J Mol Evol (1994) 39:191–199

Journal of Molecular Evolution

Rapid Evolution of a Heteroplasmic Repetitive Sequence in the Mitochondrial DNA Control Region of Carnivores

A. Rus Hoelzel,[1] Jose V. Lopez,[2] Gabriel A. Dover,[3] Stephen J. O'Brien[1]

[1] Laboratory of Viral Carcinogenesis, Bld. 560, National Cancer Institute, Frederick, MD 21702, USA
[2] Biological Carcinogenesis and Development Program, Program Resources, Inc./DynCorp, NCI-FCRDC, Frederick, MD 21702, USA
[3] Department of Genetics, Leicester University, Leicester, LE1 7RH, United Kingdom

Received: 16 November 1993 / Revised and accepted: 7 February 1994

Abstract. We describe a repetitive DNA region at the 3′ end of the mitochondrial DNA (mtDNA) control region and compare it in 21 carnivore species representing eight carnivore families. The sequence and organization of the repetitive motifs can differ extensively between arrays; however, all motifs appear to be derived from the core motif "ACGT." Sequence data and Southern blot analysis demonstrate extensive heteroplasmy. The general form of the array is similar between heteroplasmic variants within an individual and between individuals within a species (varying primarily in the length of the array, though two clones from the northern elephant seal are exceptional). Within certain families, notably ursids, the array structure is also similar between species. Similarity between species was not apparent in other carnivore families, such as the mustelids, suggesting rapid changes in the organization and sequence of some arrays. The pattern of change seen within and between species suggests that a dominant mechanism involved in the evolution of these arrays is DNA slippage. A comparative analysis shows that the motifs that are being reiterated or deleted vary within and between arrays, suggesting a varying rate of DNA turnover. We discuss the evolutionary implications of the observed patterns of variation and extreme levels of heteroplasmy.

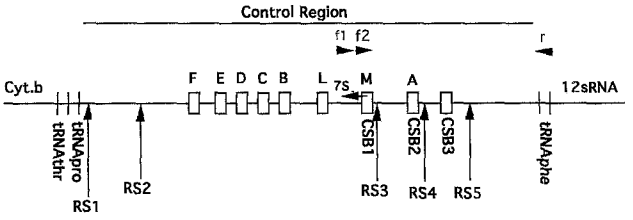
Correspondence to: A.R. Hoelzel

Key words: mtDNA — DNA slippage — Repetitive DNA — Carnivores

Introduction

The mtDNA control region is composed of a central conserved region (CCR) and variable A/T-rich flanking sequences (Fig. 1). The CCR has been implicated in the regulation of heavy-strand replication (Clayton 1982). Outside of the CCR, evidence for cryptic DNA simplicity (regions of short interspersed repetitive motifs, Tautz et al. 1986) has been described in primates and cetaceans (Hoelzel et al. 1991). The level of observed interspecific variation in the 5′ and part of the 3′ region is correlated to the level of simplicity, suggesting turnover by DNA slippage (Hoelzel et al. 1991). Evidence for DNA turnover by slippagelike mechanisms in mtDNA has also been reported by Madsen et al. (1993). The 3′ end of the control region contains several conserved sequence blocks (CSB1-3) which are located between the light-strand promoter and the origin of heavy-strand replication, and have been associated with the initiation of heavy-strand replication (Chang et al. 1985).

Repetitive sequences (RS) have been described at several positions in the control region in a number of species (RS1-5 in Fig. 1) including, among vertebrates, the white sturgeon (Buroker et al. 1990), Atlantic cod (Arnason and Rand 1992), evening bat (Wilkinson and Chapman 1991), rabbit (Biju-Duval et al. 1990; Mignotte et al. 1990), pig (Ghivizzani et al. 1993), harbor seal (Arnason and Johnsson 1992), elephant seals (Hoelzel et al. 1993a), cat (Lopez et al. in press), and



Species	Repeat length	Array length	Het.	Reference
RS1				
Sturgeon	82bp	82-328bp	M	Buroker et al. 1990
Cod	40bp	80-480bp	M	Arnason & Rand 1992
RS2				
Evening Bat	81bp	405-648bp	L	Wilkinson & Chapman 1991
Cat	80bp	240bp	-	Lopez et al. unpublished
RS3				
Rabbit	20bp	20-200bp	H	Mignotte et al 1990
Pig	10bp	140-290bp	H	Ghivizzani et al 1993
Harbor seal	10-22bp	390bp	-	Arnason & Johnsson 1992
Elephant seals	6-38bp	96-396bp	H	Hoelzel et al 1993
18 Carnivore sp.	6-30bp	150-450bp	H	this study
RS4				
Japanese monkeys	160bp	320bp	L	Hayasaka et al 1991
RS5				
Rabbit	153bp	153-612bp	M	Mignotte et al 1990

Fig. 1. Schematic of the vertebrate mitochondrial control region showing conserved blocks and the location of repetitive sequences in different species (RS1–5). *Lettered blocks* define the central conserved region (CCR), after Anderson et al. (1981). The conserved sequence blocks (CSB1–3) are after Walberg and Clayton (1981). The range of repeat motif and array lengths are given, and a relative scale indicating degree of heteroplasmy (Het.). The positions of PRC primers are given above the control region schematic, where *r* is the reverse primer and *f1* is the forward primer for all but the cats, for which *f2* was used.

