Comp. Biochem. Physiol. Vol. 91B, No. 4, pp. 639–646, 1988
Printed in Great Britain

0305-0491/88 $3.00 + 0.00

CHARACTERIZATION OF A HIGHLY REPEATED SATELLITE DNA FROM THE CYPRINID FISH *NOTROPIS LUTRENSIS*

S. P. MOYER,* D. P. MA,† T. L. THOMAS† and J. R. GOLD*

Departments of Wildlife and Fisheries Sciences* and Biology,† Texas A&M University, College Station, TX 77843, USA (Tel: (409) 845-5777)

(*Received* 16 *December* 1987)

Abstract—1. A highly repeated, satellite DNA family from the North American cyprinid fish, *Notropis lutrensis*, was identified as a fragment band following restriction endonuclease enzyme digestion and agarose gel electrophoresis of genomic DNA; evidence of a tandem arrangement of the satellite in the genome was demonstrated by the formation of "ladders" in partial restriction endonuclease digests.

2. The satellite family was estimated densitometrically to comprise 7–8% of the *N. lutrensis* genome; mapping experiments using isolated and purified monomer repeat units of the satellite uncovered nine sites for seven different restriction enzymes.

3. A monomeric repeat unit of the satellite was cloned and sequenced, and found to be 174 base pairs in length and to have a base composition of 47% G + C (guanine + cytosine); computer analysis of the sequence revealed 13 new restriction sites for 12 additional enzymes.

4. Computer analysis also revealed that a large degree of internal redundancy in the monomer unit exists in the form of both direct and inverted repeating units, and that the entire sequence, starting with one base in either orientation, constitutes an open reading frame. In all but the last characteristic, the *N. lutrensis* satellite DNA is very similar to satellite DNAs in other eukaryotes.

INTRODUCTION

Satellite DNA, a highly repeated DNA class, was first detected in the early 1960s as a minor component following density gradient centrifugation (Kit, 1962; Sueoka and Chang, 1962). More recent studies have shown that on a molecular level, satellite DNA is generally composed of tandem arrays of relatively short, repeated polynucleotide sequences which display considerable complexity and vary in size from about 10 to several hundred base pairs (John and Miklos, 1979). This was first detected by heat denaturation and reassociation experiments (Walker and McLaren, 1965; Britten and Kohne, 1968), and subsequently demonstrated by restriction endonuclease digestion of total genomic DNA (Singer, 1982). Cytologically, satellite DNA is located primarily in the constitutively heterochromatic regions of metaphase chromosomes, and in large, positively heteropycnotic regions of interphase cells (John and Miklos, 1979). One common feature of satellite DNA in almost all eukaryotes examined thus far is an apparent lack of measurable transcriptional activity (Yunis and Yasmineh, 1971; John and Miklos, 1979; Singer, 1982). The cellular or organismal function of satellite DNA is unknown, although a variety of hypotheses have been proposed (John and Miklos, 1979).

In this report, the isolation, purification, and sequencing of a repeat unit from a satellite DNA family in the cyprinid fish, *Notropis lutrensis*, is described. To our knowledge, this is the first instance where a satellite DNA from a fish species has been isolated and characterized. The presence of satellite DNAs in several fish species had been inferred from studies using buoyant density ultracentrifugation, thermal denaturation, and/or the reassociation kinetics of genomic DNA (Hanham and Smith, 1979; Hudson *et al.*, 1980; Karel and Gold, 1987, 1988; Schmidtke *et al.*, 1979). A number of fish species also possess relatively large amounts of chromosomal heterochromatin (Gold *et al.*, 1986), suggesting the presence of satellite DNA.

MATERIALS AND METHODS

The fish species used in the research was the North American cyprinid, *Notropis lutrensis*. The fish were collected by seine from the Little Brazos River near College Station, Texas, transported live to the laboratory, and maintained in aerated aquaria until sacrificed.

Genomic DNA was isolated using a modification of the method of Zimmer *et al.* (1981). Full details of DNA isolation, all other procedures, and all solutions and buffers may be found in Moyer (1986). Most of the procedures followed Maniatis *et al.* (1982). The purity of isolated genomic DNA was determined by the 260/280 nm ratio of a 1:50 dilution (DNA:buffer) in a spectrophotometer. For most preparations (usually employing 3–5 fish), ratio values ranged from 1.6 to 1.8.

