Canadian Technical Report of

Fisheries and Aquatic Sciences 2080

1996

AMPLIFICATION AND SEQUENCING OF CONTROL REGION MITOCHONDRIAL DNA FROM THE BELUGA WHALE, *Delphinapterus leucas*

by

W.R. Lillie, J.G. Brown Gladden and D.N. Tretiak

Central and Arctic Region

Canada Department of Fisheries and Oceans

Winnipeg, Manitoba, Canada, R3T 2N6

This is the 74th Technical Report

from the Central and Arctic Region, Winnipeg

Cat. no. Fs 97-6/2080E ISSN 0706-6457

Correct citation for this publication is:

W.R. Lillie, J.G. Brown Gladden, and D.N. Tretiak. 1996. Amplification and sequencing of control region mitochondrial DNA from the beluga whale, *Delphinapterus leucas*. Can. Tech. Rep. Fish. Aquat. Sci. 2080: iv + 8 p.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

W.R. Lillie, J.G. Brown Gladden, and D.N. Tretiak. 1996. Amplification and sequencing of control region mitochondrial DNA from the beluga whale, *Delphinapterus leucas*. Can. Tech. Rep. Fish. Aquat. Sci. 2080: iv + 8 p.

The polymerase chain reaction (PCR) and direct sequencing were used to obtain the beluga whale (*Delphinapterus leucas*) mitochondrial tRNA proline gene sequence, the adjacent control region left domain (L-strand), central region, and right domain sequences. Sequences are presented and compared to those of other cetaceans. The central region, conserved regions (A, B, C, D, E, F, L and M) and the right domain are similar to those found in related species. Beluga control region sequences are most similar to the odontocete sequences (dolphin (85.8 % overall similarity) and killer whale (85.4 %)). The mysticete sequences of the fin and minke whales, although comparable are less similar, 79.1 % and 77.5 % respectively. Differences are located primarily in the left domain (av. difference in four species = 28.70 %) followed by the right domain (16.12 %) then the central region (7.95 % av. difference).

Sufficient genetic variability has been found in a section of the beluga left domain to conduct population studies. Eighteen variant sites expressing 37 haplotypes have been found during routine sequencing of a 235 basepair region of the left domain. An additional four variant sites have been found in an otherwise well conserved central region.

Keywords: beluga; whale; *Delphinapterus leucas;* cetacea; mitochondrial DNA; control region; PCR, sequencing.

RÉSUMÉ

W. R. Lillie, J. G. Brown Gladden, and D. N. Tretiak. 1996. Amplification and sequencing of control region mitochondrial DNA from the beluga whale, *Delphinapterus leucas*. Can. Tech. Rep. Fish. Aquat.Sci. 2080: iv + 8 p.

Pour obtenir la séquence génétique de la proline de l'ARN de transfert mitochondrial du béluga (*Delphinapterus leucas)*, séquences du domaine gauche adjacent à la région de contrôle (chaîne L), de la région centrale et du domaine droit, on a eu recours à la réaction en chaîne à la polymérase (PCR) et au séquençage direct. Les séquences sont présentées et comparées à celles d'autres cétacés. Dans la région centrale, les régions conservées (A, B, C, D, E, D, L et M) et le domaine droit, les séquences sont similaires à celles trouvées dans des espèces apparentées. Dans l'ensemble, les séquences des régions de contrôle du béluga présentent plus de similarités avec les séquences des odoncètes: le dauphin (85,8% de similarité d'ensemble) et l'épaulard (85,4%). Bien que comparables, les séquences des mysticètes. petit rorqual et rorqual commun, sont moins similaires (79,1% et 77,5% respectivement). C'est dans le domaine gauche que se remarquent les différences les plus importantes (différences les plus importantes (différence moyenne relevée dans quatre espèces: 28,7%), puis vient le domaine droit (16,12%), et enfin la région centrale (7,95%).

On a trouvé une variabilité génétique suffisane dans une section du domaine gauche du béluga pour justifier des études de population. Lors d'un séquençage de routine d'une région de 235 paires de base du domaine gauche, on a dénombré dix-huit sites variants exprimant 37 haplotypes. En outre, quatre sites variants supplémentaires ont été trouvés dans une région centrale par ailleurs bien conservée.

