Length Variation, Heteroplasmy and Sequence Divergence in the Mitochondrial DNA of Four Species of Sturgeon (*Acipenser*)

James R. Brown,*,† Karen Beckenbach,‡ Andrew T. Beckenbach*,‡ and Michael J. Smith*,‡

**Evolutionary Biology Program, Canadian Institute for Advanced Research, † Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7 and ‡Institute of Molecular Biology and Biochemistry, Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6*

Manuscript received March 2, 1995
Accepted for publication October 23, 1995

ABSTRACT

The extent of mtDNA length variation and heteroplasmy as well as DNA sequences of the control region and two tRNA genes were determined for four North American sturgeon species: *Acipenser transmontanus, A. medirostris, A. fulvescens* and *A. oxyrhnychus*. Across the Continental Divide, a division in the occurrence of length variation and heteroplasmy was observed that was concordant with species biogeography as well as with phylogenies inferred from restriction fragment length polymorphisms (RFLP) of whole mtDNA and pairwise comparisons of unique sequences of the control region. In all species, mtDNA length variation was due to repeated arrays of 78–82-bp sequences each containing a D-loop strand synthesis termination associated sequence (TAS). Individual repeats showed greater sequence conservation within individuals and species rather than between species, which is suggestive of concerted evolution. Differences in the frequencies of multiple copy genomes and heteroplasmy among the four species may be ascribed to differences in the rates of recurrent mutation. A mechanism that may offset the high rate of mutation for increased copy number is suggested on the basis that an increase in the number of functional TAS motifs might reduce the frequency of successfully initiated H-strand replications.

THERE are two basic types of mitochondrial DNA (mtDNA) sequence polymorphism: nucleotide substitutions caused by point mutations and length variation that can result from varying copy numbers of an oligonucleotide sequence or larger scale genomic duplications. The propagation of any new mutation requires, if only briefly, a transition stage where the multiple forms of the mitochondrial genome coexist within a single individual—the so-called heteroplasmic state. Heteroplasmy for point mutations appear to be rare (HALE and SINGH 1986; MORITZ *et al.* 1987) although it must be at least a transient stage in the fixation of alternative sequences in different lineages. The time for the fixation of nucleotide substitutions appears to be rapid and, in cattle, it has been shown to take as little as a single generation (HAUSWIRTH and LAIPIS 1982; KOEHLER *et al.* 1991).

In contrast, heteroplasmy for length variation has been frequently observed (reviewed in RAND 1993). A common type of length variation in vertebrate mtDNA are tandemly repeated sequences in the origin of heavy strand replication and transcription, also known as the control or D-loop region. The latter name describes the resting state triplex structure where a short nascent D-loop strand binds to the light (L) strand thus displacing the heavy (H) strand (reviewed in CLAYTON 1982). Length variation and heteroplasmy in the control region has led to several proposals for the generation and maintenance of multiple sized mtDNA molecules including nuclear genome-like slipped-mispairing (EFSTRADIADIS *et al.* 1980), intra- and intermolecular recombination (RAND and HARRISON 1989) and competitive displacement between D-loop and H-strands (BUROKER *et al.* 1990).

North American sturgeon species of the genus Acipenser provide additional opportunities to infer mechanisms of mtDNA length variation and heteroplasmy. *A. transmontanus* (white sturgeon) were previously shown to have extensive heteroplasmy due to a varying copy number (1–6) of a perfect tandemly repeated 82-bp element in the D-loop region (BUROKER *et al.* 1990; BROWN *et al.* 1992a). It is widely held that the distinctive eastern and western assemblages of North American fish species resulted from the formation of the Continental Divide, some 10 mya (SMITH 1981). Of the sturgeon species examined here, *A. medirostris* (green sturgeon), like *A. transmontanus*, are found only in rivers and Pacific coastal waters to the west of the Rocky Mountains while *A. fulvescens* (lake sturgeon) and *A. oxyrhnychus* (Atlantic sturgeon) are limited to the Continental interior and Atlantic Coast, respectively. Although recently overfished and their habitats frequently disturbed by dams and pollution, there has been no recorded artificial transplantation of these stur-

Corresponding author: Michael J. Smith, Institute of Molecular Biology and Biochemistry, Department of Biological Sciences, Simon Fraser University, Burnaby, B.C., Canada V5A 1S6.
E-mail: msmith@darwin.mbb.sfu.ca

TABLE 1

Pairwise estimates of sequence divergence among four sturgeon species, genus Acipenser

Species	*A. transmontanus*	*A. medirostris*	*A. fulvescens*	*A. oxyrhynchus*
A. transmontanus	—	14.07 (19.70)	29.65 (16.69)	32.49 (23.89)
A. medirostis	4.14 ± 0.96	—	29.31 (15.72)	32.70 (16.95)
A. fulvescens	6.65 ± 1.29	8.27 ± 1.50	—	38.30 (18.12)
A. oxyrhynchus	7.82 ± 1.59	8.49 ± 1.69	9.07 ± 1.72	—

The lower diagonal presents percent sequence differences and standard deviations estimated from RFLP analysis of whole mtDNA using the maximum likelihood method of NEI and TAJIMA (1983). The upper diagonal shows pairwise estimates of the expected number of nucleotide substitutions per 100 sites between non-repeated portions of the control region using KIMURA's (1980) correction for multiple substitutions. In parenthesis, similar estimates are given between single copies of central repeats.

geon species. Thus their present day distributions probably reflect natural speciation events. The purpose of this study is to examine mtDNA nucleotide substitutions and length polymorphisms among these closely related but geographically isolated sturgeon species.

MATERIALS AND METHODS

Specimen collection and restriction enzyme analysis of mtDNA: Collection of *A. transmontanus* specimens and mtDNA analysis were described previously (BUROKER *et al.* 1990; BROWN *et al.* 1992a,b, 1993). *A. medirostris* specimens were caught in the lower Columbia River, WA. *A. oxyrhynchus* specimens were collected in the lower St. Lawrence River, Quebec. *A. fulvescens* specimens originated from two river systems; the Nelson River and Lake Winnipeg, Manitoba, and the upper St. Lawrence River at Lac St. Pierre, Quebec.

mtDNA for RFLP and PCR analyses were purified from liver tissue within 6–72 hr of collection using a modified protocol of Lansman *et al.* (1981), which omitted the sucrose gradient step but included an increase in CsCl-ethidium bromide sedimentation time to 60–72 hr. Typical mtDNA yields were 300–500 ng per gram of homogenized tissue. RFLP comparisons were based on digests with 12 restriction enzymes with specificity for six nucleotide (*Apa*I, *Bcl*I, *Bgl*II, *Eco*RI, *Hind*III, *Pvu*II and *Xba*I), multiple six nucleotide (*Acc*I, *Ava*I, *Hae*II and *Hin*dII) and multiple five nucleotide (*Ava*II) recognition sequences.

Southern blotting and hybridization: For all four sturgeon species, *Hae*II digests of mtDNA consistently produced a particular DNA fragment, which varied in length from 1.5 to 2.0 kb and, in some individuals, these occurred as multiple bands. To determine whether this fragment spanned the control region, *Hae*II digested mtDNA was transferred to nitrocellulose filters and probed with the mitochondrial control region cloned from *A. transmontanus* [for a description of the clone and radioactive labeling protocol see BROWN *et al.* (1992a)]. Hybridizations were done under moderately stringent conditions of 60° overnight followed by three washes in ×1.0 SSPE. Filters were left wet and wrapped in plastic wrap for autoradiography. The film (Kodak X-Omat AR) was preexposed to ensure a linear response for microdensitometry readings (LASKEY 1980). Several exposures (0.5–14 days) were made to ensure the best contrast between bands in heteroplasmic individuals.

