MAJOR HISTOCOMPATIBILITY COMPLEX CLASS II SEQUENCE VARIATION IN CETACEANS: *DQβ* AND *DRβ* VARIATION IN BELUGA (*DELPHINAPTERUS LEUCAS*) AND *DQβ* VARIATION IN NORTH ATLANTIC RIGHT WHALES (*EUBALAENA GLACIALIS*)

By

BRENT WILLIAM MURRAY, M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (1997) **McMaster University**
(Biology) Hamilton, Ontario

TITLE: Major histocompatibility complex class II sequence variation in cetaceans: *DQβ* and *DRβ* variation in beluga (*Delphinapterus leucas*) and *DQβ* variation in North Atlantic right whales (*Eubalaena glacialis*).

AUTHOR: Brent William Murray, M.Sc. (University of Alberta)

SUPERVISOR: Professor B.N. White

NUMBER OF PAGES: xviii, 163

Abstract

Allelic variation at the Major Histocompatibility Complex (Mhc) class II loci *DQβ* and *DRβ* was assessed in a large sample (~300) of beluga (*Delphinapterus leucas*) and a sample (~12) of narwhal (*Monodon monoceros*) in order to study the evolutionary significance of the Mhc in cetaceans and to compare levels of variation among beluga populations. Mhc class II *DQβ* allelic variation was also analysed in a preliminary survey of right (*Eubalaena glacialis*) and bowhead (*Balaena mysticetus*) whales. In each survey variation was assessed by analysis of class II loci ß chain exon 2 nucleotide sequences. Exon 2 encodes for the peptide binding region of the class II molecules. These sequences were amplified via the polymerase chain reaction, followed by either cloning and DNA sequencing or single-stranded conformation polymorphism analysis.

A low amount of variation was observed at the beluga *DQβ* locus while a low to moderate amount of variation was observed at the beluga *DRβ1* locus and the right whale *DQβ* loci when compared with terrestrial mammals. Two *DRβ* loci in beluga and two *DQβ* loci in right and bowhead whales were detected. Comparison among beluga *DQβ*, among beluga *DRβ1*, and among right whale *DQβ* sequences show, at each locus, a high ratio of nonsynonymous to synonymous substitutions per site. Further, in each case, the majority of substitutions did not maintain the physio-chemical properties of the residue and were found at sites implicated as being important in the selective binding of foreign antigen. This evidence of positive selection is similar to that found at functional Mhc loci and is consistent with the functional significance of Mhc class II loci in the immune response of cetaceans. Comparison of beluga and right whale *DQβ* sequences shows a reduced rate of synonymous substitutions in cetaceans. This may explain the low to moderate levels of variation found in the order.

No significant differences in Mhc allelic or genotypic frequencies were observed among beluga summering populations which are believed to share a common overwintering area in either the Hudson Strait or the Bering Sea. Significant differences ($p \leq 0.0001$) were found among sampling locations of the High Arctic/Baffin Bay beluga population. This suggests a recurrent role of Arctic polynyas as overwintering locations for some groups of High Arctic beluga, i.e. beluga sampled from Cunningham Inlet. *DRβ1* allele and genotype frequencies were also significantly different ($p \leq 0.0001$) among the overwintering locations, Hudson Strait, Bering Sea, Baffin Bay, Cunningham Inlet, and the St. Lawrence. *DQβ* allele frequencies were only significantly different ($p \leq 0.005$) between the High Arctic/Baffin Bay beluga and all other sampling locations.

Comparison of *DQβ* and *DRβ1* allele and genotype frequencies within the St. Lawrence population between dead beached beluga and live biopsy darted whales shows no evidence of a sample collection bias, however, a larger sample size is needed to detect non-trivial small effects. No reduction in the number of *DQβ* alleles was found in the St. Lawrence beluga population. A slight reduction in the number of *DRβ1* alleles was observed (six vs. eight), however, these five alleles each represent one of the five genetically distinct allelic lineages found in beluga. This indicates that the evolutionary potential of this population at these class II molecules remains, despite a recent population bottleneck. The effects of the historic bottleneck are observed upon examination of *DRβ1-DQβ* haplotypes. Evidence for linkage disequilibrium between these loci is noted in the St. Lawrence. Comparison of possible haplotypes shows the St. Lawrence has about half the haplotypes found in Arctic populations. The short term effects of this reduction are unknown and further investigation of Mhc variation within this population is warranted.

Acknowledgements

It is a warm winters day. A blue sky, a sun with penetrating warmth, and a fresh, cold, crisp air. The perfect day for reflection. During the past four years I have had the privilege of being a biologist. To study nature, to study populations of whales, and to study a small but important molecule contained within them. Throughout the course of this investigation I have had much help and guidance from a number of people. I would like to include a few lines, here at the beginning, to thank those people.

Traditionally, a general acknowledgment begins with the thesis supervisor. Of all people, he or she has provided the most help and guidance throughout the project. Bradley White has indeed been a truly excellent supervisor. Thank you, Brad, for providing me with the means to conduct the research, the always useful criticism, ideas, and direction, and most importantly, for allowing me the freedom and time to make my own decisions. I also wish to thank my thesis committee, Brian Golding and Del Harnish, who never complained (well hardly) when I gave them progress reports at the last moment, and on whose help I could always rely. Although not officially a supervisor or committee member, Jim Quinn was a great help and always available for discussion and advice.

To study a wild population of beluga from within a lab can be bit depressing. Robert Michaud and Sylvain De Guise allowed me the opportunity to experience, first hand, the reality of the questions I was trying to address. You enriched my knowledge of belugas and provided me with experiences that were remembered in the long lab hours to follow.

When I first arrived in Hamilton, the former Graduate student working on the St. Lawrence beluga project was just finishing her thesis. Although in the process of writing

her thesis, Natialie Patenuade took the time to carefully explain the project, to introduce me to her collagues, and to be a friend. Thank you Nat, I shall not forget your kindness.

I have made a number of friends during my time at McMaster. I appreciated the 'Lads' and other lab mates just only for their relaxation abilities. Paul, for his inexhaustible desire to talk science. Ruth, for her inexhaustible desire to talk about everything. Chris, for his ability to relax in the face of pressure. Jeff, for his realism. Sam, for being someone else with which to talk about beluga. Carole, for her energy. And Sobia, for her smile. A list of friends can never be complete; Joyce, Monica, Diane, Diana, Tyler, Jennifer, YuiXiang, Sonya, Paula are just a few of the people I am privileged to call friends. I thank all of you for your their time, laughter, and camaraderie. Cheers! Um, Deutsches Bier ist sehr gut!

My family has always been a constant support for me. Mom and Dad, thank you for providing that support. Thanks for the loans, the new clothes, the calling card number, the flights home for Christmas, and most importantly your gift to me in the belief that I could do whatever I wanted in life if I had the desire.

Lastly, I would like to thank Jacqueline. Jacqui, you have become a tremendous support to me, a support upon which I know I can always rely, perhaps unfairly in these last few months. Thank you for everything; for all the laundry washed, all the meals cooked, all the proofreading (especially the proofreading), and for your unconditional love.

Brent W, Murray

8, February, 1997

Tübingen.

Thesis Format

This thesis is organized into seven chapters. The first and seventh chapters are a General Introduction and the Concluding Remarks respectively, providing a general overview of the thesis and a synthesis of the results. Chapters two through six have been written as manuscripts for submission to peer-reviewed scientific journals. At the time of writing only chapter two has been published. For the ease of the reader and to avoid redundancy, the Literature Cited for all Chapters has been amalgamated into one section at the end of the thesis. The Appendix contains genetic information on each beluga sample, including the results of this thesis (Chapters 2, 3, 4 and 5) and the mitochondrial DNA information of Mancuso (1995) and Brown (1996).

The following section provides information on each paper including title, authors, reference or status of publication, and the contribution of each author and other key personnel to each paper.

Chapter 2: "Sequence variation at the Major Histocompatibility Complex locus *DQß* in beluga whales (*Delphinapterus leucas*)."

Authors: Brent W. Murray, Sobia Malik, and Bradley N. White.
Reference: Mol. Biol. Evol. **12**:582-593. 1995
Contribution: Data on the beluga *DQß* sequence variation were generated exclusively by B.W.M.. The majority of the narwhal DQß sequence variation was generated by S.M. as part of a fourth year honours thesis. The research was conducted under the supervision and guidance of B.N.W..

Chapter 3: "Sequence variation at the Major Histocompatibility Complex *DRβ* loci in beluga whales (*Delphinapterus leucas*)."

Authors: Brent W. Murray and Bradley N. White.
Status: preprint
Contribution: Data on the beluga *DRβ* sequence variation were generated exclusively by B.W.M.. Narwhal *DRβ* sequence variation was intially generated by Elizabeth Saunders, as part of her fourth year honours thesis, and subsequently finished by B.W.M.. The research was conducted under the supervision and guidance of B.N.W..

Chapter 4: "Levels of Major Histocompatibility Complex variation in dead beached and live biopsied St. Lawrence beluga (*Delphinapterus leucas).*

Authors: Brent W. Murray, Robert Michaud, Pierre Béland, and Bradley N. White.
Status: preprint
Contribution: Data on the beluga *DQB* and *DRβ1* sequence variation were generated exclusively by B.W.M.. Photo-ID and biopsy samples were collected from live St. Lawrence beluga by R.M. The research was conducted under the supervision and guidance of P.B. and B.N.W..

Chapter 5: "Comparison of allelic and haplotype variation of Major Histocompatibility Complex class II *DRβ1* and *DQβ* loci in North American beluga (*Delphinapterus leucas*) populations."

Authors: Brent W. Murray and Bradley N. White.
Status: preprint
Contribution: Data on the beluga *DRβ* sequence variation were generated exclusively by B.W.M.. The research was conducted under the supervision and guidance of B.N.W..

Chapter 6: "Preliminary survey of North Atlantic right whale (*Eubalaena glacialis*) Major Histocompatibilty Complex, class II, *DQβ* allelic variation."

Authors: Brent W. Murray and Bradley N. White.

Status: preprint

Contribution: Data on the right and bowhead whale *DQβ* sequence variation were generated exclusively by B.W.M.. Information on right whale family groups was kindly provided by Moria W. Brown and the New England Aquarium. The research was conducted under the supervision and guidance of B.N.W..

Table of Contents

List of Tables

List of Figures

Chapter 1

General Introduction

Natural History

Beluga (*Delphinapterus leucas*, Pallas 1776) are known by many common names throughout their circumpolar distribution. The name, *Delphinapterus leucas*, means 'white dolphin without a fin' and most of the common names also use the obvious descriptor. In English, beluga are also commonly known as white whales, in French as *béluga* and historically in Quebec as *marsouins blancs* (whites porpoises), in Danish as hvidhval or hvidfisk (whitewhale or whitefish), and in Russian as *belukha*. Beluga itself is thought to be derived from the Russian word for white, *byely*. Names for beluga from the peoples of the Arctic regions include, *qilaluag* or *qaqortoq* (Greenlandic), *qilalugaq* (Inuktitut), *sisuaq* (Northern Alaskan Inupiat), *situag* (Bering Sea Inupiat), and *cetuag* (Alaska mainland Yupik).

Beluga are a member of the Family Monodontidae, which is comprised of two extant monotypic genera. The narwhal (*Monodon monoceros*, Linnaeus 1758), also an arctic adapted Odontocete with a circumpolar distribution, is the other extant species of this family. Kasuya (1973) and Barnes (1984) argue, based on ear bone morphology, for the inclusion of the Irrawaddy dolphin (*Orcaella brevirostris*, Gray 1886) within the family. However, allozyme data and quantitative immunological methods place the Irrawady dolphin with the Delphinidae (Lint et al. 1990). Late Miocene and late Pliocene monodontid fossils have been found at temperate locations (Barnes 1977) indicating the monodontids are unlikely to have been of arctic origin (Gaskin 1982). Gaskin (1982), following Davies (1963), suggested that the extant arctic adapted species resulted from an

entrapment in the Arctic basin during one of the cold periods of the Tertiary. Fordyce and Barnes (1994) noted that the disappearance of monodontids from mid to low-latitudes in the late Neogene corresponds with the rapid radiation of delphinids, and suggest the possibility of ecological displacement.

Beluga have many characteristics that make them adapted for an arctic environment. Their cervical vertebrae, unlike most cetaceans, are not fused allowing for greater lateral movement of the head. They lack a dorsal fin but have a narrow, ridged back. In addition, they possess well developed echolocation, as evidenced by their prominent, rounded melon.

Full grown beluga measure up to 3 - 5 m in length and can weigh between 500 - 1500 kg. Adult males are usually larger than females. While adult beluga are a characteristic white, juveniles are a slate gray. Juveniles lose this colouration as they age and are usually white upon sexual maturity. Sexual maturity is between 4 - 7 years for females and 6 - 9 years for males (Brodie 1971; Sergeant 1973; Burns and Seaman 1985; Heide-Jørgensen and Teilmann 1994). The life span of a beluga is between 25 to 30 years (Sergeant 1973; Brodie 1971 respectively), although ages as great as 38+ have been estimated in a few whales (Burns and Seaman 1985).

Beluga females are most reproductive between the ages of 6 and 22, after which the reproductive rate slows (Burns and Seaman 1985). Mating is thought to occur from late February to May at the wintering locations, or during the spring migration. Gestation is about 12-14 months and the peak calving times range from May to June. Estimates of times of mating, gestation, and calving vary among populations (Heide-Jørgensen and Teilmann 1994). Nursing occurs for 12-18 months and weaning occurs well after the first solid food is taken (Burns and Seaman 1985) . Most female beluga conceive in a three year cycle, however, about 25% are thought to conceive during lactation and therefore follow a two year cycle (Sergeant 1973). The general reproductive cycle appears to be similar throughout the circumpolar distribution (Hazard 1988).

The data regarding feeding and feeding habitats for beluga are very diverse with over a hundred different kinds of prey (Kleineberg et al. 1964). Their diet varies according to geography and prey abundance and includes benthic invertebrates, crustaceans, various cephalopods, and numerous fish (e.g. cod species, herring, capelin, salmon, charr, and eels).

Most beluga populations migrate in response to the shifting patterns of ice cover in the Arctic. In spring they advance with the ice break up into their summering areas, actively exploiting new leads as they occur (Reeves and Mitchell 1987a). In the Arctic summer they can be found in great concentration at specific river estuaries. Estuaries are believed to be important areas for feeding and possibly nursing (Sergeant 1973), but of primary importance for molting (St. Aubin et al. 1990). Site tenacity and philopatry have been observed (Caron and Smith 1990). With autumn and the advance of ice the beluga migrate back to their wintering areas. Overwintering occurs in deep water along the edge of the flow ice. Common overwintering areas are thought to be shared by beluga from different summering locations. Breeding is believed to occur at this time. The degree of interbreeding among summering groups at the overwintering locations is presently an area of investigation.

Population Structure

Worldwide, 16 beluga populations have been identified based on summering concentrations (Table 1.1) and the total number of beluga may be around 100,000. Seven summering populations are identified in Canada. Of these, the St. Lawrence, southeast Baffin Island, and Ungava Bay are considered endangered (COSEWIC, Campbell 1993) and/or of small population size (≤ 500) and vulnerable to hunting or habitat deterioration (IWC 1992). In addition, the eastern Hudson Bay/James Bay is considered threatened (COSEWIC, Campbell 1993) and/or of medium size (500 - 3000) and exploited at rates

Table 1.1 Circumpolar beluga populations classified into summering and wintering populations (Adapted from IWC 1992, references for population estimates therein). IWC status: *Large* (3000+), *Medium* (3000-500), and *Small* (500 and less) and *Light* (lightly or sustainably exlpoited), *Concern* (exploited at rates that give cause for concern), and *Vulnerable* (vulnerable to hunting or habitat deterioration). COSEWIC status from Campbell (1993). A question mark indicates status is not known.

Wintering population * Summering population	Pop. Estimate	IWC status	COSEWIC status
St. Lawrence			
* St. Lawrence R.	500	Small & Vulnerable	endangered
Hudson Strait			
* Southeast Baffin Is.	500	Small & Vulnerable	endangered
* Ungava Bay	low	Small & Vulnerable	endangered
* E. Hudson/James Bay	1,864-3,874	Medium & Concern	threatened
* W. & N. Hudson Bay	25,000	Large & Light	
Baffin Bay			
* High Arctic	6,300-18,600	Large & Concern	vulnerable
Bering Sea			
* E. Beaufort Sea	11,500	Large & Light	
* E. Chukchi Sea	2,500-3,000	?	na
* Norton Sound	2,000	?	na
* Bristol Bay	1,000-1,500	Medium & Light	na
Bering Sea ?			
* Anadyr Gulf	2,000-3,000	Medium & Light	na
* E. Siberian W. Chukchi Sea E. Siberian Sea	2,000-3,000	?	na
Cook Inlet			
* Cook Inlet	300-400	Small & Vulnerable	na
Sea of Okhotsk			
* Sea of Okhotsk	25,000-30,000	Large & Light	na
W. Siberian Barents Sea Kara Sea Laptev Sea	7,000-10,000	?	na
White Sea ?			
* White Sea	500-1000	?	na

that give cause for concern (IWC 1992), and the High Arctic/west Greenland is considered vulnerable and/or of large size (≥ 3000) and exploited at rates that give cause for concern (IWC 1992). Alternatively, these beluga can be classified based on the location of their overwintering locations (Table 1.1). In Canada, four wintering populations have been identified, two of which are shared with either the United States and Russia (Bering Sea) or Greenland (Baffin Bay).

The discreteness of the above populations is an area of current research. Most of these populations are hunted, on a subsistence level, by the aboriginal peoples of the Arctic. To fully understand the effects of the current take on local beluga populations, an understanding of the discreteness of their population structure is essential. Morphological data has been used with limited success to investigate population structure within the eastern North American Arctic. An analysis of age-to-length relationships, augmenting the work of Doidge (1990), indicates the High Arctic (Jones Bay) and southeast Baffin Island (Cumberland Sound) whales are significantly longer than those found in the east or west Hudson Bay (Stewart 1994). Significant differences were not found between the east and west Hudson Bay summering populations or the High Arctic and southeast Baffin Island summering populations (Stewart 1994). Investigations of parasites (Measures et al. 1995) and contaminants may also be useful for the determination of population structure, however, genetic data has proven to be the most informative (Brennin 1992; Mancuso 1995; Brown 1996).

Strong site fidelity has been noted in beluga whales (Caron and Smith 1990). Beluga in the Nastapoka estuary showed strong site tenacity by returning to the estuary 40 hr after incidents of hunting, and site philopatry by returning to the estuary in the following year (Caron and Smith 1990). Analysis of mitochondrial DNA (mtDNA) variation among summering populations shows evidence, in both males and females, of long term philopatry to summering estuaries (Brennin 1992; Mancuso 1995; Brown 1996). Genetic structure of mtDNA variation was first noted between the eastern and western Hudson Bay

summering populations (Brennin 1992). More detailed geographic analyses of mtDNA sequence variation in the Hudson Bay (Mancuso 1995) and throughout North America (Brown 1996) found evidence of nine distinct summering populations in North America; St. Lawrence, eastern Hudson Bay/James Bay, western/northern Hudson Bay, southern Hudson Strait/Ungava Bay, southeast Baffin Island, Baffin Bay, eastern Beaufort Sea, Eastern Chukchi Sea, and Norton Sound.

Although mtDNA variation clearly shows long term site fidelity to estuaries, analysis of microsatellite variation indicates that mating is occurring among beluga from different summering populations but which share a common wintering area (Brown 1996). In addition, Brown only finds a clear separation of the western (i.e. Bering Sea) and eastern (i.e. Baffin Bay, Hudson Strait, and St. Lawrence) North American populations. This interpretation is based on a neighbor joining dendrogram, constructed from average delta-μ distances (Goldstein 1995) of five microsatellite loci, that shows a western and an eastern North American cluster. Because the two sample locations within the Baffin Bay wintering population and the St. Lawrence samples are clustered together with the nine Hudson Strait sampling locations, Brown (1996) concludes a high amount of genetic exchange is occurring among these populations.

I disagree with the interpretation of this dendrogram as evidence for high levels of genetic exchange, and feel the evidence is inconclusive. Brown (1996) clearly states that, "Standard errors of the genetic distances among whales from different sampling sites indicate that some of the relationships constructed by the clustering program may not be significant.", however, the significance of the branches is not given, except for a single statement that the standard errors between the eastern and western North American sampling locations are relatively small. A visual inspection of the pairwise distances used to construct the dendrogram show that the smallest genetic distance between the St. Lawrence and the Hudson Strait sampling locations (Churchill) has a very high standard error (delta-μ 0.195 SE. 0.156). Genetic distance estimates of beluga populations have

three major biases due to sampling which will cause the distance to deviate from the theoretical expectations. These are: 1) nonrandom sample collection, i.e. nonrandom hunting practices, 2) small sample size, i.e. in most cases less than 50 alleles sampled, and 3) the small number of loci examined, i.e. five. The frequency of the alleles of each locus in each population will change due to mutation and genetic drift. For closely related populations, random drift will be the major force affecting allele frequency. In order to correctly observe drift, the effects of small sample sizes have to be taken into account. Although, the measure of standard error gives an indication of the existence of such biases, it is unclear in this case how it was used to assess the confidence of clusters.

Brown (1996) also conducted a pairwise comparison of the allele frequencies at each locus using exact tests. Exact tests are useful in the assessment of differences in allele frequencies from small sample sizes, as the probability of the observed data, given the null hypothesis of allele frequencies being independent of sample location, is assessed by a comparison to an estimation of the probabilities of other allele frequency distributions with the same row and column totals (see Materials and Methods Chapters 4 and 5). Analysis of the pairwise exact tests of allele frequencies at five loci, among sampling locations, conducted by Brown (1996) does not support the conclusions based on the dendrogram. Within the wintering populations, the mean and mode number of loci, which show significant differences in allele frequency ($p < 0.05$) at each pairwise comparison, are equal or very close to 1 (Bering Sea (10 comparisons) range 0-2, mean=0.9, mode=1 (n=7); Hudson Strait (36 comparisons) range 0-3, mean=1, mode=1 (n=18); Baffin Bay (one comparison) number of loci = 1). This value increases as comparisons between proposed wintering populations are made (e.g. Hudson Strait to Baffin Bay (18 comparisons), range 0-3, mean=1.6, mode=2 (n=7); Hudson Strait to St. Lawrence (nine comparisons), range 2-4, mean=3, mode=3 (n=5); Baffin Bay to St. Lawrence (two comparisons) mean=3, mode=3 (n=2)). The largest number of differences is observed between the western and eastern North American sampling locations ((60 comparisons) range 2-5, mean 3.7, mode

4 (n=25)). Even though over half the loci in the St. Lawrence population have significantly different allele frequencies than the Hudson Strait and Baffin Bay sampling locations, Brown (1996) concludes that there is no evidence for genetic structure. It seems more likely that, due to large standard errors in the estimation of genetic distance, the dendrogram is only detecting the most genetically distant populations, with the relationships among the genetically more similar population being unresolved.

The genetic distinction between the Hudson Bay, the eastern Beaufort Sea, and the St. Lawrence whales is supported by analysis of minisatellite variation (Mancuso 1995). Average levels of minisatellite band sharing within and between populations can be used to test the null hypothesis (H_0) of 'no reduction in band sharing because of the sharing of common bands due to random mating between the populations'. Average band sharing coefficients within the St. Lawrence were significantly greater ($p < 0.01$) than those observed between either the St. Lawrence and Hudson Bay or the St. Lawrence and eastern Beaufort Sea. In addition, average band sharing coefficients within either the Hudson Bay or the eastern Beaufort Sea beluga were significantly greater ($p < 0.05$) than those between the Hudson Bay and eastern Beaufort samples. In each case, the H_0 can be rejected, indicating genetic structure among the populations.

In summary, the genetic data, to date, indicate that beluga are highly philopatric but that mating does appear to be occurring among summering populations that share a common wintering area. The relationships among the wintering locations is less clear. The western North American Arctic populations are genetically distinct from those in the eastern North American Arctic. The St. Lawrence population appears to be genetically distinct, based on analysis of minisatellite variation and with, on average, over half the microsatellite loci examined being significantly different in allele frequency from the other eastern North American wintering populations. The genetic relationship between the Baffin Bay and Hudson Strait populations is unclear.

The St. Lawrence beluga

The St. Lawrence beluga have been the most thoroughly studied of any beluga population. It is currently considered an endangered population (COSEWIC, Campbell 1993), and is believed to be geographically isolated from the closest Arctic populations residing in either Hudson Bay and Hudson Strait or the high Arctic and Baffin Bay. No beluga populations are found along the coast of Labrador and only occasional migrants have been noted. The occurrence of beluga along the central Labrador coast is poorly known, but beluga have been recorded on the Atlantic coast of Newfoundland (Sergeant et al. 1970; Béland et al. 1992). Southward, beluga are occasionally spotted along the continental coast from the Bay of Fundy to New Jersey (Reeves and Katona 1980; Michaud et al. 1990). The beluga sightings south of the Gulf of St. Lawrence are believed to be St. Lawrence animals that have followed the Gaspé and Labrador currents, while the Newfoundland sightings may be Arctic whales that have followed southward the Labrador current (Sergeant et al. 1970). Based on fishermen's claims of a dramatic increase in the number of beluga in the St. Lawrence at the end of the 1920's, Vladykov (1944, 1946) speculated that a large number of whales, presumably from the west coast of Greenland, migrated into and then left the St. Lawrence in the late 1920's to mid 1930's. Upon a reanalysis of the arguments of Vladykov, however, Reeves and Mitchell (1984) find this scenario unlikely.

The vast majority of the St. Lawrence beluga are found year round in the waters of the St. Lawrence estuary (Michaud 1993). In summer, the St. Lawrence population is centred around the Saguenay river and makes use of the St. Lawrence estuary from Battures-aux-Loups-Marins to Riviére-Porneuf and Iles du Bic. Habitat use is correlated with social groups, as mothers with calves tend to use the upper areas of the river, while groups of adults (presumably male) are more common in the downstream waters (Michaud 1993).

Beluga fossils found in the St. Lawrence area have been radiocarbon dated to approximately 10,000 years ago, a date which roughly coincides with the formation of the proglacial Champlain Sea (Harington 1989). The association of beluga bones at Iroquoian sites indicate that beluga were used as a food source by some native groups (Roland Tremblay, pers. comm.). Beluga were hunted, primarily for their oil and hides, from the early 1800's until the 1960's, at which point harvesting was no longer economically feasible (Reeves and Mitchell 1984). The Canadian government issued a bounty of 15 dollars on beluga, which were blamed for declining fish stocks, between the dates 1932-35 and 1937-38. In that period 2,233 bounties were paid (Reeves and Mitchell 1984). Through a careful examination of harvest records and a backcalculation of the population size needed to sustain recorded take, it has been estimated that there were at least 5000 St. Lawrence beluga prior to 1885 (Reeves and Mitchell 1984).

In 1979, the St. Lawrence beluga were officially protected by an amendment of the Canada Fisheries Act. Despite a protected status and a lack of significant exploitation for over 25 years, numerous population surveys from 1973 on have indicated that the population is either stable or slowly increasing in number (e.g. Pippard 1985; Sergeant 1986; Sergeant and Hoek 1988; Kingsley 1993; Michaud 1993). A number of hypotheses have been put forward to explain this lack of significant population recovery including harassment, habitat degradation, contamination by toxic chemicals, and the deleterious effects of inbreeding.

A long term investigation of the St. Lawrence beluga began in 1982 and has grown to include studies of distribution, habitat, photo-identification, behaviour, population dynamics, genetic variability, toxicology, and pathology. The results of the first nine years are reviewed by Béland et al. (1993). A large amount of information has come from the analysis of beluga found dead, floating or beached, along the shores of the St. Lawrence, when compared to other St. Lawrence marine mammals and arctic beluga. Population modelling (Béland et al. 1988), supported by field studies (Michaud 1993), indicate that the

observed population stability may be due to low calf production and/or low survivorship to adulthood. Toxicological studies of the dead beached whales show high levels of mercury, lead (Wagemann et al. 1990), PCB's, DDT, Mirex (Martineau et al. 1987; Muir et al. 1990), and benzo[a]pyrene metabolites (Martineau et al. 1988). Necropsies on over 45 dead beached whales reveal large numbers of tumors and non-neoplastic lesions (Martineau et al. 1988, 1995; De Guise et al. 1994a, 1994b, and 1995a). Thirty neoplasms have been described from 20 of the 47 St. Lawrence beluga carcasses examined, representing 39% of the 77 tumors reported in cetaceans worldwide (Martineau et al. 1995). No correlation of tumors was observed with regard to sex or age distribution. Although papilloma virus-like particles have been linked with eight cases of gastric papillomas within the St. Lawrence (De Guise et al. 1995b), a common viral etiology does not seem likely due to the diversity of the tumors described (Béland et al. 1995).

The correlation with high loads of toxic chemicals, and the diversity of the neoplasms and other lesions have led Martineau et al. (1987) and others (e.g. De Guise et al. 1994a) to suggest that the contaminants are either acting directly as carcinogens, or having an adverse effect on the immune system. Ongoing studies are attempting to quantify the immune functions of the St. Lawrence beluga by comparison to arctic whales in order to test these hypotheses (De Guise et al. 1995b). Alternatively, genetic factors within the St. Lawrence may be leading to a predisposition to the observed pathologies. Studies on the DNA minisatellite variation of the beached whales found a reduced amount of variation in the St. Lawrence beluga when compared to either the Beaufort Sea (Patenaude et al. 1994) or the Hudson Bay populations (Mancuso 1995). Studies on captive populations of 38 mammal species have shown that, on average, the rate of mortality is 33% higher in the progeny of first degree relatives when compared to the progeny of unrelated pairs (Ralls et al. 1988). Thus it may be a combination of environmental and genetic factors that are responsible for the lack of population recovery.

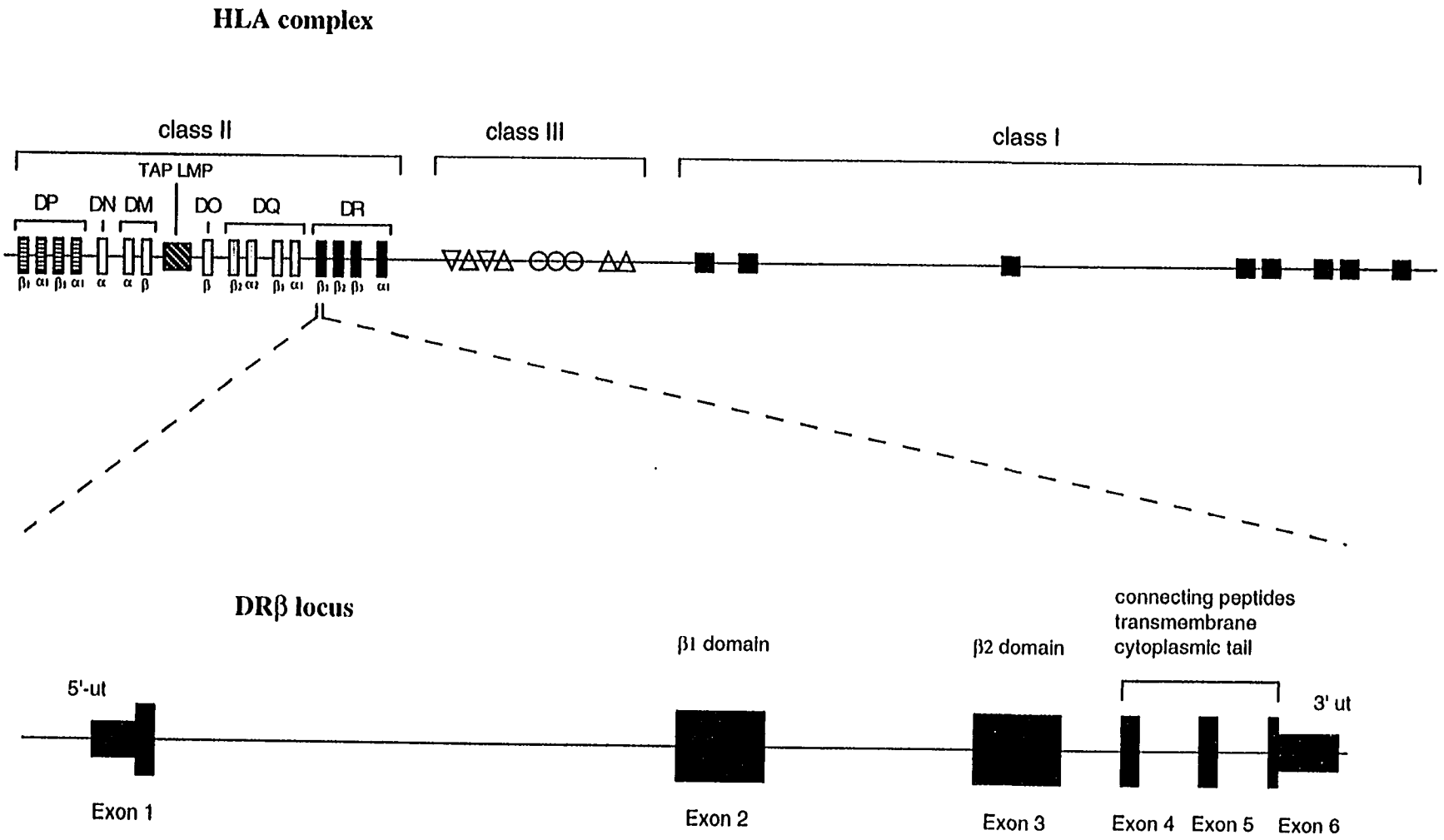
A major objective of the following studies is to compare levels of variation among beluga populations at loci directly involved with the immune response.

The Major Histocompatibility Complex

The Major Histocompatibility Complex (Mhc) is a large, physically linked complex of loci of about four million base pairs, in humans (Trowsdale 1995). The majority of loci are directly or indirectly involved with antigen presentation to T-cells which subsequently causes the initiation of an immune response. Mhc loci have been divided into three classes based on their function (Fig. 1.1): class I, presentation of endogenous antigens (e.g. viruses); class II, presentation of exogenous antigens (e.g. extracellular bacteria); and class III, non-classical function, i.e. many not directly involved in the immune system (Klein 1986). Classical Mhc genes, i.e. those involved in antigen presentation, seem to be found only in vertebrates. All mammals, birds, amphibians and most fish examined are found to contain Mhc genes (Trowsdale 1995). Despite attempts, Mhc genes have yet to be identified in jawless fish (Powis and Geraghty 1995). No conclusive evidence of antigen presenting Mhc loci has been found in invertebrates (Klein 1986, Humphreys and Reinherz 1994). Comparative studies among vertebrate Mhc loci indicate that the physical linkage of class I, II and III loci found in mammals may be over 300-500 Million years (Myr) old (Trowsdale 1995).

The classical antigen presenting class I and II molecules are both cell surface glycoproteins. The close physical linkage and similarity in form and function have led some to speculate that these two types of molecules are the result of a duplication of a class II like loci early in the evolutionary history of vertebrates (Hughes and Nei 1993; Klein and O'hUigin 1993). Both molecules are heterodimers with a cytoplasmic tail, a transmembrane domain, a conserved Membrane-Proximal Domain (MPD), and a

Figure 1.1 Organization of the human Mhc (HLA) and of a class II *DRβ* locus (adapted from Trowsdale 1995). Genes or exons are identified by boxes, triangles, or circles. Names of class II loci only are shown.


HLA complex
class II
class III
class I
TAP LMP
DP
DN
DM
DO
DQ
DR
DRβ locus
β1 domain
β2 domain
connecting peptides
transmembrane
cytoplasmic tail
5'-ut
3' ut
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Exon 6

distal polymorphic Peptide Binding Region (PBR) (Klein 1986). Class II molecules are composed of two equal peptide chains, α and ß, which combine to form equal portions of the MPD and PBR. In contrast, class I molecules are composed of a large α chain, which forms the entire PBR, and a smaller ß2-microglobulin (encoded outside the Mhc) which combines with the α chain to form the conserved MPD.

Both molecules bind foriegn and self peptides and are subsequently expessed on cell surfaces. Class I molecules are found on all nucleated cell surfaces, and most commonly act to signal intracellular infection by the presentation of intracellular peptides to $CD8^+$ T-cells (cytotoxic T-cells). A class I molecule, foreign peptide, and $CD8^+$ T-cell complex, in conjunction with appropraite secondary signals, leads to the release of cytotoxic chemicals from the T-cell which kill the infected cell presenting the foreign peptide (Germain 1994; York and Rock 1996). Class II molecules are constitutively expressed on antigen presenting cells such as macrophages, which engulf, endocytose, and degrade foreign extracellular particles. These foreign peptides are displayed on the cell surface by class II molecules where they are recognized by $CD4^+$ T-cells (helper T-cells). This complex leads to the release of lymphokines which modulate the immune response by the activation of B-cells to release antibodies against the foreign pathogen, or by the direction of cytotoxic T-cells to attack infected cells (Germain 1994; Germain et al 1996).

Both class I and II molecules are translated and assembled in the endoplasmic reticulum (ER), however, the location of peptide loading differs (reviewed by Benham et al. 1995). Peptides degraded in the cytosol and transported into the lumen of the ER by TAP molecules (Transporters associated with Anitigen Processing) are loaded onto class I molecules at this point. Due to the nature of the class I PBR, a short grove enclosed at either end, only peptides between the sizes of 8-11 amino acids are bound by class I molecules. Class II molecules are blocked from the binding of peptides in the ER by a close association with the invariant chain (Ii). The Ii either directly interferes with peptide binding or keeps the molecule in a partially unfolded state. In either case, class II

molecules associated with Ii are needed for efficient release from the ER, and transportation to the endosomes. Once in the endosomes, the Ii chain is degraded and peptides from the endosomal/lysosomal pathway are loaded through an association with another class II Mhc molecule, *DM* (Kelly et al. 1991; Morris et al. 1994; Sanderson et al. 1994). Class II molecules have an open PBR and can bind peptides between 12-25 amino acids in length. Once bound with peptide, Mhc molecules are transported to the cell surface. Antigen processing, peptide loading, and presentation to T-cell takes about one hour, and the half life of a cell surface Mhc molecule is about 30 hours (Lanzavecchia and Watts 1992).

In humans, about six class I and six class II molecules are expressed on the cell surface (Rammensee et al. 1995). It has been estimated that each class I molecule can present $\geq$ 10,000 peptides (Engelhard 1994). However, each molecule, and the alleles of each molecule, have been shown to bind different peptide motifs (e.g. Falk et al. 1994; Rammensee et al. 1995; Friede et al. 1996). In this way, a large number of peptides are presented to T-cells.

Mhc loci are among the most polymorphic functional loci within the animal kingdom. The role of balancing selection, in the form of overdominance, in the maintenance of this diversity was first proposed by Doherty and Zinkernagel (1975). Overdominance, driven by the need to respond to a wide range of parasites, would give heterozygotes a higher fitness than homozygotes. Alternatively, a second form of balancing selection, frequency dependent selection, may also be responsible for the maintenance of the high levels of allelic variation. However, current evidence does not allow for the exclusion of either hypothesis (Takahata and Nei 1990). Recently, with the study of disease association to HLA alleles, a third form of balancing selection based on fluctuating selection has been implicated (Hill et al. 1994).

Analysis of DNA sequence variation, amino acid variation, allele and genotype frequency, and the persistence of allelic lineages provide strong evidence for the key role of balancing selection in the maintenance of Mhc variation. The type of selection (i.e. positive

[directional], negative [balancing], or neutral) that a Mhc locus has encountered may be deduced by analyzing directly the PBR. Comparison of the number of nonsynonymous to synonymous substitutions at the PBR of both class I and class II loci shows a significantly greater number of nonsynonymous substitutions (Hughes and Nei 1988, 1989; Hughes et al 1994). As neutral evolution would predict an equal number of each type of substitution, this provides evidence of positive directional selection at the PBR. Nonsynonymous substitutions at the PBR appear to have a selective advantage, presumably due to changes in PBR specificity. Further evidence for this type of directional selection was noted in the type of nonsynonymous substitutions at the class I PBR. As predicted by the theory, the proportion of nonconservative changes was greater than the conservative changes (Hughes et al. 1990). Analysis of allele and genotype frequencies also show deviation for neutral theory expectations. Mhc loci have a lower level of homozygosity for a given number of alleles than would be predicted by neutral theory (i.e. Ewens-Watterson test) (Klitz et al. 1986; Klein et al. 1993a). Further, heterosis has been observed in human (Degos et al. 1974; Black and Salzano 1981) and natural mice populations (Ritte et al. 1991). Finally, the existence of trans-species allelic lineages first proposed by Arden and Klein (1982) on serological evidence, and more recently supported by sequence analysis (e.g. Lawlor et al. 1988; Mayer et al. 1992; Figueroa et al. 1994), also indicates the existence of balancing selection. Allelic lineages are groups of alleles which can be shared among species and are more similar to each other than to the alleles found within a single species. Within primates, allelic lineages 23 to 40 Myr old have been identified (Figueroa et al. 1994). The maintenance of allelic lineages over such large time scales would be extremely unlikely given neutral evolution, i.e. random fixation of alleles (Klein et al. 1993a).

Other theories have been proposed to explain the high level of Mhc variation, one of which is a high mutation rate. Klein et al. (1993a) used the variation found within primate allelic lineages to measure nucleotide substitution rates. They found that the synonymous substitution rate was similar to that found at other, nonpolymorphic primate genes.

Further, they found the majority of the Mhc molecule, i.e. PMD, was under a moderate negative or purifying selection. An increased mutation rate for the PBR is also not likely. Eventhough the PBR shows an elevated rate of nonsynonymous substitution, the synonymous substitution rate is equal throughout the molecule (Klein et al. 1993a; Hughes and Hughes 1995).

