J. Appl. Ichthyol. 20 (2004), 1–7

ISSN 0175–8659

Received: May 10, 2004
Accepted: August 12, 2004

Characterization of nuclear 18S rRNA gene sequence diversity and expression in an individual lake sturgeon (*Acipenser fulvescens*)

By J. Krieger[1] and P. A. Fuerst[2,3]

[1]*Department of Ecology and Evolutionary Biology, Brown University, Providence, RI; Departments of* [2]*Evolution, Ecology and Organismal Biology and* [3]*Molecular Genetics, The Ohio State University, Columbus, OH, USA*

Summary

The unexpected discovery of extensive intraindividual variation of the 18S rRNA gene in nine North American species of sturgeon (and its absence in the North American paddlefish) was described in a previous paper by Krieger and Fuerst (2002b). As part of the study, genomic 18S rRNA genes from a lake sturgeon (*Acipenser fulvescens*) individual were amplified by polymerase chain reaction (PCR), cloned and sequenced, resulting in the identification of at least 17 different sequence variants. Because of the number of variants detected and the improbability that such a large number of different alleles are required for survival, these sequence variants were subjected to further analyses (sequence comparison, phylogenetic and relative rate) to examine the possibility that some variants may be pseudogenes. The cDNA produced by reverse transcriptase (RT)-PCR amplification of 18S rRNA isolated from lake sturgeon liver tissue was also sequenced. The topology of the phylogenetic tree produced suggests a split of the sequence variants into two groups: paddlefish-like (PL) alleles and non-paddlefish-like (NPL) alleles. Relative rate comparisons of these two groups indicate that the alleles within the NPL allele group are evolving more quickly than those in the paddlefish-like (PL) group, and so are likely under relaxed functional constraints. Less than one-third of nucleotide changes in the non-paddlefish-like (NPL) alleles occur at positions that are considered to be highly conserved in a universal eukaryotic model of rRNA structure. No substitutions at eukaryotic universally conserved sites occur in the PL alleles. Analysis of the sequence of the RT-PCR product shows that a single lake sturgeon 18S rDNA sequence variant is expressed in major quantities in the liver, and that this single product is identical to the allele most similar in sequence to the paddlefish. These results provide support for the idea that many variants are unimportant for proper cellular function or are pseudogenes. Although concerted evolution has apparently failed to homogenize the many rDNA sequence variants found within sturgeon, the detection of a single major expressed sequence variant in lake sturgeon indicates that the expression of the variants is in fact under the control of selective factors.

Introduction

With rare exceptions, intraindividual variation in sequence of the rDNA genes is very unusual (Dame et al., 1984; Carranza et al., 1996; Muir et al., 2001). Concerted evolution is generally believed to homogenize the many copies of rDNA units often found in eukaryotic genomes within individuals as well as within species (Elder and Turner, 1995). We first described intraindividual variation for the 18S rRNA gene in nine species of sturgeon (Krieger and Fuerst, 2002b), discovered while conducting a molecular phylogenetic study of the North American Acipenseriformes. Polymerase chain reaction (PCR) products of the 18S rRNA gene were examined for variation in the North American paddlefish (Family Polyodontidae) and nine North American sturgeon species (Family Acipenseridae). The presence of sequence variation within individual sturgeon of all species studied prevented the determination of complete gene sequences from the original PCR products. As part of the same study, in order to investigate the amount and degree of 18S rDNA variation that existed in an individual, 17 18S rRNA gene sequence variants were isolated from an individual lake sturgeon (*Acipenser fulvescens*) and characterized by PCR amplification, cloning, and DNA sequencing. It seemed unlikely that this large number of sequence variants would be biologically necessary for sturgeon, especially as only one sequence was apparently needed in paddlefish. This prompted additional investigation into this unusual phenomenon in sturgeon. We present in this paper, further comparison and analysis of these lake sturgeon variant sequences, as well as a relative rate test and reverse transcriptase (RT)-PCR experiments with the aim of exploring the possibility that some of these sequence variants are pseudogenes.

Materials and methods

Samples

Lake sturgeon tissue samples (fin snips for DNA analysis and liver for RNA analysis) were collected from fish from the Wolf River in Wisconsin with the help of researchers from the Wisconsin Department of Natural Resources and the Ohio State University.

Sequence alignment and analysis

The 17 lake sturgeon 18S rRNA gene sequence variants (GenBank accession numbers AF188384–AF188400) were aligned with the paddlefish (*Polyodon spathula*) (accession number AF188371) and *Polypterus senegalus* (accession number AF188367) sequences by eye using the program ESEE (Cabot and Beckenbach, 1989) and the aid of secondary structure information for the 18S rRNA of *Polyo. spathula* based on currently accepted models (Gutell, 1994; Cannone et al., 2002). All sequences included in these phylogenetic analyses were previously determined and reported in Krieger and Fuerst (2002b). To examine the relationships among lake sturgeon 18S rRNA intraindividual gene sequence variants,

JAI 6 1 0 B
Journal Name | Manuscript No.
Dispatch: 21.9.04 | Journal: JAI | CE: Benjamin
Author Received: | No. of pages: 7 | PE: K.S. Kavitha

phylogenetic analyses were conducted using the sequence from *Polyp. sengalus* as the outgroup. The program MODELTEST 3.06 (Posada and Crandall, 1998) was used to optimize a model of DNA evolution to the 18S rRNA gene data set, and PAUP* v.4.0b10 (Swofford, 2002) was utilized for phylogenetic reconstruction using the maximum likelihood method based on that model and the associated parameters.

