

Structure and Conservation of Tandem Repeats in the Mitochondrial DNA Control Region of the Least Brook Lamprey (*Lampetra aepyptera*)

Matthew M. White · Holly R. Martin

Received: 22 January 2009 / Accepted: 27 April 2009 / Published online: 16 May 2009
© Springer Science+Business Media, LLC 2009

Abstract Tandemly repeated sequences are a common feature of vertebrate mitochondrial DNA control regions. However, questions still remain about their mode of evolution, function, and phylogenetic distribution. We report phylogenetic and geographic patterns of variation of control region repeat sequence and number in a nonparasitic lamprey, *Lampetra aepyptera*. A survey of populations from throughout the species' range revealed remarkably low repeat sequence polymorphism but some interpopulation variation in repeat number. The high sequence similarity extended to repeats observed in other species in the genus *Lampetra* and other lamprey genera. The very low levels of variation suggest a high copy turnover. Our data are consistent with the illegitimate elongation model of repeat gain and loss and further suggest that repeat change occurs at internal copies. However, the limited variation across some species of lamprey suggests that functional constraints may further limit variation.

Keywords Mitochondrial DNA · Tandem repeats · Control region · Illegitimate elongation model · Lamprey · Geographic variation

Introduction

Numerous studies have documented DNA sequence variation in the mitochondrial control region. Despite the presence of several highly conserved sequences (Saccone et al. 1987), this region is known to exhibit some of the

highest rates of evolution of any region on the mitochondrial DNA (e.g., McMillan and Palumbi 1997). Much of this variation consists of nucleotide substitutions and small insertions and deletions. However, considerable length variation has been observed in an ever-growing list of species (Lunt et al. 1998). This length variation is due to variation in the number of tandemly repeated sequences. These repeats can be observed near the 5' end of the control region (e.g., Arnason and Rand 1992; Starner et al. 2004; Ludwig et al. 2000; Faber and Stepien 1998) or near the 3' end (e.g., Broughton and Dowling 1994; Mundy and Helbig 1994), vary in size from four to hundreds of nucleotides, and may number from 2 to more than 100 copies (Lunt et al. 1998).

Length variation is thought to arise via various mechanisms that incorporate misalignment of tandem repeats in the control region during mtDNA replication (Buroker et al. 1990; Broughton and Dowling 1994; Mundy and Helbig 1994). The misalignment is stabilized by secondary structures that form in the displaced strand. Among those species that exhibit mtDNA length variation, many are heteroplasmic, though the frequency of heteroplasmy can vary among populations (Wilkinson and Chapman 1991; Arnason and Rand 1992; Starner et al. 2004) and species (Ludwig et al. 2000). Some recent studies have suggested that mtDNA recombination can generate tandem repeat variation (Piganeau et al. 2004), particularly in species with heteroplasmy (Hoarau et al. 2002; Mjelle et al. 2008).

One of the great strengths of utilizing mtDNA in population genetics and systematics is the conserved gene arrangement (Meyer 1993; Boore 1999), particularly among vertebrates. One exception to this arrangement is the organization of the lamprey mitochondrial genome, which differs from the consensus vertebrate order (Lee and Kocher 1995; Boore 1999) including that of the hagfish

M. M. White (✉) · H. R. Martin
Department of Biological Sciences, Ohio University, Athens,
OH 45701, USA
e-mail: whitem@ohio.edu



Fig. 1 Structure of the lamprey mitochondrial control region identifying the approximate location of the repeat region in noncoding region 1. The lamprey control region is flanked by the ND6 and

CYTB genes. Noncoding regions 1 and 2 are separated by the Glu-tRNA and Thr-tRNA genes. The approximate locations of the presumed conserved sequence blocks (CSBs) are shown

(Delarbre et al. 2001). The lamprey control region is located between ND6 and CYTB instead of between the tRNA^{Pro} and the tRNA^{Phe} genes (Fig. 1). The lamprey control region is also distinctive in that it is divided into two noncoding regions (Lee and Kocher 1995; Delarbre et al. 2000) by the tRNA^{Glu} and tRNA^{Thr} genes. Noncoding region 1 (NC1) is thought to be the control region due to the presence of conserved sequence blocks (CSB2 and CSB3) known to be important in the regulation of mtDNA replication (Lee and Kocher 1995). At the 5' end of NC1 there are three copies of a 39-base repeat whose secondary structures may be involved in regulation of D-loop DNA replication (Lee and Kocher 1995). Unlike other lamprey species, which have three copies of the 39-base repeat, the Least Brook lamprey, *Lampetra aepyptera*, has 4–6 copies (Martin and White 2008).

In this study, we assessed repeat variation in Least Brook lamprey NC1. We examined sequence variation among repeats and within and among populations and species. The geographic distribution of repeat variants was mapped on a phylogeny of populations sampled from throughout the distribution of the Least Brook lamprey. Patterns of variation among repeats and among populations and species allowed us to infer the mode of evolution of repeat number.

Materials and Methods

The Least Brook lamprey is a nonparasitic species with a distribution in the eastern United States. Specimens were collected from throughout the range (Table 1) and preserved in 95% ethanol. DNA was purified using DNeasy (Qiagen, Inc.). Polymerase chain reaction (PCR) was used to amplify the control region and ND3 gene of the lamprey mitochondrial DNA. The ND3 gene region was amplified using light-strand (5' ACG TGA ATT CTA TAG TTG GGT TCC AAC CA 3') and heavy-strand (5' ATG CGG ATC CTT TTG AGC CGA AAT CA 3') primers (Docker et al. 1999). NC1 (Fig. 1), from the ND6 gene to Glu-tRNA, was amplified with the primers LACR3F [(5' TACCCCATGCTACAAAATAACA 3') and LACR3R (5' CTGGTTTACAAGACCAGTGCTTT 3').

