BioOne COMPLETE

The Occurrence, Detection, and Avoidance of Mitochondrial DNA Translocations in Mammalian Systematics and Phylogeography

Authors: Triant, Deborah A., and DeWoody, J. Andrew

Source: Journal of Mammalogy, 88(4) : 908-920

Published By: American Society of Mammalogists

URL: https://doi.org/10.1644/06-MAMM-A-204R1.1

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Journal of Mammalogy, 88(4):908–920, 2007

THE OCCURRENCE, DETECTION, AND AVOIDANCE OF MITOCHONDRIAL DNA TRANSLOCATIONS IN MAMMALIAN SYSTEMATICS AND PHYLOGEOGRAPHY

DEBORAH A. TRIANT* AND J. ANDREW DEWOODY

Department of Forestry & Natural Resources, Purdue University, West Lafayette, IN 47907-1159, USA

The generation and analysis of mitochondrial DNA (mtDNA) sequence data has become routine in mammalogy. Unfortunately, these analyses can be confounded because fragments of the mitochondrial genome are contained in the nucleus of most eukaryotes. Furthermore, these nuclear fragments of mitochondrial genes, or numt pseudogenes, are often represented hundreds of times in mammalian nuclear genomes. Most modern analyses of mtDNA rely on the polymerase chain reaction to generate a population of molecules that can be sequenced. Templates for DNA sequencing reactions should be homogenous, and in the case of mtDNA, cytoplasmic in origin. The unwanted (and often unwitting) amplification of numts results in a heterogenous mixture of nuclear and cytoplasmic amplicons or, if a numt is preferentially amplified, a near-homogenous mixture of the wrong (nuclear) template. These nuclear sequences can cause major—although often cryptic—problems in the analyses of systematic or phylogeographic data. Here, we review the occurrence, detection, and avoidance of numts in mammals. Furthermore, we isolate a cytochrome-*b* numt and its corresponding mitochondrial sequence in the North American prairie vole (*Microtus ochrogaster*) to illustrate various methods to detect numts. Finally, we present approaches by which numts, once identified, can be utilized in molecular studies.

Key words: cytochrome *b*, genetic species concept, mitochondrial DNA, *Microtus*, numt, pseudogene

The mammalian mitochondrial genome possesses a number of characteristics that have made it a valuable molecular marker in population genetics, conservation biology, phylogeography, and systematics (Avise 1991). Since the publication of the genetic species concept of Bradley and Baker (2001), mammalogists have increasingly relied upon mitochondrial DNA (mtDNA) sequences in evolutionary studies. Unfortunately, many of these mtDNA appraisals have failed to consider the potentially confounding influence of nuclear sequences on their analyses. Despite the physical separation of the mitochondrion and the nucleus within the cell, fragments of the mtDNA genome have been found within the nuclear genome. Because these nuclear sequences can be highly similar to their mtDNA counterparts, they may be isolated along with mtDNA sequences during molecular studies.

Nuclear mitochondrial translocations (numt pseudogenes or numts—Lopez et al. 1994) are mtDNA fragments that have been incorporated into the nuclear genome (see Bensasson et al. 2001; Leister 2005; Zhang and Hewitt 1996 for reviews). First documented in the 1960s (du Buy and Riley 1967), they have subsequently been identified in a wide variety of plants and animals, including many mammalian lineages (Table 1). More than 600 numt pseudogenes have been documented in the human genome alone (Woischnik and Moraes 2002). Numts can originate from different portions of the mitochondrial genome, can vary in their degree of similarity to their corresponding mtDNA fragments, and can encompass multiple genes or mere fragments of genes. Richly and Leister (2004) analyzed numts within sequenced mtDNA and nuclear eukaryotic genomes and found that there was no apparent correlation between numt size or abundance with genome size or gene density. Fragments of the mtDNA control region are underrepresented in the human genome sequence (Mourier et al. 2001; Woischnik and Moraes 2002), but it is still unclear whether or not this is true in other organisms.

Although the original numts that colonized mammalian genomes probably originated via independent translocation events, once integrated into a chromosome these DNA sequences are subject to the same gene duplication processes that give rise to gene families (Triant and DeWoody, in press). The exact proportion of numts that are derived from independent translocations versus nuclear duplications is unclear (Bensasson et al. 2003; Hazkani-Covo et al. 2003; Mirol et al. 2000), but whatever the absolute contribution of insertions versus

* *Correspondent: dtriant@cvm.tamu.edu*

© **2007 American Society of Mammalogists**
www.mammalogy.org

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

TABLE 1.—A sample of nuclear mitochondrial translocations (numt pseudogenes or numts) that have been isolated from a wide variety of mammalian taxa. Intervening tRNA sequences, if present, are not listed. In most cases, the genes involved represent a minimum accounting because insertion sites and flanking sequences of numts are uncommon in published literature. In many cases, numerous numts are known to have occurred independently in the same lineage. *Cytb*, cytochrome *b*; 12S, 12S rRNA; 16S, 16S rRNA; C.R., control region; COI, cytochrome oxidase I; COII, cytochrome oxidase 2; mtDNA, mitochondrial DNA; ND1, reduced nicotinamide adenine dinucleotide (NADH) dehydrogenase subunit 1; ND2, (NADH) dehydrogenase subunit 2.

Order	Family	Species	Common name	Translocated mtDNA region	Length (kb)	Citation
Didelphimorphia	Didelphidae	*Didelphis virginiana*	Virginia opossum	*Cytb*; C.R.	~0.6	Lemos et al. 1999
Rodentia	Ctenomyidae	*Ctenomys* (several species)	Tuco-tuco	*Cytb*	~0.3	Mirol et al. 2000
Rodentia	Cricetidae	*Abrothrix jelskii*	Ornate akodont	*Cytb*	<0.4	Smith et al. 1992
Rodentia	Cricetidae	*Microtus arvalis*	Common vole	*Cytb*	≥1.14	DeWoody et al. 1999
Rodentia	Muridae	*Mus musculus*	House mouse	Multiple comprising entire mtDNA genome	12.4 (largest)	Richly and Leister 2004
Rodentia	Muridae	*Rattus norvegicus*	Brown rat	C.R.	0.5	Zullo et al. 1991
Carnivora	Felidae	*Felis catus*	Domestic cat	C.R.; 12S; 16S; ND1; ND2; CO1; COII	7.9	Lopez et al. 1994
Carnivora	Felidae	*Panthera tigris*	Tiger	*Cytb*; C.R.	2.5	Cracraft et al. 1998
Proboscidea	Elephantidae	*Elephas maximus*	Asian elephant	C.R.	<0.3	Greenwood and Pääbo 1999
Artiodactyla	Bovidae	*Kobus kob* and *K. vardonii*	Kob and puku	*Cytb*	1.1	Birungi and Arctander 2001
Artiodactyla	Bovidae	*Ovibos moschatus*	Muskox	*Cytb*	0.8	MacPhee et al. 2005
Artiodactyla	Cervidae	*Cervus elaphus*	Red deer	*Cytb*	0.4	Lister et al. 2005
Primates	Cercopithecidae	*Macaca mulatta*	Rhesus monkey	12S	~0.1	Vartanian and Wain-Hobson 2002
Primates	Hominidae	*Gorilla gorilla*	Western gorilla	C.R.	~0.4	Jensen-Seaman et al. 2004; Thalmann et al. 2004
Primates	Hominidae	*Homo sapiens*	Humans	Multiple comprising entire mtDNA genome	14.7 (largest)	Ricchetti et al. 2004
Primates	Lorisidae	*Nycticebus coucang* and *N. pygmaeus*	Slow loris	*Cytb*	~0.4	Lü et al. 2002
Perissodactyla	Rhinocerotidae	*Coelodonta antiquitatis*	Woolly rhinoceros	*Cytb*; 12S	1.2	Orlando et al. 2003
Proboscidea	Elephantidae	*Mammuthus primigenius*	Woolly mammoth	C.R.	0.08	Greenwood et al. 1999

duplications, the fact remains that mammalian nuclear genomes are replete with fragments of mtDNA.

How and why mtDNA fragments incorporate themselves into the nuclear genome has also been the subject of speculation. Some suggest that numt integration may be associated with chromosomal repair mechanisms (Willett-Brozick et al. 2001), whereas others postulate that mobile elements may somehow facilitate numt transfer (Mishmar et al. 2004). Numts are generally thought to be nonfunctional once incorporated into the nucleus because of the different genetic codes utilized by the mitochondrial and nuclear genomes. Although their accumulation may have helped shape the evolution of mammalian genomes, numts may not be completely benign. For example, at least 27 numts have colonized the human genome in the last 4–6 million years (Ricchetti et al. 2004). Of these, 23 (85%) are inserted directly into genes (usually introns), including critical tumor suppressor genes. Turner et al. (2003) described a numt insertion in humans that caused a stop codon in a functional gene, leading to a truncated polypeptide that could potentially impede normal development.