The significance of any difference in the relative rates of sequence evolution between the two lake sturgeon sequence variant-type groups as discovered in the phylogenetic analyses were evaluated using the two-cluster relative rate test of Takezaki et al. (1995) as implemented in PHYLTEST (Kumar, 1996). In the comparison, the *Polyp. senegalus* 18S rRNA gene sequence was used as the outgroup and the Kimura (1980) two-parameter distance was used as the measure of sequence divergence. PHYLTEST (Kumar, 1996) was also used to conduct a two-tailed normal deviate test (Takezaki et al., 1995) for the relative rate comparison to determine if the difference in evolutionary rates between the two variant groups was statistically different from zero at the 5% level.

To estimate the potential number of different allelic sequences that occur in the strurgeon genome, a resampling simulation was performed. Resampling (randomization without replacement) was used to estimate the mean number of alleles detected for sample sizes N = 2, 3,..., 30 clones, based on the observed number of copies of alleles, and the distribution of the number of copies observed in the sample of 30 clones. Each resampling estimate was based on 100 iterations. The data was then fit using an asymptotic regression equation to the mean values obtained from resampling, to estimate the asymptotic number of alleles. Specifically, the equation used was $y = b1 + b2 \times \exp(-b3 \times x)$, where y is the number of alleles, x is the number of clones, $b1$ is the asymptotic number of alleles, $b2$ is replaced with $-b1$ (to force the equation through the origin), and $b3$ is the decay constant.

RNA extraction, RT-PCR, and cDNA sequencing

Total cellular RNA was extracted from fresh lake sturgeon (*A. fulvescens*) liver tissue using the SNAP^{TM}. Total RNA Isolation Kit, utilizing a spin column purification method and digestion with DNase to ensure that the final extraction was not contaminated with genomic DNA (Invitrogen, Carlsbad, CA). Extracted RNA was quantified using a spectrophotometer and stored at −20°C for a maximum of 5 days before carrying out RT-PCR for 18S rRNA. Primers were developed (18SAf-F: ttgtctcaaagattaagccatg and 18SAf-R: acctacggaaaccttgttacg) internal to the 5′- and 3′-ends of 18S rRNA sequence that were complimentary to all of the 18S rRNA gene variants characterized from lake sturgeon in an effort to amplify any sequence variants present in the total cellular RNA. RT-PCR was carried out in 50 μl reactions using the ProSTAR^{TM} HF Single Tube RT-PCR System (Stratagene, La Jolla, CA), with the following conditions: 5 min at 65°C, addition of reverse transcriptase and DNA polymerase enzymes, then 45 min at 45°C, 1 min at 95°C, followed by 40 cycles of 1 min at 95°C, 2 min at 55°C, 3 min at 68°C and finally one cycle of 10 min at 68°C. Dimethyl sulfoxide (DMSO) was added to the reactions at 1 or 2% concentration to destabilize ribosomal RNA secondary structure (Varadaraj and Skinner, 1994) so that the reaction could proceed at the relatively low temperature (45°C) required for the MMLV RT enzyme to function. Negative control RT-PCR reactions lacking the reverse transcriptase enzyme were carried out to test for DNA contamination. Lake sturgeon 18S rRNA RT-PCR products from four separate reactions were band isolated using a SNAP^{TM} nucleotide gel purification kit (Invitrogen) and then pooled for sequencing. DNA sequencing of the RT-PCR products was carried out as described in Krieger and Fuerst (2002b). The sequence was scored manually, recorded and has been deposited in GenBank (accession number AF198113).

Results

As described in Krieger and Fuerst (2002b), a minimum of 17 different gene sequence variants were present in the 30 clones isolated from a lake sturgeon individual. In order to characterize the types and amounts of differences that exist among sequence variants within a lake sturgeon individual and search for evidence of 18S rDNA pseudogene sequences, these 17 sequences were aligned with the paddlefish sequence and compared.

Phylogenetic analyses

New phylogenetic analyses of the lake sturgeon and paddlefish 18S rRNA gene sequences using *Polyp. senegalus* as the outgroup were carried out to determine if there were any patterns of relationship among lake sturgeon clone sequences. If mutational events were occurring independently in different members of a multigene array, we would expect little phylogenetic structure to appear in such an analysis. To reconstruct the phylogenetic history of the lake sturgeon alleles with greatest accuracy, we determined the model of DNA evolution that best fits the data based on hierarchical likelihood ratio tests in MODELTEST 3.06 (Posada and Crandall, 1998). The best-fit model of sequence change was determined to be the TIMef + I + G model, a variant of the General Time Reversible model (Rodríguez et al., 1990). This model was then used in a maximum likelihood analysis in PAUP* v4.0b10 with the following associated parameters: equal base frequencies; proportion of invariable sites (I) = 0.6222; gamma distribution shape parameter = 0.7875; and the rate matrix of R(a) [A-C] = 1.0000, R(b) [A-G] = 2.3218, R(c) [A-T] = 0.4849, R(d) [C-G] = 0.4849, R(e) [C-T] = 3.9304, R(f) [G-T] = 1.0000.

