Heteroplasmy, Length and Sequence Variation in the mtDNA Control Regions of Three Percid Fish Species (*Perca fluviatilis*, *Acerina cernua*, *Stizostedion lucioperca*)

Camilla Lothe Nesbø,[1] Mohammed O. Arab and Kjetill S. Jakobsen

Division of General Genetics, Department of Biology, University of Oslo, N-0315 Oslo, Norway

Manuscript received October 12, 1997
Accepted for publication January 7, 1998

ABSTRACT

The nucleotide sequence of the control region and flanking tRNA genes of perch (*Perca fluviatilis*) mtDNA was determined. The organization of this region is similar to that of other vertebrates. A tandem array of 10-bp repeats, associated with length variation and heteroplasmy was observed in the 5′ end. While the location of the array corresponds to that reported in other species, the length of the repeated unit is shorter than previously observed for tandem repeats in this region. The repeated sequence was highly similar to the Mt5 element which has been shown to specifically bind a putative D-loop DNA termination protein. Of 149 perch analyzed, 74% showed length variation heteroplasmy. Single-cell PCR on oocytes suggested that the high level of heteroplasmy is passively maintained by maternal transmission. The array was also observed in the two other percid species, ruffe (*Acerina cernua*) and zander (*Stizostedion lucioperca*). The array and the associated length variation heteroplasmy are therefore likely to be general features of percid mtDNAs. Among the perch repeats, the mutation pattern is consistent with unidirectional slippage, and statistical analyses supported the notion that the various haplotypes are associated with different levels of heteroplasmy. The variation in array length among and within species is ascribed to differences in predicted stability of secondary structures made between repeat units.

MITOCHONDRIAL DNA heteroplasmy is at least a transient stage in the propagation of any new mutation in the organelle DNA. While heteroplasmy for point mutations is rare (Moritz *et al.* 1987), heteroplasmy for length variants frequently occurs in natural populations (*e.g.*, Bentzen *et al.* 1988; Wilkinson and Chapman 1991; Árnason and Rand 1992; Brown *et al.* 1992; Broughton and Dowling 1994; Stewart and Baker 1994; Fumagalli *et al.* 1996; Mundy *et al.* 1996). These intraspecific and intraindividual length polymorphisms are most often observed in tandem repeated structures in the control region (reviewed in Moritz *et al.* 1987; Rand 1993). The tandemly repeated arrays are typically located in the parts of the control region that coincide with the 5′ end of the D-loop DNA where replication is initiated, and the 3′ end where the D-loop DNA is terminated (Lee *et al.* 1995). Moreover, the repeated sequences are usually associated with secondary structures (*e.g.*, Wilkinson and Chapman 1991; Árnason and Rand 1992; Lee *et al.* 1995). The precise mechanisms causing mtDNA length variation heteroplasmy are not known; however, several models have been suggested (*e.g.*, Buroker *et al.* 1990; Brown *et al.* 1996).

Corresponding author: K. S. Jakobsen, Division of General Genetics, Department of Biology, University of Oslo, P.O. Box 1031 Blindern, N-0315 Oslo, Norway. E-mail: k.s.jakobsen@bio.uio.no

[1]*Present address:* Department of Toxicology, National Institute of Occupational Health, P.O. Box 8149 Dep., N-0033 Oslo, Norway.

MtDNA length variations, caused by tandem repeats, have previously been identified in a number of fish species: several species of sturgeon (*e.g.*, *Acipenser transmontanus*, *Acipenser oxyrhynchus desotoi*) (Brown *et al.* 1992; Miracle and Campton 1995); cod (*Gadus morhua*) (Johansen *et al.* 1990; Árnason and Rand 1992); European sea bass (*Dicentrarchus labrax*) (Cecconi *et al.* 1995); and American shad (*Alosa sapidissima*) (Bentzen *et al.* 1988). With few exceptions, the repeats in these species tend to be long (*e.g.*, 81 bp in sturgeon, 1.5 kbp in American shad), and located in the 5′ end of the control region flanking the *trnP* gene.

The order Perciformes is among the most species-rich vertebrate groups, and thus provides an opportunity to study the evolutionary stability of mitochondrial arrays. In a previous study, Refseth *et al.* (1998) have investigated the population structure and biogeographic history of perch (*Perca fluviatilis*) by sequencing the control region of perch mtDNA from several Norwegian and Swedish populations. The results revealed relatively low levels of mtDNA sequence variation and high levels of geographical structuring among Scandinavian perch populations. In this paper, using direct sequencing and TA-cloning of PCR products, we demonstrate the presence of length variation heteroplasmy in perch mtDNA due to a 10-bp tandemly repeated sequence. We have addressed the origin and maintenance of heteroplasmy among perch populations by analyzing a large sample ($n = 149$) for the occurrence of heteroplasmy, the phylogenetic relationship among different units and arrays,

and the distribution of array lengths. Furthermore, to investigate the mechanisms maintaining heteroplasmy, we have analyzed mature oocytes from heteroplastic females. The evolutionary stability of the array was assessed by sequencing the 5′ part of the control regions of two other species belonging to the Percidae family, zander (or pike perch; *Stizostedion lucioperca*) and ruffe (*Acerina cernua*), as well as other fish species both within and outside the order Perciformes.

Our data demonstrate that most perch show heteroplasmy for different array lengths. We also observed similar arrays in both zander and ruffe. Moreover, results from cloning of PCR products suggest that units are added and deleted by unidirectional slippage, and that maternal transmission plays a major role in the maintenance of heteroplasmy.

MATERIALS AND METHODS

Sample collection and preparation: A total of 143 perch, four ruffe, and two zander were used. The perch used in this study were collected from 18 east Norwegian and four Swedish populations. 132 of the perch samples were also used in Refseth *et al.* (1998). In addition, three oocytes were analyzed from each of two perch.

Tail fin or liver was removed in field and stored in 96% ethanol or at −80°. Mature oocytes were removed from two perch specimens and stored in 96% ethanol. Total genomic DNA was extracted using a standard proteinase K phenol-chloroform method with ethanol precipitation (Sambrook *et al.* 1989), or by the DNA-direct kit (Dynal AS, Oslo, Norway) (Rudi *et al.* 1997).

PCR on total DNA: Amplifications were carried out in a 50-μl final volume containing 1–5 ng template DNA, 1× PCR buffer, 2.5 mm $MgCl_2$, 1 μmol of each primer, and 0.5 units Taq-polymerase (Advanced Biotechnology, Surrey, UK). The following primers were used to amplify segments of the mtDNA control region: HV2: TTCCCCGGTCTTGTAAACC (modified from Hoelzel *et al.* 1991); CSB-2: AAACCCCCC TACCCCCC (Shedlock *et al.* 1992); 12SR: CGGTGACTTG CATGTGTAATGTCA (Refseth *et al.* 1998); CSB-3: TAT TCCTGTTTCCGGGG; and CSB-D: GGAACCAAATGCCAG GAA (designed from *Champsochromis spilorhychus*, Lee *et al.* 1995). The locations of the primers in the control region are shown in Figure 1. The HV2 primer is located in the *trnT* gene, the CSB-2, -3, and -D span the conserved sequence block (CSB) 2, 3, and D, respectively, and the 12SR primer is located in the 12SRNA gene. For direct sequencing of PCR products, one of the primers was biotinylated. The reactions were submitted to an initial denaturation at 96° for 5 min, and then 30 cycles each consisting of denaturation at 96° for 1 min, annealing at 51° (HV2-CSB3), 54° (HV2-CSBD), 55° (CSB2-12SR), or 46° (HV2-12SR) for 1 min, and extension for 2 min at 72°. The reaction was performed on a Biometra Trio-Thermo block TB1 (Biomedizinische Analytik GmbH, Göttingen, Germany). Biotinylated PCR products were sequenced by a direct solid phase approach using Dynabeads M-280 Streptavidin (Dynal AS) (Hultman *et al.* 1989), applying the nonbiotinylated PCR-primer as sequencing primer.

Cloning of PCR products and detection of heteroplasmy: Nonbiotinylated PCR products from 11 adult perch, three mature perch oocytes, and two ruffe were cloned using the pGEM-T Vector System (Promega, Madison, WI). From each ligation between eight and 10 individual clones were PCR-amplified. Sequencing of the PCR products was performed either on an ABI 373 DNA Sequencer using ABI PRISM Dye Primer Cycle Sequencing Kit (Applied Biosystems, Foster City, CA), or on a Vistra DNA Sequencer 725 using Thermo Sequenase sequencing kit (Molecular Dynamics and Amersham Life Science, Buckinghamshire, UK).