Mots clés: béluga; baleine; *Delphinapterus leucas*; cétacé; ADN mitochondrial; région de contrôle; PCR; séquençage.

INTRODUCTION

Beluga whales, *Delphinapterus leucas*, range throughout much of the Arctic. In Canadian arctic waters, beluga whales migrate from wintering areas and congregate in shallow estuaries during the brief arctic summer. In some areas, including Ungava Bay and Cumberland Sound, the current numbers of summering beluga appear to be sharply reduced from historic levels (Richard 1991). This may indicate the existence of isolated summering stocks. Accurate estimates of stock size and trends require that stocks be identified and delineated to ensure their sound management and conservation. DNA analysis is being used to determine the genetic structure of these stocks.

This paper describes the use of the polymerase chain reaction (PCR)(Mullis and Faloona 1987) to amplify mitochondrial DNA, asymmetric PCR to produce single stranded DNA, and sequencing using the dideoxynucleotide chain termination method. The mitochondrial DNA control region sequence is presented for the beluga, *Delphinapterus leucas*, and is compared to that of the dolphin, *Cephalohynchus commersoni* (Southern et al. 1988), fin whale, *Balaenoptera physalus* (Arnason et al. 1991), killer whale, *Orcinus orca* (Hoelzel and Dover 1991) and minke whale, *Balaenoptera acutorostrata* (Hoelzel and Dover 1991).

Variant sites found in routine sequencing of a 235 basepair region of the L-strand left domain are shown to illustrate the type and location of variant sites.

MATERIALS AND METHODS

Beluga tissues from aboriginal hunts in the Canadian Arctic have been collected by Canada Department of Fisheries and Oceans biologists over many years. These have been archived as frozen whole tissue, tissue preserved in saturated NaCl/20% DMSO solutions or as DNA extracted from muscle or skin using the protocol of Sambrook et al. (1989). The sequence shown in this study is that of an individual taken from an aboriginal hunt in 1989 at Kendall Island in the Beaufort Sea.

The initial amplification was performed with the universal primers L15926 and H0651 (Kocher et al. 1989). The Perkin Elmer/Cetus protocol was used with the universal primers L15926 and H0651 (Kocher et al. 1989). Optimum results were obtained with 2.0 μM L15926 and 0.1 μM H0651 and a nucleotide concentration of 100 μM. The thermocycle parameters were: five cycles with denaturation at 94°C for 1.5 minutes, annealing at 37°C for 1 minute, and extension at 72°C for 1 minute with an additional 5 seconds per cycle; 20 cycles with denaturation at 92°C for 1 minute, annealing at 37°C for 1 minute, and extension at 72°C for 1.5 minutes with an additional 5 seconds per cycle. A final cycle at 72°C for 7 minutes finished the thermocycle regime. This amplified the entire mitochondrial control region, the tRNA proline gene, and parts of the tRNA threonine and tRNA phenylalanine genes producing a product approximately 1200 base pairs in length.

The initial amplification product was separated by gel electrophoresis using 1% NuSieve™ agarose and a gel plug (approximately 3 μl) containing only the desired amplification product was removed. This plug was reamplified asymmetrically with 0.005 μM L15926 and 0.5 μM of a nested primer Phh (Y C A T C T A R R C A T T T T C A G T G) (IUPAC coding, Y = C and/or T, R = A and/or G)) using the same thermocycler regime. The product of this reaction was purified and concentrated using Geneclean 2™ before sequencing.

Dideoxynucleotide termination sequencing was performed using Sequenase Version 2™ (United States Biochemicals Corp.) and a ^{35}S-ATP radiolabel. The Sequenase™ protocol was used with the following modifications to reduce reannealing of the template: 0.5% Nonidet P40 (NP40) was added to the templates before they were denatured at 94°C for 10 min. then quench-cooled on ice; Sequenase™, ^{35}S-ATP, dithiothreitol (DTT), Mn. buffer, and enzyme dilution buffers were premixed then added directly to the cold template mixture. This cold mixture was then added to the termination mixtures and heated at 50°C for 5 minutes. Parallel sequencing reactions were done using inosine 5' triphoshate(ITP) in place of ddGTP to remove gel compressions. Primers used for sequencing from the 5' end(L-strand) were L15926 (Kocher et al.,1989), Bel5' (A C A T T T T A C T G T G A C T A T T G) exclusively for beluga, dWl (A Y A T R C T A T G T A T W A T Y G T G) exclusively for whales, dFl (C A T G C C G C G T G A A A C C A G C A), dCl (C T T A A A T A A G A C A T C T C G A T) and dBl (C A T A C A T T T G G T A T T T T T T T). In most cases primer locations correspond to conserved regions throughout the mitochondrial control region. Conserved regions, primer locations and the direction of sequencing are shown in Figure 1.