Figure 1 shows the maximum likelihood tree that was obtained from our analysis, including bootstrap support values resulting from 1000 replications. The larger cluster found in this tree includes 15 of the 17 lake sturgeon clones. There is little significant structure within this major clade, as reflected by low bootstrap support at most branch points, and there is no suggestion from the structure of the tree that variants are not being independently derived from various members of a large array of repeated sequences. The primary clade of lake sturgeon alleles is significantly differentiated from a smaller group that includes the paddlefish and two remaining lake sturgeon clones. The two lake sturgeon clones that cluster with the paddlefish sequence are Af-7 and Af-13: clone Af-7 has a single base difference from paddlefish, while Af-13 has six differences when compared with paddlefish. This division among the lake sturgeon sequences suggests a classification of lake sturgeon sequence variants into PL and NPL allele types. It also prompts consideration of the possibility that the lake sturgeon sequence(s) most closely related to the presumably functional paddlefish sequence have retained function, while it is possible that some or all of the NPL sequences may be non-essential or

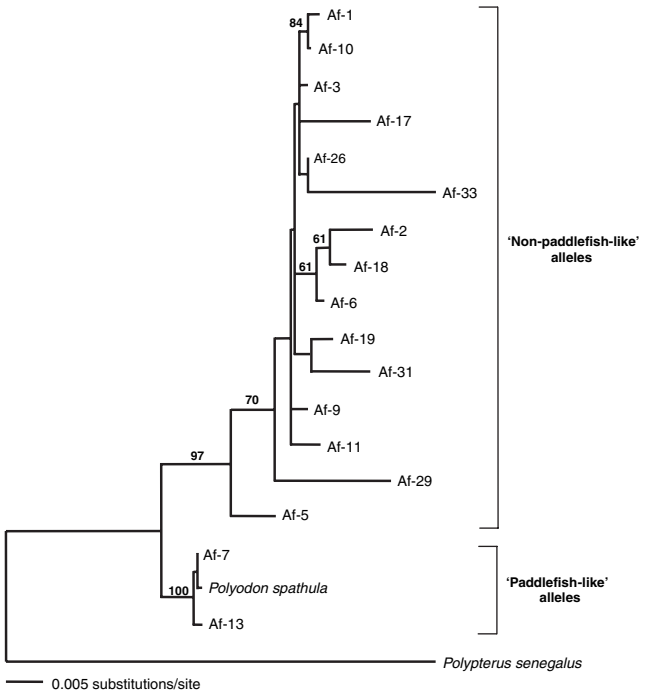

Fig. 1. Phylogenetic tree based on 18S rRNA gene sequences (1786 sites) constructed using the maximum likelihood method in PAUP* v.4.0b10 (Swofford, 2002), with *Polypterus senegalus* as the designated outgroup. Parameters used to obtain the maximum likelihood tree were determined using MODELTEST 3.06 (Posada and Crandall, 1998). Sequences labeled Af-n represent the 18S rDNA clone sequences isolated from lake sturgeon (*Acipenser fulvescens*). Numbers at branch points indicate the bootstrap values for 1000 replications, and those nodes without numbers have bootstrap values $< 50\%$

pseudogenes. Alternative phylogenetic analyses of trees produced by maximum parsimony and distance methods in PAUP* v.4.0b10 (Swofford, 2002) (not shown) indicate the presence of the same two main clusters of sequences.

Given that we have observed 17 alleles among 30 lake sturgeon rRNA clones, we speculated about the number of alleles that are present within the genome of lake sturgeon. To get a rough estimate of the number of alleles that may exist, we used the frequency of repeated occurrence of the same sequence in different to estimate the number of potential different alleles that would exist in a large random set of clones from lake sturgeon. The distribution of the number of clones potentially sharing the same sequence variants was: one sequence type was represented by six clones; a second sequence type was represented by four clones; two different sequence types were found in three clones each; a fourth was found in two clones, and the remaining 12 sequence types were unique. A resampling procedure was used to obtain an asymptotic estimate for the number of alleles in the large sample of clones. The asymptotic estimate of the number of alleles was 23.57, with a 95% confidence interval ranging from 23.063 to 24.076. The final fitted equation obtained by resampling was $y = 23.570 - 23.570 \times \exp(-0.0416 \times x)$, with a fitted $r^2 = 0.999$. This indicates that the number of alleles actually present in the lake sturgeon genome may not have been much greater than the number of alleles that we observed.

Analysis of rRNA structural factors

There are 1783 sites within the paddlefish + sturgeon alignment. The number of differences among various lake sturgeon sequence variants ranged from 6 to 72. There are 135 variable sites because of either substitutions or occasionally substitutions and indels (three sites). Four of the sites had three different nucleotides occurring in the collection of sequences. Twenty-eight sites involve indel events only. Discounting indel differences, the number of nucleotide differences (and proportional differences) between pairs of lake sturgeon variants ranged from 3 (0.002) to 59 (0.034) nucleotides. Among base substitutions, transitions were more common than transversions by 3 : 1 (104 : 35), with C↔T transitions (57) being the most common types of change. This is in agreement with a study of rates and patterns of base change in small subunit rRNAs by Vawter and Brown (1993), who also found that C↔T transitions were the most common type of change. The number of A↔G transitions in the lake sturgeon variants was 47. Transversions occurred in the following frequencies: 13 A↔C, nine G↔T, seven A↔T and six G↔C. The total number of base substitutions (139) is larger than the total number of variable sites as a result of substitutions (135) because occasionally more than one type of base substitution was found at a particular site.