Analyses of mtDNA should always consider numts as potential sources of contamination. Unrecognized numts can appear to be unique mtDNA haplotypes, which can then confound downstream molecular analyses (Collura and Stewart 1995; Hirano et al. 1997; Parr et al. 2006). For example, Jensen-Seaman et al. (2004) re-examined mtDNA control region sequences of gorillas (*Gorilla*) and found that multiple haplotypes deposited in the GenBank database were, in fact, numts. Another study reported that divergent primate mtDNA haplotypes thought to have contaminated polio vaccines were later discovered to be macaque (genus *Macaca*) numts (Vartanian and Wain-Hobson 2002).

Systematic and phylogeographic assessments can be compromised if sequences with different evolutionary histories are utilized without knowledge of their true ancestry. Because the mtDNA genome is effectively haploid, is under different selection pressures than the nuclear genome, and exhibits a faster rate of nucleotide substitution, the inclusion of nuclear DNA in mtDNA data sets can lead to inaccurate species identifications, divergence estimates, phylogenetic groupings, and population breaks (Zhang and Hewitt 1996). Possible signs of numt contamination include multiple bands during electrophoresis or in restriction profiles and ambiguities in chromatograms, but numts may appear as unambiguous sequences if they are amplified preferentially. Fortunately, there are many ways by which numts can be identified and thus excluded from analyses

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

as confounding factors. Several approaches to avoid and detect numts are described below.

Avoiding numts

Isolation of entire mtDNA genomes.—Traditionally (i.e., before the advent of polymerase chain reaction [PCR]), mtDNA genomes were isolated in their entirety by ultracentrifugation in a salt gradient such as cesium chloride (Lansman et al. 1981). Unfortunately, this method is labor-intensive and expensive, but entire intact mtDNA molecules can now be isolated using spin chromatography. In brief, genomic DNA, isolated using conventional procedures, is passed through a column containing a filter matrix that preferentially binds DNA of a given size. Many such columns will accommodate mammalian mtDNA (~16–17 kilobases [kb]) and yield virtually pure product. This procedure is less labor intensive and less expensive than salt-gradient extractions and many commercial companies make spin-chromatography extraction kits (e.g., Bio-Rad, Hercules, California; Promega, Madison, Wisconsin; and Qiagen, Valencia, California). However, if primers preferentially amplify nuclear contaminants rather than mtDNA target sequences, purified mtDNA can still yield numt sequences (Collura and Stewart 1995).

Tissue sources.—Tissue sources that are rich in mtDNA relative to nuclear DNA (e.g., muscle and liver) can also help to reduce numt occurrence. Numts have been found in equal proportion to mtDNA sequences in avian blood samples, which have nucleated red blood cells (Sorenson and Quinn 1998). Conversely, Greenwood and Pääbo (1999) found that mammalian blood samples, which contain anucleated red blood cells, were a good source of mtDNA. The authors used primers that amplified a portion of the control region in elephant blood but found that those same primers amplified nuclear insertions in hair samples taken from the same individual. This study illustrates that primers may amplify mtDNA in one tissue type but numts in another. If multiple tissue types are available, mtDNA sequences can be verified from multiple sources.

Long-distance PCR.—Numts can vary in size, but most are small (<1 kb—Richly and Leister 2004). Thus, numts often can be avoided by using PCR primers that amplify substantial portions of the mtDNA molecule (i.e., more than 1 mtDNA gene). This approach supplies a long amplicon from which smaller amplicons of interest can be isolated using internal primers (e.g., Thalmann et al. 2004; Triant and DeWoody 2006). Long amplicons, when sequenced, should reveal only open reading frames if they are mitochondrial in origin. If the sequence is nuclear in origin, longer fragments will likely reveal numt features such as stop codons or frame-shift mutations (see below). Sequencing a long amplicon beyond the mtDNA gene(s) of interest can also reveal numt insertion sites as indicated by the sequence abruptly falling out of alignment with its corresponding mtDNA sequence. A related approach takes advantage of the circularity of the mtDNA genome; mtDNA fragments can be isolated with outward-extending primers in a manner similar to inverse PCR (Ochman et al. 1988). Using this approach, numts can be avoided regardless of their length because such inverse primers should not amplify linear fragments.

Detecting numts

Restriction digests.—Signs of numt contamination often include unexplained banding patterns during electrophoresis and restriction assays. If bands appear to be pure mtDNA during electrophoresis, the PCR product then can be digested with restriction enzymes that cut within the mtDNA fragment to further identify potential numts. Sequencing should distinguish any suspect numts that exhibit spurious bands. With large data sets, digesting PCR products with restriction enzymes before sequencing can alert researchers to possible numt contamination and save valuable time and resources.

Cloning.—Theoretically, PCR amplification and subsequent cloning of haploid mtDNA should yield recombinants identical in sequence. In practice, this is seldom the case because of polymerase (e.g., *Taq*) errors, heteroplasmy, and cloning artifacts (Baker et al. 1999). Despite this background level of nucleotide variation, numts have been isolated via T/A cloning of PCR products. T/A cloning can identify numts that appear as ambiguities in chromatograms or at such low levels that they are not otherwise detected through conventional sequencing (DeWoody et al. 1999). This procedure takes advantage of the tendency of some DNA polymerases, such as *Taq*, to add a 3′-A overhang to the end of PCR products. These PCR products can then be ligated into a vector with a complimentary 3′-T overhang, cloned, and sequenced (Sambrook and Russell 2001).

Gene expression.—Most interspecific systematic assays in mammalogy rely on protein-coding genes (e.g., cytochrome *b*), but numts are usually thought to be unexpressed. Thus, one can potentially confirm the mitochondrial origin of sequences via reverse-transcriptase PCR. This approach relies upon the isolation of polyadenylated rRNAs and mRNAs (Fernández-Silva et al. 2003; Ojala et al. 1981) using a standard poly-T or random priming protocol (Sambrook and Russell 2001) and the subsequent conversion of mRNA into cDNA using reverse transcriptase; the resulting cDNA can then be used for PCR. This approach avoids unexpressed templates (i.e., most numts) but has the disadvantage of requiring fresh tissue for RNA extraction. Although some numts are occasionally transcribed (Blanchard and Schmidt 1996), this procedure can enrich for mtDNA template.

Fluorescent in situ hybridization.—Another time-consuming but powerful method to identify potential numts and their chromosomal location within the nuclear genome is fluorescent in situ hybridization (Rudkin and Stollar 1977). Fluorescent in situ hybridization employs fluorescent microscopy to detect labeled DNA probes that have been hybridized to metaphase chromosomes and is useful in assessing numt position and copy number (Kim et al. 2006; Lopez et al. 1994). The suspect numt sequence can be used as a probe and hybridized to chromosome spreads resulting in a fluorescent signal visible at the sites of probe hybridization (i.e., putative numt integration—

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Trask 1991). Alternatively, the entire mtDNA genome can be used as a probe to assess whether the nuclear genome contains multiple numt copies. Probe-labeling kits are commercially available (e.g., Roche, Pleasanton, California; and Sigma, St. Louis, Missouri) but probes often need to be at least ~5 kb in length to capture any fluorescent signal (Trask 1991) and the preparation of chromosomal slides requires fresh cells from live-captured animals (Baker et al. 2003).

Comparative sequence analysis, translations, and secondary structures.—Numts are no longer under the strong selective constraints found in the mtDNA genome; thus, they should not exhibit codon position bias (e.g., selection against changes at the 2nd codon position). Numts also should lack the pronounced transitional bias found in animal mtDNA, depending upon their date of translocation to the nucleus (Bensasson et al. 2001). Substitution patterns inferred from pairwise comparisons of putative numts with known mtDNA sequences from related taxa can reveal whether a sequence is indeed nuclear, but this approach is not foolproof as some numts and their mtDNA complements have highly similar nucleotide compositions (Kim et al. 2006; Lopez et al. 1996; Triant and DeWoody 2007).

Functional protein-coding genes require open reading frames and most DNA analysis software can easily search for the presence of open reading frames in a sequence. Stop codons, insertions–deletions (indels), or frame-shift mutations within a coding mtDNA sequence are likely indicative of a numt, although recent translocations that have not yet accumulated such degenerative mutations could still possess an open reading frame. Unlike protein-coding genes, the control region, rRNA, and tRNA genes are not constrained by open reading frames and thus can be particularly problematic. However, numts derived from rRNAs and tRNAs can be identified through the evaluation of secondary structures (Sorenson and Quinn 1998).

The alignment of suspected numt sequences with a model based upon secondary structure can determine whether nucleotide substitutions cause structural abnormalities (e.g., disruptions to conserved stem-and-loop structures in tRNA—Hickson et al. 1996). Such comparisons are facilitated by the availability of public databases that curate secondary structure information (e.g., Van de Peer et al. 2000). For example, Pereira and Baker (2004) used the secondary structures of tRNA genes to detect numt pseudogenes in the chicken genome and found that most tRNA numt sequences were not predicted to fold into the proper secondary structure. On the other hand, Olson and Yoder (2002) attempted to use secondary structures to identify 12S rRNA numts but were unable to do so; they advocate using other methods (such as those we discussed herein) in addition to secondary structure analyses.