Densitometry and analysis of autoradiographs were done as previously described (BROWN *et al.* 1992a). mtDNA prepared from an *A. transmontanus* individual heteroplasmic for five different mtDNA size classes was included as a size standard in all interspecific Southern blots. mtDNA size classes observed in *A. medirostris*, *A. fulvescens* and *A. oxyrhynchus* were scored relative to repeat copy numbers known to exist in *A. transmontanus*. The distributions of mtDNA size classes were compared between species using a chi-square analysis of 1000 Monte Carlo randomizations (ROFF and BENTZEN 1989). For this analysis, the sums of repeat proportions within individuals were rounded off to the nearest integer value.

Partitioning diversity: Estimates of genetic diversity were partitioned for variation within and between fish of different species using the statistical approach of BIRKY *et al.* (1989). For each fish, the probability of sampling two different genes (K_b) is equal to $1 - \Sigma x_i^2$ where x_i is the fractional representation of each size variant within each fish. For homoplasmic fish, $K_b = 0$. For each species, a mean value of K_b was calculated including both homoplasmic and heteroplasmic fish. The probability of sampling two genes which are different from anywhere in the population (K_c) is equal to $1 - \Sigma \bar{x}_i^2$, where $\bar{x}_i$ is the mean frequency of the ith size variant in the population. K_c obtained this way includes heteroplasmy (K_c* of BIRKY *et al.* 1989). G_{IP} is the proportion of gene diversity within a species attributable to between fish diversity.

PCR amplification: PCR primers were designed to amplify D-loop and tRNA genes from all four species with the exception of AfDL1PH (5′ AGTACCAACAGTCCGTGA 3′), which was specific to *A. fulvescens*. Control region primers are designated by the letters "L" and "H" (for light and heavy strand, respectively) and by a number corresponding to the position of the primer's 3′ base in the reference 1.6-kb fragment from *A. transmontanus* (BUROKER *et al.* 1990). Three of the D-loop region primers used, L185, L506 and H740, have been described earlier (BROWN *et al.* 1993). New primers are H308 (5′ CAGTTGTGAATCCCTACAG 3′) and H482 (5′ AGGAGTTCTACATGTCAGTG 3′) in the D-loop region; tPRO123 (5′ CACCCTTAACTCCCAAAGC 3′) in tRNA^{Pro} gene (GILBERT *et al.* 1988) and CyB1102 (5′ TGACCGGCTGACTAG 3′) in the gene for cytochrome B (BROWN *et al.* 1989).

PCR amplification, DNA fragment isolation and sequencing protocols have been previously described (BROWN *et al.* 1993). Occasionally, PCR amplified DNA fragments extracted from agarose gels were blunt-end ligated into the plasmid vector pUC19 (MESSING and VIEIRA 1982). Selected recombinants were then sequenced using either universal sequencing primers or internal primers to the cloned fragment. A minimum of three plasmid isolates of each target clone were sequenced.

DNA sequence analysis: Nucleotide substitutions per site (d) between pairs of mtDNA genotypes were estimated from RFLP data using the maximum likelihood method of NEI and TAJIMA (1983). A distance tree of percent sequence divergence (d) was constructed using the unweighted pair-group method with arithmetic mean (UPGMA) and standard errors about the nodes were calculated (NEI *et al.* 1985). Wagner parsimony analyses of RFLP data were bootstrap replicated 100 times using the program BOOT of the PHYLIP v3.5c package (FELSENSTEIN 1993).

DNA sequences were aligned by eye using the program ESEE v3.0 (CABOT and BECKENBACH 1989). Distance estimates and neighbor-joining trees were determined using MEGA v1.01 software (KUMAR *et al.* 1993) as well as the PHYLIP programs DNADIST, SEQBOOT and NEIGHBOR (FELSENSTEIN 1993). Pairwise sequence differences, omitting indels, were estimated using the two parameter model for nucleotide substitutions of KIMURA (1980). Confidence intervals for internal branching points in neighbor-joining trees were estimated by 500 bootstrap replications as implemented by MEGA.

Thermal stabilities of repeat secondary structures were determined for one or more repeat elements using the PCFOLD algorithm of ZUKER and STIEGLER (1981). A search for secondary structures that would base-pair to the limits of repeat elements was done by ascertaining the thermal stability of duplex structures that were off-set by one nucleotide per determination and forced to fold to the end of the fragment.

RESULTS

Restriction enzyme analysis: The 12 restriction enzymes detected between 40 (*A. oxyrhynchus*) and 59 (*A. fulvescens*) sites per mtDNA genome. Restriction enzyme maps and sequence alignments available upon request from J. R. BROWN [GenBank Accession Nos U30601 (Am5), U30728 (Am12), U32308 (Ao8), and U32309 (Af14)]. The size range, between 16.1 to 16.5 kb, of *A. medirostris, A. fulvescens* and *A. oxyrhynchus* mtDNA were similar to that reported for *A. transmontanus* (BROWN *et al.* 1992a,b). Sequence divergence estimates between four sturgeon species (Table 1) were at least fourfold greater than intraspecific differences (BROWN *et al.* 1992b). *A. transmontanus* and *A. medirostris,* found west of the Continental Divide, were the most closely related species pair (4.14% sequence difference). The two eastern species, *A. fulvescens* and *A. oxyrhynchus,* were highly diverged (6.65–8.49% difference) from the western species pair although *A. fulvescens* was marginally more closely related. The separation of species pairs was well supported by standard error and bootstrap statistical tests of the respective distance (UPGMA) and parsimony (Wagner) phylogenies (Figure 1).

RFLP data suggested that genetic diversity within each species is low. Out of 10 *A. oxyrhnychus,* nine individuals had identical restriction enzyme site maps while the remaining individual differed by only a single *Ava*II site (0.174% difference). Among 12 *A. fulvescens* individuals, two closely related genotypes (0.314% difference) were detected by *Hind*II and *Ava*II restriction site polymorphisms. Six and four *A. fulvescens* individuals from the Nelson River, Manitoba, and Lac St. Pierre, Quebec, respectively, shared one genotype while two individuals from the Nelson River had a second genotype.

mtDNA length variation: In Southern blots of *Hae*II digested mtDNA, the cloned control region of *A. transmontanus* (BUROKER *et al.* 1990) strongly hybridized to fragments varying in size from 1.5 to 2.0 kb in all four species, an indication that these variable length fragments did indeed span the D-loop region (Figure 2). Sequence analysis later revealed that one *HaeII* site was

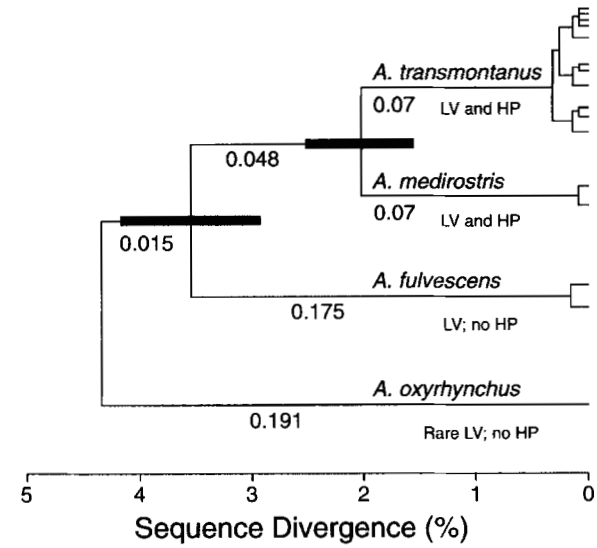

FIGURE 1.—Phylogenetic relationships and the occurrences of length variation (LV) and heteroplasmy (HP) among studied Acipenser species. The UPGMA tree was based on DNA sequence distances estimated from whole mtDNA RFLP data. Branching points are plotted as $1/2d$. Solid bars represent 1 standard error on either side of the branching point. Identical tree topologies were obtained using the maximum parsimony method as well as the neighbor-joining distance (2-parameter model; KIMURA 1980) method applied to unique control region sequences. Numbers in parentheses are branch lengths estimated from control region comparisons. Statistical significance of the node separating the species pairs *A. medirostris/A. transmontanus* and *A. fulvescens/A. oxyrhynchus* was ≥95% according to bootstrap analyses (500 replicates) and the nonoverlap of nodal standard error bars.

located ~74 bases upstream (on the light strand) of the control region in the gene for $tRNA^{Thr}$. The second *Hae*II site was not found in the sequenced 3′ terminus of the control region and, therefore, must be located further downstream.