The purified *N. lutrensis* genomic DNA was screened initially by single digestions with 32 different restriction endonuclease enzymes. Digested DNAs were electrophoresed on 1.5% agarose gels, stained with ethidium bromide, and the resulting fragment patterns photographed using Kodak technical pan 2415 film developed in Diafine (Acufine). A putative satellite DNA family was identified by the presence of a distinct, sharp band of unit size (*ca* 180 base pairs or bp) in four of the restriction digests and by a series or ladder of bands representing approximate multiples of 180 bp in partial digests with these same enzymes. It was assumed that the 180 bp fragments were members of a highly repeated, tandemly arranged satellite DNA family and represented the

monomeric repeat units of that sequence family (Singer, 1982). The fragments contained within the 180 bp band were isolated and purified in bulk by digesting purified genomic DNA with one of the enzymes (*Mbo* I) which had produced the band, electrophoresing the mixture on 1.5% agarose gels, and recovering the 180 bp fragments from the gels using BRL Nacs-Prepac columns. The fragments were eluted from the Nacs columns, washed, air dried and resuspended in buffer.

The 180 bp fragments (assumed to be the monomeric repeat units of the satellite DNA family) were mapped by restriction enzyme digestion of the purified *Mbo* I produced fragments. Both single and pairwise digestions were radioactively end-labelled using the Klenow fragment of DNA polymerase I (Brown, 1980), separated electrophoretically on 3.5% polyacrylamide gels, and autoradiographed using X-omat RP film at room temperature for several hours. Most of the methodologies followed Carr *et al.* (1986). Maps were constructed following Brown and Vinograd (1974).

For cloning and DNA sequencing procedures, the purified, *Mbo* I produced fragments were first ligated into the *Bam* HI site of the plasmid vector pUC8 (Vieira and Messing, 1982), and cloned in *E. coli* TB1. Possible recombinant plasmids were screened by ampicillin resistance and lac$^-$ (white colony) phenotypes. Eight recombinant plasmids, termed pSM1-8, were isolated and their inserts subsequently screened for size. All eight had inserts in the size range expected for the *Mbo* I produced monomers. Mapping experiments indicated that most contained inserts of the 180 bp satellite DNA. The inserts from pSM6 and pSM8 were sequenced by the dideoxy chain termination method (Messing, 1983) using linearized and denatured plasmid templates with synthetic sequencing primers for both orientations (D. P. Ma, unpublished).

Table 1. Restriction endonuclease enzymes used in initial screening of *N. lutrensis* genomic DNA

Enzyme	Recognition site (5′→3′)	Digestion pattern*
Acc I	GT/$^{AT}_{CG}$AC	—
Alu I	AG/CT	—
Ava II	G/G$^{A}_{T}$CC	Band
Bal I	TGG/CCA	—
Bam HI	G/GATCC	—
Bcl I	T/GATCA	—
Bgl II	A/GATCT	—
Bst EII	G/GTNACC	—
Cla I	AT/CGAT	—
Dde I	C/TNAG	Complete
Eco RI	G/AATC	—
Hae III	GG/CC	Band
Hin cII	GT Py/PuAC	—
Hin dIII	A/AGCTT	—
Hin fI	G/ANTC	Complete
Hha I	GCG/C	—
Hpa II	C/CGG	—
Kpn I	GGTAC/C	—
Mbo I (*Sau* 3AI)†	/GATC	Band
Msp I	C/CGG	Complete
Pst I	CTGCA/G	—
Pvu II	CAG/CTG	—
Sal I	G/TCGAC	—
Sau 96I	G/GNCC	Band
Sma I	CCC/GGG	—
Sst I	GAGCT/C	—
Sst II	CCGC/GG	—
Taq I	T/CGA	Complete
Tha I	CG/CG	—
Xba I	T/CTAGA	Complete
Xho I	C/TCGAG	—
Xor II	CGATC/G	—

*—: Complete digestion same as uncut pattern; band: distinct band of 175–185 bp in size formed upon complete digestion; complete: smear formed upon complete digestion.
†Insensitive to methylation at the G residue.

RESULTS AND DISCUSSION

The 32 restriction enzymes used initially to digest purified *N. lutrensis* genomic DNA are shown in Table 1. Complete digestion with four of the enzymes (*Ava* II, *Hae* III *Sau* 96I and *Mbo* I) produced distinct bright fragment bands of *ca* 180 bp in length following agarose gel electrophoresis (Table 1; Fig. 1). This indicated the presence in the *N. lutrensis* genome of a satellite-type DNA with a repeat unit or monomer length of *ca* 180 bp and with restriction sites for these four enzymes (Singer, 1982). In *Hae* III digests, a second band was observed just below the 180 bp band (lane 4, Fig. 1). The smaller size of the second band (*ca* 150 bp) suggested the presence of either a second satellite DNA family in the *N. lutrensis* genome, or that a second, variable *Hae* III site occurs within the 180 bp monomer (i.e. some members of the 180 bp satellite DNA family have a second *Hae* III site, while others do not). Complete digestion with the remaining enzymes yielded either uncut or complete smear patterns (Table 1). It was assumed that the enzymes yielding the latter pattern, but not the former, could have sites among some members of the 180 bp satellite family.