Japanese monkeys (Hayasaka et al. 1991). Repeats at RS1, RS2, RS4, and RS5 involve motifs of approximately 40, 80, or 160 bp (multiples of 40) showing variable-number tandem repeat (VNTR) variation, and up to 50% of sampled individuals were heteroplasmic. The locations of the various repeats and a description of the array structures are given in Fig. 1. Arrays at repeat site 3 (RS3) are distinctive in that the level of heteroplasmy is highest, and the repetitive elements are comparatively small. The elephant seal repeats at RS3 can be further distinguished from the rabbit and pig repeats at RS3 by the variable and nested character of the elephant-seal repetitive elements (Hoelzel et al. 1993a; Fig. 2). In this study we show that the elephant-seal-array characteristics are shared, so far exclusively, by all 21 species of carnivores investigated to date, three published elsewhere, and 18 presented in this study. Differences in the organization and level of heterogeneity of these arrays in different species provide insight into the mechanisms that may have influenced their evolution.

Methods

Control region DNA between the central conserved region and the tRNA for phenylalanine was amplified from whole-cell DNA (extracted from various tissues) by PCR (Mullis and Faloona 1987) using the following primers: TCA TTT ACC AAC AT CATA (forward relative to the light strand) or GCA TCT GGT TCT TAC TTC AGG (forward for cats only) and ATT TTC AGT GTC TTG CTTT (reverse). Amplified DNA was cloned into bluescript sk+ phagemid or pT7Blue T-vector (Novagen) and transformed into TG1 or NovaBlue strains of *E. coli*. Cloned DNA was sequenced in both directions by a modified chain termination method (Tabor and Richardson 1987). One to four clones of the repetitive region from one individual from each of 18 species were sequenced. One clone from the giant panda was used as a template for PCR amplification, and the amplified DNA was cloned as described above. Sequences of these clones were compared to determine the extent to which polymerase slippage during PCR and recombination during cloning had altered the natural sequence. (Three clones were sequenced and all were identical—data not shown.) As a further test, the repeat was amplified from whole-cell DNA and from clones from four taxa (giant panda, black-footed ferret, dog, and cat), run on a 2% agarose gel, and stained with ethidium bromide. DNA amplified from clones migrated as a single band while DNA amplified from whole-cell preparations migrated as a smear of various-size fragments (data not shown). These tests do not rule out the possibility of slippage or recombination during amplification, but suggest that such events are rare.

Code	Seq.	Change	Code	Seq.	Change
1	ACGT		2	ACACGT	
c	G...	ts	e	C	ts
h	...C	ts	j	CC	tv,ts
m	..A.	ts	p	.T....	ts
n	.G..	tv	q	.T.T..	ts
o	.T..	ts	s	.T...C	ts

Code	Seq.	Change	Code	Seq.	Change
3	ACACACGT		4	ACACACACGT	
b	C	ts	i	C	ts
r	.T.....	ts			
t	...T....	ts			

Fig. 2. Relationship between repetitive elements. A *dot* indicates a conserved residue. The type of change is indicated by *ts* (transition) or *tv* (transversion).

Five micrograms of DNA from six black-footed ferrets, four giant pandas, eight dogs, and eight cats were digested with *AluI* (except cats which were digested with *HinfI*) and electrophoresed through 1% agarose. DNA was transferred by Southern blotting (Southern 1975) in 10× SSC onto nylon membranes. Blots were probed with amplified species-specific repeat sequence probes. Each blot was then stripped and probed with homologous 16s RNA mtDNA amplified by PRC using primers described in Hoelzel and Green (1992). Probe DNA was labeled with ^{32}P dCTP by random priming and hybridized to the membranes in a phosphate/SDS-based solution (after Westneat et al. 1988) overnight at 62°C and washed in 1× SSC, 0.1% SDS at the same temperature.

DNA secondary structure was investigated using the FOLD program in the sequence analysis package GCG (Zuker and Stiegler 1981). FOLD finds secondary structure based on published values of stacking and loop-destabilizing energies.

Results

All 18 carnivore species sequenced in this study had a repetitive array in the RS3 position (Fig. 1). In addition, DNA from another four mammalian species (short-tailed shrew, *Blarina brevicauda,* South American fur seal, *Arctocephalus australis,* raccoon, *Procyon lotor,* and European badger, *Meles meles*) was amplified by PCR across the repetitive region, and in each case a

fragment larger than the expected 1 kb (typical for mammalian species in the absence of repetitive regions) was observed, suggesting the presence of an array (data not shown). Each of the sequenced arrays have in common structural motifs that are related to each other in a hierarchical manner and are based on alternating purines and pyrimidines. The fundamental repeat element is ACGT. The three other basic repeat elements can be derived by repetition of "AC." We have encoded repeat elements based on the number of repetitions of the dinucleotide "AC" in the motif (1 = ACGT, 2 = $(AC)_2GT$, 3 = $(AC)_3GT$, and 4 = $(AC)_4GT$, see Fig. 2). The four basic repeat elements are also modified by base substitutions (Fig. 2), and most of these changes are transitions. The one-letter codes representing these modified motifs (Fig. 2) correspond to the code presented in Hoelzel et al. (1993). A comparison of one sequence from each species indicated that the smallest elements are most common. Including variants, element "1" comprises 35.7%, "2" comprises 25.9%, "3" comprises 21.9%, and "4" comprises 1.5% of the motifs from those 21 representative sequences.