Another intriguing possibility is the role of mating preference. Potts et al. (1991) found that mice in a seminatural population did not choose their mates at random but were choosing mates, based on odour, that had different Mhc genotypes than themselves. Although Mhc disassortative mating preferences are consistent with (i.e. would reinforce) a pathogen driven balancing selection, Potts et al. (1994) find further evidence to suggest that the mating preference may be more important in the avoidance of inbreeding. Some (Hughes and Hughes 1995), however, find the evidence inconclusive and feel that this mechanism is unlikely to be a general phenomenon due to the decreased importance, or total lack of smell, in some taxa, e.g. birds and fish. In addition, it does not explain the strong evidence for positive selection observed at the PBR (Hughes and Hughes 1995).

Mhc, species conservation, and marine mammals

It has been suggested that Mhc allelic variability should be taken into account in the management of captive breeding programs (Hughes 1991a) and that lack of variation at this locus in an isolated population may increase the chance of extinction of that population (Yukhi and O'Brien 1990). The association of disease resistance to malaria and HLA alleles in African human populations indicates a functional significance of Mhc variation (Hill et al. 1991; Hill et al. 1994). A low degree of polymorphism at this complex, as in the case of the African cheetah, has been implicated as the cause of the increase in the populations' susceptibility to pathogens (O'Brien et al. 1985). The existence of trans-

species allelic lineages and the slow rate of Mhc evolution (Klein et al. 1993a) suggest that once lost, the variation at this complex will only very slowly be replaced.

Most mammalian species examined, notably primates, ungulates, and rodents, have been found to have high levels of Mhc variation. Comparative studies, however, show that levels of polymorphism are variable between similar loci in different species. Most primates (human, Marsh and Bodmer 1993; non-human, O'hUigin et al. 1993) and ungulates (cattle (*Bos taurus*), Ammer et al. 1992; Sigurdardottir et al. 1991; Mikko and Andersson 1995a, red deer (*Cervus elphus*), Swarbrick et al. 1995, sheep (*Ovis spp)* , Schwaiger et al. 1994, goat (*Capra aegagrus*), Schwaiger et al. 1993) have high levels of *DRβ* variation. In rodents, the *DR* region seems to be less important with some mice strains lacking this locus (Figueroa et al. 1990). The *DR* region has been completely lost in the mole rat (*Spalax ehrenbergi*) and its function replaced by a duplication of the *DP* locus (Nizetic et al. 1987). Variability in class II polymorphism also exists among primates (reviewed by Bergström and Gyllensten 1995). These, and similar observations, have been explained by the shifting accordion hypothesis (Klein et al. 1993b). This hypothesis asserts that variation is reduced in small founding populations through random drift, and is generated by selective pressure and the duplication of Mhc loci.

Because of the variability of locus polymorphism among species, it is important to characterize and quantify the variation found at a number of Mhc loci before the overall level of Mhc variation can be assessed. For example, in the cotton-top tamarin (*Saguinus oedipus*) low levels of variation were found at some class I and class II loci, while *DRβ* loci contained a high amount of variation (Gyllensten et al. 1994). In addition, it is important to consider the evolutionary history of the group in question. Comparisons should be made in light of the variation found in closely related species and the overall genetic variation within the species. In this way the 'normal' amount and type of variation within a taxa, i.e. closely related group of organisms, can be assessed, and the effects of any previous bottlenecks identified.

Examination of Mhc variation within marine mammals has been limited to only a few evolutionarily separated taxa. Trowsdale et al. (1989) examined a small sample of two baleen whale species; i.e. nine fin whale, *Balaenoptera physalus*, and five sei whale, *B. borealis*, samples. They examined the restriction fragment length polymorphism identified by various human Mhc probes. They found no class II *DQα*, *DRα*, or *DNα* variation and only limited amounts of *DQβ*, *DRβ* and class I variation. The low amount of α chain variation is not surprising given that most human α chains, except *DQα*, are not polymorphic (Marsh and Bodmer 1993). However, the amount of variation at the class I and class II ß chain loci was lower than expected based on human or mice samples and indicates a reduced amount of Mhc variation in these baleen whales. Slade (1992) also found limited amounts of class I and class II (*DQα*, *DQβ*, and *DRβ*) variation in the southern elephant seal (*Mirounga leonina*), and speculated that the general reduction in Mhc variation found in these three marine mammal species may be characteristic of all marine mammals, possibly due to a lower amount of micro-parasites encountered in the marine environment which has led to a reduction in the balancing selection pressure.

Other work on marine mammal Mhc is preliminary or incomplete. Slade et al. (1994) reported *DQα* sequences from a number of pinniped species. This study used the sequences for inferring species phylogenies and no survey of population variation was conducted. Preliminary analysis of the Hawaiian monk seal (*Monachus schauinslandi*) reveals no *DQα* variation and a low amount of class I variation in accordance with the very low levels of overall genetic variation (Armstrong 1995). In contrast, a preliminary survey of class I polymorphism in the European harbour seal (*Phoca vitulina*) found levels of variation comparable to terrestrial mammals. The high amount of variation and the positive selection observed at the PBR suggest no reduction in the balancing selection pressure in this species (Goodman and Slade 1995).

Objectives

This thesis will investigate the hypothesis that reduced levels of genetic variation may be a factor leading to the observed immunosuppression of the St. Lawrence beluga. Previous studies have noted a reduction in the minisatellite diversity in the St. Lawrence beluga (Patenaude et al. 1994; Mancuso 1995). This study expands on the sample size used in those previous studies and focuses on loci whose products are involved directly in the immune response. Mhc class II loci were chosen for this analysis because of their role in the upregulation of the immune response by the presentation of foreign peptides to T-cells. A decrease in the allelic diversity of these loci may decrease the number of peptide motifs being efficiently presented to T-cells. Speculatively, the inability of individual to efficiently respond to wide range of pathogens, due to a decreased level of Mhc heterozygosity, may lead to relative decrease in the immunocompetency of the St. Lawrence beluga.

Before levels of variation can be compared however, more information on the nature of Mhc variation in cetacea is needed. Most of the previous studies on Mhc variation in marine mammals have not addressed the functional significance of the molecules (although see Goodman and Slade 1995). By examining directly the nucleotide sequences responsible for encoding the PBR the type of selection these loci have encountered can be addressed and the functional significance of the loci indirectly implied. Analogy to the Mhc variation of terrestrial mammals would predict this region to be under strong positive selection (Klein et al. 1993a; Hughes and Hughes 1995). A similar finding in cetaceans would indirectly imply these loci are acting as peptide presentation molecules. In contrast, most proteins are subject to negative selection (selection against change of the amino acid sequence) (Li et al. 1985) and nonfunctional loci, such as pseudogenes, evolve neutrally.

Mhc variation will be estimated by DNA sequence analysis of class II ß chain loci. A large majority of the polymorphism of class II molecules can be attributed to ß chain

amino acids involved in the PBR (Marsh and Bodmer 1993). Variation of *DQ* and *DR* molecules will be assessed by analysis of the nucleotide sequences responsible for encoding the ß chain PBR, located in the second exon of a typical ß chain locus (Fig 1.1).

The two major goals of this study are: 1) To assess levels of Mhc variation in the small and endangered St. Lawrence beluga population, and to compare that level of variation to beluga populations from throughout North America; 2) To assess the role of selection in the evolution of cetacean Mhc loci by comparing the levels and type of variation found in a large survey of beluga, and in smaller surveys of narwhal, right and bowhead (*Balaena mysticetus*) whales.

The arrangement of the following chapters is as follows: Chapter two analyzes the *DQß* variation found in beluga and narwhal and presents the results of a survey of allelic variation within beluga; Chapter three describes the *DRß* variation found in beluga and narwhal and assesses the evolutionary implications of this variation; Chapters four and five compare levels of *DQß* and *DRß1* allelic variation between sample types collected in the St. Lawrence and among beluga populations respectively; and, Chapter six describes the results of a preliminary survey of *DQß* variation in right and bowhead whales.

Chapter 2

Sequence Variation at the Major Histocompatibility Complex locus *DQβ* in Beluga Whales (*Delphinapterus leucas*).

Abstract

Genetic variation at the Major Histocompatibility Complex locus *DQβ* was analyzed in 233 beluga whales (*Delphinapterus leucas*) from seven populations: St. Lawrence Estuary, eastern Beaufort Sea, eastern Chukchi Sea, western Hudson Bay, eastern Hudson Bay, southeastern Baffin Island and High Arctic, and in 12 narwhals (*Monodon monoceros*) sympatric with the High Arctic beluga population. Variation was assessed by amplification of the exon coding for the peptide binding region via the polymerase chain reaction, followed by either cloning and DNA sequencing or single stranded conformation polymorphism analysis. Five alleles were found in the beluga populations and one in the narwhal. Pairwise comparisons of these alleles showed a 5:1 ratio of nonsynonymous to synonymous substitutions per site leading to eight amino acid differences, five of which were nonconservative substitutions, centered around positions previously shown to be important for peptide binding. Although the amount of allelic variation is low when compared with terrestrial mammals, the nature of the substitutions in the peptide binding sites is consistent with an important role for the *DQβ* locus in the cellular immune response of beluga whales. Comparisons of allele frequencies among populations show the High Arctic population to be different ($p \leq 0.005$) from the other beluga populations surveyed. In these other populations an allele, *Dele-DQβ**0101-2, was found in 98% of the animals while in the High Arctic it was only found in 52% of the animals. Two other alleles were found at high frequencies in the High Arctic population, one being very similar to the single allele found in narwhal.

Introduction

Cell surface glycoproteins encoded by the Major Histocompatibility Complex (Mhc) play a key role in the initiation of an immune response by binding foreign peptides and presenting them to T-cells. The high levels of Mhc class I and class II genetic variation found in most mammals at this gene complex have been proposed to be an adaptation resulting from the large number of pathogens encountered by natural populations (Klein and Takahata 1990). A low degree of polymorphism at this complex, as in the case of the African cheetah, has been suggested to be the cause of the increase in the population's susceptibility to pathogens (O'Brien et al. 1985). It has been suggested that Mhc allelic variability should be taken into account in management of captive breeding programs (Hughes 1991a) and that lack of variation at this locus in an isolated population may increase the chance of extinction of that population (Yuhki and O'Brien 1990).

The Peptide Binding Region (PBR) shows a large amount of the functional allelic variation expressed in the beta chains of most terrestrial mammal Mhc class II cell surface glycoproteins (e.g. Hughes and Nei 1989). This region is postulated to be directly involved with the interaction and association of foreign peptides which are subsequently presented to T-cells. The type of selection (i.e. positive [directional], negative [balancing], or neutral) that a Mhc locus has encountered may be deduced by analyzing directly the PBR.

Previous studies on Mhc variation in marine mammals suggested less polymorphism than in mice or humans. Trowsdale (1989) found limited restriction fragment length polymorphism (RFLP) variation at a number of Mhc loci in a sample of nine fin (*Balenoptera physalus*) and five sei (*B. borealis*) whales. A study on RFLP variation in the southern elephant seal (*Mirounga leonina*) gave a similar result (Slade 1992). The small amount of Mhc polymorphism found in these different groups suggested

that this may be a common feature of all marine mammals and be due to a decreased exposure to parasite diversity in marine as compared to terrestrial mammals (Slade 1992).

In Canadian waters three beluga (*Delphinapterus leucas*) populations are considered endangered, southeastern Baffin Island, Ungava Bay, and the St. Lawrence Estuary, one threatened, the eastern Hudson Bay population, and one vulnerable, the High Arctic population (COSEWIC, Campbell 1993). Detailed information on all aspects of the biology of these populations is needed for future management decisions. The southernmost and geographically isolated population in the St. Lawrence Estuary has been the most thoroughly studied population. Although these resident whales are currently protected, their population, which numbered over 5000 animals at the turn of the century (Béland et al. 1988), has failed to increase in number. Several surveys since 1973 have indicated the population (around 500) is either stable or slowly increasing in number (Pippard 1985; Sergeant 1986; Kingsley pers. comm.; Michaud 1993). In an attempt to understand the lack of population growth, studies have been conducted on beluga carcasses found beached along the St. Lawrence river.

Toxicological studies indicate high levels of contamination by PCB, DDT and mirex while necropsies have revealed a large number of unusual tumours and pathological conditions (Béland et al. 1993). These findings suggest that the contaminants are having an adverse effect on the immune system. At the same time, genetic analysis of minisatellite loci have found the mean allele frequency in the St. Lawrence animals (.31) is significantly higher than in the Beaufort sea population (.21) indicating a reduction in genetic variation (Patenaude et al. 1994). Thus it may be a combination of environmental and genetic factors that are responsible for the lack of population recovery. By examining the amount of allelic variation present in the Mhc of the St. Lawrence stock the question of inbreeding depression and its possible effect on immune response can be addressed.

The objective of this study is to characterize and quantify the amount of genetic variation at a locus involved in the immune response, Mhc class II *DQβ*, in beluga whale

populations in order to address two questions. First, through direct nucleotide sequencing of a functionally important region the evolutionary significance of this locus in beluga will be assessed. Second, allelic variation of the Mhc class II locus, *DQβ*, will be compared among beluga whale populations: St. Lawrence Estuary (endangered), eastern Beaufort Sea, eastern Chukchi Sea, western Hudson Bay, eastern Hudson Bay (threatened), southeastern Baffin Island (endangered), and High Arctic (vulnerable), and a narwhal (*Monodon monoceros*) population sympatric with the High Arctic beluga.

Materials and Methods

Samples

Tissue samples were collected over a number of years from beluga populations and a narwhal population (Table 2.1). These served as the base of comparison for Mhc allelic variation. The Hudson Strait samples were collected in October and may be part of either the eastern or western Hudson Bay populations. Most of the samples were collected as a result of aboriginal subsistence hunting and are considered a random sample of the population. Some of the Hudson Bay and High Arctic samples were collected from whales live-captured for other scientific purposes. The St. Lawrence samples are taken from dead beached whales.

Sequencing of Mhc alleles

Seven beluga were chosen randomly from the samples listed in Table 2.1 for the sequencing analysis; one each from the St. Lawrence Estuary, the eastern Beaufort Sea, and the High Arctic; and two each from the eastern and western Hudson Bay. Four Narwhal samples were also analyzed; two from each High Arctic location sampled. In addition, two beluga samples, one each from the eastern Hudson Bay and eastern Beaufort

Table 2.1 List of samples analyzed including species, population, sample location, year of sampling, number of individuals sampled and origin of sample.

Species	Population -- Sample location	Year	Sample size	Origin of Sample[1]
Beluga				
	St. Lawrence Estuary	88 - 91, 95	21	B
	Eastern Beaufort Sea			
	-- Mackenzie Delta[2]	84, 87 - 90	30	H
	Eastern Chukchi Sea			
	-- Point Lay	88 - 90	19	H
	Western Hudson Bay			
	-- Churchill	89 - 90, 93	26	LC
		91	3	H
	-- Arviat	86, 87	20	H
	Eastern Hudson Bay			
	-- Nastapoka River	87	34	H
		93	2	LC
	-- Little Whale River	92, 93	3	H
		92, 93	6	LC
	Hudson Strait			
	-- Wakeham Bay	83	14	H
	Southeastern Baffin Island			
	-- Cumberland Sound	86	12	H
	High Arctic			
	-- Grise Fiord	84, 85, 87	20	H
	-- Creswell Bay	93	5	LC
	-- Cunningham Inlet	88	14	H
		90	4	LC
Narwhal				
	High Arctic			
	-- Arctic Bay	85, 87	10	H
	-- Grise Fiord	87	2	H

[1] Under origin of sample a 'B' indicates a dead beached whale, a 'LC' indicates a whale live-captured for other scientific purposes (i.e. radio telemetry) and an 'H' indicates an aboriginal subsistence hunter killed whale.

[2] Mackenzie Delta area includes 6 locations; East Whitefish Station, Hendrickson Island, Tuktoyuktuk, Single Point, Kendall Island, and West Whitefish Station.

Sea, were chosen for sequencing analysis based on their containing a unique Single Stranded Conformation Polymorphism (SSCP) pattern (see below).

DNA sequencing of the PBR of this Mhc gene was carried out by enzymatic amplification of the region in question followed by cloning and sequencing of the resulting products. The *DQβ* PBR was amplified using the primers (CTG GTA GTT GTG TCT GCA CAC) and (CAT GTG CTA CTT CAC CAA CGG) (Tsuji et al. 1992). The reaction conditions for the enzymatic reaction were 10 mM Tris-HCl (pH-8.3), 50 mM KCl, 2.0 mM $MgCl_2$, 0.2 mM dNTP's, 0.2 µM of each primer, 1 unit of Taq DNA polymerase (Perkin-Elmer-Cetus), and 50-100 ng of template DNA carried out in a 50 µl volume. Thermal cycling was conducted on a Perkin-Elmer-Cetus model 480 and involved 30 cycles of 94 ^{o}C for 1 min., 55 ^{o}C for 1 min., and 72 ^{o}C for 2 min.. The products of this reaction were isolated from a 1.5% agarose gel using the GlassMAX (BRL) system, cloned into the pGEM^{T} vector system (Promega), and transformed into DH5 alpha (BRL) competent *Escherichia coli* cells.

Four to five clones were sequenced for most animals. By sequencing 5 clones (assuming that every allele was amplified at an equal frequency) there was low probability ($p = .0625$) of not detecting both alleles if the animal was a heterozygote. At least 2 identical clones per unique PBR, usually from separate amplifications (except *Dele-DQβ**0102), were obtained to confirm the nucleotide sequence. In this way DNA sequence differences arising during the Polymerase Chain Reaction (PCR) could be detected. Nomenclature of the alleles is based on the proposed rules for nomenclature of the Mhc of different species (Klein et al. 1990) and are based on sequence similarity. Standard nomenclature is a four letter species code (i.e. *Dele* and *Momo* for beluga and narwhal respectively), a locus code, an *, and a four digit allele code, i.e. *Mhc Dele-DQβ**0101. The first two numbers represent the allelic lineage while the last two designate the unique sequence.

Analysis of allelic variation

The relationship of alleles was estimated using the computer package Phylip 3.5c (Felsenstein 1993). Pairwise genetic distance measures were generated using the program DNAdist. This program estimated distance based on Kimura's two-parameter model (Kimura 1980). A distance tree was constructed from the distance matrix using the Neighbor-Joining program which is based on the neighbor-joining method of Saitou and Nei (1987). A right whale (*Eubalaena glacialis, Eugl-DQβ**0301) *DQβ* sequence (Chapter 6) was used for reference.

Pairwise comparisons of nucleotide substitutions between alleles were conducted on 171 bp of sequence (57 complete codons) according to the method of Nei and Gojobori (1986). Mean values of nonsynonymous (d_n) and synonymous (d_s) substitutions per site were compared with a student T-test.

Single Stranded Conformation Polymorphism analysis

SSCP analysis was conducted via PCR amplification (concentrations as above except reactions were carried out in a 10 µl volume) of the PBR for all the samples listed in Table 2.1 (See Hayashi 1992 for review). To ease analysis only one primer was radioactively end labelled with γ^{33}P-dATP (ICN). PCR products were electrophoresed through a nondenaturing acrylamide gel (5% acrylamide (59 acrylamide: 1 bisacrylamide), 10% glycerol and 1/2 TBE) for 13 hours at room temperature. Cloned *DQβ* sequences were used as conformation polymorphism standards from which DNA samples of the whales were typed. Any unique conformation polymorphism encountered was subsequently cloned and sequenced as described above. Pairwise comparisons of beluga population allele frequencies were compared with a maximum-likelihood-ratio chi-square test of independence (Kennedy 1992). The null hypothesis was that the observed allele frequencies were independent of the population sampled.

Results

DQβ sequence analysis

DQβ alleles from nine beluga samples, representing a wide geographic range of populations, and 4 narwhal samples were sequenced. In the nine beluga samples, five alleles were found (Fig. 2.1) while in the narwhal samples a single allele was found. Through the use of SSCP analysis to type allelic variation in a larger number of samples from each population, the prevalence of these alleles has been confirmed (see below). These alleles have been classified into two lineages, 01 and 02, based on nucleotide similarity (Fig. 2.2). The alleles within both lineages vary by 1-3 nucleotides. Comparison of the *Dele-DQβ**0101 and 0201 sequences to all sequences in genbank (release date, Oct. 20/94) reveals sequence similarity to *DQβ* alleles from other species. The most similar sequences (90%) were with sheep and swine *DQβ* genes.

All beluga and narwhal alleles sequenced contain a continuous open reading frame for the 172 bp examined. In total, three first, five second and three third within codon position substitutions were observed in nine codons (Fig. 2.1). Of the 11 nucleotide substitutions 10 were nonsynonymous and one was synonymous. These differences led to 8 amino acid substitutions (Fig. 2.3). Pairwise comparisons of nucleotide substitutions per site between alleles are shown in Table 2.2. In the majority of comparisons (13 of 15) d_n is greater than d_S. The average d_n, 4.12 (± SEM 0.5), is significantly greater than the average d_S, 0.8 (± SEM 0.5) ($\alpha = .0002$). From these means a 5:1 ratio of nonsynonymous to synonymous substitutions per site is observed.

A close examination of the amino acid sequences shows the substitutions tend to be clustered around sites that have been postulated to be responsible for selective peptide recognition (indicated by an asterisk in Fig. 2.3). Further, 5 of the 8 amino acid differences are nonconservative (i.e. do not conserve the physio-chemical properties of the residue) which may indicate that they would be responsible for causing a shift in selective

Figure 2.1 *DQβ* nucleotide sequences. A '-' is used in codons containing a variable position to indicate nucleotides identical to *Dele-DQβ**0101. In nonvariable codons no character is used to display identity to ease comparison. Exon 2 amino acid positions are given for reference. Codons marked by underlining or with an asterisks indicate the encoded amino acids either face the peptide binding groove or may be part of a peptide binding pocket respectively in the human *DR* molecule (Brown et al. 1993; Stern et al. 1994). Nucleotide sequences have been submitted to genbank and can be located under the accession numbers U16986-U16991 respectively.

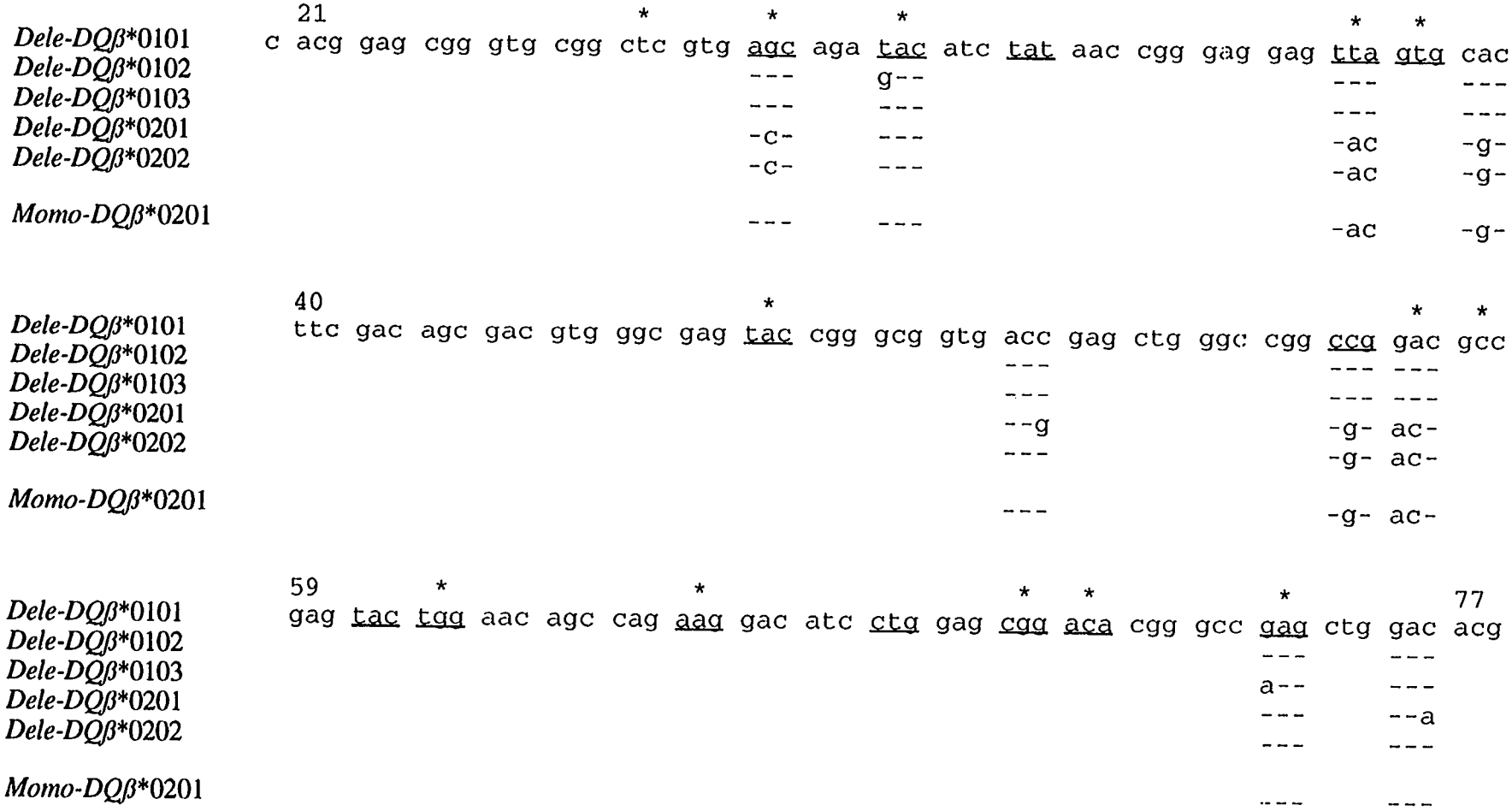

21 * * * * *
Dele-DQβ*0101 c acg gag cgg gtg cgg ctc gtg agc aga tac atc tat aac cgg gag gag tta gtg cac
Dele-DQβ*0102 --- g-- --- --- ---
Dele-DQβ*0103 --- --- --- --- ---
Dele-DQβ*0201 -c- --- -ac -g-
Dele-DQβ*0202 -c- --- -ac -g-
Momo-DQβ*0201 --- --- -ac -g-
40 * * *
Dele-DQβ*0101 ttc gac agc gac gtg ggc gag tac cgg gcg gtg acc gag ctg ggc cgg ccg gac gcc
Dele-DQβ*0102 --- --- ---
Dele-DQβ*0103 --- --- ---
Dele-DQβ*0201 --g -g- ac-
Dele-DQβ*0202 --- -g- ac-
Momo-DQβ*0201 --- -g- ac-
59 * * * * * 77
Dele-DQβ*0101 gag tac tgg aac agc cag aag gac atc ctg gag cgg aca cgg gcc gag ctg gac acg
Dele-DQβ*0102 --- --- ---
Dele-DQβ*0103 a-- --- ---
Dele-DQβ*0201 --- --a ---
Dele-DQβ*0202 --- --- ---
Momo-DQβ*0201 --- --- ---

Figure 2.2 Neighbor-joining distance tree of beluga and narwhal *DQß* alleles. The tree was constructed using the computer programs DNAdist and Neighbor (in the computer package Phylip 3.5c, Felsenstein 1993) which use the neighbor-joining method of Saitou and Nei (1987) to construct a tree from a pairwise distance matrix estimated by Kimura's two-parameter model (Kimura 1980). A right whale (*Eugl-DQß**0301) *DQß* sequence (Chapter 6) was used for refernce. The scale represents genetic distance estimated by the total length of horizontal bars between any two taxa.

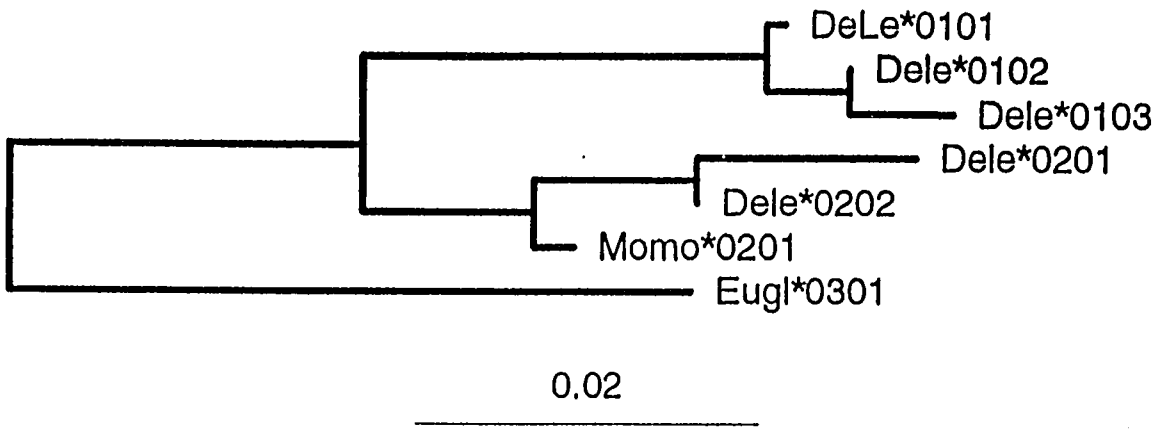

DeLe*0101
Dele*0102
Dele*0103
Dele*0201
Dele*0202
Momo*0201
Eugl*0301
0.02

Figure 2.3 *DQβ* amino acid sequences based on the nucleotide sequences (Fig. 2.1). The number above the consensus sequence represents the amino acid position based on human *DQβ* sequences. As with Fig. 2.1, identity to the consensus is only indicated by a '-' in variable positions to ease comparison. Codons marked by underlining or with an asterisks indicate the encoded amino acids either face the peptide binding groove or may be part of a peptide binding pocket respectively in the human *DR* molecule (Brown et al. 1993; Stern et al. 1994). Amino acid substitutions resulting in strong, moderate, and weak conservation of physio-chemical properties (polar or nonpolar, size, shape and charge) are indicated by '*s*, *m*, and *w*' respectively (McLachlan 1972 *in* Taylor 1986).

	21		30				40			50			60			70		
		*	*	*			**		*			**	*		*	**	*	
	T	E	R	V	R	L	V	S	R	Y	I	Y	N	R	E	E	Y	V	R	F	D	S	D	V	G	E	Y	R	A	V	T	E	L	G	R	P	T	A	E	Y	W	N	S	Q	K	D	I	L	E	R	T	R	A	E	L	D	T
*Dele-DQβ**0101				-	-					L	H																									-D																				-	-			
*Dele-DQβ**0102				-	D					L	H																									-D																				-	-			
*Dele-DQβ**0103				-	-					L	H																									-D																				K	-			
*Dele-DQβ**0201				T	-					-	-																									R-																				-	E			
*Dele-DQβ**0202				T	-					-	-																									R-																				-	-			
*Momo-DQβ**0201				-	-					-	-																									R-																				-	-			
Physio-chemical Conservation				*s*						*w*																																																										*m*			

Table 2.2 Pairwise comparisons of nonsynonymous and synonymous substitutions.

Allele	*Dele-DQß* *0101	*Dele-DQß* *0102	*Dele-DQß* *0103	*Dele-DQß* *0201	*Dele-DQß* *0202	*Momo-DQß* *0201
*Dele-DQß**0101		0.77 : 0	0.77 : 0	6.45 : 2.45	5.61 : 0	4.77 : 0
*Dele-DQß**0102	1 / 0		1.55 : 0	7.29 : 2.45	6.45 : 0	5.60 : 0
*Dele-DQß**0103	1 / 0	2 / 0		7.29 : 2.45	6.45 : 0	5.60 : 0
*Dele-DQß**0201	8 / 1	9 / 1	9 / 1		0.78 : 2.47	1.57 : 2.47
*Dele-DQß**0202	7 / 0	8 / 0	8 / 0	1 / 1		0.78 : 0
*Momo-DQß**0201	6 / 0	7 / 0	7 / 0	2 / 1	1 / 0	

Ratio of nucleotide substitutions per site (d_n) to synonymous substitution per site (d_s), $d_n : d_s$, expressed as percentages above the diagonal. Observed numbers of nonsynonymous (n_d) and synonymous (s_d) substitutions, n_d / s_d, below the diagonal.

peptide recognition in the PBR (Fig. 2.3). Of the four substitutions found between, but shared within, the allelic lineages 01 and 02 (including narwhal), three are nonconservative and one is weakly conserved. For example, at position ß57 tyrosine is replaced by aspartic acid in the 01 lineage resulting in the substitution of an aromatic amino acid for a negatively charged amino acid. Similarly, the *Dele-DQß*0103* allele differs from the *Dele-DQß**0101 allele by a single nucleotide change which results in the substitution, at a position important for peptide binding, of a negative (glutamic acid) for a positive (lysine) amino acid (position ß74, Fig. 2.3). Four of the five nonconservative replacements involve a change in charge, an observation also noted for mammalian class I genes (Hughes et al. 1990)

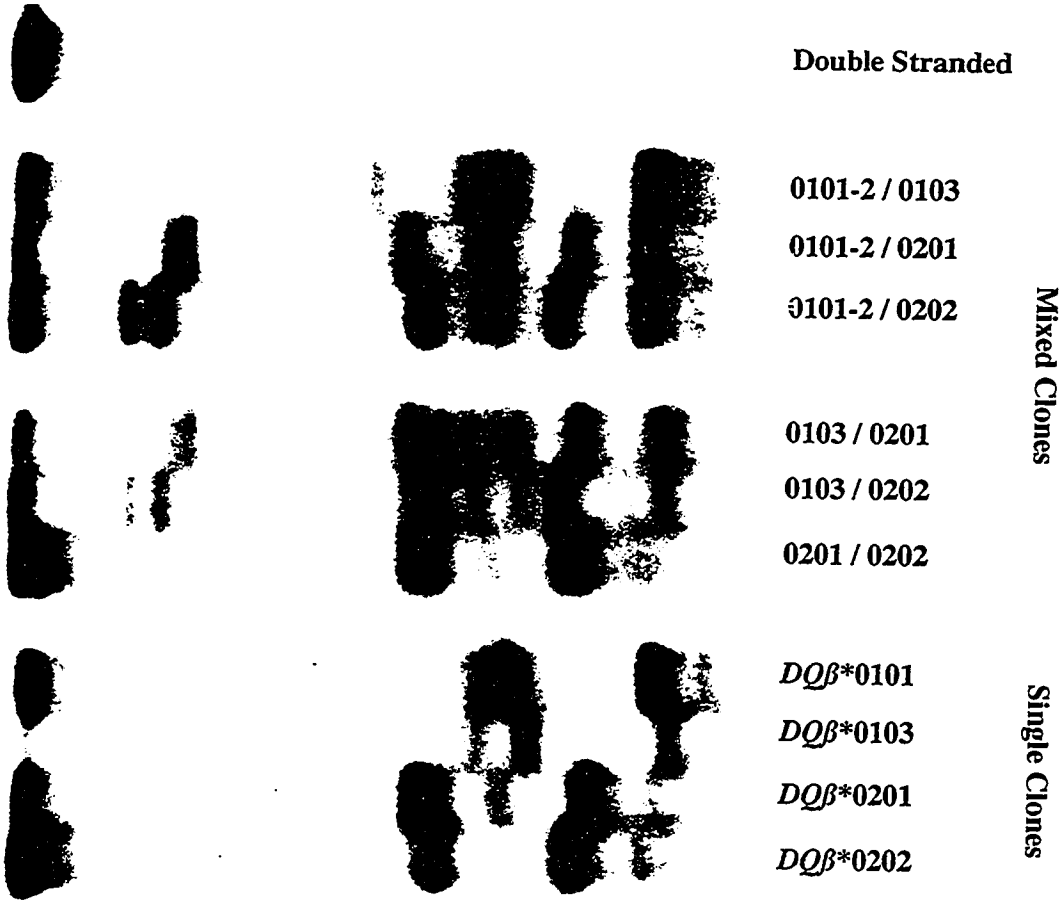
SSCP survey

SSCP analysis was used as a rapid and efficient method (detecting > 90% of single base pair differences in 200 bp fragments (Hayashi 1992)) to detect the presence of known and new alleles in 233 beluga and 12 narwhal samples. Allele *Dele-DQß**0103 was initially detected by the SSCP analysis and was subsequently sequenced. Genotype frequency was estimated for each population (Table 2.3a) through interpretation of the SSCP pattern for each of the animals listed in Table 2.1. Alleles were identified in each sample by comparison with the SSCP phenotypes derived from cloned alleles (Fig. 2.4). The phenotype of a single allele is composed of three to four bands. In each case the lowest band is double stranded DNA (far left sample in Fig. 2.4), i.e. DNA that has re-annealed to its complementary strand after denaturation, while upper bands are stable single stranded conformations. Multiple single stranded conformational variants have been noted previously (Michaud et al. 1992). Phenotypes of heterozygotes were reconstructed by the mixing of amplified cloned *DQß* alleles. In most cases additional bands are present intermediate to the double stranded DNA and the single stranded conformations found in the separate *DQß* clones. They may be explained by the formation of heteroduplex DNA. The ^{33}P-dATP labelled DNA strand not only re-anneals with its complementary strand to

Table 2.3 **a:** *DQß* genotype frequency in beluga populations. **b:** *DQß* allele frequency in beluga populations.

	Population							
	St. Lawrence Estuary	E. Beaufort Sea	E. Chukchi Sea	E. Hudson Bay	W. Hudson Bay	Hudson Strait	S.E. Baffin Island	High Arctic
a: *DQß* * Genotype								
0101-2 / 0101-2	12	21	13	32	38	8	12	3
0101-2 / 0103	1		5	3	1	2		11
0101-2 / 0201	4	5		4	8	2		
0101-2 / 0202	4	4	1	3	2	2		10
0103 / 0103								5
0103 / 0201				1				
0103 / 0202								7
0201 / 0202				1				
0202 / 0202				1				5
Σ	21	30	19	45	49	14	12	43
b: *DQß* * Allele								
*Dele-DQß**0101-2	33	51	32	74	87	22	24	27
*Dele-DQß**0103	1	4	5	4	1	2		32
*Dele-DQß**0201	4	5		6	8	2		
*Dele-DQß**0202	4		1	6	2	2		27
Σ	42	60	38	90	98	28	24	86

Figure 2.4 PCR-SSCP phenotypes of cloned *DQβ* alleles. The double stranded band is the PCR product of a *DQβ**0101 clone not denatured prior to electrophoresis. The samples to the right prefixed with *DQβ* are the SSCP phenotypes of clones containing a sequenced *DQβ* allele. The six middle samples are derived from the mixing of the PCR products of the previous four samples prior to denaturation. These represent the reconstructed phenotypes for each of the heterozygote genotypes.


Double Stranded
0101-2 / 0103
0101-2 / 0201
0101-2 / 0202
0103 / 0201
0103 / 0202
0201 / 0202
DQβ*0101
DQβ*0103
DQβ*0201
DQβ*0202
Mixed Clones
Single Clones

form native double-stranded DNA but anneals with the complementary strand from the other allele to form a heteroduplex product with an altered electrophoretic mobility. Note, heteroduplex mobility is most altered when clones of the 01 and 02 lineages are combined (Fig. 2.4).

A SSCP analysis of animals from two High Arctic locations shows a wide range of genotypes (Fig. 2.5). Genotype was determined by comparison of SSCP patterns with those of the cloned *DQβ* alleles and the reconstructed phenotypes (Fig. 2.4). Of the 11 animals analyzed four are homozygotes, three for the *DQβ**0103 and one for the *DQβ**0101-2, and seven are heterozygotes, three for both *DQβ**0103 / 0202 and *DQβ**0101-2 / 0202 and one for *DQβ**0101-2 / 0103.

All of the bands amplified from genomic DNA were found either in the cloned alleles or reconstructed genotypes (Fig. 2.4). No more than two alleles were detected in a sample by either sequencing of *DQβ* alleles or SSCP analysis. This shows that only one locus was amplified. Although *DQβ* alleles may exist that are not amplified by the primers used in this survey, their presence seems unlikely since every sample examined yielded at least one product. Further, pooled genotype frequencies (excluding the High Arctic samples) were not significantly different from the Hardy-Weinberg expectations, i.e. an excess of heterozygotes was not observed. No conformational difference was detected between the *Dele-DQβ**0101 and *Dele-DQβ**0102 alleles under various gel conditions ranging from 4 to 8 % acrylamide with 0 to 10 % glycerol. In the analysis the *Dele-DQβ**0101 and *Dele-DQβ**0102 alleles have been grouped together (i.e. *Dele-DQβ**0101-2). Extrapolation from the sequencing results indicates that the majority of the alleles in this group would be *Dele-DQβ**0101 (11/12).

Genotype frequency data (Table 2.3a) was converted to allele frequencies (Table 2.3b) and compared among populations using the maximum-likelihood-ratio chi-square tests of independence (Table 2.4). The High Arctic population was found to be different from all populations ($p = 0.001$, except Hudson Strait, $p = 0.005$). The *Dele-DQβ**0101-2

Figure 2.5 PCR-SSCP analysis of *DQβ* genotypes from High Arctic beluga. Five animals from Creswell Bay, 1993, and six from Grise Fiord, 1984, are shown. Genotypes, listed above each sample, are inferred by comparison to clones containing a sequenced *DQβ* allele (to the right of samples and prefixed by *DQβ*).