Partial digests of *N. lutrensis* genomic DNA using *Mbo* I produced a series or ladder of bands representing approximate multiples of 180 bp (Fig. 2). This ladder or type A pattern (Singer, 1982) demonstrates that the 180 bp monomer is tandemly arranged in the *N. lutrensis* genome and is further evidence that it represents a satellite-type DNA sequence. Complete *Mbo* I digestion of *N. lutrensis* DNA, however, invariably produced a second, lighter band of *ca* 370 bp in length in addition to the band at 180 bp (lane 6, Fig. 1). Extended incubations and use of increased units of enzyme yielded similar results, indicating that the larger *Mbo* I produced band is probably a dimer of the 180 bp repeat unit which has lost an *Mbo* I site within the tandem repeat. Partial digests of *N. lutrensis* DNA using *Hae* III consistently revealed the presence of the second, smaller band of *ca* 150 bp (data not shown). In each case, the smaller band was consistently brighter under u.v. illumination than the 180 bp band, suggesting that the former comprises many more sequences (or copies) than the latter in the satellite family.

A Joyce–Loebl recording microdensitometer was used to scan the 35 mm negative of the gel shown in Fig. 1. In each of the four lanes (3–6) which show the satellite band, the curve areas of the most intense band and of the entire lane were quantitated by summing the graph units under their respective curves. An example of a densitometric tracing (lane 6) is shown in Fig. 3. The relative amount of the 180 bp monomer in each lane was estimated as the ratio of peak areas calculated as the curve area of the most intense band divided by the curve area of total DNA for that lane. In the *Hae* III digest (lane 4), the two bands could not be separated and were quantitated as a single unit. The estimated relative amounts of the 180 bp monomer were: 7% (lane 3), 7.5% (lane 4), 6.5% (lane 5) and 8% (lane 6). Since a low frequency of ladder fragments can still be seen even

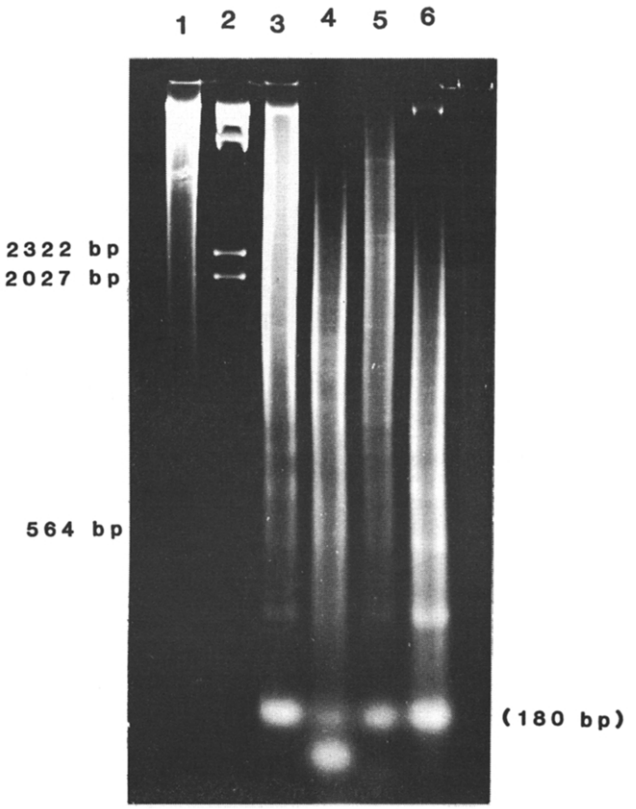

Fig. 1. Agarose gel containing uncut and restriction endonuclease digested *N. lutrensis* genomic DNA. Lane 1 = uncut *N. lutrensis* DNA; lane 2 = *Hin*d III-digested lambda DNA; lane 3 = *Ava* II-digested *N. lutrensis* DNA; lane 4 = *Hae* III-digested *N. lutrensis* DNA; lane 5 = *Sau* 961-digested *N. lutrensis* DNA; and lane 6 = *Mbo* I-digested *N. lutrensis* DNA. The 564 bp lambda DNA fragment is not visible in the photograph, but was apparent on the original gel. Fragment sizes in parentheses were estimated from a standard calibration curve.

in complete digests with all four enzymes (Fig. 1), the figure of 7–8% should be considered a minimum estimate. Given that the DNA content of *N. lutrensis* is 2.37 pg (Gold and Price, 1985), it can be estimated that there are minimally from 2.1×10^4 to 2.4×10^4 copies of the 180 bp monomer in the *N. lutrensis* genome.