Polymerase chain reactions (PCRs) consisted of an initial denaturation step at 94°C for 3 min, followed by 30 cycles at 94°C for 30 s, 50°C for 30 s, and 72°C for 30 s. All amplification products were verified on 0.8% agarose gels. PCR products were purified using Qiaquick PCR purification kits (Qiagen Inc.). Between 5 and 30 ng of purified PCR product was used as template for the sequencing PCR. Sequences were determined on a 3730 DNA Analyzer (Applied Biosystems, Inc.). Between one and four individuals were sequenced from each population.

Chromatograms were read using FinchTV v1.3.0 (Geospiza, Inc.) and aligned for comparison using CLUSTALX v1.81 (Thompson et al. 1997). ND3 and control region sequences were aligned with *Lampetra fluviatilis* (GenBank accession number NC001131), the sister taxon of *L. aepyptera* (Docker et al. 1999). Due to the presence of the tandem repeats and some short indels, sequence alignments were further adjusted manually. Control-region sequences of four additional lamprey species, *L. richardsoni* (AY205574), *Entosphenus lethophagus* (AY205570), *E. similis* (AY205565), and *E. tridentatus* (AY205567), were included in the comparisons. Potential secondary structures of the tandem repeats and their free energies were evaluated using MFOLD (Zucker 2003). To test the significance of the free energy estimates, 20 random DNA sequences of lengths corresponding to different numbers of repeats (Table 3) were generated. Folding energies were estimated for each of the random sequences and means and standard deviations determined. Free energies for the lamprey repeats were compared to the estimates for the random sequences by a Z-test.

To evaluate the geographic and phylogenetic distribution of NC1 repeat numbers and motifs within *L. aepyptera*, ND3 sequences were used to construct a phylogenetic hypothesis. The ND3 gene region was chosen for construction of the phylogeny to avoid the possibility that sequence variation in other parts of the control region influenced repeat copy number. ND3 sequences for *L. richardsoni* and three species of *Entosphenus* were from Docker et al. (1999). Relationships among haplotypes were evaluated with maximum parsimony in MEGA 3.1 (Kumar et al. 2004) using the Close Neighbor Interchange option with random tree addition. Support for nodes and branches

Table 1 Sample collection numbers, locations, sample sizes, and voucher numbers of *Lampetra aepyptera* (samples 1–28), *L. appendix* (samples 29–30), and *Ichthyomyzon unicuspis* (sample 31)

Sample	County, state	Drainage	N	Voucher ^a
1. Davis Mill Creek	Dorchester, MD	Chesapeake	2	OUVVC11001
2. Neuse River	Johnston, NC	Atlantic	1	NCSM40904
3. Burnt Mill Creek	Wicomico, MD	Chesapeake	1	OUVVC11003
4. Tar Creek	McNairy, TN	Mississippi	4	OUVVC11004
5. Gaylor Creek	Hardeman, TN	Mississippi	3	OUVVC11005-07
6. Schultz Creek	Bibb, AL	Mobile	3	OUVVC11008
7. Terrapin Creek	Graves, KY	Mississippi	5	OUVVC11009-10
8. Middle Fork Obion	Henry, TN	Mississippi	1	OUVVC110011
9. Bear Creek	Henry, TN	Tennessee	1	UAIC11140.01
10. McCollough Fork	Calloway, KY	Tennessee	1	OUVVC110013
11. Weatherford Creek	Wayne, TN	Tennessee	1	UAIC10649.01
12. Little Bear Creek	Franklin, AL	Tennessee	1	UAIC11140.01
13. Strouds Run	Athens, OH	Ohio	1	OUVVC110016
14. Indian Run	Hocking, OH	Ohio	2	OUVVC110017
15. Big Sinking Creek	Carter, KY	Ohio	1	OUVVC110018
16. Big Caney Creek	Elliott, KY	Ohio	1	OUVVC 110019
17. Camp Creek	Pike, OH	Ohio	1	OUVVC 110020
18. Shade River	Athens, OH	Ohio	2	OUVVC 110021
19. Raccoon Creek	Athens, OH	Ohio	1	OUVVC 110025
20. Cooney Creek	Scioto, OH	Ohio	1	OUVVC 110026
21. Sunday Creek	Athens, OH	Ohio	1	OUVVC 110027
22. Turkey foot Run	Scioto, OH	Ohio	1	OUVVC 110028
23. Wolfpen Run	Athens, OH	Ohio	1	OUVVC 110029
24. Hocking River	Athens, OH	Ohio	1	OUVVC 110030
25. Long Creek	Gallia, OH	Ohio	1	OUVVC 110031
26. Little Black River	Ripley, MO	White	3	OUVVC 110022
27. Yellow Leaf Creek	Lafayette, MS	Yazoo	2	OUVVC 110023
28. Kettle Creek	Lafayette, MS	Yazoo	3	OUVVC 110024
29. Zekiah Swamp Run	Charles, MD	Potomac	1	UAIC 12263.09
30. Big Sycamore Creek	Claiborne, TN	Tennessee	2	UAIC 9838.13
31. Ohio River	Meigs, OH	Ohio	1	OUVVC 110025

^a OUVVC Ohio University Vertebrate Collection; NCSM North Carolina State Museum; UAIC University of Alabama Ichthyology Collection

was obtained by bootstrapping (Felsenstein 1985) with 1,000 pseudoreplicates. *Ichthyomyzon unicuspis* (DQ889805) was used as outgroup. We tested for phylogenetic signal in repeat number on the ND3 phylogeny using Mesquite v2.5 (Maddison and Maddison 2008). The tandem repeat number states were randomly shuffled on the terminal taxa of the 50% majority consensus tree. Ten thousand trees were simulated.