Phylogenetic analysis.—With a prior knowledge of phylogenetic relationships, numt paralogs can often be detected by atypical branch lengths or incorrect placement within a clade (Arctander 1995). Once integrated into the nuclear genome, the substitution rate of the translocated fragment should decelerate because the substitution rate within the mitochondrial genome is higher than that found in nuclear DNA. The mtDNA genome lacks the proofreading and repair mechanisms found in the nuclear genome, suffers from cumulative oxidative damage, and replicates more frequently than the nuclear genome; thus, mammalian mtDNA can accumulate ~10 times as many mutations as nuclear DNA (Brown et al. 1979). Most of these changes occur at 3rd codon positions or within intergenic regions (Avise 1991). Upon transfer of mtDNA to the nuclear genome, the decreased substitution rate can affect phylogenetic results and appear as shorter numt branch lengths. In contrast, Lopez et al. (1997) reported that some numts might be evolving at the same rate or faster than their mtDNA paralogs.

Because numt sequences and mtDNA sequences may not have the same evolutionary history, the position of a numt within a lineage can reveal when the transfer to the nuclear genome took place. If a numt insertion predated a speciation event, the nuclear sequence might appear basal to older lineages. Conversely, if the insertion event was recent, the numt might not have substantially diverged from its mtDNA counterpart. In any event, diagnosing numts within a phylogeny requires some knowledge of evolutionary relationships, but because many phylogenetic studies are conducted to establish relationships without previous knowledge, detecting numts via phylogenetic approaches can be challenging.

Of course, the most robust phylogenetic inferences rely upon analyses of multiple gene sequences, as single-gene trees are not equivalent to species trees (Avise 2000; Pamilo and Nei 1988; Tajima 1983). Thus, it is generally good practice to use multiple genes in attempts to recover phylogenetic relationships (Maddison and Knowles 2006). In so doing, however, one may encounter incongruities (deQueiroz et al. 1995) that are due to different modes of evolution among genes (or genomes) sampled. For example, if multiple mtDNA genes are utilized in an analysis, a discordance involving a single gene (amplicon) could indicate the presence of a numt in the data set. Alternatively, if both mtDNA and nuclear genes are utilized, putative mtDNA genes that exhibit phylogenetic signatures similar to nuclear sequences (e.g., cluster within nuclear clades) should be closely inspected, as this may be a clear indication that a given sequence is nuclear in origin.

Although the preventative measures described above can be effective in identifying numts, none are guaranteed. Herein, we illustrate the use of some methods outlined above in isolating a cytochrome-*b* numt and its corresponding mtDNA sequence in the North American prairie vole (*Microtus ochrogaster*). Nuclear copies of the mitochondrial cytochrome-*b* gene have previously been described in voles and other arvicoline rodents (DeWoody et al. 1999; Jaarola and Searle 2004; Jaarola et al. 2004; Triant and DeWoody 2007). We use comparative sequence analyses, mRNA expression assays, and phylogenetic analysis to highlight numt detection methods and offer cautionary suggestions.

Additionally, we further investigate numt representation within mammals using mammalian nuclear genome sequences available within the GenBank database. We use mtDNA protein-coding genes, ribosomal RNA genes, and control region sequences to examine whether or not certain regions of the mtDNA are more or less prone to translocation as numts.

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Materials and Methods

Isolation of mitochondrial and nuclear sequences.—Genomic DNA was extracted from cardiac and skeletal muscle tissue of a local specimen of *M. ochrogaster* with a standard proteinase K/phenol–chloroform protocol (Sambrook and Russell 2001). To mimic a typical mammalian evolutionary study, we amplified the mitochondrial cytochrome-*b* gene using the universal primers L14724/H15915 (Irwin et al. 1991). In parallel, we isolated a nuclear copy of the cytochrome-*b* gene using the numt-specific primers $PcytbF_2$, PcytbR, and $PcytbR_2$ (Triant and DeWoody 2007) in 2 separate reactions: $PcytbF_2$/PcytbR and $PcytbF_2$/$PcytbR_2$. These 2 sets of primers were originally designed to isolate a cytochrome-*b* numt sequence in *Microtus*. PCRs for both mitochondrial (cytochrome-*b*) and nuclear (numt) amplifications were performed in a final volume of 25 μl and included 1X ThermoPol Buffer (New England BioLabs, Ipswich, Massachusetts), 2 mM $MgSO_4$ 0.2 mM deoxynucleoside triphosphates, 0.25 μM each primer, 1.5 U *Taq* DNA polymerase (New England Biolabs), and 0.015 U *Pfu* DNA polymerase (Stratagene, La Jolla, California) to reduce polymerase infidelity (Cline et al. 1996). The thermal profile consisted of an initial denaturation at 94°C for 2 min; 32 cycles of 94°C for 1 min, 50°C for 30 s, and 72°C for 1 min; and a final elongation step for 4 min at 72°C. PCR products were cleaned with sodium acetate–ethanol precipitation and sequenced in both directions with the amplification primers and 2 internal sequencing primers (mitochondrial cytochrome-*b* gene: M.och_Cytb_Int1: 5′-TCACACGATTCTTCGCCT-3′, M.och_Cytb_Int2: 5′-GGAATAGTAGATGGACTA-3′; numt: PcytbSeq1 and PcytbSeq2—Triant and DeWoody 2007) using BigDye v.3.1 (Applied Biosystems, Foster City, California) following the manufacturer's protocol modified to one-eighth reactions. This study was conducted according to the guidelines of the American Society of Mammalogists (Animal Care and Use Committee 1998).

Comparative sequence alignment.—Putative mitochondrial and nuclear amplicons were aligned using Sequencher 4.1 (GeneCodes, Ann Arbor, Michigan). The sequence was considered mitochondrial in origin if, using the mammalian mtDNA genetic code, it possessed an open reading frame terminated by a stop codon. Alternatively, a sequence was considered nuclear in origin if it possessed premature stop codons or frame-shift mutations that disrupted the reading frame. We then used the mitochondrial cytochrome-*b* sequence of *M. rossiaemeridionalis*, the sibling vole (GenBank accession DQ015676), and aligned it with each of the sequences from *M. ochrogaster* to compare mitochondrial and nuclear substitution rates.

Phylogenetic analyses.—We performed phylogenetic analyses of *Microtus* mitochondrial cytochrome-*b* data sets that included both mtDNA and numt sequences from *M. ochrogaster*. Included in the analysis were 12 North American *Microtus* species and 3 Asian species. Maximum-likelihood trees were generated with PAUP 4.0b10* (Swofford 2003) under the GTR+I+G model as determined by Modeltest 3.7 (Posada and Crandall 1998) under the hLRT and Akaike information criteria. We used heuristic searches with 100 bootstrap replicates. *Myodes* (formerly *Clethrionomys*) *rutilus* and *M. glareolus* were used as outgroups because *Myodes* is the putative sister taxon of *Microtus* (Conroy and Cook 2000; Jaarola et al. 2004).

Gene expression.—Total RNA was isolated from fresh skeletal and cardiac muscle tissue of the same individual of *M. ochrogaster* described above. We used TRIzol (Invitrogen, Carlsbad, California) for RNA isolation and, to avoid numt contamination, removed trace amounts of genomic DNA from RNA products using DNase (Deoxyribonuclease I). SuperScript III First-Strand Synthesis System (Invitrogen) was used during reverse-transcriptase PCR to synthesize cDNA using the oligo $(dT)_{20}$. Protocols were followed as per manufacturers' suggestions. The mitochondrial cytochrome-*b* primers of Irwin et al. (1991) are located within mitochondrial tRNA sequences. Thus, we designed novel primers within the mitochondrial cytochrome-*b* gene to amplify its transcripts from cDNA: M.och_RNA_F (forward primer) 5′-ATGACAATCATCCGAAAA-3′ and M.och_RNA_R (reverse primer) 5′-GGATGTTGTTTTCGATTATA-3′. PCR and thermal profile were the same as those used for genomic DNA. To test for possible (but unexpected) numt expression, we also attempted to isolate the numt sequence from cDNA using the same numt primer pair combinations and conditions described above. PCR products were bidirectionally sequenced with amplification primers and internal sequencing primers (M.och_Cytb_Int1/ M.och_Cytb_Int2) and comparative sequence analyses were performed with sequences generated from genomic DNA. Sequences generated in this study have been deposited into the GenBank database under the accession numbers DQ432006–DQ432008.

Numt surveys.—We conducted NCBI-BLASTN searches (Altschul et al. 1997) on the sequenced mammalian genomes available for BLAST searching within the GenBank database as of November 2006 ($n = 11$), which included *Bos taurus*, *Canis familiaris*, *Felis catus*, *Homo sapiens*, *Macaca mulatta*, *Mus musculus*, *Oryctolagus cuniculus*, *Ovis aries*, *Pan troglodytes*, *Rattus norvegicus*, and *Sus scrofa*. We separately BLASTed each of their mtDNA protein-coding gene, ribosomal RNA gene, and control region (D-loop) sequences against their nuclear genomes, discounting matches with E values greater than 10×10^{-4}. The genomes for *Felis*, *Oryctolagus*, *Ovis*, and *Sus* were draft genomes and our searches revealed few numt sequences. In light of the extensive numt integration that has been described for felids (Cracraft et al. 1998; Kim et al. 2006; Lopez et al. 1994), we attributed this paucity of numts to incomplete nuclear genome sequences. Venkatesh et al. (2006) demonstrated that misleading results and spurious matches can be generated when searching for numts in draft genomes; therefore, we removed these 4 taxa from further analyses. For each remaining taxon, we estimated the number and total length of numts per mtDNA genome region and the proportion of all numt sequences represented by each region.