The restriction endonucleases used to map the purified 180 bp *Mbo* I-produced monomers included seven of the enzymes which produced "band" or "complete" digestion patterns in the initial screening of *N. lutrensis* genomic DNA (Table 1). *Sau* 96I was omitted since its recognition site is almost identical to that of *Ava* II (Table 1), and it was assumed that the two sites were synonymous. *Sst* II, which apparently does not have a site in the monomer unit (Table 1), was used as a control.

The estimated fragment sizes produced from single and double digests using the eight enzymes are shown in Table 2. In many digests, particularly those involving two enzymes, additional bands larger than those indicated in Table 2 were observed. In almost every case, these were assumed to be incomplete digests since (i) the lane totals in each digest exceeded 225 bp if these fragments were included in the calculations, and (ii) most of these larger bands could be accounted for by summing the sizes of two of the smaller fragments observed in that lane. Of the 36 digests, only four produced totals that significantly exceeded the size range of 174–188 bp. These included two double digests using *Dde* I (with *Sst* II or *Ava* II), the double digest using *Hae* III + *Hin* fI, and the single digest using *Hae* III (Table 2). The two digests using *Dde* I were inconsistent to varying degrees with preliminary restriction maps generated from other single and double digests using *Dde* I, *Sst* II and *Ava* II, and were therefore disregarded and omitted from further consideration. The two *Hae* III digests (*Hae* III and *Hae* III + *Hin* fI) revealed a large fragment of 145 bp which was assumed to be an incomplete digest of a 117–120 bp fragment and a 25–28 bp fragment. This assumption was based in part on the observation that other *Hae* III digests (e.g. *Hae* III + *Msp* I) revealed only smaller fragment bands in the 117–127 bp range, and in part on the lane totals which became 186–187 bp when the 145 bp fragment was considered to represent a 125 bp fragment. Finally, in digests using *Taq* I and *Hae* III (Table 2), the largest fragment scored often appeared as two fragment bands *ca* 6–10 bp apart. In *Taq* I digests, a small fragment band of *ca* 10 bp was often observed, suggesting the larger of the two large bands was an incomplete digest. As a consequence, only the smaller of the two large bands was used in calculating lane totals and in generating restriction maps. A similar, very small (< 10 bp) fragment band was not observed

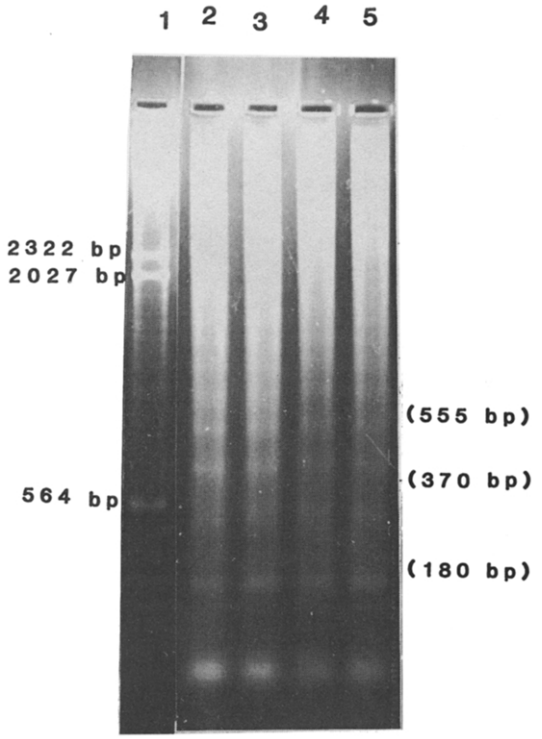

Fig. 2. Agarose gel showing "ladder" pattern produced in partial digests of *N. lutrensis* genomic DNA. Lane 1 = *Hin*d III-digested lambda DNA; lane 2 = *Mbo* I-digested (10u, 3 hr) *N. lutrensis* DNA; lane 3 = *Mbo* I-digested (5u, 3 hr) *N. lutrensis* DNA; lane 4 = *Mbo* I-digested (5u, 2 hr) *N. lutrensis* DNA; and lane 5 = *Mbo* I-digested (5u, 1 hr) *N. lutrensis* DNA. Fragment sizes in parentheses were estimated from a standard calibration curve.

in most *Hae* III digests, and as a result only the larger of the two large fragments was used in calculating lane totals and in generating restriction maps.