Data analysis: Sequence analysis was performed using the GCG package of computer programs (Version 7.0; Genetics Computer Group Inc., Madison, WI). Database searches were carried out using the FASTA program. Localization of tRNA genes and conserved sequences were done using BESTFIT with sequences from a cichlid (*Champsochromis spilorhychus*; acc. no. U12553), European sea bass (*Dicentrarchus labrax*; acc. no. X81472), tuna (*Thunnus thynnus*; acc. no. X82653), cod (*Gadus morhua*; acc. no. X99772), minnow (*Cyprinella spiloptera*; acc. no. L07753), rainbow trout (*Oncorhynchus mykiss*; acc. no. S68946), Arctic char (*Salvelinus alpinus*; acc. no. X68659), white sturgeon (*Acipenser transmontanus*; acc. no. X54348), a lungfish (*Protopterus dolli*; acc. no. L42813), *Xenopus laevis* (acc. no. M13046), chicken (*Gallus domesticus*; acc. no. X52392), and humans (*Homo sapiens*; acc. no. V00662).

Potential secondary structures in the DNA were analyzed using the FOLDRNA program in GCG and visualized using *loopDloop* (Gilbert 1992a). The multiple alignment of nucleotide sequences was carried out with the PILEUP program in GCG and SeqApp (Gilbert 1992b). Statistical analyses were performed using the insight mode of the SAS program package (SAS 6.12 for windows; SAS Institute Inc., Cary, NC).

Phylogenetic analyses were performed in PAUP Version 3.1.1 (Swofford 1991). The maximum parsimonious relationship among the three percid species was estimated using the "branch and bound"-search algorithm in PAUP. To estimate the phylogenetic relationship among perch control region haplotypes we used the algorithm described in Templeton *et al.* (1992). This cladogram estimation procedure is specifically designed to estimate intraspecific gene trees where most of the nodes (haplotypes) are present in the population under investigation and are few mutations apart. A set of cladograms that have a high probability (>95%) of being true is obtained. The resulting cladograms can be converted into a nested design as described in Templeton *et al.* (1987) and Templeton *et al.* (1993). The units defining the various nested branches of the cladogram are called *n*-step clades, where *n* indicates the number of mutational steps necessary to define the clade. 0-step clades refer to haplotypes. Given the *n*-step clades, the $n + 1$ clades are defined by the union of all *n*-step clades that can be joined together by moving back one mutational step from the terminal *n*-step clades. Hence, haplotypes are grouped into 1-step clades (*i.e.*, 1-1, 1-2, 1-3, and 1-4), and these clades are further grouped into 2-step clades (*i.e.*, 2-1 and 2-2 in Figure 3).

RESULTS

The perch mtDNA control region; conserved sequences and homology to other vertebrate mtDNA: The complete sequence of the perch control region and its flanking tRNA genes has been determined. Figure 1 depicts overall structure of the perch control region including the conserved sequence blocks (CSBs), tRNA genes, and the repeated array. The nucleotide sequence of the perch control region is aligned with the 5′ parts of the zander and ruffe control regions in Figure 2. The perch mtDNA control region revealed a length of 882 bp. The general organization of the perch control re-

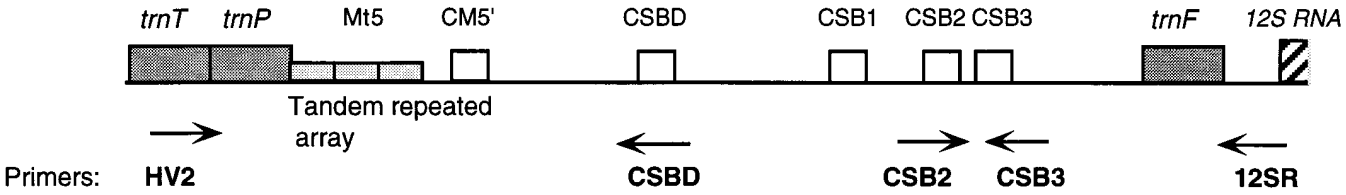

Figure 1.—Overview of the perch mtDNA control region and flanking tRNA genes. Conserved sequence elements and tRNAs are indicated by boxes. Primer positions are indicated by arrows. The HV2-CSB3 primer combination gave rise to an approximately 800-bp long fragment, and the HV2-CSBD combination gave rise to an approximately 550-bp fragment. The PCR product obtained using the CSB2 and 12SR primer was 300 bp. For ruffe and zander only the HV2-CSBD combination was used.

gion was similar to that reported for other vertebrates (*e.g.*, Chang and Clayton 1984; Foran *et al.* 1988; Buroker *et al.* 1990; Johansen and Johansen 1993; Broughton and Dowling 1994).

Conserved regulatory elements: Three TAS (Termination Associated Sequence) motifs—TACAT—were found in the 5′ part of the control region, indicated in Figure 2 as TAS-1, -2, and -3. A motif similar to the conserved motif in the 5′ region of lungfish and other vertebrates (Zardoya and Meyer 1996) was also identified (CM5′ in Figure 2 and Table 1). This motif forms a potential stemloop ($\Delta G = -4.4$ kcl/mol) and the most conserved parts are the nucleotides predicted to form the stem, suggesting that this is a conserved secondary structure that may contribute to TAS function.

A motif with high similarity (52% total identity and 100% identity at the last five bases) to human CSB1 is present at position 772–794 (Figure 2 and Table 1). Somewhat surprisingly, the CSB1 element has been reported present in only a few fish species, *e.g.*, minnow (Broughton and Dowling 1994) and sturgeon (Buroker *et al.* 1990). However, alignment of the perch D-loop with other fish demonstrates that this element is also found in other species (Table 1), and thus seems to be universally distributed among fish species.

In vertebrates, the H- and L-strand promoters are located near the *trnF* gene. In perch, only one segment with homology to previously reported promoter sequences (reviewed in Tracy and Stern 1995) was found (Figure 2; Table 1), indicating a single putative bidirectional promoter. The only mitochondrial bidirectional promoter reported so far is in chicken (L'Abbé *et al.* 1991; Tracy and Stern 1995), and the perch promoter is similar to this sequence (Table 1). Moreover, the perch sequence is able to form a cruciform structure similar to that in chicken. The distance between CSB3 and the *trnF* gene is only 68 bases, which is considerably shorter than what is previously reported for other fish (*e.g.*, Lee *et al.* 1995).

Finally, a tandemly repeated array located between the TAS 1 and *trnP*, composed of two to five repeated units in perch, six units in zander, and six to 23 units in ruffe, was observed. The repeated sequence is highly similar to a putative control element, Mt5 (Figure 2; Table 1), observed in several vertebrates including humans (Ohno *et al.* 1991). This element has been shown to specifically bind a protein possibly involved in regulation of D-loop DNA termination (Kumar *et al.* 1995).

The anatomy of the tandem repeats of percid fish: *A short 10-bp tandem repeated sequence in the 5′ end of the percid control region:* All individuals examined among the three percid species studied possessed the tandemly repeated array consisting of 10-bp repeats, between the *trnP* gene and TAS 1 (Figure 2). Throughout this paper the repeat unit nearest the *trnP* gene is referred to as the first repeat. The remaining perfect repeats are referred to as second, third, fourth, or array repeats. Furthermore, in all perch sequenced, the array was flanked by a 3′ degenerated repeat unit (hereafter referred to as the last repeat) that differed by several mutations from the "standard" perch array unit (Figure 2).

The first repeat unit extends four bases into the *trnP* gene. In general, the repeat units are highly similar both within and among the three species (identical in ruffe and zander; see Figure 2 and Table 2). The repeats contain perfect palindromic sequences (TTGCAA) and imperfect palindromes (AA(G)TATT).