Results

The ND3 gene region is 351 nucleotides long. Sequences were obtained for 29 haplotypes from 47 individuals of

L. aepyptera (GenBank numbers DQ532785–DQ532804, EU350075–EU350082). The parsimony analysis, rooted to *Ichthyomyzon* (Fig. 2), identifies a monophyletic *L. aepyptera* with several well-supported clades whose relationships are poorly resolved. Average sequence divergence among *L. aepyptera* haplotypes was 3.5% (0.3–6.3%). Sequence divergence among taxa varied from 6.3% to 19.3% (Table 4).

Approximately 650 bases of NC1 were resolved in the same 47 individuals of *L. aepyptera*, 3e individuals of *L. appendix*, and 1 *Ichthyomyzon unicuspis*. There was no evidence from either sequencing or electrophoresis to suggest the presence of heteroplasmy. Electrophoresis yielded a single amplification product in all individuals,

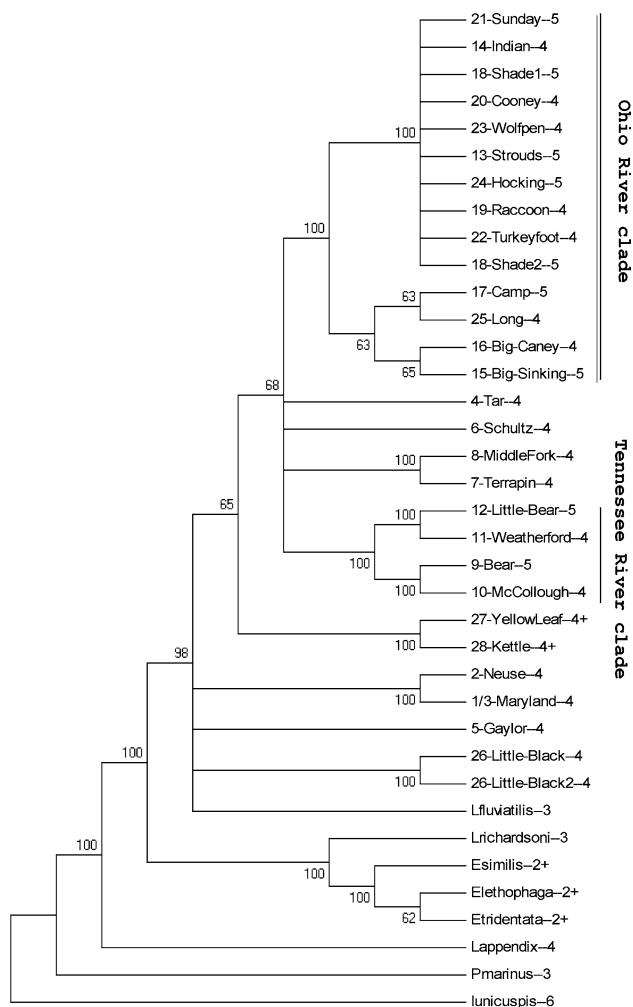


Fig. 2 Maximum parsimony analysis of 29 mitochondrial ND3 haplotypes of *Lampetra aepyptera*. The tree is rooted to *L. fluviatilis*. Numbers at the nodes are bootstrap support values. The number preceding a haplotype refers to the locality in Table 1. The number following a haplotype is the number of 39-base repeats in NC1 of the control region. A plus sign refers to a fragment of an additional copy

and the size of the amplification product varied with the number of tandem repeats. Tandem repeat sequence was similar to that observed in sea lamprey (*Petromyzon marinus* [Lee and Kocher 1995]) and identical to that in river lamprey [*Lampetra fluviatilis* (Delarbre et al. 2000)]. Each repeat was 39 bases long and included two copies of a 12-base subrepeat (5' ATGGCATAGGTA 3'; Table 2). *Lampetra aepyptera* had 4–6 copies of the repeat; *L. appendix*, four copies; *L. richardsoni*, three copies; *L. fluviatilis*, three copies; *Entosphenus* species, three copies; *P. marinus*, three copies; and *I. unicuspis*, six copies. We designated repeat copies as 5', internal, or 3' based on their position on the light strand relative to the ND6 gene.

The number of NC1 repeats was mapped onto the ND3-based phylogeny (Fig. 3). Most *aepyptera* populations displayed four copies. Some populations had a complete

fifth copy (Bear Creek, Hocking River, Sunday Creek, Shade River, Strouds Run, Camp Creek, and Big Sinking Creek). The Little Bear Creek population had six copies. Two populations, Yellowleaf Creek and Kettle Creek, in the Yazoo River drainage, had a fragment of a fifth copy in the 5' position. This copy fragment included the first 21 bases and was missing 6 bases plus the second 12-base subrepeat. The three species of *Entosphenus* also exhibited a fragment of the 5' copy, though this fragment sequence is different from that observed in the Yazoo River *L. aepyptera* fragment (Table 2). In *Entosphenus* there was also a 5-base sequence (ATACC or CTACC) in nucleotide positions 8–12 not seen in any other species. In the test for phylogenetic signal in the distribution of repeat number, 97.6% of the simulated trees were significantly different from the observed tree. This suggests that there is a phylogenetic signal of tandem repeat number distribution in the ND3 phylogeny.

