Characteristics of a Conserved 1,579-bp Highly Repetitive Component in the Killer Whale, *Orcinus orca*[1]

Bengt Widegren, Úlfur Árnason,* and Göran Akusjärvi†*
*University of Lund, Sweden; and †University of Uppsala, Sweden

A tandemly organized, highly repetitive DNA component of the killer whale was sequenced. The length of the repeat was 1,579 bp. This unit, which characterizes all delphinids, shows stringent hybridization homology with a 1,740-bp repeat that is characteristic of all other cetacean families. The 1,579-bp component comprises ~15% of the killer-whale genome, in which it is repeated $4–5 \times 10^5$ times. Computer analysis of the sequence showed no linear repetition within the component. This indicates that the 1,579-bp unit has not evolved by amplification of shorter repeats. Several inverted repeats of substantial length were found in the 1,579-bp unit. The most conspicuous of these was a 72-bp sequence that deviated from matching in only three positions. The 72-bp sequence occurs within an open reading frame 330 bp in length. Transcriptional activity was registered in the cloned repeat in a cell-free system. The length of the transcript was ~340 nucleotides. The chromosomal localization of the 1,579-bp repeat was determined by in situ hybridization. The repeat was present in eight of 21 autosomal pairs and was found in almost all C-band-positive (constitutive heterochromatin) regions of the karyotype.

Introduction

An evolutionarily conservative, highly repetitive component is present in all cetaceans (whales, dolphins, and porpoises). The odontocete (toothed whale) and mysticete (whalebone whale) lineages separated ~40–50 Myr ago, but it is conceivable that the component is more ancient than the separation of the odontocete and mysticete lineages. The length of the repeat is ~1,740 bp in all cetacean families except the Delphinidae, in which the length of the repeat is ~1,570 bp (Árnason et al. 1984). Hybridization between the 1,740-bp repeat and the delphinid repeat occurs under stringent conditions. It is logical to assume that the 1,740-bp repeat was replaced by the delphinid repeat prior to the radiation of the delphinids and that the age of the repeat thus exceeds 20 Myr.

As judged on the basis of restriction analysis, the length of the delphinid repeat and the relative localization of restriction sites have been preserved among the various delphinid genera. These characteristics of the repeat have thus been maintained despite different degrees of amplification of the repeat in different delphinid genera.

The repeat is particularly abundant in the killer whale, and this species also has a C-band pattern that distinguishes it from the pattern in other delphinid karyotypes. Thus, in the killer whale, C-bands are more prominent than in other delphinids and occur in chromosome positions unique to this species.

The present study was undertaken to characterize the highly repetitive delphinid

1. Key words: highly repetitive DNA, base composition, cetacean evolution.

Address for correspondence and reprints: Dr. Úlfur Árnason, Department of Genetics, University of Lund, Sölvegatan 29, S-223 62 Lund, Sweden.

Mol. Biol. Evol. 2(5):411–419. 1985.

0737-4038/85/0205-0355$02.00

component as represented by the killer whale, with respect to both its composition and its chromosomal localization. The in situ hybridization technique was applied to clarifying whether the delphinid repeat was present in C-bands in positions unique for the killer whale as well as in positions common to other delphinids.

Experimental Procedures

Single sites for several restriction enzymes suitable for cloning occur in the presently studied delphinid repeat (Árnason et al. 1984). Cloning was performed after cleaving total DNA of the killer whale with *Eco*RV. The DNA was loaded on 1% agarose gel, and the highly repetitive fragment was electroeluted. After purification, the fragment was ligated into the *Eco*RV site of pBR322. Both strands of the cloned fragment were sequenced following the Maxam and Gilbert (1980) procedure after 3′ and 5′ end labeling. Positions 1 and 1,579, respectively, refer to axis of the *Eco*RV site in the repeat.

Computer analysis of the repeat was performed using three different programs. One program was provided by Drs. Petter Gustafsson and Robert Harr (Dept. of Microbiology, Umeå, Sweden), another by Dr. Philip Taylor (MRC Virology Unit, Glasgow, Scotland), and a third was that of the University of Wisconsin Genetics Computer Group (Deveraux et al. 1984).

Transcription of the cloned repeat in a cell-free system was studied according to the method of Manley et al. (1980). For discriminating between RNA polymerase II and III, transcription was performed either in absence or presence of α-amanitin (1 μg/ml).