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

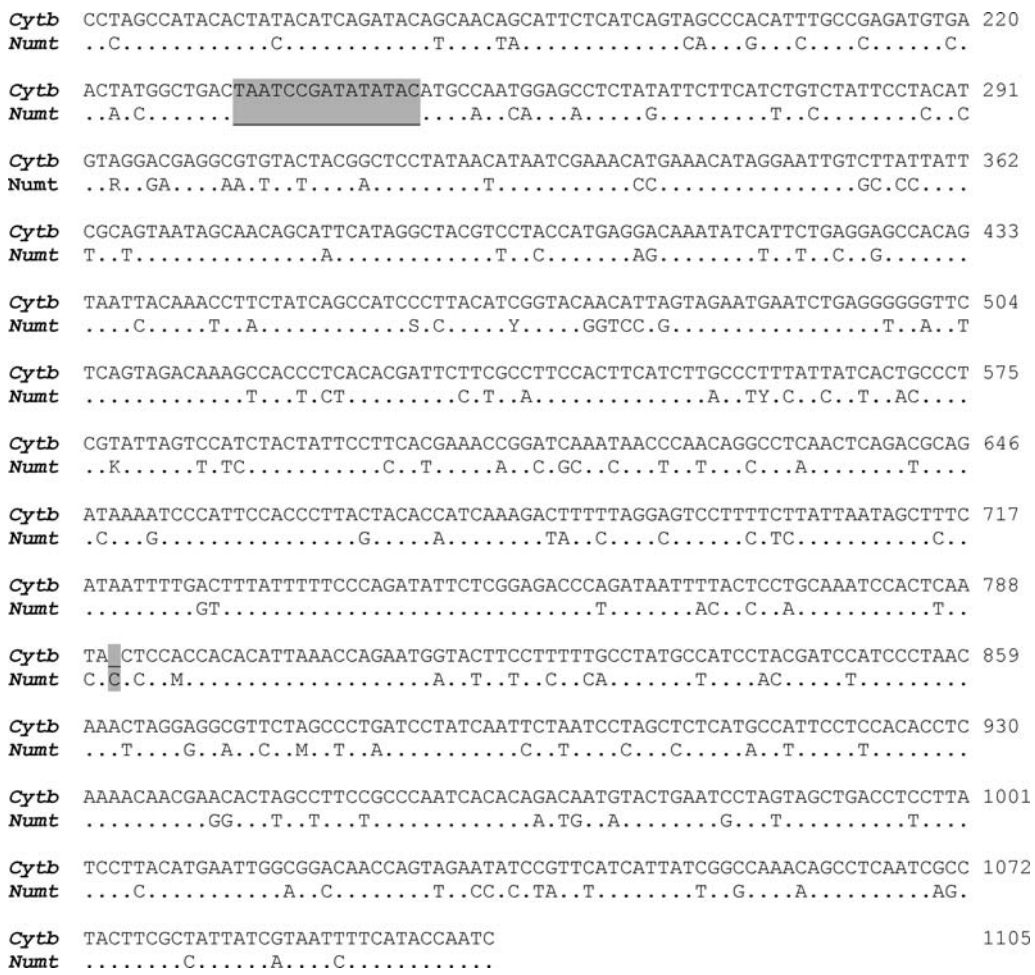


FIG. 1.—Sequence alignment of cytochrome-*b* nuclear mitochondrial translocation (numt pseudogene or numt) from *Microtus ochrogaster* and corresponding portion of the mitochondrial cytochrome-*b* gene in the same individual. Insertions–deletions (indels) are highlighted in gray and are represented by underscores. Numbers represent positions within cytochrome-*b* sequence corresponding to numt sequence. Mitochondiral DNA (mtDNA) cytochrome-*b* sequence begins at codon position 3 within the mtDNA reading frame.

RESULTS

We used the universal cytochrome-*b* primers L14724/H15915 (Irwin et al. 1991) to amplify and sequence the cytochrome-*b* gene from genomic DNA of *M. ochrogaster* and subsequently translated it with the mammalian mitochondrial genetic code to confirm that it was mitochondrial in origin. The 1,143-base pair (bp) sequence possessed an open reading frame, an initiation codon, and a terminal stop codon. The chromatograms were clean with no noticeable secondary peaks.

We also used numt-specific primers to amplify and sequence a 941-bp putative cytochrome-*b* pseudogene that differed from the mitochondrial sequence at 17.5% of its sites (20% at 1st codon positions, 13% at 2nd codon positions, and 67% at 3rd codon positions) and had a transition : transversion ratio of 1.8:1.0. Both pseudogene primer pairs isolated the same numt sequence. The numt pseudogene contained frame-shift mutations and at least 10 stop codons in each of 3 possible reading frames as translated with both the universal and mammalian mitochondrial genetic codes. Indels included a 15-base deletion and a single base insertion (Fig. 1). Substitution rates calculated from the pairwise comparisons of *M. ochrogaster* and *M. rossiaemeridionalis* are listed in Table 2. Chi-square analysis revealed that they were significantly different ($\chi^2 = 10.56$, $d.f. = 2$, $P = 0.005$).

Phylogenetic analyses.—The placement of the sequences from *M. ochrogaster* within the maximum-likelihood trees depended upon whether the sequence was mitochondrial or nuclear in origin (Fig. 2). The mitochondrial sequence was

TABLE 2.—Percent differences, number of inserted–deleted base pairs (indels), and transition : transversion ratio (Ts:Tv) between mitochondrial cytochrome-*b* genes from *Microtus ochrogaster* and *M. rossiaemeridionalis* (*Cytb*/*Cytb*), and cytochrome-*b* nuclear mitochondrial translocation (numt pseudogene or numt) from *M. ochrogaster* and mitochondrial cytochrome-*b* gene from *M. rossiaemeridionalis* (numt/*Cytb*).

		Codon position				
	Total	1st	2nd	3rd	Ts:Tv	Indels
Cytb/*Cytb*	12.6	17.4	2.1	80.6	2.6	0
Numt/*Cytb*	18.4	23.7	12.7	63.6	1.6	16

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

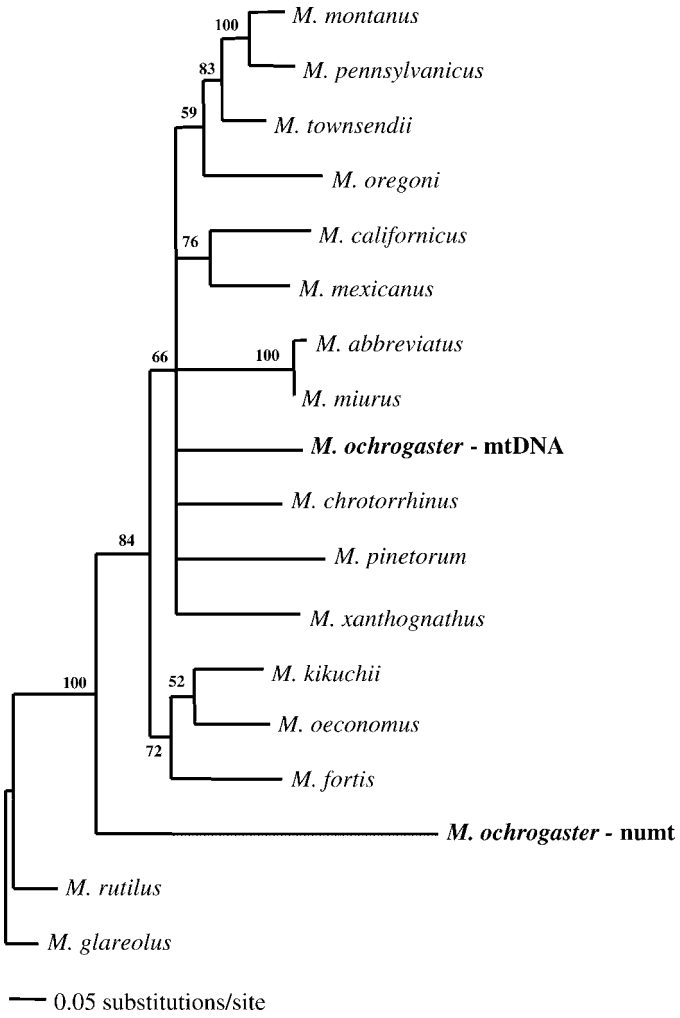


Fig. 2.—Cytochrome-*b* maximum-likelihood tree for *Microtus* with mitochondrial DNA (mtDNA) and nuclear mitochondrial translocation (numt pseudogene or numt) from *M. ochrogaster*. Numbers at nodes indicate bootstrap support. *Myodes* (=*Clethrionomys*) *rutilus* and *M. glareolus* were used as outgroups. Sequences from *M. ochrogaster* were from this study. mtDNA cytochrome-*b* sequences were downloaded from GenBank: North American species—*Microtus abbreviatus*, AF163890; *M. californicus*, AF163891; *M. chrotorrhinus*, AF163893; *M. mexicanus*, AF163897; *M. miurus*, AF163899; *M. montanus*, AF119280; *M. oregoni*, AF163903; *M. pennsylvanicus*, AF119279, *M. pinetorum*, AF163904; *M. townsendii*, AF163906; *M. xanthognathus*, AF163907; Asian species—*M. fortis*, AF163894; *M. oeconomus*, AF163902; *M. kikuchii*, AF163896; *M. glareolus*, AY309421; *M.rutilus*, AY309424.