In most *L. aepyptera*, the sequences of the 5' and internal copies of NC1 repeats were identical. Most instances of sequence variation were limited to the 3' copy, where a polymorphic site was detected in the second 12-base subrepeat at position 36. Thirty-four of 47 samples (72%) exhibited a G/A substitution at this site. A single fixed difference was observed in all complete repeat copies from the Yazoo River system (Table 2). Only two other substitutions were observed in any *L. aepyptera* repeat copies (position 25 in copy 3 in one individual from locality 7 and position 24 in copy 4 in one individual from locality 5). Single fixed differences were also seen in all copies from *L. appendix* and *L. richardsoni*. Two substitutions were observed in all repeats of *P. marinus*. Four substitutions were observed in all repeat copies of *I. unicuspis* (Table 2). Sequence divergences of the internal 39 base repeats (Table 2) between taxa varied from 2.5% to 12.8% (Table 4).

The MFOLD analysis indicated that the 39-base repeats are capable of stable secondary structures (Fig. 3). Different repeat copy numbers and a combination of whole and partial repeats were also capable of secondary structures (Table 3). Increasing the number of repeats increased the potential stability of the folding. All repeat sequence arrays produced free energy estimates significantly lower than random sequences ($P < 0.001$). The MFOLD analysis also identified a possible hairpin structure involving the first 17 bases of a repeat (Table 2).

Discussion

Variation in repeat copy number was observed among populations of *L. aepyptera*. Most haplotypes had four copies of the repeat. Haplotypes with five copies are

Table 2 Sequences of the complete 39-base repeat for populations and species of lamprey

Taxon	*																																														
<i>L. aepyptera</i>	<u>I</u>	<u>A</u>	<u>I</u>	<u>G</u>	<u>C</u>	<u>C</u>	T	C	T	A	T	<u>G</u>	<u>G</u>	<u>C</u>	<u>A</u>	<u>I</u>	<u>A</u>	G	G	T	A	T	A	T	A	T	A	A	T	G	G	C	A	T	A	G	G	T	A								
Yazoo River			
<i>L. appendix</i>			
<i>L. fluviatilis</i>			
<i>L. richardsoni</i>		
<i>E. similis</i>		
<i>E. tridentatus</i>	
<i>E. lethophaga</i>	
<i>I. unicuspis</i>	
<i>P. marinus</i>

Nucleotides identical to the standard *L. aepyptera* sequence are designated by a *period*. Yazoo River refers to haplotypes from localities 27 and 28 of *L. aepyptera*. The *asterisk* designates the polymorphic nucleotide site at position 36 in the 3' repeat copy. *Boldfaced* nucleotide sequences (CTATGCCAT and ATGGCATAG) and *underlined* nucleotide sequences (TATGCC and GGCATA) are capable of forming stem structures

distributed in two clades, Ohio River and Tennessee River (Fig. 2). Although localities with multiple samples were few, we did not observe any instances of intrapopulation variation in repeat number. Within the Ohio River clade, different streams had either four or five copies. It is possible that four or five repeat copies might be widespread throughout the range of *L. aepyptera*. Larger sample sizes at each locality might reveal instances of intrapopulation variation in copy number.

The test for phylogenetic signal of repeat number variation suggests that a signal is present. Thus repeat number is not randomly distributed on the phylogeny (Fig. 2). Several observations are worth noting. The ancestral condition (*P. marinus*) in the clade including *L. aepyptera* is three repeats. With the exception of the loss of a repeat fragment in the *L. richardsoni*/*Entosphenus* clade (3→2+), all instances of repeat variation appear to involve repeat gain (3→4→5, 4→4+, 4→6). Further, the *L. aepyptera* clades that exhibited repeat variation (Ohio River, Yazoo River, and Tennessee River clades) are geographically and genetically distant, suggesting independent origins for repeat gains (Purdue et al. 2006).