For the in situ hybridization analysis, long-term fibroblast cultures were used. Preparations were denatured in 0.1 × SSC (1 × SSC = 0.15 M NaCl and 0.015 M Na citrate) just below boiling temperature (Singh et al. 1977). A ^{3}H nick-translated (Rigby et al. 1977) recombinant plasmid was used as a labeled probe. C-banded preparations were from the same specimen.

Results

Sequencing of the killer-whale repeat was based on readings of both strands of the repeat. The sequencing strategy is shown in figure 1. The length of the repeat was 1,579 bp. The sequence of the repeat is given in figure 2. The central A of the *Eco*RV site was chosen as position 1 of the sequence. The G + C content of the repeat measured 44%. This figure is close to the G + C content of cetacean mainband DNA as judged from isopycnic ultracentrifugation (Árnason et al. 1977, 1978).

The 1,579-bp sequence was analyzed with respect to presence and distribution of repetitions having the same orientation. Five 9-bp sequences occurred twice in the repeat. In no case were these sequences adjacent to each other. The distribution of these sequences showed no particular pattern, and thus these identities did not indicate that the 1,579-bp repeat had evolved by means of an accumulation of short repeats in direct orientation.

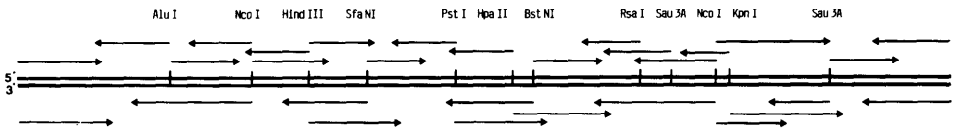

FIG. 1.—Sequencing strategy of the 1,579-bp repeat of the killer whale, *Orcinus orca*. Restriction sites used for labeling are indicated. Both strands of the repeat were sequenced according to the figure.

```
   1 ATCTTATGATGTTTCTTTTTTTGTGTGACTTATTTCATGTAGAATCATCGTACCTGAATC
  61 CACTCATTATGCTGCTACGGGCCTGATGACATAGATTTCATTGCTGAGTGATATTGCATT
 121 GTACGTAAGTACCACAAAGTCTTTATCCATTTTTCGCTTTCTGCGATACTGAACTTGTAC
 181 CATAAACGAGGTTCTTGTAAACAGAGCCGTCACAAACTTTGGGATGGCTGTGTCTTTTTG
 241 ATTTTAATTTCAACTAAGCTATAGGACCATAAGTGGAAGTGCCCTAGGCTCTGTTGCTTT
 301 GTTGTTTAGATGTTTCAGGAAACACCATACACTTCTCCAGAGTGGCTGTTGGCAATTTAC
 361 ATCCCGCCCATCAGCATAACAAGGCTCCCAATTCTCCATGGCCTGTCCCGCCTTTCTGGA
 421 TTTTACACTTTTTTCAGATGGCCCTTTTGACCGTGGGGCAGTGAGACTTCATGTAGTGCA
 481 GATTTCCTTTGCAAGCTTGCTTGGTTGGCCAAAAAGTTCGTATGCGTTTCTTTCTGAATA
 541 TATTCAGGAAAAAACGCATAAGACCTTTTTGGCCAAGTGCATCATTGTGGACGTTCTGCC
 601 TGTTTTCCTATGCTTTACATGCAATTCCAGTCTACCTCCTGAAATCGGTGTCTGGCAATT
 661 TTGCCCCGCTTTCAAGTCCTCTTGGCAGCCTTACTTCAATATATTTTTGGACGATAGCTG
 721 TCATTTATAACTCTGCAGGTTTGTGAATCACAGTGCCCCTGAGCTCCTTTCTTCAACTCG
 781 CTTTCTTGTGAGCTGGCTGCAACACCGCAGGATTGCTTGAGGCCCTCGTGTGGTTCCGGC
 841 AGGGCACGCTGAGCCTTTTGTTAATTCCTCTTCCTGGTGGGAAATGAGGGTTAAATTTGC
 901 CCGTCCAGACACCTCCAGCTAGTCTCTCATTGATTCTCCCTATTCCTGTTCATTTTCCGC
 961 AGAAATTGCAAACTGGGCCAAACAGGAGGTTAAAGGCACTGACTCTCCAAGTGGGGAGAG
1021 TGTTAGTAAAGCGTCTGGAATGTTGCACCCGAGTACCAGGGGACGAAAACTGAGACACAT
1081 TTGAACACGTTTCCCGATCACACGGTGGATCATACTCTGGGTTCCACATGCATGTTTTAG
1141 CTGAAGGAAGAATCCCTTAAACCTGGAGAGTTGAGACCCATGGAATGGGTACCATGCAAT
1201 ATGACTTCAAAGGGTCTGTATTTGCTAACTGAAACTCACTAATCCTATCACTGCTGCGTT
1261 TATGCCACTGTACACACGCTTGATTCCCTTTCGGAGACATATAAGTCCACAGGTTTTAAG
1321 ATTCTTACTAGTCAGGTATATTCTTAGGAGTTTAATATGGAGTGTTGAGTCCACTCGATC
1381 AGCAAGGAGTAGCTCTTGTCTATTAAATATTTGGCTTATGGAACGGTATCTGTGCTGATT
1441 TCAATCTCTGGTTTTATGCAGCACCTCAACTCACCTTTCCCCTTAAGCAAGCATAAGTTG
1501 GTGTTCTACATTTGAGACCCTGTTCTGTTTTGTAATTCAGTTCCTGTGTAGCCAAGTTTA
1561 CATTCCGTGTAGTAGTGAT
```