The variable sites in lake sturgeon were relatively evenly distributed throughout the 18S rRNA gene. The nucleotides of the 18S rRNA may be classified into several categories, based on their location within the secondary structure of the molecule. The structural classification scheme used here is adapted from Cannone et al. (2002). 'Stems' are base paired regions that form helices consisting of two or more consecutive base pairs. 'Hairpin loops' are single stranded regions capping the end of a single helix. 'Internal loops' are unpaired bases within stem regions. 'Other regions' consist of long single-stranded sections that interact with ribosomal proteins (Woese et al., 1983). A final class used for the analysis was 'unknown regions', which are those segments of the rRNA molecule for which an accepted secondary structure has not been determined.

The distribution of variable sites in the sample of 30 cloned sequences that were found in the specific structural region classes of the molecule are: 43 sites in hairpin loop regions (26.4% of all changes); 54 sites in stem regions (33.1% of all changes); 28 sites in internal loop regions (17.1% of all changes); 22 sites in other regions (13.5% of all changes); and 16 changes are located in unknown regions (10.0% of all changes). If the number of variable sites is considered relative to the total number of sites of the standard rRNA structure classified into a particular structural type, however, hairpin loop regions contain the largest relative proportion of sites that are variable (18.1%), followed by internal loop regions (13.7%), unknown regions (12.5%), other regions (9.1%), and stem regions (5.5%). Analysis of these results, using a chi-square test of independence in a 2×5 table, indicates that there is significant heterogeneity among the five classes of sites in the percentage of variable positions ($\chi^2 = 42.4$ with 4 d.f., $P < 0.0001$). This is consistent with hypotheses that relate the importance of structural characteristics with the rate of change in regions of the rRNA molecule. The stem and other groupings possess the smallest percentage of variable sites, and these sequences are considered crucial for secondary structure and protein interactions, respectively. The stem segments are significantly less variable all other structural types. At the other extreme, hairpin loop regions are significantly more variable than either stem or other important single-stranded regions of the rRNA molecule. Finally, the great majority of sites identified with indel events (24 of 28) are

located in the two structural classes with the most variability overall, hairpin loop, and internal loop regions.

Another aspect of structure that can be examined with respect to the occurrence of sequence changes in lake sturgeon alleles is the preservation of complementary pairing within stem regions. Within stem regions there were 44 instances in the 17 clone sequences in which a nucleotide change maintained base pairing. These were distributed among 22 sites. This compares with approximately 90 instances of variation at 30 sites at which nucleotide changes destroyed base pairing. Only a single pair of sites showed compensatory changes on both strands of a stem that maintained base pairing, with compensatory pairing at this pair of sites occurring in 11 of the 17 sequence variants (change being classified relative to the paddlefish sequence). At this same pair of sites, two of the remaining variants had a change at one site that produced G-U pairing in the rRNA, while the other four variants showed no change at either site (compared with paddlefish) and thus were base paired.

None of the sequence variants have extensive regions of non-homology or indels compared with the paddlefish sequence that would immediately suggest that they are pseudogenes. The conclusion of this first analysis of the distribution of changes with respect to structural characteristics suggests that changes are under the control of natural selection and that they have occurred in a pattern that would be consistent with occurrence of changes that allow the continued functioning of the rRNA molecule.

In order to further examine the potential impact of sequence variation on structure and function, we next compared the locations of changes in the lake sturgeon alleles with the degree of site conservation in the generalized eukaryotic 18S rRNA secondary structure constructed by Cannone et al. (2002). The site conservation criteria are based on the phylogenetic comparison of 1939 eukaryotic small subunit rRNA sequences and identifies highly conserved sites within the secondary structure as well as regions of the molecule that are highly variable in either sequence and/or length. Sixty-three percentage of the sites that show variation in lake sturgeon alleles fall into site classes that are classified as either minimally conserved or non-conserved in the general eukaryotic structure. Minimally conserved sites are positions present in 95% of eukaryotic sequences but with sequence conserved in $<80\%$ of sequences. Non-conserved sites are structural positions present in $<95\%$ of eukaryotic sequences examined. Somewhat more than one-third of sites showing variation in the lake sturgeon alleles (37%) are at positions that were considered universally conserved, highly conserved, or conserved sites in the general eukaryotic structure. Universally conserved sites are present in 95% of sequences, with the same nucleotide present in $>98\%$ of sequences, while the other two classes are present in 95% of sequences with between 90 and 98% or 80 and 90% of sequences having the same nucleotide present, respectively. For comparison, the paddlefish sequence was found to differ from the sequence of the generalized eukaryotic phylogenetic conservation diagram (Cannone et al., 2002) at only one of the variable sites occurring in lake sturgeon, a site that is considered highly conserved (sequence identical in 90–98% of species examined). Twenty-two of the 28 indel sites in lake sturgeon occur in variable non-conserved regions (ones that tend to differ in length and sequence among taxa), but six of them were found in a highly conserved (sequence conserved in $>98\%$ of sequences examined) stem and hairpin loop. These six sites were found linked as a 6 bp deletion in one clone (Af-2), where the deletion probably has a significant effect on the secondary structure of the region, as it destroys the highly conserved 2 bp stem and hairpin loop.