Electrophoresis was performed on a Bio-Rad sequencing system using wedge spacers, double sharks tooth combs, and BRL Gelmix 8™ acrylamide. Band spacing was further enhanced by adding one part 10 M ammonium acetate to seven parts 1xTBE buffer in the lower reservoir and by using 0.5xTBE in the upper reservoir (modified from Wilkinson and Chapman 1991). Gels were run between 45 and 55°C for 3

hours.

The beluga sequence was aligned with other cetacean species using the MacVector™ 4.1.4 sequence analyses program. Parsimony analyses, including computation of a distance matrix and determination of a phylogenetic tree, was performed using PAUP version 3.1 (Swofford 1993)

RESULTS AND DISCUSSION

The control region right domain (L strand) between conserved region C and the tRNA phenylalanine gene (method not shown) was amplified first. A double stranded PCR product was readily obtained using primers Cl and Phh or H0651. We were initially unable to sequence the dsDNA and only obtained sequences after agarose gel electrophoresis, asymmetric PCR with Phh or H0651 limiting, and a final phenol/ chloroform cleanup. Although the genetic variability expressed in the right domain allows one to easily differentiate among different whale species, the sequences are relatively well conserved and little variation was found within potential beluga populations. This was inadequate for population studies, and the left domain was examined next.

The dsDNA PCR product of the entire control region was easily obtained in the first amplification although it was necessary to reduce the level of primer H0651 much below that of L15926 (0.025 and 0.5 μM respectively). Asymmetric PCR was not possible unless L15926 was the limiting primer, perhaps showing that this primer may be less than ideal. Asymmetric PCR was also not possible when H0651 was reused in the second amplification. The nested primer Phh (1 μM) and a reduced level of L15926 (.005 μM) provided a good asymmetric product suitable for sequencing the L-strand after a Geneclean™ purification procedure.

Frequent stops (bands in all four lanes) were initially encountered while sequencing the left domain. This is a complex region containing considerable secondary structure involved in various complex protein/DNA interactions and termination of the developing H strand D-loop (Saccone et al. 1987, 1991). Our sequences show considerable potential for various secondary structures which could cause the DNA polymerase to pause or stop during sequencing reactions, forming nonspecific terminations in all four lanes. Our methods have been modified to reduce reannealing and prevent post denaturation secondary structure formation and have largely overcome banding in all four lanes (stops were only encountered before the long run of Cs in the right domain - This area was resolved by sequencing in both directions). With this problem removed we were able to selectively sequence throughout the control region including the region sequenced earlier. The versatility offered by amplifying the entire control region make this the preferred method, at least for cetacea. Agreement of overlapping sequences obtained from different amplification procedures did confirm the validity of both methods for cetacea.

The beluga mitochondrial DNA control region sequence was aligned to the dolphin (Southern et al. 1988), fin whale (Arnason et al. 1991), killer whale, and minke whale (Hoelzel and Dover 1991) sequences (Figure 2). Variant sites found among beluga are entered using the IUPAC code to illustrate the type and location of variant sites. The overall length of the beluga control region (917 bp) is similar to that found in other cetacea (dolphin = 900 bp, killer whale = 921 bp, finwhale = 930 bp, and minke whale = 946 bp). The central region (Figure 2, positions 332 to 591) and right domain (positions 592 to 952) sequences show few deletions or additions with length differences confined to the 3' end adjacent to the tRNAphe gene. Alignment of this region is straightforward. The conserved regions A, B, C, D, E, F, L and M(CSB 1) outlined by Southern et. al. 1989, were easily distinguished and appear in the expected locations. These regions are shown on the beluga sequence using the terminology adopted by Southern et al. 1989. These sequences have been entered into Genbank with the following accession numbers; Control region - U18117; tRNA pro - U18118.