FIG. 2.—Base sequence of the killer-whale 1,579-bp repeat.

The presence of inverted repeats within which intrastrand pairing might occur was also investigated. Several such regions of high quality (UWGCG stemloop program; Deveraux et al. 1984) were observed. The most conspicuous of these was a 72-bp-long sequence in position 505–576. The quality value for complementarity for this sequence was 76.0. The sequence was preceded in position 491–500 by a 10-bp sequence having full intrastrand complementarity. A 100-bp sequence (position 485–584) that included the two inverted repeats, displayed as cruciform structures, is shown in figure 3. Other inverted repeats having high-quality values were found in positions 201–233, 251–304, and 717–754 (fig. 4). The 251–304 region is interrupted in position 269–286 by a sequence without intrastrand complementarity.

Several regions with an open reading frame were found in the repeat. The longest segment was a 330-bp sequence in position 398–728. Other regions of substantial length were located in positions 522–747, 1,262–952, and 1,112–942 (fig. 4).

Transcription in a cell-free system (Manley et al. 1980) was recorded in the repeat. The length of the transcript was ~340 nucleotides (fig. 5). The results were repeatable in the sequenced repeat but could not be verified in other clones.

Screening for sequence similarities between the 1,579-bp repeat and primate Alu sequences revealed sequence similarities in some cases. The most conspicuous similarities were found in position 1,143–1,168 of the killer-whale repeat and in position 163–188 of a cloned Alu sequence (Rubin et al. 1980) in which a 26-bp sequence had 22 matches. An 18-bp sequence, position 807–824, showed 17-bp identities with position 298–315 of an Alu sequence of the African green monkey (Saffer and Lerman 1983). These similarities were not elaborated on further, since they might be coincidental.

A highly repetitive component of the bowhead (*Balaena mysticetus*), a whalebone whale, has been sequenced in most of its length. The length of the repeat was 1,746 bp. Comparison between the killer-whale 1,579-bp repeat and the bowhead repeat showed substantial similarities between the two components. Figure 6 shows graphically the results of an alignment of the two sequences. Positions 1,247–1,426 in the bowhead