We next determined the distribution of sites among the cloned lake sturgeon sequence variants falling into the five different classes of site conservation, following the scheme detailed above for phylogenetic conservation within the generalized eukaryotic rRNA structure diagram of Cannone et al. (2002). Table 1 shows this distribution. It is evident that on average the great majority of variable sites ($>68\%$) in any given clone sequence occur at sites that are highly variable among eukaryotic taxa. More than half of the lake sturgeon alleles show no variation at sites considered universally conserved. Only four cloned alleles have more than a single change in such highly conserved sites. The two alleles that show sequence most similar to the paddlefish sequence (Af-7 and Af-13) show no changes at sites that are considered to be conserved at any level $>80\%$. Close examination of the data in Table 1 suggests that as lake sturgeon allele sequences have diverged from the paddlefish sequence, they have gradually increased the percentage of changes in conserved positions. For example, the most divergent allele (Af-33, with 3.4% divergence from paddlefish) is also the allele with the most change in conserved regions. There is a significant positive relationship between the percentage of all changes for an allele that occur at conserved sites and the divergence from the paddlefish sequence, but the relationship, while still positive, is not significant if we exclude the two alleles Af-7 and Af-13.

Continuing the analysis of the relationship between structure, function and change, we examined the conservation of several regions that are believed important for the normal function of the small subunit rRNA within the ribosome in the process of translation. These regions were examined to obtain additional evidence about the active function of the allele sequences. We considered the portions of the rRNA molecule that make up the A-site (including the three universally conserved sites 530, 1492 and 1493 of the decoding region; see Mears et al., 2002), P-site and E-sites and helix 27 (numbering based in *Escherichia coli* sequence and structure). The

Table 1
Distribution among clone sequences of the lake sturgeon 18S rRNA gene for different classes of variable sites

	Nucleotide sites present in 95% of rRNA sequences in database				
	Degree of conservation				
Clone name	>98%	90–98%	80–90%	<80%	Non-conserved sites
Af-1		5	3	9	38
Af-2	5	5	3	9	35
Af-3		5	2	8	38
Af-5	3	4	1	8	30
Af-6		4	4	9	43
Af-7					1
Af-9		3	2	8	36
Af-10	1	5	2	9	37
Af-11	1	4	2	7	34
Af-13					6
Af-17	3	7	2	12	40
Af-18	1	4	5	10	38
Af-19			2	4	30
Af-26		4	2	8	39
Af-29		7	5	10	51
Af-31	1	5	2	8	28
Af-33	6	7	2	9	47

sequences that compose the A-, P- and E-sites in the ribosome, sites important for mRNA and tRNA interactions, are found in various regions throughout the molecule, and are brought together by secondary and tertiary structure (see Carter et al., 2000), and helix 27 is believed to be a conformational switch responsible for accuracy in translation. Comparing the various lake sturgeon alleles with the presumptively functional paddlefish sequence revealed no sequence variation in the stem region corresponding to helix 27 in any of the lake sturgeon alleles. The stem sequence was also consistent with that indicated by the eukaryotic phylogenetic conservation structure diagram (Cannone et al., 2002). Similarly, comparison of the A-, P- and E-site sequences from the lake sturgeon alleles with the paddlefish sequence found almost no variation in these regions, and sequences were again consistent with that shown in the eukaryotic phylogenetic conservation structure diagram. This shows a general pattern of sequence conservation in functional sites in the lake sturgeon 18S rRNA gene variants. Three exceptions occurred: one variable site was found in the A-site sequence and two variable sites were found in the P-site sequence. The variation at the A-site was found in a single clone (Af-5) and is located in a hairpin loop, so it does not disrupt base paring. Both P-site mutations occur in one clone (Af-33), with one of the changes also found in a second clone, Af-17. A P-site mutation that disrupted a G-C base pairing, changing it to G-G, was only found in Af-33, while the mutation shared by both Af-17 and Af-33 altered a G-C base pair to G-U. All three changes found in these functional regions occur at sites that are considered to be universally conserved sites (>98% sequence conservation). The three sequences with changes in the A- or P-sites (Af-5, Af-17 and Af-33) are among the four alleles with the most changes in universally conserved sites in the entire molecule. This makes them among the most likely candidates to be non-functional pseudogenes.

Relative rate analysis

Finally, to further investigate whether the NPL alleles seen in lake sturgeon are under reduced selective constraints when compared with the PL alleles, sequences were subjected to a relative rates test using the Two-cluster method of Takezaki et al. (1995). This method allows indirect comparison of the amount of change found in the lineages leading to each of the allele-type groups since they diverged from one another, by utilizing comparisons with an outlying control sequence. In addition, PHYLTEST (Kumar, 1996) was also used to conduct a two-tailed normal deviate test (Takezaki et al., 1995) for the relative rate comparison to determine if the differences in evolutionary rates between the two groups are statistically significant. The 15 NPL sequences were compared with the sequences of PL variants Af-7 and Af-13, using the *Polypterus* 18S rRNA gene sequence as an outlier. The Kimura (1980) two-parameter distance was used as the measure of sequence divergence. The comparison showed the NPL variants evolving more quickly than the PL variants, with the result statistically significant at the 5% level. This is consistent with relaxed selective constraints on NPL variants, which would result in an increased rate of change of the NPL alleles.

