Heteroplasmy of Short Tandem Repeats in Mitochondrial DNA of Atlantic Cod, *Gadus morhua*

Einar Árnason*,† and David M. Rand*,‡

**Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, †Institute of Biology, University of Iceland, Reykjavik, Iceland, and ‡Division of Biology and Medicine, Brown University, Providence, Rhode Island 02912*

Manuscript received May 10, 1991
Accepted for publication May 23, 1992

ABSTRACT

The mitochondrial DNA of the Atlantic cod (*Gadus morhua*) contains a tandem array of 40-bp repeats in the D-loop region of the molecule. Variation among molecules in the copy number of these repeats results in mtDNA length variation and heteroplasmy (the presence of more than one form of mtDNA in an individual). In a sample of fish collected from different localities around Iceland and off George's Bank, each individual was heteroplasmic for two or more mtDNAs ranging in repeat copy number from two (common) to six (rare). An earlier report on mtDNA heteroplasmy in sturgeon (*Acipenser transmontanus*) presented a competitive displacement model for length mutations in mtDNAs containing tandem arrays and the cod data deviate from this model. Depending on the nature of putative secondary structures and the location of D-loop strand termination, additional mechanisms of length mutation may be needed to explain the range of mtDNA length variants maintained in these populations. The balance between genetic drift and mutation in maintaining this length polymorphism is estimated through a hierarchical analysis of diversity of mtDNA length variation in the Iceland samples. Eighty percent of the diversity lies within individuals, 8% among individuals and 12% among localities. An estimate of $\theta = 2N_{eo}\mu > 1$ indicates that this system is characterized by a high mutation rate and is governed primarily by deterministic dynamics. The sequences of repeat arrays from fish collected in Norway, Iceland and George's Bank show no nucleotide variation suggesting that there is very little substructuring to the North Atlantic cod population.

A number of recent studies of animal mitochondrial DNA (mtDNA) have shown that the length of the molecule can vary between, and even within, individuals of a species (SOLIGNAC *et al.* 1983; HARRISON, RAND and WHEELER 1985; BERMINGHAM, LAMB and AVISE 1986; BENTZEN, LEGGETT and BROWN 1988; RAND and HARRISON 1989; BOYCE, ZWICK and AQUADRO 1989; BUROKER *et al.* 1990; MIGNOTTE *et al.* 1990; WILKINSON and CHAPMAN 1991; GJETVAJ, COOK and ZOUROS 1992). In view of the small size and efficient organization of animal mitochondrial genomes (ATTARDI 1985), the molecular basis of mtDNA length variation and its maintenance in natural and cytoplasmic populations poses a variety of intriguing questions. A characterization of the sequences associated with mtDNA size variation may shed light on the mechanisms responsible for the generation of this variation. Moreover, the existence of heteroplasmic cytoplasms (those carrying two or more mtDNA molecules of different size) provide an opportunity to examine the action of genetic drift, mutation and selection. A thorough understanding of the dynamics of "intracellular population genetics" (BIRKY 1983) is essential for the interpretation of mtDNA variation in natural populations as all mtDNA variants that ultimately become fixed must pass through an initial heteroplasmic stage.

MORITZ, DOWLING and BROWN (1987) identified three general types of mtDNA size variation: 1) variation in the number of nucleotides in a homopolymer run, 2) variation in copy number of tandemly repeated elements and 3) duplication or deletion of large regions of the molecule. Strand slippage and mispairing has been suggested as a mechanism responsible for the first type of variation (HAUSWIRTH *et al.* 1984; LEVINSON and GUTTMAN 1987). Sequence analyses of the repeats in tandem arrays has revealed that these regions contain motifs capable of forming secondary structures (RAND and HARRISON 1989; BUROKER *et al.* 1990; LA ROCHE *et al.* 1990) and that these structures may contribute to the slippage and mispairing of strands across entire repeat units (LA ROCHE *et al.* 1990). RAND and HARRISON (1989) also suggested recombination as one possible mechanism generating length variants, but there is, as yet, little convincing evidence for recombination in animal mtDNA (MORITZ, DOWLING and BROWN 1987). In cases where large duplications result in mtDNA size variation, it appears that the ends of these duplications lie near transfer RNA genes (MORITZ, DOWLING and BROWN 1987). It may be that the tRNA sequences fold into their characteristic secondary structures (or approximations of these structures) while mtDNA is replicating and that

these conformations contribute to the insertion or deletion of DNA, leading to length variation. Further implications for a role of secondary structures in mtDNA length variation is the observation that this type of variation usually maps to the D-loop (*i.e.*, control) region of animal mtDNA, a region commonly containing sequence motifs capable of forming secondary structures (MORITZ and BROWN 1987).

In a recent paper on sturgeon mtDNA, BUROKER *et al.* (1990) proposed a model in which tandem repeat arrays that have the potential to form secondary structures, and lie in the D-loop region of mtDNA, can generate mtDNA length mutations. The proposed mechanism takes advantage of the unique, triple-stranded nature of the D-loop to allow for both insertion and deletion of repeats. This is an improvement over models of slip-mispairing of repeats at replication forks (*e.g.*, EFSTRADIADIS *et al.* 1980) which are more likely to lead to the deletion than the insertion of a repeat. An explicit prediction of BUROKER *et al.*'s (1990) model is that the smallest mtDNA molecule will carry three repeats, and their population survey indicates that this prediction is upheld.

In this paper we demonstrate the existence of a very similar case of mtDNA size variation and heteroplasmy in Atlantic cod (*Gadus morhua*). However, our molecular analyses clearly show that mtDNAs carrying two repeats are found in almost every fish sampled from nature. Furthermore, sequence data suggest that the cod repeats may not be included in the D-loop strand. While the model proposed by BUROKER *et al.* (1990) may hold in cases where a molecule contains three or more repeats, additional mechanisms may operate when fewer repeats are present and in species where D-loop strands terminate prior to repeat arrays. Such differences can alter the mutational component of the population genetic forces governing the dynamics of this polymorphism. To estimate the nature of this dynamic, a hierarchical analysis of the diversity for mtDNA size within and among individual fish and collecting localities in Iceland is performed. The balance between genetic drift and length mutation is quantified using this statistical approach, and it is shown that this polymorphic system is maintained primarily by deterministic dynamics.