In perch, two types of first and second repeat were observed (Table 2). Among the 149 individuals sequenced, this variable position defining the A- and T-type arrays (*i.e.*, A or T at position 6 in the first repeat, Table 2), and an identical substitution in the second repeat unit, were the only sequence polymorphisms observed in the array (see also Refseth *et al.* 1998). Similar levels of intraspecific conservation have also been observed in several other species (*e.g.*, Buroker *et al.* 1990; Brown *et al.* 1996). The sequence of the last degenerated repeat also varies between individuals, and, using the same haplotype definitions as in Refseth *et al.* (1998), can be separated into two major classes of units: GC (last unit 1) and TT (last unit 2J; 2D, G, L, M; 3H; and 3I) (see Table 2). The two substitutions defining the TT and GC groups (G_{130} $C_{131} \Leftrightarrow T_{130}$ T_{131}) apparently represent convergent mutations. The T_{130} T_{131} are present in the haplotypes D, G, L, and M (Figure 3). These haplotypes are phylogenetically separated by haplotypes possessing GC arrays (Figure 3), indicating recurrent mutations at these sites. Supporting this, both GC and TT arrays were found in about equal amounts within the same individual (Eikeren #3) (see below; Table 3). The last

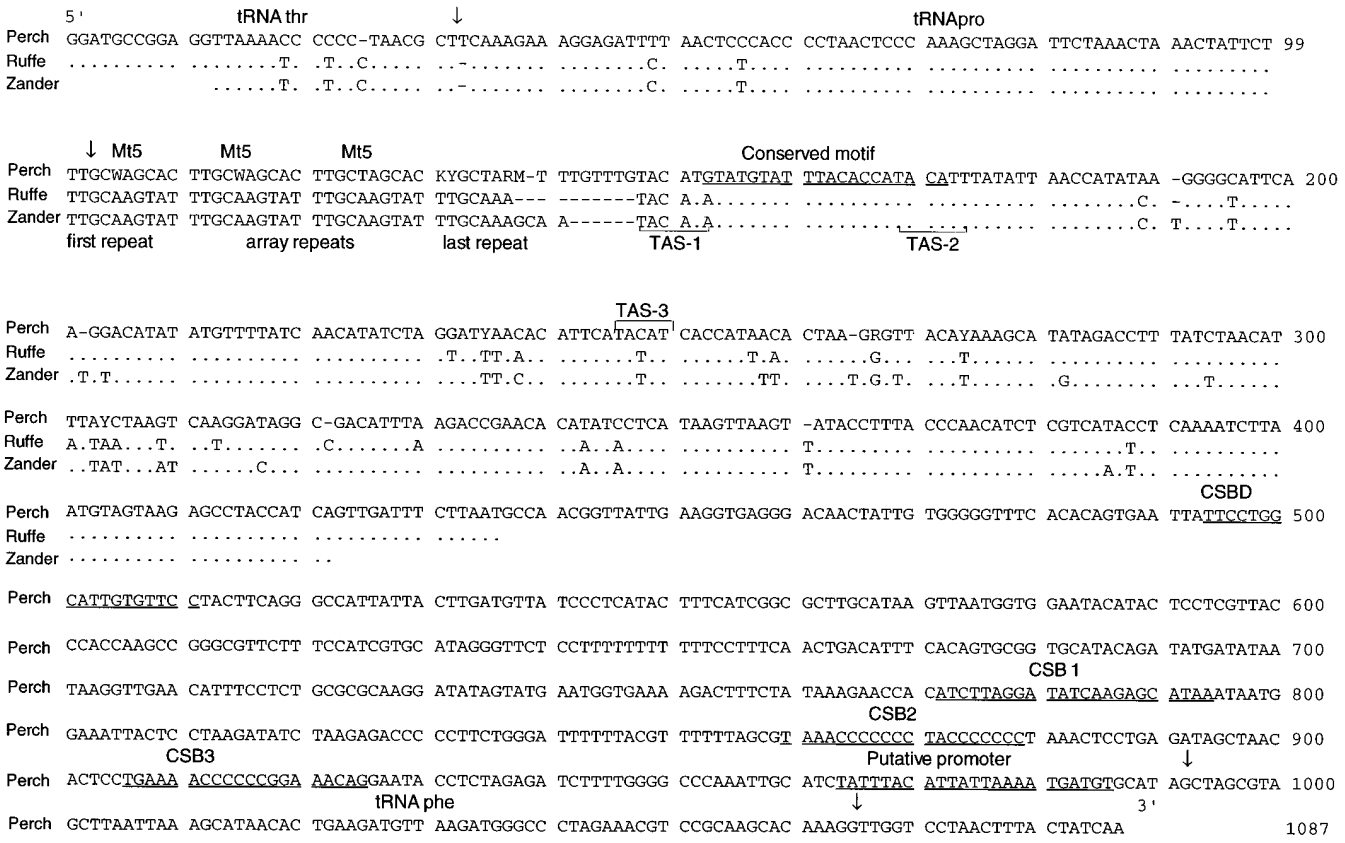

Figure 2.—The perch mtDNA control region, including the flanking tRNAs, aligned with the sequenced fragments from ruffe and zander. Identical positions are indicated by dots (except for the tandem repeated array), and indels by dashes. The sequence corresponds to the L-strand of the human mtDNA (Anderson *et al.* 1981). The perch sequence is 1080 bp. Conserved sequence motifs are underlined. The borders of the tRNAs are indicated by arrows. In ruffe, the repeated array between *trnP* and TAS1 varied between six and 23 units, and in zander the array consisted of six units. Only four units (including the 3′ degenerated unit) are shown to obtain the optimal alignment with the perch sequence. The sequences are submitted to GenBank's databases under acc. no. Y14724-6.

repeat unit in the ruffe and zander array did not differ as much from the consensus unit as the perch last unit (Table 2).

Arrays containing more than one repeat unit are predicted to form stable secondary structures: Since secondary structures have been implied in formation and maintenance of tandemly repeated structures (see Buroker *et al.* 1990; Wilkinson and Chapman 1991; Árnason and Rand 1992; Brown *et al.* 1992; Broughton and Dowling 1994; Fumagalli *et al.* 1996), the stability of potential secondary structures was predicted by the FOLDRNA algorithm in GCG program package. As can be inferred from the presence of palindromic sequences in the repeats, two or more single-stranded units were able to form stable secondary structures in all three species studied (Figure 4). The stability of the predicted structures varied, however. The perch internal units are predicted to produce more stable structures compared to the zander / ruffe units (Figure 4A). Moreover, in perch, the T-type first repeat forms stronger secondary structures than the A-type repeat (Figure 4B). The degenerated last units are also able to form stable secondary structures when folded with an internal unit, but the estimated ΔG values suggest lower stability (Figure 4B).

Phylogenetic topologies based on the repeat unit and the control region are congruent: In order to investigate the evolutionary stability of the percid tandem repeated sequence, we estimated the phylogenetic relationship both among the flanking control region sequences and the different repeat unit sequences observed.

Topologies based on control region sequences: In this analysis we also sequenced the control region of American yellow perch (*Perca flavescens*; acc. no. Y14728), as well as spotted wolffish (*Anarhichas minor*; family Anarhicadidae, acc. no. Y14775), and flounder (*Plantichthys flesus*; order Pleuronectiformes, acc. no. Y14730), to obtain outgroup information. The tree obtained by applying the "branch and bound"-search algorithm in PAUP, with results from 1000 branch and bound bootstrap replicates is presented in Figure 5A. High bootstrap values were obtained for both perch-yellow perch and ruffe-zander monophyletic groups. However, when omitting the outgroup species, only the perch-yellow perch cluster was supported (bootstrap value at 100%, data not

TABLE 1

Comparison of censerved sequence elements in the percid mtDNA control region to elements in other species

Element	Species	Sequence	Position[a]
Mt5	Perch	Y T T G C A W G C A	100–129[b]
	Ruffe and zander	T T T G C A A G T A	100–129[b]
	Human[c]	A T G C T T A C A A G C A A G	16194–16208
CM5′[d]	Perch	G T A T G T A T T T A C A C C A T A C A	152–171
	Ruffe	A T A T G T A T T T A C A C C A T A C A	152–171
	Zander	A T A T G T A T T T A C A C C A T A C A	152–171
	Lungfish[c]	C T A T G T A T A T C G T A C A T T A A	15537–15557
CSB D	Perch	T T C C T G G C A T T T G G T T C C	494–511
	"Cichlid-a"[c]	T T C C T G G C A T C T G G T T C C	371–388
CSB 1	Perch	A T C T T A G G A T A T C A A G A G C A T A A	772–794
	"Cichlid-a"	C A T A A C T G A T A T C A A G A G C A T A A	640–662
	Salmonids[c]	A T A C T T G G A T A T C A A G T G C A T A A	726–742
	Cod	A T T A A A G T T T T T C A A G A G C A T A A	761–783
	Lungfish	A T C A T A T T T C A C A G T G A A C A T A A	16217–16239
	Human[c]	T A A T T A A T G C T T G T A G G A C A T A A	209–231
CSB 2	Perch	T A A A C C C C C C C T A C C C C C C C	860–879
	Human	C A A A C C C C C C C T – C C C C C	299–315
CSB 3	Perch	T G A A A A C C C C C C G G A A A C A	906–924
	Human	T G C C A A A C C C – – A A A A A C A	346–363
Promoter	Perch	T A T T T A C A – T T A T T A A A A T G A T G T	964–976
	Xenopus laevis[d]	A C A R T T A T A	2130–2132 and 2183–2190
	Chicken[d,e]	A T A T A C A – T T A T T	1072–1083

[a] The position given refers to the original sequence (accession numbers are given in materials and methods), or the sequence in Figure 2 for the percids.