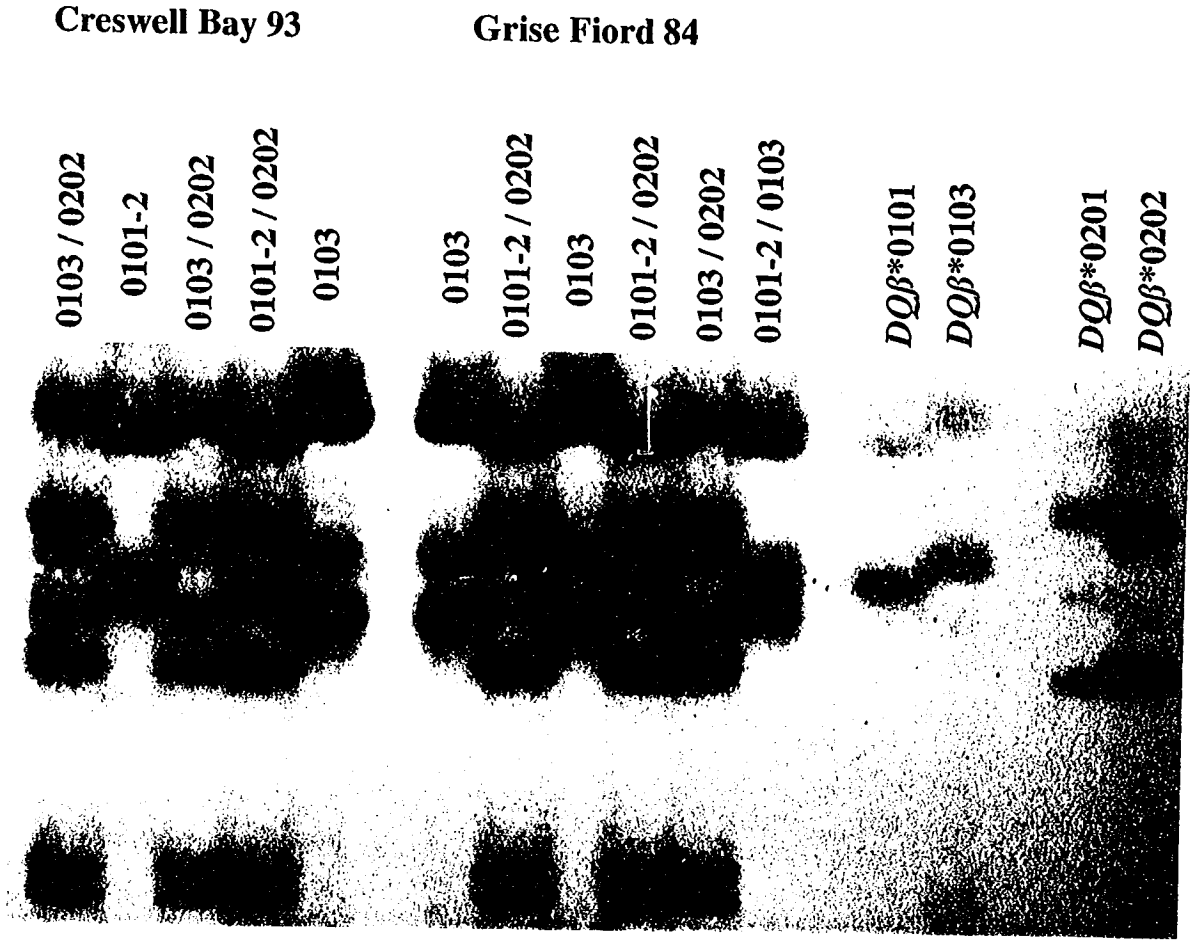

Creswell Bay 93
Grise Fiord 84
0103 / 0202
0101-2
0103 / 0202
0101-2 / 0202
0103
0103
0101-2 / 0202
0103
0101-2 / 0202
0103 / 0202
0101-2 / 0103
DQß*0101
DQß*0103
DQß*0201
DQß*0202

Table 2.4 Results of maximum-likelihood-ratio chi-square tests of independence beluga population and allele frequency (Table 2.3b).

Population	E.Beaufort Sea	E.Chukchi Sea	E.Hudson Bay	W.Hudson Bay	Hudson Strait	S.E.Baffin Island	High Arctic
St.Lawrence Estuary	1.061	4.429	0.331	2.010	0.497	3.926	** 22.723
E. Beaufort Sea		7.599	2.127	1.518	2.359	3.239 [1]	** 36.639
E. Chukchi Sea			3.875	6.675	2.391	3.138 [1]	** 16.762 [1]
E. Hudson Bay				2.509	0.166	4.008	** 35.105
W. Hudson Bay					2.237	2.540	** 52.984
Hudson Strait						4.049	* 14.958
S.E. Baffin Island							** 22.444 [1]

A double ** indicates $L^2 > \chi^2$ (3) = 16.27 or χ^2 (2) = 13.82 (p = .001). A single * indicates $L^2 > \chi^2$ (3) = 12.84 (p = .005).

[1] Maximum-likelihood-ratio chi-square test with 2 degrees of freedom. All others performed with 3 degrees of freedom.

allele is the most common, being found at a frequency of about 0.85 in most populations, and is present in 98% of those animals, except the High Arctic and southeastern Baffin Island (Table 2.3a). The High Arctic population has a *Dele-DQβ**0101-2 frequency of 0.31 with the allele present in only 52% of the samples. Conversely, the southeastern Baffin Island samples contain only the *Dele-DQβ**0101-2 allele. Due to the small sample size for this population the presence of the alleles *Dele-DQβ**0103, 0201 and 0202 can not be dismissed.

Twelve narwhal samples, including the four used in the DNA sequence analysis, were examined for SSCP's. In every case a uniform pattern was observed indicating that the samples are homozygous for the *Momo-DQβ**0201 allele. Assuming the 12 samples are random, the population is at Hardy-Weinberg equilibrium, and that every allele amplifies at an equal frequency, a simple examination of probability indicates a minimum allele frequency for *Momo-DQβ**0201 of ≥ 0.883 ($p = .95$), i.e. if x = allele frequency then $p > 0.05$ for x^{24} when $x \geq 0.883$.

Discussion

The small amount of *DQβ* allelic variation in beluga whales is similar with that found in other marine mammals (Trowsdale et al. 1990; Slade 1992). In a comparison of alleles of the mammalian *DQβ1* locus identifiable at exon 2 sequences, five *DQβ* alleles in beluga compare with 17 in *Homo sapien* (Marsh and Bodmer 1991), 34 in the *Mus musculus* complex (She et al. 1991), and nine diverse alleles in *Bos taurus* (Sigurdardóttir et al. 1992). In addition *DQβ1* exon 2 sequences are known from a number of nonhuman primates (Bontrop 1994). Comparison with this compilation also indicates a lower amount of allelic variation in belugas. Although some nonhuman primates also have a small number of known *DQβ1* alleles, these comparisons should be interpreted with caution as only a small number of samples for each species were examined in most cases (e.g.

Gyllensten et al. 1990; Otting et al. 1992). It is also important to judge the amount of Mhc diversity in contrast to the amount of overall genetic diversity. Analysis of minisatellite loci in eastern Beaufort Sea beluga did not indicate a substantial reduction of genetic variation (Patenaude et al. 1994). Similar low levels of Mhc variation are also seen in southern elephant seals, sei whales, and fin whales all of whom have high to moderate levels of allozyme variation when compared with the mammalian average (Slade 1992).

Slade (1992) put forward four hypotheses to explain the reduction of Mhc diversity in marine mammals: i) it is a methodological artifact; ii) it is due to random drift acting on small populations; iii) Mhc loci are non-functional in marine mammals; and iv) it is due to a reduced balancing selection pressure in the marine environment. Slade found the fourth hypothesis to be the simplest explanation, arguing the reduction of balancing selection pressure was due to the relatively lower abundance of parasites encountered in the marine as compared to the terrestrial environment. Variation at the beluga *DQβ* locus is consistent with this interpretation.

The investigation of additional marine mammals, the increase in sample size, and the use of different techniques decreases the possibility of the results being due to methodological artifacts. The consistency of low *DQβ* diversity in odontocetes (this study), mysticetes (Trowsdale et al. 1990) and pinnipeds (Slade 1992) argues against a bias in the species sampled. Further, the larger sample size examined here, including a number of populations, reduces the possibility of a stochastic within species sampling artifact. Finally, the use of more precise techniques allows the functional variation to be characterized directly.

Random drift operates most strongly on small populations. Although a number of the beluga populations in this survey have gone through a recent reduction in size, comparisons of allele frequency show most populations to be similar. The large eastern Beaufort Sea and the endangered St. Lawrence beluga populations have similar *DQβ* allele frequencies and yet DNA fingerprinting data suggests the St. Lawrence population has

overall a lower amount of genetic diversity (Patenaude et al. 1994). The eastern Beaufort Sea population appears stable (Norton-Fraker and Fraker 1982 *in* Finley et al. 1987), is estimated to be over 11,500 individuals (Finley et al. 1987), and no documented evidence exists of a substantial reduction of population size in historical time. Thus, if a genetic bottleneck reduced the Mhc variation in belugas it must have taken place before the establishment of the present species.

Observations on the type and position of nucleotide substitutions support positive Darwinian selection as the evolutionary force shaping the variation of the cetacean *DQβ* alleles and indicate the functional importance of the locus. Analysis of the type of nucleotide substitutions shows a mean d_n to mean d_s ratio of 5:1 (Table 2.2). A high ratio of d_n to d_s has been used as evidence of positive selection in a number of studies, for example: class I Mhc (Hughes and Nei 1988; Hughes et al. 1990; Imanishi and Gojobori 1992); class II Mhc (Hughes and Nei 1989; Schwaiger et al. 1994); the circumsporozoite antigen in *Plasmodium* (Hughes 1991b); human influenza A virus (Fitch et al. 1991); and abalone sperm lysins (Lee and Vacquier 1992). In the majority of proteins the number of $d_s > d_n$ indicates a constraint on the change of amino acid sequence, i.e. negative selection (Li et al. 1985). In contrast pseudogenes evolve with no constraints, i.e. neutral evolution. Variation at a Mhc class I pseudogene was similar to other mammalian pseudogenes and not to the patterns observed in functional class I sequences (Imanishi and Gojobori 1992). Further analysis of the type and position of nucleotide differences leading to amino acid substitutions show the differences are nonconservative and clustered around polymorphic positions responsible for selective peptide binding (Fig. 2.3). The occurrence of nonconservative amino acid substitutions in sites responsible for the selective binding of foreign peptide may lead to the recognition and presentation of different peptide motifs.

Although *DQβ* variation is low, evidence for the importance of this diversity is seen in comparison among beluga populations. The most striking difference occurs between the High Arctic and all other beluga populations (Table 2.4). The main ecological

difference between these populations is that the southern populations, with the exception of southeastern Baffin Island, spend part of the year in the relatively warmer waters of continental estuaries. This ecological difference may correspond with differences in pathogens encountered. Evidence for pathogen driven differences in allele frequencies is observed in the high frequency of the *Dele-DQß**0202 allele in the High Arctic population. *Dele-DQß**0202 is the beluga allele most similar to the narwhal *Momo-DQß**0201 allele, differing only by a single amino acid substitution. The fact that the sympatric High Arctic beluga and High Arctic narwhal populations share a similar allelic type may be evidence of a need for an immune response to a similar pathogen.

The large differences in allele frequencies between the High Arctic and the other beluga populations is also evidence for the distinctness of this population. The High Arctic population is believed to over-winter in Davis Strait along with the whales from west Greenland (Reeves and Mitchell 1987a; Doidge and Finley 1994), while the Hudson Bay, Ungava Bay and southeastern Baffin Island populations are thought to over-winter in the Hudson Strait (Finley et al. 1982; Reeves and Mitchell 1987b). The difference in Mhc diversity argues against a substantial gene flow between these populations. In support, Sergeant and Brodie (1969) found differences in size between the Hudson Bay (small), High Arctic and St. Lawrence (medium) and west Greenland beluga whales (Large). A reexamination of size relationships (Doidge 1990) found differences in age-length and length-weight comparisons among the Hudson Bay animals and other whale populations, but not to the extent reported by Sergeant and Brodie (1969). No difference was observed between the west Greenland and St. Lawrence populations. Examination of Mhc *DQß* variation in west Greenland animals may show if these animals are part of the same stock as the High Arctic belugas. This information would be useful in the management of the High Arctic belugas due to the large catch of belugas off west Greenland (Reeves and Mitchell 1987a, Doidge and Finley 1994).

All of the southern beluga populations have similar allele frequencies. The endangered St. Lawrence Estuary and southeastern Baffin Island, and the threatened eastern Hudson Bay populations are not significantly different from most of the other populations, except the High Arctic. This indicates the *DQβ* variation has been maintained in the endangered and threatened populations, however, analyses of other Mhc loci is required before variation at the complex can be quantified. Although not significantly different in allele frequency, the endangered southeastern Baffin Island population may be of concern. All individuals examined possess only the *Dele-DQβ**0101-2 allelic type. At this time the sample size is too small to draw conclusions, however loss of the rarer alleles in this population would lower the populations' ability to respond to other pathogens encountered by belugas. Further analysis of this population is required.

Conclusions

Beluga *DQβ* allelic variation is consistent with the low amount of *DQβ* variation observed in other marine mammals. This may be due to a reduction in pathogens encountered in the marine environment, however, evidence for positive Darwinian selection indicates that the ability to respond to these pathogens has been an important evolutionary force shaping the variation.

Comparison of beluga population allele frequencies find the High Arctic population to be different from the other populations examined. This difference may be due to a difference in the pathogens encountered by the High Arctic population and indicates that this population should be treated as a discrete management stock.

Acknowledgments

Samples used in this study were obtained through the co-operation of the Canada Department of Fisheries and Oceans, the Fisheries Joint Management Committee (Inuvik), the Alaska Department of Fish and Game (North Slope Borough, Department of Wildlife Management), the Makivik Corporation, the native communities listed in Table 2.1, and the St. Lawrence National Institute of Ecotoxicology. Financial and graduate student support were provided by grants from NSERC, NSERC-DFO Science Subvention Program, the Endangered Species Recovery Fund (co-sponsored by the Canadian Wildlife Service of Environment Canada and World Wildlife Fund Canada), and a Northern Studies Training Program award. We would also like to thank N. J. Patenaude, S. Mancuso, D. P. Singal, J. S. Quinn, J. J. Dockray, J. E. Minor, P. D. Robinson, P. Wilson, and two anonymous reviewers for their useful suggestions and help with research, data analysis, and manuscript preparation.

Chapter 3

Sequence Variation at the Major Histocompatibility Complex *DRβ* loci in Beluga Whales (*Delphinapterus leucas*).

Abstract

The variation at loci with homology to *DRβ* class II Major Histocompatibility Complex loci was assessed in 313 beluga collected from 13 sampling locations across North America, and 11 narwhal collected in the Canadian high Arctic. Variation was assessed by amplification of exon 2, which codes for the peptide binding region, via the polymerase chain reaction, followed by either cloning and DNA sequencing or single-stranded conformation polymorphism analysis. Two *DRβ* loci were identified in beluga; *DRβ1*, a polymorphic locus, and, *DRβ2*, a monomorphic locus. Eight alleles representing five distinct lineages (based on sequence similarity) were found at the beluga *DRβ1* locus. Although the relative number of alleles is low when compared to terrestrial mammals, the amino acid variation found amongst the lineages is moderate. At the *DRβ1* locus, the average number of nonsynonymous substitutions per site is greater than the average number of synonymous substitutions per site (0.0806 : 0.0207 respectively, $p < 0.01$). The majority of 31 amino acid substitutions do not conserve the physiochemical properties of the residue, and 21 of these are located at positions implicated as forming pockets responsible for the selective binding of foreign peptides side chains. Only *DRβ1* variation was examined in 11 narwhal revealing a low amount of variation. These data are consistent with an important role for the *DRβ1* locus in the cellular immune response of beluga.

Introduction

One of the main functions of the genes within the Major Histocompatibility Complex (Mhc) is to encode for cell surface glycoproteins which play a key role in the initiation of an immune response. Mhc class II loci are found on the surface of cells of the immune system (e.g. B-cells and antigen presenting cells). These glycoproteins bind, with varying degrees of affinity, extracellular foreign peptides (e.g. peptides from bacteria engulfed and degraded by the immune system cells) in an antigen binding groove, also known as the Peptide Binding Region (PBR). Class II molecules, in association with foreign peptide are displayed on the cell surface where they are presented to T-cells, which subsequently modulate an immune response (Germain 1994; Germain et al 1996).

This study reports the type and amount of allelic variation at loci with homology to Mhc class II *DRβ* loci in beluga whales (*Delphinapterus leucas*) and the closely related narwhal (*Monodon monoceros*). Allelic variation was assessed by analysis of nucleotide sequences within exon 2, which encodes the PBR. The PBR shows a large amount of the functional allelic variation expressed in the beta chains of most terrestrial mammal Mhc class II cell surface glycoproteins (e.g. Hughes and Nei 1989; Trowsdale 1995). The high levels of genetic variation found in the PBR of most mammals have been proposed to be an adaptation resulting from the large number of pathogens encountered by natural populations (Klein and Takahata 1990). The type of selection (i.e. positive [directional], negative [balancing], or neutral) that has shaped the variation present at a Mhc locus may be deduced by analyzing directly the PBR.

Within marine mammals, previous studies of Mhc variation have found limited amounts of variation. Limited restriction fragment length polymorphism (RFLP) variation has been found at Mhc loci in a sample of nine fin (*Balenoptera physalus*) and five sei (*B. borealis*) whales (Trowsdale et al. 1989) and in the southern elephant seal (*Mirounga leonina*) (Slade 1992). The small amount of Mhc polymorphism found in these different

groups, as compared to terrestrial mammals, suggested that this may be a common feature of all marine mammals and is due to either a decreased exposure to parasite diversity (Slade 1992) and/or to the homogeneity of the marine environment. An analysis of PBR variation at the *DQβ* locus of beluga also shows a low level of polymorphism, however, strong evidence of positive selection is consistent with the functional significance of this locus in the immune response (Chapter 2).

Variation at one locus is not a measure of variation for the entire Mhc. In the cotton-top tamarin (*Saguinus oedipus*) low levels of variation were found at some class I and class II loci while *DRβ* loci contained a high amount of variation (Gyllensten et al. 1994). This, and similar observations, have been explained by the shifting accordion hypothesis (Klein et al. 1993b). In this hypothesis, variation is reduced in small founding populations, through random drift, and is generated by selective pressure and the duplication of Mhc loci. Therefore, it is important to characterize and quantify the variation found at a number of Mhc loci before the hypothesis of reduced levels of Mhc variation in marine mammals can be accepted.

The study of Mhc variation in cetaceans, a mammalian group that has undergone adaptive radiation within the marine environment, may lead to insights into Mhc evolution. The objective of this study is to characterize and quantify the amount of Mhc class II *DRβ* variation in beluga. The evolutionary significance of the *DRβ* locus will be assessed and the hypothesis of reduced amounts of Mhc variation in marine mammals addressed.

Materials and Methods

Samples

Tissue samples were collected over a number of years from a narwhal and a number of beluga populations. Sampling locations have been classified into wintering populations, and the number of animals analyzed given (Table 3.1).

Table 3.1 List of samples analyzed including species, population, sample location, year of sampling, number of individuals sampled.

Species	Wintering population – Sample location	Sampling year(s)	Sample size
Beluga			
	St. Lawrence Estuary	88-91,94, 95	47
	Bering Sea		
	– Mackenzie Delta[1]	84, 87 - 90	48
	– Point Lay	88 - 90	24
	Hudson Strait		
	– Churchill	89 - 90, 93	29
	– Arviat	86, 87	20
	– Nastapoka River	87,93	37
	– Little Whale River	92, 93	9
	– Cumberland Sound	86	12
	Baffin Bay		
	– West Greenland	90	43
	– Grise Fiord	84, 85, 87	20
	– Creswell Bay	93	5
	– Cunningham Inlet	88, 90	19
			$\Sigma = 313$
Narwhal			
	High Arctic		
	– Arctic Bay	85, 87	9
	– Grise Fiord	87	2
			$\Sigma = 11$

[1] Mackenzie Delta area includes 6 locations: East Whitefish Station, Hendrickson Island, Tuktoyuktuk, Single Point, Kendall Island, and West Whitefish Station.

SSCP analysis

Two primer sets were used to amplify *DRβ*-like sequences. The first primer set, *DRβ*-5b, CTC GCC GCT GCA TGA AAC (Ammer et al. 1992), and DRBAMP-A, C CCC ACA GCA CGT TTC TTG (Tsjui et al. 1992), were used to survey the entire sample of whales at the *DRβ1* locus. *DRβ*-5b is located at the 5' end of exon 2 covering the intron/exon boundary, while DRBAMP-A is located at the 3' end of exon 2. This primer set amplifies 238 bp (88%) of exon 2. A second primer set, *DRβ*-5c, TCA ATG GGA CGG AGC GGG TGC, and DRBAMP-A (from above) were used to survey the variation in a subsample (n=42) of the above sample. Primer *DRβ*-5c is located 3' of *DRβ*-5b, within exon 2 and amplifies alleles of both the *DRβ1* and *DRβ2* loci. This primer set amplifies 181 bp (67%) of exon 2. The PCR conditions were as follows; 10 mM Tris-HCl (pH-8.3), 50 mM KCl, 2.5 mM $MgCl_2$, 0.2 mM dNTP's, 0.2 µM of each primer (1 primer end-labelled with $\gamma\,^{33}$P-dATP (ICN)), 0.77 units of Taq DNA polymerase (Perkin-Elmer-Cetus), and 50-100 ng of template DNA carried out in a 10 µl volume. Thermal cycling was conducted on a Perkin-Elmer-Cetus, model 480, and involved 3 cycles of 94 ^{o}C for 3 min., 56 ^{o}C for 1 min., and 72 ^{o}C for 2 min. followed by 27 cycles of 94 ^{o}C for 15 sec., 56 ^{o}C for 30 sec., and 72 ^{o}C for 1 min.. PCR products were electrophoresed through a non-denaturing acrylamide gel (5% acrylamide (59 acrylamide: 1 bisacrylamide), 15% glycerol, and 1/2 TBE) for 13 h at room temperature, and the gel was subsequently dried onto Whatman filter paper. The SSCP's were visualized by exposure of the dried gel to PhosPhor Image screens (Molecular Dynamics). A PhosphoImager (Molecular Dynamics) was used to analyze the audioradiograph images captured on the phosphor image screens. Variant SSCP patterns were chosen for sequencing analysis, and cloned alleles were used to reconstruct the observed genotype SSCP patterns as described in Chapters 2 and 5.

Sequencing of DRβ alleles

The samples chosen for sequencing were amplified as above, except in a 50 µl volume. The products of this reaction were isolated from a 1.5% agarose gel using the GlassMAX (BRL) system, cloned into the pGEM^T vector system (Promega), and transformed into DH5 alpha (BRL) competent *Escherichia coli* cells. Clones containing inserts were analyzed with the SSCP protocol above, and variants sequenced. Every allele has been cloned and sequenced from at least two separate samples. Using this protocol, a large number of clones can be screened and artifacts due to PCR error identified. Nomenclature of the alleles is based on the proposed rules for nomenclature of the Mhc of different species (Klein et al. 1990) and is based on sequence similarity. Standard nomenclature is a four letter species code (i.e. *Dele* and *Momo* for beluga and Narwhal respectively), a locus code, an asterisk (*), and a four digit allele code, i.e. *Mhc Dele-DRβ**0101. The first two numbers represent the allelic lineage, while the last two designate the unique sequence.

Analysis of results

The relationship among alleles was estimated using the computer package Phylip 3.5c (Felsenstein 1993). Pairwise genetic distance measures were generated using the program DNAdist. This program estimated distance based on Kimura's two-parameter model (Kimura 1980). A distance tree was constructed from the distance matrix using the Neighbor-Joining program which is based on the neighbor-joining method of Saitou and Nei (1987). In order to test the significance of the branches, 1000 bootstrap replicates were conducted (Felsenstein 1985). Parsimony analysis of nucleotide substitutions was conducted using the computer program PAUP 3.1.1 (Swofford 1993). Estimation of the most parsimonious unrooted tree was conducted using the exhaustive tree searching method.

Pairwise comparisons of nucleotide substitutions between alleles were conducted according to the method of Nei and Gojobori (1986) by the computer program MEGA (Kumar et al. 1993). The number of nonsynonymous (d_n) and synonymous (d_S) substitutions per site were estimated for each pair from the Jukes-Cantor formula, and the mean d_n and d_S values compared with a Student *t-test* with infinite degrees of freedom (Kumar et al. 1993).

Results

DRβ1 variation

Through a SSCP survey of 313 beluga and 11 narwhal collected from throughout North America, eight beluga and three narwhal alleles have been identified and sequenced (Fig. 3.1). These alleles have been classified into allelic lineages based on their nucleotide similarity (Fig. 3.2). A nucleotide sequence from each lineage was compared to all sequences deposited in Genbank or EMBL (Table 3.2). In each case, the most similar sequence contained in the depositories (ranging from 85% - 90% identity) was an allele of a *DRβ* locus, and in most cases, was of ungulate origin.

All nucleotide sequences code for an uninterrupted amino acid sequence (Fig. 3.3). The majority of nucleotide substitutions lead to amino acid substitutions. A pairwise comparison of all *DRβ1* alleles shows the average number of nonsynonymous substitutions per site (d_n = 0.0806 SEM 0.0134) is significantly greater ($p < 0.01$) than the average number of synonymous substitutions per site (d_S = 0.0207 SEM 0.0130). Further, these amino acid substitutions are found primarily at positions important for peptide binding (Fig. 3.3). Of the 31 amino acid substitutions which occur among the beluga *DRβ1* alleles, 21 are at 16 of the 19 positions implicated as responsible for the selective binding of foreign peptides in side chain binding pockets (Brown et al. 1993; Stern et al. 1994). In

Figure 3.1 Nucleotide sequence of the beluga and narwhal *DRβ* alleles. A '.' indicates an identical base pair with reference to *Dele-DRβ1**0101. Exon 2 codon positions are given. Codons marked by underlining or with an asterisks indicate the encoded amino acids either face the peptide binding groove or may be part of a peptide binding pocket respectively in the human *DR* molecule (Brown et al. 1993; Stern et al. 1994).

```
                     *   10   *         *                                   20                      *
Dele-DRβ1*0101      tat cag ttt aag ggc gag tgt cgt ttc tct aat ggg aca gag cgg gtg cgg ctc
Dele-DRβ1*0102      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Dele-DRβ1*0201      ct. ... c.. ... .c. ... ... ... ... ... ... ... ... ... ... ... ... ...
Dele-DRβ1*0202      ct. ... c.. ... .c. ... ... ... ... ... ... ... ... ... ... ... ... ...
Dele-DRβ1*0301      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Dele-DRβ1*0401      .t. .g. ... ... a.. ... ... ... ... ... ... ... ... ... ... ... .a. ...
Dele-DRβ1*0402      .t. .g. ... ... a.. ... ... ... ... ... ... ... ... ... ... ... .a. ...
Dele-DRβ1*0501      .t. ... ... ... .c. ... ... ... ... ... ... ... ... ... ... ... ... t..

Momo-DRβ*0101       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Momo-DRβ*0102       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
Momo-DRβ*0103       ... ... ... ... ... ... ... .a. ... ... ... ... ... ... ... ... ... c..

Dele-DRβ2*0601                                                                       .. ta.

                   *       *  31                               *   *       40                          *
D-β1*0101      gtg acc aga cac atc tat aac ggg gag gaa ttc atg cgc tac gac agc gac gtg ggc gag tgc
D-β1*0102      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D-β1*0201      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D-β1*0202      ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D-β1*0301      ... ... ... g.. ... ... ... ... ... ..g .a. g.. ... ... ... ... ... ... ... ... ...
D-β1*0401      ... gat ... t.. ... ... ... ... ... ..g .a. g.. ... ... ... ... ... ... ... ... .a.
D-β1*0402      ... gat ... t.. ... ... ... ... ... ..g .a. g.. ... ... ... ... ... ... ... ... .a.
D-β1*0501      a.. ... ... t.. ... ... ... ... ... ..g .a. g.. ... .g. ... ... ... ... ... ... .a.

M-β*0101       ... ... ... ... ... ... ... .a. ... ... ... ... ... ... ... ... ... ... ... ... ...
M-β*0102       ... ... ... ... ... ... ... .a. ... ... ... ... ... ... ... ... ... ... ... ... ...
M-β*0103       ... ... ... ... ... ... ... .a. ... ... ... ... ... ... ... ... ... ... ... ... ...

D-β2*0601      a.. ... ... ... ... ... ... ... ... ..g .a. g.. ... .t. ... ... ... ... ... ... .a.
```

```
                      50                                   *   *       60  *           *
D-β1*0101   cgg gcg gtg acc gag ctg ggc cgg cgg acc gcc gag tcc ttg aac agc cag aag gac ttc ctg
D-β1*0102   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D-β1*0201   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D-β1*0202   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D-β1*0301   ... ... ... ... ... ... ... ... .c. ga. ... a.. .a. .g. ... ... ... ... ... c.. ...
D-β1*0401   ... ... .a. ... ... ... ... ... ... ... ... ... .a. .g. ... ... ... ... ... c.. ...
D-β1*0402   ... ... .a. ... ... ... ... ... ... ... ... ... .a. .g. ... ... ... ... ... c.. ...
D-β1*0501   ... ... ... ... ... ... ... ... ... ... ... ... ... ..c ... ... ... ... ... c.. ...

M-β*0101    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
M-β*0102    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
M-β*0103    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

D-β2*0601   ... a.. ... ... ... ... ... ... ... ... ... ... .a. .g. ... ... ... ... ... c.. ...


            69   *   *           *               *       80                  *   *  87
D-β1*0101   gag cgg.aga cgg gcc gag gtg gac acg gtg tgc aga cac aac tac ggg gtt gtg gag a
D-β1*0102   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .g. ... .
D-β1*0201   ... ... ... ... ... ... ... ... ... tac ... ... ... ... ... ... ... ... ... .
D-β1*0202   ... ... ... ... ... ... ... ... ... tac ... ... ... ... ... ... ... .g. ... .
D-β1*0301   ... ... ... ... ... ... ... ... ... tac ... ... ... ... ... ... ... ... ... .
D-β1*0401   ... .a. .ac ... ... .c. c.. ... ... tac ... ... ... ... ... ... ... ... ... .
D-β1*0402   ... .a. .ac ... ... .c. c.. ... ... tac ... ... ... ... ... ... ... .g. ... .
D-β1*0501   ... ... cac ... ... .c. c.. ... ... tac ... ... ... ... ... ... ... .g. ... .

M-β*0101    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .
M-β*0102    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .g. ... .
M-β*0103    ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .

D-β2*0601   ... .a. .a. .a. ... a.. .c. ... ... tac ... ..g ... ... ... ... .c. ... ... c
```

Figure 3.2 Neighbor-Joining radial dendrogram of beluga *DRβ1* and narwhal *DRβ* alleles. Bootstrap values greater than 70% are placed along the branches.

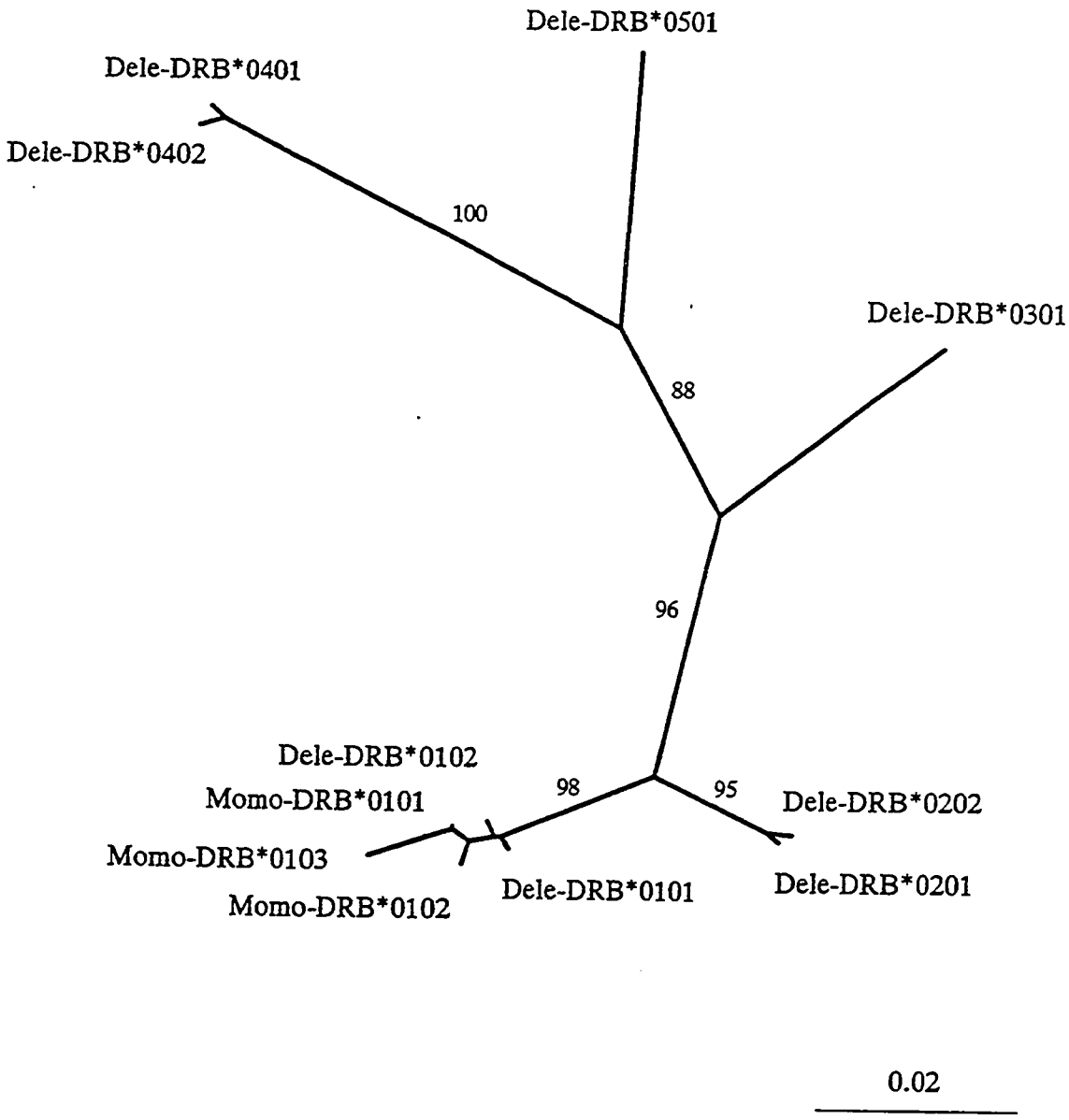
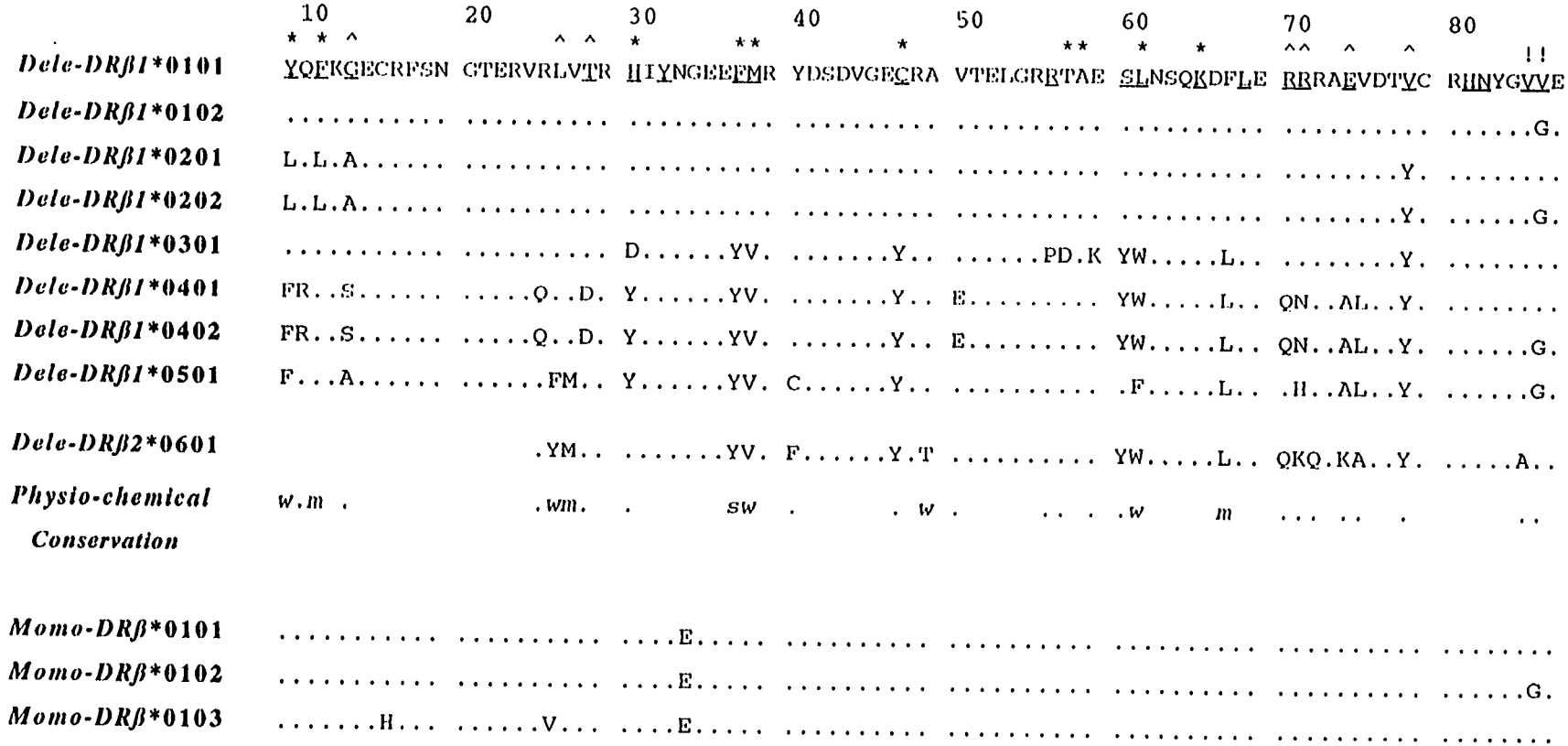

Dele-DRB*0501
Dele-DRB*0401
Dele-DRB*0402
100
Dele-DRB*0301
88
96
Dele-DRB*0102
Momo-DRB*0101
98
95
Dele-DRB*0202
Momo-DRB*0103
Dele-DRB*0101
Dele-DRB*0201
Momo-DRB*0102
0.02

Table 3.2 Results of a Genebank BLASTN search for each allelic lineage (August 26/96).

Dele DRβ1 allele	Species	Locus	allele	Identities
0101	*Bos taurus*	Class II *DRβ* gene	*BoLA-DRβ3**1101	200/229 (87%)
0201	*Alces alces*	Class II *DRβ* gene	moose 57	200/233 (85%)
0301	*Alces alces*	Class II *DRβ* gene	moose 06	207/229 (90%)
0402	*Alces alces*	Class II *DRβ* gene	moose 44,60,64	204/229 (89%)
0501	*Galago moholi*	Class II *DRβ* gene	*Gamo-DRβ1**0402	202/234 (86%)
	Alces alces	Class II *DRβ* gene	moose 30	200/229 (87%)
Dele DRβ2 allele				
0601	*Alces alces*	Class II *DRβ* gene	moose 54	165/188 (87%)
	Alces alces	Class II *DRβ* gene	moose 44	164/188 (87%)
	Pan troglodytes	Class II *DRβ* gene	*DRβ3**0211 & 0213	164/188 (87%)

Figure 3.3 Amino acid sequence of the beluga and narwhal *DRβ* alleles based on the nucleotide sequences (Fig. 3.1). Positions facing the peptide binding groove are underlined. Amino acid positions implicated in peptide side chain binding pockets of the human HLA-*DR1* molecule (Brown et al. 1993; Stern et al. 1994) are shown by !, ^, and *. The P1 peptide side chain binding pocket residues are indicated with an !, P4 with ^. Positions believed to have a role in all other probable side chain binding pockets are indicated by an *. For variable positions the conservation of physiochemical properties (i.e. polar or nonpolar, size, shape, and charge) of the residue is indicated by *s*, *m*, and *w* (strong, medium, and weak conservation respectively, McLachlan 1972). For positions with multiple substitutions the least conservative change is given.


10 20 30 40 50 60 70 80
Dele-DRβ1*0101 YQEKGECRFSN GTERVRLVTR HIYNGEEFMR YDSDVGECRA VTELGRRTAE SLNSQKDFLE RRRAEVDTVC RHNYGVVE
Dele-DRβ1*0102 ....G.
Dele-DRβ1*0201 L.L.A..... ....Y. .....
Dele-DRβ1*0202 L.L.A..... ....Y. .....G.
Dele-DRβ1*0301 .... D.....YV. .....Y.. .....PD.K YW.....L.. .....Y. .....
Dele-DRβ1*0401 FR..S..... .....Q..D. Y.....YV. .....Y.. E..... YW.....L.. QN..AL..Y. .....
Dele-DRβ1*0402 FR..S..... .....Q..D. Y.....YV. .....Y.. E..... YW.....L.. QN..AL..Y. .....G.
Dele-DRβ1*0501 F...A..... .....FM.. Y.....YV. C.....Y.. ..... .F.....L.. .H..AL..Y. .....G.
Dele-DRβ2*0601 .YM.. .....YV. F.....Y.T ..... YW.....L.. QKQ.KA..Y. .....A..
Physio-chemical w.m . .wm. . sw . . w . . . .w m ... .
Conservation
Momo-DRβ*0101 ....E..... ....
Momo-DRβ*0102 ....E..... ....G.
Momo-DRβ*0103 .....H... .....V... ....E..... ....

addition, the majority of variable positions contain substitutions that do not maintain the physiochemical properties of the residue (Fig. 3.3).

A comparison of the amino acid substitutions observed among the beluga *DRβ* sequences, to a selection of mammals, indicates that although the beluga have a relatively small number of sequences, there are a modest amount of substitutions among the sequences (Table 3.3). Amino acid substitutions unique to beluga are observed at six positions in this comparison.