At a higher level of organization, repeat elements combine to form "compound units" that are in turn reiterated (e.g., 2r in the Antarctic fur seal and 133 in the southern elephant seal, see Fig. 3). Most of the variation within arrays can be characterized as repetition at one of these three levels: subelement (repetition of AC or GT), element, or compound unit. Additional variation is generated by point mutations and the spreading of these variant elements through the array by DNA turnover (one or more of a variety of non-Mendelian mechanisms, such as DNA slippage or unequal crossing-over, see Dover 1982).

The level of heteroplasmy was assessed by digesting whole-cell DNA with a restriction enzyme that cuts close to, but not within the array, and probing with the repetitive region specific to each species. Figure 4 shows the hybridization profiles of several individuals from each of four species, representing four major families of carnivores (mustelids, ursids, canids, and felids). In all four species there is extensive length heteroplasmy (all individuals were heteroplasmic), and the mean and range of length variants differs between individuals. Interindividual variation ranges over 200 bp in the giant panda, and closer to 100 bp for each of the other three species. Intraindividual variation is as great as interindividual variation in all but the dog, where only interindividual variation is apparent from the Southern blot data. The same blots probed with 16s RNA mtDNA are shown for comparison with a nonheteroplasmic region (Fig. 4b). In addition, two to four clones were sequenced from one individual of each of six species to assess the level and range of heteroplasmy (Fig. 3). In this small sample, the overall organization of the array is conserved between clones within an individual, and length variation is based primarily on reiteration of elements or compound units. This is consistent with earlier findings for two species of elephant seals, for which a greater number of individuals and clones were sequenced (Hoelzel et al. 1993a). Hoelzel et al. (1993a) also demonstrated conservation of the overall organization of the array between individuals of a species (though the degree of pattern conservation was greater in the southern than in the northern elephant seal).

The arrays fall into two general groups—those based on the repetition of a single element (or variant of that element, see Figs. 2, 3), and more commonly, those based on the repetition of compound units. Single element (type 1) repeats either have rare variants, such as the cheetah (Fig. 3), or more common variants (e.g., the red panda and the skunk; Fig. 3). Compound unit repeats can be further classified into four types (types 2–5, though these are not discreet classes). Type 2 arrays appear to have been homogenized throughout most or all of the array (e.g., the Antarctic fur seal, polecat, ferret, brown bear, polar bear, and black bear; Fig. 3). Type 3 arrays have variants that are dispersed throughout the array (e.g., the dog, mink, sea otter, giant panda, and sloth bear; Fig. 3). In type 4 arrays, single elements within the compound unit are reiterated (e.g., southern elephant seal, cat, and tiger; Fig. 3). Type 5 arrays show alteration of the compound repeat element across the array, as well as single element slippage (e.g., northern elephant seal, harbor seal, lion, and ocelot; Fig. 3).

All of the arrays could form stable secondary structures, as demonstrated by FOLD. In each case, single compound unit motifs did not form stable structures, but short runs of each motif did (data not shown). The energy states ranged from −48.9 to −193.5 for arrays ranging in size from 160 bp (with an energy value of −48.9) to 372 bp (with an energy value of −105.1).

Discussion

Phylogeny

In most cases there is little correspondence between the DNA sequence or the pattern of repeats, and the temporal relationships indicated by published phylogenies (e.g., Wayne et al. 1989; Novacek 1992). For example, there are three species that are closely related by various criteria—the black-footed ferret, the Siberian polecat, and the mink (Eisenberg 1981; Wayne et al. 1989). However, the black-footed ferret and the polecat have repeats based on the same compound unit (2h), while the mink has a distinct compound unit sequence (3321). The compound units of the ferret/polecat and the mink are a minimum of four steps apart (2h ↔ 21 ↔ 221 ↔ 321 ↔ 3321, see Fig. 3). At the same time, the dog repeat unit (21) is closer to the ferret or polecat than the mink is to either (ferret, polecat, and dog repeats are all based on a modified "21" compound unit, while the mink repeat is based on "3321"). There are numerous similar

Southern elephant seal
1331331331331331331331331331333133333333333331333

Northern elephant seal
b3d42223d42223d4223d42223d4223d42e3d42e3d4222223d42

Harbor seal
123h123h123h123h123h12i12222i12223h12222h1222222222222222222233232222