[b] This motif was repeated two to five times between the *trnP* and TAS 1 in perch, six times in zander, and between six and 23 times in ruffe. A three-unit sequence is shown in Figure 2.

[c] The cod sequence was obtained from Johansen *et al.* (1990), the lungfish sequence from Zardoya and Meyer (1996), the cichlid sequence refers to the cichlid-a sequence from Lee *et al.* (1995), the salmonid sequence is a consensus of the sequences in Shedlock *et al.* (1992), the human CSB-elements were taken from Chang and Clayton (1984) and Foran *et al.* (1988), the human Mt5 sequence was obtained from Ohno *et al.* 1991, and the Xenopus and chicken promoters were obtained from Bogenhagen and Yoza (1986), and L'Abbé *et al.* (1991), respectively.

[d] Conserved motif in the 5′ end of the control region (Figures 1 and 2).

[e] Only the sequence homologous to the perch and Xenopus is included for clarity.

TABLE 2

Observed repeat units in the percid arrays

Repeat	Sequence	Repeat designation
Perch first repeats	T T T G C A A G C A	First A
	T T T G C T A G C A	First T
Perch array repeats	C T T G C T A G C A	Array
	C T T G C A A G C A	Second A
Perch imperfect last repeats[a]	C G C G C T A A A	Last 1 (GC)
	C T C G C T A A A	Last 2J (TT)
	C T T G C T A A A	Last 2D, G, L, and M (TT)
	C T T G C T A G A	Last H (TT)
	C T T G C T A G C	last I (TT)
Ruffe and zander first repeat and array repeat	T T T G C A A G T A	Ruffe and zander array[b]
Zander last repeat	T T T G C A A G C A A	Zander last
Ruffe last repeat	T T T G C A A G	Ruffe last

[a] Five different 3′ imperfect units were observed in perch, which are divided into two groups, GC and TT.

[b] All array repeats were identical.

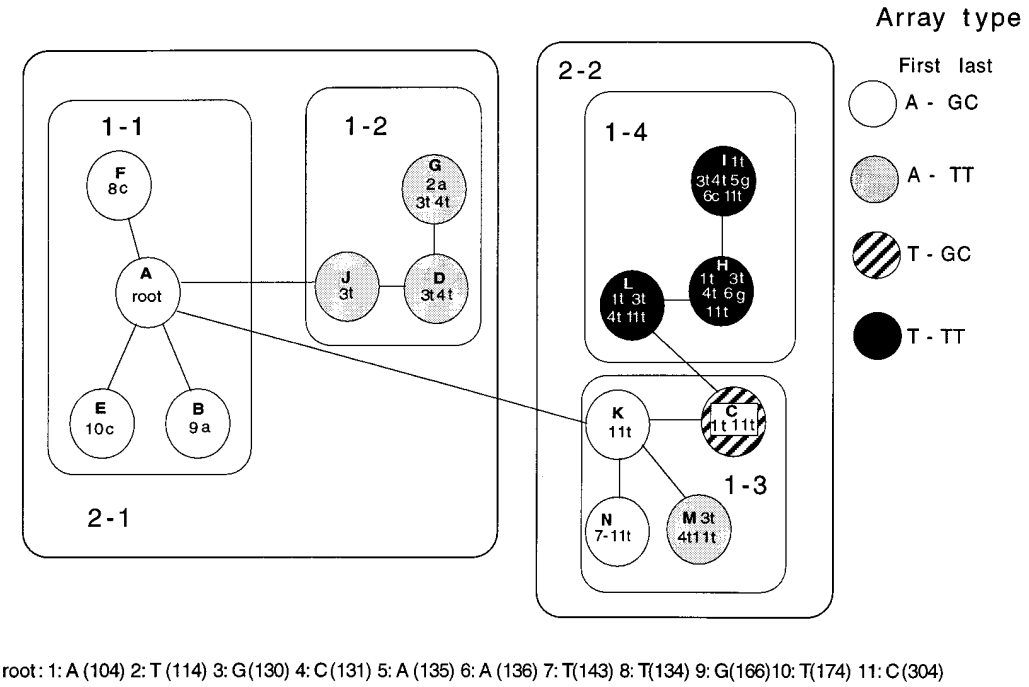

Figure 3.—The cladogram is modified from Refseth *et al.* (1998) and represents a 99% plausible set of haplotypes networks, estimated using the maximum parsimony algorithm given in Templeton *et al.* (1992). The cladogram was converted into a nested design as described in Templeton *et al.* (1987, 1993); haplotypes one mutation apart are grouped into 1-step clades (1-1, 1-2, 1-3 and 1-4), and these clades are further grouped into 2-step clades (2-1 and 2-2). The array types found in the haplotypes are indicated; *e.g.*, first A - last GC means that the first repeat was type A and the last degenerated repeat was type GC (Table 2). The A haplotype is given, and the number in brackets corresponds to the position in the sequence (Figure 2). Only variable sites are included.

shown). The same topology was also suggested when using only the flounder sequence as outgroup (bootstrap value at 85%, data not shown). In these analyses the tandem repeat was included. Leaving out the repeats, however, did not change the topology of the trees.

Phylogenetic relationships among different repeats: We also estimated the phylogenetic relationships among the different repeat units observed, combining the perch units and the ruffe/zander units (see Table 2). A total of 10 equally parsimonious trees were found, and the 50% rule consensus tree is shown in Figure 5B with the results from 1000 bootstrap replicates. The low bootstrap value, due to short sequences and low divergence, indicates a high degree of uncertainty associated with the suggested topology. However, the "majority-rule" values are high (between 67 and 100%). All perch last units are clustered together, and were connected to the remainder of the tree through the perch array unit. Moreover, the perch type A first unit is the one closest to the ruffe and zander sequences. This is in agreement with our previous findings based on intraspecific phylogenetic relationship and frequencies of perch haplotypes (see Figure 3 and Refseth *et al.* (1998). Hence, the PAUP analysis suggests that the ancestral unit resembles the ruffe/zander unit or the type A first perch repeat.

Evolutionary stability of the tandem repeats: The sequence obtained from American yellow perch shows that this species has an array similar to the perch A-GC type. The standard array length also here seems to be three units (not shown). However, sequences highly similar to the percid repeats were not found among previously reported mtDNA control regions, including other members of the order Perciformes (*e.g.*, sea bass, cichlids; as determined by FASTA searches in the GenBank data bases), possibly suggesting that this repeat is confined to the family Percidae. However, by sequencing a number of fish both within and closely related to the order Perciformes, a distantly related array was observed in flounder (Figure 6). In this species the repeated unit is 19 bases, seemingly composed of two percid units. Remnants of this repeat were observed in two other flatfishes, American plaice (*Hippoglossides platessoides*; acc. no. Y14727) and brill (*Scophtalamus rhombus*; acc. no. Y14729), in which we have sequenced this region (data not shown). Furthermore, an array, showing similarity to the flounder repeats, but composed of 74-bp units, is present in left eye flounder (*Paralichys olivaceus*; acc. no. AB000668).

Characterization of length variation and heteroplasmy associated with the tandem repeat: *Different array types show different levels of heteroplasmy:* Cloning of PCR products and sequencing of the individual clones revealed variation among molecules in copy number of the tandem repeats causing length variation and heteroplasmy in the mtDNA of perch.