Previously, it was shown that length variation of the *A. transmontanus* D-loop was due to varying numbers, from one to six copies, of an 82-bp sequence repeated in tandem (BUROKER *et al.* 1990). These repeated arrays were followed by a degenerate or partial copy of the repeat. In the present study, Southern blot analysis revealed that D-loop length variants of the new sturgeon species closely corresponded to those discrete size classes observed in *A. transmontanus* (Figure 2). This permitted a standardized ranking of mtDNA size variants across species relative to the observed *A. transmontanus* repeats (Table 2). Chi-square analysis revealed a highly significant difference between the frequencies of mtDNA size classes among all four species ($\chi^2 = 43.77$, $P < 0.01$). However, when the *A. oxyrhynchus* data were removed, no significant differences were observed among the remaining three species ($\chi^2 = 7.90$, $P = 0.61$).

The incidence of heteroplasmy was strikingly different between west and east species clades. About 42% of 174 sampled *A. transmontanus* individuals were heteroplasmic for at least two different mtDNA size variants

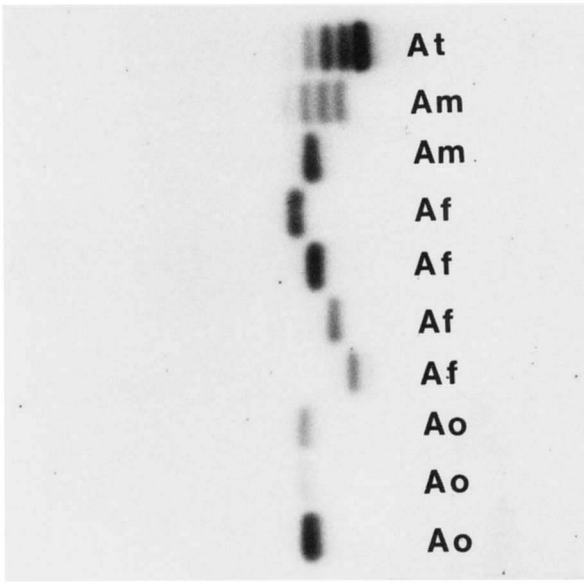

FIGURE 2.—Autoradiograph showing examples of size variation in mtDNA from various Acipenser species. CsCl-purified mtDNA from *A. transmontanus* (At), *A. medirostris* (Am), *A. fulvescens* (Af) and *A. oxyrhynchus* (Ao) was digested with the enzyme *Hae*II and probed with the D-loop region cloned from *A. transmontanus* mtDNA. Bands (1.5–2.0 kb in length from bottom to top) in the far left lane correspond to *A. transmontanus* mtDNA with 1–5 copies of a perfect 82-bp tandemly repeated sequence.

(BROWN *et al.* 1992a). Although the *A. medirostris* sample size ($n = 10$) was considerably smaller, 50% of those individuals were also heteroplasmic. The observed size range of mtDNA variants in *A. medirostris* closely corresponded to *A. transmontanus* D-loop regions with one to four repeats.

In contrast, heteroplasmic individuals were not detected in either eastern species. *A. fulvescens* individuals ($n = 21$) had one of five possible mtDNA size variants that closely corresponded to *A. transmontanus* mtDNA with one to five repeat units (Table 2). In *A. oxyrhynchus*, nearly every individual ($n = 19$) was fixed for mtDNA roughly equivalent in size to the smallest size variant (one repeat) found in the other three species. Only one *A. oxyrhynchus* individual had a larger D-loop fragment that roughly corresponded to an *A. transmontanus* D-loop with two repeats.

The mean genetic diversity within individuals (K_b) of the two western species were not significantly different (Student's $t = -0.782$; d.f. = 183; $P = 0.44$; Table 2). In the three species with extensive mtDNA length polymorphism, most of the genetic diversity could be attributed to variation between, rather than within, individuals (K_c and G_{IP}).

DNA sequences of tRNA genes: The DNA sequences of the genes for $tRNA^{Thr}$ and $tRNA^{Pro}$ of *A. medirostris*, *A. fulvescens* and *A. oxyrhynchus* differed little from those previously determined for *A. transmontanus* (GILBERT *et al.* 1988). The only mutations were a single substitution in the three nucleotide intergenic region of *A. fulvescens* and a single transitional substitution and a single base insertion in the $tRNA^{Thr}$ and $tRNA^{Pro}$ genes, respectively, of *A. oxyrhynchus*. None of these changes altered the previously proposed secondary structures for these tRNAs.

The control region: Entire and partial control regions were sequenced from selected individuals with varying numbers of putative repeats (Table 3). General organizations of control regions from the three additional sturgeon species are similar to that previously reported for *A. transmontanus* (Figure 3; BUROKER *et al.* 1990).

The overall architecture of sturgeon species' control regions closely resembles that of higher vertebrates with respect to the position of conserved sequence blocks or CSBs that have been implicated in the initiation of transcription and heavy strand replication (reviewed in CLAYTON 1982). The repeat region, responsible for both length variation and heteroplasmy is near the end of the D-loop region, separated by two or three nucleotides from the $tRNA^{Pro}$ genes. These repeats consistently have D-loop strand termination associated sequences or TAS motifs and an additional, divergent TAS-like block can be identified in the $tRNA^{Pro}$ gene. Separating the repetitive region and CSBs is a region of unique

TABLE 2

Distribution of size variants among four sturgeon species

Species	N	Copy numbers						Diversity		
		1	2	3	4	5	6	$\bar{K}_b$	K_c	G_{IP}
A. transmontanus	174	0.273	0.251	0.283	0.144	0.031	0.018	0.185 (±0.018)	0.762	0.758
A. medirostris	10	0.027	0.306	0.476	0.192	0.0	0.0	0.247 (±0.09)	0.642	0.615
A. fulvescens	21	0.381	0.143	0.238	0.190	0.048	0.0	0.0	0.739	1.0
A. oxyrhynchus	19	0.947	0.053	0.0	0.0	0.0	0.0	0.0	0.1	1.0

Size variants in all species were categorized according to distance travelled on a 1.2% agarose gel relative to mtDNA from a highly heteroplasmic *A. transmontanus* individual (see MATERIALS AND METHODS). $\bar{K}_b$ is the mean of the values, $K_b = 1 - \Sigma x_i^2$, where x_i is the proportion of the i^{th} copy number in each fish. The standard errors of $\bar{K}_b$ are in parenthesis. $K_c = 1 - \Sigma x_i^2$, where x_i is the frequency of the i^{th} size variant in each species. $G_{IP} = (K_c - \bar{K}_b)/K_c$ is the size diversity occurring between, rather than within, fish.

