. After assembling all protein-coding reading frames into one contiguous file, codon usage in each separate species (domestic cat, human, cow, mouse, and harbor seal) was determined with the UWGCG program CODON FREQUENCY. Values for each of the 64 possible codons were entered into the STATVIEW or STATGRAPH analysis program and plotted. A primary observation from this analysis is that codon usage among these mammals is very similar. The lowest correlations are consistently observed with the human sequence, representing the primate lineage.

Appendix D

Comparison of total mtDNA codon usage



Vita

Jose Victor (Quedding) Lopez was born on September 12, 1962 in Quezon City, Republic of the Philippines to Ernesto G. Lopez, a pathologist and Rosario Quedding, a medical technologist. His family immigrated to Canada in 1964, and all became naturalized U.S. citizens in 1971. After living in New York, Virginia and Ohio, Jose and his family settled in Waycross, Georgia in 1973. There, Jose graduated from high school with honors (National Honor Society, Presbyterian College and Lion's Club Academic Achievement Awards) and matriculated to the Georgia Institute of Technology in Atlanta, Georgia (1980 - 1984) for a Bachelor of Science degree in Applied biology, with an eye towards working in the relatively new field of "genetic engineering". Immediately following, Jose completed a Master of Science degree program (1985-1988) in Biological sciences at Florida State University in Tallahassee, Florida under the distinguished professor and member of the National Academy of Sciences, Dr. J. Herbert Taylor. After graduation, Jose temporarily assumed a research technician position at the National Cancer Institute's (NCI) Laboratory of Biology with Dr. Jay Doniger in Bethesda, Maryland. Eventually, the gravity and urgency of molecular evolution and conservation genetics questions drew Jose to study under Dr. Stephen O'Brien, chief of the NCI's Laboratory of Viral Carcinogenesis and editor of *Genetic Maps* and the *Journal of Heredity*, in Frederick Maryland, as well as the doctoral program in Environmental Biology and Public Policy at George Mason University in Fairfax, Virginia, since 1990.