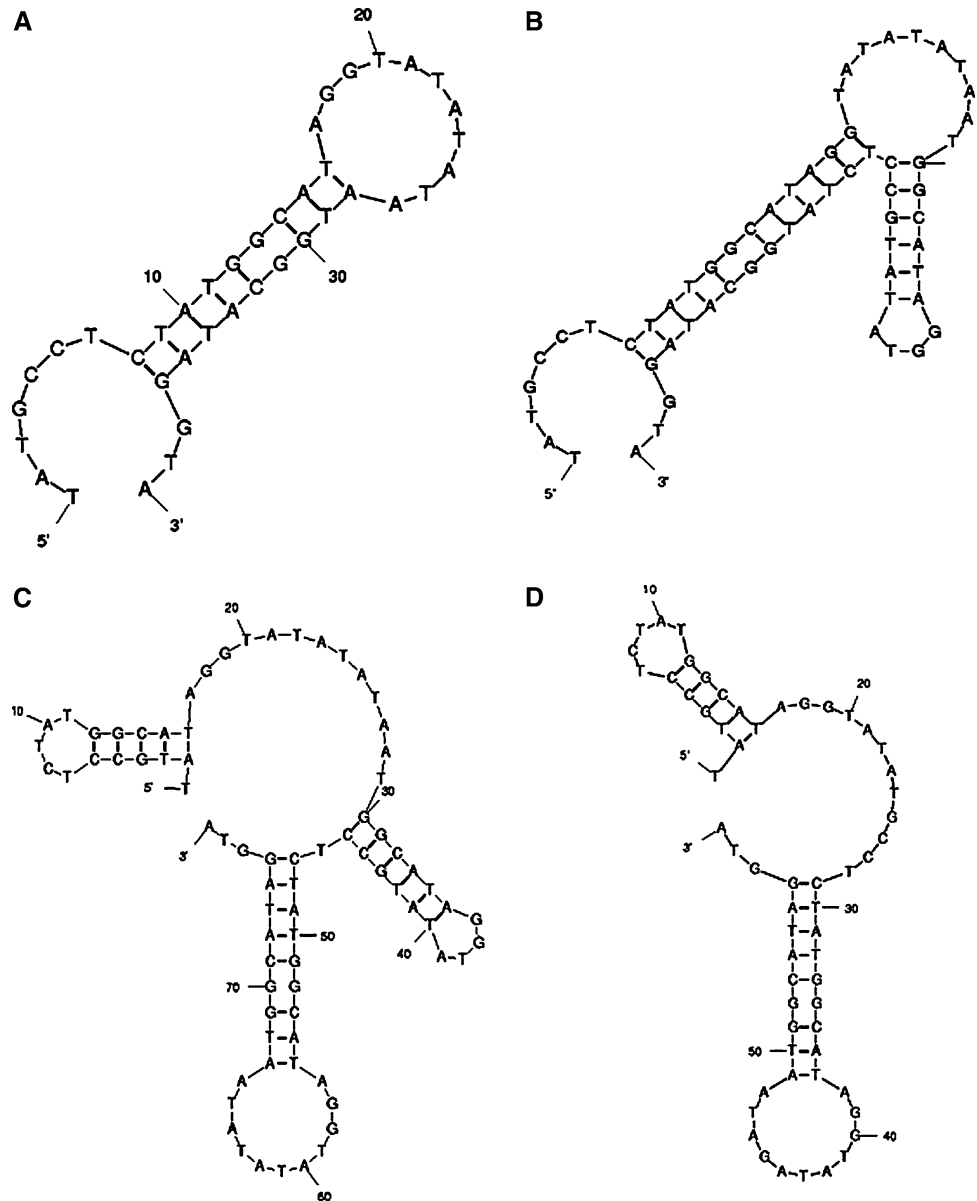
Petromyzon marinus has three copies of the repeat (Lee and Kocher 1995). *Lampetra fluviatilis*, the hypothesized sister-taxon of *L. aepyptera*, and *L. richardsoni* also have three copies. *Lampetra appendix*, a more distantly related species (Docker et al. 1999), has four. A minimum number of three repeats in the family would be consistent with one of the predictions of the illegitimate elongation model (Buroker et al. 1990) of repeat addition and loss. However, three species in the genus *Entosphenus* have only a fragment of a third copy in the 5' position, possibly due to a

loss of part of the 5' repeat. In addition, there are numerous examples of other taxa with fewer than three repeat copies (e.g., Lunt et al. 1998).

Because we did not detect heteroplasmy, the evolutionary rate of change in copy number may be slow. Differences in copy number between streams within drainages (Fig. 2) may also reflect limited gene flow and the effects of drift and lineage sorting. However, a low occurrence of heteroplasmy in some individuals and low frequencies of intrapopulation variation might not be detectable without large sample sizes.

Compared to other examples of larger tandem repeats (Broughton and Dowling 1997; Mundy and Helbig 1994; Wilkinson and Chapman 1991) there was remarkably little sequence variation among repeats within and between populations of *L. aepyptera*. Other than one fixed difference in the Yazoo River haplotypes and two single substitutions, sequence variation was restricted to the 3' repeat (position 36). This is similar to patterns of variation seen in a few other species [sturgeon (Buroker et al. 1990), shrews (Fumagalli et al. 1996), arawana (Yue et al. 2006)]. In the 3' repeat, a G was observed in 72% of the samples. Our data suggest that *L. appendix* does not have the polymorphic site but our sample size is small ($N = 3$). *Lampetra fluviatilis*, *L. richardsoni*, and *Entosphenus* species have the site but *Petromyzon* and *Ichthyomyzon* may not. The presence of the polymorphic site at position 36 only in the 3' repeat copy suggests that repeat addition or deletion occurs either at an internal position or at the 5' end. The substitution was never observed in an internal copy. Wilkinson and Chapman (1991) suggested that the direction of replication in the D-loop would result in repeat duplication at the end (5') of the

Fig. 3 Potential secondary structures for different numbers of 39-base repeats of *Lampetra aepyptera*. Unless otherwise noted, the sequence is from the Hocking River (locality 24). **a** One repeat (−5.9 kcal); **b** 1.5 repeats (−11.5 kcal); **c** 2 repeats (−15.8 kcal); **d** 1.5 Yazoo repeats (−11.1 kcal)



array. Further, the presence of the site in only position 36 of the 3' copy was observed in other species in the family. This suggests that this polymorphism is not random although a functional role can not yet be inferred.

Based on morphological characters, Gill et al. (2003) concluded that *Petromyzon* and *Ichthyomyzon* formed a monophyletic group sister to the other northern hemisphere lampreys (including *Lampetra*). However, we observed that the repeat sequence in *Petromyzon* differs by only two substitutions from *L. aepyptera* (Tables 2 and 4). The repeat sequence from *Ichthyomyzon* shows the greatest divergence. Based on ND3 sequences, *Ichthyomyzon* is more divergent from other lamprey taxa than *Petromyzon* (Table 4).