TABLE 3

***Acipenser* control region clones and repeat copy numbers**

Species	Clone label	Repeat copy no.[a]	No. of exact copies within an individual	No. of exact copies within species[b]	Repeat length
Acipenser transmontanus[c]	At1.9	4 + 1	4	5	82 bp
	At1.6	1 + 1	0		
A. medirostris	Am5	2 + 1	2	6	78 bp
	Am6	2 + 1	2		
	Am12	3 + 1	2		
A. fulvescens	Af14	3 + 1	2	2	82 bp
	Af17[d]	2 + 1	0		
A. oxyrhynchus	Ao8	2 + 1	0	2	78–79 bp
	Ao12	1 + 1	0		
	Ao17[d]	1 + 1	0		

[a] Repeat number is given as $n + 1$, where n is the number of nearly identical repeats and "+ 1" indicates a 3′ terminal partial repeat.
[b] Total number of sequenced repeats identical within a species.
[c] BUROKER *et al.* 1990.
[d] Partial D-loop sequences.

sequence, segments of which were polymorphic within a species (Figure 3).

Pairwise sequence comparisons of these nonrepeated or unique D-loop segments confirmed the species relationships suggested by RFLP analysis of whole mtDNA (Table 1). Neighbor-joining distance trees supported the separation of western and eastern species pairs in nearly 99% of 500 bootstrap replicates. Interspecific nucleotide substitution rates of unique D-loop sequences were estimated to be about four to five times greater than the average rates of the entire mtDNA genome (as estimated from RFLP comparisons), which is consistent with early calculations of intraspecific relative mutation rates for *A. transmontanus* (BROWN *et al.* 1993) and human (AQUADRO and GREENBERG 1983) hypervariable D-loop sequences.

Previous pairwise comparisons of *A. transmontanus* hypervariable sequences from 27 individuals revealed a

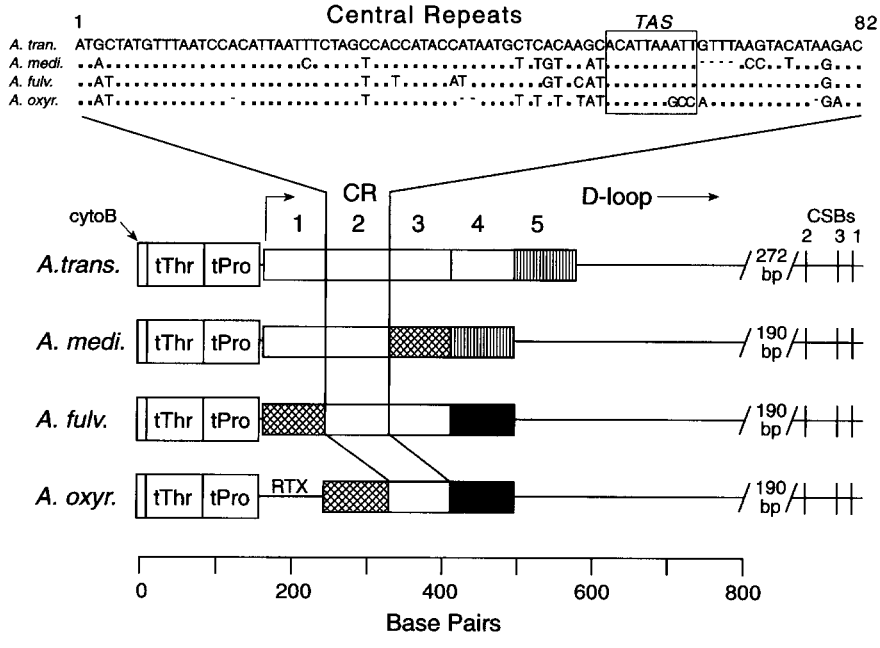

FIGURE 3.—Schematic diagram of aligned light strand sequences of *Acipenser*. An ~1000–1100-bp segment extending from the 3′end of the cytochrome *b* (cytoB) gene, through the genes for $tRNA^{Thr}$ (*tThr*) and $tRNA^{Pro}$ (*tPro*), to the conserved sequence blocks (CSB) located on the 3′ end of the control region were sequenced from the sturgeon species: *A. transmontanus* (designated *A. trans.*; BUROKER *et al.* 1990), *A. medirostris* (*A. medi.*), *A. fulvescens* (*A. fulv.*), and *A. oxyrhynchus* (*A. oxyr.*). The relative positions of tandem D-loop repeats (labeled 1–5) are shown as well as the degree of sequence conservation among repeats within each species array relative to the central repeat (sequences and position of TAS motifs shown above). On the basis of intraspecific pairwise comparisons to the central repeat, the flanking repeats differed by either 0% (white fill), <5% (cross-hatch); between 20 and 30% (vertical lines) or between 30 and 40% (black fill) nonidentical nucleotides. The number of repeats shown for each species reflects the maximum number sequenced within any individual rather than the maximum number in any species detected by Southern blot hybridizations. In the case of *A. oxyrhynchus*, the repeats are displaced from the tPro gene by a segment, RTX, of limited identity with the repeat elements.

```
          1                                                                                                  84
At1.9  CR ATGCTATGTT TAATCCACAT TAATTTCTAG CCACCATACC AT-AATGCTC ACAAGCACA- TTAAATTGTT TAAGTACATA AGAC
At1.9 RT5 .......... .......... .......... ...T....T. .C-....T.T CATCTAC..- ......-... AT.CAC...T .TTT

Am12  CR ATACTATGTT TAATCCACAT TAATCTCTAG TCACCATACC AT-AATGTTT GTAAATACA- TTAAATT--- ACCTATA-TA GGAC
Am12 RT3  .......... .......... .......... .......... ..-....... ........CA .......G-- .......-.. ....
Am12 RT4  .......... .......... .......... .......... ..-....... CATCTAC..- ......GATA TA.ACC.-.T ATCT

Am5  RT3  .......... .......... .........T GTG.AC..A. ..-.CGA... CCCG....-- ...G..ATGT .GTA.G.GCC .A..

Af14  CR ATATTATGTT TAATCCACAT TAATTTCTAG TCATCATACA TT-AATGCTC GTACATACA- TTAAATTGTT TAAGTACATA GGAC
Af14 RT1  ...C...... .......... .......... .......... .C-....... .........- .......... .......... ....
Af14 RT4  .......... .......... .......C.. ...CT.C... CAG....T.T CAT.TAC..- .C....GA.C C.CACCAT.. TCT.

Af17 RT1  ...CG..... .......... .......... ...C..C... A.-....... .........- C......A.. .......... ....
Af17 RT2  .......... .......... .......C.

Ao8   CR ATATTATGTT TAATCC-CAT TAATTTCTAG TCACCATACC ---AATGTTT ATATATACA- TTAAGCCATT TAAGTACAT- GAAC
Ao8  RTX  ....C...C. CGC-.TA..C C...G.T..- ---------- ---------- ------...- ....ATTGC. ..T.--...- .G..
Ao8  RT3  ...C...... .T....C... .......... .......... ---....... .........- .......... .........- ....
Ao8  RT4  ...C...... ..TG..C... ...C...ACT AT.A..---- --G....... CATCTAC..- ....A-TG.. AT.TAC...T TTCT

Ao12 RTX  ....C...C. CGC-.TA..C C...G.T..T ---------- ---------- ------...- ....ATTG-- CTTA.G...- .G..
Ao12 RT3  ...C...... .T....C... ...C...... C...T...A. A-GG...... CATCTAC..- ....A-TG.. AT.TAC...T TTCT

Ao17 RTX  ....C...C. GGC-.TA..C C...G.T..T ---------- ---------- ------...- ....ATTG-- CTTA.G...- .G..
Ao17 RT2  .......... ......C... .......... .......... ---....... .........- .......... .........- ....
Ao17 RT3  ...C...... .T....C... ...C...... C...T..
```

FIGURE 4.—Aligned light strand sequences of repeated elements. Sequence designations, given on the left, are defined in Table 3. Variant repeats within each species are aligned relative to the central repeat (CR), which in turn are aligned across species (hence 84 bp rather than 82 bp in length). Not shown are repeats within individuals or a species that were identical to the central repeat (see Table 3 and Figure 4). Dots represent complete sequence identity and dashes are gaps. For each individual, the last sequence contains a partial repeat added to the length of a complete repeat unit using adjacent (3′) unique control region sequence. In *A. oxyrhynchus* (Ao), repeat elements do not begin at bp 34 of the D-loop region as in other species therefore this "nonrepeated" sequence is labeled "RTX".

range of 0.0–0.043 corrected nucleotide substitutions per site (BROWN *et al.* 1993). Low levels of sequence variation were also observed for *A. medirostris* (0.0; $N = 3$), *A. fulvescens* (0.018; $N = 2$), and *A. oxyrhynchus* (0.002–0.008; $N = 3$).