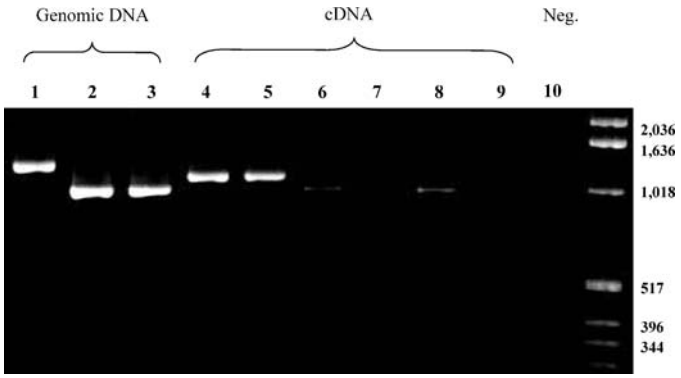


Fig. 3.—Agarose gel image visualizing products isolated from genomic DNA and cDNA. Lane 1: mitochondrial DNA (mtDNA) cytochrome-*b* gene amplified from genomic DNA; lane 2: nuclear mitochondrial translocation (numt pseudogene or numt) amplicon from genomic DNA with primers $PcytbF_2$/PcytbR; lane 3: numt amplicon from genomic DNA with primers $PcytbF_2$/$PcytbR_2$; lane 4: mtDNA cytochrome-*b* transcripts amplified via cDNA derived from cardiac tissue; lane 5: mtDNA cytochrome-*b* transcripts amplified via cDNA derived from skeletal muscle tissue; lane 6: numt transcripts amplified via cDNA derived from cardiac tissue with primers $PcytbF_2$/PcytbR; lane 7: numt transcript amplification via cDNA derived from skeletal muscle tissue with primers $PcytbF_2$/PcytbR (reaction failed); lane 8: numt transcripts amplified via cDNA from cardiac tissue with primers $PcytbF_2$/$PcytbR_2$; lane 9: numt transcript amplification via cDNA from skeletal muscle tissue with primers $PcytbF_2$/PcytbR2 (reaction failed); lane 10: negative control reaction. Numbers to the right of the size standard indicate band length (base pairs).

embedded within the *Microtus* clade among the other North American species but the placement of the nuclear sequence was at the base of the clade.

Gene expression.—We extracted mRNA from both cardiac and skeletal muscle tissue and then used reverse transcriptase to produce cDNA. We isolated 1,013-bp mitochondrial cytochrome-*b* transcripts from the cDNA obtained from both cardiac and skeletal muscle tissue. The transcripts possessed an open reading frame when translated with the mitochondrial genetic code and matched the cytochrome-*b* sequence isolated from genomic DNA. Surprisingly, we isolated putative numt transcripts from cDNA obtained from cardiac tissue, but were unable to isolate numt transcripts from skeletal muscle tissue using numt primers (Fig. 3).

Numt surveys.—The number of numts for the 7 taxa included in our search ranged from 50 to 1,030 and the summed length of all numts present ranged from 4,876 to 255,682 bp (Table 3). The mean length of individual numts ranged from 62-232 bp across taxa and the median values ranged from 49 to 160 bp (Table 3). The percentage of total numt sequences represented by each mtDNA region is shown in Fig. 4.

Discussion

We have empirically demonstrated various ways in which numt and mtDNA sequences from the same individual can be reconciled. Pairwise comparisons between the mitochondrial and nuclear sequences from *M. ochrogaster* revealed typical numt features such as stop codons, indels, and frame-shift mutations in the nuclear pseudogene. In comparisons with an mtDNA sequence from a related species, the numt from *M. ochrogaster* had less pronounced codon-position bias at the most selectively constrained site (mtDNA 2nd codon position) and a lower transition : transversion ratio (Table 2). Transition saturation may influence transition : transversion ratios and, therefore, should be used in combination with other numt detection measures, especially in species that show evidence of considerable intraspecific divergence. Differences in overall

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

TABLE 3.—Nuclear mitochondrial translocations (numt pseudogenes or numts) in 7 completely sequenced mammalian genomes. "Total" refers to the total sum length (in base pairs [bp]) of each individual numt in the genome. The mean ($\bar{X}$) is the average size, median is the value in the middle of the numt distribution, and range lists the smallest and largest numt revealed in our searches. *Cytb*, cytochrome *b*; ATP, adenosine triphosphatase; NADH, reduced nicotinamide adenine dinucleotide; ATP6, ATP synthase subunit 6; ATP8, ATP synthase subunit 8; COI, cytochrome oxidase 1; COII, cytochrome oxidase 2; COIII, cytochrome oxidase 3; ND1, NADH dehydrogenase subunit 1; ND2, NADH dehydrogenase subunit 2; ND3, NADH dehydrogenase subunit 3; ND4, NADH dehydrogenase subunit 4; ND4L, NADH dehydrogenase subunit 4L; ND5, NADH dehydrogenase subunit 5; ND6, NADH dehydrogenaese subunit 6.

	Cytb	ND6	ND5	ND4	ND4L	ND3	COIII	ATP6	ATP8	COII	COI	ND2	ND1	16S rRNA	12S rRNA	D-loop	Total
Mus musculus																	
n	20	4	14	15	1	6	12	7	8	7	22	11	10	13	14	37	201
Total (bp)	3,382	651	3,610	2,201	297	818	2,763	2,351	693	2,123	5,839	1,876	1,494	1,947	1,424	5,791	37,260
$\bar{X}$ (bp)	169	163	258	147	297	136	230	336	87	303	254	171	149	150	102	157	194
Median (bp)	109	53	139	82	297	92	162	241	75	240	174	102	154	103	66	114	138
Range (bp)	36–1,143	27–519	53–1,080	30–876	0–297	50–345	36–784	34–681	39–204	76–684	53–1,545	41–511	44–283	42–435	30–227	58–454	
Rattus norvegicus																	
n	0	1	0	2	0	0	2	3	0	3	5	1	1	5	20	7	50
Total (bp)	0	69	0	174	0	0	185	208	0	308	408	49	101	452	1,863	1,059	4,876
$\bar{X}$ (bp)	0	69	0	87	0	0	93	69	0	103	82	49	101	90	93	151	62
Median (bp)	0	69	0	87	0	0	93	65	0	60	58	49	101	47	78	85	49
Range (bp)	—	—	—	75–99	—	0	74–111	44–99	—	42–203	44–121	—	—	25–273	26–235	43–307	
Canis familiaris																	
n	19	9	22	21	2	9	8	13	3	11	27	8	13	30	23	35	253
Total (bp)	5,282	1,388	4,489	4,768	627	1,576	2,610	2,144	268	1,896	7,549	2,775	3,481	4,655	4,265	3,041	50,814
$\bar{X}$ (bp)	278	154	204	227	209	175	326	165	67	172	280	396	268	155	185	87	209
Median (bp)	222	131	88	117	239	145	304	77	55	132	148	277	215	118	113	58	152
Range (bp)	62–535	63–327	32–1,158	46–1,220	91–297	31–347	44–781	31–524	53–110	63–503	40–1,287	36–908	58–626	28–395	32–642	28–271	
Bos taurus																	
n	15	15	39	23	8	15	8	9	9	12	23	19	15	61	20	43	334
Total (bp)	1,889	2,804	6,985	6,102	954	2,097	2,219	1,701	877	2,306	5,466	4,101	3,023	8,890	3,360	4,879	57,453
$\bar{X}$ (bp)	126	187	179	265	119	140	277	189	97	192	238	216	202	142	168	113	178
Median (bp)	100	163	122	130	87	125	172	159	67	140	192	113	134	102	107	94	125
Range (bp)	37–274	40–520	37–582	32–1,066	48–290	44–346	49–743	43–521	36–197	67–498	40–1,001	56–888	36–668	34–499	57–547	30–319	
Macaca mulatta																	
n	52	38	97	66	19	20	66	38	6	24	71	51	44	178	123	102	995
Total (bp)	9,310	5,937	32,797	21,934	2,523	3,396	17,334	11,073	745	5,011	23,994	11,767	9,759	48,229	34,377	17,496	255,682
$\bar{X}$ (bp)	179	156	338	332	133	170	263	291	124	209	338	231	222	271	279	172	232
Median (bp)	135	109	184	191	93	168	200	262	126	122	206	124	150	202	186	111	160
Range (bp)	35–627	43–521	33–1,812	41–1225	56–297	36–346	55–784	50–685	53–207	41–677	32–1,539	38–1,023	38–957	19–1,436	27–949	28–883	
Pan troglodytes																	
n	66	24	73	50	27	15	60	33	11	16	85	61	55	156	111	30	873
Total (bp)	13,581	2,568	12,513	7,012	3,095	1,865	14,729	6,096	1,312	3,837	24,929	9,419	8,108	33,438	24,888	5,033	172,423
$\bar{X}$ (bp)	206	112	171	140	114	124	245	185	119	240	293	154	147	214	224	168	179
Median (bp)	165	89	135	131	108	95	204	102	74	133	206	128	102	159	106	89	127
Range (bp)	29–668	38–256	34–619	40–311	38–293	37–339	36–777	41–681	41–207	58–684	38–1,259	36–580	38–791	36–1,561	36–952	35–788	
Homo sapiens																	
n	91	36	96	85	38	16	70	26	12	19	94	44	79	158	132	34	1,030
Total (bp)	23,402	5,942	21,478	16,400	4,350	1,545	19,421	6,665	1,375	4,247	29,804	7,814	11,821	36,801	27,938	3,641	222,644
$\bar{X}$ (bp)	257	165	224	193	114	97	227	256	115	224	317	178	150	233	212	107	192
Median (bp)	183	145	143	149	106	58	192	255	83	149	202	127	118	200	109	96	145
Range (bp)	35–1,135	34–524	29–1,812	39–1,376	53–297	34–346	39–778	52–681	50–207	65–684	41–1,543	38–1,044	43–729	27–1,532	36–954	34–235	