DRβ2 variation

A second set of primers was used to investigate *DRβ* variation in a subset of beluga samples (n=42). These primers amplified all the *DRβ1* alleles and a new additional sequence belonging to a second *DRβ* locus, *DRβ2*. This sequence has been classified as the allele *Dele-DRβ2**0601 (Fig. 1.3 and 3.3). All samples analyzed via SSCP with the second primer set were found to contain a conformation polymorphism consistent with this allele. The alleles amplified by the second primer set were cloned and sequenced from three samples chosen from this survey. In every case they contained the *Dele-DRβ2**0601 allele in addition to the appropriate portion of the *DRβ1* allele(s) predicted from the first survey. Finally, a SSCP pattern consistent with the *Dele-DRβ2**0601 allele was found in eight beluga, each homozygous for one of the eight *DRβ1* alleles (Fig. 3.4a). Up to two stable single stranded conformations are found to represent the phenotype of each allele. The phenotypes of each of the eight cloned *DRB1* alleles amplified with the second primer set are run along side the original products (Fig. 3.4a) for comparison (Fig. 3.4b). This provides evidence that the *Dele-DRβ2**0601 allele is associated with each *DRβ1* allele, and indicates a conserved *DRβ* haplotype structure consisting of a variable *DRβ1* allele and a monomorphic *DRβ2* allele.

The *Dele-DRβ2**0601 allele contains nine unique substitutions (Fig. 3.1). Of these, eight are within complete codons, and seven lead to unique amino acid substitutions

Table 3.3 Comparison of the amino acid substitutions found among *DRβ* exon 2 sequences, positions 9 - 87, for 8 mammal species (Adapted from Swarbrick et al. 1995). *DRβ* sequences are from, human (Marsh and Bodmer 1993), chimpanzee (O'hUigin et al. 1993), red deer (Swarbrick et al. 1995), cattle (Ammer et al. 1992; Sigurdardottir et al. 1991), sheep (Schwaiger et al. 1993), goat (Schwaiger et al. 1994), and moose (Mikko and Anderson 1995). Positions that face the peptide binding groove or that may by involved in peptide side chains binding are indicated as in Fig. 3.1. Positions unique to beluga are underlined and in bold. Amino acids in brackets are substitutions found in the *Dele-DRβ2**0601 allele. Total number of sequences compared and amino acid substitutions observed are given at the end of each column.

Consensus	Human	Chimp	Red Deer	Cattle	Sheep	Goat	Beluga	Moose
E*	EKQW	EKLW	ELM	EQ	E	E	FLY	E
Y (10)	ELQY	LPQY	HQY	Y	Y	Y	QR	Y
A/S*	ADGLRP SV	ADEGHP SV	AFHLPST Y	ACHLST Y	AHRSTY	CHSTY	FL	AH
K	KT	KT	KT	KT	KRT	KT	K	K
S*	CFGHRS Y	ACFGHR SY	AGKRS	GKRS	KRS	GKRS	AGS	S
E	EK	EK	R	E	E	E	E	E
C	C	C	C	C	C	C	C	C
H	HQY	H	HPY	H	HR	HR	R	H
F	F	F	F	F	F	F	F	F
F	FL	FL	FPS	F	FS	FS	S	F
N	N	N	N	NS	N	N	N	N
G (20)	G	G	G	G	G	G	G	G
T	T	T	T	T	T	T	T	T
E	E	DEQ	DEQ	E	E	EG	E	E
R	R	RW	R	R	R	R	R	R
V	V	MV	MV	LV	V	V	V	V
R	R	QRW	EGQR	R	RW	GRW	QR	R
F*	FLNY	FLNY	FLSY	FLY	FLY	FLY	FL	F
L	L	L	L	L	L	L	VM (Y)	L
D*	DEIH	DEHIK	ADEGQ	DEHN	DE	DH	TD	D
R	R	R	R	R	R	R	R	R
Y* (30)	CDGHLR Y	DGHLRY	Y	CHSY	Y	FY	DHY	Y
F	FIV	FI	FIV	FY	FY	FY	I	I
Y	HY	HY	CHSY	HTY	HTY	HTY	Y	Y
N	HNW	N	NS	N	N	N	N	N
Q	Q	Q	GRW	G	G	G	G	R
E	E	E	E	E	E	E	E	E
E	E	DE	EG	E	E	E	E	E
Y*	DFLNSY	DFNSY	FTY	FLNTY	FNTY	FINTY	FY	Y
V*	ALV	AFLMV	ALV	V	ALV	LV	MV	V
R	R	R	R	R	R	LR	R	R
F (40)	FY	FY	FY	F	F	FY	CY (F)	F
D	DN	DN	D	D	D	D	D	D
S	S	S	CS	S	NS	NS	S	S
D	D	D	D	D	D	D	D	D

Table 3.3 Continued.

Consensus	Human	Chimp	Red Deer	Cattle	Sheep	Goat	Beluga	Moose
V	LV	LV	VW	W	W	RW	V	V
G	G	G	G	DG	GS	G	G	G
E	E	E	E	E	E	E	E	E
Y*	FY	FY	FY	F	FY	FY	CY	Y
R	QR	QR	QR	QR	R	R	R	R
A	A	A	A	A	A	A	A (T)	A
V (50)	V	V	V	V	V	V	EV	V
T	RT	T	T	T	A	AT	T	T
E	E	E	E	E	E	E	E	E
L	L	LP	L	L	L	LQ	L	L
G	G	G	G	G	G	G	G	G
R	R	R	LR	QR	R	QR	R	R
P	P	P	PR	PQR	PQR	EPQ	PR	P
D*	ADSV	ADSV	DSV	ADRSV	ADES	ADEST	DT	DS
A*	AE	A	A	ARV	AV	A	A	A
E	E	EKV	EK	EKV	EK	EK	EK	K
Y (60)	HSY	SY	GLY	HQY	HQY	Y	SY	Y
W*	W	CWY	LWY	CLY	W	W	FLW	W
N	N	N	N	N	N	N	N	N
S	S	S	RS	GS	S	S	S	S
Q	Q	Q	LQR	Q	Q	Q	Q	Q
K*	K	K	KR	K	K	K	K	E
D	D	DG	ED	DE	DEN	DE	D	D
L	FILY	FILY	FILY	FILT	FIL	FIL	FL	I
L	L	LV	LM	L	L	L	L	L
E	E	AE	E	EG	E	E	E	E
Q* (70)	DQR	DHQR	DEQR	DEQR	EQR	DQRS	QR	Q
R*	AEKR	AEKRST	AEHKLN RST	AEGKR	AKRT	KRS	HNR (K)	GRT
R	R	GR	GR	R	R	R	R (Q)	R
A	AG	AG	AG	A	AT	AT	A	A
A/E*	AELQR	ADEKRT Q	AEN	AENSY	AENT	AE	AE (K)	AE
V	V	EV	V	V	V	AV	LV (A)	V
D	D	D	DN	D	DN	D	D	D
T	NT	NT	RT	RT	T	T	T	T
Y*	VY	FVY	FVY	VY	VY	CFVY	VY	Y
C	C	C	C	C	C	C	C	C
R(80)	R	R	R	R	R	R	R	R
H	HY	HY	HY	H	H	HY	H	H
N	N	N	DN	DNS	N	N	N	N
Y	Y	Y	Y	Y	Y	Y	Y	Y
G	G	GR		G	G	G	G	G
V*	AV	AV		GRV	V	V	V (A)	V
G*	GV	AGLV		FGMV	DFGI	FGILV	GV	GI
E	E	E		E	E	E	E	E
	n = 69 78 aa	n = 64 98 aa	n = 34 90 aa	n = 35 71 aa	n = 35 53 aa	n = 21 59 aa	n = 9 39 aa	n = 10 5 aa

Figure 3.4 A) PCR-SSCP phenotypes of eight beluga homozygous (listed in red) at the *DRβ1* locus amplified with the second primer set. B) PCR-SSCP phenotypes of the eight beluga from A, reanalyzed with each of the *DRβ1* clones and *DRβ2*0601* (listed in black and prefixed by DRB#*) amplified with the second primer set. For clarity, only the single stranded conformations are given .

A
DRB1 Homozygotes
0501
0402
0401
0301
0202
0201
0102
0101
B
DRB clones
DRB2*0601
0101
DRB1*0101
0102
DRB1*0102
0201
DRB1*0201
0202
DRB1*0202
0301
DRB1*0301
0401
DRB1*0401
0402
DRB1*0402
0501
DRB1*0501
DRB2*0601

(Fig. 3.3). A parsimony tree of amino acid substitutions suggests a basal origin of this sequence (Fig. 3.5). It may have arisen following a duplication of an ancestral allele of present *DRβ1* lineages. Eight unique, and three possible convergent amino acid substitutions, may have occurred along this lineage, while only a single synonymous substitution is observed.

Narwhal DRβ variation

Three *DRβ* sequences were found in a sample of 11 narwhals (Table 3.4). One allele *Momo-DRβ**0101 was found at a high frequency, 0.68. The genotype frequencies are not significantly different from the Hardy-Weinberg expectations. The narwhal sequences are similar to the *Dele-DRβ**01 sequences (Fig. 3.2), however, a common substitution of Gly to Glu (position ß34) separates the beluga and narwhal sequences (Fig. 3.3). Among the narwhal sequences three amino acid substitutions are observed, two of which occur at positions implicated in peptide side chain binding pockets (Brown et al. 1993; Stern et al. 1994). No synonymous substitutions are observed (Fig. 3.1).

Discussion

Evolution of beluga DRβ1 alleles

Parsimony analysis of the nucleotide substitutions indicates that most of the substitutions are the result of point mutations (C.I. 0.825; Fig. 3.5). Among the beluga *DRβ1* allele, 57 substitutions are needed to most parsimoniously explain the evolution. Excluding position ß86 improves the C.I. to 0.868 (i.e. 46 observed substitution/53 required for maximum parsimony). At position ß86 there is an indication of intraexonic recombination similar to that observed in moose (*Alces alces*) (Andersson and Mikko 1995) and murine species (She et al. 1991). In moose, the 10 variable nucleotide positions observed among the *DRβ* alleles can be explained by 18 point mutations, or more

Figure 3.5 Phylogram of the maximum parsimony relationships of the *DRβ* allelic lineages based on nucleotide substitutions. This is an unrooted phylogram that has been midpoint rooted after the analysis. Branch lengths are proportional to the number of substitutions. This tree requires 57 substitutions and has a consistency index (C.I.) of 0.825. Possible convergent substitutions are identified on each branch with the ß-chain exon 2 amino acid position of each substitution listed (see Fig. 3.1).

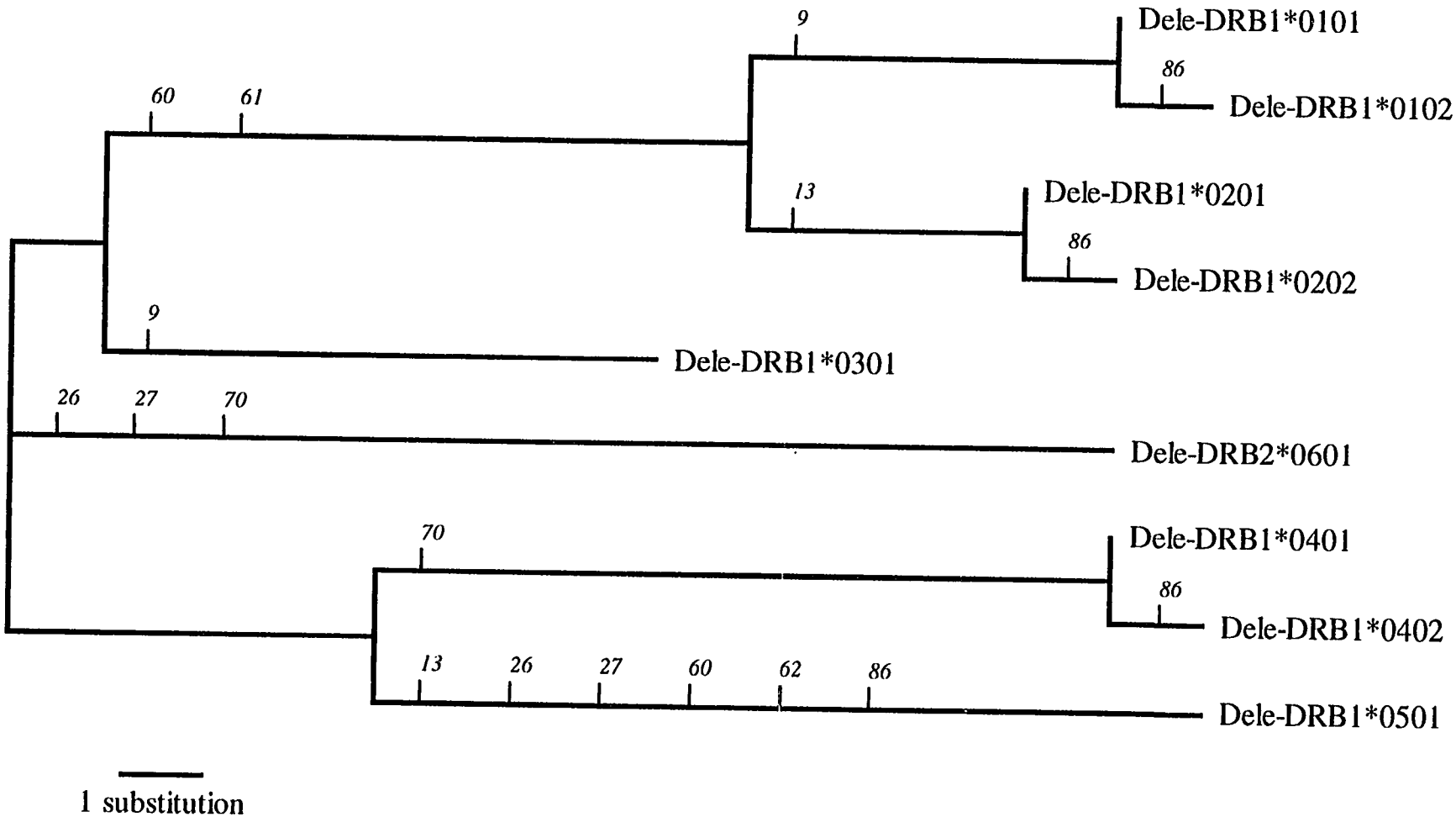

9
Dele-DRB1*0101
86
Dele-DRB1*0102
60
61
13
Dele-DRB1*0201
86
Dele-DRB1*0202
9
Dele-DRB1*0301
26
27
70
Dele-DRB2*0601
70
Dele-DRB1*0401
86
Dele-DRB1*0402
13
26
27
60
62
86
Dele-DRB1*0501
1 substitution

Table 3.4 Genotype and allele frequency of a *DRβ* locus in a survey of narwhal .

Momo-DRβ Genotype	Genotype frequency	*Momo-DRβ* allele	Allele frequency
0101 / 0101	5	0101	15
0101 / 0102	1	0102	2
0101 / 0103	2	0103	5
0102 / 0103	1		

fparsimoniously, by 10 unique point mutations and four recombination events. Three of the our recombination events involve a sequence motif which is correlated with an amino acid substitution at position ß86. The possible recombination of this motif has also been noted in murine species (She et al. 1991). Position ß86 is variable in both beluga and narwhal (Fig. 3.3), and, in beluga, it is the only difference between the *DRß1** alleles 0101 and 0102, 0201 and 0202, and 0401 and 0402 respectively. This similarity may be equally explained in beluga by four convergent nucleotide substitutions or by one substitution (possibly prior to speciation) and three recombination events.

Evidence for intraexonic recombination has also been noted at the primate *DRß1* locus (Gyllensten et al. 1991), however, see O'hUigin (1995) for an alternate interpretation (i.e. biases in replacement rates leading to convergent evolution). Gyllensten et al. (1991) report an area with sequence similarity (~10 bp out of 13) to a putative recombination signal that roughly separates the ß-sheet from the α-helical regions of the PBR. In beluga, this region has the same amount of similarity to the putative recombination signal, however, there is no evidence of recombinational events involving this location. It may be more informative to view the recombination of peptide motifs with reference to the three-dimensional structure of the molecule. It is interesting that of four or five possible peptide side chain binding pockets only one, P1, is never interrupted by members of other possible side chain pockets along the ß chain (Brown et al. 1993; Stern et al. 1994). Part of the P1 peptide side chain binding pocket is made up of ß chain amino acids of the carboxyterminal of the α-helical region and contains positions ß85, ß86, ß89 and ß90 in the human *HLA-DR1* molecule (Stern et al. 1994). The recombination of a complete side chain pocket ß chain motif may have a higher selective advantage in that the new Mhc molecule would possess a functionally tested side chain binding pocket. This may explain the seemingly high recombination rate for this portion of the ß chain.

In human class II molecules, the P1 side chain binding pocket, of which position ß86 is a key part, is thought to provide a very intense hydrophobic anchor (Rammensee et

al. 1995). The Gly/Val substitution at position ß86 observed both in beluga and humans (Table 3.3) involves the substitution of a tiny, hydrophobic amino acid with an aliphatic, hydrophobic amino acid (Table 3.5). In human *DR* molecules, the variation at P1 is limited to the same Gly/Val polymorphism (Rammensee et al. 1995) that has been correlated with differences in P1 binding specificity in very similar molecules. Friede et al. (1996) examined the natural ligand motifs of human *HLA-DR4* molecules and found that molecules with a Gly at position ß86 (i.e. *DRβ1**0401 and *0405) show a different P1 specificity than *DR4* molecules with a Val at position ß86 (i.e. *DRβ1**0402 and *0404). This suggests that the observed variation at ß86 in beluga may cause a shift in the peptide binding specificity of the alleles within the allelic lineages, *Dele-DRβ1**01, *02, and *04.

In addition to P1, between two to four other peptide side chain pockets, with a more degenerate range of binding specificity, are important in peptide binding in class II molecules (Falk et al. 1994; Rammensee et al. 1995: Friede et al. 1996). The interpretation of binding pocket motifs from DNA sequence data by comparison to positions important in other class II molecules, i.e. *DR1*, should, therefore, be made with caution (Stern et al. 1994; Friede et al. 1996). In Table 3.5, positions implicated as being involved in the *HLA-DR1*, P4 side chain binding pocket are listed (Brown et al. 1993; Stern et al. 1994). The importance of these positions in peptide side chain binding in beluga *DR* molecules is only speculative, however, such comparison may allow possible differences in peptide binding specificity to be assessed. The specificity of side chain pockets is well explained by the characteristics of the amino acids that make up the pocket , i.e. hydrophobicity, size and charge (Friede et al. 1996). A comparison of amino acid variation found at the speculated P4 binding sites among beluga alleles shows large changes in the physio-chemical characteristics of many of these sites (Table 3.5). For example, in allele *Dele-DRβ1**0401, substitutions at positions ß70 and ß71 have replaced two positively charged with two neutrally charged amino acids, while a neutral amino acid in position ß28 has been replaced with the negatively charged aspartic acid. Although the absolute positions of peptide

Table 3.5 Comparison of variation among beluga alleles at amino acid positions implicated in the P1 and P4 side chain binding pockets of the human *HLA-DR1* molecule. The possible amino acid positions in each pocket are based on the analysis of the crystal structure of human *HLA-DR1* (Brown et al. 1993; Stern et al. 1994). Positions P1 and P4 are relative to the bound peptide. Physio-chemical propeties of the amino acids are given in brackets: Hydrophobicity, Hpb=hydrophobic, Hpl=hydrophilic, Hab=ambivalant; Size, s=small, t=tiny; Structure, ar=aromatic (ring containing side chains), al=alipahtic (only those with branched side chains); Charge +'=positive, -'=negavtive (after Taylor 1986).

Allele	**P1**		**P4**						
*DRß1**	ß85	ß86	ß13	ß26	ß28	ß70	ß71	ß74	ß78
0101	V (Hpb,s,al)	V (Hpb,s,al)	G (Hpb,t)	L (Hpb,al)	T (Hab,s)	R (Hpl,+')	R (Hpl,+')	E (Hpl,-')	V (Hpb,s,al)
0102	V	G (Hpb,t)	"	"	"	"	"	"	"
0201	V	V	A (Hpb,t)	L	T	R	R	E	Y (Hab,ar)
0202	V	G	"	"	"	"	"	"	"
0301	V	V	G	L	T	R	R	E	Y
0401	V	V	S (Hpl,t)	L	D (Hpl,s,-')	Q (Hpl)	N (Hpl,s)	A (Hpb,t)	Y
0402	V	G	"	"	"	"	"	"	"
0501	V	G	A	F (Hpb,ar)	T	R	H (Hab,ar,+')	A	Y
*DRß2**									
0601	A (Hpb,t)	V	---	Y	T	Q	K (Hab,+')	K (Hab,+')	Y

binding are unknown for the beluga alleles, this comparison argues for differences in the binding specificities of the beluga alleles.

Levels of Mhc variation in Marine Mammals

Previous studies of the level of Mhc variation in marine mammals have found a low amount of allelic variation in comparison to terrestrial mammals (Trowsdale et al. 1989; Slade 1992; Chapter 2). Evidence for the functional importance of the class II Mhc in cetacean, found at the *DQβ* loci in beluga, indicates that low levels of variation may be due to a reduction in the balancing selection pressure in marine mammals (Chapter 2), as first suggested by Slade (1992). Slade (1992) argues that the marine environment has a reduced amount of micro-parasite diversity that affects marine mammals when compared with terrestrial mammals, thus reducing the balancing selection pressure maintaining Mhc diversity. A consequence of the supposed reduced Mhc variation is the occurrence of epizootics (i.e. morbillivirus, Visser et al. 1993) that have been reported for marine mammals.

Similar evidence of positive selection at the *DRβ1* locus as seen at the beluga *DQβ* locus (Chapter 2) (i.e. significantly higher levels of nonsynonymous substitutions and the position and nonconservative nature of the substitution), indicates the functional significance of a second Mhc class II locus in the immune response of beluga. In contrast, however, the modest amount of variation at the *DRβ1* locus, and the level of positive selection needed to generate such variation, do not support the hypothesis of reduced balancing selection pressure in marine mammals. Terrestrial mammals have over a four fold greater substitution rate of nonsynonymous changes (positive selection) at PBR sites than synonymous changes (neutral evolution) (Klein et al. 1993a). This difference is thought to be due to balancing selection pressure in the form of either overdominance or frequency dependent selection (Klein et al. 1993a; Hughes and Hughes 1995). Slade's (1992) hypothesis predicts that there is a reduction in balancing selection pressure in the

marine environment. This would mean that either heterozygotes or new, rare alleles would have a lower fitness when compared with terrestrial mammals. In either case, the substitution rate of nonsynonymous changes at PBR sites should be much lower in marine mammals than that observed in terrestrial mammals. In this study, an average rate of nonsynonymous substitution per site in exon 2, which contains a large number of PBR sites, is 3.9 times greater than the synonymous rate per site. The selection pressure needed to generate the variation among the beluga *DRβ1* alleles is similar to that observed in terrestrial mammals, arguing against a reduction in the marine environment.

Klein et al. (1993a) estimated a synonymous substitution rate of 1.2×10^{-9} /site/year, a nonsynonymous substitution rate at non-PBR sites of 0.4×10^{-9} /site/year, and a nonsynonymous substitution rate at PBR sites of 5.9×10^{-9} /site/year at Mhc class I and II molecules. These estimates allow for the prediction of time of divergence of allelic lineages. At the beluga *DRβ1* second exon, 71 codons, or 237 nucleotides, with an average of 59 synonymous and 179 nonsynonymous sites, were examined. From the terrestrial estimates, a rate of synonymous mutations of 0.7×10^{-7} /PBR/year is calculated. In other words, we expect to see 0.7 silent substitutions every 10 million years (Myr). Assuming the neutral rate of evolution at Mhc loci is similar in marine and terrestrial mammals, this estimate can be used to predict the average amount of time separating the *DRβ1* alleles. An average of 1.35 (range 0.5 to 2.5) synonymous substitutions are found between the *DRβ1* allelic lineages, which corresponds to approximately 20 Myr (range 7 to 35.7 Myr), or, assuming a clock-like evolution observed at other Mhc loci (Klein et al. 1993a), 10 Myr of evolution along each lineage. A similar estimate can be made based on the nonsynonymous substitutions at exon 2. O'hUigin (1995) estimated a nonsynonymous substitution rate of 0.5×10^{-6}/exon 2/year, or 0.5 substitutions per million years. The average number of nonsynonymous substitutions among the allelic lineages is 14.9 (range 7 to 22). Using O'hUigin's estimate for the partial exon 2 sequences examined here (~80% of nucleotides) an average time separating the lineages of 29.8 Myr (range 14 to 44 Myr),

or about 15 Myr of evolution, is calculated. These estimates place the average origin of beluga *DRβ1* allelic lineages in the late Miocene, a date which coincides with the origin of monodontids (Barnes et al. 1985; Fordyce and Barnes 1994) and indicates that most of the allelic lineages, with the exception of 01 and 02, probably diverged prior to the origin of modern beluga species.

It is interesting to note that the increase in allelic variation of beluga *DRβ1*, when compared with *DQβ*, is also associated with a duplication event. Parsimony analysis suggests that one synonymous and 10 nonsynonymous substitutions occurred along the *DRβ2**06 lineage. This indicates a similar selective pressure acting at this locus, and places the duplication event at about 10 Myr, a time which is contemporary with the diversification of the allelic lineages. It is not clear why this locus has remained monomorphic. Speculatively, the duplication of a functional Mhc *DRβ* locus may have lowered the selective pressure, previously maintaining generalist Mhc molecules. This may have enabled one locus, in this case *DRβ1*, to evolve in response to pathogen diversity. This hypothesis is supported by the modest amount of *DQβ* variation in right whales also being associated with a duplication (Chapter 6).

This data is in contrast with the results of Slade (1992) and Trowsdale et al. (1989). Further, recent work on seal and baleen whale Mhc is also contradictory. Hawaiian monk seals (*Monachus schauinslandi*) have no DQα variation and a low level of class I variation consistent with very low levels of overall genetic variation (Armstrong 1995). In contrast, the European harbour seal (*Phoca vitulina*) has levels of class I variation similar to terrestrial mammals, and substitution patterns which indicate a similar selective pressure in the marine and terrestrial environment (Goodman and Slade 1995). Finally, a modest amount of *DQβ* variation in a small sample of right whales (*Eubalaena glacialis*) has been described (Chapter 6). These results suggest that the reduction of Mhc variation in some marine mammals may be due to reasons other than selective pressure. Population bottlenecks have been implicated in the reduction of Mhc variation in terrestrial mammals.

Moose (*Alces alces*) and cheetahs (*Acinonyx jubatus*) both have reduced levels of Mhc and genetic variation and are believed to have gone through a population bottleneck in the last 10,000 - 50,000 years (Ellegren et al. 1996; O'Brien et al. 1985 respectively). In addition, European beavers (*Castor fiber*) also have low levels of Mhc and do not seem to be suffering the deleterious effects of inbreeding despite a dramatic reduction of genetic diversity in some populations (Ellegren et al. 1993), suggesting that a bottleneck of some form has occurred in the evolutionary past prior to the establishment of the European populations. With respect to marine mammals, however, only in the Hawaiian monk seal has a population bottleneck been correlated with a reduction in Mhc variation. In most of the other marine mammals, modest levels of genetic variation are present (e.g. Slade 1992; Patenaude et al. 1994).

Speculatively, the high proportion of marine mammals with decreased levels of Mhc polymorphism may be due to the nature of speciation events in the marine environment. The existence of trans-species polymorphism in primates suggests that effective population sizes at the time of speciation must have remained relatively large (Klein et al. 1993c). It is possible that, on average, speciation occurs in a population of smaller size in marine mammals. Marine mammals are highly mobile and live in an environment with few geographic boundaries. Small population sizes would, therefore, be more likely to become genetically isolated and possibly undergo speciation. This process, however, would not be universal as, given an average age of species of two Myr (Stanley 1975), the beluga allelic lineages would have probably passed through at least one speciation event. In contrast, the closely related narwhals must have shared a common ancestor with beluga within the age of the allelic lineages, and yet, have low levels of variation at both the *DQβ* and *DRβ1* loci. Furthermore, only one to two Myr of evolution are needed to account for the variation among the *DRβ* alleles found in this species. This provides evidence for the above hypothesis, however, a larger sample size, a survey of

DRβ2 variation, and an examination of the overall genetic variation of this species would also be needed for stronger support.

Current hypotheses of the speciation process in cetaceans, however, do not agree with the above interpretation. Gaskin (1982) believes ecological factors, i.e. behavioural isolation and trophic-niche differentiation, have been the most important isolating mechanisms leading to speciation in cetaceans. Alternatively, the low to modest amounts of Mhc variation described in cetaceans may be due to a general reduction in the rate of nucleotide substitution. Comparisons of cetacean mitochondrial sequences (Hoelzel et al. 1991; Baker et al. 1993), the flanking sequences of microsatellite loci (Schlötterer et al. 1991) and the beluga and right whale *DQβ* sequences (Chapter 6) indicate a 2x to 4x reduction in the neutral nucleotide substitution rate in comparison to terrestrial mammals. Although this data has been used as evidence of a need to reexamine the fossil record of cetaceans (Schlötterer et al. 1991; Milikovitch et al. 1993, 1994), it is also consistent with a correlation body size and substitution rate (Martin and Palumbi 1993). Martin and Palumbi (1993) argue that a combination of generation time and metabolic rate in what they termed "nucleotide generation time" may explain this correlation. Following the current understanding of the fossil record the colonization of the seas by cetaceans began approximately 55 Myr ago, with the emergence of the modern suborders, Odontoceti and Mysticeti, occurring in the late Eocene, 35-40 Myr ago (Fordyce and Barnes 1994). The colonization of this dramatically different environment has led to a large number of adaptations. It is reasonable to expect that the Mhc loci would have also undergone an adaptation to the new aquatic pathogens. A reinterpretation of the age of the *DRβ* allelic lineages based on a 3x slower rate of evolution (Chapter 6) gives an estimate of the average age at about 29 Myr. This places the origin of the *DRβ1* allelic lineages and the duplication event early in the phlyogeny of Odontoceti.

These two hypotheses may equally explain the modest amount of Mhc variation in cetaceans, but the latter may have more serious conservation implications. Although

relatively healthy species such as moose (Ellegren et al. 1996) and the European beaver (Ellegren et al. 1993) have relatively low amounts of Mhc variation, it has been argued that low amounts of Mhc variation also increase the risk of extinction (Yuhki and O'Brien 1990). The first hypothesis predicts that Mhc variation has been generated *de novo* in most cetacean species. Therefore most species would have passed through a period of low Mhc variation. In contrast, the second hypothesis predicts that, as in primates (Klein et al, 1993c), most of the Mhc variation is passed through speciation events. This implies an evolutionary importance of Mhc variation. Further, if the reduction in the substitution rate in cetaceans is due to a reduction in the mutation rate, then a loss of Mhc variation would have serious consequences (see Chapter 6).

The first hypothesis predicts minimal sharing of variation among closely related species while the second predicts most variation, i.e. allelic lineages, should be shared. In order to discriminate between the hypotheses it would be most informative to assess Mhc variation in delphinids or phocids, i.e. species closely related to monodontids (10-15 Myr), with large populations, and no evidence of historic bottlenecks. In this regard, the harbour porpoise (*Phocoena phocoena*) and the bottlenose dolphin (*Tursiops truncatus*), both of which have had comprehensive mtDNA surveys (Rosel et al. 1995; Wang et al. 1996; Curry et al. 1995), would be ideal. In addition, the analysis of *DRβ* variation in more distantly related odontocetes and mysticetes would allow for another estimation of Mhc substitution rates.

Conclusions

In summary, a modest amount of Mhc class II *DRβ1* variation is found in beluga populations. The strong evidence for positive selection at the PBR, and the analysis of the nature of amino acid substitutions at possible peptide side chain binding pockets, indicate the functional significance of this locus in the immune response of beluga whales. Further,

the ratio of nonsynonymous to synonymous substitutions is similar to that among primate alleles, arguing against a reduction in the balancing selection pressure in the marine environment. Two hypotheses may explain the modest amount of Mhc variation when compared with terrestrial mammals; small population sizes at speciation, or a reduced neutral substitution rate in cetaceans. Further work is needed to distinguish between these hypotheses.

Acknowledgments

Samples used in this study were obtained through the co-operation of the Canada Department of Fisheries and Oceans, the Fisheries Joint Management Committee (Inuvik), the Alaska Department of Fish and Game (North Slope Borough, Department of Wildlife Management), the Makivik Corporation, the native communities listed in Table 3.1, and the St. Lawrence National Institute of Ecotoxicology. Financial and graduate student support were provided by grants from NSERC, NSERC-DFO Science Subvention Program, the Endangered Species Recovery Fund (co-sponsored by the Canadian Wildlife Service of Environment Canada and World Wildlife Fund Canada), and a Northern Studies Training Program award. A preliminary analysis of the narwhal *DRβ* sequence variation was conducted by E. Saunders. We would also like to thank N. J. Patenaude, S. Mancuso, J. S. Quinn, J. J. Dockray, P. Wilson, C. Schutz, and J. Wang for their useful suggestions and help with research, data analysis, and manuscript preparation.

Chapter 4

Levels of Major Histocompatibility Complex variation in dead beached and live biopsied St. Lawrence beluga (*Delphinapterus leucas*).

Abstract

Genotypic and allelic variation at the Major Histocompatibility Complex class II *DRβ1* and *DQβ* loci were compared between beluga found dead, floating or beached, in the St. Lawrence and biopsy samples collected from live beluga residing in the St. Lawrence Estuary. No significant differences were observed in either the genotypic or allelic frequencies. Power analysis indicates that large differences in allelic frequencies due to concordance with the large number of pathologies found in the dead beached St. Lawrence beluga are unlikely, however smaller effect sizes could not be ruled out. A sample size of 200 individuals would be needed to test these effects. This study is consistant with the hypothesis that the dead beached whales are representative of the St. Lawrence population.

Introduction

The St. Lawrence beluga, *Delphinapterus leucas*, is considered an endangered population (COSEWIC, Campbell 1993), and is believed to be geographically isolated from the closest Arctic populations residing in either Hudson Bay and Hudson Strait or the high Arctic and Baffin Bay. Unlike arctic beluga, which migrate thousands of miles between their summering and wintering locations, the St. Lawrence beluga is restricted to the St. Lawrence Estuary (Michaud 1993). In summer, the St. Lawrence population is centered around the Saguenay river and makes use of the St. Lawrence estuary from Battures-aux-Loups-Marins to Riviére-Porneuf and Iles du Bic. In winter, the population is found from Taddousac to the Gulf where they overwinter along the ice edge.

The present beluga population, estimated at about 500 animals, is the remnant of what was once a large and dynamic population. St. Lawrence beluga were exploited, primarily for their oil and hides, as early as the 1800's, and continued to be until 1960, at which point harvesting was no longer economically feasible (Reeves and Mitchell 1984). Through a careful examination of harvest records and a back-calculation of the population size needed to sustain recorded yields, it has been estimated that there were at least 5000 St. Lawrence beluga prior to 1885 (Reeves and Mitchell 1984). In 1979, the St. Lawrence beluga were protected by an amendment of the Canada Fisheries Act. Despite a protected status for over 16 years, numerous population surveys since 1973 have indicated that the population is either stable or slowly increasing in number (e.g. Pippard 1985; Sergeant 1986; Sergeant and Hoek 1988; Kingsley 1993; Michaud 1993). A number of hypotheses have been put forward to explain this lack of significant population recovery including harassment, habitat degradation, contamination by toxic chemicals, and the deleterious effects of inbreeding.

In an attempt to understand the lack of population growth, studies have been conducted on beluga carcasses found beached along the St. Lawrence river. Over a nine

year period, population modelling and surveys have indicated low calf production and/or survivorship, toxicological studies have indicated high levels of contamination by Mercury, Lead, PCB, DDT and Mirex, and necropsies have revealed a large number of unusual tumours and pathological conditions (Béland et al. 1993). These findings suggest that the contaminants are having an adverse effect on the immune system (Martineau et al. 1987). In addition, studies on the DNA minisatellite variation of the beached whales found a reduced amount of variation in the St. Lawrence beluga when compared to either the Beaufort Sea (Patenaude et al. 1994) or the Hudson Bay populations (Mancuso 1995). Studies on captive populations of 38 mammal species have shown that, on average, the rate of mortality is 33% higher in the progeny of first degree relatives when compared to the progeny of unrelated pairs (Ralls et al. 1988). Thus, it may be a combination of environmental and genetic factors that are responsible for the lack of population recovery.

Although population modelling indicates that the beached whales have a normal age structure (Béland et al. 1988), a possible bias may exist in the sampling of dead beached whales. Recently, a biopsy darting program of live St. Lawrence beluga has been initiated. Throughout the summers of 1994 and 1995 beluga have been selectively darted and photo-identified. Only beluga with distinctive markings were biopsied so that the individuals could be identified from existing photographic catalogues, and the possibility of sample replication reduced. The biopsy darting of live beluga has allowed the direct comparison of dead beached and live beluga. This paper reports the genotypic and allelic variation at the Major Histocompatibility Complex (Mhc) class II loci, *DQβ* and *DRβ1*, in a sample of live biopsy darted whales, and compares it to levels found in dead beached whales (Chapters 2 and 3).

Cell surface glycoproteins encoded in the Mhc play a key role in the initiation of immune response by binding foreign antigens and presenting them to T-cells. The Mhc class II loci, *DQβ* and *DRβ1*, are believed to be responsible for presenting extracellular antigens (e.g. extracellular bacteria) to $CD4^+$ positive T-cells, which subsequently initiate

an immune response to the pathogen (Germain 1994; Germain et al 1996). Bacterial infections have been suggested to be a major cause of disease and death in cetaceans (Howard et al. 1983). By examining the amount of allelic variation present in the Mhc of the St. Lawrence population, the question of reduced genetic variation and its possible effect on immune response can be addressed (Chapters 2 and 5). This study investigates the hypothesis that the dead beached beluga represent a biased sample of the actual variation found in the St. Lawrence population. It is possible that specific Mhc alleles/genotypes have led to a predisposition to the pathologies observed in the beached beluga carcasses. The assumption and hypothesis that the dead beached whales are truly representative of the St. Lawrence beluga will be tested.

Materials and Methods

Samples

Tissue samples were collected over a number of years from dead beached whales and through tissue biopsy of live whales. The dead beached whale samples consist of 21 individuals collected in 1988 (n=8), 1989 (n=4), 1990 (n=4), 1991 (n=4) and 1995 (n=1). Biopsy samples of live St. Lawrence beluga were collected by Robert Michaud (Institut National d'Ecotoxicologie du St. Lawrence). Throughout the summers of 1994 and 1995, beluga have been selectively darted and photo-identified. Only beluga with distinctive markings were biopsied so that the individuals could be identified from the existing catalogues, and the possibility of sample replication reduced. Twenty six live beluga whales were biopsy darted; in 1994 (n=4) and 1995 (n=22).

DQβ and DRβ1 Genotyping

The *DQβ* and *DRβ1* genotypes of each sample were determined through a combination of single stranded conformation analysis and DNA sequencing as described in

Chapter 2 and Chapter 3 respectively.

Analysis of results

Analysis of population variation was conducted with the computer program GENEPOP, version 2 (Raymond and Rousset 1995a). A Markov chain approach was used to estimate the outcome of an exact test in order to test the null hypothesis (H_O) of "random union of gametes", expected from a population in Hardy Weinberg equilibrium (Rousset and Raymond 1995). The probability of the observed genotype table was compared to a distribution of the probabilities of genotype tables with the same allelic counts. The genotype tables used for the comparison distribution were constructed following a Markov chain approach. The p-value of the test corresponds to the sum of probabilities of all tables with the same or lower probability as the observed data. For each test 10,000 dememorization steps were conducted before 100 batches of 5,000 steps was used to estimate the p-value.

Comparisons of allelic and genotypic frequencies between sample types were conducted through estimations of exact tests utilizing a Markov chain approach to construct new tables with the same marginals (Raymond and Rousset 1995b). For allelic and genotypic comparisons, the H_O is "the allele or genotype distribution is independent across populations". In the genotypic comparisons, a F_{St} based exact test was performed. In this test, the p-value of the observed F_{St} is estimated from the sum of probabilities of all tables with an equal or higher F_{St} value. F_{St} values were estimated following the formula of Weir and Cockerham (1984). All comparisons were done with 10,000 dememorization steps followed by 100 batches of 10,000 steps to estimate the p-value of the observed distributions of alleles or F_{St} values (Raymond and Rousset 1995a).

Analysis of power of the comparison of allele frequencies (i.e. the probability of a rejection of the null hypothesis for a given α level with x degrees of freedom (df), sample size, and effect size) was performed by comparison to tables of the power of χ^2 test, with

an $\alpha = 0.05$ and either three, four, or five degrees of freedom (Cohen 1977). Effect size (*w)* was estimated from actual and idealized contingency tables by the formula of Cohen (1977);

$$w = \sqrt{\sum_{i=1}^{m} \frac{(P_{1i} + P_{0i})^2}{P_{0i}}},$$

where P_{1i} = the proportion in cell *i* posited by the alternate hypothesis (reflecting the effect for that cell), P_{0I} = the proportion in cell i posited by the null hypothesis, and m = the number of cells. Idealized data were used to estimate the expected w for various changes in allele frequency. In these cases, sample sizes for the beached and biopsied samples were equal to 50 alleles (total = 100) and the frequencies observed in the biopsied samples were used as the expected St. Lawrence frequencies.

Results and Discussion

The Mhc class II *DRβ1* and *DQβ* genotypes and allele frequencies of 26 biopsy darted beluga were determined and compared to those observed in the dead beached whales (Table 4.1). The genotype frequencies at each locus within each sample type were not significantly different from the Hardy-Weinberg expectations. Comparisons of allele frequencies revealed no significant difference at either the *DRβ1* (p = 0.7391, SE 0.0019) or the *DQβ* (p = 0.3193, SE 0.0016) loci. Finally, F_{St} based comparisons of either the *DRβ1* or the *DQβ* genotype frequencies revealed that the observed distributions were not significantly different from expected (Fst -0.0075, p = 0.6324, SE 0.0021 and Fst = 0.0167, p = 0.1510, SE 0.0014 respectively). In addition to the results presented here, an examination of the probable *DRβ1/DQβ* haplotypes also reveals no significant differences between the sample types. Of the nine probable haplotypes, all are observed in the beached samples and eight are observed in the biopsied samples (Chapter 5).

Table 4.1 Mhc *DRβ1* and *DQβ* genotype and allele frequencies for dead beached and live biopsy darted beluga samples.