An approximate restriction site map of the 180 bp monomer unit based on the fragment patterns shown in Table 2 is illustrated in Fig. 4. No sites were found for the enzymes *Msp* I and *Sst* II; single sites were found for *Hin* fI, *Xba* I, *Dde* I, and *Ava* II; two sites were found for *Hae* III; and three sites were found for *Taq* I. The two *Hae* III sites are located at either end of the monomer unit and corroborate the previous suggestion that the two bands observed in *Hae* III digests of genomic DNA (Figs 1 and 2) could be due to a second, variable *Hae* III site within the repeat unit. Exactly which *Hae* III site is absent from some members of the satellite family is problematic since the two sites are *ca* 25 and 35 bp from either end. One final point to note is that the restriction map indicates sites which occur in the family of monomer repeats, and not necessarily the sites within any one member (or copy) of the family.

The nucleotide sequence of the inserts in plasmids pSM6 and pSM8 was identical and is shown in Fig. 5. The sequence is 174 bp in length and has a base composition of 47% G + C (guanine + cytosine) base pairs. The predicted restriction endonuclease map of the sequence is shown in Fig. 6, and includes seven of the nine restriction sites (including both *Hae* III sites) found in the previous experiments along with thirteen new sites for twelve additional enzymes. The sites for *Dde* I and *Ava* II found in the mapping experiments (Fig. 4) were not present in this particular sequence. Putative locations for each site, however, were easily found (Fig. 5) by scanning the regions in which each site was expected to occur and looking for a sequence which differed from the appropriate recognition sequence by only a single base. The presence (or absence) of a site (or sites) within any single member (or copy) of a satellite DNA family is not surprising given the comparatively rapid rate of sequence divergence which occurs in these types of DNAs (John and Miklos, 1979; Miklos and Gill, 1982). The variable sites do, however, indicate the presence of sequence variants or subfamilies within the satellite DNA family. Finally, a *Sau* 96I site was found at position 31–35. This suggests that our previous assumption that the *Sau* 96I and *Ava* II sites were synonymous was incorrect. The *Ava* II site as shown in Fig. 4, however, is apparently real since (i) the *Ava* II enzyme will not recognize the 5′-GGGCC-3′ sequence recognized by *Sau* 96I (Table 1), and (ii) the mapping experiments placed the *Ava* II site to the right of the *Xba* I site and considerably displaced from the *Hin* fI site (Table 2; Fig. 4).

Computer analysis of the nucleotide sequence was used to elucidate the degree of repetition and the open reading frames contained within the 174 bp monomer unit. Five different direct repeats of five and six base pairs with full complementarity occur within the sequence, one of which is found three times (Table 3). The spacing or distance in base pairs between each repeat varied from 12–50 bp, although a distinct mode at 38–40 bp is plainly evident. This may indicate that the 174 bp monomer is comprised of smaller repeating units. Eighteen different inverted

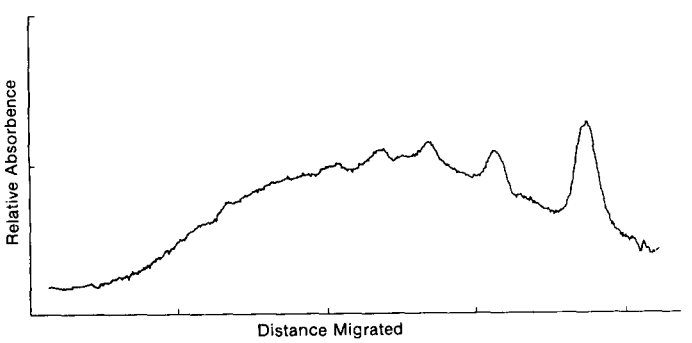

Fig. 3. Densitometric tracing of *Mbo* I digested *N. lutrensis* genomic DNA.

Table 2. Approximate fragment sizes produced from single and double digests of purified monomer (*ca* 180 bp) fragments

Restriction enzymes	Fragment sizes*					
	A	B	C	D	E	Total
Msp I	—	—	—	—	—	—
Sst II	—	—	—	—	—	—
Ava II	113	69	—	—	—	182
Hin fI	158	24	—	—	—	182
Xba I	110	65	—	—	—	175
Msp I + *Sst* II	—	—	—	—	—	—
Msp I + *Ava* II	113	69	—	—	—	182
Msp I + *Hin* fI	155	24	—	—	—	179
Msp I + *Xba* I	110	66	—	—	—	176
Sst II + *Ava* II	113	68	—	—	—	181
Sst II + *Hin* fI	155	22	—	—	—	177
Sst I + *Xba* I	110	66	—	—	—	176
Ava II + *Xba* I	(110)	70†	40	—	—	180
Ava II + *Hin* fI	90	68	24	—	—	182
Hin fI + *Xba* I	102	(70)	54	24	—	180
Taq I	(93)	83	50	43	10	186
Taq I + *Msp* I	(93)	87	44†	—	—	175
Taq I + *Sst* II	88	(50)	46	40	10	184
Taq I + *Xba*I	(83)	73	44	30†	10	187
Taq I + *Hin* fI	(83)	75	50	29	24	178
Taq I + *Ava* II	70	52	42	10†	—	184
Dde I	87†	—	—	—	—	174
Dde I + *Msp* I	87†	—	—	—	—	174
Dde I + *Hin* fI	89	68	24	—	—	181
Dde I + *Taq* I	87	45†	10	—	—	187
Dde I + *Xba* I	89	71	19	—	—	179
Dde I + *Sst* II	115	85	—	—	—	200
Dde I + *Ava* II	113	68	27	—	—	208
Hae III	145‡	34	27	—	—	206 (186)‡
Hae III + *Msp* I	122	[117]	34	27	—	183
Hae III + *Sst* II	125	[117]	⟨34⟩	⟨27⟩	—	186
Hae III + *Dde* I	127	120	34	27	—	188
Hae III + *Hin* fI	145‡	27	25	10	—	207 (187)‡
Hae III + *Ava* II	78	49	34	27	—	188
Hae III + *Xba* I	88	[81]	40	34	26	188
Hae III + *Taq* I	62	51	34	26	10	183