Repeated elements: mtDNA length variation could be attributed to different copy numbers of tandemly repeated sequences in the D-loop region (Figure 3). Each species had at least one complete copy of their respective repeated element followed by a 3′ terminal degenerate repeat (Figure 3 and Table 3). Depending on the species, individual repeats were 78–82 bp in length with a putative TAS motif similarly located in each repeated module (Figure 3).

The central repeats were well-conserved within individuals and within each species but less so between species (Table 1; Figure 3). The repeated elements in *A. fulvescens* most closely resembled those of *A. transmontanus* (BUROKER *et al.* 1990) although the range of distances among all species was quite narrow (0.157–0.239 corrected nucleotide substitutions per site). Overall, the 5′ end of the repeats showed greater sequence conservation than the 3′ end (Figures 3 and 4).

Intraspecific and intraindividual conservation of repeat sequences were greatest in *A. transmontanus* (BUROKER *et al.* 1990) and *A. medirostris*. In the latter species, the basic repeat length was 78 bp. Of the *A. medirostris* D-loops sequenced, two individuals (Am5 and Am6) each had two perfect repeats while a third (Am12) had three repeats. mtDNA from Am5 and Am6 was also identical with respect to comparisons of RFLPs and unique D-loop regions. The first two repeats in all three individuals were identical while the third repeat in individual Am12 differed from the perfect repeats at three nucleotide positions (Figure 4).

The two *A. fulvescens* individuals (Af14 and Af17) examined had four and two repeats in their respective control regions with each repeat being 82 bp in length. The second repeat element was only partially sequenced for Af17 (Figure 4). In contrast to *A. transmontanus* and *A. medirostris*, the second and third repeats in *A. fulvescens* were identical while the first repeat differed by two C-T transitions. The first repeat of Af17 individual was similar, but not identical, to the two identical repeats and the initial imperfect repeat of Af14.

Relative to the other three sturgeon species, length variation in *A. oxyrhynchus* mtDNA is rare. A larger D-loop fragment was detected in only one of the 19 individuals investigated. Of the three *A. oxyrhynchus* individuals sequenced, two (Ao12 and Ao17) had a single repeat element, 78 bp in length, the other (Ao8) had two complete copies (CR and RT3; Figure 4). The first repeats in individuals Ao8 and Ao12 were identical while that of Ao17 differed by single base insertion (Fig-

TABLE 4

Thermal stability of potential intrastrand duplex structures in sturgeon mitochondrial control region repeat sequences

Species	Number of repeats			
	1	2	3	4
A. transmontanus[a]	−14.1	−34.4	−58.0	−78.3
A. medirostris[b]	−8.5	−30.0	−46.6	−57.5
A. fulvescens[c]	−12.4	−26.8	−44.3	
A. oxyrhynchus[d]	−6.8	−23.1		

Thermal stability was kcal M^{-1}.
[a] Values from BUROKER *et al.* (1990).
[b] 78-nt element repeats starting at bp 5 in Figure 4.
[c] 82-nt element repeats starting at bp 1 in Figure 4.
[d] 79-nt element repeats starting at bp 5 in Figure 4.

ure 4). The two repeats within Ao8 also differed slightly from each other. In *A. oxyrhynchus,* the D-loop sequence adjacent to the $tRNA^{Pro}$ gene had only limited similarity to the downstream repeats, including a 27-bp deletion, and single base differences (this region is labeled RTX in Figures 3 and 4).

Secondary structure: The assignment of the initial point of repeats within each species is somewhat arbitrary because the optimal secondary structure may or may not begin at the first nucleotide of a repeated array. Only in *A. medirostris* and *A. transmontanus* did we find a DNA segment near the length of the estimated repeat unit that could form a secondary structure with a terminal hairpin stem. In the remaining species, the segments that could form a hairpin with a terminal stem were all smaller than the estimated repeat lengths. Although *A. medirostris* 78 bp long repeats will fold to form secondary structures, they will not form hairpin stems base-paired exactly to the 5′ and 3′ terminus of the repeat units. For terminus base-pairing to occur, the element is 77 bp long and the next element is removed by one nucleotide (C) for each repeat. The estimated thermal stabilities (Table 4) for the *A. medirostris* repeats are somewhat lower than those previously reported for *A. transmontanus* repeats (BUROKER *et al.* 1990).

In *A. fulvescens,* only one terminally base-paired 76-bp hairpin structure near the repeat length could be formed, from nucleotide position 37–112 of the D-loop region. In *A. oxyrhynchus,* a terminally base-paired stem segment 77 bp in length is possible from base pair 120–200 and base pair 202–283 of the D-loop region. These junction regions are themselves five nucleotide palindromes. Thermal stabilities for potential hairpins of repeat elements found in *A. oxyrhynchus* and *A. fulvescens* are also lower than those calculated previously for *A. transmontanus* (Table 4).

DISCUSSION

We determined the structure of the D-loop in four North American species of sturgeon. In all four species, multiple copies of a 78–82 basepair segment bearing the termination associated sequences (TAS) were found. Three different structural patterns were observed in these species. In the two West Coast species, repeat copy numbers varied both between and within individuals. In contrast, no heteroplasmy was found in the Eastern species. *A. fulvescens* individuals were fixed for varying numbers of repeats, while all but one *A. oxyrhynchus* were fixed for a single copy of this segment. To attempt to understand these patterns, we first examine the phylogenetic relationships of the species.

Sturgeon phylogeny: The phylogeny of the four *Acipenser* species, as inferred from either whole mtDNA RFLP or unique control region sequence data, clearly shows distinctive species assemblages on opposite sides of the Continental Divide. The Pacific Coast species, *A. transmontanus* and *A. medirostris,* are more closely related to each other than to either species living east of the Divide, *A. fulvescens* and *A. oxyrhnychus.*

A. transmontanus and *A. medirostris* apparently diverged from each other well after the genetic isolation of western fish fauna. The present-day distributions of these species overlap. *A. medirostris* inhabits marine coastal waters and migrates short distances inland to spawn in freshwater while *A. transmontanus* has a similar marine distribution but is often found further inland in lakes and rivers along the western slope of the Continental Divide (SCOTT and CROSSMAN 1973).

The separation of *A. oxyrhynchus* from the three other species was probably the most ancient divergence event and may have occurred in the Mississippi Basin. The contemporary distribution of *A. fulvescens* includes the middle to upper areas of the Mississippi drainage while *A. oxyrhynchus* are found in the lower river and coastal waters (SWIFT *et al.* 1986). The long term ecological stability of the Mississippi River throughout the Cenozoic likely facilitated a broadening of species diversity, which, perhaps, included the emergence of new species of sturgeon (SMITH 1981; ROBISON 1986).