RT-PCR and sequencing analysis in lake sturgeon

The 18S rRNA RT-PCR product obtained from total cellular RNA isolated from lake sturgeon liver was sequenced to determine which sequence variants are expressed. The sequencing reaction does not resemble that obtained after PCR from bulk genomic DNA, which clearly shows a mixture of sequencing templates with variable sequences (Krieger and Fuerst, 2002b). Genomic DNA sequences show especially the presence of indel events at several positions within the sequence, which result in an inability to accurately read the sequence on a gel. A single easily readable sequence (GenBank accession number AF198113) was obtained from the RT-PCR product (Fig. 2). The sequence was identical with that of lake sturgeon clone Af-7, one of the two sequence variants from lake sturgeon (PL alleles) that clusters with the paddlefish sequence in phylogenetic analyses. This variant differs from the paddlefish sequence by only one base substitution. If other sequence variants are expressed in lake sturgeon liver, they must be expressed at very low levels compared with that of the primary expressed 18S rRNA component (represented by clone Af-7), as they were not detectable during sequencing.

We analyzed the composition of the transcript pool in order to determine if both PL and NPL allele types are expressed in this organism, although only one allele (PL) was detectable during sequencing of the RT-PCR product. This experiment contrasted the use of a PCR primer pair that was found to amplify only NPL sequences and a pair that would amplify both PL and NPL sequences. Both primer pairs produced amplification from RNA transcripts, indicating that some NPL sequence was present among the transcripts. A final set of reactions was performed, equivalent to real-time PCR, using lake sturgeon 18S rRNA RT-PCR product as a template, to try to determine the ratio of PL : NPL transcripts. Results suggest that the ration of PL : NPL is at least 33 000 : 1 and may actually be in a ratio of $>10^6$: 1. Since the sequence of clone Af-7 was recovered only three times among the 30 clones analyzed from genomic DNA (13.3%), it is unlikely to represent

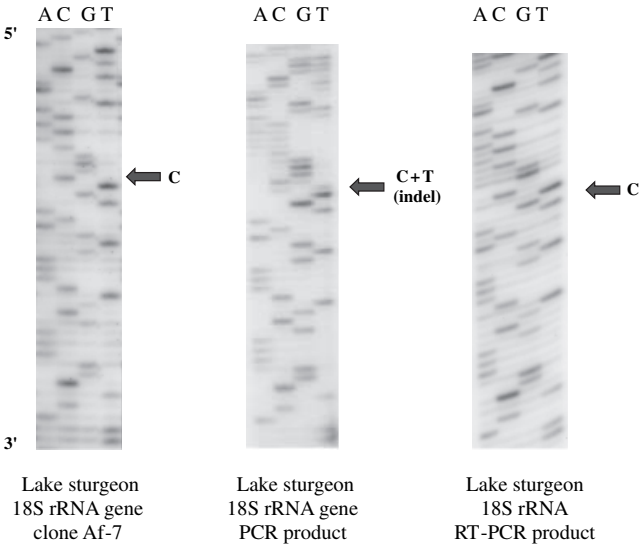

Fig. 2. DNA sequencing gels of lake sturgeon 18S rRNA gene clone Af-7, genomic polymerase chain reaction (PCR) product and 18S rRNA reverse transcriptase (RT)-PCR product, for primer 170. Sequence from this region illustrates the clarity of the RT-PCR product sequence when compared with the genomic PCR product sequence from lake sturgeon, which has multiple sequencing ladders. Other primer positions produce similar multiple sequence ladders at other points in the total sequence. These sample sequences also illustrate the sequence identity of clone Af-7 and the RT-PCR product

the most frequent genomic form of the gene. Nevertheless, it appears that an overwhelmingly large amount of PL rRNA similar in sequence to clone Af-7 is present in expressed lake sturgeon 18S rRNA when compared with the amount of NPL rRNA. This would explain why only one sequence is observed when the 18S rRNA RT-PCR product is sequenced.

Discussion

18S rDNA sequence variation in lake sturgeon

Sturgeon have an unexpected intraindividual diversity for sequence in the nuclear small subunit ribosomal RNA gene. Of 30 lake sturgeon 18S rDNA clones examined, at least 17 different sequence variants were isolated. Few other organisms have been found to possess intraindividual variation for the 18S rRNA gene, and in those cases only two or three sequence variants have been found.

Among the 17 variant sequences identified in our study, nucleotide differences (not including indels) ranged from 3 (0.002) to 59 (0.034) differences among 1783 aligned positions. Comparisons of 18S rRNA sequences from other fish indicate the unusual nature of the levels of variability seen in this study. We examined sequences retrieved from GenBank from four salmonid species (*Oncorhynchus kisutch*, AF030250; *O. masou*, AF243427; *O. mykiss*, AF308735 and *Salmo trutta*, X98839). Thirty-seven of 1613 aligned sites show differences between species, and interspecific nucleotide differences range from 12 (0.007) to 29 (0.018) differences. Looking more broadly, comparisons can be made among sequences from a group of four distantly related percomorph species (*Sebastolobus altivelis*, M91182; *Fundulus heteroclitus*, M91180; *Oreochromis esculentus*, AF337051 and *Tetraodon nigroviridis*, DS4272). Variation occurs at 99 of 1823 aligned sites; interspecific nucleotide differences range from 38 (0.021) to 68 (0.038) differences.