MATERIALS AND METHODS

Collecting: The cod used in this study were obtained through the Icelandic Marine Research Institute from small fishing boats operating out of local villages around Iceland during the 1987 and 1988 fishing season. The sampling localities were Reykjavik (RE) in the southwest, Olafsvik (OV) in the west, Husavik (HV) in the north, Thorshofn (TH) in the northeast, Hofn (HF) in the southeast, and Vestmannaeyjar (VE) in the south. One additional sample from George's Bank (GB) off the coast of North America was obtained from a local fish market in Rhode Island (see Figure 1). The sampling localities thus span a large part of the breeding grounds of Atlantic cod in the North Atlantic ocean. Ovaries with mature oocytes, which suggest that the female was breeding at the local fishing grounds, were removed by a local marine biologist who also determined the state of maturity of the oocytes (fillet was used as a source of DNA in the cod from George's Bank).

Isolation of mtDNA: Samples were shipped on ice by air to EA's laboratory at the University of Iceland in Reykjavik. There samples of oocytes were homogenized in TEK buffer (50 mM Tris, 10 mM EDTA, 1.5% KCl, pH 7.5; CHAPMAN and POWERS 1984) and mitochondrial DNA subsequently isolated by the alkaline lysis method (TAMURA and AOTSUKA 1988). The mtDNA was later further purified by proteinase K digestion and phenol/chloroform extraction (SAMBROOK, FRITCH and MANIATIS 1989). These procedures generally produced a highly enriched sample of mtDNA containing very little nuclear DNA.

Restriction fragment analysis: Restriction endonuclease digestions were carried out according to the vendors' specifications (Bethesda Research Laboratories, New England Biolabs). Digestions were treated simultaneously with RNase to facilitate visualization of fragments under 500 base pairs (bp). Fragments were separated by electrophoresis in 0.6% to 2.0% agarose gels and 5.0% native polyacrylamide (SAMBROOK, FRITCH and MANIATIS 1989). Fragments were visualized by either 1) ethidium bromide staining, 2) autoradiography of Southern blots hybridized with [α-^{32}P]dCTP-labeled DNA probes or 3) autoradiography of DNA labeled by fill-in of 3′ overhangs using the 5′ to 3′ polymerase activity of the Klenow fragment of DNA polymerase I or the T4 DNA polymerase in the presence of the appropriate [α-^{35}S]dNTP (SAMBROOK, FRITCH and MANIATIS 1989).

Cloning and sequencing: Purified mtDNA from individual VE4 was digested with *Eco* RI and the fragments were ligated into the plasmid vector pEMBL (DENTE, CESARENI and CORTESE 1983). Amplification products from polymerase chain reactions (PCR) were also ligated into pEMBL cut with *Sma*I to provide a blunt end compatible with the blunt ends of PCR fragments. The ligations were used to transform *E. coli* strain DH5α which were plated on an ampicilin medium (100 μg/ml). Recombinant plasmids containing native cod mtDNA were selected by colony hybridization (SAMBROOK, FRITCH and MANIATIS 1989) using purified mtDNA of trout (*Salmo gairdneri*) as a radioactively labeled probe. Recombinant plasmids containing PCR amplification fragments were identified by blue/white colony selection and subsequently confirmed through sequence analysis. Multiple independent isolates of each cloned PCR fragment were sequenced to identify possible cloning artifacts; none were detected. Sequencing was performed using the dideoxy chain termination method with a Sequenase kit (Version 2.0, United States Biochemical Corporation) and [α-^{35}S] dATP (1000 Ci/mmol, New England Nuclear) as a labeled nucleotide.

Sequence analysis was performed using the Genetics Computer Group's (GCG) package of computer programs. Secondary structures were plotted using a program by GILBERT (1990) which reads GCG output files from the FOLD and SQUIGGLES procedures.

PCR: Oligonucleotide primers, codrep1: 5′-AATTCTAAATTTAAACTACC-3′ and codrep2: 5′-GGACATATTATGAGGCAGGG-3′, complementary to unique DNA sequences on either side of the repeat, were designed based on our own sequence data as well as sequences from JOHANSEN, GUDDAL and JOHANSEN (1990). To confirm the specificity of the primers, each was used in a sequencing reaction of the plasmid clone containing the repeat region. The predicted sequences were generated by