Antarctic fur seal
c11-6 22r2r2r2r2r2r2r2r2r2r2r2r2r2r2r2r2r2

Mink
c3-1 33213321332133213321332132eee232132132133213321321133321

Skunk
c4-1 4t33ttttttttt333333333333

Sea otter
c5-1 j112j112j112j112j112j112j112j112j11212112112
c5-5 1j112j112j112j112j112j112j112j112j112j112j

Polecat
c1-1 k2h2h2h2h2h2h2h2h2h2h2h2h2h2h2h2h24444
c1-4 k2h2h2h2h2h2h2h2h2h2h2h2h2h2h2h2h2h2h2h444

Black-footed ferret
c2-2 k2h2h2h2h2h2h2h2h2h2h2h2h2h2hm223
c2-4 k2h2h2h2h2h2h2h2h2h2h2h2hm2334
c2-5 k2h2h2h2h2h2h2h2h2h2h2hm2334
c2-7 k2h2h2h2h2h2h2h2h2h2h2hm2334

Domestic dog
c8-3 212c212c2c2c212c21212c2c2c2c212c212c2c2c212c2c2c2c2c212c212c2121212c
c8-5 212c212c2c2c212c2c2c2c2c212c212c2c212c2c212c212c212c2121212c212
c8-6 212c212c212c21212c2c2c2c212c212c2c212c2c2c2c2121212c2121212c211
c8-7 212c212c2c212c21212c2c2c2c212c212c2c212c2c2c2c212c212c2121212c211

1=ACGT, 2=(AC)$_2$GT, 3=(AC)$_3$GT, 4=(AC)$_4$GT, b=(AC)$_3$GC, c=GCGT, d=GT, e=(AC)$_2$GC, h=ACGC, i=(AC)$_4$GC,
j=ACACCC, k=GTGC, m=ACAT, n=AGGT, o=ATGT, p=ATACGT, q=ATATGT, r=ATACACGT, s=ATACGC, t=ACATACGT

Fig. 3. Encoded repetitive array sequences from 1–4 clones from one animal of each species. *SES* (southern elephant seal) and *NES* (northern elephant seal) sequences were previously presented in Hoelzel et al. (1993), and the harbor seal pattern is interpreted from the sequence published by Arnason and Johnsson (1992). The key to the sequence codes is given in Fig. 3, with the following exceptions: k = GTGC and d = GT. Species presented (from top to bottom) are: southern elephant seal (*Mirounga leonina*), northern elephant seal (*Mirounga angustirostrus*), harbor seal (*Phoca vitulina*), Antarctic fur seal (*Arctocephalus gazella*), mink (*Mustela vison*), striped skunk (*Mephitis mephitis*), sea otter (*Enhydra lutra*), siberian polecat (*Mustela eversmanni*), black-footed ferret (*Mustela nigripes*), dog (*Canus familiarus*), giant panda (*Ailuropoda melanoleuca*), brown bear (*Ursus arctos*), polar bear (*Thalarctos maritimus*), sloth bear (*Melursus ursinus*), black bear (*Ursus americanus*), red panda (*Ailurus fulgens*), cat (*Felis catus*), lion (*Panthera leo*), tiger (*Panthera tigris*), ocelot (*Leopardus pardalis*), and cheetah (*Acinonyx jubatus*).

examples. This implies a discontinuous rate of DNA turnover and possibly a combination of overlapping turnover mechanisms affecting the evolution of these sequences. It also illustrates that a new variant can be homogenized rapidly through an array.

The ursids are exceptional in that an otherwise-rare variant (d = GT, 2.9% of elements in nonursid arrays) is common in all five species (38.5% of elements in ursid arrays), and the compound unit repeat is similar between species. This suggests that the pattern was established early in the radiation of this family, and has evolved slowly since. The discontinuous pattern of variants in two of the species (giant panda and sloth bear) could also be indicative of a slow rate of turnover. It is interesting that the giant panda array is much more like the four ursid arrays than like the red panda array, consistent with its placement among the ursids by other genetic criteria (O'Brien et al. 1985).

DNA Turnover

DNA slippage is thought to be an extremely common mechanism in the evolution of simple repetitive regions in both plant and animal genomes (Tautz et al. 1986; Levinson and Gutman 1987). For example, Moran (1993) searched the current DNA sequence databases and found short repetitive sequences in 13% of pig genes and 10% of chicken genes. The most commonly proposed mechanism involves the mispairing of a slipped strand in an array of repeated motifs, followed by polymerase and endonuclease-facilitated repair. The result is a gain or loss of repeat units. It is expected that slippage should be biased toward the duplication of shorter repeat units, since local melting and reannealing of the DNA duplex and the consequent formation of loops distort the normal configuration of the molecule, and smaller loops are likely to be less destabilizing

Fig. 3. Continued.

(Levinson and Gutman 1987). Observed rates of slippage in vitro are consistent with this expected trend (Wells et al. 1967). Some of the repeat motifs observed in the vertebrate mitochondrial control region are much larger than would normally be expected to arise by DNA slippage (up to 160 bp, see Fig. 1).