Occurrence of heteroplasmy could also be detected

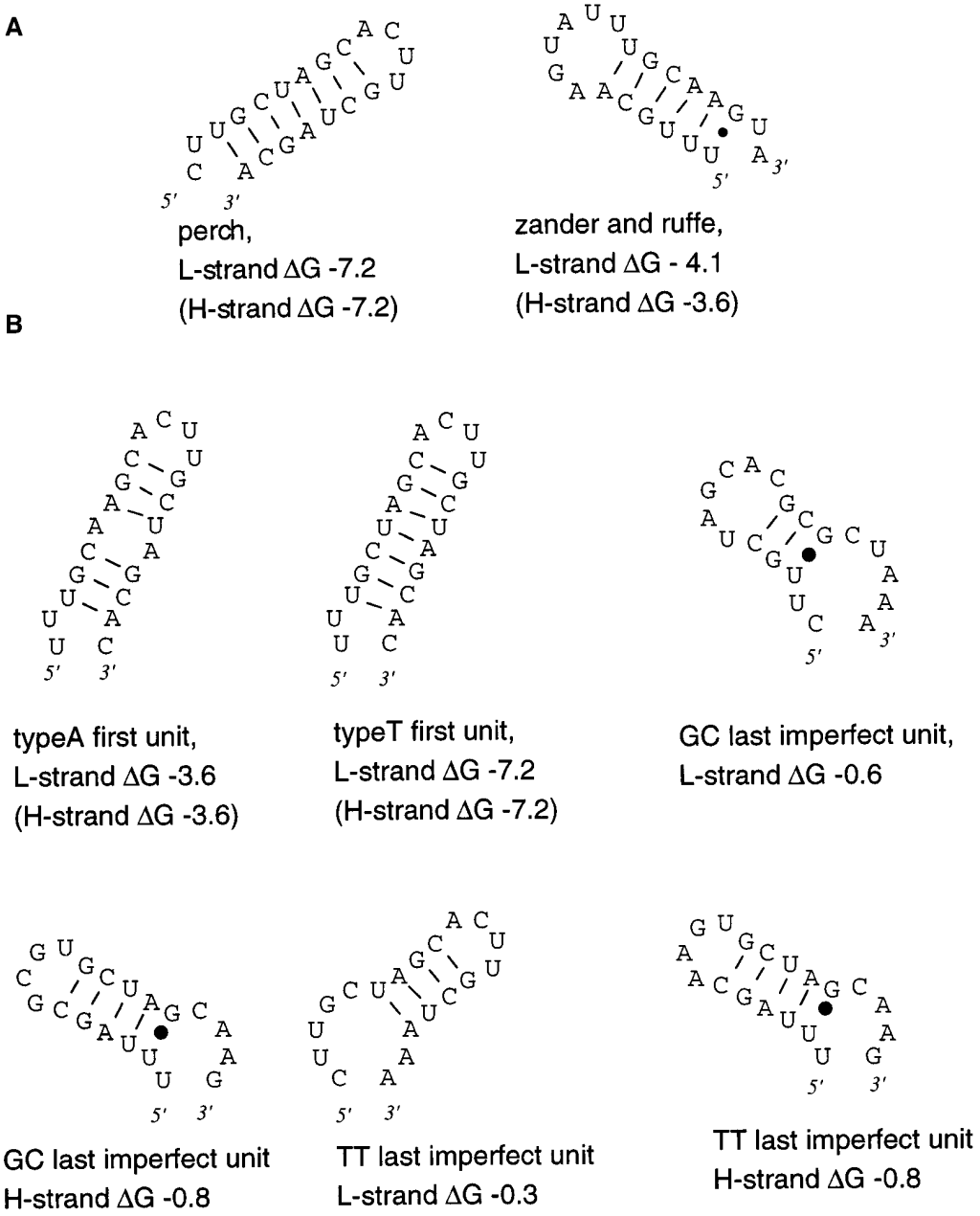

Figure 4.—Predicted secondary structures and their ΔG values (kcl/mol). (A) Two perch array-units and two zander/ruffe units. (B) Different types of perch first and last units folded with one array repeat. For all structures, except the last imperfect units, the sequences correspond to the L-strand, and the ΔG values for the H-strand are given in brackets. Note that the FOLD algorithm calculates ΔG values for RNA sequences.

by direct sequencing of PCR products. Many of the perch sequences displayed a double sequence on the autoradiogram; *i.e.*, at least two sequence ladders were superimposed on each other *after* the repeat array (Figure 7). Sequencing of individual clones obtained from double sequence PCR products always revealed at least two length variants. Furthermore, of the 20 clones obtained from two PCR products not showing double sequences, only one possessed a four-repeat array (the rest were three-repeat arrays, Suluvatn #9 and Store Lauarvann #6 in Table 3).

Analyzing the 149 directly sequenced perch PCR products revealed that 74% showed length variation heteroplasmy. Among the A-GC arrays (type A first unit and GC last degenerated unit) 80% showed heteroplasmy. 75% of the A-TT arrays, 68% of the T-GC arrays, and 50% of the T-TT arrays showed heteroplasmy. A log linear model (based on Poisson distribution and log link) suggested that these differences were significant (deviance/degrees of freedom = 1.02, $P = 0.0001$; see Table 3). The parameter estimates (Table 3) supported that there is a higher degree of heteroplasmy in individuals possessing GC last units in contrast to TT last units ($P = 0.023$). There was no significant effect of type of

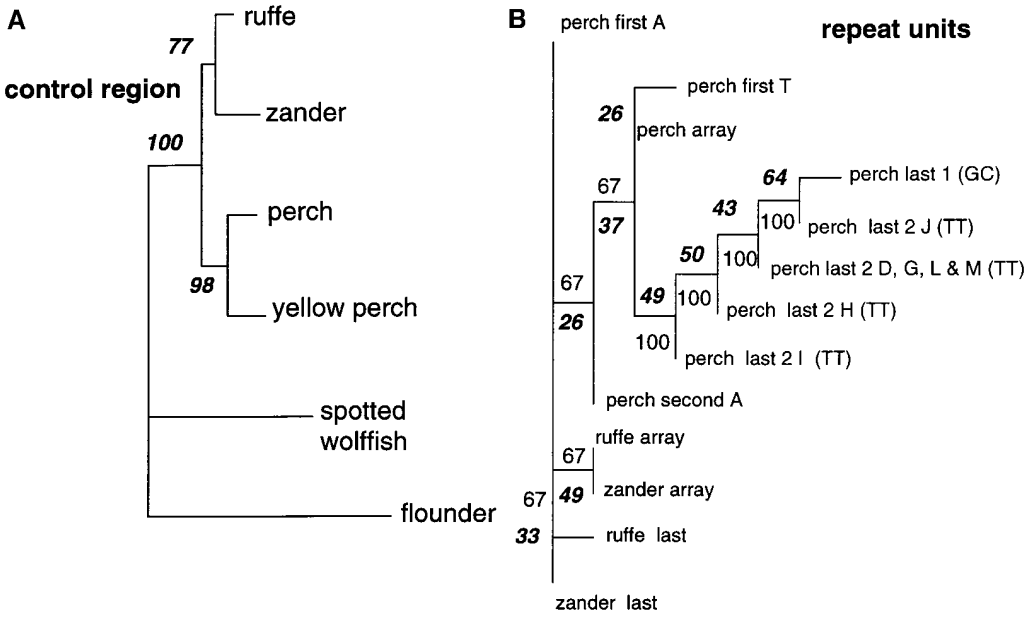

Figure 5.—Phylogenetic relationships inferred from the control region fragment sequenced in perch, zander, and ruffe (A), and among the observed repeat units (B). The bootstrap option of PAUP was used to demonstrate confidence in positions of tree nodes, and 1000 branch and bound bootstrap replicates were performed. The values obtained are shown in bold italic. (A) The most parsimonious tree based on the control region sequences. The yellow perch sequence (acc. no. Y14728) was also included. The consistency index (ci) was 0.933, and the retention index (ri) was 0.659. The flounder (acc. no. Y14730) and the spotted wolffish (acc. no. Y14775) sequences were used as outgroups. (B) The majority-rule consensus tree estimated using the repeat sequences. Ten equally parsimonious trees were obtained, and the percentages of the trees containing the branch are indicated in addition to the bootstrap values (roman and bold italic values, respectively). The ci for each most parsimonious tree was 0.8, and the ri was 0.909. The "midpoint rooting" option was used.

first repeat ($P = 0.09$) and no dependence between the first and last unit combined and the frequency of heteroplasmy (Chi square "first unit × last imperfect unit × homo/heteroplasmy" = 1.3, $P = 0.24$). The relationship between the first and last unit ($P = 0.072$) simply reflects that most of the sampled individuals possessed arrays with the A-GC combination.

We sequenced 8–10 individual clones from each of 14 different cloned PCR products and found that all individuals contain molecules with three-repeat units (Table 3). This was also generally the major array length variant. For haplotype L (see Figure 3), however, the major length variant was four- or five-unit arrays. Among the 14 PCR products, only one contained molecules with less than three repeats (one sequence from Ravalsjøen #5; Table 3). No directly sequenced PCR products showed "double sequences" before the third repeat unit, and all individuals classified as homoplastic possessed

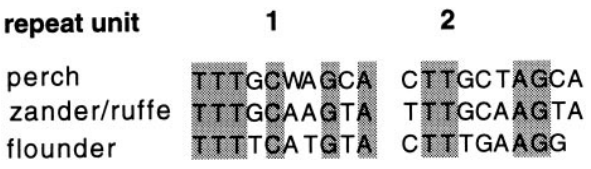

Figure 6.—Alignment of the percid and the flounder repeats. One flounder unit is aligned against two percid units.

three-unit arrays. This is consistent with the results from cloning; molecules with less than three repeats are rare.