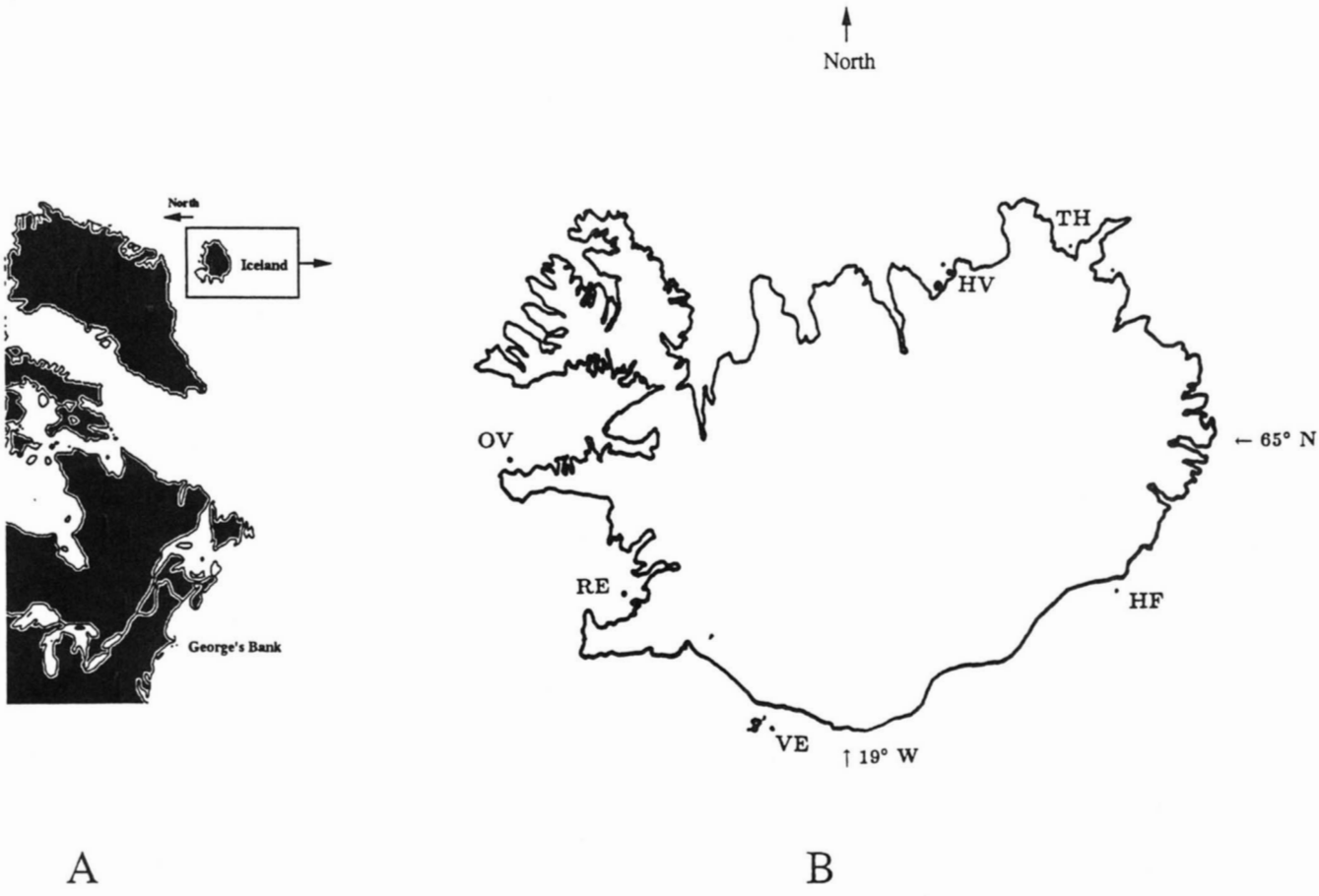

FIGURE 1.—Map of Northeastern North America (A) and an expanded view of Iceland (B) showing location of collecting localities: Reykjavik (RE), Olafsvik (OV), Husavik (HV), Thorshofn (TH), Hofn (HF) and Vestmannaeyjar (VE).

each of the primers. The primers were used in amplification reactions with the following conditions: melt at 94° for 30 sec, anneal at 45° (or 47°) for 60 sec and extend at 72° for 60 sec. Thirty cycles of this regime were performed with the time between segments set to a minimum of 30 sec. Reagents used were from the GeneAmp kit (Perkin-Elmer Cetus) and a typical reaction involved 50 ng of cod DNA, 50 to 70 pM of each primer and the proportions of nucleotides, 10× reaction buffer and *Taq* polymerase specified by the supplier.

Densitometry: To estimate the relative frequencies of mtDNA length variants within and among individual cod, the intensities of bands reflecting heteroplasmic variation were quantified with an LKB densitometer. The height of the densitometric peak generated from a given band, divided by the sum of all peak heights detectable within a lane of a gel, was taken as an estimate of the frequency of the mtDNA molecule of the corresponding size. Densitometry of photographic negatives of ethidium bromide stained gels in which PCR amplification products had been electrophoresed was also performed. It was found that different annealing temperatures in the PCR reactions gave quite different relative frequencies of bands reflecting heteroplasmic variation. Thus only native cod mtDNA was quantified for length variation.

Apportionment of mtDNA size variation: The variation in mtDNA size present in the total sample of cod from Iceland was apportioned into hierarchical components as described in RAND and HARRISON (1989). Gene diversity was calculated from the mtDNA size class frequencies as follows: $K = 1 - \Sigma x_i^2$, where x_i is the frequency estimate of mtDNA size class i within individuals, collecting localities or the total sample, as determined from densitometry. Following the notation of BIRKY, FUERST and MARUYAMA (1989) K_a is the diversity within a cell (not measured in the current study), K_b is the diversity within an individual animal, K_c is the diversity within a deme (or collecting locality) and K_t is the diversity within the total sample. Following LEWONTIN (1972) these measures of diversity at different levels of a hierarchy can be apportioned into different components. C_i = (K_b averaged over all individuals)/K_t = the within-individual component of the total diversity. C_{ip} = ([K_c averaged over all demes] − [K_b averaged over all individuals])/K_t = the among-individual-within-deme component of the total diversity. C_{pt} = (K_t − [K_c averaged over all demes])/K_t = the among-deme component of the total diversity. By definition, $C_i + C_{ip} + C_{pt} = 1.0$.

To characterize the variability in these measures from this relatively small sample of cod, a jackknife analysis was performed. Each individual sample was removed, one at a time, from the data set and the C statistics recalculated. With the sample of 15 cod, this resulted in 15 different estimates of C_i, C_{ip} and C_{pt}. Means and an unbiased estimate of the variances ($\sigma^2 = [(n - 1)/n]\Sigma_{i=1}^{n}\{C_i - \bar{C}\}^2$) of these values were then calculated (where n = sample size, C_i = the C value for the ith jackknife sample, and $\bar{C}$ = the mean C value for all samples.

RESULTS

Heteroplasmy: Figure 2 is a photograph of 10 representative cod mtDNA samples from an autora-

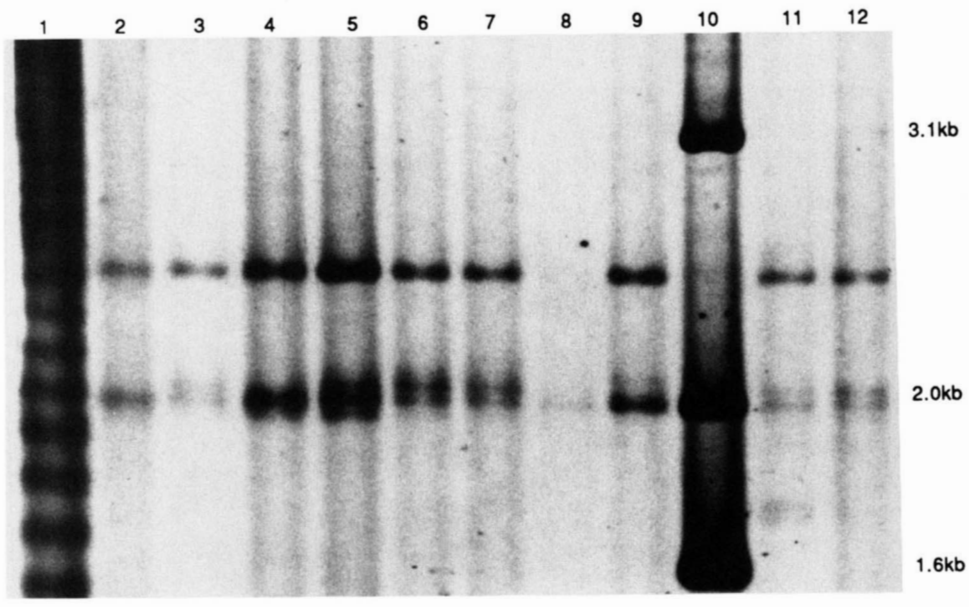

FIGURE 2.—Length heteroplasmy in *Ava*II digests of cod mtDNA. Samples of mtDNA of 10 cod obtained from three different localities were digested with *Ava*II, end-labeled with ^{35}S, and electrophoresed in 2% agarose along with labeled size standards. From left to right the lanes show 123-bp size ladder, cod 1, 2, 9, 10, and 11 from Husavik; cod 1, 3, and 4 from Thorshofn; the 1-kb size standard; and cod 3 and 9 from Hofn. Size heteroplasmy, estimated to be about 40 bp by comparison with the 123-bp ladder, is evident in a fragment of about 2.05 kb as seen by comparison with the 1-kb ladder.

diograph of [^{35}S]dATP end-labeled *Ava*II fragments. At a fragment size of approximately 2.05 kb, a ladder of substoichiometric bands is apparent. If each band observed in a lane is scored as a unique *Ava*II fragment, the sum of all fragments yields a total mtDNA size of over 20 kb, 3.5 kb larger than the 16.5-kb estimate of ours and JOHANSEN, GUDDAL and JOHANSEN (1990). However, by counting only one 2-kb band the total sum is 15.9 kb which, within the limits of resolution, agrees quite well with the estimated size of the mtDNA molecule. Based on an extrapolation plot comparison to the migration distances of the bands of the 123-bp ladder, the substoichiometric bands differ in size by 40 bp. These observations suggest that these cod were heteroplasmic for mtDNA molecules that differed in size by 40 bp.