In contrast, the rate of interhelical DNA turnover events, such as unequal crossingover, are not expected to be biased based on motif length (Smith 1976; Dover 1993). It is not clear, however, to what extent recombination may occur in the mitochondrial genome. Rand and Harrison (1989) proposed a mitochondrial recombination model involving a double molecule intermediary as a possible mechanism for the generation of VNTR variation in cricket mtDNA. However, others have suggested that mtDNA recombination is uncommon based on empirical evidence from interspecific (rat/mouse) somatic-cell hybrid and intraspecific (rat) cybrid analyses (Hayashi et al. 1985). Extensive recombination in yeast (Horak et al. 1974) mtDNA occurs following organellar fusion, which is rare in mammalian cells. Further, when HVJ (hemagglutinating virus of Japan) was used for cell hybridization (which is thought to induce mitochondrial and cell fusion), mtDNA recombination was observed in rat/mouse (Hayashi et al. 1985) and human/rodent (Horak et al. 1974) hybrid cells. This suggests that organellar fusion may be a prerequisite for recombination, and that the low frequency of this event in mammalian cells may account for the limited evidence for recombination. However, a lack of organellar fusion would not rule out recombination between mitochondrial genomes within the mitochondria. Further, there could be a qualitative difference between the large-scale genomic events investigated in the aforementioned studies and local recombination events at RS3.

Buroker et al. (1990) suggest a mechanism for the amplification of an 82-bp repeat in the white sturgeon that involves the displacement strand (a third DNA strand in a structure known as the D-loop, Clayton 1982) during replication. This involves the formation of stable stem structures by the 82-bp repeat and the addition of a motif to the array if the structure forms in the D-loop strand or a deletion if it forms in the other two strands. The carnivore RS3 repeat is upstream of where the beginning of the D-loop is expected to be, so the illegitimate elongation model proposed by Buroker et al. (1990) may not apply. However, the carnivore arrays at

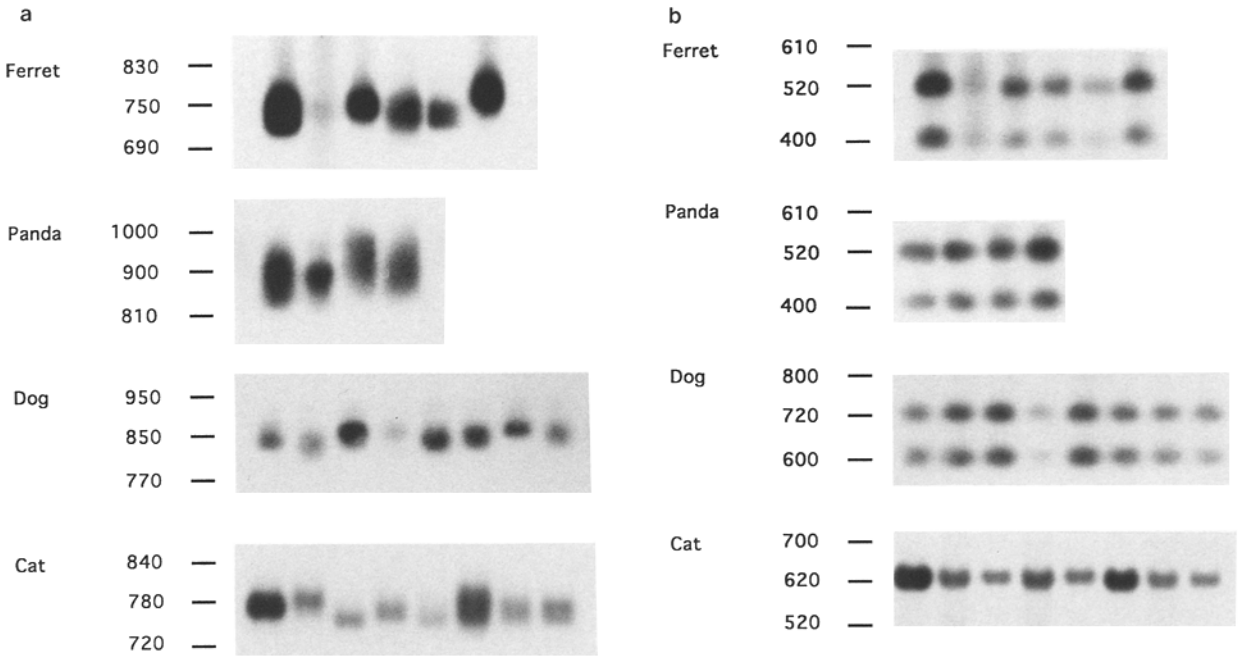


Fig. 4. **a** Southern blots showing level of heteroplasmy within and between individual arrays of black-footed ferrets (ferret), giant pandas (panda), dogs, and cats. Molecular weight is given in base pairs. **b** Same blots probed with 16s RNA mtDNA.

RS3 can form stable stem structures, which may stabilize loops and help account for the slippage of long motifs.