*Fragments in parentheses are assumed to be incomplete digests. Fragments in "[]" brackets were not included in calculating lane totals and generating restriction maps (see text). Fragments in "⟨ ⟩" brackets were not observed, but were assumed to be present.

†Two fragments (doublet) of approximate size are assumed to occur in this size class.

‡In these digests, the 145 base pair fragment was assumed to be 125 base pairs for calculating lane totals and generating restriction maps (see text). The lane totals using 125 instead of 145 base pairs are shown in parentheses.

repeats of 6–9 bp in length also occur within the sequence (Table 4). The spacing or distance in base pairs between each inverted repeat varied from 0–19 nucleotides, and may indicate that even smaller subunits occur within the monomer unit. The distribution of the inverted repeats, however, was fairly uniform throughout the sequence in contrast to the distribution of the direct repeats. This may suggest that the molecular origins of the two types of repeat are different. Regardless, the degree of repetition within the 174 bp monomer unit is striking, and is fully consistent with findings on satellite DNAs in other organisms (John and Miklos, 1979; Singer, 1982; Miklos and Gill, 1982).

Computer analysis also revealed that the entire 174 bp sequence constituted an open reading frame in two of the six possible alignments. One of these begins with the first 5′ base in one orientation; the other begins with the third 5′ base in the opposite orientation. The remaining four alignments contained open reading frames of from 9–105 nucleotides. The occurrence of long open reading frames is surprising in that by chance one might expect more DNA stop codon sequences to occur in what presumably is a non-transcribed DNA. However, given the findings on satellite DNA in other organisms (John and Miklos, 1979), it seems unlikely that the *N. lutrensis* satellite would be transcriptionally active. It is noteworthy that no translation start signals were found within the putative repeat unit by computer analysis.

As noted in the Introduction, this is, to our knowledge, the first instance where a satellite DNA sequence from a fish species has been isolated and characterized. The occurrence of satellite DNAs in fish is not surprising given their ubiquity in other eukaryotes (John and Miklos, 1979), and the inferences made from buoyant density ultracentrifugation, thermal denaturation, and reassociation kinetics of several fish genomes (Hanham and Smith, 1979; Hudson *et al.*, 1980; Karel and Gold, 1987, 1988; Schmidtke *et al.*, 1979). Many fish species also contain appreciable amounts of heterochromatin (Gold *et al.*, 1986), further indicating the presence of satellite DNAs (John and Miklos, 1979). The *N. lutrensis* satellite DNA is typical of many, if not most, eukaryote satellite DNAs in terms of size, copy number, presumed tandem arrangement, internal repetition, and the presence of subfamilies of sequences (John and Miklos, 1979; Singer, 1982; Miklos and Gill,

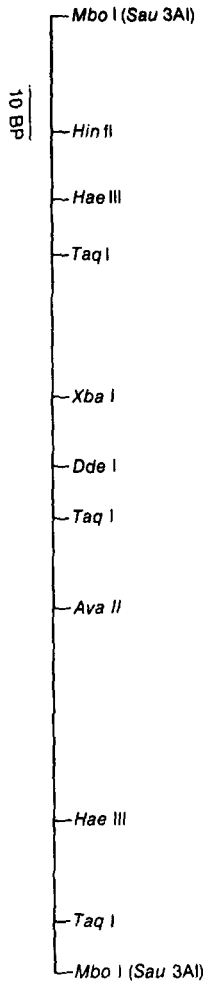

Fig. 4. Restriction endonuclease map of the *ca* 180 bp satellite DNA monomer unit.