A total of 27 cod from six different localities were characterized for mtDNA size variation and heteroplasmy, 15 by end-labeling of *Ava*II fragments and an additional 12 by PCR amplification (see below). All 27 individuals were heteroplasmic for two or more mtDNA size classes, each of which differed in size by 40 bp.

Molecular characterization: By analogy to other examples of mtDNA heteroplasmy (DENSMORE, WRIGHT and BROWN 1985; BERMINGHAM, LAMB and AVISE 1986; BENTZEN, LEGGETT and BROWN 1988; BUROKER *et al.* 1990), it was assumed that the putative 40 bp length variable region was located in or near the D-loop region of mtDNA. In an attempt to characterize the length-variable region, purified mtDNA from codVE4 was cloned in a sequencing plasmid. Two recombinant plasmids were recovered: clone "cod4" containing an 8.2-kb insert and clone "cod40" containing both a 7.5-kb insert and a 0.7-kb insert. These fragments correspond to the sizes of *Eco*RI fragments generated when native cod mtDNA is digested (E. ARNASON and D. RAND, unpublished data; JOHANSEN, GUDDAL and JOHANSEN 1990). To position the cloned fragments on the vertebrate mitochondrial map, sequencing reactions of clones cod4 and cod40 were performed using the m13 forward and reverse primers. The forward sequence of clone cod4 fortuitously revealed five copies of a 40-bp sequence at a distance of 22 bp from the *Eco*RI cloning site (see Figure 3). The reverse primer sequence of clone cod4 showed clear homology to the cytochrome oxidase II gene which lies 8 kb from the D-loop in the vertebrate mtDNA map. These observations are confirmed by the data presented in JOHANSEN, GUDDAL and JOHANSEN (1990). All five copies of the repeat in clone cod4 are identical in sequence and agree with the sequence of four identical repeats described by JOHANSEN, GUDDAL and JOHANSEN (1990).

Since the D-loop is usually a highly variable region of mtDNA, the repeat arrays of three additional fish (collected at Thorshofn (TH), Husavik (HV) in Northeastern Iceland and George's Bank (GB) in the Western Atlantic) were sequenced to examine the possibility that cod from geographically distant localities carry unique mtDNA haplotypes. From each fish the sequence of a three-repeat array was determined and all repeats were identical to those from the cod4 clone and the sequence reported by JOHANSEN, GUDDAL and JOHANSEN (1990).

The 22-bp sequence between the *Eco*RI site and the repeat array is a part of the $tRNA^{Pro}$ gene which extends 4 bases into the first repeat (JOHANSEN, GUDDAL and JOHANSEN 1990). The repeat contains 24/40 or 60% A + T which is only slightly lower than the 64% A + T of the D-loop (JOHANSEN, GUDDAL and JOHANSEN 1990). The repeat elements can form dis-

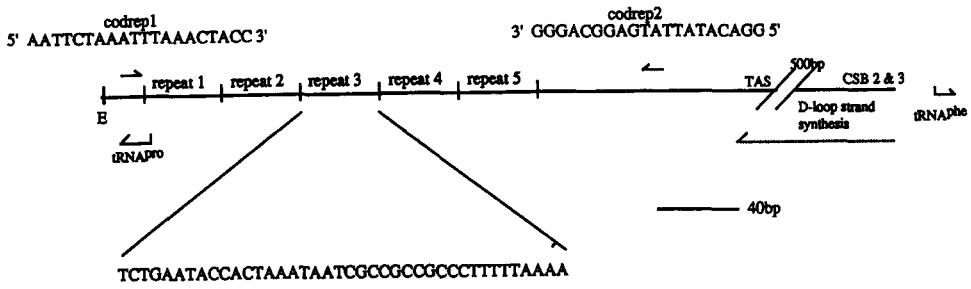

FIGURE 3.—Map of repeat region showing the sequence of one repeat. Short arrows above the map indicate the locations of the PCR primers codrep1 (3′-end is 1bp to the left of the first repeat) and codrep2 (3′- end is 46 bp to the right of the last repeat), the tRNA^{pro} and tRNA^{lys} genes, the TAS and the CSB. Based in part on data from JOHANSEN, GUDDAL and JOHANSEN (1990).