Length differences were found in the left domain (positions 1 to 331) and alignment was considerably more difficult. The MacVector™ align to folder function was used with the following parameters; match = +4, mismatch = -2, indel penalty = 12, continuing indel penalty = 4. Alternative alignments were also performed using the MacVector™ program with match = +1 and no penalty for a mismatch. Neither was entirely satisfactory. Alignment of the left domain was only possible when the region was subdivided into three and obvious indels inserted manually before a second alignment. Alignments of the left domain, particularly the placement of additions and deletions (indels) are open to interpretation until the functional constraints of this region are better understood.

Simple percent differences, calculated from the aligned sequences for the tRNApro sequence, the entire control region, the left domain(L-strand), the central region, the right domain(L-strand) and the conserved regions are presented in Table 1. The smaller central region size designation of Hoelzel et al. (1991), was used for calculations. Indels that corresponded between sequences were counted as a

match, those that did not share a corresponding position were counted as a mismatch. Differences occur primarily in the left domain where length differences(indels) are common (average difference =28.7%). The beluga central region, including conserved blocks B, C, D, E, F, and M is well conserved (differences = 7.9 and 2.4% respectively). Blocks A(av. dif = 19.1%), L(av. dif. = 20.8%), and K(av. dif. = 47.2%) are not well conserved. These differences are consistent with those found in other cetaceans, except the sperm whale (Dillon and Wright 1993) where the central region is not well conserved.

The absence of major indels in the central region and right domain plus conservation of conserved regions combined with frequent interspersed variation make this a suitable region for between species comparisons. A distance matrix for the combined central region and right domain is shown in Table 2. Figure 3 shows a phylogenetic tree using the same combined sequence. Beluga sequences are most similar to other odontocetes used in this study. However the beluga does not group closely with these whales but occupies a position between odontocetes and the two mysticetes used in this study. The branch length to the toothed whales is shorter, reflecting a closer relationship between beluga and the odontocetes. This arrangement was consistent when the entire control region, the central region, the right domain and even the left domain were analyzed independently. The left domain was excluded only because indels introduced to enhance alignment to the beluga sequence could bias results.

In contrast to the right domain and tRNA proline gene, the left domain control region is highly variable. Variants occur predominantly in the left domain in an area where the nascent h strand (D-loop) is terminated and has various secondary structures and protein/DNA interactions (Saccone et al. 1991). In the 235 basepairs routinely sequenced (position 134 to 384) we have found 18 variable sites expressing 37 haplotypes in 450 individuals. Non routine sequencing of a much smaller number of individuals has revealed four additional variant sites in the central region. It was surprising to find variant sites (positions 353, 354, 356, 383, 428, 437, 481, 482) in an otherwise well conserved central region. Figure 2 shows the location of these variable sites. The left domain has frequently been found to be the most variable portion of the control region and has been used in population studies on beluga (Hare 1990), humpback whales (Baker et al.,1993), bears (Shields and Kocher 1991), bats (Wilkinson and Chapman 1991) and shrews (Stewart and Baker 1994). Our findings are consistent with the findings of these authors.

CONCLUSIONS

Beluga control region sequences are most similar to other odontocete sequences used in this study (dolphin (85.8 % overall similarity) and killer whale (85.4 %)). The mysticete sequences of the fin and minke whales, although comparable are less similar, 79.1 % and 77.5 % respectively. Beluga sequences have conserved sequence blocks corresponding to those found in related species. This suggests a common evolutionary background and a conservation of functional elements. Between species variation is located predominantly in the left domain (av. difference in four species = 28.70 %) followed by the right domain (16.12 %) then the central region (7.95 % av. difference). This structural consistency combined with abundant variation make it possible to discriminate between even the most closely related species as well as to determine evolutionary distances using more conserved regions. Within species variation in beluga will provide valuable information on family units, site fidelity, and population structure.

ACKNOWLEDGMENTS

This study was supported in part by the Panel on Energy Research and Development (PERD), the Fisheries Joint Management Committee (FJMC), the Tungavik Federation of Nunavut (TFN) and the World Wildlife Fund (WWF). We thank Dr. J.W. Clayton for securing funding from these organizations. Laboratory assistance was provided by M.K. Friesen, D. Lint, L. Postma and D. Risbey. We also thank Dr. S. Cosens, Dr. T. Crease, Dr. B. de March, Dr. G. Klassen, R. Jestadt and D. Murray for reviewing this manuscript. We especially thank the many field personnel who supplied samples.