	Beached	Biopsied		Beached	Biopsied
DRβ1 Genotype*			*DRβ1* Allele*		
0101/0101		1	0101	8	15
0101/0301	2	4	0201	8	7
0101/0301		1	0301	2	4
0101/0401	3	8	0401	18	20
0201/0401	6	1	0402	6	5
0301/0401	1	1	0501		1
0401/0401	3	4	Total	42	52
0101/0402	3				
0201/0402		2			
0301/0402	1	1			
0401/0402	2	2			
0101/0501		1			
Total	21	26			
DQβ Genotype*			*DQβ* Allele*		
0101-2/0101-2	12	21	0101-2	33	47
0101-2/0103	1		0103	1	
0101-2/0201	4	3	0201	4	3
0101-2/0201	4	2	0202	4	2
Total	21	26	Total	42	52

These data suggest that at the Mhc class II loci examined, the dead beached beluga do not represent a biased sample of the variation found in the population. However, power analysis indicates that only a large effect size would have a high probability of rejecting the null hypothesis given the current sample size (Table 4.2). In order, to investigate the effect of changes of allele frequency on effect size, two arbitrary scenarios of an increase of a single allele frequency were examined (Table 4.2). In the first example, the frequency of an allele with intermediate frequency is doubled, trebled, and quadrupled. In the second example, the same increase in allele frequency is used at the most common allele. In both cases, only when allele frequencies increased by 21, was a high level of power reached. This would represent a substantial effect size, as at least 10 individuals, in addition to those expected by chance, would have had to have died due to an increased susceptibility to pathogens.

Observed data shows effect sizes of $w = 0.162$ and $w = 0.190$ for *DQβ* and *DRβ1* respectively. Comparison of these values to Table 4.2 and to power Tables (Cohen 1977) indicate that these effect sizes are not trivial and that they both have power ≤ 0.30. The probablity of committing a type II error (accepting the null hypothesis when it is false) is equal to 1-power. Power of ≤ 0.30 indicates that if the effect was real and not random, given the current sample size, there would be a 70% chance of accepting the null hypthesis when it was false. Analysis of power Tables (Cohen 1977) and an artificial increase in the sample size, maintaining the relative frequencies, both suggest that a total sample size of at least 300 to 400 alleles (150 to 200 in each group) would be needed to detect these effects if they were real (i.e. reject the H_0, $\alpha < 0.05$). Over 100 animals beached along the shores of the St Lawrence have been sampled with detailed necropsies being performed on ≥ 45 of those (Béland et al. 1993). Further, a goal of the ongoing biopsy darting program is to analyze over 100 individuals, and with the samples obtained in the summer of 1996, the sample size is currently approaching 50 (Michaud, pers comm). Sample sizes of 400 alleles are obtainable and would be of great use for the further testing of the hypothesis.

Table 4.2 Effect size and estimation of power (Cohen 1977) of idealized data at the *DRβ1* locus assuming an $\alpha = 0.05$ (4 df) and a total sample size of 100 alleles. Two cases are presented, both with increases of a single allele frequency, in increments of seven.

*DRβ1*allele Sample type	0401	0101	0201	0402	0301/ 0501
Biopsied	19	14	7	5	5
Increase of 0201 freq.					
$w = 0.165$; Power ~ 0.20	16	12	**14**	4	4
$w = 0.305$; Power ~ 0.66	13	9	**21**	3.5	3.5
$w = 0.430$; Power > 0.90	10	7	**28**	2.5	2.5
Increase of 0401 freq.					
$w = 0.141$; Power ~ 0.20	**26**	11	5.5	3.75	3.75
$w = 0.280$; Power ~ 0.62	**33**	7.6	3.8	2.8	2.8
$w = 0.427$; Power > 0.90	**40**	4.5	2.3	1.6	1.6

It is important to note however, that the results of the *DQβ* variation in the biopsied samples support the findings of Chapter 2 (which are based on beached samples only), indicating no significant difference between the Hudson Bay, eastern Beaufort Sea, eastern Chukchi Sea and St. Lawrence sampling locations. The similarity in allele frequencies among these geographically separated beluga populations argues against the effect size within the St. Lawrence samples at this locus being caused by a real effect, and points toward the observed effect size being due to sampling variance.

An examination of power was not conducted on the comparison of genotype frequencies, however, as power is directly related to sample size, the power of the genotype comparisons is probably lower than the allelic comparisons. Artificial increases in sample size, again maintaining genotype frequencies, finds that total sample sizes of 100 and 200 individuals, for *DQβ* and *DRβ1* respectively, would be needed to reject the null hypothesis given the current effect sizes ($\alpha < 0.05$).

A large number of toxicological and pathological studies have been conducted on the dead beached St. Lawrence beluga (reviewed in Béland et al. 1993; De Guise et al. 1994a, 1994b; De Guise et al 1995a; Martineau et al. 1995). These studies find unusual amounts and types of pathologies in the beluga carcasses. It would be advantageous for future comparisons to correlate the results of these studies with the comparisons of genotype frequencies. In this way, more specific hypotheses can be addressed and the expected effect sizes more clearly predicted. This would represent an exciting new avenue of research, important for both the ongoing assessment of immune functions in the St. Lawrence beluga (De Guise et al. 1995b) and further examination of sampling bias.

The results of the postmortem examination of 13 of the dead beached beluga examined in this study are available for comparison (De Guise et al. 1994a and 1995a). For these animals, Mhc genotypes can be compared to the non-neoplastic lesions, tumor, and bacterial infections observed (Table 4.3). These data allow more specific alternate hypotheses to be tested, such as the correlation of a bacterial infection and sample

Table 4.3 Mhc *DQβ* and *DRβ1* genotypes and pathology of 13 dead beached St. Lawrence beluga. Table adapted from De Guise et al. (1995a), Table 1. Pathology of non-neoplastic lesions and bacteriological examination of from De Guise et al. (1995a)[1]. Identification of tumors from De Guise et al. (1994a)[2].

Whale (yr-#)[1]	Sex	Age	Mhc Genotype		Pathology			
			DQβ	*DRβ1*	Non-neoplastic Lesions[1]	Bacteria[1]	Site of bacteria[1]	Tumors[2]
Dele-185 (1988-4)	F	≥21	0101-2 0101-2	0401 0401	Ciliated protozoal bronchopneu-monia, adrenal cysts, mild chronic mastitis with foci of squamous metaplasia	Escherichia coli	Lung, liver, kidney	Scirrhous gastric adenocarcinoma
Dele-187 (1988-7)	M	≥19	0101-2 0202	0101 0402	Severe, diffuse, fibrino-purulent peritonitis, left side testicular necrosis, left epididymitis, severe periodontitis, oral ulceration, focal verminous bronchopneu-monia, emaciation	Pseudomonas putrefaciens	Lung, liver, kidney, peritoneal fluid, in pure culture	
Dele-188 (1988-9)	F	≥22	0101-2 0101-2	0101 0201	Mild myocarditis, calcified intes-tinal mass, sarcocystosis, adrenal cyst, subcutaneous haematoma	Plesiomonas shigelloides	Intestinal contents; mesenteric lymph node, in pure culture	Hepatocellular carcinoma, mammary adenocarcinoma, pulmonary chondroma
Dele-189 (1988-11)	F	≥25	0101-2 0201	0201 0401	Lymphocytic enteritis with ces-todes, ulcerations of 1st and 2nd gasteric compartments, renal and adrenal haematoma, chronic mas-titis, adrenal cysts, probable toxo-plasmosis (spleen)	Morganella morganii	Intestine	

Table 4.3 continued

Whale (yr-#)[1]	Sex	Age	Mhc Genotype *DQβ*	Mhc Genotype *DRβ1*	Pathology: Non-neoplastic Lesions[1]	Pathology: Bacteria[1]	Pathology: Site of bacteria[1]	Pathology: Tumors[2]
Dele-190 (1988-13)	F	≥21	0101-2 0101-2	0401 0401	Diffuse alveolar damage of lungs, mild chronic mastitis, erosions of 1st gastric compartment, oesophageal ulcerations, intestinal polyp, subcutaneous haematoma, adrenal cystic haematoma	none		Gastric papilloma, granulosa cell tumor
Dele-342 (1989-1)	M	≥27	0101-2 0202	0101 0402	Diffuse interstitial pneumonia, parasitic ulceration of 1st gasteric compartment, mild periodontitis, fibrous plaques on intima of aorta	Klebsiella sp. Not Done	Lungs ---	Gastric papilloma, phaeochromocytoma
Dele-334 (1989-3)	F	≥19	0101-2 0101-2	0401 0402	Subacute parasitic bronocopneumonia, mild diffuse lymphocytic enteritis, ulcerative mamillitis, periodontitis, adrenal nodules and csyt.	Not Done	---	
Dele-335 (1989-4)	F	14	0101-2 0103	0101 0401	Parasitic eosinophilic pneumonia, mediastinal eosinophilic lymphadenitis, mild subacute enteritis with cestodes, chronic endocarditis, oesophageal ulceration, adrenal cyst, ovarian follicular cyst	Not Done	---	Gastric papilloma

Table 3 continued

Whale (yr-#)[1]	Sex	Age	Mhc Genotype *DQβ*	Mhc Genotype *DRβ1*	Pathology: Non-neoplastic Lesions[1]	Pathology: Bacteria[1]	Pathology: Site of bacteria[1]	Pathology: Tumors[2]
Dele-336 (1989-6)	F	≥25	0101-2 0101-2	0101 0401	Fibrinous peritonitis, necropurulent mastitis, ulceration of 2nd gasteric compartment, adrenal cyst and nodule	Not Done	---	Granulosa cell tumor
Dele-338 (1990-2)	M	≥23	0101-2 0201	0201 0401	Mild to moderate purulent bronchopneumonia, fibrous plaques on intima of aorta	Clostridium tertium	Lungs, liver, kidney	Abdominal fibroma, thyroid follicular adenoma
Dele-339 (1990-4)	F	13	0101-2 0202	0101 0402	Moderate acute eosinophilic bronchopneumonia (verminous) with mediastinal eosiniphilic lymphadenitis, mild lymphocytic enteritis, multifocal segmental glomerulopathy	none		
Dele-340 (1990-5)	F	≥13	0101-2 0101-2	0301 0401	Mild vaginitis, subcutaneous pyogranulomas	Aeromonas sp,	Liver	
Dele-341 (1990-6)	M	≥24	0101-2 0101-2	0401 0402	Ulcerative stomatitis, peridontitis, ulceration of 2nd gasteric compartment, thyroid abscess, mild purulent multifocal pneumonia, lymphocytic enteritis, abdominal mass (probably scar tissue), adrenal cyst and nodules	Clostridium perfringens Clostridium sp. Kingella kingae Aeromonas sp.	Intestinal contents Lung and thyroid Thyroid	

homozygosity. Speculatively, homozygosity at Mhc loci may lower the ability to respond to bacteria and lead to an increased likelihood of infection. In this case eight samples were found to have significant intracellular or extracellular bacterial infections. Comparison of the frequency of *DRβ1* homozygotes and heterozygotes (1:7 respectively) in this sample to the frequencies found in the biopsied sample (6:20) does not reject the null hypothesis ($p = 0.4647$, Fisher exact test). As expected, an examination of power reveals that only a large effect would be detected with this sample size (i.e. 2:6 in the beached sample), however, the observed effect is small and in the opposite direction to that predicted by the alternate hypothesis. This trend is also observed at the *DQβ* loci (i.e. 4:4 compared to 21:5).

Conclusions

The results of the present study indicate that a large effect of allelic or genotypic concordance with the observed pathologies is not present in the dead beached beluga samples. However, the smaller, more realistic effect sizes, observed in the current data, remain ambiguous and could be tested by an increase of sample size to about 100 individuals for each sampling type (400 alleles). Finally, no support is found for the alternate hypothesis that homozygosity at Mhc class II loci leads to greater susceptibility of bacterial infection, demonstrating the usefulness of correlating Mhc genotypes with observed pathologies.

Acknowledgments

Samples used in this study were obtained through the co-operation of the Canada Department of Fisheries and Oceans, Parks Canada, and the St. Lawrence National Institute of Ecotoxicology. Financial and graduate student support were provided by grants from NSERC, NSERC-DFO Science Subvention Program, the Endangered Species

Recovery Fund (co-sponsored by the Canadian Wildlife Service of Environment Canada and World Wildlife Fund Canada). We would also like to thank M. Kingsley, N. J. Patenaude, J. S. Quinn, J. J. Dockray, P. J. Wilson, R. Waldick, and J. Wang for their useful suggestions and help with research, data analysis, and manuscript preparation.

Chapter 5

Comparison of allelic and haplotype variation of Major Histocompatibility Complex class II *DRβ1* and *DQβ* loci in North American beluga (*Delphinapterus leucas*) populations

Abstract

The variation at loci with homology to *DRβ* class II Major Histocompatibility Complex loci was assessed in 313 beluga (*Delphinapterus leucas*), collected from 13 sampling locations across North America, and 11 narwhal (*Monodon monoceros*) collected in the Canadian high Arctic. Variation was assessed by amplification of exon 2, which codes for the peptide binding region, via the polymerase chain reaction, followed by either cloning and DNA sequencing or single-stranded conformation polymorphism analysis. Comparisons of allelic and genotypic distributions among all sampling locations indicates genetic structure ($p < 0.0001$). No structure is observed among the sampling locations of beluga which belong to either the Hudson Strait (Churchill River, Arviat, Nastakopa River, Little-whale River, and Cumberland Sound, $p = 0.1736$) or Bering Sea (Mackenzie Delta and Point Lay, $p = 0.0809$) wintering populations. Among the sampling locations of the Baffin Bay wintering population, there is evidence of genetic structure (Nuussuaq, Disko Bay, Grise Fiord, Creswell Bay, and Cunningham Inlet, $p > 0.0001$) which is lost when Cunningham Inlet is excluded from the comparison ($p = 0.3557$). Significant genetic structure ($p = 0.05$) exists among each pair of the following sampling location groups; St. Lawrence, Hudson Strait, Bering Sea, Cunningham Inlet (CI), and Baffin Bay (-CI). In the St. Lawrence population, six of the eight *DRβ1* alleles are present. Even though the frequencies of these alleles are significantly different from the northern populations, all five allelic lineages are present. Evidence of linkage disequilibrium between the *DRβ1* and *DQβ* is present in two populations, the St. Lawrence and Baffin Bay. Analysis of probable *DRβ1-DQβ* haplotypes among populations shows a reduction in the haplotypes found in the St. Lawrence.

Introduction

Beluga whales (*Delphinapterus leucas*) have a circumpolar distribution and are restricted to arctic or near arctic waters. Within North America, 11 populations are currently recognized based on summering distributions: St. Lawrence Estuary, eastern Hudson Bay, western Hudson Bay, Ungava Bay, southeast Baffin Island, Baffin Bay/High Arctic, eastern Beaufort Sea, eastern Chukchi Sea, Norton Sound/Yukon Delta, Bristol Bay, and Cook Inlet (IWC 1992). Samples from most of these populations, except the western Alaskan (Norton sound/Yukon Delta, Bristol Bay, and Cook Inlet) and the Ungava Bay populations will be compared in the following survey. Ice plays a very important role in the biology of this cetacean, regulating its yearly migration patterns. In summer, beluga congregate in shallow river estuaries. Strong site fidelity to estuaries has been noted in behavioural observations (Caron and Smith 1990) and in mitochondrial DNA variation (Brennin 1992; Mancuso 1995; Brown 1996). Beluga are excluded from arctic estuaries by the advancement of arctic ice. In North America, five major overwintering locations have been speculated, St. Lawrence, Hudson Strait, Baffin Bay, Bering Sea and Cooks Inlet, and beluga may be classified on this basis. A number of summering populations are thought to overwinter together in the Hudson Strait and Bering Sea. As mating is thought to occur during the late winter or spring migration (Heide-Jørgensen and Teilmann 1994) the amount of mating occurring among the summering groups is unknown.

Of the populations examined in this survey, the St. Lawrence and southeast Baffin Island populations are considered endangered (COSEWIC, Campbell 1993) and/or of small size and vulnerable to hunting or habitat deterioration (IWC 1992), the eastern Hudson Bay is considered threatened (COSEWIC, Campbell 1993) and/or of medium size but exploited at rates that give cause for concern (IWC 1992), and the High Arctic/Baffin Bay is

considered threatened (COSEWIC, Campbell 1993) and/or of large size but exploited at a rate that gives cause for concern (IWC 1992). Of these, the southernmost and geographically isolated population in the St. Lawrence Estuary has been the most thoroughly studied population. Although these resident whales have been legally protected since 1979, their population, which numbered over 5000 animals at the turn of the century (Reeves and Mitchell 1984), has failed to increase in number. Several surveys since 1973 have indicated the population (around 500) is either stable or slowly increasing in number (Pippard 1985; Sergeant 1986; Sergeant and Hoek 1988; Kingsley 1993; Michaud 1993). In an attempt to understand the lack of population growth, studies have been conducted on beluga carcasses found beached along the St. Lawrence river. Toxicological studies indicate high levels of contamination by Mercury, Lead, PCB, DDT and Mirex while necropsies have revealed a large number of unusual tumours and pathological conditions (Béland et al. 1993). These findings suggest that the contaminants are having an adverse effect on the immune system. At the same time, genetic analysis of minisatellite loci has shown significantly higher levels of band sharing in the St. Lawrence animals when compared to those in the eastern Beaufort Sea (Patenaude et al. 1994) or the Hudson Bay (Mancuso 1995) indicating a reduction in genetic variation. Thus it may be a combination of environmental and genetic factors that are responsible for the lack of population recovery. By examining the amount of allelic variation present in the Mhc of the St. Lawrence stock, the question of inbreeding depression and its possible effect on immune response can be addressed.

Cell surface glycoproteins encoded by the Major Histocompatibility Complex (Mhc) play a key role in the initiation of an immune response by binding foreign peptides and presenting them to T-cells. The Mhc class II loci, *DQβ* and *DRβ1*, are believed to be responsible for presenting extracellular antigens (e.g. bacteria peptides) to $CD4^+$ positive T-cells, which subsequently initiate an immune response to the foreign pathogen (Germain 1994; Germain et al. 1996). Bacterial infections have been suggested to be a major cause

of disease and death in cetaceans (Howard et al. 1983). The high levels of Mhc class I and class II genetic variation found in most mammals at this gene complex have been proposed to be an adaptation resulting from the large number of pathogens encountered by natural populations (Klein and Takahata 1990). A low degree of polymorphism at this complex, as in the case of the African cheetah, has been suggested to be the cause of the increase in the population's susceptibility to pathogens (O'Brien et al. 1985). It has been suggested that Mhc allelic variability should be taken into account in management of captive breeding programs (Hughes 1991a) and that lack of variation at this locus in an isolated population may increase the chance of extinction of that population (Yuhki and O'Brien 1990).

Previous studies of Mhc in beluga whales have investigated the amount and type of variation at the Peptide Binding Region (PBR) (Chapters 2, 3 and 4). The PBR shows a large amount of the functional allelic variation expressed in the beta chains of most terrestrial mammal Mhc class II cell surface glycoproteins (e.g. Hughes and Nei 1989; Trowsdale 1995). This region is postulated to be directly involved with the interaction and association of foreign peptides which are subsequently presented to T-cells, and is used as an indicator of allelic diversity. The previous studies (Chapters 2 and 3) indicate that the variation among the exon 2 sequences at the Mhc class II *DQβ* and *DRβ1* loci in beluga are consistent with the functional significance of these loci. This paper expands on the previous population survey of *DQβ* variation (Chapter 2), reports the findings of a population survey of *DRβ1* variation, examines the linkage between these two loci, and reports on levels of *DRβ1-DQβ* haplotype diversity.

Materials and Methods

Samples

Tissue samples were collected over a number of years from beluga populations (Table 5.1). Samples collected from a beluga wintering population off the west coast of

Table 5.1 List of samples analyzed including species, population, sample location, year of sampling, number of individuals sampled and origin of sample.

Wintering population * Summering population -- Sample location	Year	Sample size	Origin of Sample[1]
St. Lawrence Estuary (SL)	88 - 91,95	21	B
	94, 95	26	BD
Bering Sea			
* Eastern Beaufort Sea			
-- Mackenzie Delta[2] (MC)	84, 87 - 90	48	H
* Eastern Chukchi Sea			
-- Point Lay (PL)	88 - 90	24	H
Hudson Strait			
* Western Hudson Bay			
-- Churchill (CH)	89 - 90, 93	26	LC
	91	3	H
-- Arviat (AV)	86, 87	20	H
* Eastern Hudson Bay			
-- Nastapoka River (NA)	87	35	H
	93	2	LC
-- Little Whale River (LW)	92, 93	3	H
	92, 93	6	LC
* Southeastern Baffin Island			
-- Cumberland Sound (CS)	86	12	H
Baffin Bay			
-- West Greenland-Nuussuaq[3] (NQ)	90	28	H
-- West Greenland-Disko Bay[3] (DB)	90	15	H
* High Arctic			
-- Grise Fiord (GF)	84, 85, 87	20	H
-- Creswell Bay (CB)	93	5	LC
-- Cunningham Inlet (CI)	88, 90	19	H

[1] Under origin of sample a 'B' indicates a dead beached whale, a 'BD' indicates a biopsy sample from a live whale, a 'LC' indicates a whale live-captured for other scientific purposes (i.e. radio telemetry) and an 'H' indicates an aboriginal subsistence hunter killed whale.

[2] Mackenzie Delta area includes 6 locations; East Whitefish Station, Hendrickson Island, Tuktoyuktuk, Single Point, Kendall Island, and West Whitefish Station.

[3] Western Greenland samples were collected in the winter and the summering population is unknown.

Greenland, previously analyzed for *DRβ1* variation (Chapter 3), were analyzed for *DQβ* variation and included in a data set previously analyzed for *DRβ1* and *DQβ* variation (Chapters 2, 3, and 4). These served as the base of comparison for Mhc allelic variation. Most of the samples were collected as a result of aboriginal subsistence hunting, however, some of the Hudson Bay and High Arctic samples were collected from whales live-captured for other scientific purposes. The St. Lawrence samples are taken from dead beached whales and through tissue biopsy of live whales. Most of these samples were collected in the summer and can be considered to represent the summering populations of the species. The western Greenland samples were collected in February and, therefore, the summering location of these whales is not known.

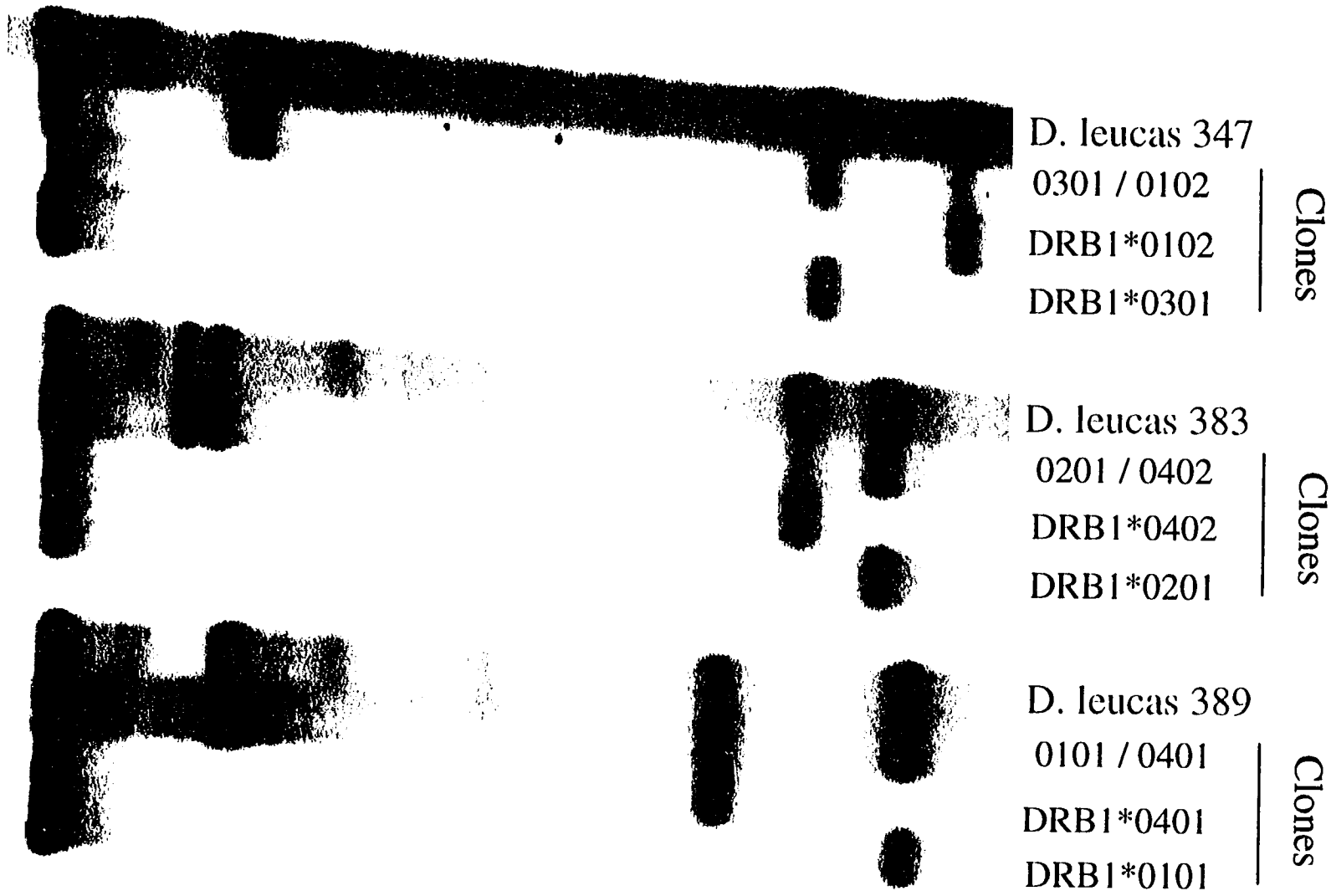
Analysis of DRβ1 and DQβ variation.

The *DRβ1* genotypes of all samples and *DQβ* genotypes of the western Greenland sample were determined according to the methods of chapters 2 and 3 respectively. The Single Stranded Conformation Polymorphism (SSCP) phenotype of each sample was compared to the phenotypes of cloned alleles and reconstructed genotypes. Samples were analyzed twice in most cases. Like phenotypes were pooled and compared to reconstructed genotypes (constructed from an equal mixture of 2 different clones. In this manner slight differences in single stranded or heteroduplex conformations were detected (Fig. 5.1). Only a single strand was end labelled with ^{33}P-dATP in order to ease comparison.

Analysis of data

Analysis of population variation was conducted with the computer program GENEPOP, version 2 (Raymond and Rousset 1995a). A Markov chain approach was used to estimate the outcome of an exact test in order to test the null hypothesis (H_0) of "random union of gametes", expected from a population in Hardy Weinberg equilibrium (Rousset and Raymond 1995). The probability of the observed genotype table was

Figure 5.1 *DRβ1* PCR-SSCP phenotypes of three beluga samples: Example of genotyping. Beluga samples are listed in red, while clones are listed in black. Individual clones are prefixed by DRB1* and reconstructed genotypes, made from an equal mixture of two clones, are identified by the two, four digit numbers of the constituent clones.


D. leucas 347
0301 / 0102
DRB1*0102
DRB1*0301
Clones
D. leucas 383
0201 / 0402
DRB1*0402
DRB1*0201
Clones
D. leucas 389
0101 / 0401
DRB1*0401
DRB1*0101
Clones

compared to a distribution of the probabilities of genotype tables with the same allelic counts. The genotype tables used for the comparison distribution were constructed following a Markov chain approach. The p-value of the test corresponds to the sum of probabilities of all tables with the same or lower probability as the observed data. For each test, 10,000 dememorization steps were conducted before 100 batches of 5,000 steps was used to estimate the p-value (Raymond and Rousset 1995a). Probability values were corrected for multiple comparisons by using the sequential Bonferroni method (Rice 1989).

Comparisons of allelic and genotypic frequencies among populations were conducted through estimations of exact tests utilizing a Markov chain approach to construct new tables with the same marginals (Raymond and Rousset 1995b). For allelic and genotypic comparisons, the H_O is "the allele or genotype distribution is independent across populations". In the genotypic comparisons, a F_{ST} based exact test was performed. In this test, the p-value of the observed F_{ST} is estimated from the sum of probabilities of all tables with an equal or higher F_{ST} value. F_{ST} values were estimated following the formula of Weir and Cockerham (1984). All comparisons were done with 10,000 dememorization steps followed by 100 batches of 10,000 steps to estimate the p-value of the observed distributions of alleles or F_{ST} values (Raymond and Rousset 1995a). Each level of comparison was corrected for multiple comparisons by using the sequential Bonferroni method (Rice 1989).

Linkage of the *DQβ* and *DRβ1* loci was tested through an estimation of an exact test (Raymond and Rousset 1995a). The H_O for this test is "genotypes at one locus are independent from genotypes at another locus". This test is similar to the analysis of allele frequencies among populations above, except a contingency table was constructed for the *DRβ1* vs *DQβ* loci for both subpopulations. Both comparisons were done with 10,000 dememorization steps followed by 100 batches of 10,000 steps to estimate the p-value of the observed contingency table.

Results

SSCP Survey of beluga DRβ1 variation

An example of the procedure used to genotype samples is shown in Fig. 5.1. In this case, three beluga samples are compared to cloned alleles and to a reconstructed genotype containing both the previous cloned alleles. The upper most bands are the most stable single strand conformation of each allele. The lower middle two bands are believed to be heteroduplex conformations formed by the annealling of a ^{33}P-dATP labelled strand with the complimentary strand from the other allele. The lower most band is normal double stranded DNA.

--Variation within sampling locations

For each sampling location, the number of individuals of each genotype is given in Table 5.2. Of the 36 possible genotypes, 34 are observed in this sample. In addition, the observed and expected number of homozygotes and heterozygotes is given for each sampling location. An exact probability test of Hardy Weinberg equilibrium was conducted for each sample location. No sampling location was significantly different from Hardy Weinberg equilibrium ($p > 0.05$, corrected for multiple comparisons by the sequential Bonferroni method (Rice 1989)). The allele frequencies within each sampling location are shown in Table 5.3. Analysis of both the genotype frequencies (Table 5.2) and the allele frequencies (Table 5.3) reveal that genetic structure exists among the sampling locations ($p > 0.0001$ and $F_{st} = 0.0606$, $p > 0.0001$, respectively).

--Variation within wintering populations

Beluga whale stocks can be classified into summering or wintering populations (Table 5.1). Excluding the St. Lawrence, three wintering populations were sampled in this survey; Hudson Strait, Bering Sea, and Baffin Bay. Table 5.4a shows a comparison of the

Table 5.2 Number of each*DRβ1* genotype at each sampling location. See Table 5.1 for two letter abbreviations.

Genotype	*Sample locations*													*Total*
	SL	NA	LW	CH	AV	CS	PL	MC	GF	CB	CI	DB	NQ	
0101/0101	1	2		1			2	1	2	1	7			17
0101/0102							1							1
0102/0102								1						1
0101/0201	6			3			1	2	2		1	3	1	19
0201/0201					1			1		1				3
0101/0202						2	3	1			1			7
0201/0202					1		2	1	1					5
0202/0202							1	1						2
0101/0301		3		1	1	4		4	1		3		2	19
0102/0301		1		1			1							3
0201/0301							2	1				1	2	6
0202/0301		1		1										2
0301/0301	1		1	1	3	1			1				1	9
0101/0401	11	4	1	2	1	1	1		1		3		1	26
0102/0401				1										1
0201/0401	7	3		1	2			1						14
0202/0401		1	1						1			1		4
0301/0401	2	6		2	4	2		3		2				21
0401/0401	7	1	1	2	2	1	1	3						18

Table 5.2 Continued

Genotype	Sample locations													Total
	SL	NA	LW	CH	AV	CS	PL	MC	GF	CB	CI	DB	NQ	
0101/0402	3	1			1			3	2	1	2		3	16
0102/0402				1									4	5
0201/0402	2			1				1	2				3	9
0202/0402				1							1	1		3
0301/0402	2		2		2			1				2		9
0401/0402	4	2		1	1			5	1				2	16
0402/0402		1							2			3	4	10
0101/0501	1	1					1		1		1			5
0102/0501		1												1
0201/0501		4		2				3				1		10
0202/0501		1		1	1		3	7						13
0301/0501				1			1	2					2	6
0401/0501		2	2	3			3	4						14
0402/0501		1	1					1	2			2	3	10
0501/0501		1		2		1	1	1	1					7
Homozygotes	9	5	2	6	6	3	5	8	6	2	7	3	5	
expected Hom.	12.2	5.9	1.6	4.1	4.2	2.7	3.8	6.6	3.5	.7	8.4	2.7	6.1	
Heterozygotes	38	32	7	23	14	9	19	40	14	3	12	12	23	
expected Het.	34.8	31.1	7.4	24.9	15.8	9.3	20.2	41.4	16.5	4.3	10.6	12.3	21.9	

Table 5.3 *DRβ1* allele frequencies at each sample locations.

Sample location	*Dele-DRβ* alleles								Total
	0101	0102	0201	0202	0301	0401	0402	0501	
St. Lawrence R.	23		15		6	38	11	1	94
Nastapoka R.	13	2	7	3	11	20	6	12	74
Little whale R.	1			1	4	6	3	3	18
Churchill R.	8	3	7	3	8	14	4	11	58
Arviat	3		5	2	13	12	4	1	40
S.E. Baffin Is.	7			2	8	5		2	24
E. Chukchi Sea	11	2	5	10	4	6		10	48
E. Beaufort Sea	12	2	11	11	11	19	11	19	96
Grise Fiord	11		5	2	3	3	11	5	40
Creswell Bay	3		2		2	2	1		10
Cunningham In.	25		1	2	3	3	3	1	38
Disko Bay	4	1	5	2	3	1	11	3	30
Nuussuaq	7	4	6		8	3	23	5	56
Total	128	14	69	38	84	132	88	73	626

Table 5.4 Results of exact test analyses of allele and genotype frequency comparisons within (A) and among (B) wintering groups.

Comparison of sample locations		p-value of allelic distribution	F_{st}	p-value of F_{st}
A) Within wintering groups				
Hudson Strait		0.1736 SE 0.0039	0.0065	0.2003 SE 0.0060
Bering Sea		0.0810 SE 0.0017	0.0068	0.2010 SE .0024
Davis Strait		< 0.0001 SE 0.0001	0.0883	< 0.0001 SE 0.0001
Davis Strait (-CI)		0.3557 SE 0.0043	0.0016	0.3929 SE 0.0061
B) Among wintering groups				
St. Lawrence	Hudson Strait	< 0.0001 SE 0.0001	0.0261	0.0005 SE 0.0002
	Bering Sea	< 0.0001 SE 0.0001	0.0525	< 0.0001 SE 0.0001
	Davis Strait (-CI)	< 0.0001 SE 0.0001	0.0701	< 0.0001 SE 0.0001
	Cunningham In.	< 0.0001 SE 0.0001	0.1832	< 0.0001 SE 0.0001
Hudson Strait	Bering Sea	< 0.0001 SE 0.0001	0.0126	0.0175 SE 0.0022
	Davis Strait (-CI)	< 0.0001 SE 0.0001	0.0393	< 0.0001 SE 0.0001
	Cunningham In.	< 0.0001 SE 0.0001	0.1083	< 0.0001 SE 0.0001
Bering Sea	Davis Strait (-CI)	< 0.0001 SE 0.0001	0.0371	< 0.0001 SE 0.0001
	Cunningham In.	< 0.0001 SE 0.0001	0.1011	< 0.0001 SE 0.0001

sampling locations within each of these wintering populations. Both the Bering Sea and Hudson Strait populations show no evidence of genetic structure. The lack of genetic structure among the Hudson Strait sample locations is of interest because previous analysis of mitochondrial DNA (mtDNA) variation has revealed genetic structure between the eastern and western Hudson Bay summering populations. Two distinct lineages of mtDNA are found to segregate generally between the eastern and western Hudson Bay populations (Brennin 1992). A large portion of the Hudson Strait wintering population analyzed here was previously analyzed for mtDNA variation (Mancuso 1995, Brown 1996). This subsample was separated based on mtDNA lineages (see appendix 1) and analyzed for genetic structure at the *DRβ1* locus. No significant differences were observed.

Evidence for genetic structure was observed within the Baffin Bay wintering population (Table 5.4 (A)). A pairwise comparison of sample locations within this wintering population reveals that all significant differences exist between the Cunningham Inlet (CI) sampling location and all other sampling locations except Creswell Bay. No genetic structure was apparent after the removal of the Cunningham Inlet samples from this group of sampling locations (Baffin Bay (-CI)). For the subsequent analyses, the Baffin Bay wintering population has been divided into Cunningham Inlet and Baffin Bay (-CI).

--Variation among wintering populations

A pairwise analysis of genetic structure among wintering populations was conducted by comparing allele and genotype frequencies of the sampling locations found within each pair of wintering populations. In all comparisons, evidence for genetic structure was observed (Table 5.4 (B)).

SSCP survey of Western Greenland DQB variation

Samples collected from along western Greenland were analyzed for *DQβ* variation

and compared to the results of the previous survey (Chapter 2). No significant differences from the expected Hardy-Weinberg genotype frequencies were found at any sampling location. Analysis of the allele frequencies of the sampling locations within the Baffin Bay wintering population (Table 5.5) shows evidence of structure ($p = 0.0039$ SE 0.0004). Pairwise comparisons of *DQβ* allele frequencies among the five sampling locations showed significant differences between only the Cunningham Inlet and the western Greenland sampling locations ($p < 0.05$).

Analysis of linkage: DRβ1 - DQβ haplotypes

Analysis of the independence of the *DRβ1* and *DQβ* genotype frequencies (Table 5.6 - 5.9) within beluga wintering populations reveals that in two cases the null hypothesis of independence can be rejected ($p < 0.005$). The genotypes at the *DRβ1* and *DQβ* loci within the St. Lawrence (Table 5.6, $p = 0.0009$) and Baffin Bay (-CI) (Table 5.9, $p = 0.0103$) show that they are not independently assorting. This provides indirect evidence for linkage disequilibrium between these two loci.

It is possible to infer *DRβ1-DQβ* haplotypes from the genotypes in Tables 6 - 9. First, when one locus is homozygous the haplotype conformation is easily determined. From the samples homozygous at the *DQβ* locus for the 0101-2 allele, we know that this allele is found with each of the six *DRβ1* alleles found in the St. Lawrence (Table 5.6). The remaining *DQβ* genotypes are all heterozygotes involving the *DQβ**0101-2 allele and each of the remaining three *DQβ* alleles, therefore we only have to deduce the alternate haplotype. An analysis of the possible *DRβ1-DQβ* haplotypes shows that only three additional *DRβ1-DQβ* haplotypes are needed to most parsimoniously explain the data (Fig. 5.2). It is interesting to note that the first two haplotypes are not the most likely given the allele frequencies. Under an assumption of no linkage disequilibrium, four *DRβ1-DQβ* haplotypes, 0401-0201, 0401-0202, 0101-0201, and 0101-0202, have a higher probability of occurrence.

Table 5.5 **a:** *DQß* genotype frequency in Baffin Bay/High Arctic beluga sampling locations.
b: *DQß* allele frequency in Baffin Bay/High Arctic beluga sampling locations.

	Sampling location				
	Disko B.	Nuussuaq	Grise Fiord	Cunningham	Creswell B.
a: *DQß* * Genotype					
0101-2 / 0101-2	3	3		2	1
0101-2 / 0103	1	3	5	6	
0101-2 / 0202	6	6	7	2	1
0103 / 0103		1	3	3	1
0103 / 0202	2	2	1	4	2
0202 / 0202	3	12	4	1	
Σ	15	27	20	18	5
b: *DQß* * Allele					
*Dele-DQß**0101-2	13	15	12	12	3
*Dele-DQß**0103	3	7	12	16	4
*Dele-DQß**0202	14	32	16	8	3
Σ	30	54	40	36	10

Table 5.6 Frequency of combined *DRβ1-DQβ* genotypes in the St. Lawrence. Genotypes are listed in order of descending probability (listed in italics) based on observed allele frequencies. Independence of *DRβ1* and *DQβ* (p = 0.0009 SE 0.0002) is rejected.

DRβ1		0101	0401	0201	0401	0101	0101	0101	0301	0201	0301	0101	0301
DQβ		0401	0401	0401	0402	0201	0101	0402	0401	0402	0402	0501	0301
	exp. freq.	*.198*	*.163*	*.129*	*.095*	*.078*	*.060*	*.057*	*.052*	*.037*	*.015*	*.005*	*.004*
0101-2/0101-2	*.724*	10	7	3	3	4	1		2		1	1	1
0101-2/0201	*.126*			4		2				1			
0101-2/0202	*.109*				1			3		1	1		
0101-2/0103	*.019*	1											

Table 5.7 Frequency of combined *DRβ1-DQβ* genotypes in the Hudson Strait wintering population. Genotypes are listed in order of descending probability (listed in italics) based on observed allele frequencies. Independence of *DRβ1* and *DQβ* (p = 0.3504 SE 0.0177) is not rejected.