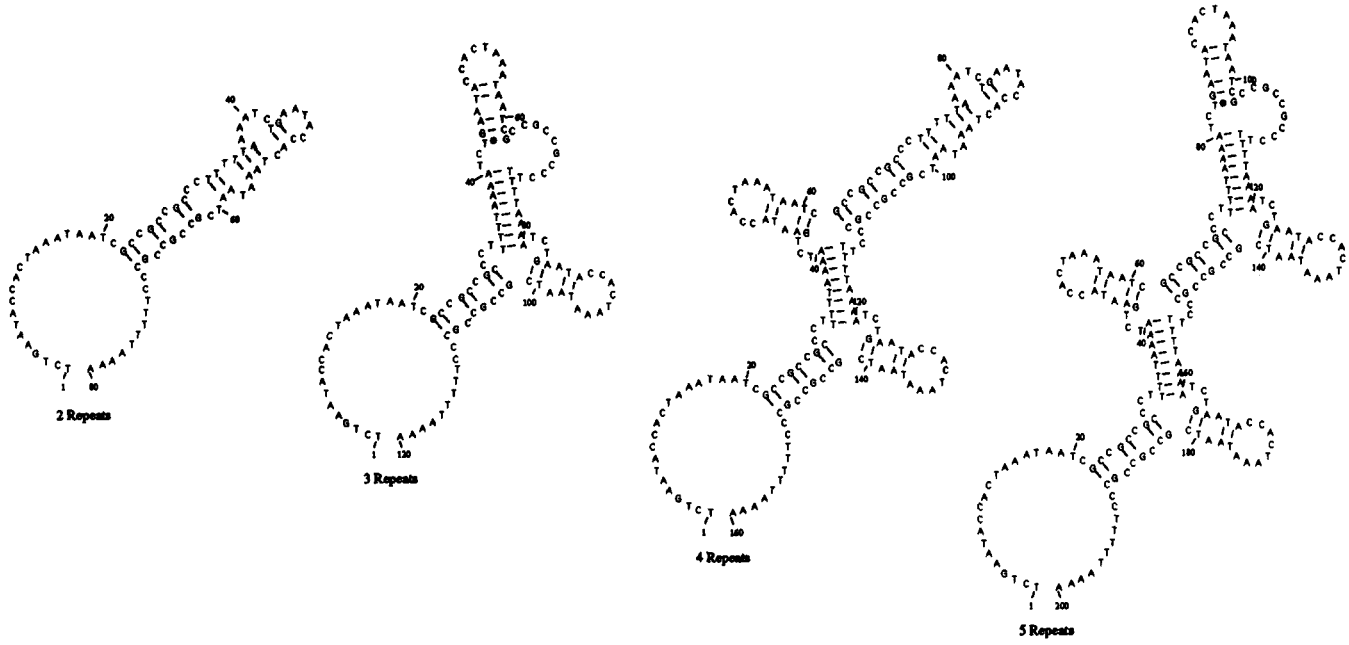

FIGURE 4.—Predicted secondary structures of the repeat array with two, three, four and five copies of the repeat. Note the similarity of stem structures in the 2- and 4-repeat sequences and in the 3- and 5-repeat sequences. The sequences in these figures are the complement to the sequence presented in Figure 3.

tinctive secondary structures (Figure 4) with free energies increasing linearly with repeat copy number: the 2-, 3-, 4-, 5- and 6-repeat arrays have folding energies of −3.4, −6.4, −9.9, −12.9 and −16.4 kcal/mole, respectively as determined by the Fold algorithm of the Genetics Computer Group sequence analysis programs, University of Wisconsin. The 22-bp unique sequence prior to the repeat elements did not participate in forming a secondary structure with the repeat elements.

To test whether the 40-bp tandem repeat element of the D-loop is responsible for the heteroplasmy, we hybridized a Southern blot of an *Ava*II digested native mtDNA cod sample with a repeat element probe from the cod4 clone. The probe was made by amplifying the repeat array of the cod4 clone using equimolar ratios of the codrep1 and codrep2 primers. The double-stranded DNA product was electrophoresed in a 1.5% SeaPlaque agarose gel, the single band cut from the gel and labeled with [α-^{32}P]dCTP. The probe hybridized to only one band: the 2.05-kb fragment which showed heteroplasmic variation on previous gels (data not shown).

Further characterization of heteroplasmy was done through PCR amplification of the repeat region from native cod DNA samples using the codrep1 and codrep2 primers. Using some of the individuals shown in Figure 2 as reference samples, 12 additional cod (collected from the six localities in Iceland, plus George's Bank, but not studied by autoradiography of end-labeled restriction fragments) were analyzed. All samples yielded amplification products indicating heteroplasmic cod mtDNA with size intervals of 40 bp (a few samples showed intermediate-sized bands at low frequency). The amplification products corresponded to bands containing two to six repeats and matched the sizes of bands predicted by the location of the primers (see Figures 3 and 5). Thus the heteroplasmy observed in end-labeled samples of purified mtDNA also was revealed in the products of the PCR.

Frequencies and hierarchical variation of length variants : Table 1 gives the frequencies of the various size classes representing two to six repeats in the 15 *Ava*II-digested native mtDNA as determined by a densitometric scanning of the autoradiograph (Figure

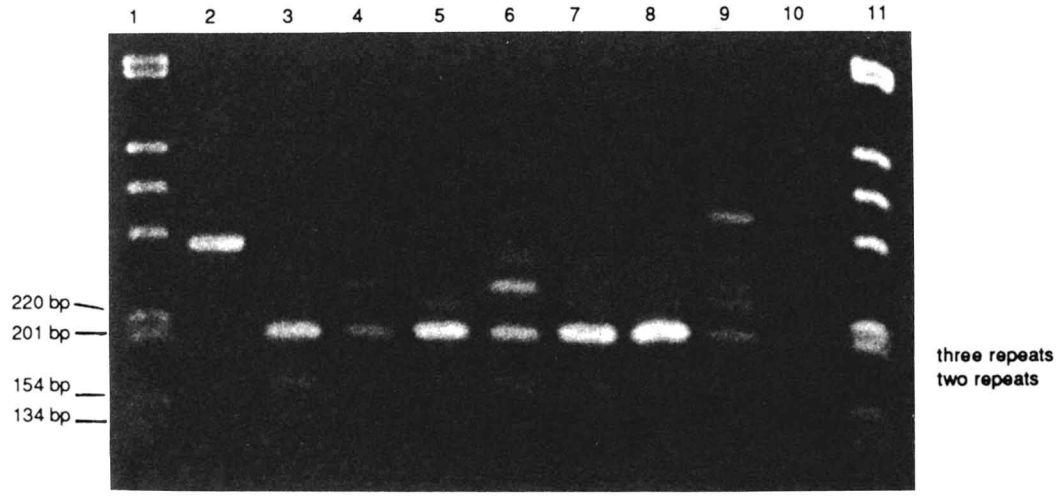

FIGURE 5.—PCR products of native cod DNA using primers that flank the repeat region. Lane 2 shows the double-stranded PCR product from an amplification of the five-copy array which has been sequenced (see Figure 3). Lanes 3, 4 and 5 show double-stranded DNA products amplified from the same DNA samples as were electrophoresed in lanes 2, 3 and 4 of Figure 2; the remaining lanes are additional samples from around Iceland. Band sizes indicate that each individual carries the two-repeat mtDNA: each primer is 20 bp long, there is a 1-bp interval between the 3′-end of codrep1 and the first base of the repeats, a 46-bp interval between the 3′ end of codrep2 and the last base of the repeats, and each repeat is 40 bp long. Thus, two-, three- or four-repeat PCR products would be 167, 207 or 247 bp long, respectively. These predictions fit well with the size standards in the left- and right-most lanes (1-kb ladder from BRL).