REFERENCES

Arnason, U., A. Gullberg, and B. Widegren. 1991. The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus*. J. Mol. Evol. 33: 556-568.

Arnason, U., A. Gullberg, and B. Widegren. 1993. Cetacean mitochondrial DNA control region: Sequences of all extant Baleen whales and two sperm whale species. Mol. Biol. Evol. 10(5): 960-970.

Baker, C.S., A. Perry, J.L. Bannister, M.T. Wienrich, R.B. Abernathy, J. Calambokidis, J. Lien, R.H. Lambertsen, J. Urbán Ramírez, O.

asquez, P. J. Clapman, A. Alling, S.J. O'Brien, and S. R. Palumbi. 1993. Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. Proc. Natl Acad. Sci. 90: 8239-8243.

Dillon, M.C., and J. W. Wright. 1993. Nucleotide sequence of the D-loop region of the sperm whale (*Physeter macrocephalus*) mitochondrial genome. Mol. Biol. Evol. 10(2):296-305.

Hare, M.. 1990. A preliminary study using mitochondrial DNA nucleotide sequence comparisons for stock identification in Alaska beluga whales (*Delphinapterus leucas*). Final Report to the Alaska Inuvialuit Beluga Whale Committee (A.I.B.W.C). 17 pgs.

Hoelzel, A.R., and G.A. Dover. 1991. Mitochondrial D-loop DNA variation within and between populations of the minke whale (*Balaenoptera acutorostrata*). Rep. Int. Whal. Commn. (Special Issue 13).

Hoelzel, A.R., J.M. Hancock and G.A. Dover. 1991b. Evolution of cetacean Mitochondrial D-loop region. Mol. Biol. Evol. 8(3): 475-493.

Kocher, T.D., W.K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca, and A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. Proc. Natl Acad. Sci. U.S.A. 86: 6196-6200.

Mullis, K. B. , and F. R. Faloona. 1987. Specific synthesis of DNA in vitro via a polymeraze catalyzed chain reaction. Methods in Enzymology 155: 335-350.

Richard, P.R. 1991. Status of the beluga, *Delphinapterus leucas*, of southeast Baffin Island, Northwest Territories. Can. Field Nat. 105(2): 206-214.

Saccone C., M. Antimonelli, and E. Sbisà. 1987. Structural elements highly preserved during the evolution of the D-loop containing region in vertebrate mitochondrial DNA. J. Mol. Evol. 26: 205-211.

Saccone, Cecilia, G. Pesole, and E. Sbisà. 1991. The main regulatory region of mammalian mitochondrial DNA structure - Function model and evolutionary pattern. J.Mol. Evol. 33: 83-91.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning; A Laboratory Manual; Second Edition. Cold Spring Harbor Laboratory Press.

Shields, G.F., and T.D. Kocher. 1991. Phylogenetic relationships of North American ursids based on analysis of mitochondrial DNA. Evolution 45: 218-221.

Southern, S.O., P.J. Southern and A.E. Dizon. 1989. Molecular characterization of a cloned dolphin mitochondrial genome. J. Mol. Evol. 28: 32-42.

Stewart, D. T., and A. J. Baker. 1994. Patterns of sequence variation in the mitochondrial D-loop region of shrews. Mol. Biol. Evol. 11(1):9-21

Swofford, D.L. 1993. PAUP: phylogenetic analyzes using parsimony, version 3.1. Illinois Natural History Survey, Champaign, IL.

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

Table 1. Simple % differences found between beluga sequences and other cetaceans within the control region and tRNApro gene. Sizes include indels.