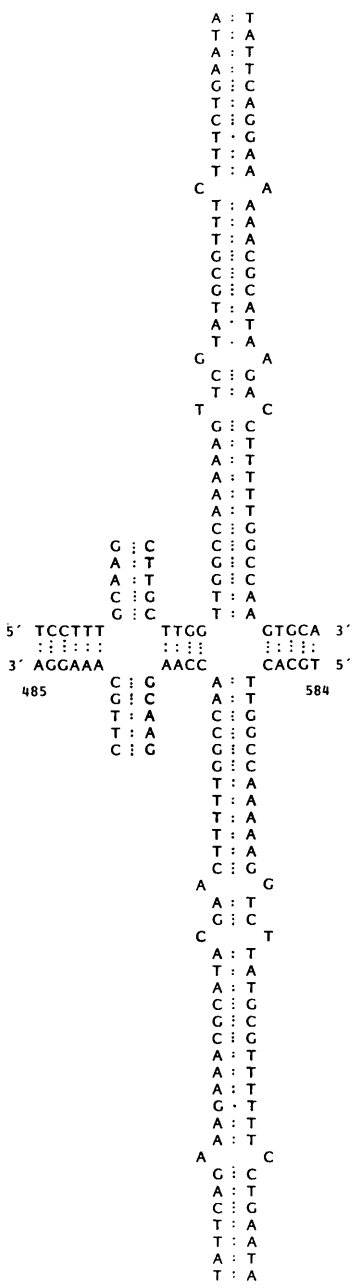

FIG. 3.—Sequence of the 485–584-bp region of the repeat. Inverted repeats in the region are depicted as cruciform structures. The consensus sequence of RNA polymerase III is located at the termination of the large cruciform structure.

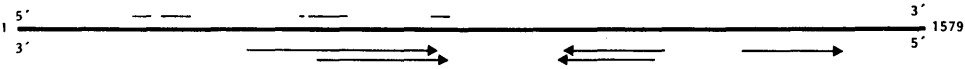

FIG. 4.—Localization of inverted repeats and open reading frames in the 1,579-bp component. Reading frames and direction of reading are indicated by arrows.

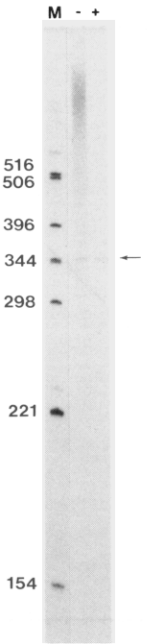

FIG. 5.—In vitro transcription of the sequenced killer-whale repeat. For discriminating between RNA polymerase II and RNA polymerase III activity, RNA synthesis was performed either in the absence (−) or in the presence (+) of α-amanitin. Electrophoretic separation was accomplished through a 6% polyacrylamide gel containing 8 M urea. A transcript, ~340 bp long, was recorded (arrow). M = pBR322 size marker.

component are missing in the killer-whale repeat. The distance between the *Pst*I and *Sac*I sites is the same in both repeats. Figure 7 shows an alignment of a 100-bp sequence, position 701–800, in which the *Pst*I and *Sac*I sites are located. Deviation from complete agreement was recorded in 14 positions.

C-bands are prominent in the killer-whale karyotype (Árnason et al. 1980). The C-bands are predominantly located in terminal and interstitial chromosome positions (fig. 8*a*). The chromosomal distribution of the 1,579-bp repeat was determined by in situ hybridization, the results of which are shown in figure 8*b*. The repeat was located in all major C-band-positive regions of the karyotype except in the long arm of chromosome sm9. Conspicuous C-band heteromorphism was observed in some chromosome pairs. The degree of heteromorphism was paralleled by different labeling intensity. This is particularly evident in the short arms of chromosomes st1 and st2 in figure 8*b*.

Discussion

Some cetacean species have been shown to harbor highly repetitive components with highly different degrees of divergence (Árnason and Widegren 1984). In the mysticete genus *Balaenoptera,* three unrelated, highly repetitive components have been studied. One variable component is specific for the genus. This component is assumed to be more recent than the other two, which occur outside the balenopterids. Of these two components, one characterizes the whalebone whales, and the other is the common 1,740-bp cetacean component. Conceivably, the presently analyzed 1,579-bp repeat