Nucleotide substitution rates: Among mammalian mtDNAs, the generalized rate of nucleotide substitution has been estimated to be ~2% per million years (reviewed in BROWN 1985; WILSON *et al.* 1985). Application of this rate to the present RFLP data would suggest that *A. fulvescens* diverged from *A. transmontanus* and *A. medirostris* ≥3.5 mya. This estimated time of divergence is much later than speciation events that are postulated to have occurred as a result of transcontinental uplifting and block-faulting during the mid- to late Miocene, some 7–12 mya (CAVENDER 1986).

In addition to the RFLP data (sampling the entire mitochondrial genome), we can examine the evolutionary rate of the D-loop sequence. Previous intraspecific analysis of *A. transmontanus* mtDNA estimated the nucleotide substitution rate of unique D-loop sequences to be four- to fivefold higher than that observed in the mitochondrial genome as a whole (BROWN *et al.* 1993).

Present pairwise comparisons among sturgeon species confirm this elevated *relative* mutation rate and comparable estimates have been made for human D-loops (AQUADRO and GREENBERG 1983). Pairwise comparisons among human hypervariable D-loop sequences suggest an *absolute* mutation rate of ~8% per million years (VILIGANT *et al.* 1989). Provisional application of this rate to sturgeon suggests that the divergence of *A. transmontanus* and *A. medirostris* from *A. fulvescens* occurred ~5.7 mya while the subsequent *A. transmontanus* and *A. medirostris* split happened 2.2 mya. These estimated times of divergence are similar to those derived from RFLP analysis and are not compatible with Miocene speciation events.

Interspecific comparisons based on a Miocene separation of East and West Coast species suggest a two- to fourfold lower mutation rate in sturgeon relative to mammalian mtDNA. A substantially slower rate of nucleotide substitution has been observed in a number of large, long-lived ectotherms (MARTIN *et al.* 1992; AVISE *et al.* 1992; MARTIN and PALUMBI 1993). If we assume that the Western clade has been isolated for 10 mya, we estimate a divergence time of ~6 mya for *A. transmontanus* and *A. medirostris*.

Concerted evolution: While *A. medirostris* and *A. transmontanus* individuals share patterns of mtDNA size variation and heteroplasmy, the actual nucleotide composition of repeat units differed between species. In two *A. medirostris* individuals with three and four repeats, the initial (5′ end) two and three repeats, respectively, were nearly identical both within and between individuals. A similar intraspecific pattern of sequence conservation of repeat elements was reported for *A. transmontanus* (BUROKER *et al.* 1990).

Secondary structure analysis suggests that 78-bp repeats in *A. medirostris* might be capable of forming stem-loop structures comparable with those proposed for 82-bp repeats in *A. transmontanus*. The illegitimate elongation model predicts that in an array of consecutive repeats, the central repeats should evolve in a concerted manner (BUROKER *et al.* 1990). Nucleotide sequences of the repeats may vary between distantly related genotypes (or species) but within a genotype, repeats should be perfectly conserved. Apparently, such is the case for repeats in *A. medirostris* and *A. transmontanus*. The high level of sequence identity among central repeats both within and between individuals of each species, suggests the existence of some on-going process that serves to homogenize the sequences of central, adjacent repeats.

In *A. fulvescens*, the nucleotide compositions of the second and third repeats were identical within a single individual but not between individuals. The first repeat was slightly different in comparison to subsequent repeats within the same array as well as to the initial repeat in another individual. *A. oxyrhynchus* individuals were consistently fixed for single repeats with the exception of an individual with two, nonidentical copies. The occurrence of nucleotide substitutions and single indels among repeat elements, suggests that sequence homogenization processes in *A. fulvescens* and *A. oxyrhynchus* are not acting as strongly as those present in heteroplasmic sturgeon species.

Several reports of heteroplasmy involving tandem repetitive sequences in the D-loop region indicate repeat segments can fold into putative stable structures, either singly or as multiples, leading to the addition or deletion of repeat elements (BUROKER *et al.* 1990; WILKINSON and CHAPMAN 1991; ÁRNASON and RAND 1992; STEWART and BAKER 1994). It must be noted, however, that models for the prediction of stable DNA secondary structure, and perhaps more importantly tertiary interactions, are inexact.

mtDNA size variation and heteroplasmy: A number of authors have suggested that heteroplasmy involving tandem repeats is dynamically maintained by high mutation rate, offsetting the effects of rapid segregation (RAND and HARRISON 1987; BUROKER *et al.* 1990). As CLARK (1988) noted, relatively subtle differences in mutation rate may determine whether a system is tipped in favor of mutations leading to heteroplasmy, or segregational elimination of heteroplasmy. As a byproduct of high levels of mutation between different size classes, we would expect concerted evolution to maintain sequence identity among the repeats in any mitochondrial lineage.

Length heteroplasmy as seen in mitochondrial D-loops evidently requires several features: the presence of tandem repeats, a high degree of sequence identity among repeats, and the capability to form secondary structure when single stranded. Loss of an ancestral heteroplasmic condition could presumably occur by any of three mechanisms: segregational loss of tandem repeats in lineages and species, point mutations in repeat elements, reducing the ability for slipped-strand pairing, and point mutations reducing the secondary structural characteristics. Gain of heteroplasmy, either from D-loop structures lacking repeats, or from a variable but fixed multicopy state as observed in *A. fulvescens*, would appear to be less likely.

The results of this study shows that two species that diverged ~6 mya share the multicopy heteroplasmic state. We also find that tandem repeats, either fixed or heteroplasmic, are present in all four taxa. The repeats are very similar in both length and sequence, suggesting that at least a multicopy D-loop structure was present in an ancestor common to all four species perhaps ≥12 mya. We suggest that the ancestral condition was heteroplasmic as well, and that heteroplasmy was secondarily lost independently in both eastern species. Under this hypothesis, *A. fulvescens*, lost the heteroplasmic condition but not the high frequency multicopy state. Most *A. oxyrhynchus* individuals bear only a single copy. Loss of heteroplasmy in this species could have occurred by any of the three mechanisms noted above. If the initial

step was fixation of a variable multicopy condition, then we must assume that multiple copies were lost through some other mechanism.

It should be noted that we cannot rule out the alternative that heteroplasmy was gained in the western species, either in the common ancestor or as recent independent events in each species. But the phylogenetic evidence indicates that the origin of the tandem repeat array is much older. When the array first arose, at least two of the three requirements for heteroplasmy would be met: multiple tandem copies and sequence identity of those copies.

Mutation and the distribution of copy number: It is clear that mutation rates altering copy number play an important role in the maintenance of heteroplasmy and the maintenance of sequence identity among repeats. At least three mutation rates must be considered: the mutation rate to increased copy number when only a single copy is present, the mutation rate to decreased copy number when multiple tandem repeats are present, and the mutation rate to increased copy number when multiple copy numbers are present. The latter two parameters are believed to be quite high in species with high levels of length heteroplasmy (RAND and HARRISON 1989; BUROKER *et al.* 1990). Mutations from a single copy are generally assumed to be rare.

Mutation rate to increase or decrease copy number from a multicopy state must be considered separately. Under the illegitimate elongation model of BUROKER *et al.* (1990), only the displaced strand is single stranded. If the D-loop strand is displaced, it will be free to form hairpins. Folding of the D-loop strand is an essential step leading to increased copy number in that model. Under recombination models, such as that proposed by RAND and HARRISON (1989), the distinction between the mechanisms for increased and decreased copy number is even more pronounced.