18S rDNA variants: functional genes or pseudogenes?

Observations on the patterns and types of sequence changes within the lake sturgeon sequence variants taken together seem neither to support nor disprove them as either functional genes or pseudogenes. The number of changes in areas of the molecule with different structural features is consistent with there continuing to be some selective constraints on sequence change. Comparison with the eukaryotic small subunit rRNA phylogenetic conservation structure map (Cannone et al., 2002) indicated that variation in lake sturgeon alleles is less likely to occur at sites that normally are considered to be conserved than in non-conserved sites. Most indels events in lake sturgeon alleles occurred in the regions considered highly variable. Examination of changes in small subunit rRNA regions known to be important for ribosome function indicate little or no variation compared with the presumably functional paddlefish sequence, as well as with the sequences found in the eukaryotic phylogenetic conservation map (Cannone et al., 2002). With the exception of a single six-base deletion event in one clone, there are no drastic changes in sequence to suggest that the lake sturgeon variants are not capable of forming a functional secondary structure.

These observations are more interesting, but less surprising in light of the results of a recent study by Olson and Yoder (2002) of 12S rRNA gene numts (mitochondrial-derived nuclear pseudogenes) in the human genome. The authors compared gene sequence and gene product secondary structure for the functional human mitochondrial 12S rRNA gene and three known human numts of the 12S rRNA gene. They found that it was unlikely that they would have been identified as pseudogene sequences, even when using a detailed visual examination of the three numt sequences with respect to secondary structure models or even comparison of the predicted stabilities of certain structural regions.

The analysis of transcripts does show that there is one 18S rRNA sequence variant (clone Af-7) that represents most or all of the expressed rRNA in lake sturgeon. The paddlefish appears to possess only one small subunit rRNA sequence; it is assumed that it must be functional by default. Thus, it is likely that the principal expressed sequence variant found in lake sturgeon, which shows only 1 bp difference from the paddlefish sequence, is functional as well. As for the remaining 16 sequence variants isolated from lake sturgeon, the degree of their sequence divergence from the presumptive functional sequence suggests that they are under reduced selection pressures. Considering all aspects of structure, only four of the 17 lake sturgeon variants show substantial combined changes that would suggest that they are likely to be non-functional pseudogenes. The results of the relative rates test provide further support of this conclusion for the NPL alleles. With the detection of a single major expressed 18S rDNA variant, it is possible that the other variants have little or no adaptive importance and may actually be pseudogenes. It appears that although incomplete homogenization of 18S rDNA sequences occurs within the sturgeon genome, there is some mechanism in place that results in the selective expression of only certain (conserved) alleles.

This lack of rRNA gene sequence homogenization may be indicative of a reduced rate of concerted evolution within the entire genome of the Acipenseridae. Interestingly, de la Herrán et al. (2001) recently discovered evidence that another tandemly repeated sequence, the *Hin*dIII satellite DNA family, also exhibits low rates of homogenization within sturgeon. A slowed rate of concerted evolution would also be consistent with the hypothesis that the order Acipenseriformes possesses a reduced rate of molecular evolution (Birstein and Vasiliev, 1987; Brown et al., 1996; Krieger and Fuerst, 2002a). It is also possible that concerted evolution may not be acting on all rDNA clusters equally in sturgeon. The possession of multiple NORs and the positions of these rDNA clusters on various chromosomes may affect the ability of the mechanisms of concerted evolution to homogenize sequences among all of the NORs. Unfortunately, with the information available at this time it is not possible to provide a definite explanation for the lack of homogenization of tandemly repeated DNA arrays observed in sturgeon. Further study of the unusual genomes of the Acipenseriformes as well as a search for the same phenomenon in other polyploid fish or vertebrates could prove to be key for understanding the complexities of concerted evolution and why it sometimes fails.

Acknowledgements

The authors would like to thank Ron Bruch and Règis Moreau who generously provided the samples necessary for this study, and Rob Haney for instruction in the use of MODELTEST and PAUP*. Also thank Tom Waite for assistance in the resampling estimate of allele number. This work was supported by a Theodore Roosevelt Memorial Fund Grant from the American Museum of Natural History to JK, a grant from the Ohio-Sea Grant to PAF and a Dissertation Improvement Grant from the National Science Foundation.

References

Birstein, V. J.; Vasiliev, V. P., 1987: Tetraploid-octoploid relationships and karyological evolution in the order Acipenseriformes (Pisces): karyotypes, nucleoli, and nucleolus-organizer regions in four acipenserid species. Genetica **72,** 3–12.

Brown, J. R.; Beckenbach, K.; Beckenbach, A. T.; Smith, M. J., 1996: Length variation, heteroplasmy and sequence divergence in the mitochondrial DNA of four species of sturgeon (*Acipenser*). Genetics **142,** 525–535.

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

Cannone, J. J.; Subramanian, S.; Schnare, M. N.; Collett, J. R.; D'Souza, L. M.; Du, Y.; Feng, B.; Lin, N.; Madabusi, L. V.; Müller, K. M.; Pande, N.; Shang, Z.; Yu, N.; Gutell, R. R., 2002: The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron and other RNAs. BioMed Central Bioinformatics 3:2 [Correction: BioMed Central Bioinformatics 3:15].