The high rate of DNA turnover seen for the carnivore RS3 repeat in somatic tissue is unusual. Of the hundreds of microsatellite and minisatellite DNA regions identified in the nuclear genomes of various vertebrate species, most show no variation within the somatic tissues of an individual. One exception is the triplet (GGC) repeat in the human FMR-1 region, thought to be responsible for the onset of fragile-X syndrome (Yu et al. 1991). Genomic DNA digests of individuals with the expanded repeat array at FMR-1 (the pathological allele condition) show a somatic distribution of numerous length variants within an individual (Reyniers et al. 1993), similar to that seen for the carnivore RS3 arrays. Another similarity between the carnivore RS3 and human FMR-1 is in the position of the array. The FMR-1 GGC repeat is located in the 5′ nontranslated region of the gene, which like the RS3 repeat is within a potentially regulatory region. RS3 repetitive arrays in pig mtDNA have been shown to be transcribed along with sequences from CSB1 and CSB2, and this RNA transcript is thought to serve as a primer in the initiation of heavy-strand DNA replication (Ghivizzani et al. 1993; Schmitt and Clayton 1993). DNA polymerase "stuttering" near the RNA-DNA transition may explain some of the turnover at RS3.

Array Organization

The repetitive elements in the arrays described in this study are organized hierarchically and are based on the motif "ACGT." AC is a common repeat motif in nuclear simple repetitive sequences. This may be explained in part by the fact that nuclear methylated C residues are subject to deamination, which leads to the conversion of a C to a T (Coulondre et al. 1978), and since most methylated C residues occur in CpG pairs (Razin and Riggs 1980), this leads to an increase in GT/AC motifs (Levinson and Gutman 1987). Methylation of C residues occurs in mammalian mtDNA, but at a much lower level than in nuclear DNA (Pollack et al. 1984; Mazin et al. 1988). Methylation in mouse mtDNA is distributed nonrandomly, and CpG pairs are underrepresented, suggesting that the pattern of methylation and deamination is similar to that of the nuclear genome (Pollak et al. 1984).

The ACGT motif is apparently restricted to the RS3 position in the mitochondrial control region (and it occurs in all arrays described at that site); however, a search through Genbank also reveals its presence in numerous short nuclear repetitive sequences (such as in the histone genes of *Urechis caupo,* Davis et al. 1992; and *Strongylocentrotus purpuratus* transposable elements, Cohen et al. 1985).

Reiteration of the AC motif generates the four basic repeat elements seen in this study (1, 2, 3, and 4 in Fig. 2). Transition type changes generate most of the variants of these basic elements (Fig. 2). The "1" and "2" elements are most common. Groups of elements combine to form repeated motifs of "compound units." The pattern of elements and compound units in the different arrays are suggestive of differential rates of turnover by DNA slippage, and possibly unequal crossingover for the larger compound units (e.g., the 38-bp repeat in the

northern elephant seal array, see Hoelzel et al. 1993a). For example, among the compound unit repeats, we describe four nondiscrete classifications. (See above.) The type 2 array shows homogenization throughout the array which implies a high rate of turnover, while the type 3 and 4 arrays show a decay of that pattern, suggesting either a higher rate of turnover for a smaller motif or an interaction between different turnover mechanisms. (See below.) Type 5 arrays show decay of the repetitive structure of the array at both the compound and single element levels, suggesting a low rate of homogenization.

The rate of turnover is not always proportional to the size of the repetitive motif. For example, the 20-bp compound unit, j112 in the sea otter arrays, is homogenized throughout most of the array, implying a high rate of turnover for that motif. No reiteration of the single-element motifs is observed. In other arrays, such as in the southern elephant seal (where the compound unit 133 is also 20 bp long), single elements are reiterated, disrupting the structure of the compound units. In this case the smaller motif is apparently being reiterated at a higher rate.

The most common types of compound unit arrays are those homogenized for a 10–20-bp motif (type 2) and those that showed the heterogeneous distribution of a variant (type 3). In the dog and giant panda arrays, the compound unit is 10 bp long and imperfectly homogenized through the array. In each of these two species a variant compound unit involving a single basepair mutation is distributed in patches throughout the array, suggesting local slippage of what was previously an alternating pattern (for example, 112122111211222 from 12121212). A high rate of turnover (high enough to homogenize the 20-bp pattern incorporating both the mutated and nonmutated compound units) may have changed over time allowing turnover at the 10-bp motif to cause the disruption of the alternating pattern. This is suggestive of an interaction between unequal crossingover (generating the initial pattern) and slippage (disrupting the pattern). Disruption of the pattern may have slowed the rate of turnover by interrupting the register for homologous pairing and led to further decay of the pattern. The four clones sequenced from the giant panda showed little similarity in the pattern of variants across the array (9–23 mutations required to convert one clone to another, Fig. 3). In contrast, the dog clones were more similar to each other (2–6 changes between clones, Fig. 3). Further, Southern blot analysis indicates greater length heteroplasmy within individuals in the giant panda than in the dog. Together these results suggest a lower rate of intraindividual turnover of repeat motifs in the dog, even though motifs are the same length in dog and panda arrays (10 bp).