The relative magnitudes of these mutation parameters may play a major role in determining the distributions of copy number in these organisms. We have previously analyzed these distributions in *A. transmontanus* (BROWN *et al.* 1992a). The distributions of length variation in all four species is shown in Figure 5. For two of the species, *A. transmontanus* and *A. fulvescens,* the distributions are evidently compounded with a mode at one copy, and a second mode at three repeats. The mode at one copy is easily explained, both as a boundary condition, and as a near-absorbing state because we assume that mutations from a single copy structure are infrequent. The mode at three repeats is more interesting. If the mutation rate to decreased copy number exceeds the rate to increased copy number, mutation pressure should eliminate multiple copy genomes. If the rate to increase copy number exceeds the rate of mutation to lower copy number, then evolution should lead to large numbers of copies, unless opposed by

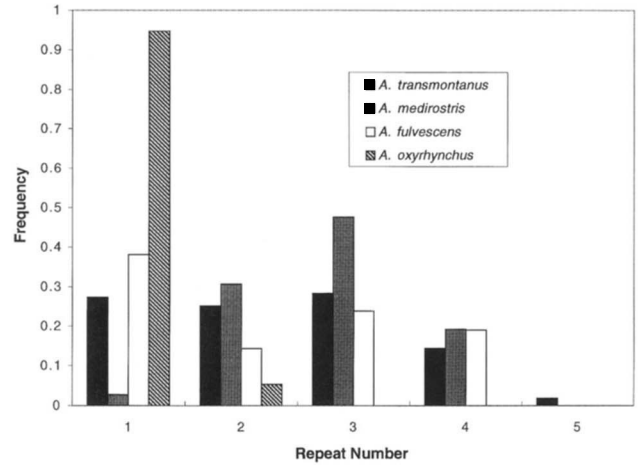

FIGURE 5.—Frequency distributions of repeat arrays in sturgeon species. Distributions of repeat copy numbers for *Acipenser* species were taken from Table 2.

other factors. If the two rates are identical, then other factors will determine the distribution.

Two forces opposing the accumulation of large numbers of repeats have been proposed: selection (reduced fitness at the organismal level) and a replication race during the buildup of mitochondria during oogenesis (RAND 1993). It is not clear how much effect either of these two mechanisms would have on a short, evidently noncoding segment in an 16-kb genome. We offer an alternative mechanism that might apply when the repeated segment includes the TAS motif.

A TAS-based replication model: MADSEN *et al.* (1993), using an *in organello* methylation protection assay, discovered an approximately 48-kD *trans*-acting protein that binds to a conserved TAS-like sequence in bovine mtDNA control regions. Although multiple TAS-like sequences were identified on the basis of sequence similarity, protein-binding was shown to be specific to only one type of TAS motif. It was suggested that the synthesis of D-loop strands could be arrested by the formation of a single TAS-protein complex although the involvement of additional factors in the initiation of H-strand synthesis cannot be ruled out. H-strand replication appears to be tightly regulated since, although D-loop strand initiation and termination continuously occur, only ~0.05 of the initiated D-loop strands lead to completed cycles of H-strand synthesis (BOGENHAGEN and CLAYTON 1978).

If TAS motifs are involved in the negative regulation of D-loop strand synthesis, then a mechanistic explanation for the replicative advantage of smaller mitochondrial genomes can be postulated. Provided that quantities of the TAS-binding *trans*-acting factor are not saturating, there would be a 0.05 probability that, at any one moment, a single TAS block will not be bound to a protein thus allowing H-strand synthesis to proceed. If more than one functional TAS-like sequence were present, as may be the case for multiple repeated

elements in sturgeon, then the successful initiation of H-strand synthesis would depend upon the probability that all TAS sequences were simultaneously unbound, defined here as p(TAS). This probability could be described by the simple exponential function $p(TAS) = (0.05)n$ where n is the number of boundable TAS sequences. This is a very simplistic model since the *in vivo* frequency of TAS-protein complex assembly and the successful completion of H-strand synthesis are likely mediated by additional factors (CLAYTON 1991; MADSEN *et al.* 1993).

However, this model of TAS-based regulation of H-strand synthesis can be reconciled with some pan-species observations on the structural arrangements and frequency distributions of repeat sequences. First, the TAS-limitation model predicts an exponential decline in successful H-strand replication with increasing repeat copy number. The resulting distribution would be skewed toward genomes with fewer repeats, which is consistent with observations in sturgeon (BROWN *et al.* 1992a) and several other species (see figure 4 in RAND 1993). The second prediction is that the replication of genomes with a large number of repeats (say more than seven copies) will be rare events. Accordingly, for vertebrate species in which mtDNA length variation consists of short (<200 bp) tandem repeats with embedded TAS-like motifs, the number of repeats in an array ranges from one to eight copies (RAND 1993). While some species, such as sturgeon, appear to require only a single repeat for a functional control region, other vertebrates, including treefrog (YANG *et al.* 1994), shrews (STEWART and BAKER 1994) and bats (WILKINSON and CHAPMAN 1991), have arrays with a minimum of four to five repeats, each with an embedded TAS motif. The higher minimal number of repeats in these species might be due to yet uncharacterized TAS-protein binding requirements for specific DNA strand configurations, such as stable secondary structures involving multiple repeats.

Figure 5 shows the frequency of repeats in different sturgeon species. While the extremes of this distribution closely correspond to the observed frequencies of genomes with either a single repeat or four or more repeats, there is a lack of correspondence between the predicted and observed frequencies of two and three repeats. Earlier, it was suggested that this distribution reflects recurrent mutation for higher repeat copy number partially offsetting selection for smaller genome size. Accordingly, the frequency of mutation for increasing size might be sufficient to overcome the frequency of assembling protein-DNA complexes at TAS motifs responsible for D-loop termination. Alternatively, secondary structural formations involving a certain number of repeats might be more predisposed to promote D-loop strand elongation than structures formed from either fewer or more repeats.

Similarly skewed distributions have been observed in invertebrates such as crickets (RAND and HARRISON 1989) and Drosophila (HALE and SINGH 1986). The repeats in crickets have been localized to the A-T rich region, which has been implicated in insect mtDNA replication. However, the regulation of H-strand synthesis in invertebrates is poorly understood and sequences motifs homologous to TAS and CSB sequences in vertebrate control regions have not been identified (reviewed in CLAYTON 1991). Therefore, the present TAS-limitation model cannot be convincingly extended to invertebrate mtDNA at this time.

We wish to thank L. BECKMAN, A. J. DEREKSEN, S. DIONNE, S. GUÉNETTE and C. TRACY for providing sturgeon liver samples and two anonymous reviewers for their comments on this manuscript. This research was supported by a Simon Fraser University (SFU) Special Research Grant, SFU President's Research Grant and Natural Sciences and Engineering Grants to A.T.B. and M.J.S. J.R.B. was supported by a British Columbia Sciences Council Graduate Research Award and currently holds a Medical Research Council of Canada Postdoctoral Fellowship in the lab of W. F. DOOLITTLE.

LITERATURE CITED

AQUADRO, C. F., and B. D. GREENBERG, 1983 Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics **103:** 287–312.

ÁRNASON, E., and D. M. RAND, 1992 Heteroplasmy of short tandem repeats in mitochondrial DNA of Atlantic cod, *Gadus morhua*. Genetics **132:** 211–220.

AVISE, J. C., B. W. BOWEN, T. LAMB, A. B. MEYLAN and E. BERMINGHAM, 1992 Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines. Mol. Biol. Evol. **9:** 457–473.

BIRKY, C. W., P. FUERST and T. MARUYAMA, 1989 Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics **121:**613–627.

BOGENHAGEN, D., and D. A. CLAYTON, 1978 Mechanism of mitochondrial DNA replication in mouse L-cells: kinetics of synthesis and turnover of the initiation sequence. J. Mol. Biol. **119:** 49–68.