The ruffe and zander arrays contained identical repeat unit sequences (Table 2). The length of the array varied, however. The two cloned ruffe PCR products possessed arrays with six, seven, and 10, and 16, 17, 18, 22, and 23 units, respectively (Table 4). Furthermore, two directly sequenced ruffe PCR products showed double sequences beyond the seventh unit. The two zanders sequenced both possessed six-unit arrays, and neither showed double sequences.

Mutation mechanisms in the repeated array inferred from the distribution of point mutations and potential secondary structures: Along with the length variation polymorphism, site heteroplasmy was observed in five of the cloned PCR products (see Table 3). Eikeren #3 possessed arrays with both TT and GC last units (Table 3). Since all other individuals from this location (and from the other populations within the same geographical region) possessed GC last units (Refseth *et al.* 1998), it seems most likely that the TT array arose from deleting the last GC unit and half the fourth unit in a four-unit array. Folding of a GC last repeat and the array repeat confirm this (Figure 4B); *i.e.*, formation of such a structure in the L-strand during H-strand replication, would result in an array with a TT last unit.

In Røysjø #8, Svartvann #4, Mjær #11 and Mjær #15

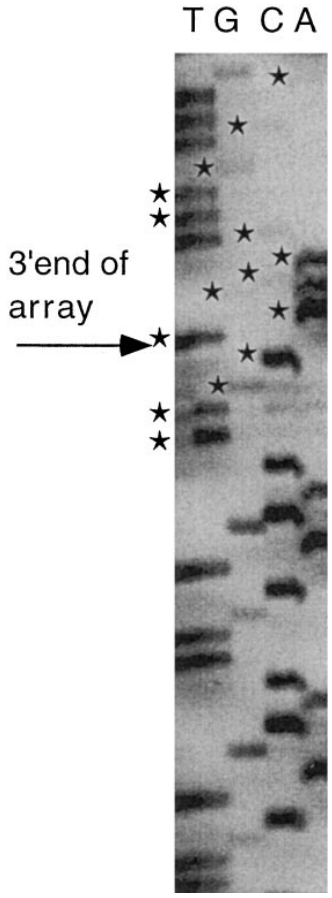

Figure 7.—Autoradiogram showing double sequences; *i.e.*, at least two sequence ladders superimposed on each other after the repeat array. The end of the array is indicated. The band of the "second" sequence is indicated by a ★ and reads: . . . TTGCTAGCACTTGCTA . . ., *i.e.*, at least one extra repeat unit.

another mutation occurred (Table 3); in Røysjø #8 all five-unit arrays possessed an A-type second unit while the three unit arrays possessed a T-type second unit. This mutation was also present in the two four-unit arrays observed in Svartvann #4 and Mjær #11, and in one of the two four-unit arrays in Mjær #15. Finally, this was the major type of array in one individual [three-unit array; haplotype G in Refseth *et al.* (1998); Figure 3]. All arrays with an A-type second repeat also had an A-type first repeat. Thus, this mutation presumably occurred by duplication of the first unit. Consequently, both incidents of site heteroplasmy can be explained by duplications and deletions of repeats.

The phylogenetic topology inferred from the different repeats (Figure 5B), the minimum spanning network in Figure 3, and Refseth *et al.* (1998) suggests that the type A perch first unit (*i.e.*, A at position 6 in the repeat, Table 2) represents the ancestral form. However, the perch standard array unit has a T in this position. Thus, the A ⇔ T transversion has spread to all units (*i.e.*, by concerted evolution), except to the first unit in 109 of the 149 perch individuals sequenced. This observation strongly suggests that the repeats are mainly duplicated and deleted by a 5′ to 3′ (relative to the sequence in Figure 2) unidirectional slippage mechanism.

Imperfect duplications or deletions of units were not observed. Moreover, the last degenerated unit was never duplicated. Taken together, all the above observations imply that the duplication and deletions of units mainly occur after two or three array units are replicated, and that the secondary structures formed primarily involve one array unit and the last degenerated unit. Furthermore, as noted by Brown *et al.* (1996) the high level of sequence conservation within the arrays implies high addition and deletion rates.

Transmission of heteroplasmy; distribution of array lengths in mature oocytes: To examine the possible mechanisms maintaining heteroplasmy, we analyzed the distribution of repeats in perch oocytes from a heteroplasmic individual by cloning of PCR products obtained from PCR on a single egg. The results are presented in Table 3. Both major length variants were transmitted to the three oocytes examined. In addition, PCR products obtained from three eggs from a different female were directly sequenced, and both the adult and the eggs showed double sequences (not shown). Thus, these results suggest that the high level of heteroplasmy observed may be caused by low levels of drift during oogenesis and maternal transmission of the heteroplastic condition.

DISCUSSION

Most repeats observed in the 5′ end of the D-loop tend to be composed of long units (Bentzen *et al.* 1988; Johansen *et al.* 1990; Brown *et al.* 1992; Miracle and Campton 1995). In contrast, the percid repeat units are only ten bases long. Due to the short length, movement of mutations in the perch array could easily be traced. This information is particularly valuable in the investigation of possible length mutation mechanisms. In addition, the short size of the units made it possible to compare the predicted stability of secondary structures associated with different arrays. Three patterns of length variation polymorphism were observed among the three species studied. In perch, no more than five units were found among the individuals studied, and three-unit variants were dominating. The two zander individuals both possessed six-unit arrays, and the ruffe possessed arrays with six or more units. Both the perch and ruffe showed extensive heteroplasmy.

Phylogeny and origin of the tandem repeat unit: The overall phylogenetic arrangement of the three species based on the repeated sequences alone is similar to that obtained using the entire control region fragment (Figure 5). The same topology was also suggested from phylogenetic analyses of the mtDNA cytochrome b genes from perch, yellow perch, zander, and ruffe (acc.

TABLE 3

Distribution of array lengths within the cloned PCR products

No. of units: Individual	2	3	4	5	6	7	10	16	17	18	22	23
Perch[a]												
Eikeren #3[b]		2	4	2								
Røysjø #8		5		4[c]								
Svartvann #4		9	2[c]									
Suluvatn #9[d]		10										
Suluvatn #8		1	8									
Stordammen #5		8	2									
Sandungen #5		3	5									
Mjær #11		6	2[c]									
Mjær #15		8	2[c]									
Store Lauarvann #6[d]		9	1									
Ravalsjøen #5[e]	1	5	4									
Ravalsjøen #5 egg #1[e]		5	4									
Ravalsjøen #5 egg #2[e]		7	2									
Ravalsjøen #5 egg #3[e]		8	1									
Ruffe												
Ruffe #10					5	3	2					
Ruffe #16								3	1	3	2	1

Values are no. of clones with arrays of a given length.

[a] All perch samples investigated for heteroplasmy by cloning of PCR products originated from lakes in southern Norway, and all except the Mjær individuals came from the western side of the Oslofjord (Refseth *et al.* 1998).

[b] The three-unit arrays possessed TT last units, while the rest possessed GC last units, *i.e.*, site heteroplasmy.

[c] The four- and five-unit arrays possessed a type A second repeat (see Table 2), *i.e.*, site heteroplasmy.

[d] These individuals did not appear heteroplastic when sequenced directly.

[e] Four PCR products were obtained from the same individual. DNA was isolated from tail fin tissue (Ravalsjøen #5) and three mature oocytes (Ravalsjøen #5 egg #1, 2, and 3).

nos. Y14776, AJ001521, AJ001512, and AJ001511, respectively) using the mackerel sequence as outgroup (*Scomber scombrus*, family Scombridae; acc. no. X81564; data not shown). This indicates that the sequence of the repeat unit possesses some phylogenetic information.

The localization of the array between two conserved secondary structures, CM5′ and the *trnP* gene, suggests that the array originated from a duplication event induced by secondary structures formed in one, or both, of these sequences. Furthermore, the relatively high level of sequence identity between the flatfish and the percids (Figure 6) indicates that the array originated either once early in the evolution of Perciformes and Pleuronectiformes lineages, or alternatively, it may have been formed several times independently by a common mechanism. We find it most likely that it has originated from a single event and has subsequently been lost in some lineages (*e.g.*, tuna, cichlids, sea bass).