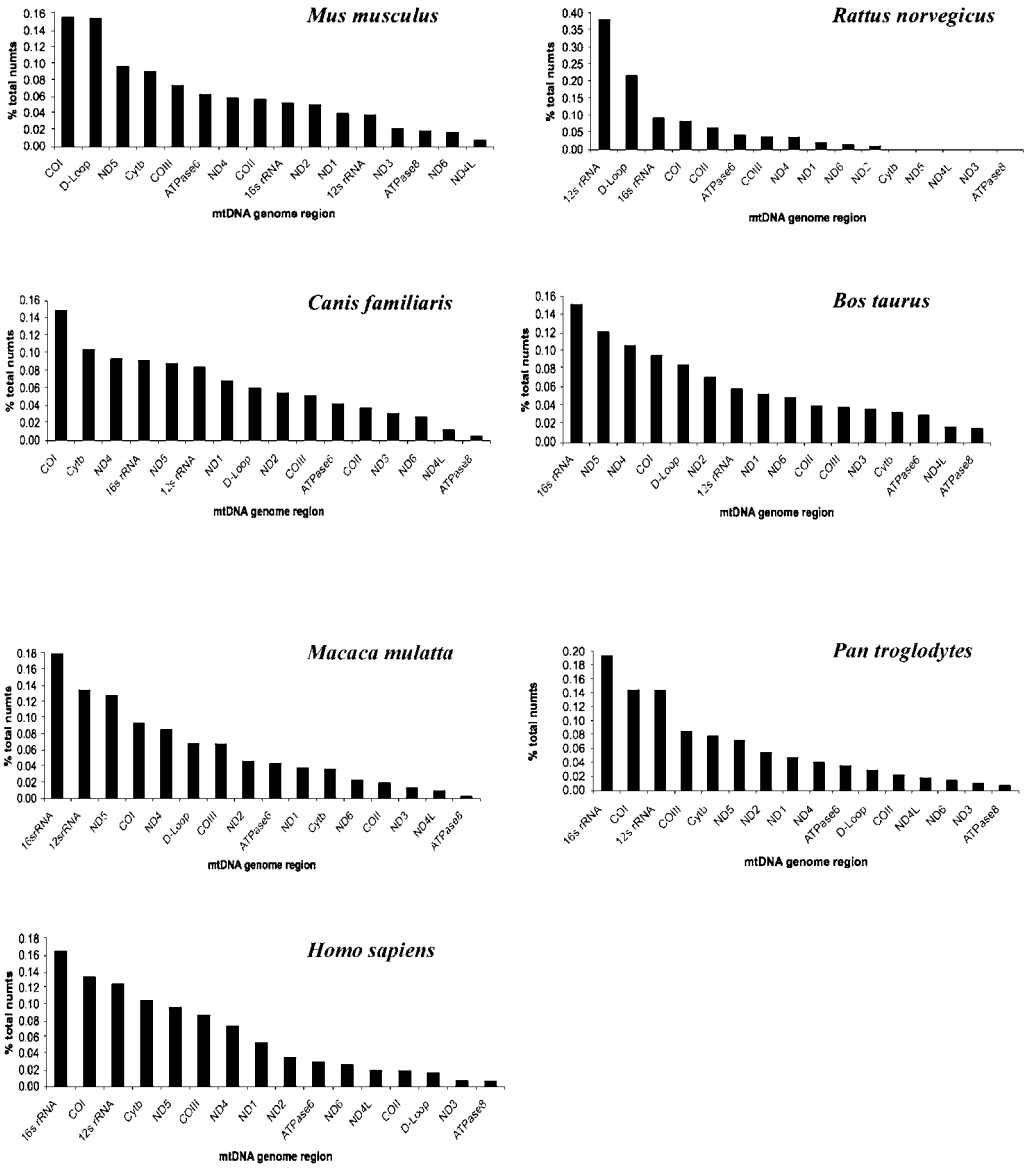


FIG. 4.—The percentage of total nuclear mitochondrial translocation (numt pseudogene or numt) sequences represented by each mitochondrial DNA region as identified via a GenBank database search.

substitution rates revealed that the numt sequence may be evolving more rapidly than the mtDNA sequence (Table 2). Without knowing the date of the nuclear insertion, it is difficult to gauge these discrepancies and compare absolute or relative rates of mitochondrial and nuclear evolution. Standardized substitution rates can be used to estimate the divergence between a mtDNA sequence and its numt pseudogene (e.g., Lopez et al. 1994), but the rapid rate of evolution within microtine mtDNA genomes (Triant and DeWoody 2006) renders such an estimate derived from a single sequence pair questionable. However, divergence estimates may be possible with larger data sets that have been calibrated with a local clock.

Mitochondrial phylogenies (cytochrome *b*) for the genus *Microtus* have been established and within those phylogenies, North American species form a monophyletic clade (Conroy and Cook 2000; Jaarola et al. 2004). However, relationships within the North American clade were poorly resolved, and our data mirror those earlier works. Within our tree, the mtDNA sequence of *M. ochrogaster* clustered with other North American species but the nuclear (numt) sequence of *M. ochrogaster* did not and instead was basal to the Asian vole lineages (Fig. 2). The genus *Microtus* is thought to have originated less than 2 million years ago (mya—Chaline 1999; Repenning 1990), which suggests that the translocation to the nucleus of this numt sequence occurred at least 2 mya and is likely present in all *Microtus* species. In this instance, the unexpected placement of the nuclear sequence of *M. ochrogaster* at the base of the tree was conspicuous because it conflicts with the studies cited above. However, numt identification would be more difficult in the absence of a consensus phylogeny and its position within a tree would be dependent upon its age. Short branch lengths leading to

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

numts could be signatures of the slower evolutionary rate found in nuclear genomes relative to mtDNA genomes resulting from the release of the selective constraints found within the mtDNA genome. However, long numt branch lengths have been reported in primate lineages (Schmitz et al. 2005; Zischler et al. 1998) and may be the result of DNA damage incurred during numt integration (Collura and Stewart 1995). Because mtDNA and nuclear sequences have different rates and modes of evolution, phylogenetic analyses can become complicated by the inclusion of both sequence types in a single analysis (Bensasson et al. 2001). The ultimate cause of long numt branch lengths is unclear, but our results coupled with the apparent higher rate of nucleotide substitution in our numt–mtDNA comparisons (Table 2) suggest that the *Microtus* numt described herein may be evolving more rapidly than its mtDNA counterpart.

Although numts are generally considered to be nonfunctional and therefore unexpressed, we provide strong evidence that the arvicoline cytochrome-*b* numt is expressed at low levels in cardiac tissue. The reverse-transcriptase PCR products from cardiac tissue provided insufficient template for direct sequencing, but the results were repeatable and can clearly be seen in Fig. 3. Despite this unexpected and potentially misleading result, we still endorse gene expression assays as a means for identifying numt sequences. PCR amplifications of mtDNA transcripts were consistently more robust than the numt transcripts, were found in more than 1 tissue type, and were easily sequenced. We cannot overlook the possibility of genomic DNA contamination in our mRNA extract despite our efforts to eliminate DNA from our samples using DNase. The elimination of genomic DNA is an optional step in cDNA library construction, but for numt avoidance we consider it a critical component. Further investigation of numt pseudogene transcription seems to be warranted (Blanchard and Schmidt 1996).