BROWN, J. R., T. L. GILBERT, D. J. KOWBEL, P. J. O'HARA, N. E. BUROKER *et al.*, 1989 Nucleotide sequence of the apocytochrome *b* gene in white sturgeon mitochondrial DNA. Nucleic Acids Res. **17:** 4389.

BROWN, J. R., A. T. BECKENBACH and M. J. SMITH, 1992a Mitochondrial DNA length variation and heteroplasmy in populations of white sturgeon (*Acipenser transmontanus*). Genetics **132:** 221–228.

BROWN, J. R., A. T. BECKENBACH and M. J. SMITH, 1992b Influence of Pleistocene glaciations and human intervention upon mitochondrial DNA diversity in white sturgeon (*Acipenser transmontanus*) populations. Can. J. Fish. Aquat. Sci. **49:** 358–367.

BROWN, J. R., A. T. BECKENBACH and M. J. SMITH, 1993 Intraspecific DNA sequence variation of the mitochondrial control region of white sturgeon (*Acipenser transmontanus*). Mol. Biol. Evol. **10:** 326–341.

BROWN, W. M., 1985 Evolution of animal mitochondrial DNA, pp. 62–88 in *Evolution of Genes and Proteins*, edited by M. NEI and R. K. KOEHN. Plenum Press, New York.

BUROKER, N. E., J. R. BROWN, T. L. GILBERT, P. J. O'HARA, A. T. BECKENBACH *et al.*, 1990 Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics **124:** 157–163.

CABOT, E. L., and A. T. BECKENBACH, 1989 Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput. Appl. Biosci. **5:** 233–234.

CAVENDER, T. M., 1986 Review of the fossil history of North American freshwater fishes, pp. 699–724 in *The Zoogeography of North American Freshwater Fishes*, edited by C. H. HOCUTT and E. O. WILEY. John Wiley and Sons, New York.

CLARK, A. G., 1988 Deterministic theory of heteroplasmy. Evolution **42:** 621–626.

CLAYTON, D. A., 1982 Replication of animal mitochondrial DNA. Cell **28:** 693–705.

CLAYTON, D. A., 1991 Replication and transcription of vertebrate mitochondrial DNA. Annu. Rev. Cell Biol. **7:** 453–478.

EFSTRATIADIS, A., J. W. POSAKONY, T. MANIATIS, R. W. LAWN, C. O'CONNELL *et al.,* 1980 The structure and evolution of the human-globin gene family. Cell **21:** 653–668.

FELSENSTEIN, J., 1993 PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, WA.

GILBERT, T. L., J. R. BROWN, P. J. O'HARA, N. E. BUROKER, A. T. BECKENBACH *et al.,* 1988 Sequence of $tRNA^{Thr}$ and $tRNA^{Pro}$ from white sturgeon (*Acipenser transmontanus*) mitochondria. Nucleic Acids Res. **16:** 11825.

HALE, L.R., and R.S. SINGH, 1986 Extensive variation and heteroplasmy in size of mitochondrial DNA among geographical populations of *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **83:** 8813–8817.

HAUSWIRTH, W. W. and P. J. LAIPIS, 1982 Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. Proc. Natl. Acad. Sci. USA **79:** 4686–4690.

KIMURA, M., 1980 A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

KOEHLER, C. M., G. L. LINDBERG, D. R. BROWN, D. C. BEITZ, A. E. FREEMAN *et al.,* 1991 Replacement of bovine mitochondrial DNA by a sequence variant within one generation. Genetics **129:** 247–255.

KUMAR, S., K. TAMURA and M. NEI, 1993 MEGA: molecular evolutionary genetics analysis, version 1.0. The Pennsylvania State University, University Park, PA.

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. J. Mol. Evol. **17:** 214–226.

LASKEY, R. A., 1980 The use of intensifying screens or organic scintillators for visualizing radioactive molecules resolved by gel electrophoresis. Methods Enzymol. **65:** 363–371.

MADSEN, C. S., S. C. GHIVIZZANI and W. M. HAUSWIRTH, 1993 Protein binding to a single termination-associated sequence in the mitochondrial DNA D-loop region. Mol. Cell. Biol. **13:** 2162–2171.

MARTIN, A. P., G. J. P. NAYLOR and S. R. PALUMBI, 1992 Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. Nature **357:** 153–155.

MARTIN, A. P., and S. R. PALUMBI, 1993 Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA **90:** 4087–4091.

MESSING, J., and J. VIEIRA, 1982 A new pair of M13 vectors for selecting either strand of double-digest restriction fragments. Gene **19:** 269–276.

MORITZ, C., T. E. DOWLING and W. M. BROWN, 1987 Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18:** 269–292.

NEI, M., J. C. STEPHENS and N. SAITOU, 1985 Methods for computing the standard errors for branching points in an evolutionary tree and their application to molecular data from human and apes. Mol. Biol. Evol. **2:** 66–85.

NEI, M., and F. TAJIMA, 1983 Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics **105:** 207–217.

RAND, D. M., 1993 Endotherms, ectotherms, and mitochondrial genome-size variation. J. Mol. Evol. **37:** 281–295.

RAND, D. M., and R. G. HARRISON, 1986 Mitochondrial DNA transmission genetics in crickets. Genetics **114:** 955–970.

RAND, D. M., and R. G. HARRISON, 1989 Molecular population genetics of mtDNA size variation in crickets. Genetics **121:** 551–569.

ROBISON, H. W., 1986 Zoogeographic implications of the Mississippi River basin, in *The Zoogeography of North American Freshwater Fishes,* edited by C. H. HOCUTT and E. O. WILEY. John Wiley and Sons, New York.

ROFF, D. A., and P. BENTZEN, 1989 The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. Mol. Biol. Evol. **6:** 539–545.

SCOTT, W. B., and E. J. CROSSMAN, 1973 Freshwater fishes of Canada. Fish. Res. Bd. Canada Bull. **184.**

SMITH, G. R., 1981 Late cenozoic freshwater fishes of North America. Annu. Rev. Ecol. Syst. **12:** 163–193.

STEWART, D. T., and A. J. BAKER, 1994 Patterns of sequence variation in the mitochondrial D-loop region of shrews. Mol. Biol. Evol. **11:** 9–21.

SWIFT, C. C., C. R. GILBERT, S. A. BORTONE, G. H. BURGESS and R. W. YERGER, 1986 Zoogeography of the freshwater fishes of the southeastern United States: Savannah River to Lake Pontchartrain, pp. 213–266 in *The Zoogeography of North American Freshwater Fishes,* edited by C. H. HOCUTT and E. O. WILEY. John Wiley and Sons, New York.

VIGILANT, L., R. PENNINGTON, H. HARPENDING, T. D. KOCHER and A. C. WILSON, 1989 Mitochondrial DNA sequences in single hairs from a southern African population. Proc. Natl. Acad. Sci. USA **86:** 9350–9354.

WILKINSON G., and A. M. CHAPMAN, 1991 Length and sequence variation in evening bat D-loop mtDNA. Genetics **128:** 607–617.

WILSON, A. C., R. L. CANN, S. M. CARR, M. GEORGE, U. B. GYLLENSTEN *et al.,* 1985 Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. **26:** 375–400.

YANG, Y .-J., Y. S. LIN, J.-L. WU and C. F. HUI, 1994 Variation in mitochondrial DNA and population structure of the Tapei treefrog *Rhacophorus taipeianus* in Taiwan. Mol. Ecol. **3:** 219–228.

ZUKER M., and P. STIEGLER, 1981 Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9:** 133–148.

Communicating editor: A. G. CLARK