DRβ1		0301	0401	0401	0101	0201	0301	0101	0401	0101	0301	0301	0201	0202	0402	0101
DQβ		0401	0501	0401	0401	0401	0501	0301	0402	0501	0301	0402	0501	0401	0501	0402
	exp. freq.	*.103*	*.076*	*.073*	*.070*	*.054*	*.053*	*.049*	*.049*	*.036*	*.036*	*.034*	*.028*	*.027*	*.025*	*.023*
0101-2/0101-2	*.733*	8	7	5	6	6	1	4	1	1	2	1	5	1	1	1
0101-2/0201	*.127*	3			1						2	1	1	1		
0101-2/0202	*.074*								1		1	1				1
0101-2/0103	*.046*	1			1				1						1	
0201/0202	*.006*			1												
0103/0201												1				

DRβ1		0501	0202	0201	0101	0102	0202	0102	0201	0202	0202	0402	0102	0101	0102
		0501	0301	0402	0101	0401	0501	0301	0202	0202	0402	0402	0501	0102	0402
	exp. freq.	*.020*	*.019*	*.018*	*.017*	*.016*	*.014*	*.011*	*.010*	.010	*.009*	*.008*	*.008*	*.008*	*.005*
0101-2/0101-2	*.733*	2	2	1	2	1	3	1	1	1	1		1	1	1
0101-2/0201	*.127*							1						2	
0101-2/0202	*.074*				1										
0202/0202	*.002*											1			

Table 5.8 Frequency of combined *DRβ1-DQβ* genotypes in the Bering Sea wintering population. Genotypes are listed in order of descending probability (listed in italics) based on observed allele frequencies. Independence of *DRβ1* and *DQβ* (p = 0.2105 SE 0.007) is not rejected.

DRβ1		0401	0101	0202	0101	0101	0201	0301	0501	0201	0101	0301	0201
DQβ		0501	0501	0501	0401	0202	0501	0501	0501	0401	0201	0401	0202
	exp. freq.	*.068*	*.064*	*.060*	*.054*	*.048*	*.044*	*.040*	*.040*	*.037*	*.035*	*.034*	*.033*
0101-2/0101-2	*.847*	3	1	7	1	3	2	1	1		1	1	1
0101-2/0103	*.092*									1	1		2
0101-2/0201	*.051*			1			1	1				1	
0101-2/0202	*.010*	1											

DRβ1		0101	0402	0401	0401	0101	0101	0202	0201	0301	0101	0102
		0301	0501	0401	0402	0101	0402	0202	0301	0402	0102	0102
	exp. freq.	*.032*	*.032*	*.029*	*.027*	*.026*	*.026*	*.023*	*.022*	*.016*	*.010*	*.001*
0101-2/0101-2	*.847*	3	1	2	2	1	1	1				1
0101-2/0103	*.092*					1	1		1	1	1	
0101-2/0201	*.051*				1							
0101-2/0202	*.010*											

Table 5.9 Frequency of combined *DRβ1-DQβ* genotypes in the Baffin Bay (-CI) wintering population. Genotypes are listed in order of descending probability (listed in italics) based on observed allele frequencies. Independence of *DRβ1* and *DQβ* (p = 0.0103 SE 0.0015) is rejected.

DRβ1		0402	0101	0201	0301	0402	0101	0401	0101	0101	0102
DQβ		0402	0402	0402	0402	0501	0201	0402	0301	0501	0402
	exp. freq.	*.130*	*.122*	*.094*	*.079*	*.072*	*.044*	*.043*	*.037*	*.034*	*.029*
0101-2/0202	*.325*	1	2		1	2	3	1	1	1	
0202/0202	*.225*	6		2		2	1				4
0103/0202	*.175*		1	1			1				
0101-2/0103	*.114*		1		1	2		1	2		
0101-2/0101-2	*.106*					1					
0103/0103	*.031*	1	1	1			1				

DRβ1		0201	0101	0201	0202	0301	0101	0101	0301	0201	0202
		0301	0101	0501	0402	0501	0401	0102	0301	0202	0401
	exp. freq.	*.029*	*.029*	*.026*	*.022*	*.022*	*.020*	*.026*	*.012*	*.008*	*.004*
0101-2/0202	*.325*	1			1		1		2		1
0202/0202	*.225*	2					1			1	
0103/0202	*.175*							1			
0101-2/0103	*.114*		2								
0101-2/0101-2	*.106*			1	1	2					1
0103/0103	*.031*										

Figure 5.2 Possible *DRβ1* - *DQβ* haplotypes in the St. Lawrence beluga population. The possible haplotypes for individuals heterozygous at the *DQβ* locus are given. Only haplotypes with the less frequent *DQβ* alleles are shown, a previously observed genotype with a *DQβ**0101 allele is implied. The number to the left of each pair indicates the number of individuals with the required genotype. In seven of the eight cases, the probable genotype is identified by either underlined italics or italics based on similarity. Only three haplotypes are needed to explain the observed variation.

Possible *DRβ1 - DQβ* haplotypes

(1) 0101 - 0103 / 0401 - 0103

(2) 0101 - 0201 / *0201 - 0201*

(4) *0201 - 0201* / 0401 - 0201

(1) *0201 - 0201* / 0402 - 0201

(3) 0101 - 0202 / *0402 - 0202*

(1) 0201 - 0202 / *0402 - 0202*

(1) 0301 - 0202 / *0402 - 0202*

(1) 0401 - 0202 / *0402 - 0202*

Using the same approach, a minimum number of *DRβ1-DQβ* haplotypes was determined for each population (Table 5.10). In addition, assuming all haplotypes were different when possible, a maximum number of haplotypes was determined. Without evidence for linkage disequilibrium the true number of haplotypes lies somewhere between these values. However, as in the case of the St. Lawrence beluga, evidence for linkage disequilibrium indicates that the true value is probably at or near the minimum value. Compared to the three other major beluga populations, the St. Lawrence beluga has a significant reduction in the number of *DRβ1-DQβ* haplotypes (Table 5.10)

Discussion

Comparison of DRβ1 variation among beluga populations

Genetic structure at the *DRβ1* locus within North American beluga was investigated by testing the existing hypotheses about population structure in an hierarchical fashion. The majority of differences in allelic or genotypic distributions could be explained by differences among the wintering populations. No structuring was observed within the sampling locations of the two wintering populations, Hudson Strait and Bering Sea. Only the Baffin Bay wintering population showed evidence of structure. In this case, a single sample location, Cunningham Inlet, was found to be significantly different from all other wintering populations.

The lack of genetic structure at microsatellite loci among sampling locations within the three Arctic wintering populations provides evidence that breeding occurs among individuals from different summer populations in late winter or early spring, when belugas are still on their common wintering grounds (Brown 1996). The results at the *DRβ1* locus lend support to this hypothesis, and indicate that the individuals collected at the sampling locations within two of the wintering populations (Bering Sea and Hudson Strait) are responding in a similar way to a common environment. Due to the possible effects of

Table 5.10 Inferred minimum and maximum number of *DRβ1-DQβ* haplotypes at each wintering population.

Wintering population	Sample size	Independ. of *DRB1/DQB*	Observed Haplotypes	Observed plus Inferred Haplotypes	
				Minimum	Maximum
St Lawrence	47	0.0009	6	9	14
Hudson Strait	92	0.3504	14	17	24
Bering Sea	49	0.2105	8	16	21
Baffin Bay (-CI)	60	0.0103	20	20	23

selection on estimates of genetic distance and migration rates, loci under selection are not ideal for the detection of genetic stocks. Similarities in allele frequencies may be maintained by similar selection pressures in genetically isolated stocks. Alternatively, strong selection at the different locations may reduce the effect of migration acting as an homogenizing force. Models using overdominant selection, however, have shown that only small selection coefficients (≤ 0.02) are needed to explain the variation observed at human Mhc loci (Klein et al. 1993a).

The evidence for genetic structure among the wintering populations is also in agreement with the findings of Brown (1996), except that the microsatellite data does not support the separation of the Baffin Bay (-CI) and Hudson Strait populations at all sampling location comparisons. Brown (1996) did not examine the Cunningham Inlet sampling location. In contrast, the results of comparisons at both the *DRβ1* and the *DQβ* (Chapter 2) loci clearly segregate the Baffin Bay and Hudson Strait sampling locations.

A number of hypotheses may explain the distinct nature of the samples collected from Cunningham Inlet from the rest of the Baffin Bay wintering population. First, the majority of the samples were collected in a single season (1988) and there may be a sample collection bias due to related individuals. Two factors argue against this hypothesis; the genotype frequencies at the sample location are not significantly different from the expected Hardy-Weinberg proportions, and in the four samples collected two years later, a similar high frequency of the *DRβ1**0101 allele is observed. In 1988, 67% of the alleles were *DRβ1**0101 compared to 63% in 1990. This is well above the next two sampling locations, Creswell Bay (3/10, 30%) and S.E. Baffin Island (7/24, 29%), with a high frequency of the *DRβ1**0101 allele. A second hypothesis is that the distinct allele frequencies are being maintained by a strong selective pressure at this summering location. Fidelity to estuarine summer locations has been noted in beluga through behavioural observations (Caron and Smith 1990) and in surveys of mtDNA variation (Brennin 1992, Mancuso 1995, Brown 1996). This hypothesis of a strong selective pressure at a single

summer estuary could be tested by an analysis of microsatellite variation. If this summering population is mating with members of other summering groups at a common wintering ground in southern Baffin Bay, no differences at neutral loci should be observed. Finally, there exists a possibility that the Cunningham Inlet animals are part of a separate wintering population. Although most High Arctic beluga are believed to overwinter in areas of open water along the southwest coast of Greenland around Disko Bay (Doidge and Finley 1993; Finley et al. 1987), areas of ice-free water can be found year round in the "North Water", an area which extends from the mouths of Lancaster and Jones Sounds up into northern Baffin Bay. Aerial surveys of the North Water, in two successive years, found about 500 beluga overwintering in this area. This may represent the overwintering location of the approximately 1500 beluga which summer in Cunningham Inlet.

An analysis of *DQβ* variation among the five Baffin Bay/high Arctic sampling locations above also supports the distinct nature of the Cunningham Inlet location. Pairwise comparisons of *DQβ* allele frequencies among the five sampling locations shows significant differences between only the Cunningham Inlet and the western Greenland sampling locations. Significant differences were not observed between Cunningham Inlet and either Grise Fiord or Creswell Bay. The lack of significant differences between Cunningham Inlet and Creswell Bay at the *DQβ* or *DRβ1* loci may be due to the small sample size of the Creswell Bay sample (n=5). The summering population utilizing Creswell Bay has been estimated at 5000 individuals, and it is unlikely that these animals would be part of the hypothesized North Water overwintering population. It is clear that a larger sample size of this population is needed to fully test any of the above hypotheses.

The possible existence of two distinct genetic stocks in the Canadian high Arctic has serious management implications and needs to be further studied. If two separate stocks are present in the High Arctic, the possibility of overhunting and local extinction of either stock is increased if they are managed as a single unit. To investigate this question the Cunningham Inlet samples should be analyzed with the microsatellite loci employed by

Brown (1996) and the results compared to her findings for the Grise Fiord, Nuussuaq, and Disko Bay sampling locations. Next, the possibility of sample bias should be corrected by the further sampling of summering locations in the high Arctic, i.e. Cunningham Inlet and Creswell Bay.

The St. Lawrence population is also significantly different in all allelic, as well as all genotypic, comparisons among wintering populations. This may reflect the unique nature of the St. Lawrence environment. Although only six of the eight *DRβ1* alleles have been found in this population, these alleles represent each of the five allelic lineages (based on sequence similarity) found in beluga (Chapter 3). Even though the allele frequencies are different, the number and type of alleles are not significantly reduced in comparison to the Arctic populations. At this time the reduction in genetic variation observed at minisatellite loci (Patenaude et al. 1994; Mancuso 1995) did not significantly affect the allele frequencies at the Mhc class II *DRβ1* or *DQβ* loci (Chapter 2). However, the significance of these results must be considered with caution, as the population is still of small size (~500 individuals) and has a low growth rate (Béland et al. 1988; Michaud 1993). Further, a significant reduction in the number of *DRβ1-DQβ* haplotypes has also occurred

DRβ1-DQβ haplotypes

The occurrence of linkage at the *DRβ1* and *DQβ* loci is not surprising since in every mammal species examined (e.g. human, mice, and cattle) these loci are located within a single gene complex (Trowsdale 1995). In humans no recombination events have been observed between the *DRβ* and *DQα* loci with the *DQβ* locus being located upstream of this group (Trowsdale 1995; Fig. 1.1). The organization of the beluga Mhc is unknown, however, the linkage disequilibrium between these loci observed in the St. Lawrence beluga does not appear to be the normal situation for beluga. No evidence for disequilibrium is observed in either the Hudson Strait or Bering Sea populations, and even though the Baffin Bay population shows evidence of linkage, it possesses a large number

of genotypes. Modest amounts of recombination appear to have occurred between these loci. It is most likely that the linkage disequilibrium of the *DRβ1* and *DQβ* loci in the St. Lawrence is related to the reduction in minisatellite variation in this population, presumably caused by the recent genetic bottleneck (Patenaude et al. 1994; Mancuso 1995). Although this reduction in variation did not significantly affect the amount and type of variation found at either the *DQβ* locus (Chapter 2) or the *DRβ1* locus, the number of *DRβ1-DQβ* haplotypes has been reduced (Table 5.10). The linkage disequilibrium observed in the St. Lawrence in combination with the historical data (Reeves and Mitchell 1984) lend support to the hypothesis that the genetic bottleneck was a recent event.

The immediate effect of the reduction in *DRβ1-DQβ* haplotypes is that there is a corresponding reduction in the frequency of occurrence of some *DRβ1-DQβ* genotypes. However, as the amount and type of *DRβ1* and *DQβ* alleles have not been significantly reduced, this situation will be rectified as recombination breaks apart the haplotype bias. It is not known how the disruption of the genetic equilibrium at the Mhc complex will affect the short term responses to foreign pathogens. Certain advantageous genotype combinations may occur less often than predicted by chance and these *DRβ1-DQβ* combinations would have a lower probability of being transmitted together to offspring. We speculate, therefore, that the disruption of the genetic equilibrium in the St. Lawrence beluga may, in the short term, lower the populations' ability to respond to certain pathogens.

A second major cause for concern in the conservation of the St. Lawrence population is how this reduction in haplotype diversity has affected the entire Mhc complex. These data indicate that a minimum of nine chromosomes have survived the population reduction. The question remaining is how other Mhc loci have been affected. Examination of Mhc class I loci would address this question. In addition, the unique situation in the St. Lawrence beluga may allow for the demonstration of the physical linkage of class I and class II loci in cetaceans.

Conclusions

For two of the overwintering populations, Hudson Strait and Bering Sea, no significant differences were found among the allele or genotype frequencies of the sampling locations within the population. In the High Arctic wintering population, a single sampling location was significantly different in allele and genotype frequency from most other sampling locations. These data support the hypothesis that mating is occurring among summering populations at overwintering locations, however, in the High Arctic, more than one overwintering location may exist. Significant differences exist among each of the wintering populations.

The St. Lawrence population has slightly fewer *DRβ1* alleles than found in the other wintering populations, but possesses representatives of each of the five allelic lineages found in beluga. The population has a significant reduction in the number of *DRβ1-DQβ* haplotypes, probably due to the historic bottleneck. This reduction may be of concern in the short term recovery of the population, but in the long term the variation should be regenerated.

Acknowledgments

Samples used in this study were obtained through the co-operation of the Canada Department of Fisheries and Oceans, the Fisheries Joint Management Committee (Inuvik), the Alaska Department of Fish and Game (North Slope Borough, Department of Wildlife Management), the Makivik Corporation, the native communities listed in Table 5.1, Parks Canada, and the St. Lawrence National Institute of Ecotoxicology. Financial and graduate student support were provided by grants from NSERC, NSERC-DFO Science Subvention Program, the Endangered Species Recovery Fund (co-sponsored by the Canadian Wildlife Service of Environment Canada and World Wildlife Fund Canada), and a Northern Studies

Training Program award. We would also like to thank M. Kingsley, N. J. Patenaude, S. Mancuso, J. S. Quinn, J. J. Dockray, J Wang, and P. Wilson for their useful suggestions and help with research, data analysis, and manuscript preparation.

Chapter 6

Preliminary Survey of North Atlantic Right Whale (*Eubalaena glacialis*) Major Histocompatibilty Complex, Class II, *DQβ* Allelic Variation

Abstract

Major Histocompatibility Complex (Mhc) class II *DQβ* variation was assessed in two North Atlantic right whale (*Eubalaena glacialis*) family groups (each consisting of a mother and three calves), three unrelated right whales, and two bowhead whales (*Balaena mysticetus*). Variation was assessed by amplification, via the polymerase chain reaction, of exon 2, which encodes for the peptide binding region, followed by either cloning and DNA sequencing or single-stranded conformation polymorphism analysis. Up to four sequences with similarity to *DQβ* loci were found in each individual, indicating the duplication of this locus in the Balaenidae. Analysis of the family groups allowed the determination of probable *DQβ1-DQβ2* haplotypes. In the 16 right whale chromosomes examined, a minimum of nine haplotypes have been found. The variation among the sequences (i.e. a high ratio of nonsynonymous to synonymous substitutions [0.0858 : 0.0140 respectively, $p < 0.005$], the majority of which are at positions thought to be involved in peptide binding), is similar to that found at expressed Mhc loci. This suggests the functional significance of these loci in the immune response. Comparison of right whale and beluga (*Delphinapterus leucas*) *DQβ* sequences indicates a three fold reduction in the rate of neutral substitution at this Mhc locus in cetaceans.

Introduction

The Major Histocompatibility Complex (Mhc) encodes for cell surface glycoproteins which, by binding foreign peptides and presenting them to T-cells, are instrumental in the initiation of an immune response to foreign pathogens (Klein 1986). Mhc genes appear to be isolated to vertebrates having been identified in most vertebrate orders, except in jawless fish (Trowsdale 1995). High levels of Mhc class I and class II genetic variation are found in most terrestrial mammals. This substantial variation has been proposed to be an adaptation resulting from the large number of pathogens encountered by natural populations (Klein and Takahata 1990), and is thought to be maintained by a form of balancing selection, i.e. overdominant or frequency dependent (Klein et al. 1993a; Hughes and Hughes 1995). A low degree of polymorphism at this complex, as in the case of the African cheetah, has been suggested to be the cause of the increase in the population's susceptibility to pathogens (O'Brien et al. 1985). It has been suggested that Mhc allelic variability should be taken into account in the management of captive breeding programs (Hughes 1991a) and that lack of variation at this locus in an isolated population may increase the chance of extinction of that population (Yuhki and O'Brien 1990).

Mhc molecules are heterodimers made of two similar peptide chains, α and ß. Both of these chains form the Peptide Binding Region (PBR), however, the ß chain shows a large amount of the functional allelic variation expressed in the Mhc class II cell surface glycoproteins of most terrestrial mammals (e.g. Hughes and Nei 1989; Trowsdale 1995). The PBR is believed to be directly involved with the interaction and association of foreign peptides which are subsequently presented to T-cells (Brown et al, 1993; Stern et al 1994). DNA sequence analysis of the second exon of a ß chain locus allows for the characterisation of the amino acids which make up the ß chain contribution to the PBR. The type of selection (i.e. positive [directional], negative [balancing], or neutral) that a Mhc locus has encountered may be deduced by analyzing, directly, substitutions at the PBR.

Few studies have been conducted on Mhc loci in cetaceans (Trowsdale et al. 1989, Chapters 2-5). These studies indicate either low (Trowsdale et al. 1989; Chapter 2) or modest levels of variation (Chapter 3). Slade (1992) also found low levels of Mhc variation in southern elephant seals (*Mirounga leonina*) and hypothesized that the general lack of variation in marine mammals may be due to a reduction in the balancing selection in the marine environment. This hypothesis, however, does not seem likely given the evidence of directional selection and the high ratio of nonsynonymous to synonymous substitutions observed at the beluga (*Delphinapterus leucas*) class II loci (Chapters 2 and 3) which indicate a similar balancing selection pressure in cetaceans and terrestrial mammals. Similar evidence has also been reported in a preliminary report on European harbour seal (*Phoca vitulina*) Mhc variation (Goodman and Slade 1995).

The objective of this preliminary survey is to describe the Mhc class II *DQβ* variation found in a small sample of North Atlantic right whales (*Eubalaena glacialis*). North Atlantic right whales were officially protected by the International Whaling Commission in 1946, after centuries of commercial whaling which severely depleted the population size (Aguilar 1986; Reeves and Mitchell 1986a, 1986b). This population is currently recognized as one of the most endangered of all cetaceans, with population estimates of about 350 individuals (NMFS 1991). Further, population studies indicate that the reproductive rate of the population is only about 2%, which is lower than the rate of about 7% observed in the South Atlantic right whale (*Eubalaena australis*) (Knowlton et al. 1994). Information on Mhc diversity may lead to insights on the current status of this populations, and its long term chances of survival.

Right whales have distinctive callosity patterns, morphological features, and scars which have enabled researchers to distinguish individuals based on a series of photographs taken in the field (Kraus et al. 1986). This has allowed for the creation of a catalogue of known individuals, and through long term association of mothers and calves, for the determination of maternal family relationships (Crone and Kraus 1990). The collection of

skin biopsies through the darting of known individuals (Brown et al. 1991a) has also enabled the correlation of the behaviour data to genetic analyses of sex (Brown et al. 1991b; 1993), mitochondrial DNA variation (Schaeff et al. 1993), and minisatellite variation (Schaeff 1993). This data set is unique among investigations of cetaceans and will allow for the pattern of inheritance of Mhc variation to be assessed.

This preliminary study of *DQβ* is the first step in the analysis of Mhc variation in the North Atlantic right whale. Analysis of maternal family groups will allow for the determination of Mhc haplotypes and for the detection of linkage disequilibrium between loci. Analysis of unrelated individuals will assess the level of population variation. Finally, comparative studies of Mhc variation in right whale with other cetaceans, i.e. beluga and the closely related bowhead whale (*Balaena mysticetus*), will allow the evolution of this complex in cetaceans to be studied.

Materials and Methods

Samples

Ten right whale samples, collected via biopsy darting of photo-identified animals, from the Northern Atlantic population, and two bowhead whales from a Baffin Bay population were used to investigate allelic variation of the Mhc *DQβ* loci in these cetacean species (Table 6.1). Two right whale family groups were analyzed. Family groups were chosen because they differ in their mitochondrial DNA matrilines (Schaeff et al. 1993). Each family group consists of a mother (listed first) and three offspring (indented) (Table 6.1).

Sequencing of DQβ-like alleles

DNA sequencing of the PBR of this Mhc gene was carried out by enzymatic amplification of the region in question followed by cloning and sequencing of the resulting

Table 6.1 Right and Bowhead whale samples, corresponding New England Aquarium number (if applicable), sex of the individual (if known), number of clones analyzed and number of *DQß* sequences detected.

Sample	NEA #	Sex	Number of clones analyzed	Number of *DQß* sequences
Right Whales				
Family 1				
Eugl-130	*1135*	F	63	4
Eugl-134	*1406*	F	56	3
Eugl-136	*2135*	M	50	2
Eugl-068	*1163*	F	51	2
Family 2				
Eugl-060	*1171*	F	48	2
Eugl-059	*1971*	M	30	2
Eugl-122	*1170*	M	30	3
Eugl-150	*2271*	M	48	2
Unrelated				
Eugl-020	*1510*	M	8	3
Eugl-056	*1242*	F	22	4
Eugl-070	*1613*	M	12	4
Bowhead Whales				
Bamy-002	na	?	16	3
Bamy-003	na	?	2	1

products. The *DQß* PBR was amplified using the primers (CTG GTA GTT GTG TCT GCA CAC) and (CAT GTG CTA CTT CAC CAA CGG) (Tsuji et al. 1992). The reaction conditions for the enzymatic reaction were 10 mM Tris-HCl (pH-8.3), 50 mM KCl, 2.0 mM $MgCl_2$, 0.2 mM dNTP's, 0.2 µM of each primer, 1 unit of Taq DNA polymerase (Perkin-Elmer-Cetus), and 50-100 ng of template DNA carried out in a 50 µl volume. Thermal cycling was conducted on a Perkin-Elmer-Cetus model 480 and involved 30 cycles of 94 °C for 1 min., 53 °C for 1 min., and 72 °C for 2 min.. The products of this reaction were isolated from a 1.5% agarose gel using the GlassMAX (BRL) system, cloned into the pGEM^T vector system (Promega), and transformed into DH5 alpha (BRL) competent *Escherichia coli* cells.

Single Stranded Conformation Polymorphism (SSCP) analysis was used to analyze the cloned PBR sequences. SSCP analysis was conducted via Polymerase Chain Reaction (PCR) amplification (concentrations as above except reactions were carried out in a 10 µl volume) of the cloned PBR and of right whale samples (See Hayashi 1992 for review). To ease analysis, only one primer was radioactively end labelled with γ^{33}P-dATP (ICN). PCR products were electrophoresed through a nondenaturing acrylamide gel (5% acrylamide (59 acrylamide: 1 bisacrylamide), 10% glycerol and 1/2 TBE) for 13 hours at room temperature, which was subsequently dried onto Whatman filter paper. The SSCP's were visualized by exposure of the dried gel to PhosPhor Image screens (Molecular Dynamics). A PhosphoImager (Molecular Dynamics) was used to analyze the audioradiograph images captured on the Phosphor Image screens. Clones with variant SSCP patterns were chosen for sequencing. Clones were sequenced via PCR cycle-sequencing utilizing fluorescently labelled terminators, and the sequences were determined using the ABI 373A Automated Sequencer (Applied Biosystems, Inc)

At least two identical clones per unique exon 2 sequence were obtained to identify a tentative nucleotide sequence. Exon 2 sequences were confirmed by the presence of identical sequences in separate individuals. In this way, DNA sequence differences arising

during the PCR could be detected. Nomenclature of the alleles is based on the proposed rules for nomenclature of the Mhc of different species (Klein et al. 1990) and are based on sequence similarity. Standard nomenclature is a four letter species code (i.e. *Eugl* and *Bamy* for right and bowhead whales respectively), a locus code, an *, and a four digit allele code, e.g. *Eugl-DQß**0101. The first two numbers represent the allelic lineage while the last two designate the unique sequence.

Analysis of variation

The relationship of the alleles was estimated using the computer package Phylip 3.5c (Felsenstein 1993). Pairwise genetic distance measures were generated using the program DNAdist. This program estimated distance based on Kimura's two-parameter model (Kimura 1980). A distance tree was constructed from the distance matrix using the Neighbor-Joining program which is based on the neighbor-joining method of Saitou and Nei (1987). In order to test the significance of the branches, 1000 bootstrap replicates were conducted (Felsenstein 1985).

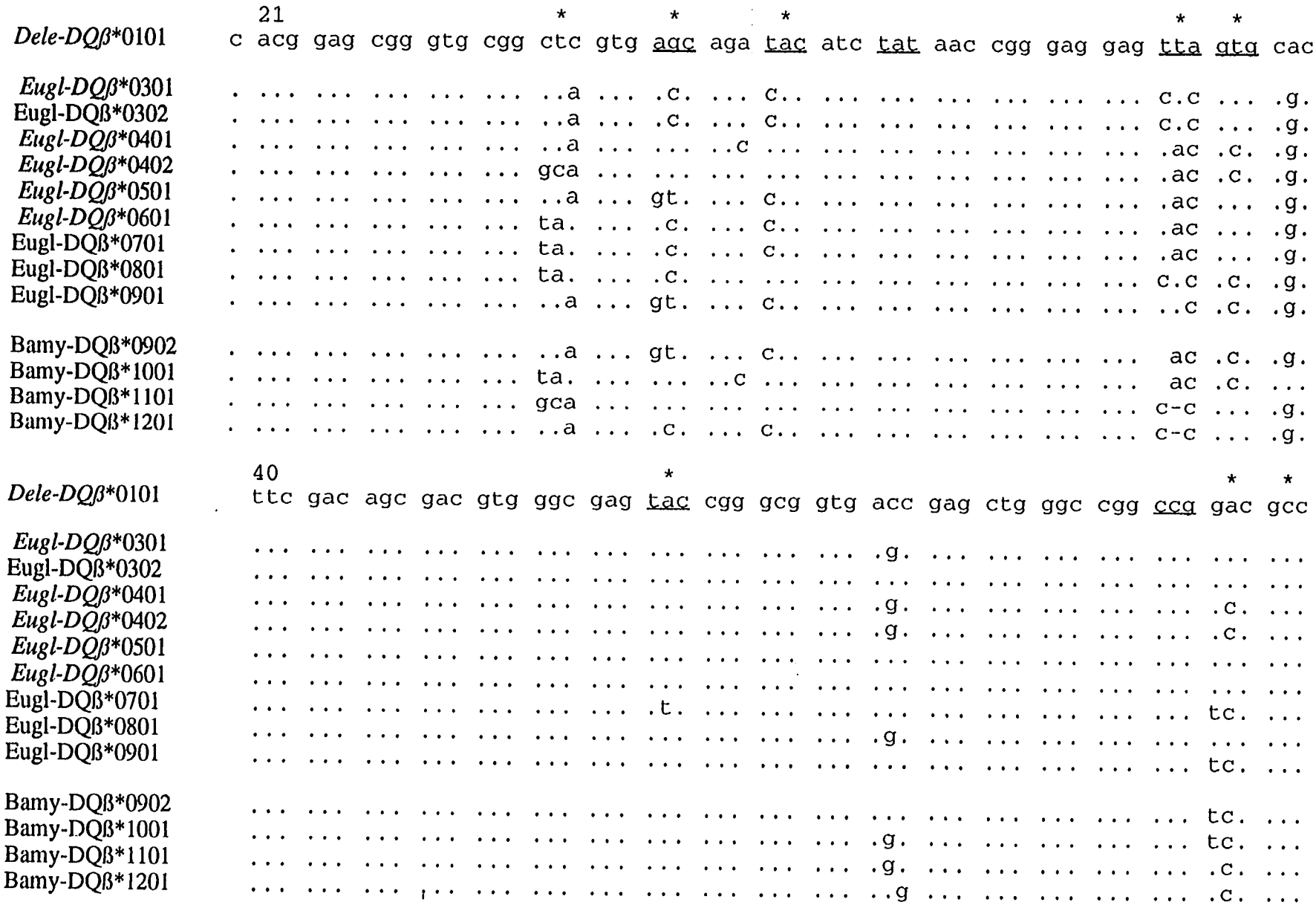
Pairwise comparisons of nucleotide substitutions between alleles were conducted according to the method of Nei and Gojobori (1986) by the computer program MEGA (Kumar et al. 1993). The number of nonsynonymous (d_n) and synonymous (d_s) substitutions per site were estimated for each pair from the Jukes-Cantor formula, and the mean d_n and d_s values compared with a Student *t-test* with infinite degrees of freedom (Kumar et al. 1993).

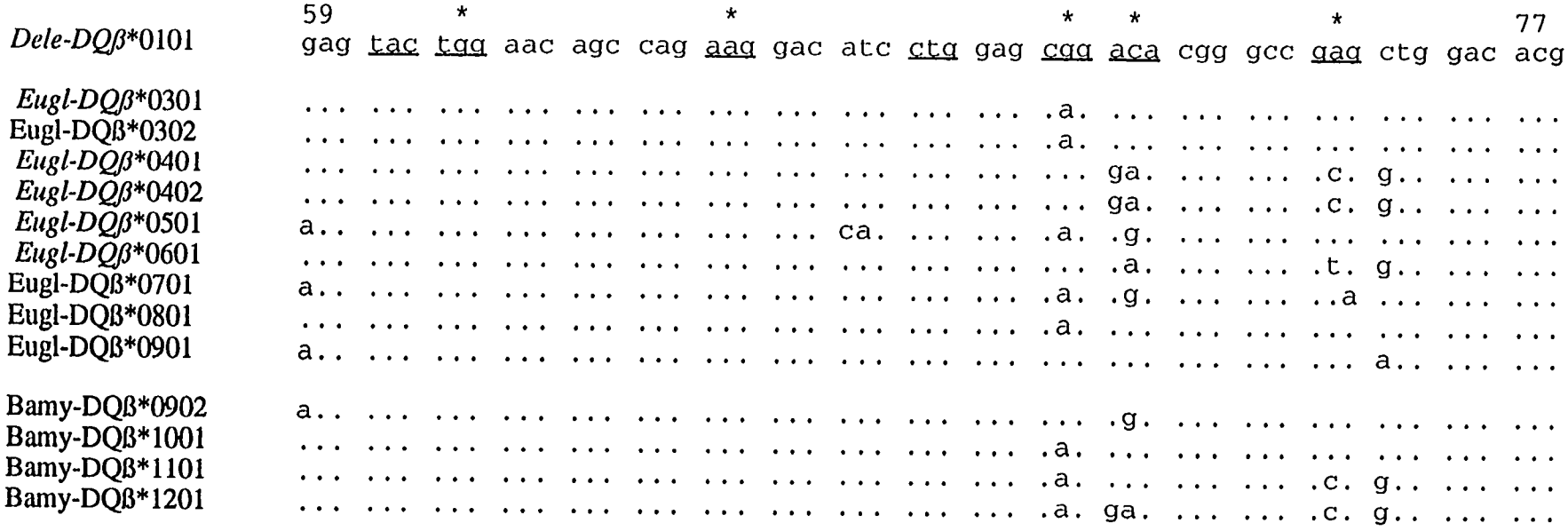
Results

Allelic variation in right and bowhead whales

Thirteen sequences (nine from right and four from bowhead whales) with similarity to the exon 2 sequences of other *DQß* loci in cetaceans (Chapter 2) were identified (Fig. 6.1).

Figure 6.1 Nucleotide sequence of the right and bowhead *DQβ* exon 2 sequences. A '.' indicates an identical base pair with reference to *Dele-DQβ1**0101. Exon 2 amino acid positions are given for reference. Codons marked by underlining or with an asterisk indicate the encoded amino acids either face the peptide binding groove or may be part of a peptide binding pocket respectively in the human *DR* molecule (Brown et al. 1993; Stern et al. 1994). Tentative sequences are given in normal text while sequences present in more than one individual are given in italics. This form is followed throughout the text.


21
*
*
*
*
*
Dele-DQß*0101
c acg gag cgg gtg cgg ctc gtg agc aga tac atc tat aac cgg gag gag tta gtg cac
Eugl-DQß*0301
Eugl-DQß*0302
Eugl-DQß*0401
Eugl-DQß*0402
Eugl-DQß*0501
Eugl-DQß*0601
Eugl-DQß*0701
Eugl-DQß*0801
Eugl-DQß*0901
Bamy-DQß*0902
Bamy-DQß*1001
Bamy-DQß*1101
Bamy-DQß*1201
40
*
*
*
Dele-DQß*0101
ttc gac agc gac gtg ggc gag tac cgg gcg gtg acc gag ctg ggc cgg ccg gac gcc
Eugl-DQß*0301
Eugl-DQß*0302
Eugl-DQß*0401
Eugl-DQß*0402
Eugl-DQß*0501
Eugl-DQß*0601
Eugl-DQß*0701
Eugl-DQß*0801
Eugl-DQß*0901
Bamy-DQß*0902
Bamy-DQß*1001
Bamy-DQß*1101
Bamy-DQß*1201


59 * * * * * 77
Dele-DQß*0101 gag tac tgg aac agc cag aag gac atc ctg gag cgg aca cgg gcc gag ctg gac acg
Eugl-DQß*0301 ... ... .a. ... ...
Eugl-DQß*0302 ... ... .a. ... ...
Eugl-DQß*0401 ... ... ga. ... ... .c. g.. ... ...
Eugl-DQß*0402 ... ... ga. ... ... .c. g.. ... ...
Eugl-DQß*0501 a.. ... ... ca. ... ... .a. .g. ... ...
Eugl-DQß*0601 ... ... .a. ... ... .t. g.. ... ...
Eugl-DQß*0701 a.. ... ... .a. .g. ... .a ... ...
Eugl-DQß*0801 ... ... .a. ... ...
Eugl-DQß*0901 a.. ... ... a.. ... ...
Bamy-DQß*0902 a.. ... ... .g. ... ...
Bamy-DQß*1001 ... ... .a. ... ...
Bamy-DQß*1101 ... ... .a. ... ... .c. g.. ... ...
Bamy-DQß*1201 ... ... .a. ga. ... ... .c. g.. ... ...

A nonredundant BlastN search of GENBANK (Oct. 16/96) found each of the sequences listed here to have the highest identity with previously reported beluga *DQβ* sequences (Chapter 2). Relationship amongst the sequences, based on genetic distance and bootstrap support, was used to establish the allelic lineages and thus the nomenclature for the sequences (Fig. 6.2). Confirmed sequences are given in italics (i.e. present in more than one animal) while tentative sequences are listed in normal text. Between two to four sequences were found in each right whale sample (Table 6.1). As the locus identity of the sequences is only tentative at present (see below), the *DQβ*-like locus for each allele was not included in the nomenclature.

Due to the large number of clones examined for each member of the two family groups (Table 6.1) and the consistency of the sequences found in each individual (Table 6.2) with the SSCP phenotypes of the family members (Fig. 6.3), it is likely that all the *DQβ* sequences amplified in these samples were determined. Based on the SSCP phenotypes, samples Eugl-068 and Eugl-136 and samples Eugl-060 and Eugl-059 should contain identical sequences respectively (Fig. 6.3). This was confirmed by analysis of cloned sequences (Table 6.2). Further, new conformations are observed in samples Eugl-150 and Eugl-122 (Fig. 6.3), which represent sequences distinct to those individuals (Table 6.2). All single stranded conformations of Family group 2 are consistent with the SSCP phenotypes of the cloned sequences (comparison not shown). A similar statement cannot be made for Family group 1, due to the similarity of the single stranded conformation of the cloned alleles (Fig. 6.3). It is not known if all *DQβ* sequences were identified for samples Eugl-020 and Eugl-070. A further examination of cloned sequences needs to be conducted.

Pairwise comparison of nonsynonymous to synonymous substitutions on a per site basis reveals evidence for positive Darwinian evolution of these sequences in both right and bowhead whales. In a pairwise comparison of all bowhead whale sequences, the average number of nonsynonymous substitutions per site (d_n = 0.1010 SEM 0.0214) is

Figure 6.2 Neighbor-Joining radial dendrogram of cetacean *DQβ* sequences. Bootstrap values greater than 50% are placed along the branches.

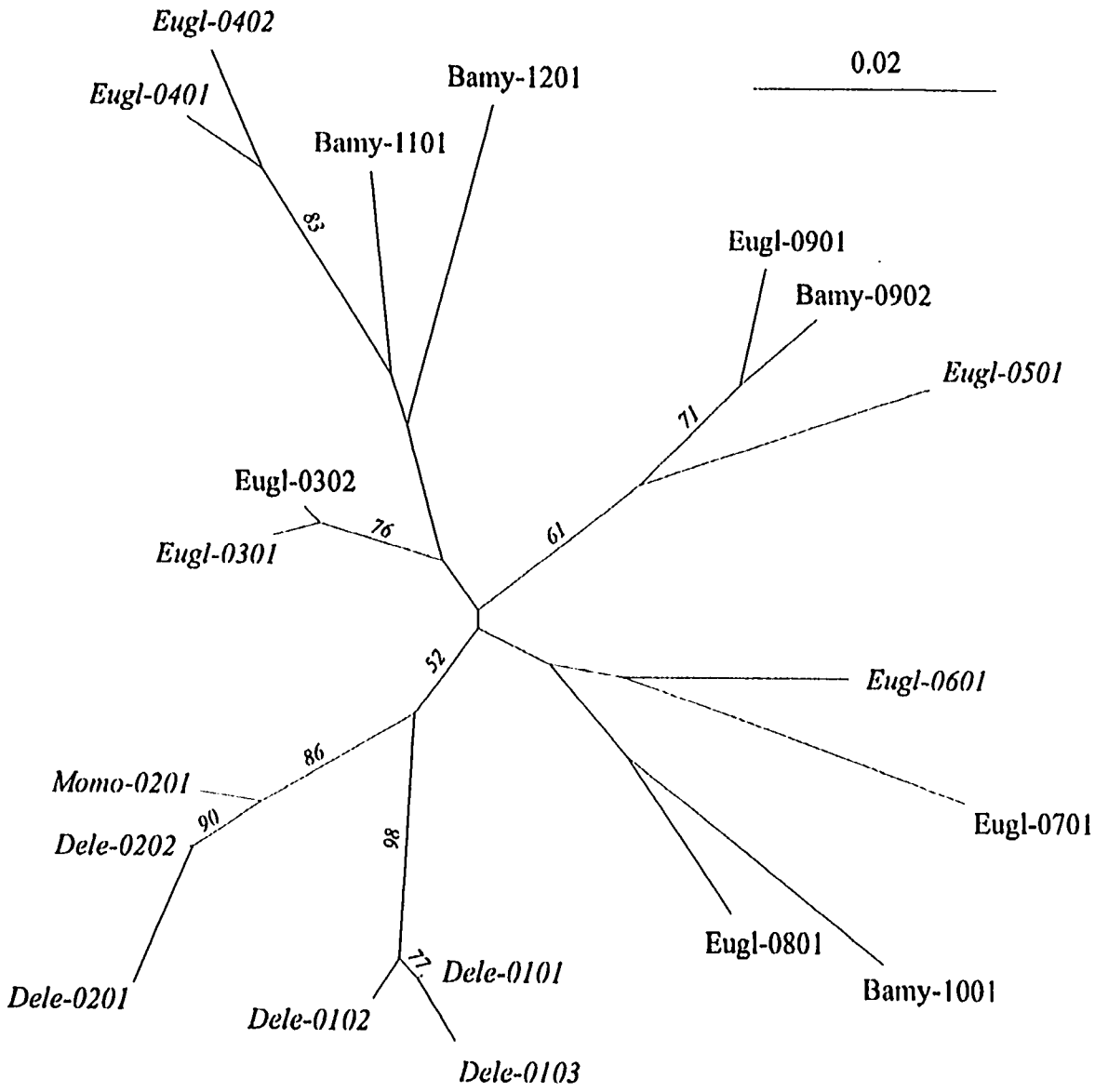

Eugl-0402
Eugl-0401
Bamy-1201
Bamy-1101
0.02
83
Eugl-0901
Bamy-0902
Eugl-0501
71
Eugl-0302
76
Eugl-0301
61
52
Eugl-0601
86
Momo-0201
90
Dele-0202
98
Eugl-0701
Eugl-0801
Bamy-1001
77
Dele-0101
Dele-0201
Dele-0102
Dele-0103

Table 6.2 *DQß*-like sequences found in each unrelated right whale and two family groups.