The high similarity in repeat sequences among populations, species, and genera of lampreys is notable. An

important process for determining copy number and variation is the rate of copy addition/deletion (turnover rate) relative to the copy sequence mutation rate. If the turnover rate is low, copies should diverge and we should see sequence variation among copies, populations, and species (Rand 1994). High copy turnover should result in repeat sequence homogenization (Broughton and Dowling 1997). Although the lack of variation among so many of the repeat copies would suggest rapid turnover, the presence of copy fragments at the 5' end and the single substitution in the 3' copy does not support rapid turnover and homogenization. However, repeat copy turnover may be more frequent in internal copies, resulting in homogenization of the internal repeats (Rand 1994). Copy number is correlated with heteroplasmy and repeat sequence variation in bats

Table 3 Free energy estimates (kcal) of folded secondary structures for combinations of tandem repeats in noncoding region I of *Lampetra aepyptera*

Repeat array	Size (bp)	Repeat free energy	Mean random sequence free energy (SD)
Hocking River, OH 1 repeat	39	−5.9	−2.3 (1.5)
1.5 repeats	60	−11.5	−4.5 (2.0)
Yazoo 1.5 repeats	60	−11.1	−4.5 (2.0)
2 repeats	78	−15.8	−5.9 (2.8)
3 repeats	117	−24.2	10.7 (3.7)
4 repeats	156	−28.6	−15.5 (3.6)
4.5 repeats	176	−35.5	−17.3 (4.0)
5 repeats	195	−38.8	19.0 (4.6)
Little Bear Creek 6 repeats	234	−47.2	−23.9 (4.6)

Notes: Hocking River, OH, refers to locality 24. Yazoo refers to locality 28. Little Bear Creek refers to locality 12. The half repeat includes the first 21 nucleotides of a second repeat. All free energy estimates for the lamprey repeat arrays were significantly lower than free energies estimated for random DNA sequences ($P < 0.001$)

Table 4 Percentage sequence divergence between lamprey taxa: values below the diagonal are divergences at the mitochondrial ND3 gene; those above are divergences at internal copies of the 39-bp repeat in noncoding region I

Taxon	1	2	3	4	5	6	7
1. <i>Lampetra aepyptera</i>		0	2.5	2.5	5.1	5.1	10.0
2. <i>L. fluviatilis</i>	6.3		2.5	2.5	5.1	5.1	10.0
3. <i>L. appendix</i>	8.2	9.1		5.1	2.5	7.7	7.7
4. <i>L. richardsoni</i>	9.8	9.4	10.5		7.7	7.7	12.8
5. <i>Entosphenus</i>	10.7	11.5	11.5	10.0		10.0	5.1
6. <i>Petromyzon</i>	14.8	15.9	15.6	16.8	17.2		5.1
7. <i>Ichthyomyzon</i>	17.2	18.8	14.5	19.3	17.3	17.3	

(Wilkinson et al. 1997). In the absence of detectable heteroplasmy and the paucity of intra- and interspecific repeat sequence variation, the applicability of this relationship to lampreys is limited. It is likely that the number of repeats is a balance between length mutation rate and selection (Wilkinson et al. 1997).

Alternatively, these repeats could be under functional constraints themselves or near sequences that are selectively constrained. The control region is the location of sequences that regulate replication and transcription of the mtDNA molecule. In mammals, the light-strand promoter is usually found near the CSB sequences and the heavy-strand promoters between the nearby 12S RNA gene and the start of the control region (Bonawitz et al. 2006). Both CSB-II and CSB-III have been observed in the lamprey control region (Lee and Kocher 1995; this study) but at the 3' end of NC1. We observed a sequence similar to CSB-I 58 bases upstream from CSB-II. There was no evidence of a light-strand promoter sequence upstream of the CSBs where it is observed in humans (Pham et al. 2006). However, the mammalian consensus promoter sequence likely shows little similarity to the lamprey sequence. Further, the

structure of the lamprey control region and the gene order are very different from the consensus vertebrate pattern (Lee and Kocher 1995). Although almost nothing is known about transcription initiation in lampreys, the lamprey repeat sequences could be involved in transcription initiation by helping to stabilize unwinding of the promoter region (e.g., Asin-Cayuela and Gustafsson 2007).

Typically, noncoding regions show higher levels of divergence than coding regions often thought to be due to constraints on protein structure and function. We compared divergences among taxa at the ND3 gene and the sequence from an internal repeat (Table 4). Divergences among the repeats were always less than those observed among ND3 sequences, suggesting that there may be functional constraints on the repeat sequence. Alternatively, concerted evolution may be homogenizing the repeats (Tatarenkov and Avise 2007). However, repeat sequence divergences between orthologous repeats are almost identical to divergences between paralogous repeats, an observation inconsistent with the predictions of concerted evolution.