Region	Size	Dolphin	Killer W.	Fin W.	Minke W.	Average difference
tRNA proline	66	3.0	N.A.	16.7	N.A.	9.9
Entire D-loop	951	14.2	14.6	20.9	22.5	18.1
Left domain	331	22.7	24.2	32.9	35.1	28.7
Central region	259	8.5	6.2	8.5	8.5	7.9
Right domain	361	10.8	12.3	19.5	21.8	16.1
Conserved Regions						
A	17	11.8	17.6	23.5	23.5	19.1
B	19	0	0	0	0	0
C	27	3.7	0	0	0	0.9
D	25	0	0	0	4.0	1.0
E	40	7.5	0	5.0	7.5	5.0
F	28	0	0	3.6	0	0.9
K	27	44.4	40.7	55.6	48.1	47.2
L	30	10.0	6.7	33.3	33.3	20.8
M	17	5.9	5.9	5.9	5.9	5.9
B,C,D,E,F,M	156	3.2	0.6	2.6	3.2	2.4

Table 2. Pairwise distances between taxa using the combined central region and left domain sequences - Above diagonal: Mean distances; Below: Absolute distances

	Beluga	Dolphin	Killer W.	Fin W.	Minke W.
Beluga	-	0.095	0.089	0.145	0.150
Dolphin	59	-	0.069	0.166	0.155
Killer W.	55	43	-	0.153	0.150
Fin W.	90	103	95	-	0.071
Minke W.	93	96	93	44	-

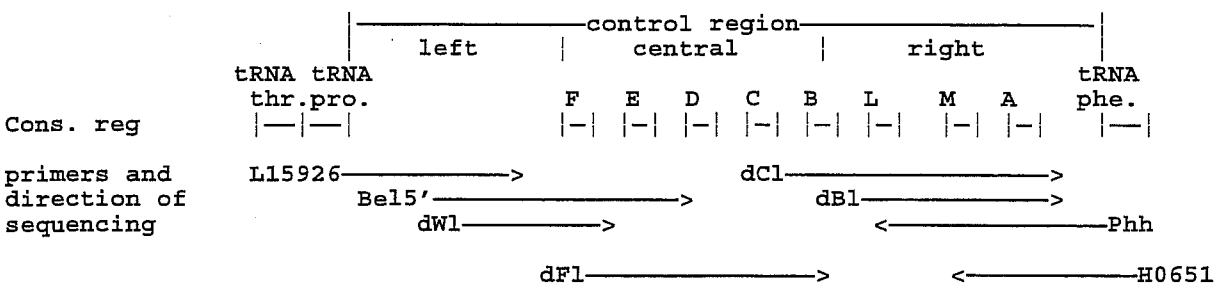

Figure 1. Schematic drawing of the control region and adjacent tRNA genes showing conserved regions, primer locations and direction of sequencing. Conserved regions are shown using the terminology of Southern et al., 1989. The configuration is that of the light strand.