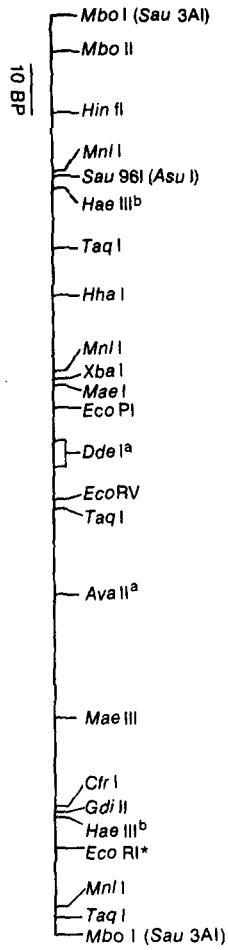

Fig. 6. Restriction endonuclease map of the 174 bp satellite DNA monomer unit. Enzymes in parentheses have the same recognition sequence. The sites for *Dde* I and *Ava* II (noted by *a*) were not found in the sequence, but are presumed to occur within some members of the satellite family (cf. text). One of the two *Hae* III sites (noted by *b*) is also variable within the satellite family. *Eco* RI* recognizes the sequence 5′-PuPuATPyPy-3′. The sites for *Mbo* II and *Mnl* I represent the locations of the recognition sites for the two enzymes and not the actual cleavage sites.

1982). The one dissimilarity is the occurrence of extensive open reading frames within the monomer repeat, although as noted above, it seems unlikely that the sequence is transcribed to form a functional product.

One final point to note is that the choice of *N. lutrensis* as a study animal may have been fortuitous. Karel and Gold (1987, 1988 and unpublished) have only been able to positively identify satellite DNA families in the genomes of nine out of the 100 different fish species thus far assayed by thermal denaturation. One of these was *Notropis lutrensis* which has a discrete, heavy melting component of

```
          10         20         30         40         50
5' - GATCTTCTTGAAACTCTGGATTCTGACAGAGGGCCACGAGACTTCGACCA
3' - CTAGAAGAACTTTGAGACCTAAGACTGTCTCCCGGTGCTCTGAAGCTGGT

          60         70         80         90        100
GCGCACAAAAATGGAGGCTCTAGACCCTTCCAGAGCCGAGATATCGAGTG
CGCGTGTTTTTACCTCCGAGATCTGGGAAGGTCTCGGCTCTATAGCTCAC

         110        120        130        140        150
TCAAAGTTTTGACCCGAAAAAAAAACATAAGGGGTAACACTGTCGTTTTT
AGTTTCAAAACTGGGCTTTTTTTTTGTATTCCCCATTGTGACAGCAAAAA

         160        170
GGCCGAAATTTTGACCCCTCGATC - 3'
CCGGCTTTAAAACTGGGGAGCTAG - 5'
```

Fig. 5. Nucleotide sequence of inserts from pSM6 and pSM8. Bases in boldface indicate putative sites for *Dde* I and *Ava* II. Note that two putative sites for *Dde* I occur between base pairs 81–90.

Table 3. Direct repeats within the 174 bp monomer sequence

Sequence	Spacing*
Five base pairs:	
37† CGAGA 41 87 CGAGA 91	50
73 GACCC 77 111 GACCC 115	38
106 GTTTT 110 145 GTTTT 149	39
107 TTTTG 111 147 TTTTG 151	40
147 TTTTG 151 159 TTTTG 163	12
Six base pairs:	
114 CCGAAA 119 153 CCGAAA 158	39

*Distance between repeat units in base pairs.
†Numbers represent position in the sequence in Fig. 5.

pairs. The similarity in % G + C between the satellite DNA family found in *N. lutrensis* by Karel and Gold (1987) and the one studied here suggests they are one and the same. Regardless, it may be that satellite DNAs are rare in fishes and that *N. lutrensis* is an exception. Future studies will address this question along with questions regarding the chromosomal distribution of the *N. lutrensis* satellite DNA and its occurrence in other North American cyprinid fishes.

Table 4. Inverted repeats within the 174 bp monomer sequence

	Sequence	Length	Spacing*
66†	GGCTCTAGA	9	4
87	CCGAGACCT		
63	GGAGGCTCT	9	0
80	CTTCCCAGA		
98	GTGTCAAAG	9	0
115	CCCAGTTTT		
38	GAGACTTCG	9	15
70	CTCGGAGGT		
151	GGCCGAAA	8	0
166	CCAGTTTT		
109	TTGACCC	7	0
137	AATGGGG		
73	GACCCTT	7	16
102	CTGTGAG		
62	TGGAGG	6	9
82	ACCTTC		
147	TTTTGG	6	0
158	AAAGCC		
121	AAAAAC	6	18
150	TTTTTG		
110	TGACCC	6	16
137	AATGGG		
94	TCGAGT	16	12
117	AGCCCA		
14	CTCTG	5	8
31	GAGAC		
151	GGCCGA	6	12
174	CTAGCT		
81	CAGAGC	6	19
111	GTTTTG		
29	GAGGGC	6	16
56	CACGCG		
29	GAGGGC	6	5
45	CTTCAG		
3	TCTTCT	6	19
33	GGGAGA		

*Distance between repeats in base pairs.
†Numbers represent position in the sequence in Fig. 5.