The Yazoo River haplotypes possess a fragment of a repeat at the 5' end. If repeats are only capable of folding in multiples of complete (39 base) repeats, we would not expect to see partial repeats. However, it is possible for incomplete repeats to be involved in stable secondary structures (Fig. 3b, d and Table 2). Since each 39-base repeat includes two 12-base subrepeats, each subrepeat could conceivably pair with any other. If 12-base subrepeats from different 39-base repeats were to pair, there would be the potential to gain or lose parts of repeats. A gain would occur when the newly synthesized D-loop strand looped; a loss would occur when the light strand looped (Buroker et al., 1990). Thus, the Yazoo River haplotypes resulted from the gain of a 21-base fragment or the loss of an 18-base fragment. The missing piece of that repeat does include the second 12-base subrepeat. The

fragment in the 5' copy in *Entosphenus* is missing most of the first 12-base subrepeat and likely arose from a different base pairing than the Yazoo River repeat. The MFOLD analysis also detected a potential hairpin in the first 16 bases of each repeat (Table 2) and between bases 30–35 of one repeat and bases 1–6 of the following repeat (Fig. 3d). This base-pairing complexity could contribute to repeat fragment formation.

Determination of whether additions or deletions occur at the 5' end or internally would require detection of certain repeat combinations. In those populations or species that possess a fragment of the 5' repeat, can we observe repeat number variation? If there is a gain or loss of a complete repeat, it would suggest that turnover involves the internal copies. If additional fragmentary repeats are observed, then additions or losses might also occur at the 5' end. Larger samples in the Yazoo River *L. aepyptera* populations and in *Entosphenus* species might detect such variation.

The NC1 repeats are capable of achieving a secondary structure, either singly or in combinations. A common characteristic of these structures is the presence of the sequence 5' TATATAA 3' (5' TAGATAA 3' in Yazoo River haplotypes) in one of the loops (Fig. 3a, b). This sequence is found in all repeats and lies between the two 12-base subrepeats. It is similar to sequences identified by Doda et al. (1981) and Wilkinson and Chapman (1991) as termination-associated sequences (TASs). Wilkinson and Chapman (1991) also detected these sequences in the loops of repeat secondary structures. These sequences direct termination of D-loop synthesis (Doda et al. 1981), which begins near CSB2 and CSB3. In *L. aepyptera*, CSB2 and CSB3 are located near the 3' end of NC1. Thus, like some other species with larger tandem repeats (e.g., Buroker et al. 1990; Wilkinson and Chapman 1991; Starner et al. 2004), each lamprey repeat may contain a TAS sequence. These multiple TASs would result in the synthesis of multiple D-loop strands that differ in size (Doda et al. 1981).

Several models have been proposed for repeat number change. The heavy-strand model of repeat expansion suggests that repeats are added at the 5' end of the repeat array (Broughton and Dowling 1997); however, the repeats in their study were located at the 3' end of the control region. Mundy and Helbig (1994) suggested that repeat expansion in shrikes (*Lanius*) occurred at the 5' end and was made possible by a secondary structure that formed at an inverted repeat sequence 3' of the tandem repeats. No equivalent inverted repeat was observed in lampreys. However, sequences in NC2, located 3' of NC1 (Fig. 1), are capable of secondary structures and could function in this capacity. Further, NC1 and NC2 are separated by two tRNA genes. These could also stabilize repeat slippage in NC1 and allow for the gain or loss of repeats at the 5' end.

Recombination can generate tandem repeat variation (Hoarau et al. 2002; Mjelle et al. 2008). In these cases, recombination was detected in species with extensive heteroplasmy and inter-repeat sequence variation. Hoarau et al. (2002) suggested that recombination would be difficult to detect when the tandem repeat arrays are perfect. Although we cannot rule out a role for recombination in lamprey repeats, the very low levels of repeat sequence variation preclude a test of this hypothesis. We believe that the structure, position, and variation of the NC1 repeats in lampreys are consistent with the illegitimate elongation model (Buroker et al. 1990). Similar to sturgeon, lamprey repeats are located in the 5' end of the control region, the repeats can achieve a secondary structure, and each repeat has an apparent TAS. Buroker et al. (1990) predicted that the minimum number of repeat copies would be three and that the internal copies should be perfectly conserved. They also predicted that the copies at each end could diverge. Despite the very limited divergence observed at 3' and 5' copies, and the presence of repeat fragments at the 5' end in two clades, our data are largely consistent with these predictions. However, the sequence homogeneity, both between repeats and among populations, species, and genera, suggests that these repeats may be under functional constraints, the nature of which is not yet known.

Acknowledgments A. Drake, D. Elliott, B. Martin, S. Reilly, D. Eisenhour, P. Lienesch, S. Adams, R. Mayden, B. Kujajida, and W. Starnes kindly provided specimens. We would like to thank the Kentucky Nature Preserves Commission, Kentucky Department of Natural Resources, Tennessee Wildlife Resources Agency, and Ohio Department of Natural Resources for collecting permits. R. Broughton, J. Faber, and two anonymous reviewers provided valuable comments on early versions of the paper. We thank D. Robinson for assistance with Mesquite. This work was conducted under Ohio University IACUC permit no. L98-08. R. Burdick, J. Rybka, M. Peters, M. Hazel, J. Keith, and J. Kunkle provided laboratory support. This research was supported by the Department of Biological Sciences at Ohio University, the Program in Molecular and Cellular Biology, the Ohio University Research Council, and an Ohio University Graduate Student Senate grant.