Local expansion by slippage is evident in several other arrays, such as the "eee" pattern in the mink array. In a number of the arrays, local reiteration of single elements is biased toward the light-strand 3′ end of the array (cf. Hoelzel et al. 1993a). A similar pattern of polarity has been reported for minisatellite arrays by Jeffreys et al. (1991). They were able to infer the process leading to this pattern by looking directly at recombinant products of pairs of parental alleles. Most often the mechanism was an unusual gene conversion event in which extra repeats are added to the acceptor allele along with a patch of the donor's repeats, so there is a tendency to grow in local regions (A.J. Jeffreys et al. personal communication). To determine if this process is happening in mtDNA VNTRs it would be necessary to analyze recombinants (assuming they exist) from identified arrays.

Although the array organization seen in different species is not easily correlated to divergence time between species, both the repeat unit structure and array organization can differ considerably between species, while showing homology within a species (Hoelzel et al. 1993a, further analysis will be required to determine the extent of intraspecific homology in different carnivore species). In this study, Southern blot data using array-specific probes suggested intraspecific homology between individuals of four carnivore species. This could be the result of drift within matrilines, leading to the expectation of divergence of motif or array structure between ancient matrilines within the same species. Hoelzel et al. (1993a) sequenced arrays at RS3 for a number of individuals from each of two Southern elephant seal populations. The populations had diverged approximately 100,000 years ago (Hoelzel et al. 1993b), but the motif of the repeated unit and the array organization have not diverged between the populations, as a drift hypothesis would predict. Another possible explanation for within-species homogeneity in RS3 organization is that interhelical DNA turnover is occurring between the relatively few mitochondrial genomes contributed in the head of a sperm and the many mitochondrial genomes in the egg during sexual recombination, followed by further interactions between genomes within the egg. Such communication between matrilines (with the male as a bridge) as a consequence of intergenome turnover (e.g., by unequal crossingover or gene conversion) could account for the homogenization of array organization within a species.

Conclusions

In summary, repeats at RS3 are unusual compared with other control region repetitive sequences. They differ in the size, character, and sequence of the repeated motif and in the extent of variation and heteroplasmy. The location and general form of this repetitive array are shared by all carnivores investigated, and by the rabbit, pig, and shrew (P. Taberlet personal communication). The degree of variation seen across most carnivore arrays distinguishes them from the noncarnivore arrays at

this site. We suggest that much of the variation seen in these arrays can be explained by the mechanism of DNA slippage, operating at different rates and over different motif sizes both within and between arrays. However, some data also suggest interhelical mechanisms of DNA turnover. A striking feature of these arrays is their species specificity, and the apparent conservation of motifs throughout the species (though determining the extent of this will require further investigation). This may imply homogenization of arrays between lineages within a species via some, as yet unspecified, mechanism of intergenomic "crosstalk," in addition to genetic drift among lineages.

Acknowledgments. We thank Ian Boyd and the British Antarctic Survey, Claudio Campagna, Patricia Majluf, and Burney Le Boeuf for assistance in the acquisition of seal samples; Tom Thorne and the Wyoming Fish and Game Department for polecat and black-footed ferret samples; the American Type Culture Collection for mink samples; Rodney Mead for spotted skunk samples; Jim Estes and the U.S. Fish and Wildlife Service for sea otter samples; Mitchell Bush and the National Zoological Park for sloth bear, raccoon, tiger, and red panda samples; Lee Simmons, Lyndsay Phillips, and the Henry Doorly Zoo for polar bear, cheetah, and brown bear samples; Melody Roelke and Florida Game for black bear and ocelot samples; Terry Burke for European badger samples; Barbarah Penny and the NIH Animal Center for dog samples; and Pieter Kat and the National Museum of Kenya for lion samples. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

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

Arnason E, Rand DM (1992) Heteroplasmy of short tandem repeats in Atlantic cod (Gadus morhua). Genetics 132:211–220

Biju-Duval C, Ennafaa H, Dennebouy N, Monnerot M, Mignotte F, Soriguer RC, El Gaaied A, El Hili A, Mounolou J-C (1990) Mitochondrial DNA evolution in lagomorphs: origin of systematic heteroplasmy and organization of diversity in European rabbits. J Mol Evol 33:92–102

Buroker E, Brown JR, Gilbert TA, O'Hara PJ, Beckenbach AT, Thomas WK, Smith MJ (1990) Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics 124:157–163

Chang DD, Wong TW, Hixson JE, Clayton DA (1985) Regulatory sequences for mammalian mitochondrial transcription and replication. In Quagliariello E, Slater EC, Palmieri F, Saccone C, Kroon AM (eds) Achievements and perspectives of mitochondrial research. Elsevier, New York, pp 135–144

Clayton DA (1982) Replication of animal mitochondrial DNA. Cell 28:693–705

Cohen JB, Hoffman-Liebermann B, Kedes L (1985) Structure and unusual characteristics of a new family of transposable elements in the sea urchin *Strongylocentrotus purpuratus.* Mol Cell Biol 5: 2804–2813

Coulondre C, Miller JH, Farabaugh PJ, Gilbert W (1978) Molecular basis of base substitution hotspots in *Escherichia coli.* Nature 274:775–780

Davis FC, Shelton JC, Ingham LD (1992) Nucleotide sequence of the *Urechis laupo* core histone gene tandem repeat. DNA Seq 2:247–256