Using the mitochondrial and nuclear genome sequences for 7 mammalian taxa, we assessed the occurrence of mtDNA protein-coding genes, mtDNA ribosomal RNA genes, and the mtDNA control region within their nuclear genomes. Because we used individual mtDNA genes and not complete mtDNA genome sequences for our searches, the totals we present do not include various insertions that are associated with numts (e.g., mobile and repetitive elements—Mishmar et al. 2004). The primates in our sample had the most numts (total length in *Homo*: 222,644 bp; *Macaca*: 255,682; *Pan*: 172,423 bp), consistent with other studies that have found extensive mtDNA tranlocations in primates (Ricchetti et al. 2004; Schmitz et al. 2005). On the other end of the spectrum were rodents (*Mus*: 37,260 bp; *Rattus*: 4,876 bp; Table 3), and there was nearly an 8-fold difference between the 2 species sampled. Note that these numbers are not absolute, because the assembly status of genomes is constantly in flux and bioinformatic searches for numts in sequenced genomes can be hampered by misalignments (Venkatesh et al. 2006).

Our data do not reveal a propensity of 1 or more mtDNA genes or regions to translocate to the nuclear genome (Fig. 4). Consistent with other findings (Mourier et al. 2001; Woischnik and Moraes 2002), we found that the control region was relatively rare in the human nuclear genome but was common within rodent genomes (Fig. 4). The presence of the control region within the nuclear genome provides evidence for DNA-mediated numt insertions as opposed to RNA-mediated insertions as the control region has no RNA intermediate (Attardi and Schatz 1989).

Numts as molecular markers.—While cryptic numts can unwittingly confound molecular analyses, once identified they do have some practical utility. There is a growing interest in numts and what they can reveal about the evolutionary history of their host. Because numts generally have a slower rate of evolution than mtDNA, they may represent an ancestral form of functional mtDNA sequences (Perna and Kocher 1996). In this respect, they can be used to compare rates of mitochondrial and nuclear evolution, as phylogenetic markers or outgroups, and to estimate divergence dates (Bensasson et al. 2001; Lopez et al. 1997; Zischler et al. 1995). Schmitz et al. (2005) used numts and mtDNA to retrace 40 million years of primate history. Numt insertion sites also can be examined to assess whether numts preferentially integrate into certain regions of the genome (Mishmar et al. 2004).

Conclusions.—Some of the recommendations herein are more tractable than others. In particular, the isolation of cDNA template may be particularly burdensome for investigators using tissues that are not amenable to RNA extraction (e.g., hair or fecal samples). Researchers using noninvasive sampling or ancient DNA extraction techniques also might find these procedures troublesome because their templates are often limiting. On the other hand, these types of preventative measures are necessary only when initially characterizing primers and amplification profiles in a given species, usually at the beginning of a study. Once the target mtDNA sequence has been identified using some or all of the above procedures, this sequence can be aligned to any anomalous sequences to identify potential numts. If the numt and mtDNA sequences have sufficiently diverged from one another to allow for unique primer-binding sites, the numt and mtDNA sequence can be amplified independently to generate both mtDNA and nuclear data sets. Once in hand, the numt sequences can be used as neutral markers to generate comparative phylogenies, be compared to mtDNA results, or both, so long as investigators are aware that multiple (nonorthologous) numts may be amplified with the same primers.

Although we have presented various methods for detecting numts, we are not implying that PCR-based analyses of mtDNA are always suspect; indeed, hundreds of studies have successfully been performed without considering mitochondrial-derived nuclear pseudogenes. This study used universal primers and genomic DNA that was not enriched for mtDNA but we were able to generate clean mtDNA sequences without any apparent numt contamination. However, there have been cases where numts have been isolated with both universal and conserved primers (Lü et al. 2002; Mirol et al. 2000; Smith et al. 1992). Thus, mammalogists should at least be aware that numts have the potential to confound analyses and lead to potentially erroneous conclusions. To avoid numts, we suggest a number

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

of preventative approaches that can be employed both before and after data collection. The precautions necessary to ensure that a data set is truly mitochondrial in origin might seem onerous, but they are well worth the effort if they later save researchers from having to reanalyze or recollect data that is numt contaminated.

Acknowledgments

We thank D. Bos, J. Detwiler, D. Glista, D. Gopurenko, E. Lach, J. Rudnick, L. Theile, S. Turner, R. Williams, M. Zavodna, and 2 anonymous reviewers for reviewing earlier versions of this manuscript. This research was supported in part by Purdue University, the National Science Foundation, and the United States Department of Agriculture's National Research Initiative. This is publication ARP2006-17891 from the School of Agriculture at Purdue University.

Literature Cited

Altschul, S. F., et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389–3402.

Animal Care and Use Committee. 1998. Guidelines for the capture, handling, and care of mammals as approved by the American Society of Mammalogists. Journal of Mammalogy 79:1416–1431.

Arctander, P. 1995. Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. Proceedings of the Royal Society of London, B. Biological Sciences 262:13–19.

Attardi, G., and G. Schatz. 1989. Biogenesis of mitochondria. Annual Review of Cell Biology 4:289–333.

Avise, J. 1991. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. Annual Review of Genetics 25:45–69.

Avise, J. 2000. Phylogeography. The history and formation of species. Harvard University Press, Cambridge, Massachusetts.

Baker, R., J. A. DeWoody, A. J. Wright, and R. K. Chesser. 1999. On the utility of heteroplasmy in genotoxicity studies: an example from Chornobyl. Exotoxicology 8:301–309.

Baker, R., M. Hamilton, and D. A. Parish. 2003. Preparations of mammalian karyotypes under field conditions. Occasional Papers, The Museum, Texas Tech University 228:1–8.

Bensasson, D., M. W. Feldman, and D. A. Petrov. 2003. Rates of DNA duplication and mitochondrial DNA insertion in the human genome. Journal of Molecular Evolution 57:343–354.

Bensasson, D., D.-X. Zhang, D. L. Hartl, and G. M. Hewitt. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology and Evolution 16:314–321.

Birungi, J., and P. Arctander. 2001. Molecular systematics and phylogeny of the Reduncini (Artiodactyla: Bovidae) inferred from the analysis of mitochondrial cytochrome *b* gene sequences. Journal of Mammalian Evolution 8:125–147.

Blanchard, J. L., and G. W. Schmidt. 1996. Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. Molecular Biology and Evolution 13: 537–548.

Bradley, R. D., and R. J. Baker. 2001. A test of the genetic species concept: cytochrome-*b* sequences and mammals. Journal of Mammalogy 82:960–973.

Brown, W. M., M. George, Jr., and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences 76:1967–1971.

Chaline, J. 1999. Anatomy of the arvicoline radiation (Rodentia): palaeogeographical, palaeoecological history and evolutionary data. Annales Zoologici Fennici 36:239–267.

Cline, J., J. C. Braman, and H. H. Hogrefe. 1996. PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Research 24:3546–3551.

Collura, R. V., and C.-B. Stewart. 1995. Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. Nature 378:485–489.

Conroy, C. J., and J. A. Cook. 2000. Molecular systematics of a holarctic rodent (*Microtus*: Muridae). Journal of Mammalogy 6:221–245.

Cracraft, J., J. Feinstein, J. Vaughn, and K. Helm-Bychowski. 1998. Sorting out tigers (*Panthera tigris*): mitochondrial sequences, nuclear inserts, systematics, and conservation genetics. Animal Conservation 1:139–150.

deQueiroz, A., M. J. Donoghue, and J. Kim. 1995. Separate versus combined analysis of phylogenetic evidence. Annual Review of Ecology and Systematics 26:657–681.

DeWoody, J. A., R. K. Chesser, and R. J. Baker. 1999. A translocated cytochrome *b* pseudogene in voles (Rodentia: *Microtus*). Journal of Molecular Evolution 48:380–382.

du Buy, H. G., and F. L. Riley. 1967. Hybridization between the nuclear and kinetoplast DNA's of *Leishmania enriettii* and between nuclear and mitochondrial DNA's of mouse liver. Proceedings of the National Academy of Sciences 57:790–797.

Fernández-Silva, P., J. A. Enriquez, and J. Montoya. 2003. Replication and transcription of mammalian mitochondrial DNA. Experimental Physiology 88.1:41–56.

Greenwood, A. D., C. Capelli, G. Possnert, and S. Pääbo. 1999. Nuclear DNA sequences from late Pleistocene megafauna. Journal of Molecular Evolution 16:1466–1473.

Greenwood, A. D., and S. Pääbo. 1999. Nuclear insertion sequences of mitochondrial DNA predominate in hair but not in blood of elephants. Molecular Ecology 8:133–137.

Hazkani-Covo, E., R. Sorek, and D. Graur. 2003. Evolutionary dynamics of large *numts* in the human genome: rarity of independent insertions and abundance of post-insertion duplications. Journal of Molecular Evolution 56:169–174.

Hickson, R. E., C. Simon, A. Cooper, G. S. Spicer, J. Sullivan, and D. Penny. 1996. Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. Molecular Biology and Evolution 13:150–169.

Hirano, M., et al. 1997. Apparent mtDNA heteroplasmy in Alzheimer's disease patients and in normals due to PCR amplification of nucleus-embedded mtDNA pseudogenes. Proceedings of the National Academy of Sciences 94:14894–14899.

Irwin, D. M., T. D. Kocher, and A. C. Wilson. 1991. Evolution of cytochrome *b* gene of mammals. Journal of Molecular Evolution 32:128–144.

Jaarola, M., and J. B. Searle. 2004. A highly divergent mitochondrial DNA lineage of *Microtus agrestis* in southern Europe. Heredity 92:228–234.