References

- Arnason E, Rand DM (1992) Heteroplasmy of short tandem repeats in mitochondrial DNA of Atlantic cod, *Gadus morhua*. *Genetics* 132:211–220
- Asin-Cayuela J, Gustafsson CM (2007) Mitochondrial transcription and its regulation in mammalian cells. *Trends Biochem Sci* 32:111–117
- Bonawitz ND, Clayton DA, Shade GS (2006) Initiation and beyond: multiple functions of the human mitochondrial transcription machinery. *Mol Cell* 24:813–825
- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27:1767–1780
- Broughton RE, Dowling TE (1994) Length variation in mitochondrial DNA of the minnow *Cyprinella spiloptera*. *Genetics* 138:179–190

- Broughton RE, Dowling TE (1997) Evolutionary dynamics of tandem repeats in the mitochondrial DNA control region of the minnow *Cyprinella spiloptera*. *Mol Biol Evol* 14:1187–1196
- Buroker NE, 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
- Delarbre C, Escriva H, Gallut C, Barriol G, Kourilsky P, Janvier P, Laudet V, Gachelin G (2000) The complete nucleotide sequence of the mitochondrial DNA of the agnathan *Lampetra fluviatilis*: bearings on the phylogeny of the cyclostomes. *Mol Biol Evol* 17:519–529
- Delarbre C, Rasmussen A, Arnason U, Gachelin G (2001) The complete mitochondrial genome of the hagfish *Myxine glutinosa*: unique features of the control region. *J Mol Evol* 53:634–641
- Docker MF, Youson JH, Beamish RJ, Devlin RH (1999) Phylogeny of the lamprey genus *Lampetra* inferred from mitochondrial cytochrome *b* and ND3 gene sequences. *Can J Fish Aquat Sci* 56:2340–2349
- Doda JN, Wright CT, Clayton DA (1981) Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proc Natl Acad Sci USA* 78:6116–6120
- Faber JE, Stepien CA (1998) Tandemly repeated sequences in the mitochondrial DNA control region and phylogeography of the pike-perches *Stizostedion*. *Mol Phylogenet Evol* 10:310–322
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Fumagalli L, Taberlet P, Favre L, Hausser J (1996) Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. *Mol Biol Evol* 13:31–46
- Gill HS, Renaud CB, Chapleau F, Mayden RL, Potter IC (2003) Phylogeny of living parasitic lampreys (Petromyzontiformes) based on morphological data. *Copeia* 2003:687–703
- Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL (2002) Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish, *Platichthys flesus*. *Mol Biol Evol* 19:2261–2264
- Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163
- Lee W, Kocher TD (1995) Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. *Genetics* 139:873–887
- Ludwig A, May B, Debus L, Jenneckens I (2000) Heteroplasmy in the mtDNA control region of sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). *Genetics* 156:1933–1947
- Lunt DH, Whipple LE, Hyman BC (1998) Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol Ecol* 7:1441–1455
- Maddison WP, Maddison DR (2008) Mesquite: a modular system for evolutionary analysis. Version 2.5. Available at: <http://mesquiteproject.org>
- Martin HR, White MM (2008) Intraspecific phylogeography of the Least Brook lamprey (*Lampetra aepyptera*). *Copeia* 2008:579–585
- McMillan WO, Palumbi SR (1997) Rapid rate of control region evolution in Pacific butterflyfishes (Chaetodontidae). *J Mol Evol* 45:473–484
- Meyer A (1993) Evolution of mitochondrial DNA in fishes. In: Hochachka PW, Mommsen P (eds) *The biochemistry and molecular biology of fishes*. Elsevier Press, Amsterdam, pp 1–38
- Mjelle KA, Karlson BO, Jorgensen TE, Moum T, Johansen SD (2008) Halibut mitochondrial genomes contain excessive heteroplasmic tandem repeat arrays involved in DNA recombination. *BMC Genomics* 9:10
- Mundy NI, Helbig AJ (1994) Origin and evolution of tandem repeats in the mitochondrial DNA control region of shrikes (*Lanius* spp.). *J Mol Evol* 59:250–257
- Pham XH, Farge G, Shi Y, Gaspari M, Gustafsson CM, Falkenberg M (2006) Conserved sequence box II directs transcription termination and primer formation in mitochondria. *J Biol Chem* 281:24647–24652
- Piganeau G, Gardner M, Eyre-Walker A (2004) A broad survey of recombination in animal mitochondria. *Mol Biol Evol* 21:2319–2325
- Purdue JR, Oleksyk TK, Smith MH (2006) Independent origins of multiple repeats in the control region of mitochondrial DNA of white-tailed deer. *J Hered* 97:235–243
- Rand DM (1994) Concerted evolution and RAPPING in mitochondrial VNTRs and the molecular geography of cricket populations. In: Schierwater B, Streit B, Wagner GP, DeSalle R (eds) *Molecular ecology and evolution: approaches and applications*. Birkhauser Verlag, Basel, pp 227–245
- Saccone C, Attimonelli M, Sbisà E (1987) Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. *J Mol Evol* 26:203–211
- Starmer H, Pahlsson C, Linden M (2004) Tandem repeat polymorphism and heteroplasmy in the mitochondrial DNA control region of threespine stickleback (*Gasterosteus aculeatus*). *Behaviour* 141:1357–1369
- Tatarenkov A, Avise JC (2007) Rapid concerted evolution in animal mitochondrial DNA. *Proc R Soc B* 274:1795–1798
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24:4876–4882
- Wilkinson GS, Chapman AM (1991) Length and sequence variation in Evening bat D-loop mtDNA. *Genetics* 128:607–617
- Wilkinson GS, Mayer F, Kerth G, Petri B (1997) Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. *Genetics* 146:1035–1048
- Yue GH, Liew WC, Orban L (2006) The complete mitochondrial genome of a basal teleost, the Asian arowana (*Scleropages formosus*, Osteoglossidae). *BMC Genomics* 7:242
- Zucker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415