Acknowledgements—We thank Mary Jane Bastiaans and Steve Carr for their assistance in various phases of this study, and Chara Ragland for helpful comments on the manuscript. The work was supported by the Texas Agricultural Experiment Station under Projects H-6187 and H-6703.

REFERENCES

Britten R. J. and Kohne D. E. (1968) Repeated sequences in DNA. *Science* **161,** 529–540.

Brown W. M. (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc. natn. Acad. Sci. USA* **77,** 3605–3609.

Brown W. M. and Vinograd J. (1974) Restriction endonuclease cleavage maps of animal mitochondrial DNAs. *Proc. natn. Acad. Sci. USA* **11,** 4617–4621.

Carr S. M., Brothers A. J. and Wilson A. C. (1986) Evolutionary inferences from restriction maps of mitochondrial DNA from nine taxa of *Xenopus* frogs. *Evolution* **41,** 176–188.

Gold J. R. and Price H. J. (1985) Genome size variation among North American minnows (Cyprinidae). I. Distribution of the variation in five species. *Heredity* **54,** 297–305.

Gold J. R., Amemiya C. T. and Ellison J. R. (1986) Chromosomal heterochromatin in North American cyprinid fishes. *Cytologia* **51,** 557–566.

Hanham A. and Smith M. J. (1979) Sequence organization in the genomic DNA of the chum salmon, *Oncorhynchus keta*. *Can. J. Zool.* **57,** 1878–1886.

Hudson A. P., Cuny G., Cortadas J., Haschemeyer A. E. V. and Bernardi G. (1980) An analysis of fish genomes by density gradient centrifugation. *Eur. J. Biochem.* **112,** 203–210.

John B. and Miklos G. L. G. (1979) Functional aspects of satellite DNA and heterochromatin. *Int. Rev. Cytol.* **58,** 1–114.

Karel W. J. and Gold J. R. (1987) A thermal denaturation study of genomic DNAs from North American minnows (Cyprinidae: Teleostei). *Genetica* **74,** 181–187.

Karel W. J. and Gold J. R. (1988) DNA base composition and nucleotide distribution among fifteen species of teleostean fishes. *Comp. Biochem. Physiol.* **90B,** 715–719.

Kit, S. (1962) Species differences in animal deoxyribonucleic acids as revealed by equilibrium sedimentation in density gradients. *Nature* **193,** 274–275.

Maniatis T., Fritsch E. F. and Sambrook J. (1982) In *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor (NY): Cold Spring Harbor Laboratory.

Messing J. (1983) New M13 vectors for cloning. *Methods Enzymol.* **101,** 20–78.

Miklos G. L. G. and Gill A. C. (1982) Nucleotide sequences of highly repeated DNAs; compilation and comments. *Genet. Res.* **39,** 1–30.

Moyer S. P. (1986) Isolation and characterization of a highly repeated satellite DNA sequence from the cyprinid fish *Notropis lutrensis*. MS thesis, Texas A&M University, College Station, Texas.

Schmidtke J., Schmitt E., Leipoldt, M. and Engel W. (1979) Amount of repeated and non-repeated DNA in the genomes of closely related fish species with varying genome sizes. Comp. Biochem. Physiol. **64B,** 117–120.

Singer M. F. (1982) Highly repeated sequences in mammalian genomes. *Int. Rev. Cytol.* **76,** 67–112.

Sueoka N. and Chang T. Y. (1962) Natural occurrence of a deoxyribonucleic acid resembling the deoxyadenylate-deoxythymidylate polymer. *Proc. natn. Acad. Sci. USA* **48,** 1851–1856.

Vieira J. and Messing J. (1982) The pUC plasmids, a M13 mp7 derived system for insertion mutagenesis and

sequencing with synthetic universal primers. *Gene* **19,** 259–268.

Walker P. M. B. and McLaren A. (1965) Fractionation of mouse deoxyribonucleic acid on hydroxyapatite. *Nature* **208,** 1175–1179.

Yunis J. J. and Yasmineh W. G. (1971) Heterochromatin, satellite DNA, and cell function. *Science* **174,** 1200–1209.

Zimmer E. A., Riven C. J. and Walbot V. (1981) Isolation of DNA and DNA recombinants from maize. *Plant molec. Biol. Newsletter* **2,** 93–96.