Right whale samples	*DQß*-like sequences								
	Possible *DQß1* alleles			Possible *DQß2* alleles					
	0301	0302	0501	0401	0402	0601	0701	0801	0901
Family 1									
Eugl-130	*		*	*		*			
Eugl-134	*		*			*			
Eugl-136			*	*					
Eugl-068			*	*					
Family 2									
Eugl-060			*		*				
Eugl-059			*		*				
Eugl-122			*		*	*			
Eugl-150			*					*	
unrelated									
Eugl-020		*	*	*					
Eugl-056	*		*	*			*		
Eugl-070	*		*		*				*

Figure 6.3 *DQβ* PCR-SSCP phenotypes of two right whale family groups from two separate audioradiographs. Right whale samples (Eugl) are listed in blue (or red) with the mother of each group at the upper right of the vertical line. Cloned sequences are listed in black and prefixed by DQB*.

Family 1

Family 2

significantly greater ($p < 0.005$) than the average number of synonymous substitutions per site ($d_S = 0.0164$ SEM 0.0147). A significant difference ($p < 0.005$) between the average number of nonsynonymous substitutions per site ($d_n = 0.0858$ SEM 0.0153) and the average number of synonymous substitutions per site ($d_S = 0.0140$ SEM 0.0115) is also found in a pairwise comparison of right whale sequences. In addition, analyses of the ratio of nonsynonymous to synonymous substitutions among the alleles of the putative *DQβ1* (d_n =0.0491 SEM 0.0166: $d_S = 0$) and *DQβ2* ($d_n = 0.0912$ SEM 0.0178:$d_S = 0.0183$ SEM 0.0145, $p < 0.005$) loci also shows evidence for positive selection.

Finally, the majority of amino acid substitutions in both right and bowhead whales do not maintain the physio-chemical properties of the residue (Fig. 6.4). Of the 15 variable positions found among right whale sequences, 10 contain amino acid substitutions that do not strongly, intermediately or weakly maintain the physio-chemical properties of the residue. A similar trend is observed among the bowhead whales sequences (9 of 14 positions). It is interesting that the vast majority (17 of 23) of amino acid substitutions observed among the right whale sequences are also observed among the sequences found in a much smaller sample of bowhead whales (Fig. 6.4).

Haplotype variation in right whales

The distribution of *DQβ*-like sequences within the two right whale family groups and the unrelated whales may allow for the determination of the *DQβ* haplotypes found in this population (Table 6.2). First, as only up to four sequences were observed in the samples, it was assumed that only two *DQβ* loci are present; *DQβ1* and *DQβ2*. Sequence *Eugl*0501* is found in every right whale sample. To begin, this high frequency sequence was assigned as an allele of the *DQβ1* locus. Next, *Eugl*0301* appears to be a second, lower frequency allele of the *DQβ1* locus. Only this sequence and *Eugl*0501* are common to every sample with four sequences. From these starting assumptions, the possible *DQβ1--DQβ2* haplotype can be deduced (Fig. 6.5). For each calf, the maternal haplotype

Figure 6.4 Amino acid sequence, based on nucleotide sequence (Fig. 6.1), of the right and bowhead whale *DQß* sequences. Positions facing the peptide binding groove are underlined. Positions implicated in the four polymorphic clusters of amino acids on the ß chain thought to each form variable pockets responsible for peptide side chain binding are shown by an * (Brown et al. 1993; Stern et al. 1994). For variable positions, the conservation of physio-chemical properties (i.e. polar or nonpolar, size, shape, and charge) of the residue is indicated by *s*, *m*, and *w* (strong, medium, and weak conservation respectively, McLachlan 1972). For positions with multiple substitutions, the least conservative change is given.

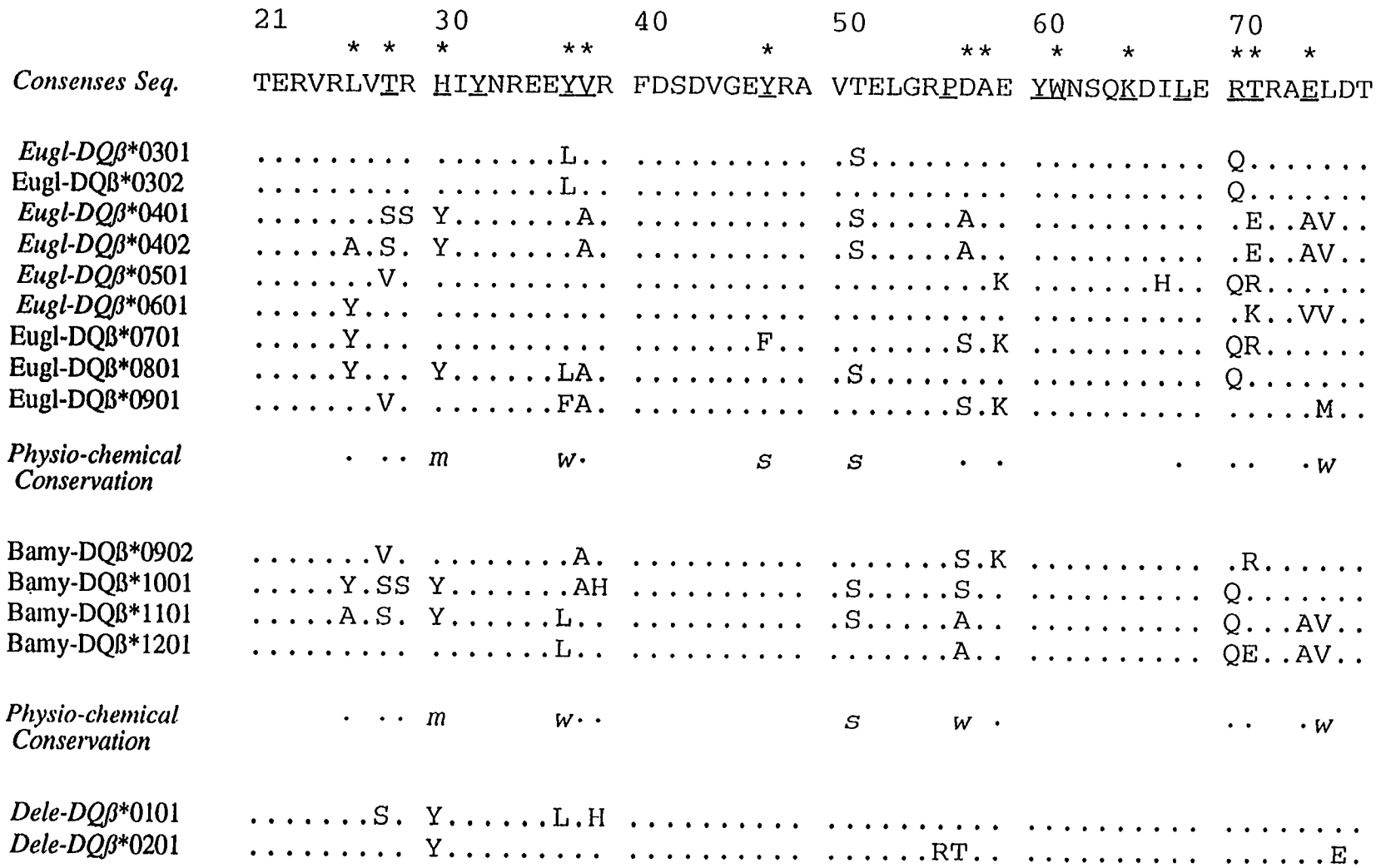

21 30 40 50 60 70
Consenses Seq. TERVRLVTR HIYNREEYVR FDSDVGEYRA VTELGRPDAE YWNSQKDILE RTRAELDT
Eugl-DQß*0301 .....L.. .S..... Q.....
Eugl-DQß*0302 .....L.. Q.....
Eugl-DQß*0401 .....SS Y.....A. .S.....A.. .E..AV..
Eugl-DQß*0402 .....A.S. Y.....A. .S.....A.. .E..AV..
Eugl-DQß*0501 .....V. .....K .....H.. QR.....
Eugl-DQß*0601 .....Y... .K..VV..
Eugl-DQß*0701 .....Y... .....F.. .....S.K QR.....
Eugl-DQß*0801 .....Y... Y.....LA. .S..... Q.....
Eugl-DQß*0901 .....V. .....FA. .....S.K .....M..
Physio-chemical Conservation
Bamy-DQß*0902 .....V. .....A. .....S.K .R.....
Bamy-DQß*1001 .....Y.SS Y.....AH .S.....S.. Q.....
Bamy-DQß*1101 .....A.S. Y.....L.. .S.....A.. Q...AV..
Bamy-DQß*1201 .....L.. .....A.. QE..AV..
Physio-chemical Conservation
Dele-DQß*0101 .....S. Y.....L.H
Dele-DQß*0201 Y..... .....RT.. .....E.

Figure 6.5 Possible right whale *DQß1--DQß2* haplotypes. Haplotypes are listed above and below each vertical solid line and identified with superscripts. The right whale sample is given to the right of each haplotype pair. The family groups are arranged vertically. In the calves, the probable maternal haplotype is listed above the solid vertical line. The possible haplotypes of three unrelated individuals is given below the family groups. A '?' indicates either the sequence is identical to the maternal haplotype, the haplotype does not contain a *DQß2* allele, or the presence of a non-amplified allele.

Possible *--DQβ1--DQβ2--* Haplotypes

	Family 1		Family 2	
Mother	--0501--0401--[1] / --0301--0601--[2]	Eugl-130	--0501--0402--[4] / --0501--?--[3]	Eugl-060
Calves	--0301--0601--[2] / --0501--?--[3]	Eugl-134	--0501--0402--[4] / --0501--?--[3]	Eugl-056
	--0501--0401--[1] / --0501--?--[3]	Eugl-136	--0501--0402--[4] / --0501--0601--[5]	Eugl-122
	--0501--0401--[1] / --0501--?--[3]	Eugl-068	--0501--?--[3] / --0501--0801--[6]	Eugl-150
Unrelated whales	--0501--0401--[1] / --0301--0701--[7]	Eugl-056		
			--0501--0401--[1] / --0302--?-- or --0501--0302--[8]	Eugl-020
	--0501--0402--[4] / --0301--0901--[9]	Eugl-070		

is listed above the solid line while the paternal haplotype is listed below. If the paternal haplotype contains a question mark at the *DQβ2* loci, then the allele or identity of this locus is unknown. It is either an allele that does not amplify with the given primers, a haplotype that does not contain a *DQβ2* locus, or a haplotype that has a *DQβ2* allele identical to the maternal allele. None of these possibilities can be ruled out, however, analysis of Family 2 supports the first two, as calf Eugl-150 must have received from his mother either a haplotype with a 'null' allele or one without a *DQβ2* locus .

These data can also be used to establish the minimum number of males responsible for the offspring. Although the first family is uninformative in this regard, Family group 2 indicates that at least two males must have mated with Eugl-060. This is the first demonstration of male contribution to progeny, and suggests a polygamous mating system.

Based on the deduced family group haplotypes, a minimum number of new haplotypes in the unrelated right whales can be deduced (Fig. 6.5). In total, 16 chromosomes were sampled, two maternal and three paternal for each of the two family groups, and six in the unrelated whales. A minimum of nine haplotypes are needed to explain the distribution of sequences within animals.

Discussion

The north Atlantic right whale has a modest amount of *DQβ* allelic variation at least equal to, and possibly greater than, the variation observed at the beluga *DRβ1* locus. Only a small number of chromosomes were examined in this survey and the presence of null alleles may indicate that *DQβ* sequences are present but are not being amplified with the current protocol. Analysis of the right whale *DQβ* sequences reveals the role of positive selection in their evolution and supports the assumption of the functional significance of this class of molecule in the immune response. Further, the ratio of nonsynonymous to synonymous substitutions (6:1) is similar to that observed in primates (Klein et al. 1993a).

This evidence is the same as that observed at the beluga *DRβ1* locus (Chapter 3) indicating that a reduction in balancing selection pressure, when compared to terrestrial mammals, as proposed by Slade (1992), is not likely.

Comparison of the beluga and right whale *DQβ* sequences allows for the estimation of the neutral substitution rate in cetacean Mhc. Synonymous substitutions are selectively neutral and are directly related to mutation rate. The average number of synonymous substitutions between the beluga sequences *Dele-DQβ**0101 and *Dele-DQβ**0201 and the right whale sequences *Eugl-DQβ**0301, *Eugl-DQβ**0401, *Eugl-DQβ**0501, and *Eugl-DQβ**0601 is 1.25 (range 0-2). Using the synonymous substitution rate of 1.2×10^{-9}/site/year estimated from primate sequences (Klein et al. 1993a), an expected rate of 0.5×10^{-7} synonymous substitutions/*DQβ* exon 2 sequences/year is calculated. Using this rate, an estimate of approximately 25 Million years (Myr) of evolution, or 12.5 Myr, separates the beluga and right *DQβ* sequences. However, based on fossil evidence, right and beluga whales have been separated, at least, since the split of toothed and baleen whales in the late Eocene, about 35-40 Myr (Fordyce and Barnes 1994). This shows the rate of neutral evolution in cetaceans is 3x slower than in terrestrial mammals. A reduction in the neutral substitution rate in cetaceans is supported by studies of mitochondrial DNA in cetacean, which also suggest a similar 4x to 2x reduction in the rate of neutral nucleotide substitution in cetacean D-loop sequences, when compared to terrestrial mammals (Hoelzel et al. 1991; Baker et al. 1993).

Using the synonymous substitution rate calculated for cetacean, an average of 34 Myr separating the right whale *DQβ* sequences can be estimated. This time indicates an average age of the sequence lineage to be 17 Myr, and means that observed variation has evolved after the split of the Balaenidae from the other mysticete taxa (Barnes et al. 1985). This date would predict that the right and bowhead whales should share a large number of allelic lineages. This study shows one pair of very similar *DQβ* sequences, Eugl-DQß*0901 and Bamy-DQß*0902, between these species and a number of interrelated

sequences. It also predicts that a number of *DQβ* allelic lineages should be shared with the more distantly related pigmy right whale (*Caperea marginata*).

A reduced rate of neutral substitution at the Mhc *DQβ* locus in cetaceans when compared to primates is significant in the interpretation of the variation at other loci (see Chapter 3). It is probable that other Mhc loci, i.e. *DRβ*, would have a similar rate of evolution, especially given the evidence of linkage disequalibrium (Chapter 5). Further, this finding questions the interpretation of Trowsdale et al. (1989) who indirectly quantified the amount of Mhc variation through restriction fragment analysis and found low amounts of variation in a small sample of fin (*Balaenoptera physalus*) and sei (*B. borealis*) whales. Trowsdale et al. (1989) did find some *DRβ* and class I variation, which was low when compared with the amount of variation found in man or mice, however, it may be consistent with the moderate levels of variation found at the beluga *DRβ1* and the right whale *DQβ* loci. Trowsdale et al. (1989) used only one, or with some probes, two, restriction enzymes to sample the variation. Given the low number of synonymous substitutions observed among the Mhc exon 2 sequences of right or beluga whales (Chapters 2 and 3), and the possible reduced rate of neutral substitution, it is unlikely that much variation would have been detected in the present or the previous studies (Chapters 2 and 3) if we had utilized this technique. It is not clear whether the fin and sei whale have a similar or lower amount of variation when compared with right and beluga whales. A direct analysis of PBR encoding sequences is needed to address this question.

Conclusions and Further Work

Two main findings are clear from the results of the preliminary survey. The right whales have a moderate and possibly a high amount of *DQβ* variation. Second, the neutral substitution rate may be up to 3x lower in cetaceans than terrestrial mammals. The reduced rate of substitution may also explain the low amount of variation observed in other

cetaceans (Trowsdale et al. 1989) when compared with terrestrial mammals.

Further studies are needed to improve the current data and confirm the results. First, additional right whale family groups and unrelated animals are needed to confirm the haplotype structure and to quantify the amount of variation in the North Atlantic right whale. Second, alternate *DQβ* primer sets should be tested in order to establish the reason for the 'null' alleles. Third, additional samples of bowhead whales and South Atlantic right whales (*E. australis*) should be analyzed so that comparative studies between these closely related groups can be carried out, and the possible reduction of North Atlantic right whale Mhc variation be assessed. Finally, analysis of more distantly related mysticetes, i.e. pigmy right, sei and fin whales, would be of use in the understanding of the evolution of the Mhc in cetaceans.

Acknowledgments

Samples used in this study were obtained through the efforts of the New England Aqurium. Behavioural data was kindly provided by Moria Brown and Scott Kraus. Financial and graduate student support were provided by grants from NSERC and the Endangered Species Recovery Fund (co-sponsored by the Canadian Wildlife Service of Environment Canada and World Wildlife Fund Canada). We would also like to thank J. S. Quinn, J. J. Dockray, S. Malik, J. Wang, P. Wilson, and R. Waldick for their useful suggestions and help with research, data analysis, and manuscript preparation.

Chapter 7

Concluding Remarks

The previous chapters set out to achieve two major goals: 1) To assess the role of selection in the evolution of cetacean Major Histocompatibility complex (Mhc) loci by comparing the levels and type of variation found in a large survey of beluga (*Delphinapterus leucas*), and in smaller surveys of narwhal (*Monodon monoceros*), right (*Eubalaena glacialis*) and bowhead (*Balaena mysticetus*) whales; 2) To assess levels of Mhc variation in the small and endangered St. Lawrence beluga population, and to compare that level of variation to beluga populations from throughout North America. This chapter is a synthesis of work conducted and suggestions for further work required to achieve these goals.

Evolution of the Mhc in cetaceans

Seventy to fifty five million years (Myr) ago, in the warm shallow waters of the Tethys Sea, mammals related to ungulates are thought have begun one of the most successful recolonizations of the marine environment (Gaskin 1982; Barnes et al. 1985; Evans 1987; Fordyce and Barnes 1994). Members of the two modern suborders of Odontoceti and Mysticeti first appear in the fossil record in the early Oligocene, and today are represented by 75-77 extant species (Fordyce and Barnes 1994). These modern cetaceans illustrate the large amount of morphological adaptations, including the development of echolocation and baleen, that have allowed this group to flourish in the marine environment. It is not known how this adaptive radiation to a dramatically different environment has affected cellular processes such as the immune system. It is probable that

earliest cetaceans were faced with a new range of pathogens associated with the marine environment. The study of the evolution of the Mhc in cetaceans presents an exciting opportunity to observe the response of the Mhc to the new pressures of the marine environment.

The examination of Mhc class II variation in beluga (Chapters 2 and 3) and right whales (Chapter 6) shows that the exon 2 sequences have a significantly higher ratio of nonsynonymous to synonymous substitutions per site and that the majority of these substitutions occur at positions important in peptide binding. These results are similar to those observed among functional alleles at primate Mhc loci (e.g. Hughes and Nei 1988, 1989), indicating the role of positive selection in the evolution of the Mhc sequences. This provides the first evidence that the cetacean Mhc loci have the same functional significance in the immune response of whales (Chapters 2, 3 and 6).

The next step should be to establish which loci are being expressed. For example, in beluga, it is important to determine if products from both the polymorphic *DRβ1* and the monomorphic *DRβ2* loci are being equally expressed. If the hypothesis that duplication events allow for the allelic diversification of one of the loci (Chapters 3 and 6) is correct, this would predict that both loci are expressed. Further, expression will provide further evidence of the functional significance of Mhc loci. Unfortunately, this type of analysis will be limited in cetaceans. For the analysis of the mRNA of class II molecules, fresh lymphocytes are needed. This tissue is accessible in cetaceans such as beluga which are hunted, captured live for other scientific purposes, and are maintained in captivity. In contrast, samples from large cetaceans, such as right whales, would depend on opportune analysis of fresh carcasses and are very hard to acquire.

Lower amounts of Mhc variation are found in beluga and possibly right whales when compared to terrestrial mammals (Chapters 2, 3 and 6). However, no evidence was found for a reduction in the balancing selection pressure in the marine environment, contrary to the hypothesis of Slade (1992). The evidence of positive selection at Mhc loci

suggests a similar level of balancing selection (Chapters 3 and 6). Given the relatively slow rate of Mhc evolution (Klein et al. 1993a) and the mounting evidence that a large amount of the Mhc variation in terrestrial mammals is of trans-species origin (e.g. Arden and Klein 1982; Lawlor et al. 1988; Mayer et al. 1992 ; Figueroa et al. 1994), two hypotheses may explain the low amount of allelic variation in cetaceans:

1) Cetaceans have experienced severe genetic bottlenecks which have reduced the amount of trans-species polymorphism. One explanation may be that, on average, speciation in cetaceans occurs in populations of smaller size than in terrestrial mammals (Chapter 3);

2) Cetaceans have a slower neutral substitution rate than terrestrial mammals. The invasion of a dramatically new environment may have required a new suite of Mhc polymorphism. This polymorphism is accumulating, just at a slower rate than in terrestrial mammals (Chapters 3 and 6).

The first hypothesis remains entirely speculative. Although a large number of cetacean populations have, in historic times, suffered severe population reductions, this does not explain the modest to low amount of variation in beluga. In beluga, the population reduction would have had to have occurred prior to the establishment of the current populations, as similar levels and types of variation are found in geographically and genetically distinct populations (Chapters 2 and 5). Comparison of beluga and right whale *DQβ* sequences (Chapter 6), however, shows a 3x reduction in the neutral substitution rate when compared to terrestrial mammals, supporting the second hypothesis. In addition, analysis of cetacean nuclear and mitochondrial DNA sequences also show a similar 2-4x reduction in the neutral substitution rate (Schlötterer et al. 1991; Hoelzel et al. 1991; Baker et al. 1993).

Further comparative studies of Mhc variation among cetacean populations are needed to confirm a reduced neutral substitution rate. If cetacean do have a reduced substitution rate, they represent an opportunity to study the process of evolution of Mhc

loci, such as in the analysis of moose (*Alces alces*) *DRβ* variation (Andersson and Mikko 1995), at time frames before the accumulation of a large number of reversions have clouded the evolutionary events. These studies may allow for the relative importance of intra or inter-allelic recombination (Gyllensten et al. 1991) versus the accumulation of point mutations (O'hUigin 1995) in the generation of new variation to be assessed. Further, the effects of duplication events on Mhc variation could be assessed through the mapping of events in the cetacean phylogeny.

A number of studies have been suggested that would improve the understanding of Mhc class II evolution in cetaceans. First, the analysis of right whale *DRβ* variation would allow for a second estimation of neutral substitution rates (Chapter 6). Next, larger surveys of class II variation in whales closely related to beluga and right whales, i.e. narwhals and bowhead whales, should be conducted to assess the amount of trans-species allelic lineages (Chapters 3 and 6). Surveys of Mhc variation in delphinids (i.e. the bottlenose dolphin (*Tursiops truncatus*)) or phocids (i.e. the harbour porpoise (*Phocoena phocoena*)), species closely related to monodontids (10-15 Myr), with large populations, and no evidence of historic bottlenecks, would also be useful in assessing Mhc evolution (Chapter 3). Third, a reassessment of Mhc variation in fin (*Balenoptera physalus*) and sei (*B. borealis*) whales is in order (Chapter 6).

Beluga Conservation

Mhc variation has also proven useful in the analysis of population structure. Although it is hard to dissect the role of selection affecting the allele frequencies, significant differences between populations do indicate a different genetic response in two environments. Variation among beluga populations supports the hypothesis that interbreeding is occurring among summering populations that share a common wintering area (Chapters 2 and 5). In addition, the regular role of polynyas (such as the North Water) as overwintering locations for High Arctic beluga is suggested (Chapter 5). The

genetic differences among each of the wintering populations indicates that the amount of genetic exchange among the wintering groups is either minimal, or that the selection pressure at each location is strong enough to overcome its effects. However, as no effects of selection were detected at any of the sampling locations, the former explanation is favoured. Modelling of the intensity of selection needed to maintain the observed allele frequencies in the face of migration among the wintering populations would aid in the assessment of genetic structure among beluga populations.

Only through an understanding of the evolutionary forces shaping the Mhc variation in cetaceans can the true implications of a loss of this variation be understood. If there is a reduction in neutral substitution rate in cetaceans, then the loss of Mhc variation may have a much greater consequence for this group. It has been suggested that a loss of Mhc variation may increase the chance of extinction of a population (Yuhki and O'Brien 1990). A slow substitution rate would mean that a loss of Mhc variation in cetaceans would only very slowly be regained.

The St. Lawrence beluga, however, does not have a significant reduction in amount of *DQβ* or *DRβ1* class II variation (Chapters 2 and 5). The evolutionary potential of this population has not been lost at these loci. Effects of the historic bottleneck, however, can be seen in the linkage disequilibrium between the two loci (Chapter 5). It is not clear what consequence the reduction in haplotype variation may have on the population. Certain genotype combinations will occur less often than by chance and the likelihood of inheritance will not be equal (Chapter 5). Further studies comparing live biopsied beluga to dead beached whales, and correlating Mhc variation with pathologies (Chapter 4), may help in this situation. Finally, it is not known how the reduction in haplotype variation has affected other Mhc loci. Studies of class I variation should be the next step in the analysis of the Mhc of beluga. Interestingly, the linkage disequalibrium, presumably caused by the population bottleneck, may allow for a glimpse at the structure of the Mhc complex in beluga, as a physical linkage between class I and class II loci may be detected.

Literature Cited

AGUILAR, A. 1986. A review of old basque whaling and its effect on right whales (*Eubalaena glacialis*) of the North Atlantic. Reports of the International Whaling Commission (Special Issue 10):191-199.

AMMER, H., F-W. SCHWAIGER, C. KAMMERBAUER, M. GOMOLKA, A. ARRIENS, S. LAZARY, and J. T. EPPLEN. 1992. Exonic polymorphic vs intronic simple repeat hypervariability in *MHC-DRß* genes. Immunogenetics **35**:332-340.

ANDERSSON, L. and S. MIKKO. 1995. Generation of MHC Class II diversity by intra- and intergenic recombination. Immunol. Rev. **143**:5-12.

ARDEN, B. and J. KLEIN. 1982. Biochemical comparison of major histocompatibility complex molecules from different subspecies of *Mus musculus*: Evidence for *trans*-specific evolution of alleles. Proc. Natl. Acad. Sci. USA **79**:2342-2346.

ARMSTRONG, P. S. 1995. MHC diversity in the Hawaiian monk seal. Eleventh Biennial Conference on the Biology of Marine Mammals. Orlando, Florida, USA (Abstract). p 4.

BAKER, C. S., A. PERRY, J. L. BANNISTER, M. T. WEINRICH, R. B. ABERNETHY, J. CALAMBOKIDIS, J. LIEN, R. H. LAMBERTSEN, J. URBAN RAMIREZ, O. VASQUEZ, P. J. CLAPHAM, 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. USA. **90**:8239-8243.

BARNES, L. G. 1977. Outline of eastern North Pacific fossil cetacean assemblages. Syst. Zool. **25**:321-43.

BARNES, L. G., D. P. DOMNING, and C. E. RAY. 1985. Status of studies on fossil marine mammals. Mar. Mamm. Sci. **1**(1):15-53.

BÉLAND, P., A. VÉRZINA, and D. MARTINEAU. 1988. Potential for growth of the St. Lawrence (Quebec, Canada) beluga whale (*Delphinapterus leucas*) population based on modelling. J. Cons. Int. Explor. Mer. **45**:22-32.

BÉLAND, P., S. DE GUISE, and R. PLANTE. 1992 Toxicology and pathology of St. Lawrence marine mammals. Rept. Wildlife Toxicology Fund, World Wildlife Fund (Canada), Toronto.

BÉLAND, P., S. DEGUISE, C. GIRARD, A. LAGACÉ, D. MARTINEAU, R. MICHAUD, D. G. G. MUIR, R. J. NORSTROM, É. PELLETIER, S. RAY, and L. R. SHUGART. 1993. Toxic compounds and health and reproductive effects in St. Lawrence beluga whales. J. Great Lakes Res. **19**(4):766-775.

BÉLAND, P., S. DE GUISE, and A. LAGACÉ. 1995. More about carcinogens and tumors. Vet. Pathol. **32**:212-213.

BENHAM, A., A. TULP and J. NEEFJES. 1995. Synthesis and assembly of MHC-peptide complexes. Immunol. Today **16**:359-362.

BERGSTRÖM, T. and U. GYLLENSTEN. 1995. Evolution of Mhc class II polymorphism: the rise and fall of class II gene function in primates. Immunol. Rev. **143**:13-31.

BLACK, F. L. and F. M. SALZANO. 1981. Evidence for heterosis in the HLA system. Am. J. Hum. Genet. **33**:894-899.

BRENNIN, R. 1992. Population genetic structure of beluga whales, *Delphinapterus leucas*: mitochondrial DNA sequence variation within and among North American populations. Master of Science thesis. McMaster University, Hamilton, ON, Canada.

BRODIE, P. F. 1971. A reconsideration of aspects of growth, reproduction and behaviour of the white whale *(Delphinapterus leucas)* with reference to the Cumberland Sound, Baffin Island population. J. Fish. Res. Board Can. **28**:1309-1318.

BROWN, J. G. 1996. Genetic population structure of North American beluga whales, *Delphinapterus leucas*. Doctoral Thesis, University of Guelph, Guelph, ON, Canada.

BROWN, J. H., T. JARDETZHY, M. A. SAPER, B. SAMRAOUI, P. J. BJORKMAN, and D. C. WILEY. 1988. A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. Nature **332**:845-850.

BROWN, J. H., T. S. JARDETZHY, J. C. GORGA, L. J. STERN, R. G. URBAN, J. L. STROMINGER and D. C. WILEY. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature **364**:33-39.

BROWN, M. W., R. HELBIG, P. T. BOAG, D. E. GASKIN and B. N. WHITE. 1991a. Sexing beluga whales (*Delphinapterus leucas*) by means of DNA markers. Can. J. Zoo. **69**:1971-1976.

BROWN, M. W., S. D. KRAUS and D. E. GASKIN. 1991b. Reaction of North Atlantic right whales (*Eubalaena glacialis*) to skin biopsy sampling for genetic and pollutant analysis. Rept. Inter. Whal. Commn (Special Issue 13):81-89.

BROWN, M. W., S. D. KRAUS, D. E. GASKIN and B. N. WHITE. 1994. Sexual composition and analysis of reproductive females in the North Atlantic right whale, *Eubalaena glacialis*, population. Mar. Mam. Sci. **10**(3):253-265.

BROWN, W. M., M. GEORGE, and A. C. WILSON. 1979. Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76**:1967-1971.

BURNS, J. J. and G. A. SEAMAN. 1985. Investigation of belukha whales in coastal waters of western and northern Alaska. II. Biology and ecology. Rept. Alaska Dept. Fish and Game, contract NA 81 RAC 0049, 129 p.

CAMPBELL, R. R. (*ed.*) 1993. Rare and endangered fishes and marine mammals of Canada: COSEWIC fish and marine mammal subcommittee status reports IX. Can. Fld. Nat. **107**:395-401.

CARON, L. M. J. and T. G. SMITH. 1990. Philopatry and site tenacity of belugas, *Delphinapterus leucas*, hunted by the Inuit at the Nastapoka estuary, eastern Hudson Bay. *In:* T. G. Smith, D. J. St.Aubin, and J. R. Geraci (*eds.*), Advances in research on the beluga whale, *Delphinapterus leucas.* Can. Bull. Fish. Aquat. Sci. **224**:69-79.

COHEN, J. 1977. Statistical power analysis for the behavioral sciences. Rev. Ed. Academic Press, New York, NY. 474 p.

CRONE, M. J., and S. D. KRAUS. 1990. Right whales (*Eubalaena glacialis*) in the western North Atlantic: A catalogue of identified individuals. New England Aquarium, Boston, Massachusetts.

CURRY, B. E. , M. MILINKOVITCH, J. SMITH and A. E. DIZON. 1995. Stock structure of bottlenose dolphins, *Tursiops truncatus.* Eleventh Biennial Conference on the Biology of Marine Mammals. Orlando, Florida, USA (Abstract). p 27.

DAVIES, J. L. 1963. The Antitropical factor in cetacean speciation. Evolution **17**:107-116.

DEGOS, L., J. COLOMBANI, A. CHAVENTRE, B. BENGTSON and A. JACQUARD. 1974. Selective pressure on HL-A polymorphism. Nature **249**:62-63.

DE GUISE, S, A. LAGACÉ, and P. BÉLAND. 1994a. Tumors in St. Lawrence beluga whales (*Delphinapterus leucas*). Vet. Pathol. **31**:444-449.

DE GUISE, S, A. LAGACÉ, and P. BÉLAND. 1994b. Gastric papillomas in eight St. Lawrence beluga whales (*Delphinapterus leucas*). J. Vet. Diagn. Invest. **6**:385-388.

DE GUISE, S, A. LAGACÉ, P. BÉLAND, C. GIRARD, and R. HIGGINS. 1995a. Non-neoplastic lesions in beluga whales (*Delphinapterus leucas*) and other marine mammals from the St. Lawrence estuary. J. Comp. Path. **112**:257-271.

DE GUISE, S, D. FLIPO, J. R. BOEHM, D. MARTINEAU, P. BÉLAND, and M. FOURNIER. 1995b. Immune functions in beluga whales (*Delphinapterus leucas*): Evaluation of phagocytosis and respiratory burst with peripheral blood leukocytes using flow cytometry. Vet. Immun. and Immunopath. **47**:351-362.

DOHERTY, P. C. and R. M. ZINKERNAGEL. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature **256**:50-52.

DOIDGE, D. W. 1990. Age-length and length-weight comparisons in the beluga, *Delphinapterus leucas. In:* T. G. Smith, D. J. St. Aubin and J. R. Geraci (*eds.*), Advances in research on the beluga whale, *Delphinapterus leucas.* Can. Bull. Fish. Aquat. Sci. **224**:59-68.

DOIDGE, D. W. and K. J. FINLEY. 1993. Status of the Baffin Bay population of beluga, *Delphinapterus leucas.* Can. Fld. Nat. **107**:533-546.

ELLEGREN, H., G. HARTMAN, M. JOHANSSON and L. ANDERSSON. 1993. Major histocompatibility complex monomorphism and low levels of DNA fingerprinting variability in a reintroduced and rapidly expanding population of beavers. Proc. Natl. Acad. Sci. USA **90**:8150-8153.

ELLEGREN, H., S. MIKKO, K. WALLIN and L. ANDERSSON. 1996. Limited polymorphism at major histocompatibility complex (MHC) loci in the Swedish moose *A. alces*. Mol. Ecol. **5**:3-9.

ENGELHARD, V. H. 1994. Structure of peptides associated with class I and class II Mhc molecules. Annu. Rev. Immunol. **12**:181-207.

EVANS, P. G. H. 1987. The natural history of whales and dolphins. Facts on File Inc. New York, NY. 343 p.

FALK, K., O. RÖTZSCHKE, S. STEVANOVIC, G. JUNG, and H-G. RAMMENSEE. 1994. Pool sequencing of natural HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules. Immunogenetics **39**:230-242.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution **39**:783-791.

FELSENSTEIN, J. 1993. PHYLIP manual version 3.5c. University Herbarium, University of California, Berkeley California.

FIGUEROA, F., J. GUTKNECHT, H. TICHY, and J. KLEIN. 1990. Class II Mhc genes in rodent evolution. Immunol. Rev. **113**:27-46.

FIGUEROA, F., C. O'hUIGIN, H. TICHY, and J. KLEIN. 1994. The origin of the primate *Mhc-DRβ* genes and allelic lineages as deduced from the study of prosimians. J. Immunol. **152**:4456-4465.

FINLEY, K. J., J. P. HICKIE, and R. A. DAVIS. 1987. Status of the beluga, *Delphinapterus leucas*, in the Beaufort Sea. Can. Fld. Nat. **101**:271-278.

FINLEY, K. J. and W. E. RENAUD. 1980. Marine mammals inhabiting the Baffin Bay North Water in winter. Arctic **33**:724-738.

FITCH, W. M., J. M. E. LEITER, X. LI, and P. PALESE. 1991. Positive Darwinian evolution in human influenza A viruses. Proc. Natl. Acad. Sci. USA **88**:4270-4274.

FORDYCE, R. E. and L. G. BARNES. 1994. The evolutionary history of whales and dolphins. Annu. Rev. Earth Planet. Sci. **22**:419-55.

FRIEDE, T., V. GNAU, G. JUNG, W. KEILHOLZ, S. STEVANOVIC and H-G. RAMMENSEE. 1996. Natural ligand motifs of closely related HLA-DR4 molecules predict features of rheumatoid arthritis associated peptides. Biochimica et Biophysica Acta **1316**:85-101.

GASKIN, D. E. 1982. The ecology of whales and dolphins. Heinemann Educational Books Inc. London and Portsmouth, New Hampshire. 459 p.

GERMAIN, R., N. 1994. Mhc-dependent antigen processing and peptide presentation: providing ligands for T lymphocytes activation. Cell **76**:287-299.

GERMAIN, R. N., F. CASTELLINO, R. HAN, C. R. E. SOUSA, P. ROMANGNOLI, S. SADEGH-NASSERI, and G-M. ZHONG. 1996. Processing and presentation of endocytically acquired protein antigens by Mhc class I and class II molecules. Immunol. Rev. **151**:5-30.

GOLDSTEIN, D. A., A. R. LINARES, L. L. CAVALLI-SFORZA, and M. W. FELDMAN. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc. Natl. Acad. Sci. USA **92**:6723-6727.

GOODMAN, S. J. and R. SLADE. 1995. DNA sequence variation in European harbour seal (*Phoca vitulina*) major histocompatibility (MHC) class I genes. Eleventh Biennial Conference on the Biology of Marine Mammals. Orlando, Florida, USA (Abstract). p. 45.

GYLLENSTEN, U. B., D. LASHARRI, and H. A. ERLICH. 1990. Allelic diversification at the class II *DQβ* locus of the mammalian major histocompatibilty complex. Proc. Natl. Acad. Sci. USA **87**:1835-1839.

GYLLENSTEN, U. B., M. SUNDVALL, I. EZCURRA, and H. A. ERLICH. 1991. Genetic diversity at Class II DRß loci of the primate MHC. J. of Immunol. **146**:4368-4376.

GYLLENSTEN, U. B., T. BERGSTRÖM, A. JOSEFSSON, M. SUNDVALL, A. SAVAGE, E. S. BLUMER, L. H. GIRALDO, L. H. SOTO, and D. I. WATKINS. 1994. The cotton-top tamarin revisited: MHC class I polymorphism of wild tamarins, and polymorphism and allelic diversity of the class II DQα1, DQß and DRß loci. Immunogenetics **40**:167-176.

HAYASHI, K. 1992. PCR-SSCP: a method for detection of mutations. GATA **9**(3):73-79.

HARINGTON, C. R. 1989. Marine mammals of the Champlain Sea, and the problems of whales in Michigan. *In:*N. R. Gadd (*ed*) The late Quaternary development of the Champlain Sea Basin: geological association of Canada. Special Paper **35**:225-240.

HAZARD, K. 1988. Beluga whale *Delphinapterus leucas. In:* J. W. Lintfer (*ed*), Selected marine mammals of Alaska. Species accounts with research and management recommendations. U.S. Marine Mammal Commission. Washington. pp 195-235.

HEIDE-JØRGENSEN, M. P. and J. TEILMANN. 1994. Growth, reproduction, age structure and feeding habits of white whales (*Delphinapterus leucas)* in west Greenland waters. Meddr. Grønland, Biosci. **39**:195-212.

HILL, A. V. S., C. E. M. ALLSOPP, D. KWIATKOWSKI, N. M. ANSTEY, P. TWUMASI, P. A. ROWE, S. BENNETT, D. BREWSTER, A. J. MCMICHAEL and B. M. GREENWOOD. 1991. Common West African HLA antigens are associated with protection from severe malaria. Nature **352**:595-600.

HILL, A. V. S., S. N. R. YATES, C. E. M. ALLSOPP, S. GUPTA, S. C. GILBERT, A. LALVANI, M. AIDOO, M. DAVENPORT and M. PLEBANSKI. 1994. Human leukocyte antigens and natural selection by malaria. Phil. Trans. R. Soc. Lond. B. **346**:379-385.

HOELZEL, A. R., J. M. HANCOCK and G. A. DOVER. 1991. Evolution of the cetacean mitochondrial D-loop region. Mol. Biol. Evol. **8**:475-493.

HOWARD, E. B., J. O. BRITT, Jr., G. K. MATSUMOTO, R. ITAHARA, and C. N. NAGANO. 1983. Bacterial Diseases. *In:* E. B. Howard (*ed*), Pathobiology of marine mammal diseases, vol 1. CRC Press Inc. Boca Raton, Florida. pp 69-118.

HUGHES, A. L. 1991a. MHC polymorphism and the design of captive breeding systems. Conserv. Biol. **5**:249-251.

HUGHES, A. L. 1991b. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. Genetics **127**:345-353.

HUGHES, A. L. and M. NEI. 1988. Pattern of nucleotide substitution at major histocompatibility complex loci reveals overdominant selection. Nature **335**:167-170.

HUGHES, A. L. and M. NEI. 1989. Nucleotide substitutions at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86**:958-962.

HUGHES, A. L. and M. NEI. 1993. Evolutionary relationships of the classes of major histocompatibility complex genes. Immunogenetics **37**:337-346.

HUGHES, A. L. and M. K. HUGHES. 1995. Natural selection on the peptide-binding regions of major histocompatibility complex molecules. Immunogenetics **42**:233-243.

HUGHES, A. L., T. OTA, and M. NEI. 1990. Positive Darwinian selection promotes charge profile diversity in the antigen binding cleft of class I MHC molecules. Mol. Biol. Evol. **7**:515-524.

HUGHES, A. L., M. K. HUGHES, C. Y. HOWELL, and M. NEI. 1994. Natural selection at the class II major histocompatibility complex loci of mammals. Phil. Trans. R. Soc. Lond. B. **345**:359-367.

HUMPHREYS, T. and E. L. REINHERZ. 1994. Invertebrate immune recognition, natural immunity and the evolution of positive selection. Immunol. Today **15**:316-320.

IMANISHI, T. and T. GOJOBORI. 1992. Patterns of nucleotide substitutions inferred from the phylogenies of the Class I major histocompatibility complex genes. J. Mol. Evol. **35**:196-204.