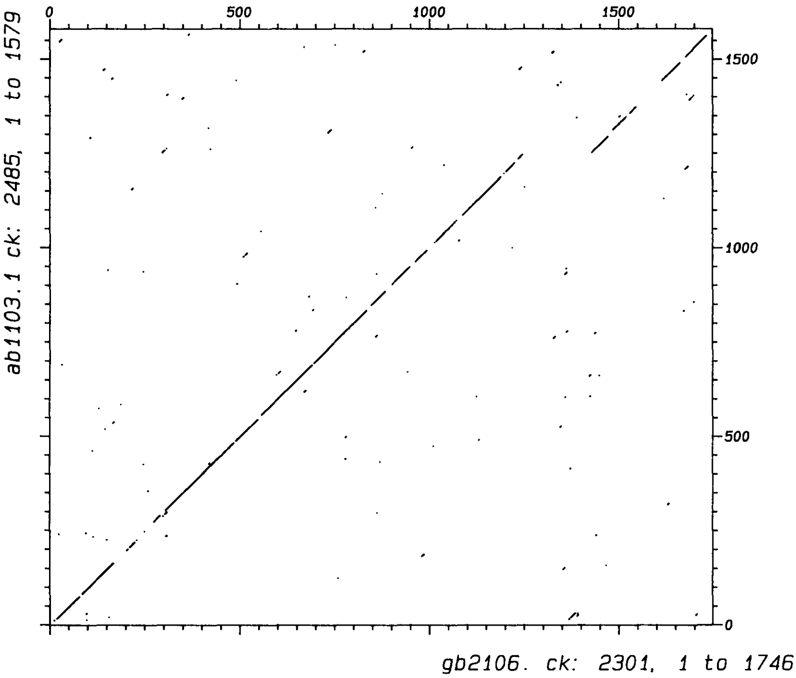

FIG. 6.—Graphic demonstration of results of alignment between the killer-whale, *Orcinus orca,* (vertical axis) and bowhead, *Balaena mysticetus,* (horizontal axis) repeats. Positions 1,533–1,605 of the bowhead have not been sequenced. The length of the killer-whale repeat is 1,579 bp, and the length of the bowhead repeat is ~1,740 bp. Region 1,247–1,426 of the bowhead component is missing in the killer-whale repeat. A dot is placed on the diagram whenever 14 of 21 consecutive aligned nucleotides match (Deveraux et al. 1984).

evolved from the 1,740-bp repeat that it has largely replaced in the delphinids. A common notion (e.g., Lewin 1982) is that highly repetitive DNA diverges rapidly, with each species having its own specific pattern of highly repetitive DNA. The 1,740- and 1,579-bp components differ radically from this concept, since their evolution is evidently very slow.

Molecular hybridization between various cetaceans has shown that the amount of the 1,740-bp repeat is limited in the family Delphinidae and that the delphinid repeat is absent in other cetacean families. It can be assumed that the delphinid repeat replaced the 1,740-bp repeat prior to the evolutionary radiation of the Delphinidae

```
                                       Pst I
701  ATATTTTTGGACGATAGCTGTCATTTATAACTCTGCAGGTTTGTGAATCA
      ||||| ||||||  || | ||||||||||||||||||||||||||||
     GTATTTGTGGACGGGAGTTATCATTTATAACTCTGCAGGTTTGTGAATTG

               Sac I
     CAGTGCCCCTGAGCTCCTTTCTTCAACTCGCTTTCTTGTGAGCTGGCTGC   800
     ||||||||||||||||| ||  |||||||||||| ||| ||||||||  |
     CAGTGCCCCTGAGCTCCATTTGTCAACTCGCTTTTTTGGGAGCTGGCCAC
```

FIG. 7.—Alignment of 100-bp regions (position 701–800) of the killer-whale (top) and the bowhead (bottom) repeats. The distance between the *Pst*I and *Sac*I sites is the same in both repeats. Nonidentical nucleotides occurred in 14 positions.

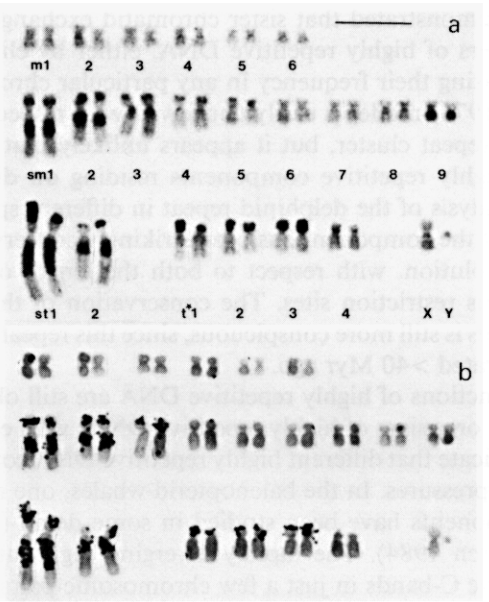

FIG. 8.—The karyotype of the killer whale. *a*, C-banded karyotype; *b*, results of in situ hybridization using a cloned repeat as labeled probe. The 1,579-bp component is present in all major C-bands. The C band heteromorphisms in *a* are reflected by different degrees of labeling in *b*. This is particularly evident in the short arms of chromosomes st1 and st2. Bar = 10 μm. Both cytological preparations are from the same specimen.

and that the repeat has been conserved during the later evolution of the delphinids, irrespective of the different degrees of amplification in different delphinid genomes. The greatest amplification has occurred in the killer-whale genome in which the number of repeats was estimated at 4–5 × 10^5, comprising more than 15% of the genome.