Addition and deletion of repeats in the array: Several models have attempted to explain the cause and persistence of heteroplasmy (Rand and Harrison 1989; Buroker *et al.* 1990; Hayasaka *et al.* 1991; Madsen *et al.* 1993a; Broughton and Dowling 1994; Brown *et al.* 1996; Mundy *et al.* 1996; Wilkinson *et al.* 1997). The distribution of point mutations in the perch arrays and the predicted secondary structures suggest that the repeats are deleted and duplicated by unidirectional slippage. The percid repeat units differ in some respects compared to repeats observed in the 5′ end of the control region among other fish species; they are shorter and do not contain the TAS element. Since the TAS is thought to be involved in termination of the D-loop DNA [*i.e.*, a nascent H-strand terminated after a couple of hundred bases (*e.g.*, Madsen *et al.* 1993b)], the occurrence of a TAS sequence within the repeated sequence has been a major point in some of the suggested mutation mechanisms (*e.g.*, Buroker *et al.* 1990; Brown *et al.* 1996). However, the percid repeats show high similarity with the Mt5 element (see Table 1), which has been associated with protein binding and regulation of the termination of D-loop DNA (Ohno *et al.* 1991; Kumar *et al.* 1995).

Brown *et al.* (1996) proposed a biochemical model emphasizing selection against tandem repeats containing binding sites for proteins involved in D-loop DNA termination. This model predicts that selection should lead to array lengths skewed around one repeat, as observed in sturgeon species (Brown *et al.* 1996). However, as also observed among different bat species (Wilkinson *et al.* 1997), the persistence of arrays over two repeats long in the four percid species investigated suggests that this model cannot account for the array length distributions observed in this study.

The most crucial feature of the illegitimate elongation

TABLE 4

Estimated effects of the 5′ first and 3′ last unit on the level of heteroplasmy

Parameter[a]	Type III Wald test[b]	Parameter estimates[c]
First unit × last imperfect unit	0.072	
First A-last GC[d]		0.82 ± 0.46
First unit × homo/heteroplasmy	0.091	
First A-heteroplastic		0.73 ± 0.43
Last imperfect unit × homo/heteroplasmy	0.023	
Last GC-heteroplastic		1.05 ± 0.46

[a] The estimates were obtained from the log linear analysis of the model: ln (observed number of individuals) = first unit + last imperfect unit + homo/heteroplasmy + first unit × last imperfect unit + first unit × homo/heteroplasmy + last imperfect unit × homo/heteroplasmy. Only nonzero values are included.
[b] *P*-values from the type III (Wald) test for each parameter.
[c] Parameter estimates with standard deviation.
[d] A-type first repeat and GC-type last repeat.

model suggested by Buroker *et al.* (1990) is the inclusion of the repeated array in the D-loop DNA. This model emphasizes the competition between the D-loop DNA and the H-strand. Due to secondary structures in the D-loop DNA, or the H- and L-strand, units will be added or deleted, respectively (see Buroker *et al.* 1990). Apart from containing the Mt5 element, the percid arrays are flanked by two TAS sequences and the CM5′ between TAS1 and TAS2 (see Figure 2). Since the D-loop DNA in most vertebrates terminates 60–80 bases 5′ (relative to the sequence in Figure 2) to the TAS sequence (*e.g.*, Clayton 1982; Foran *et al.* 1988; Madsen *et al.* 1993b), the percid repeat array is most likely included in the D-loop DNA, which would be consistent with the model of Buroker *et al.*

The illegitimate elongation model predicts that with a minimum of two repeats needed for hairpin formation, the minimum number of units in the array is three (Árnason and Rand 1992). Two-unit molecules were nevertheless observed at low frequency in perch (Ravalsjøen #5, Table 3). However, this does not invalidate the model; formation of a secondary structure involving the last imperfect unit and one array repeat results in deletion of one unit (see Figure 4B). A similar structure in the D-loop DNA may account for the addition of one unit to a three unit molecule, hence explaining that most arrays were composed of three or four units. Furthermore, the illegitimate elongation model, as opposed to replication slippage, can explain the low occurrence of duplications of the first unit. The distribution of mutations in the array indicated that most duplications and deletions involved secondary structures between the last degenerated repeat and one array unit. This, taken together with the observation that most arrays were three or four units long, would imply that most perch D-loop DNAs are terminated after the replication of the last imperfect unit and two array units. Some of the D-loop DNAs must, however, also extend to the first unit, since the A-type first repeat was duplicated in some individuals.

However, other mechanisms causing length mutations must exist as the tandem repeat of cod fails to meet several criteria of the model of Buroker *et al.* (Árnason and Rand 1992). Deviation from the model was also observed in the perch array. According to the illegitimate elongation model, mutations can only move by duplications in the 5′ to 3′ direction (relative to the sequence in Figure 2; Buroker *et al.* 1990; Wilkinson and Chapman 1991). The transversion observed in the 38 perch individuals possessing a T-type first unit is therefore either a parallel substitution, or, perhaps more likely, this mutation has "jumped" by another mechanism. One possibility is that deletion of the first A-type unit has occurred during L-strand replication. All haplotypes (C, L, H, I) with this particular mutation are connected in the maximum parsimony network in Figure 3, suggesting that this is a rare mutation, and confirming that unidirectional slippage and illegitimate elongation is the major mechanism.

Mechanisms causing differences in level of heteroplasmy and mean array lengths among percid fish: Clark (1988) showed by deterministic models that subtle differences in mutation rate can be responsible for large differences in level of heteroplasmy. Brown *et al.* (1996) suggest that the relative magnitude of the mutation rates associated with deletion and addition of units plays a major role in determining distribution of copy numbers in sturgeon. Considering the special mode of replication in mitochondria (see Clayton 1991a,b for reviews), different probabilities for addition and deletion of units is to be expected. Assuming that secondary structures preferentially will form in the D-loop strand and that illegitimate elongation is the major mutation mechanism in the arrays, the critical step will be deletion of units. The ruffe/zander units form less stable secondary structures compared to perch units (see Figure 4A). Thus, we predict that the ability to delete units should be reduced, and the equilibrium array length should be longer compared to perch—and this is exactly what we observe.

The log linear model suggested that the last imperfect unit had the most significant effect on the differences in level of heteroplasmy. Individuals possessing GC last units were more heteroplastic than individuals possessing TT last units. As demonstrated in Figure 4B, the same ΔG value is obtained for both the GC- and TT-units when folding the H-strand sequences. Folding of the L-strand sequences however, gives a lower ΔG value for the GC last imperfect unit. This implies that the elevated frequencies of heteroplasmy are caused by a higher deletion rate, which might seem contradictory to our previous findings. However, due to the weaker secondary structures (see Figure 4), the deletion rate would be much lower, and will not be able to match the duplication of array units. Hence, we would expect arrays containing a type A first unit, since these units form weaker structures compared to T-type units, and a GC last unit to be the most variable, which was what we indeed observed. This also explains the observation that four or longer arrays as major type (observed in haplotype L in Figure 3) tended to be associated with T-TT arrays.

Similar mechanisms might also account for differences in length distribution and amount of heteroplasmy among tandem repeated arrays in other species. Fumagalli *et al.* (1996) reported different levels of heteroplasmy in two species of shrews; *Crocidura russula* showed higher levels of heteroplasmy than *Sorex araneus*, where the units form stronger secondary structures (ΔG of −16 compared to −7.5 in *C. russula*). However, the shrew units also contain TAS sequences, which are suggested to affect length distribution as well (Brown *et al.* 1996). The array length and level of heteroplasmy in these species might therefore be influenced by both factors.

The results from cloning PCR products obtained from oocytes suggests that maternal transmission of molecules with different length variants is at least partly responsible for the high levels of heteroplasmy among perch populations. High degree of heteroplasmy in gonad tissue compared to other tissue types has previously been observed in rabbits, and has been attributed to stabilizing selection (Casane *et al.* 1994, 1997). Low levels of drift during oogenesis, combined with high mutation rates, might thus be a general feature in the evolution of heteroplasmy.

Concluding remarks: Since the tandemly repeated array was observed in species representing both subfamilies in the family Percidae (Percinae: ruffe, perch and yellow perch; Lucioperchinae: zander), the array and the associated length variations are general features of percid mtDNAs. Investigating more percid species might therefore further elucidate the mutation mechanisms in the repeat arrays. Similar arrays, but composed of longer repeat units, have also been observed in flounder (19 bp) and Japanese flounder (74 bp), which belong to another order, Pleuronectiformes. The flounder repeat showed high similarity to the percid repeats and the 5′ end of the Japanese flounder repeat. Hence, surveying more species among both percids and flatfishes could unravel the mechanisms organizing short repeats into higher order structures.

We thank Kjartan Østbye, Tor Erik Nesbø, Unn Hilde Refseth, Andrew Lambertsson, Bill Davies, Knut Rudi, John E. Stacy and two anonymous reviewers for comments on earlier drafts of this work. We thank Unn Hilde Refseth for sequencing some of the clones from Mjær #11. We also want to thank Emelita Karlsen for excellent lab assistance and Tonje Fossheim for sequencing the yellow perch. This work was supported by grant no. 107622/420 from the Research Council of Norway (NFR) to K.S.J.