Carranza, S.; Giribet, G.; Ribera, C.; Baguñà, J.; Riutort, M., 1996: Evidence that two types of 18S rDNA coexist in the genome of *Dugesia (Schmidtea) mediterranea* (Platyhelminthes, Turbellaria, Tricladida). Mol. Biol. Evol. **13,** 824–832.

Carter, A. P.; Clemons, W. M.; Broderson, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V., 2000: Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature **407,** 340–348.

Dame, J. B.; Sullivan, M.; McCutchan, T. F., 1984: Two major sequence classes of ribosomal RNA genes in *Plasmodium berghei*. Nucleic Acids Res. **12,** 5943–5952.

Elder, J. F.; Turner, B. J., 1995: Concerted evolution of repetitive DNA sequences in eukaryotes. Q. Rev. Biol. **70,** 297–320.

Gutell, R. R., 1994: Collection of small subunit (16S- and 16S-like) ribosomal RNA structures: 1994. Nucleic Acids Res. **22,** 3502–3507.

de la Herrán, R.; Fontana, F.; Lanfredi, M.; Congiu, L.; Leis, M.; Rossi, R.; Ruiz Rejón, C.; Ruis Rejón, M.; Garrido-Ramos, M. A., 2001: Slow rates of evolution and sequence homogenization in an ancient satellite DNA family of sturgeons. Mol. Biol. Evol. **18,** 432–436.

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

Krieger, J.; Fuerst, P. A., 2002a: Evidence for a slowed rate of molecular evolution in the order Acipenseriformes. Mol. Biol. Evol. **19,** 891–897.

Krieger, J.; Fuerst, P. A., 2002b: Evidence of multiple alleles of the nuclear 18S ribosomal RNA gene in sturgeon (Family: Acipenseridae). J. Appl. Ichthyol. **18,** 290–297.

Kumar, S., 1996: PHYLTEST: a program for testing phylogenetic hypothesis, version 2.0. Institute of Molecular and Evolutionary Genetics and Department of Biology, The Pennsylvania State University, University Park, PA, USA.

Mears, J. A.; Cannone, J. J.; Stagg, S. M.; Gutell, R. R.; Agrawal, R. K.; Harvey, S. C., 2002: Modeling a minimal ribosome based on comparative sequence analysis. J. Mol. Biol. **321,** 215–234.

Muir, G.; Fleming, C. C.; Schlötterer, C., 2001: Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. Mol. Biol. Evol. **18,** 112–119.

Olson, L. E.; Yoder, A. D., 2002: Using secondary structure to identify ribosomal numts: cautionary examples from the human genome. Mol. Biol. Evol. **19,** 93–100.

Posada, D.; Crandall, K. A., 1998: MODELTEST: testing the model of DNA substitution (version 3.06). Bioinformatics **14,** 817–818.

Rodríguez, F. J.; Oliver, J. L.; Marín, A.; Medina, J. R., 1990: The general stochastic model of nucleotide substitution. J. Theor. Biol. **142,** 485–501.

Swofford, D. L., 2002: PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Associates, Sunderland, Massachusetts, USA.

Takezaki, N.; Rzhetsky, A.; Nei, M., 1995: Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. **12,** 823–833.

Varadaraj, K.; Skinner, D. M., 1994: Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically-engineered DNA-polymerases. Gene **140,** 1–5.

Vawter, L.; Brown, W. M., 1993: Rates and patterns of base change in the small subunit ribosomal RNA gene. Genetics **134,** 597–608.

Woese, C. R.; Gutell, R.; Gupta, R.; Noller, H. F., 1983: Detailed analysis of the high order structure of 16S-like ribosomal ribonucleic acids. Microbiol. Rev. **47,** 621–669.

Author's address: Paul A. Fuerst, Departments of Evolution, Ecology and Organismal Biology and Molecular Genetics, The Ohio State University, 386 Aronoff Laboratory, 318 W. 12th Avenue, Columbus, OH 43210, USA. E-mail: fuerst.1@osu.edu

Author Query Form

Journal: JAI

Article: 610

Dear Author,
During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers on the query sheet if there is insufficient space on the page proofs. Please write clearly and follow the conventions shown on the attached corrections sheet. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Many thanks for your assistance.

Query reference	Query	Remarks
Q1	Au: Please define: MMLV.	
Q2	Au: Crandall and Posada, 1998 has been changed to Posada and Crandall, 1998 so that this citation matches the list	
Q3	Au: Please define: NOR.	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	··· under matter to remain	Stet
Insert in text the matter indicated in the margin		New matter followed by
Delete	through matter to be deleted	
Delete and close up	through matter to be deleted	
Substitute character or substitute part of one or more word(s)	/ through letter or through word	New letter or new word
Change to italics	— under matter to be changed	
Change to capitals	≡ under matter to be changed	
Change to small capitals	= under matter to be changed	
Change to bold type	~ under matter to be changed	
Change to bold italic	≈ under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Insert 'superior' character	/ through character or where required	under character e.g.
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	
Insert comma	(As above)	
Insert single quotation marks	(As above)	and/or
Insert double quotation marks	(As above)	and/or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking letters	
Insert space between letters	between letters affected	
Insert space between words	between words affected	
Reduce space between letters	between letters affected	
Reduce space between words	between words affected	