TABLE 1

Apportionment of diversity for mtDNA size in Atlantic cod

Sample No.	Specimen	Size class (No. of repeats) 2	3	4	5	6	K_b	K_c
1	T1HV	0.25	0.75				0.38	
2	T2HV	0.16	0.37	0.32	0.15		0.71	
3	T9HV	0.26	0.74				0.38	
4	T10HV	0.21	0.38	0.29	0.12		0.71	
5	T11HV	0.08	0.26	0.44	0.22		0.68	
	HV pooled	0.192	0.500	0.210	0.098			0.66
6	T1TH	0.09	0.22	0.32	0.20	0.17	0.77	
8	T4TH		0.82		0.18		0.29	
	TH pooled	0.045	0.522	0.160	0.188	0.085		0.66
9	T3HF	0.56	0.44				0.49	
10	T9HF	0.50	0.50				0.50	
	HF pooled	0.530	0.470					0.50
11	T2RE	0.36	0.64				0.46	
12	T4RE		0.95	0.05			0.10	
13	T8RE		0.58	0.42			0.49	
	RE pooled	0.120	0.723	0.157				0.44
14	T2OV	0.21	0.40	0.39			0.64	
15	T1OV		0.50	0.50			0.50	
	OV pooled	0.105	0.450	0.445				0.59
28	VE8		0.48	0.52			0.50	0.50
	Grand pool	0.179	0.535	0.217	0.058	0.011		
	K_t						0.63	
	Mean K_c						0.56	
	Mean K_b						0.51	
			Jackknife:	Mean	Variance			
	C_i	0.804		0.802	0.003			
	C_{ip}	0.078		0.073	0.007			
	C_{pt}	0.118		0.125	0.011			

The values listed under each size class are the frequencies of that size class in each cod specimen. All individuals were heteroplasmic. The mean size class frequency for each collecting locality is listed under each group of specimens ending with the same two-letter code. K_b is the diversity within an individual, K_c within a locality and K_t the diversity in the total sample. At the bottom of the table the means and estimates of variance from a jackknife analysis of the data set are presented. See text for details.

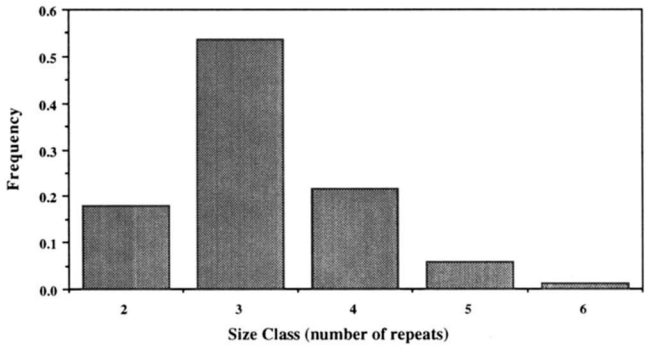

FIGURE 6.—Frequency distribution of mtDNA length variants in Icelandic samples of Atlantic cod. Since all fish were heteroplasmic each bar indicates the sum across all fish of the proportional contributions to that size class.

2). We have not included the frequencies of the size variants of the 12 individuals studied by polymerase chain reaction as different annealing temperatures in the PCR gave different relative frequencies of the size variable bands. All individuals carried the three-repeat molecule and most carried the two- and four-repeat molecule. The distribution of these three size classes is quite symmetric, but the overall distribution is clearly skewed (Table 1 and Figure 6). Although the sample is small, the hierarchical structure of the variation is clear from the analysis of the 15 individuals. The three-level statistics show that approximately 80% of the total diversity for mtDNA size lies within individuals (C_i). Only about 8% of the variation lies between individuals within a population (C_{ip}) and 12% lies between populations (sampling localities) (C_{pt}) indicating some population differentiation. The jackknife analysis (see Table 1) produced low estimates of variance indicating that the conclusions reached from these data (see DISCUSSION) would not be altered if the sample sizes were larger.

BIRKY, MARUYAMA and FUERST (1983) provide an equation for the equilibrium value of K_c assuming K_a (diversity within a cell) is small and $K_c \gg \mu$ (mutation rate):

$$K_c \sim 2N_{eo}\mu/(2N_{eo}\mu + 1) \qquad (1)$$

where N_{eo} is the effective number of organelle genomes. The data in Table 1 suggest that K_a may not be small (diversity data were determined from oocyte tissue), the effect of which is to introduce an error of about 25% in equation 1 (see Figure 3 of BIRKY, FUERST and MARUYAMA 1989). Bearing this in mind, the value of $K_c = 0.56$ from Table 1 can be applied to equation 1 to provide a reasonable estimate of $\theta = 2N_{eo}\mu = 1.27$. If one assumes that cod from Iceland actually represent one large panmictic population (ÁRNASON, PALSSON and ARASON 1991) and uses $K_t = 0.63$, instead of K_c, in this analysis, an estimate of $\theta = 2N_{eo}\mu = 1.70$ is obtained. This assumption seems justified since repeat arrays from fish collected in Norway, Iceland and George's Bank can have identical sequence.

DISCUSSION

Several independent lines of evidence confirm that heteroplasmy in Atlantic cod is caused by variation in copy number of a 40-bp repeat. Electrophoretic analysis of end-labeled, purified mtDNA from cod collected at different localities around Iceland show the characteristic substoichiometric bands in a ladder pattern. The interband interval on the autoradiograph indicates that the bands in this ladder differ in size by 40 bp. Our sequence from the repeat region of a cod from Iceland and George's Bank shows five and three copies, respectively, of the 40-bp repeat while the sequence reported by JOHANSEN, GUDDAL and JOHANSEN (1990) contains four repeats. PCR amplification of the repeat region from different cod samples also produces a variable number of bands differing in size by 40 bp. When these PCR products are used as radioactive probes of Southern blots of *Ava*II-digested total genomic cod DNA, the only band producing a signal is the 2.05-kb band which showed heteroplasmic variation on end-label gels.

In vertebrate mtDNA, the D-loop strand is a third strand of DNA which is synthesized off the light strand (L-strand) and displaces the heavy strand (H-strand) of the molecule. Synthesis of the D-loop strand is initiated near the conserved sequence blocks (CSBs), runs toward the $tRNA^{pro}$ gene and terminates beyond the terminal associated sequence (TAS; CLAYTON 1982). It has been suggested that the D-loop strand is the product of an aborted attempt at the replication of the H-strand of mtDNA (CHANG and CLAYTON 1985). If more than one TAS is present, the D-loop strand synthesis can terminate at a more distal TAS resulting in D-loop strands of different lengths on different mtDNA molecules (DODA, WRIGHT and CLAYTON 1981). However, the existence of multiple TASs alone is not sufficient to generate length mutations as no DNA is lost or gained in the nascent H-strand.