IWC (International Whaling Commission). 1992. Report of the sub-committee on small cetaceans. Rep. Int. Whal. Comm. **42**:178-234.

JUKES, T. H. and C. R. CANTOR. 1969. Evolution of protein molecules. *In*: H. N. Monro *(ed.)*, Mammalian protein metabolism III. Academic Press, New York. Pp. 21-132.

KENNEDY, J. J. 1992. Analzing qualitative data: log-linear analysis for behavioral research. Praeger, New York. 299 p.

KIMURA, M. A. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111-120.

KINGSLEY, M. C. S. 1993. Census, Trend, and Status of the St. Lawrence beluga population in 1992. Can. Tech. Rep. Fish. Aquat. Sci. **1938**.

KLEIN, J. 1986. Natural History of the Major Histocompatibility Complex. John Wiley & Sons, Inc. 775 p.

KLEIN, J. and C. O'hUIGIN. 1994. MHC polymorphism and parasites. Phil. Trans. R. Soc. Lond. B. **346**:351-358.

KLEIN, J. and C. O'hUIGIN. 1993. Composite origin of major histocompatibility complex genes. Curr. Opin. Genet. Devel. **3**:923-930.

KLEIN, J. and N. TAKAHATA. 1990. The major histocompatibility complex and the quest for origins. Immunol.. Rev. **113**:5-25.

KLEIN, J., R. E. BONTROP, R. L. DAWKINS, H. A. ERLICH, U. B. GYLLENSTEN, E. R. HEISE, P. P. JONES, P. PARHAM, E. K. WAKELAND, and D. I. WATKINS. 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics **31**:217-219.

KLEIN, J., Y. SATTA and C. O'hUIGIN. 1993a. The molecular descent of the major histocompatibility complex. Annu. Rev. Immunol. **11**:269-295.

KLEIN, J., H. ONO, D. KLEIN and C. O'hUIGIN. 1993b. The accordion model of Mhc evolution. *In:* J. Gergely, M. Benczúr, A. Erdei, A. Falus, Gy. Füst, G. Medgyesi, Gy. Petrányi, É. Rajnavölgyi (*eds*), Progress in Immunology Vol. VIII. Proceedings of the 8th International Congress of Immunology Budapest 1992. Springer-Verlag, Budapest. Pp 137-195.

KLEIN, J., TAKAHATA, N. and AYALA, F.J. 1993c. MHC polymorphism and human origins. Sci. Amer. **269:**Dec. 46-51.

KLEINENBERG, S. E., A. V.,YABLOKOV, B. M., BEL'KOVICH, and M. N. TARASEVICH. 1964. Belukha opyt monograpficheskogo issledovaniya vida. (Beluga (*Delphinapterus leucas)*: Investigation of the species) Akademiya Nauk. S.S.S.R., Inst. Morfologii Zhivotnykh in A.N. Severtsova Izdatel'stov Nauka. Moscow, 456pp. Translated from Russian by the Israel Program for Scientific Translations, Jerusalem, No. 1923. 1969, 376p.

KLITZ, W., G. THOMSON and M.P. BAUR. 1986. Contrasting evolutionary histories among tightly linked HLA loci. Am. J. Hum. Genet. **39**:340-349.

KNOWLTON, A. R., S. D. KRAUS and R. D. KENNEY. 1994. Reproduction in North Atlantic right whales (*Eubalaena glacialis*). Can. J. Zool. **72**:1297-1305.

KRAUS, S. D., K. E. MOORE, C. A. PRICE, M. J. CRONE, W. A. WATKINS, H. E. WINN and J. E. P. 1986. The use of photographs to identify individual North Atlantic right whales (*Eubalaena glacialis*). Rept. intl. Whal. Commn (Spec. Iss. 10):145-151.

KRONENBERG, M., R. BRINES and J. KAUFMAN. 1994. MHC evolution: a long term investment in defense. Immunol. Today **15**:4-6.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: Molecular evolutionary genetic analysis, version 1.01. The Pennsylvania State University, University Park, PA.

LANZAVECCHIA, A. and C. WATTS. 1993. Peptide binding to Mhc class II molecules in living cells: role of class II synthesis and stability of peptide/class II complexes. *In:* J. Gergely, M. Benczúr, A. Erdei, A. Falus, Gy. Füst, G. Medgyesi, Gy. Petrányi, É. Rajnavölgyi (*eds*), Progress in Immunology Vol. VIII. Proceedings of the 8th International Congress of Immunology Budapest 1992. Springer-Verlag, Budapest. Pp. 189-195.

LAWLOR, D. A., F. E. WARD, P. D. ENNIS, A. P. JACKSON and P. PARHAM. 1988. HLA-A and -B polymorphism predate the divergence of humans and chimpanzees. Nature **335**:268-271.

LEE, Y-H. and V. D. VACQUIER. 1992. The Divergence of species-specific abalone sperm lysins is promoted by positive Darwinian selection. Biol. Bull. **182**:97-104.

LI, W-H., C-I. WU, and C-C. LOU. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150-174.

LINT, D. W., J. W. CLAYTON, W. R. LILLIE, and L. POSTMA. 1990. Evolution and systematics of the beluga whale, (*Delphinapterus leucas*) and other odontocetes: A molecular approach. *In:* T. G. Smith, D. J. St. Aubin and J. R. Geraci (*eds.*), Advances in research on the beluga whale, *Delphinapterus leucas.* Can. Bull. Fish. Aquat. Sci. **224**:7-22.

MANCUSO, S. J. 1995. Population genetics of Hudson Bay bleuga whales (*Delphinapterus leucas*): An analysis of population structure and gene flow using mitochondrial DNA sequences and multilocus DNA fingerprinting. Master of Science Thesis, McMaster University, Hamilton, ON, Canada.

MARSH, S. G. E. and J. G. BODMER. 1993. HLA class II nucleotide sequences, 1992. Immunogenetics **37**:79-94.

MARSH, S. G. E. and J. E. BODMER. 1991. *HLA* class II nucleotide sequences, 1991. Immunogenetics **33**:321-334.

MARTIN, A. P. and S. R. PALUMBI. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA **90**:4087-4091.

MARTINEAU, D., P. BÉLAND, C. DESJARDINS, and A. LAGACÉ. 1987. Levels of organochlorine chemicals in tissues of beluga whales (*Delphinapterus leucas*) from the St. Lawrence estuary, Quebec, Canada. Arch. Environ. Contam. Toxicol. **16**:137-147.

MARTINEAU, D., A. LAGACÉ, P. BÉLAND, R. HIGGINS, D. ARMSTRONG, and L. R. SHUGART. 1988. Pathology of stranded beluga whales (*Delphinapterus leucas*) from the St. Lawrence estuary, Quebec, Canada. J. Compar. Pathol. **98**:287-311.

MARTINEAU, D., S. LAIR, S. DE GUISE, and P. BÉLAND. 1995. Intestinal adenocarcinomas in two beluga whales (*Delphinapterus leucas*) from the estuary of the St. Lawrence river. Can. Vet. J. **36**:563-565.

MAYER, W. E., C. O'hUIGIN, Z. ZALESKA-RUTCZYNSKA and J. KLEIN. 1992. Trans-species origin of Mhc-DRß polymorphism in the chimpanzee. Immunogenetics. **37**:12-23.

MCLACHLAN, A. D. 1972. Repeating sequences and gene duplication in proteins. J. Mol. Biol. **64**:417-437.

MEASURES, L. N., P. BÉLAND, D. MARTINEAU, and S. DE GUISE. 1995. Helminths of an endangered population of beluga, *Delphinapterus leucas*, in the St. Lawrence estuary, Canada. Can. J. Zool. **73**:1402-1409.

MICHAUD, J., L. C. BRODY, G. STEEL, G. FONTAINE, L. S. MARTIN, D. VALLE, and G. MITCHELL. 1992. Strand-separating conformational polymorphism analysis: efficacy of detection of point mutations in the human ornithine δ-aminotransferase gene. Genomics **13**:389-394.

MICHAUD, R. 1993. Distribution estivale du beluga du Saint-Laurent: synthèse de 1986 à 1992. Rapp. Tech. Can. Sci. Halieut. Aquat. **1906**.

MICHAUD, R., A. VÉZINA, N. RONDEAU, and Y. VIGNEAULT. 1990. Distribution annuelle et caractérisation préliminaire des habitats du béluga (*Delphinapterus leucas*) du Saint-Laurent. Rapp. tech. can. sci. halieut. aquat. **1757** 31p.

MIKKO, S. and L. ANDERSSON. 1995a. Extensive MHC class II DRß3 diversity in African and European cattle. Immunogenetics **42**:408-413.

MIKKO, S. and L. ANDERSSON. 1995b. Low major histocompatibility complex class II diversity in European and North American moose. Proc. Natl. Acad. Sci. USA **92**:4259-4263.

MILINKOVITCH, M. C., G. ORTI, and A. MEYER. 1993. Revised phyolgeny of whales suggested by mitochondrial DNA sequences. Nature **361**:346-348.

MILINKOVITCH, M. C., A. MEYER, and J. R. POWELL. 1994. Phyolgeny of all major groups of cetaceans based on DNA sequences from three mitochondrial genes. Mol. Biol. Evol. **11**:939-948.

MUIR, D. C. G., C. A. FORD, R. E. A. STEWART, T. G. SMITH, R. F. ADDISON, M. E. ZINCK and P. BÉLAND. 1990. Organochlorine contaminants in beluga, *Delphinapterus leucas*, from Canadian waters. Can. Bull. Fish. Aquat. Sci. **224**:165-190.

NATIONAL MARINE FISHERIES SERVICE (NMFS) 1991. Final recovery plan for the Northern right whale (*Eubalaena glacialis*). National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Washington, D.C. 86 p.

NEI, M. and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418-426.

NIZETIC, D., F. FIGUEROA, Z. DEMBIC, E. NEVO and J. KLEIN. 1987. Major histocompatibility complex gene organization in the mole rat *Spalax ehrenbergi*:: Evidence for transfer of function between class II genes. Proc. Natl. Acad. Sci. USA **84**:5828-5832.

O'BRIEN, S. J., M. E. ROELKE, L. MARKER, A. NEWMAN, C. A. WINKLER, D. MELTZER, L. COLLY, J. F. EVERMAN, M. BUSH, and D. E. WILDT. 1985. Genetic basis for species vulnerability in the cheetah. Science **227**:1428-1434.

O'hUIGIN, C. 1995. Quantifying the degree of convergence in primate *Mhc-DRβ* genes. Immun. Rev. **143**:123-140.

O'hUIGIN, C., R. BONTROP, and J. KLEIN. 1993. Nonhuman primate *Mhc-DRβ* sequences: a compilation. Immunogenetics **38**:165-183.

OTTING, N., M. KENTER, P. VAN WEEREN, M. J. JONKER, and R. E. BONTROP. 1992. *Mhc-DQβ* repertoire variation in hominiod and Old World primate species. J. Immunol. **149**:461-470.

PATENAUDE, N., J. S. QUINN, P BÉLAND, M. KINGSLEY, and B. N. WHITE. 1994. Genetic variation of the St. Lawrence beluga whale population assessed by DNA fingerprinting. Mol. Ecol. **3**:375-381.

PIPPARD, L. 1985. Status of the St. Lawrence river population of beluga, *Delphinapterus leucas*. Can. Fld. Nat. **99**:438-450.

POTTS, W. K., C. J. MANNING and E. K. WAKELAND. 1991. MHC genotype influences mating patterns in semi-natural populations of *Mus*. Nature **352**:619-621.

POTTS, W. K., C. J. MANNING and E. K. WAKELAND. 1994. The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. Phil. Trans. R. Soc. Lond. B. **346**:369-378.

POWIS, S. H. and D. E. GERAGHTY. 1995. What is the MHC? Immunol. Today **16**:466-468.

RALLS, K., J. D. BALLOU, and A. TEMPLETON. 1988. Estimates of lethal equivalents and the cost of inbreeding in mammals. Conserv. Biol. **2**:185-193.

RAMMENSEE, H-G., T. FRIEDE and S. STEVANOVIC. 1995. MHC ligands and peptide motifs: first listing. Immunogenetics **41**:178-228.

RAYMOND, M.and R. ROUSSET. 1995a. GENEPOP (version 2.0): A population genetics software for exact tests and ecumenicism. Originally descibed (version 1.2) in J. Heredity 1995, **86**:248-249, Institut des Sciences de l'Evolution, Montpellier, France.

RAYMOND, M. and R. ROUSSET. 1995b. An exact test for population differentiation. Evolution **49**:12380-1283.

REEVES, R. R. and S. K. KATONA. 1980. Extralimital records of white whales (*Delphinapterus leucas*) in eastern North American waters. Can. Fld. Nat. **94**:239-247.

REEVES, R. R. and E. MITCHELL. 1984. Catch history and initial population of white whales (*Delphinapterus leucas*) in the River and Gulf of St. Lawrence, eastern Canada. Naturaliste can. (Rev. Écol. Syst.) **111**:63-121.

REEVES, R. R. and E. MITCHELL. 1986a. The Long Island, New York, right whale fishery: 1650-1924. Rept. Intern. Whal. Commn (Special Issue **10**):201-220.

REEVES, R. R., and E. MITCHELL. 1986b. American pelagic whaling for right whales in the North Atlantic. Rept. Intern. Whal. Commn (Special Issue **10**):221-254.

REEVES, R. R. and E. MITCHELL. 1987a. Distribution and migration, exploitation and former abundance of white whales (*Delphinapterus leucas*) in Baffin Bay and adjacent waters. Can. Spec. Publ. Fish. Aquat. Sci. **99**: 34 p.

REEVES, R. R. and E. MITCHELL. 1987b. History of white whales (*Delphinapterus leucas*) exploitation in eastern Hudson Bay and James Bay. Can. Spec. Publ. Fish. Aquat. Sci. **95**: 45 p.

RICE, W.R. 1989. Analyzing tables of statistical tests. Evolution **43**:223-225.

RITTE, U., E. NEUFELD, C. O'hUIGIN, F. FIGUEROA and J. KLEIN. 1991. Origins of H-2 polymorphism in the house mouse. II. Characterization of a model population and evidence for heterozygous advantage. Immunogenetics **34**:164-173.

ROUSSET, R and M. RAYMOND. 1995. Testing heterozygote excess and deficiency. Genetics **140**:1413-1419.

ROSEL, P. E., A. E. DIZON and M. G. HAYGOOD. 1995. Variability of the mitochondrial control region in populations of the harbour porpoise, *Phocoena phocoena*, on interoceanic and regional scales. Can. J. Fish. Aquat. Sci. **52**:1210-1219.

SAITOU, N. and M. NEI. 1987. The neighbor-joining method. A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406-425.

SCHAEFF, C. 1993. Genetic analysis of right whales (*Eubalaena*) in the Atlantic Ocean. Doctoral thesis, Queen's University, Kingston, ON, Canada.

SCHAEFF, C., S. KRAUS, M. BROWN, J. PERKINS, R. PAYNE, D. GASKIN, P. BOAG and B. WHITE. 1991. Preliminary analysis of mitochondrial DNA variation within and between the right whale species *Eubalaena glacialis* and *Eubalaena australis*. Rept. Intern. Whal. Commn (Special Issue **13**):217-223.

SCHAEFF, C., S. KRAUS, M. BROWN and B. WHITE. 1993. Assessment of the population structure of western North Atlantic right whales (*Eubalaena glacialis*) based on sighting and mtDNA RFLP data. Can. J. Zool. **71**:339-345.

SCHLÖTTERER, C., B. AMOS, and D. TAUTZ. 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature **354**:63-65.

SCHWAIGER, F.-W., E. WEYERS, C. EPPLEN, J. BRüN, G. RUFF, A. CROWFORD, and J. T. EPPLEN. 1993. The paradox of *MHC-DRß* exon/intron evolution: α-helix and ß-sheet encoding regions diverge while hypervariable intronic simple repeats coevolve with ß-sheet codons. J. Mol. Evol. **37**:260-272.

SCHWAIGER, F.-W., E. WEYERS, J. BUITKAMP, A. J. EDE, A. CROWFORD, and J. T. EPPLEN. 1994. Interdependent *MHC-DRß* exon-plus-intron evolution in artiodactyls. Mol. Biol. Evol. **11**:239-249.

SERGEANT, D. E. 1973. Biology of white whales (*Delphinapterus leucas*) in western Hudson Bay. J. Fish. Res. Board Can. **30**:1065-1090.

SERGEANT, D. E. 1986. Present status of white whales in the St. Lawrence estuary. Naturaliste can. (Rev. Ecol. Syst.) **113**:61-81.

SERGEANT, D. E. and P. F. BRODIE. 1969. Body size in white whales, *Delphinapterus leucas*. J. Fish. Res. Board. Can. **26**:2561-2580.

SERGEANT, D. E. and HOEK, W. 1988. An update of the status of white whales *Delphinapterus leucas* in the Saint Lawrence estuary, Canada. Biol. Conn. **45**:287-302.

SERGEANT, D. E., A. W. MANSFIELD and B. BECK. 1970. Inshore records for Cetacea for eastern Canada, 1949-68. J. Fish. Res. Board. Can. **27**:1903-1915.

SHE, J. X., S. A. BOEHME, T. W. WANG, F. BONHOMME and E. K. WAKELAND. 1991. Amplification of major histocompatibility complex class II gene diversity by intraexonic recombination. Proc. Natl. Acad. Sci. USA **88**:453-457.

SIGURDARDÓTTIR, S., C. BORSCH, K. GUSTAFSSON and L. ANDERSSON. 1991. Cloning and sequence analysis of 14 DRß alleles of the bovine major histocompatibility complex by using the polymerase chain reaction. Anim. Genet. **22**:199-209.

SIGURDARDÓTTIR, S., C. BORSCH, K. GUSTAFSSON and L. ANDERSSON. 1992. Gene duplications and sequence polymorphism of bovine class II *DQß* genes. Immunogenetics **35**:205-213.

SLADE, R. 1992. Limited MHC polymorphism in the southern elephant seal: implications for MHC evolution and marine mammal population biology. Proc. R. Soc. Lond. B **249**:163-171.

SLADE, R. W., C. MORITZ and A. HEIDEMAN. 1994. Multiple nuclear-gene phylogenies: Application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny. Mol. Biol. Evol. **11**(3):341-356.

STANLEY, S. M. 1975. A theory of evolution above the species level. Proc. Nat. Acad. Sci. USA **72**:646-650.

ST. AUBIN, D. J., T. G. SMITH, and J. R. GERACI. 1990. Seasonal epidermal moult in beluga whales, *Delphinapterus leucas*. Can. J. Zool. **68**:359-364.

STERN, L. J., J. H. BROWN, T. S. JARDETZKY, J. C. GORGA, R. G. URBAN, J. L. STROMINGER and D. C. WILEY. 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature **368**:215-221.

STEWART, R. E. A. 1994. Size-at age relationships as discriminators of white whale (*Delphinapterus leucas)* stocks in the eastern Canadian Arctic. Meddr. Grønland., Biosci. **39**:217-225.

SWARBRICK, P. A., F-W. SCHWAIGER, J. T. EPPLEN, G. S. BUCHAN, J. F. T. GRIFFEN and A. M. CRAWFORD. 1995. Cloning and sequencing of expressed *DRβ* genes of the red deer (*Cervus elaphus*) MHC. Immunogenetics **42**:1-9.

SWOFFORD, D. L. 1993. PAUP Phylogentic analysis using parsimony, version 3.1.1 (© Smithsonian Institute). Illinios Natural History Survey, Champaign, Illinios.

TAKAHATA, N. and M. NEI. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics **124**:967-978.

TAYLOR, W. R. 1986. The classification of amino acid conservation. J. Theor. Biol. **119**:205-218.

TROWSDALE, J. 1995. "Both man & bird & beast": comparative organization of MHC genes. Immunogenetics **41**:1-17.

TROWSDALE, J., V. GROVES, and A. ARNASON. 1989. Limited MHC polymorphism in whales. Immunogenetics **29**:19-24.

TSUJI, K., M. AIZAWA, and T. SASAZUKI. 1992. HLA 1991: Proceedings of the eleventh international histocompatibility workshop and conference. Oxford University Press.

VISSER, I. K., M. F. VAN BRESSEM, T. BARRETT and A. D. OSTERHAUS. 1993. Morbillivirus infections in aquatic mammals. Veterinary Research **24**:169-178.

VLADYKOV, V. D. 1944. Études sur les mammifères aquatiques. III. Chasse, biologie et valeur économique du marsouin blanc ou béluga (*Delphinapterus leucas*) du fleuve et du golfe Saint Laurent. Dép. Pêcheries, Prov. Québec. 194 p.

VLADYKOV, V. D. 1946. Études sur les mammifères aquatiques. VI. Nourriture du marsouin blanc ou béluga (*Delphinapterus leucas*) du fleuve Saint Laurent. Contr. Dép. Pêcheries, Prov. Québec. **#17** 160 p.

WAGEMANN, R., R. E. A. STEWART, P. BÉLAND and C. DESJARDINS. 1990. Heavy metals and selenium in tissues of beluga whales, *Delphinapterus leucas*, from the Canadian Arctic and the St. Lawrence estuary. Can. Bull. Fish. Aquat. Sci. **224**:191-206.

WANG, J. Y., D. E. GASKIN, and B. N. WHITE. 1996. Mitochondrial DNA analysis of the harbour porpoise, *Phocoena phocoena*, subpopulations in North American waters. Can. J. Fish. Aquat. Sci. **53**:685-698.

WEIR, B. S. and C. C. COCKERHAM. 1984. Estimating F-statistics for the analysis of popuation structure. Evolution **38:**1358-1370.

YUHKI, N. and S. J. O'BRIEN. 1990. DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. Proc. Natl. Acad. Sci. USA **87**:836-840.

Appendix 1

Genetic information available on each beluga sample

Beluga samples are listed according to summering population, sampling location, identification number, and year of collection. Mhc genotypes are listed for each of the beluga samples analyzed. [1] Allele names have been shortened to a three digit number (e.g. *Dele-DQβ**0101 is 101). [2] Mitochondrial haplotype information is also given and determined by Mancuso (1995), prefixed by a 'D' (e.g. D05), and by Brown, prefixed by a 'J ' (e.g. J3). Haplotype lineage, as first described by Brennin (1992), is also given. [3] The existence of microsatellite information (Brown 1996) on this sample is indicated by an asterisk. In each column a dash indicates no genetic information was collected for that locus. NP indicates an attempt to genotype a sample was made at this locus but no amplifiable product was obtained.

Population	Sample Info.		Mhc Genotype[1]		MtDNA haplotype[2]		Micro[3]
– Sample location	Dle #	Year	DQß	DRß1	haplotype	lineage	Data
St. Lawrence							
Dead beached samples	185	88	101 / 101	401 / 401	J29	A	*
	186	88	101 / 101	101 / 201	-	-	-
	187	88	101 / 202	101 / 402	-	-	-
	188	88	101 / 101	101 / 201	J18	A	*
	189	88	101 / 201	201 / 401	-	-	-
	190	88	101 / 101	401 / 401	J29	A	*
	192	88	101 / 202	301 / 402	J18	A	*
	239	88	101 / 101	401 / 401	-	-	-
	334	89	101 / 101	401 / 402	J18	A	*
	335	89	101 / 103	101 / 401	J18	A	*
	336	89	101 / 101	101 / 401	J29	A	*
	338	90	101 / 201	201 / 401	J29	A	*
	339	90	101 / 202	101 / 402	J29	A	*
	340	90	101 / 101	301 / 401	J28	B	*
	341	90	101 / 101	401 / 402	J29	A	*
	342	89	101 / 202	101 / 402	J29	A	*
	368	91	101 / 101	101 / 401	J18	A	*
	369	91	101 / 201	201 / 401	J29	A	*
	370	91	101 / 101	201 / 401	J18	A	*
	371	91	101 / 101	201 / 401	J29	A	*
	503	95	101 / 201	201 / 401	-	-	-
Biopsy darted samples	477	94	101 / 201	201 / 402	-	-	-
	478	94	101 / 101	101 / 401	-	-	-
	479	94	101 / 101	101 / 401	-	-	-
	480	94	101 / 101	401 / 401	-	-	-
	481	95	101 / 101	101 / 401	-	-	-
	482	95	101 / 101	301 / 401	-	-	-
	483	95	101 / 101	301 / 301	-	-	-
	484	95	101 / 101	401 / 401	-	-	-
	485	95	101 / 201	101 / 201	-	-	-
	486	95	101 / 101	101 / 401	-	-	-
	487	95	101 / 101	101 / 401	-	-	-
	488	95	101 / 101	401 / 401	-	-	-
	489	95	101 / 201	101 / 201	-	-	-
	490	95	101 / 101	101 / 501	-	-	-
	491	95	101 / 101	101 / 401	-	-	-
	492	95	101 / 101	201 / 401	-	-	-
	493	95	101 / 101	101 / 201	-	-	-
	494	95	101 / 101	401 / 402	-	-	-
	495	95	101 / 101	101 / 401	-	-	-
	496	95	101 / 202	401 / 402	-	-	-
	497	95	101 / 101	101 / 201	-	-	-
	498	95	101 / 202	201 / 402	-	-	-
	499	95	101 / 101	101 / 401	-	-	-
	500	95	101 / 101	301 / 402	-	-	-
	501	95	101 / 101	101 / 101	-	-	-
	502	95	101 / 101	401 / 401	-	-	-

Population	Sample Info.		Mhc Genotype[1]		MtDNA haplotype[2]		Micro[3]
-- Sample location	Dle #	Year	DQß	DRß1	haplotype	lineage	Data
E. Beaufort Sea							
-- Mackenzie Delta	48	84	101 / 101	201 / 501	J4	B	*
	54	84	101 / 101	101 / 201	J7	B	*
	58	87	-	401 / 501	-	-	-
	59	87	101 / 101	401 / 501	-	-	-
	60	87	101 / 201	401 / 402	J14	B	*
	62	88	101 / 201	201 / 501	J4	B	*
	63	88	-	101 / 402	J1	B	*
	64	88	101 / 101	201 / 501	J7	B	*
	65	88	-	101 / 101	J7	B	*
	66	88	-	202 / 202	J7	B	*
	67	88	101 / 101	101 / 301	J7	B	*
	68	88	101 / 201	301 / 401	J7	B	*
	69	87	-	401 / 501	-	-	-
	70	88	101 / 101	101 / 202	J7	B	*
	72	88	101 / 101	401 / 401	J2	B	*
	73	88	101 / 101	101 / 301	J7	B	*
	82	88	101 / 101	301 / 401	J6	B	*
	88	88	101 / 101	202 / 501	J6	B	*
	89	88	101 / 101	202 / 501	-	-	-
	90	88	101 / 103	201 / 202	-	-	-
	91	88	101 / 101	402 / 501	J3	B	*
	92	88	101 / 101	401 / 501	J6	B	*
	240	89	101 / 101	102 / 102	J4	B	*
	241	89	101 / 103	101 / 402	J8	B	*
	242	89	-	301 / 401	J7	B	*
	243	89	-	401 / 402	-	-	-
	247	89	-	401 / 402	J10	B	*
	249	89	-	301 / 501	J2	B	*
	251	89	-	401 / 401	J4	B	*
	252	89	-	101 / 301	J4	B	*
	253	89	101 / 101	101 / 301	J2	B	*
	256	89	101 / 101	401 / 402	JI	B	*
	259	89	-	201 / 402	J6	B	*
	260	89	101 / 201	202 / 501	J6	B	*
	261	89	101 / 103	201 / 401	J2	B	*
	263	89	101 / 101	202 / 501	J8	B	*
	264	89	-	201 / 301	J7	B	*
	265	89	-	202 / 501	J7	B	*
	267	89	101 / 201	301 / 501	J7	B	*
	268	89	101 / 101	202 / 501	J7	B	*
	269	89	-	101 / 201	J2	B	*
	270	89	-	401 / 401	J2	B	*
	271	89	101 / 101	401 / 402	J3	B	*
	272	89	-	501 / 501	J2	B	*
	280	90	-	201 / 201	J7	B	*
	281	90	101 / 101	202 / 501	J4	B	*
	285	90	101 / 103	301 / 402	J7	B	*
	298	90	101 / 101	101 / 402	JI	B	*

Population	Sample Info.		Mhc Genotype[1]		MtDNA haplotype[2]		Micro[3]
-- Sample location	Dle #	Year	DQß	DRß1	haplotype	lineage	Data
E. Chukchi Sea							
-- Point Lay	94	88	101 / 101	101 / 401	-	-	-
	98	88	101 / 103	201 / 301	-	-	-
	100	88	101 / 101	202 / 202	-	-	-
	102	88	101 / 103	101 / 102	-	-	-
	103	88	101 / 101	202 / 501	J7	B	*
	105	88	101 / 101	101 / 202	J1	B	*
	107	88	101 / 101	201 / 202	J3	B	*
	109	88	101 / 101	401 / 501	J1	B	*
	115	88	101 / 202	401 / 501	J6	B	*
	303	89	-	202 / 501	J11	B	*
	304	89	101 / 103	101 / 201	J2	B	*
	306	89	-	201 / 301	J6	B	*
	307	89	101 / 101	101 / 501	J6	B	*
	311	89	101 / 101	202 / 501	J3	B	*
	312	89	101 / 101	101 / 202	J1	B	*
	314	89	-	102 / 301	-	-	-
	316	90	101 / 103	101 / 101	-	-	-
	319	90	101 / 101	401 / 401	J2	B	*
	321	90	101 / 101	501 / 501	J9	B	*
	323	90	101 / 101	301 / 501	J9	B	*
	326	90	-	101 / 202	J9	B	*
	328	90	-	401 / 501	J7	B	*
	330	90	101 / 103	201 / 202	J6	B	*
	331	90	101 / 101	101 / 101	J6	B	*
W. Hudson Bay							
-- Arviat	152	86	101 / 101	401 / 401	-	-	-
	153	86	101 / 101	101 / 402	-	-	-
	155	86	101 / 202	401 / 402	-	-	-
	156	86	101 / 202	301 / 301	-	-	-
	157	86	101 / 201	301 / 402	-	-	-
	158	86	101 / 101	201 / 202	-	-	-
	159	86	101 / 101	201 / 401	-	-	-
	160	86	101 / 101	101 / 401	-	-	-
	161	86	101 / 101	201 / 201	-	-	-
	162	86	101 / 101	202 / 501	-	-	-
	163	87	101 / 101	301 / 401	-	-	-
	164	87	101 / 201	301 / 401	J13	B	*
	165	87	101 / 201	301 / 401	J5	B	*
	166	87	101 / 101	301 / 301	J5	B	*
	167	87	101 / 101	301 / 402	J23	B	*
	168	87	101 / 201	301 / 301	-	-	-
	169	87	101 / 101	101 / 301	J2	B	*
	170	87	101 / 103	301 / 401	J6	B	*
	171	87	101 / 101	401 / 401	J7	B	*
	172	87	101 / 101	201 / 401	J1	B	*

Population	Sample Info.		Mhc Genotype[1]		MtDNA haplotype[2]		Micro[3]
-- Sample location	Dle #	Year	DQß	DRß1	haplotype	lineage	Data
W. Hudson Bay							
-- Churchill	344	90	101 / 201	201 / 501	J5	B	*
	345	90	101 / 201	101 / 201	J18	A	*
	346	90	101 / 101	202 / 501	J2	B	*
	347	90	101 / 201	102 / 301	J2	B	*
	348	90	101 / 101	102 / 402	J2	B	*
	349	90	101 / 101	401 / 401	J22	B	*
	350	90	101 / 101	501 / 501	J2	B	*
	351	90	101 / 101	401 / 401	J2	B	*
	352	90	101 / 101	101 / 401	J2	B	*
	353	89	101 / 101	401 / 501	J2	B	*
	354	89	101 / 101	201 / 401	J2	B	*
	355	89	101 / 101	101 / 201	-	-	-
	356	89	101 / 101	401 / 501	-	-	-
	357	89	101 / 101	301 / 401	-	-	-
	358	89	101 / 101	102 / 401	-	-	-
	359	90	101 / 101	301 / 301	-	-	-
	360	90	101 / 101	101 / 301	-	-	-
	365	91	101 / 101	401 / 501	-	-	-
	366	91	101 / 101	301 / 401	-	-	-
	367	91	101 / 201	101 / 201	-	-	-
	382	93	101 / 101	301 / 501	-	-	-
	383	93	101 / 101	201 / 402	-	-	-
	384	93	-	401 402	-	-	-
	385	93	101 / 101	201 / 501	-	-	-
	386	93	101 / 101	202 / 301	-	-	-
	388	93	101 / 101	202 / 402	-	-	-
	389	93	101 / 101	101 / 401	-	-	-
	390	93	101 / 101	101 / 101	-	-	-
	391	93	101 / 101	501 / 501	-	-	-
E. Hudson Bay							
-- Nastapoka R	1	87	101 / 101	301 / 401	D01	A	-
	2	87	101 / 101	202 / 501	D01	A	-
	7	87	101 / 101	301 / 401	D01	A	-
	8	87	101 / 103	402 / 501	D01	A	-
	9	87	101 / 101	201 / 501	D01	A	-
	10	87	101 / 101	101 / 301	D01	A	-
	11	87	101 / 202	101 / 402	D01	A	-
	13	87	101 / 103	401 / 402	D02	A	-
	14	87	101 / 101	201 / 501	D03	A	-
	15	87	101 / 101	101 / 101	D04	A	-
	16	87	101 / 101	101 / 401	D05	A	-
	17	87	-	501 / 501	D01	A	-
	18	87	101 / 101	301 / 401	D04	A	-
	19	87	101 / 101	202 / 401	D04	A	-
	21	87	101 / 101	101 / 501	DO4	A	-
	22	87	101 / 101	301 / 401	D01	A	-
	23	87	101 / 101	101 / 401	D06	B	-

Population	Sample Info.		Mhc Genotype[1]		MtDNA haplotype[2]		Micro[3]
– Sample location	Dle #	Year	DQβ	DRβ1	haplotype	lineage	Data
E. Hudson Bay							
– Nastapoka R	24	87	201 / 202	401 / 401	D07	B	-
	25	87	101 / 101	102 / 501	D03	A	-
	26	87	101 / 101	201 / 401	D04	A	-
	28	87	101 / 101	201 / 501	D01	A	-
	29	87	101 / 101	202 / 301	D01	A	-
	30	87	101 / 101	201 / 401	D01	A	-
	33	87	101 / 202	101 / 101	D01	A	-
	34	87	101 / 101	101 / 301	D01	A	-
	35	87	101 / 101	201 / 501	D01	A	-
	36	87	101 / 101	101 / 401	D01	A	-
	37	87	202 / 202	402 / 402	D04	A	-
	38	87	101 / 103	101 / 401	D01	A	-
	39	87	-	101 / 301	D04	A	-
	41	87	101 / 101	401 / 501	D01	A	-
	42	87	101 / 101	401 / 501	D01	A	-
	43	87	101 / 101	201 / 401	D07	B	-
	44	87	101 / 101	102 / 301	D03	A	-
	45	87	101 / 101	401 / 402	D01	A	-
	426	94	101 / 101	301 / 401	D08	B	-
	427	94	101 / 201	301 / 401	D03	A	-
E. Hudson Bay							
– Little Whale R	379	92	101 / 202	301 / 402	D05	A	-
	380	92	101 / 101	402 / 501	D02	A	-
	381	92	103 / 201	301 / 402	D02	A	-
	392	93	101 / 101	401 / 501	D08	B	-
	393	93	101 / 201	301 / 301	D09	B	-
	422	93	101 / 101	401 / 501	D08	B	-
	423	93	101 / 101	401 / 401	D08	B	-
	424	93	101 / 201	101 / 401	D12	A	-
	425	93	101 / 201	202 / 401	D12	A	-
S. E. Baffin Is							
– Cumberland Snd	132	86	101 / 101	301 / 301	J7	B	*
	133	86	101 / 101	101 / 202	J4	B	*
	134	86	101 / 101	501 / 501	J6	B	*
	135	86	101 / 101	101 / 301	J2	B	*
	136	86	101 / 101	101 / 301	J6	B	*
	137	86	101 / 101	301 / 401	J1	B	*
	139	86	101 / 101	101 / 301	J6	B	*
	141	86	101 / 101	101 / 401	J6	B	*
	145	86	101 / 101	301 / 401	J6	B	*
	146	86	101 / 101	401 / 401	J2	B	*
	150	86	101 / 101	101 / 301	J11	B	*
	151	86	101 / 101	101 / 202	J3	B	*

Population	Sample Info.		Mhc Genotype[1]		MtDNA haplotype[2]		Micro[3]
– Sample location	Dle #	Year	DQß	DRß1	haplotype	lineage	Data
High Arctic							
-- Grise Fiord	173	85	202 / 202	402 / 402	J2	B	*
	174	85	103 / 103	201 / 402	J2	B	*
	175	87	101 / 103	401 / 402	J7	B	*
	176	87	101 / 202	101 / 402	J7	B	*
	177	87	101 / 202	101 / 401	J2	B	*
	178	87	101 / 103	101 / 301	J2	B	*
	180	87	101 / 202	301 / 301	J2	B	*
	181	87	101 / 103	101 / 101	J7	B	*
	182	87	101 / 202	202 / 401	J7	B	*
	183	87	202 / 202	201 / 202	J5	B	*
	432	84	103 / 103	101 / 201	J7	B	*
	433	84	101 / 202	402 / 402	J3	B	*
	434	84	103 / 103	101 / 402	J5	B	*
	435	84	101 / 202	101 / 501	J5	B	*
	436	84	103 / 202	201 / 402	J5	B	*
	437	84	101 / 103	402 / 501	J5	B	*
	438	85	202 / 202	402 / 501	J30	B	*
	439	85	101 / 103	101 / 101	J2	B	*
	440	85	202 / 202	101 / 201	J2	B	*
High Arctic							
-- Creswell Bay	441	93	103 / 202	101 / 101	J7	B	*
	442	93	101 / 101	301 / 401	J7	B	*
	443	93	103 / 202	301 / 401	J37	B	*
	444	93	101 / 202	101 / 402	J5	B	*
	445	93	103 / 103	201 / 201	J6	B	*
High Arctic							
-- Cunningham Inlet	193	88	103 / 202	101 / 201	-	-	-
	196	88	101 / 202	101 / 401	-	-	-
	199	88	101 / 101	101 / 202	-	-	-
	201	88	103 / 103	101 / 402	-	-	-
	203	88	103 / 103	101 / 301	-	-	-
	208	88	101 / 202	101 / 401	-	-	-
	211	88	103 / 202	101 / 101	-	-	-
	214	88	101 / 103	101 / 101	-	-	-
	200	88	-	101 / 402	-	-	-
	222	88	101 / 103	101 / 101	-	-	-
	226	88	101 / 103	101 / 101	-	-	-
	228	88	103 / 202	202 / 402	-	-	-
	230	88	103 / 103	101 / 101	-	-	-
	233	88	103 / 202	101 / 501	-	-	-
	236	88	101 / 103	101 / 101	-	-	-
	275	90	101 / 103	101 / 101	-	-	-
	277	90	101 / 103	101 / 401	-	-	-
	278	90	101 / 101	101 / 301	-	-	-
	279	90	202 / 202	101 / 301	-	-	-

Population	Sample Info.		Mhc Genotype[1]		MtDNA haplotype[2]		Micro[3]
– Sample location	Dle #	Year	DQß	DRß1	haplotype	lineage	Data
High Arctic/ Baffin Bay							
– Nuussuaq	521	90	101 / 202	301 / 301	-	-	-
Winter samples	522	90	101 / 101	301 / 501	-	-	-
W. Greenland	523	90	202 / 202	201 / 402	-	-	-
	525	90	103 / 103	402 / 402	-	-	-
	527	90	103 / 202	NP	-	-	-
	529	90	202 / 202	102 / 402	J2	B	*
	530	90	202 / 202	402 / 501	-	-	-
	531	90	202 / 202	201 / 301	-	-	-
	532	90	101 / 103	101 / 301	-	-	-
	533	90	202 / 202	201 / 402	-	-	-
	535	90	101 / 202	401 / 402	-	-	-
	541	90	202 / 202	102 / 402	-	-	-
	543	90	202 / 202	102 / 402	-	-	-
	544	90	101 / 103	101 / 402	-	-	-
	545	90	101 / 103	402 / 501	-	-	-
	548	90	202 / 202	402 / 402	-	-	-
	550	90	101 / 103	402 / 501	-	-	-
	551	90	101 / 101	301 / 501	-	-	-
	552	90	-	201 / 402	-	-	-
	553	90	202 / 202	402 / 402	J5	B	*
	555	90	101 / 202	101 / 301	-	-	-
	556	90	202 / 202	101 / 401	-	-	-
	558	90	-	401 / 402	J27	B	*
	560	90	101 / 202	101 / 201	J2	B	*
	561	90	101 / 202	201 / 301	-	-	-
	565	90	103 / 202	101 / 402	-	-	-
	567	90	202 / 202	402 / 402	-	-	-
	568	90	101 / 202	101 / 402	-	-	-
	569	90	202 / 202	102 / 402	-	-	-
– Disko Bay	504	90	101 / 202	202 / 402	-	-	-
Winter samples	506	90	101 / 101	202 / 402	-	-	-
W. Greenland	509	90	202 / 202	201 / 301	J7	B	*
	513	90	101 / 202	402 / 501	-	-	-
	515	90	101 / 101	202 / 401	-	-	-
	516	90	101 / 202	402 / 501	-	-	-
	517	90	101 / 101	201 / 501	-	-	-
	518	90	101 / 202	301 / 402	-	-	-
	519	90	202 / 202	402 / 402	-	-	-
	520	90	202 / 202	402 / 402	-	-	-
	536	90	101 / 103	301 / 402	-	-	-
	537	90	101 / 202	101 / 201	J2	B	*
	538	90	103 / 202	101 / 201	-	-	-
	539	90	101 / 202	101 / 201	-	-	-
	540	90	103 / 202	101 / 102	J7	B	*