Jaarola, M., et al. 2004. Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. Molecular Phylogenetics and Evolution 33:647–663.

Jensen-Seaman, M. I., E. E. Sarmiento, A. S. Deinard, and K. K. Kidd. 2004. Nuclear integrations of mitochondrial DNA in gorillas. American Journal of Primatology 63:139–147.

Kim, J.-H., et al. 2006. Evolutionary analysis of a large mtDNA translocation (*numt*) into the nuclear genome of the *Panthera* genus species. Gene 366:292–302.

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

LANSMAN, R. A., R. O. SHADE, J. F. SHAPIRA, AND J. C. AVISE. 1981. The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. III. Techniques and potential applications. Journal of Molecular Evolution 17:214–226.

LEISTER, D. 2005. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. Trends in Genetics 21:655–663.

LEMOS, B., F. CANAVEZ, AND M. A. M. MOREIRA. 1999. Mitochondrial DNA–like sequences in the nuclear genome of the opossum genus *Didelphis* (Marsupialia: Didelphidae). Journal of Heredity 90: 543–547.

LISTER, A. M., ET AL. 2005. The phylogenetic position of the 'giant deer' *Megaloceros giganteus*. Nature 438:850–853.

LOPEZ, J. V., S. CEVARIO, AND S. J. O'BRIEN. 1996. Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA tandem repeat (*numt*) in the nuclear genome. Genomics 33:229–246.

LOPEZ, J. V., M. CULVER, J. C. STEPHENS, W. E. JOHNSON, AND S. J. O'BRIEN. 1997. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Molecular Biology and Evolution 14:277–286.

LOPEZ, J. V., N. YUKI, R. MASUDA, W. MODI, AND S. J. O'BRIEN. 1994. *Numt,* a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. Journal of Molecular Evolution 39:174–190.

LÜ, X.-M., Y.-X. FU, AND Y.-P. ZHANG. 2002. Evolution of mitochondrial cytochrome *b* pseudogene in genus *Nycticebus*. Molecular Biology and Evolution 19:2337–2341.

MACPHEE, R. D. E., A. N. TIKHONOV, D. MOL, AND A. D. GREENWOOD. 2005. Late Quaternary loss of genetic diversity in muskox (*Ovibos*). BMC Evolutionary Biology 5:49.

MADDISON, W. P., AND L. L. KNOWLES. 2006. Inferring phylogeny despite incomplete lineage sorting. Systematic Biology 55:21–30.

MIROL, P. M., S. MASCHERETTI, AND J. B. SEARLE. 2000. Multiple nuclear pseudogenes of mitochondrial cytochrome *b* in *Ctenomys* (Caviomorpha, Rodentia) with either great similarity to or high divergence from the true mitochondrial sequence. Heredity 84: 538–547.

MISHMAR, D., E. RUIZ-PESINI, M. BRANDON, AND D. C. WALLACE. 2004. Mitochondrial DNA–like sequences in the nucleus (numts): insights into our African origins and the mechanism of foreign DNA integration. Human Mutation 23:125–133.

MOURIER, T., A. J. HANSEN, E. WILLERSLEV, AND P. ARCTANDER. 2001. The human genome project reveals a continuous transfer of large mitochondrial fragments to the nucleus. Molecular Biology and Evolution 18:1833–1837.

OCHMAN, H., A. S. GERBER, AND D. L. HARTL. 1988. Genetic applications of inverse polymerase chain reaction. Genetics 120: 621–623.

OJALA, D., J. MONTOYA, AND G. ATTARDI. 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature 290: 470–474.

OLSON, L. E., AND A. D. YODER. 2002. Using secondary structure to identify ribosomal numts: cautionary examples from the human genome. Molecular Biology and Evolution 19:93–100.

ORLANDO, L., J. A. LEONARD, A. THENOT, V. LAUDET, C. GUERIN, AND C. HÄNNI. 2003. Ancient DNA analysis reveals woolly rhino evolutionary relationships. Molecular Phylogenetics and Evolution 28:485–499.

PAMILO, P., AND M. NEI. 1988. Relationships between gene trees and species trees. Molecular Biology and Evolution 5:568–583.

PARR, R. L., ET AL. 2006. The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. BMC Genomics 7:185.

PEREIRA, S. L., AND A. J. BAKER. 2004. Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*): implications for molecular inference of population histories and phylogenetics. BMC Evolutionary Biology 4:17.

PERNA, N. T., AND T. D. KOCHER. 1996. Mitochondrial DNA: molecular fossils in the nucleus. Current Biology 6:128–129.

POSADA, D., AND K. A. CRANDALL. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.

REPENNING, C. A. 1990. Of mice and ice in the late Pliocene of North America. Arctic 43:314–323.

RICCHETTI, M., F. TEKAIA, AND B. DUJON. 2004. Continued colonization of the human genome by mitochondrial DNA. Public Library of Science Biology 2:1313–1324.

RICHLY, E., AND D. LEISTER. 2004. Numts in sequenced eukaryotic genomes. Molecular Biology and Evolution 21:1081–1084.

RUDKIN, G. T., AND B. D. STOLLAR. 1977. High resolution detection of DNA–RNA hybrids in situ by indirect immunofluorescence. Nature 265:472–473.

SAMBROOK, J., AND D. W. RUSSELL. 2001. Molecular Cloning: a laboratory manual. 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

SCHMITZ, J., O. PISKUREK, AND H. ZISCHELR. 2005. Forty million years of independent evolution: a mitochondrial gene and its corresponding nuclear pseudogene in primates. Journal of Molecular Evolution 61:1–11.

SMITH, M. F., W. K. THOMAS, AND J. L. PATTON. 1992. Mitochondrial DNA–like sequence in the nuclear genome of an akodontine rodent. Molecular Biology and Evolution 9:204–215.

SORENSON, M. D., AND T. W. QUINN. 1998. Numts: a challenge for avian systematics and population biology. Auk 115:214–221.

SWOFFORD, D. L. 2003. PAUP*: phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Inc., Publishers, Sunderland, Massachusetts.

TAJIMA, F. 1983. Evolutionary relationships of DNA sequences in finite populations. Genetics 105:437–460.

THALMANN, O., J. HEBLER, H. N. POINAR, S. PÄÄBO, AND L. VIGILANT. 2004. Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. Molecular Ecology 13:321–335.

TRASK, B. 1991. Fluorescent in situ hybridization: applications in cytogenetics and gene mapping. Trends in Genetics 7:149–1547.

TRIANT, D. A., AND J. A. DEWOODY. 2006. Accelerated molecular evolution in *Microtus* (Rodentia) as assessed via complete mitochondrial genome sequences. Genetica 128:95–108.

TRIANT, D. A., AND J. A. DEWOODY. 2007. Molecular analyses of mitochondrial pseudogenes within the nuclear genome of arvicoline rodents. Genetica.

TRIANT, D. A., AND J. A. DEWOODY. In press. Extensive numt transfer in a rapidly evolving rodent has been mediated by independent insertion events and by duplications. Gene.

TURNER, C., ET AL. 2003. Human genetic disease caused by de novo mitochondrial–nuclear DNA transfer. Human Genetics 112: 303–309.

VAN DE PEER, Y., P. D. RIJK, J. WUYTS, T. WINKELMANS, AND R. D. WACHTER. 2000. The European small subunit ribosomal RNA database. Nucleic Acids Research 28:175–176.

VARTANIAN, J.-P., AND S. WAIN-HOBSON. 2002. Analysis of a library of macaque nuclear mitochondrial sequences confirms macaque origin of divergent sequences from old oral polio vaccine. Proceedings of the National Academy of Sciences 99:7566–7569.

VENKATESH, B., N. DANDONA, AND S. BRENNER. 2006. Fugu genome does not contain mitochondrial pseudogenes. Genomics 87:307–310.

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use

WILLETT-BROZICK, J. E., S. A. SAVUL, L. E. RICHEY, AND B. E. BAYSAL. 2001. Germ line insertion of mtDNA at the breakpoint junction of a reciprocal constitutional translocation. Human Genetics 109:216–223.

WOISCHNIK, M., AND C. T. MORAES. 2002. Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Research 12:885–893.

ZHANG, D. X., AND G. M. HEWITT. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. Trends in Ecology and Evolution 11:247–251.

ZISCHLER, H., H. GEISERT, AND J. CASTRESANA. 1998. A hominoid-specific nuclear insertion of the mitochondrial D-loop: implications for reconstructing ancestral mitochondrial sequences. Molecular Biology and Evolution 15:463–469.

ZISCHLER, H., H. GEISERT, A. VON HAESELER, AND S. PÄÄBO. 1995. A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. Nature 378:489–492.

ZULLO, S., L. C. SIEU, J. L. SLIGHTOM, H. I. HADLER, AND J. M. EISENSTADT. 1991. Mitochondrial D-loop sequences are integrated in the rat nuclear genome. Journal of Molecular Biology 221:1223–1235.

Associate Editor was Carey W. Krajewski.

Downloaded From: https://bioone.org/journals/Journal-of-Mammalogy on 28 Mar 2024
Terms of Use: https://bioone.org/terms-of-use