Tandem repeats in the D-loop region: The cod repeats are comparable to a number of other repeated sequences found in animal mtDNA in that they lie in the D-loop region and are the source of mtDNA length variation both within and between individuals [see MORITZ, DOWLING and BROWN (1987) for a review]. The most comparable example is the sturgeon (BUROKER *et al.* 1990) where the repeats are 82 bp long, lie between the 5′-end of the $tRNA^{Pro}$ gene and the CSBs, and each repeat contains a TAS. In cod the repeats are 40 bp long and lie in virtually the same position, showing a four bp overlap with the 5′-end of the $tRNA^{Pro}$ gene. However, the TAS lies outside the repeats approximately 100 bp from the end of the

array, and no TAS-like motif can be found within the repeat elements (JOHANSEN, GUDDAL and JOHANSEN 1990, Figure 2a).

BUROKER *et al.* (1990) present a model of how an mtDNA molecule might change size through competitive displacement of D-loop and H-strand sequences. One prediction stated by BUROKER *et al.* (1990) is that the shortest mtDNA should be a three-repeat molecule (p161). While their data support this prediction, our molecular analyses, as well as those of RAND and HARRISON (1989), clearly show that two-repeat molecules are quite frequent in natural populations. While this does not invalidate the mechanics of the model for the generation of heteroplasmy in sturgeons, the model may not apply to all examples of heteroplasmy from tandem repeat copy number variation.

The nature of the secondary structures in the repeat array are an important component of the model presented by BUROKER *et al.* (1990). In cod, a single repeat does not form any convincing secondary structures, and two repeats are needed to provide the CGCCGCCGCCCTTTTTAAAA motif with an opportunity for complimentary pairing. In sturgeon, a single repeat, as well as a dimer, form striking hairpins where the 5′- and 3′-ends of the repeat(s) align and more than 60% of the nucleotides are paired (BUROKER *et al.* 1990, Figure 3). These observations present a paradox. With a minimum of two repeats needed for hairpin formation in the cod D-loop, one would expect a minimum of three repeats under the competitive displacement model: two in the hairpin and one as an invading strand. Yet most cod individuals studied to date carry two-repeat D-loop regions (Table 1). Conversely in sturgeon, the ability to form a hairpin within a single repeat would seem to allow for the deletion of all but two repeats: one in a hairpin and the other as the invading strand (see Buroker *et al.* 1990, Figure 4b). That no two-repeat D-loop regions were detected in 128 sturgeon strongly suggests that deletion events involving two-repeat hairpins occur very rarely (or not at all) when the D-loop region contains three or four repeats.

Relative position of the TAS: A critical component of the BUROKER *et al.* (1990) model is that the D-loop strand runs through the repeat region. In cod the only TAS clearly identified lies upstream from the repeats (in the direction of D-loop strand synthesis; JOHANSEN, GUDDAL and JOHANSEN 1990). It is therefore possible that the D-loop strand terminates before the repeat region and does not contain repeated sequences. If this were the case in cod [and in crickets where two-repeat molecules are also common in nature (RAND and HARRISON 1989)], the mechanics of the competitive displacement model (BUROKER *et al.* 1990) might not apply. Evidence that the D-loop strand runs through the repeat region should therefore be taken as a necessary (but not sufficient) condition for the operation of this mechanism. It is possible that additional mechanisms of slip-mispairing (*e.g.*, EFSTRADIADIS *et al.* 1980; LEVINSON and GUTTMAN 1987) work in concert with the competitive displacement model to produce heteroplasmic variation. However, at the very least it appears that the prediction of a minimum of three repeats needs to be relaxed for the competitive displacement model to fit other known systems.

The nature of the mechanism(s) involved in insertion/deletion of repeats can affect the shape of the frequency distribution of length variants sampled from natural populations. A skewed frequency distribution of mtDNA length variants has been interpreted as evidence for selection (HALE and SINGH 1986). However, it is virtually impossible, using only samples from natural populations, to assign the source of this skewness to selection rather than some asymmetry in the rates of insertion/deletion between size classes. Given the small difference between cod mtDNA size classes, it seems unlikely that a race for replication favoring the smaller molecules is sufficiently great to overcome the effects of drift. Selection may operate on an "optimal" number of repeats for the initiation of certain steps in replication (RAND and HARRISON 1989). This could increase certain size classes to higher frequencies than others assuming a completely symmetric mutation process. Alternatively, equal mutation rates between the two-, three- and four-repeat molecules with low mutation rates to the five- and six-repeat molecules could also result in the observed frequency distribution (see Figure 6). The increased stability of secondary structures containing more repeats may contribute to such an asymmetry of mutation rates (see Figure 4). Direct functional studies are the only means by which the effects of selection and mutation can be disentangled.

Evolutionary forces and mtDNA length variation: The hierarchical apportioning of diversity for mtDNA size shows that the vast majority of variation lies within individuals (heteroplasmy). This implies that the generation of variation by length mutations overcomes the loss of variation due to random drift during cell division. That all fish studied were heteroplasmic for two or more length variants is consistent with this interpretation. An alternative means of quantifying this balance of genetic drift and mutation is with the estimate of $\theta = 2N_{eo}\mu$. The current data indicate that $\theta > 1$. This implies that the effects of mutation slightly outweigh the effects of drift and that this polymorphism is governed primarily by mutation-driven, deterministic dynamics (WRIGHT 1940).

At higher organizational levels, the low values of C_{ip} and C_{pt} show that little variation among individuals within collecting localities, and among collecting lo-

calities in the total sample, is maintained in the face of rapid shuffling of mtDNA size from insertion and deletion of repeats. This is largely an outcome of the finite number of "alleles" (or length variants). A high mutation rate for insertion/deletion in a tandem array rapidly reintroduces any length variants lost by random drift at either the cellular, individual or population level (CLARK 1988; RAND and HARRISON 1989). If mutations produced unique haplotypes, such as in an infinite alleles model, the apportionment of diversity would be rather different, with much of the variation being among individuals and localities.

A high migration rate could also contribute to low C_{pt} values as dispersal of genotypes would tend to reduce any local differentiation. In a finite alleles model, however, it is difficult to separate the relative contributions of homogenization due to migration from homogenization due to a high rate of insertion/deletion. The limited sequence data available can shed some light on this problem. With the exception of the variation in copy number, the sequence of the repeat array from a Norwegian cod (JOHANSEN, GUDDAL and JOHANSEN 1990) is identical to that of two Icelandic cod and one cod from George's Bank (this paper). The vertebrate D-loop is known to have a high rate of nucleotide substitution (UPHOLT and DAWID 1977; MORITZ, DOWLING and BROWN 1987). With a high mutation rate for insertion/deletion in the D-loop region, base substitutions could be swept through each repeat in the array. With sufficient substructuring to the population, D-loop sequences from different regions would tend to acquire new sequences, resulting in the pattern of intraspecific concerted evolution. That cod collected on opposite sides of the North Atlantic ocean can carry identical sequences in this rapidly evolving region lend support to the suggestion that there is substantial mixing of, and hence, little substructure to, the North Atlantic cod population (*e.g.*, MORK *et al.* 1985; JOHANSEN, GUDDAL and JOHANSEN 1990). Given the dynamics of this molecular polymorphism, as characterized above, a larger sample of sequences from the repeat region of cod mtDNA sampled from within and among adjacent and distant localities should provide a sensitive assay of for substructuring of the North Atlantic cod population.