The results of the in situ hybridization showed that the amplification of the 1,579-bp repeat in the killer-whale genome had taken place both at ancestral delphinid C-band sites and also at C-band sites unique to the killer whale.

The way by which repetitive DNA elements spread through eukaryotic genomes is still not known in detail, but substantial information is available on transposable elements in *Drosophila* (Rubin 1983) and the Alu sequences in various mammals. Transposable DNA elements are often flanked by tandem repeats of DNA, but no such structures were found in the 1,579-bp repeat. These observations do not, however, exclude the possibility that such structures may exist in a limited number of repeats and may promote the introduction of the repeat at new chromosomal sites. It has been proposed (Jagadeeswaran et al. 1981; Van Arsdell et al. 1981) that transposition of Alu DNA may be directly linked to its transcription. Transcription was registered in the now sequenced 1,579-bp repeat, and a TGGCCAAGTG sequence corresponding to the consensus promoter sequence of RNA polymerase III (Galli et al. 1981) occurs in the 1,579-bp repeat. The 10-bp sequence is in position 570–579 of the repeat at the very end of the inverted 72-bp repeat. The sequence is within two open reading frames of substantial length. It is probable that the transcriptional activity observed in the repeat is connected with the presence of the RNA polymerase III promoter consensus sequence, but the localization and the general occurrence of trancriptional activity is yet to be determined.

Smith (1976) demonstrated that sister chromatid exchange would impose uniformity within clusters of highly repetitive DNA, either by eliminating new repeat varieties or by increasing their frequency in any particular chromosome region. The validity of Smith's (1976) model is easily conceived with respect to maintaining homogeneity within a repeat cluster, but it appears unlikely that it would prevent divergence between highly repetitive components residing on different chromosome pairs. Restriction analysis of the delphinid repeat in different species (Árnason et al. 1984) has shown that the component has been strikingly conserved during the 20–24 Myr of delphinid evolution, with respect to both the length of the repeat and the localization of various restriction sites. The conservation of the 1,740-bp repeat in other cetacean families is still more conspicuous, since this repeat has been maintained in lineages that separated >40 Myr ago.

The possible functions of highly repetitive DNA are still obscure. However, the presence in the same organism of highly repetitive DNAs with entirely different rates of evolution may indicate that different highly repetitive DNA components are subject to different selective pressures. In the balenopterid whales, one rapidly diverging and two conserved components have been studied in some detail (Árnason et al. 1978; Árnason and Widegren 1984). The rapidly diverging (light-satellite) component is present in centromeric C-bands in just a few chromosome pairs. Compared with the other two components, the light satellite appears to evolve quite freely and to be subject to limited selective pressure with respect to its composition. However, it does not follow from this that such components are immaterial to the organism that carries them. C-band heteromorphism is striking in most cetaceans (Árnason 1974), and based on these findings and the observations of Nankivell (1976) and Miklos and Nankivell (1976) that chiasmata positions are affected by the distance to C-band sites, Árnason et al. (1978) suggested that the different amounts of highly repetitive DNA in heteromorphic C-bands would put adjacent euchromatin out of parity and thus reduce the meiotic recombination in the vicinity of highly repetitive DNA. This effect would be achieved just by the presence of the heterochromatin, irrespective of the composition of the highly repetitive DNA.

The conservation of the presently studied delphinid repeat and the 1,740-bp cetacean repeat makes it highly plausible that these components are subject to selective pressure. Such pressure might be based on transcriptional activity in particular tissues or at certain developmental stages, or on structural properties of these components that might affect the spatial organization of the chromosomes in meiosis or mitosis. Evolutionary conservation of these components might also be imposed by their being selfish DNA (Orgel and Crick 1980) whose structure is optimal for being maintained and propagated in cetacean genomes. However, since it is hardly possible to test the selfish DNA hypothesis experimentally, we feel that possible functions of this repeat in cetacean genomes need to be thoroughly studied before other explanations are sought.