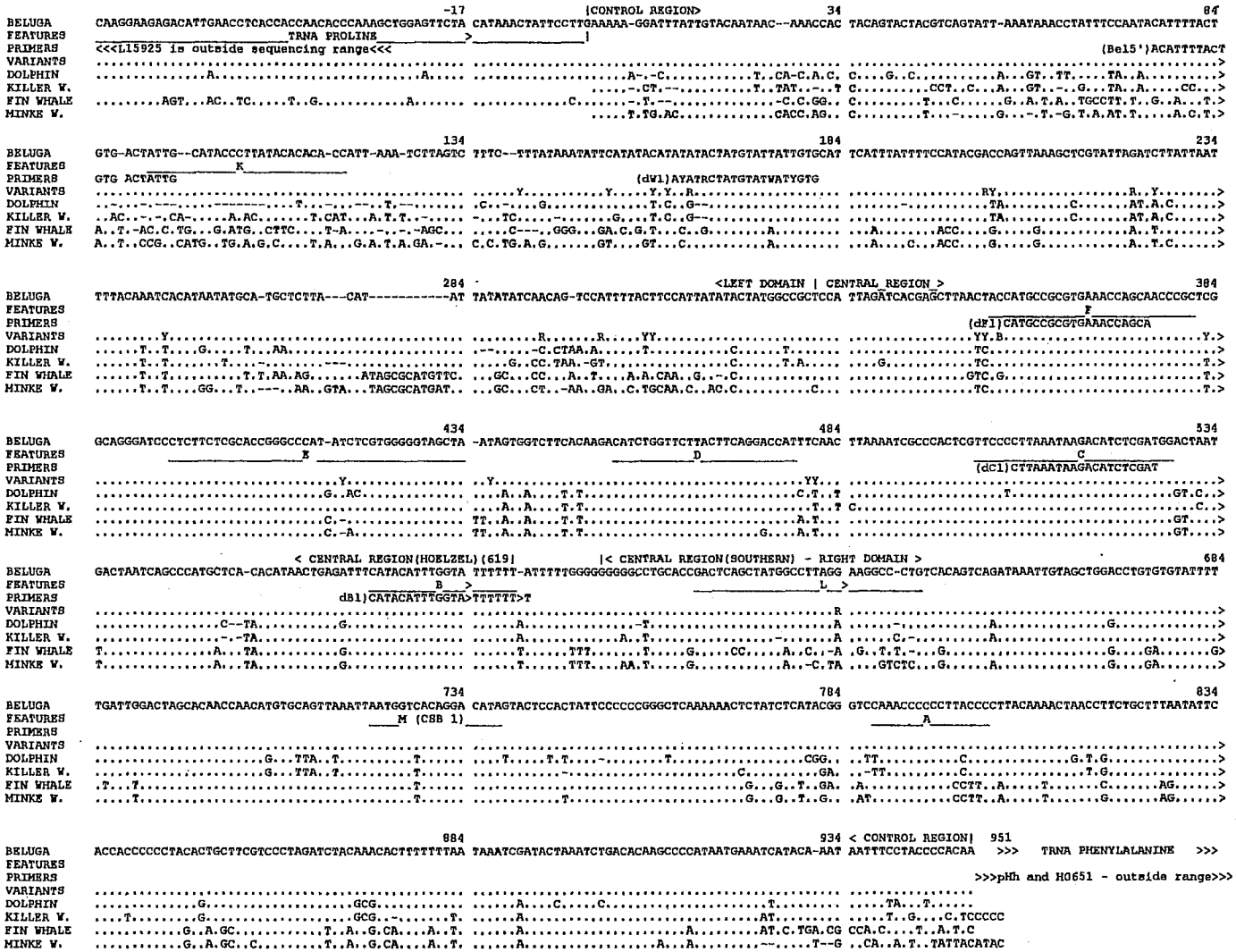

FIGURE 2 Beluga Whale mitochondrial control region sequence (L strand) aligned to dolphin(Southern et al., 1989), killer whale (Hoelzel et al., 1991), fin whale (Arnason et al.,1991) and minke whale sequences(Hoelzel et al., 1991). Variant sites are shown using the IUPAC code (B=C,G or T)(R= A or G)(Y=C or T).

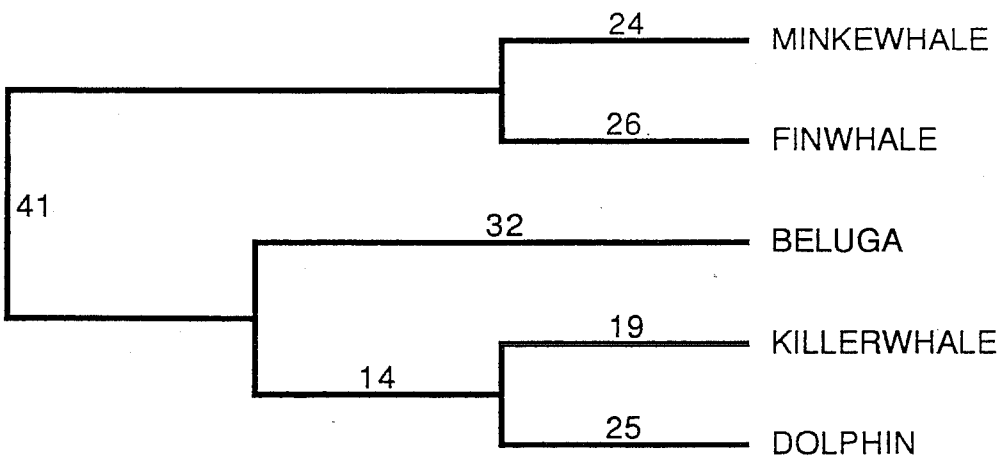

Figure 3. Heuristic phylogram (PAUP, Swofford, 1993) with branch length values showing the relationship between beluga, other odontocete and mystocete combined central region and right domain control region sequences. Bootstrap values (500 replications, 50% majority rule consensus) are 100(beluga to mystocete branch), and 98(beluga to odontocete branch).