We thank R. C. LEWONTIN in whose laboratory major parts of this work were done. We also thank VILHJALMUR THORSTEINSSON for help in obtaining samples and ADALGEIR ARASON and SNAEBJORN PALSSON for technical assistance. This work was supported by National Institutes of Health grant GM21179 to R. C. LEWONTIN, NIH Postdoctoral NRSA GM12357 to D.M.R., research funds from the Division of Biology and Medicine at Brown University provided to D.M.R., a Fulbright-Hayes grant to E.A., a University of Iceland travel grant to E.A., and by a University of Iceland Research Fund grant to E.A.

LITERATURE CITED

ÁRNASON, E., S. PALSSON and A. ARASON, 1991 Gene flow and lack of population differentiation in Atlantic cod, *Gadus morhua* L., from Iceland, and comparison of cod from Norway and Newfoundland. J. Fish Biol. **40:** 751–770.

ATTARDI, G., 1985 Animal mitochondrial DNA: an extreme example of genetic economy. Int. Rev. Cytol. **93:** 93–145.

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

BERMINGHAM, E., T. LAMB and J. C. AVISE, 1986 Size polymorphism and heteroplasmy in the mitochondrial DNA of lower vertebrates. J. Hered. **77:** 249–252.

BIRKY, C. W., JR., 1983 Relaxed cellular controls and organelle heredity. Science **222:** 468–475.

BIRKY, C. W., JR., T. MARUYAMA and P. FUERST, 1983 An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics **103:** 513–527

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

BOYCE, T. M., M. E. ZWICK and C. F. AQUADRO, 1989 Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics **123:** 825–836.

BUROKER, N. E., J. R. BROWN, T. E. GILBERT, P. J. O'HARA, A. T. BECKENBACH, W. K. THOMAS and M. J. SMITH, 1990 Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics **124:** 157–163.

CHANG, D. D., and D. A. CLAYTON, 1985 Priming of mammalian mitochondrial DNAs begins at the light strand promoter. Proc. Natl. Acad. Sci. USA **82:** 351–355.

CHAPMAN, R. W., and D. A. POWERS, 1984 A method for the rapid isolation of mitochondrial DNA from fishes. Technical Report. Maryland Sea Grant Program. UM-SG-TS-84-05. College Park, Md.

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

DENSMORE, L. D., J. W. WRIGHT and W. M. BROWN, 1985 Length variation and heteroplasmy in mitochondrial DNA from parthenogenetic and bisexual lizards (genus *Cnemidophorus*). Genetics **110:** 689–707.

DENTE, L., G. CESARENI and R CORTESE, 1983 pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11:** 1645–1655.

DODA, J. N, C. T. WRIGHT and D. A. CLAYTON, 1981 Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA **78:** 6116–6120.

EFSTRADIADIS, A, J. W. POSAKONY, T. MANIATIS, R. M. LAWN, C. O'CONNELL, R. A. SPRITZ, J. K. DE RIEL, B. G. FORGET, S. W. WEISMANN, J. L. SIGHTOM, A. E. BLECHEL, O. SMITHIES, F. E. BARALLE, C. C. SHOULDERS and N. J. PROUDFOOT, 1980 The structure and evolution of the human β-globin gene family. Cell **21:** 653–668.

GILBERT, D. G., 1990 LoopViewer, a Macintosh program for visualizing RNA secondary structure. Published electronically on the Internet, available via anonymous ftp to ftp.bio.indiana.edu.

GJETVAJ, B., D. I. COOK and E. ZOUROS, 1992 Repeated sequences and large-scale size variation of mitochondrial DNA: a common feature among scallops (Bivalvia: Pectinidae). Mol. Biol. Evol. **9:** 106–124.

HALE, L. R., and R. S. SINGH, 1986 Extensive variation and heteroplasmy in size of mitochondrial DNA among geographic populations of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **78:** 8813–8817.

HARRISON, R. G., D. M. RAND and W. C. WHEELER, 1985 Mitochondrial DNA size variation within individual crickets. Science **228:** 1446–1448.

HAUSWIRTH, W. W., M. J. VAN DE WALLE, P. J. LAIPIS and P. D. OLIVO, 1984 Heterogeneous mitochondrial DNA D-loop sequences in bovine tissue. Cell **37:** 1001–1007.

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.

LA ROCHE, J. M. SNYDER, D. I. COOK, K. FULLER and E. ZOUROS, 1990 Molecular characterization of a repeat element causing large-scale variation in the mitochondrial DNA of the sea scallop, *Placopecten magellanicus*. Mol. Biol. Evol. **7:** 45–64.

LEVINSON, G., and G. A. GUTMAN, 1987 Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4:** 203–221.

LEWONTIN, R. C., 1972 The apportionment of human diversity. Evol. Biol. **6:** 381–398.

MIGNOTTE, F., M. GUERIDE, A.-M. CHAMPAGNE and J.-C. MOUNOLOU, 1990 Direct repeats in the noncoding region of rabbit mitochondrial DNA. Eur. J. Biochem. **194:** 561–571.

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

MORK, J., N. RYMAN, G. STAHL, F. M. UTTER and G. SAUNDERS, 1985 Genetic variation in Atlantic cod (*Gadus morhua*) throughout its range. Can. J. Fish. Aquat. Sci. **42:** 1580–1587.

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

SAMBROOK, J., E. F. FRITCH and T. MANIATIS, 1989 *Molecular Cloning*, Ed. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SOLIGNAC, M., M. MONNEROT and J. C. MOUNOLOU, 1983 Mitochondrial DNA heteroplasmy in *Drosophila mauritiana*. Proc. Natl. Acad. Sci. USA **80:** 6942–6946.

TAMURA, K., and T. AOTSUKA, 1988 Rapid isolation of animal mitochondrial DNA by the alkaline lysis procedure. Biochem. Genet. **26:** 815–819.

UPHOLT, W. I., and I. B. DAWID, 1977 Mapping of mitochondrial DNA in sheep and goats: rapid evolution in the D-loop region. Cell **11:** 571–583.

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

WRIGHT, S., 1940 The breeding structure of populations in relation to speciation. Am. Nat. **74:** 232–248.

Communicating editor: A. G. CLARK