Acknowledgments

This work was supported by the Swedish Natural Science Research Council (grants B-BU 3655-105 and B-BU 3655-106) and by The Erik Philip-Sörensen Foundation.

LITERATURE CITED

ÁRNASON, Ú. 1974. Comparative chromosome studies in Cetacea. Hereditas **77**:1–36.

ÁRNASON, Ú., M. HÖGLUND, and B. WIDEGREN. 1984. Conservation of highly repetitive DNA in cetaceans. Chromosoma **89**:238–242.

ÁRNASON, Ú., A. LIMA-DE-FARIA, H. GRANSTRÖM, and M. ISAKSSON. 1977. Analysis of cetacean DNA. Localization of 18S and 28S ribosomal RNA cistrons in a heavy DNA component. Hereditas **87**:67–76.

ÁRNASON, Ú., R. LUTLEY, and B. SANDHOLT. 1980. Banding studies on six killer whales: an account of C-band polymorphism and G-band patterns. Cytogenet. Cell Genet. **28**:71–78.

ÁRNASON, Ú., I. F. PURDOM, and K. W. JONES. 1978. Conservation and chromosomal localization of DNA satellites in balenopterid whales. Chromosoma **66**:141–159.

ÁRNASON, Ú., and B. WIDEGREN. 1984. Different rates of divergence in highly repetitive DNA of cetaceans. Hereditas **101**:171–177.

DEVERAUX, J., P. HAEBERLI, and O. SMITHIES. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.

GALLI, G., H. HOFSTETTER, and M. C. BIRNSTIEL. 1981. Two conserved sequence blocks within eukaryotic tRNA genes are major promotor elements. Nature **294**:626–631.

JAGADEESWARAN, P., B. G. FORGET, and S. M. WEISSMAN. 1981. Short interspersed repetitive DNA elements in eucaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts? Cell **26**:141–142.

LEWIN, R. 1982. Repeated DNA still in search of a function. Science **217**:621–623.

MANLEY, J. L., A. FIRE, A. CANO, P. A. SHARP, and M. L. GESTER. 1980. DNA dependent transcription of adeno virus genes in a soluble whole cell extract. Proc. Natl. Acad. Sci. USA **77**:3855–3859.

MAXAM, A. M., and W. GILBERT. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.

MIKLOS, G. L. G., and R. N. NANKIVELL. 1976. Telomeric satellite DNA functions in regulating recombination. Chromosoma **56**:143–167.

NANKIVELL, R. N. 1976. Karyotype differences in the crenaticeps group of Atractomorpha (Orthoptera, Acridoidea, Pyrgomorphidea). Chromosoma **56**:127–142.

ORGEL, L. E., and F. H. C. CRICK. 1980. Selfish DNA: the ultimate parasite. Nature **284**:604–607.

RIGBY, P. W., M. DIECKMANN, C. RHODES, and P. BERG. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. **113**:236–251.

RUBIN, G. M. 1983. Dispersed repetitive DNAs in Drosophila. Pp 329–361 *in* J. A. SHAPIRO, ed. Mobile genetic elements. Academic Press, New York.

RUBIN, G. M., C. M. HOUCK, P. L. DEININGER, T. FRIEDMANN, and C. W. SCHMID. 1980. Partial nucleotide sequence of the 300 nucleotide interspersed repeated human DNA sequences. Nature **284**:372–374.

SAFFER, J. D., and M. I. LERMAN. 1983. Unusual class of Alu sequences containing a potential z-DNA segment. Mol. Cell Biol. **3**:960–964.

SINGH, L., I. F. PURDOM, and K. W. JONES. 1977. Effect of different denaturing agents on the detectability of specific DNA sequences of various base compositions by in situ hybridization. Chromosoma **60**:377–389.

SMITH, G. P. 1976. Evolution of repeated DNA sequences by unequal crossover. Science **191**: 528–535.

VAN ARSDELL, S. W., R. A. DENISON, L. B. BERNSTEIN, A. M. WEINER, T. MANSER, and R. F. GESTELAND. 1981. Direct repeats flank three small nuclear RNA pseudogenes in the human genome. Cell **26**:11–17.

KEN W. JONES, reviewing editor

Received February 14, 1985; revision received April 9, 1985.