Dover GA (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111–117

Dover GA (1993) Evolution of repetitive DNA for advanced players. Curr Opin Genet Dev 3:

Eisenberg JF (1981) The mammalian radiations; an analysis of trends in evolution, adaptation and behavior. University of Chicago Press, Chicago, Illinois

Ghivizzani SC, Mackay SLD, Madsen CS, Laipis PJ, Hauswirth WW (1993) Transcribed heteroplasmic repeated sequences in the porcine mitochondrial DNA D-loop region. J Mol Evol 37:36–47

Hayasaka K, Ishida T, Horai S (1991) Heteroplasmy and polymorphism in Japanese monkeys: association with tandemly repeated sequences. Mol Biol Evol 8:399–415

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

Hoelzel AR, Green A (1992) Analysis of population-level variation by sequencing PCR-amplified DNA. In: Hoelzel AR (ed) Molecular genetic analysis of populations, a practical approach. Oxford University Press, Oxford, pp 159–224

Hoelzel AR, Hancock JM, Dover GA (1991) Evolution of the cetacean mitochondrial D-loop region. Mol Biol Evol 8:475–493

Hoelzel AR, Hancock JM, Dover GA (1993a) 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 AR, Halley J, O'Brien SJ, Campagna C, Arnbom T, Le Boeuf, Ralls K, Dover GA (1993b) Elephant seal genetic variation and the use of simulation models to investigate historical population bottlenecks. J Hered 84:443–449

Horak I, Coon HG, Dawid IB (1974) Interspecific recombination of mitochondrial DNA molecules in hybrid somatic cells. Proc Natl Acad Sci USA 71:1828–1831

Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG (1991) Minisatellite repeat coding as a digital approach to DNA typing. Nature 354:204–209

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

Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ (in press) *Numt,* a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol

Madsen CS, Ghivizzani SC, Hauswirth WW (1993) *In vivo* and *in vitro* evidence for slipped mispairing in mammalian mitochondria. Proc Natl Acad Sci USA 90:7671–7675

Mazin AL, Boiko LM, Ogarkova OA, Vanyushin BF (1988) Loss of CpG dinucleotides from DNA. VI. Methylation of mitochondrial and chloroplast genes. Mol Biol (Mosk) 22:1688–1696

Mignotte FM, Gueride M, Champagne A-M, Mounolou J-C (1990) Direct repeats in the non-coding region of rabbit mitochondrial DNA: involvement in the generation of intra and inter-individual heterogeneity. Eur J Biochem 194:561–571

Moran C (1993) Microsatellite repeats in pig and chicken genomes. J Hered 84:274–280

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

Novacek MJ (1992) Mammalian phylogeny: shaking the tree. Nature 356:121–125

O'Brien SJ, Nash WG, Wildt DE, Bush ME, Benveniste RE (1985) A molecular solution to the riddle of the giant panda's phylogeny. Nature 317:140–144

Pollack Y, Kasir J, Shemer R, Metzger S, Szyf M (1984) Methyla-

tion pattern of mouse mitochondrial DNA. Nucleic Acids Res 12: 4811–4824

Rand DM, Harrison RG (1989) Molecular population genetics of mtDNA size variation in crickets. Genetics 121:551–569

Razin A, Riggs AD (1980) DNA methylation and gene function. Science 210:604–610

Reyniers E, Vits L, De Boulle K, Van Roy B, Van Velzel D, de Graaf E, Verkerk AJMH, Jorens HZJ (1993) The full mutation in the FMR-1 gene of male fragile X patients is absent in their sperm. Nature Genet 4:143–146

Schmitt ME, Clayton DA (1993) Conserved features of yeast and mammalian mitochondrial DNA replication. Curr Opin Genet Dev 3:769–774

Smith GP (1976) Evolution of repeated DNA sequences by unequal crossing-over. Science 191:528–535

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

Tabor S, Richardson CC (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc Natl Acad Sci USA 84:4767–4771

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

Walberg MW, Clayton DA (1981) Sequence and properties of the human KB cell and the mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res 9:5411–5421

Wayne RK, Benveniste RE, Janczewski DN, O'Brien SJ (1989) Molecular and biochemical evolution of the carnivora. In: Gittleman JL (ed) Carnivore behavior, ecology and evolution. Cornell University Press, Ithaca, New York, pp 465–494

Wells RD, Jacobs TM, Narang SA, Khorana HG (1967) Studies on polynucleotides, LXIX. Synthetic deoxyribonucleotides as templates for the DNA polymerase of Escherichia coli: DNA-like polymers containing repeating trinucleotide sequences. J Mol Biol 27:237–263

Westneat DF, Noon WA, Reeve HK, Aquadro CF (1988) Improved hybridization conditions for DNA 'fingerprints' probed with M13. Nucleic Acids Res 16:161

Wilkinson GS, Chapman AM (1991) Length and sequence variation in evening bat D-loop mtDNA. Genetics 128:607–617

Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, Baker E, Holman K, Mulley JC, et al. (1991) Fragile X genotype characterized by an unstable region of DNA. Science 252:1179–1181

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