Noted added in proof: The complete control region of several percid species, among them also yellow perch and zander, has been published recently (Faber and Stepien 1997; Turner 1997) The repeat was found in all ancestral genera. In species with repeat unit similar to the ruffe and zander (A-type repeat), the motif is repeated at least seven times. The yellow perch in Faber and Stepien (1997), which had similar repeat sequence to perch, possessed three repeat units and one GC-imperfect last unit.

LITERATURE CITED

Anderson, S., A. T. Bankier, B. G. Barell, M. H. L. de Bruijn, A. R. Coulson *et al.*, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–465.

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

Bentzen, P., W. C. Leggett and G. G. Brown, 1988 Length and restriction site heteroplasmy in the mitochondrial DNA of American Shad (*Alosa sapidissima*). Genetics **118:** 509–518.

Bogenhagen, D. F., and B. K. Yoza, 1986 Accurate in vitro transcription of the *Xenopus laevis* mitochondrial DNA from two bi-directional promoters. Mol. Cell. Biol. **6:** 2543–2550.

Broughton, R. E., and T. E. Dowling, 1994 Length variation in mitochondrial DNA of the Minnow *Cyprinella spiloptera*. Genetics **138:** 179–190.

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

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

Buroker, N. E., J. R. Brown, T. A. Gilbert, P. J. O'Hara, A. T. Beckenbach *et al.*, 1990 Length heteroplasmy of Sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics **124:** 157–163.

Casane, D., N. Dennebouy, H. de Rochambeau, J. C. Mounolou and M. Monnerot, 1994 Genetic analysis of systematic mitochondrial heteroplasmy in rabbits. Genetics **138:** 471–480.

Casane, D., N. Dennebouy, H. de Rochambeau, J. C. Mounolou and M. Monnerot, 1997 Nonneutral evolution of tandem repeats in the mitochondrial DNA control regions of Lagomorphs. Mol. Biol. Evol. **14:** 779–789.

Cecconi, F., M. Giorgi and P. Mariottini, 1995 Unique features in the mitochondrial D-loop region of the European sea bass *Dicentrarchus labrax*. Gene **160:** 149–155.

Chang, D. D., and D. A. Clayton, 1984 Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. Cell **36:** 635–643.

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

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

Clayton, D. A., 1991a Nuclear gadgets in mitochondrial DNA replication and transcription. Trends Biochem. Sci. **16:** 107–111.

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

Faber, J. E., and C. A. Stepien, 1997 The utility of mitochondrial DNA control region sequence for analysing phylogenetic relationships among populations, species, and genera of the Percidae, pp. 129–143 in *Molecular systematics of fishes*, edited by T. D. Kocher and C. A. Stepien. Academic Press, San Diego.

Foran, D. R., J. E. Hixson and W. M. Brown, 1988 Comparison of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res. **16:** 5841–5861.

Fumagalli, L., P. Taberlet, L. Favre and J. Hausser, 1996 Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. Mol. Biol. Evol. **13:** 31–46.

Gilbert, D. G., 1992a *loopDloop*, a Macintosh program for visualizing RNA secondary structure. Published electronically on the Internet, available via anonymous ftp to ftp.bio.indiana.edu.

Gilbert, D. G., 1992b SeqApp; A biosequence editor and analysis application. Preliminary release 1.9a169. Apple Computer Inc.

Hayasaka, K., T. Ishida and S. Horai, 1991 Heteroplasmy and polymorphism in the major noncoding region of mitochondrial DNA of Japanese monkeys: association with tandemly repeated sequences. Mol. Biol. Evol. **8:** 399–415.

Hoelzel, A. R., J. M. Hancock and G. A. Dover, 1991 Evolution of the cetacean mitochondrial D-loop region. Mol. Biol. Evol. **8:** 475–493.

Hultman, T., S. Ståhl, E. Hornes and M. Uhlén, 1989 Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. Nucleic Acids Res. **17:** 4937–4946.

Johansen, S., and T. Johansen, 1993 The putative origin of heavy strand replication (oriH) in mitochondrial DNA is highly conserved among the teleost fishes. DNA Sequencing **3:** 397–399.

Johansen, S., P. H. Guddal and T. Johansen, 1990 Organization of the mitochondrial genome of Atlantic cod, *Gadus morhua*. Nucleic Acids Res. **18:** 411–419.

Kumar, S., H. Suzuki, S. Onoue, S. Suzuki, N. Hattori *et al.*, 1995 Rat mitochondrial mtDNA binding proteins to inter-specifically conserved sequences in the displacement loop region of vertebrate mtDNAs. Biochem. Mol. Biol. Int. **36:** 973–981.

L'abbé, D., J.-F. Duchainu, B. F. Lang and R. Morais, 1991 The transcription of DNA in chicken mitochondria initiates from one major bi-directional promoter. J. Biol. Chem. **266:** 10844–10850.

Lee, W.-J., J. Conroy, W. H. Howell and T. D. Kocher, 1995 Structure and evolution of teleost mitochondrial control region. J. Mol. Evol. **41:** 54–66.

Madsen, C. S., S. C. Ghivizzani and W. W. Hauswirth, 1993a *In vivo* and *in vitro* evidence for slipped mispairing in mammalian mitochondria. Proc. Natl. Acad. Sci. USA **90:** 7671–7675.

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

Miracle, A. L., and D. E. Campton, 1995 Tandem repeat sequence variation and length heteroplasmy in the mitochondrial DNA D-loop of the threatened Gulf of Mexico sturgeon, *Acipenser oxyrhynchus desotoi*. J. Hered. **86:** 22–27.

Moritz, C., T. E. Dowling and W. M. Brown, 1987 Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann. Rev. Ecol. Syst. **18:** 269–292.

Mundy, N. I., C. S. Winchell and D. S. Woodruff, 1996 Tandem repeats and heteroplasmy in the mitochondrial DNA control region of the loggerhead shrike (*Lanius ludovicianus*). J. Hered. **87:** 21–26.

Ohno, K., M. Tanaka, H. Suzuki, T. Ohbayashi, S. Ikebe *et al.*, 1991 Identification of a possible control element, Mt5, in the major noncoding region of mitochondrial DNA by intraspecific nucleotide conservation. Biochem. Int. **24:** 263–271.

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

Rand, D. M., and R. G. Harrison, 1989 Molecular population genetics of mtDNA size variation in crickets. Genetics **121:** 551–559.

Refseth, U. H., C. L. Nesbø, J. E. Stacy, L. A. Vøllestad, E. Field and K. S. Jakobsen, 1998 Genetic evidence for different migration routes of freshwater fish into Norway revealed by analysis of current perch (*Perca fluviatilis*) populations in Scandinavia. Mol. Ecol. (in press).

Rudi, K., M. Kroken, O. Dahlberg, A. Deggerdal, K. S. Jakobsen *et al.*, 1997 A rapid, universal method to isolate PCR-ready DNA using magnetic beads. BioTechniques **22:** 506–511.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Shedlock, A. M., J. D. Parker, D. A. Crispin, T. W. Pietsch and G. C. Brumer, 1992 Evolution of the salmonid mitochondrial control region. Mol. Phylogenet. Evol. **1:** 179–192.

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

Swofford, D. L., 1991 PAUP: Phylogenetic analysis using parsimony, Version 3.1.1. Computer program distributed by Illinois Natural History Survey, Champaign, Illinois.

Templeton, A. R., and C. F. Sing, 1993 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analysis with cladogram uncertainty and recombination. Genetics **134:** 659–669.

Templeton, A. R., E. Boerwinkle and C. F. Sing, 1987 A cladistic analysis of the phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila. Genetics **117:** 343–351.

Templeton, A. R., K. A. Crandall and C. F. Sing, 1992 A cladistic analysis of the phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:** 619–633.

Tracy, R. L., and D. B. Stern, 1995 Mitochondrial transcription initiation: promoter structures and RNA polymerases. Curr. Genet. **28:** 205–216.

Turner, T. F., 1997 Mitochondrial control region sequences and phylogenetic systematics of darters (Teleostei: Percidae). Copeia **2:** 319–338.

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

Wilkinson, G. S., F. Mayer, G. Kerth and B. Petri, 1997 Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. Genetics **146:** 1035–1046.

Zardoya, R., and A. Meyer, 1996 The complete nucleotide sequence of the mitochondrial genome of the lungfish (*Protopterus dolli*) supports its phylogenetic position as a close relative of land vertebrates. Genetics **142:** 1249–1263.

Communicating editor: A. G. Clark